January 2015

Structure-Function Analysis of the DNA Damage Repair Complex STR in Saccharomyces cerevisiae

Jessica Ashley Kennedy
University of South Florida, kennedy5@mail.usf.edu

Structure-Function Analysis of the DNA Damage Repair Complex STR in *Saccharomyces cerevisiae*

by

Jessica A. Kennedy

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Cell and Molecular Biology
Department of Cell Biology, Microbiology, and Molecular Biology
College of Arts and Sciences
University of South Florida

Co-Major Professor: Kristina Schmidt, Ph.D.
Co-Major Professor: Gary Daughdrill, Ph.D.
Stanley Stevens, Ph.D.
Patrick Bradshaw, Ph.D.

Date of Approval:
June 29, 2015

Keywords: RecQ Helicases, Intrinsic Disorder, Unstructured Proteins

DEDICATION

I would like to dedicate this dissertation to all those that played a hand in educating me to this point. It is only because of their dedication to education that I am the student I am today. I owe the faculty of Sanders Memorial Elementary, Pine View Middle, Land O' Lakes High, and University of South Florida for the tools and drive to reach this point. Thank you for challenging me and keeping me hungry.

I would also like to dedicate this to my parents, Frederick and Nancy Kennedy, and particularly to my sisters Erin Kennedy and Mary Ghazarian, who are the perfect partners in academic crime. Thank you for listening to my complaints on the most rough days of the bench. Having someone who knows the struggle to succeed matters (even if they don't really know what I do. C-3PO is not a gene).

Lastly, I dedicate at least a small part of this to Amy Morrison and Charles Slowik, who kept me fed throughout the writing process. Your stress-eating sessions of delicious food make the whole process just a little easier. You are the truest friends.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

4-NQO	4-nitroquinoline 1-oxide
ADR1	Alcohol dehydrogenase regulator
ALT	Alternative lengthening of telomeres
APC	Adenomatous polyposis coli
ATM	Ataxia telangiectasia mutated
ATP	Adenosine triphosphate
ATR	Ataxia telangiectasia and Rad3-related
BER	Base excision repair
BLM	Bloom
BRCA1	Breast cancer-associated protein 1
BS	Bloom's syndrome
CAN1	Canavanine resistance gene 1
CPT	Campthothecin
D.m.	Drosophilia melanogaster
D2O	Deuterated water
dHJ	Double Holliday junction
DNA	Deoxyribonucleic acid
DNA2	DNA synthesis defective protein 2
DOX	Doxorubicin
DSB	Double strand break
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EST2	Ever shorter telomeres 2
EXO1	Exonuclease 1
FANC	Fanconi anemia
FHA1	Forkhead-associated 1
GCR	Gross-chromosomal rearrangement
GST	Glutathione S-transferase
H.s.	Homo sapiens
HIS	Histidine
HJ	Holliday junction
HR	Homologous recombination
HRDC	RNaseD C-terminal
HSQC	Heteronuclear single quantum correlation

HU	Hydroxyurea
IDP/R	Intrinsically disordered protein/region
IPTG	Isopropyl-beta-D-thiogalactopyranoside
IR	Irradiation
LB	Luria Broth
LEU	Leucine
LiAc	Lithium acetate
LIG4	Ligase 4
LOH	Loss of heterozygosity
MLH1	MutL homolog 1
MMS	Methylmethanesulfonate
MoRF	Molecular Recognition Element
MRC1	Mediator of the replication checkpoint
MRE11	Meiotic recombination
MRN	Mre11-Rad50-Nbs1 protein complex
MRX	Mre11-Rad50-Xrs2 protein complex
MSH2	MutS homolog 2
MSH6	MutS homolog 6
mtDNA	Mitochondrial DNA
MYC	Myelocytomatosis gene
NEIL1	Nei endonuclease VIII-like
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
NHNOE	Heteronuclear Overhauser effect
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
OB	Oligonucleotide-binding
PARP1	Poly(ADP-ribose)polymerase 1
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PML	Promyelocytic leukemia
PVDF	Polyvinylidene fluoride
RAD16	Radiation (sensitive) protein 16
RAD2	Radiation (sensitive) protein 2
RAD50	Radiation (sensitive) protein 50
Rad51	Radiation (sensitive) protein 51
RAD52	Radiation (sensitive) protein 52
Rad53	Radiation (sensitive) protein 53
RAD54	Radiation (sensitive) protein 54
RAD9	Radiation (sensitive) protein 9
RecQL1	Human RecQ protein 1

RecQL4	Human RecQ protein 4
RecQL5	Human RecQ protein 5
RMI1	RecQ mediated genome instability
RNA	Ribonucleic acid
RPA	Replication protein A
RQC	RecQ C conserved
RQH1	RecQ helicase 1
RRM3	rDNA recombination mutation
RTR	RecQ-Top3-Rmi1 complex
RTS	Rothmund-Thomson syndrome
S.c.	Saccharomyces cerevisiae
SCE	Sister chromatid exchange
SE	Strand exchange
SF2	Super family 2
SGD	Saccharomyces genome database
SGS1	Slow-growth suppressor 1
SLD2	Synthetically lethal with Dpb11-1
SLX1	Synthetic lethal of unknown function 1
SLX4	Synthetic lethal of unknown function 4
SRS2	Suppressor of Rad6
SSA	Single strand annealing
SSB	Single-stranded binding
STR	Sgs1/Top3/Rmi1
ssDNA	Single-stranded DNA
TBST	Tris-buffered saline w/ Triton
TCA	Trichloroacetic acid
TLC1	Telomerase component 1
TOP2	Topoisomerase 2
TOP3	Topoisomerase 3
TRP	Tryptophan
URA	Uracil
URA3	Uracil requiring gene 3
UV	Ultraviolet
VSV	Vesicular stomatitis viral glycoprotein
WRN	Werner
WS	Werner’s syndrome
WT	Wildtype
XRS2	X-ray sensitive
YPD	Yeast extract with peptone and dextrose
ZIP1	Molecular Zipper

