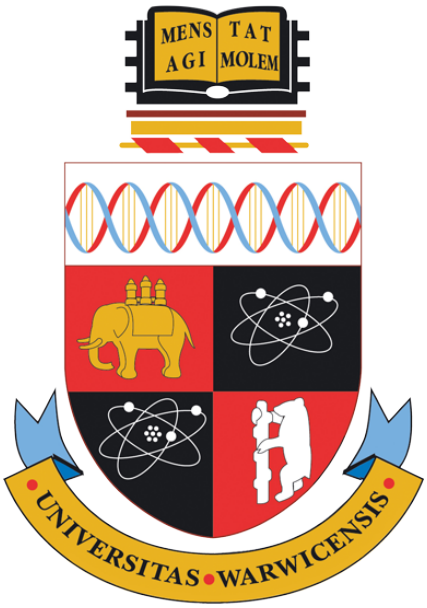
Analysis of the mechanism of recruitment of the helicase Sen1 at the replisome and its role in the maintenance of genomic stability

Emma Lones
Supervisors: Dr Giacomo De Piccoli, Dr Aga Gambus



Thesis

Submitted to the University of Warwick
in partial fulfilment for the degree of

Doctor of Philosophy in Life Sciences

September 2020

Contents

List of figures

List of tables

Acknowledgements

I would like to thank Dr Giacomo De Piccoli, who has been an incredibly kind and supportive mentor throughout the course of my PhD, providing valuable scientific advice and feedback. I am also extremely grateful to all members of the De Piccoli lab, who welcomed me into the group with open arms and made it a very enjoyable place to work. First, Rowin, whose own PhD project started the story of *sen1-3*; your knowledge and willingness to listen to ideas or answer questions was invaluable! I would also like to thank Alicja for always being available for a chat, whether it be scientific or otherwise. Nick, your help with that pesky cloning will never be forgotten and it was a pleasure to share a write up bench with you. Finally, Katy, in addition to your help in the lab, I will always be grateful for the glasses of wine shared; while celebrating the highs, and for lending an ear when things didn't quite go to plan!

I would also like to thank my wonderful family for their support over the last four years, and especially my partner James; I couldn't have done it without your unwavering belief and encouragement!

Finally, I am also very thankful to MIBTP and the BBSRC for funding my PhD project.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented in this thesis was carried out by the author, except in the cases of the following figures: the experiments described in figure 1.16, 3.1 and 3.6 were carried out by R.Appanah and G. De Piccoli, and the experiments in figure 6.3 were carried out by our collaborators in the Libri lab at L'Institut Jacques Monod. This is acknowledged within the text and figure legends.

Some of the work presented in this thesis has been published as part of the peer reviewed article 'Sen1 Is Recruited to Replication Forks via Ctf4 and Mrc1 and Promotes Genome Stability' (Appanah et al., 2020). The paper is included at the end of this thesis as Appendix 1. In this article, I contributed to the biochemical work analysing the interaction between Sen1 and its replisome binding partners, as well as investigating the phenotypes of the *sen1-3* mutant.

Abstract

During the S phase of the cell cycle, the processes of RNA transcription and DNA replication occur on the same template DNA; thus, can interfere with one another. In addition, co-transcriptional DNA:RNA hybrids (R loops) can impede processive DNA replication. When these obstacles are not removed, clashes with replication forks can result in their stalling, affecting genome stability.

The highly conserved DNA:RNA helicase Sen1, required for transcription termination and R loops resolution, travels with replisomes via an interaction with Ctf4 and Mrc1. I have analysed the role of Sen1 at replication forks in *S. cerevisiae* using a novel separation of function mutant, *sen1-3;* which no longer binds replisomes, but is fully competent in its transcription termination function. s*en1-3* is synthetic lethal combined with deletion of the RNase H enzymes, which digest R loops; and displays synthetic defects combined with the RNA metabolism mutant *hpr1Δ*, which are suppressed by RNase H overexpression. This suggests one function of Sen1 at forks may be to remove R loops that they encounter. I have also shown *sen1-3* is synthetic defective with the S phase checkpoint mutants *mrc1Δ*, *ctf18Δ* and *rad53Δ,* however the growth defects and increased recombination are not sensitive to RNase H. This indicates Sen1 may also remove other toxic obstacles encountered by replisomes.

Additionally, I have explored further the biochemistry of Sen1 association at forks. While full length Sen1 binding to its replisome partners is restricted to S phase, the N-terminal domain can interact with Mrc1 and Ctf4 throughout the cell cycle. I have identified that the interaction appears to be regulated by S phase cells extracts but does not depend on phosphorylation. I have also observed that the C-terminal of Sen1 interacts with DNA polymerase epsilon outside of the context of the replisome, independently from the cell cycle.

Abbreviations

ACS: ARS consensus sequence
AID: Auxin degron
ALS4: Amyotrophic lateral sclerosis type 4
AOA2: Ataxia oculomotor apraxia type 2
APC/C: Anaphase promoting complex/cyclosome
ARS: Autonomously replicating sequence
ATP: Adenosine triphosphate
BER: Base excision repair
BSA: Bovine serum albumin
CD: Co-directional
CDK: Cyclin-dependent kinase
CFS: Common fragile site
ChIP: Chromatin Immunoprecipitation
CIP: Ctf4 interacting peptide
CKI: Cyclin-dependent kinase inhibitor
CMG: Cdc45-MCM2-7-GINS
CPF: Cleavage and polyadenylation factor
Cryo-EM: Cryogenic electron microscopy
CTD: C-terminal domain
CUT: Cryptic unstable transcript
DDC: DNA damage checkpoint
DDK: Dbf4-dependent kinase
dNTP: Deoxynucleotide triphosphate
DRC: DNA replication checkpoint
DRIP: DNA:RNA hybrid immunoprecipitation
DSB: Double strand break
dsDNA: Double stranded DNA
FACT: Facilitates chromatin transcription
FHA1: Fork-head associated 1
FPC: Fork protection complex (Tof1-Csm3-Mrc1)
G1: Gap 1
G2: Gap2
GAL: Galactose
GINS: Sld5 (Go), Psf1 (Ichi), Psf2 (Ni), Psf3 (San)
HBD: Hybrid binding domain
HO: Head-on
HR: Homologous recombination
HU: Hydroxyurea
IAA: Indole-3-acetic acid (Auxin)
Ig: Immunoglobulin
IVTT: In vitro transcription termination assay
lncRNA: Long non-coding RNA
M: Mitosis
Mass spec/MS: Mass spectrometry
MCM: Minichromosome maintenance
miRNA: MicroRNA
MMS: Methyl methane sulfonate
MO: MCM-ORC

MOC-MC: MCM-ORC-Cdc6-MCM-Cdt1
mRNA: Messenger RNA
mRNP: Messenger ribonucleoprotein
MRX: Mre11-Rad50-Xrs2
NNS: Nrd1-Nab3-Sen1
OC: ORC-Cdc6
OOCM: ORC-Cdc6-Cdt1-Mcm2-7
ORC: Origin recognition complex
ORF: Open reading frame
Ori: Origin of replication
PAS: Polyadenylation signal
PCNA: Proliferating cell nuclear antigen
PIKK-3: Phosphoinositide 3-kinase related kinase
Pol ε: DNA Polymerase epsilon
Pol α: DNA Polymerase alpha
Pol δ: DNA Polymerase delta
Poly(A): Poly-adenylation
Pre-RC: Pre-replicative complex
RAF: Raffinose
rDNA: Ribosomal DNA
RER: Ribonucleotide excision repair
RFB: Replication fork barrier
RFC: Replication factor C
RFP: Replication fork pause
RNA: Ribonucleic acid
RNAPII: RNA polymerase II
Rnh1: RNase H1
Rnh2: RNase H2
RNR: Ribonucleotide reductase
rNTP: Ribonucleoside triphosphate
RPA: Replication protein A
SCF: Skp, Cullin, F-box containing complex
Sen1: Splicing endonuclease I
Ser2-P/S2-P: Serine 2 phosphorylated
Ser5-P/S5-P: Serine 5 phosphorylated
SETX: Senataxin
snoRNA: Small nucleolar RNA
snRNA: Small nuclear RNA
ssDNA: Single stranded DNA
ssRNA: Single strand RNA
TAM: Transcription-associated mutagenesis
TAR: Transcription-associated recombination
TC-NER: Transcription-coupled nucleotide excision repair
***td*:** Temperature degron
TERRA: Telomeric repeat-containing RNA
Top1: Topoisomerase I
Top2: Topoisomerase II
TRC: Transcription replication collision
UV: Ultraviolet

Chapter 1: Introduction

To maintain genomic stability, the entire genetic content of a cell needs to be replicated faithfully once per cell cycle. The tightly controlled process of DNA replication occurs during S phase, before mitotic division separates the newly synthesised chromosomes. A specialised machine termed the 'replisome' is responsible for chromosome duplication; comprising a complex eleven subunit helicase (CMG; Cdc45-Mcm2-7-GINS) which unwinds the template DNA, along with several accessory proteins that ensure regulated and processive DNA synthesis. Three specialised DNA polymerases (Pol α, Pol δ and Pol ε) are responsible for catalysing the formation of new DNA molecules, two of which physically associate as part of the replisome. These will be discussed further in chapter 1.3 (Burgers and Kunkel, 2017).

During S phase, replication forks must negotiate their way through a number of obstacles or barriers. These include protein complexes bound to the DNA, secondary DNA structures that are difficult to unwind (such as G4 quadruplexes), and damaged DNA templates. Critically, a major obstacle for replication is RNA transcription; both processes occur on the same template DNA, and so can clash both in 'head-on' or 'co-directional' orientations. Increasing evidence also suggests that, although they play some important physiological roles within the cell, stable R loop structures (where nascent mRNA reanneals with the template DNA strand behind elongating RNA polymerase II (RNAPII)) can also interfere with processive DNA synthesis.

In this introduction, I will first describe the eukaryotic cell cycle; with particular emphasis on S phase, DNA replication and the various components of the eukaryotic replisome. I will also discuss in more detail the obstacles that replication forks may encounter, including the effects of collisions between forks and transcription bubbles or R loops. I will also consider the mechanisms employed by cells to minimise and deal with these collisions, before focusing on the known functions of DNA:RNA helicase Splicing Endonuclease 1 (Sen1) in both the removal of R loops and non-canonical transcription termination. Finally, I will present the evidences of the role of Sen1 during DNA replication.

1.1 The use of *Saccharomyces cerevisiae* as a model organism for eukaryotic DNA replication

Saccharomyces cerevisiae, more colloquially known as 'budding yeast' due to its method of division, is a widely used model organism for the study of fundamental eukaryotic cellular processes. In fact, in 1996, the 16-chromosome genome of *S. cerevisiae* (comprising approximately 6,000 genes) was the first eukaryotic genome to be sequenced in its entirety (Goffeau et al., 1996); and has many pathways in common with humans, including the cell division process. This study uses this organism as model for eukaryotic DNA replication, as the many proteins involved are highly conserved. All components of the replisome (described in chapter 1.3.3) have single orthologs in other eukaryotic organisms, including higher eukaryotes and humans (Gambus et al., 2006). The genetic tractability; which allows easy deletion, insertion, mutation or overexpression of genes, along with rapid growth cycle (~90mins) compared with animal models or human cells makes *S. cerevisiae* a powerful tool for such studies. In addition, the stable existence of haploid yeast cells allows for the synchronous release of cultures into S phase following cellular arrest in G1 by the addition of mating pheromone. Though recent developments in gene editing, i.e. CRISPR-Cas9 allows disruption, mutation and tagging of endogenous genes in mammalian cells with relative ease, yeast has the advantage that haploid strains of different genotypes can be quickly and easily mated, and the desired genotypes selected following meiosis. Thus, budding yeast remains a fast and easy tool to study fundamental processes.

By convention, specific nomenclature is used to distinguish between proteins, wild type and mutated genes in budding yeast. Wild type genes are designated by uppercase italic (e.g. *SEN1*); whereas mutant genes are lowercase italic, followed by a number or symbol (e.g. *sen1-1* or *sen1Δ*) if recessive, or uppercase italic if dominant. The first character of a protein name is presented in uppercase e.g. (Sen1 or Sen1-1) and is not italicised.

1.2 Overview of the eukaryotic cell cycle

A single cell follows a tightly controlled, pre-programmed pattern of events to replicate its DNA and produce two genetically identical daughter cells. The division cycle of eukaryotic cells is separated into four distinct phases. Interphase is comprised of: Gap 1 (G1), Synthesis phase (S) and Gap 2 (G2), which are followed by the process of mitosis (M) (fig 1.1). During interphase, cells grow and prepare to divide (G1), replicate their DNA (S), check for any errors (G2), before finally, division occurs (M). Cells commit to progress through the cycle in G1 at a stage known as the 'restriction point' (R) in higher eukaryotes or 'Start' in yeast. Beyond this, external growth stimuli are no longer required. Under certain circumstances, cells that are not committed to division enter a phase termed G0. In this phase, 'quiescent' cells; for example, mature hepatocytes, retain the ability to re-enter the cell cycle in response to specific growth factor stimuli. However, some somatic cells enter a permanent 'senescent' state, where they are unable to do so. This can be due to cellular ageing or induced in response to stresses such as irreversible DNA damage.

The progression of cells through this program of events is tightly regulated by the cyclin family of proteins, in order to ensure that they do not enter the next phase before they are ready. Accurate duplication is achieved through the exquisite interplay between the control exerted by this protein family and monitoring by the various cell cycle checkpoints.

Cyclin dependent kinases (CDKs) are serine threonine kinases that are activated by the binding of co-factor cyclins. These cyclin-CDK complexes phosphorylate specific targets to control the cell cycle. Budding yeast has a single 'master-regulator' CDK, Cdc28 (Mendenhall and Hodge 1998). Nine cyclins (Cln1-3, Clb1-6), with different properties or phase specific expression are responsible for the targeting of various proteins to initiate certain cell cycle events (reviewed in (Bloom and Cross, 2007)) (fig 1.1). The specificity of these cyclins is controlled both at the transcriptional level, and at the protein level by ubiquitin mediated proteolysis. In addition, CDK inhibitors (CKIs) and inhibitory phosphorylation of CDK controls their activity.

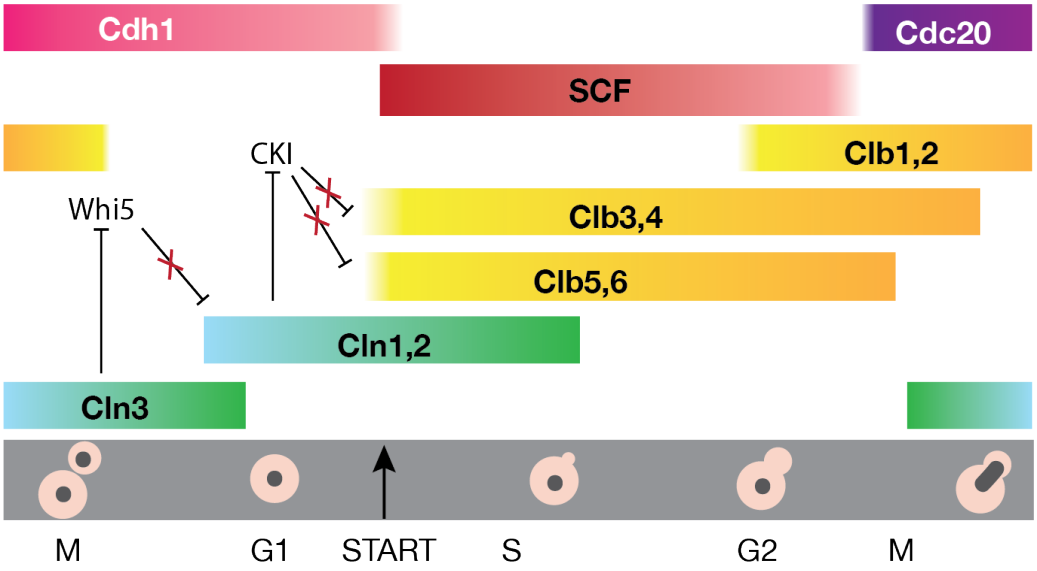


Figure 1.1 Schematic of *S. cerevisiae* cell cycle. In order to replicate their DNA and produce two identical daughter cells, eukaryotic cells follow a specific program of events during the mitotic cycle. Various cyclins and the cyclin dependent kinase, Cdc28, are responsible for the regulation of these events. A schematic overview of the *S. cerevisiae* cell cycle and the regulation of these cell cycle events by cyclin activity is shown. The figure is adapted from (Diffley, 2004).

Cln1-3 are G1 specific cyclins, responsible for commitment to cell cycle 'Start', duplication of the spindle pole body, cell growth and activating the transcription of later expressed cyclins (Bloom and Cross, 2007). The transcription of Cln3 occurs throughout the cell cycle, however its levels peak late M to early G1 (Tyers et al., 1993). At this stage, Cln3-Cdc28 phosphorylates the Whi5 transcriptional repressor, targeting it for nuclear export. This allows the transcription of Cln1 and Cln2, the levels of which peak during the G1-S transition (Koch et al., 1996, Costanzo et al., 2004, de Bruin et al., 2004) (fig 1.1). This also true for the transcription of the B-type S phase cyclins Clb5 and Clb6, which during G1 are inactivated by the cyclin dependent kinase inhibitor Sic1 (Schwob et al., 1994). Once Sic1 is phosphorylated by Cdc28-Cln1 and Cdc28-Cln2 for proteasomal degradation, Clb-Cdc28 complexes and the firing kinase DDK (Cdc7-Dbf4-dependent kinase) promote the firing of early origins and the cells enter S phase (Nishizawa et al., 1998). Importantly, the activity of Cdc28-Clb complexes also plays a crucial role in prohibiting the licensing of new origins to prevent re-replication (by targeting free key players for degradation), and promotes centrosome duplication and budding. Following Start, the G1 cyclins, Cln1 and Cln2, are ubiquitylated by the SCF Ubiquitin ligase for proteasomal degradation (Barral et al., 1995, Skowyra et al., 1997).

The mitotic cyclins Clb3 and Clb4 are also transcribed during S phase, and their levels remain elevated until late anaphase. This pair of cyclins may have some redundancy of function in S phase regulation with Clb5 and 6; however, they also contribute to early mitotic processes such as spindle formation, pole separation and chromosome condensation. Finally, the transcription of Clb1 and Clb2 peaks just before anaphase. These cyclins are required for mitotic events, in addition to activating the anaphase promoting complex/cyclosome (APC/C). The APC/C is a multi-subunit E3 Ubiquitin ligase, which following phosphorylation by mitotic cyclins binds to its co-factor Cdc20 (Rudner and Murray, 2000). This renders the APC/C^{Cdc20} functional; whereby it is responsible for ubiquitylating Dbf4 (the regulatory subunit of the origin firing kinase DDK) and the S phase/mitotic B-type cyclins, which targets them for degradation. It also targets securin for degradation which enables mitotic exit. Securin prevents the digestion of cohesin (which holds sister chromatids together) by binding to and inhibiting the action of the protease separase (Shirayama et al., 1999, Wasch and Cross, 2002, Peters, 2006, Thornton and Toczyski, 2003). Once separase is released, sister chromatids are able to move to opposite poles of the cell, thus progressing the cells from metaphase to anaphase.

At a later stage of mitosis, the APC/C^{Cdc20} mediated reduction in the levels of mitotic cyclins leads to dephosphorylation of the APC/C, causing Cdc20 to dissociate; thus, it promotes its own inactivation. This leaves the APC/C free to bind another of its cofactors, Cdh1. During late mitosis to G1, the APC/C^{Cdh1} contributes to proteolysis of Cdc20 and the mitotic cyclins to prevent their re-accumulation. It is during this low CDK, high APC/C^{Cdh1} window that origin licensing is permitted to occur (Peters, 2006, Bloom and Cross, 2007). However, as the G1 cyclins are not substrates for the APC/C^{Cdh1}, their accumulation ultimately leads to phosphorylation and inactivation of the APC/C^{Chd1}. This, along with the targeting of CKIs for degradation by the E3 Ubiquitin ligase SCF promotes S phase (Diffley, 2004).

While progression through the cell cycle is driven by these cyclin dependent kinase complexes, cell cycle checkpoints play crucial roles in monitoring the order and fidelity of these events, surveying for any DNA damage and arresting progression to allow time for DNA repair to occur before entering the next phase. This is mediated by signal transduction cascades and complex regulatory networks, where sensors

detect damage or defects and transduce the signal to effectors, coordinating the appropriate physiological response (Putnam et al., 2009). The various cell cycle checkpoints in *S. cerevisiae* have been reviewed extensively (see (Putnam et al., 2009, Barnum and O'Connell, 2014) and references therein).

In brief, cell size checkpoints exist at various points during G1 and G2 to ensure that cells reach the correct size at the correct time for DNA replication and division (Barnum and O'Connell, 2014). Critically, both intrinsic and extrinsic sources of DNA damage result in activation of the DNA damage checkpoints, which are responsible for delaying the transition between G1/S or blocking the G2/M transition, in order to allow time for repair processes to occur. In addition, the mitotic spindle checkpoint ensures that sister chromatids are correctly aligned at the metaphase plate and are attached to microtubules under tension before anaphase can occur. APC/C^{Cdc20} activity promotes anaphase by the degradation of mitotic cyclins and securin, however mitotic checkpoint proteins prevent its activation until the checkpoint is satisfied (Barnum and O'Connell, 2014).

Importantly, as DNA replication is a particularly perilous time for cells, two arms of the S phase checkpoint exist which differ in their mediator proteins but that, in *S. cerevisiae*, converge on a common pathway and response. The DNA damage checkpoint recognises double strand breaks and gaps behind replication forks. On the other hand, the DNA replication checkpoint specifically signals a critical threshold of arrested replication forks. Following detection, both pathways arrest cell cycle progression, regulate transcription and target replisomes so to promote their ability to resume DNA replication following removal of the obstacle (Branzei and Foiani, 2007, Putnam et al., 2009). Later in this chapter, after describing the process of DNA replication and introducing various components of the replisome, I will discuss in more detail the methods of activation and role of the S phase checkpoint response in the maintenance of genome stability.

1.3 The Process of Eukaryotic DNA Replication

1.3.1 Origin licensing

DNA replication is limited to S phase, and the synthesis of an entire new copy of the genome needs to occur within an appropriate time frame. Prokaryotes, which have smaller and less complex genomes, initiate replication bidirectionally from a single origin. Eukaryotes on the other hand, have larger and more complex genomes and initiate bidirectional replication at multiple loci, activated specifically via a temporal program. In *S. cerevisiae*, these sites are known as origins of replication (Oris) or autonomously replicating sequences (ARS elements). These are approximately 100-200 base pairs long and include an AT-rich, 11 base pair ARS consensus sequence (ACS) along with other auxiliary elements (Broach et al., 1983). However, in other eukaryotes and metazoans; replication origins do not contain consensus sequences and are less well defined, but features such as chromatin architecture or DNA topology and structure have been shown to influence their location (Mechali, 2010, Prioleau and MacAlpine, 2016, Ekundayo and Bleichert, 2019).

Throughout the cell cycle, Oris are recognised by a multi-subunit DNA binding protein called the origin recognition complex (ORC). Comprised of Orc1-6, where subunits 1-5 contain AAA+ ATPase motifs, ORC binds to Oris in a manner that is dependent on ATP. During late mitosis and G1, when S phase/mitotic cyclins are inactive; the Mcm2-7 holoenzyme (minichromosome maintenance) is loaded at origins by using ORC as a scaffold, aided by other mediator proteins. Assembly of Mcm2-7 double hexamers onto origin DNA forms pre-replicative complexes (pre-RCs), which marks these origins as licensed for replication (fig 1.2) (reviewed (Remus and Diffley, 2009, Bleichert, 2019, Amin, 2019)).

Mcm2-7 is the core of the replicative helicase (CMG). Its six subunits are organised in a specific order to form a ‘ring-like’ toroidal shape that encircles DNA, and although not identical, all subunits contain an AAA+ ATPase motif in their C-terminal domain (Li et al., 2015, Forsburg, 2004). As such, Mcm2-7 is the ‘motor’ that provides energy for the unwinding of double stranded DNA at forks via ATP hydrolysis.

Importantly, two Mcm2-7 enzymes are loaded at Oris onto double stranded DNA as inactive head-to-head double hexamers, ready for the formation of two replication forks that travel in opposite directions from that single origin (Evrin et al., 2009, Remus et al., 2009, Gambus et al., 2011). Many interactions are responsible for linking the two hexamers together, largely between the conserved-zinc finger domains within the N-termini of the MCM subunits (Li et al., 2015). Loading of the hexamers onto the DNA is achieved by the presence of a 'gate' between the Mcm2 and Mcm5 subunits, permitting the ring to adopt an open or closed conformation (Bochman and Schwacha, 2008). Importantly, although the double hexamer is loaded onto dsDNA; following origin melting and separation of the double hexamer, the single hexamer within each CMG encircles ssDNA. Thus, opening and closing of this gate is also important for Mcm2-7 remodelling as part of the active helicase (Bochman and Schwacha, 2008). The mechanism of how MCM double hexamers are loaded onto DNA has been widely debated. One model hypothesised that a single molecule of ORC is responsible for directing the binding of both MCMs, while another suggested that the first MCM is responsible for the seconds recruitment. Another proposed that a second ORC may bind in the opposite orientation to an auxiliary element within Oris that is similar to the ACS for second MCM recruitment, however this element is not present at all Oris (reviewed by (Bell and Labib, 2016), see references therein).

A recent study from the Costa lab used time-resolved cryo-EM to elucidate the precise mechanism responsible for loading MCM-double hexamers at Oris. Here, they used linear DNA containing an Ori and purified yeast proteins to examine the intermediates formed during this process. In addition to ORC, the assembly of Mcm2-7 double hexamers requires the mediator proteins Cdt1 and the AAA+ ATPase Cdc6. First, the binding of ORC and subsequently Cdc6 to origins results in the recruitment of a complex of Cdt1 and the first Mcm2-7 hexamer, whereby the C-termini of ORC and MCM interact. Following the threading of the DNA through the central channel of Mcm2-7, this results in the formation of the observed OCCM (ORC-Cdc6-Ctd1-Mcm2-7) intermediate (fig 1.2) (Miller et al., 2019).

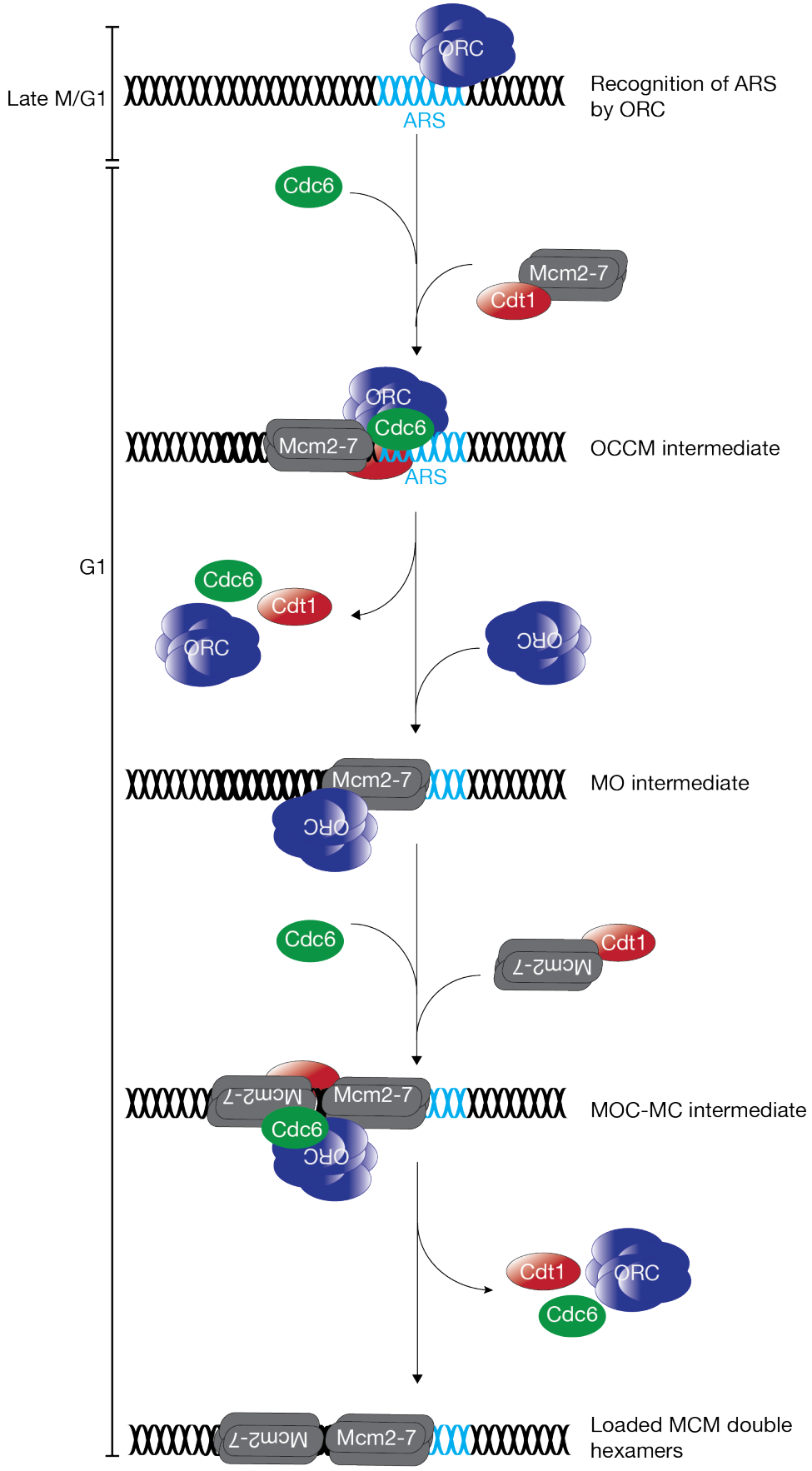


Figure 1.2 The licensing of replication origins in *S. cerevisiae*. During late mitosis/G1, the origin recognition complex (ORC) recognises and binds to autonomously replicating

sequences across the genome. During G1, Cdc6 binds to the ORC and these recruit a complex of Mcm2-7 and Cdt1 to achieve loading of an inactive Mcm2-7 double hexamer, forming an OCCM intermediate. Following dissociation of Ctd1 and Cdc6, a second ORC engages with the loaded MCM in an inverted orientation to recruit another Cdc6 and Cdt1-Mcm2-7. The loaded double hexamer marks origins as licensed for DNA replication.

Cdt1 is thought to stabilise Mcm2-7 in a conformation where the gate remains open to allow DNA access to the central channel. In this conformation, MCM ATPase activity is inhibited and it remains bound to ATP. Subsequent binding to ORC-Cdc6 may then disrupt this conformation, promoting release of Cdtl, thus allowing MCM to hydrolyse ATP resulting in closure of the gate (Frigola et al., 2017). At this stage, Cdc6 also dissociates. Next, an MO (MCM-ORC) species was observed; where, through an interaction with the N-terminal homodimerization interface of the first loaded Mcm2-7, a second ORC is engaged at the Ori in an inverted orientation. Interestingly, recruitment of this second ORC can occur both while the first ORC is still engaged, or after it is released. This second ORC subsequently recruits Cdc6 and the second Cdt1-Mcm2-7 complex, resulting in the observed MOC-MC species (MCM-ORC-Cdc6-MCM-Cdt1). Ultimately, following release of Cdc6, Cdt1, ORC and closure of the gate, the double hexamer is formed (fig 1.2) (Miller et al., 2019).

1.3.2 Origin firing

Although helicases are loaded at many prospective replication origins across the genome, this step does not commit all of these origins to initiate DNA replication. Instead, the 'firing' of a licensed origin is a result of formation of the full Cdc45-Mcm2-7-GINS (CMG) helicase and activation of the loaded MCMs (Gambus et al., 2006). GINS (named **g**o-**i**chi-**n**i-**s**an for its subunits) is made up of Sld5, Psf1, Psf2 and Psf3.

Throughout S phase, some origins are 'early' firing, and some are 'late firing', while others act as 'back-up' dormant origins. These are inactivated if a fork from a neighbouring origin passes through. However, in situations where two converging forks collapse and cannot finish DNA synthesis; activation of a dormant origin within that region ensures complete replication of the genome. In *S. cerevisiae*, a model has been proposed where low levels of firing factors are limiting compared to the number of Oris, which results in this temporal program of origin firing. Following initiation

of early origins, these factors are recycled to initiate the next wave of later origins (Mantiero et al., 2011). The rate-limiting abundance of Sld3, Sld7 and Cdc45, which associate with Oris during G1 in a manner that also depends on DDK, defines early-firing origins. Increasing the levels of Sld3, Sld7, Cdc45 and DDK results in the firing of late origins sooner during S phase (Tanaka et al., 2011).

It is thought that firing occurs in a stochastic manner, though licensed origins have different efficiencies. For example, accessibility of firing factors to the origin can be affected by nucleosome positioning and local chromatin structure (Rodriguez et al., 2017). Domains with 'open' euchromatin over the central regions of chromosomes tend to be early replicating, while highly repetitive regions such as rDNA and heterochromatic telomeres are late replicating (Fragkos et al., 2015). Interestingly, competition for initiation factors between 'unique' single-copy protein-coding regions and highly repetitive rDNA arrays has been shown in yeast. Sir2 is responsible for repressing origin firing of rDNA in wild type cells, to protect the cells from using up their pool of initiation factors at the many Oris within these repetitive regions. Deletion of *SIR2* results in increased rDNA initiation, which is detrimental to the replication of unique regions, resulting in unreplicated regions in G2/M cells (Foss et al., 2017).

It has also been suggested Sir2 plays a separate, direct role in modulating the potential imbalance in the distribution of licensed origins across the genome that could result from variable accessibility of different regions to MCMs. This achieves equal spatial distribution of licensed origins across early/late firing euchromatin and heterochromatin. Deacetylation of histones adjacent to an Ori by Sir2 inhibits its ability to become licensed (no MCM binding is observed). Sir2 dependent attenuation of origin licensing has been shown both in early-replicating regions and at telomeres. This prevents a high density of licensed Oris becoming localised in these regions to the detriment of others. It was shown that in cells lacking Sir2, a shift in licensed Ori density to early replicating chromatin and telomeres results in incomplete replication of some late-replicating regions by the end of S phase. This is likely due to the recruitment of a greater proportion of the limited firing factors to licensed early regions and telomeres, which affects their availability for timely firing and complete replication of later origins (Hoggard et al., 2020).

Importantly, cells have evolved mechanisms to separate origin licensing and firing in order to prevent re-replication over a single cell cycle, which would be catastrophic for genome integrity; leading to genome rearrangements or cell death. Primarily, the loading of helicases onto origins (licensing) and activation of the helicase (origin firing) is separated temporally by restricting these events to G1 and S phase, respectively. This is achieved through the activity of cyclin dependent kinases. During the G1 to S transition (after helicase loading has occurred), S phase CDK activity prevents the licensing of origins that have already been fired by phosphorylating three of the key players. Free Mcm2-7 proteins are exported from the nucleus as a result of Mcm3 phosphorylation (Labib et al., 1999, Nguyen et al., 2000), Cdc6 is targeted for proteasomal degradation (Drury et al., 2000), and helicase loading by ORC is inhibited by phosphorylation of Orc2 and Orc6 (Chen and Bell, 2011). In addition to this, the proteins involved in these events are controlled at the transcriptional level. For example, transcription of Cdc6, which opens up the binding interface on ORC for helicase loading, is limited to G1. Moreover, transcriptional repressors such as Whi5 ensure cyclin levels oscillate with the cell cycle as required.

Initiation of origins is a complex, multistep process that requires the action of the Cdc7 Dbf4-dependent kinase (DDK) and CDKs which become active at the G1/S transition. Dbf4 is degraded by the APC/C^{Cdc20} between the time of chromosome segregation and pre-start G1 to prevent premature activation of loaded helicases before S phase (Ferreira et al., 1999).

After ‘Start’, individual subunits of loaded MCMs are either phosphorylated or bound by the ‘firing kinase’ DDK. This triggers the recruitment of Sld3, which binds to the phosphorylated peptides within Mcm4 and Mcm6 (Sheu and Stillman, 2006, Francis et al., 2009, Randell et al., 2010, Sheu and Stillman, 2010). Throughout the cell cycle, Sld3 exists as a complex with Cdc45, which as a consequence is also recruited to origins (Kamimura et al., 2001). In addition, Sld7 associates with Sld3 to stabilise Sld3 interaction with the helicase (Tanaka et al., 2011).

As the G1/S transition approaches and S phase CDK activity increases, Sld2 and Sld3 are phosphorylated in a CDK-Clb5/6-dependent manner. Away from the Oris, phosphorylated Sld2 binds to a BRCA1 C-terminus repeat (BRCT domain) within

Dpb11. Sld2-Dpb11 associates with GINS and DNA pol ε to form the pre-loading complex (pre-LC) (Muramatsu et al., 2010). The pre-LC is then recruited to MCMs via an interaction between another, separate BRCT domain within its Dpb11 subunit and phosphorylated Sld3 (fig 1.3). Yeast 2 hybrid assays (Y2H) have also shown that additional subsequent interactions between various constituents at origins also contributes to CMG formation (Bell and Labib, 2016). Furthermore, the interaction between Dpb2 (a subunit of Pol ε) and GINS is essential, suggesting Pol ε is important for initiation of DNA replication in addition to its role in DNA synthesis (Sengupta et al., 2013).

In order for translocation of the bidirectional forks to become possible, the loaded MCMs need to be remodelled from dsDNA to ssDNA. Then, each Mcm2-7 hexamer encircles only the ssDNA of the leading strand so each sister replication fork can travel in their respective directions. Therefore, the double hexamers must dissociate from one another, DNA at the origin needs to be separated (an event termed origin melting), and the lagging strand needs to be released from the central MCM channel. However, the exact order and mechanism of these events is only recently becoming clearer.

A cryo-EM study examining the structure of Mcm2-7 on G1 chromatin found that the central channel is kinked at the interface between double hexamers, which may distort the tightly bound dsDNA. The authors hypothesised a slight subsequent rotation between the two single hexamers could melt the origin DNA (Remus and Diffley, 2009, Li et al., 2015). More recent cryo-EM studies suggest that indeed, double hexamer formation itself does not cause melting, but a second step inside the channel is required. It has been proposed that one of the DNA filaments is stretched, leading to untwisting (Ali et al., 2017). Another study found that both strands are positioned near the Mcm2-5 gate, and that ATP dependent tilt and lateral shift of the MCM rings extrudes the lagging strand of dsDNA from the central channel (Noguchi et al., 2017).

Importantly, Mcm10, which associates with MCM double hexamers, is thought to play a key role in final activation of the CMG (van Deursen et al., 2012, Watase et al., 2012, Douglas et al., 2018). In a CDK dependent manner, at the onset of S phase; Mcm10 accumulates at origins following assembly of the CMG. Initial data showed

that this association does not require melting of the origin (van Deursen et al., 2012, Watase et al., 2012). Furthermore, Mcm10 inactivation does not affect assembly of the CMG, but does lead to a loss of replication protein A (RPA) recruitment (a protein that binds/stabilises ssDNA) to replication origins (van Deursen et al., 2012, Watase et al., 2012). These data highlighted Mcm10 as a requirement for the final step of initiation, and a speculative model was proposed whereby Mcm10 may stabilise the Mcm2-7 ring in an open conformation for dsDNA to ssDNA remodelling (van Deursen et al., 2012). A more recent study, which used cross linking mass spectrometry found that in fact, Mcm10 interacts with 6 of the 11 CMG subunits (Mayle et al., 2019). Using recombinant CMG and Mcm10, the two proteins together on dsDNA were demonstrated to exert sufficient force to melt the origin DNA duplex to allow the transition of head-to-head MCMs to ssDNA (Langston and O'Donnell, 2019). This is in line with data from the Diffley lab, which showed firing factors provoke the stabilisation of CMG and initial DNA untwisting, however the helicase remains inactive. Further DNA unwinding and CMG activation occurs in an Mcm10 and ATP dependent manner, after which the hexamers translocate past one another in the 3' to 5' direction, to establish the bidirectional forks (Douglas et al., 2018). As Mcm10 binds ssDNA with high affinity, the authors speculate it may play a direct role in aiding exclusion of the lagging strand during origin melting (Robertson et al., 2008, Douglas et al., 2018).

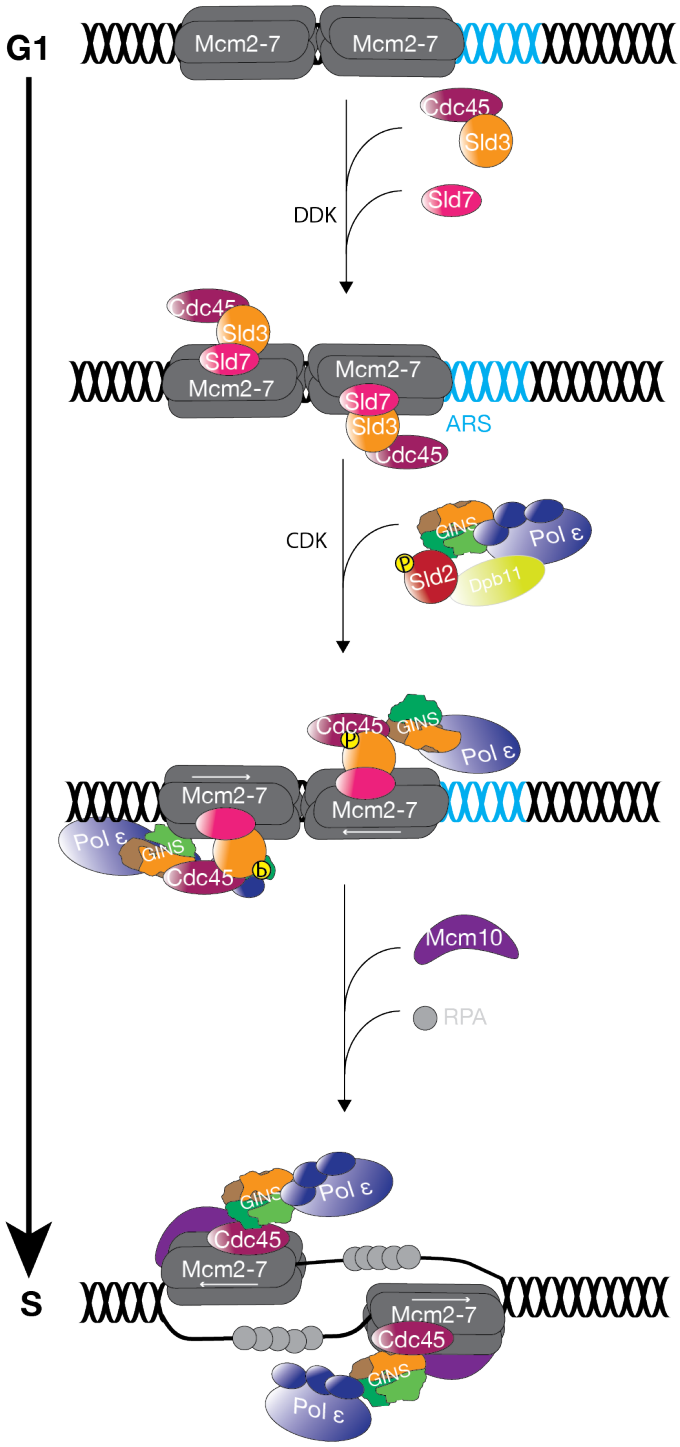


Figure 1.3 The firing of replication origins in *S. cerevisiae*. During G1, the firing kinase DDK phosphorylates Mcm2-7. Phosphorylated Mcm4 and Mcm6 recruits a complex of Sld3-Cdc45; then Sld7 binding with Sld3 stabilises its interaction with the MCM hexamer. CDK-Clb5/6 then phosphorylates Sld2 and Sld3 to bind to separate Dpb11 BRCT repeats. The pre-LC is formed away from origins where Dpb11-Sld2 associates with GINS and DNA pol ε. The pre-LC is subsequently recruited to origins by the interaction between Dpb11 and phosphorylated Sld3. Further interactions between various components then occur, and the MCM double hexamers are remodelled and activated, where they disconnect from one another and expel the lagging strands to transition from dsDNA to ssDNA. Interactions between Mcm10 and the CMG results in melting of the origin, where the template DNA separates for 3'-5' directional translocation of the individual MCMs.

1.3.3 S phase and DNA replication

Following the initiation step of replication, the unwound DNA at fired origins is used as a template for bi-directional replication forks to carry out DNA synthesis. The forks will continue until they meet with another converging fork; thus, resulting in complete duplication of the cells entire genetic content. Other factors are assembled around the CMG helicase to form the replisome (fig 1.4), which is specialised to ensure accurate, processive and timely DNA synthesis.

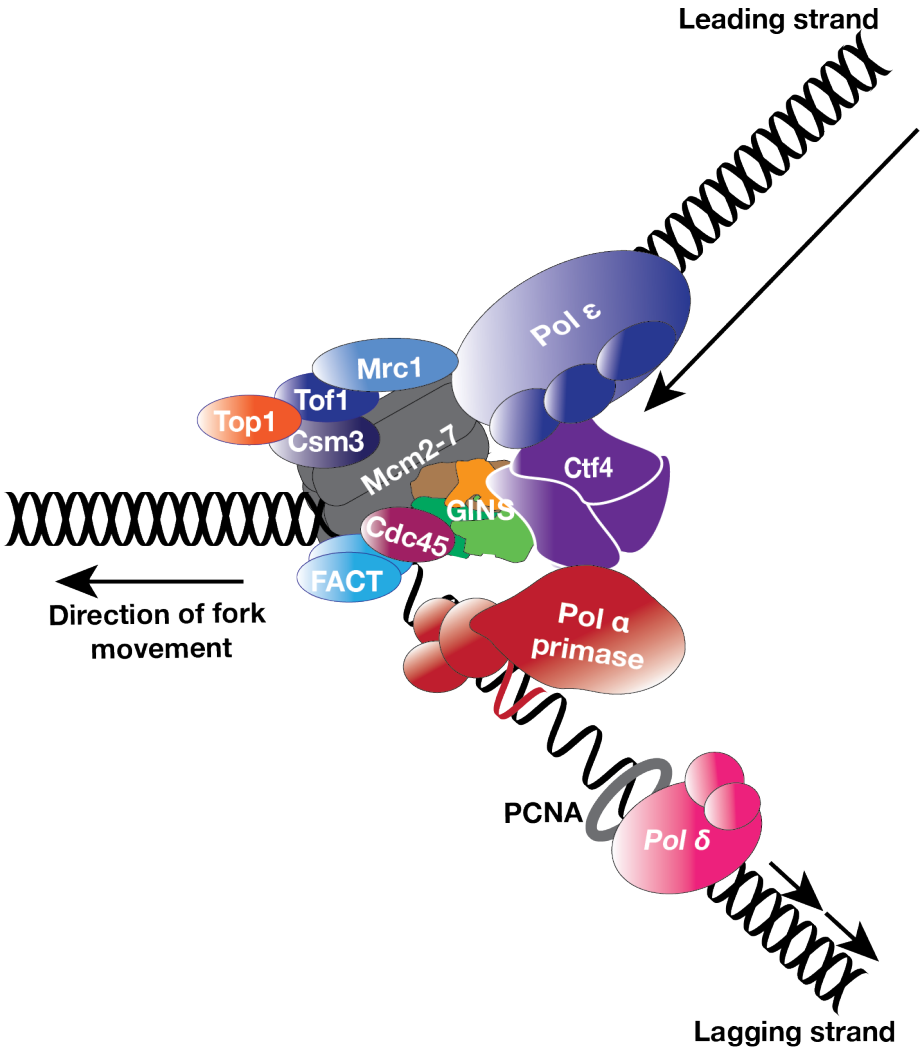


Figure 1.4 Schematic of the eukaryotic replisome. The eukaryotic replisome is comprised of an eleven-subunit helicase (CMG) to unwind the template DNA, three specialised DNA polymerases (Pol α, Pol δ and Pol ε) responsible for catalysing DNA synthesis, along with several accessory proteins.

Following initiation, several regulatory factors associate with the CMG to make up the replisome, the existence of which is specific to S phase and has been shown by Chromatin Immunoprecipitation (ChIP) to travel with forks (Calzada et al., 2005, Gambus et al., 2006). Key components of the replisomes are: the protein interaction hub Ctf4, the S phase checkpoint protein Mrc1, the fork protection complex Tof1-Csm3, Topoisomerase 1 (Top 1, which relieves torsional strain in DNA), the histone chaperone FACT, Mcm10, the DNA polymerases Pol α and Pol ε, and SCF^{Dia2}, an E3 Ubiquitin ligase which ubiquitylates Mcm7 to disassemble the CMG upon completion of DNA replication (Gambus et al., 2006, Sengupta et al., 2013, Morohashi et al., 2009). Other proteins are also fundamental for high fidelity DNA replication but fail to co-purify with the rest of the replication machinery. These include: the ssDNA binding protein RPA, the DNA polymerase Pol δ, the RFC (replication factor C) clamp loader and the PCNA (proliferating cell nuclear antigen) clamp. This section will describe these key players and their functions in more detail.

The DNA polymerases α, δ and ε

DNA polymerases are the enzymes responsible for catalysing the synthesis of new DNA molecules during the process of replication. Importantly, these polymerases can only carry out DNA synthesis in the 5' to 3' direction, by catalysing the formation of a covalent bond between the free hydroxyl group (-OH) of a deoxyribose (located at the 3' end) and the 5' phosphate group of an incoming nucleotide. Thus, as a result of the antiparallel nature of the DNA double helix, only one strand of the unwound DNA can be synthesised continuously behind the CMG helicase. This is termed the leading strand, as the 3' to 5' orientation of the template allows continuous 5' to 3' synthesis of the nascent DNA strand, following the direction of CMG unwinding and the fork's movement. Replication of the other 'lagging' strand is more complicated, as the antiparallel 5'-3' orientation of its template means short fragments of DNA (Okazaki fragments) need to be replicated in a discontinuous fashion by DNA polymerase as more DNA is unwound and exposed (fig 1.5). Furthermore, an additional complication is that DNA polymerases can only begin synthesis if there is a pre-existing strand annealed to the template with a free 3' -OH to attach the new deoxynucleotides to. Therefore, a short RNA primer is required to initiate the synthesis of each new DNA molecule (fig 1.5).

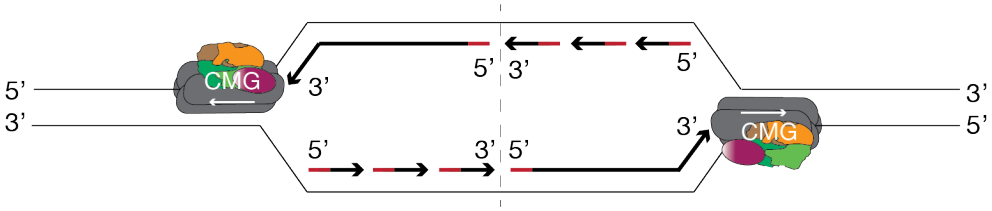


Figure 1.5 Schematic of the leading and lagging strands at replication forks. The CMG helicase translocates along the leading strand in a 3' to 5' direction to unwind the DNA. As the DNA is unwound, this strand can be replicated continuously as DNA polymerases have a 5' to 3' directionality. Due to this specific directionality, the opposite lagging strand, which has a 5' to 3' orientation, must be replicated in short discontinuous Okazaki fragments. To begin synthesis of any new DNA molecule, a short RNA-DNA primer is required (shown in red) from which the DNA polymerase can then extend.

Thus, in *S. cerevisiae,* three separate polymerases with individual roles are responsible for nascent DNA synthesis. DNA polymerase epsilon (Pol ε) is the main leading strand polymerase, continuously carrying out DNA synthesis in the 5'-3' direction behind the CMG; and DNA polymerase delta (Pol δ) is responsible for the synthesis of Okazaki fragments on the lagging strand, which are usually ~150-200 base pairs in length (Kunkel and Burgers, 2014). Importantly however, DNA polymerase alpha (Pol α) is the only enzyme that has the ability to actually initiate synthesis of a new DNA strand. This is due to its unique ability to produce a short RNA primer using the cells ribonucleotide pools, from which it can extend the chain a short number of deoxynucleotides before 'handing over' synthesis to the other more processive polymerases.

Pol α primase – priming the synthesis of new DNA strands

Dependence on the generation of a short RNA primer to initiate DNA synthesis is common across all organisms. This activity is universally carried out by an enzyme called a primase, which is a DNA-dependent RNA polymerase (reviewed (Pellegrini, 2012)). Importantly, *de novo* synthesis of these short RNA primers not only initiates DNA replication during origin firing, but is also required continuously at replication forks, as due to the 5'-3' directionality of DNA polymerases, synthesis of Okazaki fragments necessitates constant re-priming as more of the DNA template becomes available following helicase unwinding.

In eukaryotic cells, the primase is part of a heterotetramer termed the Pol α-primase complex, which possesses both RNA primase and DNA polymerase activity. The primase activity of Pol α was first discovered in studies of the small double stranded SV40 virus, which utilises host proteins for its replication and chromatin organisation. The only exceptions are those required for initiation and replicative helicase activity, which is carried out by its own T4 antigen (reviewed in (Burgers, 2009). Pri1 and Pri2 make up the small (catalytic) and large (regulatory) primase subunits of the complex, respectively. These are coupled with DNA polymerase α, comprised of its Pol1 (catalytic) and Pol12 (B regulatory) subunits; which are responsible for extending the RNA primer by ~10-15 deoxynucleotides, forming a hybrid RNA-DNA primer (reviewed (Pellegrini, 2012, Bell, 2019)). Crystallographic and biochemical evidence suggests that Pol α is responsible for the transfer from RNA primer synthesis to deoxynucleotide addition by its recognition of the 'A-form' conformation of the RNA primer/DNA template hybrid. The polymerase then extends the chain with dNTPs until approximately one full turn of the double DNA helix is synthesised, at which point it loses contact with the A-form helix, causing it to stall and disengage from the template (Perera et al., 2013). As Pol α is not a very processive enzyme, it is not suitable for sustained DNA replication. The RNA-DNA primers make up the pre-existing strand from which the more processive Pol ε and Pol δ extend. Another factor that makes Pol α unsuitable for ongoing replication is that compared with Pol δ and Pol ε, it does not possess any 3' exonuclease activity; which is responsible for proof reading for any errors, and so it presents a greater hazard for genomic stability (Pavlov et al., 2006).

The replisome component Ctf4 has been found to play a central role in the association of Pol α with the CMG helicase. Using epitope tagged CMG subunits, immunoprecipitation experiments and mass spectrometry, Gambus et. al observed that Pol α co-purifies with a complex of Mcm2-7 and GINS, specifically during S phase; demonstrating that it travels with forks as part of the replisome. This association is lost in the absence of Ctf4, suggesting that Ctf4 plays a key role in tethering Pol α to the CMG as part of the replisome. Interestingly, a fraction of Ctf4 was found to associate with GINS throughout the cell cycle and the recruitment of Ctf4 to origins depends on GINS (Gambus et al., 2009, Gambus et al., 2006). A direct interaction

between GINS and Ctf4, as well as a direct interaction between Ctf4 and the amino terminus of Pol α has been reconstituted in *E. coli* (Gambus et al., 2009). It has since been shown the interaction occurs through the Sld5 subunit of GINS, and Sld5 and Pol α both share a common conserved Ctf4-interacting peptide motif (CIP-motif) (Simon et al., 2014). ChIP experiments have shown that destabilisation of the interaction between GINS and Pol α occurs in *ctf4Δ* cells, and this affects the distribution, but not the binding of other fork components to chromatin, both in HU and unperturbed S phase. Thus, the bridging of Pol α with the CMG by Ctf4 appears to be important for coordinating the progression of the helicase with DNA synthesis (Tanaka et al., 2009).

Pol ε and Pol δ, the leading and lagging strand polymerases

Pol ε and Pol δ are responsible for replicating the bulk of the length of the genome, following 'handover' of both the leading and lagging strand substrates from Pol α. Like Pol α, both of these enzymes are members of the B family of polymerases; however, they associate with different proteins and have different properties and structures (Kunkel and Burgers, 2014). Much evidence (discussed in more detail later in this section) suggests Pol ε is the main leading strand polymerase, while Pol δ replicates the lagging strand.

Pol ε is heterotetramer, comprised of Dpb2, Dpb3, Dpb4 and the catalytic subunit Pol2. The N-terminal domain (NTD) of the Pol2 subunit possesses both its polymerase and exonuclease activity, while the C-terminal (CTD) is non catalytic and has been proposed to play a structural role (Tahirov et al., 2009). A recent study has determined the cryo-EM structure of the yeast Pol ε holoenzyme as a whole. The Dpb3 and Dpb4 subunits were found to be located in the centre of the enzyme, bridging the NTD and CTD of Pol2 together in a rigid state. Finally, Dpb2 forms close contacts with the CTD, while the NTD interacts with PCNA. Considering this structure, the authors propose an atomic model of the leading strand replisome; where following exit of the leading strand from the CMG, Dpb2 is oriented to direct the DNA into the catalytic Pol2 NTD (Yuan et al., 2020).

Pol δ is a heterotrimer, comprised of Pol31, Pol32 and the catalytic subunit Pol3 (which like Pol ε also possesses both polymerase and exonuclease activity). The cryo-EM structure of the yeast Pol δ holoenzyme has also recently been reported. This study found that the Pol3 exists as a globular module which engages the primer/template double helix for catalysis. The regulatory module, consisting of Pol31 and Pol32 is flat and does not contact the DNA. It is bridged to the catalytic domain, but the structure appears to be flexible; which is thought to aid its exonuclease activity and the binding of diverse DNA substrates, as it is involved in other pathways, including mis match repair (Jain et al., 2019).

Importantly, in addition to being more processive, both of these enzymes are capable of higher fidelity DNA synthesis, as in contrast to Pol α, they possess intrinsic 3'-5' exonuclease activity. Thus, they have the ability to 'proofread' the nascent DNA by removing any mis incorporated nucleotides. Indeed, evidence suggests that the exonuclease activity of Pol δ may also correct errors generated by Pol α, as the reduced fidelity of the *pol1-L868M* mutant (which retains normal processivity) is exacerbated by inactivation of Pol δ 3' exonuclease activity (Pavlov et al., 2006).

Similar to Pol α, Pol ε is also tethered to the replisome by binding to the CMG helicase (in this case via the Psf2 subunit of GINS) (Sengupta et al., 2013). The Dpb2 subunit of Pol ε contains the GINS-binding domain and is responsible for its association with the replisome during S phase. Also, as previously mentioned in Chapter 1.3.2, this interaction is essential for initiation (Sengupta et al., 2013). Furthermore, additional CMG subunits have also been shown to crosslink with Pol ε (Sun et al., 2015). In contrast, it does not appear that Pol δ associates with the CMG.

Many studies investigating the division of labour between the three different polymerases have demonstrated that they are arranged asymmetrically at forks (reviewed (Lujan et al., 2016)). A seminal study examined the exonuclease deficient *pol2-4* and *pol3-01* mutants, which found that Pol ε and Pol δ correct errors on different strands of the template DNA (Shcherbakova and Pavlov, 1996). Subsequent studies used the *pol2-M644G* and *pol3-L612M* mutants, where both alleles have substitutions in their catalytic site which result in specific stable mismatches that lead

to transversion mutation 'hotspots'. Examining their mutation signatures found that the leading and lagging strands are primarily synthesised by Pol ε and Pol δ, respectively (Pursell et al., 2007, Kunkel and Burgers, 2008). Reconstitution of a replisome with the ability to perform both leading and lagging strand DNA synthesis *in vitro* identified that in order to achieve asymmetric synthesis of the leading and lagging strands, the polymerases are involved in multiple suppressive reactions, to ensure each polymerase extends the correct strand (Georgescu et al., 2015).

However, the division of labour between Pol δ and ε may be more promiscuous than the canonical equal model. Interestingly, in *S. cerevisiae,* deletion of Pol2 N-terminal (which contains the active polymerase domain) is not lethal (though S phase is severely compromised), and Pol ε appears to be dispensable for SV40 replication in human cells, suggesting Pol α and δ can, in certain situations, be responsible for replication of both the leading and lagging strand (Dua et al., 1999, Zlotkin et al., 1996). One model proposed that at forks, both the leading and lagging strand are synthesised by Pol δ, while Pol ε simply proofreads the leading strand (Johnson et al., 2015). It has since been demonstrated that Pol δ plays a role in the establishment of DNA synthesis on the leading strand, before handing over to Pol ε, which has the unique ability to stimulate CMG unwinding for maximal elongation rates. Interestingly, this polymerase switch may also help re-establish leading strand synthesis following replication stress and uncoupling of Pol ε and the CMG (Yeeles et al., 2017). A recent study, which tracked the strand activity of the three polymerases by mapping mis-incorporated ribonucleotides (in a ribonucleotide excision repair deficient background) showed that in both *S. cerevisiae* and *Sz. pombe,* initiation of the leading strand indeed occurs by passing synthesis from Pol α to Pol δ, before Pol ε takes over, and this is ubiquitous across the genome. Interestingly, they also observed that a hand over from Pol ε to Pol δ occurs on the leading strand as two replisomes converge for termination (Zhou et al., 2019).

The DNA clamp proliferating cell nuclear antigen and the Replication factor C clamp loaders

The sliding DNA clamp proliferating cell nuclear antigen (PCNA) is composed of a homotrimer that forms a ring-like structure that encircles DNA. It is particularly important for DNA synthesis by Pol δ, however also binds Pol ε via an association with the catalytic NTD of Pol2 (Chilkova et al., 2007, Yuan et al., 2020). It acts as a processivity factor to anchor the leading and lagging strand polymerases onto DNA, interacting differentially with them in a manner that reflects their individual properties and the division of labour at the fork (Boehm et al., 2016). Alone, Pol ε is an inherently more processive enzyme than Pol δ; thus, PCNA interacts weakly with Pol ε and only stimulates its processivity by a factor of ~6. Conversely, Pol δ interacts strongly with PCNA; which increases its processivity by approximately 100-fold, resulting in both polymerases having similar processivity on ssDNA primed with PCNA (Boehm et al., 2016). In addition, the association of other proteins with PCNA can play an important role for genome stability; for example, specialised DNA polymerases involved in translesion synthesis are recruited via their PIP box (PCNA interacting peptide) by ubiquitylated PCNA to sites of DNA damage that cannot be repaired by the classical polymerases (Boehm et al., 2016).

PCNA is loaded onto DNA by the RFC (Replication factor C) family of protein complexes. RFC clamp loaders are comprised of the subunits Rfc2-5 in combination with another larger subunit (including either Rfc1, Ctf18 or Elg1). At replication forks, after the RNA-DNA primer is generated by Pol α, RFC binds to the 3' end of the primer-template junction, and loads PCNA in an ATP-dependent manner (Bowman et al., 2004). A recent study has found that PCNA is loaded in a preferential manner onto the leading and lagging strands by the RFC^{Ctf18} and RFC^{Rfc1}, respectively (Liu et al., 2020). RFC^{Ctf18} has an additional two subunits (Ctf8 and Dcc1) (Mayer et al., 2001). Dcc1 and Ctf18 directly bind to Pol ε (Stokes et al., 2020), and this binding, while not affecting PCNA-mediated chromosome cohesion, is essential for proficient DNA replication and checkpoint activation (Crabbé et al., 2010).

During the process of origin firing, a single PCNA is loaded onto the leading strand for synthesis by Pol ε. Yeeles at al. have suggested that the presence of this PCNA, in

addition to helping promote maximal replication rates, forms a bridge between the leading-strand 3' end and Pol ε-CMG to help prevent uncoupling of DNA unwinding and synthesis (Yeeles et al., 2017). Indeed, the N-terminal domain of Pol2 directly interacts with PCNA, through which the nascent DNA is extruded (Yuan et al., 2020).

During synthesis of the lagging strand, PCNA needs to be loaded repeatedly as each individual Okazaki fragment is primed. Following the loading of PCNA by RFC^{Rfc1}, Pol δ is recruited for extension of the Okazaki fragment (Boehm et al., 2016). Eventually, Pol δ reaches the 5' end of the downstream RNA primer, which needs to be removed before DNA ligase can seal the two DNA fragments together. The main pathway for Okazaki fragment maturation sees Pol δ continue synthesis for an extra 1-2 nucleotides in order to displace the 5' end of the preceding RNA primer and generate a short flap, which is then removed by Rad27 (FEN1) 5' flap exonuclease activity. Cycles of strand displacement and removal are repeated until all of the RNA primer is removed. Importantly, each of the 3 different PCNA subunits associate specifically with Pol δ, Rad27 and the Cdc9 DNA ligase (Burgers, 2009). The RNase H enzyme, Rnh2, has also been shown to interact with PCNA via its PIP box motif, and may present an alternative pathway of Okazaki fragment maturation; where Rnh2 cleaves the RNA portion within the primer until the last ribonucleotide, which is removed by Rad27 (Qiu et al., 1999). The fusion of Okazaki fragments is completed by the DNA ligase Cdc9 (Johnston and Nasmyth, 1978). This is a requirement for subsequent PCNA unloading by RFC^{Elg1}, which is then recycled for synthesis of another Okazaki fragment (Kang et al., 2019, Kubota et al., 2015).

Due to the constant re-priming during Okazaki fragment synthesis; enrichment of the number of PCNA molecules on the lagging versus leading strand is observed, but interestingly, this distribution changes following fork stalling. PCNA is unloaded from the lagging strand at stalled forks in a manner that is dependent on RFC^{Elg1} and the checkpoint kinases Mec1 and Rad53. The authors speculate this may assist checkpoint activation by allowing the loading of the 9-1-1 complex, or promote the association of translesion polymerases with ubiquitylated PCNA on the leading strand (Yu et al., 2014).

Other members of the eukaryotic replisome

A leading strand replisome with the ability to perform DNA synthesis *in vitro* on a primed, forked substrate minimally requires: a loaded CMG helicase, Pol ε, RPA and PCNA. However, it is unable to reach physiological replication speeds. In fact, the observed average rate of 0.26 kb min^{-1} is ~ 4 to 8 times slower than *in vivo* forks, which are thought to replicate between ~1.6 to 1.9 kb min^{-1} (Georgescu et al., 2014). An elegant study by the Diffley lab purified various replication factors, and in doing so defined the minimum set of proteins required for origin-dependent initiation of an *in vitro* substrate (linear or circular DNA with an associated origin, attached to beads). These include various firing factors, the CMG helicase, Ctf4, Top II, Pol ε and Pol α. Notably, although this system was competent for replication, it did not recapitulate the coupled DNA synthesis displayed by *in vivo* forks and was much slower. Additionally, other important replication proteins such as Pol δ, RFC, PCNA, Mrc1 and Tof1-Csm3 were omitted (Yeeles et al., 2015).

Subsequently, radiolabelled nucleotides were used to identify the leading and lagging strands within the system. Pulse chase experiments demonstrated that while the template leading strand was almost completely replicated (albeit at a slower rate); replication of the lagging strand was incomplete, even in the presence of PCNA (Yeeles et al., 2017). Altogether these observations suggested that additional replisome components are necessary for forks to successfully complete DNA replication and reach the higher physiological speed. The rate of leading strand DNA synthesis was considerably increased by the addition of the purified replisome components Mrc1, Csm3-Tof1, FACT and Top 1, reaching a rate similar to *in vivo* replication forks (Yeeles et al., 2017). Excluding each component individually found that Mrc1 is critical for the forks to achieve physiological speeds *in vitro*, however Csm3-Tof1 is also required to achieve the highest rate, perhaps by stabilising Mrc1 (Yeeles et al., 2017). The further addition of Pol δ to the reaction resulted in complete synthesis of the lagging strand (Yeeles et al., 2017).

Mrc1, Tof1 & Csm3

Mrc1 plays a key role in activation of the S phase checkpoint, specifically in response to replication stress events that lead to fork stalling. It is part of a protein signalling cascade that results in the activation of the effector kinase Rad53, which elicits the downstream checkpoint responses. Mrc1 contains multiple S/TQ residues which are hyperphosphorylated by Mec1, promoting a direct interaction of Mrc1 with Rad53 for its activation. In addition to Mrc1, Tof1 may also play a minor role in activation of the replication checkpoint under certain circumstances (the checkpoint is described in more detail in chapter 1.4). Although deletion of *MRC1* is not lethal; during normal S phase, a higher degree of spontaneous replication fork damage is observed in its absence. Indeed, Rad9, which undergoes phosphorylation by Mec1 in response to DNA damage, is hyperphosphorylated in *mrc1Δ* cells during late S phase (Alcasabas et al., 2001). Additionally, *mrc1Δ* is also synthetic lethal with *rrm3Δ*, a DNA helicase involved in resolution of stalled or damaged replication fork structures. Clearly, cells become reliant on Rrm3 to resolve fork defects resulting from the loss of Mrc1 (Szyjka et al., 2005).

Commonly, the ribonucleotide reductase inhibitor hydroxyurea (HU) is used to study the S phase checkpoint, as the depleted levels of dNTPs results in replication fork stalling and Rad53 activation. Forks that stall as a result of HU exposure lose the ability to resume replication in the absence of Mrc1. This is only partially due to the failure to activate Rad53 and the checkpoint (Katou et al., 2003, Lou et al., 2008). Therefore, in response to replication stress, Mrc1 appears to both serve as a platform for Rad53 activation, and independently promote stabilisation of stalled forks by coupling helicase activity with DNA synthesis, so replication can be resumed successfully once the checkpoint in satisfied.

In addition to its fork protection/checkpoint role, Mrc1 also appears to play a role at forks that is important for DNA replication. Mrc1 is loaded onto origins with the DNA polymerases, which is a requirement for forks to achieve physiological speeds *in vitro*. As the $mrc1^{AQ}$ mutant (which cannot be phosphorylated by Mec1 for Rad53 activation) is still proficient in DNA replication, this suggests that its checkpoint and DNA synthesis rate roles are separate (Lou et al., 2008, Yeeles et al., 2017).

Mrc1 has been shown to interact with Pol2, the catalytic subunit of Pol ε (another protein implicated in signalling replication stress (Navas et al., 1995)), during both G1 and S phase. Mrc1 itself has two independent interaction domains for Pol2 that exist within its N- and C-termini, that bind separate sites within Pol2 N- and C-termini, respectively. This association appears to be important for DNA replication, potentially helping to stabilise Pol ε. Interaction with the Pol2 N-terminal polymerase/exonuclease domain also implies Mrc1 could be an accessory for Pol2 function in this respect. Alternatively, placement of Mrc1 near Pol2 active site may aid in sensing replication blocks. Moreover, following phosphorylation of Mrc1 in response to replication stress; Mrc1 C-terminal remains bound, but the N-terminal dissociates from Pol2. This may cause a conformational change in Pol2, allowing interaction of phosphorylated Mrc1 with other partners, or play a role in regulating the checkpoint response e.g. recruiting/stabilising Rad53 (Lou et al., 2008). Mrc1 also specifically interacts with coiled-coil containing central region of Mcm6. This interaction appears to be important for sensing alkylating agent methyl methane sulfonate (MMS) induced DNA damage at forks (but interestingly not HU), activating the replication checkpoint through Mrc1. In a mutant where the Mrc1-Mcm6 interaction is abolished, cells are reliant on the Rad9 pathway and activation of the DNA damage checkpoint for viability (Komata et al., 2009).

Tof1 and Csm3, as well as Mrc1, are members of the fork protection complex (FPC). Tof1-Csm3 form a tight, stable complex that can associate with replisomes independently from Mrc1. However, efficient association of Mrc1 with replisomes appears to depend on Tof1-Csm3 (Bando et al., 2009). Importantly, the FPC promotes the sensing of protein barriers ahead of the replication fork and induces pausing, such as at Fob1, which binds to replication fork barriers (RFBs) in rDNA (Calzada et al., 2005, Tourrière et al., 2005). It has been suggested that the ability of replication forks to sense and pause at RFBs in highly transcribed rDNA limits the collisions between replisomes and the transcription machinery (Takeuchi et al., 2003). A recent study used cryo-EM and cross-linking mass spectrometry to model the topology of the FPC along with Ctf4 and the CMG in the context of forked DNA. It appears that Tof1/Csm3 are located at the head of the replisome in front of the CMG, binding both MCM and dsDNA. Tof1/Csm3 “grip” the double stranded DNA duplex via DNA-binding motifs;

these domains are important for both the stabilisation of FPC (including Mrc1) within the replisome, and for efficient replisome pausing in response to replication fork barriers (in this case in response to Fob1). The authors hypothesise the location of Tof1/Csm3 at the head of the CMG may position them to sense abnormalities in the template DNA structure or protein barriers for efficient pausing and/or stabilisation. Tof1 is also reported to bind Topoisomerase I (Top1), thus it may bring Top1 to the front of forks to resolve topological stress. Finally, Mrc1 was modelled to extend across a single side of the replisome, forming cross-links with Tof1/Csm3 at the front through to contacts with Pol ε at the back. This is consistent with its role in coordinating efficient fork progression, and the authors speculate that Mrc1 could link leading strand synthesis with events occurring ahead of the fork, e.g. the sensing of RFBs by Tof1/Csm3. Finally, several amino acids within Mrc1 also cross-link to Mcm2 and Mcm6, where the leading strand template DNA exits from the helicase. This region of Mrc1 also interacts with the flexible catalytic domain of Pol ε, which could provide a potential mechanism for how Mrc1 stimulates optimal fork elongation rates; by tethering the helicase and polymerase in close proximity for synthesis of the daughter DNA (Baretic et al., 2020).

Overall, Mrc1 appears to perform the distinct functions of 1) promoting proficient DNA replication (to achieve the maximum rate) and 2) stabilising stalled forks and stimulating the inhibition of DNA replication during periods of replication stress. It has been suggested that these roles could be mediated through a shared mechanism, whereby the molecular interactions of Mrc1 with its replisome partners results in helicase/polymerase coupling and allows it to perform its replicative and checkpoint functions (Lou et al., 2008). However, while there is clear phenotypic evidence for the role of Mrc1 during DNA replication; the biochemical function within the replisome requires further clarification.

Ctf4

Ctf4 is an important protein interaction hub within the replisome and is responsible for coupling other processes to DNA replication. It has orthologues in both *Sz. pombe* (Mcl1) and higher eukaryotes/humans (AND-1) and shows a high level of conservation (Williams and McIntosh, 2002, Zhu et al., 2007, Bermudez et al., 2010). Although deletion of *CTF4* is not lethal in *S. cerevisiae*, it is required for efficient DNA replication/metabolism, sister chromatid cohesion and unperturbed cell cycle progression (Kouprina et al., 1992, Simon et al., 2014, Tanaka et al., 2009, Zhu et al., 2007, Bermudez et al., 2010, Williams and McIntosh, 2002).

Ctf4 exists *in vivo* as a trimer, the formation of which is mediated through a β-propeller (WD40) domain in its C-terminus. In line with its role as an interaction hub, this self-trimerization enables it to associate with three different binding partners concurrently. A bundle consisting of six alpha helices is fused to the C-terminal of the β-propeller domain. These α-helical bundles protrude away from the plane of the trimer, providing a site for interacting proteins to dock on to (Simon et al., 2014). A second WD40-domain is also present in the N-terminal of Ctf4 (from amino acids 2-383). This domain also extends out from the trimer and may mediate other protein-protein interactions (fig 1.6) (Gambus et al., 2009).

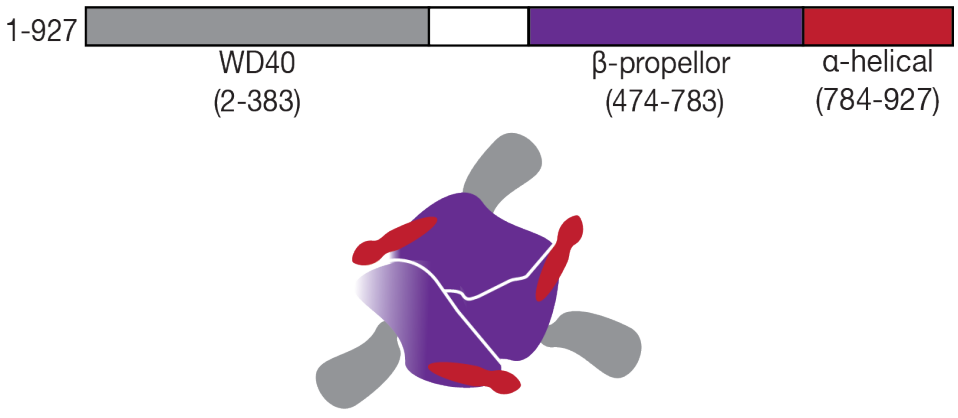


Figure 1.6 The structural organisation of Ctf4. Schematic of the Ctf4 trimer showing its domain organisation as determined by Simon et. al (2014) using single particle electron microscopy. Formation of the trimer is mediated by the Ctf4 β-propeller domain, from which the associated α-helical bundles extend upward to facilitate protein-protein interactions. The N-terminal WD40 domains extend radially from the trimer (figure adapted from (Simon et al., 2014)).

As mentioned previously (page 32), Ctf4 tethers Pol α to the CMG through a common shared Ctf4-binding motif within these two proteins, termed a CIP box (Ctf4-interacting peptide) that is conserved from yeast to humans (Gambus et al., 2009, Simon et al., 2014). Thus, like Mrc1 it has a role in coupling DNA unwinding with DNA synthesis.

The interaction between Ctf4 and Pol α is mediated through the Ctf4 protruding C-terminal α-helical bundle and the CIP box in the N-terminus of Pol1 (the catalytic subunit of the Pol α/primase complex). Ctf4 is connected to the CMG by binding the CIP-box containing N-terminal of the Sld5 subunit of GINS. In terms of the strength of these interactions, it appears that Ctf4 binds to the CMG more tightly than Pol α, as it is maintained at higher salt concentrations (700mM vs. 300mM). Further electron microscopy experiments showed that Ctf4 is able to bind both CMG and Pol α at the same time, with various different stoichiometries. In line with replisomes containing a single CMG, it is likely that Ctf4 remains bound to GINs throughout S phase, which leaves two other protomers to coordinate association of other accessory replication factors, including one or two Pol α, which is continuously required at forks to prime Okazaki fragments (Simon et al., 2014).

Among other proteins that have been shown to associate with the C-terminal of Ctf4 is the flap endonuclease Dna2, which is involved in the maturation of Okazaki fragments, acting first on RPA-coated flaps to allow subsequent cleavage by Rad27. Abolishing this association, which is mediated by a Pol1/Sld5-like CIP-box results in a smaller chromosome 12, which contains a large array of rDNA repeats (Villa et al., 2016). Ctf4 is also responsible for the recruitment of Chl1, a helicase involved in sister chromatid cohesion, to replication forks; coordinating this process with ongoing DNA synthesis during S phase (Samora et al., 2016). In addition, Tof2 and Dpb2 (a subunit of Pol ε) also interact with Ctf4, although these associations are mediated through a second, separate CIP-box which binds to another site in Ctf4. Tof2 is associated with rDNA biology, and like Dna2, abolishing the interaction with Tof2 results in a smaller chromosome 12, thus Ctf4 is important for coupling DNA replication with maintenance of rDNA copy number (Villa et al., 2016).

Finally, there are several other accessory factors that have been identified to dynamically interact with the core replisome (these factors are summarised as part of table 1.1). Mms22 associates with the replisome via Ctf4 WD40, and is a DNA repair protein implicated in homologous recombination-mediated repair of stalled replication forks (Buser et al., 2016). Dia2 also associates with replisomes via both Ctf4 and Mrc1, and is part of the SCF ubiquitin ligase, that during termination, efficiently disassembles the CMG by ubiquitylating Mcm7 for efficient degradation (Maculins et al., 2015).

Table 1.1 Summary of replisome accessory factors

Accessory Protein	Interacts via	Role
Ctf4	Sld5 subunit of GINS	Protein interaction hub within the replisome. Co-ordinates replication with various other processes. Tethers Pol alpha to the replisome.
Mrc1	Pol ε, Mcm6, Tof1/Csm3	Part of fork protection complex for stable fork pausing in response to replication stress or replication barriers. Activation of the DNA replication checkpoint during S phase. Required for optimal replisome speeds.
Tof1	CMG	Part of fork protection complex for stable fork pausing in response to replication stress or replication barriers. Activation of the DNA replication checkpoint during S phase.
Csm3	CMG	Part of fork protection complex for stable fork pausing in response to replication stress or replication barriers.
Top1	Tof1	Resolve topological stress and prevents excessive fork rotation.
Mcm10	Pol α and MCM	Activation of the CMG helicase.
Spt16	Currently unknown	Subunit of the FACT complex. Helps the retention of parental histones at forks to re-deposit on nascent DNA behind the elongating replisome.
Pob3	Currently unknown	Subunit of the FACT complex. Helps the retention of parental histones at forks to re-deposit on nascent DNA behind the elongating replisome.
Dna2	Via Ctf4	Processing of Okazaki fragments. Couples replication with maintenance of rDNA copy number.
Tof2	Ctf4 and Top1	Couples replication with maintenance of rDNA copy number.
Mms22	Ctf4	Rad52-dependent HR restart of replication at stalled forks.
Chl1	Ctf4	Contacts cohesin during S phase to facilitate histone acetylation, helping to establish cohesion.
Dia2	Ctf4 and Mrc1	Disassembly of the CMG at the end of replication by ubiquitylating Mcm7 for degradation.
Rrm3	Pol ε	Promotes fork progression past stable protein blocks.

1.4 The S phase checkpoint

1.4.1 Overview of the S phase checkpoint

S phase is a particularly perilous time for cells; any replication errors, structural DNA damage or interference with fork progression needs to be recognised and corrected before entry into mitosis/chromosome segregation, so to minimise genome instability. The S phase checkpoint can be triggered directly at and behind replication forks via complex signal cascades to maintain complete and accurate genome duplication. Importantly, two main signal transduction pathways exist to activate the S phase checkpoint. The **D**NA **R**eplication **C**heckpoint (DRC) pathway is active specifically during S phase and is signalled directly at stalled replication forks as a result of intrinsic-replisome problems. The intra-S **D**NA **D**amage **C**heckpoint (DDC) pathway is signalled by post-replicative DNA damage behind elongating forks, that is a direct result of template lesions that induce stalling (fig 1.7) (Garcia-Rodriguez et al., 2018).

There are many exogenous and endogenous insults that can disrupt faithful DNA replication. Various environmental factors; for example, UV light or ionizing radiation can cause damage to the DNA by inducing strand breakage or base alterations. Endogenous elements can also affect replication, including insufficient nucleotides, abasic sites, oxidation and methylation of DNA. Additionally, abnormal DNA structures that are difficult to unwind or proteins bound tightly to DNA could interfere with fork progression, causing them to stall. If these stalled forks are not stabilised and resolved, unreplicated regions/DNA breaks can occur. Common fragile sites also exist within the genome, which are more prone to breakage under conditions of replication stress.

To mimic these conditions in the lab and study the S phase checkpoint responses, the genotoxic drugs hydroxyurea (HU) and methyl methane sulfonate (MMS) are commonly used. Hydroxyurea is a chemotherapy drug, the primary cellular target of which is the enzyme ribonucleotide reductase (RNR). This enzyme is involved in the synthesis of the building blocks for replication, catalysing the reduction of precursor ribonucleoside diphosphates to deoxyribonucleotides (dNTPs). Thus, reduction in the active site of RNR by HU limits the cellular pool of dNTPs, causing the slowed

progression of forks and fork stalling (Singh and Xu, 2016). This activates the fork-associated DRC (Osborn and Elledge, 2003). Importantly, the effect of HU is reversible; once washed out of the media, stalled forks can resume replication, thus it is also a useful tool for synchronous S phase arrest.

MMS is an alkylating agent that is used to induce DNA damage. MMS methylate's DNA bases, predominantly guanine and adenine (on oxygen or nitrogen residues). Replication through these bulky DNA adducts causes base mispairing and slows down replication or blocks forks. It has been shown that the DRC is dispensable for survival in cells exposed to MMS, however the DDC is crucial, thus MMS is used mainly to study this pathway (Osborn and Elledge, 2003). Replication forks usually bypass blocking MMS lesions via re-priming downstream/ DNA damage tolerance pathways; resulting in ssDNA gaps behind forks on the daughter strand (Xiao et al., 1996, Garcia-Rodriguez et al., 2018). In the presence of DNA damage bypass pathways, timely activation of the DDC in response to polymerase stalling MMS lesions requires extension of post-replicative ssDNA gaps by Exo1-mediated nucleolytic processing (Garcia-Rodriguez et al., 2018). However, at higher doses of MMS, when fork stalling is rapidly induced, the DRC may also be activated (Bacal et al., 2018).

Both the DRC and DDC pathways share a common mechanism, whereby; following sensing of the replication stress or damage, the signal is transduced by different mediators from the sensory kinase Mec1 to the effector kinase Rad53. Importantly, strains with deletion of mediators in either pathway remain viable, suggesting some level of functional redundancy between the two different surveillance mechanisms. However, in budding yeast, as well as higher eukaryotes/humans, mutation of single checkpoint genes can severely impact genome stability and render cells more susceptible to genotoxic drugs.

The S phase checkpoint/signal cascade is conserved throughout evolution. Mec1, Tel1 and Rad53 are the budding yeast homologues of human ATR, ATM and Chk2, respectively. All of these are members of the phosphoinositide 3-kinase related kinase (PIKK-3) family, which act by phosphorylating target proteins. Mec1/Tel1 are sensory kinases that respond to a stress signal and once activated, transduce it to the effector kinase Rad53 (and to Chk1, which in budding yeast plays a minor role in the

checkpoint response) through phosphorylation of various mediator proteins. Phosphorylated/ activated Rad53 then acts to elicit the various downstream checkpoint responses (fig 1.7). Mec1 is the main sensory kinase; shown to respond to both replication stress and DNA damage signals, whereas Tel1 appears to be responsible only for responding to unprocessed double strand breaks (Longhese et al., 2003).

1.4.2 Sensing defects during DNA replication

The hallmark of defects during DNA replication is accumulation or stabilisation of ssDNA coated with RPA; which is used as the signal for checkpoint activation, recruiting the kinase Mec1. For activation of the DRC, ssDNA increases as a result of forks stalling. This can result from the exonucleolytic cleavage of the nascent DNA at blocked forks to generate checkpoint recognisable structures, or functional uncoupling of the helicase and polymerase. This generates longer stretches of unreplicated template in the proximity of the replisome. For activation of the DDC during S phase, ssDNA must accumulate away from the replication fork; a blocking lesion on the lagging strand, bypassed by re-priming downstream could leave a stretch of ssDNA behind the fork that can be targeted by exonucleolytic cleavage. In addition, formation of ssDNA intermediates is common amongst repair pathways, and exonucleolytic processing of DNA gaps, single strand breaks or DSBs results in accumulation of RPA and consequently recruitment of the kinase Mec1 to sites of DNA lesions (reviewed (Saldivar et al., 2017, Moriel-Carretero et al., 2019), see references therein).

Mec1/ATR acts in combination with an accessory regulatory protein. In *S. cerevisiae,* Ddc2 (ATRIP in humans) is stably associated with Mec1. Mec1 is recruited directly to sites of DNA damage/stalled forks via a direct interaction between Ddc2 and the Rfa1 subunit of RPA. Similarly, in human cells; following DNA damage, an accumulation of ATR-ATRIP foci is observed within the nucleus in an RPA dependent manner (Zou and Elledge, 2003, Costanzo et al., 2003).

The secondary, non-canonical checkpoint pathway depends on Tel1. This is similar to the canonical Mec1 pathway as it also converges on Rad53, however it requires other proteins for its recruitment to sites of unprocessed DSBs for its subsequent activation,

namely the MRX complex (Mre11-Rad50-Xrs2). MRX is recruited to DNA ends and processes DSBs to generate checkpoint recognisable DNA structures, activating the Tel1-dependent DDC (Usui et al., 2001). Usually, Tel1 activates the DDC in response to DSBs during G2. However, it also has some redundancy with Mec1 for activation of the intra-S DDC (Pardo et al., 2017).

1.4.3 The activation of Mec1

Importantly, other factors are required to activate Mec1 following its recruitment to ssDNA by RPA and Ddc2. During S phase, there are three activating proteins/complexes that have been identified as stimulators of Mec1 activity, in a partially redundant manner. The 9-1-1- complex is recruited independently from Mec1-Ddc2 to ssDNA and acts a as 'co-sensor'. It is comprised of Ddc1, Rad17 and Mec3 in a complex that is structurally similar to the DNA clamp PCNA; forming a heterotetramer that encircles DNA (Kondo et al., 1999). It is loaded onto sites of DNA damage/stalled forks at 5' ss-dsDNA junctions by the alternative clamp loader RFC^{Rad24} (Majka and Burgers, 2003). Subsequently, Mec1 is activated by the Ddc1 subunit (Navadgi-Patil and Burgers, 2008). Following phosphorylation of 9-1-1 by Mec1, Ddc1 is then also able to associate with Dpb11. Dpb11 physically interacts with Mec1-Ddc2, thus forming a bridge between the Mec1-Ddc2 and 9-1-1 complexes. Phosphorylation of Dp11 by Mec1 strongly enhances Mec1 activity (Puddu et al., 2008). Finally, Dna2 has also been implicated in enhancing the catalytic activity of Mec1; for example, it was shown to help activate Mec1 in response to HU-induced fork stalling to induce the DRC checkpoint. Dna2 binds to and is activated by the Sgs1 helicase which travels with forks (Wanrooij and Burgers, 2015). The stimulation of Mec1 by these activators appears to be a two-step process. Unstructured regions within Dpb11, Ddc1 and Dna2 contain two aromatic residues, the loss of which eliminates the activation of Mec1. These aromatic amino acids appear to bind the Mec1 activation site with low affinity, where subsequent binding with surrounding sequences increases the affinity and stability of the interaction to stimulate Mec1 (Navadgi-Patil and Burgers, 2009, Wanrooij et al., 2016).

Unexpectedly, it was observed that Mec1 is highly active during unperturbed S phase in a replication-dependent manner. This "replication-correlated" form of Mec1

appears to be uncoupled from activation of Rad53 and the checkpoint response. Mass spectrometry was used to identify substrates that are phosphorylated by Mec1 (and functionally redundant Tel1) independently from Rad53. The phospho-proteome of "replication-correlated" Mec1 involves targets which, paradoxically to its checkpoint function, facilitate fork progression during normal S phase. Those identified are involved in processes such as chromatin remodelling or RNA biogenesis. Importantly, in response to HU induced replication stress, unlike targets with a Rad53-dependent site which were strongly phosphorylated, the majority of Mec1/Tel1 targets were unchanged, or showed only minor stimulation or reduction in response to HU compared with unperturbed conditions. Thus, Mec1 appears to have two modes 1) facilitating the progression of replication during normal S phase and 2) activation of the checkpoint in response to DNA damage or replication stress, the downstream events of which would likely antagonise any of its replication-correlated effects. The study shows that both Dna2 and 9-1-1 are important for the activation of this replication mode of Mec1, but are not essential to activate Mec1 in response to MMS induced replication stress, suggesting additional, currently unidentified factors can also activate Mec1 as part of the checkpoint response (Bastos de Oliveira et al., 2015).

1.4.4 Relaying the signal to activate the S phase checkpoint

Following the recruitment and activation of Mec1 at stalled forks or sites of DNA damage, mediator proteins are responsible for relaying the signal to Rad53, which amplifies the signal and elicits the downstream checkpoint responses.

Mediators of the DNA Replication Checkpoint

A key mediator for activating the DRC is the fork component Mrc1, mutants of which show delayed Rad53 activation (Alcasabas et al., 2001, Tanaka and Russell, 2001). Mrc1 has multiple S/TQ motifs (17 in total) and is hyperphosphorylated by Mec1 following Mec1-Ddc2 recruitment to ssDNA within an abnormal fork structure (Kim et al., 1999b). This Mec1-dependent phosphorylation of Mrc1 results in activation of Rad53; whereby phosphorylated Mrc1 physically interacts with the FHA1 (fork-head associated) domain of Rad53, promoting its activation (Smolka et al., 2006). In HU treated cells, mutations eliminating all Mrc1 S/TQ residues (*mrc1*^{AQ}) results in

synthetic lethality when combined with deletion of *RAD9* (which can act as a backup pathway, activating the DDC to deal with damage accrued in the absence of Mrc1) (Naylor et al., 2009). Specifically, in a background with *rad9Δ* and upregulation of *RNR (*to increase dNTPs, which restores viability), *mrc1^{AQ}* is defective in activation of Rad53 and the DRC, failing to arrest the cell cycle in response to treatment with HU (Osborn and Elledge, 2003). Tof1, another member of the FPC has also been implicated in activation of the DRC, where *tof1Δ* cells antagonise the synthetic defects of *rad9Δ* in response to various genotoxic agents, and are defective for activation of Rad53 during HU perturbed S phase (Foss, 2001).

At forks, the alternative clamp loader RFC^{Ctf18} (Ctf18-Dcc1-Ctf8-Rfc2-5) is also required for Rad53 activation (Kubota et al., 2011, Crabbé et al., 2010, García-Rodríguez et al., 2015, Stokes et al., 2020). RFC^{Ctf18} has been shown to have the ability to both load and unload PCNA at ssDNA-dsDNA junctions (Bylund and Burgers, 2005). It interacts with Pol ε via the N-terminal catalytic domain of Pol2, and this interaction further stimulates the loading of PCNA (Stokes et al., 2020, García-Rodríguez et al., 2015, Fujisawa et al., 2017). In HU treated cells, mutations which only slightly perturb the binding between RFC^{Ctf18} and Pol ε result in synthetic genetic defects and impaired Rad53 activation in the absence of the DDC, thus highlighting the importance of this interaction for DRC activation (Stokes et al., 2020). However, the precise mechanism through which Mrc1 and Ctf18 activate Rad53 remains to be elucidated, however they may be responsible for recruiting Rad53 to the vicinity of Mec1 for its activation (Chen and Zhou, 2009).

Finally, the DNA helicase Sgs1 has also been implicated in activation of the DRC. Sgs1 also interacts with RPA, and following its phosphorylation by Mec1, like Mrc1, it interacts with the FHA1 domain of Rad53. Under conditions of replication stress (HU exposure), mutation within Sgs1 Mec1 phosphorylation sites results in defects in Rad53 activation in the absence of the DDC (Hegnauer et al., 2012).

Importantly, for intra-S checkpoint activation to occur, a critical threshold of arrested forks needs to be reached. Destabilising a subunit of ORC *(orc2-1)* results in a reduction of the number of functional replication forks to 30% of the usual cellular

value. Interestingly, this lower number of forks leads to severely reduced activation of the DRC in response to HU and MMS. This perhaps unexpectedly high level of tolerance for unusual DNA structures likely reflects the fact that even during a normal S phase, pausing of replication forks naturally occurs often. However, sister chromatids are available for quick, high fidelity repair of parental strand breaks by HR during this period. Therefore, the S phase checkpoint only needs to be activated if the amount of damage reaches a level that is higher than the 'background noise' of transiently paused forks, i.e. high levels of genotoxic insult that result in critical levels of stalled or collapsed forks, leading to greatly increased RPA accumulation on DNA (Shimada et al., 2002).

Mediators of the DNA damage checkpoint

Unlike the DRC, which is specifically signalled at replication forks, crucially the DDC pathway can be activated throughout the cell cycle as damage to the DNA can occur in any phase. In *S. cerevisiae,* the intra-S DDC halts the progression of cells into anaphase, to allow time for damage during DNA replication to be repaired before entry into mitosis. The mediator protein Rad9 activates Rad53 as part of the DDC by functioning as an adaptor for Mec1 (which is separately recruited to lesions by RPA coated ssDNA, see page 47). It also supports the autophosphorylation of Rad53 by serving as a scaffold protein (Sweeney et al., 2005).

Rad9 is recruited to chromatin at damaged DNA sites through interactions with Mec1/Tel1 phosphorylated histone H2A (γH2A; phosphorylated on serine 129) and methylated histone H3 (on lysine 79 by Dot1) (Lee et al., 2014, Giannattasio et al., 2005). Following its phosphorylation at multiple sites by Mec1, Rad9 acts as an adaptor and a scaffold protein, binding Rad53 FHA domains, promoting Rad53 recruitment, local accumulation, autophosphorylation and activation (Emili, 1998, Vialard et al., 1998, Sweeney et al., 2005, Smolka et al., 2007). Rad9 may be stabilised at lesions via an interaction with Dpb11, which occurs following Rad9 phosphorylation by CDKs (active in S/G2/M) (Pfander and Diffley, 2011). Once activated, Rad53 is released from Rad9, resulting in the amplification of the checkpoint signal and the activation the downstream response.

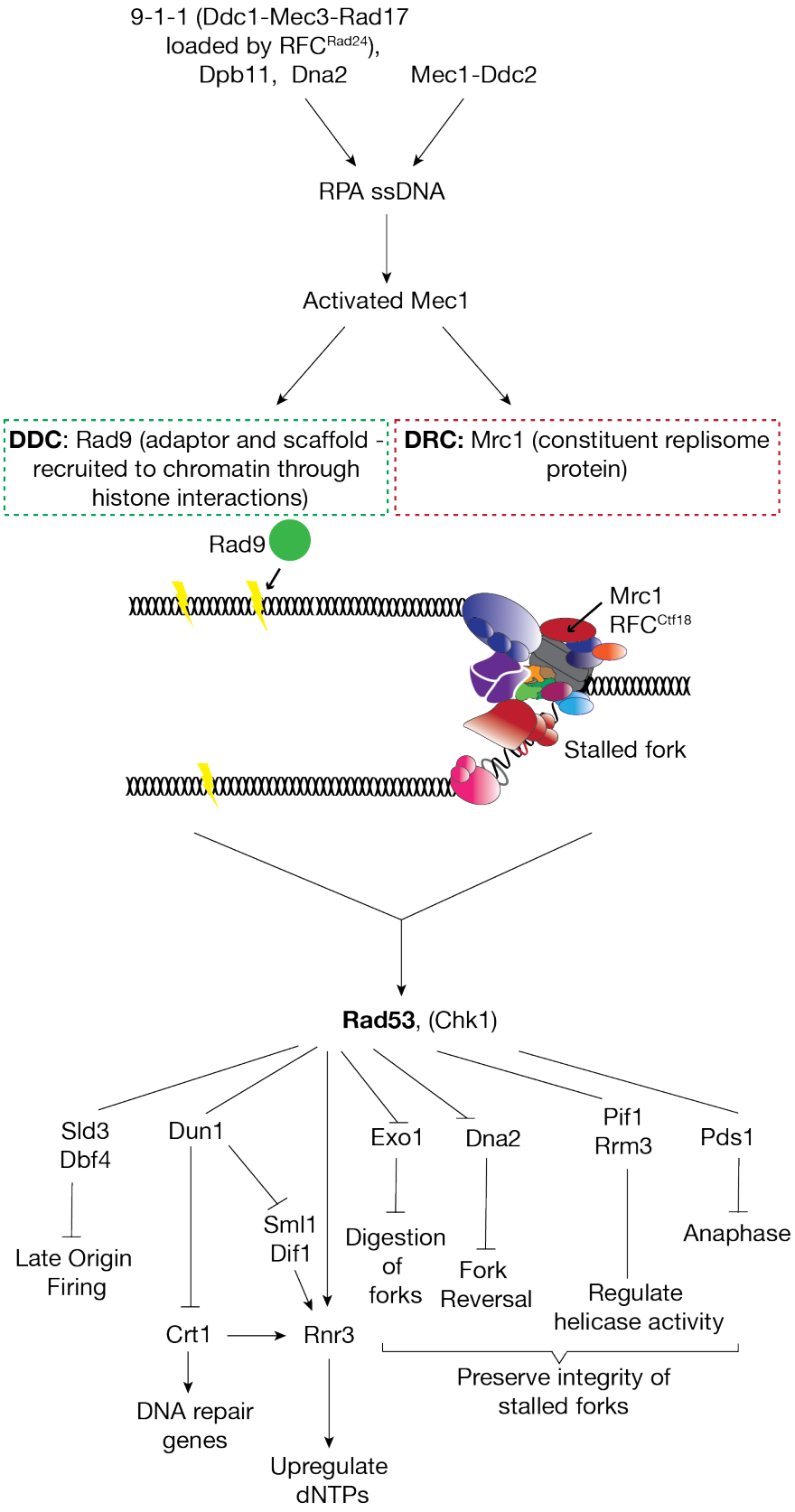


Figure 1.7 The S phase checkpoint. A schematic of the signal cascades induced at and behind forks by replication stress or DNA damage. The DNA damage checkpoint (DDC) is mediated by Rad9, and the DRC (DNA replication checkpoint) is mediated by fork component Mrc1. Activation of Rad53 by these signal cascades phosphorylates various target proteins to elicit the downstream checkpoint responses.

1.4.5 Rad53: Activation of the main checkpoint effector kinase

Rad53 is the main effector kinase of the S phase checkpoint and is phosphorylated/activated in a Mec1/Tel1 dependent manner through the action of the previously described DRC/DDC adaptor proteins. It is an essential protein in budding yeast due to its vital function in preserving dNTP levels during replication. However similar to Mec1, deletion of *SML1*, a ribonucleotide reductase inhibitor, can rescue the synthetic lethality of the null mutants. Sml1 is normally regulated by Mec1/Rad53 and its deletion upregulates dNTP synthesis in their absence (Zhao et al., 2001).