ABSTRACT

The RecQ family of helicases has been termed the “Caretakers of the Genome,” and rightfully so. These proteins are highly conserved from bacteria to humans and have been implicated in functions from homologous recombinatorial repair to damage checkpoint response to telomere maintenance and more. Mutant genes of three of the human RecQ helicases lead to syndromes characterized by a high incidence of cancer, premature aging and early death. Despite their implications in several biological functions and importance to the integrity of the human genome and suppression of cancer, many aspects of the RecQ family structure and function remain unknown. To date, much is known about the catalytic function of the helicase domain and accompanying domains, but considerably less is known about the non-catalytic N-terminus in these proteins, which, in many cases, including those human orthologs involved in disease, can make up about half of the total protein length. While experiments have been able to identify protein partners that interact with the N-terminal region, few are able to narrow the binding sites to minimally functional parts and fewer still describe any detail regarding the structural features of these binding areas. In fact, some reviews have generally described the N-terminus as “featureless,” a concept we challenge in our studies.

Many of the N-termini of these RecQs have long been known to contain large stretches of acidic residues, a feature of intrinsically disordered regions. These

regions/proteins are rich in charged and polar residues, lack compactness that makes crystallography possible, and have flexible and dynamic conformations that are prevalent in “high specificity, low affinity” interactions. Disordered proteins are well-known to be hot spots for protein/protein interactions and post-translational modifications, amongst other functions. Considering these facts, and recognizing the ties between these and what we know about the N-termini of the RecQs, we hypothesized that these proteins likely have long disordered termini. In Chapter 3, we confirm the presence of disorder at the Top3/Rmi1 binding site on Sgs1, the *Saccharomyces cerevisiae* RecQ helicase. We show that even in a disordered state, this binding region is not “featureless,” but in fact contains a transient alpha-helical molecular recognition element that is necessary to facilitate complex formation between Sgs1, Top3 and Rmi1. Loss of helical structure at this site leads to increased genomic instability and sensitivity to DNA damaging agents. Based on these results, we suggest that there are likely many more such elements in the N-terminus that that are important for other Sgs1 protein/protein interactions and provide an estimate for the number of interactions in this region.

In Chapter 4, we evaluate the prevalence of disorder in a set of Chromatin Processes proteins in an effort to establish a role for disorder with regards to maintaining chromatin integrity. In our bioinformatics study, we found that disorder is overrepresented in the Chromatin Processes proteins, and that a major driving force for disorder in these proteins is protein/protein interaction and post-translational modification. We also show a biological connection to disorder and increased protein/protein interaction by investigating these parameters in the context of the DNA

damage checkpoint response and in complex formations. Mediators between highly structured kinases in the checkpoint were the most interactive proteins and over half of all predicted interaction sites occurred in disordered areas. Complexed proteins often contained one protein with a high number of disordered sites and a high number of predicted interactions, while the rest were considerably more ordered.

Chapter 5 explores a Sgs1 interaction partner, Rmi1 and uses bioinformatics to design structurally-based point mutations in an effort to further elucidate Rmi1 function in yeast, which remains largely unknown outside of its enhancement of Top3/Sgs1 catalytic function. Using AGADIR, which predicts alpha-helical structure and is particularly useful in our hands for guided-mutagenesis in disordered regions, we identified several point mutations that lead to Δ*rmi1* phenotypes or intermediate growth on hydroxyurea. We hypothesize that these mutants are important in maintaining Rmi1 stability.

Together, these studies suggest an important change in how the field approaches further studies into the RecQ helicases; traditional methods of primary sequence comparisons and crystal structures limit the study of disordered regions that are still functionally important. Future care should be given to consider the conservation of structure or structural elements in the RecQs over strict alignments when comparing functional regions between orthologs. Our studies also suggest that it is highly likely that structural motifs for important protein interactions in RecQs are being overlooked because they are not readily obvious using traditional methods. By understanding these motifs and the interactions they facilitate, we may be able to more easily identify polymorphisms in patients with genomically unstable conditions like cancer and, having

better understood the biological process these structures facilitate, design drugs to counteract detrimental effects.

CHAPTER ONE:

INTRODUCTION

The RecQ family of helicases is highly conserved from bacteria to humans and functions at the interface of DNA replication and repair [1]. It is instrumental in maintaining genomic stability by directing homologous recombinatorial repair and preventing crossover events with non-sister chromatids during repair [2]. Some organisms like *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* contain only one RecQ helicase, while others can contain many--like the human WRN, BLM, RecQL4, RecQL1, and RecQL5 homologs--with a maximum of seven homologs discovered in *Arabidopsis* [3]. Homology arises from the similarity of the catalytic domains [4]; differences in protein primary sequence usually arises in extended N-terminal regions of the protein, most of which are predicted to have an appreciable degree of protein disorder.