Rad53 consists of N- and C-terminal Forkhead associated domains (FHA), which are responsible for the recognition of phosphorylated molecules; and a central serine/threonine kinase domain (Durocher et al., 1999). It also has two S/TQ cluster domains (SCDs). Enrichment of these S/TQ residues presents numerous targets for phosphorylation by Mec1/Tel1 (Traven and Heierhorst, 2005, Wybenga-Groot et al., 2014). Rad53 exists as inactive homodimers, where the kinase domain is in a closed conformation; however, following phosphorylation of the adaptor proteins Mrc1 or Rad9 by Mec1, it is recruited to sites of DNA damage/stalled forks via interactions between its FHA domains and these adaptors proteins. Rad53 is then phosphorylated at multiple S/TQ residues within its N-terminal SCD by Mec1 (Chen et al., 2014a, Wybenga-Groot et al., 2014), which primes it for extensive autophosphorylation. *In trans* autophosphorylation of Thr354 (within the activation segment of the kinase domain) removes a self-inhibitory loop and results in full exposure of Rad53 catalytic site, and hence, full catalytic activation (Wybenga-Groot et al., 2014). Thus, the checkpoint response is amplified and Rad53 is released and phosphorylates its target proteins to elicit the desired downstream responses. The control of the pathways and genes targeted by the checkpoint are summarised figure 1.7 and will be described briefly in the next section.

1.4.6 The downstream targets of S phase checkpoint

The S phase checkpoint targets a number of genes in order to; arrest the cell cycle, halt origin firing, organise DNA repair and upregulate the levels of dNTPs. It also plays

the crucial role of preserving stalled replication forks in a state where they can resume DNA replication.

Repression of replication origin firing and cell cycle arrest

As described in section 1.3.2, origin firing occurs via a temporal program; where some origins are activated later in S phase, while others remain as 'back-ups' and are only activated in the event two converging forks collapse. To delay the cell cycle and allow time for DNA repair processes to occur, the firing of late origins is inhibited by Rad53-mediated phosphorylation of the firing factors Sld3 and Dbf4 (Zegerman and Diffley, 2010, Lopez-Mosqueda et al., 2010). Preventing new origin initiation avoids the replication of templates harbouring damaged regions. It also means that late-firing and back-up origins do not encounter unresolved damage, which could result in their stalling or collapse, thus preserving them for firing once the checkpoint is satisfied. Mutation of the Sld3 and Dbf4 residues phosphorylated by Rad53 results in unchecked progression through MMS perturbed S phase, and preservation of late origins is crucial in mutants with unstable forks when dNTPs are depleted (Zegerman and Diffley, 2010). Interestingly, Rad53-mediated phosphorylation of Sld3 has been shown *in vivo* to require Cdc45, which targets Rad53 to replication complexes. Phosphorylation of Cdc45 by Rad53 at T189 or T195 creates a site with which the FHA1 domain of Rad53 can bind, and this interaction promotes the inhibitory phosphorylation of Sld3 (Can et al., 2019). It is less clear how Rad53 inhibits Dbf4, however it has been shown to interact directly with its N-terminus (Matthews et al., 2014).

A second kinase, Chk1 (which is also activated by Mec1), prevents entry into anaphase by targeting Pds1 to block the degradation of securin and therefore sister chromatid separation. This ensures cells do not enter mitosis with damaged or unreplicated DNA (Agarwal et al., 2003). Finally, multiple other factors involved in mitotic exit have also been identified as enriched in the Rad53 phospho-proteome (Zhou et al., 2016).

Effects on gene expression

Following activation of the checkpoint, the control of transcription is altered to either induce or maintain the expression of genes involved in DNA replication and DNA repair. In one pathway, phosphorylation of the protein kinase Dun1 by Rad53 leads to

the inactivation of the transcriptional repressor Crt1. This results in the induction of various DNA repair genes. Additionally, loss of Crt1 results in the expression of *RNR* (Ribonucleotide reductase; involved in the pathway for dNTP synthesis), which therefore increases the levels of dNTPs within the cell. This ensures the pool of dNTPs is sufficient for replication to resume successfully (Huang et al., 1998, Poli et al., 2012). In a second pathway, Rad53 phosphorylates Nrm1, which is a transcriptional repressor of MBF-regulated promoters (characteristic of G1/S phase genes). These are also involved in various replication and repair processes and include the Clb6 cyclin. This results in the continued expression of these genes during periods of damage/stress (Travesa et al., 2012).

Preserving the integrity of stalled replication forks

Most importantly the S phase checkpoint acts to preserve the integrity of stalled forks so that they are able to restart replication once the stress is removed. Checkpoint null cells with mutations in either Rad53 or Mec1 cannot resume replication following exposure to HU or MMS, even after removal of the drug (Desany et al., 1998, Morafraile et al., 2015, Tercero and Diffley, 2001). These restart incompetent forks are defined as collapsed, though replisomes seem to remain intact. Thus, it appears the checkpoint acts to stabilise replisomes by regulating their function rather than preventing them from breaking apart. In the absence of checkpoint kinases, replisomes remain on chromatin but may move away from sites of DNA synthesis (De Piccoli et al., 2012).

Indeed, evidence is accumulating to suggest that the replisome itself is a target of Rad53 and Mec1-dependent regulation. In the presence of HU, both IPs and ChIP of HA tagged Rad53 revealed that Rad53 is recruited to stalled forks via an interaction with Cdc45, a core replisome component. Loss of this interaction did not affect the viability of cells exposed to various genotoxic agents. However, synergistic defects were observed in combination with an allele of the replisome component Mrc1 that cannot bind Rad53. Loss of an interaction between Cdc45, Mrc1 and Rad53 did not affect Rad53 activation, but the recruitment of Rad53 to forks by Mrc1 and Cdc45 was shown to be important for the ability of stalled forks to re-start DNA replication, independent from their role in the regulation of origin firing. Finally, other currently

unidentified interactions between the replisome and Rad53 appear to exist (Can et al., 2019).

Data from an even more recent study suggests that Rad53 functions at replication forks to prevent the CMG helicase from becoming uncoupled from DNA synthesis, thus preserving the stability of the forks for replication re-start. In a reconstituted replication system using purified *S.* cerevisiae proteins, uncoupling of DNA unwinding and DNA synthesis was achieved by depleting the levels of dNTPs, or by inactivating the leading strand polymerase by both mutation and chemical inhibition. Excessive unwinding of DNA ahead of the polymerase was prevented by Rad53. Intriguingly however, this was found to be independent from both Mrc1 and Cdc45, thus the importance of the reported interactions and the precise mechanism through which Rad53 controls the stability of stalled forks is an interesting topic for future studies (Devbhandari and Remus, 2020).

Under conditions of replication stress that lead to fork stalling, Rad53 activity also helps maintain forks as competent for restart by inhibiting exonucleases, such as Exo1, which are involved in generating ssDNA to prime the process of homologous recombination (HR) at DSBs. This prevents DNA digestion at stalled forks that could compromise their integrity, thus preventing the accumulation of ssDNA - which potentially increases toxicity by using up all of the cells RPA, fork collapse, formation of DSBs and potentially deleterious recombination (Cotta-Ramusino et al., 2005). In *RAD53* null cells, deletion of *EXO1* rescues terminal arrest and fork restart defects in response to MMS and HU, respectively, suggesting its checkpoint-mediated inhibition is crucial for preserving the integrity of stalled forks (Morafraile et al., 2015).

During fork stalling, the S phase checkpoint also inhibits HR. The central HR repair protein Rad52 re-localises to form nuclear repair foci during S phase, where it binds to ssDNA and mediates the formation of Rad51 nucleofilaments; these invade homologous regions of a sister chromatid for the high-fidelity repair of DSBs. Rad52-GFP foci are used as a molecular marker for HR-mediated DSB repair (Lisby et al., 2001). When replication forks stall as a result of HU exposure, the formation of Rad52-GFP foci is actively suppressed by the S-phase checkpoint (Lisby et al., 2004). However, in the absence of the checkpoint proteins Rad53 or Mec1, HU stalled forks

are subject to collapse, resulting in the formation of one-ended DSBs. In these cells, Rad52 foci are observed, suggesting it is recruited to collapsed replication forks (Tercero and Diffley, 2001).

Interestingly, cells become hypersensitive to fork-stalling agents including MMS and HU in the absence of certain genes involved in the HR pathway, including *RAD52* (Chang et al., 2002). This data suggests that in response to some lesions, recombination may be critical for the re-start of stalled forks to prevent their irreversible collapse. It has been proposed that 'beneficial' recombination can occur using ssDNA already at a stalled fork to prime the HR pathway to promote fork re-start, while the Mrc1-mediated branch of the checkpoint suppresses excess fork cleavage by protecting the ssDNA from resection by exonucleases. It also functions to prevent genomic instability by restricting recombination at chromosomal DSBs while conditions of replication stress persist. Supporting this idea, in HU-treated cells, the generation of ssDNA tails at DSB ends is much slower following S-phase checkpoint activation, which likely allows time for HR-mediated fork restart before excess resection can occur (Alabert et al., 2009).

In *Sz. pombe* it has also been identified that phosphorylation of Dna2 promotes nucleolytic cleavage of any leading or lagging strand precursors to prevent the formation of reversed forks, which could be recognised by exonucleases and cleaved, preventing fork restart (Hu et al., 2012). The action of the Pif1 and Rrm3 helicases is also inhibited. These helicases usually help to promote replication across stable protein blocks, facilitating lagging strand synthesis, however at stalled forks, their helicase activity could lead to abnormal fork structures that result in fork reversal (Rossi et al., 2015).

The checkpoint is also thought to play a role in the control of gene gating, where under normal conditions, transcribed chromatin is coupled to the cell's periphery for timely export. Rad53 phosphorylates nucleoporins, such as Mlp1, thus releasing transcribed genes from the nuclear pores and alleviating any topological stress on the DNA (Chen et al., 2010, Smolka et al., 2007). Once again, this helps to protect the stability of stalled forks (Bermejo et al., 2011).

1.5 The process of transcription as a barrier to DNA replication and a threat to genomic stability

Replication forks encounter a number of barriers that can impede their progression. These include regions of damaged DNA, protein complexes that are tightly bound to the template DNA and secondary non-B form DNA structures that are difficult to unwind; for example, hairpins or G4 quadruplexes (reviewed (Lerner and Sale, 2019, Mirkin and Mirkin, 2007)).

Significantly, during S phase, DNA replication is not the only process that uses the DNA as its substrate. RNA transcription occurs on DNA throughout the cell cycle, and during S phase competes with the replication machinery for this same template. Inevitably, as both of these processes are essential for viability and occur frequently, collisions between replication forks and transcription bubbles (transcription replication clashes, referred to as TRCs from this point forward) are a common event. In human cells, transcription of certain long genes can last across a number of cell cycles, making TRCs in these regions hard to evade (Helmrich et al., 2011). Moreover, near early origins, early replicating fragile sites coincide with clusters of highly transcribed genes, which is likely due to increased TRCs in these regions (Barlow et al., 2013). Importantly these clashes have been linked to genomic instability in both prokaryotes and eukaryotes (Srivatsan et al., 2010, Hamperl et al., 2017, Liu and Alberts, 1995, Helmrich et al., 2011, Helmrich et al., 2013, Prado and Aguilera, 2005).

TRCs in prokaryotes

Collisions between the transcription and replication machineries were first identified in bacteria, which initiate DNA replication bi-directionally from a single origin. Depending on the orientation of a gene relative to the origin of replication, TRCs occur either in a co-directional (CD) or a head-on (HO) orientation. These CD and HO clashes occur when a gene is encoded on the replicative leading and lagging strand, respectively (fig 1.8). Interestingly, it appears that HO collisions are particularly deleterious for genome stability. An experiment which placed an Ori flanking either side of a highly transcribed rRNA operon in *E. coli* observed that HO collisions between forks and transcription complexes significantly delayed the rate of DNA

replication (French, 1992). Concordantly, in an evolutionary effort to avoid these more deleterious HO collisions, a significant co-orientation bias of transcription and replication within prokaryotic genomes is observed (Blattner et al., 1997). Indeed, genes that are most essential or highly transcribed are even further enriched for this co-orientation bias (Srivatsan et al., 2010). However, though to a lesser extent, CD clashes can also be a source of instability; particularly in situations where forks encounter a RNAP that is persistently blocked, or in a more stable conformation than during normal elongation. For example, clashes with a backtracked RNAPII molecule, which is no longer actively transcribing but adopts a highly stable conformation, can lead to formation of DSBs (Dutta et al., 2011).

TRCs in eukaryotes

Accumulating evidence in eukaryotes also suggests that TRCs are a source of genome instability (Prado and Aguilera, 2005, Kim et al., 2007, Helmrich et al., 2011, Helmrich et al., 2013, Hamperl et al., 2017, Barlow et al., 2013, Garcia-Rubio et al., 2018, Lang et al., 2017). Interestingly, in the human genome, though this is complicated by the existence of multiple Oris, analysis of high-resolution Okazaki fragment sequencing has also shown that transcription bubbles and replication forks are significantly co-oriented (Petryk et al., 2016).

A seminal study from the Aguilera lab examined the levels of recombination in *S. cerevisiae* cells harbouring plasmid-borne constructs, where transcription (under control of an inducible promoter) is designed to clash in either a HO or CD orientation with replication. The constructs contained two *leu2* direct repeat motifs, the recombination of which results in the generation of wild type *LEU2* that can be selected for by growth on media lacking leucine. These repeats were oriented either toward (IN) or away from (OUT) the *ARSH4* Ori. Comparing these constructs under control of the *GAL1* promoter, the frequency of recombination was only ~1.6x higher for the OUT construct when transcription was active versus inactive. This suggests that CD orientation has little influence on recombination levels. However, in the presence of galactose, recombination frequency was increased ~5.5x for the IN (HO) construct. Importantly, using cell cycle specific promoters in the place of *GAL1,* this increase in recombination was observed specifically when transcription was active

during S phase. Thus, replication is required to observe these increased levels of transcription-associated recombination. The higher levels of recombination in the presence of HO TRCs, and to a lesser extent CD TRCs, was linked to the appearance of a replication fork pause (RFP). This suggests the increase in recombination is possibly a result of a blockage in fork progression by transcription (Prado and Aguilera, 2005). Indeed, in a chromosomal context, genome-wide ChIP analysis to map the occupancy of Pol2 (Pol ε) found fork progression is impeded at highly transcribed RNAPII genes. These highly transcribed ORFs made up the majority of sites identified as having high Pol2 binding in wild type cells (Azvolinsky et al., 2009).

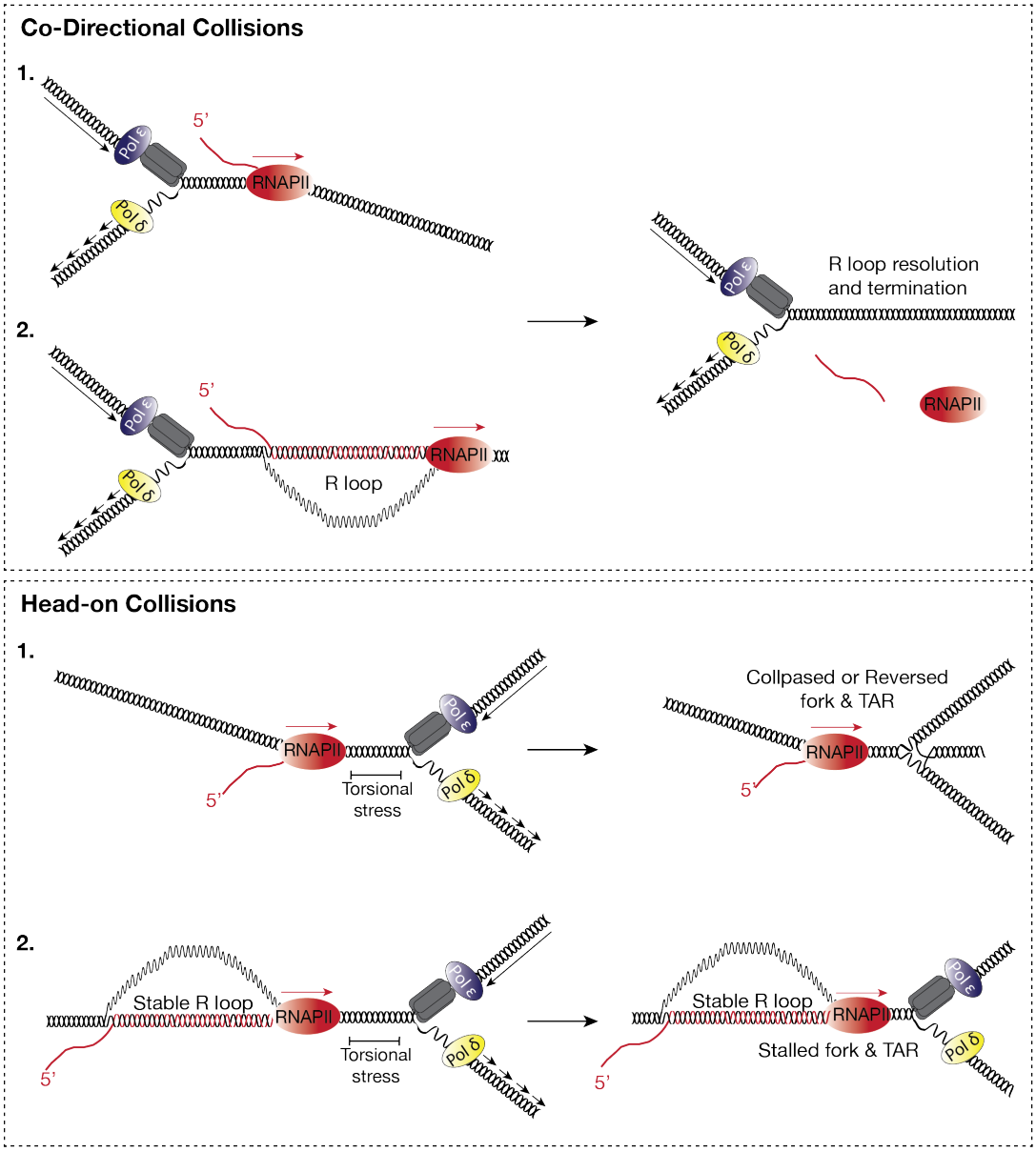


Figure 1.8 Collisions between transcription bubbles and replication forks. Collisions between transcription bubbles and replication forks can happen in both head-on and co-

directional orientations. Co directional clashes are less deleterious for genome stability, and result in termination of the RNAPII transcription complex. In this orientation, co-directional R loops are also resolved. Head-on collisions are more deleterious and can result in collapsed or reversed forks, leading to transcription associated recombination and may also promote the formation of stable R loops behind the elongating RNAPII.

Another study in *S. cerevisiae* found that at increased levels of transcription, there is a proportional increase in the rate of spontaneous mutations. A system was used where a reporter *lys2*^{frameshift} gene was placed in either the HO or CD orientation relative to the *ARS306* origin. The reversion of this allele to lysine prototrophy occurs as a result of mutagenesis within a 'reversion window'. This found that although the overall rate of mutations was not affected by the orientation, the mutation spectrum was different. Compared with the same orientation, in the opposite orientation; large deletions were enriched at low transcription levels, and complex insertions and deletions were increased at high transcription levels. This suggests that clashes in different orientations lead to distinct mutational events (Kim et al., 2007).

Importantly, the transcription and replication machineries may not have to physically come into contact with one another to impair the progression of forks. Both of these processes have a significant effect on the topology of the template DNA. During transcription, unwinding of the template DNA results in positive supercoiling (over-winding) of the DNA ahead of elongation complexes, and negative supercoiling (underwinding) behind. Resolving this topological stress relies heavily on the action of topoisomerases, as mRNA processing and gene gating (tethering of transcribed genes to nuclear pores) generates further topological constraints. DNA replication also results in an accumulation of positive supercoiling ahead of forks, which can be resolved either by passive fork rotation or by topoisomerases. However, as advancing replication forks and transcription elongation complexes converge, the increased torsional stress before they reach one another may be responsible for the generation of genomic instability. The accumulation of positively supercoiled DNA before forks can lead to their stalling or reversal (fig 1.8) (Lin and Pasero, 2012).

In addition, fork progression can also be impeded by the formation of co-transcriptional R loops. High levels of R loops result in increased levels of DNA damage and recombination. Interestingly, their formation may be favoured by HO

TRCs (Hamperl et al., 2017, Lang et al., 2017). These R loop structures and their effect on the process of DNA replication will be discussed in more detail in the next chapter.

1.6 The structure, formation and physiological role of R loops

1.6.1 The structure of R loops

During the process of transcription, it is possible for nascent RNA to re-anneal with its template DNA strand. This results in displacement of the non-template DNA strand as exposed ssDNA. This three-stranded structure is referred to as an R loop, and can extend up to a length of 1kb in higher eukaryotes (fig 1.9) (Aguilera and García-Muse, 2012).

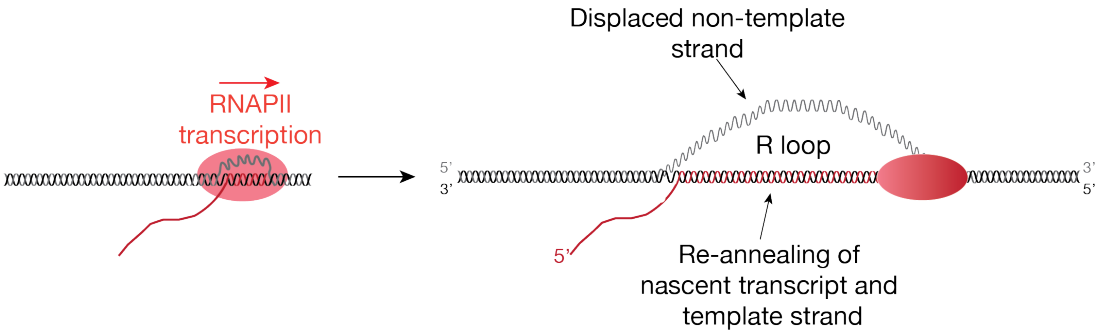


Figure 1.9 The structure of an R loop. R loops are extended RNA:DNA hybrids with a displaced ssDNA strand. They are formed as a result of transcription, where nascent RNA re-anneals with the template DNA strand. The displaced non template strand is exposed as ssDNA.

A thread-back model for the formation of R loops has been suggested, whereby invasion of the template DNA strand is thought to occur following exit of the newly transcribed RNA from the polymerase. This is the most widely accepted model and is supported by crystallographic data regarding the structure of RNAPII. This showed that a short 8 base pair RNA:DNA hybrid is formed within the transcription bubble (fig 1.9), however separate channels are responsible for the independent exit of DNA and RNA from the polymerase (Westover et al., 2004). An alternative model proposes that in some cases, the hybridised RNA:DNA could exit through the same channel, thus is extended as the transcription bubble elongates (Aguilera and García-Muse, 2012).

Regardless of the exact mechanism responsible, once formed R loops are more thermodynamically stable than double stranded DNA (Roberts and Crothers, 1992). This is thought to be because the conformation adopted is an intermediate between the A- and B- forms of dsRNA and dsDNA, respectively (Shaw and Arya, 2008). Due to this inherent stability, the action of enzymes is required to resolve these structures for restoration of the dsDNA helix. The next section will discuss the physiological roles of R loops and the factors affecting their formation. I will also consider the consequences of their aberrant formation for genome stability and finally, the mechanisms employed by cells to cope with them.

1.6.2 The physiological roles of R loops

Though unscheduled accumulation of R loops can have a negative impact on genome stability (discussed in detail later), physiological R loops are required as intermediates for certain cellular processes. Consequently, they have been coined a 'double-edged sword' for genome stability (Skourti-Stathaki and Proudfoot, 2014). Many physiological roles have been described, such as: priming the initiation of bacterial plasmid and mitochondrial DNA replication, the facilitation of Immunoglobulin class switch recombination in activated B-cells, the maintenance of telomeres and the control of gene expression.

In extracts of *E. coli,* replication initiation of the ColE1 plasmid minimally requires purified RNA polymerase, DNA polymerase and RNase H. Treatment with rifampicin, an RNA polymerase inhibitor affects ColE1 replication *in vivo*, which was also shown to be dependent on the presence of rNTPs *in vitro* (Tomizawa, 1975, Sakakibara and Tomizawa, 1974). At the *E. coli* chromosomal origin (*oriC*), this dependence on RNA synthesis for replication is not observed as it is initiated by the canonical *dnaA* primase. However, for the initiation of ColE1 replication, a 550bp plasmid encoded transcript was found to form an R loop on the leading strand across the Ori. Processing of the R loop by RNase H generates a 3'-hydroxyl end, which acts to prime further strand extension by the DNA polymerase (Sakakibara and Tomizawa, 1974, Itoh and Tomizawa, 1980). Interestingly, similar to ColE1, in the absence of the primase *dnaA*, or following inactivation of *oriC*, *rnhA* mutants maintain the ability to replicate. This has been attributed to the persistent formation of R loops, whereby

unprocessed DNA:RNA hybrids can also be used to prime DNA synthesis both *in vivo* and *in vitro* (Dasgupta et al., 1987). Similarly, R loops have also been implicated in priming the synthesis of mitochondrial DNA. Both in *S. cerevisiae* and in human cells, RNase H sensitive R loops were found to form *in vitro* during transcription over GC-rich elements at mitochondrial origins of replication. Again, once processed by RNase H, the 3' end can be used as a primer for DNA synthesis (Xu and Clayton, 1996, Pohjoismäki et al., 2010, Baldacci et al., 1984).

R loops are also believed to play a role in activated B-cells when switching synthesis from IgM to other Immunoglobulin (Ig) classes. This occurs by the removal of DNA between two 'switch' regions so only one isotype gene remains (reviewed by (Chi et al., 2020), see references therein). In brief, Ig switch regions (which have high GC skew) are located adjacent to areas of the genome encoding each of the various antibody heavy chains. At two selected switch regions, the activity of various enzymes, including activation-induced cytidine deaminase, which deaminates dC bases to dU on ssDNA, thus causing its removal) results in DNA nicks on both strands, which ultimately results in a double strand break. The intervening segment between the dsDNA breaks is lost, and the gap is repaired by non-homologous end joining. The G-rich switch regions are ideal to promote the formation of R loops, and therefore these structures are a possible candidate to provide access to ssDNA for activation-induced cytidine deaminase activity. R loops formation would result in exposure of the displaced non template strand as ssDNA, and subsequent RNase H cleavage of the RNA moiety within the RNA:DNA hybrid would also expose the template strand as ssDNA (Roy et al., 2008).

R loops also play a role in the maintenance of telomeres, which are highly repetitive protein-associated regions of heterochromatin that protect the ends of linear chromosomes from degradation, end fusions or deleterious recombination (Toubiana and Selig, 2018). As the ends of DNA cannot be fully replicated, they become gradually shorter after each successive cell cycle. Once a 'critical' length is reached, the cells enter replicative senescence (Hemann et al., 2001). The reverse transcriptase telomerase can enzymatically re-extend shortened telomeric repeats, however it is not active in most somatic cells, but is present in germ and stem cells. Importantly, studies in *S. cerevisiae* have been a large part of understanding the biology of telomeres, as

telomerase is expressed by wild-type yeast. The transcription of telomeres by RNAPII results in the production of a G-rich, conserved, telomeric repeat-containing long noncoding RNA, termed TERRA. In *S. cerevisiae* (as well as many other species), TERRA has been shown to reanneal with the complementary strand to form R loops *in cis* (reviewed (Toubiana and Selig, 2018, Lalonde and Chartrand, 2020), see references therein). In the absence of telomerase, these telomeric R loops have been implicated in the maintenance of telomeres by promoting homology directed repair (a process termed alternative lengthening of telomeres; ALT). In this background, a small proportion of yeast cells (termed Type II survivors) are able to escape senescence in a manner that depends on Rad52-mediated recombination. Down- or up- regulating the levels of TERRA decreases and increases the levels of recombination *in cis,* respectively. Importantly, it appears only short telomeres are prone to R loop mediated-HR, while long telomeres are less so (Lundblad and Blackburn, 1993, Balk et al., 2013). As such, under normal circumstances, an accumulation of R loops at telomeres may have negative consequences for genome stability; they could interfere with processive DNA replication, leading to double strand breaks and potentially deleterious levels of recombination. Their levels are regulated in several ways; an accumulation of R loops at telomeres is observed in the absence of mRNA biogenesis factors and the RNase H enzymes (which resolve R loops) (Pfeiffer et al., 2013, Balk et al., 2013). In addition, in a cell cycle dependent manner, Rat1, a 5'-3' RNA exonuclease degrades TERRA prior to the replication of telomeres to restrict R loop formation, and RNase H2 is recruited by Rif2 for their degradation to clear the way for forks in these regions. However, in cells with critically short telomeres, this cell cycle dependent control of TERRA is overcome, and R loops persist to activate the ALT pathway. It has been suggested the persistence of R loops at critically short telomeres may cause oncoming replisomes to stall, triggering DSB formation and therefore recombination (Luke et al., 2008, Graf et al., 2017).

R loops have also been implicated in the process of transcription and are enriched both at the 5' and 3' end of genes. In human cells, a region of high GC skew lies in the 5' untranslated region, directly downstream of CpG promoters which are silenced by DNA methylation. Protection against methylation requires transcription, and it has been suggested that in these regions, the exposed ssDNA within R loops may serve as a signal to recruit the defensive H3K4 trimethyl mark or recruit DNA demethylating

complexes (Ginno et al., 2012). The exposed ssDNA template within R loops has also been shown to potentially act as a promoter for widespread RNAPII transcription of antisense long non-coding (lnc) RNAs (Tan-Wong et al., 2019). The lncRNAs themselves can also form regulatory R loops; which play a role in controlling gene expression by functioning as epigenetic markers to affect the local chromatin environment, promoting histone modifications or recruiting regulatory proteins (reviewed (Niehrs and Luke, 2020)).

Finally, R loops are thought to play a role in the efficient termination of RNAPII transcription. In certain human genes, the formation of R loops downstream of the poly(A) site at G-rich regions has been shown to facilitate the pausing of RNAPII. Indeed, these 3' G-rich pause sites are common across many mammalian genes (Skourti-Stathaki et al., 2011, Salisbury et al., 2006). Resolution of R loops by overexpression of RNase H1 results in transcription-readthrough, suggesting the hybrids are required for efficient termination. Conversely however, a failure to remove these R loops also results in transcription-readthrough. Thus, their resolution which allows subsequent exonucleolytic cleavage of the nascent RNA is a second step that is also required for efficient pause-site dependent RNAPII termination (Mischo et al., 2011, Skourti-Stathaki et al., 2011).

1.6.3 Factors promoting the formation of R loops

Several factors are known to promote the formation of R loops. Regions within DNA that have a high GC-skew (where the template strand is enriched for C vs G) and highly transcribed genes are particularly prone. High levels of R loops are observed at both ends of genes, namely at the 5' end immediately downstream of the promoter and at the 3' end across termination sites. At these regions, the G-rich transcripts reanneal with the C-rich template, displacing the non-template strand (Ginno et al., 2012, Ginno et al., 2013). In addition, increases in local or unresolved negative supercoiling behind elongating transcription bubbles results in transient separation of the DNA strands, thus providing an opportunity for the nascent RNA to invade the template strand (El Hage et al., 2010). The integrity of DNA can also have an effect; a nick in the non-template DNA strand at a site downstream of a promoter was shown to increase the formation of R loops. This is likely due to an increased competition between the non-

template strand and nascent RNA for template hybridisation at the nick site, where rewinding of the DNA duplex is less efficient (Roy et al., 2010).

Finally, some evidence suggests that the orientation of RNA transcription and replication fork movement affects the formation of R loops. In bacteria, DNA:RNA immunoprecipitation and qPCR (DRIP-seq) revealed that head-on TRCs promote the formation of R loops across the region where the collisions were engineered to occur (Lang et al., 2017).This echoes a study in human cells, which demonstrated that the direction in which replisomes and transcription bubbles clash has opposite effects on the stability of R loops. In both a plasmid-based system and in a native genomic context, in the presence of transcription and following release from G1/S block, enrichment of R loops was observed specifically during DNA replication in the HO orientation. In this case, it is likely that any convergence of the two machineries could result in increased topological stress and stalling, thus promoting R loop formation. The resultant R loops could further stabilise the RNA polymerases, exacerbating the block. Interestingly, in the CD orientation, the level of R loops actually decreased, suggesting they can be cleared by CD replisomes. Many replication factors have 3' to 5' RNA:DNA helicase activity. Alternatively, as in the CD orientation R loops would be formed on the leading strand, the MCM helicase itself could play a role in their resolution (Hamperl et al., 2017).

Though most studies suggest that R loops are formed co-transcriptionally (*in cis)* (Huertas and Aguilera, 2003, Hamperl et al., 2017)*,* it has been suggested that R loops can also form *in trans*, where their post-transcriptional formation depends on Rad51 via a stand exchange reaction. One study found that the deletion of *RAD51* results in a reduction of R loop signal in several mRNA biogenesis mutants, and it co-localises to sites of R loops before any detectable γH2AX signal implies the formation of a DSB. They also demonstrated *in vivo* that R loops are able to form *in trans,* away from the transcription site in a Rad51-dependent manner. An independently replicating yeast artificial chromosome (YAC) with a human DNA sequence was transformed into a strain encoding a homologous YAC sequence on chromosome III placed under control of a promoter initiating aberrant transcription. This showed that these aberrant transcripts from chromosome III could invade the YAC and lead to R loop-mediated

instability (Wahba et al., 2013). Importantly, both the hybrid signal at these sites and YAC instability was suppressed in the absence of Rad51. Additionally, the deletion of *SRS2*, a helicase that inhibits Rad51 by removing its filaments from ssDNA led to an increase in R loops signal and YAC instability. Importantly however, deletion of *RAD51* had no effect on the R loops signal or YAC instability in *rnh1Δ rnh201Δ* cells, suggesting it is not always a requirement for R loop formation, but one mechanism. In the absence of Srs2, cells accumulate R loops across highly repetitive rDNA loci, and this study postulates that this homologous repeat containing region may be a good substrate for Rad51 mediated *in trans* R loop formation in situations when its strand invasion activity is not regulated. This also implicates Srs2 as a factor protecting cells against the formation of R loops (other mechanisms are discussed in further detail in section 1.8) (Wahba et al., 2013).

1.7 R loops as threats to genome stability

Though R loops are formed as natural by-products of the transcription process, and are involved in the previously described physiological events, evolving evidence suggests that their occurrence is more frequent than has previously been suggested (Aguilera and García-Muse, 2012). Their unscheduled or unchecked formation can have disastrous consequences. Indeed, many studies show increased genomic instability in yeast lacking factors that either prevent the formation of or resolve R loops. These cells show increased mutagenesis, recombination and gross chromosomal rearrangements (Huertas and Aguilera, 2003, Wahba et al., 2011, Li and Manley, 2005, Mischo et al., 2011). Several, non-exclusive models have been proposed regarding how R loops may lead to genome instability.

Firstly, the regions of exposed ssDNA within R loops are more unstable than dsDNA, thus are more susceptible to chemical modifications or DNA damage. These lesions can lead to transcription associated mutagenesis (TAM) or transcription associated recombination (TAR) (Skourti-Stathaki and Proudfoot, 2014). For example, *in vivo,* C-G to T-A transitions spontaneously occur 140x more frequently on ssDNA, and its increased susceptibility to events such as spontaneous deamination of dC to dU could lead to TAM or TAR (Frederico et al., 1990, Skourti-Stathaki and Proudfoot, 2014, Li and Manley, 2005). Furthermore, in human cells, the processing of R loops to double

strand breaks by transcription coupled nucleotide excision repair proteins can result in recombination and genome rearrangements (Sollier et al., 2014).

Secondly, R loop-dependent genome instability is also thought to occur by interfering with processive DNA replication. In particular, co-transcriptional R loops may stabilise transcription complexes on the DNA, blocking the passage of forks. Additionally, if forks encounter any unrepaired DNA lesions, bound proteins or abnormal DNA structures on the displaced ssDNA of an R loop, their progression may be obstructed. The stalled forks are more susceptible to collapse or nucleolytic cleavage, leading to double strand breaks and potentially deleterious recombination. Furthermore, any impediments could also interfere with the elongation of other transcription complexes, exacerbating the blockage (Skourti-Stathaki and Proudfoot, 2014).

As discussed in section 1.6.3, studies in bacterial and human cells have suggested that HO collisions between transcription and replication are more deleterious and promote the formation of R loops (Saldivar et al., 2017, Hamperl et al., 2017). In human cells, collisions in the HO orientation were linked to increased genome instability, and the different orientations were shown to elicit distinct DNA damage responses. HO collisions are more harmful, leading to an increased plasmid loss, which was partially suppressed by RNase H overexpression to resolve any R loops. In addition, a robust γH2AX signal was detected in these cells, which is an early marker for DSB formation, resulting in robust activation of the ATR-dependent DNA damage response, which responds to ssDNA at blocked replication forks. On the other hand, a greater number of CD collisions were required to induce a detectable response, where conversely, ATM (Tel1), which responds to double strand breaks was activated by CD clashes. In this case, the formation of double strand breaks could be a result of collisions between the replication machinery and an R loop associated backtracked RNA polymerase, which is more stably associated to the DNA (Hamperl et al., 2017).

A recent study in yeast, also examined the effects of CD and HO collisions in the formation of R loops and their effect on genome stability (Garcia-Rubio et al., 2018). Using the pGAL IN/pGAL OUT recombination system (described in section 1.5), the

elevated levels of hyperrecombination observed in the HO orientation is suppressed by RNase H. Consistent with the bacterial/human cells studies, this suggests that in this orientation, R loops are a source of the increased genome instability. Meanwhile, no suppression of the slightly elevated CD levels was observed. Interestingly however, overexpression of Yra1 (which binds to transient R loops, resulting in their stabilisation) resulted in greatly increased levels of recombination in both the HO and CD constructs, suppressed by RNase H. This conversely suggests that R loops actually form independently from the orientation of collisions. However, in the CD orientation, they occur on the leading strand and are likely cleared by some replisome-associated helicase function, thus do not pose a threat to genome stability unless stabilised. On the other hand, in the head on orientation, RNAPII could be stabilised by the R loop and consequently cannot backtrack or be removed by the replisome, causing fork stalling (Garcia-Rubio et al., 2018).

Recently, R-loop-induced DNA damage in yeast was mapped genome-wide using ChIP of the DNA-repair protein Rad52 in strains that accumulate persistent R loops. Lethal R loop-associated DNA damage events were determined as follows: the Rad52-ChIP signature in cells depleted of the helicase Sen1 (using an auxin degron; *sen1-AID)*, or *rnh1*Δ *rnh201*Δ cells, was compared with the triple *sen1-AID rnh1*Δ *rnh201*Δ mutant to identify new Rad52 accumulation surrounding R loop forming regions. The Rad52 occupancy in the triple mutant (where both R loops removal pathways are removed) was unique only at a subset of R loops, suggesting only a fraction induce lethal, unrepairable DNA damage. This suggests that the local chromosomal context may also influence whether an R loop is deleterious. Comparing the observed signature with a lethal R loops accumulating strain lacking Top1 and RNase H showed that Sen1 and Top1 act at distinct genomic loci to prevent lethal R loop mediated damage. Interestingly, in the Sen1 triple mutant, the Rad52 binding signature indicated that formation of DSBs occurs next to the 3' end of the R loop. This pattern is consistent with the model that head-on collisions occur when forks approach the 3' end of an R loop (fig 1.8). Moreover, much of the lethal R loop-induced damage was located within regions where forks originating from the closest origin would clash with the R loop in a head-on orientation. Finally, at each loci, Rad52 was only present on a single strand of the DNA double helix, which provides a possible mechanistic insight for how R loops can induce genome instability. The presence of the R loop

may block the resection of one end of the DSB (which occurs as a result of the TRC), while the other is resected generating a stretch of ssDNA. The authors postulate that this could result in extensive resection in unique regions, or in repetitive regions invasion of a homologous sequence, both of which can result in gross chromosomal rearrangements (Costantino and Koshland, 2018). Interestingly, R loops have been shown to form across a number of common fragile sites within the human genome (which coincide with highly transcribed genes). These regions are prone to double strand breaks, leading to chromosomal translocations, which are commonly observed in cancer cells. Thus, conflicts between transcription and replication within these regions could stabilise any R loops, resulting in fork stalling, double strand breaks and deleterious recombination events (Helmrich et al., 2011).

Importantly, further evidence supports the idea that local chromatin environment may affect whether an R loop induces lethal damage (Costantino and Koshland, 2018, García-Pichardo et al., 2017). A study has shown that R loop formation alone is not sufficient to compromise genome stability; a subsequent step that results in remodelled chromatin is also required. Unlike the R loop accumulating *hpr1Δ* and *sen1-1* strains, which present high levels of recombination and genetic instability, histone mutants with increased levels of R loops do not show this hyperrecombination phenotype. The altered chromatin in these histone mutants facilitates the formation of R loops, however the uncoupling of their formation from the induction of genome instability was linked to their reduced levels of histone H3 serine-10 phosphorylation, compared with *hpr1Δ* and *sen1-1* mutants. Combining *hpr1Δ* or *sen1-1* with these histone mutants suppressed their hyperrecombination phenotypes. Thus, for R loops to induce genome instability, a second chromatin-modifying step is necessary that includes but may not be limited to phosphorylation of H3S10. The authors suggest that as H3 phosphorylation is linked to chromatin condensation, this environment may act as a barrier to fork progression, causing them to stall (García-Pichardo et al., 2017).

1.8 Mechanisms for the prevention and removal of R loops

Regardless of the precise mechanism leading to R loop-dependent genome instability, it is clear that if their levels are unchecked, the consequences would be catastrophic. Accordingly, cells have evolved several mechanisms to both prevent the formation of

and resolve deleterious R loops (fig 1.10), which will be described in the following section. In brief, these processes ensure prompt maturation/export of nascent RNA, processive RNA polymerase activity, digestion of the RNA within the hybrids or their clearance by helicases.

1.8.1 Mechanisms preventing the formation of R loops

Efficient mRNP biogenesis: THO/TREX

One of the major pathways for the prevention of R loops involves the efficient packaging and processing of the nascent RNA for export to the cytoplasm, which helps to prevent it from re-annealing to the template DNA strand. In yeast this is carried out by THO/TREX; where during elongation, this complex couples the formation of messenger ribonucleoproteins (mRNPs) with their export through the nuclear pore. The yeast THO complex is comprised of four subunits: Hpr1, Tho2, Mft1 and Thp2, and is involved in mRNP biogenesis, packaging the nascent RNA. THO physically interacts with the transcription-export complex (TREX; comprised of Tex1, and nuclear export factors Sub2 and Yra1) through an interaction between Hpr1 and Sub2.

THO mutants are characterised by a transcription associated-hyperrecombination phenotype (first identified in in Hpr1 and Tho2 mutants) and show defects in elongation and mRNA export, particularly at genes with long or GC rich templates (Chávez and Aguilera, 1997, Chávez et al., 2000). Importantly, the hyperrecombination phenotype of *hpr1Δ* cells was linked to the nascent mRNA and formation of co-transcriptional R loops, where the high level of recombination was suppressed by overexpression of *RNH1* (Huertas and Aguilera, 2003). The authors propose that the elongation defect can be explained by R loops impeding the progression of the next RNAPII complex; and blockage of replication forks by R loops, as well as the unstable exposed ssDNA within these structures sensitises the cells to recombinogenic DNA breaks. This mRNP biogenesis-associated protection against R loops is conserved from yeast to humans. In human cells, depletion of THO has also been shown to induce defects in elongation and mRNA export, as well as an R loop-dependent increase in DSBs and recombination. The increased levels of recombination in these THO mutants was suppressed by *RNH1* overexpression, and

exacerbated by overexpression of activation-induced cytidine deaminase which deaminates ssDNA within R loops, generating DSBs) (Dominguez-Sanchez et al., 2011).

Splicing, export and degradation factors

Another study in yeast explored multiple factors involved in other areas of mRNP biogenesis in addition to the previously described elongation factors (THO). Interestingly, a subset of genes involved in diverse aspects of mRNA biogenesis, including transcription repression, initiation, elongation, RNA export and degradation also results in the formation of deleterious RNase H sensitive R loops that lead to increased genetic instability (Wahba et al., 2011). In both chicken DT40 and human HeLa cells, a hypermutagenic phenotype is observed in the absence factors such as *ASF/SF2,* that are involved in mRNA splicing or spliceosome assembly. Crucially, ASF/SF2 and other splicing factors have also been shown to function as adaptors for mRNA export. Disruption of these genes results in a significant increase in recombinogenic DSBs, which importantly can be suppressed by RNase H (Li and Manley, 2005, Paulsen et al., 2009).

Both Rrp6 and Trf4 mutants were among the factors identified in yeast as driving R loop-dependent genome instability (Wahba et al., 2011). Rrp6 is a component of the exosome, and Trf4 is a subunit of its co-factor TRAMP. A subsequent study also showed that in addition to a transcription-associated hyperrecombination phenotype, *trf4Δ* cells have an increased frequency of mutations. Importantly, overexpression of RNase H1 and overexpression of activation-induced cytidine deaminase suppressed and exacerbated these phenotypes, respectively. This suggests that mRNA surveillance and metabolism by the nuclear exosome may help maintain genomic instability by preventing R loops accumulation (Gavalda et al., 2013). Interestingly, the exosome also appears to be important in human cells at sites of TRCs. SUMOylated Senataxin (the human orthologue of the budding yeast helicase Sen1) associates with the Rrp45 subunit of the exosome, where they co-localise to form nuclear foci following transcription-related DNA damage. Interestingly, Senataxin has been shown to metabolise R loops at sites of TRCs. Considering the transcription-associated hyperrecombination phenotypes in yeast exosome mutants, the authors

speculate one possibility is that the exosome plays a role together with Senataxin to resolve R loops at these sites. Alternatively, the exosome could have a currently unidentified role in a DNA repair process (Richard et al., 2013, Yüce and West, 2013).

Topoisomerase I

Another pathway that protects cells from the formation of deleterious R loops is by controlling the accumulation of negative supercoiling behind the elongating RNA polymerase. Negative supercoiling increases the chance of R loop formation; as the underwound DNA results in transient strand separation, increasing the opportunity for the nascent RNA to invade the template (Roy et al., 2010). DNA topoisomerases are enzymes that have the ability to relax supercoiled DNA and relieve the torsional strain by cleaving and re-joining DNA strands. Their role in the prevention of R loop formation has been demonstrated in multiple species, including bacteria (Drolet et al., 1995), yeast (El Hage et al., 2010) and humans (Tuduri et al., 2009). In *S. cerevisiae* cells, in the absence of either *TOP1* or *TOP2*, R loops accumulate across Pol I transcribed rDNA, and interference with Pol I elongation is observed in strains lacking both of these enzymes (El Hage et al., 2010). Concurrently, it has been shown that the rate of recombination and replication fork stalling is increased in strains with hybrids accumulating across rDNA (Amon and Koshland, 2016, Wahba et al., 2011, Wahba et al., 2013, Stuckey et al., 2015). In human cells, in the absence of Top1, gene rich regions accumulate stalled forks and DNA breaks during S phase in an R loops dependent manner (Tuduri et al., 2009).

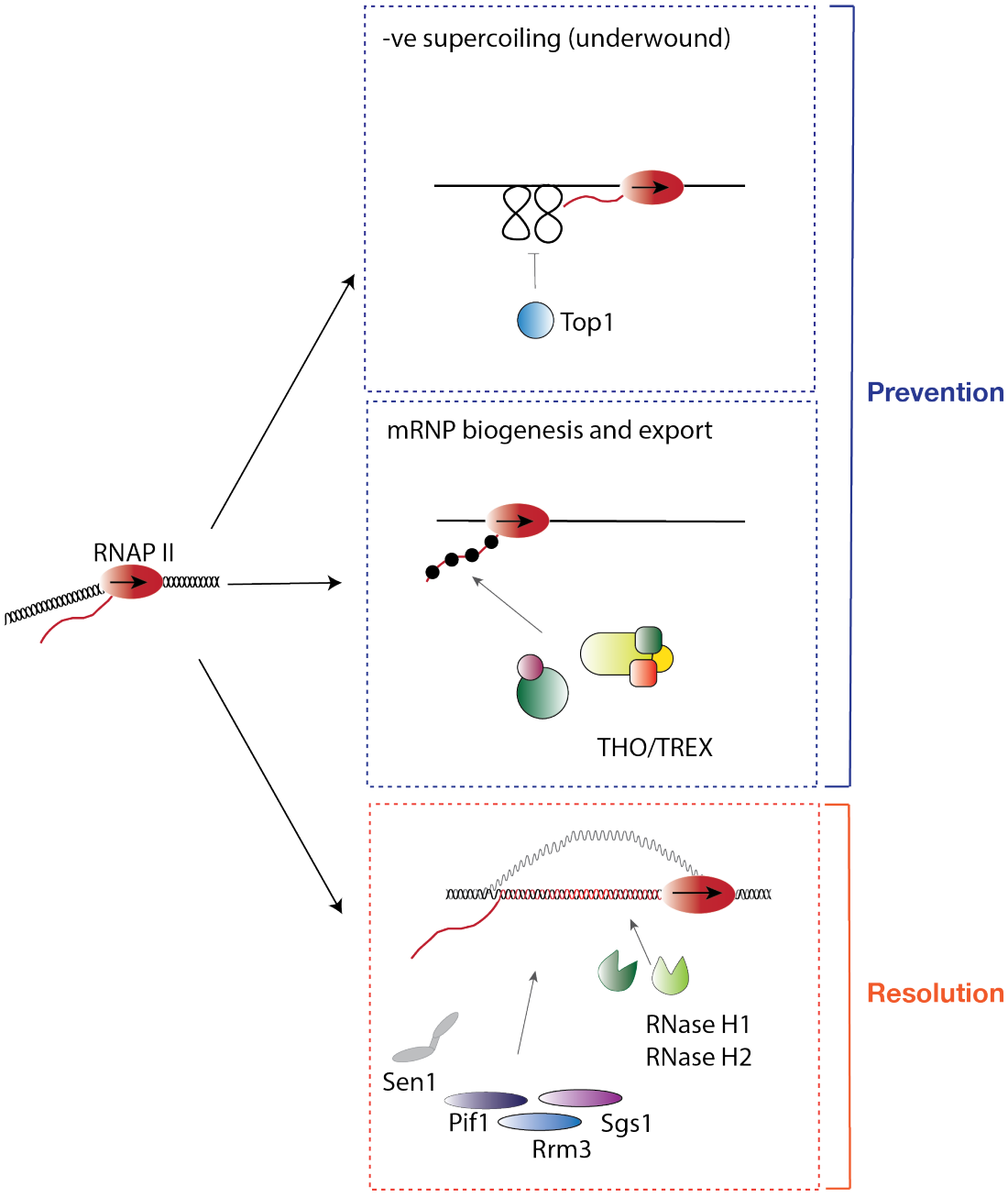


Figure 1.10 Mechanisms for the prevention or removal of R loops. Organisms have evolved several mechanisms that help prevent the formation of R loops. These include the resolution of negative suporcoiling behind the elongating polymerase by topoisomerases; and promoting the timely processing and nuclear export of nascent mRNA, to prevent it from reannealing to the template DNA strand. Once formed, the RNAse H enzymes are key players in resolving stable R loops by cleaving the RNA portion of the DNA:RNA hybrid. Other helicases, including Sen1 which possesses 5'-3' DNA:RNA unwinding activity, are also involved in clearing the hybrids from the genome.

1.8.2 Digestion of R loops by the RNase H enzymes

In addition to minimising the formation of R loops, several mechanisms in cells actively remove stable R loops from DNA. The evolutionarily conserved Ribonuclease H (RNase H) enzymes play a key role in R loops removal, whereby they recognise and endonucleolytically cleave the RNA portion within DNA:RNA hybrids. The RNase H enzymes can be separated into two classes, RNase H1 and RNase H2. The two classes are divided by their structural features. In addition, they have different, though overlapping, substrate specificities. Indeed, the levels of genome instability exhibited by either single mutant is exacerbated in the double mutant, suggesting some degree of functional redundancy (Wahba et al., 2011). While these enzymes are not required for viability in prokaryotes or yeast (though their absence results in increased sensitivity to DNA damaging agents), they are essential in higher eukaryotes (Cerritelli and Crouch, 2009).

In eukaryotes, RNase H1 presents an N-terminal hybrid binding domain (HBD), a central connection domain (CD), and a C-terminal RNase H domain (HD). In higher eukaryotes, a mitochondrial targeting sequence is also present at the N-terminus, which; for example, reflects its essential role in mitochondrial DNA replication during development. Both the N-terminal HBD and C-terminal HD are highly conserved among eukaryotes, however the middle CD has a more variable sequence and is thought to provide flexibility. The HBD shows a much greater preference for RNA:DNA duplexes over dsRNA and binds both the RNA and DNA strand. Its engagement with the hybrid is thought to enhance the catalytic activity of the protein. For cleavage of the RNA to occur, the HD (which contains the catalytic site) requires a substrate where it can interact with the 2' OH group of a minimum of four consecutive ribonucleotides (reviewed (Cerritelli and Crouch, 2009), see references therein). The ssDNA within an R loop is coated with RPA, which likely recruit's RNase H1 to these sites via a direct interaction (Nguyen et al., 2017). The RNase H1 enzyme is mainly responsible for the removal of transcription-associated R loops. In the context of backgrounds with high levels of R loops due to mutation of mRNA biogenesis factors, the absence of *RNH1* is more deleterious than that of *RNH2* (Wahba et al., 2011). Phenotypic suppression by the more well-defined RNase H1 is widely used as an experimental marker for R loops, and its overexpression has been

shown to suppress defects in both *rnh1Δ* and *rnh201Δ* cells (Li and Manley, 2005, Wahba et al., 2011).

RNase H2, on the other hand, is a multi-subunit enzyme comprised in budding yeast of Rnh202, Rnh203 and the catalytic subunit Rnh201, all of which are required for its *in vivo* activity (Cerritelli and Crouch, 2009). The neuro-inflammatory disease, Aicardi-Goutières syndrome as well as certain cancers have been linked to mutations within these subunits (Crow et al., 2006, Zimmermann et al., 2018). Importantly, unlike RNase H1, RNase H2 has the ability to recognise and cleave a single ribonucleotide within dsDNA. Thus, it is involved in the removal of ribonucleotides mis-incorporated into DNA (rNMPs) during replication (Eder et al., 1993). If RNase H2 is defective for this ribonucleotide excision repair, aberrant processing of these lesions by Top1 can result in short 2-5 bp deletions and double strand breaks, contributing to genome instability (Kim et al., 2011). It has also been demonstrated as providing an alternative pathway for the removal of RNA primers for the maturation of Okazaki fragments, though this is mainly carried out by Rad27/Fen1 and Dna2 (Qiu et al., 1999).

RNase H2 can also degrade the RNA within R loop structures, and a recent study examining the regulation of RNase H1 and 2 aimed to tease apart why two separate RNases have evolved for R loops removal (Lockhart et al., 2019). Interestingly, immunoprecipitation of TAP tagged versions of the endogenous proteins found that RNase H1 is expressed consistently throughout the cell cycle and remains weakly but constantly associated to chromatin throughout. On the other hand, the levels of Rnh201 (the catalytic subunit of RNase H2) increases during late S phase/G2, and concurrently the full heterotrimeric complex accumulates on chromatin. Consequently, the *RNH1* and *RNH202* alleles were placed under the control of cell cycle regulated promoters and their degrons to specifically limit their activity to during either S phase (using the Clb6 S phase cyclin promoter/degron) or G2/M (using the promoter/degron of the Clb2 cyclin, expressed during G2/M). Interestingly, the synthetic lethality of the triple mutant *rnh1Δ rnh201Δ sen1-1* (which accumulates high levels of toxic R loops (Mischo et al., 2011)) is rescued when Rnh202 is expressed during G2, but not S phase. This suggests that RNase H2 activity is crucial for post-replicative repair of toxic R loops. In addition, its expression during S phase was in

fact found to be detrimental to genome stability in a background with high levels of rNMP misincorporation. This is likely because during S phase, any nicks resulting from ribonucleotide removal will be encountered by the replisome and converted to recombinogenic double strand breaks. On the other hand, both S and G2 restricted expression of RNase H1 was able to restore the viability of the *sen1-1 rnh1Δ rnh201Δ* triple mutant. Importantly, in backgrounds of R-loop induced stress, RNase H1 is further enriched on chromatin, independently from the cell cycle (Lockhart et al., 2019). These results posit RNase H1 as a good responder to R-loop induced replication stress, as it does not have a role in ribonucleotide excision repair, so cannot induce double strand breaks during S phase. RNase H2 on the other hand is more of a "house-keeping" enzyme, that under normal circumstances deals with the majority of R loops and mis-incorporated ribonucleotides during G2, after DNA synthesis is complete (Lockhart et al., 2019).

1.8.3 Clearance of R loops by helicases

There are also certain helicases that have been implicated in the removal of R loops. In *S. cerevisiae,* these include Sen1, Pif1, Sgs1 and Rrm3, some of which are thought to process R loops at specific sites on the genome.

The majority of eukaryotes, including humans, encode a DNA helicase from the Pif1 family. Indeed, *S. cerevisiae* possesses two members, Pif1 and Rrm3. Rrm3 travels with replication forks and is thought to help forks progress past stable-protein blocks such as Fob1 and tRNA genes (Foury and Kolodynski, 1983, Ivessa et al., 2000, Ivessa et al., 2002). Pif1 is involved in various other processes such as maintenance of mitochondrial DNA, regulation of telomere length and resolution of G-quadruplex secondary structures (reviewed (Geronimo and Zakian, 2016), see references therein). Its homolog in *Sz. pombe* (Pfh1), like Rrm3, has been proposed to travel with replication forks, interacting with several replisome components (McDonald et al., 2016). Importantly, one of the overlapping functions of Pif1 and Rrm3 appears to be contributing to the removal of R loops across tRNA genes, for which they show some redundancy. The absence of these helicases results in increased DNA damage at tRNA genes in an R loops dependent manner. Stabilisation of R loops by deletion of *RNH1* increases the association of these helicases to tRNA genes, and overexpression of

RNase H1 in their absence suppresses the increased levels of DNA damage (Tran et al., 2017).

The RECQ-like helicase Sgs1 has a well-documented role in homologous recombination, both promoting the resection of DSB ends and the resolution of recombination intermediates (Zhu et al., 2008, Cejka and Kowalczykowski, 2010). In addition, evidence suggests that it also plays a role in protecting cells from R-loop mediated genome instability. An increase in the signal for R loops was detected in chromosome spreads of *sgs1Δ* cells (stained using the S9.6 antibody targeted against RNA:DNA hybrids), and these cells become sensitised to head-on collisions between transcription and replication complexes. ChIP-chip and DRIP-chip experiments found that loss of Sgs1 results in increased R loops and activation of the DNA damage response, particularly at longer genes compared with wild type cells (Chang et al., 2017).

Finally, the DNA:RNA helicase Sen1 plays a key role in the removal of R loops. The following chapter will first discuss in detail the structure of Sen1, followed by its role in the termination of RNAPII transcription, the resolution of R loops and in transcription-coupled DNA repair. Critically for this thesis, I will also discuss the role of Sen1 during DNA replication.

1.9 The Structure and function of the DNA:RNA helicase Sen1

1.9.1 The structure of Sen1

Sen1 (Splicing endonuclease 1) is a 250 kDa protein encoded by the essential *SEN1* gene in *S. cerevisiae*. It was first identified in a screen for mutants with altered tRNA splicing activity (Winey and Culbertson, 1988, DeMarini et al., 1992). However, later evidence suggests that Sen1 plays a number of roles within the cell, including the termination of short RNAPII transcripts, removal of deleterious R loops and transcription-coupled repair, all of which will be discussed in this chapter.

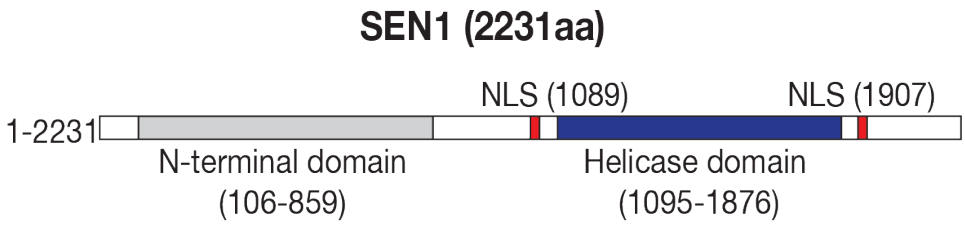


Figure 1.11 Schematic of the domain structure of the Sen1 helicase. Sen1 is comprised of an extended N-terminal domain and a conserved helicase domain. It is targeted to the nucleus via two nuclear localisation signals.

Sen1 is a member of the Upf1-like 1B superfamily of helicases (Jankowsky, 2011). It is localised to the nucleus via two nuclear localisation signals, and is comprised of an N-terminal domain, which is predicted to encode armadillo repeat motifs that fold to form a super-helix of alpha helices, and a conserved C-terminal helicase domain (fig 1.11).

The essential region for viability was mapped to the C-terminal domain of the protein (residues 1089-1907) encoding the helicase motif. *In vitro,* both full length Sen1 and purified Sen1-helicase domain have been shown to possess ATP dependent 5'-3' helicase activity. Nucleic acid duplex unwinding assays revealed it is able to unwind both DNA:DNA and RNA:DNA duplexes, provided they contain a region of single stranded overhang at their 5' end. Once engaged with the substrate, Sen1 shows poor processivity as a translocase, and *in vitro* dissociates after approximately 40

nucleotides (Han et al., 2017). Compared to the helicase domain alone, the full-length protein unwinds RNA:DNA duplexes with much greater efficiency, suggesting other regions are required for its optimum activity on this substrate. Interestingly, both constructs unwound DNA:DNA with much greater efficiency than RNA:DNA duplexes (Han et al., 2017). While Sen1 binds ssDNA and ssRNA with a similar affinity, it is more processive on ssDNA, which correlates with the greater helicase activity. In general however, data indicates that even on ssDNA, Sen1 is a poorly processive translocase (Martin-Tumasz and Brow, 2015, Han et al., 2017).

The crystal structure of the Sen1 helicase domain has been solved. Along with the helicases Upf1 and IGHMBP2, Sen1 is a member of the superfamily-1B Upf1-like family. All three of these helicases have a similar motif structure within their helicase domain, however Sen1 was found to have an additional, unique feature (Leonaitė et al., 2017). Members of this family possess two classical RecA domains (RecA1 and RecA2) separated by a short linker. These make up the core helicase domain, and contain motifs responsible for its ATPase activity, as well as nucleic acid binding. An additional two accessory subdomains (1B and 1C) are responsible for regulating the binding of the nucleic acids. In terms of polarity, the 3' end and 5' end of bound nucleic acids are located at RecA1 and RecA2 respectively. Subdomain 1B (referred to as a "stalk" and flexible "barrel") and 1C (which adopts a rigid "prong" structure) are thought to play a role in melting the duplex DNA. These are located on top of the RecA1 domain near the 3' end of the bound nucleic acid where it enters the helicase channel (fig 1.12) (Ozgur et al., 2015, Cheng et al., 2007, Lim et al., 2012, Chakrabarti et al., 2011, Leonaitė et al., 2017). They are thought to adopt a conformation which surrounds the nucleic acid and melts the incoming bases apart via insertion of the rigid prong into the duplex. In Sen1, elimination of the rigid prong was shown to prevent duplex unwinding. The structural feature unique to Sen1 is encoded by highly conserved residues which the authors termed a "brace". This brace appears to hold the barrel, stalk and RecA1 domains together in a rigid structure that shapes and fixes the conformation of the helicase for optimal nucleic acid binding and prong insertion for duplex melting (fig 1.12) (Leonaitė et al., 2017).

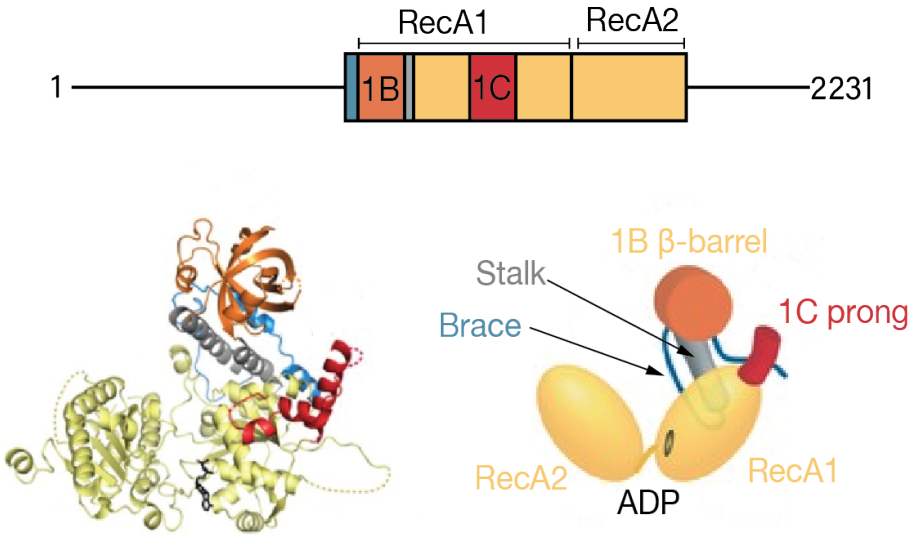


Figure 1.12 The architecture of Sen1 helicase domain. The helicase domain of Sen1 is comprised of the RecA1 and A2 domains with the accessory 1B and 1C subdomains. The 3D organisation of Sen1 helicase domain as determined by crystallography is shown. This figure is adapted from (Leonaitė et al., 2017).

While the deletion of the N-terminal domain (residues 1-975) results in a slow growth phenotype, the cells remain viable (Ursic et al., 2004). The N-terminal is believed to be responsible for mediating important protein-protein interactions, including those with RNAPII, Rnt1 and Rad2 (which are discussed in more detail later).

In general, Sen1 is maintained at relatively low levels within the cell (~125 molecules/cell (Ghaemmaghami et al., 2003)) by the ubiquitin-proteasome system, likely to avoid any toxicity that results from its excessive activity. However, its levels do vary throughout the cell cycle. Although the level of mRNA remains constant throughout, synchronous release of cells from G1 arrest and immunoblotting for Myc-tagged Sen1 found that Sen1 protein is somewhat reduced in G1 and increases towards S/G2. This fluctuation may reflect the cells changing needs; for example, higher levels in S phase may be required as during this period, the transcription machinery may encounter replication forks and require removal (Mischo et al., 2018).

1.9.2 The role of Sen1 in the termination of transcription

Transcription is the process where DNA-associated RNA polymerases produce exact RNA copies of the template DNA. Termination events result in removal of the RNA polymerase from the DNA template and release of the nascent transcript. In eukaryotes, three different RNA polymerases are responsible for synthesising many different RNA species at defined genomic loci (table 1.2, reviewed (Richard and Manley, 2009)). The various species are terminated via specific pathways, which influences their fate and specificity (Porrua and Libri, 2015).

Table 1.2 A brief overview of the three eukaryotic RNA polymerases and the RNA species that they transcribe

Enzyme	Transcript species	Role of transcript species
RNAPI	rRNA	Non-coding rRNA that is bound to ribosomal proteins to assemble the ribosome for protein synthesis.
RNAPII	mRNA	Coding RNA that is transported out of the nucleus and translated into protein.
	snoRNA	Non-coding RNAs that assist the chemical modification of other RNA species.
	snRNA	Non-coding RNAs that assist in splicing
	snRNA	Non-coding RNAs that are involved in post-transcriptionally regulating the expression of certain genes and in gene silencing.
	miRNA	Highly unstable non-coding RNAs that result from pervasive transcription. Their function is largely unknown.
	lncRNA	Non-coding RNAs involved in the regulation of gene expression.
RNAPIII	5S rRNA	Non-coding RNA that is a structural/functional component of the large ribosomal subunit.
	tRNA	Non-coding RNA that recognises codons to deliver the correct amino acid to mRNA for protein synthesis.

Coding regions of the genome are transcribed into pre-mRNA, which is exported to the cytoplasm for translation into protein. Other RNA species however are not translated; stable non-coding RNAs (which are usually retained in the nucleus) have roles in RNA processing and the control of gene expression (see table 1.2). Interestingly, it has recently been shown that transcription is far more widespread and promiscuous than previously thought, now referred to as 'pervasive transcription'. The transcription of non-coding intergenic regions and in antisense to genes; initiated either from promoter regions or from nucleosome depleted regions, results in production of a class of highly unstable RNAs dubbed cryptic unstable transcripts

(CUTs) (reviewed (Porrua and Libri, 2015)). In particular, the bi-directional nature of transcription promoters, though not symmetrical (as local chromatin environment and histone modifications may influence their specificity toward productive transcription), is a considerable contributor to the production of CUTs (Marquardt et al., 2014, Whitehouse et al., 2007, Tan-Wong et al., 2012). Pervasive transcription needs to be controlled, as it may interfere with productive transcription. In yeast, a divergently initiated CUT may silence an upstream gene as a result of readthrough transcription at its 3' end (Neil et al., 2009, Schulz et al., 2013, Xu et al., 2009).

Under normal conditions, CUTs are rapidly terminated and degraded by the nuclear exosome. CUTs become more stable in exosome mutants; observed as short transcripts with heterogenous 3' ends. Analysis of a model CUT found that short RNA elements, comprising Nrd1 and Nab3 binding sites, confer instability to the transcribed sequence. These proteins are part of the non-canonical transcription termination complex (NNS; Nrd1-Nab3-Sen1). This pathway terminates transcription of the CUT and facilitates its degradation by nuclear exosome. Importantly, Nrd1 and Nab3 binding sites are ubiquitous throughout the genome, thus ensuring rapid termination and exosomal degradation of non-productive and potentially deleterious transcription. Readthrough of these Nrd1/Nab3 sites results in termination of CUT transcripts by the canonical pathway at a downstream region. This can result in production of a stable longer RNA, thus having a significant effect on the fate of gene transcription (Thiebaut et al., 2006).

In *S. cerevisiae*, two pathways exist to terminate RNAPII transcription (fig 1.13, 1.14). The canonical pathway, which terminates the transcription of genes that code for mRNA, will be briefly described in the next section. This mechanism depends on the cleavage and polyadenylation factor but may also possibly involve Sen1 in some cases. The major focus of this chapter however will be on the non-canonical pathway, which depends on the NNS complex, comprised of Nrd1, Nab3 and Sen1. This pathway is involved in the termination of CUTs and their targeting for exosomal degradation; thus, it helps to control the levels of pervasive transcription within the cell. It is also involved in the biogenesis of short non-coding RNA species, including small nuclear and small nucleolar RNAs important for splicing and modification of rRNA (Porrua and Libri, 2013)

Termination of protein-coding genes: The canonical transcription termination pathway in eukaryotes

In eukaryotes, the CPF-CF (cleavage and polyadenylation factor-cleavage factor) pathway is responsible for terminating protein-coding genes, where a poly(A) signal (PAS) acts as both a 3' end processing and transcription termination signal (fig 1.13). The PAS is found downstream of open reading frames in the 3' untranslated region of a gene. The conserved cleavage and polyadenylation factor (CPF) together with the cleavage factors 1A and 1B (CF) make up the CPF-CF complex, which is recruited by recognition of the PAS on nascent RNA following its transcription. In addition, an interaction between the Pcf11 subunit of CPF-CF and Ser2-phosphorylated (Ser2-P) RNAPII helps this recruitment (Richard and Manley, 2009, Mischo and Proudfoot, 2013, Porrua and Libri, 2015). Importantly, the largest subunit of RNAPII, Rpo21, contains a protruding, flexible C-terminal heptapeptide repeat domain (consensus Tyr_1-Ser_2-Pro_3-Thr_4-Ser_5-Pro_6-Ser_7; 26 repeats in yeast and 52 repeats in humans) that is differentially phosphorylated/dephosphorylated by various kinases throughout the transcription cycle (Egloff and Murphy, 2008). Its phosphorylation pattern affects which processing and termination factors bind to the polymerase. Phosphorylation of Ser5 dominates at the 5' end of genes, whereas phosphorylation of Tyr1 and Ser2 increases as the transcription machinery elongates towards the 3' end. At the PAS, the levels of phosphorylated Tyr1 decline abruptly, allowing recruitment of the CPF-CF and Rtt103 and the binding to Ser2-P (Phatnani and Greenleaf, 2006, Lunde et al., 2010). Endonucleolytic cleavage of the nascent RNA occurs at the PAS, and polyadenylation of the cleaved 3' end by Pap1 serves both to promote export of the mRNA for translation, and protect it against degradation (Porrua and Libri, 2015).

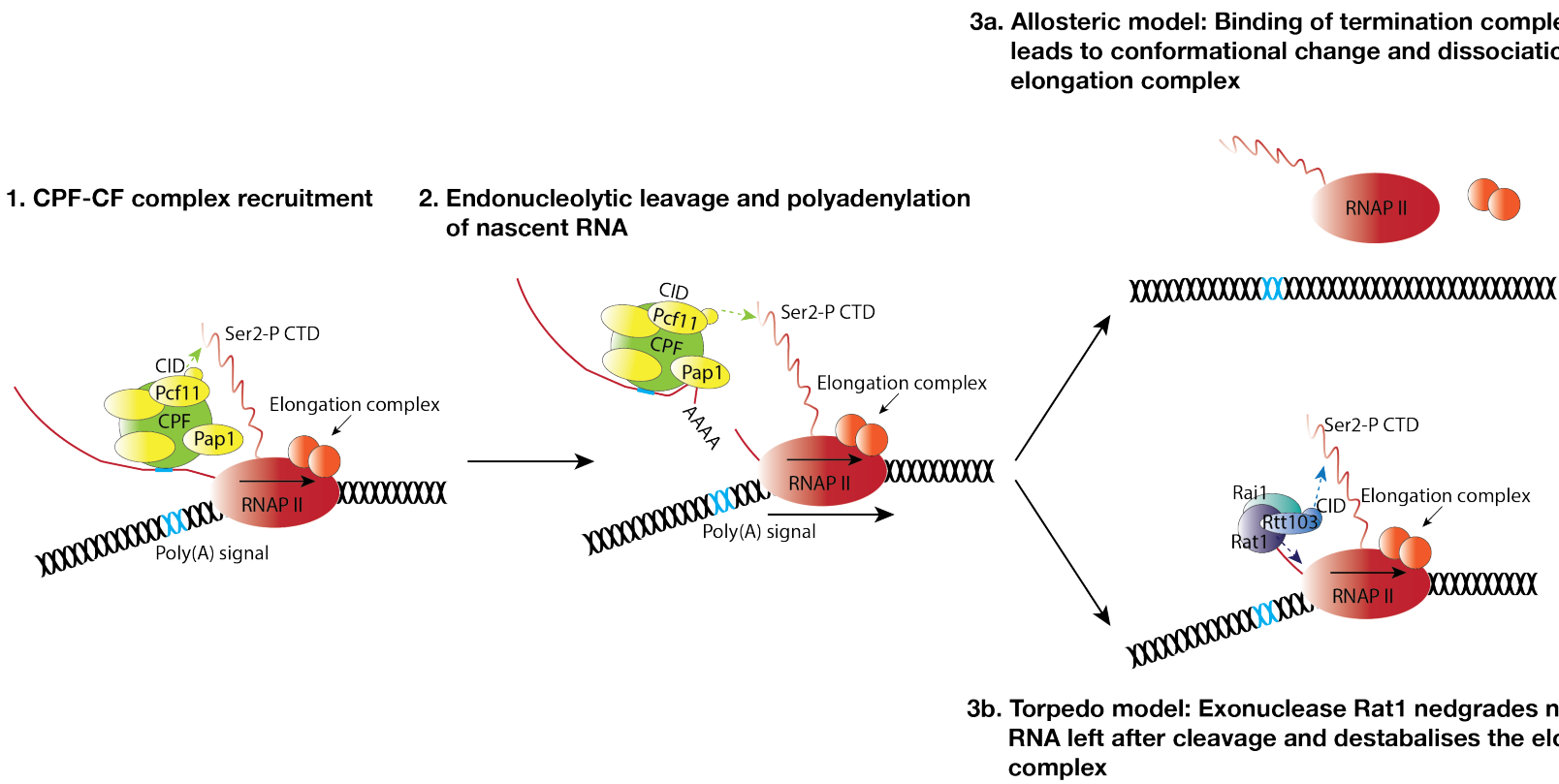


Figure 1.13 The canonical pathway for RNAPII transcription termination. The canonical termination pathway mainly acts at protein coding genes, where a polyadenylation signal is recognised by CPF-CF. Following endonucleolytic cleavage and polyadenylation of the nascent transcript, RNAPII dissociates from the template DNA. The allosteric model suggests this is due to a conformational change within the enzyme that occurs as a result of PAS transcription, whereas the alternative model suggests exonucleolytic degradation of the polymerase associated 3' transcript by Rat1 'torpedoes' the polymerase off of the DNA.

Following mRNA cleavage, the polymerase is thought to continue transcribing up to 150 nucleotides before dissociating. Currently, there are two different models that have been proposed to elicit canonical RNAPII termination. The allosteric model suggests that commitment to termination is a result of a conformational change or loss of elongation factors following transcription of the PAS. These could possibly be a result of either stochastic pausing of RNAPII or the recruitment of the CPF-CF. In addition, R loops have been shown to accumulate at terminator regions, downstream of the PAS. These structures may also contribute to the pausing of RNAPII for termination (Proudfoot, 2016). The torpedo model proposes the involvement of the 5'-3' exonuclease Rat1 (the yeast homolog of human Xrn2), which is recruited in a complex with Rai1 to the 3' end of genes by Rtt103. Following cleavage of the nascent mRNA, the exposed 5' end of the cleaved transcription product still associated with the polymerase provides access for Rat1 to degrade the nascent RNA. Upon catching the polymerase, it is suggested that Rat1 induces termination, perhaps by eliciting a conformational change. In fact, disruption of Rat1 (using the temperature sensitive *rat1-1* allele) leads to termination defects and stabilisation of mRNA downstream of the PAS (Kim et al., 2004, West et al., 2004).

Interestingly it has been shown across two protein-coding genes that the helicase activity of Sen1 is also important for efficient termination following transcription of a PAS. The termination defects of *rat1-1* cells were further exacerbated by *sen1-1*, which raised the possibility that these two proteins may co-operate (Kawauchi et al., 2008). Following cleavage at the PAS, Sen1 could expose the polymerase associated transcript to allow Rat1 access, by unwinding secondary RNA structures, such as R loops, that may prevent Rat1 from accessing the RNA to carry out its function (Kawauchi et al., 2008). Interestingly, the same study showed that Sen1 and Rat1 activity are also important for efficient termination of rDNA transcribed by RNAP1. Rat1 is able to elicit termination following cleavage of the rDNA transcript by Rnt1, which gives rise to the exposed 5' end of the transcript. Once again, Sen1 may help to promote Rat1 access, and indeed, Sen1 has been shown to interact with Rnt1 through its N-terminal domain (Kawauchi et al., 2008, Ursic et al., 2004).

Evidence for both models exists (Mischo and Proudfoot, 2013, Porrua and Libri, 2015), and they may not be mutually exclusive. It is possible that allosteric

termination can occur when transcription of PAS elicits some kind of conformational change that is sufficient to promote slowing and dissociation of the elongation complex. However, in some cases, if this is not sufficient, Rat1 may also be required to help dissociate the polymerase. In situations where the access of Rat1 to the 5' end of the RNA is prevented, Sen1 helicase activity may then also be necessary.

Termination of non-coding RNA: The non-canonical transcription termination pathway in eukaryotes

As mentioned previously, not all RNA products are transcribed into protein, but instead can play important functions within the cell (table 1.2). Also, more promiscuous pervasive transcription gives rise to short lived, non-coding RNAs (CUTs) that may have no productive role (Porrua and Libri, 2015). Indeed, this widespread transcription can pose a threat to genome stability, as it may overlap and interfere with coding or functional transcription units, both in the sense and antisense orientation. Thus, these RNAs need to be terminated and rapidly degraded to prevent their toxic accumulation. In *S. cerevisiae,* termination of these short non-coding RNA species (typically less than 500 base pairs long) depends on the NNS (Nrd1-Nab3-Sen1) complex, which is linked to the nuclear exosome; either for their processing or complete degradation, thus determining their fate (fig 1.14).

The NNS components Nrd1 and Nab3 are both essential for viability in *S. cerevisiae.* They are RNA binding proteins that exist as heterodimers, and recognise specific sequences on the nascent transcript (termed RNA recognition motifs; GUAA/G and UCUU respectively) (Carroll et al., 2007, Conrad et al., 2000). The binding of both Nrd1 and Nab3 within the heterodimer is required for optimal mRNA association (Carroll et al., 2007). These motifs are enriched at sn/snoRNA terminators, and in some cases, multiple copies are present successively. Their recognition by Nrd1-Nab3 heterodimers has been linked to the specificity of the NNS pathway for short RNA species. Mutations within these sequences reduce the affinity of the heterodimer for RNA and results in readthrough termination of the *SNR13* reporter gene (Carroll et al., 2004, Carroll et al., 2007). Furthermore, while the CPF-CF complex interacts with the Ser2-P form of RNAPII CTD, Nrd1 has been shown to bind preferentially to the Ser5-P CTD (Vasiljeva et al., 2008). This phosphorylation state is a marker for early,

promoter-proximal elongation at the 5' end of genes, and the interaction with Nrd1 likely contributes to the specificity of NNS termination at short transcription units. Interestingly, a CUT terminator has been identified that only contains binding sites for Nab3. However, interaction between Nrd1 and the CTD is still required for efficient termination, even though at this transcription unit, Nrd1 binding to RNA is dispensable (Carroll et al., 2007).

The final component of the NNS complex is Sen1, which is thought to promote the release of RNAPII and the nascent RNA from the DNA substrate (Porrua and Libri, 2013, Han et al., 2017). A study using the temperature sensitive *sen1-E1597K* mutant (substitution in the helicase domain) used ChIP-chip to map the occupancy of RNAPII genome-wide. Comparing the RNAPII profiles of wild type and mutant Sen1 cells, transcription readthrough was observed at the majority of small nuclear RNA (*SNR*) genes, as well as at a subgroup of protein coding genes when Sen1 activity was impaired. Interestingly, like *SNR* genes, the identified protein-coding genes were all short (ORFs between 205-554 base pairs) (Steinmetz et al., 2006). Another study used the *sen1-1* mutant, where transcription run on analysis revealed that at non-permissive temperature, the mutant shows strong termination defects that can be rescued by overexpression of the helicase domain. In addition, at the strong *CYC1* polyA terminator, though *sen1-1* cells exhibited no defects in mRNA 3' end formation, stabilisation of the 3' end cleavage product was observed, suggesting Sen1 is directly involved in removal of the polymerase (Mischo et al., 2011).

The Libri lab developed an *in vitro* termination assay (IVTT) to assess the contribution of Sen1 in the removal of RNAPII. In this assay, a biotinylated DNA template immobilized on streptavidin beads was used, on which purified RNAPII was assembled using a short RNA primer (with a 5' label). In this system, polymerases engaged in transcription remain associated with the beads, whereas any RNAPII detected in the supernatant is a result of termination. Here, it was found that Sen1 alone is sufficient to elicit termination of naturally paused and roadblocked RNAPII, independently from Nrd1, Nab3 or the RNA sequence (Porrua and Libri, 2013). In fact, just the helicase domain is sufficient to elicit termination (Han et al., 2017). The ATPase activity of Sen1, which is strongly stimulated in the presence of nucleic acids, is crucial (Porrua and Libri, 2013). The *sen1-G1747D* mutant, which has no ATPase

activity, is defective for termination both *in vivo* and *in vitro* (Porrua and Libri, 2013, Kim et al., 2006).

Though Sen1 can translocate along both ssDNA and ssRNA, it does not appear to interact with the template DNA to elicit termination but requires at least 15 nucleotides of accessible nascent RNA upstream of the polymerase. Thus, it likely translocates along the nascent RNA toward the polymerase (Porrua and Libri, 2013, Han et al., 2017). The termination ability of Sen1 is species specific, as it cannot terminate transcription by *E. coli* RNAP. This suggests that specific contacts between Sen1/RNAPII are important.

Interestingly, it appears that Sen1 translocation is in direct kinetic competition with the rate of RNAPII elongation for termination. Transcription read through of various *SNR* reporter genes in the *sen1-1* and *Sen1-E1597K* catalytic mutant backgrounds was examined in combination with RNAPII mutants that exhibit either slow (*rpb1-N488D)* or fast *(rpb1-E1103G)* elongation rates. The *rpb1-E1103G*^{FAST} mutant demonstrates increased transcription readthrough at *SNR* genes, and this phenotype is aggravated in the Sen1 mutant backgrounds. This suggests mutations within the Sen1 helicase domain compromise it's activity by reducing the ability to translocate along the RNA and catch the polymerase to promote termination. Conversely, in the *rpb1-N488D*^{SLOW} mutant, termination appears to occur in a shorter window, and the slower rate can partially complement the termination defects of the Sen1 mutant backgrounds. In this case, a slower polymerase may allow a longer time-frame for the mutant Sen1 to catch-up and promote dissociation (Hazelbaker et al., 2013). Indeed, several studies observe that RNAPII pausing is critical for efficient termination (Han et al., 2017, Collin et al., 2019, Han et al., 2020).

Sen1 has been shown to interact with the Ser5-P form of the RNAPII CTD via its N-terminal domain (Han et al., 2020). Interestingly, in the *in vitro* system, the CTD of RNAPII was not a specific requirement for Sen1 termination, however the stalling of RNAPII was essential (Porrua and Libri, 2013). Conversely, *in vivo,* disruption of the interaction between Sen1 and the Ser5-P CTD by deletion of Sen1 N-terminal incites strong termination defects at Sen1 targets. Interestingly, the recruitment of Sen1 to these genes was not affected, suggesting that this interaction functions downstream to

help elicit termination. The authors propose that the interaction helps Sen1 in terms of the kinetic competition with RNAPII elongation. It appears that in the absence of the Sen1 CTD interaction, Sen1 inefficiently tracks the polymerase, as it accumulates downstream of the early terminator regions. Once again, reducing the rate of elongation suppressed the observed termination defects (Han et al., 2020). Concordant with this kinetic competition, a recent genome-wide study has shown that phosphorylation of RNAPII CTD at Tyr1 results in pausing of the polymerase at the 5' end of genes, a step which is required globally for efficient termination by the NNS pathway (Collin et al., 2019).

Many studies have favoured a model whereby Sen1 is recruited to NNS terminators by interactions with Nrd1 and Nab3 homodimers (Porrua and Libri, 2013). However, disruption of Sen1 Nrd1 interacting motifs (NIMs) does not impair its recruitment and only results in termination defects at a small subset of its target genes (Han et al., 2020). Thus, the physical interaction between Sen1 and Nrd1-Nab3 does not appear to be a global requirement for efficient termination. However, the presence of Nrd1 and Nab3 at NNS termination sites is required for efficient termination. Thus, these proteins must confer some role in specificity outside of a physical interaction with Sen1, for example, they could play role in slowing down RNAPII transcription to promote termination (Carroll et al., 2004, Carroll et al., 2007, Han et al., 2020).

The precise mechanism through which Sen1 stimulates dissociation of RNAPII has been the subject of extensive debate. It was speculated that this could occur in one of three ways: 1) the allosteric model proposes contacts between the polymerase and the nascent transcript are disrupted by the helicase, resulting in a conformational change; 2) the hyper-translocation/torpedo model proposes that a pushing force by the helicase causes the polymerase to move forward and dissociate; and 3) the 'hybrid-shearing' model suggests that the helicase disrupts the DNA:RNA hybrid within the catalytic centre of the polymerase by exerting a pulling force on the RNA. The requirement for ATP hydrolysis to provoke removal of RNAPII is reminiscent of the bacterial Rho factor. Rho translocates 5'-3' along nascent RNA using ATP-hydrolysis and dissociates the bacterial polymerase off the DNA. In favour of the torpedo model, the Libri lab have demonstrated that Sen1 is also able to "push" stalled RNAPII, promoting its forward translocation. This ability for Sen1 to apply mechanical force

depends on the presence of both nascent RNA and ATP hydrolysis. Further experiments identified that the translocation of Sen1 along the nascent RNA specifically positions it relative to the polymerase in such a way that it can elicit termination (Han et al., 2017).

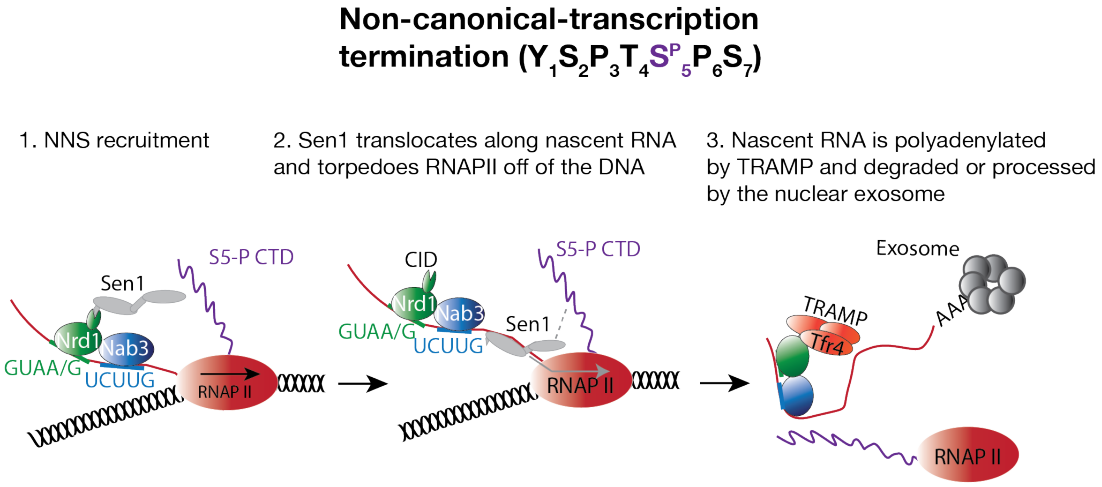


Figure 1.14 The non-canonical pathway for RNAPII transcription termination. The non-canonical pathway terminates short noncoding RNA species and cryptic unstable transcripts. Here, specific RNA sequences are recognised by RNA binding proteins Nrd1 and Nab3. The helicase Sen1, which interacts with the S5-P form of RNAPII is recruited to the termination site, and translocates along the nascent RNA leading to dissociation of the polymerase. The released nascent RNA is either trimmed or degraded by the nuclear exosome.

Coupling of NNS termination with the nuclear exosome: processing and degradation and short noncoding RNAs

Regardless of the precise mechanism leading to termination, the NNS complex plays a crucial role in coupling termination with exosome dependent processing/degradation of the released nascent RNA. By virtue of the exosome bearing the nuclear-specific Rrp6 exonuclease, snRNA and snoRNAs undergo 3' end trimming, thus become mature species ready to carry out their biological function. On the other hand, CUTs are rapidly degraded to prevent their toxic accumulation, as unlike the small nuclear RNA species, they are not protected from complete exosomal degradation by associated ribonucleoproteins (reviewed (Ogami et al., 2018)).

The CID of Nrd1, in addition to interacting with Ser5P-RNAPII CTD also interacts with TRAMP (a co-factor of the exosome) in a mutually exclusive manner, thus

separating these interactions temporally. Like Sen1, the Trf4 subunit of TRAMP contains a NIM (Nrd1 interacting motif) that mimics the CTD. TRAMP polyadenylates the nascent RNA to direct it for processing, and this activity is stimulated by the binding of Nrd1 (Tudek et al., 2014). A subsequent study found that Nrd1 also interacts with another co-factor of the exosome, Mpp6, via a NIM, and the binding of one factor precludes the other, suggesting these reflect two parallel Nrd1 mediated pathways that couple termination to the exosome (Kim et al., 2016). Indeed, this coupling is important for efficient termination, as termination defects at NNS targets are observed in *rrp6Δ* cells (Fox et al., 2015).

The role of the NNS complex in fail-safe termination and attenuation

Though NNS mediated termination predominates at sn/snoRNA genes, it is also thought to act as a backup pathway when canonical termination by CPF-CF fails at a PAS. Evidence suggests that there is some functional overlap between the two pathways, as they have the ability to recognise similar sequence elements (Porrua et al., 2012). Transcription run-on assays have shown that mutations in both Sen1 and Nrd1 exacerbate readthrough transcription of a pA terminator in the absence of functional Rat1. In addition, Nrd1 appears to localise to the 3' end of genes in *rat1-1* cells and is associated with Ser5-P CTD; the levels of which remain elevated in the mutant. Moreover, in a *rat1-1* background, the 3' end of the readthrough transcript matched Nrd1 and Nab3 RNA recognition motifs (RRMs) in the 3' UTR of the template (Rondón et al., 2009).

Conversely, the NNS pathway is also thought to be involved in premature termination of certain transcripts before the PAS, which in turn regulates gene expression. For example, the gene coding for Nrd1 is autoregulated by NNS-dependent attenuation. Both Nrd1 and Nab3 RRMs are present within the both 5' UTR and the coding sequence of the *NRD1* gene. These direct premature termination events, depending on the availability of NNS components within the cell for early termination. It was estimated that only approximately 20% of RNAPII molecules translocate to the 3' end of this gene under wild type conditions, and disruption of the RRMs results in an increase in the levels of steady-state *NRD1* mRNA (Arigo et al., 2006). Other genes

have since been identified that have attenuator regions that also promote premature termination by Sen1.

1.9.2 The role of Sen1 in the resolution of R loops

Accumulating evidence suggests that the absence of Sen1 helicase activity leads to the stabilisation of R loops, triggering transcription associated recombination and genome instability.

Work from the Proudfoot lab indicates that Sen1 is a key player in the removal of R loops formed during transcription (Mischo et al., 2011). Cells carrying the *sen1-1* allele (G1747D mutation in the helicase domain) show a hyperrecombination phenotype, which was exacerbated by increases in both the length and transcription rate of a gene. R loops were observed across the recombining sequences via DRIP-seq, and the increased levels of recombination were suppressed by overexpression of RNase H (Mischo et al., 2011).

Importantly, overexpression of the Sen1 helicase domain was also able to suppress the hyperrecombination phenotype of both *sen1-1* and *mft1*Δ cells (a mutant which has increased R loops but no termination defect), suggesting that Sen1 helicase activity is able to directly constrain the formation of R loops in addition to its role in termination (Mischo et al., 2011). Interestingly, the genetic interactions of *sen1-1,* which is synthetic defective in combination with genes involved in homologous recombination, suggests that the sensing and repair of DSBs is essential for viability in this background. Synthetic lethality was observed in combination with *rad50*Δ and *mre11*Δ, and increased sensitivity to replication stress with *sgs1*Δ, *srs2*Δ and *rad52*Δ (Mischo et al., 2011).

Interestingly, staining the nuclei of *rnh1Δ rnh2Δ* cells with an antibody against R loops shows that they form across the whole genome (Wahba et al., 2011). Mapping the genome wide distribution of R loops determined hybrid-prone regions, particularly across highly transcribed genes (El Hage et al., 2014). However, under normal conditions i.e. in the presence of RNase H, any R loops are quickly removed. A study from the Koshland lab found that in a *rnh1Δ rnh2Δ* background, only a subset of cells

experience DNA damage, indicating R loops are not converted to damage efficiently in this background (Amon and Koshland, 2016). On the other hand, when Sen1 is also depleted (using an auxin degron); in every cell, an accumulation of irreparable DNA damage occurs at multiple loci, resulting in lethality. The authors conclude this is likely due to the fact that the depletion of Sen1 is removing a second potential mechanism for R loops removal, thus resulting in their stabilisation, sensitising them to damage. They argue it is unlikely to be due to defects in termination, as *sen1-AID* cells show termination defects but remain viable following transient depletion of Sen1 when the RNase H enzymes are present. In *rnh1Δ rnh2Δ sen1-AID* cells, the damage was shown to be accrued during S phase, and often in regions where replication forks are likely to clash with transcription head-on. The fact that not all hybrid forming-regions accumulated DNA damage in *sen1-AID rnh1Δ rnh2Δ* cells suggests that additional features, for the example proximal chromatin environment, may predispose some R loop forming regions as hotspots for DNA damage (Costantino and Koshland, 2018).

As *sen1-1* shows both an accumulation of R loops and termination defects, this suggests the helicase domain of Sen1 may mediate two important activities; resolution of co-transcriptional R loops and its terminator function (Porrua and Libri, 2013, Han et al., 2017). Thus, several important questions are yet to be resolved; including whether these two functions are genetically separable, and to what extent the increase in R loops observed in Sen1 mutants is a consequence of R-loops resolution defects, or instead a consequence of transcription termination defects (Porrua et al., 2016).

1.9.4 The Role of Sen1 in DNA repair

Sen1 has also been implicated in the process of transcription-coupled nucleotide excision repair (TC-NER). This pathway removes bulky lesions induced by UV irradiation and other agents that are encountered by the transcription machinery on the template strand of a transcribed gene. A strong interaction between the N-terminal of Sen1 and the endonuclease Rad2 (responsible for 3' cleavage of the lesion site) was initially shown by yeast-two-hybrid. Interestingly, at semi-permissive temperature, the *sen1-1* mutant does not display sensitivity to UV light, however its dysfunction exacerbates the UV-induced growth defects of *rad2Δ* cells. This suggests that it may play some redundant role in TC-NER (Ursic et al., 2004). Indeed Rad26, a DNA-dependent ATPase has been shown to indirectly facilitate TC-NER by antagonising proteins involved in suppressing the pathway (Li, 2015). However, evidence suggests that compared with Rad26, Sen1 has a more direct role in TC-NER (Li et al., 2016).

It appears that the helicase activity of Sen1 does not play a major role in TC-NER, as mutations within the catalytic domain that compromise its ATPase activity only result in extremely mild TC-NER defects. However, cells lacking the N-terminal of Sen1 (*NtΔ*) display a significant reduction in the repair of UV induced lesions on the transcribed strand, highlighting this region as important. In addition, cells lacking the extreme C-terminal (downstream of the essential region and NLS, *CtΔ*) also showed defects, though milder. In *sen1NtΔCtΔ* cells, the defects in TC-NER are partially restored by deletion of the TC-NER pathway repressor *SPT4*. As deletion of *SPT4* rescues the TC-NER defects in *rad26Δ* cells (where Rad26 is an antagonist of Spt4), this suggests that part of Sen1 function in TC-NER may be to similarly antagonise repressors, but it could also have another more direct role as the defects are not fully complemented. In addition, when combined with *rpb9Δ* (a subunit of RNAPII, shown to be essential for an alternative TC-NER sub pathway that is independent from Rad26), *sen1NtΔCtΔ* cells are inviable (Li et al., 2016).

Interestingly, the *sen1NtΔCtΔ* strain increased the sensitivity to UV irradiation in cells defective in other NER sub pathways, including a *rad7Δ rad26Δ* double mutant, where Rad7 is essential for all other global genomic NER. Thus, Sen1 may

also play a role in a TC-NER-independent pathway to repair UV induced lesions (Li et al., 2016).

A more recent study has shown that Sen1 is also involved in promoting high fidelity repair of DSBs. It was shown to be recruited to double strand break sites by the Mre11 subunit of the MRX complex (which is involved in end resection). Using the HO endonuclease to generate a single double strand break in the yeast *MAT* locus (HO-DSB), R loops were found to accumulate on both sides of the HO-DSB following the depletion of Sen1. Interestingly, in *sen1-1* cells, these elevated hybrid levels lead to increased resection and generation of ssDNA at the break site; priming a non-canonical mechanism for end resection that depends on Mre11 and Dna2. Though this alternative resection pathway did not significantly limit homology directed repair when compared with functional Sen1 cells, the authors observed there was a significant increase in mutagenic non-homologous end joining events as a result of Ku70 accumulation (which acts as a scaffold for NHEJ proteins until it is outcompeted by MRX to promote HR) (Rawal et al., 2020). Thus, they suggest that Sen1, which physically interacts with Mre11, resolves DNA:RNA hybrids which otherwise hinder the activity of Mre11 in overcoming the Ku barrier to promote HR over NHEJ (Yüce and West, 2013, Rawal et al., 2020). Whether the potential DNA repair roles of Sen1 are conserved in humans and are relevant to the pathology of the associated diseases is the subject of ongoing study.

1.9.5 The role of Sen1 during DNA replication

Importantly, in addition to its role in transcription, Sen1 and its human orthologue Senataxin have also been linked to DNA replication. In HeLa cells, an accumulation of nuclear Senataxin foci has been observed during S/G2, the time at which DNA synthesis is occurring. Their formation is increased in response to DNA damage induced by various agents, as well as during replication stress resulting in impaired fork progression. These foci were found to colocalise with the DNA damage markers 53BP1 and γH2AX . Both transcription shut-off and global digestion of R loops by RNase H1 diminished these nuclear foci, and conversely; increasing the levels of R loops by inhibiting Topoisomerase I results in concomitant increase of Senataxin foci, thus linking these foci to the formation of transcription induced R loops. Taken

together, this suggests that Senataxin plays a role in coupling the DNA damage response with transcription and replication stress. As such, it may help to maintain genome stability at a time when the transcription and replication machinery are likely to collide with one another and induce the formation of deleterious R loops (Yüce and West, 2013).

Further studies in yeast also point towards a replication-specific role for Sen1. Evidence for the localisation of Sen1 to replication forks was first described by Alzu et al., where chromatin immunoprecipitation and DNA microarrays (ChIP-chip) was used to analyse the distribution patterns of Sen1 and Nrd1 across the *S. cerevisiae* genome (Alzu et al., 2012). To identify regions of actively replicating chromosomal DNA, indicative of replication fork progression, BrdU-IP-ChIP was used; where newly synthesised DNA is labelled via BrdU incorporation for immunoprecipitation (Viggiani et al., 2010). These analyses showed that in both perturbed (HU-treated) and unperturbed conditions, Sen1 clusters correlate with replicating chromatin, and interestingly, this localisation with forks is independent from Nrd1 (Alzu et al., 2012). In addition, though origin firing is unaffected in *sen1-1* cells, the level of BrdU incorporation was reduced at the highly transcribed *PDC1* locus. 2D gel analysis was used to observe forks oriented head-on relative to *PDC1* transcription in Sen1 depleted cells. An accumulation of arrested forks, concomitant with an increase in ssDNA suggests that Sen1 is important for facilitating fork progression across this highly transcribed gene, thus explaining the decreased BrdU incorporation exhibited by helicase impaired *sen1-1* cells. Finally, the Sen1 depleted cells rely on either functional fork protection mechanisms or the local firing of neighbouring back-up origins to rescue *PDC1* stalled forks. Interestingly, within the same replicon, advancement of the sister fork was also shown to be affected (Brambati et al., 2018). Moreover, immunoprecipitation of DNA:RNA hybrids and qPCR showed R loops are enriched at the *PDC1* locus in Sen1 depleted cells, both in unperturbed and HU-treated conditions (Alzu et al., 2012).

In agreement with the proposed role of Sen1 in facilitating fork progression, at non-permissive temperature, *sen1-1* cells exhibit slower kinetics through S phase, terminal G2 arrest and activation of the Rad53 checkpoint kinase (Alzu et al., 2012). Finally, *sen1-1* cells depend on a number of DNA repair proteins for viability, including HR

genes, and genetically interacts with various replisome components (Alzu et al., 2012, Mischo et al., 2011). Considering this data, the authors concluded that a fraction of Sen1 travels with forks and helps to maintain genomic stability, particularly when forks collide head on with transcription bubbles at highly transcribed RNAPII genes and generate stable R loops (Alzu et al., 2012). However, examining the role of Sen1 at replication forks comes with extreme difficulty. So far, it has been impossible to tease apart whether the defects observed during DNA replication in Sen1 mutants are a direct result of its role at replication forks, or whether they are an indirect result of deregulating its role in the removal of R loops or the termination of transcription.

Our lab has since identified that Sen1 co-purifies with the eukaryotic replisome during S phase in *S. cerevisiae*, and this introductory section will briefly describe the previous work carried out in the lab (by R. Appanah, G. De Piccoli, published in (Appanah et al., 2020)), from which the experiments and results described this thesis originated. Sen1 was initially identified as one of the proteins pulled down with the core replisome by mass spectrometry (MS) screens of S phase extracts using the CMG components Sld5 and Mcm4 as bait. This interaction was confirmed by immunoprecipitation (IP) of the GINS component Sld5 and immunoblotting for Sen1. Furthermore, pulldown of N-terminally TAP-tagged Sen1 from G1, S and G2 cell extracts revealed that Sen1 co-IPs with replisome components specifically during S phase. Importantly, the interaction of Sen1 with these replisome components does not depend on Nrd1 or Nab3, and IPs of the *rpb1-1* allele (where RNAPII is no longer present on chromatin (Zanton and Pugh, 2006, Kim et al., 2010)) also revealed that Sen1 replisome binding is independent from RNAPII transcription (Appanah et al., 2020).

To further characterise this interaction, several fragments of Sen1 (summarised in fig 1.15), N-terminally fused with TAP were tested to identify the minimal domain required for association of Sen1 with the replication machinery. These were designed to span either the N-terminal domain, the essential helicase domain, or both. IPs of S phase cells expressing these fragments revealed that the N-terminal domain of Sen1, comprising residues 2-931, is both necessary and sufficient for binding to the replisome during S phase (Appanah et al., 2020).

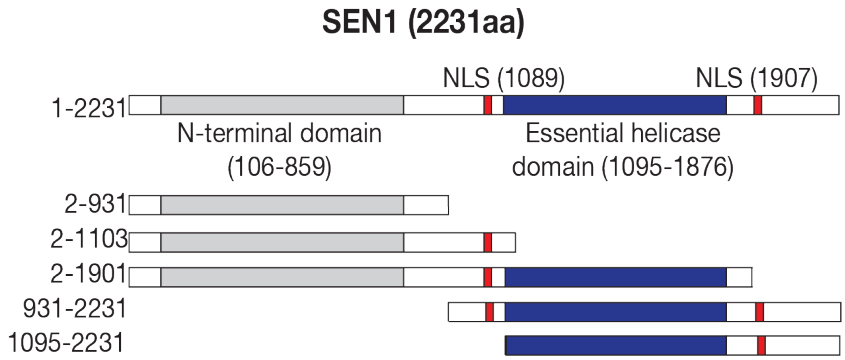


Figure 1.15 Schematic of Sen1 fragments assessed for replisome binding in G1 and S phase. Various fragments of Sen1 N-terminally tagged with TAP were ectopically expressed at the *LEU2* locus under control of the strong inducible *GAL1* promoter

Additional experiments were then carried out in order to identify the potential Sen1 binding partners. MS analysis to compare IPs of Sen1 (2-931) during G1 and S phase found that as expected, Sen1 N-terminal co-purifies with all replisome components during S phase, however it also interacts with Ctf4 and GINS during G1. Importantly, IPs showed that the interaction with GINS is lost in the absence of Ctf4, however Ctf4 association is unaffected by inactivation of GINS (Appanah et al., 2020). Hence, Ctf4 was identified as a candidate Sen1 binding partner.

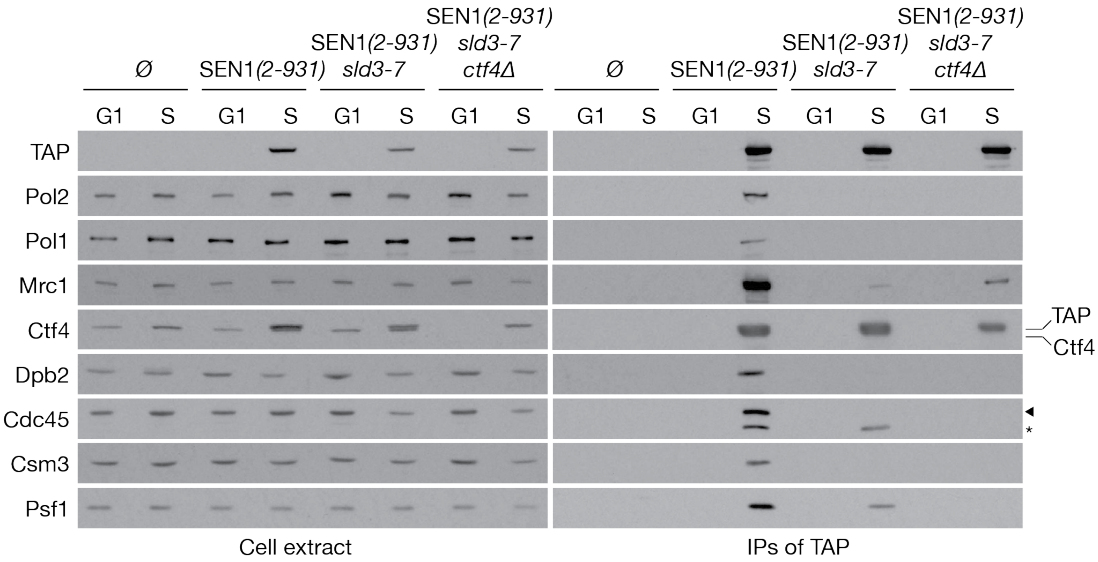


Figure 1.16 The N-terminal domain of Sen1 interacts with the replisome during S phase via Ctf4 and Mrc1. The described strains were grown in YPRaf and synchronised in G1 by the addition of alpha factor. At this stage the G1 samples were collected. For S phase samples, cells were maintained in G1 for 35mins in YPGal to induce expression of the E3 ubiquitin ligase Ubr1 and the Sen1 fragment. The arrested cells were then shifted to 37°C to degrade the degron tagged *sld3-7* allele, before releasing into S phase and harvesting after 20mins. Immunoblot analysis probing for various replisome components is shown. During S phase, the N-terminal domain of Sen1 binds to the replisome. In the absence of origin firing,

Sen1 maintains some level of affinity for Mrc1, which is increased in the absence of Ctf4. (Note, G1 samples were collected before induction of the Sen1 construct). Asterix indicates a non-specific band, arrow indicates protein of interest (Appanah et al., 2020).

However, during S phase, although deletion of *CTF4* significantly reduces co-precipitation of the replisome with Sen1 (2-931), the association is not completely lost, indicating other binding partners may exist. Consequently, *sld3-7-td* cells, where the cell extract enters S phase, but origin firing does not occur (meaning the replisome does not form) were used to search for other possible candidates. In this background, Sen1 (2-931) was also found to co-purify with the replisome component Mrc1 in addition to Ctf4 and GINs, implicating Mrc1 as another potential binding partner. The binding of Mrc1 was maintained in the absence of Ctf4 (fig 1.16). Interestingly, in *ctf4Δ mrc1-AID* cells (the degron was used as the double deletion mutant is lethal (Gambus et al., 2009), though co-purification of the replisome was significantly reduced, a low level of binding is still observed {Appanah, 2020 #762). This suggests that the association between Sen1 and the replisome occurs mainly via Ctf4 and Mrc1, but other, currently unidentified partners may also exist.

1.9.6 The *sen1-3* allele: teasing apart the replication and termination roles of Sen1

As Sen1 has both low cellular abundance and processivity, travelling with replisomes offers an attractive mechanism through which it can be quickly recruited to sites where its biological function is required. However, simply disrupting the activity of Sen1 would not allow illumination of the significance of its localisation to forks, as any replication defects could be an indirect result of disrupting its other functions; for example, termination. With the aim of developing a mutant to help understand the mechanism and significance of Sen1 at forks, as a first step, our lab mapped the minimal interacting domain between Sen1 and the replisome to the N-terminal of the protein (as discussed in the previous section). Further mini-truncations within this region found residues 622-931 to be the shortest fragment still able to support replisome binding (Appanah et al., 2020). The N-terminal domain of Sen1 is predicted to contain several Armadillo repeat motifs (ARMs). Though the primary sequence of these motifs is degenerate, each repeat typically spans approximately 40 residues that

fold in a highly conserved tertiary structure comprising three alpha helices. The tandem ARM repeats further fold and interact with one another to form a compact super-helix of the helices (Huber et al., 1997). Residues conserved among close yeast relatives within the minimal interacting 622-931 fragment, that are predicted to occur on the surface of the ARM superhelix, were mutated and assessed for replisome binding. Excitingly, the *sen1-3* mutant (*W773A E774A W777A*) was exclusively unable to support replisome binding (fig 1.17). Crucially, this mutant still maintained binding to RNAPII^{Rpo21} at a level equivalent to the wild type protein (Appanah et al., 2020).

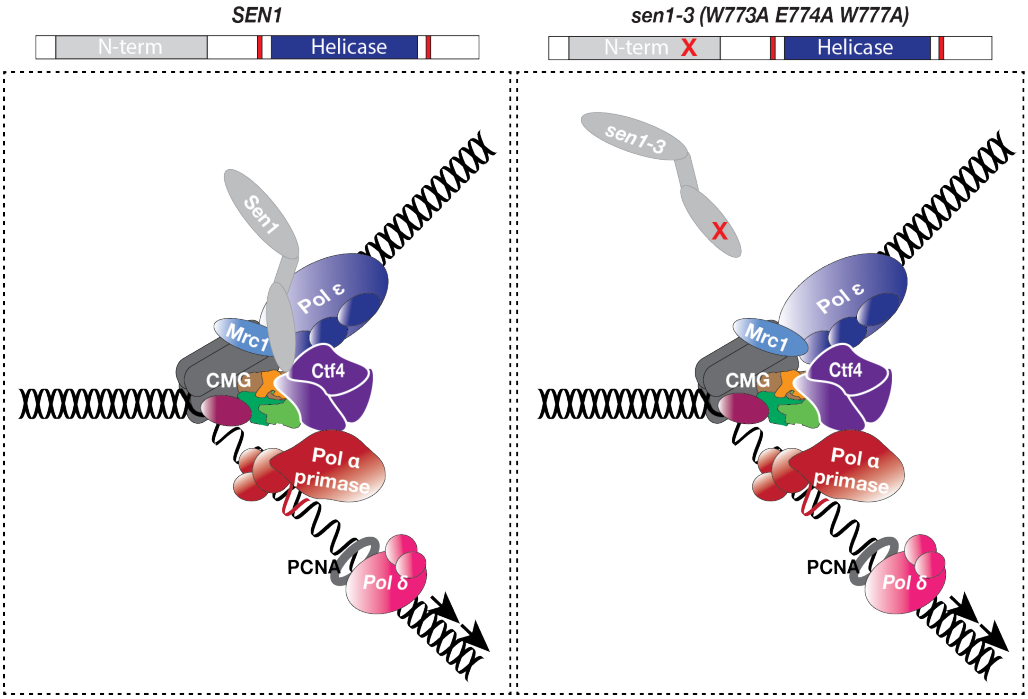


Figure 1.17 Schematic of Sen1 replisome association and the *sen1-3* allele. During S phase, Sen1 binds to the replisome via Ctf4 and Mrc1 (and possibly other replisome components). The *sen1-3* mutant which has a mutation in the N-terminal domain of the protein (*W773A E774A W77A*) is no longer able to associate with replication forks.

To further identify if *sen1-3* is a separation of function mutant, the ability of this allele to support RNAPII transcription termination was examined. Two model NNS-target genes were selected (the small nucleolar RNA *SNR13* and the CUT *NEL025*c) for analysis by qRT-PCR in *SEN1, sen1-1* and *sen1-3* cells. In contrast to *sen1-1* cells, which show strong defects in transcription termination at non-permissive temperatures, no significant difference in readthrough transcription was detected in the *sen1-3* background compared with wild type, even in the absence of *UPF1* (which

can direct longer readthrough transcripts for degradation via the nonsense mediated decay pathway). Finally, CRAC experiments (crosslinking and analysis of cDNA) were performed to map the distribution of RNAPII genome wide. UV crosslinked protein-RNA complexes were purified using IgG beads directed against Rpo21 and fractionated to obtain the RNAPII containing fraction. Following digestion of any protein in the sample, the purified RNAs were reverse transcribed, and the cDNA sequenced and quantified. Metagene analysis was carried out; revealing no difference in the mean CRAC read-counts between the *SEN1* and *sen1-3* profiles at a set of validates CUTs. On the other hand, a control strain depleted of Nrd1 shows a strong termination defect, with elevated RNAPII levels downstream of the transcription end site (Appanah et al., 2020). Thus, the *sen1-3* allele is a useful tool for studying the role of Sen1 at replication forks without affecting the catalytic activity of the protein.

1.10 Conservation of Sen1

Sen1 is evolutionarily conserved among most eukaryotes. The preserved domain structure, where N- and C-terminal domains mediate protein interactions, along with its highly conserved ATPase helicase domain suggests Sen1 may have evolved to carry out similar functions. Indeed, in *Sz. pombe*, the *sen1+* gene encodes an orthologue of Sen1 with 31% sequence homology. Importantly, $Sen1^{Sz.pombe}$ has been shown to possess ATP-dependent 5'-3' DNA and RNA helicase activity *in vitro*. *SEN1* is not an essential gene in *Sz. pombe*, however it has a paralog Dbl8 (which also has ~30% sequence homology with *S. cerevisiae* Sen1 (Kim et al., 1999a).

Senataxin is the human orthologue of *S. cerevisiae* Sen1, encoded by the *SETX* gene. Mutations within Senataxin have been associated with the neurodegenerative diseases; juvenile onset autosomal recessive ataxia oculomotor apraxia type 2 (AOA2) and autosomal dominant amyotrophic lateral sclerosis type 4 (ALS4). However, though some of the mutations resulting in these diseases have been mapped to the N-terminal and helicase domain of the protein, the exact mechanism leading to these pathologies has yet to be elucidated. Between *S. cerevisiae* Sen1 and Senataxin, there is approximately 30% sequence identity within the highly conserved helicase domain. This suggests some functions of the helicase domain are likely to be conserved in

higher eukaryotes (Leonaitė et al., 2017, Bennett and La Spada, 2015). Indeed, though Senataxin is not required for termination of *SNR* genes as in yeast, its helicase activity has been implicated in the termination of some RNAPII transcribed genes. As mentioned previously, studies in *S. cerevisiae* have shown that Sen1 can cooperate with the Rat1 exonuclease for efficient termination across some protein coding genes, and mutation of Sen1 helicase domain exacerbates transcription readthrough in a background where Rat1 is also disrupted (Kawauchi et al., 2008). In HeLa cells, the knockdown of *SETX* also results in readthrough transcription, and indeed alters the distribution of RNAPII, leading to its accumulation downstream of the poly(A) site. Interestingly, in this study, R loops were shown to form in a transcription dependent manner across the gene body, and in Senataxin depleted cells, R loops accumulated upstream at the G-rich pause elements. Somewhat surprisingly, overexpression of RNase H to resolve R loops also resulted in termination defects. Thus, it appears that both the formation of R loops at terminator regions (likely to promote polymerase pausing), and then their subsequent resolution by Senataxin are both crucial steps for termination at these G-rich pause sites. The recruitment of Xrn2 (the human orthologue or Rat1) to these regions is severely compromised in *SETX* knockdown cells, implying the resolution of R loops at these regions by Setx promotes Xrn2-mediated degradation by releasing the nascent transcript, allowing access to the 5' hydroxy group for the 5'-3' exonuclease (Skourti-Stathaki et al., 2011). Thus, similar to Sen1, it appears Senataxin plays a role in canonical transcription termination by Rat1/Xrn2 in human cells.

Finally, of direct interest to the work carried out in this thesis, Senataxin has been demonstrated to form nuclear foci during S/G2, which co-localise with 53BP1, a protein that responds to DNA damage. These foci are increased following treatment with aphidicolin, an inhibitor of the replicative polymerases which leads to fork stalling; thus, linking them to replication stress. Interestingly, the Senataxin foci were reduced by treatment with RNase H, and inhibition of RNAPII transcription using α-amanitin suppressed their induction. Thus, their formation appears to be linked to transcription-induced R loops and replication stress. This is consistent with the proposed model that Sen1 may remove R loops at sites of fork-stalling transcription replication collisions (Yüce and West, 2013, Alzu et al., 2012).

Aims of this thesis

My PhD project has focused on the role of the conserved DNA:RNA helicase Sen1 during DNA replication. Much evidence suggests that Sen1 is important for genome stability. So far, cells with mutant Sen1 variants have been shown to:

- Exhibit defects in the termination of both coding and noncoding RNAPII transcription
- Accumulate deleterious R loops
- Accumulate abnormal replication fork intermediates
- Exhibit a prolonged S phase, terminal G2 arrest and activation of the Rad53 checkpoint kinase
- Genetically interact with genes involved in DNA repair

Interestingly, in *S. cerevisiae,* Sen1 has been shown to localise to replication forks. Little is known about how eukaryotic forks deal with barriers that impede their progression, including the transcription machinery and co-transcriptional R loops. Also, understanding the biological significance of this localisation could be interesting in light of Senataxin involvement in human disease.

My PhD project aimed to answer two important questions:

1) **What is the biological relevance of Sen1 association with the replisome?**

Previous work in the lab generated the *sen1-3* mutant, which can no longer bind to the replisome, but retains full capability of its transcription termination function. Thus, this mutant was used as a tool to investigate the loss of Sen1 from forks without affecting its catalytic activity. I aimed to elucidate any fork-specific role by analysing the phenotypes of *sen1-3* combined with various DNA replication and RNA metabolism mutants.

2) **What is the mechanism of interaction between Sen1 and the replisome and how is the interaction regulated by the cell cycle?**

It has already been demonstrated that Sen1 associates with replication forks via an interaction between its N-terminal domain and the replisome components Ctf4 and Mrc1. However, several aspects of the interaction between Sen1 and the replisome are still poorly understood. In fact, while the N-terminal domain of Sen1 can interact with Ctf4 and Mrc1 throughout the cell cycle, full length Sen1 binding is restricted to S phase. Moreover, additional components of the replisome are predicted to interact with Sen1, but their identity and mechanism of interaction is currently unknown. Thus, through biochemical analysis, I aimed to further characterise this interaction and understand how it is regulated by the cell cycle.

Chapter 2: Materials and methods

2.1 Yeast methods and techniques

2.1.1 Growth of yeast strains

S. Cerevisiae strains were stored as 25% (v/v) glycerol suspensions at -80ºC. Stocks were grown by streaking a small amount onto solid non-selective YPD medium at 30ºC (or 24ºC for heat sensitive strains) and incubated until colonies of the desired size were obtained. Individual colonies were then inoculated into liquid YP medium supplemented with the appropriate sugar (glucose, galactose or raffinose) to 2% (v/v) for subsequent experiments.

Transformants containing constructs under control of the inducible *GAL1* promoter were grown in 2% (v/v) YP raffinose as a neutral carbon source to suppress induction of the constructs, or 2% (v/v) YP galactose to induce expression of the constructs. Strains were derived from W303-1a (*leu2-3,112 trp1-1 kan1-100 ura3-1 ade2-1 his3-11,15, rad5*-535), and so require medium supplemented with the additional amino acids for growth, allowing for the use of auxotrophic markers. Transformants with a marker gene, which encodes an enzyme that is necessary for the synthesis of a selected amino acid are able be selected for by their growth on medium deficient of that amino acid. *hphNT* or *kanMX* cassettes were used to select for antibiotic resistance, where cells were grown on YPD medium supplemented with either HygromycinB (Hygromycin B Gold™, InvivoGen) or G418 (Invitrogen) respectively.

Table 2.1 List of strains used in this study

Strain No.	Genotype	Source
1	*MAT*a	Lab collection
6	*MAT*α	Lab collection
16	*MAT*a *rad51Δ::kanMX*	Lab collection
22	*MAT*a *TAP-SLD5 (kanMX) pep4Δ::URA3 ADE2+*	Lab collection
55	*MAT*a/*MAT*α	Lab collection

74	*MATa pep4Δ::ADE2+*	Lab collection
239	*MATα rad18Δ::hphNT*	Lab collection
716	*MATα mph1Δ::kanMX*	Lab collection
792	*MATα rad52Δ::kanMX*	Lab collection
1175	*MATα top1Δ::kanMX*	Lab collection
1353	*MATa SEN1-TAP (kanMX) pep4Δ::ADE2+*	Lab collection
1852	*MATa leu2-3,112:: GAL_TAP_empty (LEU2+) pep4Δ::ADE2+*	Lab collection
1941	*MATa leu2-3,112 :: GAL-TAP-Sen1 (2-2231) (LEU2+) pep4Δ::ADE2+*	Lab collection
1942	*MATa leu2-3,112 :: GAL-TAP-Sen1 (2-1901) (LEU2+) pep4Δ::ADE2+*	Lab collection
1943	*MATa leu2-3,112 :: GAL-TAP-Sen1 (931-2231) (LEU2+) pep4Δ::ADE2+*	Lab collection
1956	*MATa leu2-3,112 :: GAL-TAP-Sen1 (2-1103) (LEU2+) pep4Δ::ADE2+*	Lab collection
1957	*MATa leu2-3,112 :: GAL-TAP-Sen1 (2-931) (LEU2+) pep4Δ::ADE2+*	Lab collection
2201	*MATa SEN1-TAP (kanMX) pep4Δ::ADE2+*	Lab collection
2222	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) pep4Δ::ADE2+*	Lab collection
2277	*MATα hpr1Δ::kanMX*	Lab collection
2337	*MATa mph1Δ::kanMX*	Lab collection
2584	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) (LEU2+)*	Lab collection
2607	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2- 2231) W773A E774A W777A (LEU2+)*	Lab collection
2609	*MATa sen1Δ::URA3-CP ACT1-3HA-Sen1 (2-2231) D850A E851G V852A L853G L854A (LEU2+)*	Lab collection
2734	*MATa rnh1Δ:: hphNT rnh201Δ::HISMX*	Lab collection
2736	*MATa rnh1Δ:: hphNT rnh201Δ::HISMX sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) W773A E774A W777A (LEU2+)*	Lab collection
2737	*MATα rnh1Δ:: hphNT rnh201Δ::HISMX sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) W773A E774A W777A (LEU2+)*	Lab collection
2738	*MATa rnh1Δ:: hphNT rnh201Δ::HIS3MX sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) D850A E851G V852A L853G L854A (LEU2+)*	Lab collection
2794	*MATa td-sld3-7 (kanMX) GAL1-UBR1 (HIS3MX) pep4Δ:: ADE2+ SEN1-TAP (KanMX)*	Lab collection
2795	*MATa td-sld3-7 (kanMX) GAL1-UBR1 (HIS3MX) pep4Δ:: ADE2+ MCM4-5FLAG (hphNT)*	Lab collection
2801	*MATa CTF4-9MYC (kanMX) pep4Δ:: ADE2+*	Lab collection
2808	*MATa SEN1-TAP (kanMX)*	Lab collection
2810	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX)*	Lab collection
2859	*MATa SEN1-TAP (kanMX) mrc1Δ::hphNT*	This study
2861	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) mrc1Δ::hphNT*	This study
2863	*MATa SEN1-TAP (kanMX) rrm3Δ::hphNT*	This study
2865	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) rrm3Δ::hphNT*	This study
2867	*MATa SEN1-TAP (kanMX) tof1Δ::HIS3MX*	This study
2868	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) tof1Δ::HIS3MX*	This study
2876	*MATa SEN1-TAP (kanMX) RAD52-GFP (K.l.TRP1+)*	This study
2878	*MATa SEN1-TAP (kanMX) pif1Δ::HIS3MX*	This study
2880	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) pif1Δ::HIS3MX*	This study

2882	*MATa SEN1-TAP (kanMX) sgs1Δ::URA3-CP*	This study
2884	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) sgs1Δ::URA3-CP*	This study
2890	*MATa SEN1-TAP (kanMX) top1Δ::kanMX*	This study
2892	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) top1Δ::kanMX*	This study
2894	*MATa SEN1-TAP (kanMX) srs2Δ::kanMX*	This study
2896	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) srs2Δ::kanMX*	This study
2897	*MATa SEN1-TAP (kanMX) rad51Δ::kanMX Pol2(K.l.TRP1+)*	This study
2898	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) rad51Δ::kanMX Pol2(K.l.TRP1+)*	This study
2904	*MATa SEN1-TAP (kanMX) rad50Δ::kanMX*	This study
2906	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) rad50Δ::kanMX*	This study
2934	*MATa SEN1-TAP (kanMX) mec1-100 (LEU2+, HIS3MX) sml1Δ::kanMX*	This study
2936	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) mec1-100 (LEU2+, HIS3MX) sml1Δ::kanMX*	This study
2945	*MATa SEN1-TAP (kanMX) rad53Δ::ADE2+ sml1Δ::HIS3MX*	This study
2947	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) rad53Δ::ADE2+ sml1Δ::HIS3MX*	This study
2947	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) sml1Δ::HIS3MX rad53Δ::ADE2+*	This study
2953	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) RAD52-GFP (K.l.TRP1+)*	This study
2955	*MATa SEN1-TAP (kanMX) ctf18Δ::K.l.TRP1+*	This study
2957	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) ctf18Δ::K.l.TRP1+*	This study
2959	*MATa SEN1-TAP (kanMX) fob1Δ::HIS3MX*	This study
2961	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) fob1Δ::HIS3MX*	This study
2963	*MATa SEN1-TAP (kanMX) rad9Δ::HIS3MX*	This study
2965	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) rad9Δ::HIS3MX*	This study
2967	*MATa SEN1-TAP (kanMX) rad24Δ::hphNT*	This study
2969	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) rad24Δ::hphNT*	This study
3049	*MATa SEN1-TAP (kanMX) mec1Δ::ADE2+ sml1Δ::HIS3MX*	This study
3051	*MATα SEN1 (W773A E774A W777A)-TAP (kanMX) mec1Δ::ADE2+ sml1Δ::HIS3MX*	This study
3057	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) RAD52-GFP (K.l.TRP1+) mrc1Δ::hphNT*	This study
3062	*MATa SEN1-TAP (kanMX) RAD52-GFP (K.l.TRP1+) ctf18Δ::K.l.TRP1+*	This study
3126	*MATa SEN1-TAP (kanMX) GAL-TIR1 (K.l.TRP1+) rnh1-AID (hphNT) rnh201-AID (HIS3MX)*	This study
3128	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) GAL-TIR1 (K.l.TRP1+) rnh1-AID (hphNT) rnh201-AID (HIS3MX)*	This study
3134	*MATa SEN1-TAP (kanMX) RAD52-GFP (K.l.TRP1+) mrc1Δ::hphNT*	This study
3148	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) RAD52-GFP (K.l.TRP1+) ctf18Δ::K.l.TRP1+*	This study
3168	*MATa sen1Δ::URA3-CP leu2-3,112 :: ACT1-3HA-SEN1 (2-2231) (LEU2+) rnh1Δ (hphNT) rnh201Δ (hphNT)*	This study
3169	*MATα sen1Δ::URA3-CP leu2-3,112 :: ACT1-3HA-SEN1 (2-2231) (LEU2+) rnh1Δ (hphNT) rnh201Δ (hphNT)*	This study
3186	*MATa leu2-3,112 :: GAL1-TAP-SEN1 (2-931) (LEU2+) mrc1Δ::hphNT pep4Δ::ADE2+*	This study
3187	*MATa leu2-3,112 :: GAL1-TAP-SEN1 (2-1901) (LEU2+) mrc1Δ::hphNT pep4Δ::ADE2+*	This study
3188	*MATa leu2-3,112 :: GAL1-TAP-SEN1 (2-1901) (LEU2+) ctf4Δ::hphNT pep4Δ::ADE2+*	
3224	*MATa SEN1-TAP (kanMX) mec1Δ::ADE2+ sml1Δ::HIS3MX) tel1Δ::hphNT*	This study
3226	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) mec1Δ::ADE2+ sml1Δ::HIS3MX) tel1Δ::hphNT*	This study
3228	*MATa SEN1-TAP (kanMX) tel1Δ::hphNT*	This study

3230	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) tel1Δ::hphNT*	This study
3232	*MATa SEN1-TAP (kanMX) sml1Δ::HIS3MX*	This study
3234	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) sml1Δ::HIS3MX*	This study
3281	*MATa leu2-3,112 :: GAL1-TAP-SEN1 (2-931) (LEU2+) ctf4Δ::kanMX pep4Δ::ADE2+*	Lab collection
3283	*MATa SEN1-TAP (kanMX) ctf4Δ::kanMX*	This study
3285	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) ctf4Δ::kanMX*	This study
3299	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) pif1Δ::kanMX rrm3Δ (HIS3MX)*	This study
3313	*MATa leu2,3-112 :: GAL1-3HA-MRC1 (567-1096) (LEU2+) pep4Δ::ADE2+ mrc1Δ::hphNT*	This study
3466	*MATa SEN1-TAP (kanMX) pif1Δ::kanMX rrm3Δ (HIS3MX)*	This study
3563	*MATa/α leu2-3,112 :: GAL1-TAP-SEN1 (2-1901) (LEU2+) leu2-3,112 :: GAL1-3HA-CTF4 (2-927) pep4Δ::ADE2+*	This study
3566	*MATa/α leu2-3,112 :: GAL1-TAP-SEN1 (2-1901) (LEU2+) leu2-3,112 :: GAL1-3HA-MRC1 (1-418) pep4Δ::ADE2+*	This study
3567	*MATa/α leu2-3,112 :: GAL1-TAP-SEN1 (2-1901) (LEU2+) leu2-3,112 :: GAL1-3HA-MRC1 (1-843) pep4Δ::ADE2+*	This study
3568	*MATa/α leu2-3,112 :: GAL1-TAP-SEN1 (2-1901) (LEU2+) leu2-3,112 :: GAL1-3HA-MRC1 (1-1096) pep4Δ::ADE2+*	This study
3569	*MATa/α leu2-3,112 :: GAL1-TAP-SEN1 (2-1901) (LEU2+) leu2-3,112 :: GAL1-3HA-MRC1 (140-1096) pep4Δ::ADE2+*	This study
3570	*MATa/α leu2-3,112 :: GAL1-TAP-SEN1 (2-1901) (LEU2+) leu2-3,112 :: GAL1-3HA-MRC1 (312-1096) pep4Δ::ADE2+*	This study
3571	*MATa/α leu2-3,112 :: GAL1-TAP-SEN1 (2-1901) (LEU2+) leu2-3,112 :: GAL1-3HA-MRC1 (752-1096) pep4Δ::ADE2+*	This study
3572	*MATa/α leu2-3,112 :: GAL1-TAP-SEN1 (2-931) (LEU2+) leu2-3,112 :: GAL1-3HA-CTF4 (2-927) pep4Δ::ADE2+*	This study
3573	*MATa/α leu2-3,112 :: GAL1-TAP-SEN1 (2-931) (LEU2+) leu2-3,112 :: GAL1-3HA-CTF4 (351-927) pep4Δ::ADE2+*	This study
3574	*MATa/α leu2-3,112 :: GAL1-TAP-SEN1 (2-931) (LEU2+) leu2-3,112 :: GAL1-3HA-CTF4 (426-927) pep4Δ::ADE2+*	This study
3575	*MATa/α leu2-3,112 :: GAL1-TAP-SEN1 (2-931) (LEU2+) leu2-3,112 :: GAL1-3HA-CTF4 (783-927) pep4Δ::ADE2+*	This study
3587	*MATa/α leu2-3,112 :: GAL1-TAP-SEN1 (2-1901) (LEU2+) leu2-3,112 :: GAL1-3HA-MRC1 (1-219) pep4Δ::ADE2+*	This study
3588	*MATa/α leu2-3,112 :: GAL1-TAP-SEN1 (2-1901) (LEU2+) leu2-3,112 :: GAL1-3HA-MRC1 (1-655) pep4Δ::ADE2+*	This study
3589	*MATa/α leu2-3,112 :: GAL1-TAP-SEN1 (2-1901) (LEU2+) leu2-3,112 :: GAL1-3HA-MRC1 (567-1096) pep4Δ::ADE2+*	This study
3591	*MATa/α leu2-3,112 :: GAL1-TAP-SEN1 (2-931) (LEU2+) leu2-3,112 :: GAL1-3HA-MRC1 (2-1096) pep4Δ::ADE2+*	This study
3669	*MATa SEN1-TAP (kanMX) rpb1Δ::hphNT leu2-3,112 :: RPB1 (LEU2+)*	Lab collection
3672	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) rpb1Δ::hphNT leu2-3,112 :: RPB1 (LEU2+)*	Lab collection
3675	*MATa SEN1-TAP (kanMX) rpb1Δ::hphNT leu2-3,112 :: rpb1-E1103G (LEU2+)*	Lab collection
3679	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) rpb1Δ::hphNT leu2-3,112 :: rpb1-E1103G (LEU2+)*	Lab collection
3690	*MATα SEN1-TAP (kanMX) rpb11Δ::URA3-CP trp1-1 :: rpb11-E108G (K.l.TRP1+)*	Lab collection
3698	*MATα SEN1-TAP (kanMX) leu2,3-112 :: GAL1-RNH1 (LEU2+)*	This study
3700	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) leu2,3-112 :: GAL1-RNH1 (2-348) (LEU2+)*	This study
3709	*MATα SEN1 (W773A E774A W777A)-TAP (kanMX) rpb1Δ::URA3-CP trp1-1 :: rpb11-E108G (K.l.TRP1+)*	Lab collection
3715	*MATa SEN1-TAP (kanMX) rnh201Δ::HIS3MX*	This study
3717	*MATa SEN1-TAP (kanMX) rnh1Δ::hphNT*	This study
3719	*MATα SEN1 (W773A E774A W777A)-TAP (kanMX) rnh201Δ::HIS3MX*	This study
3720	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) rnh201Δ::HIS3MX*	This study
3721	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) rnh1Δ::hphNT*	This study

3732	*MATa SEN1-TAP (kanMX) rnh1Δ::hphNT rnh201Δ::HIS3MX leu2,3-112 :: GAL1-RNH1 (LEU2+)*	This study
3734	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) rnh1Δ::hphNT rnh201Δ::HIS3MX leu2,3-112 :: GAL1-RNH1 (LEU2+)*	This study
3741	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) hpr1Δ::kanMX*	This study
3742	*MATa SEN1-TAP (kanMX) hpr1Δ::kanMX*	This study
3745	*MATa sen1Δ::URA3-CP leu2,3-112 :: SEN1 (W773A E774A W777A)-TAP (LEU2+) hpr1Δ::kanMX*	This study
3782	*MATa SEN1-TAP (kanMX) mlp1Δ::hphNT*	This study
3790	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) mlp1Δ::hphNT*	This study
3804	*MATa SEN1-TAP (kanMX) RAD52-GFP (K.l.TRP1+) leu2,3-112 :: GAL1-RNH1 (LEU2+)*	This study
3808	*MATa SEN1-TAP (kanMX) RAD52-GFP (K.l.TRP1+) leu2,3-112 :: GAL1-RNH1 (LEU2+) mrc1Δ::hphNT*	This study
3813	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) RAD52-GFP (K.l.TRP1+) leu2,3-112 :: GAL1-RNH1 (LEU2+) mrc1Δ::hphNT*	This study
3816	*MATa SEN1-TAP (kanMX) mlp1Δ::hphNT mrc1Δ::hphNT*	This study
3817	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) mlp1Δ::hphNT mrc1Δ::hphNT*	This study
3847	*MATa SEN1-TAP (kanMX) RAD52-GFP (K.l.TRP1+) mrc1Δ::hphNT GPD_empty (HIS3MX)*	This study
3848	*MATa SEN1-TAP (kanMX) RAD52-GFP (K.l.TRP1+) mrc1Δ::hphNT GPD-hRNH1 (HIS3MX)*	This study
3849	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) RAD52-GFP (K.l.TRP1+) ctf18Δ::K.l.TRP1+ GPD_empty (HIS3MX)*	This study
3850	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) RAD52-GFP (K.l.TRP1+) ctf18Δ::K.l.TRP1+ GPD-hRNH1 (HIS3MX)*	This study
3851	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) RAD52-GFP (K.l.TRP1+) GPD_empty (HIS3MX)*	This study
3852	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) RAD52-GFP (K.l.TRP1+) GPD-hRNH1 (HIS3MX)*	This study
3853	*MATa SEN1-TAP (kanMX) RAD52-GFP (K.l.TRP1+) GPD_empty (HIS3MX)*	This study
3854	*MATa SEN1-TAP (kanMX) RAD52-GFP (K.l.TRP1+) GPD-hRNH1 (HIS3MX)*	This study
3855	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) RAD52-GFP (K.l.TRP1+) mrc1Δ::hphNT GPD_empty (HIS3MX)*	This study
3856	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) RAD52-GFP (K.l.TRP1+) mrc1Δ::hphNT GPD-hRNH1 (HIS3MX)*	This study
3857	*MATa SEN1-TAP (kanMX) RAD52-GFP (K.l.TRP1+) ctf18Δ::K.l.TRP1+ GPD_empty (HIS3MX)*	This study
3858	*MATa SEN1-TAP (kanMX) RAD52-GFP (K.l.TRP1+) ctf18Δ::K.l.TRP1+ GPD-hRNH1 (HIS3MX)*	This study
3902	*MATα SEN1-TAP (kanMX) leu2,3-112 :: GAL1-RNH1 (LEU2+) hpr1Δ::kanMX*	This study
3904	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) leu2,3-112 :: GAL1-RNH1 (LEU2+) hpr1Δ::kanMX*	This study
3927	*MATa SEN1-TAP (kanMX) RAD52-GFP (K.l.TRP1+) hpr1Δ::kanMX*	This study
3930	*MATa SEN1-TAP (kanMX) RAD52-GFP (K.l.TRP1+) hpr1Δ::kanMX leu2,3-112 :: GAL1-RNH1 (LEU2+)*	This study
3932	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) RAD52-GFP (K.l.TRP1+) hpr1Δ::kanMX*	This study
3935	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) RAD52-GFP (K.l.TRP1+) hpr1Δ::kanMX leu2,3-112 :: GAL1-RNH1 (LEU2+)*	This study
3936	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) RAD52-GFP (K.l.TRP1+) leu2,3-112 :: GAL1-RNH1 (LEU2+)*	This study
3955	*MATa MRC1-18MYC (K.l.TRP1+) pep4Δ::ADE2+*	Lab collection
4018	*MATα leu2,3-112 :: GAL1-3HA-CTF4 (426-927) (LEU2+) pep4Δ::ADE2+ ctf4Δ::kanMX*	This study
4022	*MATa leu2,3-112 :: GAL1-3HA-MRC1 (752-1096) (LEU2+) pep4Δ::ADE2+ mrc1Δ::hphNT*	This study
4037	*MATa/α leu2-3,112 :: GAL1-TAP-EMPTY (LEU2+) leu2-3,112 :: GAL1-3HA-CTF4 (426-927) pep4Δ::ADE2+*	This study
4038	*MATa/α leu2-3,112 :: GAL1-TAP-EMPTY (LEU2+) leu2-3,112 :: GAL1-3HA-MRC1 (2-418) pep4Δ::ADE2+*	This study
4039	*MATa/α leu2-3,112 :: GAL1-TAP-EMPTY (LEU2+) leu2-3,112 :: GAL1-3HA-MRC1 (2-655) pep4Δ::ADE2+*	This study

4040	*MATa/α leu2-3,112 :: GAL1-TAP-EMPTY (LEU2+) leu2-3,112 :: GAL1-3HA-MRC1 (2-843) pep4Δ::ADE2+*	This study
4041	*MATa/α leu2-3,112 :: GAL1-TAP-EMPTY (LEU2+) leu2-3,112 :: GAL1-3HA-MRC1 (752-1096) pep4Δ::ADE2+*	This study
4042	*MATα leu2,3-112 :: GAL1-3HA-MRC1 (2-655) (LEU2+) pep4Δ::ADE2+ mrc1Δ::hphNT*	This study
4081	*MATa leu2,3-112 :: GAL1-3HA-MRC1 (2-843) (LEU2+) pep4Δ::ADE2+ mrc1Δ::hphNT*	This study
4083	*MATa leu2,3-112 :: GAL1-3HA-MRC1 (2-418) (LEU2+) pep4Δ::ADE2+ mrc1Δ::hphNT*	This study
4107	*MATa/α leu2-3,112 :: GAL1-TAP-SEN1 (2-931) (LEU2+) leu2-3,112 :: GAL1-3HA-CTF4 (426-782) pep4Δ::ADE2+ ctf4Δ::kanMX*	This study
4108	*MATa/α leu2-3,112 :: GAL1-TAP-SEN1 (2-931) (LEU2+) leu2-3,112 :: GAL1-3HA-CTF4 (426-782) pep4Δ::ADE2+*	This study
4110	*MATa/α leu2-3,112 :: GAL1-TAP-SEN1 (2-931) (LEU2+) leu2-3,112 :: GAL1-3HA-CTF4 (426-782) pep4Δ::ADE2+*	This study
4189	*MATα sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) (LEU2+) rad52Δ:kanMX*	This study
4191	*MATα sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) (LEU2+) top1Δ:kanMX*	This study
4206	*MATα sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) (LEU2+) rad18Δ:hphNT*	This study
4208	*MATα sen1Δ::URA3-CP ACT1-3HA-SEN1 (W773A E774A W777A) (LEU2+) rad18Δ:hphNT*	This study
4228	*MATα leu2,3-112 :: GAL1-3HA-CTF4 (426-782) (LEU2+) pep4Δ::ADE2+ ctf4Δ::kanMX*	This study
4290	*MATα rpb1-1 (ts)*	Gift from Domenico Libri
4296	*MATa SEN1-TAP (kanMX) chl1Δ::kanMX*	This study
4299	*MATα SEN1 (W773A E774A W777A)-TAP (kanMX) chl1Δ::kanMX*	This study
4325	*MATa SEN1-TAP (kanMX) rpb1-1 (ts) RAD52-GFP (K.l.TRP1+) mrc1Δ::hphNT*	This study
4327	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) rpb1-1 (ts) RAD52-GFP (K.l.TRP1+) mrc1Δ::hphNT*	This study
4329	*MATa SEN1-TAP (kanMX) rpb1-1 (ts) RAD52-GFP (K.l.TRP1+)*	This study
4331	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) rpb1-1 (ts) RAD52-GFP (K.l.TRP1+)*	This study
4334	*MATa rad27Δ::kanMX*	This study
4367	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (W773A E774A W777A) (LEU2+) mph1Δ:hphNT*	This study
4369	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) (LEU2+) mph1Δ:hphNT*	This study
4371	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (W773A E774A W777A) (LEU2+) rad52Δ:kanMX*	This study
4372	*MATa SEN1-TAP (kanMX) rad27Δ:kanMX*	This study
4374	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) rad27Δ:kanMX*	This study
4376	*MATa leu2,3-112 :: GAL1-3HA-MRC1 (418-843) (LEU2+) pep4Δ::ADE2+ mrc1Δ::hphNT*	This study
4377	*MATa leu2-3,112 :: GAL-TAP-Sen1 (931-2231) (LEU2+) pep4Δ::ADE2+ mrc1Δ::hphNT*	This study
4378	*MATa SEN1-TAP (kanMX) rad18Δ:hphNT*	This study
4380	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) rad18Δ:hphNT*	This study
4382	*MATa SEN1-TAP (kanMX) mph1Δ::hphNT*	This study
4384	*MATa SEN1 (W773A E774A W777A)-TAP (kanMX) mph1Δ::hphNT*	This study
4386	*MATa sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) (LEU2+) hpr1Δ::kanMX*	This study
4391	*MATa leu2,3-112 :: GAL1-3HA-CTF4 (2-782) (LEU2+) pep4Δ::ADE2+ ctf4Δ::kanMX*	This study
4397	*MATα sen1Δ::URA3-CP ACT1-3HA-SEN1 (2-2231) (LEU2+) rad27Δ::kanMX*	This study
4399	*MATα sen1Δ::URA3-CP ACT1-3HA-SEN1 (W773A E774A W777A) (LEU2+) rad27Δ::kanMX*	This study
4400	*MATa TAP-SLD5 (kanMX) pep4Δ::URA3-CP MCM7-AID (HIS3MX) GAL1-TIR1 (K.l.TRP1+) ADE2+*	This study
4402	*MATa SEN1-TAP (kanMX) pep4Δ::ADE2+ MCM7-AID (HIS3MX) GAL1-TIR1 (K.l.TRP1+)*	This study
4410	*MATa leu2,3-112 :: GAL1-3HA-CTF4 (351-927) (LEU2+) pep4Δ::ADE2+ ctf4Δ::kanMX*	This study

4428	*MAT*a *leu2,3-112 :: GAL1-3HA-CTF4 (351-927) L351G A352G V355A V356G (LEU2+) pep4Δ::ADE2+ ctf4Δ::kanMX*	This study
4429	*MAT*a *leu2,3-112 :: GAL1-3HA-CTF4 (351-927) D376A S377G E278A S379G D380A L381G (LEU2+) pep4Δ::ADE2+ ctf4Δ::kanMX*	This study
4430	*MAT*a *leu2,3-112 :: GAL1-3HA-CTF4 (351-927) E385A N387A D388G D389G N391A K392A D394G (LEU2+) pep4Δ::ADE2+ ctf4Δ::kanMX*	This study
4431	*MAT*a *leu2,3-112 :: GAL1-3HA-CTF4 (351-927) E403A A404G N405A A406G E407A D408A V409A F410A (LEU2+) pep4Δ::ADE2+ ctf4Δ::kanMX*	This study
4432	*MAT*a *leu2,3-112 :: GAL1-3HA-CTF4 (351-927) Y426A F428A E429G D4230A E4231G E4232A D4233G I4235G D4236A D4237A D4238A D4239A (LEU2+) pep4Δ::ADE2+ ctf4Δ::kanMX*	This study
4433	*MAT*a *leu2,3-112 :: GAL1-3HA-CTF4 (351-927) K447A K448A (LEU2+) pep4Δ::ADE2+ ctf4Δ::kanMX*	This study
4434	*MAT*a *leu2,3-112 :: GAL1-3HA-CTF4 (351-927) S456G H459A S463G (LEU2+) pep4Δ::ADE2+ ctf4Δ::kanMX*	This study

2.1.2 Crossing of yeast strains

A haploid yeast cell has the ability to mate with another haploid yeast cell of the opposite mating type (i.e. *MAT*a x *MAT*α), allowing cells of different genotypes to be crossed. From cells grown on solid non-selective media, two parents were mixed in 150µl of sterile deionised water. For parent strains harbouring different markers, a ratio of approximately 50:1 was used. For parents harbouring the same markers, equal amounts were mixed. After vortexing; 50µl of the cell suspension was plated onto solid non-selective YPD and grown overnight at either 24ºC or 30ºC in order for them to mate and form diploid cells. The diploid containing mixture was then streaked onto the appropriate media to select against the marker of the most abundant strain and obtain individual colonies. After overnight incubation at the appropriate temperature, individual colonies were streaked onto sporulation medium (RSM) and typically grown for 3-5 days, until asci were visible for tetrad dissection. Asci were digested using β-glucoronidase from *Helix pomatia* (Sigma) for 30-35 mins, and the spores dissected on YPD plates via micromanipulation using a Singer MSM400 tetrad dissector. Following germination and growth for 2-3 days, the spores were replica plated onto the relevant selective media. Tetrads were either scanned and scored for viability, or strains of the desired genotype were saved as 25% (v/v) glycerol stocks for subsequent experimental use.

2.1.3 Mating type of yeast strains

The mating type of budding yeast is determined by two non-homologous alleles, *MAT*a and *MAT*α. This renders *S. cerevisiae* a very useful tool for studying the cell cycle, as haploid cells respond to mating pheromone of the opposite mating type. Cells of mating type 'a' respond to the pheromone alpha (α)-factor and arrest in G_1 phase of

the cell cycle, thus allowing the synchronous release of haploid cycling cultures into S phase. The arrested cells undergo a morphological change know as a 'schmoo' where cells produce a nodule. For newly made haploid strains, following growth in liquid YPD overnight, the mating type was assessed by the addition of α-factor pheromone to a final concentration of 7.5µg/ml for 2.5-3 hrs. The cells were viewed under a microscope to examine their morphology. While *MAT*a type strains form 'schmoo' protrusions following α-factor addition, *MAT* α type strains are unaffected and are seen as asynchronously budding populations.

2.1.4 Transformation of yeast strains by the Lithium acetate method

Exponentially growing cultures of yeast at a density of $1.0x10^7$ were harvested by centrifugation, then washed with 1ml of sterile dH_20 and 1ml of a solution of 0.1M lithium acetate [pH 7.5] 1M Tris-HCl [pH 7.5], then resuspended at $2x10^9$ cells/ml in the same solution. 500µg of denatured carrier ssDNA (salmon sperm) and 1-2µg of target DNA (plasmid or PCR) in a volume of 10µl was added to 50µl of the cell suspension. After vortexing vigorously, 40% PEG 4000 in 0.1M lithium acetate [pH 7.5] 1M Tris-HCl [pH 7.5] was added to a final concentration of 33.3% (w/v) PEG, followed again by vigorous vortexing. This was then incubated for 30-60min at 24°C on a rotating platform. Sterile DMSO was then added to a final concentration of 10%, and the mixture was heat shocked at 42°C for 15min. The cell suspension was cooled rapidly on ice for 2 mins, before centrifugation to remove the supernatant. For transformants with markers rendering them autotrophic, the pellet was resuspended in 100µl of sterile dH20, where 10% and 90% of the suspension was plated on the appropriate selective media. For antibiotic resistance markers, e.g. *G418* or *hphNT*, the pellet was resuspended in 1ml of non-selective medium and grown for a further 3-4hrs to allow time for expression of the antibiotic resistance gene, before plating on the appropriate selective media.

2.1.5 Dilution spotting of yeast strains

The desired strains were streaked from -80°C glycerol stocks and grown for 2-3 days at 24°C on non-selective medium to obtain single colonies. Single colonies were diluted in 1ml of dH_20 to a final concentration of $5x10^6$ cells/ml, then serially diluted 10-fold to produce suspensions of $5x10^5$, $5x10^4$ and $5x10^3$ cells/ml. 10µl of each cell

suspension was then spotted onto 100mm square plates (Sterilin) in a grid-like configuration, left to dry and then grown at the desired temperatures. YP agar plates of the desired conditions were freshly made, including different sugars, addition of 0.5mM IAA, or the presence of various concentrations of the genotoxic compounds MMS and HU (as in table 2.2). Every 24 hours, plates were scanned (800dpi, 8-bit greyscale, Epson V700) and the image saved as a TIFF file.

2.1.6 Direct repeat recombination assay

Cells were transformed with the pL and pLYΔNS plasmids (Mischo et al., 2011) as described in section 2.1.4 and plated on -uracil plates to select for the plasmid (plates were supplemented with extra leucine to prevent premature activation of the leucine cassette). 8 individual clones per strain/plasmid were then selected, resuspended in dH20 and counted. The dilutions of the cells were plated on both -leucine plates to select for recombinants and -uracil plates. Cells were then grown at 24ºC until individual colonies were suitably large for counting. The experiments were performed in triplicate and are presented as a ratio of the number of recombinants (*LEU2+*) over the total number of cells carrying a plasmid (*URA+*).

2.1.7 Harvesting of yeast strains for IPs

Overnight yeast cultures were diluted to between 0.3-0.4x10^7 cells/ml in a volume of either 250mls (dilute samples) or 1L (concentrated samples) at 24ºC. For asynchronous samples, either haploid or diploid cells were grown to a density of 2x10^7 cells/ml for harvesting. For synchronous cultures, *MAT*a haploid cells were grown to a density of 0.7x10^7 cells/ml and subsequently arrested in G_1 using α-factor pheromone to a final concentration of 7.5μg/ml for at least 3 hrs. After the first 90mins, α-factor was added every 30mins to a final concentration of 3.25μg/ml to maintain cells in G_1. When fully arrested, G_1 cells were harvested. Alternatively, arrested cells were washed twice with fresh YPD media (containing no α-factor) and released for 30 mins at 24ºC for S phase (20mins at 37ºC). FACS samples were taken at each stage to ensure cells were collected in the correct phase of the cell cycle.

For any strains with constructs controlled under the inducible *GAL1* promoter, cells were first grown overnight in YPRaf and diluted to 0.3-0.4x10^7 cells/ml.

Asynchronous samples were grown to $1.0x10^7$, then induced for 2hrs in YPGal before harvesting. For synchronous samples, cultures grown to $0.7x10^7$ cells/ml in YPRaf were arrested in G_1 by the addition of α-factor as described above, then switched for 35 mins to YPGal medium containing alpha factor to maintain cells in G_1 while inducing expression of the constructs. Cells were then washed and released in fresh YPGal media. Similarly, strains carrying an auxin degron tagged allele and the auxin receptor *TIR1* under control of the *GAL1* promoter were arrested in YPRaf as described, then shifted to YPGal in the presence of alpha factor for 35 minutes to allow expression of *TIR1*. 0.5mM IAA was then added for a further hour to degrade the protein of interest. Cells were washed and released into YPGal medium supplemented with 0.5mM IAA. For strains with a temperature sensitive degron allele and the E3 ligase *UBR1* under control of the *GAL1* promoter, cells were arrested in YPRaf as previously described, then shifted to YPGal supplemented with alpha factor for 35mins to induce expression of *UBR1*, before shifting to 37ºC for 1 hour to degrade the degron tagged allele. Cells were then washed twice with fresh YPGal and released into S phase at 37ºC. The cultures were harvested 20mins following release.

To harvest, cell pellets were collected by centrifugation at 3,000rpm for 3mins at 4ºC, and washed twice, first with a solution of 20mM Hepes-KOH [pH 7.9] (Sigma), then with a solution of 100mM Hepes-KOH [pH 7.9] (Sigma), 50mM potassium acetate (Fischer), 10mM magnesium acetate (Sigma) and 2mM EDTA-KOH at 4°C (henceforth referred to as 'popcorn lysis buffer'). For dilute samples, cell pellets were resuspended at a volume three times greater than their mass in popcorn lysis buffer supplemented with 2mM glycerophosphate (Johnson Matthey), 2mM sodium fluoride (Fischer), 1mM DTT, 1% (v/v) Sigma protease inhibitor cocktail (Sigma) and 0.24% (w/v) EDTA-free Complete Protease Inhibitor Cocktail (Roche). For concentrated samples, the pellets were re-suspended at a quarter volume of their mass in popcorn lysis buffer, supplemented with 8mM glycerophosphate, 8mM NaF, 1mM DTT, 4% (v/v) Sigma protease inhibitor cocktail and 0.48% (w/v) EDTA-free Complete Protease Inhibitor Cocktail. All steps and solutions were maintained at 4ºC. The cell suspensions were thoroughly vortexed to homogeneity, and pipetted drop-wise into liquid nitrogen to snap-freeze. Once all of the liquid nitrogen had evaporated, the cell 'popcorn' was stored at -80ºC for later use. To alter the stringency for the

immunoprecipitation, the salt concentration of the above buffers was adjusted by altering the KoAC concentration of the popcorn lysis buffer.

2.1.8 Harvesting of yeast cells for FACS/TCA

Yeast strains were inoculated overnight in YPD media at 24ºC, then diluted down to a density of 0.3-0.4x10^7 cells/ml in the appropriate volume of fresh media. Once the cultures reached a density of 0.7x10^7 cells/ml (approximately 2 hrs later), asynchronous samples were collected and the cells were arrested in G_1 by the addition of α-factor to a final concentration of 7.5µg/mL for 3 hours (adding 3.75µg/mL α-factor every 40mins after 1.5hrs). To release from G_1, cells were washed 2x with fresh YPD media, so no trace of α-factor remained. For experiments carried out at non-permissive temperature, cells were arrested in G_1 as described, and then shifted to 37ºC for 1 hour in the presence of 7.5µg/ml of α-factor. Then, cells were washed 2x with pre-warmed YPD to release at 37ºC. For strains harbouring an auxin degron tagged allele and the galactose inducible *TIR1* auxin receptor, cells were grown in YPRaf and arrested as previously described. Cells were then induced in YPGal supplemented with α-factor for 35mins. Auxin was then added to a final concentration of 0.5mM for a further 1hr at 24ºC. Cells were washed in fresh YPGal supplemented with 0.5mM auxin and released into S phase in this same media.

For fluorescence activated sell sorting (FACS) samples, 1 ml of cells (~ 10^7) were aliquoted and collected by centrifugation at 3000xg for 30 sec. The supernatant was discarded, and the cell pellet fixed by vortexing in 1ml of 70% ethanol (v/v). Cells were the stored at 4ºC until processing. For trichloroacetic acid (TCA) protein precipitation from cycling cultures, 10ml of cultures were centrifuged at 3000xg at 4ºC for 3 mins, then washed with 1ml of ice-cold PBS. The cells were then collected by centrifugation and resuspended in 300µl of 20% TCA (w/v) at -20ºC until processing. For testing the expression of a new construct by TCA analysis, a single colony was inoculated in 5ml of YPD and grown overnight, and the pellet processed as described above.

2.1.9 Preparation of cells and formaldehyde fixation for microscopy

Strains harbouring the *RAD52-GFP* allele were grown to a density of 0.7x10^7 cells/ml and arrested in G_1, using α-factor as previously described. For experiments carried out

at non-permissive temperature (37°C), cells were arrested at 24°C, and then maintained in G_1 at 37°C for 1 hour, before being released by washing twice with fresh pre-warmed medium. Following release from G1 block, cells were harvested at various time points during the cell cycle. Cells were fixed by adding 900µl of 16% paraformaldehyde to 900µl of the cycling culture to give a final concentration of 3% (v/v) and incubating at room temperature for 10mins on a rotating wheel. The cell pellets were collected by centrifugation at 3000rpm for 2mins, then washed twice with cold PBS, before resuspending in 1ml of fresh PBS. Cell suspensions were then stored at 4°C overnight in the dark. To minimise any GFP signal loss, imaging was carried out 24 hours or less following cell fixation. Samples were re-suspended in 500µl of fresh PBS supplemented with 1µg/ml of DAPI for 10 mins. The cells were then spun down at 3000rpm for 3mins and resuspended in a small volume of PBS to achieve a desirable cell density, then spotted onto a glass slide before applying a coverslip. The cells were imaged using an Olympus IX71 (brightfield, ~ 510 nm emission (GFP), ~460 nm (DAPI)). The images were processed and analysed using Fiji (2.0.0-rc-68/1.52g), and the number of Rad52-foci counted using adobe photoshop CC (19.1.7). The number of repeats is shown in the figure legends. Population means were compared using t-tests and the P values for statistical significance are reported in the figures.

2.1.10 R loops chromosome spreads

The protocol was performed as described in (Appanah et al., 2020, Wahba et al., 2011, Grubb et al., 2015). Asynchronous cultures were diluted to 0.5-0.7x10^7 cells/ml and grown to exponential phase in YPD at 30°C for 2 hours (~1.5x10^7 cells/ml). A total of 2x10^8 Cells were harvested, then spheroplasted in a solution of 0.1M potassium phosphate [pH 7.4], 1.2M sorbitol, 0.5mM MgCl2, 40mM DTT, supplemented with 20U zymolyase (zymo research #E1005) for 1 hour at 30°C. Sheroplasting was defined as complete when >90% of cells lysed following the addition of 2% sarcosyl. To halt the spheroplasting reaction, cells were then washed and resuspended in an ice-cold solution of 1M sorbitol [pH 6.4], 0.1M MES, 0.5mM MgCl2, 1mM EDTA. Subsequently, 20µl of cell suspension was dropped onto a slide, followed by the addition of 40µl of fixative (4% paraformaldehyde (w/v), 3.4% sucrose (w/v)), then lysed by the addition of 80µl of 1% lipsol (v/v), swirling the slide for 2 mins. 80µl of

fixative was then added, and after briefly swirling to mix, the chromosomes were spread homogenously across the surface of the slide using the side of a glass pipette, then left to dry overnight. Prior to immunostaining, slides pre-treated with RNase H were incubated for 1hr at 37°C in a humidity chamber with 4U of RNase H (Invitrogen #18021071) diluted in 400µl of 5mg/ml BSA. As a control, untreated slides were also incubated in the humidity chamber for 1hr at 37°C. Slides were then immunostained for DNA:RNA hybrids in a room temperature humidity chamber for 1 hour, using a 1:2,000 (0.25µg/ml) dilution of mouse monoclonal antibody S9.6 (kerafast #ENH001) in blocking buffer (5% BSA, 0.2% milk, 1XPBS). After washing 2x10mins in PBS, a 1:2,000 dilution of Cy3-conjucated goat anti-mouse (Jackson laboratories #115165003) in blocking buffer was added to the slides and incubated in a dark humidity chamber for 1 hour at room temperature. After a further 2x 10min washes in PBS, 50µl of in situ duolink mounting medium with DAPI (Sigma DUO82040) was added to the slide, a coverslip applied, and the edges sealed using clear nail varnish. Slides were then stored in the dark overnight at 4°C. A deltavision 1 microscope with a 100×/NA 1.4 objective was used to observe indirect immunofluorescence. Image analysis was performed using Fiji (2.0.0-rc-68/1.52g) and adobe photoshop CC (19.1.7). The figures show an average of three biological repeats.

2.1.11 R loops slot blot for quantification of R loops

Overnight cultures in YPD liquid medium were diluted to a density $0.35x10^7$ and grown for approximately 2 hours to $0.7x10^7$ cells/ml, then arrested in G1 as previously described. Cells were harvested 30mins after release, during S phase. The genomic DNA was extracted as described in section 2.4.1 (note, using nuclease free TE and RNase free dH20). To determine the concentration of genomic material, the absorbance of the samples was measured at 260nm. Serial dilutions of each sample (1µg/µl, 0.5µg/µl and 0.25µg/µl) were prepared in nuclease free dH20. 2µl of each dilution was treated with 1U of RNase H (Invitrogen, #18021071) with or without 1U of RNase III (to digest any double stranded RNA (Invitrogen, #AM2290)). These showed similar results after incubation at 37°C for 1hr. The untreated samples were also incubated at 37°C for 1hr as a control. 200µl of SSC2X hybridisation buffer (0.3M NaCl, 30mM trisodium Citrate [pH 7.0]) was added to the remaining DNA, and then transferred using a slot blot manifold under vacuum to a pre-equilibrated hybond-N+

nylon membrane (GE healthcare, #RPN203B), before UV crosslinking the DNA to the membrane. The membranes were incubated in blocking buffer (5% (w/v) milk for anti S9.6, 5% (w/v) BSA for anti-ds-DNA) at 24°C for 1 hr, then incubated at 4°C overnight with the appropriate primary antibody diluted 1:2,000 in the respective blocking buffer (S9.6 (Kerafast #ENH001), anti dsDNA (abcam #ab27156)). After washing 3x10mins with TBS-T, the membranes were incubated with anti-mouse IgG-HRP at a concentration of 1:5,000 for 1 hr at 24°C. Membranes were then washed 3x10mins with TBS-T, incubated with ECL solution (GE healthcare) for 90seconds, before removing excess ECL and visualising the chemiluminescent signal using a geldoc (G:BOX, Chemi XRG, Syngene).

Table 2.2 List of media used in this study

Media	Recipe
YPD	1% (w/v) yeast extract (Bacto), 2% (w/v) peptone (Oxoid), 2% (w/v) sugar (glucose, raffinose or galactose), (2% (w/v) Formedium agar for solid agar plates). YPD was supplemented with either 0.2 mg/ml Geneticin (G418, Invitrogen) or 0.3 mg/ml HygromycinB (Hygromycin B Gold, InvivoGen) to select for antibiotic resistance tagged genes. For experiments using the genotoxic agents MMS or HU, YPD was supplemented with 0.005-0.075% MMS (Sigma) or 25-100mM HU (Sigma). For auxin experiments, IAA was added to a final concentration of 0.5mM (where the auxin stock solution was dissolved in ethanol).
RSM	0.25% (w/v) yeast extract (Bacto), 1.5% (w/v) K(C2H3CO2), 0.1% (w/v) sugar (glucose, raffinose or galactose), 2.5% (v/v) amino acid mix
Amino acid deficient	0.17% (w/v) yeast nitrogen base (Difco), 0.5% (w/v) NH4SO4, 0.2% (w/v) glucose, 0.2% (w/v), Kaiser SC single Drop-out (Formedium), (2% (w/v) Formedium agar for solid agar plates).
Amino Acid mix	0.4% (w/v) adenine, 0.2% (w/v) arginine 0.4% (w/v), histidine, 0.2% (w/v) leucine, 0.2% (w/v) lysine, 0.2% (w/v) methionine, 1% (w/v) phenylalanine, 0.2% (w/w) tryptophan, 0.08% (w/w) tyrosine
Lysogeny broth (LB)	1% (w/v) bacto-tryptone, 0.5% (w/v) yeast extract (Bacto), 1% (w/v) NaCl [pH 7.0], (2% (w/v) agar (Formedium) for solid agar plates). To select for antibiotic resistance, this medium was supplemented with 100 µg/ml ampicillin or 50 µg/ml kanamycin.

2.2 *E. coli* methods

2.2.1 Preparation of competent DH5α cells

A small amount of DH5α 25% (v/v) glycerol stock was grown on LB agar plates at 37°C overnight (~16hrs) to obtain single colonies. 277.5ml of SOC media was inoculated with 20-30 colonies of the bacteria, and grown for 22-24hrs at 25°C, shaking. Once the cultures reached mid log phase (OD_{600} of 0.6), the cultures were incubated on ice for 10 mins and then centrifuged at 1200xg for 10 mins at 4°C. Following removal of the supernatant, cells were resuspended in 80ml of ice-cold TB buffer (10mM PIPES pH7.9, 15mM CaCl2, 250mM KCl, 55mM MnCl2, [pH 6.7]), then placed on ice for a further 10 mins. The cell suspension was then centrifuged as

previously described, and the pellet resuspended in 10ml of TB buffer supplemented with DMSO to a final concentration of 7%. Following another 10 min incubation on ice, the cell suspension was divided into 200µl aliquots and snap frozen in liquid nitrogen for storage at -80°C for future use.

2.2.2 Transformation of chemically competent *E. Coli*

An aliquot of chemically competent DH5α cells were taken from -80°C and left to thaw on ice for 5mins. 100µl of the cell suspension per reaction was aliquoted into a pre-chilled Eppendorf. The DNA to be transformed was then added (plasmid or ligation product (typically 5ng) and tap-mixed, then incubated on ice for 60-90mins. The cells were heat shocked for 2 mins at 42°C, then cooled on ice for a further 5 mins. To allow expression of the antibiotic resistance gene, cells were then incubated shaking at 37°C for 30mins with 0.5ml of pre-warmed LB or SOC medium. Typically, 10% and 90% dilutions of the cells were then plated on LB agar supplemented with the appropriate antibiotic and grown at 37°C for ~16 hrs.

2.3 Molecular Biology

2.3.1 Tagging or deletion of genes

A PCR based approach was used for the tagging and deletion of yeast genes (fig 2.1). For the deletion of genes, primers were designed to amplify various expression cassettes encoding autotrophy or antibiotic resistance, with contiguous sequences of homology to regions immediately flaking the gene of interest. For the tagging of genes, the protein tag was PCR amplified, and regions of the primers were designed with homology to the desired integration site. The PCR products were transformed using the lithium acetate method, and successful transformants were selected for by growth on the appropriate selective media and sequencing.

2.3.2 Checking genotype by PCR

To check correct integration of a tag or deletion of a gene, a PCR based approach was used. Oligos with homology either upstream or downstream of the intended loci as well as primers with homology within the marker cassette were used (fig 2.1, green

arrows). The PCR products were analysed by agarose gel electrophoresis (0.8% agarose gel) and in some cases, the bands were excised and purified for subsequent sequencing.

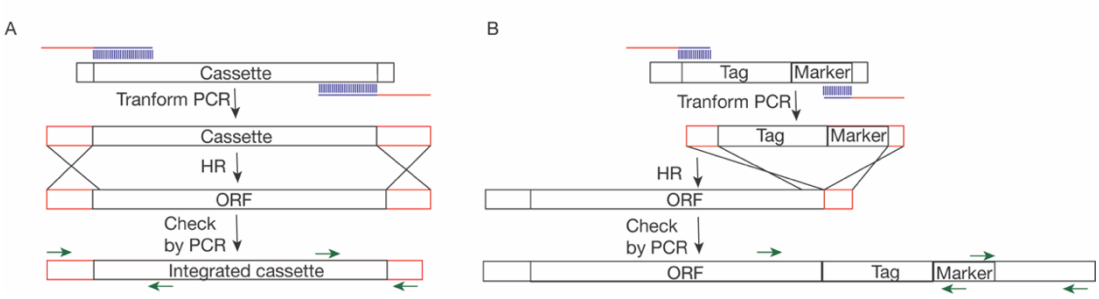


Figure 2.1 Schematic of checking tagging or deletion of a yeast gene by PCR

2.3.3 Cloning inducible constructs of *CTF4* and *MRC1*

Primers were designed to amplify the desired regions of either *CTF4* or *MRC1* from genomic DNA with a SalI restriction site added at the 5' end and PspXI restriction site to the 3' end. The fragments were then cloned into the pCS25 plasmid (pRS305-*GAL1-3HA*) via restriction digest at the aforementioned sites and transformed into chemically competent DH5α cells. Plasmid DNA was then extracted from the transformants, analysed by test digestion then sequenced. Positive plasmids were linearised by a single digestion at the *LEU2* marker and transformed into diploid yeast. Integration of the constructs at the ectopic *LEU2* loci was confirmed by extracting genomic DNA and checking by PCR as previously described. Expression of the constructs was also tested via TCA protein extraction and polyacrylamide gel electrophoresis.

2.3.4 Generation of *CTF4* mutants

Mutants were designed by selecting conserved residues within the Ctf4 fragment found to bind Sen1 N-terminal with the highest affinity. The DNA of the mutants was synthesised and cloned into the pMA-T plasmid (GeneART Gene synthesis, Thermofisher). Plasmid pCS169 (*GAL1-3HA*-Ctf4 (351-927)) was digested using the restriction enzymes SalI and XhoI to cut out the wild type Ctf4 (351-927) fragment. For each mutant, PCRs were carried out using a forward and reverse primer designed to amplify the mutation containing region from the DNA sequence, with homology to pCS25 to tag the mutant fragment with 3HA under the *GAL1* promoter. The cut plasmid, mutant PCR and a second PCR containing the remaining wild type sequence

of Ctf4 after the mutated region (as detailed in fig 2.2) were mixed together with NEBuilder HiFi Assembly mastermix, according to manufacturer's instructions. In brief, in a single tube reaction, the 5' ends of the DNA fragments are chewed back by exonucleases, and the 3' overhangs anneal at the regions of homology to be filled by DNA polymerase and ligated together. The assembled product was then transformed into chemically competent DH5α cells and plated on selective medium. The correct constructs were identified by diagnostic digests and sequencing.

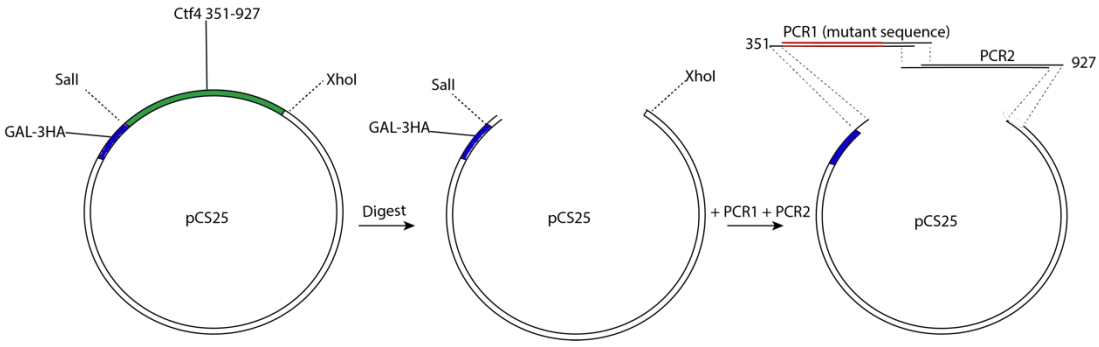


Figure 2.2 Schematic of PCR approach used to generate the *CTF4* mutants

Table 2.3 List of plasmids used in this study

Plasmid	Insert	Backbone	Purpose	Source
pcs25	*GAL1-3HA-Ø*	pRS305	Integration of *GAL1-3HA-Ø* into the yeast genome at the *LEU2*, and as a backbone for cloning various Mrc1 and Ctf4 fragments.	Lab collection
pCS216	*GAL1-3HA-CTF4 (426-782)*	pRS305-*GAL1-3HA* (pCS25)	Integration of *GAL1-3HA-CTF4 (426-782)* into yeast genome at *LEU2*	This study
pCS217	*GAL1-3HA-CTF4 (2-782)*	pRS305-*GAL1-3HA* (pCS25)	Integration of *GAL1-3HA-CTF4 (2-782)* into yeast genome at *LEU2*	This study
pCS168	*GAL1-3HA-CCTF4 (2-927)*	pRS305-*GAL1-3HA* (pCS25)	Integration of GAL1-3HA-CTF4 (2-927) into yeast genome at LEU2	This study
pCS169	*GAL1-3HA-CTF4 (351-927)*	pRS305-*GAL1-3HA* (pCS25)	Integration of *GAL1-3HA-CTF4 (351-927)* into yeast genome at *LEU2*	This study
pCS170	*GAL1-3HA-CTF4 (426-927)*	pRS305-*GAL1-3HA* (pCS25)	Integration of *GAL1-3HA-CTF4 (426-927)* into yeast genome at *LEU2*	This study
pCS171	*GAL1-3HA-CTF4 (783-927)*	pRS305-*GAL1-3HA* (pCS25)	Integration of *GAL1-3HA-CTF4 (783-927)* into yeast genome at *LEU2*	This study
pCS172	*GAL1-3HA-MRC1 (2-219)*	pRS305-*GAL1-3HA* (pCS25)	Integration of *GAL1-3HA-MRC1 (2-219)* into yeast genome at *LEU2*	This study
pCS173	*GAL1-3HA-MRC1 (2-418)*	pRS305-*GAL1-3HA* (pCS25)	Integration of *GAL1-3HA-MRC1 (2-418)* into yeast genome at *LEU2*	This study

pCS174	*GAL1-3HA-MRC1 (2-655)*	pRS305-*GAL1-3HA* (pCS25)	Integration of *GAL1-3HA-MRC1 (2-655)* into yeast genome at *LEU2*	This study
pCS175	*GAL1-3HA-MRC1 (2-843)*	pRS305-*GAL1-3HA* (pCS25)	Integration of *GAL1-3HA-MRC1 (2-843)* into yeast genome at *LEU2*	This study
pCS176	*GAL1-3HA-MRC1 (2-1096)*	pRS305-*GAL1-3HA* (pCS25)	Integration of *GAL1-3HA-MRC1 (2-1096)* into yeast genome at *LEU2*	This study
pCS177	*GAL1-3HA-MRC1 (140-1096)*	pRS305-*GAL1-3HA* (pCS25)	Integration of *GAL1-3HA-MRC1 (140-109)* into yeast genome at *LEU2*	This study
pCS178	*GAL1-3HA-MRC1 (312-1096)*	pRS305-*GAL1-3HA* (pCS25)	Integration of *GAL1-3HA-MRC1 (312-1096)* into yeast genome at *LEU2*	This study
pCS179	*GAL1-3HA-MRC1 (567-1096)*	pRS305-*GAL1-3HA* (pCS25)	Integration of *GAL1-3HA-MRC1 (567-1096)* into yeast genome at *LEU2*	This study
pCS180	*GAL1-3HA-MRC1 (752-1096)*	pRS305-*GAL1-3HA* (pCS25)	Integration of *GAL1-3HA-MRC1 (752-1096)* into yeast genome at *LEU2*	This study
pCS230	*GAL1-3HA-MRC1 (418-843)*	pRS305-*GAL1-3HA* (pCS25)	Integration of *GAL1-3HA-MRC1 (418-843)* into yeast genome at *LEU2*	This study
pCS196	*hRNH1*	pRS423	Expression of human RNH1 under control of the GPD promoter	Gift from Domenico Libri
pCS271	*GPD-Empty*	pRS423	Control plasmid with GPD promoter for hRNH1 expression experiments	Lab collection
pCS1344	*3x mini IAA*	pMK152	3x mini auxin degron C-terminal tagging	Lab collection
pCS1345	*3x mini IAA*	pMK153	3x mini auxin degron C-terminal tagging	Lab collection
pCS231	*Ctf4 L351G A352G V355A V356G*	pMA-T	Gibson cloning of Ctf4 mutants	This study, GeneART (sigma)
pC232	*Ctf4 D376A S377G E278A S379G D380A L381G*	pMA-T	Gibson cloning of Ctf4 mutants	This study, GeneART (sigma)
pCS233	*Ctf4 E385A N387A D388G D389G N391A K392A D394G*	pMA-T	Gibson cloning of Ctf4 mutants	This study, GeneART (sigma)
pCS234	*Ctf4 E403A A404G N405A A406G E407A D408A V409A F410A*	pMA-T	Gibson cloning of Ctf4 mutants	This study, GeneART (sigma)
pCS235	*Y426A F428A E429G D4230A E4231G E4232A D4233G I4235G D4236A D4237A D4238A D4239A*	pMA-T	Gibson cloning of Ctf4 mutants	This study, GeneART (sigma)
pCS236	*Ctf4 K447A K448A*	pMA-T	Gibson cloning of Ctf4 mutants	This study, GeneART (sigma)
pCS237	*GAL-3HA-Ctf4 S456G H459A S463G*	pMA-T	Gibson cloning of Ctf4 mutants	This study, GeneART (sigma)
pCS238	*GAL-3HA-Ctf4 L351G A352G V355A V356G*	pCS25	Cloning of Ct4 mutants under control of GAL1 promoter with 3HA tag	This study
pCS239	*GAL-3HA-Ctf4 D376A S377G E278A S379G D380A L381G*	pCS25	Cloning of Ct4 mutants under control of GAL1 promoter with 3HA tag	This study
pCS240	*GAL-3HA-Ctf4 E385A N387A D388G D389G N391A K392A D394G*	pCS25	Cloning of Ct4 mutants under control of GAL1 promoter with 3HA tag	This study
pCS241	*GAL-3HA-Ctf4 E403A A404G N405A A406G E407A*	pCS25	Cloning of Ct4 mutants under control of GAL1 promoter with 3HA tag	This study

pCS242	*GAL-3HA Ctf4-Y426A F428A E429G D4230A E4231G E4232A D4233G I4235G D4236A D4237A D4238A D4239A*	pCS25	Cloning of Ct4 mutants under control of GAL1 promoter with 3HA tag	This study
pCS243	*GAL-3HA-Ctf4 K447A K448A*	pCS25	Cloning of Ct4 mutants under control of GAL1 promoter with 3HA tag	This study
pCS244	*GAL-3HA-Ctf4 S456G H459A S463G*	pCS25	Cloning of Ct4 mutants under control of GAL1 promoter with 3HA tag	This study

Table 2.4 List of Oligos used in this study

Oligo Number	Purpose	5'-3' sequence
CS291	Tagging/gene deletion of the sen1 gene in W303 (Saccharomyces cerevisiae) for CIP tag	TATACACCAATATATATGCAGGTATAATTCCTAACACTTTTAC TTCAAGATCAATCGATGAATTCGAGCTCG
CS292	Tagging/gene deletion of the sen1 gene in W303 (Saccharomyces cerevisiae) for CIP tag	CGGAATGCTTCATCTAGCCCATTTATCCCAAAAAAAAGAAA GCCTAGATCACGTACGCTGCAGGTCGAC
CS1079	GAL3-HA-CTF4 C-terminal fragments 2_frw	ttactgtcgacGTTTCAGTTATAGACAAGCTTG
CS1080	GAL3-HA-CTF4 C-terminal fragments 927_rev	aatatcctcgaggTTATTTCAATTGCTGTTCATATC
CS1081	GAL3-HA-CTF4 C-terminal fragments 426_frw	ttactgtcgacTTGGCAGAGTCCGTAGTATC
CS1082	GAL3-HA-CTF4 C-terminal fragments 783_frw	ttactgtcgacAACTTCGAAGATGAAGAAGA
CS1083	GAL3-HA-CTF4 C-terminal fragments 2_frw	ttactgtcgacAAGAGTAAATTACTAGAGGAAAA
CS1090	Cloning GAL-3HA-MRC1 fragments 2_frw	ttactgtcgacGATGATGCCTTGCATGCTTTG
CS1091	Cloning GAL-3HA-MRC1 fragments 219_rev	aatatcctcgaggtcaAGGAATTTTCGGCGACAGA
CS1092	Cloning GAL-3HA-MRC1 fragments 418_rev	atatcctcgaggtcaTCATCATCTCCATATTCGTC
CS1093	Cloning GAL-3HA-MRC1 fragments 655_rev	atatcctcgaggtcaGTTTTGTGTATCTAGAACAT
CS1094	Cloning GAL-3HA-MRC1 fragments 843_rev	atatcctcgaggtcaTGGGTTTTTAACAAGCTTTG
CS1095	Cloning GAL-3HA-MRC1 fragments 1096_rev	aatatcctcgaggCTAATTATCAAAGCTATCTTG
CS1096	Cloning GAL-3HA-MRC1 fragments 140_frw	ttactgtcgacGTACCTATCCATTCCGTTAAT
CS1097	Cloning GAL-3HA-MRC1 fragments 312_frw	ttactgtcgacGAATACAAAAAACCGCAAAAGC
CS1098	Cloning GAL-3HA-MRC1 fragments 567_frw	ttactgtcgacTATGAATCATCTGGTAGTGAAA
CS1099	Cloning GAL-3HA-MRC1 fragments 752_frw	ttactgtcgacATCGATGATTATTCCAAGAAC
CS1115	make construct to tag N-terminally with TPR domain from Dia2	CGGCCGCTCTAGAACTAGTGGATCCcATTAAGGCGCGCCA GATCTG
CS1116	make construct to tag N-terminally with TPR domain from Dia2	GCCTCGAGGCCAGAAGACta
CS1117	make construct to tag N-terminally with TPR domain from Dia2	cacctcttaGTCTTCTGGCCTCGAGGCAAATTTTTTGTACTTTTA ATGATAAAA
CS1118	make construct to tag N-terminally with TPR domain from Dia2	GCCTGAATTCCCTGGGGAAGACATTTATAATAAACAGATGC GCTTT
CS1119	make construct to tag N-terminally with TPR domain from Dia2	ATGTCTTCCCCAGGGAATTC
CS1120	make construct to tag N-terminally with TPR domain from Dia2	CGAGGTCGACGGTATCGATAAGCTTAAAATTTTTAAAATTTA TTGGTGCC
CS1121	tag N-terminally with TPR domain from Dia2 to SEN1	GTATGGCATCTATCTCTATATATATAAAAAAGCGCATCTGTTT ATTATAAcATTAAGGCGCGCCAGATC
CS1122	tag N-terminally with TPR domain from Dia2 to SEN1	TTATTATTATTAATGTTGTTGCTATTATTATTATCAGGATTGTT GGAATTAAAATTTTTAAAATTTATTGGTGC
CS1123	tag N-terminally with TPR domain from Dia2 to SEN1	CGAAAGCAAATCTTGACAAAG

CS1124	tag N-terminally with TPR domain from Dia2 to SEN1	GTGAGTCCTTTTTATAGGAAAG
CS1125	tag N-terminally with TPR domain from Dia2 to SEN1	GCATATTAACCTACTAGAAAAG
CS1126	tag N-terminally with TPR domain from Dia2 to SEN1	ATACGATGTACAAGGACGTTG
CS1229	Cloning Gal-HA CTF4 fragments 782_rev	aatatcctcgaggTTAAACAAATACTGGCATTCTAA
CS1243	delete rad27 5', KanMX cassette from pYM18	ATATATGCCAAGGTGAAGGACCAAAAGAAGAAAGTGGAAAAAGAACCCCCatcgatgaattcgagctcg
CS1244	delete rad27 3', KanMX cassette from pYM18	CAGCATACATTGGAAAGAAATAGGAAACGGACACCGGAAGAAAAAATATGattaaggcgcgccagatctg
CS1267	frw tag Mcm7 Ct with IAA, hphNT cassette	CTTCCGCTAATGTGAGCGCCCAAGATTCTGATATCGATCTACAAGACGCTcgtacgctgcaggtcgac
CS1268	rev tag Mcm7 Ct with IAA. hphNT cassette	GGGAAAAAGAATAAAGAATGAAGGCCCTGTTGCTTTTTTTTTTAGAACTTatcgatgaattcgagctcg
CS1285	Cloning GAL-3HA-MRC1 fragments 418_frw	ttactgtcgacGATATGGATAGCATTAAATTATCC
CS1286	making CTF4 mutants, Gibson cloning	CTGGTGCCGGTGCTGGTGCCGGTGCCGGTGtcg
CS1287	making CTF4 mutants, Gibson cloning	GAAATTTTCCTGTGTTTGCCAAACTGATTGGAA
CS1288	making CTF4 mutants, Gibson cloning	TTCCAATCAGTTTGGCAAACac
CS1289	making CTF4 mutants, Gibson cloning	GGGAACAAAAGCTGGGTACCGGGCCCCCCCtcga

Table 2.5 Sequence of the CIP and Dia2-TPR tags fused to *sen1-3* N-terminal to attempt to recapitulate replisome binding

Tag	Sequence
CIP	GAGCAGAAGCTGATTAGCGAGGAAGATCTGGGAGGCTCAGGCGGGCTCCGTAATATAGATGATATTCTCGATGATATAGAGGGCAGCGGCGGATCTATCGATAATTTTGATGACATTCTTGGTGAGTTTGAATCCGGTGGTTCAGGTGACATTAACATAGACGATATTCTTGCAGAACTGGATTAA
Dia2-TPR	ATGTCTTCCCCAGGGAATTCAGGCGTGGCTATAGATTCTACCGTTCTGAAAGCAATTGAATTAGGAACGAGGCTTTTTAAATCTGGAGAATATTTGCAAGCGAAAAGGATATTCACAAATGCTTTGAGAGTATGTGATTCTTATTCTCAGGAGCAGATTATGCGCATTAGGAACGCGTATCAATTAGATACTGCGAGACCGGATAACAAACGATTATATCATCCTAGATACATAAAGATATTGGATAACATCTGTGCGTGCTACGAAAAATTGAATGATTTGAAATCTTGTTTGGACGTATCACAAAGGTTACTTAAGTTAGAACCAGGTAATATCAAATGCTATATACGATGTACAAGGACGTTGATAAAATTAAAAGACTGGAAAAGGGCATATAAAACATGTTCTCGTGGACTGCAATTATGCAATAACGATAGCAATCATCTAAGACAACAAAAACAATTCATTAAAAATAACATGGTTCAAAAACAGGACGGCAAGAGAAGTTATATAGATCCATTAGAGGAGACCAAAATAGCAAAAAAAAAAAAGAATAATAATGTTCTAGAATCGTTACCAAAGAAGAAGATTAAAGGTAGTACCAAGAAAACTGATTTAGTTGGCAATTTGCCGATAGAAATACTACCTATCATATTCCAAAGGTTTACCACTAAGGAGCTTGTTACGTTGTCGCTAGTTTGTAACAAATGGAGGGACAAAATTTTGTATCATTTAGACTGCTTCCAAGAATTCAACTTGGCACCAATAAATTTTAAAAATTTT

2.4 Biochemistry

2.4.1 Genomic DNA extraction

Overnight cultures of single colonies inoculated in 5ml of YPD were collected by centrifugation and washed with 1ml of sterile dH20, then 1ml of yeast lysis buffer (10 mM Tris-HCl [pH 7.5], 1 mM EDTA, 3% (w/v) SDS). After washing, cells were resuspended in 200µl of yeast lysis buffer, followed by addition of 200µl sterile dH20 and 200µl of phenol/chloroform/isoamyl alcohol (25:24:1) (Acros organics). The cells were lysed by the addition of a small amount of 0.5mm glass beads, with vigorous

vortexing for 2 mins. After centrifuging at 12,000xg for 5 mins, 380µl of the aqueous phase was then transferred to a tube containing 760µl of 100% ethanol to precipitate the DNA. The DNA was then recovered by centrifugation as previously described, and the DNA pellet washed with 1ml of fresh 70% ethanol (v/v) to remove any contamination by residual phenol/chloroform. After removing the supernatant, the pellet was then dried for 30 mins at 37°C, resuspended in 50µl of TE buffer [pH 8.0] supplemented with 50µg/ml RNase A and incubated at 37°C for 2 hrs. The purified DNA was then stored at -20°C.

2.4.2 TCA protein extraction

Cells pellets were collected as previously described in section 2.1.8. After thawing, the TCA sample was thoroughly resuspended, and approximately 300µl of 0.5mm glass beads was added. Cells were lysed mechanically by vigorous vortexing for 5 mins. After quickly spinning down in a microfuge, the supernatant and any protein-precipitates were transferred to a fresh Eppendorf tube, and 300µl of 5% TCA (w/v) was added. The solution was then centrifuged at 3000xg for 10 mins, and the protein pellet thoroughly resuspended in 200µl of high pH laemmli buffer (1x laemmli buffer supplemented with 150mM Tris base), before boiling for 4 mins. Following centrifugation at 3000xg for 10 mins, the supernatant was transferred to a fresh Eppendorf tube and stored at -20°C.

2.4.3 FACS sample processing and analysis

200-500µl of the ethanol fixed cells was added to 1ml of 50mM Na citrate buffer and vortexed, before spinning down at 3,000xg for 3mins to remove the supernatant. The cell pellet was washed twice more with 50mM Na citrate buffer, then resuspended in 500µl of Na citrate supplemented with 0.1mg/ml RNase A. The suspension was incubated at 37°C for 2 hours. The cell pellet was then collected as previously described, and resuspended in 500µl of 50mM HCl, supplemented with 5mg/ml of pepsin and incubated at 37°C for 30mins. Subsequently, the supernatant was removed, and the pellet resuspended in 50mM Na citrate with 2µg/ml of propidium iodide, then stored in the dark at 4°C overnight. The cell suspension was then sonicated 2x 5

seconds at 8 microns, and vortexed to homogeneity. The cell suspension was run using the Becton Dickinson FACScan flow cytometer. 40,000 cells were acquired per sample, gated to accept populations of cells between G_1 and G_2 phase. A 488nm laser was used to induce excitation of the sample, and a 650nm long pass filter was used to detect the emission.

2.4.4 Immunoprecipitation of TAP tagged proteins

For the immunoprecipitation of TAP tagged proteins, anti-sheep IgG conjugated beads were used. These were prepared in advance using M-270 Epoxy Dynabeads (Invitrogen) resuspended in dimethyl formamide. 425µl of Dynabead slurry was washed twice with 0.1M sodium phosphate [pH 7.4] for 10mins on a rotating wheel. The beads were then incubated with 425µl of 0.1M sodium phosphate [pH 7.4], 300µl of 3M ammonium sulphate (in a solution of 0.1M sodium phosphate) and 300µg of Anti-Sheep IgG from rabbit (Sigma S1265) for 2 days at 4°C on a rotating wheel. The beads were than washed four times with PBS, once with PBS/0.5% NP-40 for 10 mins, followed by two washes with 5 mg/ml BSA (Sigma). (For long term storage, the beads were stored in 5 mg/ml BSA supplemented with 0.02% sodium azide to prevent any bacterial growth, which was removed by washing in PBS/BSA before use in IP experiments).

Dilute and concentrated yeast popcorn were prepared as previously described in section 2.1.7. The frozen cells were lysed to a powder at -80°C using a cryomill (Spex Sample Prep, 6870) filled with liquid nitrogen, grinding 4-6 times for 1 minute at 14 cycles per second. The powdered cell lysate was thawed to 4°C, where 1g of powdered mass was considered equivalent to 1ml for calculation of the subsequent volumes of buffer to add.

For dilute IP samples, one quarter of the volume plus 50µl of glycerol buffer (50% (v/v) glycerol, 50mM Hepes-KOH [pH 7.9], 50mM KoAC, 50mM MgOAC, 0.5% Igepal® CA-630 (Sigma)) was added, supplemented with 2mM EDTA, 2mM β-glycerophosphate, 2mM sodium fluoride, 1mM DTT, 1% (v/v) Sigma protease inhibitor cocktail and 0.24% (w/v) EDTA-free Complete Protease Inhibitor Cocktail (Roche). Pierce Universal Nuclease (ThermoFisher) was added to a final concentration

of 0.4 U/µl, and the samples were rotated at 4ºC for 30 minutes to degrade any DNA and RNA in sample.

For concentrated samples, a quarter volume of the above glycerol buffer was added, supplemented with 8mM EDTA, 8mM β-glycerophosphate, 8mM sodium fluoride, 8mM DTT, 4% (v/v) Sigma protease inhibitor cocktail and 0.96% (w/v) EDTA-free Complete Protease Inhibitor Cocktail (Roche). In addition, 1ml of a solution of popcorn lysis buffer (as detailed in section 2.1.7), supplemented with 8mM β-glycerophosphate, 8mM NaF, 8mM DTT, 4% (v/v) Sigma protease inhibitor cocktail and 0.96% (w/v) EDTA-free Complete Protease Inhibitor Cocktail was added to reduce viscosity of the lysate. Pierce Universal Nuclease (ThermoFisher) was added to a final concentration of 0.8 U/µl and rotated at 4ºC for 30 minutes.

The cell lysate was then clarified by centrifuging at 4ºC, first at 8,700xg for 30 mins, followed by a further 1 hour at 126,600xg under vacuum in an ultra-centrifuge. The supernatant fraction between the pellet of cellular debris and the lipid layer was collected using a needle and syringe to puncture through the tube wall, and 50µl boiled in 100µl of 1.5x Laemmli buffer as the whole cell extract sample. The remainder of each cell extract sample was split between 2 aliquots of 100µl of the pre-prepared anti-Sheep IgG conjugated Dynabeads. After rotating for 2 hours at 4ºC, the beads were washed once with a solution of 100mM HEPES-KOH [pH 7.9], 50mM KOAc, 50mM MgOAc, 2mM EDTA, 0.1% (v/v) Igepal® CA-630, supplemented with 2mM β-glycerophosphate, 2mM NaF, 1mM DTT, 1% (v/v) Sigma protease inhibitor cocktail and 0.24% (w/v) EDTAfree Complete Protease Inhibitor Cocktail, then twice with the same solution without the inhibitors. Each aliquot of beads was then resuspended in 50µl of 1x Laemmli buffer and boiled for 4 minutes. The supernatant was then collected and snap frozen on dry ice and stored -80ºC for subsequent western blotting.

For dephosphorylation IPs, the extract was obtained as described above. The extract for each strain was split equally and added to 6 aliquots of TAP beads for incubation. Following this, the beads were washed once with a solution of 100mM HEPES-KOH [pH 7.9], 50mM KOAc, 50mM MgOAc, 2mM EDTA, 0.1% (v/v) Igepal® CA-630, supplemented with 2mM β-glycerophosphate, 2mM NaF, 1mM DTT, 1% (v/v) Sigma protease inhibitor cocktail and 0.24% (w/v) EDTAfree Complete Protease Inhibitor

Cocktail, then three times in the above solution without any inhibitors. The beads were then washed once with 300µl of PMP reaction buffer (NEB) and $MnCl_2$ (prepared according to manufacturer's instructions). Two tubes of beads were subjected to mock treatment with 50µl of PMP reaction buffer/$MnCl_2$. Two tubes of beads were incubated in 50µl of buffer with 400U of Lambda Protein Phosphatase (NEB P0753S), and the remaining two tubes with Lamba protein phosphatase + phosphatase inhibitors (50mM NaF, 20mM $NaVO_4$). The beads were incubated at 30°C for 30 mins shaking gently. For each pair of tubes, the supernatant was removed and one set of beads was boiled immediately in 50µl of 1x Laemmli buffer, while the other was washed 4x with 100mM HEPES-KOH [pH 7.9], 50mM KOAc, 50mM MgOAc, 2mM EDTA, 0.1% (v/v) Igepal® CA-630, before boiling in 50µl of 1x Laemmli buffer. For IPs of different stringencies, the salt concentration was altered by adjusting the concentration of potassium acetate in all of the above buffers.

2.4.5 Detection of proteins by Immunoblotting

According to the size of the proteins of interest, 6, 8, 10 or 12% denaturing polyacrylamide (National Diagnostics) gels were prepared. Protein samples were electrophoretically separated at 110V in 1X Tris/Glycine buffer, then transferred to pre-equilibrated nitrocellulose blotting membranes (Amersham) using semi-dry transfer apparatus at 13V for 90 mins. For TCA samples, membranes were stained in 0.1% (w/v) Ponceau S in 5% (v/v) acetic acid solution as a loading control, washed with dH20 and scanned. Membranes were blocked with a solution containing 5% (w/v) skimmed milk dissolved in TBS-T for 1 hr at room temperature to reduce non-specific binding. After blocking, the membranes were incubated with the appropriate concentration of primary antibody diluted in blocking buffer at 4°C on a shaking platform overnight. Following this, the membranes were subjected to 3x 10-minute washes in TBS-T, and then incubated for 1 hour at room temperature in the appropriate secondary antibody diluted in 5% (w/v) milk. Subsequently, the membranes were washed 3x for 10 minutes in TBS-T and then incubated with ECL solution (GE healthcare) for 2mins. Any excess ECL was then removed, and the membrane sealed in plastic. Chemiluminescent films (Amersham Hyperfilm™, GE Healthcare) were exposed to the membranes for various lengths of time in a dark room to detect the chemiluminescent signal, before developing. The films were then scanned using an Epson V700 scanner at 300dpi in 8-bit greyscale and saved as a TIFF file. For band

quantification, the bands were analysed using ImageJ. The appropriate bands were selected as a region of interest (ensuring the signal was not saturated) and the intensity was measured using the analyse tool. The bands were then normalised to the intensity of a specified band e.g. the TAP and the ratio plotted in Microsoft Excel.

Table 2.6 Primary antibodies used in this study

Primary Antibody	Host	Typical Concentration used	Source
Anti-TAP-HRP (PAP35)	Rabbit	1:2,000 – 1:100,000	Sigma (M4439)
Anti-HA (12CA5)	Mouse	1:500 – 1:1,000	Sigma (11583816001)
Anti-Mrc1	Mouse	1:500 - 1:1,000	
Anti-Ctf4	Sheep	1:30,000	Gifted by K. Labib
Anti-RNAPII	Mouse	1:1,000	Novus Biologicals (NB-200-598)
Anti-Cdc45	Sheep	1:2,500	Gifted by K. Labib
Anti-Mcm6	Sheep	1:20,000	Gifted by K. Labib
Anti-Mcm5	Sheep	1:1,000	Gifted by K. Labib
Anti-Mcm3	Sheep	1:1,000	Gifted by K. Labib
Anti-Pol1	Sheep	1:1,000	Gifted by K. Labib
Anti-Pol2	Sheep	1:1,000	Gifted by K. Labib
Anti-Psf1	Sheep	1:500	Gifted by K. Labib
Anti-Dpb2	Sheep	1:1,000	Gifted by K. Labib
Anti-Csm3	Sheep	1:1,000	Gifted by K. Labib
Anti-Rad53	Mouse	1:1,500	Abcam (EL7.E1)
Anti-DNA:RNA hybrid S9.6	Mouse	1:1,000-1:2,000	Kerafast ENH001
Anti-c-myc	Mouse	1:1,000	Sigma (M4439)

Table 2.7 Secondary antibodies used in this study

Secondary Antibody	Host	Typical Concentration used	Source
Anti-Sheep IgG-HRP	Donkey	1:5000 – 1:10,000	Sigma (A3415)
Anti-Mouse IgG-HRP	Horse	1:10,000	Cell Signalling Technology (#7076)
Anti-Rabbit IgG-HRP	Mouse	1:1,000	Rockland (18-8816-33)
Cy3-conjucated goat anti-mouse	Goat	1:2,000	Jackson laboratories #115165003

Chapter 3: Results

INVESTIGATING THE IMPORTANCE OF SEN1 REPLISOME TETHERING FOR THE RESOLUTION OF R LOOPS

3.1 Background

During S phase, DNA serves as the substrate for two vital cellular processes, transcription and replication. As they compete for the same template; inevitably, replication forks and transcription bubbles will collide with one another. If unresolved, these collisions can lead to genomic instability. An additional barrier to replication is the stable three stranded nucleic structures, termed R loops. These form as a result of transcription; where the nascent RNA reanneals with the template DNA behind the elongating polymerase, displacing the non-template strand. Clashes in the head-on orientation between forks and transcription complexes are the most deleterious and are thought to promote the formation of stable R loops, further blocking the fork (Aguilera and García-Muse, 2012, Hamperl et al., 2017, Lang et al., 2017, Garcia-Rubio et al., 2018).