Founding Member of the RecQ Family of Helicases

RecQ was first identified in *E. coli* as a mutated gene in the RecF pathway that imparts death-resistance in thymineless media, but makes the organism more sensitive to UV damage and results in a deficiency in recombination [5]. Purification efforts of the gene product found that the gene produces a helicase with 3' to 5' unwinding activity and DNA-dependent ATPase activity [6]. This activity is sensitive to ATP and Mg^{+2}

levels, to the presence of single-stranded DNA (ssDNA) and is facilitated by the presence of single stranded DNA binding (SSB) protein and [7]. RecQ promotes and disrupts recombinatorial events, and thus is an important mediator in preventing aberrant recombination events (i.e. Between homeologous sequences) [8].

RecQ Helicase Structure and Function

Structurally, all RecQ family members contain a C-terminal helicase region with a DEAD/H motif that spans about 400 residues with seven sequence motifs common in most DNA and RNA helicases, including the Walker-A motif, which is required for the binding and hydrolysis of ATP [9]. Crystallography has revealed that the helicase region of RecQ consists of two lobes separated by a cleft lined with highly conserved amino acid residues that bind ATP and potentially ssDNA [10]. Single point mutations in and around this cleft in the human RecQ variant, BLM, have been found to lead to either full- or partial-loss-of-function of the helicase [11].

Many RecQ helicases also contain domains that are essential for the protein's catalytic activity including a RecQ conserved (RQC) domain and a helicase and RNaseD C-terminal (HRDC) domain [12]. The RQC domain, while not present in all members, is exclusive to the RecQ family. The function of this domain is not fully understood, but it has been implicated as an important region for DNA binding and processing. Crystallography reveals that the region has two subdomains. One subdomain is crucial for Zn^{+2} binding via a set of four conserved cysteines; this domain was also found to be essential for proper BLM core folding and ATPase activity [13]. The second subdomain is a winged-helix domain, which has been shown to bind DNA

in the human WRN homolog [10, 14]. NMR structural analysis of this region in the human BLM protein reveals that it is comprised of 4 alpha helices and 4 beta strands with a flexible loop between the first and second helices not present in other RecQs, implying that the full functionality of this winged helix domain may vary between orthologs [15]. In WRN, this domain also contains a nucleolar targeting sequence [16].

The functionality of the HRDC domain has been implicated through biophysical studies; the structure of this domain has been resolved via NMR and crystallography, and has been shown to resemble DNA binding domains from other helicase and DNA-binding proteins [17, 18]. This domain in the *S. cerevisiae* ortholog was found to be comprised of 5 helices, and, since the hydrophobic residues responsible for packing these helices into a functional core are conserved, it is hypothesized that the structure is likely similar in other orthologs with this domain. On the surface of this core are a series of positively-charged residues that can facilitate binding to negatively-charged DNA [17]. The HRDC domain is required for dissolution of double Holliday junctions (dHJs) in BLM protein, and can bind DNA in the RecQ protein [18, 19].

A study has suggested that the RecQ helicases be divided into two classes, depending on the length of the protein and overall structure (Figure 1.1) [9]. One class, called the 'long-form' RecQ helicases, contains a large N-terminal stretches of acidic amino acids and/or extended C-terminal regions of poorly described function, or, in the case of WRN, an exonuclease domain. Members of the RecQ family included in this class include BLM, WRN, Sgs1, *D.m.*BLM, and RecQL4. The 'short-form' of RecQ helicases is missing this long N-terminal region but still contains the typical helicase domain and may contain the RQC and HRDC domains. An example of this class

includes *E. coli* RecQ. Opresko et al suggests that the presence or absence of this additional protein sequence on either the N- or C-terminal end may reflect the nature of DNA substrate-binding specificities in the RecQ family; in general, *E. coli* RecQ is the least structurally complex member but preferentially binds the broadest number of DNA duplex types. Sgs1, BLM and WRN, with their long stretches of additional sequence on both the N- and C-termini, tend to unwind mostly those DNA duplexes containing junctions, such as Y-structures and HJs [20]. It is probably more likely that the extended domains in the N-terminus are to serve as protein/protein interaction sites and modifications sites, as discussed at length below.

Figure 1.1: Domain comparison of long- and short-form RecQ helicases. Data for *H. sapiens* sand *S. cerevisiae* come from genetic studies; *C. albicans* domains are estimated by a CD search on NCBI BLASTp.

RecQ Helicase in *S. cerevisiae*, Sgs1

Sgs1 is the only RecQ helicase in the budding yeast Saccharomyces cerevisiae; it contains 1447 amino acid residues with a helicase domain (residues 674-1017), an RQC domain (residues 1017-1085) and an HRDC domain (residues 1272-1351) [21]. Purification of residues 400-1268 (including the helicase and RQC domain) showed a protein able to hydrolyze ATP, but only in the presence of DNA substrates of various forms [22]. This fragment is able to bind DNA, with a preference for ssDNA over dsDNA, particularly forked substrates. This same fragment is capable of unwinding G4 quadruplex structures [23]. Full-length Sgs1 is capable of binding Y-structure, ssDNA, 3' overhang, 5' overhang, HJs, and dsDNA. The binding of substrates with 5' and 3' overhangs was found to be similar, unlike the results found in the 400-1268 fragment. This implies that the N-terminal and C-terminal domains of the protein, including the HRDC and mostly undefined N-terminus, may facilitate the binding of a wider array of substrates [24]. Like the founding family member, Sgs1 unwinds DNA in a 3'-5' direction and requires ATP (and preferably, Mg^{+2}) [22]. The rate of ATPase activity of Sgs1 is approximately 10 fold greater than RecQ and is inhibited by the presence of RPA, possibly as a function of competition between the two proteins for ssDNA [24]. The first 674 amino acids lack any well-defined or well-conserved domains, but do contain two extended regions of acidic residues (residues 400-474 and 510-596; [12]) that are present in other yeast homologs and in human BLM. Deletion of these regions can suppress Δ*top3* slow growth like a Δ*sgs1* mutant, but are mostly like wildtype growth on hydroxyurea (HU) and methylmethanesulfonate (MMS), implying that there is

a separation of function in this region, though the specific function of these regions remains to be discovered [12].

Characterization of the Δ*sgs1* Phenotype

Sgs1 was first classified by Gangloff et al. in 1994 during a search for suppressors of the slow-growth phenotype common to strains lacking Topoisomerase 3 (Top3). The importance of Sgs1 in maintaining genomic stability is apparent as cells lacking Sgs1 show an increase in mitotic recombination in the form of interchromosomal homologous recombination, intrachromosomal excision recombination, and ectopic recombination [25, 26]. This recombination has been found to be improperly induced and concluded in Δ*sgs1* cell lines, as cells lacking Sgs1 have been found to have an increase in gross chromosomal rearrangements (GCRs), most notably in the form of translocation/deletions and telomere additions [27]. Meiotically, Δ*sgs1* cells are also abnormal, as an increased number of multichromatid molecules and an increase in homeologous recombination between non-identical sequences (though not homologous recombination) have been observed [25, 28, 29]. Cells lacking Sgs1 were found to undergo more synapsis in meiosis and have an increased number of crossing over events over wildtype cells [30]. Exposure of Δ*sgs1* cells to a diverse range of DNA damage has revealed several sensitivities, including UV light, hydroxyurea and MMS [31-33]. In addition to being sensitive to damage, sgs1Δ cells are overall unhealthy under normal conditions, and have an average life span only 40% that of wildtype cells. In older Δ*sgs1* cells, the nucleoli appear to age faster than wildtype cells as well [34].

Role of Sgs1 in DNA Repair

Catalytically, Sgs1 acts in many steps in DNA repair (Figure 2). At the beginning of double strand break processing, Sgs1 works in tandem with Dna2 and Exo1 to resect the ends of the break beyond the actions of the MRX complex, which is needed to produce a length of ssDNA long enough for Sgs1 to work on [35, 36]. Sgs1 is needed in order to achieve resection efficiency needed for further steps in the repair process; in cells missing Sgs1, only a fraction of cells are able to achieve the same degree of resection as wildtype cells [37]. The ends produced by Sgs1 and partners are then acted upon by Rad51, which is able to invade a homologous sequence and use it as a template for error-free recombinatorial repair. Sgs1 later works in conjunction with the topoisomerase Top3 and accessory protein Rmi1 to dissolve double dHJs in a non-crossover event [38]. The result is a repaired gap that is error-free. Loss of Sgs1 results in a greater percentage of these breaks being repaired in an alternative pathway that promotes crossing over by allowing HR to occur between homologous chromosomes rather than sister chromatids [39]. This may result in a loss of heterozygosity (LOH). In addition to properly processing ends in the first steps and resolving branched structures in the last steps of recombination, Sgs1 has been found to be instrumental in preventing homeologous recombination by acting as an anti-recombinase [27]. Because there is an increase in recombination between sequences of less than perfect homology in cells lacking Sgs1, it is hypothesized that Sgs1 plays a role in dismantling the d-loop created by Rad51 during strand invasion if the site of recombination on the donor strand is not homologous. Though the method of divergent strand rejection is not fully known, it has been shown that mismatch repair proteins often drive the suppression of homeologous

recombination, and that helicase activity is likely a necessary addition to the proteins that mediate the dissolution [27, 29, 40].

Sgs1 has also been shown to be required in the intra-S-phase checkpoint that activates in response to stalled DNA replication. This checkpoint is activated in part by the phosphorylation of Rad53, and cells that were lacking Sgs1 were found to have substantially lower amounts of phosphorylated Rad53 in S-phase under damaging conditions [41]. This implies that, in addition to its role in the physical repair of DNA damage, Sgs1 has been found to promote genome integrity by activating replication checkpoints that allow for sound repair of DNA damage and by stabilizing DNA polymerase ε at these stalled forks along with Mec1 [42-44]. Sgs1 has also been implicated in the phosphorylation of Rad53 in conjunction with Exo1 in G1 cells, and both are needed to induce a checkpoint delay at G1 in the presence of irradiating damage [45]. Sgs1 interaction with Rad53 has been confirmed via mutagenic co-immunoprecipitation and is believed to be reliant on the phosphorylation of Sgs1 at residue T451; this gives support to Sgs1 working as an upstream element of Rad53 in the checkpoint cascade, a theory supported by the experiments in G1 cells described above [46].

Sgs1 has also been implicated in telomere maintenance, since in cells that lack telomerase (Δ*tlc1*, the gene for telomerase RNA) loss of Sgs1 results in more rapid aging, an increased rate of telomere shortening, and arrest in G2/M phase. Though the exact nature of Sgs1's role in telomere maintenance is still unknown, it is hypothesized that the phenotypes observed in Δ*tlc* Δ*sgs1* cells may be the result of a) formation of DNA quadruplex structures at shortened telomere ends that require Sgs1 to be