Importantly, replication is limited to S phase and high-fidelity duplication of the cells entire genetic content within this window is essential to maintain genome stability. Conversely, as transcription occurs throughout interphase; it has been hypothesised that during S phase, the progression of replication forks is favoured by the cell. To this end, it is likely that at sites of collisions, mechanisms have evolved to quickly resolve any R loops and disengage the transcription machinery from the DNA, clearing the way for forks. In this scenario, the key players would need to be either quickly recruited to the required sites, or it is tempting to speculate they may even travel with replication forks or transcription bubbles for their local enrichment.

Previously, it has been reported that the DNA:RNA helicase Sen1 plays a vital role in protecting the genome from R loop mediated DNA damage and instability by removing these extended DNA:RNA hybrids (Mischo et al., 2011, Costantino and Koshland, 2018). A mutation within the catalytic domain of Sen1 (the *sen1-1* allele) leads to increased levels of R loops. This mutant also presents a hyperrecombination phenotype, increased genetic instability, and a delayed S phase with terminal arrest in G2. Its viability also depends on several DNA repair proteins. This suggests it plays some important role during DNA replication. Significantly, overexpression of RNase H, which specifically digests R loops, reduces the high levels of recombination. Finally, *sen1-1* is synthetic lethal in the absence of both of the RNase H enzymes, while any combination of the double mutants are viable (Mischo et al., 2011). Taken together, this suggests Sen1 may have some level of functional redundancy with these enzymes for the removal of R loops. Indeed, in cells carrying the *sen1-1* allele, Rnh1 is enriched on chromatin during both S phase and G2 at semi-permissive temperature (Lockhart et al., 2019). In addition, Sen1 is also responsible for terminating various species of short noncoding RNAs transcribed by RNAPII. However, though Sen1 appears to work in concert with the RNA binding proteins Nrd1 and Nab3 as part of the NNS complex at NNS terminators, *in vitro,* Sen1 activity alone is sufficient to dissociate RNAPII from DNA (Porrua and Libri, 2013, Han et al., 2017).

Sen1 has been shown to travel with replication forks, interacting with the replisome components Ctf4 and Mrc1 during S phase. This association is independent from RNAPII transcription (Alzu et al., 2012, Appanah et al., 2020). Due to its function in both R loops resolution and transcription termination, it is an ideal candidate to perform a role in clearing the way for forks at transcription-replication conflict sites. Our lab developed the *sen1-3* allele, which is a useful tool for investigating the role of Sen1 at forks. As the mutant is defective in Sen1 replisome binding activity, but importantly, is unaffected in its transcription termination function (see chapter 1.9.5,6) (Appanah et al., 2020), this enables us to investigate the phenotypic consequences of Sen1 loss from forks *in vivo,* while Sen1 catalytic activity remains unaffected.

Interestingly, similar to *sen1-1*; analysis of meiotic progeny revealed that the *sen1-3 rnh1*Δ *rnh201*Δ triple mutant is also synthetic lethal, though double mutants lacking either of the RNase H enzymes in combination *sen1-3* are viable (fig 3.1) (Appanah et

al., 2020). These genetic interactions suggest that there may be some level of functional redundancy between Sen1 and the RNase H enzymes specifically at replication forks. Alternatively, Sen1 activity and its tethering to the replisome is critical to overcome any defects resulting from the absence of RNase H and the increase in R loops. Considering this data, our lab is particularly interested in whether the biological relevance of Sen1 traveling with the replisome is to resolve R loops encountered by the replication machinery.

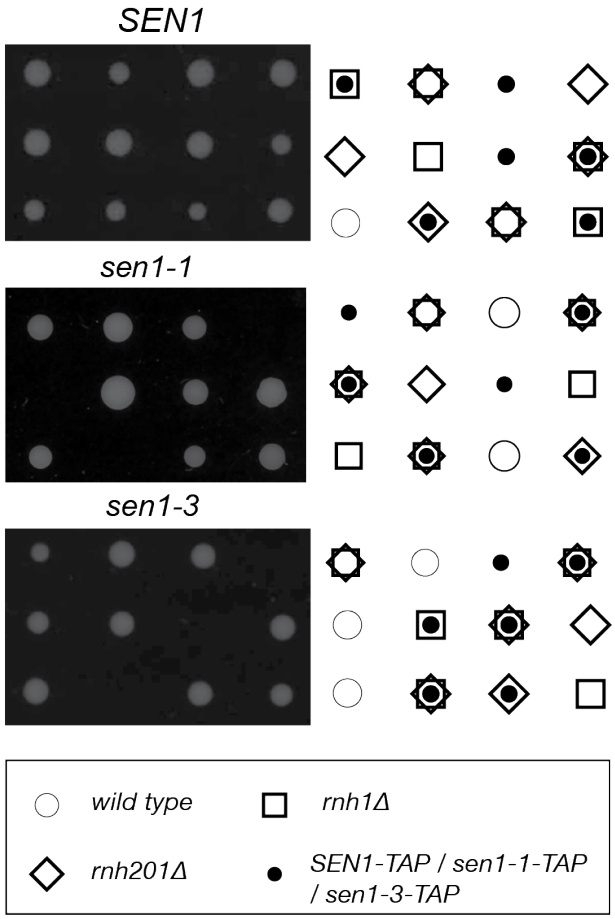


Figure 3.1 *The sen1-1* and *sen1-3* alleles are synthetic lethal when combined with the deletion of both RNase H enzymes. Analysis of meiotic progeny of haploid *SEN1/sen1-1/sen1-3* cells crossed with *rnh1Δ rnh201Δ* cells. The genotype of each allele is signified by the described symbols (Appanah et al., 2020).

Therefore, the work in this chapter aims to further understand the functional relevance of the Sen1 replisome tethering mechanism in light of its proposed role in R loop removal. Various experiments were carried out to further analyse the phenotype of *sen1-3* in the absence of RNase H, as well as with other RNA metabolism mutants, and assess the level of functional redundancy at the fork.

3.2 *sen1-3* is synthetic defective in the absence of the RNase H enzymes

Importantly, *sen1-3* cells exposed to replication stress via HU, which depletes dNTP pools leading to stalled and eventually collapsed forks, or DNA damage by alkylating agent MMS, do not present any growth defects. They also show no increased temperature sensitivity (fig 3.2a, 3.7) or abnormal DNA replication dynamics (fig 3.4) compared with wild type *SEN1* cells. This suggests that Sen1 may play a function at replication forks that is redundant with the activity of other factors. Considering the observed lethality of *sen1-3* in the absence of RNase H activity, we hypothesised the function of Sen1 at forks may be to remove R loops encountered by the replication machinery, perhaps in a partially redundant manner.

To investigate this genetic interaction further, we attempted to utilise the inducible auxin degron system to degrade the Rnh1 and Rnh201 proteins in cycling *SEN1* and *sen1-3* cells. This would allow us to analyse the effect of the triple mutant during a single cell cycle and observe if the defects leading to cell death are replication specific. In this system, addition of the plant hormone auxin induces proteasomal degradation of any degron tagged proteins. Auxin promotes an interaction between the degron and the Tir1 receptor, which is coupled to the E3 ubiquitin ligase SCF. This subsequently results in the tagged proteins polyubiquitylation and proteasomal degradation (Nishimura et al., 2009). *RNH1* and *RNH201* were tagged at their C-terminal with a 3x mini auxin degron (*AID*) and crossed into cells carrying either *SEN1* or *sen1-3* with the *TIR1* allele under control of the *GAL1* promoter.

Initial dilution spotting of these strains revealed that while *sen1-3 rnh1-AID rnh201-AID* cells were sick in the presence of galactose and auxin, they remained viable (fig 3.2a), suggesting that depletion of the desired proteins was incomplete. In addition, growth of the triple mutant was reduced in both glucose and galactose, with or without auxin. Defects in glucose and in the absence of auxin suggests that tagging of the enzymes rendered them less functional or stable (fig 3.2a). FACS analysis of these strains was carried out, which found that depletion of the RNase H enzymes resulted in slightly altered replication dynamics, where in the presence of auxin, cells displayed an accumulation in G2 and delayed exit from mitosis, which was further exacerbated

in the *sen1-3* background (fig 3.2b). However, as incomplete degradation of Rnh1 and Rnh201 could potentially mask any phenotype or render it less severe, and possible defects caused by tagging the target genes could make cells chronically sick, we decided not to continue with this approach.

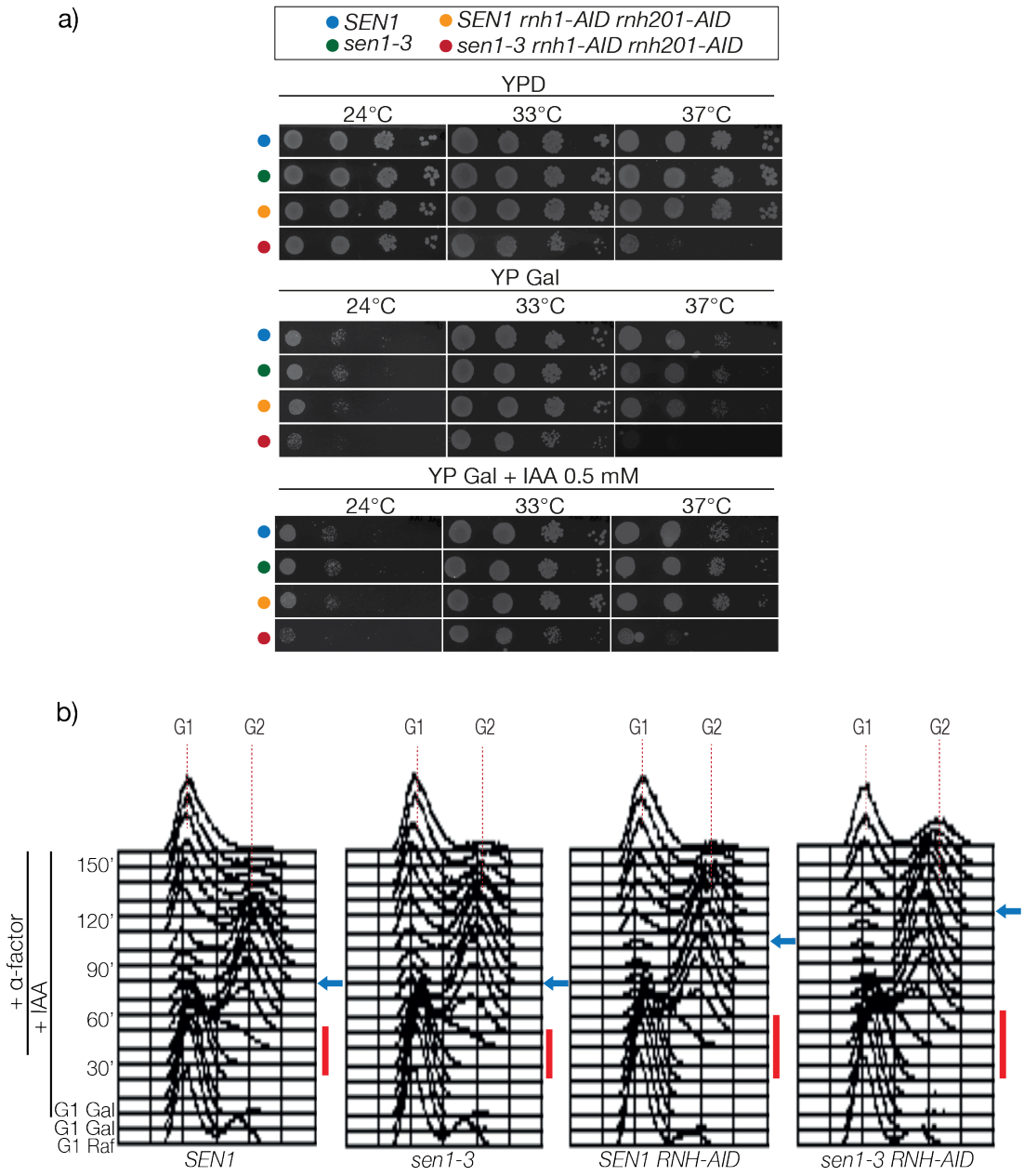


Figure 3.2 Using the Auxin inducible degron system to deplete the RNase H enzymes. A 3xmini *IAA* degron was used to tag *RNH1* and *RNH201*. Upon addition of 0.5mM auxin to the media, the tag binds the galactose inducible Tir1 E3-ubiquitin ligase, resulting in ubiquitination and proteasome mediated degradation of the protein. a) Strains were serially diluted 10-fold to obtain concentrations of $5x10^6$, $5x10^5$, $5x10^4$ and $5x10^3$ cells/ml, then 10µl of each suspension was spotted on YPD, or YPGal medium with or without 0.5mM IAA. Plates were imaged every 24 hours. In the presence of galactose and 0.5mM auxin, the triple mutant,

although sick, remained viable suggesting that depletion of the RNase H proteins was not complete. b) Cells were grown in YPRaf to a density of $0.7x10^7$ at 24ºC and arrested in G1 by the addition of alpha factor for 3 hours. Following arrest, cells were maintained in G1 and induced with YPGal for 35 mins, before supplementing the medium with 0.5mM IAA for 1 hr to degrade the IAA tagged Rnh1 and Rnh201. Cultures were then washed to remove any trace of alpha factor and released into S phase at 24ºC. Cells were collected every 10 mins and fixed in 70% (v/v) ethanol for FACS analysis. The red bar indicates the length of S phase and the blue arrow the beginning of mitotic exit. Strains used in this experiment: CS2808, CS2810, CS3126, CS3128.

To explore the nature of the lethality in *sen1-3 rnh1*Δ *rnh201*Δ cells, we attempted to test whether the synthetic defects could be rescued by re-establishing the interaction with the replisome. Hence, we cloned our *sen1-3* allele with a Ctf4 interacting peptide (CIP)-tag (table 2.5). Many of the Ctf4 replisome binding partners, such as pol α and Sld5 of the CMG helicase contain these CIP-boxes (Villa et al., 2016), so we reasoned that tagging with these motifs could recapitulate *sen1-3* binding to the replisome. In addition, we also fused a Dia2-TPR domain (table 2.5) to *sen1-3* N-terminal (which is responsible for fusing SCF to the replisome (Morohashi et al., 2009)). However, neither approach suppressed the lethality of the *RNH1 RNH201* deletions and was not analysed any further (data not shown).

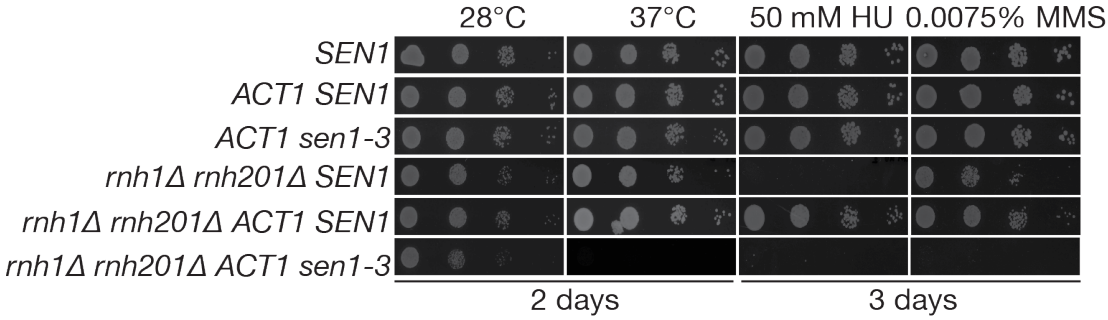


Figure 3.3 Overexpression of *sen1-3* by the strong constitutive *ACT1* promoter rescues the synthetic lethality of *sen1-3 rnh1Δ rnh201Δ* triple mutant. Serial dilution spotting of strains was carried out under conditions of replication stress, DNA damage, or on YPD medium grown at various different temperatures. The plates were imaged every 24 hours to assess growth. Strains used in this experiment: CS2808, CS2584, CS2607, CS2734, CS3168, CS2736.

Interestingly, overexpression of *sen1-3* under the control of the strong constitutive *ACT1* promoter is able to rescue the synthetic lethality of the *sen1-3* allele in a background lacking the RNase H enzymes, suggesting that increasing the global levels of the protein can compensate for loss of the mechanism whereby Sen1 specifically travels with the forks as part of the replisome (fig 3.3). Moreover, this poses the possibility that Sen1 could operate at regions with fork clashes or transcription

complexes independently from the replisome. However, cells overexpressing *sen1-3* in the *rnh1*Δ *rnh201*Δ background are temperature sensitive and analysis of their DNA replication dynamics at 37ºC by FACS shows the triple mutant cells accumulate in G2/M (fig 3.3 & 3.4).

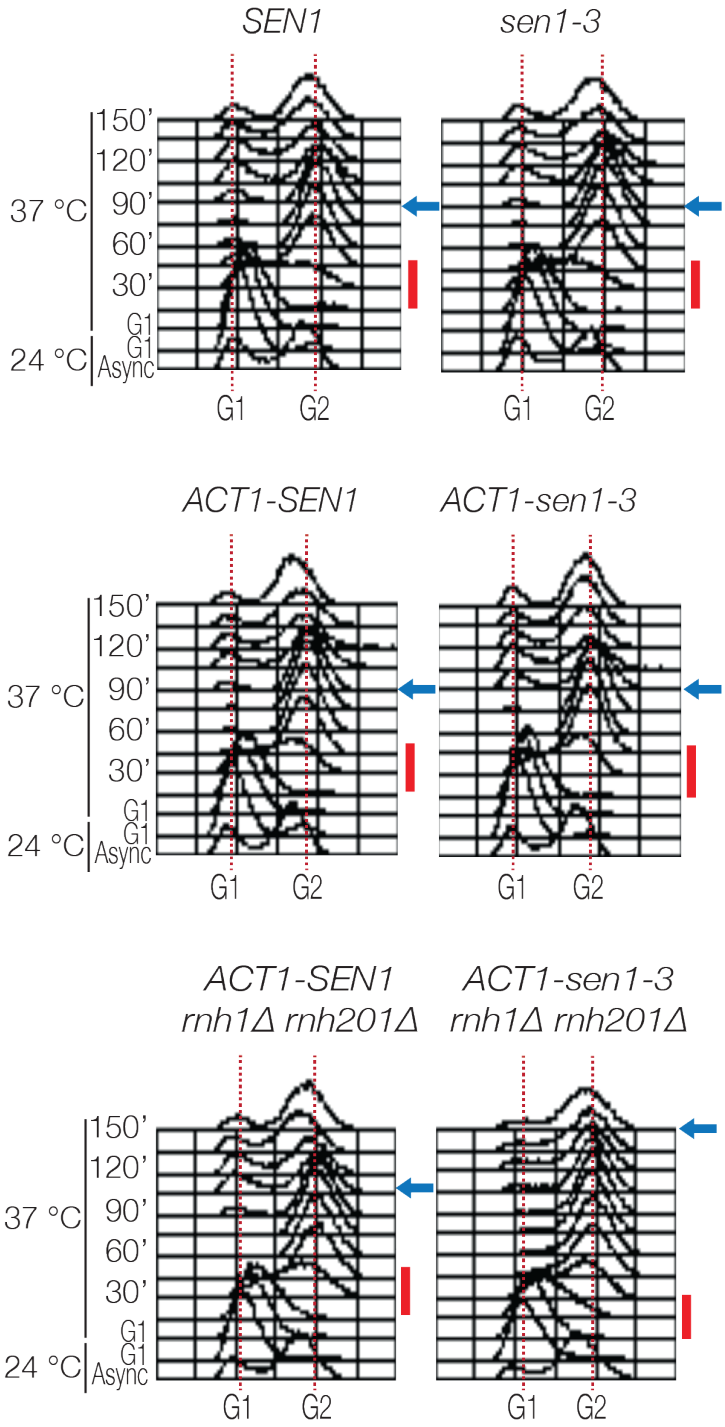


Figure 3.4 Cells overexpressing *sen1-3* in a *rnh1Δ rnh201Δ* background have altered DNA replication dynamics. a) Cells were grown to a density of 0.7x10^7 then arrested in G1 by the addition of alpha factor for 3 hours. Cultures were shifted to 37ºC for 1 hour before

release into S phase at non-permissive temperature. Samples were collected every 15mins and fixed using 70% (v/v) ethanol. The red bar indicates the length of S phase, and the blue arrow the beginning of the exit of the cell population from mitosis. The FACS profiles show that the triple mutant accumulates in G2/M. Strains used in this experiment: CS2808, CS2810, CS2584, CS2607, CS3168, CS2736.

Moreover, the *ACT1-sen1-3 rnh1*Δ *rnh201*Δ cells also show increased activation of the checkpoint response (fig 3.5). This is demonstrated by immunoblotting for hyperphosphorylation of the checkpoint kinase Rad53, involved in checkpoint arrest, stabilising stalled forks, induction of DNA repair genes and inhibition of late replication origin firing. Following DNA damage, Rad53 is phosphorylated/activated through Rad9 by the Mec1/Tel1 pathway, or following replicative stress independently from Rad9, via Mrc1 (Chen and Zhou, 2009).

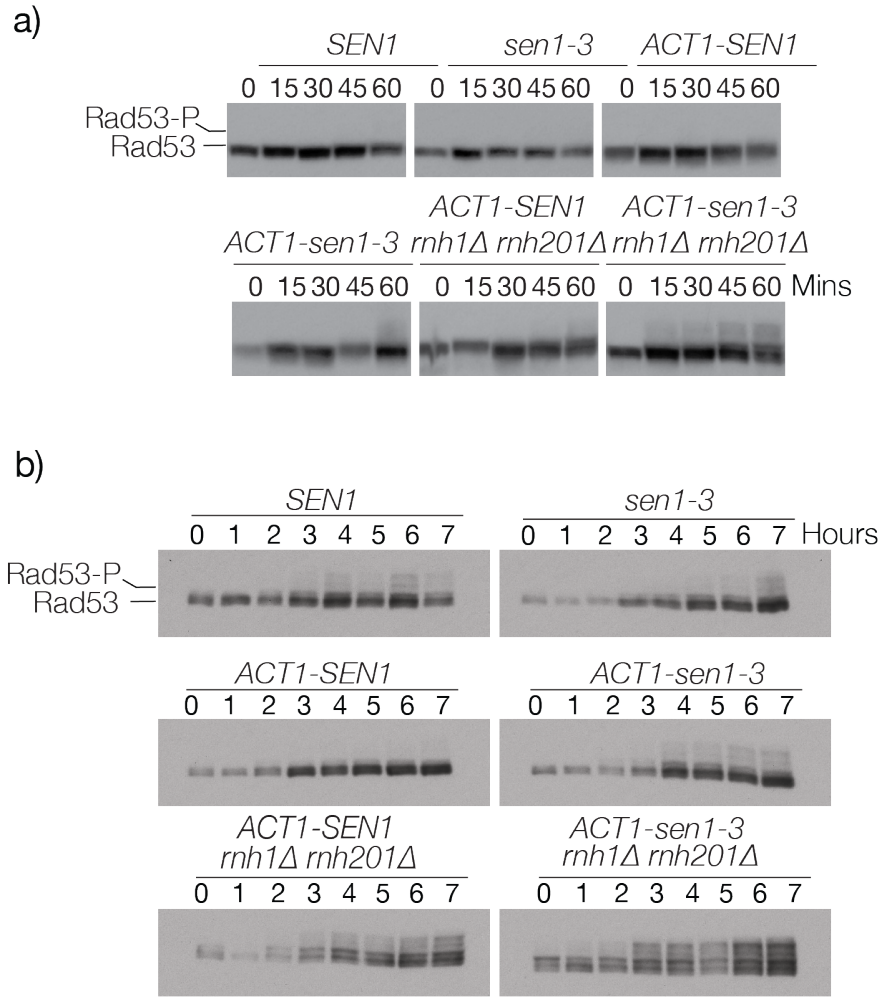


Figure 3.5 Cells overexpressing *sen1-3* in a *rnh1Δ rnh201Δ* background show increased activation of the checkpoint response compared with *SEN1* a) Protein extraction and

immunoblotting of the samples from figure 3.4 (where samples were taken at the described time points following release from G1 arrest at 37°C) show activation of the checkpoint kinase Rad53 during DNA replication. The higher degree of Rad53 hyperphosphorylation in the triple mutant was also observed in b) where asynchronously growing cells were analysed by TCA extraction and western blot for Rad53 activation over the course of several hours at 24°C. Strains used in this experiment: CS2808, CS2810, CS2584, CS2607, CS3168, CS2736.

Interestingly, the dilution spotting shows that overexpression of *SEN1* is able to suppress the hyper-sensitivity of *rnh1*Δ *rnh201*Δ cells to replication stress induced by HU, while *ACT1-sen1-3* is unable to do so. Moreover, loss of the specific replisome tethering mechanism sensitises the cells to MMS induced DNA damage (fig 3.3). Importantly, analysis of the levels of *ACT1-SEN1* and *ACT1-sen1-3* tagged with 3HA showed the observed phenotypes are not due to any differences in the level of protein expression (R. Appanah, G. De Piccoli) (fig 3.6).

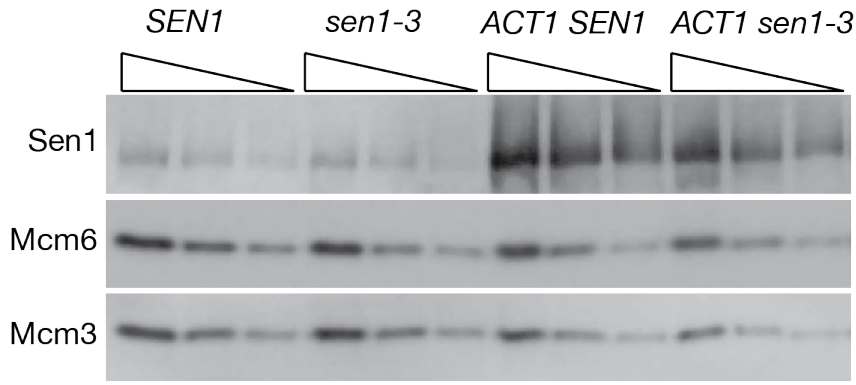


Figure 3.6 Analysing the protein expression levels of various *3HA* tagged *SEN1* and *sen1-3* alleles. In asynchronous samples grown at 24°C, the levels of Sen1 versus Sen1-3 protein expression was compared by immunoblotting for the 3HA tag. The loading controls Mcm3 and Mcm6 are also shown. (R. Appanah, G. De Piccoli) (Appanah et al., 2020). Strains used in this experiment: CS2808, CS2810, CS2584, CS2607.

3.3 *sen1-3* is synthetic defective with strains accumulating high levels of R loops

To further assess whether the synthetic defects found in *sen1-3* cells are the result of an accumulation of R loops, we crossed our *sen1-3* allele with *hpr1*Δ. Hpr1 is one of four components comprising the THO complex, which along with additional mRNA export factors becomes THO/TREX, involved in ensuring efficient mRNA metabolism and export (Chávez et al., 2000). This complex prevents R loop formation behind elongating RNAPII by stabilising the nascent RNA as messenger ribonucleoprotein particles (mRNP), and couples their efficient export with transcription elongation (Chávez et al., 2000). Correspondingly, *hpr1*Δ mutants have been shown to present defects in transcription elongation and gene gating, as well as exhibiting high levels of R loops due to the increased availability of the nascent RNA for reannealing to the template strand (Chávez et al., 2000, Bermejo et al., 2011, García-Benítez et al., 2017).

Dilution spotting of *SEN1* and *sen1-3 hpr1*Δ strains revealed that the double mutant is synthetic defective at elevated temperatures and under conditions of HU induced replication stress (fig 3.7). The suppression of a defect or phenotype by overexpression of *RNH1* is widely used as an indicator for R loops and interestingly, overexpression of *RNH1* under the *GAL1* promoter suppresses the growth defects of *sen1-3 hpr1*Δ cells back to the level of the single *hpr1*Δ mutant (fig 3.7). This suggests that accumulation of R loops does indeed contribute to the defects observed in this *sen1-3 hpr1*Δ background.

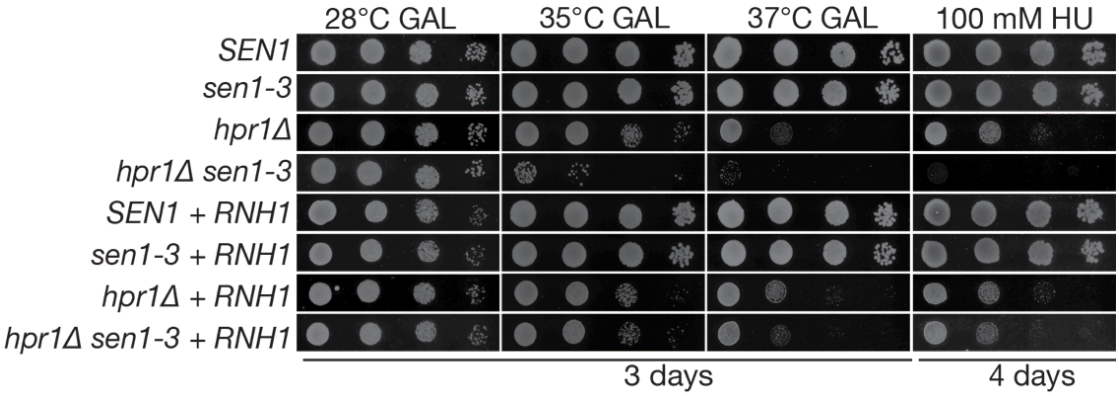


Figure 3.7 Synthetic defects of the *sen1-3* allele in a *hpr1Δ* background are suppressed by overexpression of *RNH1* to resolve any accumulating R loops. Dilution spotting of cells

carrying a *GAL1-RNH1* construct inserted at the *LEU2* locus and the relevant controls was carried out. Cells were plated on medium containing galactose to induce expression of the construct and grown at various temperatures or under conditions of replication stress by supplementing the media with HU. Plates were scanned every 24 hours to assess growth. Strains used in this experiment: CS2808, CS2810, CS3742, CS3741, CS3698, CS3700, CS3902, CS3904.

Next, we crossed a *RAD52-GFP* allele into *SEN1* and *sen1-3 hpr1*Δ cells to observe the induction/level of Rad52 foci formation in cultures synchronously released from G1 arrest. This allowed us to analyse whether the defects seen in the double mutant occur during S phase and are linked to DNA replication. Rad52 is a central DNA repair protein that re-localises to sites of DNA damage by binding ssDNA, forming nuclear foci. Rad52-GFP foci are widely used as an *in vivo* molecular marker for homologous recombination and the repair of double strand breaks (Lisby et al., 2001, Aylon and Kupiec, 2004).

These experiments were carried out at 28ºC, as *hpr1*Δ cells could not be released synchronously at non permissive temperatures (35ºC and 37ºC) due to budding defects. We observed that during the later stages of replication, the *sen1-3* allele stimulates a small but statistically significant increase ($p<0.01$) in recombination compared to *SEN1* (fig 3.8a,b). In addition, the double mutant *sen1-3 hpr1*Δ cells also show a statistically significant increase ($p<0.01$) in recombination compared with their *SEN1* counterpart (fig 3.8a,b). Similar to the *GAL1-RNH1* mediated suppression observed by dilution spotting (fig 3.7), the increase in recombination in *sen1-3 hpr1*Δ cells, but interestingly not the *sen1-3* single mutant, was reduced following overexpression of *RNH1* (fig 3.8a,b). This again indicates that accumulation of R loops is a source of toxicity in the *hpr1*Δ *sen1-3* background.

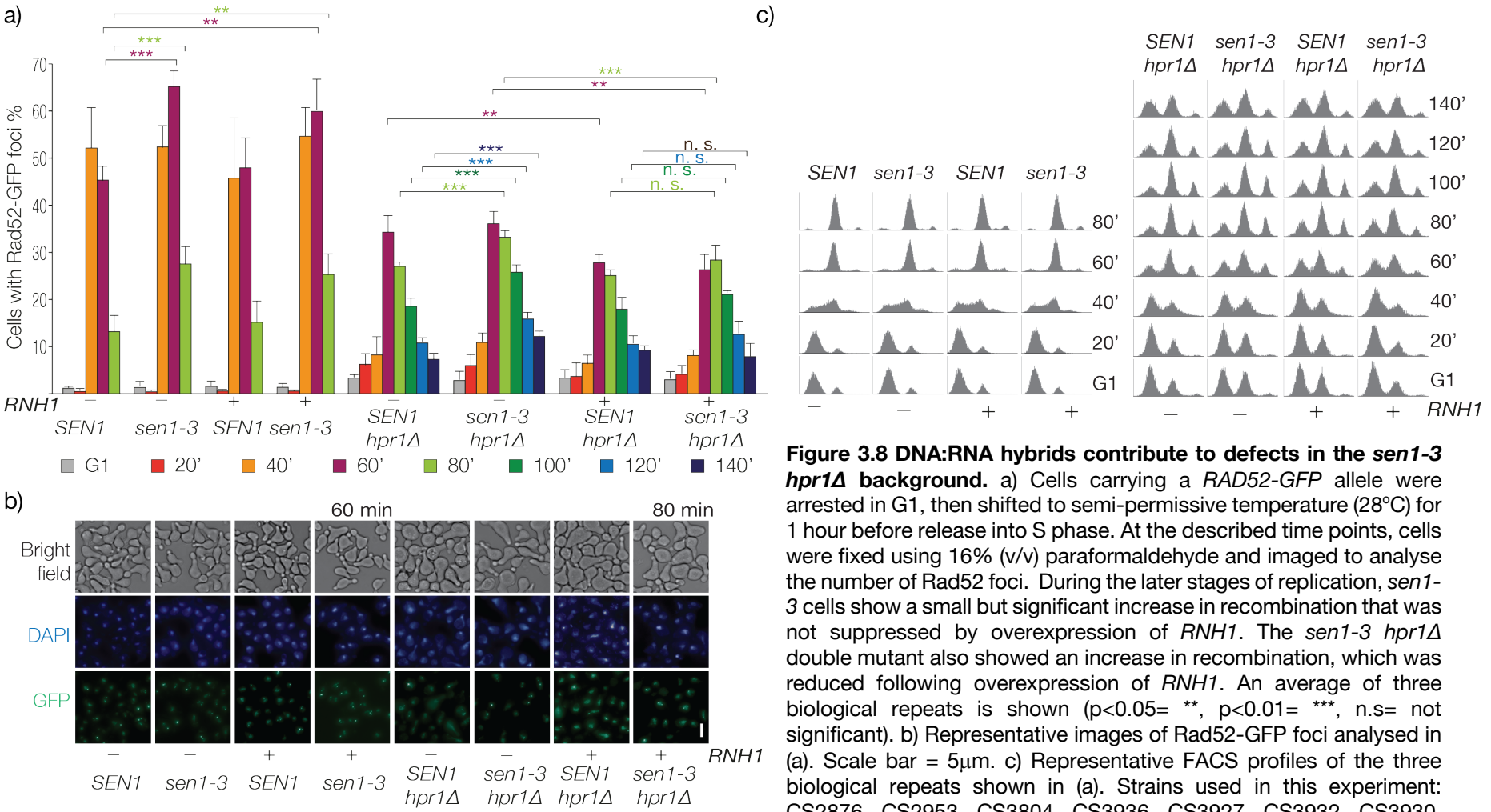


Figure 3.8 DNA:RNA hybrids contribute to defects in the *sen1-3 hpr1Δ* background. a) Cells carrying a *RAD52-GFP* allele were arrested in G1, then shifted to semi-permissive temperature (28°C) for 1 hour before release into S phase. At the described time points, cells were fixed using 16% (v/v) paraformaldehyde and imaged to analyse the number of Rad52 foci. During the later stages of replication, *sen1-3* cells show a small but significant increase in recombination that was not suppressed by overexpression of *RNH1*. The *sen1-3 hpr1Δ* double mutant also showed an increase in recombination, which was reduced following overexpression of *RNH1*. An average of three biological repeats is shown ($p<0.05$= **, $p<0.01$= ***, n.s= not significant). b) Representative images of Rad52-GFP foci analysed in (a). Scale bar = 5μm. c) Representative FACS profiles of the three biological repeats shown in (a). Strains used in this experiment: CS2876, CS2953, CS3804, CS3936, CS3927, CS3932, CS3930, CS3935.

3.4 Direct analysis of R loops accumulation in *sen1-3* mutants

Although, as previously mentioned, suppression of a defect or phenotype (such as recombination) by *RNH1* overexpression is a commonly used diagnostic for R loops, we also aimed to confirm the previous results by using a method to directly analyse and compare the levels of DNA:RNA hybrids within these strains.

A cytological chromosome spread assay was adopted, where cells were digested with the cell-wall degrading enzyme Zymolyase to generate spheroplasts. The cell and nuclear membranes were then lysed via osmotic shock aided by a detergent, in the presence of fixative to preserve protein and DNA:RNA structures. The lysed cells were spread across the surface of a slide, allowing the insoluble nuclear components to settle and bind homogenously (Grubb et al., 2015, Wahba et al., 2011). These spread chromosomes were then analysed for DNA:RNA hybrid levels by indirect immunofluorescence using the S9.6 antibody, which is directed against R loops. Before immunostaining; the slides were either mock treated or treated with commercial Rnh1 to confirm that the signal is specific to R loops. In addition, *rnh1*Δ *rnh201*Δ cells were used as a positive control; and as previously described by Wahba et al., displayed an extremely high level of R loops signal, which was significantly reduced following treatment with commercial RNase H1 (fig 3.9a,b) (Wahba et al., 2011). Surprisingly, however, we did not observe any increase in the number of cells with detectable DNA:RNA hybrids between *SEN1 hpr1*Δ cells or *sen1-3 hpr1*Δ cells (fig 3.9a,b).

Similarly, slot blot analysis to examine R loops *ex vivo* using the S9.6 antibody also did not appear to detect any difference between the two strains (fig 3.9c). In this experiment, the genomic material was extracted from S phase cells, then treated with RNase A to degrade any single stranded RNA in the samples. Serial dilutions were prepared, and initial test experiments treated the material with either RNase H to degrade any DNA:RNA hybrids, or RNase H plus RNase III (which cleaves dsRNA), to ensure the S9.6 antibody was not non-specifically recognising any double stranded RNA in the samples. These two conditions yielded similar results (data not shown). Therefore, serial dilutions of mock treated, and RNase H treated samples were UV

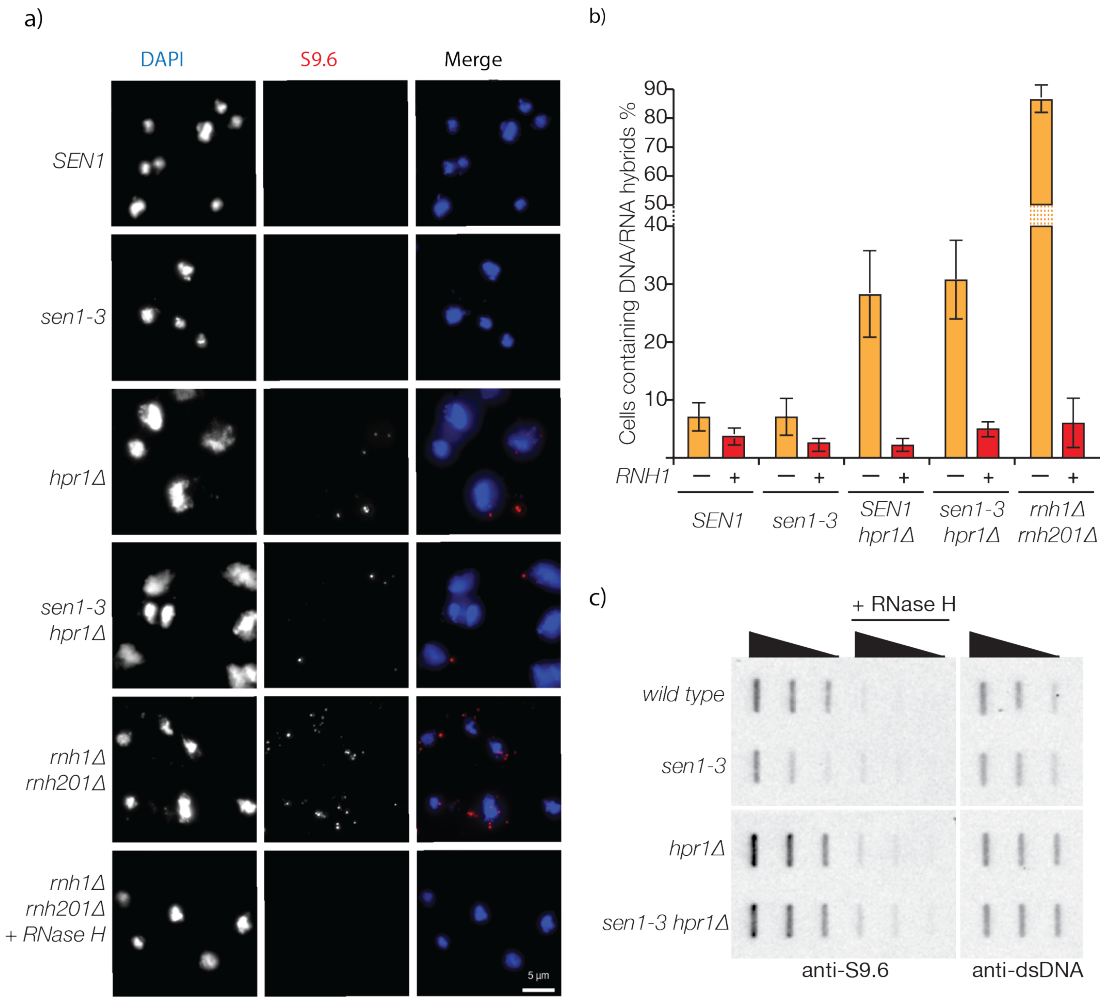


Figure 3.9 Accumulation of DNA:RNA hybrids in *sen1-3* may only be genotoxic at specific loci or under certain conditions a) Chromosome spreads were performed from cells grown to exponential phase at 28ºC for 2 hours. Prior to immunostaining, cells were either mock treated or pre-treated with RNase H1 at 37ºC for 1 hour as a control. R loops were then visualised by immunostaining with the S9.6 antibody directed against RNA:DNA hybrids followed by incubation with Cy3-conjugated secondary antibody. a) Representative images of the chromosome spreads. Scale bar = 5µm. b) High levels of signal were detected in a *rnh1Δ rnh201Δ* background, which was specifically reduced in Rnh1 treated cells, indicating the signal is R loop specific. No difference in the level of signal was detected between *SEN1* and *sen1-3* cells, even in an *hpr1Δ* background. An average of three biological repeats is shown c) To analyse R loops *ex vivo*, exponentially growing cultures were arrested in G1, then washed and released at 24ºC for 30 mins (S phase). The genomic material, including DNA:DNA and DNA:RNA hybrids was collected by phenol/chloroform extraction in nuclease free water. Serial dilutions (1, 0.5, 0.25µg/µl) of the samples were either mock treated or treated with RNase H1 at 37ºC for 1 hour. The samples were transferred under vacuum using a slot blot manifold and crosslinked to the nylon membrane. The membranes were then probed for DNA:RNA hybrids using S9.6, or dsDNA using anti-dsDNA antibody. The image is representative of three biological repeats. Strains used in these experiments: CS2808, CS2810, CS3742, CS3741, CS2734.

crosslinked to hybond-N+ nylon membranes, which were then probed with S9.6 and anti-dsDNA antibody. Spectrophotometric analysis confirmed correct serial dilution of the samples; however, unfortunately, the signal using the anti-dsDNA did not decrease in a linear fashion as expected, making it impossible to quantify the observed S9.6 R loops bands by normalising them to the dsDNA signal. Visually however, immunoblotting revealed a similarly increased S9.6 signal in both *hpr1*Δ strains compared with *SEN1* or *sen1-3* cells, consistent with the increased R loops described in this background and the results of our chromosome spreads. This signal was significantly reduced by RNase H treatment (fig 3.9c).

3.5 Testing the genetic interactions of *sen1-3* with other mutant backgrounds that display an enrichment of R loops

As the experiments in the previous section contrastingly showed no difference in the levels of R loops between *SEN1* and *sen1-3 hpr1*Δ cells compared with the Rad52-GFP foci experiments, we decided to analyse the genetic interactions of *sen1-3* with other genes that have a known role in R loop resolution, or deletion of which results in elevated hybrid levels. In some cases, the absence of these proteins results in the enrichment of R loops at specific genomic loci.

These genes included the DNA helicases *RRM3* or *PIF1,* which resolve R loops at *tDNAs* (Tran et al., 2017, Boulé and Zakian, 2007), and *SGS1*, which acts at long genes and hard-to replicate regions (Chang et al., 2017). *SRS2* on the other hand has been described as a novel anti-hybrid mechanism, which counteracts deleterious Rad51 R loop-forming strand invasion activity (Wahba et al., 2013). Finally, deletion of the enzyme *TOP1* (which resolves negative supercoiling behind elongating RNAPII) results in enrichment of R loops particularly across rDNA arrays (El Hage et al., 2010). Surprisingly, in these backgrounds, no *sen1-3* dependent synthetic genetic defects were observed (fig 3.10).

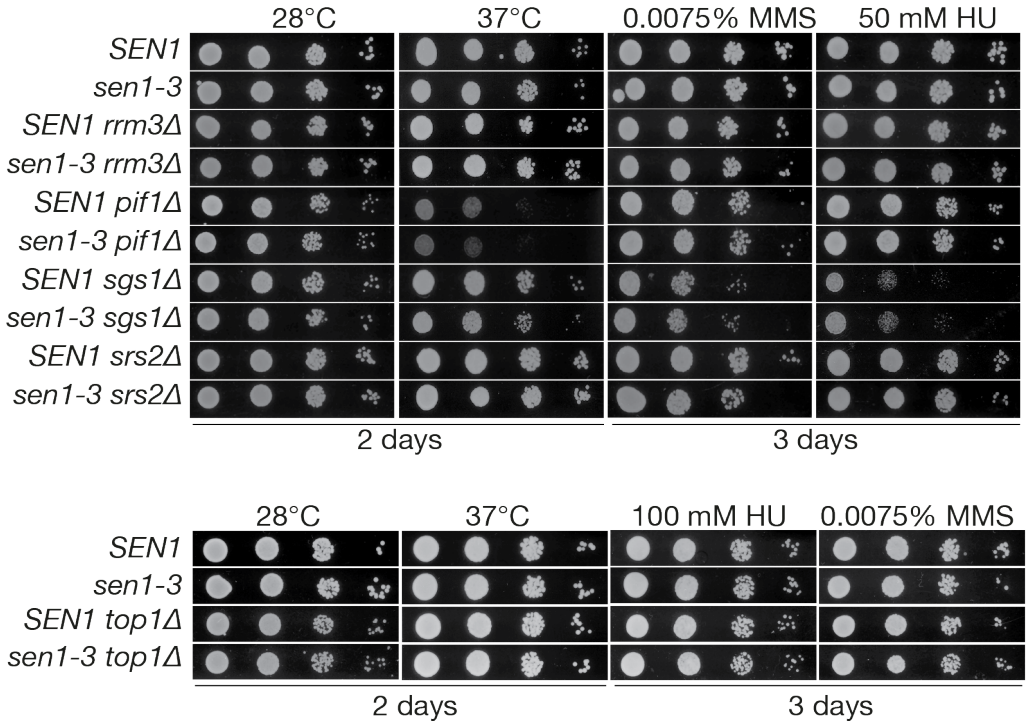


Figure 3.10 Tethering Sen1 to the replisome may only be critical for the resolution of R loops under certain circumstances. Serial dilution spotting of yeast strains was performed, where 10-fold dilutions of the various cell suspensions were grown on YPD medium at different temperatures, or under conditions of DNA damage or replication stress induced by addition of MMS or HU to the media. Plates were imaged every 24 hours. s*en1-3* cells reveal no synthetic defects in other backgrounds with high levels of R loops. Strains used in this experiment: CS2808, CS2810, CS2863, CS2865, CS2878, CS2880, CS2882, CS2884, CS2894, CS2896, CS2890, CS2892.

3.6 Investigating the functional redundancy between Sen1 and RNase H at replication forks

As *sen1*-3 is synthetic lethal in the absence of RNase H; this suggests that there is some level of functional redundancy between these enzymes for the removal of R loops, specifically at replication forks. However, it is also clear that the activity of these pathways does not fully overlap; for example, the RNase H enzymes are not required for transcription termination. Consequently, we next aimed to assess the level of functional redundancy between Sen1 and RNase H at forks.

In the context of damage induced by R loops in a *rnh1*Δ *rnh201*Δ background, many different pathways are essential for cell viability. These include genes involved in the prevention of R loops, as well as homologous recombination (HR) and post replicative repair (table 3.1). Though it appears there is some functional overlap between Sen1 and RNase H for the digestion of R loops; in these lethal backgrounds, it is possible

endogenous levels of Sen1 cannot remove R loops with kinetics sufficient to maintain viability without these factors. Alternatively, while playing redundant roles in R loop metabolism, RNase H and Sen1 may have other specific essential functions that cannot be complemented by one another. Thus, we decided to survey the levels of Sen1 redundancy in a *rnh1*Δ *rnh201*Δ background in combination with deletion of various genes involved in R loop prevention or metabolism and different DNA repair pathways.

To do this, we analysed whether overexpression of *SEN1* (where we crossed in the *ACT1-SEN1* allele) could suppress the lethality in strains lacking RNase H activity. If the answer was positive; we would then repeat the overexpression of *SEN1* using an *ACT1-sen1-3* allele, so to understand whether the viability depends on the tethering of Sen1 to the replisome. We reasoned that if overexpression of *SEN1,* but not *sen1-3* was able to suppress any growth defects or rescue cell lethality of *rnh1*Δ *rnh201*Δ cells crossed with the other candidate genes, this may provide an indication of the level of Sen1 redundancy in these genetic backgrounds specifically at replication forks. Rescuing the lethality could either be a result of overlapping R loops resolution function in that background, or a unique redundancy of Sen1 with the other candidate pathway. Interestingly, for some of the genes tested, *ACT1-SEN1,* but not *ACT1-sen1-3* was able to suppress the lethality or partially improve the growth defects of the triple mutants, however the cells still remained sick (fig 3.11a,b, results summarised table 3.1).

ACT1-SEN1 was able to rescue the lethality of *rnh1*Δ *rnh201*Δ combined with *hpr1*Δ*,* where defects in mRNP biogenesis further increase the levels of R loops in the cell. However, these cells still displayed growth defects (fig 3.11a). This suggests that Sen1 does indeed have some functional redundancy for the removal of R loops in this background, and that by increasing its availability and consequently its activity, the levels of these DNA:DNA hybrids are reduced to a point where viability is restored. However, this is perhaps not with kinetics sufficient to fully complement deletion of the other alleles. Interestingly, *hpr1*Δ cells have been shown to accumulate telomeric R loops, suggesting the R loop activity of Sen1 at forks may be particularly important at these regions (Pfeiffer et al., 2013). In a background where the RNase H enzymes are present, the loss of Sen1 from forks exacerbates the growth defects of *hpr1Δ* cells

(fig 3.7), suggesting Sen1 plays an important function at forks that cannot be fully complemented by RNase H, or that RNase H cannot resolve R loops with the same kinetics as Sen1 at forks in this background.

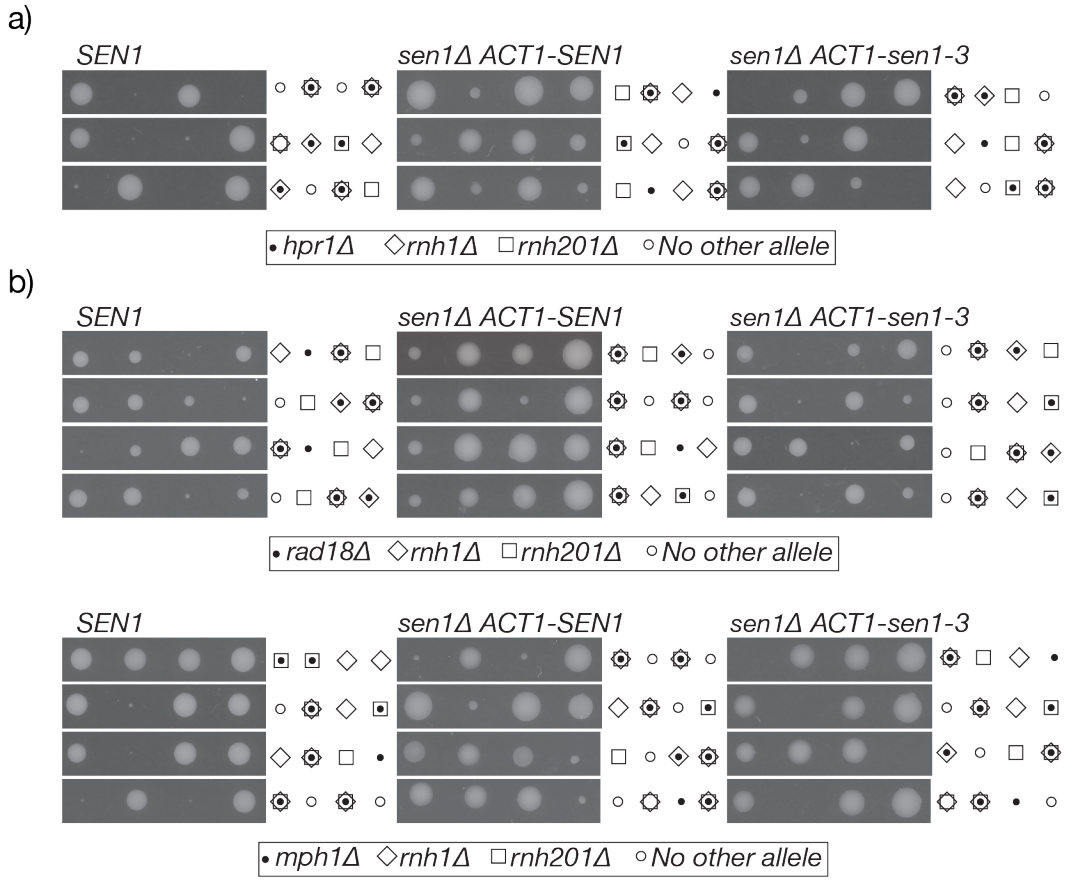


Figure 3.11 Genetic interactions of *sen1-3 rnh1Δ rnh201Δ* cells. Meiotic progeny of *SEN1, ACT1-SEN1* or *ACT1-sen1-3 rnh1Δ rnh201Δ* cells crossed with various candidate genes involved in different pathways including a) mRNA biogenesis and b) DNA lesion bypass were scored by replica plating on selective media at 24ºC. Genotype is indicated by the symbols described in the key. Strains used for crosses: CS2734, CS2736, CS3168, CS2277, CS4386, CS3745, CS239, CS4206, CS4208, CS716, CS4369, CS4367.

Interestingly, similar to *hpr1Δ,* cells lacking the Mph1 helicase have also been shown to accumulate R loops at telomeres (Lafuente-Barquero et al., 2017). Telomeres are areas of the genome that are especially hard to replicate. Their existence as heterochromatin (enriched with heterochromatic proteins and markers for histone remodelling), together with their repetitive R loop-forming sequence, means fork stalling is common across these regions (Ozer and Hickson, 2018). Hpr1 and Mph1 control the levels of R loops via different mechanisms; the helicase activity of Mph1 appears resolve any that accumulate, whereas Hpr1 prevents their formation.

Suppression of *mph1*Δ *rnh1*Δ *rnh2011*Δ lethality by *ACT1-SEN1* (fig 3.11b) once again suggests that Sen1 may be important at forks for the maintenance of telomeres and accurate replication of chromosome ends in an R loops dependent manner. The inability of *ACT1-sen1-3* to suppress the lethality of both *mph1*Δ and *hpr1*Δ backgrounds (fig 3.11) suggests that when Sen1 is not already present at forks through its replisome tethering mechanism; it cannot be recruited to any replisome/hybrid collision sites, such as at telomeres, with kinetics sufficient to remove the R loop before lethal damage occurs.

Finally, both *RAD18* and *MPH1* are involved in bypassing damaged DNA with the purpose of promoting replication through lesions; thus, reducing fork collapse. However, they are each part of distinct biochemical pathways. Rad18 is an E3 Ubiquitin ligase, involved translesion synthesis. Mph1 is also involved in error-free bypass of lesions, but as part of a separate non-canonical HR pathway (Schurer et al., 2004). Both of these pathways have been shown to be critical for tolerating ribonucleotides mis incorporated into DNA as a result of the absence of RNase H, which plays a main role in their removal (Lazzaro et al., 2012). Partial suppression of the defects by *ACT1-SEN1* (fig 3.11b) suggests that the presence of Sen1 at replication forks may either play a partially redundant role with RNase H in the removal of ribonucleotides mis incorporated by the replicative polymerases, reducing the cells reliance on these lesion bypass pathways for viability, or that Sen1 has some post-replicative repair function. The latter may be more likely, as *ACT1-SEN1* is unable to rescue the lethality of *rnh1Δ rnh201Δ top1Δ* cells (fig 3.12c). Top1 provides an alternative pathway in cells lacking RNase H for the removal of rNMPs (Lazzaro et al., 2012, Potenski and Klein, 2014). Importantly, once again, *ACT1-sen1-3* was unable rescue the lethality in these genetic backgrounds.

The synthetic lethality of *rnh1*Δ *rnh201*Δ combined with deletion of *RAD52* indicates a critical need for HR in this background to repair DNA damage and maintain cell viability (fig 3.12a). Rad52 is responsible for displacing RPA from exposed ssDNA to facilitate loading of Rad51, which forms the nucleofilaments responsible for strand invasion of a homologous template (Sugiyama et al., 1997, Sugiyama and Kowalczykowski, 2002). The lethality is likely due to the fact that high levels of DNA lesions occur as a result of both R loops interference with replication, leading to

collapsed forks (which require HR for their repair), and the vulnerability of the displaced single stranded DNA within the hybrid. However, overexpression of *SEN1* or *sen1-3* under the *ACT1* promoter did not rescue this lethality (fig 3.12a). This suggests that *SEN1* cannot resolve R loops in this background with kinetics sufficient to reduce the levels of damage enough to complement this reliance on HR. It also suggests that *SEN1* activity has no redundancy with the HR pathway for the repair of any R loops induced DNA damage. Concurrently, dilution spotting experiments showed that any growth defects in *rad51Δ* cells were not exacerbated by the *sen1-3* allele, providing further evidence Sen1 does not play a functionally redundant role in the HR pathway at forks (fig S.1)

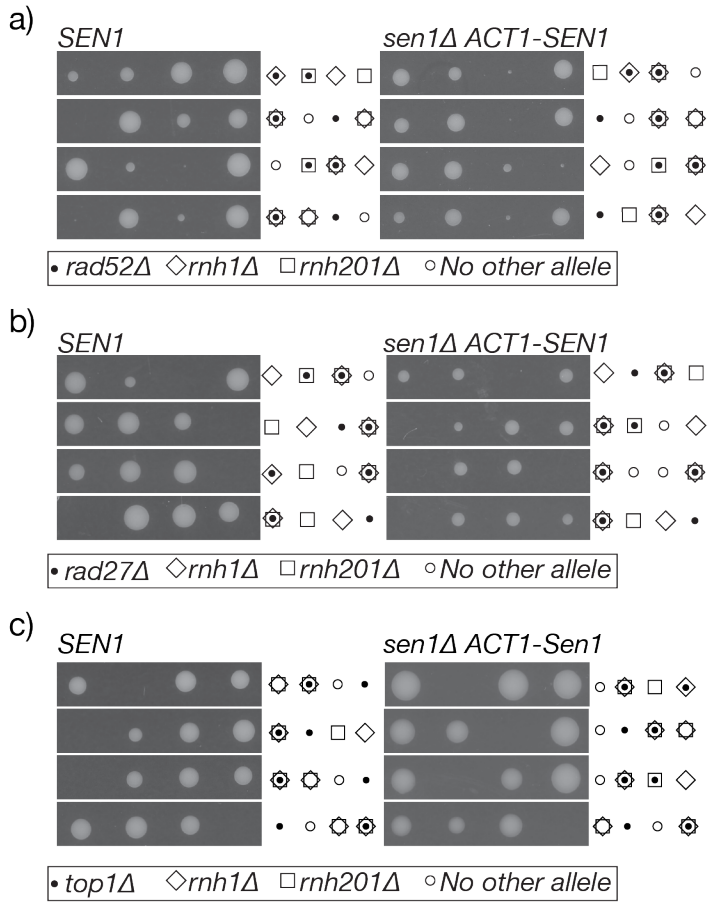


Figure 3.12 Meiotic progeny of *SEN1, ACT1-SEN1* or *ACT1-sen1-3 rnh1Δ rnh201Δ* cells crossed with various candidate genes involved in different pathways including a) Homologous recombination b) Base excision repair c) topological stress relief and ribonucleotide excision repair were scored by replica plating on selective media grown at 24°C. The genotypes are indicated by the symbols described in the key. Strains used for crosses: CS2734, CS3168, CS792, CS4189, CS4334, CS4397, CS1175, CS4191.

Similar results were obtained combining these alleles with deletion of the nuclease *RAD27* (fig 3.12b). Rad27 is involved in base excision repair by resolving flaps generated during the repair process. It is also involved in maintaining genome fidelity, due to its role in Okazaki fragment maturation. Though a major pathway of maturation depends on the strand displacement activity and flap removal activity of Pol δ and Rad27, respectively, a second major pathway may also involve co-operative action of RNase H2 (Qiu et al., 1999, Burgers, 2009). The lack of suppression by *ACT1-SEN1* (fig 3.12b) and the absence of any synergistic synthetic defects in *sen1-3 rad27Δ* cells (fig S.1) suggests that Sen1 does not play a redundant role in the maturation of Okazaki fragments or Rad27-dependent long patch base excision repair.

Finally, *ACT1-SEN1* did not suppress synthetic lethality with deletion of *TOP1* (fig 3.12c). Aside from providing an alternative pathway for rNMP removal, it is involved in relieving torsional stress surrounding forks (Potenski and Klein, 2014). It has also been shown to be particularly important for maintaining genome stability across rDNA, which accumulates R loops in its absence (El Hage et al., 2010). The inability of *ACT1-SEN1* to suppress lethality of *rnh1Δ rnh201Δ top1Δ* cells (fig 3.12c) is in agreement with a study from the Koshland lab which observed that sites of R loop-associated lethal damage in *rnh1Δ rnh201Δ Top1*-depleted cells (across rDNA) are not present in those depleted of Sen1, and vice versa. Thus, their R loops function appears to be important at non-overlapping regions of the genome (Costantino and Koshland, 2018). In agreement, no synergistic defects are observed between the *top1Δ* allele and *sen1-3* (fig S.1).

Table 3.1 Summary of genetic interactions examining the redundancy between Sen1 and RNase H activity at replication forks

		Synthetic lethality rescued	
Gene	**Pathway/ function**	***ACT1-SEN1***	***ACT1-sen1-3***
HPR1	mRNP biogenesis and telomere maintenance	✓✓✓	✕✕✕
RAD18	Post-replicative repair	✓✓✓	✕
MPH1	Error-free lesion bypass and telomere maintenance	✓✓✓	✕✕✕
RAD52	Homologous recombination	✕	
RAD27	Okazaki fragment processing and base excision repair	✕✕✕	
TOP1	Torsional stress relief and removal of ribonucleotides mis incorporated into DNA	✕✕✕	

Chapter Summary

Considering all of the data, it appears that Sen1 has some functional redundancy with the RNase H enzymes for the resolution of R loops at replication forks. Interestingly, though overexpression of *sen1-3* under the strong constitutive *ACT1* promoter rescues the lethality of the *sen1-3 rnh1Δ rnh201Δ* triple mutant, unlike *ACT1-SEN1,* these cells are sensitive to elevated temperatures, MMS and HU (fig 3.3). This suggests that under conditions of DNA damage or replication stress, the tethering of Sen1 to forks is particularly important to cope with R loops in cells lacking RNase H activity. The association between Sen1 and the replisome may provide a kinetic advantage by ensuring it is readily available at the required collision sites to resolve the R loops and promote progression of the replication fork. Any prolonged stalling of the fork could lead to deleterious recombination events and genome instability. Analysing the ability of the *ACT1-SEN1* and *ACT1-sen1-3* alleles to rescue the synthetic lethality of double *RNH* deletion in various different genetic backgrounds has shown that *SEN1* activity at forks may be important particularly for telomere maintenance or post-replicative repair, as well as R loops metabolism (fig 3.11). Interestingly, it appears that Sen1 R loops activity may only be important under specific conditions or at particular genomic loci, as the *sen1-3* allele does not exacerbate any defects in cells lacking various other helicases involved in R loops removal (fig 3.10). These may act at non-overlapping regions of the genome or on DNA:RNA hybrid substrates with different biochemical features. However, *sen1-3* was also shown to display synthetic growth defects and increased recombination when combined with mRNP biogenesis mutant *hpr1Δ* (fig 3.7,8). Importantly, these phenotypes were linked to the formation of R loops, as they could be suppressed by *RNH1* overexpression (fig 3.7,8). However, direct analysis of R loops, using both a cytological assay and slot blot analysis did not detect any difference in the levels of DNA:RNA hybrids between *SEN1* and *sen1-3 hpr1Δ* cells (fig 3.9). It is possible that although the levels of R loops are similarly increased in these strains, a subset of these could be more efficiently converted to damage in the absence of Sen1 at forks, leading to the increased levels of recombination.

Chapter 4: Results

INVESTIGATING THE ROLE OF SEN1 AT REPLICATION FORKS BY ANALYSING THE GENETIC INTERACTIONS OF *SEN1-3*

4.1 Background

In addition to causing defects in transcription termination, increased levels of R loops and hyperrecombination, the *sen1-1* allele (which has a mutation in the essential helicase domain) is synthetic lethal with deletion of various genes involved in replication fork stability and DNA repair. Among those essential for viability are the homologous recombination proteins Srs2, Mre11, Sgs1, Rad51, Rad52; and the replisome components Ctf4, Tof1 and Mrc1 (fig 4.1) (Alzu et al., 2012, Mischo et al., 2011). This suggests that the activity of Sen1 is critical for maintaining the integrity of the genome by playing a role in preventing events that lead cells to rely on the fork protection complex (Tof1-Mrc1-Csm3) or require DNA repair mechanisms such as HR.

Accordingly, we were interested whether any of these pathways are also critical for maintaining viability in cells where Sen1 is no longer tethered to replication forks (*sen1-3).* Surveying for synthetic genetic interactions of *sen1-3* with various mutants and analysing whether any defects found arise during DNA replication provides further information toward elucidating the specific role of Sen1 at forks.

4.2 Genetic interactions of the *sen1-3* allele

First, a number of mutants, including those that are synthetic lethal with *sen1-1,* were screened for genetic interactions with *sen1-3* by performing dilution spotting experiments. Strains carrying deletion of the gene of interest crossed with either the *SEN1 or sen1-3* allele were assessed for sensitivity to temperature, replication stress and DNA damage. In combination with *sen1-3,* several of the mutants tested displayed no synthetic genetic defects, however a portion were found to negatively affect cell growth (summarised fig 4.1).

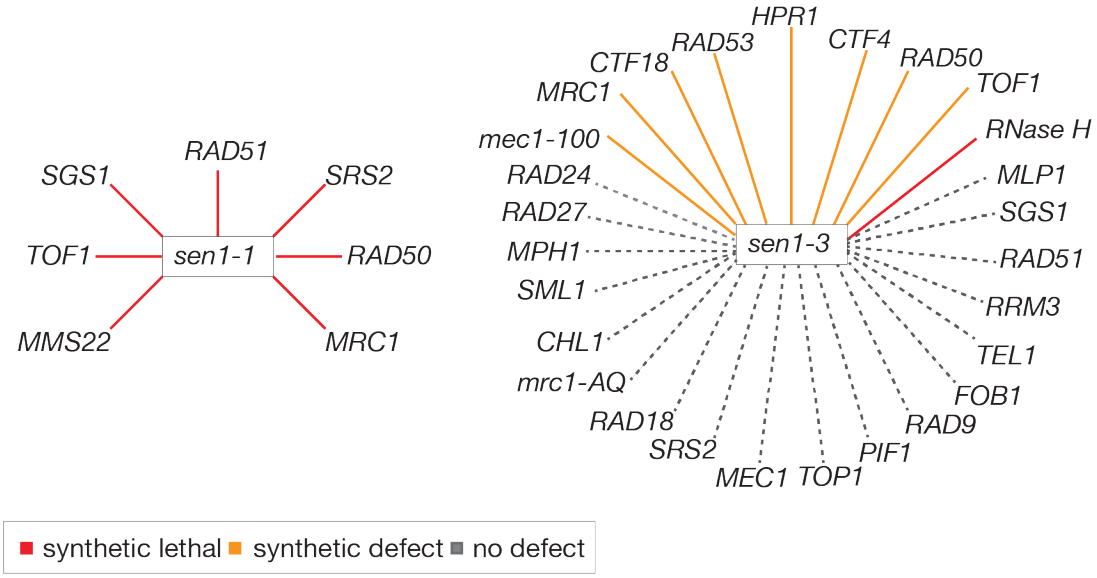


Figure 4.1 Schematic of *sen1-1* and *sen1-3* genetic interactions. *sen1-1* cells are synthetic defective in combination with deletion of various genes involved in replication fork stability and DNA repair. *sen1-3* shows genetic defects in combination with deletion of a fraction of these genes, as well as various other replication proteins. (Dilution spotting of genes that presented no genetic interaction is shown in fig S.1).

4.2.1 Ctf4

Similar to *sen1-1, sen1-3* is synthetic defective when combined with deletion of *CTF4,* a key interaction hub within the eukaryotic replisome (fig 4.2). Ctf4 not only couples the CMG helicase to Pol α/primase (Gambus et al., 2009, Simon et al., 2014), helping to ensure the processes of DNA unwinding and lagging strand synthesis are co-ordinated, but also mediates the interaction of the replisome with multiple other accessory proteins (Villa et al., 2016). These proteins, characterised by the presence

of a CIP-box include: Sld5, Dna2 and Tof2. The Ctf4-mediated recruitment of some of these factors not only contributes to chromosome duplication, but also coordinates other important processes with DNA replication, for example: rDNA copy-number regulation, DNA repair and sister chromatid cohesion (Villa et al., 2016, Samora et al., 2016, Simon et al., 2014, Zhu et al., 2007, Hanna et al., 2001). As deletion of *CTF4* results in its absence from the replisome, as a consequence, the coupled processes are no longer coordinated with DNA synthesis. The *sen1-3 ctf4Δ* double mutant is sensitive to both temperature and replication stress, suggesting that in this background, the loss of Sen1 from forks is genotoxic (fig 4.2).

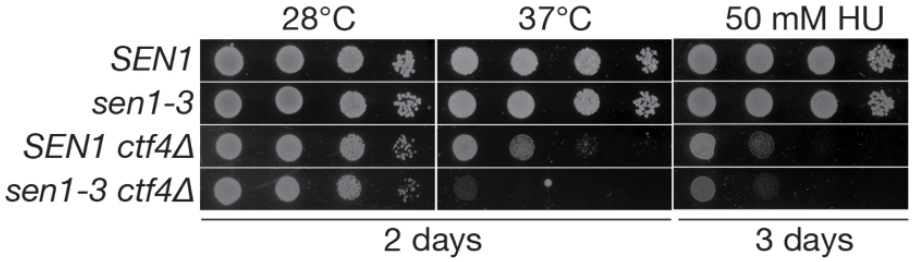


Figure 4.2 Lack of coordination between replication and Ctf4 mediated processes is genotoxic in *sen1-3*. Strains were serially diluted 10-fold to obtain concentrations of $5x10^6$, $5x10^5$, $5x10^4$ and $5x10^3$ cells/ml, then 10µl of each suspension was spotted on YPD medium of various conditions, including media supplemented with DNA damaging agent MMS or dNTP depleting agent HU. Cells were incubated at the described temperatures and imaged every 24 hours. This revealed that *sen1-3 ctf4Δ* mutant is sensitive to both temperature and replication stress. Strains used in this experiment: CS2808, CS2810, CS3283, CS3285.

4.2.2 Checkpoint proteins

The *sen1-3* allele was also identified as synthetic defective with deletion mutants of the S phase checkpoint genes *MRC1*, *CTF18* and *RAD53*. Rad53 is the main checkpoint effector kinase, and Mrc1 and Ctf18 are key mediator proteins. All of the double mutants exhibited growth defects at 37ºC and *sen1-3 mrc1*Δ and *ctf18*Δ cells are also sensitive to replication stress. Moreover, *sen1-3 rad50*Δ cells are also synthetic defective at 37ºC (fig 4.3). Rad50 is a component of the yeast MRX complex (Mre11/Rad50/Xrs2), which is involved in the processing of DSBs and has been implicated in activation of both the G1 and S phase checkpoint in response to DSBs (Grenon et al., 2001). Also, Rad50 is thought to be involved in the recruitment of cohesin to replicating chromatin. At stalled forks, increased accessibility to chromatin is achieved by Rad50 in cooperation with chromatin remodelling factors, where they

promote resection of the nascent DNA by exonucleases. The resultant ssDNA stimulates the loading of Cohesin, where its accumulation at stalled forks presumably serves to maintain close association of sister chromatids, either for the repair of one ended DSBs at collapsed forks, or to facilitate HR-mediated fork restart (Delamarre et al., 2020, Tittel-Elmer et al., 2012).

sen1-3 exacerbated HU sensitivity was also observed with *tof1*Δ (fig 4.3), a member of the fork protection complex (FPC), which stabilises HU stalled replication forks by preventing helicase and polymerase uncoupling (Katou et al., 2003). Tof1 is also thought to play a role in activation of the DNA replication checkpoint, as its deletion aggravates checkpoint defects in *rad9*Δ cells (Foss, 2001). However, its role is less clear than that of Mrc1. Finally, *sen1-3 mec1-100* cells are also sensitive to replication stress (fig 4.3). *mec1-100* is a hypomorphic mutant allele of the checkpoint sensory kinase *MEC1* and is defective in the G1/S and S phase checkpoint; showing only weak activation of Rad53 during S phase. *mec1-100* cells also activate excess replication origins (Paciotti et al., 2001, Zhong et al., 2013).

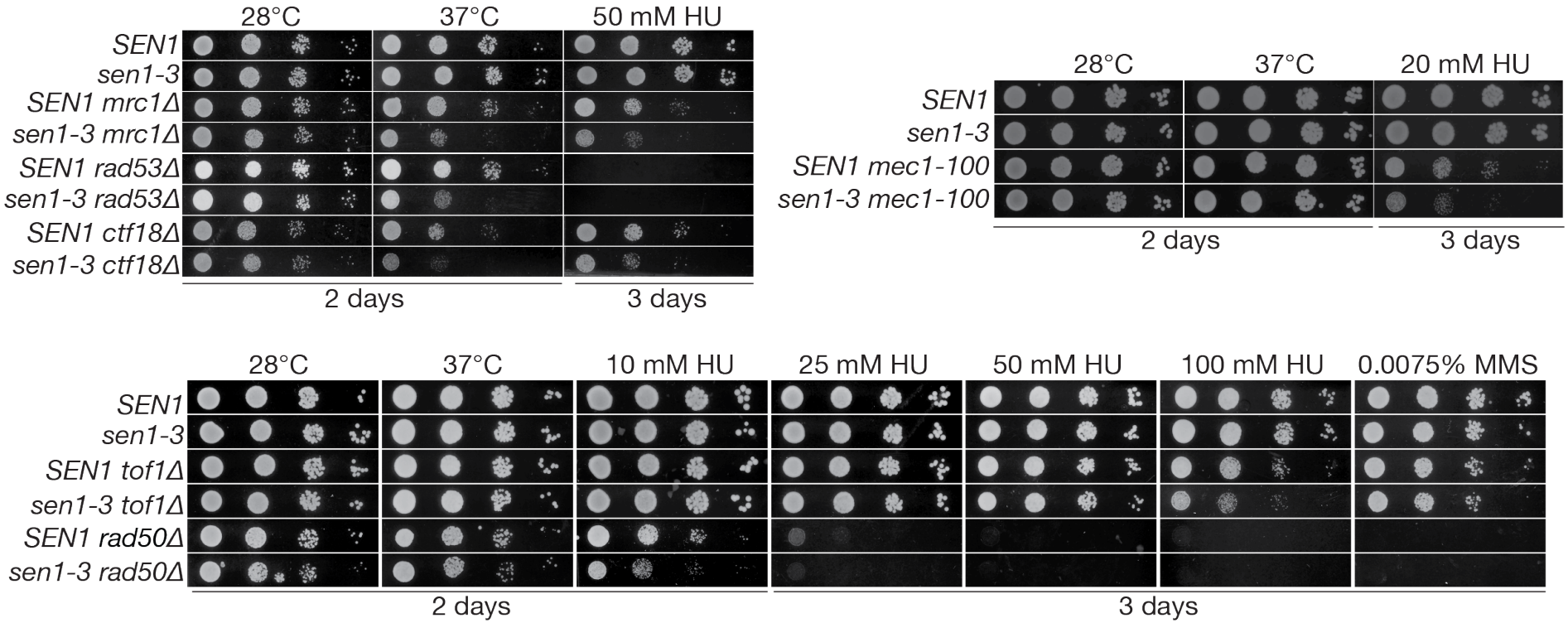


Figure 4.3 The *sen1-3* allele displays synthetic genetic defects in combination with deletion of various S phase checkpoint proteins. Serial dilution spotting of yeast strains was carried out on YPD media under conditions of DNA damage, replication stress or at the indicated temperatures. Plates were imaged every 24 hours. Stains used in this experiment: CS2808, CS2810, CS2859, CS2861, CS2945, CS2947, CS2955, CS2957, CS2934, CS2936, C2867, CS2868, CS2904, CS2906.

To analyse whether any of the defects observed in these checkpoint mutants were due to gene gating, we crossed the *mlp1Δ* allele (which is the nuclear pore involved in establishing gene gating) into *sen1-3 mrc1Δ* cells. At sites where transcription and replication collide, one of the proposed functions of the S phase checkpoint is to promote fork passage by controlling gene gating. Checkpoint dependent phosphorylation of nucleoporins inhibits gene gating; thus, results in release of nascent RNA from nuclear pores, which relieves some of the torsional stress on the template DNA (Bermejo et al., 2011). However, instead of suppressing any defects, *mlp1Δ* further exacerbates the synthetic growth defects of *sen1-3 mrc1Δ* cells (fig 4.4). Deletion of *MLP1* has been shown to result in an increase in R loops (García-Benítez et al., 2017), which is likely responsible for the exacerbation of the growth defects in *sen1-3 mrc1Δ* cells (fig 4.4).

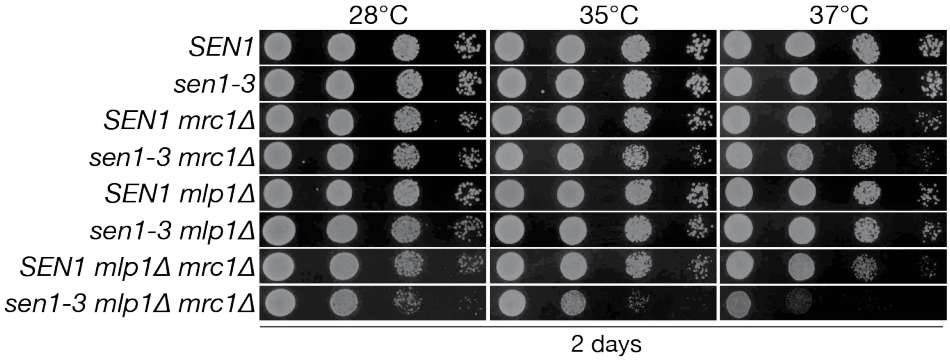


Figure 4.4 Inhibition of gene gating does not rescue the synthetic defects of *sen1-3 mrc1Δ* cells. 10-fold serial dilutions of the various strains *were spotted* onto YPD media and incubated at the described temperatures. The cells were imaged every 24 hours. Strains used in this experiment: CS2808, CS2810, CS2859, CS2861, CS3782, CS3790, CS3816, CS3817.

Interestingly, and puzzlingly, no synthetic genetic defects were observed between *sen1-3* and deletion mutants of genes involved in the canonical activation of the S phase checkpoint pathway: *mec1*Δ *sml1*Δ*, tel1*Δ*,* and *mec1*Δ *sml1*Δ *tel1*Δ (fig 4.5) (note, *mec1*Δ strains require deletion of the ribonucleotide reductase inhibitor *SML1* for viability. Sml1 is usually degraded via the Mec1 pathway to upregulate synthesis of dNTPs in response to replication stress, which is an essential function (Zhao et al., 1998)). In addition, *sen1-3* does not genetically interact with $mrc1^{AQ}$, a mutant where all 17 S/TQ motifs are mutated to AQ (fig 4.5). Therefore, $mrc1^{AQ}$ cannot be

phosphorylated by Mec1 for Rad53 activation, thus, is defective in the downstream S phase checkpoint response, but is still proficient in its DNA replication role (Osborn and Elledge, 2003).

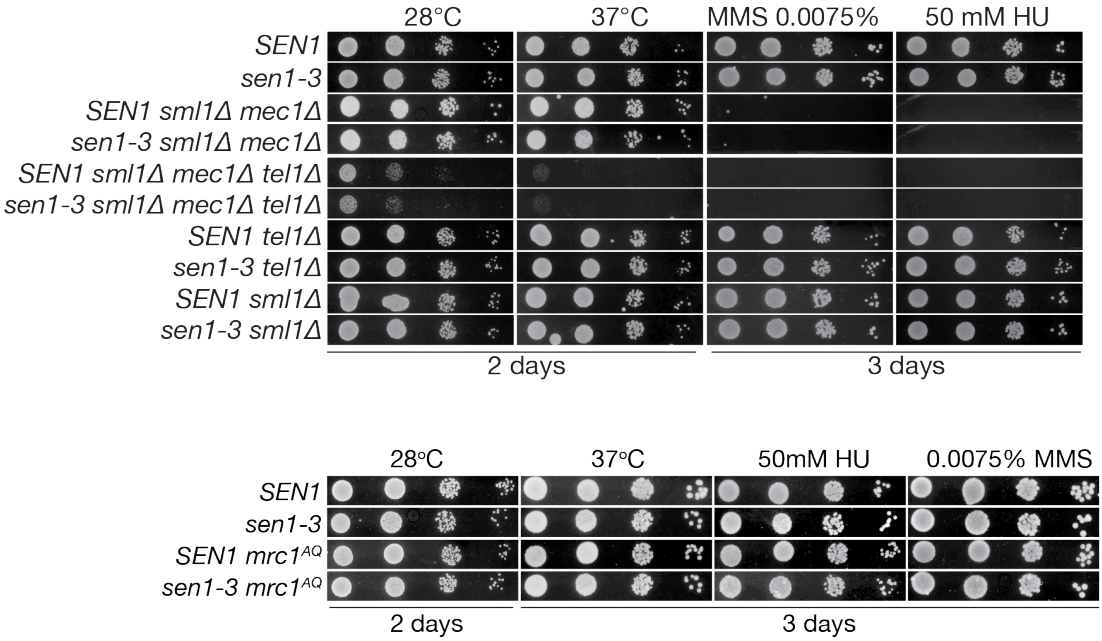


Figure 4.5 *sen1-3* does not display synthetic genetic defects in combination with deletion mutants of proteins involved in canonical activation of the S phase checkpoint response. Dilution spotting of these strains was carried out, where 10-fold dilutions of cell suspensions were spotted on YPD plates of various conditions and grown for several days at various temperatures. Cells were imaged every 24 hours and assessed for *sen1-3* allele dependent growth defects. Strains used in this experiment: CS2808, CS2810, CS3049, CS3051, CS3224, CS3226, CS3228, CS3230, CS3232, CS3234.

4.3 The *sen1-3* mutants show slower cell cycle progression in the absence of the S phase checkpoint proteins Mrc1 and Ctf18

To further investigate the defects in *sen1-3 mrc1*Δ *and ctf18*Δ mutants and whether these are linked to DNA synthesis, we once again examined the replication dynamics and kinetics of Rad52-GFP foci formation in these strains.

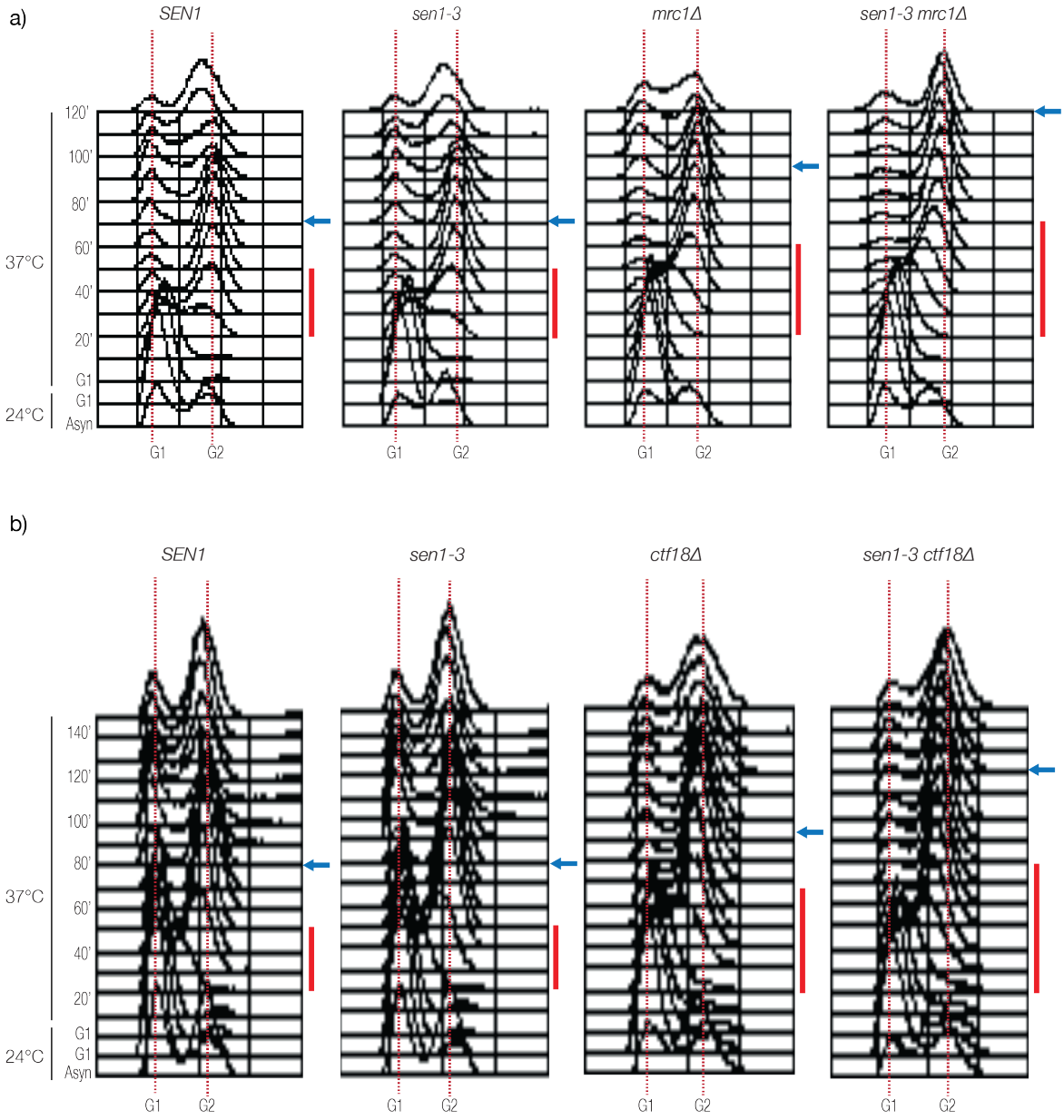


Figure 4.6 *sen1-3 mrc1*Δ or *ctf18*Δ double mutants show a delayed progression through S phase, with a significant proportion of cells arrested in G2. Cells were arrested in G1, shifted to 37°C for 1 hour and then released at non permissive temperature (37°C). Samples

were taken every 10 minutes and fixed in 70% (v/v) ethanol for analysis by flow cytometry. FACS profiles for a) *mrc1Δ* and b) *ctf18Δ* are shown. The red bar indicates the length of DNA replication, the blue arrow indicates the beginning of the exit from mitosis. Strains used in these experiments: CS2808, CS2810, CS2859, CS2861, CS2955, CS2957.

Populations of cells were synchronously released from G1 arrest into S phase at non permissive temperature (37ºC); samples were taken every 10 minutes, fixed in 70% (v/v) ethanol and analysed by flow cytometry. The FACS profiles revelated that both of the double mutant's exhibit a longer S phase (indicated by the red bar) than their respective *SEN1* controls, with an accumulation of cells blocked in G2/M (fig 4.6 a,b).

Samples of these cells were also collected at the described time points and fixed using 16% (v/v) formaldehyde to examine the kinetics of Rad52-GFP foci formation. The fixed cell nuclei were stained with DAPI and then imaged using widefield microscopy. As seen in previous experiments (fig 3.8), cells carrying the *sen1-3* allele display a small statistically significant increase ($p<0.01$) in Rad52-GFP foci compared with *SEN1* cells. Corresponding with the FACS profiles shown in fig 4.6, during late S phase, the *sen1-3 mrc1*Δ and *ctf18*Δ double mutants both showed a statistically significant increase in Rad52 foci (fig 4.7a), with a larger proportion of cells carrying two or more distinct foci, indicating damage at multiple different loci across the genome (fig 4.7b).

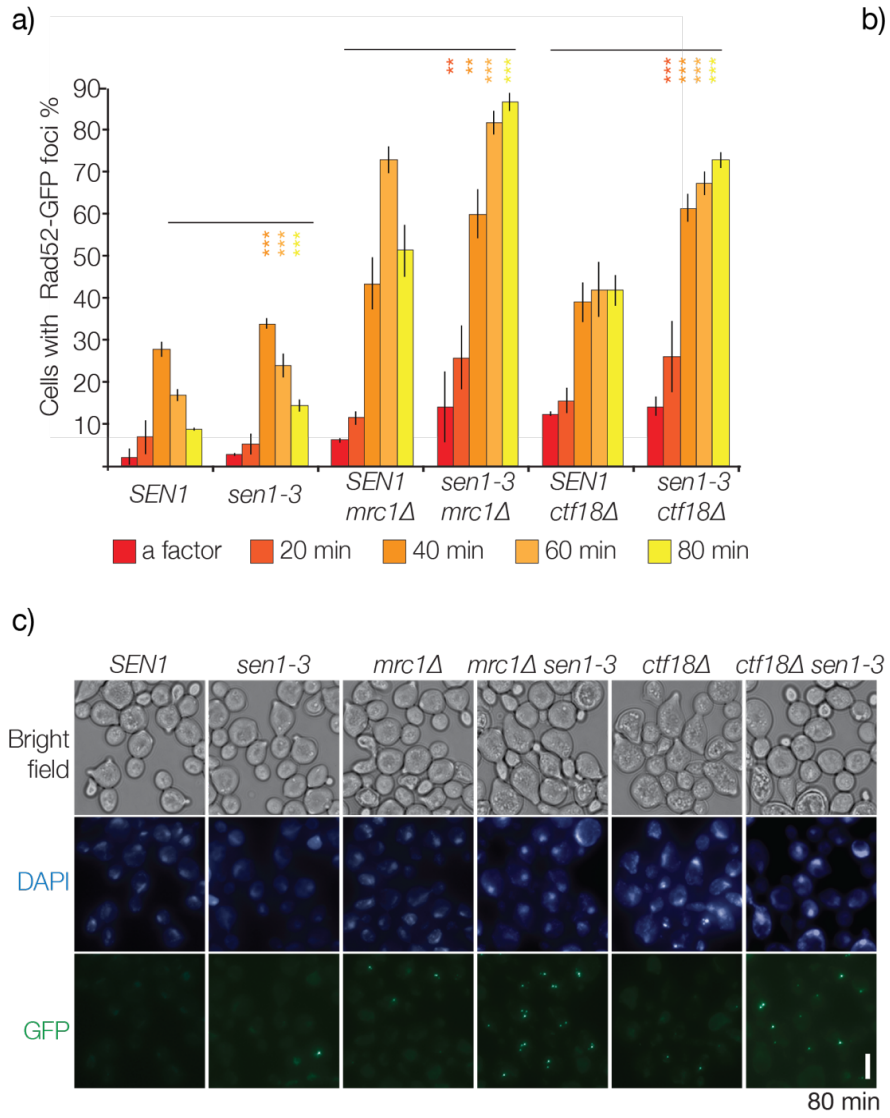


Figure 4.7 *sen1-3* shows an increase in recombination with *mrc1Δ* and *ctf18Δ*. Cells were arrested in G1, shifted to 37ºC for 1 hour and then released synchronously into S phase at non permissive temperature. The cells were fixed using 16% (v/v) formaldehyde at various time points during S phase and analysed for the number of Rad52-GFP foci a) *sen1-3*, and *sen1-3* in combination with *mrc1Δ* or *ctf18Δ* cells show an increase in recombination in the later stages of S phase, b) The single and double mutants also show an increase in cells with multiple Rad52 foci. An average of three biological repeats is shown ($p<0.05$=**, $p<0.01$=***). c) Representative images of Rad52-GFP foci analysed in (a). Strains used in this experiment: CS2876, CS2953, CS3134, CS3057, CS3062, CS3148.

Considering the previous data for *sen1-3 rnh1*Δ *rnh201*Δ and *hpr1*Δ cells; non-metabolised R loops due to loss of Sen1 from the replisome could be a source of replication fork stalling and DNA damage in these mutants. To investigate whether the observed growth defects, altered replication dynamics and increased Rad52 foci are due to increased levels of DNA:RNA hybrids as a result of abrogating the Sen1 replisome interaction, we tested whether overexpression of *RNH1* was able to suppress any of the phenotypes observed in *sen1-3 mrc1*Δ cells.

The *GAL1-RNH1* allele was crossed into *mrc1*Δ *SEN1* or *sen1-3* cells, and initial dilution spottong of the strains on YP glucose (to suppress *GAL1-RNH1* induction) versus YP galactose (to induce *GAL1-RNH1)* revelealed no *RNH1* mediated suppression of the double mutants sensitivity to increased temperatures or replication stress (fig 4.8a).

Additionally, the microscopy experiments were repeated to examine the kinetics of Rad52-GFP foci formation in these new strains during S phase. Cells grown in YP raffinose were arrested in G1, before changing the medium to YP galactose for 35 mins to induce expression of *GAL1-RNH1.* Cultures were then shifted to 37^{o}C for 1 hour while maintaining G1 arrest, before being released synchronously into S phase. Similar to the previous experiments (fig 4.7), in these condtions, a stastically significant increase in the levels of Rad52 foci between *SEN1 mrc1*Δ and *sen1-3 mrc1*Δ cells was observed. In agreement with the dilution spotting experiments, overexpression of *RNH1* did not suppress the increased levels of Rad52-GFP foci (fig 4.8a,b).

To ensure that this was not the result of the yeast having some ability to self-regulate the amount of *RNH1* within the cell, we also overexpressed the human orthologue of *RNH1* (Wahba et al., 2011, Chang et al., 2017) under the strong *GPD* promoter from a multi-copy plasmid. This obtained similar results, suggesting that the removal of R loops cannot suppress the defects observed in *sen1-3 mrc1*Δ cells (fig 4.9).

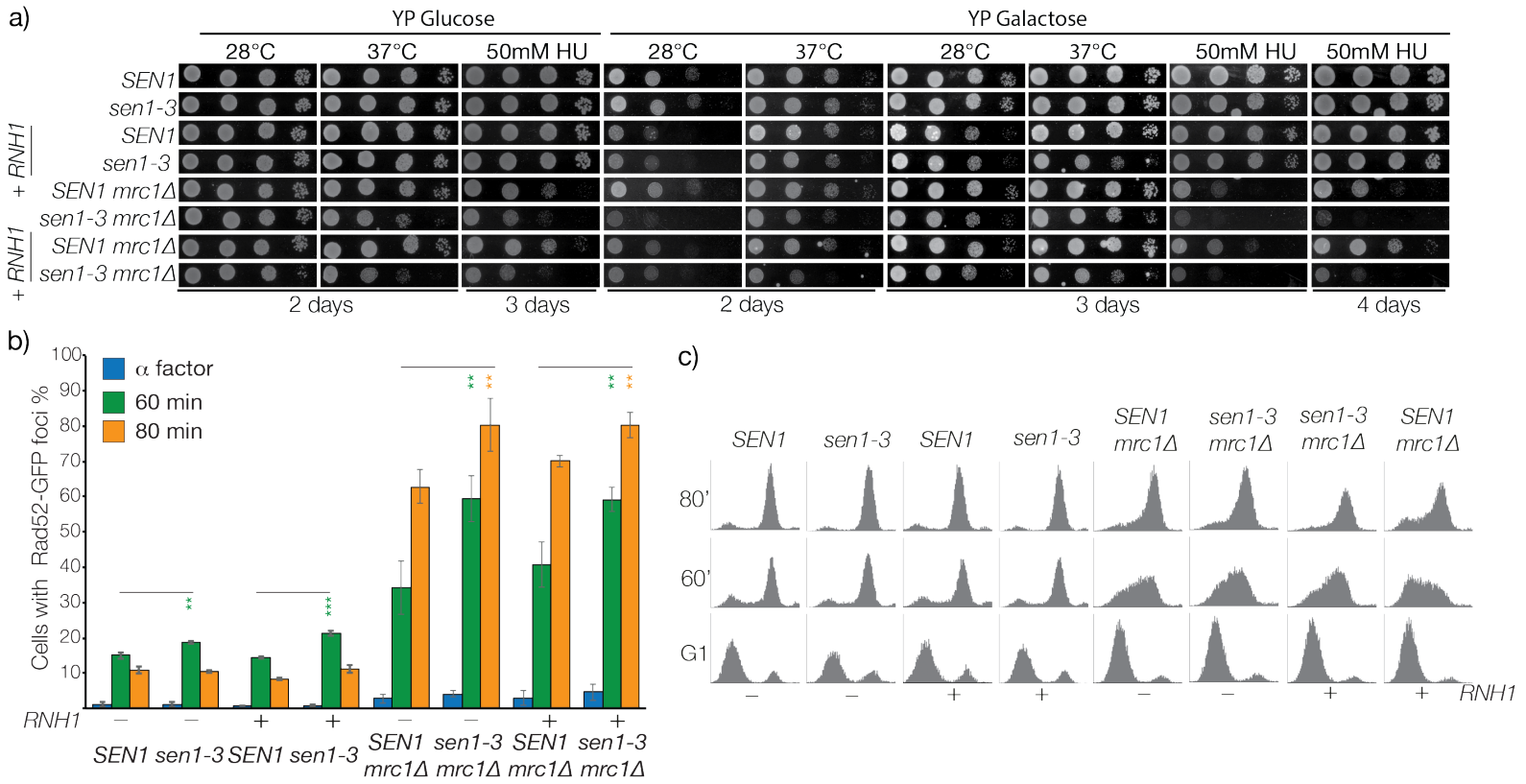


Figure 4.8 *sen1-3* shows synthetic defects and increased recombination in an *mrc1Δ* background, independently from R loops. a) Dilution spotting of cells carrying a *GAL1-RNH1* construct inserted at the *LEU2* locus and the relevant controls was carried out. Cells were plated at various conditions on medium containing either glucose to suppress or galactose to induce expression of the construct. Overexpression of *RNH1* under the *GAL1* promoter does not suppress the synthetic defects in *mrc1Δ sen1-3* cells. b) Cells were arrested in G1, shifted to 37ºC for 1 hour and then released synchronously into S phase at non permissive temperature and analysed for the number Rad52-GFP foci. The levels of recombination in the single *sen1-3* or *sen1-3 mrc1Δ* cells were unaffected by overexpression of *RNH1*. An average of three biological repeats is shown ($p<0.05$=**, $p<0.01$=***). c) Representative FACS profiles of the experiments analysed in (b). Strains used in these experiments: CS2876, CS2953, CS3134, CS3607, CS3804, CS3936, CS3808, CS3813.

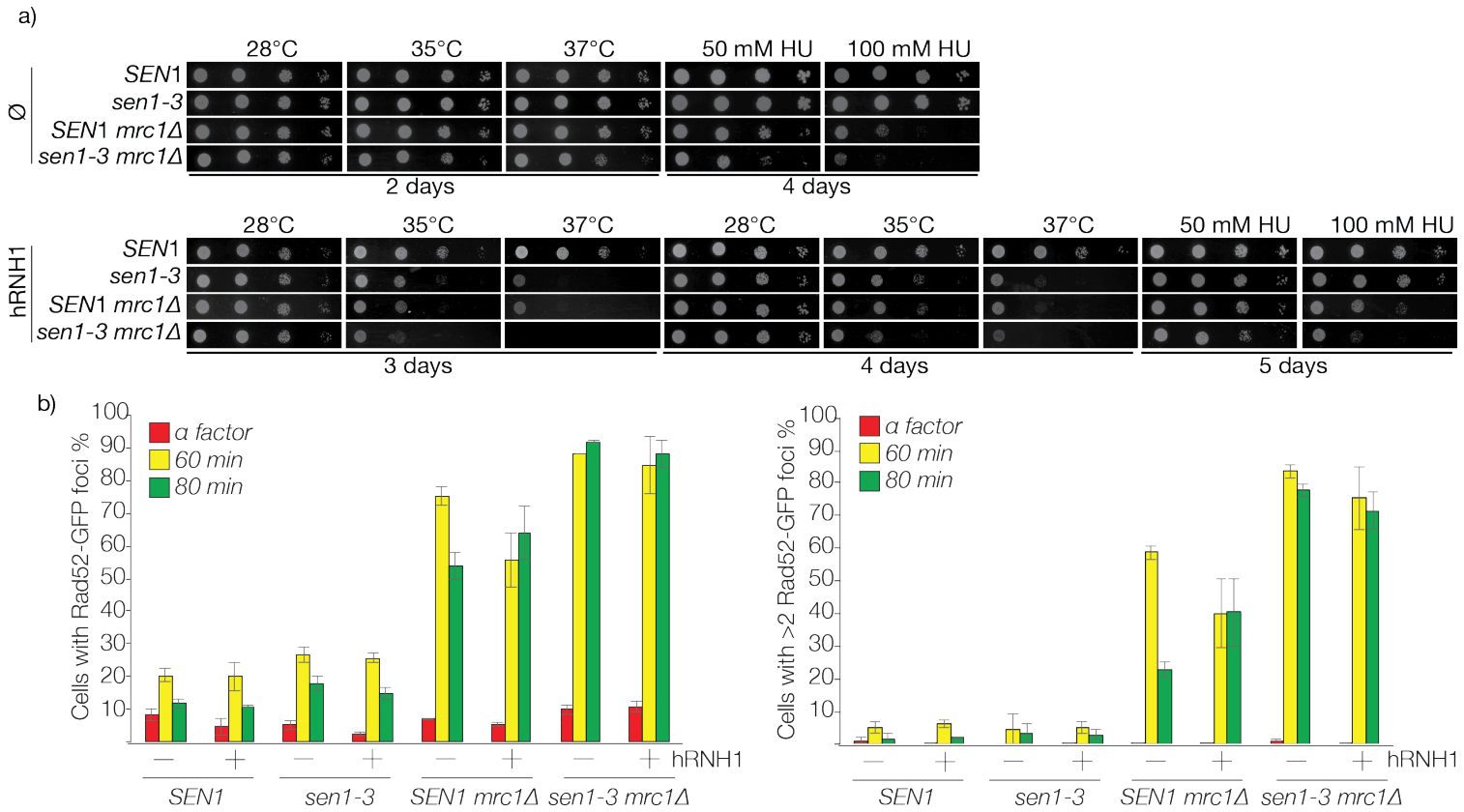


Figure 4.9 Overexpression of the human *RNH1* orthologue under control of the strong constitutive *GPD* promoter from a multi copy plasmid does not suppress synthetic defects or increased recombination in *sen1-3 mrc1Δ* cells. a) Dilution spotting of strains transformed with either the empty or *GPD-hRNH1* plasmid were grown at various temperatures and conditions. Overexpression of *hRNH1* does not suppress the growth defects of *sen1-3 mrc1Δ* cells. b) Cells transformed with either the *GPD-empty* or *GPD-hRNH1* plasmid were arrested in G1, shifted to 37ºC for 1 hour and then released synchronously into S phase at non permissive temperature and analysed for Rad52-GFP foci. Similar to *GAL-RNH1* overexpression, the levels of recombination in the single *sen1-3* or *sen1-3 mrc1Δ* cells were not suppressed by overexpression of *hRNH1*. An average of three biological repeats is shown. Strains used in these experiments: CS3853, CS3851, CS3854, CS3852, CS3847, CS3848, CS3855, CS3856.

4.4 The increased recombination of *sen1-3* mutant cells during DNA replication leads to increased genome instability

A direct repeat recombination assay was used to investigate directly whether the increased levels of Rad52 foci and growth defects observed in *sen1-3* mutants during DNA replication leads to increased recombination and genomic instability.

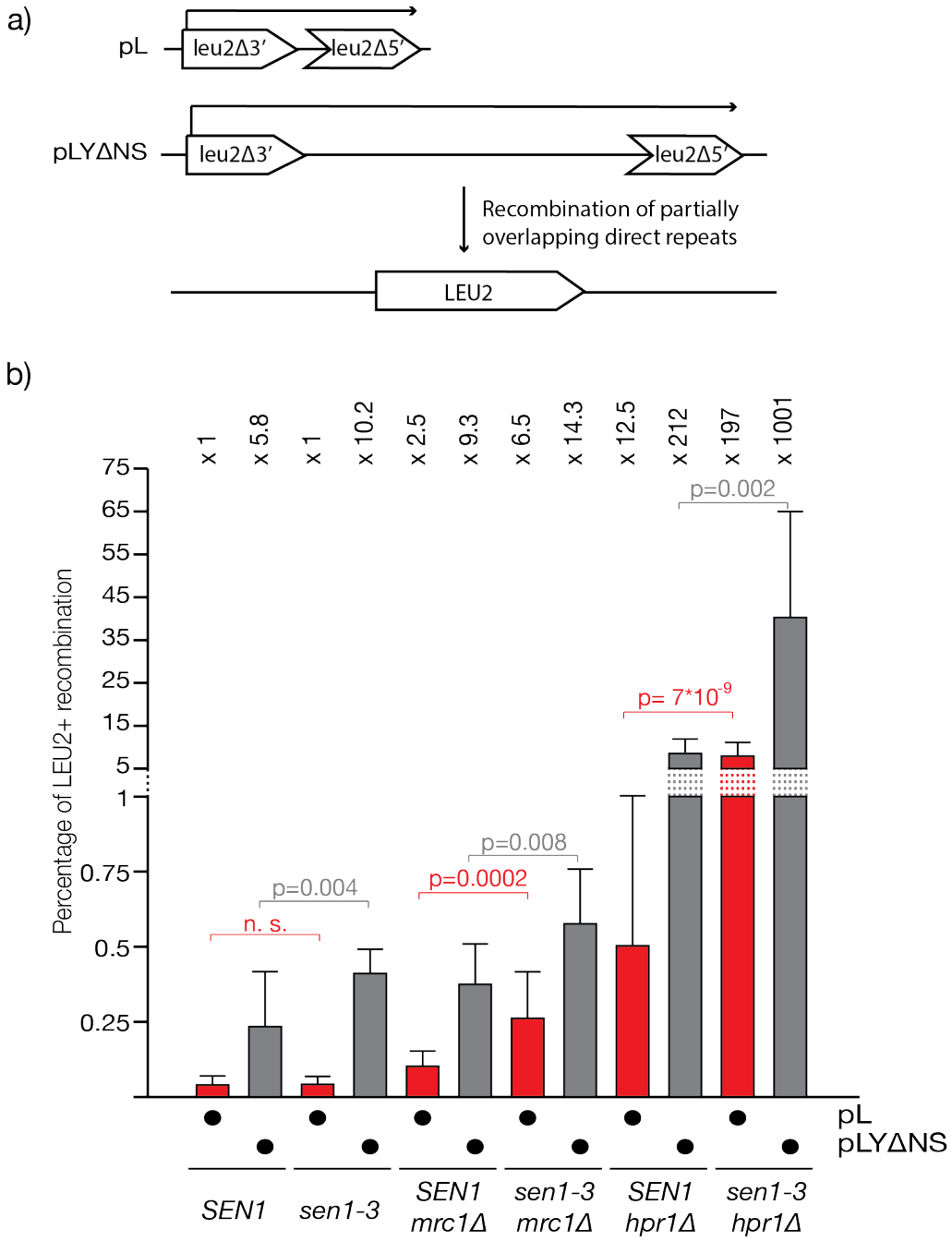


Figure 4.10 *sen1-3* causes increased recombination. Cells were transformed with the direct repeat recombination substrates, pL or pLYΔNS, shown in (a) and grown at 24°C. b) The level of recombination is presented as a ratio of the number of recombinant cells (*LEU2*+) over the total number of plasmid carrying cells (*URA*+). The percentage increase vs. wild type

Sen1 for each substrate is shown above the bars. An average of three biological repeats is shown. Strains used in this experiment: CS2808, CS2810, CS2859, CS2861, CS3742, CS3741.

In this plasmid-based system, two truncated *leu2* direct repeat motifs with partial overlap are separated either by 39 (pL) or 3,900 (pLYΔNS) nucleotides (Mischo et al., 2011, González-Aguilera et al., 2008). Following transformation of the plasmids into the various mutants, any recombination between the repeats results in a functional *LEU2*, thus allowing quanitifcation of the number of recombinant cells by analysing growth on media lacking leucine. The level of recombination is presented as the ratio of the number of recombinants (*LEU2+*) over the total number of *URA+* cells, which is the marker for the plasmid. In this system, the *sen1-3* allele showed a small but statistically significant increase in direct-repeat recombination for the pLYΔNS plasmid. In addition, for both pL and pLYΔNS, *sen1-3 mrc1*Δ double mutant cells showed a modest statistically significant increase compared with their *SEN1* counterpart. Interestingly, *sen1-3 hpr1*Δ cells showed a much higher increase. For all mutants, the levels of recombination were greater for the pLYΔNS plasmid compared with pL, thus the levels of recombination and genome instability increases with the length of the gene (fig 4.10).

4.5 Investigating the role of Sen1 at replication forks outside of R loops removal

As it appears that the presence of Sen1 at forks may promote DNA replication in a manner that is independent from the removal of R loops, this begs the question of what other obstacles it may be removing. Sen1 has a well-documented function in the termination of transcription, and as DNA replication can only happen during a short, tightly controlled period once per cell cycle, one attractive possibility is that Sen1 could play a role in the dissociation of actively transcribing or stalled RNA polymerases that are encountered by replication forks during S phase. To this end, we aimed to address whether transcription in general is the cause of the defects observed in *sen1-3 mrc1*Δ cells. We utilised the *rpb1-1* allele, which has a temperature sensitive mutation in *RPO21/RPB1,* the catalytic subunit of RNAPII; where following inactivation at 37ºC, *rpb1-1* is lost from chromatin (Zanton and Pugh, 2006, Kim et al., 2010). Therefore, at non permissive temperature this should allow us to switch off

transcription and observe whether we are able to suppress the increased recombination in *sen1-3* and *sen1-3 mrc1*Δ cells compared with the *SEN1* controls.

Cells were arrested in G1 and released into HU at permissive temperature in order to achieve arrest during S phase (as release from G1 is not possible without RNAPII). The cultures were then shifted to 37ºC and released into fresh medium to complete DNA replication, where samples of cells were fixed using 16% (v/v) formaldehyde to analyse the levels of Rad52-GFP foci. As observed previously, the levels of Rad52-GFP foci were increased in *sen1-3* cells compared to *SEN1* (fig 3.8, 4.7, 4.8, 4.9, 4.10). However, in the *rpb1-1* background, the levels were increased independently of the *SEN1* and *sen1-3* alleles in HU, meaning the observation of any possible suppression of *sen1-3* related defects is not possible in this background (fig 4.11a). Indeed, the *rpb1-1* allele sensitises both *SEN1* and *sen1*-3 cells to replication stress and DNA damage to the same extent (fig 4.11b).

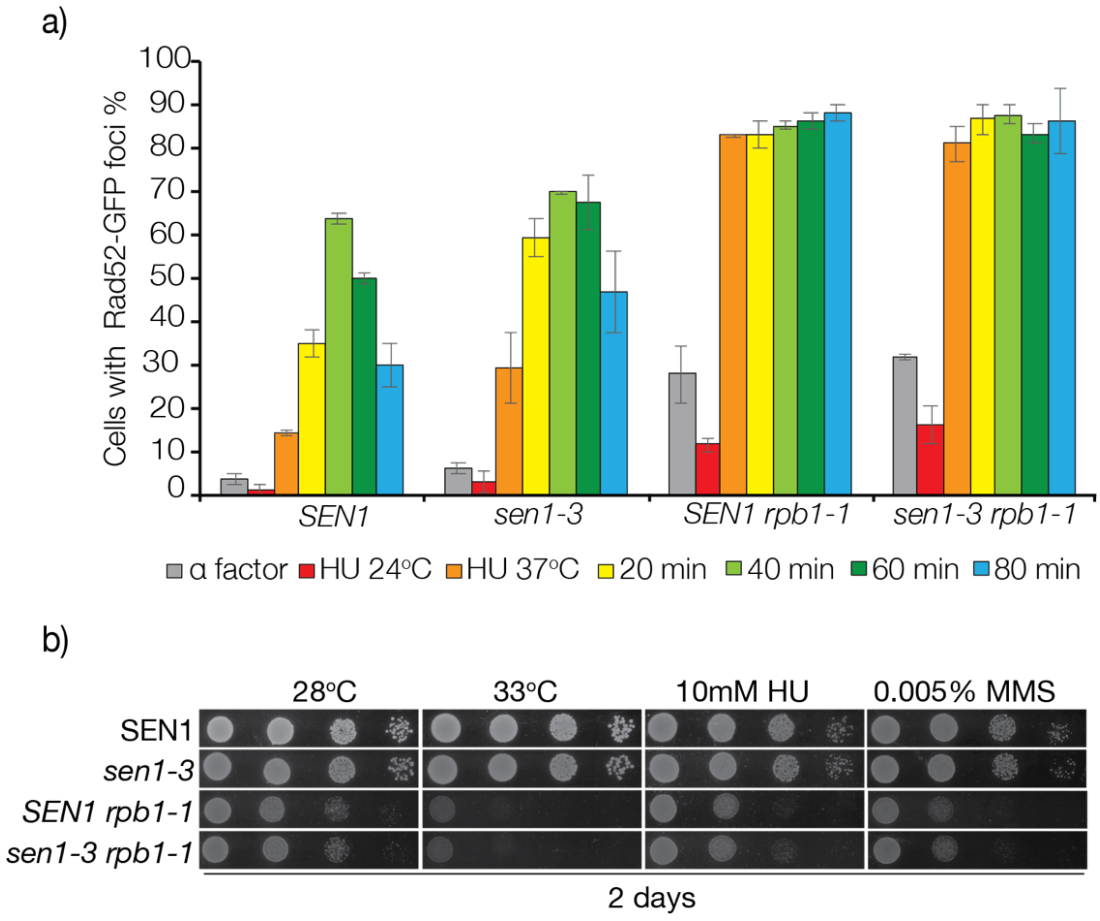


Figure 4.11 In the *rbp1-1* background, levels of Rad52-GFP foci are increased independently of the *SEN1* and *sen1-3* alleles. a) Cells were arrested in G1 at 24ºC, then released into 0.2 M HU at permissive temperature in order to achieve arrest during S phase

(as release from G1 is not possible without RNAPII). The cultures were then shifted to 37ºC to inactivate Rpb1-1 and released at 37ºC into fresh medium to complete DNA replication. Samples of cells were fixed using 16% (v/v) formaldehyde to analyse the levels of Rad52-GFP foci. As previously observed, the levels of recombination were increased in *sen1-3* cells compared to the wt. However, in the *rpb1-1* background, the levels of recombination were increased independently of the *SEN1* and *sen1-3* alleles. An average of two experiments is shown. b) Dilution spotting of *SEN1* and *sen1-3* cells carrying the *rpb1-1* allele, which has a temperature sensitive mutation in *RPO21/RPB1*, shows both *SEN1* and *sen1-3* cells display growth defects in a *rpb1-1* background. Strains used in these experiments: CS2876, CS2953, CS4329, CS4331.

An alternative approach is required to try and understand whether Sen1 could be involved in removing active transcription complexes at replisome collision sites. To examine whether the tethering of Sen1 to the replisome is important in a background with mutant RNAPII elongation rates, we crossed *sen1-3* with the mutant RNAPII alleles *rpb1-E1103G*^{FAST} and *rpb1-N488G*^{SLOW}. Interestingly, the *rpb1-E1103G*^{FAST} mutant, which has been shown to exhibit increased transcription readthrough, is temperature sensitive in combination with *sen1-3* (fig 4.12). Conversely, the *sen1-3* allele did not sensitise *rpb1-N488G*^{SLOW} cells to replication stress or increased temperatures (fig 4.12).

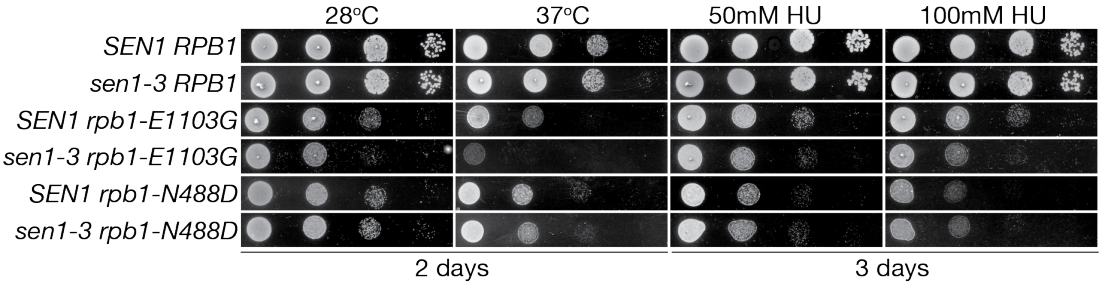


Figure 4.12 *sen1-3* is temperature sensitive in combination with *rpb1-E1103G* mutant. 10-fold dilutions of cell suspensions were spotted on YPD plates of various conditions and grown for several days at the described temperatures. Cells were imaged every 24 hours and assessed for *sen1-3* allele dependent growth defects. Strains used in this experiment: CS3669, CS3672, CS3675, CS3679, CS3771, CS3773.

Chapter Summary

In addition to the R loop accumulating strains described in the previous chapter, the *sen1-3* allele displays synthetic genetic defects in combination with various other deletion mutants, namely proteins involved in DNA replication and repair. These include Ctf4, which couple's DNA replication with other key processes; as well as components of the fork protection complex and the S phase checkpoint. Interestingly, it appears that Sen1 may play a role at replication forks that is independent from its role in R loops removal, as the growth defects and increased recombination in *sen1-3 mrc1*Δ cells cannot be suppressed by overexpression of either yeast or human *RNH1*. The reliance of *sen1-3* cells on these checkpoint or fork protection proteins suggests that when Sen1 is not readily available at the fork to carry out its required function, this could lead to more persistent fork stalling, recombination and genome instability. Considering the widely reported role of Sen1 in the termination of RNAPII transcription, we attempted to examine whether Sen1 could act at replication forks to remove RNAPII from DNA at sites of transcription-replication collisions, clearing the way for the replication fork, however further work is required to test this hypothesis.

Chapter 5: Results

MAPPING THE INTERACTION BETWEEN SEN1, CTF4 AND MRC1

5.1 Background

As detailed in chapter 1.9.5 & 6, previous work in the lab identified that Sen1 co-purifies with the eukaryotic replisome specifically during S phase. This interaction appears to be mediated by Ctf4 and Mrc1, however other currently unidentified binding partners may also exist. An important question is framed by the observation that Sen1 N-terminal is able to bind Ctf4 independently from the cell cycle, but full length Sen1 does not interact with Ctf4 outside of S phase. This suggests that there might be some kind of cell cycle dependent regulation of the interaction, the mechanism of which is currently unknown. It could, for example; be the result of an S phase dependent conformational change, a post translational modification of Sen1 or its binding partners such as phosphorylation by an S phase kinase, or an event that only occurs downstream of replisome formation. It is also currently unclear whether the recruitment of Sen1 to forks is a result of co-operation between Ctf4, Mrc1 and any other potential partners to tether a single Sen1 molecule to the replisome, or if several Sen1 molecules are individually recruited by different binding partners, which may have different affinities or compete with one another.

Consequently, I carried out the work described in this chapter to further elucidate the mechanism of the association between Sen1 and the replisome, with the aim to answer some of the many outstanding questions, including how the interaction is regulated by the cell cycle. In addition, work was carried out to map the interaction from the side of the replisome, by identifying the minimal interacting domains of both Ctf4 and Mrc1 for the binding of Sen1 N-terminal. This was conducted with the aim to generate mutants within these domains that break the interaction, which would allow us to test if these mutants phenocopy *sen1-3*.

5.2 Sen1 N-terminal interacts with Ctf4 and Mrc1 during G1

First, immunoprecipitation experiments were performed to validate that the N-terminal domain of Sen1 is able to bind Ctf4 and Mrc1 during G1. Exponentially growing cells carrying the various N-terminal Sen1 fragments (fig 5.1a) were arrested in YPRaf at 24ºC, before expression of the constructs was induced by shifting to YPGal for 35 minutes while maintaining G1 arrest.

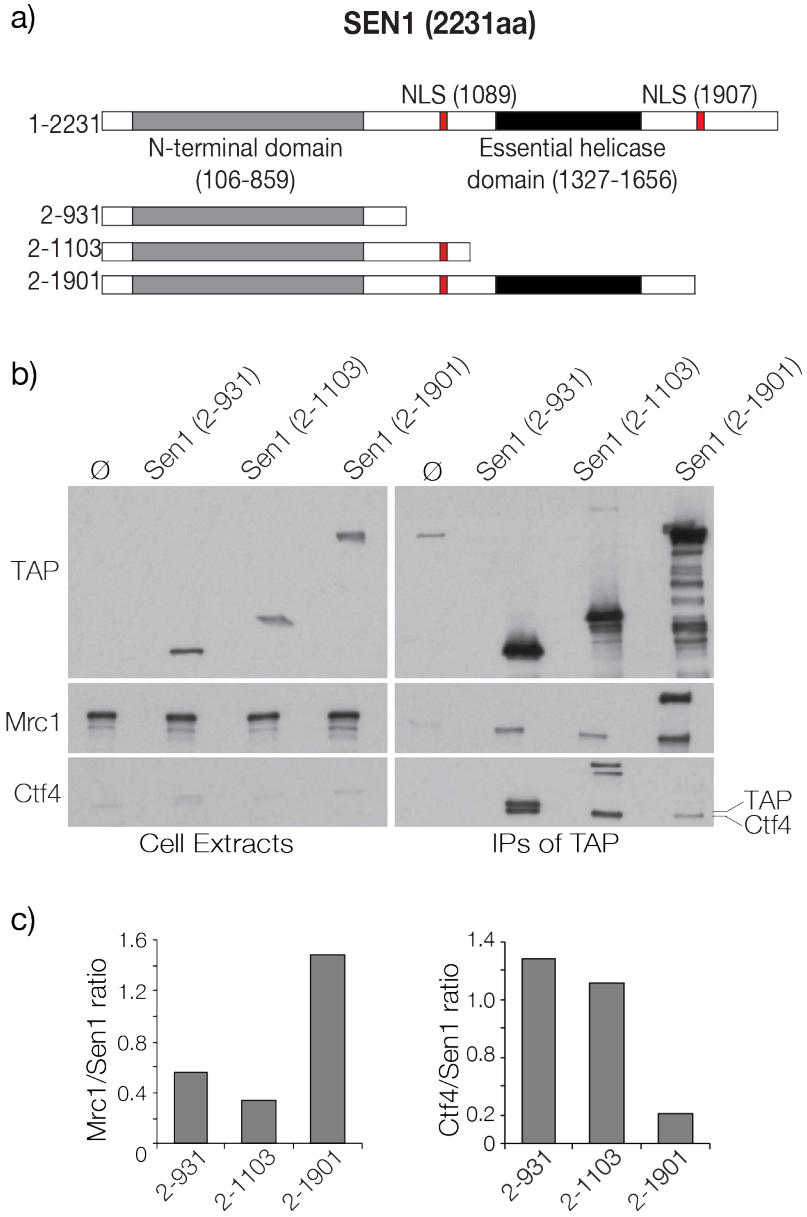


Figure 5.1 The N-terminal of Sen1 is both necessary and sufficient for the interaction with replisome components Mrc1 and Ctf4 during G1. a) Schematic of the N-terminal Sen1 fragments b) 250ml cultures of cells carrying the different Sen1 N-terminal fragments tagged with TAP under control of the *GAL1* promoter (cloned at the ectopic *LEU2* locus) were grown

at 24ºC in YPRaf to a density of 0.7x10^7 cells/ml. Cultures were arrested in G1 by the addition of alpha factor for 3 hours. Next, the medium was substituted for YPGal supplemented with 7.5μg/ml alpha factor for 35 mins to induce expression of the constructs. The cells were then harvested in G1, and IPs performed at 50mM salt using TAP beads to assess replisome component binding. Immunoblot analysis of Ctf4 and Mrc1 interaction with Sen1 (2-931), (2-1103) and (2-1901) during G1 is shown. c) Quantification of the immunoblot signal shown in (b), presented as the ratio of Mrc1 or Ctf4 signal versus the TAP signal. Strains used in this experiment: CS1852, CS1957, CS1956, CS1942.

IP and immunoblot analysis of the cell lysates revealed that during G1, Sen1 (2-931) is indeed necessary and sufficient for co-purification of both Ctf4 and Mrc1. Interestingly, these blots also showed that the affinity for Mrc1 appears to increase as the length of the Sen1 fragment increases and vice versa for Ctf4 (fig 5.1b). This was confirmed by quantification of the relative IP signal, presented as a ratio of either Mrc1 or Ctf4 IP versus the TAP IP (fig 5.1c). These pulldowns were then repeated in strains with deletion of either *CTF4* or *MRC1*, to confirm whether one partner still co-purifies with Sen1 in the absence of the other.

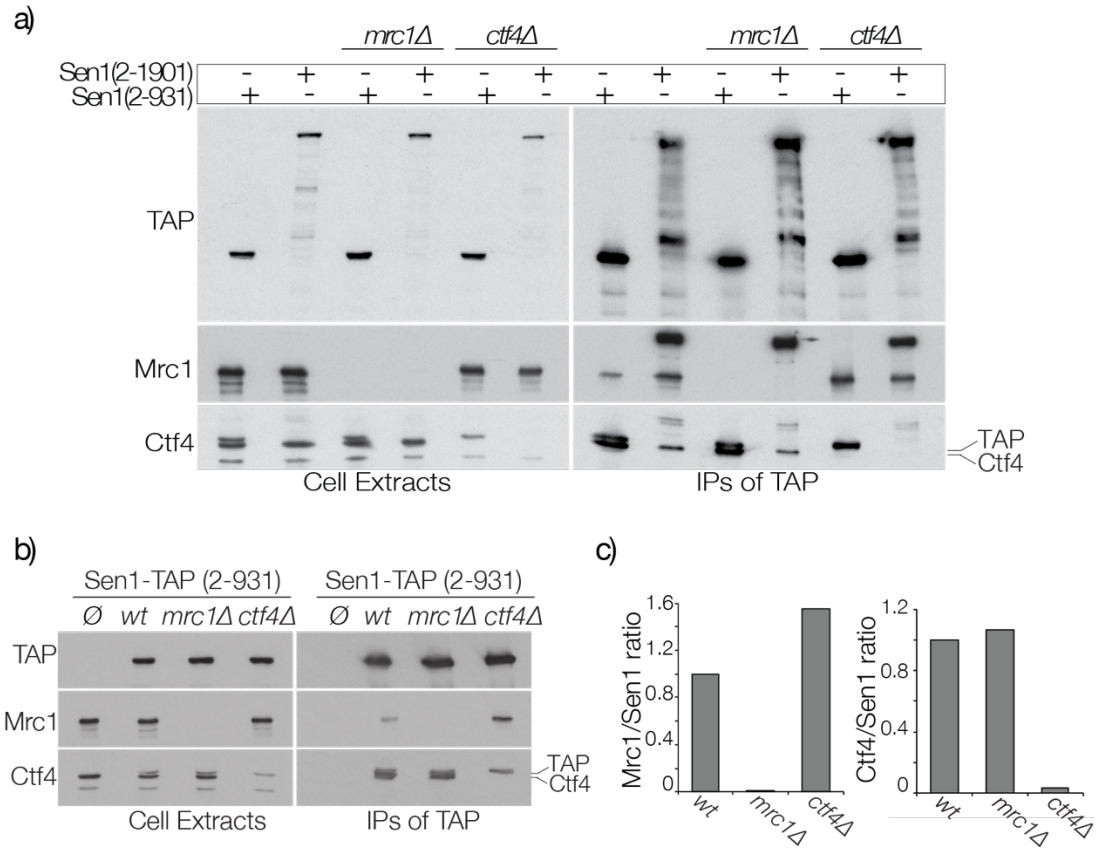


Figure 5.2 The affinity of Sen1 N-terminal for Ctf4 is unaffected by deletion of *MRC1*, however the binding of Mrc1 to Sen1 N-terminal increases in the absence of *CTF4*. Sen1 N-terminal fragments (2-931) and (2-1901) were grown at 24ºC in YPRaf and arrested in G1. Cells were shifted to YPGal for 35 mins to induce expression of the constructs while being maintained in G1. Cells were harvested to perform IPs using magnetic TAP beads a) IPs of

Sen1 (2-931) and (2-1901) in the absence of either Ctf4 or Mrc1 were electrophoretically separated and probed for Mrc1 and Ctf4 binding. b) Immunoblot analysis of whole cell extract and IPs of Sen1 (2-931) in a wt, *mrc1Δ* or *ctf4Δ* background. c) Quantification of the immunoblot signal shown in (b), where the ratio of Ctf4 or Mrc1 signal versus the TAP signal is normalised against the wild type. Strains used in these experiments: CS1957, CS1942, CS3186, CS3187, CS3281, CS3188.

In agreement with the previous experiment (fig 5.1), these IPs again show that Sen1 (2-931) has a greater affinity for Ctf4 compared to Sen1 (2-1901), while the opposite is true for Mrc1 (fig 5.2a). In a background with deletion of one binding partner, the other is still able to bind, indicating that one is not associating indirectly with Sen1 through the other.

Interestingly, the affinity of Sen1 (2-931) for Mrc1 appears to increase in the absence of *CTF4* (fig 5.2a), which was also observed in S phase cell extracts (Appanah et al., 2020). Conversely, the affinity of Ctf4 appears to be unaffected by deletion of *MRC1* (fig 5.2a). To confirm this result, Sen1 (2-931) G1 IPs were repeated and the bands quantified. The relative signal is shown as a ratio of Ctf4 or Mrc1 versus TAP, normalised to the wild type. This revealed that Mrc1 binding is increased by ~60% in *ctf4Δ* cells versus wild type, while the level of Ctf4 binding remains similar to the wild type in an *mrc1Δ* background (fig 5.2 b,c). This raises the question of whether these two proteins are competing with one another for Sen1 binding.

5.3 Mapping the minimal interacting domains of Mrc1 and Ctf4 for Sen1 binding

5.3.1 Mapping the minimal interacting domain of Ctf4

In order to further characterise the mechanism of interaction, a series of truncated forms of both Mrc1 and Ctf4 were generated to map the minimal domains responsible for the association with Sen1.

As an important interaction hub within the replisome, the structure of Ctf4 has been of great interest. Bioinformatics, X-ray crystallography and electron microscopy have shown that Ctf4 consists of an N-terminal WD40 domain, which is thought to mediate protein-protein interactions (Gambus et al., 2009) and a C-terminal β-propeller domain

that is fused to an α-helical bundle (fig 1.6, 5.3) (Simon et al., 2014). Ctf4 exists *in vivo* as a trimer, where self-association is mediated by the β-propeller domain. The interactions with DNA polymerase α and the CMG are a result of these proteins docking onto the C-terminal helical extensions which project from the trimer (Simon et al., 2014).

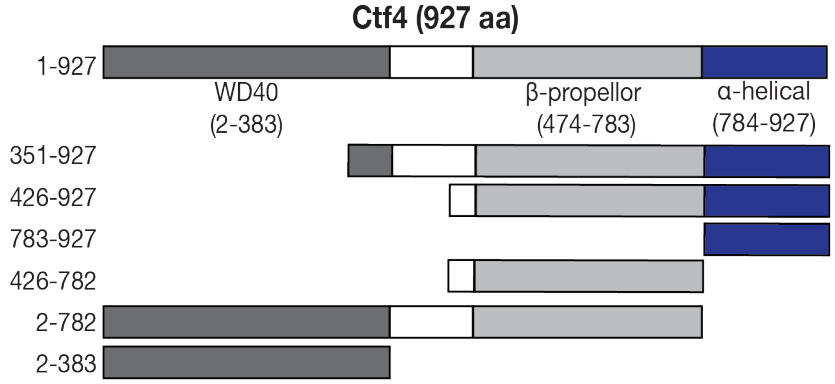


Figure 5.3 Schematic of the various N-terminally tagged Ctf4 fragments. Fragments were cloned under the *GAL1* promoter at the ectopic *LEU2* locus.

We aimed to determine which of these domains is responsible for the tethering of Sen1 (via the interaction with its N-terminal) to the replisome. Consequently, fragments of Ctf4 were designed to span individual or combinations of these domains, with a *3HA* tag fused to their N-terminal (fig 5.3). These constructs were then cloned at the ectopic *LEU2* locus under control of the *GAL1* promoter.

As the Sen1 (2-931) fragment appears to bind Ctf4 with the highest affinity (fig 5.1, 5.2), the Cft4 fragments were crossed into a strain carrying *GAL1*-TAP-Sen1 (2-931) to obtain diploid cells carrying both constructs. To achieve a sufficient volume of cells for subsequent immunoprecipitation, 250mls of asynchronously cycling cultures were grown to a density of $1.0x10^7$ cells/ml, before shifting to YPGal for 2 hours to induce expression of the constructs. The cells were then harvested and IPed using magnetic TAP beads. Immunoblot analysis was performed using an antibody targeted against the 3HA.

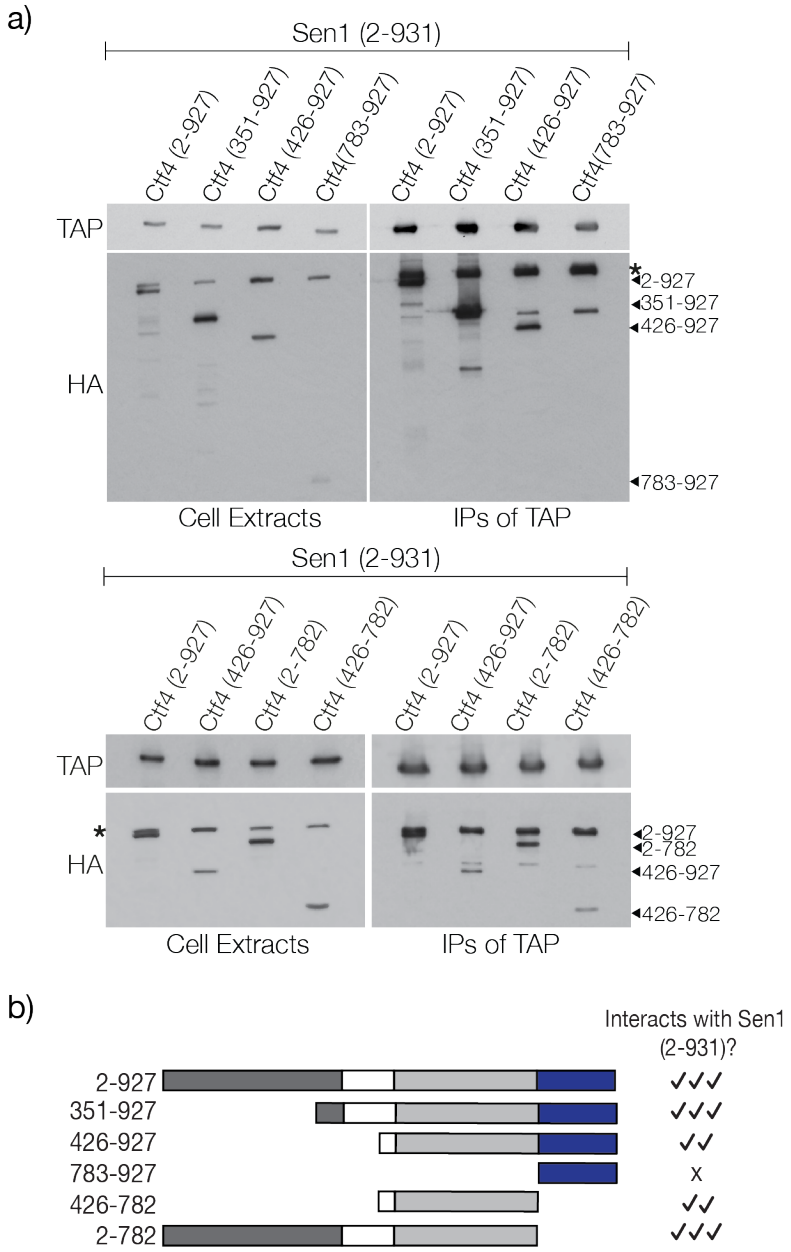


Figure 5.4 Mapping the minimal domain of Ctf4 necessary and sufficient to bind Sen1 N-terminal. a) 250ml cultures of diploid cells carrying Sen1 (2-931) with the different Ctf4 fragments were grown at 24°C to a density of 1.0x10^7 cells/ml in YPRaf. Cells were then induced for 2 hours in YPGal, before the asynchronous cultures were harvested and IPed using TAP beads. b) Immunoblot analysis of Sen1 (2-931) ability to bind the various different Ctf4 fragments. The * indicates non-specific recognition of the TAP by the HA antibody. Arrows indicate the Ctf4 fragments. b) Schematic of the ability of the various Ctf4 fragments to bind Sen1 (2-931). Strains used in this experiment: CS3563, CS3573, CS3574, CS3575, CS4110, CS4108.

Previous data indicated that the N-terminal WD40 domain of Ctf4 is not responsible for Sen1 binding (residues 2-383, R. Appanah, G. De Piccoli, unpublished). Interestingly, all of the positive fragments in these initial IPs spanned the β-propeller domain, raising the possibility this domain may mediate the binding (fig 5.4a,b). Only

Ctf4 (783-927) comprising the α-helical domain appeared unable to bind to Sen1 (fig 5.4a,b). Although this fragment was less well expressed than the others, even at longer exposures, no IP signal was detected. IPs of a candidate fragment crossed with *GAL1*-TAP_empty confirmed that the 3HA tagged fragments do not interact non-specifically with the TAP tag in the absence of the Sen1 construct (fig S.2).

In these cells, while the fragments are overexpressed under control of the strong *GAL1* promoter (from the ectopic *LEU2* locus), full length Ctf4 is also still expressed from its endogenous loci at a comparably lower physiological level. As the β-propeller domain is responsible for the formation of Ctf4 trimers, the positive IPs were repeated in a *ctf4Δ* background to ensure that binding was not a result of the 3HA tagged fragments self-associating with and binding to Sen1 through endogenously expressed full length Ctf4. Asynchronous cultures of haploid cells carrying Sen1 (2-931), Sen1 (2-931) *ctf4Δ* or the various Ctf4 fragments in a *ctf4Δ* background were induced and harvested. The Sen1 and Ctf4 lysates were then mixed in 1:1 ratio for TAP pulldown to compare the binding of the fragments with or without the presence of endogenous full length Ctf4 (see fig 5.5a for schematic). All of the fragments tested were expressed at similar levels, as shown by the whole cell extracts (fig 5.5b). In a *ctf4Δ* background, Ctf4 (351-927) appeared to IP the strongest with Sen1 (2-931), followed by a weaker interaction with Ctf4 (2-782). In comparison, Ctf4 (426-927) and Ctf4 (426-782), which extend from the β-propeller, IPed very weakly; the interaction was barely detectable, even at longer exposures. This discrepancy, where binding of these fragments in an endogenous *CTF4+* background was more easily observed in the previous IPs (fig 5.4a), could a be result of this experimental set up, where 1:1 mixing of the lysates meant that the samples were diluted by one half.

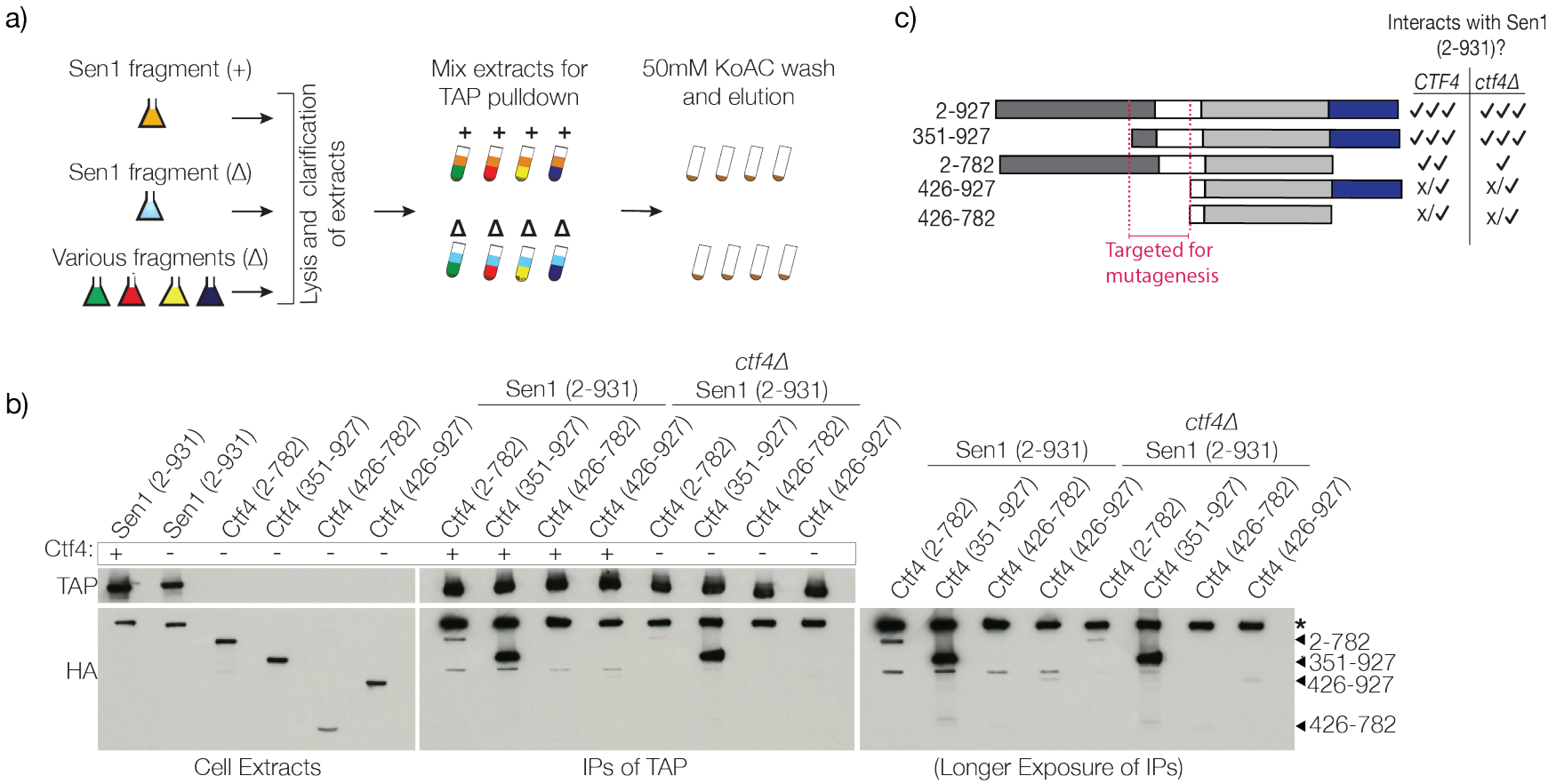


Figure 5.5 Identification of Ctf4 fragments that retain the ability to bind Sen1 N-terminal in a *ctf4Δ* background. a) Schematic of the IPs. 1L cultures of haploid cells carrying Sen1 (2-931) in a background with or without deletion of endogenous *CTF4*, and 250ml cultures of the different Ctf4 fragments in a *ctf4Δ* background were grown at 24°C in YPRaf to a density of $1x10^7$, then induced for 2 hours in YPGal. The cells were harvested, followed by lysis and clarification of the extracts. Each Sen1 extract was split into 4 aliquots, and an equal volume of each Ctf4 fragment was mixed with the Sen1 (2-931) and Sen1 (2-931)

ctf4Δ lysates. The mixed extracts were incubated with TAP beads for 2 hours at 4°C, before washing with 50mM KoAC buffer and eluting. b) Immunoblot analysis of the cell extracts and IPs described in (a). The * indicates non-specific recognition of the TAP by the HA antibody. Arrows indicate the Ctf4 fragments. c) Schematic of the ability of the various fragments to bind Sen1 (2-931). Strains used in this experiment: CS1957, CS3281, CS4391, CS4410, CS4228, CS4018.

Ctf4 (351-927) and Ctf4 (2-782) uniquely contain the residues 351-426 compared with the other fragments (fig 5.5c). Thus, in *ctf4Δ* cells where there is no interference from the endogenous full-length protein, residues 351-426 may be important for binding Sen1 (2-931), not the β-propeller as previously hypothesised. With the aim to create a mutant of Ctf4 that is no longer able to bind Sen1, this region was screened for conserved residues. Conservation was assessed by performing multiple sequence alignments of the *S. cerevisiase* Ctf4 sequence with the closest orthologues in other yeasts as well as mice and humans (fig 5.6).

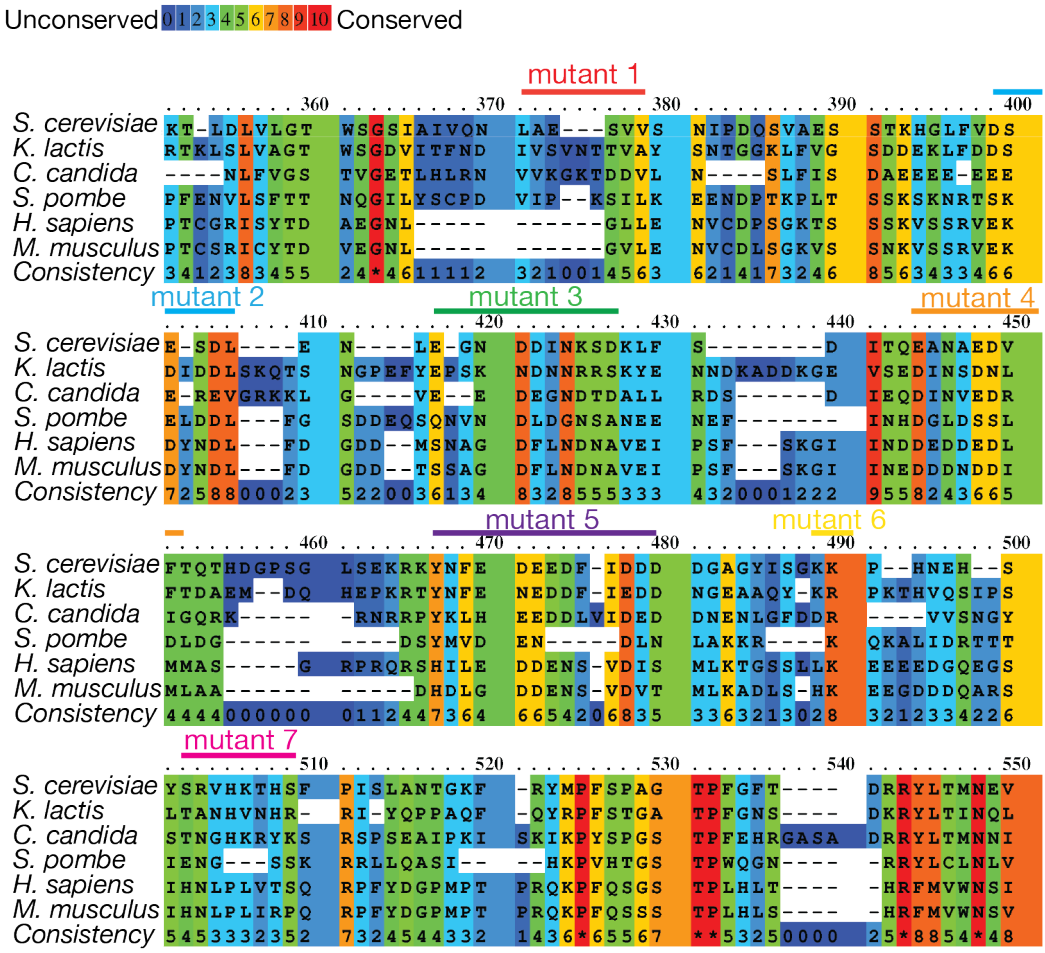


Figure 5.6 Alignment of Ctf4 with its orthologues in other eukaryotes. Mutations were designed against conserved residues within the proposed minimal interacting fragment of Ctf4. This alignment was performed using the Praline software.

In an attempt to disrupt the interaction between Ctf4 and Sen1, mutants were designed within the identified conserved regions. The conserved amino acids at the described locations (table 5.1, fig 5.6) were mutated interchangeably to either glycine or alanine.

Table 5.1 Mutations designed against the conserved residues within the proposed Ctf4 minimal interacting domain

Mutant	Mutated residues
ctf4-1	*L351G A352G V355A V356G*
ctf4-2	*D376A S377G E278A S379G D380A L381G*
ctf4-3	*E385A N387A D388G D389G N391A K392A D394G*
ctf4-4	*E403A A404G N405A A406G E407A D408A V409A F410A*
ctf4-5	*Y426A F428A E429G D4230A E4231G E4232A D4233G I4235G D4236A D4237A D4238A D4239A*
ctf4-6	*K447A K448A*
ctf4-7	*S456G H459A S463G*

A strategy was employed whereby the mutant DNA was commercially synthesised, then cloned into a plasmid carrying the *GAL1-3HA*-Ctf4 (351-927) sequence, previously shown to interact strongly with Sen1 (2-931) (fig 5.5). The mutants were transformed into yeast and integrated at the ectopic *LEU2* locus. The *ctf4Δ* allele was then crossed into these cells to eliminate the wild type endogenous protein. Expression of the constructs was confirmed by TCA protein extraction (data not shown), and each clone verified by sequencing.

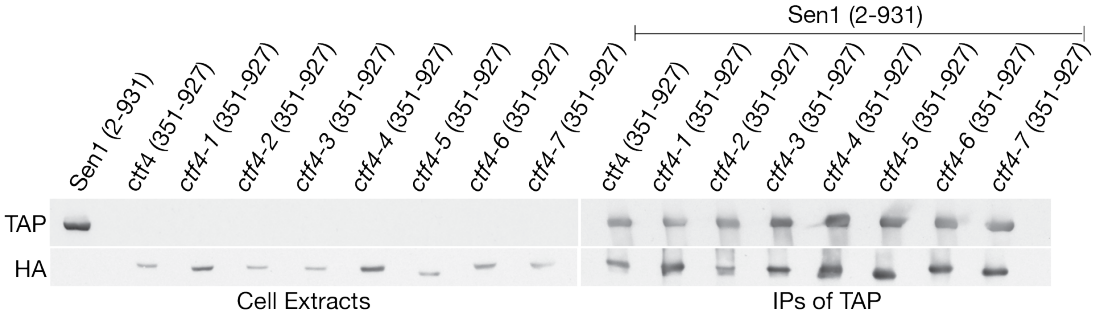


Figure 5.7 Attempting to break the interaction between Sen1 (2-931) and Ctf4. Asynchronous cultures of cells carrying either *GAL1*-TAP-Sen1 (2-931) or the *GAL1*-3HA-Ctf4 (351-927) mutant variants in a *ctf4Δ* background were grown to a density of 1.0x107 in YPGal at 24ºC before being harvested for IP. Mutant cell lysates were mixed in a 1:1 ratio with the TAP lysates and IPed using magnetic TAP beds. The IPs were assessed by immunoblotting for TAP and HA. Strains used in this experiment: CS3281, CS4410, CS4428, CS4429, CS4430, CS4431, CS4432, CS4433, CS4434.

Each mutant was screened by IP to determine whether they maintained the ability to bind Sen1 (2-931). Following the same strategy as described in fig 5.5a, asynchronous

cultures of induced cells carrying either *GAL1*-TAP-Sen1 (2-931) or the various 3HA tagged Ctf4 mutant fragments were mixed in a 1:1 ratio, and IPed using magnetic TAP beads. Unfortunately, probing for the 3HA tag revealed that none of the mutations disrupted the interaction with Sen1 (fig 5.7). Further work will be required to identify the region of Ctf4 important for binding to Sen1.

5.3.2 Mapping the minimal interacting domain of Mrc1

The strategy used to map the association of Ctf4 with Sen1 was also used to identify the region of Mrc1 responsible for binding. A series of Mrc1 constructs, truncated at ether the N-terminal, C-terminal or both, fused to an N-terminal *3HA* under control of the *GAL1* promoter were cloned at the ectopic *LEU2* loci (fig 5.8). However, in this case, these fragments were crossed with cells carrying Sen1 (2-1901), as this was previously identified as the fragment with the highest affinity for Mrc1 (fig 5.1, 5.2).

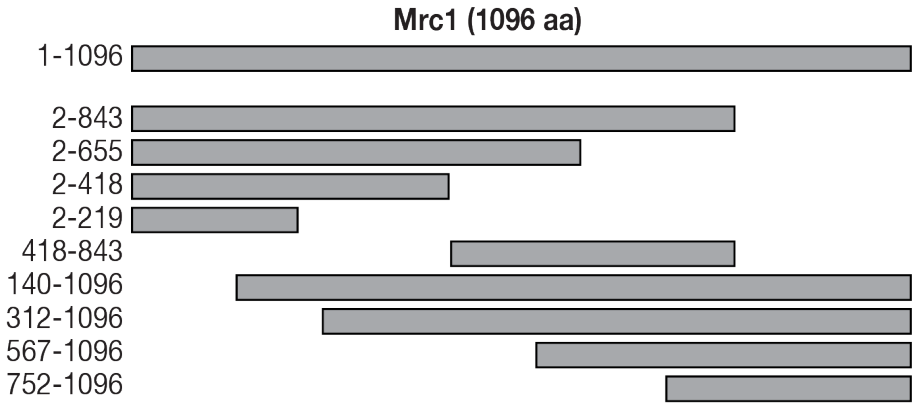


Figure 5.8 Schematic of the various N-terminally tagged Mrc1 fragments. Fragments were cloned under the *GAL1* promoter at the ectopic *LEU2* locus.

Asynchronous cultures of the diploids were induced and harvested for IP, as described in section 5.3.1. Probing the samples for 3HA revealed that all of the fragments were expressed to a reasonably similar extent, as shown by the whole cell extracts (fig 5.9). The IPs showed that Mrc1 (2-219) was negative for binding. All other fragments co-purified with Sen1 (2-1901) to varying extents. Mrc1 (567-1096) was the strongest interacting fragment, with affinity to similar to the full-length protein (fig 5.9). Considering both N- and C- terminally truncated fragments were positive for binding,

of which some do not have any overlapping region, this raised the possibility that Mrc1 is able to bind to Sen1 (2-1901) through both its N- and C-terminal.

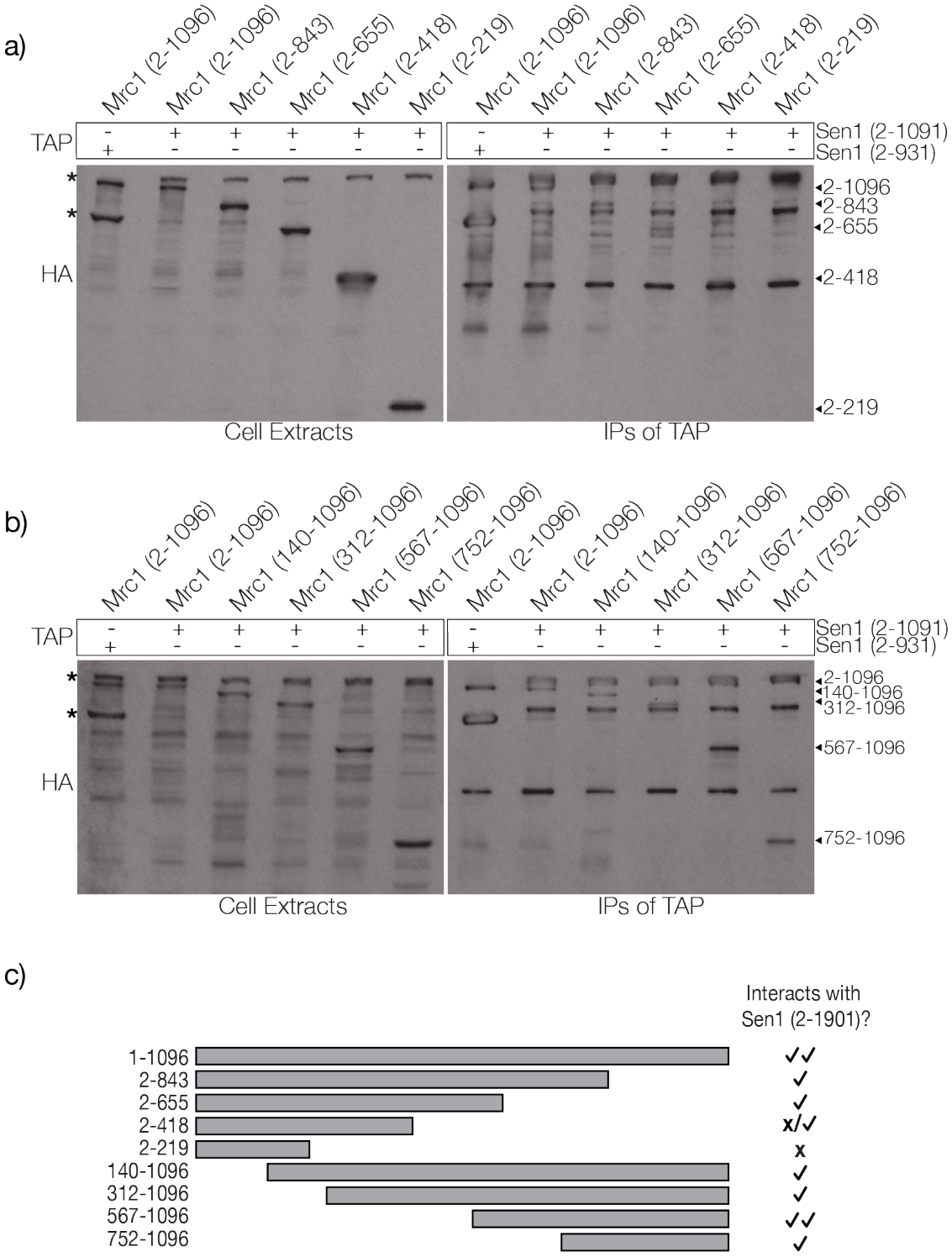


Figure 5.9 Mapping the minimal domain of Mrc1 necessary and sufficient to bind Sen1 N-terminal. 250ml cultures of diploid cells carrying Sen1 (2-1901) with the different Mrc1 fragments were grown at 24ºC to a density of $1.0x10^7$ cells/ml in YPRaf. Cells were then induced for 2 hours in YPGal, before the asynchronous cultures were harvested for IPs using TAP beads. The ability of the a) C-terminal and b) N-terminal Mrc1 fragments to bind Sen1

(2-1901) was analysed by immunoblotting. Arrows indicate the Mrc1 fragments. c) Schematic of the ability of the various Mrc1 fragments to bind Sen1 (2-1901). Strains used in this experiment: CS3591, CS3568, CS3567, CS3588, CS3566, CS3587, CS3569, CS3570, CS3589, CS3571.

IPs of the positive fragments were then repeated in an *mrc1Δ* background (experimental setup as in fig 5.5a), to test whether binding is still observed in the absence of the endogenous full-length protein. In addition, after considering these data; a new fragment was generated, Mrc1 (418-843), to investigate whether any binding is retained following loss of both the extreme N-terminal (2-418) and C-terminal residues (843-1096), both of which were found to bind.

All of the previously tested positive fragments co-purified with Sen1 (2-1901) in an *mrc1Δ* background, and interestingly, deletion of the endogenous full-length protein appeared to somewhat increase the binding of the N-terminal 3HA tagged fragments, suggesting there could be some intramolecular competition (fig 5.10). The Mrc1 (418-843) fragment did not co-precipitate, which suggests that indeed, both the N- and C-terminal of Mrc1 are able to bind to Sen1, though it appears the N-terminal does so with higher affinity in an *mrc1Δ* background.

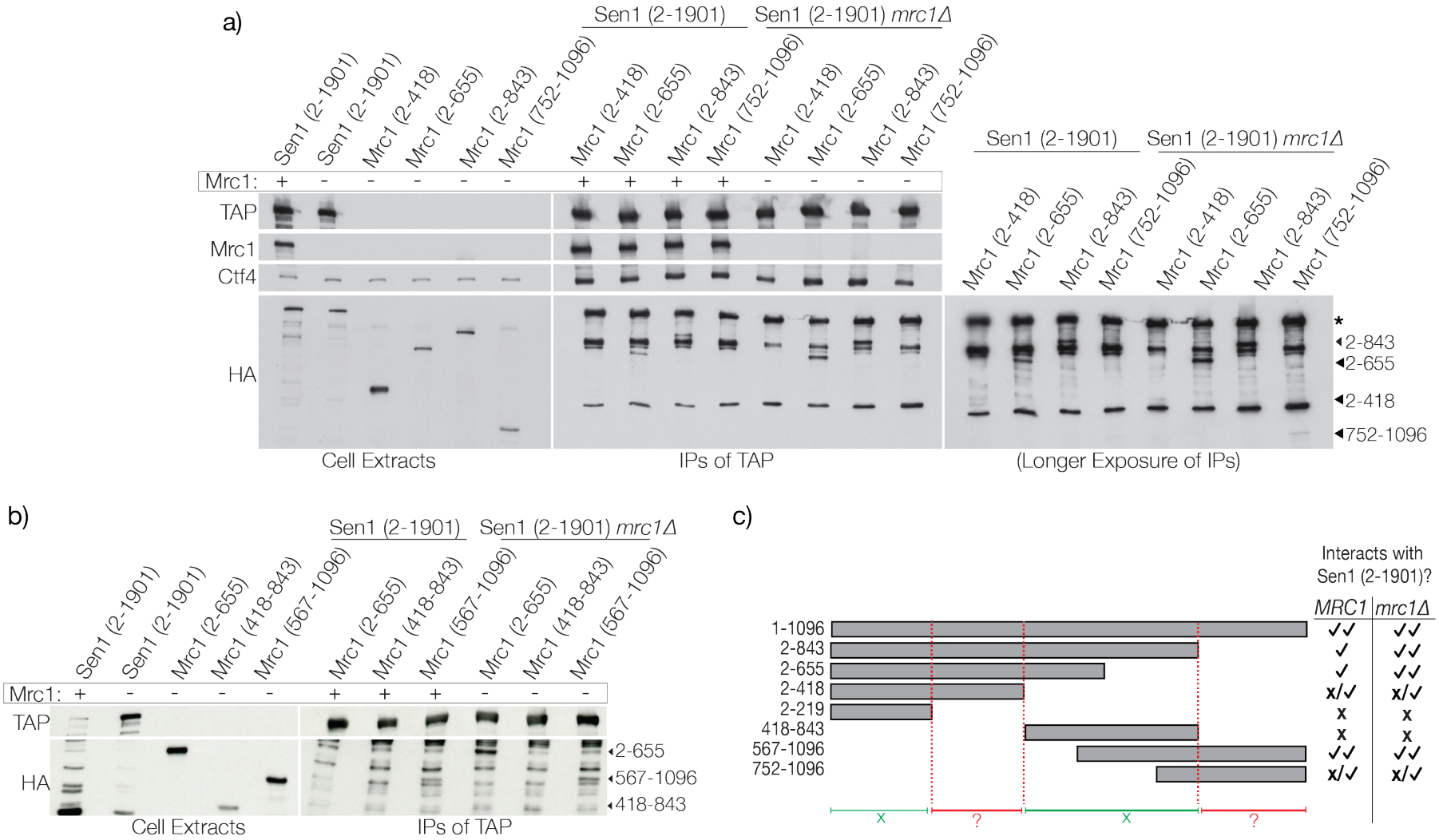

a)
Sen1 (2-1901)
Sen1 (2-1901) mrc1Δ
Sen1 (2-1901)
Sen1 (2-1901)
Mrc1 (2-418)
Mrc1 (2-655)
Mrc1 (2-843)
Mrc1 (752-1096)
Mrc1:
TAP
Mrc1
Ctf4
HA
Cell Extracts
IPs of TAP
(Longer Exposure of IPs)
*
2-843
2-655
2-418
752-1096
b)
Mrc1 (418-843)
Mrc1 (567-1096)
2-655
567-1096
418-843
c)
Interacts with Sen1 (2-1901)?
MRC1
mrc1Δ
1-1096
2-843
2-655
2-418
2-219
418-843
567-1096
752-1096
✓✓
✓
x/✓
x
?

Figure 5.10 Identification of Mrc1 fragments that retain the ability to bind Sen1 N-terminal in a *mrc1Δ* background. 1L cultures of haploid cells carrying Sen1 (2-1901) in a background with or without deletion of endogenous *MRC1*, and 250ml cultures of the different Mrc1 fragments in a *mrc1Δ* background were grown at 24°C to a density of $1x10^7$, then induced for 2 hours in YPGal. The cells were harvested, followed by lysis and clarification of the extracts. Each Sen1 extract was split into 4 aliquots, and an equal volume of each Mrc1 fragment was mixed with the Sen1 (2-1901) and Sen1 (2-1901) *mrc1Δ* lysates. The mixed extracts were incubated with TAP beads for 2 hours at 4°C, before washing with 50mM KoAC buffer and eluting. Immunoblot analysis of the cell extracts and IPs are shown in (a) and (b). The * indicates non-specific recognition of the TAP by the HA antibody. Arrows indicate the Mrc1 fragments. c) Schematic of the ability of the various Mrc1 fragments to bind Sen1 (2-1901). Strains used in this experiment: CS1942, CS3187, CS4083, CS4042, CS4081, CS4022, CS4376, CS3313.

5.3.3 Full length Sen1 retains some affinity for Mrc1 during G1, and Sen1 lacking its N-terminal domain retains some affinity for DNA polymerase epsilon during G1

Interestingly, new data has also shown that full length Sen1 retains some affinity for Mrc1 during G1, which is comparatively stronger compared with G1 IPs of Sen1 lacking its C-terminal (fig 5.11). This adds a further level of complexity and leads to the question of whether Mrc1 may also be able to bind a region in the C-terminal of Sen1, in addition to the N-terminal replisome binding site. Indeed, multiple Mrc1 binding sites throughout the Sen1 protein could be a possible explanation for the higher affinity of Sen1 (2-1901) for Mrc1, compared with the shorter Sen1 (2-931) fragment (fig 5.1 & 5.2).

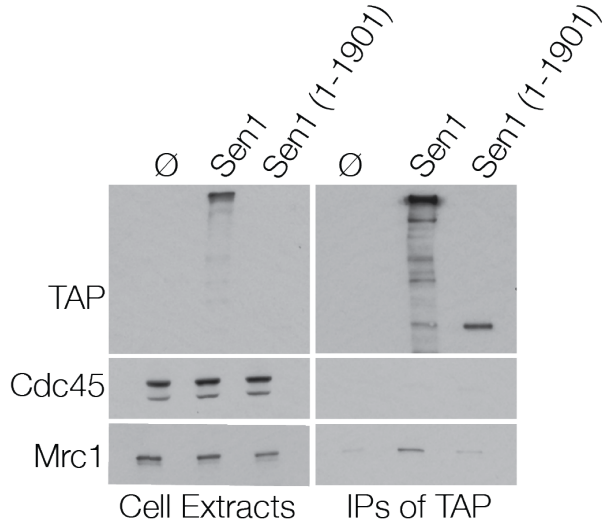


Figure 5.11 Full length Sen1 has some affinity for Mrc1 during G1. a) 1L cultures of cells carrying TAP tagged Sen1 or Sen1 (2-1901) inserted at the chromosomal *SEN1* locus were grown at 24°C to a density of $0.7x10^7$ in YPD, then arrested in G1 for harvest. IPs were performed at 50mM salt using magnetic TAP beads, and the samples were analysed by immunoblotting. Strains used in this experiment: CS74, CS2222, CS2201.

IPs of G1 and S phase cell extracts carrying either Sen1 N-terminal (Sen1 (2-931)) or an N-terminally truncated version of Sen1 (Sen1 (931-2231)) were probed for various replisome components to assess binding. As expected, no interaction with Ctf4 was observed in G1 or S phase extracts of Sen1 (931-2231), and candidate components of the core replisome only co-purified with the N-terminal domain of Sen1 during S phase (fig 5.12a).

Assessing Mrc1 binding to the 931-2231 fragment was inconclusive in this first IP, as the TAP tagged Sen1 (931-2231) fragment runs at a similar molecular weight to Mrc1, and so some of the observed signal could be a result of non-specific recognition of the TAP by the Mrc1 antibody (fig 5.12a). Repetition of these IPs showed that the Mrc1 signal in the Sen1 (931-2231) IP was unaffected by deletion of *MRC1*, thus this signal appears to be a result of non-specific recognition of the TAP and so may not be indicative of Mrc1 binding (fig 5.12b). Interestingly, we observed that Sen1 (931-2231) co-precipitates with Pol2 and Dpb2, subunits of DNA polymerase epsilon, outside of the context of the replisome during both G1 and S phase (fig 5.12a,b). Pol ε and Mrc1 interact throughout the cell cycle, and it appears that Pol ε binds Sen1 (931-2231) independently from Mrc1 and the replisome, as its binding is still observed in *mrc1Δ* cells.

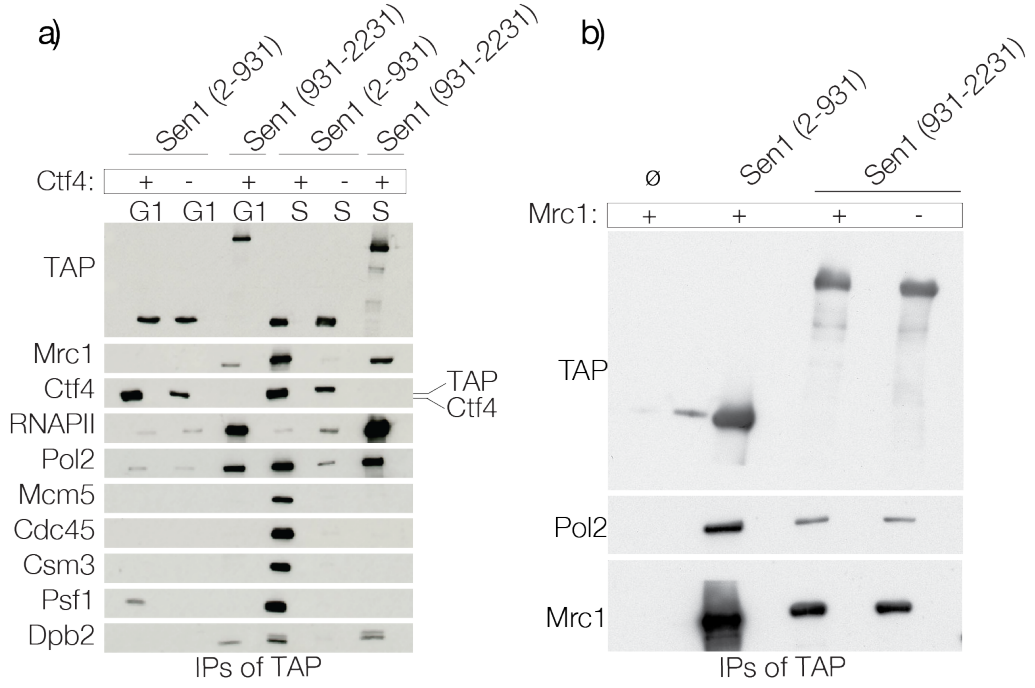


Figure 5.12 Sen1 (931-2231) binds to Pol ε outside of the context of the replisome a) 1L cultures of various TAP tagged Sen1 fragments under control of the *GAL1* promoter were grown to a density of 0.7x10^7 cells/ml at 24°C in YPRaf. Cells were then arrested in G1 by the

addition of alpha factor for 3 hours before shifting to YPGal supplemented with 7.5μg/ml alpha factor for 35 mins to induce expression of the constructs. At this stage, G1 cells were harvested for IP. The remaining cultures were washed twice with YPGal to remove any trace of alpha factor, released into S phase at 24°C for 30mins in fresh YPGal, then harvested for IP using TAP beads. These strains were assessed for binding of various replisome components via immunoblot analysis. b) IPs of S phase extracts of Sen1 (2-931) and (931-2231) with or without Mrc1 were carried out as in (a). Strains used in these experiments: CS1957, CS3281, CS1943, CS4377.

In these IPs, we also observed a strong interaction between Sen1 (931-2231) and RNAPII (Rpo21), during both G1 and S phase (fig 5.12a). It has previously been reported that Sen1 N-terminal is responsible for the interaction with RNAPII (Ursic et al., 2004, Chinchilla et al., 2012), however a recent study also detected a robust interaction between RNAPII and Sen1 lacking its N-terminal domain (Han et al., 2020), which in agreement with this IP, suggests that regions outside of Sen1 N-terminal can also mediate an interaction with RNAPII.

5.4 Investigating how the interaction between Sen1, Ctf4 and Mrc1 is regulated by the cell cycle

As detailed in the introductory section 1.9.5, previous work within the lab observed that Sen1 only co-IPs with the core replisome during S phase, predominantly through Ctf4 and Mrc1, and possibly also through other currently unidentified components.

It is possible that this cell cycle regulation is the result of either an S-phase dependent conformational change of Sen1; for example, a region of the C-terminal could block binding of the N-terminal outside of S phase. Indeed, it has recently been reported that the C-terminal domain of Sen1 contains a Nrd1-interacting motif that mimics phosphorylated CTD; and as the N-terminal domain recognises Ser5P-CTD, the N- and C- terminal can mediate intramolecular interactions (Han et al., 2020). This conformational change may be driven by post-translational modifications; interestingly, it has been reported that Sen1 is highly phosphorylated (Swaney et al., 2013, Bodenmiller et al., 2010, Helbig et al., 2010), and when immunoblotting for Sen1, the band is often observed as a smear rather than a sharp band. It could be that phosphorylation is a requirement for the interaction between Sen1 and the replisome, and as this is S phase specific, potential candidates could be the S phase kinases CDK/DDK. Alternatively, the binding of Ctf4 and Mrc1 could be a result of an S phase specific conformation of these proteins, where perhaps Sen1 binding sites are

exposed as a result of their conformation within the context of the replisome. Moreover, the association may be controlled by the binding or degradation of a regulatory factor.

First, to distinguish whether Sen1 itself, or something within the extract is responsible for mediating the binding, a 2 step IP was performed. In this experiment, TAP-tagged Sen1 was collected during G1 and S phase. A s*ld3-7-td* background, which at 37^{o}C, cannot initiate chromosome replication (as Cdc45 cannot be recruited to origins) was used (fig 5.13a). This renders the first step 'cleaner', as no additional replisome components should co-purify with Sen1 without replisome formation. Therefore, any components pulled down during the second step are not the result of a residual interaction from the first pulldown. The G1 or S phase conformations of Sen1-TAP were bound to magnetic beads and washed with a high salt buffer. Finally, the washed, Sen1-bound beads were incubated with G1 or S phase extracts of Ctf4 and Mrc1 tagged with *9MYC* (fig 5.13b).

Interestingly, the binding of Mrc1-9Myc and Ctf4-9Myc to Sen1 appears to depend on the extract being in S phase, independently of whether Sen1 was collected during G1 or S phase. This suggests that it is likely the association does not depend on some self-regulatory mechanism or intrinsic feature of Sen1, such as a conformation where the C-terminal blocks the N-terminal during G1. In agreement with the previous data, affinity for Mrc1 is maintained regardless of the cell cycle phase (fig 5.13c).

As the cell cycle specific association does not appear to depend on a distinct conformation of Sen1, it is possible that an S phase specific post-translational modification occurring at replication forks is required for the binding of Sen1 and its partners. Alternatively, a conformational change of Ctf4 or Mrc1 may occur in the context of the replisome during S phase that exposes their Sen1 binding sites in a manner that supports its association.

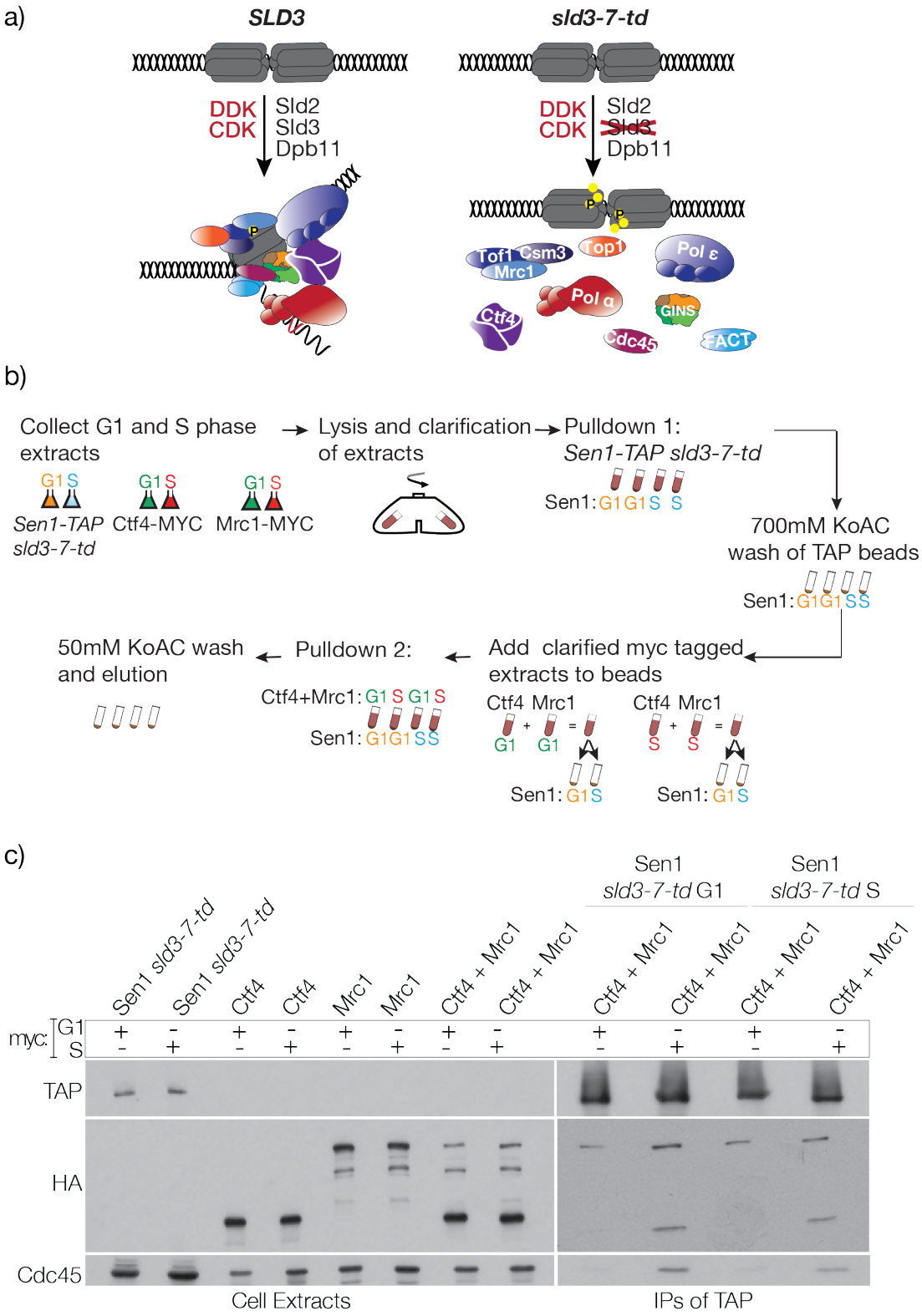


Figure 5.13 The interaction between Sen1 and the replisome appears to depend on the extract entering S phase. a) Schematic of the *sld3-7-td* allele, which prevents origin firing and replisome formation. b) Schematic of the IP shown in (c). 1L cultures of cells carrying the *SEN1-TAP*, *sld3-7-td* and *GAL1-UBR1* alleles were grown to a density of 0.7×10^7 in YPRaf, then arrested in G1 by the addition of alpha factor for 3 hours. Following this, cells were maintained in G1 by further addition of alpha factor to the media. The cells were first resuspended in YPGal for 35 mins to induce expression of Ubr1, then shifted to 37°C for 1

hour to degrade the degron tagged Sld5. G1 cells were harvested at this stage, or the cultures were washed twice with fresh YPGal to remove any trace of alpha factor and released for 20 mins at 37ºC to harvest during S phase. G1 and S phase cells carrying Mrc1 or Ctf4 tagged with *9MYC* were collected by arresting exponentially growing cells in YPD for 3 hours. G1 cells were harvested at this stage, or the cultures were washed and released for 30mins to collect the S phase cells. Two step IPs were then performed as in (b), where magnetic TAP beads were incubated with the Sen1 lysates for 2 hours before washing with 700mM KoAC buffer so only the TAP tagged Sen1 remained bound. The G1 or S Sen1 beads were then incubated with either G1 or S phase Ctf4 and Mrc1 extracts, before washing at 50mM salt. Immunoblot analysis of the cell extracts and IPs described in (b) are shown. Strains used in this experiment: CS2794, CS2801, CS3955.

In an attempt to understand whether formation of the replisome is required for a downstream event that promotes binding, or if the conformation of the replisome itself modulates the interaction; we tried to disassemble the replisome during S phase, to test whether Sen1 is still able to bind Ctf4 and Mrc1. To this aim, we fused the core replicative helicase component *MCM7* with an auxin degron. The construct was then crossed into a strain carrying the *TAP-SLD5* allele, and IPed to test the ability of the degron to destabilise the replisome. A 2L culture of cells was first grown in YPRaf and arrested in G1, before shifting to YPGal, maintaining G1 arrest for 35 mins to induce expression of the *TIR1* auxin receptor. Cells were then released into YPGal supplemented with 0.2M HU to arrest the cells during S phase. The S phase arrested cells were then split into two 1L cultures, where one was supplemented with 0.5mM auxin (in 70% ethanol) for 1 hour to degrade the *mcm7-IAA,* while the other was mock treated (equal volume of 70% ethanol was added). After 1 hour, the cells were harvested for IP. Encouragingly, in the presence of auxin; Cdc45, Pol2 and Pol1 no longer co-purified with Sld5, and the association with Mcm3 was greatly reduced (fig 5.14a). In addition, dilution spotting of the strain on YPGal supplemented with auxin revealed the cells were extremely sick (fig 5.14c). Therefore, the degron was crossed into a strain carrying *SEN1-TAP*. As before, cells were harvested during S phase, with or without auxin, and IPed to assess replisome binding. Unfortunately, when combined with *SEN1-TAP*; although dilution spotting showed growth in the presence of auxin was similar to that of *SLD5-TAP MCM7-IAA* cells, IPs revealed that the destabilisation of the replisome was no longer convincing. The experiment was repeated several times to exclude any experimental artefact and gave the same result each time. Although dilution spotting of the strain showed the cells were extremely sick, we speculate that the conformation within the replisome in cells carrying a *TAP-SLD5* allele might

favour the accessibility to the degron, thus causing a more efficient degradation of Mcm7 in the *TAP-SLD5* strain, sufficient to degrade most replisomes within 1 hour.

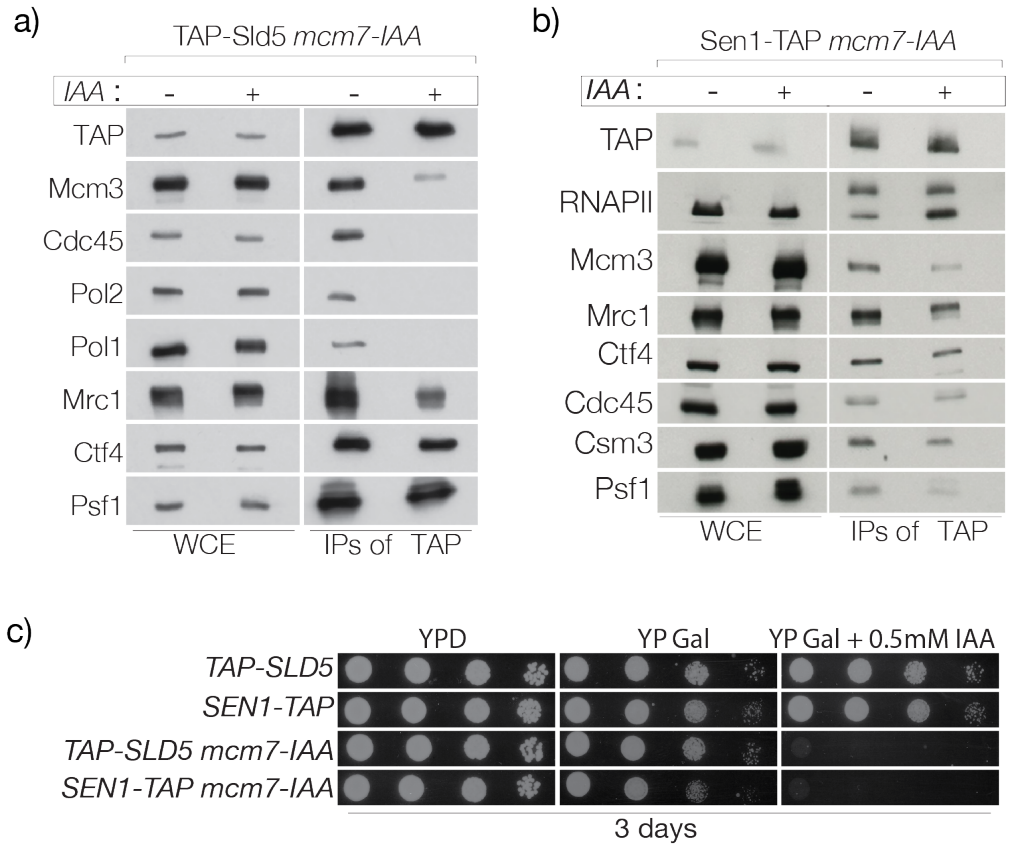


Figure 5.14 Using the auxin degron system to destabilise the replisome during S phase. 2L cultures of cells carrying core replicative helicase component *MCM7* tagged with a 3x mini auxin degron and the *GAL1-TIR1* receptor were grow in YPRaf at 24°C to a density of 0.7x107 cells/ml. Cells were then arrested by the addition of alpha factor for 3 hours, before shifting to YPGal supplemented with alpha factor for 35 mins to induce expression of the Tir1. Cells were subsequently released in YPGal supplemented with 0.2M HU to arrest the cells in S phase. The culture was split into two; 0.5mM auxin was then added to one culture for 1 hour to degrade IAA tagged Mcm7 and destabilise the replisome, while the other was mock treated. IPs were performed using TAP beads to assess replisome component binding. a) Immunoblot analysis of TAP-Sld5 IPs using this system, which show loss of replisome components Cdc45, Mcm3, Pol1 and Pol2 following the addition of auxin. b) Immunoblot analysis of Sen1-TAP IPs using this system. Strains used in these experiments: CS4400, CS4402.

We next decided to use an alternative approach and test whether phosphorylation, for example by an S phase specific kinase (CDK/DDK), is responsible for the association. All of the previous IPs were carried out in the presence of the phosphatase inhibitors sodium fluoride and sodium β-glycerophosphate to ensure the phosphorylation state of the proteins was preserved. To examine if phosphorylation is important for Sen1 replisome association, *SEN1-TAP* cells were arrested in S phase by the addition of HU

and harvested for IP. The lysate was incubated with TAP beads, then washed with 50mM KoAC buffer. The beads were then either: mock treated, treated with lambda phosphatase (to dephosphorylate any proteins bound to the beads), or lambda phosphatase supplemented with phosphatase inhibitors. One tube per condition was boiled immediately, while another tube was once again washed with 50mM KoAC buffer before boiling (fig 5.15a). Our hypothesis was that, if the binding of Sen1 to the replisome were dependent on phosphorylation, binding to the replisome should have been lost following treatment with the phosphatase and extensive washes.

Probing for RNAPII using the Pol2RA phosphoantibody, which recognises both hyper- and hypo-phosphorylated RNAPII indicated successful dephosphorylation of the extract, as no signal was detected in the phosphatase treated IP. Importantly, the co-purification of Mrc1, Ctf4 and candidate replisome components Pol2 and Cdc45 with Sen1 during S phase was not affected by dephosphorylation (fig 5.15b). Thus, the observed cell cycle regulation does not appear to be dependent on phosphorylation by a kinase.

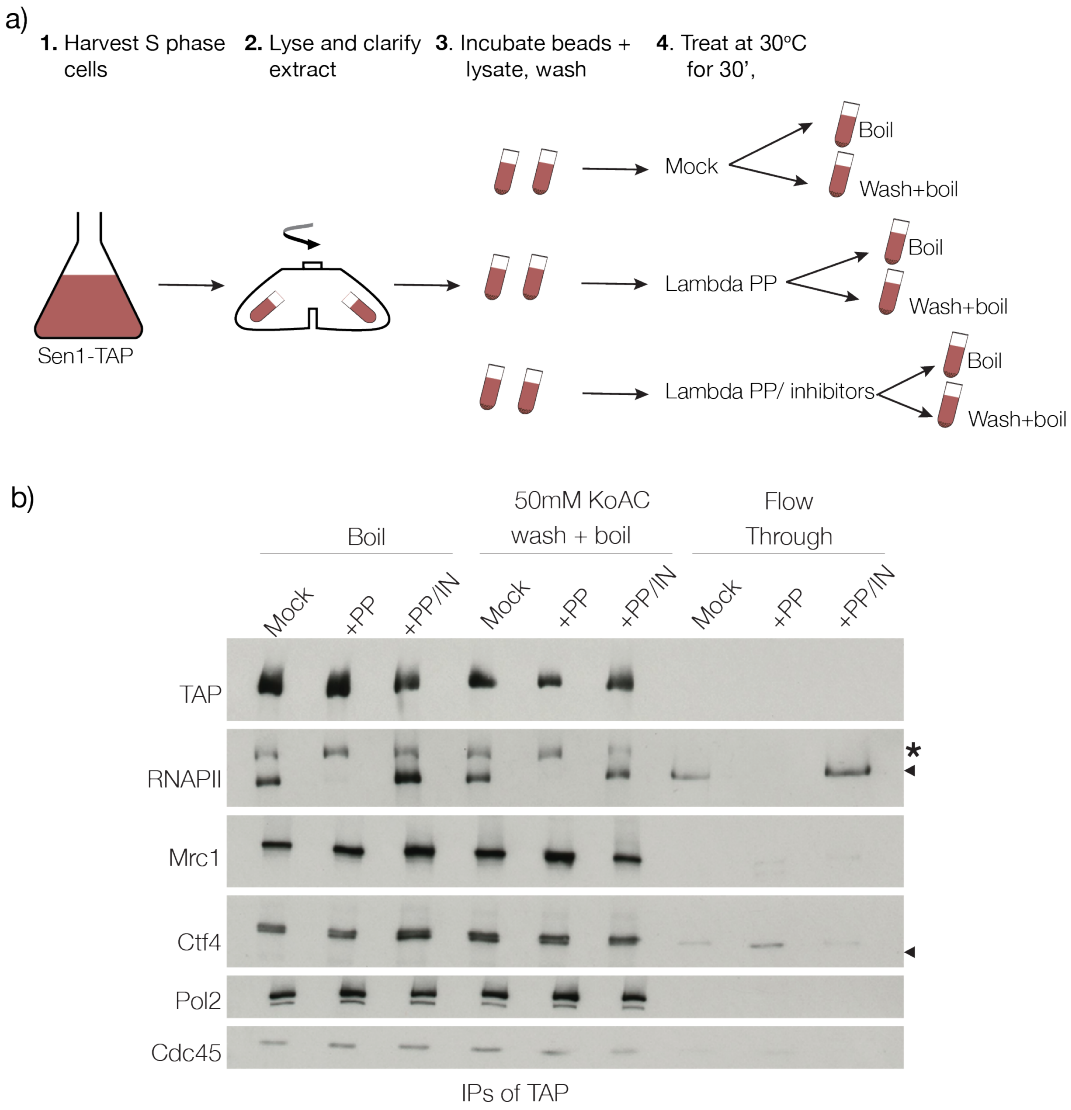


Figure 5.15 The S phase specific interaction between Sen1 and the replisome is not regulated by a kinase. a) Schematic of the dephosphorylation IP. 1L cultures of cells carrying the *SEN1-TAP* allele inserted at the *SEN1* chromosomal locus were grown in YPD at 24°C to a density of $0.7x10^7$ cells/ml. Cultures were then arrested by the addition of alpha factor for 3 hours. Following successful arrest, cells were washed with fresh YPD and released for 30 mins to harvest during S phase. IPs of these cells were performed, where the clarified extract was split into six aliquots, each added to 100µl of TAP beads. Following a 2-hour incubation at 4°C, the beads were washed with 50mM KoAC buffer, and the tubes were split into pairs. Pairs were either treated with 400U of lambda protein phosphatase, 400U of lambda protein phosphatase supplemented with phosphatase inhibitors (50mM NaF, 20mM $NaVO_4$), or mock treated for 30 mins at 30°C. After removing the supernatant from the beads, one tube per pair was boiled immediately in 50µl of 1x laemmli buffer, and the other washed three times with 50mM KoAC buffer before boiling. b) Immunoblot analysis of IPs described in (b). The * indicates non-specific recognition of the TAP. Strains used in this experiment: CS1941.

Chapter summary

Experiments described in this chapter show that the N-terminal domain of Sen1 is able to bind to Ctf4 outside of S phase, unlike the full-length protein. Analysing the binding of this Sen1 fragment to its replisome partners Ctf4 and Mrc1 during G1 found that in the absence of Ctf4, the affinity for Mrc1 is increased. This raises the possibility that these two proteins may compete for the binding of individual Sen1 molecules. Interestingly, during G1 and S phase (outside of the context of the replisome) a fragment of Sen1, lacking its N-terminal, was found to interact with DNA polymerase ε. IPs of Sen1 fragments during G1 also showed that full length Sen1 has a greater affinity for Mrc1 compared to the N-terminal fragment (which contains the Mrc1 replisome binding site). The interaction of Sen1 C-terminal with Pol ε may account for the observed cell cycle independent binding between full length Sen1 and Mrc1, as Mrc1 and Pol ε form a complex throughout the cell cycle. Thus, the increased Mrc1 signal during G1 in full length Sen1 IPs may be an indirect result of this association. Further experiments showed that the cell cycle dependent regulation of Sen1 binding to Ctf4 and replisomes does not depend on Sen1 itself, but instead appears to be regulated by the cell extract. Finally, the binding is not a result of a phosphorylation event during S phase.

Chapter 6: Discussion

6.1 *sen1-3* mutants are sensitive to the accumulation of R loops

6.1.1 Sen1 and the RNase H enzymes

In addition to its role in the termination of RNAPII transcription, the DNA:RNA helicase Sen1 has been demonstrated to remove R loops (Mischo et al., 2011, Costantino and Koshland, 2018). We have shown a mutant version of Sen1 that fails to be recruited at the replisome cannot sustain cell growth in absence of RNase H1 and H2 (fig 3.1). These enzymes are also both involved in the metabolism of R loops and display some functional overlap with one another in this respect (Lockhart et al., 2019). Thus, we hypothesised that the anchoring of Sen1 to the replisome may be essential to facilitate its fast recruitment or local enrichment to sites where its R loops resolving activity is required, namely when these structures are encountered by elongating replication forks (fig 6.1).

Interestingly, increasing the global levels of *sen1-3* by overexpressing it under control of the strong, constitutive *ACT1* promoter suppresses the synthetic lethality of *sen1-3 rnh1Δ rnh201Δ* triple mutant cells (fig 3.3). As Sen1 is usually maintained at relatively low levels (Ghaemmaghami et al., 2003), it appears that increasing the level and consequently the activity of *sen1-3* can lead to its presence at forks in the absence of the specific replisome anchoring mechanism. This also suggests that it could be recruited independently from its interaction with the replisome to any R loop collision sites, however this would likely occur with much slower kinetics (fig 6.1). Congruently, the overexpression of wild type Sen1 using the *ACT1-SEN1* allele suppresses the HU sensitivity of *rnh1Δ rnh201Δ* cells (fig 3.3). However, though the *ACT1-sen1-3* allele restores cell viability, in contrast, it is not able to suppress the HU-sensitivity of RNase H double deletion, which has been linked to the formation of R loops as well as defective ribonucleotide excision repair (Lockhart et al., 2019). The inability of overexpressed *sen1-3* to bind replisomes also sensitises *rnh1Δ rnh201Δ*

cells to MMS induced-DNA damage (fig 3.3). Recent slot blot analysis has shown that the levels of R loops are increased in both wild type and *rnh1Δ rnh201Δ* cells as a result of treatment with MMS (Lockhart et al., 2019). Additionally, at non-permissive temperatures, *ACT1-sen1-3 rnh1Δ rnh201Δ* cells accumulate in G2/M and show greater activation of the checkpoint kinase Rad53, suggesting an increase in the levels of DNA damage in this background (fig 3.4, 3.5).

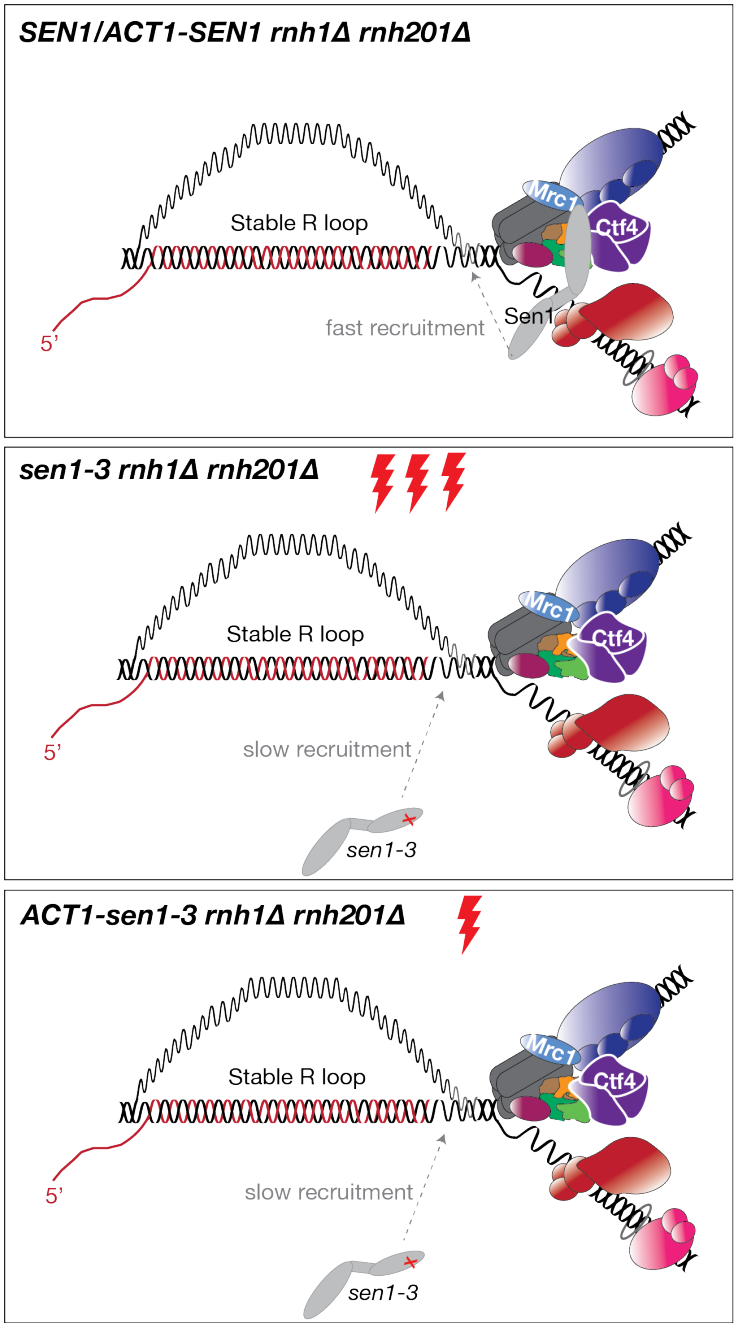


Figure 6.1 Schematic of the role of Sen1 at replication forks in the absence of the RNase H enzymes. Both the *SEN1* and *ACT1-SEN1* alleles are able to compensate for the absence of RNase H1 and H2 activity at replication forks. However, *rnh1Δ rnh201Δ* cells are inviable

when combined with the *sen1-3* allele. Overexpression of *sen1-3* is able to rescue the lethality suggesting that increasing its levels can compensate for its loss from forks, or that it can be recruited more slowly to any collision sites, independently from the replisome tethering mechanism. However, the cells are sensitive to MMS and HU, suggesting that anchoring Sen1 to replication forks is particularly important under conditions of DNA damage or replication stress. Persistent fork stalling at sites of unresolved R loops, while *sen1-3* or another redundant factor is recruited more slowly, can lead to DNA damage.

Overall, these data indicate that under normal circumstances, the activities of Sen1 and the RNase H enzymes are able to compensate for the lack of one another at forks (fig 6.1). However, although *ACT1-sen1-3* suppresses the triple mutant lethality, at higher temperatures or in response to HU or MMS exposure, the cells rely on the tethering of Sen1 to replisomes to maintain viability (fig 3.3). This may confer a kinetic advantage, where under conditions of DNA damage or replication stress, the defects observed in the triple mutant could result from slower recruitment and enrichment of Sen1 at forks due to loss of the replisome tethering mechanism; leading to deficient removal of R loops in the absence of redundant RNase H activity. Alternatively, it could be the case that the tethering of Sen1 to replisomes becomes crucial for it to deal with some defect at forks that is a direct result of R loops accumulating in the absence of RNase H.