resolved, b) the loss of S-phase checkpoint ability that results in the delayed G2/M arrest seen in these cells, c) Sgs1 playing an active, yet unknown role in telomerase-independent repair pathways (ALT pathway) as it was shown to co-localize with a few proteins in this pathway, or a combination of the above [47]. Deletion of SGS1 in a Δ*est2* background (a catalytic subunit of telomerase) also saw rapid senescence beyond the deletion of EST2 alone. The double mutant also exhibited the increased rate of telomere loss found in the Δ*tlc1* Δ*sgs1* strain, and this loss was further exacerbated by the loss of another protein in the recombination pathway, Rad52. When senescent cells were allowed to recover via recombination-mediated telomere lengthening, Δ*sgs1* Δ*est2* cells were unable to recover wildtype growth rates like the Δ*est2* mutant, implying a role for recombination and, subsequently, Sgs1 in telomere lengthening as a means of maintenance. Indeed, evaluation of the telomere structure of Δ*est2* versus Δ*est2* Δ*sgs1* cells found that the former contained longer type II structures with Y' elements followed by extended tracts of $C_{1-3}A/TG_{1-3}$ repeats, while the latter contained shorter type I structures, which contains tandem Y' elements followed by a short tract of repeats [48]. It has been shown that cells with type II structures tend to eventually reach a wildtype growth rate, while type I cells never recover wildtype growth. The findings imply a role for Sgs1 in suppressing preference for type I structure formation in the absence of telomerase [49, 50]. Part of Sgs1's role in maintaining type II telomere structure is due to a sumoylation event at K621, as mutation of this residue to an arginine incapable of being sumoylated shows the increase in type I telomere structure found in strains lacking telomerase and full-length Sgs1 [51].

Recently, Sgs1 has also been implicated in single strand annealing (SSA) and strand exchange (SE) activities [52]. The region in the N-terminus spanning residues 103-322 has been identified as the minimum domain for SSA and SE. *In vivo* assays have shown that in a sgs1 mutant lacking this region, recombination rates are elevated on par with a Δsgs1 strain, implicating this region in suppressing hyperrecombination, though the mechanism, and whether or not deletion of the domain inactivates the whole protein, is unknown. Loss of this region also leads to a greater amount of homeologous annealing *in vivo*, implying that the region is needed in order to reject duplex DNA formation between mismatched sequences. *In vitro*, a peptide containing this region can perform SA with homeologous DNA, but not SE, implying that the SE function of this region in Sgs1 is what is used to prevent aberrant recombination between nonhomologous sequences *in vivo* [52].

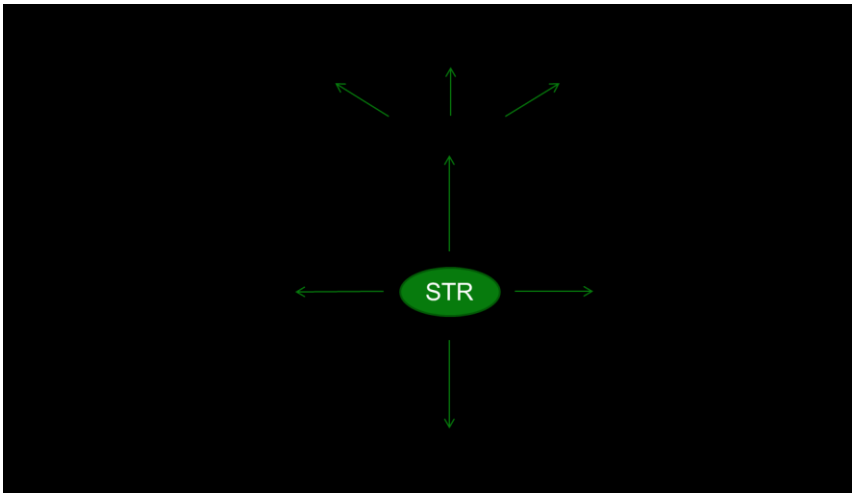


Figure 1.2: Roles for STR complex in yeast.

N-Terminal and Putative N-terminal Sgs1 Protein/Protein Interactions

While the catalytic core serves as the driver for the catalytic functions of Sgs1, only half of the protein is actually dedicated to helicase and helicase-related activities. The entire N-terminus of Sgs1 has no known activity outside of SE/SSA, but has been implicated in several protein/protein interactions; the diversity of these interacting proteins illustrates the numerous pathways and steps Sgs1 has been implicated in with regards to DNA repair (Figure 1.3).

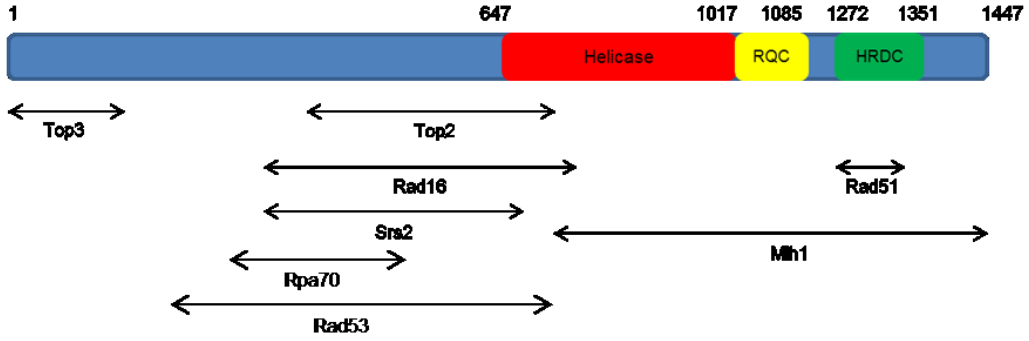


Figure 1.3: Putative domains of Sgs1 binding partners. Domains of selected partners important in DNA repair are present over whole protein, but are preferentially located in the N-terminus.