6.1.2 *sen1-3* and other RNA metabolism mutants

We have also observed synthetic defects in another background with high endogenous levels of R loops. Combining the *sen1-3* allele with mRNA biogenesis mutant *hpr1Δ* revealed that the double mutant is sensitive to both temperature and replication stress (fig 3.7). In addition, *sen1-3 hpr1Δ* cells show an increase in Rad52-GFP foci formation during late S phase, indicative of increased levels of recombination linked to DNA damage or fork stalling (fig 3.8). Although in these experiments the levels of Rad52-GFP foci appear lower in the *hpr1Δ* mutants compared with wild type and *sen1-3* cells (which is surprising considering the increased levels of R loops), the FACS profiles show that the *hpr1Δ* cells progress slower and less synchronously through the cell cycle. Correspondingly, the foci accumulate slower in the *hpr1Δ* mutants but persist through to later time points. Indeed, transcription defects have been observed in *HPR1* null cells (Chávez and Aguilera, 1997), and these could affect the synchronous transition from G1 to S phase by interfering with cell cycle

progression. Conversely, in the synchronous wild type and *sen1-3* single mutant cells, the increase in foci observed during late S phase/G2 is quickly resolved (fig 3.8). Finally, the phenotype is likely weakened by the fact that this experiment was carried out at permissive temperature due a complete failure of *SEN1* and *sen1-3 hpr1Δ* cells to arrest and bud synchronously at non-permissive temperature. At 35ºC and 37ºC, dilution spotting experiments clearly show *sen1-3 hpr1Δ* cells are synthetic defective compared to the wild type control, while this is less clear at 28ºC (fig 3.7).

Analysing the level of Rad52-GFP foci is widely used as a molecular marker for HR-mediated repair of DSBs and could also signal HR-mediated restart of stalled replication forks (Lisby et al., 2001, Alabert et al., 2009). However, though no direct evidence exists, it is possible that undamaged replication forks could be transiently recognised by HR factors such as Rad52, even during normal S phase. Indeed, the anti-recombinase Srs2 acts to remove Rad51 nucleofilaments from ssDNA (the formation of which is mediated by Rad52), which under normal circumstances could lead to toxic recombination events (Krejci et al., 2003). Therefore, we also used a direct-repeat system to analyse the levels of recombination in the cells directly. In agreement with the increased Rad52-GFP foci, *sen1-3 hpr1Δ* cells show significantly greater levels of direct-repeat recombination compared with their *SEN1* counterpart, which increases with the length of the intervening transcript, suggesting some transcription dependence (fig 4.10). This reflects *sen1-1* cells, which also show a link between the intervening transcript length and recombination of direct-repeat substrates (Mischo et al., 2011). Importantly, overexpression of *RNH1* to digest any R loops was able to suppress the growth defects and level of Rad52-GFP foci formation in *sen1-3 hpr1Δ* cells back to the level of the single *hpr1Δ* mutant (fig 3.7, 3.8). This suggests that the absence of Sen1 tethering to replication forks is toxic in a *hpr1Δ* background as a result of R loops accumulation or stabilisation. This increased genomic instability is also demonstrated by the fact that *sen1-3 hpr1Δ* cells are totally defective in the maintenance of mini chromosomes (Appanah et al., 2020). Defects in the maintenance of plasmids can be linked to chromosome resolution defects, whereby increased recombination, defects in origin firing, incomplete DNA replication or defective chromosome segregation can result in plasmid loss. In this case, overexpression of *RNH1* moderately suppresses the defects of the double mutant *sen1-3 hpr1Δ* cells,

suggesting that an accumulation of R loops is partly responsible for chromosome loss in this background (Appanah et al., 2020).

In contrast to the Rad52-GFP experiments, analysing the levels of R loops directly revealed that while they are elevated in cells where both *SEN1* and *sen1-3* are combined with *hpr1Δ,* there was no significant difference in DNA:RNA hybrid signal between the two (fig 3.9). It is possible that the discrepancy may arise due to limitations in the sensitivity of these assays, where the S9.6 antibody could fail in the detection of labile or short R loops. These types of R loops may be responsible for the increased level of recombination observed in *sen1-3 hpr1Δ* cells, shown to be specifically reduced by RNase H digestion (fig 3.8a). In addition, once again, the budding defects of *hpr1Δ* cells required these experiments to be carried out at permissive temperature, thus likely weakening any phenotype.

Genome-wide mapping of R loop-prone loci has observed that not all R loops formed are automatically toxic, congruent with the idea some play important physiological roles. Only a subset of R loops actually result in DNA damage and the accumulation of Rad52 (Costantino and Koshland, 2018). It is possible that though the levels of R loops are similarly increased in *SEN1* and *sen1-3 hpr1Δ* cells, the levels of Rad52-foci are elevated in the *sen1-3* background because certain R loops are more efficiently converted to damage when Sen1 is absent from forks, perhaps as they are cleared with slightly different kinetics than if it was already present.

In different genetic backgrounds, R loops may accumulate on different areas of the genome or have distinct biochemical features, some of which may require the presence of Sen1 at forks, while others could be cleared just as efficiently by redundant factors. In agreement with this idea, one study has shown that in a *rnh1Δ rnh201Δ* background, cells depleted of Sen1 or Top1 accumulate hybrid induced damage on different areas of the genome (Costantino and Koshland, 2018). I have observed that unlike *hpr1Δ* or *rnh1Δ rnh201Δ* cells, other R loops accumulating strains did not show any synergistic synthetic growth defects in combination with *sen1-3.* These included deletion of; *TOP1, RRM3, PIF1, SGS1* and *SRS2,* which resolve R loops at various different loci (fig 3.10). As *sen1-3* does not exacerbate any defects in these backgrounds; it is likely

that in these cases, another redundant factor can resolve R loops at those particular sites, or *sen1-3* could be recruited to these sites via another mechanism with kinetics sufficient for their timely resolution. Finally, the presence of Sen1 at forks may only be critical if the levels of R loops within the cells reaches a certain threshold, such as in *rnh1Δ rnh201Δ* cells, after which other pathways do not have sufficient kinetics to compensate for its absence.

Future experiments should be carried out in order to clarify further whether loss of Sen1 from forks results in the enrichment of deleterious R loops, and at what areas of the genome this may occur. One possible experiment could be to carry out DNA:RNA immunoprecipitation using the S9.6 antibody followed by sequencing (DRIP-seq) in *SEN1* and *sen1-3* cells, to identify if R loops are enriched as a result of *sen1-3*. In addition, it would be particularly interesting to look at telomeres, in light of the tetrad analysis carried out to look at the functional redundancy of Sen1 at forks. The suppression of defects in *rnh1Δ rnh201Δ mph1Δ* and *hpr1Δ* cells by *ACT1-SEN1* suggests that its presence at forks may be particularly important at telomeres (fig 3.11).

Another potential experiment could adopt the strategy developed by the Koshland lab and look genome-wide for differences in accumulation of lethal R loop associated damage (Costantino and Koshland, 2018). The ChIP-profile of DNA repair protein Rad52 in a *hpr1Δ* background with either wild type *SEN1* or *sen1-3* could be compared at the previously identified R loop-prone loci (Wahba et al., 2016, Costantino and Koshland, 2018). Looking for unique Rad52 accumulation at sites surrounding DNA:RNA hybrids may provide further information regarding whether the absence of Sen1 from forks leads to increased conversion of R loops to DNA damage, and at which regions of the genome this occurs. This analysis would also be interesting in a background without the RNase H enzymes, however as they are lethal in combination with *sen1-3,* an approach to efficiently degrade Rnh1 and Rnh2 would be required.

Work has suggested that head-on collisions between the transcription and replication machinery are particularly deleterious in the context of R loops. In yeast, a study has suggested that R loops form independently from any collisions, but they can lead to stabilisation of RNAPII, blocking oncoming forks in the HO orientation (Garcia-

Rubio et al., 2018). However, in bacteria and human cells, studies have suggested that R loops may actually form as a result of these HO collisions (Hamperl et al., 2017, Lang et al., 2017). Either way, the R loops at these sites present a threat to genome stability and it would be interesting to understand whether *sen1-3* mutants are defective for R loops removal at TRC sites using simultaneous DRIP-seq and ChIP analysis of replisome components. Finally, it would be interesting to examine whether there is any correlation between the location of any *sen1-3* dependent R loop-induced damage (mapped using the Rad52-seq method) being closer to an origin that would encounter the R loop head-on rather than co-directionally, as has been described in Sen1 depleted cells (Costantino and Koshland, 2018).

7.2 Sen1 and the S phase checkpoint

MRC1, CTF18 and *RAD53* were among the factors identified while screening for genes that genetically interact with the *sen1-3* allele (fig 4.1, 4.3). Rad53 is the main effector kinase of both the DDC and DRC, while Mrc1 and Ctf18 are key mediators of the DRC. It may be the case that when replisomes collide with either R loops or transcription bubbles, the absence of Sen1 from forks leads to more prolonged fork stalling, as it is not immediately present at collision sites to remove the offending obstacle. Instead, *sen1-3* or another redundant factor needs to be recruited, likely with slower kinetics. By removing the checkpoint mediators/effectors, this could lead to increased genome instability, whereby the checkpoint response is required in the *sen1-3* background to preserve the integrity of stalled forks, or deal with the consequences of the defects accumulating in *sen1-3* cells. In addition, *sen1-3* is also mildly sensitive to replication stress in combination with *tof1Δ* and *rad50Δ* (fig 4.3), which have been implicated in activation of the DRC and DDC, respectively, as well as playing independent roles in promoting fork stabilisation and re-start (Foss, 2001, Delamarre et al., 2020, Tittel-Elmer et al., 2012).

Surprisingly, however, though Mrc1, Ctf18 and Rad53 appear to be important for genome stability in *sen1-3* cells, we did not detect similar defects in combination with deletion mutants of genes involved in canonical activation of the S phase checkpoint response (fig 4.5). This included Mec1 and Tel1. How to explain this? One possibility is that, in the absence of the sensor kinases Mec1 and Tel1, the defects caused by the

sen1-3 allele might be resolved in an alternative manner by cells at the end of DNA replication. In the presence of Mec1, however, *sen1-3* might trigger its activation that requires an intact checkpoint response to promote viability; in the absence of the downstream effectors and mediator, the activation of Mec1 might become toxic for the cell. This hypothesis might explain why a hypomorphic allele of *MEC1* (*mec1-100*) might still show synthetic defects with *sen1-3* (fig 4.3), while the full deletion appears to be epistatic (although further analysis using the appropriate amount of HU and MMS might be required to fully assess the epistasis). Thus, Mec1 sensing any defects and being unable to elicit a full checkpoint response could be more deleterious for the cells than it not sensing them at all. Alternatively, the synthetic defects observed between *sen1-3* and *mrc1*Δ*, ctf18*Δ and *rad53*Δ might not be linked to their function in the checkpoint response. In fact, $mrc1^{AQ}$, an allele that cannot be phosphorylated by Mec1 so is deficient in its checkpoint function, did not exacerbate the defects (fig 4.5). Moreover, all of these proteins also have important functions outside of the checkpoint, which could be important in the *sen1-3* background, particularly as hyperphosphorylation of Rad53 is not observed in *sen1-3* cells (fig 3.5).

Mrc1 has been assigned a DNA replication role whereby its presence in the replisome is important for forks to achieve the top speed for DNA synthesis (Hodgson et al., 2007, Yeeles et al., 2017). It has been suggested that Mrc1 couples the CMG helicase and leading strand polymerase in close proximity for optimal replication speeds via interactions with the catalytic domain of Pol ε and the region of MCM where the leading strand exits the CMG after being unwound (Baretic et al., 2020). Though the biochemical function of Mrc1 during DNA synthesis needs further clarification, it could be that its role in promoting efficient DNA synthesis is important in the *sen1-3* background. It has also been suggested that Mrc1, which bridges Pol ε at the rear of the replisome to Tof1/Csm3 at the front, may somehow help the cells to coordinate leading strand synthesis with any oncoming blocks sensed by Tof1/Csm3 (Baretic et al., 2020). Finally, Mrc1 has also been directly implicated in helping prevent instability resulting from transcription-replication conflicts, outside of checkpoint. This has been termed the "Mrc1 transcription-replication safeguard mechanism", where it was shown during cellular stress (which induces a massive transcriptional response in the cell), phosphorylation of Mrc1 leads to the slowing down of replication

fork progression to reduce the frequency of TRCs and transcription associated recombination (Duch et al., 2018).

Ctf18 localises to forks via its interaction with Pol ε as part of RFC^{Ctf18}. Outside of the checkpoint, it has been shown to be important for the establishment of sister chromatid cohesion and nuclear organisation by positioning telomeres in proximity to the nuclear periphery (García-Rodríguez et al., 2015, Stokes et al., 2020, Hiraga et al., 2006). Ctf18 also has a separate role during DNA replication, perhaps helping to promote fork restart following stalling, a function which interestingly may be shared by Mrc1 (Gellon et al., 2011, Stokes et al., 2020). In a *sen1-3* background, it is possible that cells rely on damage-free restart of forks that stall as result of the defects. Finally, Rad53 limits the level of histones within the cell independently from its checkpoint activity, which prevents any deleterious effects as a result of their excess accumulation (Gunjan and Verreault, 2003). However, whether the defects in *sen1-3* are associated with the checkpoint or other functions of these proteins is currently not clear.

7.3 Obstacles other than R loops are toxic in *sen1-3* mutants

Both the *sen1-3 ctf18Δ* and *mrc1Δ* double mutants showed growth defects, altered DNA replication dynamics and increases in recombination during late S/G2, as indicated by Rad52-GFP foci formation (fig 4.3, 4.6, 4.7). Taken together, this suggests that the defects arise during DNA synthesis. As we observed in *sen1-3 rnh1Δ rnh201Δ* and *sen1-3 hpr1Δ* cells that the synthetic defects could be somewhat reversed by overexpression of *RNH1* (fig 3.7, 3.8), we wanted to understand whether the defects in the S phase checkpoint mutants combined with *sen1-3* were linked to R loops. We used *sen1-3 mrc1Δ* cells as a candidate strain, and interestingly overexpression of *RNH1* was unable to suppress the growth defects and increased foci in these cells (fig 4.8, 4.9). Additionally, higher levels of plasmid loss in *sen1-3* or *sen1-3 mrc1Δ* cells are not rescued by *RNH1* overexpression (Appanah et al., 2020). This suggests that Sen1 has another role at replication forks that promotes DNA replication and genome stability independently from any role it may also play in the removal of R loops.

As Sen1 has an important role in the termination of RNAPII transcription at NNS target genes, as well as some protein coding genes (Porrua and Libri, 2013, Han et al.,

2017), an attractive possibility is that Sen1 may terminate actively transcribing or stalled RNAPII at sites of replication fork collisions to allow forks to resume replication. Some mechanisms exist within the cell to try and reduce collisions; for example, replication fork barriers in highly transcribed rDNA (which are conserved from yeast to humans) prevent replication from occurring in a direction where it would collide head on with transcription. However, on other areas of the genome collisions are inevitable and, in eukaryotes, it is currently not well understood how replication forks negotiate their way past any elongating transcription complexes. It has been shown *in vitro* that ahead of bacterial replication forks, the Mfd protein allows stably stalled replication forks to continue replication by displacing RNAPII from the DNA. Interestingly, similar to Sen1 in yeast, Mfd is part of the bacterial transcription-coupled nucleotide excision repair pathway and possesses low processivity DNA translocase activity (Pomerantz and O'Donnell, 2010, Le et al., 2018). It is possible that in cells where Sen1 is not present at the replisome to quickly remove transcribing RNAPII, the replication fork stalls more persistently, and thus it relies on the fork protection complex and checkpoint/repair genes for their stabilisation and eventual restart once the polymerase is removed (fig 6.2).

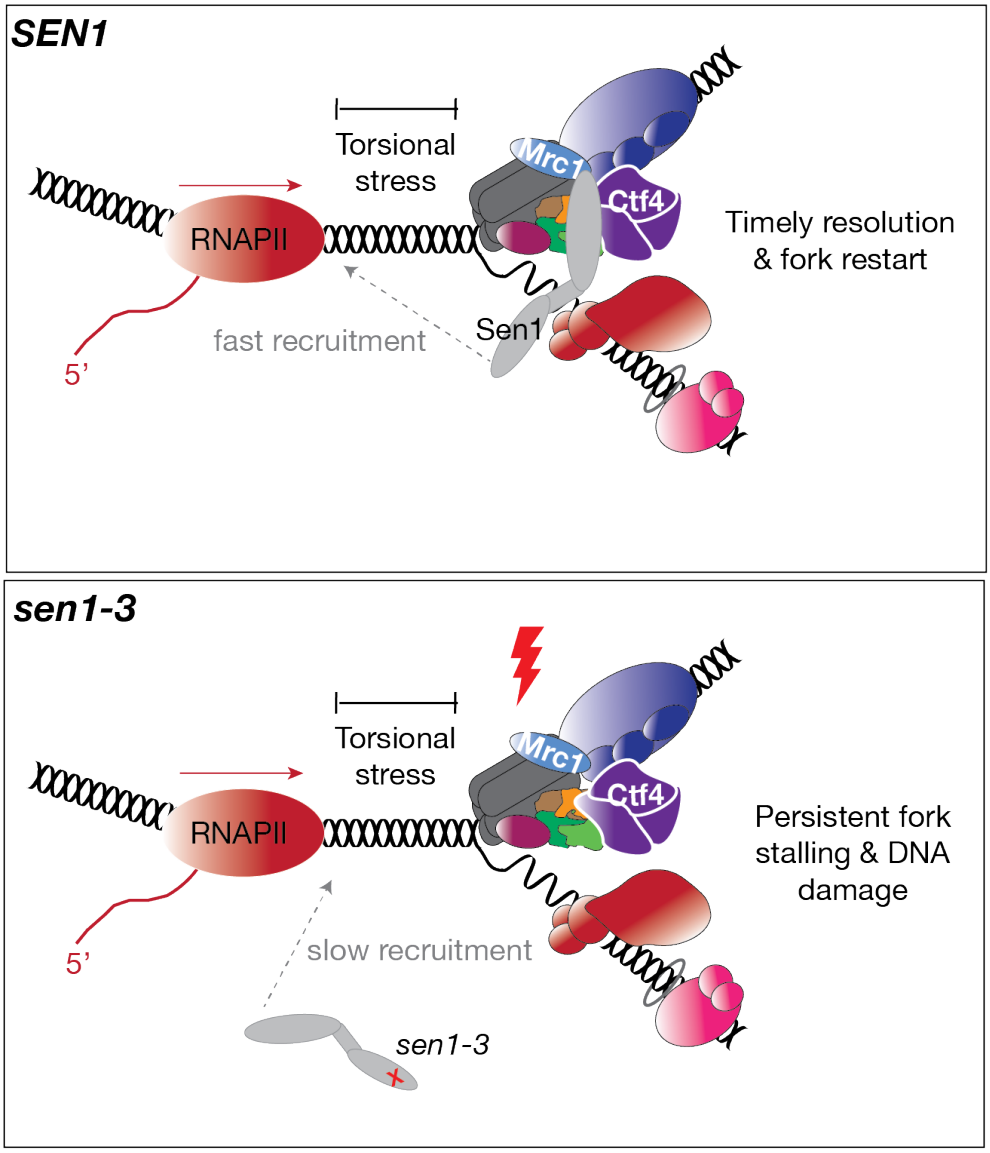


Figure 6.2 Speculative model for the action of Sen1 at sites of head-on transcription-replication conflicts. Sen1 may travel with replication forks for its fast recruitment to sites of transcription-replication conflicts. This reduces replication fork stalling and deleterious recombination events. When Sen1 is not readily available at forks via its replisome tethering mechanism, prolonged fork stalling or collapse may occur as a result of the requirement for alternative slower mechanisms to remove the obstacle. These cells become reliant on S phase checkpoint components and DNA repair proteins to maintain viability and promote replication fork restart.

Interestingly, preliminary metagene-analysis data from our collaborators has shown that there are differences in the occupancy profiles of RNAPII across both CUTs and protein coding genes in exponentially growing *SEN1* and *sen1-3* cells. In particular, the occupancy of RNAPII is significantly increased at the 5' end of genes near the transcription start site in exponentially growing *sen1-3* cells (fig 6.3). At the 5' end of genes, replication forks are likely to encounter promoter proximal paused RNAPII

molecules. Importantly, the difference between *SEN1* and *sen1-3* cells is not observed in G1 arrested cultures (Libri lab, unpublished).

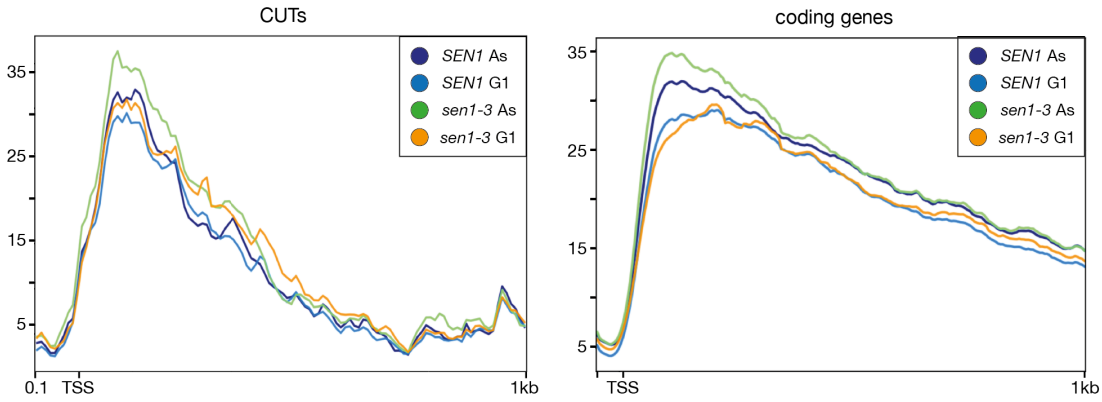


Figure 6.3 Meta-gene analysis of RNAPII occupancy across CUTs and protein coding genes in G1 vs exponentially growing *SEN1* and *sen1-3* cells, (Libri lab, unpublished).

Further experiments are required to investigate whether Sen1 helps replication forks negotiate their way past elongating or stalled transcription complexes. Ideally, these experiments would not use any RNAP mutants, as any effects on the transcriptional program of the cell could lead to defects independently from TRCs. However, obtaining a robust and quantitative analysis of any changes in RNAPII occupancy is complicated, as DNA replication dynamics and the orientation and activity of genes needs to be integrated into the analyses. In addition, as the differences likely occur only during S phase, sharp differences in the RNAPII peaks may be hard to detect when averaging millions of cells, even if synchronised. Finally, as the delay caused by *sen1-3* may only be short-lived, either due to alternative slower recruitment of *sen1-3* or recruitment of another redundant factor, the resolution of this method may not be sufficient to observe any differences during S phase.

A single molecule study, although technically challenging, could address this question by using purified yeast proteins to assess the effect of *sen1-3* on TRCs *in vitro*. A reconstituted replication fork and transcription machinery could be oriented either head-on or co-directionally to one another on a template DNA duplex. In the presence of ATP, the rate of transcription and replication could be monitored in real time by measuring the fluorescence of incorporated dNTPs and rNTPs. Analysing the length of the newly synthesised DNA and RNA at the timescale of seconds would allow us

to understand further the effect of losing the Sen1 replisome tethering mechanism on fork progression and RNA transcription when TRCs occur.

If Sen1 helps forks to negotiate past RNAPII transcription, we would also like to understand the mechanism of how this may occur. Key questions would be whether Sen1 uses its reported torpedo activity (Han et al., 2017) to push the RNAP off of the DNA in the case of head-on collisions, and whether it could push stalled polymerases forward when collisions occur in the codirectional orientation, promoting the RNAP to continue transcribing ahead of the fork (fig 6.4). The aforementioned *in vitro* system could be used to track fluorescent-tagged RNA polymerase in the presence of Sen1 and Sen1-3 to determine whether it remains bound to the template DNA.

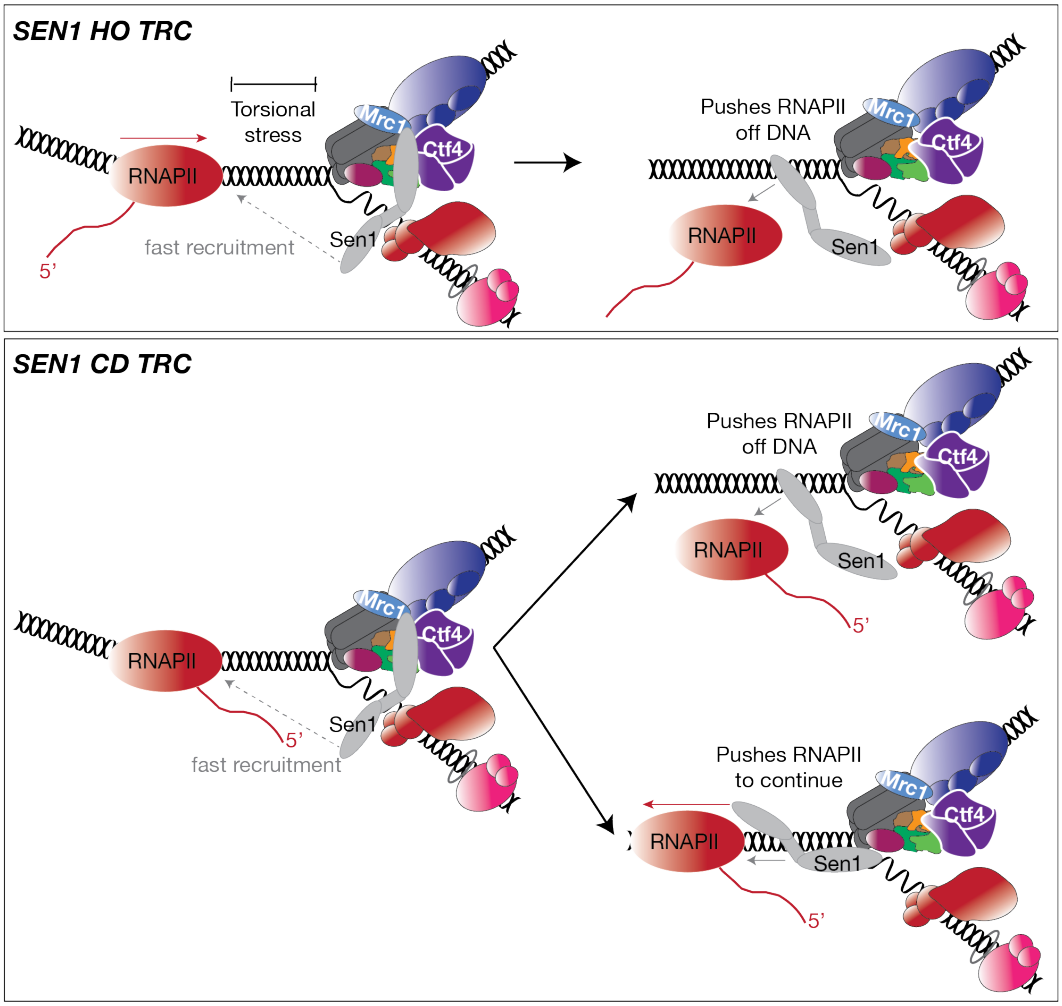


Figure 6.4 Schematic of the ways in which Sen1 could help replication forks to negotiate their way past the transcription machinery. At site of Head-on collisions, we speculate that Sen1 may act to torpedo oncoming RNAPII off of the DNA. As Sen1 is able to push stalled RNAPII, it could act at sites of co-directional collisions in one of two ways. It could either terminate any elongation complexes encountered by the replisome or could 'push' stalled polymerases forward to complete transcription of the gene ahead of the replication fork.

7.4 Analysis of Sen1 N-terminal binding with Ctf4 and Mrc1

Sen1 binds to replisomes mainly via Ctf4 and Mrc1, though other binding partners may also exist. Three single amino acid substitutions within the N-terminal domain of Sen1 (*W773A E774A W777A*) eliminates the ability for it to interact with these proteins and travel with replication forks during S phase. There are two possible explanations as to how the mutation of three closely spaced residues within Sen1 can disrupt its binding with multiple factors. First, the mutated region could be the site at which both Ctf4 and Mrc1 bind directly. Alternatively, Ctf4 and Mrc1 may actually have separate binding sites within Sen1, but mutation of the aforementioned residues alters the conformation of a larger portion of the protein in such a manner that both are disrupted.

There are also a number of possibilities that could explain why Sen1 has multiple different partners within the replisome. First, these may act co-operatively to tether a single molecule of Sen1; where binding with more than one replisome component serves to strengthen the affinity of the interaction. It is also conceivable that each partner binds Sen1 with different affinities, and that they could compete with one another. Alternatively, each partner may be able to bind separate molecules of Sen1 at the same time, thus tethering multiple copies to the replisome (fig 6.5). In some scenarios, for example during collisions with either R loops or transcription bubbles; this could help Sen1 to deal with the different possible orientations, making it more efficient in carrying out its function at these sites. Otherwise, recruitment by different subunits of the replisome could be redundant; where there are several possible options simply to ensure the presence of Sen1 at the fork, particularly if one site is occluded for some reason.

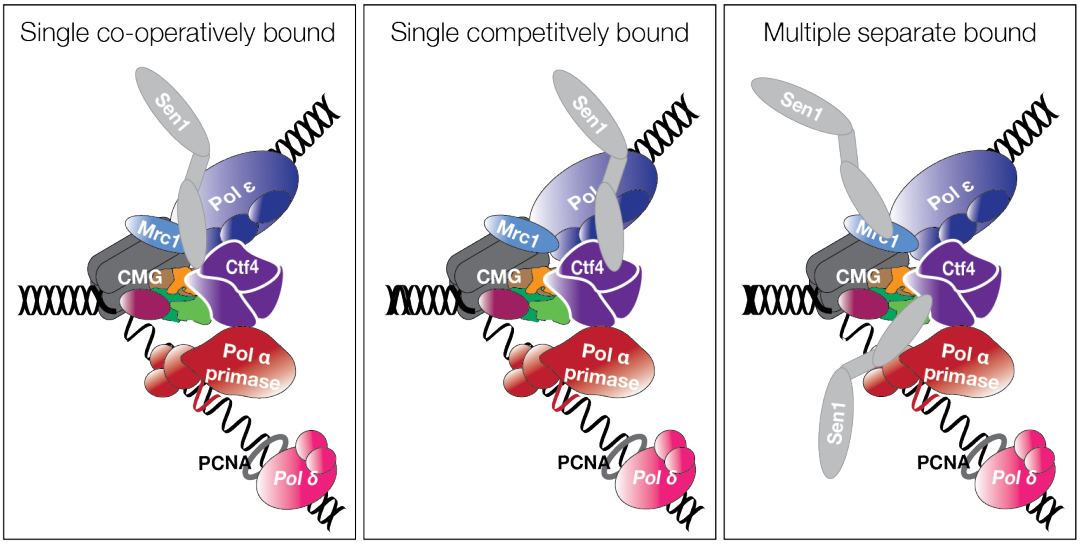


Figure 6.5 Schematic of the potential ways in which Sen1 may bind replisomes. Sen1 has multiple replisome binding partners, where it associates with Ctf4 and Mrc1, as well as other potential factors which are currently unidentified. It is possible that these factors work together to cooperatively bind a single molecule of Sen1 to the replisome with high affinity, or they may compete with one another for Sen1 binding. Another possibility is that each partner associates with separate Sen1 molecules, thus tethering several to the replisome at once.

In this thesis, I have observed that the affinity of Sen1 N-terminal for Mrc1 increases in the absence of Ctf4 during G1 (fig 5.2), which has also been confirmed in S phase cell extracts (fig 1.16) (Appanah et al., 2020). This data goes some way toward supporting the hypothesis they bind Sen1 independently, as it appears Mrc1 may compete with Ctf4. Conversely however, the binding of Ctf4 is not affected by the absence of Mrc1 (fig 5.2). Moreover, IPs, with or without crosslinking, have shown that the replisome association of Sen1 is reduced in the absence of either of the proteins (Appanah et al., 2020). This data is more in line with the idea that there is no competition for Sen1 binding and that Ctf4 and Mrc1 may act co-operatively.

In order to answer the important question of whether the proteins are competing with one another, we aimed to reconstitute the interaction between Sen1, Mrc1 and Ctf4 outside of yeast extract, thus in the absence of any other eukaryotic proteins. We used *E. coli* as an expression vector in an attempt to purify the minimal interacting fragments of the three proteins. However, we were only able to successfully express fragments of Ctf4 in *E. coli.* Unfortunately, the expression of various fragments of Mrc1, as well as Sen1 (Daniel Garbazski, personal communication), was unsuccessful; despite attempts to optimise the purification protocol to solubilise the protein (data not

shown). Future experiments should attempt to express fragments of Mrc1 and Sen1 using another host, e.g. baculovirus, to obtain the recombinant protein. This would then allow us to answer several key questions. First, we could confirm whether the proteins are able to interact directly with one another by assessing whether they co-purify in the absence of any other eukaryotic proteins. This would eliminate the possibility that a currently unidentified factor is also involved in the interaction as part of a complex. Second, we could fractionate the mixed protein lysates using a gel filtration column and analyse the fractions by western blot. This would allow us to determine whether a single complex of the three proteins is formed, or whether two subcomplexes of Sen1-Mrc1 and Sen1-Ctf4 are formed. Finally, we could also titrate the concentrations of either Ctf4 or Mrc1 and look at the effects via IPs and immunoblotting, to understand whether they may be competing with one another for Sen1 binding. If so, it is likely that increasing the concentration of one protein would affect the binding and thus co-precipitation of the other.

Critically, structural biology techniques will also be required to ultimately understand how Sen1 interacts with the replisome. One possibility could be to try and elucidate the cryo-EM structure of a fork associated replisome with Sen1. As Sen1 may transiently bind and dissociate from the replisome, this should be carried out in the presence of a cross-linking agent to obtain a 3D reconstitution of the interaction. This would help elucidate whether single or multiple copies of Sen1 are tethered to the replisome and get a picture of the replisome orientation of its partners binding sites.

7.5 Sen1 retains some affinity for Mrc1 and Pol ε during G1

The observation that full length Sen1 is able to bind Mrc1 during G1, while it does not interact with Ctf4 outside of S phase is of great interest. We wondered whether Sen1 could possess multiple binding sites for Mrc1 throughout the protein, as during G1, the affinity for Mrc1 increases from the N-terminal (2-931), (2-1901) fragments and full length Sen1, respectively (fig 5.1, 5.11). If this were true, this raises the possibility that binding at one site during G1, perhaps in the C-terminal of Sen1, could help with the recruitment of Sen1 to forks, where molecular hand off to its separate N-terminal site in the context of the replisome could be responsible for its tethering during S phase. Indeed, Mrc1 appears to have two separate Sen1 binding sites, where analysing

the minimal interacting fragments of Mrc1 with Sen1 (2-1901) found that non-overlapping regions of its N- and C- terminal are able to support Sen1 binding (fig 5.9, 5.10). This is reminiscent of the binding of Mrc1 to Pol2, where independent interaction domains in both the N- and C-terminal of each protein dynamically interact depending on their phosphorylation state, possibly to regulate the binding of Mrc1 to other factors (Lou et al., 2008). A mechanism such as this could be responsible for the apparent cell cycle regulation of Sen1 binding to Ctf4 and the replisome. However, analysing IPs of Sen1 (2-931) vs. (931-2231) for Mrc1 binding were inconclusive, as a band that appears to be non-specific recognition of the TAP is present at the same molecular weight as Mrc1 (fig 5.12b). For this experiment, we attempted to maximise the separation of the two bands by using the TEV protease to cleave the TAP (leaving the tag bound to the beads) and immunoblotting the TEV eluates for Mrc1 and calmodulin binding protein (the remaining part of the tag following cleavage). However, the signal was too weak to be detected due to both the efficiency of cleavage and the volume of buffer required to for the cleavage reaction. These experiments should be repeated using a different tag in order to confirm whether or not Sen1 (931-2231) is unable to bind to Mrc1.

Interestingly, these IPs showed that during both G1 and S phase, there is an interaction between Sen1 (931-2231) and DNA polymerase epsilon, which is both independent from replisomes and from Mrc1 (fig 5.12). Mrc1 and Pol ε are known to associate with one another throughout the cell cycle (Lou et al., 2008), however the interaction of Sen1 (931-2231) with Pol ε is maintained in *mrc1Δ* cells. Future experiments should determine whether the increased interaction of full length Sen1 with Mrc1 during G1 is a direct result of their binding, or whether the signal is increased due to an indirect result of the association between Sen1 (931-2231) and Pol ε. Further work also needs to be carried out to determine the functional significance of the interaction between Sen1 (931-2231) and Pol ε. Speculatively, it is possible that an interaction between Sen1 (931-2231) and Pol ε may help to recruit Sen1 to the proximity of replisomes where it then binds to Mrc1 and Ctf4 via its N-terminal domain.

6.5 Mapping the interaction from the side of Ctf4 and Mrc1

Experiments regarding the minimal interacting domains of Ctf4 and Mrc1 for Sen1 are currently inconclusive. Successfully mapping the interaction sites would go some way towards helping us to understand the mechanism of interaction. Analysing the binding pattern of various Ctf4 fragments by IP indicated that the possible region mediating the interaction with Sen1 N-terminal is encoded by residues 351-426 (fig 5.5). Surprisingly, however, mutation of conserved residues within this region did not break the interaction. It is possible that the mutated amino acids are not important for binding, thus did not disrupt it. Alternatively, several of the mutated regions may be important for binding; where disrupting a combination of these sites is required to break the interaction. It is also possible that another region of the protein might contribute to the binding. In the absence of endogenous Ctf4, other fragments spanning the beta-barrel showed a comparably very weak signal for Sen1 binding at long exposures (fig 5.5). Perhaps residues within that area are important for binding, but residues (351-426) are important for the stability of the interaction. As residues (351-426) were deduced as key by virtue of it being a unique overlapping region of larger strongest binding fragments, creating the smaller (351-426) fragment and assessing it by IP would help to understand whether this is indeed the region responsible or not. Further mutants could then be designed accordingly, or different combinations of the mutants could be tested.

Non-overlapping N- and C-terminal fragments of Mrc1 were identified as able to bind Sen1 (2-1901). Mutations of any conserved residues within these regions were not designed or tested yet, as final experiments were being carried out with the aim of understanding whether a region within the C-terminal/more than one binding site for Mrc1 exists within Sen1, which would further complicate matters. Once it has been confirmed whether Sen1 has Mrc1 binding sites outside of its N-terminal domain (fig 5.12), future work should look for conserved regions within the minimal domains and target them for mutagenesis.

6.5.1 The importance of breaking the association between Sen1 and replication forks from the side of the replisome

As the *sen1-3* mutant retains wild type affinity for RNAPII and does not appear to be deficient in its transcription termination function (section 1.9.5,6) (Appanah et al., 2020), we hypothesise that the phenotypes observed in *sen1-3* cells are a result of losing the mechanism whereby it is specifically tethered to replisomes. However, we cannot currently exclude that the *sen1-3* mutation is affecting other unknown factors. Further work is needed to provide more evidence that the phenotypes observed in *sen1-3* cells are a direct result of its loss from forks.

Mapping the minimal interacting domains of Sen1, Mrc1 and Ctf4; allowing us to design mutations that break the interaction with Sen1 from the side of the replisome, would allow us to repeat experiments to examine whether these mutants phenocopy *sen1-3.* Another approach would also be to reattach *sen1-3* to the replisome and analyse whether this suppresses the observed defects. We have already tried using tags comprising either the CIP (Ctf4-interacting peptide; responsible for tethering pol alpha/the CMG to the replisome), or the TPR domain (which fuses SCF to the replisome); however, these attempts were unsuccessful. Interestingly, the Dia2-TPR domain interacts with a region within the C-terminal of Ctf4 (348-927), and the interaction of Sld5/Pol1-type CIP box requires the alpha helical bundle (Morohashi et al., 2009, Villa et al., 2016). The data generated in this thesis investigating the minimal interacting fragments of Ctf4 and Sen1 suggests the alpha helical region of Ctf4 is not responsible for Sen1 binding. It is possible that these tags may not be able recapitulate Sen1 replisome tethering in such a conformation that; for example, supports its engagement with the required substrates.

If future work shows that 1) *sen1-3* successfully re-attached to the replisome can suppress the observed phenotypes of *sen1-3,* or that 2) when the interaction is successfully abrogated from the side of the replisomes these cells phenocopy *sen1-3,* this would prove that the observed phenotypes are indeed specific to the loss of Sen1 activity at replication forks.

6.6 The binding of Sen1 to the replisome depends on the cell cycle phase of the extract, not Sen1

As a fragment comprising the first 931 residues of Sen1 is able to bind Ctf4 during G1 but the full-length protein cannot interact with it outside of S phase, this suggests that something within the last 1300 residues may be responsible for mediating the interaction with the replisome. There are a number of possibilities regarding how this could occur (fig 6.6).

From the side of Sen1, firstly, the interaction could be regulated by a cell cycle specific post-translational modification of Sen1. Interestingly, it is a highly phosphorylated protein, with a number of identified phosphorylation sites, many of which are in the C-terminal within the last 1300 residues (Bodenmiller et al., 2010, Swaney et al., 2013). Thus, phosphorylation of Sen1, perhaps by an S phase kinase at a site within the last 1300 amino acids, is an attractive possibility. Alternatively, a currently unidentified regulatory factor may inhibit or promote Sen1 binding to the replisome by either binding or dissociating from Sen1 in a cell cycle dependent manner (fig 6.6). A recent study has also shown that Sen1 is capable of intramolecular interactions *in vitro*, where the N-terminal expressed in yeast co-purifies with recombinant Sen1 C-terminal (Han et al., 2020). It is possible that Sen1 may exist in different conformations in G1 and S phase, whereby the replisome binding site(s) for Ctf4 and Mrc1 in the N-terminal are blocked by the C-terminal of the protein during G1 (fig 6.6).

IPs were carried out to test whether the observed cell cycle dependence of the binding depends on some intrinsic feature/ conformation of Sen1. Sen1 was isolated from G1 and S phase extracts to capture the protein in any phase specific conformation/ post-translationally modified state (our IPs are carried out in the presence of phosphatase inhibitors). Following a high salt wash, G1 and S phase extracts of tagged Mrc1 and Ctf4 were added to the Sen1 bound beads (fig 5.13b). As observed previously, Mrc1 binding was detected independently of the cell cycle phase. However, Ctf4 was able to bind both G1 and S phase forms of Sen1, as long as the Ctf4/Mrc1 cell extract was isolated during S phase (fig 5.13c). This suggests that in fact, it is not Sen1 itself that precludes Ctf4 binding in G1, but an event that occurs within cell extracts during S

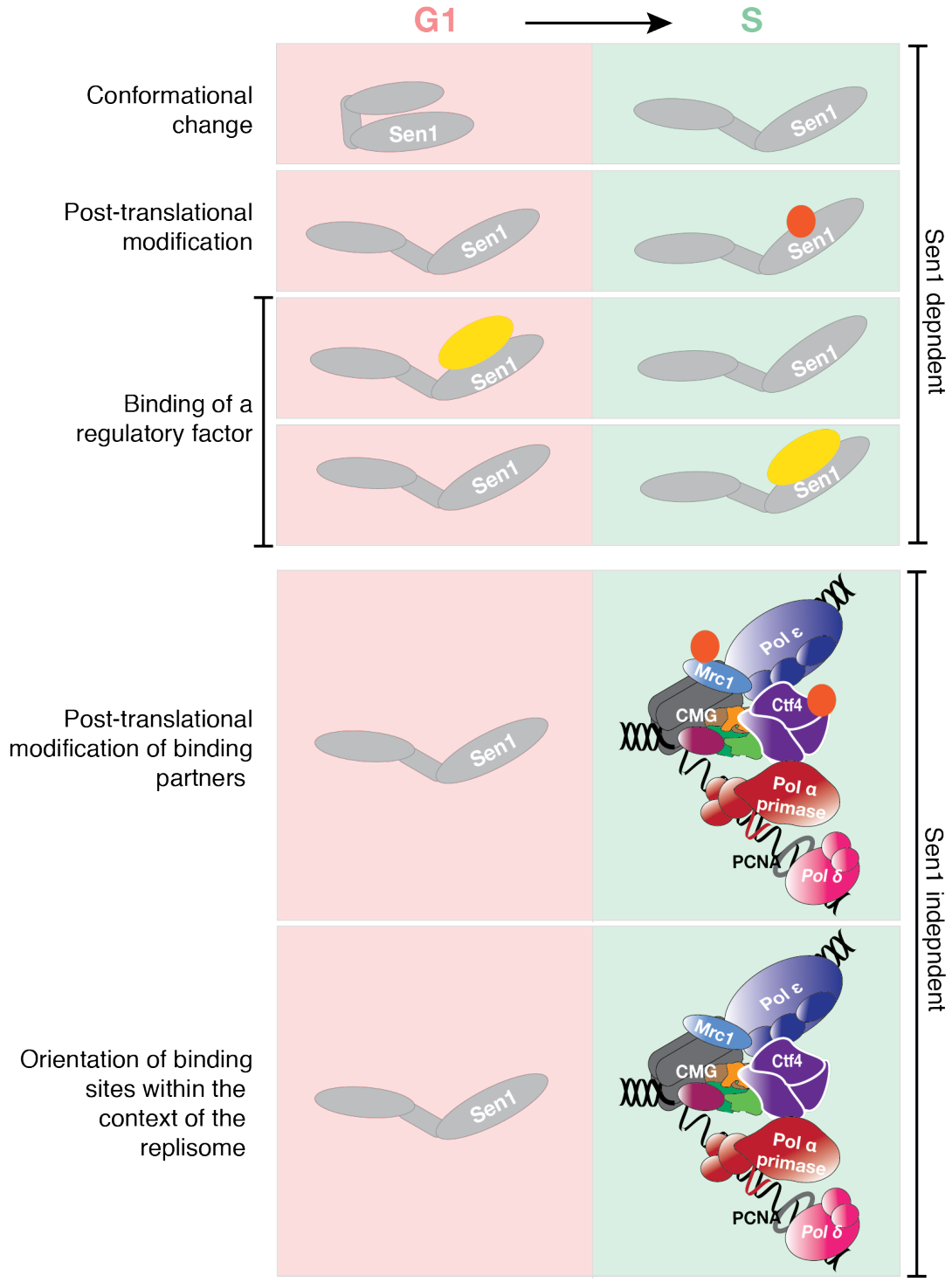


Figure 6.6 Schematic of possible Sen1 binding regulation. There are numerous ways in which the apparent cell cycle dependent regulation of Sen1 binding to Ctf4 and Mrc1 could occur. From the side of Sen1, it is possible that during G1, intramolecular interactions between the N- and C-terminal of the protein block the N-terminal replisome binding sites. An S phase dependent conformational change may expose these, thus allowing binding. Alternatively, a post translational modification of Sen1 may also responsible for regulating the binding. Finally, the presence or absence of a regulatory factor may be involved in either blocking or promoting the binding. From the side of the replisome, during S phase, post-translational modification of Sen1 replisome binding partners may promote its binding. Alternatively, the orientation of Sen1 partners within the context of the replisome may be the only conformation in which the proteins can support Sen1 binding.

phase. Possibilities include a post-translational modification of its binding partners specifically at replisomes to promote binding. Alternatively, the conformation of its binding partners within the context of the replisome could expose their binding surfaces in such a way that supports the interaction.

Though this data suggests that the binding regulation doesn't depend on Sen1 itself, there are several caveats to this experiment. While it appears unlikely that a regulatory factor blocks association of Sen1 with Ctf4 and Mrc1, we cannot exclude that adding the Sen1 to extracts with replisomes could cause a regulatory factor to dissociate. For example, the affinity of Sen1 for Mrc1 and Ctf4 might be much higher in the context of the replisome and thus displace some other regulatory factor. In addition, as other modifying factors are also present in the S phase extract, we cannot exclude that Sen1 itself undergoes a post-translational modification specifically at replication forks to promote its binding.

An approach was taken to try and disrupt the replisome during S phase; in order to determine if the conformation of Sen1 binding partners within the context of the replisome is specifically required to support the interaction, or whether it is an event that occurs downstream of replisome formation. We hypothesised that if the binding of Sen1 to Mrc1 and Ctf4 was lost following disassembly of the replisome during S phase, a specific conformation of these proteins within that context exposing their Sen1 binding sites is required. Alternatively, if the binding was maintained, it could be an event downstream of replisome formation, such as a post-translational modification at forks that is required. Unfortunately, tagging the core helicase component Mcm7 with an auxin degron did not efficiently break apart the replisome in *SEN1-TAP* cells to assess this any further (fig 5.14b). Future experiments could look at tagging other core replisome components with the aim of destabilising it during S phase in *SEN1-TAP* cells.

6.7 The interaction between Sen1 and the replisome does not depend on phosphorylation

As the binding between Sen1 and the replisome appears to be regulated by the cell extract, an obvious candidate is the S phase kinases CDK and DDK, which become active during early S phase to phosphorylate their target proteins. Both Mrc1 and Ctf4 have also been identified as part of the budding yeast phosphoproteome (Li et al., 2007, Albuquerque et al., 2008, Holt et al., 2009). However, IPs in the presence of phosphatase had no effect on the ability of Sen1 to bind to the replisome during S phase (fig 5.15). Thus, we conclude that phosphorylation of Sen1 or its binding partners at the fork does not appear to regulate the interaction. It is possible that other post-translational modifications may be responsible, such as SUMOylation, and further experiments could be carried out to test their effects.

6.8 Conservation of Sen1 binding to the replisome

The catalytic domain of Sen1 is highly conserved among eukaryotes; within this region there is approximately 30% sequence homology between yeast Sen1 and its human orthologue Senataxin (Leonaitė et al., 2017). The presence of an extended N-terminal domain connected to the helicase domain is also common to all orthologues. However, though the primary sequence of the N-terminal is conserved among other yeasts and marine vertebrates, including *S. pombe*, Zebrafish and *Xenopus*, a high degree of divergence at the primary amino acid level has occurred in mammals (Bennett and La Spada, 2015). Arguably however, as this domain has been retained as part of the protein, this suggests that it may have some conservation of function, as it is likely any mutations affecting essential activities or interactions of the N-terminal have not been selected for throughout the course of evolution.

Like *S. cerevisiae* Sen1, the N-terminal domain of human Senataxin also serves to mediate protein-protein interactions. In HeLa cells, Senataxin has been shown to interact with RNAPII, however this occurs via the RPB1, RPB2 and RPB3 subunits, rather than via the CTD as in yeast (Yüce and West, 2013). In addition, Senataxin also appears to have retained its role in coupling the nuclear exosome to RNAPII transcription, where SUMOylation of the N-terminal domain allows it to bind to the

Rrp45 subunit of the exosome, resulting in its recruitment to sites of R loop forming TRCs and transcription-induced DNA damage (Richard et al., 2013, Richard and Manley, 2014). Conversely, unlike budding yeast Sen1, no interaction with either RNAPI or RNAPIII was detected, suggesting the activity of Senataxin may be restricted to RNAPII genes in humans (Yüce and West, 2013). In addition, Senataxin does not appear to interact with Rad2 and no evidence exists to suggest it plays a role in transcription coupled-nucleotide excision repair (Bennett and La Spada, 2015).

It is not currently known whether Senataxin is recruited/tethered to active replication forks in human cells in a manner similar to Sen1, however several lines of evidence suggest that it plays an important role for the accurate duplication of chromosomes. Senataxin depleted cells are more sensitive to DNA damaging agents (Lavin et al., 2008), and like Sen1, it interacts with several DNA repair proteins. In human cells, interactors identified include RAD50 and Mre11, involved in responding to double strand breaks (Yüce and West, 2013). Senataxin has also been demonstrated to form nuclear foci during S/G2 phase, the levels of which are increased in response to replication stress induced by treatment with the replicative polymerase inhibitor aphidicolin; as well as DNA damage caused by various genotoxic agents. Moreover, these foci co-localise with the DNA damage response proteins 53BP1 and γH2A. In light of our data which suggests one of the functions of Sen1 at forks may be to resolve R loops at sites of TRCs; it is interesting that in human cells, formation of these nuclear Senataxin foci during S phase depends on transcription, and is increased or decreased by exacerbating or reducing the levels of R loops within cells, respectively. This suggests that in human cells, Senataxin may also act to clear R loops at sites of TRCs to prevent the deleterious consequences that result from their persistence (Yüce and West, 2013). Future studies should examine if Senataxin tethering to replisomes is conserved in human cells and if so, by what mechanism it occurs. Analysing any potential conservation of Sen1's role at forks would be interesting in light of Senataxin involvement in human disease. Finally, the various paralogues of Senataxin should also be investigated. These include RENT1 and IGHMBP2, which like Senataxin have also been linked to neurological disease, as well as Aquarius and ZNFx1 (Barmada et al., 2015, Guenther et al., 2009, Bennett and La Spada, 2015, Sollier et al., 2014, Chen et al., 2004).

6.9 The deregulation of Senataxin function and human disease

Fifty mutations in the *SETX* gene are known to result in two progressive neurodegenerative disorders. The most common is the juvenile onset (3-30yrs) inherited recessive disorder Ataxia with oculomotor apraxia type 2 (AOA2). This disease is progressively debilitating as a result of damage to the part of the brain that controls movement (cerebellar ataxia) with characteristic defects in voluntary eye movements (oculomotor apraxia) (Moreira and Koenig, 1993). Inserting the conserved AOA2-causing mutations within the helicase domain into yeast showed a common outcome was deregulation of Sen1 transcription termination function (Chen et al., 2014b). Of the AOA2-causing mutations in the N-terminal domain (which cannot be introduced into yeast due to the primary sequence divergence), three have been shown to prevent SUMOylation of Senataxin. This stops it from targeting the nuclear exosome to sites of DNA damage, as it can no longer bind Rrp45 (Richard et al., 2013). This is particularly interesting in light of data which has shown that during S phase, the nuclear co-localisation of Rrp45 and Senataxin during replication stress has been linked to the formation of co-transcriptional R loops and TRCs (Yüce and West, 2013).

The second, unrelated disease, juvenile amyotrophic lateral sclerosis (ALS4) is far rarer and is a result of dominant mutations. Two of the identified disease-causing mutations are within the N-terminal of Senataxin, and two are within the helicase domain. Its clinical features include early onset and slow progression, with severe muscle wasting and degeneration of motor neurones in the brain and spinal cord. One study examined the most common variant, L389S, via yeast two hybrid of a human brain expression library, comparing the interactome with that of wild type Senataxin. A specific interaction between L3896 and the antisense transcript of the noncoding RNA BCYRN was reproducibly detected. This raised the question of whether aberrant interactions could interfere with normal protein function, leading to disease (Bennett et al., 2013).

One study generated *Setx* knockout mice in an attempt to examine whether R loops could contribute to neuropathy. In this model, male infertility and germ cell apoptosis

was observed as a result of R loops accumulation in the *Setx*$^{-/-}$ mice. Surprisingly, contrary to these proliferating cells, no R loops were observed in post-mitotic nerve cells, which have high transcriptional activity (Yeo et al., 2014). It is important to note however, this murine model did not exhibit any phenotypes related to ataxia or neurodegeneration, therefore does not rule out the contribution of R loops to degeneration of post-mitotic neurones in humans (Bennett and La Spada, 2015). How mutations within Senataxin actually lead to the pathophysiology of these diseases remains an interesting topic for future study. Although most of the clinical characteristics of ALS4/AOA2 are associated with post-replicative neurons, investigating whether any R loops linked defects are present in the patients actively replicating cells would be of interest.

Finally, there is some indirect evidence which has led to the tentative suggestion that dysfunction of Senataxin could contribute to the development of some cancers (Zhao et al., 2010, Ruiz-Ballesteros et al., 2005). It has also been linked to the DNA damage response proteins BRCA1 and BRCA2, mutations of which are common across multiple cancers. The BRCA pathway is involved in both the stabilisation of stalled forks by protecting them from nucleolytic cleavage, as well as playing a role in promotion of HR mediated repair at sites of DNA damage (reviewed (Byrum et al., 2019)). It has been shown that Senataxin interacts with BRCA1 at R loop dependent terminator sites, where they act together to resolve R loops. High levels of R loop dependent DNA damage leading to insertion or deletion mutations is observed when the interaction between Senataxin and BRCA1 is disturbed. Interestingly, these types of genome rearrangements are seen near BRCA1 R loop binding sites in BRCA1 tumours (Hatchi et al., 2015, Brambati et al., 2020).

Conclusions and future perspectives

Based on current literature and the work presented in this thesis, the proposed role of Sen1 at replication forks is illustrated by the model shown in figure 6.7. The anchoring of Sen1 to replisomes looks to be important for the resolution of R loops only under certain conditions; for example, collisions with R loops that either have distinct biochemical features or are located at certain areas on the genome. As it appears Sen1 is also important to promote fork progression in other circumstances, outside of any R loops resolving activity, we suggest that it could travel with forks to terminate stalled RNA polymerases at sites of transcription replication conflicts. Sen1 acts with both Nrd1 and Nab3 as part of the NNS complex for termination, neither of which travel with the fork (Appanah et al., 2020). However, *in vitro,* the helicase activity of Sen1 alone is sufficient to dissociate RNAPII from DNA, and its presence at forks uniquely positions it to resolve TRCs (Porrua and Libri, 2013). We suggest the growth defects, increased recombination and genome instability of *sen1-3* mutants is linked to more persistent fork stalling at sites of replisome collisions, while *sen1-3* or another redundant factor is recruited more slowly, independently from the replisome to remove the blockage (fig 6.7). The stalled forks are more susceptible to cleavage and collapse, or could be recognised as deleterious recombinogenic intermediates, where unscheduled recombination leads to genome instability. Correspondingly, *sen1-3* cells are more reliant on certain genes involved in replication fork progression and stability as well as mediators/effectors of the S phase checkpoint. Taken together, it appears that *sen1-3* cells require these factors to maintain stability of the stalled forks and promote their restart, likely following collision events between forks and R loops or transcription complexes. Future experiments should be carried out to clearly show whether Sen1 acts at sites of TRCs to dissociate RNAPII from the DNA. Finally, the data in this thesis goes some way towards beginning to elucidate the mechanism of Sen1 interaction with replication forks. We have shown that the apparent cell cycle regulation of full length Sen1 binding to Ctf4 and the replisome, specifically during S phase, does not depend on the conformation of Sen1, or on phosphorylation. The future experiments described in chapter 6 should be carried out to further understand the mechanism of association.

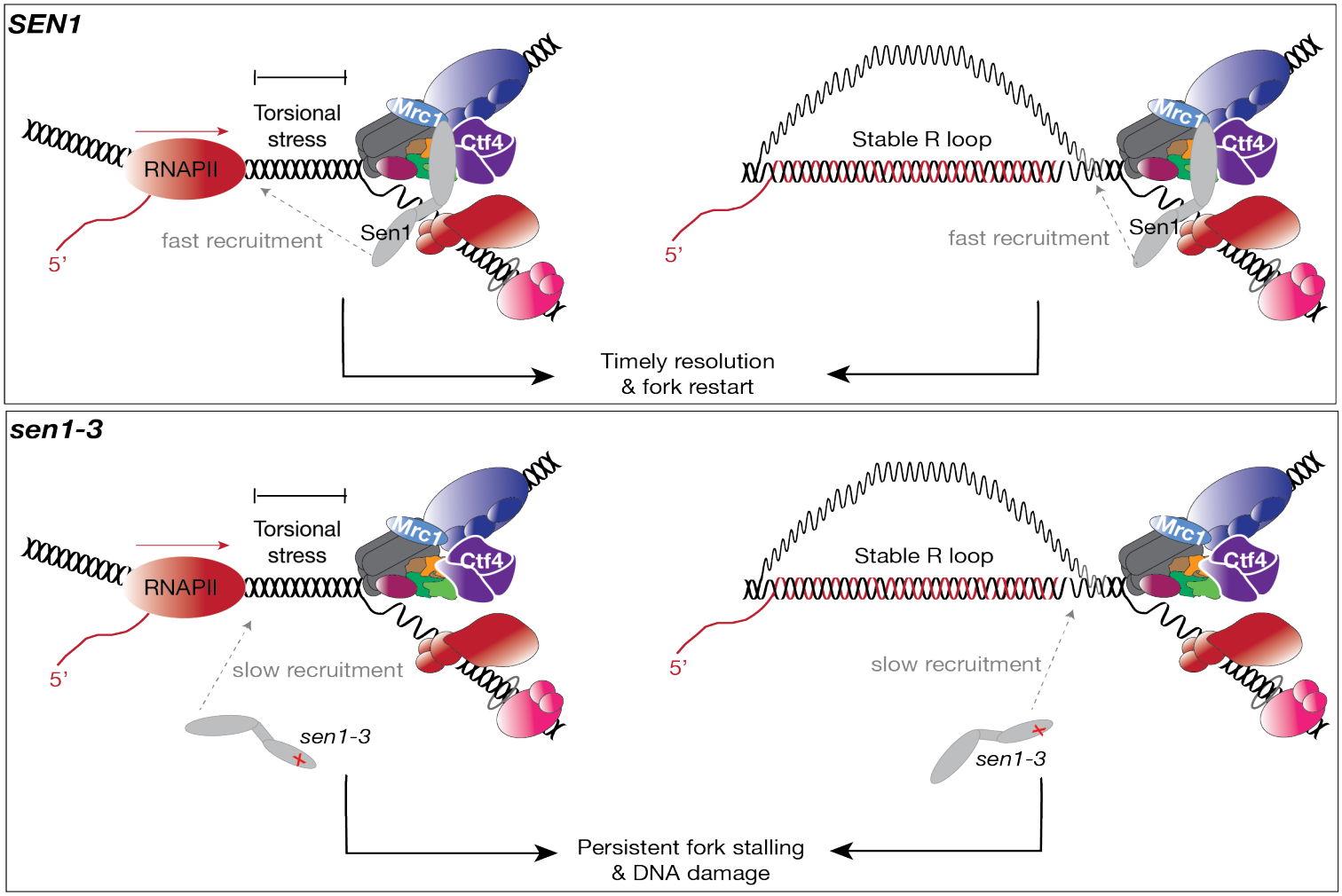

SEN1
Torsional stress
RNAPII
Mrc1
Ctf4
Sen1
fast recruitment
5'
Stable R loop
Timely resolution & fork restart
sen1-3
slow recruitment
Persistent fork stalling & DNA damage

Figure 6.7 Speculative model for the role of Sen1 at replication forks. The tethering of Sen1 to the replisome by Ctf4 and Mrc1 results in its fast recruitment to sites where its biological activity is required, namely sites where replication forks encounter barriers that impede their progression and lead to fork stalling. Sen1 may be involved in the removal of R loops at these sites, or in the termination of RNAPII transcription, thus allowing the forks to continue. In the absence of the Sen1 replisome tethering mechanism, the forks may rely on redundant factors or recruitment of the *sen1-3* allele by an alternative mechanism, which likely occurs with comparatively slower kinetics. This increases the chance of defects occurring as a result of prolonged fork stalling, leading to increased DNA damage and recombination. When Sen1 is absent from the replisome, cells rely on various S phase checkpoint and fork stability genes to promote replication re-start and maintain the stability of the genome.

Supplementary data

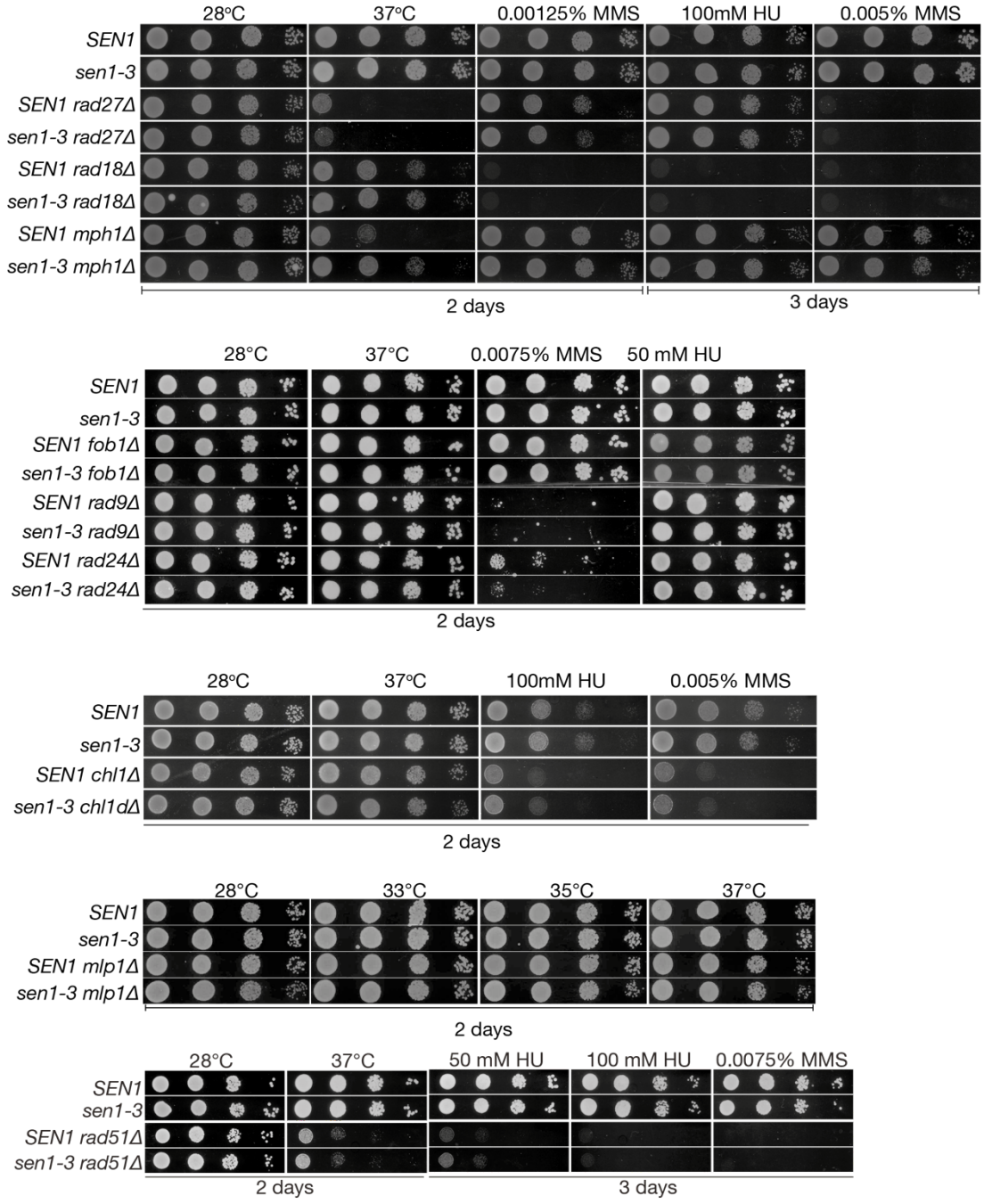


Figure S.1 *sen1-3* displays no defects in combination with deletion of various genes involved in DNA replication and repair. Dilution spotting assays were carried out, where serial dilutions of the described strains were spotted onto various different media and grown at different conditions. Plates were imaged and analysed for growth every 24 hours. Strains used in these experiments: CS2808, CS2810, CS4372, CS4374, CS4378, CS4380, CS4382, CS4384, CS2959, CS2961, CS2963, CS2965, CS2967, CS2969, CS4296, CS4299, CS3782, cS3790, CS2897, CS2898.

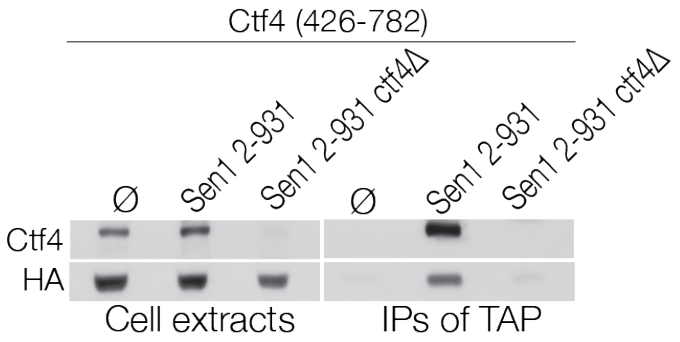


Figure S.2 3HA tagged Ctf4 fragments do not interact non-specifically with the TAP tag. Asynchronous cultures of diploid cells carrying either *GAL1-TAP_empty or GAL1-*Sen1 (2-931) with or without endogenous *CTF4* combined with the candidate *3HA* tagged Ctf4 fragment (426-782) also under control of the *GAL1* promoter were induced for 2 hours and harvested for IP to assess binding of the fragments. Strains used in this experiment: CS4037, CS4108, CS4107

References

AGARWAL, R., TANG, Z., YU, H. & COHEN-FIX, O. 2003. Two distinct pathways for inhibiting pds1 ubiquitination in response to DNA damage. *J Biol Chem,* 278**,** 45027-33.

AGUILERA, A. & GARCÍA-MUSE, T. 2012. R-Loops: From transcription byproducts to threats to genome stability. *Molecular Cell,* 46**,** 115- 124.

ALABERT, C., BIANCO, J. N. & PASERO, P. 2009. Differential regulation of homologous recombination at DNA breaks and replication forks by the Mrc1 branch of the S-phase checkpoint. *EMBO J,* 28**,** 1131-41.

ALBUQUERQUE, C. P., SMOLKA, M. B., PAYNE, S. H., BAFNA, V., ENG, J. & ZHOU, H. 2008. A multidimensional chromatography technology for in-depth phosphoproteome analysis. *Mol Cell Proteomics,* 7**,** 1389-96.

ALCASABAS, A. A., OSBORN, A. J., BACHANT, J., HU, F., WERLER, P. J., BOUSSET, K., FURUYA, K., DIFFLEY, J. F., CARR, A. M. & ELLEDGE, S. J. 2001. Mrc1 transduces signals of DNA replication stress to activate Rad53. *Nat Cell Biol,* 3**,** 958-65.

ALI, A. F., DOUGLAS, M. E., LOCKE, J., PYE, V. E., NANS, A., DIFFLEY, J. F. X. & COSTA, A. 2017. Cryo-EM structure of a licensed DNA replication origin. *Nat Commun,* 8**,** 2241.

ALZU, A., BERMEJO, R., BEGNIS, M., LUCCA, C., PICCINI, D., CAROTENUTO, W., SAPONARO, M., BRAMBATI, A., COCITO, A., FOIANI, M. & LIBERI, G. 2012. Senataxin associates with replication forks to protect fork integrity across RNA-Polymerase-II-transcribed genes. *Cell,* 151**,** 835- 846.

AMIN, A. C., M. LIANG, C. 2019. DNA Replication-Initiation Proteins in Eukaryotic Cells. *Biomedical Journal of Scientific & Technical Research,* 23.

AMON, J. D. & KOSHLAND, D. 2016. RNase H enables efficient repair of R-loop induced DNA damage. *Elife,* 5.

APPANAH, R., LONES, E. C., AIELLO, U., LIBRI, D. & DE PICCOLI, G. 2020. Sen1 Is Recruited to Replication Forks via Ctf4 and Mrc1 and Promotes Genome Stability. *Cell Rep,* 30**,** 2094-2105 e9.

ARIGO, J. T., CARROLL, K. L., AMES, J. M. & CORDEN, J. L. 2006. Regulation of yeast NRD1 expression by premature transcription termination. *Mol Cell,* 21**,** 641-51.

AYLON, Y. & KUPIEC, M. 2004. New insights into the mechanism of homologous recombination in yeast. *Mutat Res,* 566**,** 231-48.

AZVOLINSKY, A., GIRESI, P. G., LIEB, J. D. & ZAKIAN, V. A. 2009. Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae. *Mol Cell,* 34**,** 722-34.

BACAL, J., MORIEL-CARRETERO, M., PARDO, B., BARTHE, A., SHARMA, S., CHABES, A., LENGRONNE, A. & PASERO, P. 2018. Mrc1 and Rad9 cooperate to regulate initiation and elongation of DNA replication in response to DNA damage. *EMBO J,* 37.

BALDACCI, G., CHERIF-ZAHAR, B. & BERNARDI, G. 1984. The initiation of DNA replication in the mitochondrial genome of yeast. *The EMBO Journal,* 3**,** 2115- 2120.

BALK, B., MAICHER, A., DEES, M., KLERMUND, J., LUKE-GLASER, S., BENDER, K. & LUKE, B. 2013. Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. *Nature Structural Molecular Biology,* 20**,** 1199- 1205.

BANDO, M., KATOU, Y., KOMATA, M., TANAKA, H., ITOH, T., SUTANI, T. & SHIRAHIGE, K. 2009. Csm3, Tof1, and Mrc1 form a heterotrimeric mediator complex that associates with DNA replication forks. *The Journal of Biological Chemistry,* 284**,** 34355- 34365.

BARETIC, D., JENKYN-BEDFORD, M., ARIA, V., CANNONE, G., SKEHEL, M. & YEELES, J. T. P. 2020. Cryo-EM Structure of the Fork Protection Complex Bound to CMG at a Replication Fork. *Mol Cell,* 78**,** 926-940 e13.

BARLOW, J., H., FARYABI, R., B., CALLÉN, E., WONG, N., MALHOWSKI, A., CHEN, H., T., GUTIERREZ-CRUZ, G., SUN, H., W., MCKINNON, P., WRIGHT, G., CASELLAS, R., ROBBIANI, D., F., STAUDT, L., FERNANDEZ-CAPETILLO, O. & NUSSENZWEIG, A. 2013. Identification of early replicating fragile sites that contribute to genome instability. *Cell,* 152**,** 620- 632.

BARMADA, S., J., JU, S., ARJUN, A., BATARSE, A., ARCHBOLD, H., C., PEISACH, D., LI, X., ZHANG, Y., TANK, E., M., QIU, H., HUANG, E., J., RINGE, D., PETSKO, G., A. & FINKBEINER, S. 2015. Amelioration of toxicity in neuronal models of amyotrophic lateral sclerosis by hUPF1. *Proceedings for the National Academy of Sciences of the USA,* 112**,** 7821-7826.

BARNUM, K., J. & O'CONNELL, M., J. 2014. Cell cycle regulation by checkpoints. *Methods in Molecular Biology,* 1170**,** 29- 40.

BARRAL, Y., JENTSCH, S. & MANN, C. 1995. G1 cyclin turnover and nutrient uptake are controlled by a common pathway in yeast. *Genes Dev,* 9**,** 399-409.

BASTOS DE OLIVEIRA, F. M., KIM, D., CUSSIOL, J. R., DAS, J., JEONG, M. C., DOERFLER, L., SCHMIDT, K. H., YU, H. & SMOLKA, M. B. 2015. Phosphoproteomics reveals distinct modes of Mec1/ATR signaling during DNA replication. *Mol Cell,* 57**,** 1124-1132.

BELL, S. D. 2019. Initiating DNA replication: a matter of prime importance. *Biochem Soc Trans,* 47**,** 351-356.

BELL, S. P. & LABIB, K. 2016. Chromosome Duplication in Saccharomyces cerevisiae. *Genetics,* 203**,** 1027-67.

BENNETT, C., L., CHEN, Y., VIGNALI, M., LO, R., S., MASON, A., G., UNAL, A., HUQ SAIFEE, N., P., FIELDS, S. & LA SPADA, A., R. 2013. Protein interaction analysis of Senataxin and the ALS4 L389S mutant yields insights into Senataxin post-translational modification and uncovers mutant-specific binding with a brain cytoplasmic RNA-encoded peptide. *PLoS ONE,* 8**,** e78837.

BENNETT, C., L. & LA SPADA, A. R. 2015. Unwinding the role of senataxin in neurodegeneration. *Discovery Medicine,* 19**,** 127- 136.

BERMEJO, R., CAPRA, T., JOSSEN, R., COLOSIO, A., FRATTINI, C., CAROTENUTO, W., COCITO, A., DOKSANI, Y., KLEIN, H., GÓMEZ-GONZÁLEZ, B., AGUILERA, A., KATOU, Y., SHIRAHIGE, K. & FOIANI, M. 2011. The replication checkpoint protects fork stability by releasing transcribed genes from nuclear pores. *Cell,* 146**,** 233-246.

BERMUDEZ, V., P., FARINA, A., TAPPIN, I. & HURWITZ, J. 2010. Influence of the human cohesion establishment factor Ctf4/AND-1 on DNA replication. *The Journal of Biological Chemistry,* 285**,** 9493- 9505.
BLATTNER, F., R., PLUNKETT, G., 3RD., BLOCH, C., A., PERNA, N., T., BURLAND, V., RILEY, M., COLLADO-VIDES, J., GLASNER, J., D., RODE, C., K., MAYHEW, G., F., GREGOR, J., DAVIS, N., W., KIRKPATRICK, H., A., GOEDEN, M., A., ROSE, D., J., MAU, B. & SHAO, Y. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science,* 277**,** 1453- 1462.
BLEICHERT, F. 2019. Mechanisms of replication origin licensing: a structural perspective. *Curr Opin Struct Biol,* 59**,** 195-204.
BLOOM, J. & CROSS, F. R. 2007. Multiple levels of cyclin specificity in cell-cycle control. *Nat Rev Mol Cell Biol,* 8**,** 149-60.
BOCHMAN, M., L. & SCHWACHA, A. 2008. The Mcm2-7 complex has *in vitro* helicase activity. *Molecular Cell,* 31**,** 287- 293.
BODENMILLER, B., WANKA, S., KRAFT, C., URBAN, J., CAMPBELL, D., PEDRIOLI, P., G., GERRITS, B., PICOTTI, P., LAM, H., VITEK, O., BRUSNIAK, M., Y., ROSCHITZKI, B., ZHANG, C., SHOKAT, K., M., SCHLAPBACH, R., COLMAN-LERNER, A., NOLAN, G., P., NESVIZHSKII, A., I., PETER, M., LOEWITH, R., VON MERING, C. & AEBERSOLD, R. 2010. Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. *Science Signalling,* 3**,** rs4.
BOEHM, E. M., GILDENBERG, M. S. & WASHINGTON, M. T. 2016. The Many Roles of PCNA in Eukaryotic DNA Replication. *Enzymes,* 39**,** 231-54.
BOULÉ, J., B. & ZAKIAN, V., A. 2007. The yeast Pif1p DNA helicase preferentially unwinds RNA–DNA substrates. *Nucleic Acids Research,* 35**,** 5809-5818.
BOWMAN, G. D., O'DONNELL, M. & KURIYAN, J. 2004. Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. *Nature,* 429**,** 724-30.
BRAMBATI, A., ZARDONI, L., ACHAR, Y. J., PICCINI, D., GALANTI, L., COLOSIO, A., FOIANI, M. & LIBERI, G. 2018. Dormant origins and fork protection mechanisms rescue sister forks arrested by transcription. *Nucleic Acids Res,* 46**,** 1227-1239.
BRAMBATI, A., ZARDONI, L., NARDINI, E., PELLICIOLI, A. & LIBERI, G. 2020. The dark side of RNA:DNA hybrids. *Mutat Res,* 784**,** 108300.
BRANZEI, D. & FOIANI, M. 2007. Interplay of replication checkpoints and repair proteins at stalled replication forks. *DNA Repair (Amst),* 6**,** 994-1003.
BROACH, J. R., LI, Y. Y., FELDMAN, J., JAYARAM, M., ABRAHAM, J., NASMYTH, K. A. & HICKS, J. B. 1983. Localization and sequence analysis of yeast origins of DNA replication. *Cold Spring Harb Symp Quant Biol,* 47 Pt 2**,** 1165-73.
BURGERS, P. M. 2009. Polymerase dynamics at the eukaryotic DNA replication fork. *J Biol Chem,* 284**,** 4041-5.
BURGERS, P. M. J. & KUNKEL, T. A. 2017. Eukaryotic DNA Replication Fork. *Annu Rev Biochem,* 86**,** 417-438.
BUSER, R., KELLNER, V., MELNIK, A., WILSON-ZBINDEN, C., SCHELLHAAS, R., KASTNER, L., PIWKO, W., DEES, M., PICOTTI, P., MARIC, M., LABIB, K., LUKE, B. & PETER, M. 2016. The Replisome-

Coupled E3 Ubiquitin Ligase Rtt101Mms22 Counteracts Mrc1 Function to Tolerate Genotoxic Stress. *PLoS Genet,* 12**,** e1005843.
BYLUND, G. O. & BURGERS, P. M. 2005. Replication protein A-directed unloading of PCNA by the Ctf18 cohesion establishment complex. *Mol Cell Biol,* 25**,** 5445-55.
BYRUM, A. K., VINDIGNI, A. & MOSAMMAPARAST, N. 2019. Defining and Modulating 'BRCAness'. *Trends Cell Biol,* 29**,** 740-751.
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 and Development,* 19**,** 1905- 1919.
CAN, G., KAUERHOF, A. C., MACAK, D. & ZEGERMAN, P. 2019. Helicase Subunit Cdc45 Targets the Checkpoint Kinase Rad53 to Both Replication Initiation and Elongation Complexes after Fork Stalling. *Mol Cell,* 73**,** 562-573.e3.
CARROLL, K., L., GHIRLANDO, R., AMES, J., M. & CORDEN, J., L. 2007. Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *RNA,* 13**,** 361- 373.
CARROLL, K. L., PRADHAN, D. A., GRANEK, J. A., CLARKE, N. D. & CORDEN, J. L. 2004. Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts. *Mol Cell Biol,* 24**,** 6241-52.
CEJKA, P. & KOWALCZYKOWSKI, S. C. 2010. The full-length Saccharomyces cerevisiae Sgs1 protein is a vigorous DNA helicase that preferentially unwinds holliday junctions. *J Biol Chem,* 285**,** 8290-301.
CERRITELLI, S. M. & CROUCH, R. J. 2009. Ribonuclease H: the enzymes in eukaryotes. *FEBS J,* 276**,** 1494-505.
CHAKRABARTI, S., JAYACHANDRAN, U., BONNEAU, F., FIORINI, F., BASQUIN, C., DOMCKE, S., LE HIR, H. & CONTI, E. 2011. Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol Cell,* 41**,** 693-703.
CHANG, E., Y-C., NOVOA, C., A., ARISTIZABAL, M., J., COULOMBE, Y., SEGOVIA, R., CHATURVEDI, R., SHEN, Y., KEONG, C., TAM, A., S., JONES, S., J, M., MASSON, J.-Y., KOBOR, M., S. & STIRLING, P., C. 2017. RECQ-like helicases Sgs1 and BLM regulate R-loop-associated genome instability. *The Journal of Cell Biology,* 216**,** 3991-4005.
CHANG, M., BELLAOUI, M., BOONE, C. & BROWN, G. W. 2002. A genome-wide screen for methyl methanesulfonate-sensitive mutants reveals genes required for S phase progression in the presence of DNA damage. *Proc Natl Acad Sci U S A,* 99**,** 16934-9.
CHÁVEZ, S. & AGUILERA, A. 1997. The yeast *HPR1* gene has a functional role in transcriptional elongation that uncovers a novel source of genome instability. *Genes and Development,* 11**,** 3459- 3470.
CHÁVEZ, S., BEILHARZ, T., RONDÓN, A., G., ERDJUMENT-BROMAGE, H., TEMPST, P., SVEJSTRUP, J., Q., LITHGOW, T. & AGUILERA, A. 2000. A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. *The EMBO Journal,* 19**,** 5824-5834.
CHEN, E. S., HOCH, N. C., WANG, S. C., PELLICIOLI, A., HEIERHORST, J. & TSAI, M. D. 2014a. Use of quantitative mass spectrometric analysis to elucidate the mechanisms of phospho-priming and auto-activation of the checkpoint kinase Rad53 in vivo. *Mol Cell Proteomics,* 13**,** 551-65.

CHEN, S. & BELL, S. P. 2011. CDK prevents Mcm2-7 helicase loading by inhibiting Cdt1 interaction with Orc6. *Genes Dev,* 25**,** 363-72.
CHEN, S. H., ALBUQUERQUE, C. P., LIANG, J., SUHANDYNATA, R. T. & ZHOU, H. 2010. A proteome-wide analysis of kinase-substrate network in the DNA damage response. *J Biol Chem,* 285**,** 12803-12.
CHEN, S. H. & ZHOU, H. 2009. Reconstitution of Rad53 activation by Mec1 through adaptor protein Mrc1. *J Biol Chem,* 284**,** 18593-604.
CHEN, X., MULLER, U., SUNDLING, K. E. & BROW, D. A. 2014b. Saccharomyces cerevisiae Sen1 as a model for the study of mutations in human Senataxin that elicit cerebellar ataxia. *Genetics,* 198**,** 577-90.
CHEN, Y., Z., BENNETT, C., L., HUYNH, H., M., BLAIR, I., P., PULS, I., IROBI, J., DIERICK, I., ABEL, A., KENNERSON, M., L., RABIN, B., A., NICHOLSON, G., A., AUER-GRUMBACH, M., WAGNER, K., DE JONGHE, P., GRIFFIN, J., W., FISCHBECK, K., H., TIMMERMAN, V., CORNBLATH, D., R. & CHANCE, P., F. 2004. DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *American Journal of Human Genetics,* 74**,** 1128- 1135.
CHENG, Z., MUHLRAD, D., LIM, M. K., PARKER, R. & SONG, H. 2007. Structural and functional insights into the human Upf1 helicase core. *EMBO J,* 26**,** 253-64.
CHI, X., LI, Y. & QIU, X. 2020. V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. *Immunology,* 160**,** 233-247.
CHILKOVA, O., STENLUND, P., ISOZ, I., STITH, C., M., GRABOWSKI, P., LUNDSTRÖM, E., B., BURGERS, P., M. & JOHANSSON, E. 2007. The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic Acids Research,* 35**,** 6588- 6597.
CHINCHILLA, K., RODRIGUEZ-MOLINA, J., B., URSIC, D., FINKEL, J., S., ANSARI, A., Z. & CULBERTSON, M., R. 2012. Interactions of Sen1, Nrd1, and Nab3 with multiple phosphorylated forms of the Rpb1 C-terminal domain in *Saccharomyces cerevisiae*. *Eukaryotic Cell,* 11**,** 417- 429.
COLLIN, P., JERONIMO, C., POITRAS, C. & ROBERT, F. 2019. RNA Polymerase II CTD Tyrosine 1 Is Required for Efficient Termination by the Nrd1-Nab3-Sen1 Pathway. *Mol Cell,* 73**,** 655-669 e7.
CONRAD, N., K., WILSON, S., M., STEINMETZ, E., J., PATTURAJAN, M., BROW, D., A., SWANSON, M., S. & CORDEN, J., L. 2000. A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. *Genetics,* 154**,** 557- 571.
COSTANTINO, L. & KOSHLAND, D. 2018. Genome-wide Map of R-Loop-Induced Damage Reveals How a Subset of R-Loops Contributes to Genomic Instability. *Mol Cell,* 71**,** 487-497 e3.
COSTANZO, M., NISHIKAWA, J. L., TANG, X., MILLMAN, J. S., SCHUB, O., BREITKREUZ, K., DEWAR, D., RUPES, I., ANDREWS, B. & TYERS, M. 2004. CDK activity antagonizes Whi5, an inhibitor of G1/S transcription in yeast. *Cell,* 117**,** 899-913.
COSTANZO, V., SHECHTER, D., LUPARDUS, P. J., CIMPRICH, K. A., GOTTESMAN, M. & GAUTIER, J. 2003. An ATR- and Cdc7-dependent DNA damage checkpoint that inhibits initiation of DNA replication. *Mol Cell,* 11**,** 203-13.

COTTA-RAMUSINO, C., FACHINETTI, D., LUCCA, C., DOKSANI, Y., LOPES, M., SOGO, J. & FOIANI, M. 2005. Exo1 processes stalled replication forks and counteracts fork reversal in checkpoint-defective cells. *Mol Cell,* 17**,** 153-9.

CRABBÉ, L., THOMAS, A., PANTESCO, V., DE VOS, J., PASERO, P. & LENGRONNE, A. 2010. Analysis of replication profiles reveals key role of RFC-Ctf18 in yeast replication stress response. *Nat Struct Mol Biol,* 17**,** 1391-7.

CROW, Y. J., LEITCH, A., HAYWARD, B. E., GARNER, A., PARMAR, R., GRIFFITH, E., ALI, M., SEMPLE, C., AICARDI, J., BABUL-HIRJI, R., BAUMANN, C., BAXTER, P., BERTINI, E., CHANDLER, K. E., CHITAYAT, D., CAU, D., DERY, C., FAZZI, E., GOIZET, C., KING, M. D., KLEPPER, J., LACOMBE, D., LANZI, G., LYALL, H., MARTINEZ-FRIAS, M. L., MATHIEU, M., MCKEOWN, C., MONIER, A., OADE, Y., QUARRELL, O. W., RITTEY, C. D., ROGERS, R. C., SANCHIS, A., STEPHENSON, J. B., TACKE, U., TILL, M., TOLMIE, J. L., TOMLIN, P., VOIT, T., WESCHKE, B., WOODS, C. G., LEBON, P., BONTHRON, D. T., PONTING, C. P. & JACKSON, A. P. 2006. Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutieres syndrome and mimic congenital viral brain infection. *Nat Genet,* 38**,** 910-6.

DASGUPTA, S., MASUKATA, H. & TOMIZAWA, J. 1987. Multiple mechanisms for initiation of ColE1 DNA replication: DNA synthesis in the presence and absence of ribonuclease H. *Cell,* 51**,** 1113-22.

DE BRUIN, R. A., MCDONALD, W. H., KALASHNIKOVA, T. I., YATES, J., 3RD & WITTENBERG, C. 2004. Cln3 activates G1-specific transcription via phosphorylation of the SBF bound repressor Whi5. *Cell,* 117**,** 887-98.

DE PICCOLI, G., KATOU, Y., ITOH, T., NAKATO, R., SHIRAHIGE, K. & LABIB, K. 2012. Replisome stability at defective DNA replication forks is independent of S phase checkpoint kinases. *Mol Cell,* 45**,** 696-704.

DELAMARRE, A., BARTHE, A., DE LA ROCHE SAINT-ANDRE, C., LUCIANO, P., FOREY, R., PADIOLEAU, I., SKRZYPCZAK, M., GINALSKI, K., GELI, V., PASERO, P. & LENGRONNE, A. 2020. MRX Increases Chromatin Accessibility at Stalled Replication Forks to Promote Nascent DNA Resection and Cohesin Loading. *Mol Cell,* 77**,** 395-410 e3.

DEMARINI, D., J., WINEY, M., URSIC, D., WEBB, F. & CULBERTSON, M., R. 1992. *SEN1*, a positive effector of tRNA-splicing endonuclease in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology,* 12**,** 2154- 2164.

DESANY, B. A., ALCASABAS, A. A., BACHANT, J. B. & ELLEDGE, S. J. 1998. Recovery from DNA replicational stress is the essential function of the S-phase checkpoint pathway. *Genes Dev,* 12**,** 2956-70.

DEVBHANDARI, S. & REMUS, D. 2020. Rad53 limits CMG helicase uncoupling from DNA synthesis at replication forks. *Nat Struct Mol Biol,* 27**,** 461-471.

DIFFLEY, J. F. 2004. Regulation of early events in chromosome replication. *Curr Biol,* 14**,** R778-86.

DOMINGUEZ-SANCHEZ, M. S., BARROSO, S., GOMEZ-GONZALEZ, B., LUNA, R. & AGUILERA, A. 2011. Genome instability and transcription elongation impairment in human cells depleted of THO/TREX. *PLoS Genet,* 7**,** e1002386.

DOUGLAS, M. E., ALI, F. A., COSTA, A. & DIFFLEY, J. F. X. 2018. The mechanism of eukaryotic CMG helicase activation. *Nature,* 555**,** 265-268.

DROLET, M., PHOENIX, P., MENZEL, R., MASSE, E., LIU, L. F. & CROUCH, R. J. 1995. Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. *Proc Natl Acad Sci U S A,* 92**,** 3526-30.

DRURY, L. S., PERKINS, G. & DIFFLEY, J. F. 2000. The cyclin-dependent kinase Cdc28p regulates distinct modes of Cdc6p proteolysis during the budding yeast cell cycle. *Curr Biol,* 10**,** 231-40.

DUA, R., LEVY, D., L. & CAMPBELL, J., L. 1999. Analysis of the essential functions of the C-terminal protein/protein interaction domain of *Saccharomyces cerevisiae* Pol ε and its unexpected ability to support growth in the absence of the DNA polymerase domain. *Journal of Biological Chemistry,* 274**,** 222283- 222288.

DUCH, A., CANAL, B., BARROSO, S. I., GARCÍA-RUBIO, M., SEISENBACHER, G., AGUILERA, A., DE NADAL, E. & POSAS, F. 2018. Multiple signaling kinases target Mrc1 to prevent genomic instability triggered by transcription-replication conflicts. *Nat Commun,* 9**,** 379.

DUROCHER, D., HENCKEL, J., FERSHT, A. R. & JACKSON, S. P. 1999. The FHA domain is a modular phosphopeptide recognition motif. *Mol Cell,* 4**,** 387-94.

DUTTA, D., SHATALIN, K., EPSHTEIN, V., GOTTESMAN, M. E. & NUDLER, E. 2011. Linking RNA polymerase backtracking to genome instability in E. coli. *Cell,* 146**,** 533-43.

EDER, P. S., WALDER, R. Y. & WALDER, J. A. 1993. Substrate specificity of human RNase H1 and its role in excision repair of ribose residues misincorporated in DNA. *Biochimie,* 75**,** 123-6.

EGLOFF, S. & MURPHY, S. 2008. Cracking the RNA polymerase II CTD code. *Trends Genet,* 24**,** 280-8.

EKUNDAYO, B. & BLEICHERT, F. 2019. Origins of DNA replication. *PLoS Genet,* 15**,** e1008320.

EL HAGE, A., FRENCH, S. L., BEYER, A. L. & TOLLERVEY, D. 2010. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev,* 24**,** 1546-58.

EL HAGE, A., WEBB, S., KERR, A. & TOLLERVEY, D. 2014. Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genetics,* 10**,** e1004716.

EMILI, A. 1998. MEC1-dependent phosphorylation of Rad9p in response to DNA damage. *Mol Cell,* 2**,** 183-9.

EVRIN, C., CLARKE, P., ZECH, J., LURZ, R., SUN, J., UHLE, S., LI, H., STILLMAN, B. & SPECK, C. 2009. A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication. *Proceedings for the National Academy of Sciences of the USA,* 106**,** 20240-20245.

FERREIRA, M. G., SANTOCANALE, C., DRURY, L. S. & J.F., D. 1999. Dbf4p, an Essential S Phase-Promoting Factor, Is Targeted for Degradation by the Anaphase-Promoting Complex. *Molecular and Cellular Biology,* 20**,** 242-248.

FORSBURG, S., L. 2004. Eukaryotic MCM proteins: Beyond replication initiation. *Microbiology and Molecular Biology Reviews,* 68**,** 109–131.

FOSS, E. J. 2001. Tof1p regulates DNA damage responses during S phase in Saccharomyces cerevisiae. *Genetics,* 157**,** 567-77.
FOSS, E. J., LAO, U., DALRYMPLE, E., ADRIANSE, R. L., LOE, T. & BEDALOV, A. 2017. SIR2 suppresses replication gaps and genome instability by balancing replication between repetitive and unique sequences. *Proc Natl Acad Sci U S A,* 114**,** 552-557.
FOURY, F. & KOLODYNSKI, J. 1983. pif mutation blocks recombination between mitochondrial rho+ and rho- genomes having tandemly arrayed repeat units in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A,* 80**,** 5345-9.
FOX, M., J., GAO, H., SMITH-KINNAMAN, W., R., LIU, Y. & MOSLEY, A., L. 2015. The exosome component Rrp6 is required for RNA polymerase II termination at specific targets of the Nrd1-Nab3 pathway. *PLoS Genetics,* 11**,** e1004999.
FRAGKOS, M., GANIER, O., COULOMBE, P. & MECHALI, M. 2015. DNA replication origin activation in space and time. *Nat Rev Mol Cell Biol,* 16**,** 360-74.
FRANCIS, L. I., RANDELL, J. C., TAKARA, T. J., UCHIMA, L. & BELL, S. P. 2009. Incorporation into the prereplicative complex activates the Mcm2-7 helicase for Cdc7-Dbf4 phosphorylation. *Genes Dev,* 23**,** 643-54.
FREDERICO, L., A., KUNKEL, T., A. & SHAW, B., R. 1990. A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry,* 29**,** 2532-2537.
FRENCH, S. 1992. Consequences of replication fork movement through transcription units in vivo. *Science,* 258**,** 1362-5.
FRIGOLA, J., HE, J., KINKELIN, K., PYE, V. E., RENAULT, L., DOUGLAS, M. E., REMUS, D., CHEREPANOV, P., COSTA, A. & DIFFLEY, J. F. X. 2017. Cdt1 stabilizes an open MCM ring for helicase loading. *Nat Commun,* 8**,** 15720.
FUJISAWA, R., OHASHI, E., HIROTA, K. & TSURIMOTO, T. 2017. Human CTF18-RFC clamp-loader complexed with non-synthesising DNA polymerase ε efficiently loads the PCNA sliding clamp. *Nucleic Acids Res*.
GAMBUS, A., JONES, R., C., SANCHEZ-DIAZ, A., KANEMALO, M., VAN DUERSEN, F., EDMONDSON, R., D. & LABIB, K. 2006. GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks. *Nature Cell Biology,* 8**,** 358- 366.
GAMBUS, A., KHOUDOLI, G., A., JONES, R., C. & BLOW, J., J. 2011. MCM2-7 form double hexamers at licensed origins in X*enopus e*gg extract. *The Journal of Biological Chemistry,* 286**,** 11855-11864.
GAMBUS, A., VAN DUERSEN, F., POLYCHRONOPOULOS, D., FOLTMAN, M., JONES, R., C., EDMONDSON, R., D. & LABIB, K. 2009. A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase α within the eukaryotic replisome. *The EMBO Journal,* 28**,** 2992-3004.
GARCÍA-BENÍTEZ, F., GAILLARD, H. & AGUILERA, A. 2017. Physical proximity of chromatin to nuclear pores prevents harmful R loop accumulation contributing to maintain genome stability. *Proceedings of the National Academy of Sciences of the USA,* 114**,** 10942-10947.
GARCÍA-PICHARDO, D., CAÑAS, J., C., GARCÍA-RUBIO, M., L., GÓMEZ-GONZÁLEZ, B., RONDÓN, A., G. & AGUILERA, A. 2017. Histone mutants separate R loop formation from genome instability induction. *Molecular Cell,* 66**,** 597-609.

GARCÍA-RODRÍGUEZ, L. J., DE PICCOLI, G., MARCHESI, V., JONES, R. C., EDMONDSON, R. D. & LABIB, K. 2015. A conserved Polϵ binding module in Ctf18-RFC is required for S-phase checkpoint activation downstream of Mec1. *Nucleic Acids Res,* 43**,** 8830-8.

GARCIA-RODRIGUEZ, N., MORAWSKA, M., WONG, R. P., DAIGAKU, Y. & ULRICH, H. D. 2018. Spatial separation between replisome- and template-induced replication stress signaling. *EMBO J,* 37.

GARCIA-RUBIO, M., AGUILERA, P., LAFUENTE-BARQUERO, J., RUIZ, J. F., SIMON, M. N., GELI, V., RONDON, A. G. & AGUILERA, A. 2018. Yra1-bound RNA-DNA hybrids cause orientation-independent transcription-replication collisions and telomere instability. *Genes Dev,* 32**,** 965-977.

GAVALDA, S., GALLARDO, M., LUNA, R. & AGUILERA, A. 2013. R-loop mediated transcription-associated recombination in trf4Delta mutants reveals new links between RNA surveillance and genome integrity. *PLoS One,* 8**,** e65541.

GELLON, L., RAZIDLO, D. F., GLEESON, O., VERRA, L., SCHULZ, D., LAHUE, R. S. & FREUDENREICH, C. H. 2011. New functions of Ctf18-RFC in preserving genome stability outside its role in sister chromatid cohesion. *PLoS Genet,* 7**,** e1001298.

GEORGESCU, R., E., LANGSTON, L., YAO, N., Y., YURIEVA, O., ZHANG, D., FINKELSTEIN, J., AGARWAL, T. & O'DONNELL, M., E. 2014. Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nature Structural and Molecular Biology,* 21**,** 664-670.

GEORGESCU, R. E., SCHAUER, G. D., YAO, N. Y., LANGSTON, L. D., YURIEVA, O., ZHANG, D., FINKELSTEIN, J. & O'DONNELL, M. E. 2015. Reconstitution of a eukaryotic replisome reveals suppression mechanisms that define leading/lagging strand operation. *Elife,* 4**,** e04988.

GERONIMO, C. L. & ZAKIAN, V. A. 2016. Getting it done at the ends: Pif1 family DNA helicases and telomeres. *DNA Repair (Amst),* 44**,** 151-158.

GHAEMMAGHAMI, S., HUH, W. K., BOWER, K., HOWSON, R. W., BELLE, A., DEPHOURE, N., O'SHEA, E. K. & WEISSMAN, J. S. 2003. Global analysis of protein expression in yeast. *Nature,* 425**,** 737-41.

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**,** 9879-86.

GINNO, P., A., LIM, Y., W., LOTT, P., L., KORF, I. & CHÉDIN, F. 2013. GC skew at the 5' and 3' ends of human genes links R-loop formation to epigenetic regulation and transcription termination. *Genome Research,* 23**,** 1590-1600.

GINNO, P., A., LOTT, P., L., CHRISTENSEN, H., C., KORF, I. & CHÉDIN, F. 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Molecular Cell,* 45**,** 814-825.

GOFFEAU, A., BARRELL, B. G., BUSSEY, H., DAVIS, R. W., DUJON, B., FELDMANN, H., GALIBERT, F., HOHEISEL, J. D., JACQ, C., JOHNSTON, M., LOUIS, E. J., MEWES, H. W., MURAKAMI, Y., PHILIPPSEN, P., TETTELIN, H. & OLIVER, S. G. 1996. Life with 6000 genes. *Science,* 274**,** 546, 563-7.

GONZÁLEZ-AGUILERA, C., TOUS, C., GÓMEZ-GONZÁLEZ, B., HUERTAS, P., LUNA, R. & AGUILERA, A. 2008. The THP1-SAC3-SUS1-CDC31

complex works in transcription elongation-mRNA export preventing RNA-mediated genome instability. *Mol Biol Cell,* 19**,** 4310-8.
GRAF, M., BONETTI, D., LOCKHART, A., SERHAL, K., KELLNER, V., MAICHER, A., JOLIVET, P., TEIXEIRA, M. T. & LUKE, B. 2017. Telomere Length Determines TERRA and R-Loop Regulation through the Cell Cycle. *Cell,* 170**,** 72-85 e14.
GRENON, M., GILBERT, C. & LOWNDES, N. F. 2001. Checkpoint activation in response to double-strand breaks requires the Mre11/Rad50/Xrs2 complex. *Nat Cell Biol,* 3**,** 844-7.
GRUBB, J., BROWN, M. S. & BISHOP, D. K. 2015. Surface Spreading and Immunostaining of Yeast Chromosomes. *J Vis Exp***,** e53081.
GUENTHER, U., P., HANDOKO, L., LAGGERBAUER, B., JABLONKA, S., CHARI, A., ALZHEIMER, M., OHMER, J., PLÖTTNER, O., GEHRING, N., SICKMANN, A., VON AU, K., SCHUELKE, M. & FISCHER, U. 2009. IGHMBP2 is a ribosome-associated helicase inactive in the neuromuscular disorder distal SMA type 1 (DSMA1). *Human Molecular Genetics,* 18**,** 1288-1300.
GUNJAN, A. & VERREAULT, A. 2003. A Rad53 kinase-dependent surveillance mechanism that regulates histone protein levels in *S. cerevisiae*. *Cell,* 115**,** 537-549.
HAMPERL, S., BOCEK, M., J., SALDIVAR, J., C., SWIGUT, T. & CIMPRICH, K., A. 2017. Transcription-replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. *Cell,* 170**,** 774-786.
HAN, Z., JASNOVIDOVA, O., HAIDARA, N., TUDEK, A., KUBICEK, K., LIBRI, D., STEFL, R. & PORRUA, O. 2020. Termination of non-coding transcription in yeast relies on both an RNA Pol II CTD interaction domain and a CTD-mimicking region in Sen1. *EMBO J,* 39**,** e101548.
HAN, Z., LIBRI, D. & PORRUA, O. 2017. Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. *Nucleic Acids Research,* 45**,** 1355–1370.
HANNA, J. S., KROLL, E. S., LUNDBLAD, V. & SPENCER, F. A. 2001. Saccharomyces cerevisiae CTF18 and CTF4 are required for sister chromatid cohesion. *Mol Cell Biol,* 21**,** 3144-58.
HATCHI, E., SKOURTI-STATHAKI, K., VENTZ, S., PINELLO, L., YEN, A., KAMIENIARZ-GDULA, K., DIMITROV, S., PATHANIA, S., MCKINNEY, K., M., EATON, M., L., KELLIS, M., HILL, S., J., PARMIGIANI, G., PROUDFOOT, N., J . & LIVINGSTON, D., M. 2015. BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Molecular Cell,* 57**,** 636-647.
HAZELBAKER, D., Z., MARQUARDT, S., WLOTZKA, W. & BURATOWSKI, S. 2013. Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. *Molecular Cell,* 49**,** 55-66.
HEGNAUER, A. M., HUSTEDT, N., SHIMADA, K., PIKE, B. L., VOGEL, M., AMSLER, P., RUBIN, S. M., VAN LEEUWEN, F., GUENOLE, A., VAN ATTIKUM, H., THOMA, N. H. & GASSER, S. M. 2012. An N-terminal acidic region of Sgs1 interacts with Rpa70 and recruits Rad53 kinase to stalled forks. *EMBO J,* 31**,** 3768-83.
HELBIG, A., O., ROSATI, S., PIJNAPPEL, P., W., VAN BREUKELEN, B., TIMMERS, M., H., MOHAMMED, S., SLIJPER, M. & HECK, A., J. 2010.

Perturbation of the yeast N-acetyltransferase NatB induces elevation of protein phosphorylation levels. *BMC Genomics,* 11**,** 685.
HELMRICH, A., BALLARINO, M., NUDLER, E. & TORA, L. 2013. Transcription-replication encounters, consequences and genomic instability. *Nature Structural and Molecular Biology,* 20**,** 412-418.
HELMRICH, A., BALLARINO, M. & TORA, L. 2011. Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Molecular Cell,* 44**,** 966-977.
HEMANN, M. T., STRONG, M. A., HAO, L. Y. & GREIDER, C. W. 2001. The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. *Cell,* 107**,** 67-77.
HIRAGA, S., ROBERTSON, E. D. & DONALDSON, A. D. 2006. The Ctf18 RFC-like complex positions yeast telomeres but does not specify their replication time. *EMBO J,* 25**,** 1505-14.
HODGSON, B., CALZADA, A. & KARIM, L. 2007. Mrc1 and Tof1 regulate DNA replication forks in different ways during normal S phase. *Molecular Biology of the Cell,* 18**,** 3894–3902.
HOGGARD, T., MULLER, C. A., NIEDUSZYNSKI, C. A., WEINREICH, M. & FOX, C. A. 2020. Sir2 mitigates an intrinsic imbalance in origin licensing efficiency between early- and late-replicating euchromatin. *Proc Natl Acad Sci U S A,* 117**,** 14314-14321.
HOLT, L. J., TUCH, B. B., VILLEN, J., JOHNSON, A. D., GYGI, S. P. & MORGAN, D. O. 2009. Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science,* 325**,** 1682-6.
HU, J., SUN, L., SHEN, F., CHEN, Y., HUA, Y., LIU, Y., ZHANG, M., HU, Y., WANG, Q., XU, W., SUN, F., JI, J., MURRAY, J. M., CARR, A. M. & KONG, D. 2012. The intra-S phase checkpoint targets Dna2 to prevent stalled replication forks from reversing. *Cell,* 149**,** 1221-32.
HUANG, M., ZHOU, Z. & ELLEDGE, S., J. 1998. The DNA replication and damage checkpoint pathways induce transcription by inhibition of the Crt1 repressor. *Cell,* 94**,** 595-605.
HUBER, A. H., NELSON, W. J. & WEIS, W. I. 1997. Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell,* 90**,** 871-82.
HUERTAS, P. & AGUILERA, A. 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Molecular Cell,* 12**,** 711-721.
ITOH, T. & TOMIZAWA, J., I. 1980. Formation of an RNA primer for initiation of replication of ColE1 DNA by ribonucleotide H. *Proceedings for the National Academy of Sciences of the USA,* 77**,** 2450-2454.
IVESSA, A. S., ZHOU, J. Q., SCHULZ, V. P., MONSON, E. K. & ZAKIAN, V. A. 2002. Saccharomyces Rrm3p, a 5' to 3' DNA helicase that promotes replication fork progression through telomeric and subtelomeric DNA. *Genes Dev,* 16**,** 1383-96.
IVESSA, A. S., ZHOU, J. Q. & ZAKIAN, V. A. 2000. The Saccharomyces Pif1p DNA helicase and the highly related Rrm3p have opposite effects on replication fork progression in ribosomal DNA. *Cell,* 100**,** 479-89.
JAIN, R., RICE, W. J., MALIK, R., JOHNSON, R. E., PRAKASH, L., PRAKASH, S., UBARRETXENA-BELANDIA, I. & AGGARWAL, A. K. 2019. Cryo-EM structure and dynamics of eukaryotic DNA polymerase delta holoenzyme. *Nat Struct Mol Biol,* 26**,** 955-962.

JANKOWSKY, E. 2011. RNA helicases at working: binding and rearranging. *Trends in Biochemical Sciences,* 36**,** 19-29.
JOHNSON, R., E., KLASSEN, R., PRAKASH, L. & PRAKASH, S. 2015. A major role of DNA Polymerase δ in replication of both the leading and lagging DNA strands. *Molecular Cell,* 59**,** 163-175.
JOHNSTON, L. H. & NASMYTH, K. A. 1978. Saccharomyces cerevisiae cell cycle mutant cdc9 is defective in DNA ligase. *Nature,* 274**,** 891-3.
KAMIMURA, Y., TAK, Y., S., SUGINO, A. & ARAKI, H. 2001. Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in Saccharomyces cerevisiae. *The EMBO Journal,* 20**,** 2097-2107.
KANG, M. S., RYU, E., LEE, S. W., PARK, J., HA, N. Y., RA, J. S., KIM, Y. J., KIM, J., ABDEL-RAHMAN, M., PARK, S. H., LEE, K. Y., KIM, H., KANG, S. & MYUNG, K. 2019. Regulation of PCNA cycling on replicating DNA by RFC and RFC-like complexes. *Nat Commun,* 10**,** 2420.
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**,** 1078-83.
KAWAUCHI, J., MISCHO, H., BRAGLIA, P., RONDON, A. & PROUDFOOT, N., J. 2008. Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes and Development,* 22**,** 1082-1092.
KIM, H., D., CHOE, J. & SEO, Y., S. 1999a. The s*en1*$^{+}$gene of S*chizosaccharomyces pombe,* a homologue of budding yeast S*EN1,* encodes an RNA and DNA helicase. *Biochemistry,* 38**,** 14697-14710.
KIM, K., HEO, D., H., KIM, I., SUH, J., Y. & KIM, M. 2016. Exosome cofactors connect transcription termination to RNA processing by guiding terminated transcripts to the appropriate exonuclease within the nuclear exosome. *Journal of Biological Chemistry,* 291**,** 13229-13242.
KIM, M., KROGAN, N., J., VASILJEVA, L., RANDO, O., J., NEDEA, E., GREENBLATT, J., F. & BURATOWSKI, S. 2004. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature,* 432**,** 517-522.
KIM, M., VASILJEVA, L., RANDO, O., J., ZHELKOVSKY, A., MOORE, C. & BURATOWSKI, S. 2006. Distinct pathways for snoRNA and mRNA termination. *Molecular Cell,* 24**,** 723-734.
KIM, N., ABDULOVIC, A., L., GEALY, R., LIPPERT, M., J. & JINKS-ROBERTSON, S. 2007. Transcription-associated mutagenesis in yeast is directly proportional to the level of gene expression and influenced by the direction of DNA replication. *DNA Repair,* 6**,** 1285-1296.
KIM, N., HUANG, S. N., WILLIAMS, J. S., LI, Y. C., CLARK, A. B., CHO, J. E., KUNKEL, T. A., POMMIER, Y. & JINKS-ROBERTSON, S. 2011. Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. *Science,* 332**,** 1561-4.
KIM, S. T., LIM, D. S., CANMAN, C. E. & KASTAN, M. B. 1999b. Substrate specificities and identification of putative substrates of ATM kinase family members. *J Biol Chem,* 274**,** 37538-43.
KIM, T. S., LIU, C. L., YASSOUR, M., HOLIK, J., FRIEDMAN, N., BURATOWSKI, S. & RANDO, O. J. 2010. RNA polymerase mapping

during stress responses reveals widespread nonproductive transcription in yeast. *Genome Biol,* 11**,** R75.
KOCH, C., SCHLEIFFER, A., AMMERER, G. & NASMYTH, K. 1996. Switching transcription on and off during the yeast cell cycle: Cln/Cdc28 kinases activate bound transcription factor SBF (Swi4/Swi6) at start, whereas Clb/Cdc28 kinases displace it from the promoter in G2. *Genes Dev,* 10**,** 129-41.
KOMATA, M., BANDO, M., ARAKI, H. & SHIRAHIGE, K. 2009. The direct binding of Mrc1, a checkpoint mediator, to Mcm6, a replication helicase, is essential for the replication checkpoint against methyl methanesulfonate-induced stress. *Mol Cell Biol,* 29**,** 5008-19.
KONDO, T., MATSUMOTO, K. & SUGIMOTO, K. 1999. Role of a complex containing Rad17, Mec3, and Ddc1 in the yeast DNA damage checkpoint pathway. *Mol Cell Biol,* 19**,** 1136-43.
KOUPRINA, N., KROLL, E., BANNIKOV, V., BLISKOVSKY, V., GIZATULLIN, R., KIRILLOV, A., SHESTOPALOV, B., ZAKHARYEV, V., HIETER, P., SPENCER, F. & LARIONOV, V. 1992. *CTF4* (*CHL15*) mutants exhibit defective DNA metabolism in the yeast *Saccharomyces cerevisiae*. *Molecular and Cellular Biology,* 12**,** 5736-5747.
KREJCI, L., VAN KOMEN, S., LI, Y., VILLEMAIN, J., REDDY, M. S., KLEIN, H., ELLENBERGER, T. & SUNG, P. 2003. DNA helicase Srs2 disrupts the Rad51 presynaptic filament. *Nature,* 423**,** 305-9.
KUBOTA, T., HIRAGA, S., YAMADA, K., LAMOND, A. I. & DONALDSON, A. D. 2011. Quantitative proteomic analysis of chromatin reveals that Ctf18 acts in the DNA replication checkpoint. *Mol Cell Proteomics,* 10**,** M110.005561.
KUBOTA, T., KATOU, Y., NAKATO, R., SHIRAHIGE, K. & DONALDSON, A. D. 2015. Replication-Coupled PCNA Unloading by the Elg1 Complex Occurs Genome-wide and Requires Okazaki Fragment Ligation. *Cell Rep,* 12**,** 774-87.
KUNKEL, T., A. & BURGERS, P., M. 2008. Dividing the workload at a eukaryotic replication fork. *Trends in Cell Biology,* 18**,** 521-527.
KUNKEL, T. A. & BURGERS, P. M. 2014. Delivering nonidentical twins. *Nat Struct Mol Biol,* 21**,** 649-51.
LABIB, K., DIFFLEY, J. F. & KEARSEY, S. E. 1999. G1-phase and B-type cyclins exclude the DNA-replication factor Mcm4 from the nucleus. *Nat Cell Biol,* 1**,** 415-22.
LAFUENTE-BARQUERO, J., LUKE-GLASER, S., GRAF, M., SILVA, S., GOMEZ-GONZALEZ, B., LOCKHART, A., LISBY, M., AGUILERA, A. & LUKE, B. 2017. The Smc5/6 complex regulates the yeast Mph1 helicase at RNA-DNA hybrid-mediated DNA damage. *PLoS Genet,* 13**,** e1007136.
LALONDE, M. & CHARTRAND, P. 2020. TERRA, a Multifaceted Regulator of Telomerase Activity at Telomeres. *J Mol Biol*.
LANG, K., S., HALL, A., N., MERRIKH, C., N., RAGHEB, M., TABAKH, H., POLLOCK, A., J., WOODWARD, J., J., DREIFUS, J., E. & MERRIKH, H. 2017. Replication-transcription conflicts generate R-loops that orchestrate bacterial stress survival and pathogenesis. *Cell,* 170**,** 787-799.
LANGSTON, L. D. & O'DONNELL, M. E. 2019. An explanation for origin unwinding in eukaryotes. *Elife,* 8.

LAVIN, M. F., GUEVEN, N. & GRATTAN-SMITH, P. 2008. Defective responses to DNA single- and double-strand breaks in spinocerebellar ataxia. *DNA Repair (Amst),* 7**,** 1061-76.
LAZZARO, F., NOVARINA, D., AMARA, F., WATT, D., L., STONE, J., E., COSTANZO, V., BURGERS, P., M., KUNKEL, T., A., PLEVANI, P. & MUZI-FALCONI, M. 2012. RNase H and postreplication repair protect cells from ribonucleotides incorporated in DNA. *Molecular Cell,* 45**,** 99-110.
LE, T. T., YANG, Y., TAN, C., SUHANOVSKY, M. M., FULBRIGHT, R. M., JR., INMAN, J. T., LI, M., LEE, J., PERELMAN, S., ROBERTS, J. W., DEACONESCU, A. M. & WANG, M. D. 2018. Mfd Dynamically Regulates Transcription via a Release and Catch-Up Mechanism. *Cell,* 172**,** 344-357 e15.
LEE, C. S., LEE, K., LEGUBE, G. & HABER, J. E. 2014. Dynamics of yeast histone H2A and H2B phosphorylation in response to a double-strand break. *Nat Struct Mol Biol,* 21**,** 103-9.
LEONAITĖ, B., HAN, Z., BASQUIN, J., BONNEAU, F., LIBRI, D., PORRUA, O. & CONTI, E. 2017. Sen1 has unique structural features grafted on the architecture of the Upf1-like helicase family. *The EMBO Journal,* 36**,** 1590-1604.
LERNER, L. K. & SALE, J. E. 2019. Replication of G Quadruplex DNA. *Genes (Basel),* 10.
LI, N., ZHAI, Y., ZHANG, Y., LI, W., YANG, M., LEI, J., TYE, B. K. & GAO, N. 2015. Structure of the eukaryotic MCM complex at 3.8 A. *Nature,* 524**,** 186-91.
LI, S. 2015. Transcription coupled nucleotide excision repair in the yeast Saccharomyces cerevisiae: The ambiguous role of Rad26. *DNA Repair (Amst),* 36**,** 43-48.
LI, W., SELVAM, K., RAHMAN, S., A. & LI, S. 2016. Sen1, the yeast homolog of human senataxin, plays a more direct role than Rad26 in transcription coupled DNA repair. *Nucleic Acids Research,* 44**,** 6794–6802.
LI, X., GERBER, S. A., RUDNER, A. D., BEAUSOLEIL, S. A., HAAS, W., VILLEN, J., ELIAS, J. E. & GYGI, S. P. 2007. Large-scale phosphorylation analysis of alpha-factor-arrested Saccharomyces cerevisiae. *J Proteome Res,* 6**,** 1190-7.
LI, X. & MANLEY, J., L. 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell,* 122**,** 365-378.
LIM, S. C., BOWLER, M. W., LAI, T. F. & SONG, H. 2012. The Ighmbp2 helicase structure reveals the molecular basis for disease-causing mutations in DMSA1. *Nucleic Acids Res,* 40**,** 11009-22.
LIN, Y. L. & PASERO, P. 2012. Interference between DNA replication and transcription as a cause of genomic instability. *Curr Genomics,* 13**,** 65-73.
LISBY, M., BARLOW, J. H., BURGESS, R. C. & ROTHSTEIN, R. 2004. Choreography of the DNA damage response: spatiotemporal relationships among checkpoint and repair proteins. *Cell,* 118**,** 699-713.
LISBY, M., ROTHSTEIN, R. & MORTENSEN, U., H. 2001. Rad52 forms DNA repair and recombination centers during S phase. *Proceedings for the National Academy of Sciences of the USA,* 98**,** 8276-8282.
LIU, B. & ALBERTS, B. 1995. Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science,* 267**,** 1131-1137.

LIU, H. W., BOUCHOUX, C., PANAROTTO, M., KAKUI, Y., PATEL, H. & UHLMANN, F. 2020. Division of Labor between PCNA Loaders in DNA Replication and Sister Chromatid Cohesion Establishment. *Mol Cell,* 78**,** 725-738 e4.

LOCKHART, A., PIRES, V. B., BENTO, F., KELLNER, V., LUKE-GLASER, S., YAKOUB, G., ULRICH, H. D. & LUKE, B. 2019. RNase H1 and H2 Are Differentially Regulated to Process RNA-DNA Hybrids. *Cell Rep,* 29**,** 2890-2900 e5.

LONGHESE, M. P., CLERICI, M. & LUCCHINI, G. 2003. The S-phase checkpoint and its regulation in Saccharomyces cerevisiae. *Mutat Res,* 532**,** 41-58.

LOPEZ-MOSQUEDA, J., MAAS, N., L., JONSSON, Z., O., DEFAZIO-ELI, L., G., WOHLSCHLEGEL, J. & TOCZYSKI, D., P. 2010. Damage-induced phosphorylation of Sld3 is important to block late origin firing. *Nature,* 467**,** 479-483.

LOU, H., KOMATA, M., KATOU, Y., GUAN, Z., REIS, C. C., BUDD, M., SHIRAHIGE, K. & CAMPBELL, J. L. 2008. Mrc1 and DNA polymerase epsilon function together in linking DNA replication and the S phase checkpoint. *Mol Cell,* 32**,** 106-17.

LUJAN, S. A., WILLIAMS, J. S. & KUNKEL, T. A. 2016. DNA Polymerases Divide the Labor of Genome Replication. *Trends Cell Biol,* 26**,** 640-654.

LUKE, B., PANZA, A., REDON, S., IGLESIAS, N., LI, Z. & LINGNER, J. 2008. The Rat1p 5' to 3' exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in Saccharomyces cerevisiae. *Molecular Cell,* 32**,** 465-477.

LUNDBLAD, V. & BLACKBURN, E., H. 1993. An alternative pathway for yeast telomere maintenance rescues e*st-* senescence. *Cell,* 73**,** 347-360.

LUNDE, B. M., REICHOW, S. L., KIM, M., SUH, H., LEEPER, T. C., YANG, F., MUTSCHLER, H., BURATOWSKI, S., MEINHART, A. & VARANI, G. 2010. Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat Struct Mol Biol,* 17**,** 1195-201.

MACULINS, T., NKOSI, P. J., NISHIKAWA, H. & LABIB, K. 2015. Tethering of SCF(Dia2) to the Replisome Promotes Efficient Ubiquitylation and Disassembly of the CMG Helicase. *Curr Biol,* 25**,** 2254-9.

MAJKA, J. & BURGERS, P. M. 2003. Yeast Rad17/Mec3/Ddc1: a sliding clamp for the DNA damage checkpoint. *Proc Natl Acad Sci U S A,* 100**,** 2249-54.

MANTIERO, D., MACKENZIE, A., DONALDSON, A. & ZEGERMAN, P. 2011. Limiting replication initiation factors execute the temporal programme of origin firing in budding yeast. *EMBO J,* 30**,** 4805-14.

MARQUARDT, S., ESCALANTE-CHONG, R., PHO, N., WANG, J., CHURCHMAN, L. S., SPRINGER, M. & BURATOWSKI, S. 2014. A chromatin-based mechanism for limiting divergent noncoding transcription. *Cell,* 157**,** 1712-23.

MARTIN-TUMASZ, S. & BROW, D., A. 2015. *Saccharomyces cerevisiae* Sen1 helicase domain exhibits 5'- to 3'- helicase activity with a preference for translocation on DNA rather than RNA. *Journal of Biological Chemistry,* 290**,** 22880-22889.

MATTHEWS, L. A., SELVARATNAM, R., JONES, D. R., AKIMOTO, M., MCCONKEY, B. J., MELACINI, G., DUNCKER, B. P. & GUARNE, A. 2014. A novel non-canonical forkhead-associated (FHA) domain-binding

interface mediates the interaction between Rad53 and Dbf4 proteins. *J Biol Chem,* 289**,** 2589-99.
MAYER, M. L., GYGI, S. P., AEBERSOLD, R. & HIETER, P. 2001. Identification of RFC(Ctf18p, Ctf8p, Dcc1p): an alternative RFC complex required for sister chromatid cohesion in S. cerevisiae. *Mol Cell,* 7**,** 959-70.
MAYLE, R., LANGSTON, L., MOLLOY, K. R., ZHANG, D., CHAIT, B. T. & O'DONNELL, M. E. 2019. Mcm10 has potent strand-annealing activity and limits translocase-mediated fork regression. *Proc Natl Acad Sci U S A,* 116**,** 798-803.
MCDONALD, K. R., GUISE, A. J., POURBOZORGI-LANGROUDI, P., CRISTEA, I. M., ZAKIAN, V. A., CAPRA, J. A. & SABOURI, N. 2016. Pfh1 Is an Accessory Replicative Helicase that Interacts with the Replisome to Facilitate Fork Progression and Preserve Genome Integrity. *PLoS Genet,* 12**,** e1006238.
MECHALI, M. 2010. Eukaryotic DNA replication origins: many choices for appropriate answers. *Nat Rev Mol Cell Biol,* 11**,** 728-38.
MENDENHALL , M., D. & HODGE , A., E. 1998. Regulation of Cdc28 cyclin-dependent protein kinase activity during the cell cycle of the yeast *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews,* 62**,** 1191–1243.
MILLER, T. C. R., LOCKE, J., GREIWE, J. F., DIFFLEY, J. F. X. & COSTA, A. 2019. Mechanism of head-to-head MCM double-hexamer formation revealed by cryo-EM. *Nature,* 575**,** 704-710.
MIRKIN, E., V. & MIRKIN, S., M. 2007. Replication fork stalling at natural impediments. *Microbiology and Molecular Biology Reviews,* 71**,** 13-35.
MISCHO, H., E., CHUN, Y., HARLEN, K., M., SMALEC, B., M., DHIR, S., STIRLING, L., S. & BURATOWSKI, S. 2018. Cell-cycle modulation of transcription termination factor Sen1. *Molecular Cell,* 70**,** 312-326.
MISCHO, H., E., GÓMEZ-GONZÁLEZ, B., GRZECHNIK, P., RONDÓN, A., G., WEI, W., STEINMETZ, L., AGUILERA, A. & PROUDFOOT, N., J. 2011. Yeast Sen1 helicase protects the genome from transcription-associated instability. *Molecular Cell,* 41**,** 21-32.
MISCHO, H. E. & PROUDFOOT, N. J. 2013. Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. *Biochim Biophys Acta,* 1829**,** 174-85.
MORAFRAILE, E. C., DIFFLEY, J. F., TERCERO, J. A. & SEGURADO, M. 2015. Checkpoint-dependent RNR induction promotes fork restart after replicative stress. *Sci Rep,* 5**,** 7886.
MOREIRA, M. C. & KOENIG, M. 1993. Ataxia with Oculomotor Apraxia Type 2. *In:* ADAM, M. P., ARDINGER, H. H., PAGON, R. A., WALLACE, S. E., BEAN, L. J. H., STEPHENS, K. & AMEMIYA, A. (eds.) *GeneReviews((R)).* Seattle (WA).
MORIEL-CARRETERO, M., PASERO, P. & PARDO, B. 2019. DDR Inc., one business, two associates. *Curr Genet,* 65**,** 445-451.
MOROHASHI, H., MACULINS, T. & LABIB, K. 2009. The amino-terminal TPR domain of Dia2 Tethers SCF^{Dia2} to the replisome progression complex. *Current Biology,* 19**,** 1943–1949.
MURAMATSU, S., HIRAI, K., TAK, Y. S., KAMIMURA, Y. & ARAKI, H. 2010. CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol (epsilon}, and GINS in budding yeast. *Genes Dev,* 24**,** 602-12.

NAVADGI-PATIL, V. M. & BURGERS, P. M. 2008. Yeast DNA replication protein Dpb11 activates the Mec1/ATR checkpoint kinase. *J Biol Chem,* 283**,** 35853-9.
NAVADGI-PATIL, V. M. & BURGERS, P. M. 2009. The unstructured C-terminal tail of the 9-1-1 clamp subunit Ddc1 activates Mec1/ATR via two distinct mechanisms. *Mol Cell,* 36**,** 743-53.
NAVAS, T. A., ZHOU, Z. & ELLEDGE, S. J. 1995. DNA polymerase epsilon links the DNA replication machinery to the S phase checkpoint. *Cell,* 80**,** 29-39.
NAYLOR, M. L., LI, J. M., OSBORN, A. J. & ELLEDGE, S. J. 2009. Mrc1 phosphorylation in response to DNA replication stress is required for Mec1 accumulation at the stalled fork. *Proc Natl Acad Sci U S A,* 106**,** 12765-70.
NEIL, H., MALABAT, C., D'AUBENTON-CARAFA, Y., XU, Z., STEINMETZ, L. M. & JACQUIER, A. 2009. Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature,* 457**,** 1038-42.
NGUYEN, H. D., YADAV, T., GIRI, S., SAEZ, B., GRAUBERT, T. A. & ZOU, L. 2017. Functions of Replication Protein A as a Sensor of R Loops and a Regulator of RNaseH1. *Mol Cell,* 65**,** 832-847 e4.
NGUYEN, V. Q., CO, C., IRIE, K. & LI, J. J. 2000. Clb/Cdc28 kinases promote nuclear export of the replication initiator proteins Mcm2-7. *Curr Biol,* 10**,** 195-205.
NIEHRS, C. & LUKE, B. 2020. Regulatory R-loops as facilitators of gene expression and genome stability. *Nat Rev Mol Cell Biol,* 21**,** 167-178.
NISHIMURA, K., FUKAGAWA, T., TAKISAWA, H., KAKIMOTO, T. & KANEMAKI, M. 2009. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods,* 6**,** 917-22.
NISHIZAWA, M., KAWASUMI, M., FUJINO, M. & TOH-E, A. 1998. Phosphorylation of sic1, a cyclin-dependent kinase (Cdk) inhibitor, by Cdk including Pho85 kinase is required for its prompt degradation. *Mol Biol Cell,* 9**,** 2393-405.
NOGUCHI, Y., YUAN, Z., BAI, L., SCHNEIDER, S., ZHAO, G., STILLMAN, B., SPECK, C. & LI, H. 2017. Cryo-EM structure of Mcm2-7 double hexamer on DNA suggests a lagging-strand DNA extrusion model. *Proc Natl Acad Sci U S A,* 114**,** E9529-E9538.
OGAMI, K., CHEN, Y. & MANLEY, J. L. 2018. RNA surveillance by the nuclear RNA exosome: mechanisms and significance. *Noncoding RNA,* 4.
OSBORN, A., J. & ELLEDGE, S., J. 2003. Mrc1 is a replication fork component whose phosphorylation in response to DNA replication stress activates Rad53. *Genes and Development,* 17**,** 1755-1767.
OZER, O. & HICKSON, I. D. 2018. Pathways for maintenance of telomeres and common fragile sites during DNA replication stress. *Open Biol,* 8.
OZGUR, S., BUCHWALD, G., FALK, S., CHAKRABARTI, S., PRABU, J. R. & CONTI, E. 2015. The conformational plasticity of eukaryotic RNA-dependent ATPases. *FEBS J,* 282**,** 850-63.
PACIOTTI, V., CLERICI, M., SCOTTI, M., LUCCHINI, G. & LONGHESE, M. P. 2001. Characterization of mec1 kinase-deficient mutants and of new hypomorphic mec1 alleles impairing subsets of the DNA damage response pathway. *Mol Cell Biol,* 21**,** 3913-25.
PARDO, B., CRABBÉ, L. & PASERO, P. 2017. Signaling pathways of replication stress in yeast. *FEMS Yeast Res,* 17.

PAULSEN, R., D., SONI, D., V., WOLLMAN, R., HAHN, A., T., YEE, M., C., GUAN, A., HESLEY, J., A., MILLER, S., C., CROMWELL, E., F., SOLOW-CORDERO, D., E., MEYER, T. & CIMPRICH, K., A. 2009. A genome-wide siRNA screen reveals diverse cellular processes and pathways that mediate genome stability. *Molecular Cell,* 35**,** 228-239.

PAVLOV, Y. I., FRAHM, C., NICK MCELHINNY, S. A., NIIMI, A., SUZUKI, M. & KUNKEL, T. A. 2006. Evidence that errors made by DNA polymerase alpha are corrected by DNA polymerase delta. *Curr Biol,* 16**,** 202-7.

PELLEGRINI, L. 2012. The Pol alpha-primase complex. *Subcell Biochem,* 62**,** 157-69.

PERERA, R., L., TORELLA, R., KLINGE, S., KILKENNY, M., L., MAMAN, J., D. & PELLEGRINI, L. 2013. Mechanism for priming DNA synthesis by yeast DNA Polymerase a. *eLife,* 2**,** e00482.

PETERS, J., M. 2006. The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nature Reviews Molecular Cell Biology,* 7**,** 644-656.

PETRYK, N., KAHLI, M., D'AUBENTON-CARAFA, Y., JASZCZYSZYN, Y., SHEN, Y., SILVAIN, M., THERMES, C., CHEN, C. L. & HYRIEN, O. 2016. Replication landscape of the human genome. *Nat Commun,* 7**,** 10208.

PFANDER, B. & DIFFLEY, J. F. 2011. Dpb11 coordinates Mec1 kinase activation with cell cycle-regulated Rad9 recruitment. *EMBO J,* 30**,** 4897-907.

PFEIFFER, V., CRITTIN, J., GROLIMUND, L. & LINGNER, J. 2013. The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening. *The EMBO Journal,* 32**,** 2861-2871.

PHATNANI, H. P. & GREENLEAF, A. L. 2006. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev,* 20**,** 2922-36.

POHJOISMÄKI, J., L,O., HOLMES, J., B., WOOD, S., R., YANG, M., Y., YASUKAWA, T., REYES, A., LAURA, J., B., CLUETT, T., J., GOFFART, S., WILLCOX, S., RIGBY, R., E., JACKSON, A., P., SPELBRINK, J., N., GRIFFITH, J., D., CROUCH, R., J., JACOBS, H., T. & HOLT, I., J. 2010. Mammalian mitochondrial DNA replication intermediates are essentially duplex, but contain extensive tracts of RNA/DNA hybrid. *Journal of Molecular Biology,* 397**,** 1144-1155.

POLI, J., TSAPONINA, O., CRABBÉ, L., KESZTHELYI, A., PANTESCO, V., CHABES, A., LENGRONNE, A. & PASERO, P. 2012. dNTP pools determine fork progression and origin usage under replication stress. *EMBO J,* 31**,** 883-94.

POMERANTZ, R. T. & O'DONNELL, M. 2010. Direct restart of a replication fork stalled by a head-on RNA polymerase. *Science,* 327**,** 590-2.

PORRUA, O., BOUDVILLAIN, M. & LIBRI, D. 2016. Transcription Termination: Variations on Common Themes. *Trends Genet,* 32**,** 508-522.

PORRUA, O., HOBOR, F., BOULAY, J., KUBICEK, K., D'AUBENTON-CARAFA, Y., GUDIPATI, R. K., STEFL, R. & LIBRI, D. 2012. In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. *EMBO J,* 31**,** 3935-48.

PORRUA, O. & LIBRI, D. 2013. A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nature Structural and Molecular Biology,* 20**,** 884-891.

PORRUA, O. & LIBRI, D. 2015. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol,* 16**,** 190-202.

POTENSKI, C. J. & KLEIN, H. L. 2014. How the misincorporation of ribonucleotides into genomic DNA can be both harmful and helpful to cells. *Nucleic Acids Res,* 42**,** 10226-34.
PRADO, F. & AGUILERA, A. 2005. Impairment of replication fork progression mediates RNA Pol II transcription-associated recombination. *The EMBO Journal,* 24**,** 1267-1276.
PRIOLEAU, M. N. & MACALPINE, D. M. 2016. DNA replication origins-where do we begin? *Genes Dev,* 30**,** 1683-97.
PROUDFOOT, N., J. 2016. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science,* 352**,** aad9926.
PUDDU, F., GRANATA, M., DI NOLA, L., BALESTRINI, A., PIERGIOVANNI, G., LAZZARO, F., GIANNATTASIO, M., PLEVANI, P. & MUZI-FALCONI, M. 2008. Phosphorylation of the budding yeast 9-1-1 complex is required for Dpb11 function in the full activation of the UV-induced DNA damage checkpoint. *Mol Cell Biol,* 28**,** 4782-93.
PURSELL, Z. F., ISOZ, I., LUNDSTROM, E. B., JOHANSSON, E. & KUNKEL, T. A. 2007. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science,* 317**,** 127-30.
PUTNAM, C. D., JAEHNIG, E. J. & KOLODNER, R. D. 2009. Perspectives on the DNA damage and replication checkpoint responses in Saccharomyces cerevisiae. *DNA Repair (Amst),* 8**,** 974-82.
QIU, J., QIAN, Y., FRANK, P., WINTERSBERGER, U. & SHEN, B. 1999. Saccharomyces cerevisiae RNase H(35) functions in RNA primer removal during lagging-strand DNA synthesis, most efficiently in cooperation with Rad27 nuclease. *Mol Cell Biol,* 19**,** 8361-71.
RANDELL, J. C., FAN, A., CHAN, C., FRANCIS, L. I., HELLER, R. C., GALANI, K. & BELL, S. P. 2010. Mec1 is one of multiple kinases that prime the Mcm2-7 helicase for phosphorylation by Cdc7. *Mol Cell,* 40**,** 353-63.
RAWAL, C. C., ZARDONI, L., DI TERLIZZI, M., GALATI, E., BRAMBATI, A., LAZZARO, F., LIBERI, G. & PELLICIOLI, A. 2020. Senataxin Ortholog Sen1 Limits DNA:RNA Hybrid Accumulation at DNA Double-Strand Breaks to Control End Resection and Repair Fidelity. *Cell Rep,* 31**,** 107603.
REMUS, D., BEURON, F., TOLUN, G., GRIFFITH, J., D., MORRIS, E., P. & DIFFLEY, J., F. 2009. Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. *Cell,* 139**,** 719-730.
REMUS, D. & DIFFLEY, J. F. 2009. Eukaryotic DNA replication control: lock and load, then fire. *Curr Opin Cell Biol,* 21**,** 771-7.
RICHARD, P., FENG, S. & MANLEY, J., L. 2013. A SUMO-dependent interaction between Senataxin and the exosome, disrupted in the neurodegenerative disease AOA2, targets the exosome to sites of transcription-induced DNA damage. *Genes and Development,* 27**,** 2227-2232.
RICHARD, P. & MANLEY, J., L. 2009. Transcription termination by nuclear RNA. *Genes and Development,* 23**,** 1247-1269.
RICHARD, P. & MANLEY, J. L. 2014. SETX sumoylation: A link between DNA damage and RNA surveillance disrupted in AOA2. *Rare Dis,* 2**,** e27744.
ROBERTS, R., W. & CROTHERS, D., M. 1992. Stability and properties of double and triple helices: Dramatic effects of RNA or DNA backbone composition. *Science,* 258**,** 1463-1466.
ROBERTSON, P. D., WARREN, E. M., ZHANG, H., FRIEDMAN, D. B., LARY, J. W., COLE, J. L., TUTTER, A. V., WALTER, J. C., FANNING, E. &

EICHMAN, B. F. 2008. Domain architecture and biochemical characterization of vertebrate Mcm10. *J Biol Chem,* 283**,** 3338-48.
RODRIGUEZ, J., LEE, L., LYNCH, B. & TSUKIYAMA, T. 2017. Nucleosome occupancy as a novel chromatin parameter for replication origin functions. *Genome Res,* 27**,** 269-277.
RONDÓN, A., G., MISCHO, H., E., KAWAUCHI, J. & PROUDFOOT, N., J. 2009. Fail-safe transcriptional termination for protein-coding genes in *S. cerevisiae*. *Molecular Cell,* 36**,** 88-98.
ROSSI, S. E., AJAZI, A., CAROTENUTO, W., FOIANI, M. & GIANNATTASIO, M. 2015. Rad53-Mediated Regulation of Rrm3 and Pif1 DNA Helicases Contributes to Prevention of Aberrant Fork Transitions under Replication Stress. *Cell Rep,* 13**,** 80-92.
ROY, D., YU, K. & LIEBER, M., R. 2008. Mechanism of R-loop formation at immunoglobulin class switch sequences. *Molecular and Cellular Biology,* 28**,** 50-60.
ROY, D., ZHANG, Z., LU, Z., HSIEH, C., L. & LIEBER, M., R. 2010. Competition between the RNA transcript and the nontemplate DNA Strand during R-Loop formation i*n vitro:* a nick can serve as a strong R-Loop initiation site. *Molecular and Cellular Biology,* 30**,** 146-159.
RUDNER, A. D. & MURRAY, A. W. 2000. Phosphorylation by Cdc28 activates the Cdc20-dependent activity of the anaphase-promoting complex. *J Cell Biol,* 149**,** 1377-90.
RUIZ-BALLESTEROS, E., MOLLEJO, M., RODRIGUEZ, A., CAMACHO, F. I., ALGARA, P., MARTINEZ, N., POLLAN, M., SANCHEZ-AGUILERA, A., MENARGUEZ, J., CAMPO, E., MARTINEZ, P., MATEO, M. & PIRIS, M. A. 2005. Splenic marginal zone lymphoma: proposal of new diagnostic and prognostic markers identified after tissue and cDNA microarray analysis. *Blood,* 106**,** 1831-8.
SAKAKIBARA, Y. & TOMIZAWA, J., I. 1974. Replication of colicin E1 plasmid DNA in cell extracts. *Proceedings for the National Academy of Sciences of the USA,* 71**,** 802-806.
SALDIVAR, J. C., CORTEZ, D. & CIMPRICH, K. A. 2017. The essential kinase ATR: ensuring faithful duplication of a challenging genome. *Nat Rev Mol Cell Biol,* 18**,** 622-636.
SALISBURY, J., HUTCHISON, K. W. & GRABER, J. H. 2006. A multispecies comparison of the metazoan 3'-processing downstream elements and the CstF-64 RNA recognition motif. *BMC Genomics,* 7**,** 55.
SAMORA, C. P., SAKSOUK, J., GOSWAMI, P., WADE, B. O., SINGLETON, M. R., BATES, P. A., LENGRONNE, A., COSTA, A. & UHLMANN, F. 2016. Ctf4 Links DNA Replication with Sister Chromatid Cohesion Establishment by Recruiting the Chl1 Helicase to the Replisome. *Mol Cell,* 63**,** 371-84.
SCHULZ, D., SCHWALB, B., KIESEL, A., BAEJEN, C., TORKLER, P., GAGNEUR, J., SOEDING, J. & CRAMER, P. 2013. Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell,* 155**,** 1075-87.
SCHURER, K. A., RUDOLPH, C., ULRICH, H. D. & KRAMER, W. 2004. Yeast MPH1 gene functions in an error-free DNA damage bypass pathway that requires genes from Homologous recombination, but not from postreplicative repair. *Genetics,* 166**,** 1673-86.

SCHWOB, E., BOHM, T., MENDENHALL, M. D. & NASMYTH, K. 1994. The B-type cyclin kinase inhibitor p40SIC1 controls the G1 to S transition in S. cerevisiae. *Cell,* 79**,** 233-44.
SENGUPTA, S., VAN DEURSEN, F., DE PICCOLI, G. & LABIB, K. 2013. Dpb2 integrates the leading-strand DNA polymerase into the eukaryotic replisome. *Current Biology,* 23**,** 543-552.
SHAW, N., N. & ARYA, D., P. 2008. Recognition of the unique structure of DNA:RNA hybrids. *Biochimie,* 90**,** 1026-1039.
SHCHERBAKOVA, P. V. & PAVLOV, Y. I. 1996. 3'-->5' exonucleases of DNA polymerases epsilon and delta correct base analog induced DNA replication errors on opposite DNA strands in Saccharomyces cerevisiae. *Genetics,* 142**,** 717-26.
SHEU, Y. J. & STILLMAN, B. 2006. Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. *Mol Cell,* 24**,** 101-13.
SHEU, Y. J. & STILLMAN, B. 2010. The Dbf4-Cdc7 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. *Nature,* 463**,** 113-7.
SHIMADA, K., PASERO, P. & GASSER, S. M. 2002. ORC and the intra-S-phase checkpoint: a threshold regulates Rad53p activation in S phase. *Genes Dev,* 16**,** 3236-52.
SHIRAYAMA, M., TOTH, A., GALOVA, M. & NASMYTH, K. 1999. APC(Cdc20) promotes exit from mitosis by destroying the anaphase inhibitor Pds1 and cyclin Clb5. *Nature,* 402**,** 203-7.
SIMON, A. C., ZHOU, J. C., PERERA, R. L., VAN DEURSEN, F., EVRIN, C., IVANOVA, M. E., KILKENNY, M. L., RENAULT, L., KJAER, S., MATAK-VINKOVIC, D., LABIB, K., COSTA, A. & PELLEGRINI, L. 2014. A Ctf4 trimer couples the CMG helicase to DNA polymerase alpha in the eukaryotic replisome. *Nature,* 510**,** 293-297.
SINGH, A. & XU, Y. J. 2016. The Cell Killing Mechanisms of Hydroxyurea. *Genes (Basel),* 7.
SKOURTI-STATHAKI, K. & PROUDFOOT, N. J. 2014. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev,* 28**,** 1384-96.
SKOURTI-STATHAKI, K., PROUDFOOT, N. J. & GROMAK, N. 2011. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell,* 42**,** 794-805.
SKOWYRA, D., CRAIG, K. L., TYERS, M., ELLEDGE, S. J. & HARPER, J. W. 1997. F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell,* 91**,** 209-19.
SMOLKA, M. B., ALBUQUERQUE, C. P., CHEN, S. H. & ZHOU, H. 2007. Proteome-wide identification of in vivo targets of DNA damage checkpoint kinases. *Proc Natl Acad Sci U S A,* 104**,** 10364-9.
SMOLKA, M. B., CHEN, S. H., MADDOX, P. S., ENSERINK, J. M., ALBUQUERQUE, C. P., WEI, X. X., DESAI, A., KOLODNER, R. D. & ZHOU, H. 2006. An FHA domain-mediated protein interaction network of Rad53 reveals its role in polarized cell growth. *J Cell Biol,* 175**,** 743-53.
SOLLIER, J., STORK, C., T., GARCÍA-RUBIO, M., L., PAULSEN, R., D., AGUILERA, A. & CIMPRICH, K., A. 2014. Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Molecular Cell,* 56**,** 777-785.

SRIVATSAN, A., TEHRANCHI, A., MACALPINE, D. M. & WANG, J. D. 2010. Co-orientation of replication and transcription preserves genome integrity. *PLoS Genet,* 6**,** e1000810.
STEINMETZ, E. J., WARREN, C. L., KUEHNER, J. N., PANBEHI, B., ANSARI, A. Z. & BROW, D. A. 2006. Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol Cell,* 24**,** 735-746.
STOKES, K., WINCZURA, A., SONG, B., PICCOLI, G. & GRABARCZYK, D. B. 2020. Ctf18-RFC and DNA Pol form a stable leading strand polymerase/clamp loader complex required for normal and perturbed DNA replication. *Nucleic Acids Res*.
STUCKEY, R., GARCÍA-RODRÍGUEZ, N., AGUILERA, A. & WELLINGER, R., E. 2015. Role for RNA:DNA hybrids in origin-independent replication priming in a eukaryotic system. *Proceedings for the National Academy of Sciences of the USA,* 112**,** 5779-5784.
SUGIYAMA, T. & KOWALCZYKOWSKI, S. C. 2002. Rad52 protein associates with replication protein A (RPA)-single-stranded DNA to accelerate Rad51-mediated displacement of RPA and presynaptic complex formation. *J Biol Chem,* 277**,** 31663-72.
SUGIYAMA, T., ZAITSEVA, E. M. & KOWALCZYKOWSKI, S. C. 1997. A single-stranded DNA-binding protein is needed for efficient presynaptic complex formation by the Saccharomyces cerevisiae Rad51 protein. *J Biol Chem,* 272**,** 7940-5.
SUN, J., SHI, Y., GEORGESCU, R. E., YUAN, Z., CHAIT, B. T., LI, H. & O'DONNELL, M. E. 2015. The architecture of a eukaryotic replisome. *Nat Struct Mol Biol,* 22**,** 976-82.
SWANEY, D., L., BELTRAO, P., STARITA, L., GUO, A., RUSH, J., FIELDS, S., KROGAN, N., J. & VILLÉN, J. 2013. Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. *Nature Methods,* 10**,** 676-682.
SWEENEY, F. D., YANG, F., CHI, A., SHABANOWITZ, J., HUNT, D. F. & DUROCHER, D. 2005. Saccharomyces cerevisiae Rad9 acts as a Mec1 adaptor to allow Rad53 activation. *Curr Biol,* 15**,** 1364-75.
SZYJKA, S. J., VIGGIANI, C. J. & APARICIO, O. M. 2005. Mrc1 is required for normal progression of replication forks throughout chromatin in S. cerevisiae. *Mol Cell,* 19**,** 691-7.
TAHIROV, T. H., MAKAROVA, K. S., ROGOZIN, I. B., PAVLOV, Y. I. & KOONIN, E. V. 2009. Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol epsilon and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors. *Biol Direct,* 4**,** 11.
TAKEUCHI, Y., HORIUCHI, T. & KOBAYASHI, T. 2003. Transcription-dependent recombination and the role of fork collision in yeast rDNA. *Genes and Development,* 17**,** 1497-1506.
TAN-WONG, S. M., DHIR, S. & PROUDFOOT, N. J. 2019. R-Loops Promote Antisense Transcription across the Mammalian Genome. *Mol Cell,* 76**,** 600-616 e6.
TAN-WONG, S. M., ZAUGG, J. B., CAMBLONG, J., XU, Z., ZHANG, D. W., MISCHO, H. E., ANSARI, A. Z., LUSCOMBE, N. M., STEINMETZ, L. M. & PROUDFOOT, N. J. 2012. Gene loops enhance transcriptional directionality. *Science,* 338**,** 671-5.

TANAKA, H., KATOU, Y., YAGURA, M., SAITOH, K., ITOH, T., ARAKI, H., BANDO, M. & SHIRAHIGE, K. 2009. Ctf4 coordinates the progression of helicase and DNA polymerase α. *Genes to Cells,* 14**,** 807-820.
TANAKA, K. & RUSSELL, P. 2001. Mrc1 channels the DNA replication arrest signal to checkpoint kinase Cds1. *Nature Cell Biology,* 3**,** 966-972.
TANAKA, T., UMEMORI, T., ENDO, S., MURAMATSU, S., KANEMAKI, M., KAMIMURA, Y., OBUSE, C. & ARAKI, H. 2011. Sld7, an Sld3-associated protein required for efficient chromosomal DNA replication in budding yeast. *EMBO J,* 30**,** 2019-30.
TERCERO, J., A. & DIFFLEY, J., F,X. 2001. Regulation of DNA replication fork progression through damaged DNA by the Mec1/Rad53 checkpoint. *Nature,* 412**,** 553-557.
THIEBAUT, M., KISSELEVA-ROMANOVA, E., ROUGEMAILLE, M., BOULAY, J. & LIBRI, D. 2006. Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the Nrd1-Nab3 pathway in genome surveillance. *Molecular Cell,* 23**,** 853-864.
THORNTON, B. R. & TOCZYSKI, D. P. 2003. Securin and B-cyclin/CDK are the only essential targets of the APC. *Nat Cell Biol,* 5**,** 1090-4.
TITTEL-ELMER, M., LENGRONNE, A., DAVIDSON, M. B., BACAL, J., FRANÇOIS, P., HOHL, M., PETRINI, J. H. J., PASERO, P. & COBB, J. A. 2012. Cohesin association to replication sites depends on rad50 and promotes fork restart. *Mol Cell,* 48**,** 98-108.
TOMIZAWA, J., I. 1975. Two distinct mechanisms of synthesis of DNA fragments on colicon E1 plasmid DNA. *Nature,* 257**,** 253-254.
TOUBIANA, S. & SELIG, S. 2018. DNA:RNA hybrids at telomeres - when it is better to be out of the (R) loop. *FEBS J,* 285**,** 2552-2566.
TOURRIÈRE, H., VERSINI, G., CORDÓN-PRECIADO, V., ALABERT, C. & PASERO, P. 2005. Mrc1 and Tof1 promote replication fork progression and recovery independently of Rad53. *Mol Cell,* 19**,** 699-706.
TRAN, P., L,T., POHL, T., J ., CHEN, C., F., CHAN, A., POTT, S. & ZAKIAN, V., A. 2017. PIF1 family DNA helicases suppress R-loop mediated genome instability at tRNA genes. *Nature Communications,* 8**,** 15025.
TRAVEN, A. & HEIERHORST, J. 2005. SQ/TQ cluster domains: concentrated ATM/ATR kinase phosphorylation site regions in DNA-damage-response proteins. *Bioessays,* 27**,** 397-407.
TRAVESA, A., KUO, D., DE BRUIN, R. A., KALASHNIKOVA, T. I., GUADERRAMA, M., THAI, K., ASLANIAN, A., SMOLKA, M. B., YATES, J. R., 3RD, IDEKER, T. & WITTENBERG, C. 2012. DNA replication stress differentially regulates G1/S genes via Rad53-dependent inactivation of Nrm1. *EMBO J,* 31**,** 1811-22.
TUDEK, A., PORRUA, O., KABZINSKI, T., LIDSCHREIBER, M., KUBICEK, K., FORTOVA, A., LACROUTE, F., VANACOVA, S., CRAMER, P., STEFL, R. & LIBRI, D. 2014. Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Mol Cell,* 55**,** 467-81.
TUDURI, S., CRABBE, L., CONTI, C., TOURRIERE, H., HOLTGREVE-GREZ, H., JAUCH, A., PANTESCO, V., DE VOS, J., THOMAS, A., THEILLET, C., POMMIER, Y., TAZI, J., COQUELLE, A. & PASERO, P. 2009. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat Cell Biol,* 11**,** 1315-24.

TYERS, M., TOKIWA, G. & FUTCHER, B. 1993. Comparison of the *Saccharomyces cerevisiae* G1 cyclins: Cln3 may be an upstream activator of Cln1, Cln2 and other cyclins. *The EMBO Journal,* 12**,** 1955-1968.

URSIC, D., CHINCHILLA, K., FINKEL, J., S. & CULBERTSON, M., R. 2004. Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. *Nucleic Acids Research,* 32**,** 2441-2452.

USUI, T., OGAWA, H. & PETRINI, J. H. 2001. A DNA damage response pathway controlled by Tel1 and the Mre11 complex. *Mol Cell,* 7**,** 1255-66.

VAN DEURSEN, F., SENGUPTA, S., DE PICCOLI, G., SANCHEZ-DIAZ, A. & LABIB, K. 2012. Mcm10 associates with the loaded DNA helicase at replication origins and defines a novel step in its activation. *The EMBO Journal,* 31**,** 2195-2206.

VASILJEVA, L., KIM, M., MUTSCHLER, H., BURATOWSKI, S. & MEINHART, A. 2008. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nature Structural and Molecular Biology,* 15**,** 795-804.

VIALARD, J. E., GILBERT, C. S., GREEN, C. M. & LOWNDES, N. F. 1998. The budding yeast Rad9 checkpoint protein is subjected to Mec1/Tel1-dependent hyperphosphorylation and interacts with Rad53 after DNA damage. *EMBO J,* 17**,** 5679-88.

VIGGIANI, C. J., KNOTT, S. R. & APARICIO, O. M. 2010. Genome-wide analysis of DNA synthesis by BrdU immunoprecipitation on tiling microarrays (BrdU-IP-chip) in Saccharomyces cerevisiae. *Cold Spring Harb Protoc,* 2010**,** pdb prot5385.

VILLA, F., SIMON , A., C., BAZAN, M., A,O., KILKENNY, M., L., WIRTHENSOHN, D., WIGHTMAN, M., MATAK-VINKOVÍC, D., PELLEGRINI, L. & LABIB, K. 2016. Ctf4 is a hub in the eukaryotic replisome that links multiple CIP-box proteins to the CMG helicase. *Molecular Cell,* 63**,** 385-396.

WAHBA, L., AMON, J. D., KOSHLAND, D. & VUICA-ROSS, M. 2011. RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. *Mol Cell,* 44**,** 978-88.

WAHBA, L., COSTANTINO, L., TAN, F. J., ZIMMER, A. & KOSHLAND, D. 2016. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev,* 30**,** 1327-38.

WAHBA, L., GORE, S., K. & KOSHLAND, D. 2013. The homologous recombination machinery modulates the formation of RNA-DNA hybrids and associated chromosome instability. *eLife,* 2**,** e00505.

WANROOIJ, P. H. & BURGERS, P. M. 2015. Yet another job for Dna2: Checkpoint activation. *DNA Repair (Amst),* 32**,** 17-23.

WANROOIJ, P. H., TANNOUS, E., KUMAR, S., NAVADGI-PATIL, V. M. & BURGERS, P. M. 2016. Probing the Mec1ATR Checkpoint Activation Mechanism with Small Peptides. *J Biol Chem,* 291**,** 393-401.

WASCH, R. & CROSS, F. R. 2002. APC-dependent proteolysis of the mitotic cyclin Clb2 is essential for mitotic exit. *Nature,* 418**,** 556-62.

WATASE, G., TAKISAWA, H. & KANEMAKI, M., T. 2012. Mcm10 plays a role in functioning of the eukaryotic replicative DNA helicase, Cdc45-Mcm-GINS. *Current Biology,* 22**,** 343-349.

WEST, S., GROMAK, N. & PROUDFOOT, N., J. 2004. Human 5' → 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature,* 432**,** 522-525.
WESTOVER, K., D., BUSHNELL, D., A. & KORNBERG, R., D. 2004. Structural basis of transcription : nucleotide selection by rotation in the RNA Polymerase II active center. *Cell,* 119**,** 481-489.
WHITEHOUSE, I., RANDO, O. J., DELROW, J. & TSUKIYAMA, T. 2007. Chromatin remodelling at promoters suppresses antisense transcription. *Nature,* 450**,** 1031-5.
WILLIAMS, D., R. & MCINTOSH, J., R . 2002. m*cl1*+, the Schizosaccharomyces pombe homologue of C*TF4,* is important for chromosome replication, cohesion, and segregation. *Eukaryotic Cell,* 1**,** 758-773.
WINEY, M. & CULBERTSON, M. R. 1988. Mutations affecting the tRNA-splicing endonuclease activity of Saccharomyces cerevisiae. *Genetics,* 118**,** 609-17.
WYBENGA-GROOT, L. E., HO, C. S., SWEENEY, F. D., CECCARELLI, D. F., MCGLADE, C. J., DUROCHER, D. & SICHERI, F. 2014. Structural basis of Rad53 kinase activation by dimerization and activation segment exchange. *Cell Signal,* 26**,** 1825-36.
XIAO, W., CHOW, B. L. & RATHGEBER, L. 1996. The repair of DNA methylation damage in Saccharomyces cerevisiae. *Curr Genet,* 30**,** 461-8.
XU, B. & CLAYTON, D., A. 1996. RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: an implication for RNA-DNA hybrids serving as primers. *The EMBO Journal,* 15**,** 3135-3143.
XU, Z., WEI, W., GAGNEUR, J., PEROCCHI, F., CLAUDER-MUNSTER, S., CAMBLONG, J., GUFFANTI, E., STUTZ, F., HUBER, W. & STEINMETZ, L. M. 2009. Bidirectional promoters generate pervasive transcription in yeast. *Nature,* 457**,** 1033-7.
YEELES, J., T., JANSKA, A., EARLY, A. & DIFFLEY, J., F. 2017. How the eukaryotic replisome achieves rapid and efficient DNA replication. *Molecular Cell,* 65**,** 105-116.
YEELES, J. T., DEEGAN, T. D., JANSKA, A., EARLY, A. & DIFFLEY, J. F. 2015. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature,* 519**,** 431-5.
YEO, A., J ., BECHERE, L. O., J., LUFF, J., E ., CULLEN, J., .K ., WONGSURAWAT, T., JENJAROENPUN, P., KUZNETSOV, V., A ., MCKINNON, P., J . & LAVIN, M., F . 2014. R-loops in proliferating cells but not in the brain: implications for AOA2 and other autosomal recessive ataxias. *PLoS ONE,* 9**,** e90219.
YU, C., GAN, H., HAN, J., ZHOU, Z. X., JIA, S., CHABES, A., FARRUGIA, G., ORDOG, T. & ZHANG, Z. 2014. Strand-specific analysis shows protein binding at replication forks and PCNA unloading from lagging strands when forks stall. *Mol Cell,* 56**,** 551-63.
YUAN, Z., GEORGESCU, R., SCHAUER, G. D., O'DONNELL, M. E. & LI, H. 2020. Structure of the polymerase epsilon holoenzyme and atomic model of the leading strand replisome. *Nat Commun,* 11**,** 3156.
YÜCE, Ö. & WEST, S., C. 2013. Senataxin, defective in the neurodegenerative disorder Ataxia with Oculomotor Apraxia 2, lies at the interface of transcription and the DNA damage response. *Molecular and Cellular Biology,* 33**,** 406 -417.

ZANTON, S. J. & PUGH, B. F. 2006. Full and partial genome-wide assembly and disassembly of the yeast transcription machinery in response to heat shock. *Genes Dev,* 20**,** 2250-65.
ZEGERMAN, P. & DIFFLEY, J., F. 2010. Checkpoint-dependent inhibition of DNA replication initiation by Sld3 and Dbf4 phosphorylation. *Nature,* 467**,** 474-478.
ZHAO, Q., KIRKNESS, E. F., CABALLERO, O. L., GALANTE, P. A., PARMIGIANI, R. B., EDSALL, L., KUAN, S., YE, Z., LEVY, S., VASCONCELOS, A. T., REN, B., DE SOUZA, S. J., CAMARGO, A. A., SIMPSON, A. J. & STRAUSBERG, R. L. 2010. Systematic detection of putative tumor suppressor genes through the combined use of exome and transcriptome sequencing. *Genome Biol,* 11**,** R114.
ZHAO, X., CHABES, A., DOMKIN, V., THELANDER, L. & ROTHSTEIN, R. 2001. The ribonucleotide reductase inhibitor Sml1 is a new target of the Mec1/Rad53 kinase cascade during growth and in response to DNA damage. *EMBO J,* 20**,** 3544-53.
ZHAO, X., MULLER, E. G. & ROTHSTEIN, R. 1998. A suppressor of two essential checkpoint genes identifies a novel protein that negatively affects dNTP pools. *Mol Cell,* 2**,** 329-40.
ZHONG, Y., NELLIMOOTTIL, T., PEACE, J. M., KNOTT, S. R., VILLWOCK, S. K., YEE, J. M., JANCUSKA, J. M., REGE, S., TECKLENBURG, M., SCLAFANI, R. A., TAVARE, S. & APARICIO, O. M. 2013. The level of origin firing inversely affects the rate of replication fork progression. *J Cell Biol,* 201**,** 373-83.
ZHOU, C., ELIA, A. E., NAYLOR, M. L., DEPHOURE, N., BALLIF, B. A., GOEL, G., XU, Q., NG, A., CHOU, D. M., XAVIER, R. J., GYGI, S. P. & ELLEDGE, S. J. 2016. Profiling DNA damage-induced phosphorylation in budding yeast reveals diverse signaling networks. *Proc Natl Acad Sci U S A,* 113**,** E3667-75.
ZHOU, Z. X., LUJAN, S. A., BURKHOLDER, A. B., GARBACZ, M. A. & KUNKEL, T. A. 2019. Roles for DNA polymerase δ in initiating and terminating leading strand DNA replication. *Nat Commun,* 10**,** 3992.
ZHU, W., UKOMADU, C., JHA, S., SENGA, T., DHAR, S., K., WOHLSCHLEGEL, J., A., NUTT, L., K., KORNBLUTH, S. & DUTTA, A. 2007. Mcm10 and And-1/CTF4 recruit DNA polymerase alpha to chromatin for initiation of DNA replication. *Genes and Development,* 21**,** 2288-2299.
ZHU, Z., CHUNG, W. H., SHIM, E. Y., LEE, S. E. & IRA, G. 2008. Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. *Cell,* 134**,** 981-94.
ZIMMERMANN, M., MURINA, O., REIJNS, M. A. M., AGATHANGGELOU, A., CHALLIS, R., TARNAUSKAITE, Z., MUIR, M., FLUTEAU, A., AREGGER, M., MCEWAN, A., YUAN, W., CLARKE, M., LAMBROS, M. B., PANEESHA, S., MOSS, P., CHANDRASHEKHAR, M., ANGERS, S., MOFFAT, J., BRUNTON, V. G., HART, T., DE BONO, J., STANKOVIC, T., JACKSON, A. P. & DUROCHER, D. 2018. CRISPR screens identify genomic ribonucleotides as a source of PARP-trapping lesions. *Nature,* 559**,** 285-289.
ZLOTKIN, T., KAUFMANN, G., JIANG, Y., LEE, M. Y., UITTO, L., SYVAOJA, J., DORNREITER, I., FANNING, E. & NETHANEL, T. 1996. DNA

polymerase epsilon may be dispensable for SV40- but not cellular-DNA replication. *EMBO J,* 15**,** 2298-305.
ZOU, L. & ELLEDGE, S. J. 2003. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science,* 300**,** 1542-8.

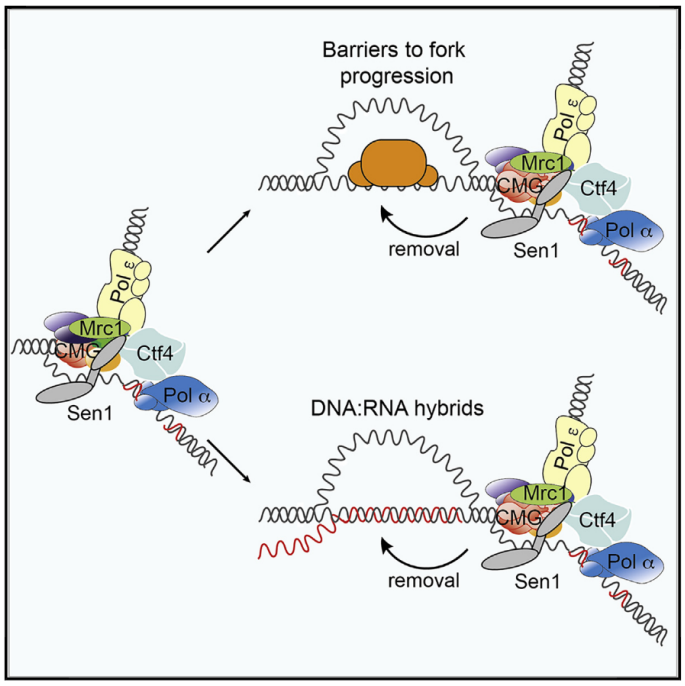
Appendix 1

Report

Cell Reports

Sen1 Is Recruited to Replication Forks via Ctf4 and Mrc1 and Promotes Genome Stability

Graphical Abstract



Authors

Rowin Appanah, Emma Claire Lones, Umberto Aiello, Domenico Libri, Giacomo De Piccoli

Correspondence

g.de-piccoli@warwick.ac.uk

In Brief

Appanah et al. identify the transcription termination helicase Sen1 as a bona fide component of the replisome. Sen1 binds the replisome via its N-terminal domain and Ctf4 and Mrc1. The allele *sen1-3* breaks this interaction without affecting transcription termination. *sen1-3* cells show sensitivity to R-loops levels and increased genomic instability.

Highlights

- The N-terminal domain of Sen1 mediates replisome association
- Ctf4 and Mrc1 bind to Sen1 and promote its recruitment to the replisome
- The *sen1-3* allele abrogates replisome binding, but not transcription termination
- *sen1-3* cells are sensitive to high levels of R-loops and defects of S phase checkpoint

Appanah et al., 2020, Cell Reports *30*, 2094–2105
February 18, 2020
https://doi.org/10.1016/j.celrep.2020.01.087

CellPress

OPEN ACCESS
CellPress

Cell Reports
Report

Sen1 Is Recruited to Replication Forks via Ctf4 and Mrc1 and Promotes Genome Stability

Rowin Appanah,[1] Emma Claire Lones,[1] Umberto Aiello,[2] Domenico Libri,[2] and Giacomo De Piccoli[1,3,*]
[1]Warwick Medical School, University of Warwick, CV4 7AL Coventry, UK
[2]Institut Jacques Monod, CNRS, UMR7592, Université Paris Diderot, Paris Sorbonne Cité, Paris, France
[3]Lead Contact
*Correspondence: g.de-piccoli@warwick.ac.uk
https://doi.org/10.1016/j.celrep.2020.01.087

SUMMARY

DNA replication and RNA transcription compete for the same substrate during S phase. Cells have evolved several mechanisms to minimize such conflicts. Here, we identify the mechanism by which the transcription termination helicase Sen1 associates with replisomes. We show that the N terminus of Sen1 is both sufficient and necessary for replisome association and that it binds to the replisome via the components Ctf4 and Mrc1. We generated a separation of function mutant, *sen1-3*, which abolishes replisome binding without affecting transcription termination. We observe that the *sen1-3* mutants show increased genome instability and recombination levels. Moreover, *sen1-3* is synthetically defective with mutations in genes involved in RNA metabolism and the S phase checkpoint. *RNH1* overexpression suppresses defects in the former, but not the latter. These findings illustrate how Sen1 plays a key function at replication forks during DNA replication to promote fork progression and chromosome stability.

INTRODUCTION

The maintenance of genome stability requires the complete and faithful duplication of DNA in every cell cycle. Yet several obstacles impede the progression of replication forks (RFs), and these must be removed to avoid stalling and increased chromosome instability. A significant barrier to RF progression is transcription. First identified in bacteria, collisions between RFs and transcription bubbles also represent a major obstacle for DNA synthesis in eukaryotes, leading to defects in chromosome maintenance and an increase in levels of recombination (Liu and Alberts, 1995; Helmrich et al., 2011, 2013; Prado and Aguilera, 2005; Kim et al., 2010; Hamperl et al., 2017; Tran et al., 2017). In order to complete the full duplication of the chromosomes, replisomes must therefore overcome transcriptional barriers, removing both the DNA-bound RNA polymerase subunits and any DNA:RNA hybrids formed during transcription. These hybrids, usually limited to eight base pairs, occur naturally during RNA transcription and are typically removed when the RNA polymerase is disengaged from the DNA (Aguilera and García-Muse, 2012; Westover et al., 2004).

At specific chromosomal loci, extended DNA:RNA hybrids can also form behind the site of RNA synthesis, through the re-annealing of nascent RNA to the template DNA and the displacement of the non-template DNA. These structures, named R-loops, form preferentially at highly transcribed genes with a high GC skew and can extend up to 1 kb in higher eukaryotes (Aguilera and García-Muse, 2012; Skourti-Stathaki et al., 2014). Formation of R-loops is favored by head-on collisions between RFs and actively transcribing complexes (Hamperl et al., 2017; Lang et al., 2017), and their non-physiological accumulation, coupled to chromatin modification, is deleterious for genome stability (García-Pichardo et al., 2017). Several pathways minimize the formation and stability of R-loops. For instance, the promotion of transcription processivity (Hazelbaker et al., 2013), transcription termination (Kim et al., 2004; Luke et al., 2008), timely processing, export or degradation of nascent mRNA (Huertas and Aguilera, 2003; Pfeiffer et al., 2013), or preventing torsional stress that arises during transcription (El Hage et al., 2010, 2014) all minimize R-loops' levels. Nevertheless, once formed, R-loops must be removed. A key role in R-loop removal is fulfilled by the RNase H enzymes that specifically digest RNA molecules within DNA:RNA hybrids (Cerritelli and Crouch, 2009). In addition, several helicases can unwind DNA:RNA hybrids *in vitro*, including Sgs1 (Chang et al., 2017) and Pif1 (Boulé and Zakian, 2007). One such helicase, Sen1, is believed to play an essential role in the removal of R-loops from the DNA in yeast (Mischo et al., 2011).

Sen1 is an Upf1-like helicase that plays a key role in transcription termination (Jankowsky, 2011; Steinmetz et al., 2006; Ursic et al., 1997; Porrua and Libri, 2013). Sen1 binds to the free 5′ ends of either RNA or DNA substrates and unwind both double-stranded DNA (dsDNA) and DNA:RNA hybrids (Han et al., 2017; Leonaitė et al., 2017; Martin-Tumasz and Brow, 2015; Porrua and Libri, 2013). *In vitro* analysis shows that Sen1 has high activity but limited processivity on DNA:RNA hybrid substrates (Han et al., 2017). Mechanistically, when Sen1 engages with nascent RNA exiting from a stalled RNA polymerase II (RNAPII), the helicase seemingly exerts a force on the polymerase to "push" it, either overcoming the stalling of RNAPII or disengaging it from the template DNA (Porrua and Libri, 2013; Han et al., 2017). *In vivo* data also suggest that Sen1 is capable of removing RNAPII from the DNA it is bound to, thus terminating transcription (Steinmetz et al., 2006; Schaughency et al., 2014; Hazelbaker et al., 2013). In fact, a mutation in the catalytic domain of Sen1 (*sen1-1*) confers defects in transcription termination at non-permissive temperatures, leading to extensive readthrough

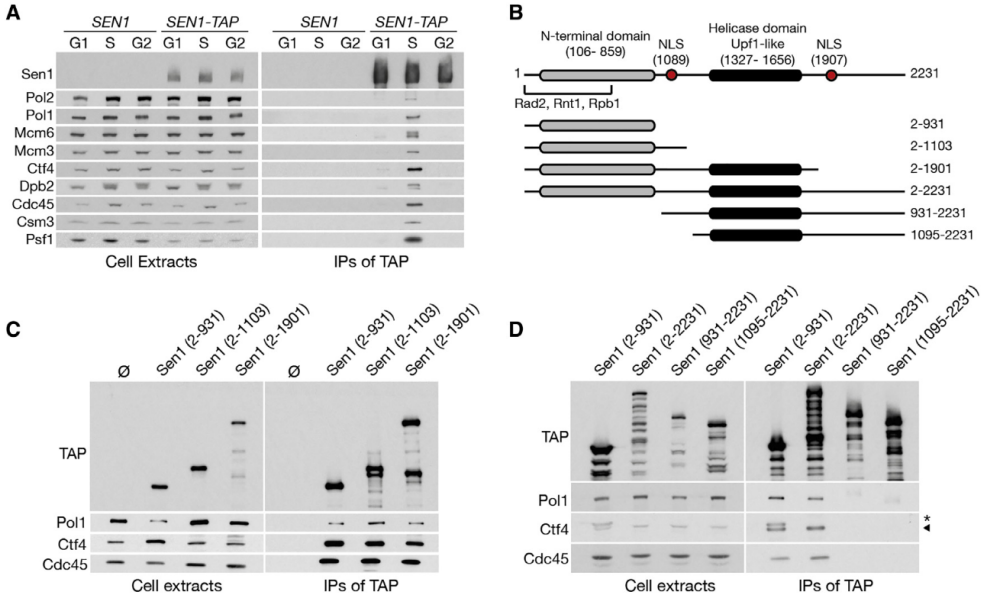


Figure 1. Sen1 Interacts with the Replisome during S Phase through Its N-Terminal Domain

(A) *SEN1* or *SEN1-TAP* cells were arrested in G1, harvested immediately, or released for either 30 min (S phase) or 60 min (G2 phase). Cell extracts and IP material were analyzed by immunoblotting (IB).

(B) Schematic of Sen1 constructs used.

(C) TAP-tagged fragments of Sen1, IPed from cells in S phase, were analyzed by IB.

(D) TAP-tagged fragments of Sen1 were analysed as above, except 4× cells were used for the IP of the fragments containing the last 330 C-terminal amino acids.

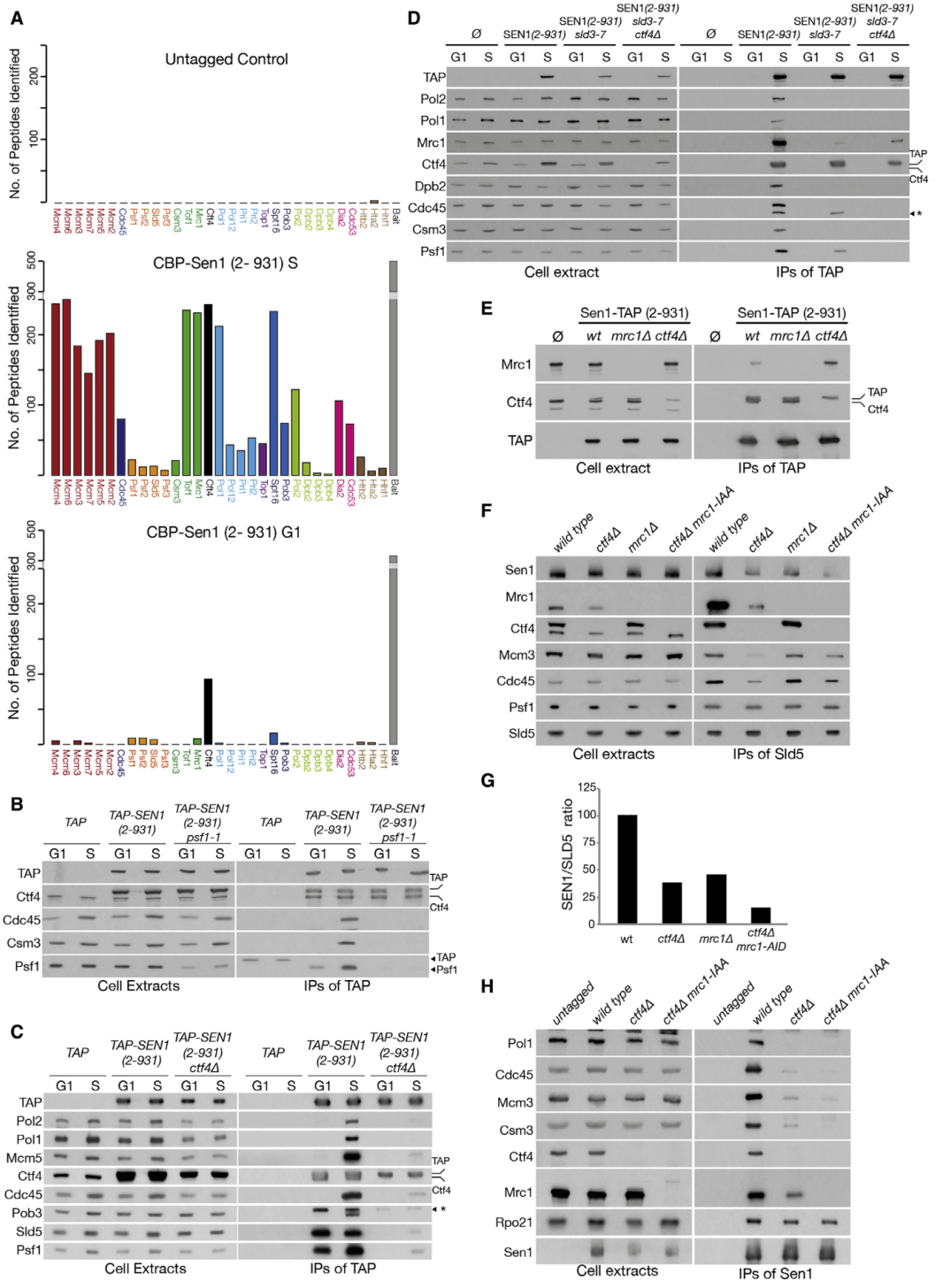
of several transcription units (Steinmetz et al., 2006), accumulation of R-loops, and increased recombination (Mischo et al., 2011). Because of these defects, the viability of *sen1-1* cells depends on several repair factors (Mischo et al., 2011; Alzu et al., 2012). Moreover, depletion of Sen1 leads to slow DNA replication and the accumulation of abnormal structures on 2D gels (Alzu et al., 2012; Brambati et al., 2018).

Given its relatively low abundance and processivity (Mischo et al., 2018; Han et al., 2017), Sen1 needs to be recruited at, or close to, sites where it can enact its biological function. Sen1 is recruited to the termination sites of cryptic-unstable transcripts (CUTs) and small nucleolar RNAs (snoRNAs) by binding to Nab3 and Nrd1, which both dock onto nascent RNA (Arigo et al., 2006; Porrua et al., 2012; Creamer et al., 2011). Nrd1 also interacts with $Rpo21^{Rpb1}$ (the largest subunit of RNAPII) early in the transcription cycle (Vasiljeva et al., 2008), thus restricting Sen1-dependent termination to short transcription units (Gudipati et al., 2008). Sen1 also promotes termination of some genes downstream of the polyadenylation site, acting with Rat1 (Mischo et al., 2011; Rondón et al., 2009), possibly by directly binding Rpo21 via its N-terminal domain (Chinchilla et al., 2012). Finally, it is likely that Sen1 is recruited at other genomic sites in a transcription-independent fashion. The human ortholog of Sen1 (Senataxin) co-localizes with 53BP1 to sites of DNA damage in a checkpoint-dependent manner (Yüce and West, 2013). Moreover, in *S. cerevisiae*, Sen1 co-localizes with replisome components and sites of bromodeoxyuridine (BrdU) incorporation (Alzu et al., 2012). However, the mechanism through which Sen1 is recruited at RFs has yet to be described. The significance of recruiting Sen1 to RFs is also poorly understood, as it has been impossible thus far to determine whether the defects in DNA replication upon inactivation of Sen1 are an indirect consequence of deregulated transcription termination, of a failure in R-loop removal, or the direct result of an important function of Sen1 at RFs. Here, we show that Sen1 binds the replisome during S phase through its N-terminal domain, map its binding site, generate a mutant that breaks this interaction, and explore the consequences of the loss of the helicase from RFs on chromosome stability.

RESULTS

Sen1 Interacts with the Replisome via Its N-Terminal Domain

The replisome is a complex and dynamic machine that relies on multiple interactions between its constituent proteins (Bell and Labib, 2016; Burgers and Kunkel, 2017). As part of a mass spectrometry (MS) screen to identify factors transiently or weakly associated to the core replisome, we observed that Sen1 co-purifies with the CMG helicase in *S. cerevisiae* (Figure S1A). To verify the MS data, we immunoprecipitated (IPed) Sen1 from extracts of yeast cells synchronized in G1, S, and G2. We observed that Sen1 interacted with replisome components only in S phase (Figure 1A). Immunoprecipitation (IP) of the GINS component



(legend on next page)

Sld5 corroborated this observation (Figure S1B). Sen1 interacts with replisomes independently of either Nrd1 or Nab3 (Figures S1C and S1D) and independently of ongoing transcription (Figures S1E and S1F), as previously observed (Alzu et al., 2012). To further explore this interaction and its biological function, we mapped the interaction sites both in the replisome and Sen1.

Sen1 contains an extended N-terminal domain and an essential and conserved helicase domain (Leonaitė et al., 2017). To identify a region of Sen1 that is sufficient for binding replisomes, we generated TAP-tagged constructs of Sen1, expressed under an inducible *GAL1* promoter (Figure 1B). All fragments containing the helicase domain folded correctly and rescued *sen1-1* lethality at non-permissive temperatures, despite constructs containing the last 330 amino acids of the protein being highly labile (Figures S1G and S1H). We then assessed the ability of the various fragments to interact with the replisome and observed that the N-terminal domain (residues 2–931) of Sen1 was both sufficient and necessary for association with replisomes (Figures 1C and 1D). Similarly, Sen1 (2–931) co-precipitated specifically with replisomes isolated from S phase cells by IP of Mcm3 (a subunit of the CMG helicase) (Figures S1I and S1J). Thus, Sen1 (2–931) contains an interaction site for replisome components.

Sen1 Binding to the Replisome Depends on Ctf4 and Mrc1

To identify specific proteins to which Sen1 binds within the replisome, we compared the G1 and S phase interactome of Sen1 (2–931) via MS analysis. As expected, Sen1 (2–931) IPed with replisomes in S phase (Figure 2A). Interestingly, Ctf4 and GINS co-purified with the bait in G1 as well. This was confirmed by immunoblotting (Figures 2B and 2C). Because Ctf4 and GINS interacts throughout the cell cycle (Gambus et al., 2009), we next analyzed whether Sen1 binds preferentially to one of the components. The interaction between Ctf4 and Sen1 in G1 was unaffected by inactivating GINS via the *psf1-1* allele (Figure 2B; Takayama et al., 2003), but GINS no longer IPed with Sen1 (2–931) in G1 in the absence of Ctf4 (Figure 2C). These data indicate that Sen1 (2–931) binds to Ctf4 in the absence of other replisome components.

Interestingly, Sen1 (2–931) retained some affinity to the replisome in the absence of Ctf4 (Figure 2C, right panel), independently of DNA (Figure S2A). This suggests that Sen1 interacts with at least another subunit of the replisome. To screen for such factors, we analyzed whether any component of the replisome binds to Sen1 (2–931) in cells progressing into S phase in the absence of origin firing. We used *td-sld3-7* cells that cannot initiate chromosome replication at 37°C following inactivation and degradation of td-Sld3-7 (Kamimura et al., 2001; Kanemaki and Labib, 2006; Figure S2B). In control cells, Sen1 (2–931) co-purified with all tested replisome components in S phase (Figure 2D). In *td-sld3-7* cells, Sen1 IPed predominantly with Ctf4 and GINS but also weakly with the replisome component Mrc1. Strikingly, Sen1 (2–931)'s affinity for Mrc1 increased in a td-*sld3-7 ctf4Δ* background. We confirmed this in cells arrested in G1 as well (Figure 2E). These observations suggest that both Ctf4 and Mrc1 are binding partners of Sen1 in the replisome. Deletion of either replisome component leads to a decrease in replisome association to Sen1, even following crosslinking to capture weak interactions (Figures S2C and S2D). Because *ctf4Δ mrc1Δ* cells are inviable (Gambus et al., 2009), we generated a *ctf4Δ mrc1-AID* strain, with the auxin-degron fused to Mrc1 (Nishimura et al., 2009) to allow rapid depletion of the protein. The association of Sen1 with the replisome was greatly reduced, although not entirely abolished, in cells with no Ctf4 and Mrc1 (Figures 2F–2H). These data indicate that, although other accessory binding partners might exist within the replisome, Sen1 mainly binds via Ctf4 and Mrc1.

sen1-3 Fails to Bind the Replisome and Is Sensitive to Increased Levels of DNA:RNA Hybrids

Deletion of the N-terminal domain of Sen1 causes pronounced defects in cell growth (Figure S3A). Thus, to investigate the role of Sen1 at RFs, we sought to generate a separation of function allele that is specifically defective for binding to replisomes. By generating truncations of the N-terminal domain, we identified that Sen1 (410–931) was the fragment with the highest affinity for replisomes although Sen1 (622–931) was the smallest construct still able to bind (Figures 3A–3C). By comparison with yeast orthologs of Sen1 (Figure S3B), we identified conserved residues within this region and targeted them for mutagenesis, creating hemagglutinin (HA)-tagged alleles of *SEN1* that were expressed under the strong *ACT1* promoter in *sen1Δ* cells. All the tested mutations supported cell growth, but one allele, combining mutations W773A E774A W777A (henceforth referred to as *sen1-3*) was uniquely defective for interaction with replisomes (Figures 3D and S3C). Similar results were obtained when the *sen1-3* mutation was

Figure 2. Sen1 Binds the Replisome Components Ctf4 and Mrc1

(A) MS analysis of the proteins co-purifying with Sen1 (2–931) was conducted in S and G1 phases.

(B) IB analysis of the proteins IPed with Sen1 (2–931) and an empty control in strains carrying the *PSF1* or *psf1-1* allele. Cells were arrested in G1, shifted to 37°C for 1 h (G1), and then released into S phase for 20 min at 37°C (S).

(C) Sen1 (2–931) binding of GINS in G1 depends on Ctf4. IB analysis of the proteins IPed with Sen1 (2–931) and an empty control, with or without *CTF4*. Cells were arrested in G1 and released in S phase for 20 min at 30°C. Ctf4 and TAP-Sen1 (2–931) have similar sizes and run closely in gel electrophoresis.

(D) IB analysis of the proteins interacting with TAP-Sen1 (2–931) in the presence or absence of origin firing and *CTF4*. Cells were treated as described in Figure S2B. G1 samples were collected before galactose induction.

(E) Wild-type, *mrc1Δ*, or *ctf4Δ* cells expressing TAP-Sen1 (2–931) were arrested in G1. IB analysis of cell extracts and IPs is shown.

(F) Wild-type, *ctf4Δ*, *mrc1Δ*, and *ctf4Δ mrc1-AID* strains were arrested in G1, treated for 1 h with 0.5 mM auxin indole-3-acetic acid (IAA) final concentration, and released in S phase. IB analysis of cell extracts and IPs is shown.

(G) Quantification of the relative signal of Sen1-9MYC versus the TAP-Sld5 signal, normalized against the wild type.

(H) Experiments were conducted as in (F). Wild-type, *ctf4Δ*, and *ctf4Δ mrc1-AID* strains, carrying an untagged or a *SEN1-TAP* allele, were used. Asterisk indicates a non-specific band.

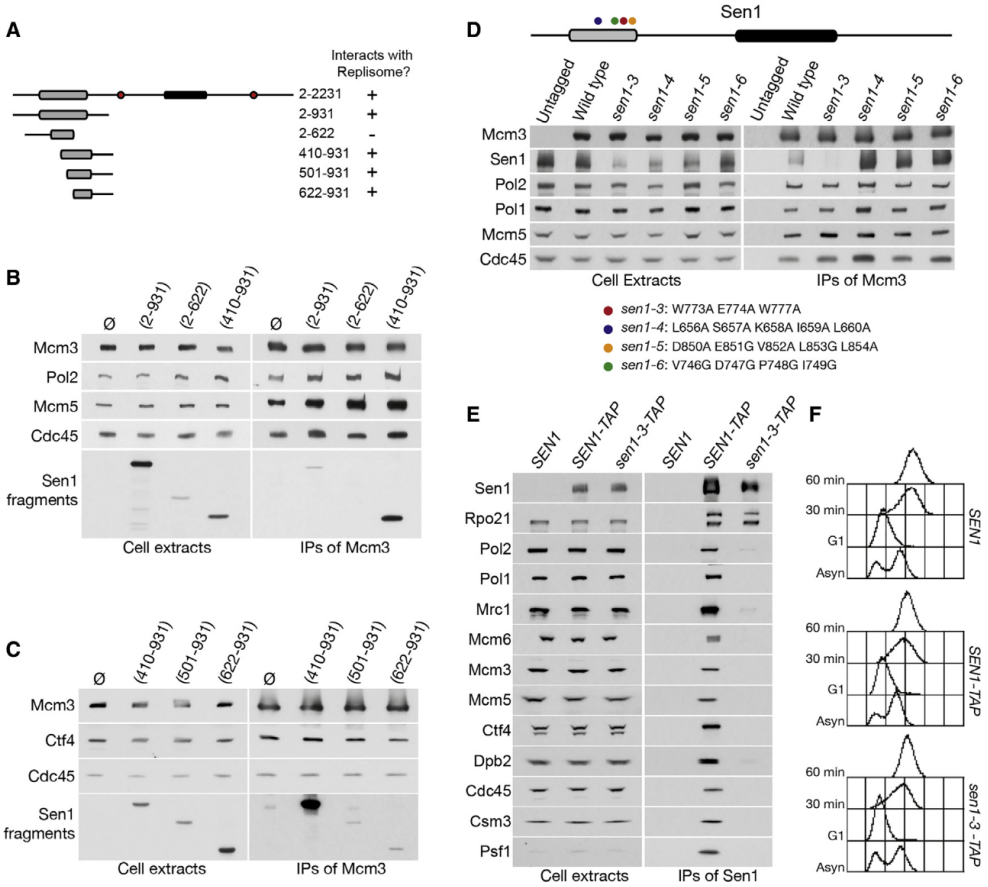


Figure 3. Sen1-3 Does Not Interact with the Replisome

(A) Summary of the ability of N-terminal fragments of Sen1 to interact with the replisome.

(B) Cells carrying different *GAL1-3HA-SEN1* fragments and a *TAP-MCM3* allele were arrested in G1 and released into S phase. The samples were then used for IPs.

(C) Sen1 fragments were analysed as in (B).

(D) Cells carrying *ACT1-3HA-SEN1* wild-type or mutated alleles at an ectopic locus were synchronously released into S phase. IB analysis of cell extracts and IPs is shown.

(E) Cells carrying a *SEN1*, *SEN1-TAP*, or *sen1-3-TAP* allele were arrested in G1 and released into S phase. IB analysis of cell extracts and IPs is shown.

(F) Fluorescence-activated cell sorting (FACS) samples for the experiment in (E).

introduced at the endogenous *SEN1* locus (Figures 3E and 3F), even following crosslinking (Figure S3D). Importantly, Sen1-3 retained wild-type affinity for RNAPII (Rpo21). Hence, *sen1-3* is an allele that abrogates the interaction between Sen1 and replisomes.

Next, we assessed whether the *sen1-3* mutation affects transcription termination, similarly to *sen1-1* cells (Mischo et al., 2011). We assayed the efficiency of termination at two model Nrd1-Nab3-Sen1 target genes, coding for a snoRNA (*SNR13*) and a CUT (*NEL025c*) (Thiebaut et al., 2006; Ursic et al., 1997). Because termination defects lead to longer RNAs that can be targeted by the nonsense-mediated decay, strains lacking *UPF1* were also tested (Culbertson and Leeds, 2003). Cells with the *sen1-3* allele presented no defects in transcription termination, unlike *sen1-1* at 37°C (Figures 4A and S3E). Defects in transcription termination were also analyzed genome-wide by mapping the distribution of RNAPII via the sequencing of nascent RNAs using CRAC (crosslinking and analysis of cDNAs) (Granneman et al., 2009; Candelli et al., 2018). Metagene analysis using a set of validated CUTs (Table S1) shows very similar RNAPII profiles between *SEN1* and *sen1-3* cells, although a clear general termination defect is observed upon depletion of Nrd1 (Figures 4B and S3F). These data indicate that *sen1-3* is proficient in terminating RNAPII transcription.

We then analyzed how the loss of Sen1 from replisomes affects cells. *SEN1* and *sen1-3* cells displayed comparable cell

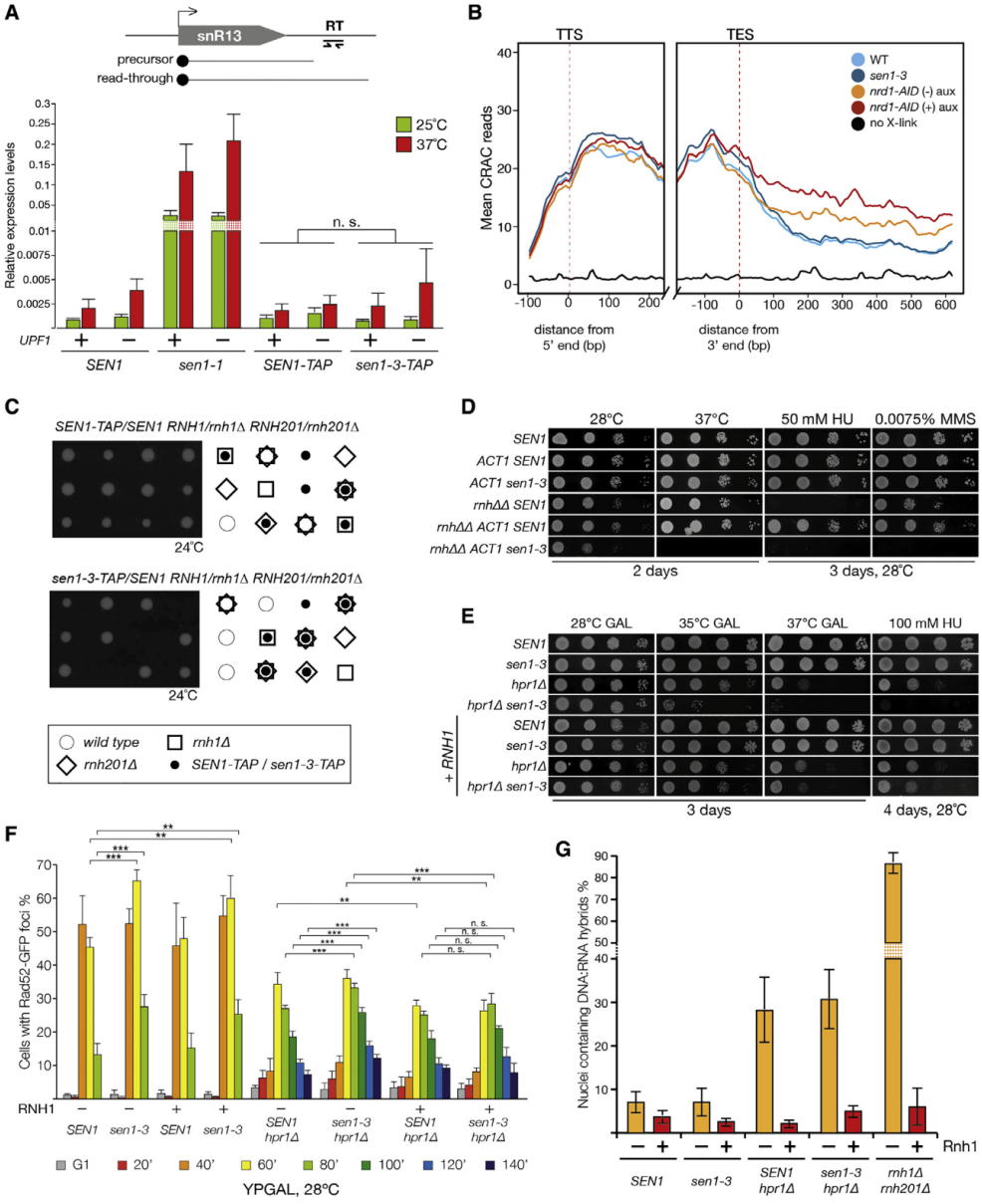


Figure 4. The *sen1-3* Allele Is Proficient in RNAPII Termination but Is Essential in the Absence of RNase H Activity

(A) *sen1-3* cells are proficient for transcription termination. qRT-PCR analysis of RNAs derived from the strains indicated is shown. The ratio of the readthrough fraction (position RT) over the total amount of *SNR13* RNA is shown (triplicate biological repeats). n.s., not significant.

(B) Metagene analysis of RNAPII density detected by CRAC on CUTs. Average read counts are plotted on regions aligned to both the transcription start site (TSS) (left) and the transcript end site (TES) (right) of the CUTs (reads count in Table S1). The profiles of RNAPII density following Nrd1 depletion (*nrd1-AID* + auxin) are included for comparison (dataset from Candelli et al., 2018). *nrd1-AID* strain behaves as a hypomorphic allele.

(C) Examples of the meiotic progeny of the indicated diploids strains are shown.

(D) Serial dilution spotting of the indicated strains is shown. *rnh1Δ rnh201Δ* is abbreviated as *rnhΔΔ*.

(E) Serial dilution spotting of the indicated strains is shown. Cells (+*RNH1*) carry *GAL-RNH1* inserted ectopically.

(legend continued on next page)

growth kinetics and sensitivity to both hydroxyurea (HU) and methyl methanesulfonate (MMS). One possibility might be that Sen1 at RFs is redundant with the enzymatic activity of other factors, such as the RNase H1 and H2 enzymes. We crossed *rnh1Δ rnh201Δ* cells with *SEN1 or sen1-3* strains and analyzed their meiotic progeny. Although single deletion of either *RNH1* or *RNH201* combined with *sen1-3* did not present any synthetic defects, *sen1-3 rnh1Δ rnh201Δ* cells were inviable (Figure 4C), similarly to *rnh1Δ rnh201Δ sen1-1* mutants (Figure S3G). Overexpression of *sen1-3* under the strong *ACT1* promoter suppresses the synthetic lethality of *sen1-3* with *rnh1Δ rnh201Δ*, suggesting that higher levels of Sen1 activity can compensate for lack of the specific replisome-tethering mechanism. Yet these cells display growth defects at 37°C, with cells accumulating in G2/M and triggering checkpoint activation (Figures S3H–S3J). Moreover, *ACT1-sen1-3* is unable to suppress the hyper-sensitivity of *rnh1Δ rnh201Δ* to HU and is synthetic defective for MMS sensitivity (Figure 4D). Altogether, these findings suggest that Sen1 at RF might either be redundant with RNases H1 and H2 or become essential to deal with the DNA:RNA hybrids accumulating in the absence of RNase H.

To explore whether increased levels of DNA:RNA hybrids lead to synthetic defects in *sen1-3* cells, we generated *hpr1Δ sen1-3* cells. Hpr1 is a component of the THO complex involved in the processing and export of mRNA (Chávez et al., 2000). *hpr1Δ* mutants accumulate R-loops and show defects in transcription elongation (García-Benítez et al., 2017; Chávez and Aguilera, 1997; Chávez et al., 2000). *hpr1Δ sen1-3* double mutants showed growth defects at higher temperatures and increased sensitivity to replication stress (Figure 4E). To explore whether defects arise during DNA replication, we analyzed the kinetics of Rad52 foci formation in cells released in S phase. The experiment was conducted at permissive temperatures (28°C) as *hpr1Δ* cells failed to synchronously bud at 35°C and 37°C. We observed that *sen1-3* causes a small but statistically significant increase in recombination in late S phase, although *hpr1Δ sen1-3* cells showed synthetic defects and an increase in recombination (Figures 4F and S4A–S4D). Interestingly, the increased rates of recombination and growth defects in *hpr1Δ sen1-3* cells were suppressed by overexpression of *RNH1* (Figures 4E and 4F), thus suggesting that DNA:RNA hybrids are toxic in these mutants.

To directly test the levels of DNA:RNA hybrids, we visualized them in chromosome spreads (Wahba et al., 2011). As previously observed, both *rnh1Δ rnh201Δ* and *hpr1Δ* mutants showed high levels of DNA:RNA hybrids (Figures 4G and S4E; Chan et al., 2014). Surprisingly, we did not observe any increase in the levels of DNA:RNA hybrids in *hpr1Δ sen1-3* cells. Similar results were observed by slot-blot analysis (Figure S4F). Given that phenotypic suppression by RNase H overexpression is accepted as a marker for R-loops, these results suggest that the suppression of *hpr1Δ sen1-3* by overexpression of *RNH1* might occur by removing short or labile DNA:RNA hybrids, not readily detectable by the S9.6 antibody used in our analysis.