Topoisomerase 2 (Top2) is a type II topoisomerase responsible for DNA decatenation events in both mitosis and meiosis; loss of Top2 can result in failure to complete mitosis and meiosis [53]. Because of this, Top2 has been described as the only essential topoisomerase in yeast. Top2 has been shown to interact with full-length Sgs1 via co-immunoprecipitation, and yeast-two-hybrid assays have narrowed down the

minimal binding area for Top2 on Sgs1 from residues 466-746 and for Sgs1 on Top2 to residues 1109-1163, a potential leucine zipper [53, 54].

Rad16 was also discovered via yeast two-hybrid to bind with the N-terminal region between residues 421-792 [55]. This region overlaps the Top2 binding domain. Rad 16 has been implicated in nucleotide excision repair (NER) in conjunction with Rad7; the two form the NEF4 complex, which is an ATP-dependent DNA damage sensor [56]. It is instrumental in repairing UV damage, specifically pyrimidine dimers induced in transcriptionally inactive DNA. Yeast cells lacking Rad16 cannot perform NER but retain transcription-coupled repair [55]. Loss of Sgs1 in a Δ*rad16* background leads to increased sensitivity to UV damage, implying that the two proteins may work in redundant repair pathways, despite the fact that they interact physically. Because of this evidence of physical interaction, the authors of the primary study of the Sgs1/Rad16 interaction suggest that in addition to the several functions of Sgs1 outlined above, it may also play a role in repairing specific DNA damage with Rad16, thus making Sgs1 an important protein in the absence of Rad16 [55].

Srs2 is a second 3'-5' helicase present in *S. cerevisiae.* It has similarities to bacterial UvrD/Rep helicases and is implicated in early-stage recombinatorial regulation [57, 58]. Though Srs2 and Sgs1 have been shown to physically interact [59], the double mutant is inviable unless upstream recombination elements are also knocked out, implying that in the absence of these helicases, lethal recombination events occur. The area mapped to Sgs1 for Srs2 binding spans residues 422-722, an overlap with both the proposed domains for Top2 and Rad16 binding. It should be considered, also, that Sgs1, Srs2 and Mre11 form a complex, though the interaction between Sgs1 and Mre11

alone has not been elaborated on; it is possible that the binding region for Srs2 also houses a binding domain for Mre11 [59].

Deletion of Sgs1 contributes to a decrease in Rad53 phosphorylation [41]. Interaction with Sgs1 via Rad53's FHA1 domain has been localized to Sgs1 residues 292-661 [44, 60]. Mutagenesis of four phosphorylation sites in this regions to unphosphorylatable residues shows an approximately 50% reduction in Rad53 activation [60]. The authors of the study suggest that interaction between Rad53 and Sgs1 at these sites in the acidic region is necessary for Rad53 recruitment and activation.

Rpa70 binding to the acidic region of the N-terminus was recently identified using beta-galactosidase binding assays; the binding area is between residues 292-661 on Sgs1--the same binding site proposed for Rad53. Isothermal titration calorimetry further narrowed this region to residues 404-560 [46]. Interestingly, the presence of Rpa70 controls the nuclease/helicase functions of another Sgs1 binding partner, Dna2. Dna2 and Sgs1 work together in conjunction with Exo1 to resect DNA at double strand breaks [61]. The authors of the study found that the presence of RPA enhances 5'-3' Dna2 resection, and impairs 3'-5' resection. It is possible, then, that the Rpa70 interaction on Sgs1 serves, in part, to regulate strand selection of the Dna2/Sgs1 interaction at breaks. Physical interaction between Sgs1 and Dna2 has been confirmed [61], but it is yet unknown where the interaction occurs on Sgs1. Considering the role of Rpa70 in resection activity of Sgs1/Dna2, and considering where Rpa70 has been confirmed to interact, it is plausible that the N-terminus contains a binding site for Dna2 as well.

The RTR Complex: RecQ Interaction with Top3 and Rmi1

The most well-studied interaction with Sgs1 involves the formation of a RecQ/Topoisomerase/Rmi protein (RTR) complex, an assembly that is conserved in many of the other RecQ homologs. In *S.cerevisiae*, Top3 and Rmi1 partner with Sgs1 to perform the catalytic functions necessary to facilitate early and late stage DNA repair.

Role and Interaction of Topoisomerase III (Top3) in the RTR complex

Top3 was first purified and classified in a 1992 study as a type IA topoisomerase [62]. Strains lacking Top3 suffer from a slow-growth phenotype under normal conditions, hyperrecombination, inability to form asci, and an increased GCR rate [27, 63, 64]. Structurally, the yeast Top3 has not been solved, but as the human variety, Top3α, is said to closely resemble the canonical topoisomerase IA structure, it stands to reason that *S. cerevisiae* Top3 likely contains the domains I-IV, Toprim, and acidic cluster found in other orthologs [65, 66]. Top3 is responsible for resolving negatively supercoiled DNA structures in the cell, and has been implicated in resolving recombination-dependent X-shaped molecules (HJs) in an Sgs1-dependent manner [67]. This topoisomerase dependency on RecQ function has also been seen in bacterial systems, as *E. coli* RecQ has been found to stimulate catenation of dsDNA via TopoIII [68]. The first yeast two-hybrid study implicated the first 282 amino acids of Sgs1 in Top3 binding [54], and the minimum functional binding area has been further narrowed down to the first 158 amino acid residues via co-immunoprecipitation and ELISA assay [69]. Yeast mutants lacking the N-terminal 158 residues of Sgs1 demonstrate hyperrecombination and a greater DNA damage sensitivity as compared to a Δ*top3*

mutant. It has been hypothesized that the loss of the N-terminus results in a "hyperactive" Sgs1 helicase which introduces more chromosomal instability than a wildtype Sgs1 in this same background [70]. This effect is described by the authors as a "toxic effect" that is specific to a Sgs1 lacking the interaction domain for Top3 but still catalytically active. The nature of this toxic effect, however, is still unknown. The phenotype observed in these mutants can be overcome by increasing the concentration of Top3 in the cell or by fusing the Top3 open reading frame to the N-terminus of Sgs1 [69, 71]; as Sgs1/Top3 interaction has been shown to be independent of DNA binding, it has been implied that Sgs1 may be responsible for recruiting Top3 to areas of DNA it needs to act upon [72].

Role and Interaction of Rmi1 in the RTR Complex

Rmi1 was first discovered in a screen for genes that, like Sgs1, were required for viability of cells that lacked Mus81, with hopes that this would lead them to other candidates for genes in the Sgs1-Top3 pathway [73]. RMI1 deletion produced a synthetically lethal phenotype in a Δ*mus81* mutant that could be rescued when RAD51 was also deleted; as this lethality could also be repeated in conjunction with deletion of several other genes with roles in replication fork restarting (RRM3, SLX1, SLX4, etc.), and that this lethality could also be rescued when genes responsible for homologous recombination were also deleted (RAD51, RAD52, RAD54), a role for Rmi1 in homologous recombination was proposed [74]. Strains lacking Rmi1 are slow-growing and sensitive to HU and MMS; these phenotypes could be suppressed by also deleting SGS1, similar to Δ*top3*. Rmi1 strains are also genetically unstable; Δ*rmi1* strains have

an increased recombination rate over both wildtype and Δ*sgs1* cells. Rad52 foci are increased in response to a loss of Rmi1, implying that there is an increased rate of spontaneous damage in Δ*rmi1* cells that likely explains the elevated GCR rates [73, 74]. A role for Rmi1 in checkpoint response has been suggested as well, as Rad53 phosphorylation is incomplete in the presence of HU or MMS when Rmi1 is absent [74]. Rmi1 has been co-precipitated with both Top3 and Sgs1, but Sgs1/Rmi1 and Sgs1/Top3 interaction appears to require the third partner to bind stably [73, 75]. Rmi1 has no known catalytic function, but has been implicated in binding small amounts of DNA, particularly dHJs [76]. Rmi1 is a known enhancer of Top3 DNA relaxation and has been found to stimulate the decatenation of dHJs by Sgs1 and Top3, particularly at late stages of junction dissolution, a function retained in the human complex [24, 76, 77]. Interestingly, yeast cells lacking Sgs1, Top3, or Sgs1 and Top3 have been found to retain DNA resection ability (albeit slow and inefficient compared to wildtype), but strains lacking Rmi1 lack the ability to resect at all, perhaps suggesting an important supporting role for Rmi1 in resection function both in the presence and absence of Sgs1 [37]. Rmi1 may also play a role in sister chromatid cohesion in conjunction with Top3, as cells lacking either of these proteins show an increase in separated chromatids over wildtype cells [78]. Notably, Δ*sgs1* mutants did not appear to have a cohesion defect; the authors suggest that, because Δ*rad51* and Δ*sgs1* in conjunction with Δ*rmi1* and Δ*top3* alleviated the cohesion defect, Rad51 and Sgs1 act upstream in a pathway that drives Top3 and Rmi1 to facilitate cohesion.

The structure of the yeast ortholog of Rmi1 is still unknown, but alignment of the human N-terminal portion of the protein best approximates the yeast structure. Human

Rmi1 contains a cluster of three alpha helices at the far N-terminus that are predicted to stabilize a central β-barrel structure containing an OB-fold [79]. This barrel affixes to Top3α in domain II, with the alpha helices facing away from the domain. From this barrel extends a 23-residue loop with no appreciable structure; this loop is believed to be a functional region for modulating Top3α activity by controlling the mechanistics of the catalytic "gate" of the topoisomerase, and is necessary for Top3α/Rmi1 interaction [65, 79]. Studies with the human Rmi1/Top3α found that this loop is needed for the enhancement of Top3α dissolution of dHJs and helps facilitate interaction between the two proteins, lending support to the gatekeeper model [79]. In human RTR, this process may be more complicated, as the Rmi1 protein is double the size of the yeast variety and the complex contains a fourth member, Rmi2, which binds to the C-terminal portion of the human Rmi1 protein and to RPA [79, 80]. $Rmi1^{-/-}$ models in mammalian systems reveal similar phenotypes to those found in yeast, as $rmi1^{-/-}$ mice stop developing in the embryonic stage due to a decrease in cell proliferation and decrease in replication and accumulate genomic instability in the form of aneuploidy and fragmented chromosomes [81].

RTR Complex Member in Humans: Rmi2

The human RTR complex contains a fourth member in Rmi2, a protein with no known homolog in yeast. Considering the fact that Rmi2 interaction in the complex is due at least in part to interaction with the C-terminal OB fold of Rmi1, which has no equivalent in yeast, it is unlikely that this protein exists at all in yeast, and could be a function of the more complex RecQ network in humans [82]. Disrupting the interface

that is responsible for Rmi1/Rmi2 binding contributes to an increase in sister chromatid exchange (SCE) in human cells, a sign of genomic instability [82]. Deleting Rmi2 leads to an overall decrease in both Top3α and Rmi1 levels, sensitivity to MMS, and an increase in chromosome breaks [83]. The role for Rmi2, like the role for Rmi1, is still being fully discovered, but it has been suggested that Rmi2 is responsible for proper Rmi1 folding, RTR complex stability, and enhancement of dHJ dissolution via RTR [82, 83].