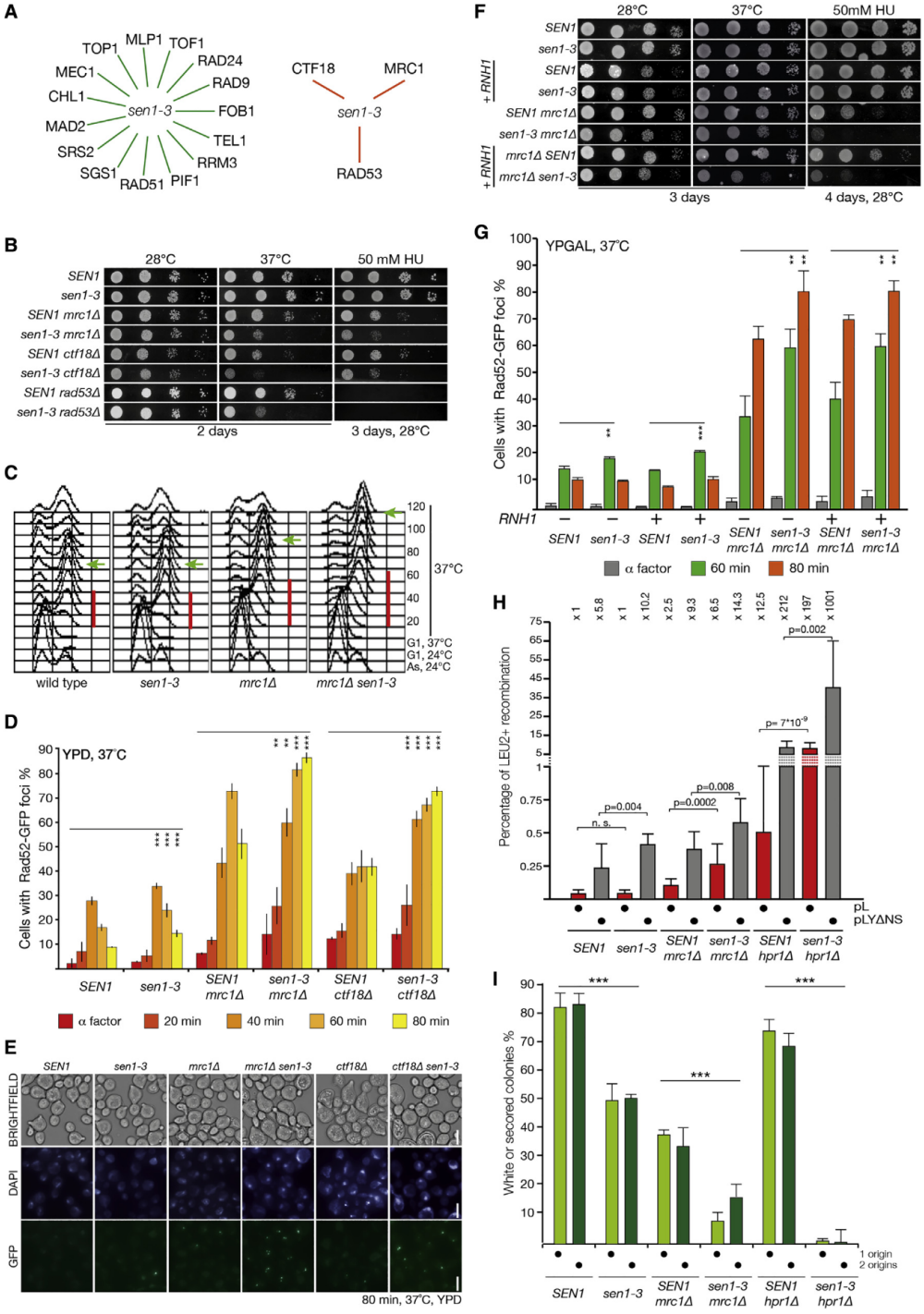
sen1-3 Cells Are Defective in Replication Fork Progression and Genome Stability in the Absence of *MRC1*

Because both *sen1-1* and *sen1-3* are synthetically lethal in the absence of *RNH1* and *RNH201*, we wanted to explore whether other pathways, essential for maintaining cell viability in *sen1-1* (Alzu et al., 2012; Mischo et al., 2011), are also important in *sen1-3*. Only a subset of deletion mutants described to negatively affect viability in *sen1-1* cells showed robust defects in cell viability in *sen1-3* cells (summarized in Figure 5A). Namely, we observed temperature sensitivity and increased sensitivity to DNA-damaging agents when *sen1-3* was crossed with either *mrc1Δ*, *ctf18Δ*, or *rad53Δ sml1Δ* (Figure 5B).

Mrc1, Ctf18, and Rad53 are key components of the S phase checkpoint, and all three mutants confer temperature sensitivity in *sen1-3* cells. To further explore the defects of *mrc1Δ sen1-3* mutants, we analyzed the DNA replication dynamics of cells arrested in G1 and then released in S phase at 37°C. The *mrc1Δ sen1-3* cells show a delay during DNA replication and accumulation of cells arrested in G2/M (Figure 5C). Correspondingly, we observed an increase in Rad52-GFP foci accumulating during the later stages of DNA replication, both in the *mrc1Δ sen1-3* and *ctf18Δ sen1-3* cells released in S phase at 37°C (Figures 5D and 5E). In addition, *mrc1Δ sen1-3* and *ctf18Δ sen1-3* cells showed an increase in cells carrying multiple foci of Rad52 (Figure S5A). Similar to what is seen in Figure 4F, we also observed a small but statistically significant increase in Rad52 foci in *sen1-3* mutants compared to wild-type. To determine whether DNA:RNA hybrids contribute to the phenotypes observed in *sen1-3 mrc1Δ*, we repeated the experiments following the overexpression of *RNH1*. This failed to suppress the growth defects and the increase in recombination during S phase (Figures 5F, 5G, and S5B). Similar results were obtained when overexpressing the human ortholog of *RNH1* (Figures S5C and S5D; Wahba et al., 2011; Bonnet et al., 2017).

To analyze whether the increased recombination observed during replication in *sen1-3*, *mrc1Δ sen1-3*, and *hpr1Δ sen1-3* compared to *SEN1* leads to an increase in genomic instability, we measured the rate of direct-repeat recombination using plasmids carrying partially overlapping fragments of the *LEU2* gene separated by 39 or 3,900 nt (plasmids pL and pLYΔNS, respectively) (Mischo et al., 2011; González-Aguilera et al., 2008). We observed, as previously described, that recombination increased with the length of the transcript. Moreover, although *mrc1Δ sen1-3* showed a modest increase in recombination compared to *mrc1Δ* for both plasmids, *hpr1Δ sen1-3* showed greater increases in the rate of recombination (Figure 5H). Furthermore, we tested for defects in mini-chromosome maintenance by transforming a single-copy plasmid carrying an *ADE2* gene and scoring for the rate of plasmid loss in the absence of selective pressure by measuring the rate of white colonies (carrying the plasmid) and red (without the plasmid). Cells carrying the *sen1-3* allele showed higher levels of plasmid loss,

(F) The indicated strains, carrying a *RAD52-GFP* allele with or without the *GAL-RNH1* construct, were grown as shown in Figures S4A–S4D. Samples were taken at the indicated time points, fixed, and analyzed for the presence of Rad52 foci (triplicate biological repeats). n.s., not significant; **p < 0.05; ***p < 0.01.
(G) The indicated strains were grown to exponential phase at 28°C; DNA:RNA hybrids were analyzed by immunofluorescence of chromosome spreads (triplicate biological repeats). Samples were treated in parallel with RNase H.



(legend on next page)

exacerbated in the absence of *MRC1* and *HPR1* (Figures 5I and S5E). Strikingly, *sen1-3 hpr1Δ* completely failed to retain the plasmid. The addition of a second origin of replication did not rescue the chromosome maintenance defects, and overexpression of *RNH1* only partially suppressed the defects in *hpr1Δ sen1-3* cells (Figure S5F).

DISCUSSION

Here, we have shown that Sen1 is a bona fide partner of the yeast replisome. The N-terminal domain mediates binding to replisomes, mainly via Ctf4 and Mrc1. Additional binding partners of Sen1 are likely because Sen1 shows some residual interaction with replisomes in the absence of Ctf4 and Mrc1. It is not yet clear whether multiple Sen1 molecules are recruited to RFs by independently binding separate subunits of the replisome with different affinities or whether multiple replisome components coordinately bind a single Sen1 to increase its strength of interaction. IPs of the N-terminal domain of Sen1 suggest a competition between Mrc1 and Ctf4 for Sen1 as Mrc1 binding increases in *ctf4Δ* cells (Figures 2D and 2E). This supports the multiple independent binding hypothesis. However, deletion of either *CTF4* or *MRC1* decreases overall binding of Sen1 to the replisome (Figures 2F, 2H, S2C, and S2D), compatible with a cooperative recruitment of Sen1. Interestingly, the mutation of three amino acids in *sen1-3* abolishes binding to both Ctf4 and Mrc1. Thus, the mutated residues either correspond to the direct interaction site for both Ctf4 and Mrc1 or they cause a change in conformation of a larger section of Sen1, thus affecting two distinct binding surfaces for Ctf4 and Mrc1. Both hypotheses are compelling, and further work is needed to determine which is correct.

The *sen1-3* allele is a separation of function mutant that breaks the interaction with the replisome without affecting the binding to RNAPII or transcription termination (Figures 3E, 4A, and 4B). However, we cannot exclude that *sen1-3* might affect other Sen1 interactions beyond the replisome. In addition, minimal levels of interaction with replisomes might be retained in *sen1-3* cells, thus weakening the severity of the phenotype observed. Nevertheless, this allele provides us with a tool to dissect the function of the helicase at RFs without affecting its catalytic activity and the bulk of its transcription functions.

It has been previously proposed (using the *sen1-1* allele or Sen1 depletion) that Sen1's presence at RFs is required to quickly remove the R-loops accumulating and interfering with RF progression (Alzu et al., 2012; Brambati et al., 2018; Mischo et al., 2011). In our experimental setting, however, loss of Sen1 from RFs did not show increases in DNA:RNA hybrids or dramatic defects in RF progression (Figures 4G and 5C). In fact, the loss of Sen1 from the replisome only leads to modest defects (small increases in post-replicative recombination and instability of mini-chromosomes; Figures 4F, 5D, and 5I). This suggests that when Sen1 is proficient in transcription termination, there might be enough redundancy at the RFs to deal with DNA:RNA hybrids. However, we observe lethality or severe growth defects when the *sen1-3* allele is present in genetic backgrounds with high endogenous levels of R-loops, such as *rnh1Δ rnh201Δ* and *hpr1Δ*. This supports the idea of an important role for Sen1 in dealing with DNA:RNA hybrids at RFs. Surprisingly, we do not observe an increase in DNA:RNA hybrids levels in *sen1-3* and in *hpr1Δ sen1-3* cells (Figures 4G, S4E, and S4F). Moreover, although increased levels of R-loops have been described for *top1Δ* (El Hage et al., 2010), *pif1Δ* (Boulé and Zakian, 2007; Tran et al., 2017), *sgs1Δ* (Chang et al., 2017), or *mlp1Δ* (García-Benítez et al., 2017), these deletions do not show defects in cell viability or DNA damage sensitivity in combination with *sen1-3* (Figure 5A). This suggests that not all increases in R-loops might be necessarily toxic in *sen1-3* cells. One possibility is that different mutations might lead to dissimilar levels or distinct biochemical features of the R-loops. Moreover, different genetic backgrounds might lead to the accumulation of DNA:RNA hybrids at different sites of the genome (as recently observed; Costantino and Koshland, 2018). Therefore, Sen1 association with the replisome might become critical for the timely resolution of some DNA:RNA hybrids in specific circumstances.

The recruitment of Sen1 at RFs also appears to promote DNA replication independently of R-loops. In fact, in *sen1-3* cells, overexpression of *RNH1* fails to suppress the higher levels of recombination observed in *sen1-3* (Figures 4F and 5D). Given the prominent role of Sen1 in transcription termination described in the literature, it is tempting to speculate that Sen1 might remove transcribing or stalled RNA polymerases at RFs (Han et al., 2017; Porrua and Libri, 2013). Alternatively, Sen1 might

Figure 5. *sen1-3* Presents Synthetic Defects with *mrc1Δ*, *ctf18Δ*, and *rad53Δ*, Leading to Increased Recombination and Mini-chromosome Loss

(A) Summary of the genetic interactions tested with the *sen1-3* allele. Some double mutants (orange line) showed marked differences in temperature sensitivity and DNA damage sensitivity although others did not (green line).

(B) Examples of the defects observed with *sen1-3*. Serial dilution spotting of the indicated strains is shown. The double mutant *rad53Δ sml1Δ* is indicated as *rad53Δ*.

(C) The indicated strains were arrested in G1, shifted to 37°C for 1 h, and released in S phase at 37°C. FACS samples were taken at the indicated times. Red bar, length of DNA replication; green arrow, beginning of the end of mitosis.

(D) Cells, carrying a *RAD52-GFP* allele, were treated as in (C). Samples were taken at the indicated time points, fixed, and analyzed for the presence of Rad52 foci (triplicate biological repeats). $^{**}p < 0.05$; $^{***}p < 0.01$.

(E) Examples of the microscopy data of the experiment in (D). Scale bars represent 5 μm.

(F) Serial dilution spotting of the indicated strains is shown. Cells (+*RNH1*) carry an ectopic *GAL1-RNH1* construct.

(G) *RNH1* overexpression does not suppress the increase in recombination in *mrc1Δ sen1-3* cells. Cell cultures were treated as in (C), except they were grown in YPGAL medium (triplicate biological repeats). $^{**}p < 0.05$; $^{***}p < 0.01$.

(H) The *sen1-3* allele causes an increase in recombination. The cells were transformed with the plasmids pL or pLYΔNS. The ratio of the number of the colonies carrying a recombinant plasmid (*LEU2*) over the total number of cells carrying a plasmid (*URA3*) is shown.

(I) Cells were transformed with plasmids carrying an *ADE2* marker and 1 or 2 origins. Percentage of white colonies over the total number of colonies scored is shown (a measure of genome stability; $^{***}p < 0.5\ 10^{-7}$).

be required to remove other barriers to fork progression, or *RNH1* overexpression might not be sufficient to remove DNA:RNA hybrids present at RF with kinetics similar to Sen1, thus leading to increased fork stalling. In either case, we observe that cells rely on the functions of Mrc1 to promote fork progression and minimize DNA recombination in a *sen1-3* background (Figures 5B–5G). Interestingly, we observe that three key mediators and effectors of the S phase checkpoint (*MRC1*, *CTF18*, and *RAD53*) genetically interact with *sen1-3*. We did not observe any synthetic defects between *sen1-3* and either *mec1Δ sml1Δ*, *tel1Δ*, or *mec1Δ sml1Δ tel1Δ* (not shown). This raises the possibility that Mrc1, Ctf18, and Rad53 might be involved in the response to defects arising in *sen1-3* cells independently of Mec1 and Tel1. Alternatively, the synthetic defects observed are the consequence of other deficiencies in these cells, independent of the S phase checkpoint response. For example, Mrc1 has a key role in RF progression (Yeeles et al., 2017; Hodgson et al., 2007; Duch et al., 2018).

Given that eukaryotic orthologs of Sen1 contain an extended non-catalytic N-terminal sequence (the function of which is still largely unknown), it will be interesting to investigate further whether Senataxin or any of its paralogs (Aquarius, IGHMBP2, RENT1, and ZNFx1) associate with replisomes in higher eukaryotes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Yeast Strains and Growth Conditions
  - Cell cycle experiments
  - Harvesting cells for IP
  - Western Blots
  - IP
  - MS Analysis of IPs
  - Counting of Rad52-foci to assess DNA damage
  - Chromosome spreads and microscopy
  - Quantification of R-loops
  - Reverse transcription followed by quantitative PCR
  - Cross-linking and analysis of cDNA (CRAC)
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.01.087.

ACKNOWLEDGMENTS

We thank Karim Labib, Benoit Palancade, Andres Aguilera, and Nicholas Proudfoot for strains, antibodies, and plasmids. We are grateful to Karim Labib, Jordi Torres-Rosell, Jonathan Millar, and Andrew McAinsh for feedback. The authors acknowledge CAMDU (Computing and Advanced Microscopy Unit) and Media Preparation Facility at the University of Warwick for their assistance in this work, as well as the help of the high-throughput sequencing core facility of I2BC (Centre de Recherche de Gif-sur-Yvette, France). G.D.P. and R.A. were funded by Cancer Research UK Career Development Fellowship C44595/A16326. R.A. has been funded by Chancellor International Scholarship at the University of Warwick. E.C.L. is funded by the BBSRC MIBTP program. This work was also supported by the Centre National de la Recherche Scientifique (CNRS), l'Agence National pour la Recherche (ANR) grants ANR-12-BSV8-0014-01 and ANR-16-CE12-0022-01 to D.L. and the Labex Who Am I? (ANR-11-LABX-0071 et Idex ANR-11-IDEX-0005-02 to D.L. U.A. is supported by a fellowship from the French Ministry of Research.

AUTHOR CONTRIBUTIONS

R.A., D.L., and G.D.P. conceived the study; R.A., E.C.L., U.A., and G.D.P. performed experiments; all authors analyzed experiments; and R.A. and G.D.P. wrote the manuscript. All authors discussed the data and commented and helped improve the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 5, 2019
Revised: September 6, 2019
Accepted: January 24, 2020
Published: February 18, 2020

REFERENCES

Aguilera, A., and García-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. Mol. Cell *46*, 115–124.

Alzu, A., Bermejo, R., Begnis, M., Lucca, C., Piccini, D., Carotenuto, W., Saponaro, M., Brambati, A., Cocito, A., Foiani, M., and Liberi, G. (2012). Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes. Cell *151*, 835–846.

Arigo, J.T., Eyler, D.E., Carroll, K.L., and Corden, J.L. (2006). Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. Mol. Cell *23*, 841–851.

Bell, S.P., and Labib, K. (2016). Chromosome duplication in Saccharomyces cerevisiae. Genetics *203*, 1027–1067.

Bonnet, A., Grosso, A.R., Elkaoutari, A., Coleno, E., Presle, A., Sridhara, S.C., Janbon, G., Géli, V., de Almeida, S.F., and Palancade, B. (2017). Introns protect eukaryotic genomes from transcription-associated genetic instability. Mol. Cell *67*, 608–621.e6.

Boulé, J.B., and Zakian, V.A. (2007). The yeast Pif1p DNA helicase preferentially unwinds RNA DNA substrates. Nucleic Acids Res. *35*, 5809–5818.

Brambati, A., Zardoni, L., Achar, Y.J., Piccini, D., Galanti, L., Colosio, A., Foiani, M., and Liberi, G. (2018). Dormant origins and fork protection mechanisms rescue sister forks arrested by transcription. Nucleic Acids Res. *46*, 1227–1239.

Burgers, P.M.J., and Kunkel, T.A. (2017). Eukaryotic DNA replication fork. Annu. Rev. Biochem. *86*, 417–438.

Candelli, T., Challal, D., Briand, J.B., Boulay, J., Porrua, O., Colin, J., and Libri, D. (2018). High-resolution transcription maps reveal the widespread impact of roadblock termination in yeast. EMBO J. *37*, e97490.

Cerritelli, S.M., and Crouch, R.J. (2009). Ribonuclease H: the enzymes in eukaryotes. FEBS J. *276*, 1494–1505.

Chan, Y.A., Aristizabal, M.J., Lu, P.Y., Luo, Z., Hamza, A., Kobor, M.S., Stirling, P.C., and Hieter, P. (2014). Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. PLoS Genet. *10*, e1004288.

Chang, E.Y., Novoa, C.A., Aristizabal, M.J., Coulombe, Y., Segovia, R., Chaturvedi, R., Shen, Y., Keong, C., Tam, A.S., Jones, S.J.M., et al. (2017).

RECQ-like helicases Sgs1 and BLM regulate R-loop-associated genome instability. J. Cell Biol. *216*, 3991–4005.

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

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

Chinchilla, K., Rodriguez-Molina, J.B., Ursic, D., Finkel, J.S., Ansari, A.Z., and Culbertson, M.R. (2012). Interactions of Sen1, Nrd1, and Nab3 with multiple phosphorylated forms of the Rpb1 C-terminal domain in *Saccharomyces cerevisiae*. Eukaryot. Cell *11*, 417–429.

Costantino, L., and Koshland, D. (2018). Genome-wide map of R-loop-induced damage reveals how a subset of R-loops contributes to genomic instability. Mol. Cell *71*, 487–497.e3.

Creamer, T.J., Darby, M.M., Jamonnak, N., Schaughency, P., Hao, H., Wheelan, S.J., and Corden, J.L. (2011). Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. PLoS Genet. *7*, e1002329.

Culbertson, M.R., and Leeds, P.F. (2003). Looking at mRNA decay pathways through the window of molecular evolution. Curr. Opin. Genet. Dev. *13*, 207–214.

De Piccoli, G., Katou, Y., Itoh, T., Nakato, R., Shirahige, K., and Labib, K. (2012). Replisome stability at defective DNA replication forks is independent of S phase checkpoint kinases. Mol. Cell *45*, 696–704.

Duch, A., Canal, B., Barroso, S.I., García-Rubio, M., Seisenbacher, G., Aguilera, A., de Nadal, E., and Posas, F. (2018). Multiple signaling kinases target Mrc1 to prevent genomic instability triggered by transcription-replication conflicts. Nat. Commun. *9*, 379.

El Hage, A., French, S.L., Beyer, A.L., and Tollervey, D. (2010). Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. Genes Dev. *24*, 1546–1558.

El Hage, A., Webb, S., Kerr, A., and Tollervey, D. (2014). Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. PLoS Genet. *10*, e1004716.

Gambus, A., van Deursen, F., Polychronopoulos, D., Foltman, M., Jones, R.C., Edmondson, R.D., Calzada, A., and Labib, K. (2009). A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase α within the eukaryotic replisome. EMBO J. *28*, 2992–3004.

García-Benítez, F., Gaillard, H., and Aguilera, A. (2017). Physical proximity of chromatin to nuclear pores prevents harmful R loop accumulation contributing to maintain genome stability. Proc. Natl. Acad. Sci. USA *114*, 10942–10947.

García-Pichardo, D., Cañas, J.C., García-Rubio, M.L., Gómez-González, B., Rondón, A.G., and Aguilera, A. (2017). Histone mutants separate R loop formation from genome instability induction. Mol. Cell *66*, 597–609.e5.

González-Aguilera, C., Tous, C., Gómez-González, B., Huertas, P., Luna, R., and Aguilera, A. (2008). The THP1-SAC3-SUS1-CDC31 complex works in transcription elongation-mRNA export preventing RNA-mediated genome instability. Mol. Biol. Cell *19*, 4310–4318.

Granneman, S., Kudla, G., Petfalski, E., and Tollervey, D. (2009). Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. Proc. Natl. Acad. Sci. USA *106*, 9613–9618.

Grubb, J., Brown, M.S., and Bishop, D.K. (2015). Surface spreading and immunostaining of yeast chromosomes. J. Vis. Exp., e53081.

Gudipati, R.K., Villa, T., Boulay, J., and Libri, D. (2008). Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. Nat. Struct. Mol. Biol. *15*, 786–794.

Hamperl, S., Bocek, M.J., Saldivar, J.C., Swigut, T., and Cimprich, K.A. (2017). Transcription-replication conflict orientation modulates R-loop levels and activates distinct DNA damage responses. Cell *170*, 774–786.e19.

Han, Z., Libri, D., and Porrua, O. (2017). Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. Nucleic Acids Res. *45*, 1355–1370.

Hazelbaker, D.Z., Marquardt, S., Wlotzka, W., and Buratowski, S. (2013). Kinetic competition between RNA polymerase II and Sen1-dependent transcription termination. Mol. Cell *49*, 55–66.

Helmrich, A., Ballarino, M., and Tora, L. (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. Mol. Cell *44*, 966–977.

Helmrich, A., Ballarino, M., Nudler, E., and Tora, L. (2013). Transcription-replication encounters, consequences and genomic instability. Nat. Struct. Mol. Biol. *20*, 412–418.

Hodgson, B., Calzada, A., and Labib, K. (2007). Mrc1 and Tof1 regulate DNA replication forks in different ways during normal S phase. Mol. Biol. Cell *18*, 3894–3902.

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

Janke, C., Magiera, M.M., Rathfelder, N., Taxis, C., Reber, S., Maekawa, H., Moreno-Borchart, A., Doenges, G., Schwob, E., Schiebel, E., and Knop, M. (2004). A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. Yeast *21*, 947–962.

Jankowsky, E. (2011). RNA helicases at work: binding and rearranging. Trends Biochem. Sci. *36*, 19–29.

Kamimura, Y., Tak, Y.S., Sugino, A., and Araki, H. (2001). Sld3, which interacts with Cdc45 (Sld4), functions for chromosomal DNA replication in *Saccharomyces cerevisiae*. EMBO J. *20*, 2097–2107.

Kanemaki, M., and Labib, K. (2006). Distinct roles for Sld3 and GINS during establishment and progression of eukaryotic DNA replication forks. EMBO J. *25*, 1753–1763.

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F., and Buratowski, S. (2004). The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature *432*, 517–522.

Kim, T.S., Liu, C.L., Yassour, M., Holik, J., Friedman, N., Buratowski, S., and Rando, O.J. (2010). RNA polymerase mapping during stress responses reveals widespread nonproductive transcription in yeast. Genome Biol. *11*, R75.

Lang, K.S., Hall, A.N., Merrikh, C.N., Ragheb, M., Tabakh, H., Pollock, A.J., Woodward, J.J., Dreifus, J.E., and Merrikh, H. (2017). Replication-transcription conflicts generate R-loops that orchestrate bacterial stress survival and pathogenesis. Cell *170*, 787–799.e18.

Leonaitė, B., Han, Z., Basquin, J., Bonneau, F., Libri, D., Porrua, O., and Conti, E. (2017). Sen1 has unique structural features grafted on the architecture of the Upf1-like helicase family. EMBO J. *36*, 1590–1604.

Liu, B., and Alberts, B.M. (1995). Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. Science *267*, 1131–1137.

Luke, B., Panza, A., Redon, S., Iglesias, N., Li, Z., and Lingner, J. (2008). The Rat1p 5′ to 3′ exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in *Saccharomyces cerevisiae*. Mol. Cell *32*, 465–477.

Martin-Tumasz, S., and Brow, D.A. (2015). Saccharomyces cerevisiae Sen1 helicase domain exhibits 5′- to 3′- helicase activity with a preference for translocation on DNA rather than RNA. J. Biol. Chem. *290*, 22880–22889.

Mischo, H.E., Gómez-González, B., Grzechnik, P., Rondón, A.G., Wei, W., Steinmetz, L., Aguilera, A., and Proudfoot, N.J. (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. Mol. Cell *41*, 21–32.

Mischo, H.E., Chun, Y., Harlen, K.M., Smalec, B.M., Dhir, S., Churchman, L.S., and Buratowski, S. (2018). Cell-cycle modulation of transcription termination factor Sen1. Mol. Cell *70*, 312–326.e7.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat. Methods *6*, 917–922.

Pfeiffer, V., Crittin, J., Grolimund, L., and Lingner, J. (2013). The THO complex component Thp2 counteracts telomeric R-loops and telomere shortening. EMBO J. *32*, 2861–2871.

Porrua, O., and Libri, D. (2013). A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. Nat. Struct. Mol. Biol. *20*, 884–891.

Porrua, O., Hobor, F., Boulay, J., Kubicek, K., D'Aubenton-Carafa, Y., Gudipati, R.K., Stefl, R., and Libri, D. (2012). In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. EMBO J. *31*, 3935–3948.

Prado, F., and Aguilera, A. (2005). Impairment of replication fork progression mediates RNA polII transcription-associated recombination. EMBO J. *24*, 1267–1276.

Rondón, A.G., Mischo, H.E., Kawauchi, J., and Proudfoot, N.J. (2009). Fail-safe transcriptional termination for protein-coding genes in *S. cerevisiae*. Mol. Cell *36*, 88–98.

Schaughency, P., Merran, J., and Corden, J.L. (2014). Genome-wide mapping of yeast RNA polymerase II termination. PLoS Genet. *10*, e1004632.

Skourti-Stathaki, K., Kamieniarz-Gdula, K., and Proudfoot, N.J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. Nature *516*, 436–439.

Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z., and Brow, D.A. (2006). Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. Mol. Cell *24*, 735–746.

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

Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., and Libri, D. (2006). Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. Mol. Cell *23*, 853–864.

Tran, P.L.T., Pohl, T.J., Chen, C.F., Chan, A., Pott, S., and Zakian, V.A. (2017). PIF1 family DNA helicases suppress R-loop mediated genome instability at tRNA genes. Nat. Commun. *8*, 15025.

Ursic, D., Himmel, K.L., Gurley, K.A., Webb, F., and Culbertson, M.R. (1997). The yeast *SEN1* gene is required for the processing of diverse RNA classes. Nucleic Acids Res. *25*, 4778–4785.

Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S., and Meinhart, A. (2008). The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. Nat. Struct. Mol. Biol. *15*, 795–804.

Wahba, L., Amon, J.D., Koshland, D., and Vuica-Ross, M. (2011). RNase H and multiple RNA biogenesis factors cooperate to prevent RNA:DNA hybrids from generating genome instability. Mol. Cell *44*, 978–988.

Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. Cell *119*, 481–489.

Yeeles, J.T.P., Janska, A., Early, A., and Diffley, J.F.X. (2017). How the eukaryotic replisome achieves rapid and efficient DNA replication. Mol. Cell *65*, 105–116.

Yüce, Ö., and West, S.C. (2013). Senataxin, defective in the neurodegenerative disorder ataxia with oculomotor apraxia 2, lies at the interface of transcription and the DNA damage response. Mol. Cell. Biol. *33*, 406–417.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-mouse-HRP	Cell Signaling Technology	#7076; RRID:AB_330924
Anti-sheep-HRP	Sigma	A3415; RRID:AB_258076
Anti-Cdc45	Labib Lab	N/A
Anti-Csm3	Labib Lab	N/A
Anti-Ctf4	Labib Lab	N/A
Anti-Dpb2	Labib Lab	N/A
Anti-FLAG	Sigma	F3165; RRID:AB_259529
Anti-HA (12CA5)	Sigma	11583816001; RRID:AB_514505
Sheep IgG	Sigma	S1265; RRID:AB_261431
Anti-Mcm3	Labib Lab	N/A
Anti-Mcm4	Labib Lab	N/A
Anti-Mcm5	Labib Lab	N/A
Anti-Mcm6	Labib Lab	N/A
Anti-Mrc1	Labib Lab	N/A
Anti-MYC	Sigma	M4439; RRID:AB_439694
Anti-Nrd1	Libri Lab	N/A
Anti Nab3	Libri Lab	N/A
Anti-Psf1	Labib Lab	N/A
Anti-Pob3	Labib Lab	N/A
Anti-Pol1	Labib Lab	N/A
Anti-Pol2	De Piccoli Lab	N/A
Anti-Rad53	Abcam	ab166859; RRID:AB_2801547
Anti-Rpo21	Novus Biologicals	NB200-598SS; RRID:AB_2252678
Anti-Sld5	K. Labib	N/A
Anti-TAP-HRP	Sigma	P1291; RRID:AB_1079562
Anti DNA:RNA hybrids S9.6	Kerafast	ENH001; RRID:AB_2687463
Cy3-conjucated anti-mouse	Jackson laboratories	#115165003; RRID:AB_2338680
Anti-dsDNA	Abcam	ab27156; RRID:AB_470907
Chemicals, Peptides, and Recombinant Proteins		
α-factor	Pepceuticals	N/A
AcTEV protease	Thermo-Fischer	12575015
Calmodulin	Sigma	A6112
Complete protease inhibitor cocktail	Roche	11 836 153 001
Dithiothreitol	Sigma	D0632
Dynabeads	Invitrogen	14302D
Ethidium bromide	Sigma	E1510
Hydroxyurea	Sigma	H8627
Methyl methanesulfonate	Sigma	129925
Propidium iodide	Sigma	P4864
Protease Inhibitor Cocktail	Sigma	P8215
Sodium fluoride	Thermo-Fischer	S299500
Zymolyase	Zymo research	#E1005
RNase H	Invitrogen	#18021071

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Sodium glycerophosphate	Johnson Matthey	170096
Universal Nuclease	Pierce	88700
Critical Commercial Assays		
Amersham ECL Western Blotting Detection Reagent	GE Healthcare	RPN2108
LightCycler® FastStart DNA Master SYBR Green I	Roche	03003230001
MLV-Reverse Transcriptase	Thermofischer	28025013
QuikChange Lightning Site-Directed Mutagenesis Kit	QIAGEN	#210519
Experimental Models: Organisms/Strains		
S. cerevisiae (from W303) CS1*MATa*	Rothstein's lab	N/A
S. cerevisiae (from W303) CS74*MATa pep4Δ::ADE2*	Lab strain	N/A
S. cerevisiae (from W303) CS1125*MATa TAP-SLD5 (kanMX) SEN1-9MYC (hphNT) pep4Δ::URA3 ADE2*	This study	N/A
S. cerevisiae (from W303) CS1126*MATa SEN1-9MYC (hphNT) pep4Δ::URA3 ADE2*	This study	N/A
S. cerevisiae (from W303) CS1187*MATa TAP-SLD5 (kanMX) SEN1-9MYC (hphNT) pep4Δ::URA3 ADE2 ctf4Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS1353*MATa SEN1-TAP (kanMX) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1416*MATa TAP-MCM3 (kanMX) SEN1-9MYC (hphNT) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1711*MATa TAP-MCM3 (kanMX) GAL1-3HA-ø (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1714*MATa TAP-MCM3 (kanMX) leu2-3,112::GAL1-3HA-SEN1 (2-931) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1534*MATa TAP-SLD5 (kanMX) SEN1-9MYC (hphNT) pep4Δ::URA3 ADE2 mrc1Δ::hphNT*	This study	N/A
S. cerevisiae (from W303) CS1852*MATa leu2-3,112::GAL1-TAP-ø (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1933*MATa leu2-3,112::GAL1-TAP-SEN1 (1095-2231) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1941*MATa leu2-3,112::GAL1-TAP-SEN1 (2-2231) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1942*MATa leu2-3,112::GAL1-TAP-SEN1 (2-1901) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1943*MATa leu2-3,112::GAL1-TAP-SEN1 (931-2231) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1956*MATa leu2-3,112::GAL1-TAP-SEN1 (2-1103) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS1957*MATa leu2-3,112::GAL1-TAP-SEN1 (2-931) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS2030*MATa TAP-MCM3 (kanMX) leu2-3,112::GAL1-3HA-SEN1 (2-622) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS2032*MATa TAP-MCM3 (kanMX) leu2-3,112::GAL1-3HA-SEN1 (410-931) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS2056*MATα td-MYC-sen1-1 (klTRP1) GAL1-UBR1 (HISMX) leu2-3,112::GAL1-TAP-ø (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2058*MATα td-MYC-sen1-1 (klTRP1) GAL1-UBR1 (HISMX) leu2-3,112::GAL1-TAP-SEN1 (2-931) (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2061*MATα td-MYC-sen1-1 (klTRP1) GAL1-UBR1 (HISMX) leu2-3,112::GAL1-TAP-SEN1 (2-1901) (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2062*MATα td-MYC-sen1-1 (klTRP1) GAL1-UBR1 (HISMX) leu2-3,112::GAL1-TAP-SEN1 (1095-2231) (LEU2)*	This study	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
S. cerevisiae (from W303) CS2148*MATa TAP-MCM3 (kanMX) leu2-3,112::GAL1-3HA-SEN1 (501-931) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS2150*MATa TAP-MCM3 (kanMX) leu2-3,112::GAL1-3HA-SEN1 (622-931) (LEU2) pep4Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS2184*MATα td-MYC-sen1-1 (klTRP1) GAL1-UBR1 (HISMX) leu2-3,112::GAL1-TAP-SEN1 (2-1103) (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2188*MATα td-MYC-sen1-1 (klTRP1) GAL1-UBR1 (HISMX) leu2-3,112::GAL1-TAP-SEN1 (2-2231) (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2451*MATα td-MYC-sen1-1 (klTRP1) GAL1-UBR1 (HISMX) leu2-3,112::GAL1-TAP-SEN1 (931-2231) (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2458*MATa/MATα SEN1/SEN1 (931-2231) (HISMX)*	This study	N/A
S. cerevisiae (from W303) CS2582*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (931-2231) (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2584*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2603*MATa leu2-3,112::GAL1-TAP-SEN1 (2-931) (LEU2) pep4Δ::ADE2 ctf4Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS2607*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) W773A E774A W777A (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2609*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) D850A E851G V852A L853G L854A (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2611*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) V858A R859A I862A (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2615*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) D876G D877G V880G (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2617*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) V746G D747G P748G I749G (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2623*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) L656A S657A K658A I659A L660 (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2636*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) L656A S657A K658A I659A L660A (LEU2) NRD1- 9MYC (HIS3MX) pep4Δ:: ADE2 TAP-MCM3 (kanMX)*	This study	N/A
S. cerevisiae (from W303) CS2638*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2- 2231)W773A E774A W777A (LEU2) NRD1-9MYC (HIS3MX) pep4Δ:: ADE2 TAP-MCM3 (kanMX)*	This study	N/A
S. cerevisiae (from W303) CS2640*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2- 2231) D850A E851G V852A L853G L854A (LEU2) NRD1- 9MYC (HIS3MX) pep4Δ:: ADE2 TAP-MCM3 (kanMX)*	This study	N/A
S. cerevisiae (from W303) CS2642*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2- 2231) V746G D747G P748G I749G (LEU2) NRD1-9MYC (HIS3MX) pep4Δ:: ADE2 TAP-MCM3 (kanMX)*	This study	N/A
S. cerevisiae (from W303) CS2669*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) (LEU2) NRD1-9MYC (HIS3MX) pep4Δ:: ADE2 TAP-MCM3 (kanMX)*	This study	N/A
S. cerevisiae (from W303) CS2670*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2- 2231) (LEU2) NRD1-9MYC (HIS3MX) pep4Δ:: ADE2*	This study	N/A
S. cerevisiae (from W303) CS2716*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) D876G D877G V880G (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS2718*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) T782G I783G Y784G (LEU2)*	This study	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
S. cerevisiae (from W303) CS2734*MATa rnh1Δ:: hphNT rnh201Δ::HISMX*	Lab strain	N/A
S. cerevisiae (from W303) CS2735*MATα rnh1Δ:: hphNT rnh201Δ::HISMX*	Lab strain	N/A
S. cerevisiae (from W303) CS2791*MATa td-sld3-7 (kanMX) GAL1-UBR1 (HIS3MX) leu2-3,112::GAL1-TAP-SEN1 (2-931) (LEU2+) pep4Δ:: ADE2*	This study	N/A
S. cerevisiae (from W303) CS2808*MATa SEN1-TAP (kanMX)*	This study	N/A
S. cerevisiae (from W303) CS2810*MATa sen1-3-TAP (kanMX)*	This study	N/A
S. cerevisiae (from W303) CS2853*MATa SEN1-TAP (kanMX) pep4Δ:: ADE2*	This study	N/A
S. cerevisiae (from W303) CS2854*MATa sen1-3-TAP (kanMX) pep4Δ:: ADE2*	This study	N/A
S. cerevisiae (from W303) CS2859*MATa SEN1-TAP (kanMX) pep4Δ:: URA3 mrc1Δ::hphNT*	This study	N/A
C *S. cerevisiae* (from W303) S2861*MATa sen1-3-TAP (kanMX) pep4Δ:: URA3 mrc1Δ::hphNT*	This study	N/A
S. cerevisiae (from W303) CS2903*MATa td-sld3-7 (kanMX) GAL1-UBR1 (HIS3MX) leu2-3,112::GAL1-TAP-SEN1 (2-931) (LEU2) pep4Δ:: ADE2 ctf4Δ:: kanMX*	This study	N/A
S. cerevisiae (from W303) CS2938*MATα SEN1-TAP (kanMX) hpr1Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS2941*MATα sen1-3-TAP (kanMX) hpr1Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS2945*MATa SEN1-TAP (kanMX) sml1Δ::HISMX rad53Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS2947*MATa sen1-3-TAP (kanMX) sml1Δ::HISMX rad53Δ::ADE2*	This study	N/A
S. cerevisiae (from W303) CS2955*MATa SEN1-TAP (kanMX) ctf18Δ::klTRP1*	This study	N/A
S. cerevisiae (from W303) CS2957*MATa sen1-3-TAP (kanMX) ctf18Δ::klTRP1*	This study	N/A
S. cerevisiae (from W303) CS3167*MATa leu2-3,112::GAL1-TAP-SEN1 (2-931) (LEU2) pep4Δ::ADE2 psf1-1 (ts)*	This study	N/A
S. cerevisiae (from W303) CS3186*MATa leu2-3,112::GAL1-TAP-SEN1 (2-931) (LEU2) pep4Δ::ADE2 mrc1Δ::hphNT*	This study	N/A
S. cerevisiae (from W303) CS3321*MATα leu2-3,112::GAL1-RNH1 (2-348) (LEU2) SEN1-TAP (kanMX) mrc1Δ::hphNT*	This study	N/A
S. cerevisiae (from W303) CS3322*MATa leu2-3,112::GAL1-RNH1 (2-348) (LEU2) sen1-3-TAP (kanMX) mrc1Δ::hphNT*	This study	N/A
S. cerevisiae (from W303) CS3499*MATa SEN1-TAP (kanMX) pep4Δ::ADE2 ctf4Δ::kanMX mrc1-3IAA (HISMX) ADH1-OsTIR1 (klTRP1,URA3)*	This study	N/A
S. cerevisiae (from W303) CS3545*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-sen1-3 (2-2231) (LEU2) rnh1Δ:: hphNT rnh201Δ::HISMX*	This study	N/A
S. cerevisiae (from W303) CS3547*MATa sen1Δ::URA3-CP leu2-3,112::ACT1-3HA-SEN1 (2-2231) (LEU2) rnh1Δ:: hphNT rnh201Δ::HISMX*	This study	N/A
S. cerevisiae (from W303) CS3562*MATa SEN1-TAP (kanMX) pep4Δ::ADE2 ctf4Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS3662*MATa SEN1-TAP (kanMX) mrc1Δ::hphNT leu2-3,112::GAL1-RNH1 (2-348) (LEU2+)*	This study	N/A
S. cerevisiae (from W303) CS3664*MATa sen1-3-TAP (kanMX) mrc1Δ::hphNT leu2-3,112::GAL1-RNH1 (2-348) (LEU2)*	This study	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
S. cerevisiae (from W303) CS3702*MATa TAP-SLD5 (kanMX) SEN1-9MYC (hphNT) pep4Δ::URA3 ADE2 ctf4Δ::kanMX mrc1-3IAA (HISMX) ADH1-OsTIR1 (klTRP1,URA3)*	This study	N/A
S. cerevisiae (from W303) CS3731*MATα SEN1-TAP (kanMX) leu2-3,112::GAL1-RNH1 (2-348) (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS3733*MATa sen1-3-TAP (kanMX) leu2-3,112::GAL1-RNH1 (2-348) (LEU2)*	This study	N/A
S. cerevisiae (from W303) CS3796*MATa SEN1-TAP (kanMX) mad2Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS3797*MATα sen1-3-TAP (kanMX) mad2Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS3903*MATa leu2-3,112::GAL1-RNH1 (2-348) (LEU2) SEN1-TAP (kanMX) hpr1Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS3905*MATa leu2-3,112::GAL1-RNH1 (2-348) (LEU2) sen1-3-TAP (kanMX) hpr1Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS4296*MATa SEN1-TAP (kanMX) chl1Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS4298*MATα sen1-3-TAP (kanMX) chl1Δ::kanMX*	This study	N/A
S. cerevisiae (from W303) CS4312*MATa NRD1-TAP (kanMX) SEN1-9MYC (hphNT) pep4Δ::URA3-CP ADE2*	This study	N/A
S. cerevisiae (from W303) CS4314*MATa SEN1-TAP (kanMX) pep4Δ::ADE2 rpb1-1 (ts)*	This study	N/A
S. cerevisiae (from W303) DLY2057*MATa sen1-1 (ts)*	Lab strain	N/A
S. cerevisiae (from W303) DLY2281*MATa upf1Δ::TAP::klTRP1*	Lab strain	N/A
S. cerevisiae (from W303) DLY3111*MATa sen1-1 (ts) upf1Δ::TAP::klTRP1*	This study	N/A
S. cerevisiae (from W303) DLY3190*MATa SEN1-TAP (kanMX) upf1Δ::TAP::klTRP1*	This study	N/A
S. cerevisiae (from W303) DLY3191*MATa sen1-3-TAP (kanMX) upf1Δ::TAP::klTRP1*	This study	N/A
Oligonucleotides		
DL377ATGTTCCCAGGTATTGCCGA	This study	N/A
DL378ACACTTGTGGTGAACGATAG	This study	N/A
DL474GCAAAGATCTGTATGAAAGG	This study	N/A
DL475CGCAGAGTTCTTACCAAACG	This study	N/A
DL481TAAATGGCCAACCGCTGTTG	This study	N/A
DL482CCAGCGTACTGCACGCCAGG	This study	N/A
DL1119AAGTGACGAAGTTCATGCTA	This study	N/A
DL1120TCCGTGTCTCTTGTCCTGCA	This study	N/A
Recombinant DNA		
pYM-N24	Janke et al., 2004	Euroscarf
pCS14pRS305-*GAL1-TAP-Ø*	This study	N/A
pCS25pRS305-*GAL1-3HA-Ø*	This study	N/A
pCS26pRS305-*GAL1-3HA-SEN1 (2-931)*	This study	N/A
pCS30pRS305-*GAL1-TAP-SEN1 (2-931)*	This study	N/A
pCS31pRS305-*GAL1-TAP-SEN1 (2-1103)*	This study	N/A
pCS32pRS305-*GAL1-TAP-SEN1 (931-2231)*	This study	N/A
pCS33pRS305-*GAL1-TAP-SEN1 (1095-2231)*	This study	N/A
pCS39pRS305-*GAL1-TAP-SEN1 (2-2231)*	This study	N/A
pCS40pRS305-*GAL1-TAP-SEN1 (2-1901)*	This study	N/A
pCS42pRS305-*GAL1-3HA-SEN1 (2-622)*	This study	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pCS43pRS305-*GAL1-3HA-SEN1 (410-931)*	This study	N/A
pCS59pRS305-*GAL1-3HA-SEN1 (501-931)*	This study	N/A
pCS61pRS305-*GAL1-3HA-SEN1 (622-931)*	This study	N/A
pCS118pRS305-*ACT1-3HA-SEN1 (931-2231)*	This study	N/A
pCS120pRS305-*ACT1-3HA-SEN1 (2-2231)*	This study	N/A
pCS123pRS305-*ACT1-3HA-SEN1 (2-2231) W773A E774A W777A*	This study	N/A
pCS124pRS305-*ACT1-3HA-SEN1 (2-2231) L656A S657A K658A I659A L660A*	This study	N/A
pCS125pRS305-*ACT1-3HA-SEN1 (2-2231) D850A E851G V852A L853G L854A*	This study	N/A
pCS127pRS305-*ACT1-3HA-SEN1 (2-2231) D876G D877G V880G*	This study	N/A
pCS128pRS305-*ACT1-3HA-SEN1 (2-2231) V746G D747G P748G I749G*	This study	N/A
pCS129pRS305-*ACT1-3HA-SEN1 (2-2231) T782G I783G Y784G*	This study	N/A
pCS188pRS305-*GAL1-RNH1 (2-348)*	This study	N/A
pCS196pRS424-*GPD-hsRNASEH1* (2-286)	From Palancade's lab	N/A
pCS197pRS315-*ADE2*	This study	N/A
pCS198pRS315-*ADE2-ARS306*	This study	N/A
pLpRS316-*leu2Δ3′-39bp-leu2Δ5′*	From Aguilera's lab	N/A
pLYΔNSpRS316-*leu2Δ3′-3900bp-leu2Δ5′*	From Aguilera's lab	N/A
Software and Algorithms		
Excel	Microsoft	RRID:SCR_016137
Illustrator	Adobe	RRID:SCR_014198
ImageJ	NIH	https://imagej.nih.gov/ij/; RRID:SCR_003070
Photoshop	Adobe	RRID:SCR_014199
PredictProtein.org		https://www.predictprotein.org/
RStudio	RStudio	RRID:SCR_000432

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed and will be fulfilled by the Lead Contact, Dr Giacomo De Piccoli (g.de-piccoli@warwick.ac.uk). All unique/stable reagents generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Saccharomyces cerevisiae is the experimental model used in this study. All strains are isogenic to W303, and are listed in the Key Resources Table.

METHOD DETAILS

Yeast Strains and Growth Conditions

All yeasts were grown in YP medium supplemented with either glucose (YPD) or galactose (YPGAL) or raffinose (YPRAF) to a final concentration of 2% (w/v). For solid media, the same formulation was used, but with a final concentration of 1% (w/v) agar. Yeasts were grown at 24, 28, 30 and 37°C, depending on their viability at the different temperatures and as required by the experimental design. For all experiments, the control and test strains were subjected to the same conditions, including temperature.

For cell spotting experiments, cells were grown on non-selective media until colonies were judged to be sufficiently big. Five discrete colonies from individual strains were added to sterile deionised water to create a cell suspension. From this suspension, serial dilutions (0.5 x10^6, 0.5 x10^5, 0.5 x10^4 and 0.5 x10^3 cells/ml) were generated. 10 μL of each suspension was pipetted onto the appropriate media and grown for up to 5 days at the required temperatures.

To assess the genetic interaction between two or three genes, parents carrying the appropriate alleles were first crossed. Analysis of the meiotic progeny was conducted by inducing sporulation of the diploid strains in sporulation medium for 3-5 days at 24°C. Asci

were treated with a 1:10 dilution of β-glucoronidase from *Helix pomatia* (Sigma) for 30 minutes, followed by tetrad dissection onto a YPD plate using a Singer MSM400 micromanipulator. Plates were incubated for 3-4 days at the appropriate temperature.

For the plasmid recombination assay, eight independent clones carrying the appropriate plasmid (pL or pLYΔNS) were each plated in medium lacking leucine (to select for recombination) or lacking uracil (marker for the presence of the plasmid) at 24°C. The experiment was repeated in triplicate. For plasmid loss assays, cells were transformed with the required plasmid (pCS197 or pCS198) and plated on minimum medium lacking leucine and incubated at 24°C. Colonies were left to grow until single isolated colonies were sufficiently big. Five to seven colonies for each strain were then picked, resuspended in sterile water and counted. Around 200-150 cells were then plated onto YPD and incubated at 24°C until red/white coloring was clearly visible. Cells were then incubated at 4°C for three days. We considered white and sectored colonies as white while only fully red colonies were scored as red. The experiment was repeated twice. The plasmid loss assay with or without *GAL-RNH1* was conducted in a similar manner, except that cells were grown and transformed in medium containing galactose and selected in medium lacking adenine (*LEU2* is the reporter gene for the *GAL1-RNH1* construct). Colonies were grown for longer periods of time before colonies were sufficient size big and were plated onto non-selective medium containing galactose.

Cell cycle experiments

Cells were diluted from an inoculum to a density of 0.35×10^7 cells/ml in a suitable volume and left to grow to a final density of 0.7×10^7 cells/ml. The cells were then synchronized in G1 by adding α-factor to a final concentration of 7.5 μg/ml. After the first 90 min, α-factor was added every 30 min to a 3.25 μg/ml final concentration to maintain the cells in G1. When the cultures were shifted to 37°C, cells were spun down and resuspended in pre-warmed medium containing 7.5 μg/ml α-factor. Cells were released from the arrest by washing the cells twice with medium without α-factor. In all experiments in which cells were collected for IPs, cells were grown at 24°C and released into S phase for 30 min, unless stated otherwise in the figure legend. For expressing constructs under the *GAL1* promoter, strains were grown in YPRAF, arrested in G1 using α-factor, upon which YPGAL was substituted for YPRAF. Alternatively, YPGAL was used throughout the experiment (appropriate for constructs that were labile).

Harvesting cells for IP

Harvested cells were immediately cooled to 4°C by washing with an ice-cold solution of HEPES-KOH (pH 7.9), followed by a wash in a solution of 100 mM HEPES-KOH (pH 7.9), 50 mM potassium acetate, 10 mM magnesium acetate and 2 mM EDTA-KOH, still at 4°C. After the wash, the solution was discarded and the cells were re-suspended in a fresh quantity of the same solution supplemented with protease and phosphatase inhibitors, so that the ratio of wet mass of the cells to the final mass of the suspension was either 1:4 (for 250 mL cultures) or 4:5 (for 1 l cultures). The re-suspended cells were immediately flash-frozen by pipetting into a flask holding liquid nitrogen. The frozen cells were kept at −80°C until use for IP. Before freezing, some cells were fixed in 70% (v/v) ethanol to test that cells did not progress through the cell cycle during sample preparation.

For cells with inducible constructs, cultures were grown as described above in YPRAF. After the cells were arrested in G1, the culture was substituted with YPGAL (supplemented with α-factor) to induce transcription from the *GAL1* promoter. Harvesting of G1 cultures can be performed prior to or after induction according to the experimental setup. After 35 min or 1 h of induction, the cells were released in S phase as described above and harvested either 30 min (24°C) or 20 min (30°C or 37°C) post-release. For temperature-sensitive strains or strains tagged with a temperature-degron (e.g *psf1-1*, *td-sld3-7* and *rnh1Δ rnh201Δ ACT1-sen1-3*), the strains were grown and synchronized in G1 at 24°C as described above. Once synchronized and, (optionally) constructs transcriptionally induced, the cells were shifted to 37°C for 1 h. α-factor was added every 20 min to maintain the cells in G1 to a final concentration of 7.5 μg/ml for 1 h. Synchronicity was monitored visually using a microscope and by harvesting a 1 mL sample of the culture by fixing in 70% (v/v) ethanol for flow-cytometric analysis. The cells were then washed and released in S phase. The cells were harvested 20 min after release, including for the *psf1-1* strains that do not actually undergo DNA replication at 37°C as the GINS complex is destabilized. For crosslinking IPs, cells cultures were incubated with formaldehyde for 25 min and treated as in De Piccoli et al. (2012).

Western Blots

Protein samples (TCA-precipitated and non-treated cell extracts, as well as IPs) were run on 5, 6, 7, 8 or 10% polyacrylamide gels. The protein bands were then transferred onto nitrocellulose or PVDF membranes. The proteins bands were then probed with the appropriate primary antibodies for 1 h in a solution of 5% (w/v) skimmed milk in TBST, washed thrice for 5 min in fresh TBST, probed with the appropriate HRP-bound secondary antibody (if any, refer to Key Resources Table) and washed thrice again for 5 min in fresh TBST. The membrane was then treated with the western blotting reagents and the resulting chemiluminescent signal was captured using either films or a digital camera (G:BOX, Chemi XRG, Syngene).

IP

IPs were conducted as previously described (De Piccoli et al., 2012). In brief, cells previously harvested were lysed using a mechanised pestle and mortar at −80°C (Spex Sample Prep, 6870). 1 g of lysate is considered equivalent to 1 ml. To 1 volume of thawed lysate, ¼ volume of a solution of 50% (v/v) glycerol, 100 mM HEPES-KOH (pH 7.9), 50 mM potassium acetate, 50 mM magnesium acetate, 0.5% (v/v) Igepal® CA-630, 2mM EDTA supplemented with protease and phosphatase inhibitors was added. Pierce

Universal Nuclease was added to a final concentration of 0.4 U/μl and samples were left on a rotating platform at 4°C for 30 min. After incubation, the sample was clarified by stepwise centrifugation at 18,700*g* and then at 126,600*g*. The supernatant was isolated, 50 μL of which was added to 100 μL of 1.5 x Laemmli buffer (cell extract). The remaining cell extract was incubated with 100 μL of TAP-beads for 2 h (M-270 Dynabeads® Epoxy beads bound to an anti-sheep IgG). Beads were washed with solutions of 100 mM HEPES-KOH (pH 7.9), 50 mM potassium acetate, 50 mM magnesium acetate, 2 mM EDTA, 0.1% (v/v) Igepal® CA-630 thrice. After washing, 100 μL of 1 x Laemmli buffer was added to the 100 μL of TAP-beads and boiled for 4 min. Crosslinking IPs were conducted as in De Piccoli et al. (2012).

When scaling up was necessary (using 1 l of cells instead of 250 ml), a few changes were implemented to the protocol. Notably, the concentration of the Pierce Universal Nuclease was increased four-fold to a final concentration of 1.6 U/μl and incubation with the nuclease was increased from 30 to 40 min.

MS Analysis of IPs

The samples were processed as above. Following the washes, TAP-Sen1 (2-931) protein was released by using the AcTEV® protease at 24°C for 2 h. Thereafter, the resultant CBP-Sen1 (2-931) (CBP: calmodulin-binding protein) and its specific interactors were incubated with pre-washed calmodulin beads at 4°C for 2 h. After washing, 30 μL of 1 X Laemmli was added to the calmodulin beads and boiled for 4 min. The samples from the four biological replicates were pooled together, flash-frozen on dry ice and stored at −80°C. The samples were then run on commercially sourced 4%–12% acrylamide gel for a short distance (~1 cm). The gel was then cut in thin slices and processed and analyzed by MS Bioworks, USA.

Counting of Rad52-foci to assess DNA damage

Cells carrying the *RAD52-GFP* allele were first grown in liquid medium and synchronized in G1. Cells were released and harvested at different times after release, corresponding to different phases of the cell cycle. Paraformaldehyde was added to the cell suspensions to a final concentration of 3% (w/v) and the samples were incubated at room temperature for 10 min. The cells were then washed with PBS at room temperature. Finally, the samples were re-suspended in fresh PBS and kept at 4°C overnight.

Less than 24 h after fixation (to minimize signal lost due to alteration of the GFP protein), the samples were re-suspended in 500 μL of fresh PBS to which the DNA stain DAPI was added to a final concentration of 1 μg/ml. The samples were incubated at room temperature for 10 min to allow for staining of the DNA. The cells were then washed with PBS to improve the signal-to-noise ratio of the DAPI staining. The cells were then brought to a suitable dilution prior to pipetting on a glass slide onto which a coverslip is applied. Images of cells were acquired (brightfield, ~510 nm emission (GFP), ~460 nm (DAPI)) using a Personal DeltaVision (Applied Precision). The images were analyzed using ImageJ and the number of Rad52-foci were counted. An average of three experiments is shown in the figures.

Chromosome spreads and microscopy

Chromosome spreads were performed as previously described (Wahba et al., 2011; Grubb et al., 2015). Exponentially growing asynchronous cultures were grown in YPD at 28°C. 2x10^8 cells were harvested and spheroplasted (0.1M potassium phosphate (pH 7.4), 1.2 M sorbitol, 0.5 mM MgCl2, 40 mM DTT, 20 U zymolyase at 30°C for 1 h or until > 90% of cells lysed following addition of 2% sarcosyl. Cells were then washed and resuspended in ice cold 1 M sorbitol (pH 6.4), 0.1 M MES, 0.5 mM MgCl2, 1 mM EDTA to stop spheroplasting reaction. 20 μL of cell suspension was placed onto a slide, followed by 40 μL of fixative (4% paraformaldehyde (w/v), 3.4% sucrose (w/v)), then lysed using 80 μL of 1% lipsol (v/v) for 2 min, followed by addition of 80 μL of fixative and spread across the surface of the slide to dry overnight. Slides pre-treated with RNase H were incubated with 4U of RNase H diluted in 400 μL of 5mg/ml BSA for 1 h at 37°C prior to immunostaining. Slides were immunostained for DNA:RNA hybrids using mouse monoclonal antibody S.96 (Kerafast) diluted 1:2000 (0.25 μg/ml) in blocking buffer (5% BSA, 0.2% milk, 1XPBS) for 1 h. The secondary antibody, Cy3-conjucated goat anti-mouse (Jackson laboratories) was diluted 1:2000 in blocking buffer and incubated in the dark for 1 h. Indirect immunofluorescence was observed using a Deltavision 1 microscope with a 100 × /NA 1.4 objective. Image analysis was performed using ImageJ and Adobe Photoshop. An average of three experiments is shown in the figures.

Quantification of R-loops

Cells growing in liquid culture was harvested and re-suspended in lysis solution (100 mM NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 3% (w/v) SDS). To a volume of cell suspension, an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) (Acros Organics) and another volume of nuclease-free deionised water were added. The cells were then lysed mechanically using glass beads and DNA was isolated by incubating the soluble cell extract to ethanol to a final concentration of 70% (v/v). The DNA was washed with fresh ethanol and re-suspended in nuclease-free TE supplemented with 50 μg/ml RNase A and incubated at 37°C for 1 h only.

The concentration of genomic material was estimated by measuring absorbance at 260 nm. For each sample, 1 μg/μl, 0.5 μg/μl and 0.25 μg/μl dilutions of DNA was prepared, using nuclease-free water. 2 μL of each dilution was treated with either 1U of RNaseH (Invitrogen, #18021071) or 1 U of RNaseH and 1 U of RNase III (Invitrogen, #AM2290) with similar results at 37°C for 1h. As a control, untreated samples were also incubated at 37°C for 1 h. The remaining DNA was then added to 200 μL of 2 X SSC hybridization buffer (0.3M NaCl, 30mM trisodium Citrate, pH 7.0) and transferred to a pre-equilibrated hybond-N+ nylon membrane (GE healthcare, #RPN203B) under vacuum. The DNA was cross-linked to the membrane using UV prior to blocking in either 5% (w/v) milk (anti-R

loops) or 5% (w/v) BSA (anti ds-DNA) at 24°C for 1 h. The membranes were then incubated overnight in 5% milk supplemented the primary antibody at 4°C overnight. After thrice washing in TBST for 30 min, the membrane was incubated with anti-mouse IgG-HRP at 24°C for 1 h. The membranes were treated with ECL, and chemiluminescent signal was visualized using a camera (G:BOX, Chemi XRG, Syngene).

Reverse transcription followed by quantitative PCR

Cells were grown to exponential phase and incubated at permissive (24°C) or non-permissive temperatures (37°C) for 3 h to induce the *sen1-1* phenotype before collection. Analysis was performed in parallel in an *upf1Δ* background for detecting elongated RNA species derived from termination failure that might be degraded in the cytoplasm. The ratio of the read-through fraction over the total amount of *SNR13* RNA is shown as a proxy of transcription termination levels. The mean of three experiments is shown. Error bars represent the standard deviation. Cells were grown in logarithmic phase, and 6 OD_{600} worth of cells were pelleted. Total RNAs were extracted by resuspending cell pellets in 1 volume of acidic phenol (pH 4.3) supplemented with 1 volume of AES Buffer (50 mM Sodium Acetate pH 5.5, 10 mM EDTA, 1% SDS). Mixtures were incubated at 70°C with agitation (1,400 rpm) for 30 min in a thermomixer (Eppendorf), before being centrifuged at *20,000 g* at 4°C for 10 min. Aqueous phases were recovered and subjected to one extra round of hot acidic phenol extraction, followed by one round of chloroform extraction. Total RNAs were finally precipitated with absolute ethanol and sodium acetate pH 5.5, washed once with 70% Ethanol, dried on a SpeedVac (Thermo) and resuspended in 30 μL of RNase-free H_2O. 60-120 μg of total RNAs were recovered routinely.

Reverse transcription was performed using random hexamer-primers annealing at multiple *loci* in the *S. cerevisiae* genome and with oligos dT. 4 μg of total RNAs were mixed to 200 ng of random hexamers and 0.5 μM of oligos dT in a 20 μL reaction containing 50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM MgCl2 and 5 mM DTT. Samples were first incubated for 15 min at 70°C to allow RNA denaturation. Then temperature was slowly decreased to 37°C to allow annealing of primers. Lastly, synthesis of cDNAs was performed by adding 200 units of MLV-reverse transcriptase for 45 min at 37°C.

To assess the amount of cDNAs reverse transcribed, quantitative PCR (qPCR) was carried out using two different primer pairs for each target (*SNR13, NEL250c, ACT1*). These allowed the amplification of a product covering either ~300 bp of the 3′ end of *ACT1* (DL377/DL378 primer pair) or ~70 bp in the read-through region of *SNR13* (DL1119/DL1120 primer pair) or ~140 bp in the body of *NEL025c* (DL474/DL475 primer pair) or ~70 bp in the read-though region of *NEL025c* (DL481/DL482 primer pair). qPCR was performed in a 10 μL reaction by mixing 2 μL of the reverse transcribed cDNAs to 5 μL of LightCycler® 480 SYBR Green I Master and 2.5 pmol of both the forward and the reverse primer.

Cross-linking and analysis of cDNA (CRAC)

The CRAC protocol used in this study is derived from Granneman et al. (2009) with a few modifications as described in Candelli et al. (2018). Raw data processing has been performed as described in Candelli et al. (2018). Metagene analysis has been performed as follows: for the CUTs presented in Table S1, we retrieved the polymerase reads count at every position around the features (3′ or 5′ end) and plotted the mean over all the values for these positions in the final aggregate plot. Analysis has been performed in the R Studio environment.

QUANTIFICATION AND STATISTICAL ANALYSIS

Where applicable, data was presented as the average ± standard deviation. t tests were used to compare population means. Statistically significant differences were indicated as such by indicating the value range of the p values.

DATA AND CODE AVAILABILITY

The published article includes all datasets generated or analyzed during this study. The raw data of the metagene analysis of the CUTs shown in Figure 4B are included in Table S1. This study did not generate any unique code.

Cell Reports, Volume *30*

Supplemental Information

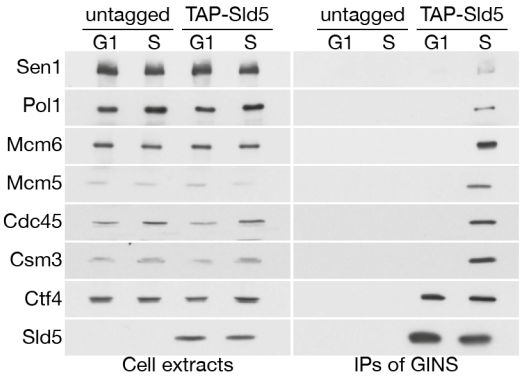
Sen1 Is Recruited to Replication Forks via Ctf4 and Mrc1 and Promotes Genome Stability

Rowin Appanah, Emma Claire Lones, Umberto Aiello, Domenico Libri, and Giacomo De Piccoli

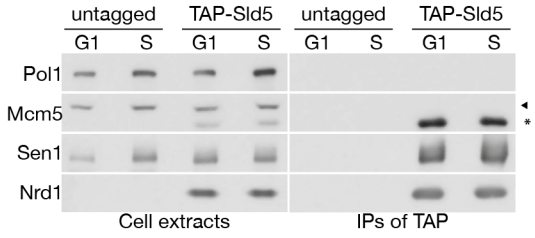
A)

Protein ID	MW (kDa)	TAGGED (peptide N)	UNTAGGED (peptide N)
CTF4	104	365	0
MCM4	115	123	0
POL2	256	97	0
SEN1	253	96	0
TOP1	90	85	0
DPB3	59	83	17
MCM3	108	61	0
SLD5	40	34	3
MCM6	113	34	0
DPB2	61	27	6
MRC1	124	25	0
MCM5	86	24	0
HHF1	11	22	0
TOF1	141	22	0

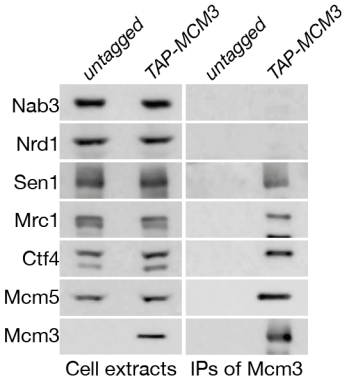
B)



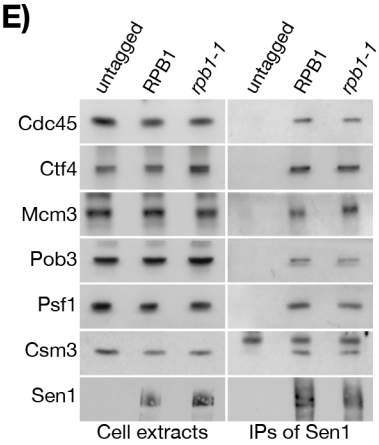
C)



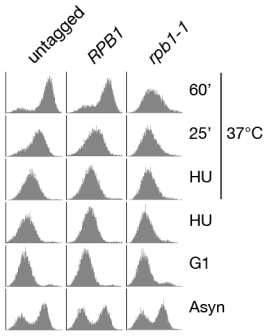
D)



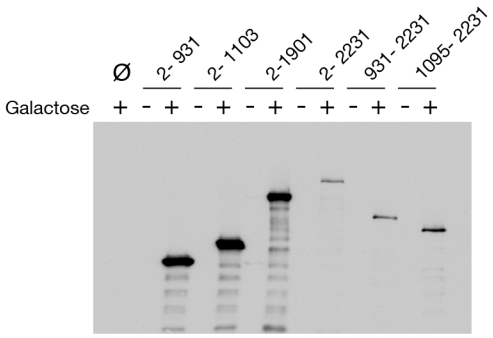
E)



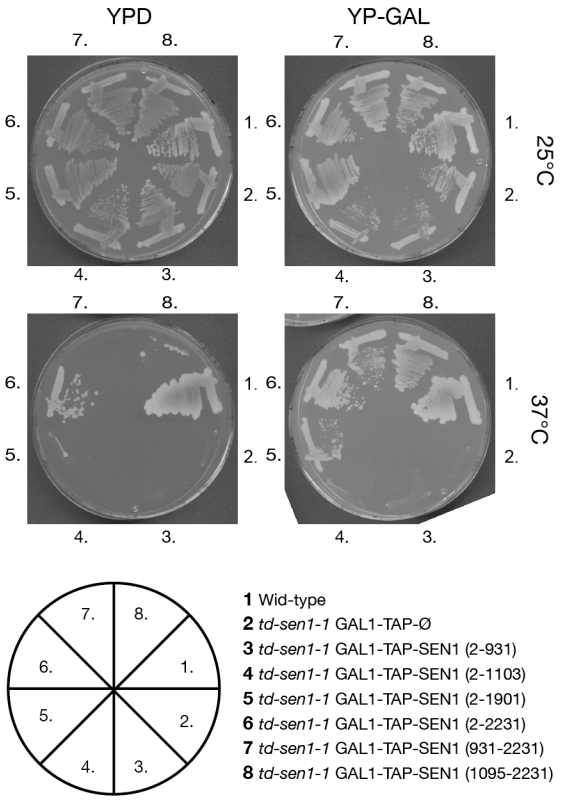
F)



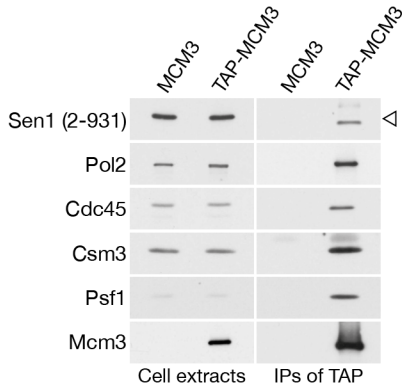
G)



H)



I)



J)

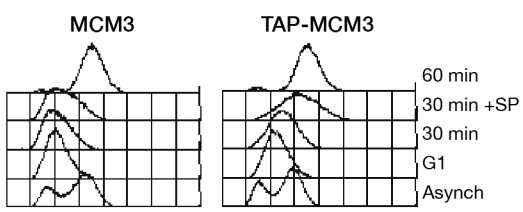


Figure S1. The replisome binds Sen1 in S phase (related to Fig 1)**. A)** Example of the mass spectrometry analysis obtained from the double purification of Sld5 and Mcm4. **B)** Cells carrying the *SEN1-9MYC* allele with a *SLD5* or *TAP-SLD5* allele were synchronously released from G1 into S phase for 30 min at 24°C. Cell extracts were incubated with anti-TAP beads and analysed by immunoblotting. **C)** Nrd1 does not interact with the replisome. *NRD1* or *NRD1-TAP* cells were released from G1 arrest into S phase for 30 min at 24°C. Cell extracts were incubated with anti-TAP beads and analysed by immunoblotting. **D)** Mcm3 immunoprecipitates Sen1 but neither Nrd1 nor Nab3. *MCM3* or *TAP-MCM3* cells were arrested in G1 and released into S phase for 30 min at 24˚C. Cell extracts were incubated with anti-TAP beads and analysed by immunoblotting. **E)** Sen1 interacts *in vivo* with the replisome independently of RNAPII transcription. Wild type or *rpb1-1* cells, either carrying an untagged or TAP-tagged allele of *SEN1*, were arrested in G1 and released in medium containing 0.2 M HU for 75 min at 24˚C. Cultures were then shifted to 37˚C for 1 h. Inactivation of *rpb1-1* cells at 37˚C for 1 h has been shown to lead to a substantial loss of Rpb1-1 from chromatin (Zanton and Pugh, 2006; Kim *et al.*, 2010), to a loss of elongation factors Spt5 and Spt16 (Tardiff, Abruzzi and Rosbash, 2007), to a loss of Sen1 recruitment at highly transcribed genes (Alzu *et al.*, 2012) and to the termination of transcription (Nonet *et al.*, 1987). Cells were then released for 25 min at 37˚C in fresh medium so to allow the synthesis of the bulk of the DNA. Cell extracts were incubated with anti-TAP beads and analysed by immunoblotting. **F)** FACS analysis of the experiment in **E)**. **G)** Cells carrying several different N-terminally tagged truncations of *SEN1* under the *GAL1* promoter were grown to exponential phase in YPRAF, divided in two cultures and transferred to either fresh YPRAF or to YPGAL for 2 h. Protein extracts were analysed by immunoblotting with an anti-HA antibody. **H)** (Top) Cells carrying a temperature-sensitive allele *td-sen1-1* and different fragments of *SEN1* under the *GAL1-3HA* promoter were plated, according to the schematic presented, on YPD or YPGAL and incubated at either 24°C or 37°C. (Bottom) schematic of the plated strains. **I)** The interaction of Sen1 (2-931) with TAP-Mcm3 during S phase is specific. Experiments carrying an untagged or a TAP-tagged allele of *MCM3* were conducted as in Fig 1C. **J)** FACS samples from the experiment are shown.

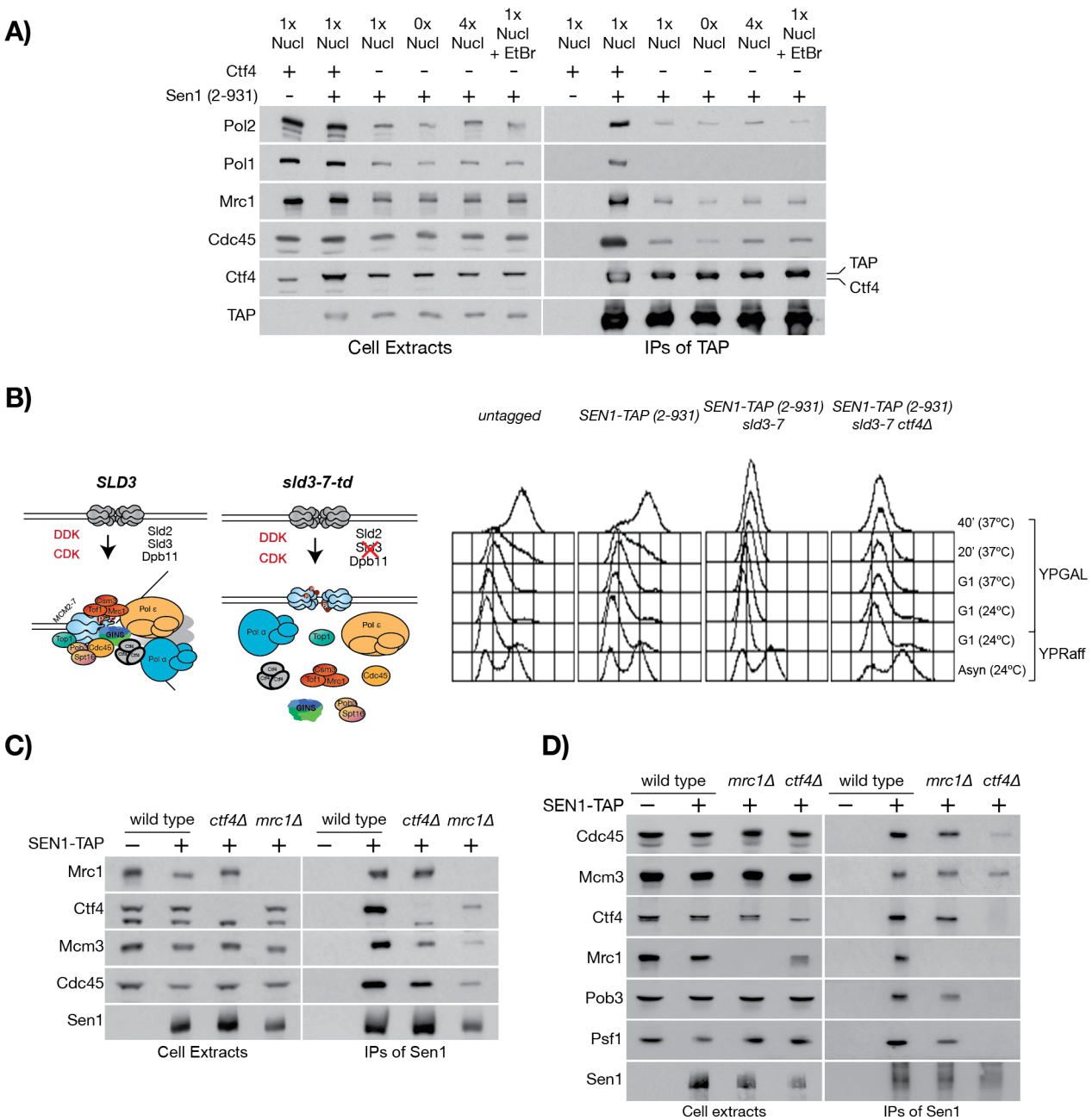

A)
1x Nucl
0x Nucl
4x Nucl
1x Nucl + EtBr
Ctf4
Sen1 (2-931)
Pol2
Pol1
Mrc1
Cdc45
Ctf4
TAP
TAP
Ctf4
Cell Extracts
IPs of TAP
B)
untagged
SEN1-TAP (2-931)
SEN1-TAP (2-931) sld3-7
SEN1-TAP (2-931) sld3-7 ctf4Δ
SLD3
sld3-7-td
DDK
CDK
Sld2
Sld3
Dpb11
40' (37°C)
20' (37°C)
G1 (37°C)
G1 (24°C)
G1 (24°C)
Asyn (24°C)
YPGAL
YPRaff
C)
wild type
ctf4Δ
mrc1Δ
SEN1-TAP
Mrc1
Ctf4
Mcm3
Cdc45
Sen1
Cell Extracts
IPs of Sen1
D)
wild type
mrc1Δ
ctf4Δ
SEN1-TAP
Cdc45
Mcm3
Ctf4
Mrc1
Pob3
Psf1
Sen1
Cell extracts
IPs of Sen1

Figure S2. **Mrc1 and Ctf4 mediate Sen1 binding to the replisome** (related to Fig 2). **A)** Cells carrying the *GAL1-TAP-SEN1 (2- 931)* construct or the empty control were grown in YPGAL, arrested in G1 phase using α-factor and released in S phase for 20 min at 30°C. The samples were then used for IPs using TAP beads and treated with the indicated amount of nuclease or ethidium bromide (50 μg/ml). Ctf4 and TAP-Sen1 (2-931) have similar sizes and run closely in gel electrophoresis. **B)** (Left) Schematic representation of the system used in the experiment shown in Fig 2D; (Right) FACS profile of the experiment conducted. Cells were grown in YPRAF at 24°C, arrested in G1 and either harvested, or resuspended in YPGAL at 24°C for 35 min to induce the expression of Sen1 (2-931) and Ubr1, shifted to 37°C for 1 h to inactivate/degrade td-Sld3-7 and then released in S phase for 20 min at 37°C. **C)** Wild type, *ctf4*Δ and *mrc1*Δ cells, carrying a TAP-tagged or untagged allele of *SEN1*, were arrested in G1 and synchronously released in S phase for 30 min at 24˚C. Cell extracts were incubated with anti-TAP beads and the immunoprecipitated material was analysed by immunoblotting. **D)** Immunoblotting analysis of cell extracts and IP material from anti-TAP beads. The experiment was conducted an in **C)**, except cultures were incubated with formaldehyde before collection.

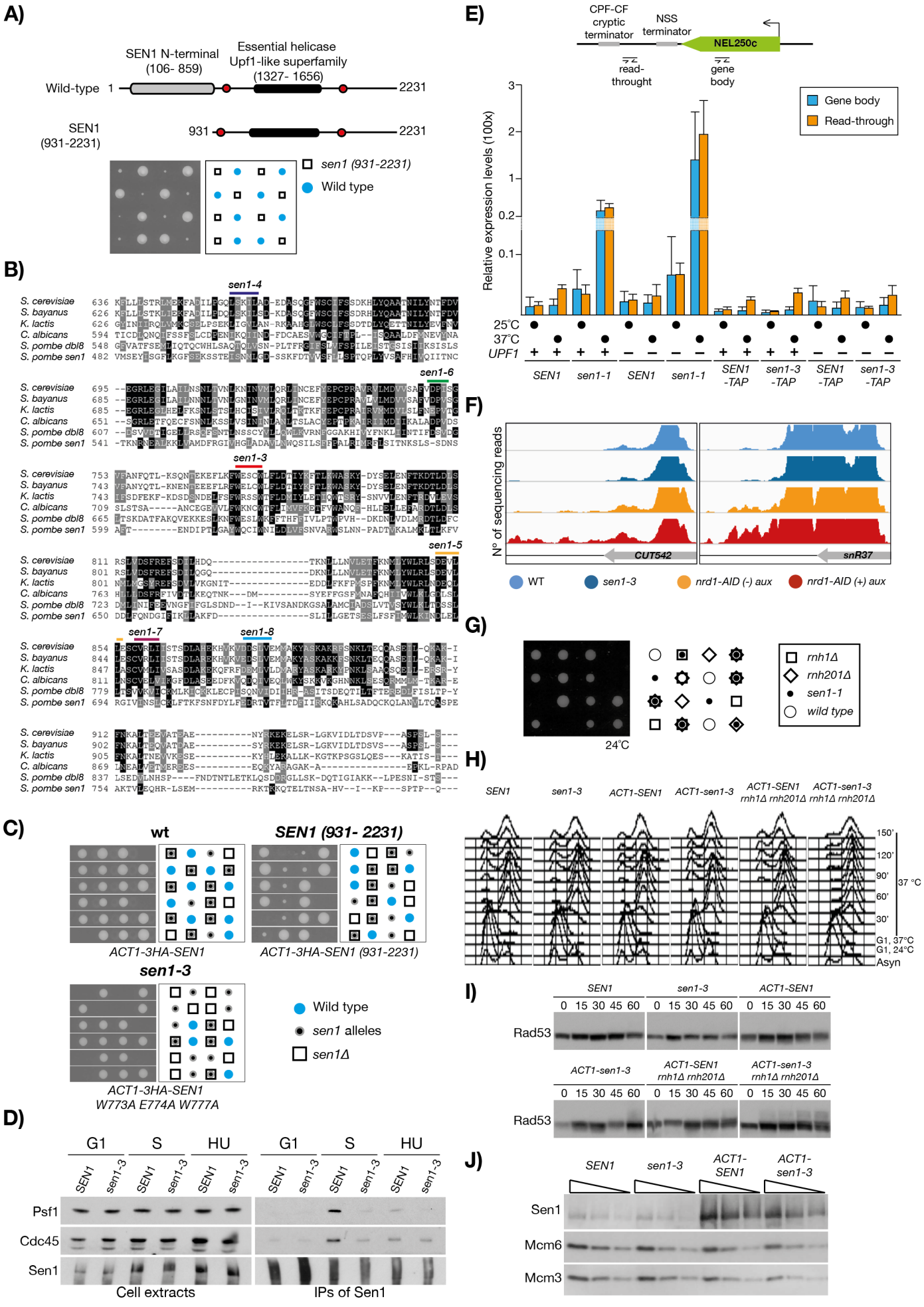

A)
SEN1 N-terminal (106- 859)
Essential helicase Upf1-like superfamily (1327- 1656)
Wild-type 1 2231
SEN1 (931-2231) 931 2231
sen1 (931-2231)
Wild type
B)
sen1-4
sen1-6
sen1-3
sen1-5
sen1-7
sen1-8
S. cerevisiae
S. bayanus
K. lactis
C. albicans
S. pombe dbl8
S. pombe sen1
C)
wt
SEN1 (931- 2231)
ACT1-3HA-SEN1
ACT1-3HA-SEN1 (931-2231)
sen1-3
ACT1-3HA-SEN1
W773A E774A W777A
Wild type
sen1 alleles
sen1Δ
D)
G1
S
HU
SEN1
sen1-3
Psf1
Cdc45
Sen1
Cell extracts
IPs of Sen1
E)
CPF-CF cryptic terminator
NSS terminator
NEL250c
read-throught
gene body
Gene body
Read-through
Relative expression levels (100x)
25°C
37°C
UPF1
SEN1
sen1-1
SEN1 -TAP
sen1-3 -TAP
F)
N° of sequencing reads
CUT542
snR37
WT
sen1-3
nrd1-AID (-) aux
nrd1-AID (+) aux
G)
24°C
rnh1Δ
rnh201Δ
sen1-1
wild type
H)
SEN1
sen1-3
ACT1-SEN1
ACT1-sen1-3
ACT1-SEN1 rnh1Δ rnh201Δ
ACT1-sen1-3 rnh1Δ rnh201Δ
150'
120'
90'
60'
30'
37 °C
G1, 37°C
G1, 24°C
Asyn
I)
SEN1
sen1-3
ACT1-SEN1
ACT1-sen1-3
ACT1-SEN1 rnh1Δ rnh201Δ
ACT1-sen1-3 rnh1Δ rnh201Δ
0 15 30 45 60
Rad53
J)
SEN1
sen1-3
ACT1-SEN1
ACT1-sen1-3
Sen1
Mcm6
Mcm3

Figure S3. The N-terminal of Sen1 is important for cell growth and is conserved in yeasts (related to Fig 3 and 4)**. A)** Tetrad analysis of a diploid yeast strain carrying the *SEN1/sen1 (1-931Δ)* alleles. Plates were imaged after 5 days of growth on YPD at 24°C. **B)** Alignment of *Saccharomyces cerevisiae* Sen1 domain interacting with the replisome (636-931) with its orthologues from *Saccharomyces bayanus*, *Kluyveromyces lactis*, *Candida albicans* and *Schizosaccharomyces pombe dbl8* and *sen1*. The mutations used in the screen were selected to mutate conserved amino acids predicted to be on the surface (www.predictprotein.org). **C)** Analysis of the *ACT1-3HA-SEN1* mutants. Tetrad analyses were conducted on diploids yeast strains carrying *SEN1*/*sen1*Δ, and ectopically integrated *ACT1-3HA-SEN1* alleles at the *leu2-3,112* locus. Plates were imaged after 3 days of growth on YPDA at 24°C. **D)** Sen1-3 show greatly reduced interaction with RFs. Wild type and *sen1-3* cells were arrested in G1 and synchronously released for 30 min in fresh medium (S) or for 90 min in medium containing 0.2 M HU. At the indicated times, cultures were treated with formaldehyde before being collected. The cross-linked cell extracts and the immunoprecipitated material from anti-TAP beads were analysed by immunoblotting**. E)** (Top) Schematic representation of the gene analysed and the probes used to assess defects in transcription termination; (Bottom) RT-qPCR analysis of RNAs derived from the indicated strains. *NEL025c* is a non-coding region as described in (Wyers *et al.*, 2005). Cells were grown to exponential phase and incubated for 3 h at the indicated temperature before being collected. The signal is presented as the expression level relative to the housekeeping gene *ACT1* (triplicate biological repeats). **F)** Snapshots illustrating RNA Pol II density detected by CRAC on two NNS complex targets (one CUT and one snoRNA) in the indicated strains. An *nrd1-AID* strain grown in the presence or absence of auxin is included as a control for transcription termination (dataset from (Candelli *et al.*, 2018)). **G)** *sen1-1* is lethal in the absence of *RNH201* and *RNH1*. Examples are shown of tetrad analyses conducted from yeast diploids strains with the *SEN1*/*sen1-1 RNH1/rnh1*Δ and *RNH201/rnh201*Δ genotype. Plates were imaged after 4 days of growth on YPDA at 24°C. **H)** FACS analysis of the cell cycle progression in cells *SEN1*, *sen1-3, ACT1-SEN1, ACT1-sen1-3, ACT1-SEN1 rnh1*Δ *rnh201*Δ and *ACT1-sen1-3 rnh1*Δ *rnh201*Δ. Cells were grown to the exponential phase at 24°C, arrested in G1, shifted to 37°C for 1 hour in G1, and released in S phase at 37°C. The samples were collected at the indicated time points. **I)** *ACT1-sen1-3 rnh1*Δ *rnh201*Δ cells show activation of Rad53 during DNA replication at 37°C. Western blot analysis of samples taken from the experiment shown in **H). J)** Analysis of the protein levels of the alleles *SEN1*, *sen1-3*, *ACT1-SEN1* and *ACT1-sen1-3*, all carrying a 3HA tag, and loading controls Mcm3 and Mcm6.

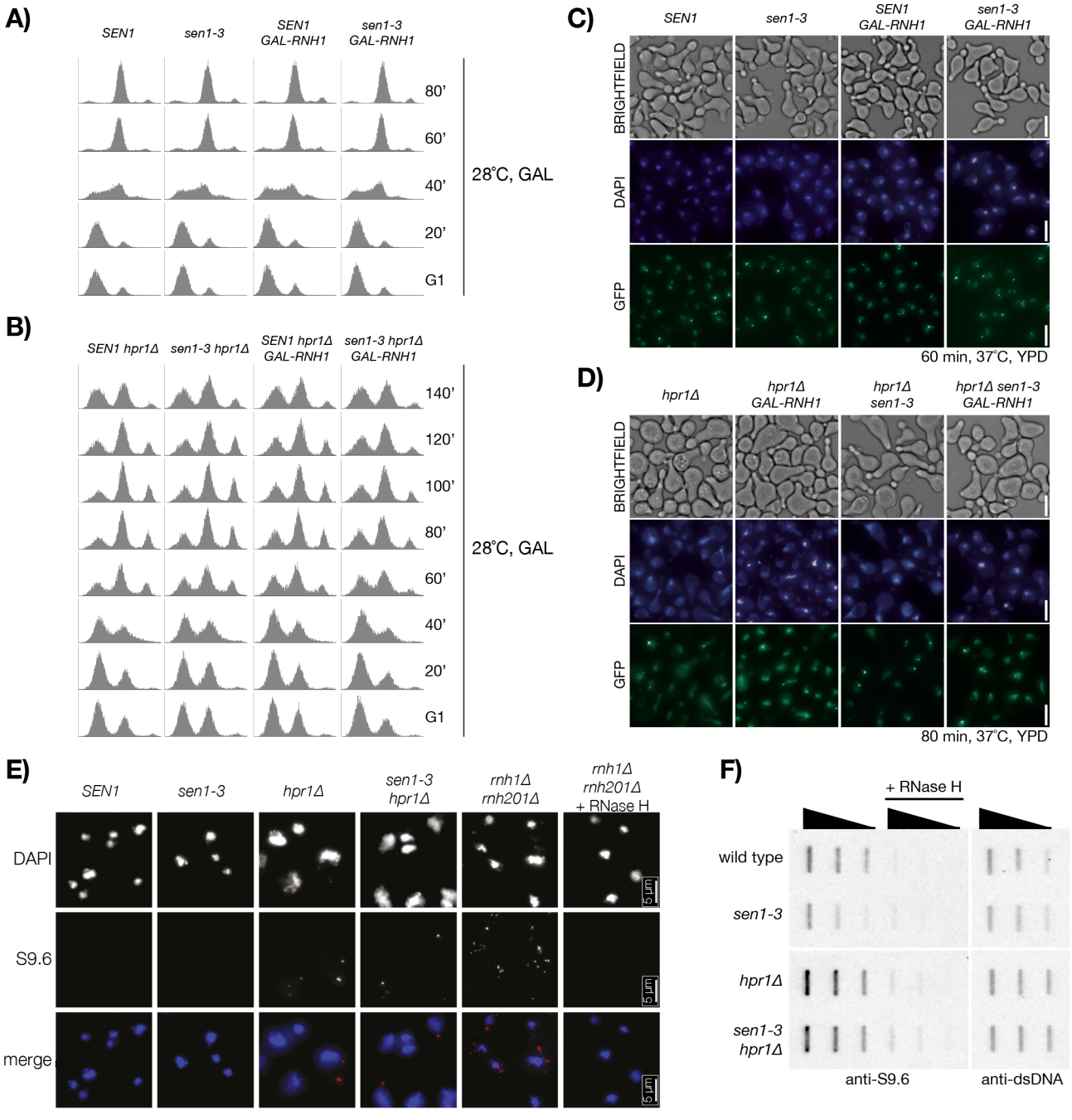

A)
SEN1
sen1-3
SEN1 GAL-RNH1
sen1-3 GAL-RNH1
80'
60'
40'
20'
G1
28°C, GAL
B)
SEN1 hpr1Δ
sen1-3 hpr1Δ
SEN1 hpr1Δ GAL-RNH1
sen1-3 hpr1Δ GAL-RNH1
140'
120'
100'
80'
60'
40'
20'
G1
28°C, GAL
C)
SEN1
sen1-3
SEN1 GAL-RNH1
sen1-3 GAL-RNH1
BRIGHTFIELD
DAPI
GFP
60 min, 37°C, YPD
D)
hpr1Δ
hpr1Δ GAL-RNH1
hpr1Δ sen1-3
hpr1Δ sen1-3 GAL-RNH1
BRIGHTFIELD
DAPI
GFP
80 min, 37°C, YPD
E)
SEN1
sen1-3
hpr1Δ
sen1-3 hpr1Δ
rnh1Δ rnh201Δ
rnh1Δ rnh201Δ + RNase H
DAPI
S9.6
merge
5 μm
F)
+ RNase H
wild type
sen1-3
hpr1Δ
sen1-3 hpr1Δ
anti-S9.6
anti-dsDNA

Figure S4. Analysis of the recombination and DNA:RNA hybrids in *hpr1*Δ and *hpr1*Δ *sen1-3* (related to Fig 4)**. A-D)** FACS analysis of the DNA replication dynamics and examples of the microscopy data of the experiments shown in Fig 4F are shown. Cells were grown to exponential phase in YPRAF at 28°C, arrested in G1, resuspended in YPGAL for 1 h and synchronously released in S phase in YPGAL (triplicate biological repeats). Scale bar = 5 μm **E)** Examples of the immunohistochemistry analysis of DNA:RNA shown in Fig 4G (triplicate biological repeats). Scale bar = 5 μm **F)** Analysis of R-loops *ex vivo***.** Cells were grown to exponential phase, arrested in G1 and then synchronously released in S phase for 30 min at 24°C. DNA:RNA hybrids double-stranded DNA were recovered in nuclease-free water and 1, 0.5 and 0.25 μg/μl dilutions of nucleic acid samples were prepared. The samples were then either treated with a commercially-sourced RNase H (or mock-treated), transferred onto nylon membrane and probed against using either the S9.6 antibody (that recognise R-loops) or an anti-dsDNA antibody.

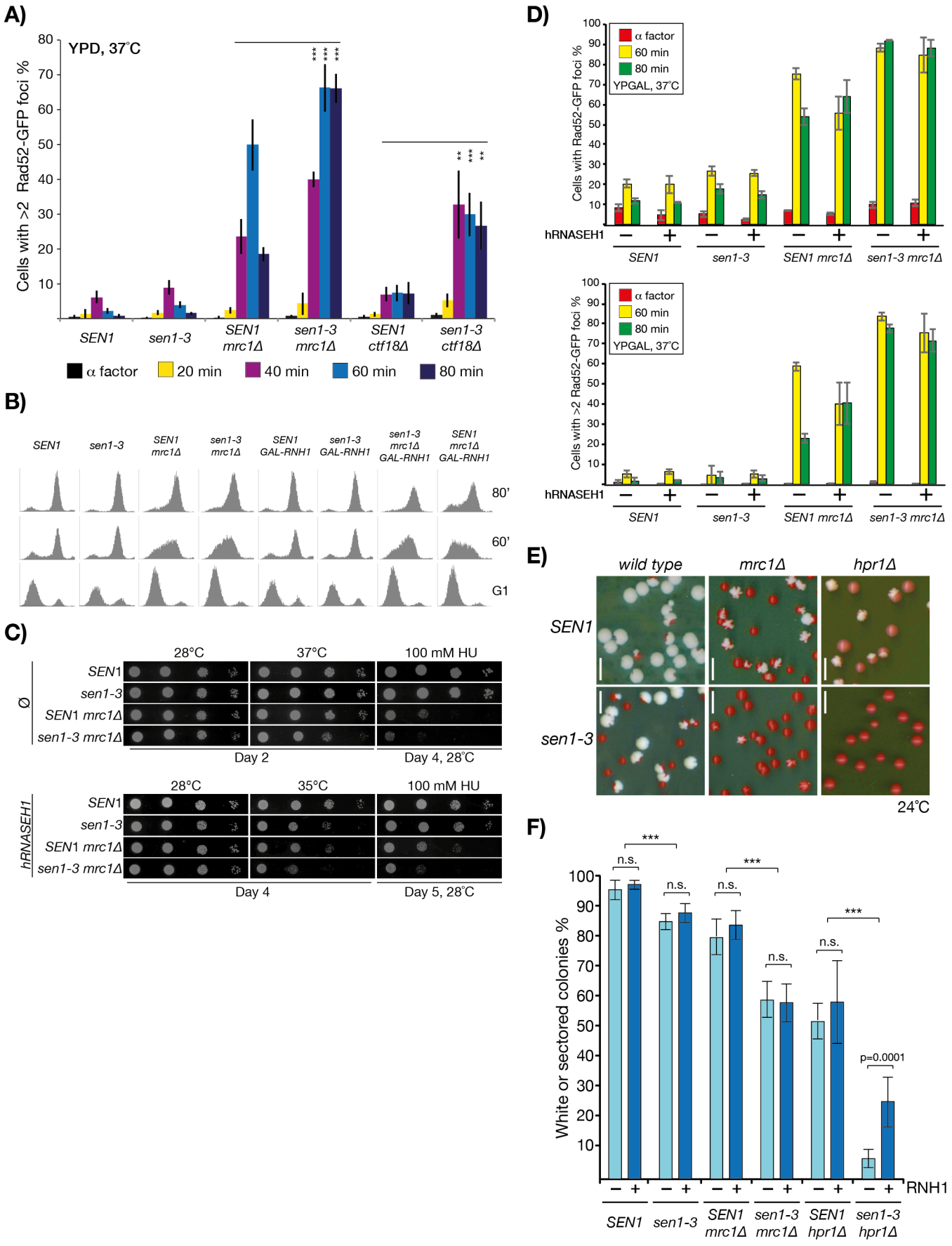

A)
YPD, 37°C
Cells with >2 Rad52-GFP foci %
SEN1
sen1-3
SEN1 mrc1Δ
sen1-3 mrc1Δ
SEN1 ctf18Δ
sen1-3 ctf18Δ
α factor
20 min
40 min
60 min
80 min
B)
SEN1
sen1-3
SEN1 mrc1Δ
sen1-3 mrc1Δ
SEN1 GAL-RNH1
sen1-3 GAL-RNH1
sen1-3 mrc1Δ GAL-RNH1
SEN1 mrc1Δ GAL-RNH1
80'
60'
G1
C)
28°C
37°C
100 mM HU
Ø
SEN1
sen1-3
SEN1 mrc1Δ
sen1-3 mrc1Δ
Day 2
Day 4, 28°C
28°C
35°C
100 mM HU
hRNASEH1
SEN1
sen1-3
SEN1 mrc1Δ
sen1-3 mrc1Δ
Day 4
Day 5, 28°C
D)
α factor
60 min
80 min
YPGAL, 37°C
Cells with Rad52-GFP foci %
hRNASEH1
SEN1
sen1-3
SEN1 mrc1Δ
sen1-3 mrc1Δ
E)
wild type
mrc1Δ
hpr1Δ
SEN1
sen1-3
24°C
F)
White or sectored colonies %
n.s.
***
p=0.0001
RNH1
SEN1
sen1-3
SEN1 mrc1Δ
sen1-3 mrc1Δ
SEN1 hpr1Δ
sen1-3 hpr1Δ

Figure S5. Overexpression of *RNH1* does not suppress the defects in *mrc1*Δ *sen1-3* (related to Fig 5)**. A)** Samples from experiments show in Fig 5D were scored for the presence of two or more foci per cell. ** p<0.05, *** p<0.01). **B)** FACS analysis of the experiment shown in Fig 5G. **C)** *hRNASEH1* overexpression does not suppress the defects observed in *sen1-3 mrc1*Δ cells. *hRNASEH1* was overexpressed from a 2 micron multicopy plasmid under the strong GPD^{TDH3} promoter. *SEN1*, *sen1-3*, *mrc1*Δ and *mrc1*Δ *sen1-3* cells were transformed with an empty or GPD^{TDH3}-*hRNASEH1* plasmid. Eight independent clones were pooled together and used for dilution spotting in medium lacking histidine, so to maintain the selective pressure for the plasmid. The cells carrying *GDP-hRNH1* grew more slowly and scans of their growth were taken at later times (Bottom panel). Serial dilution spotting (1:10) of the indicated strains is shown. **D)** *hRNASEH1* overexpression does not suppress the increase in recombination in *sen1-3 mrc1*Δ. Cell cultures were grown overnight at 24°C in medium lacking histidine to the exponential phase. Cells were diluted, resuspended in YPD, and left to grow for the length of one cell cycle. Cells were arrested in G1, shifted to 37°C for 1 h still with α-factor, and released in S phase at 37°C. Cells were taken at the indicated times, fixed, and analysed (triplicate biological repeats). **E)** Examples of the plasmid loss phenotype observed in the strains shown in Fig 5I (plasmid with 1 origin). Scale bar: 5 mm. **F)** The indicated strains, either carrying the *GAL1-RNH1* construct integrated at *leu2-3,112* or not, were transformed with the pRS315-*ADE2* plasmid. The experiment was performed as in Fig 5I, except that all media used contained galactose. n.s. = not significant, *** p<0.001.