RecQ Orthologs in Humans

Unlike yeast, humans have five known RecQs, [84], all with activity in different pathways, interactions with different proteins, and, in the case of three of the five, different syndromes that result from mutation in one of the RecQ genes. This part of the introduction will discuss the five human homologs, starting with the two not known to cause syndromes, and following with the clinically-relevant homologs.

RecQL1

RecQL1 is the most robustly expressed of all of the RecQs in humans, and is most prevalent in the heart, lungs, skeletal muscles and kidney [85]. It is capable of unwinding forked substrates in an ATP-dependent manner and of single strand annealing in the absence of ATP [86]. Deficiency in RecQL1 has no known disease phenotype, but cells lacking the protein have decreased replication origin firing and replication fork rates. Since RecQL1 is recruited to origins at the G1/S border, it is reasonable to assume that the protein plays some role in modulating proper DNA

replication *in vivo* [87]. RecQL1 interacts with PARP1 [88], a protein implicated in responding to replication stress, and mutants lacking RecQL1 are sensitive to agents that impair replication via inhibition of topoisomerase activity. RecQL1 has the capability of restoring synthetically-created replication fork structures from regressed "chicken-foot" structures, implying that the protein is needed to restore forks undergoing repair and restart proper replication [89]. Mouse fibroblast cells lacking RecQL1 also have increased sensitivity to IR damage, an increase in spontaneous γH2AX foci (an indicator of DNA damage), aneuploidy, and an increase in SCEs, all indicators of genomic instability [90].

RecQL5

The human RecQL5 gene locus encodes for three isomers, generated via alternative splicing. The α- and γ-isoforms are largely understudied, but a study into one of the Drosophila small-form isomers shows that it shares the catalytic characteristics of the other RecQs [91]. The small isomers, however, lack any nuclear localization signal, and are thus not present in the nucleus [92]. The sub-cellular localization of these isoforms and specific function has yet to be explained. The most commonly-studied variant is RecQL5β, a 991 amino acid protein present in both the nucleus and cytoplasm. It is expressed in many different tissue types in appreciable amounts, with strong expression in the testis [92, 93]. Cells lines lacking RecQL5 are enriched in γH2AX and Rad51 foci, have an increased number of broken chromatids/chromosomes, tri- and quadriradials, and other aberrations, and are associated with an increase in HR-mediated DSB repair [94]. While deletion of RecQL5 does not lead to an increase in

spontaneous SCEs, the deficient cells are campothecin (CPT)-sensitive and have an increase in SCEs when treated versus wildtype cell lines [95]. It is suggested that the cause of RecQL5 CPT sensitivity is due to the inability of the cells to restart forks post-treatment, leading to apoptosis [96]. RecQL5 is a 3'-5' helicase found to interact with the MRN complex *in vivo* and *in vitro*. RecQL5 co-localizes with the complex at the site of DNA double strand breaks, and requires the complex *in vivo* for recruitment to damage. RecQL5 has been implicated in regulating MRN exonuclease activity by impeding resection of break ends [97]. The helicase also interacts with RNA polymerase II via its RPBI subunit; because the interaction occurs only when the mapped interaction site on RPBI is phosphorylated, and because this phosphorylation is indicative of transcript elongation, it has been proposed that RecQL5 also plays a role in transcription [98]. The most fully studied interaction is between Rad51 and RecQL5, where the interaction site of RecQL5 on Rad51 at the stretch between residues 654-725 is needed for RecQL5 prevention of d-loop formation by displacing Rad51 on ssDNA [94, 99]. This prevention is believed to be an active mechanism for regulating Rad51 activity preventing HR in wildtype cells, a parallel function to the anti-recombinase function seen in Sgs1. While no known human disease is connected to RecQL5 mutation, RecQL5 deficient mice were found to be more prone to cancer than their wildtype counterparts, with 46% of those tested developing a cancer by 22 months of age [94]. The most common cancers in this study were lymphomas, and the most common solid tumor was lung adenocarcinoma. In a study of RecQ homolog levels in primary colorectal cancer, RecQL5 mRNA expression and protein levels were found to be reduced in tumors, particularly in those tissues described to be microsatellite instable [100]. Overall, this

implies that RecQL5 may be tumor-suppressing in the colon, so while there may not be a specific syndrome equated to mutant RecQL5, it is feasible that the protein levels are misregulated in other cancer types as well.

RecQL4

RecQL4 displays the same ATPase and single strand annealing properties as other RecQs, but compared to some of the other human members, it is not particularly robust as a helicase since it is unable to efficiently unwind the diversity of DNA structures found from BLM or WRN protein [101, 102]. Unlike these other helicases, however, RecQL4 seems to be adept at unwinding long duplex DNA, suggesting a specialized biological function [103]. Structurally, RecQL4 is also remarkably different from other “long” human RecQs in that it does not contain the RQC or HRDC accessory domains typical of this family of proteins. Instead, it is the only member to feature a domain with similarities to *S. cerevisiae* Sld2, a protein necessary for proper DNA replication [104, 105]. The ortholog in *Xenopus laevis*, xRTS, has been found to be recruited to DNA early in replication initiation and deficient cells are delayed in DNA replication [104]. Drosophila RecQ4 also has replication defects, including loss of DNA polymerase alpha on chromatin and S-phase arrest; the Sld2 domain has been found to be necessary to rescue RecQ4 deficient cells from replication defects [106]. These findings suggest a role for RecQL4 in maintaining proper replication fidelity in the cell. Interestingly, RecQL4 has also been detected in the mitochondria and is necessary for proper mtDNA replication [106, 107]. RecQL4-deficient fibroblasts are sensitive to HU, CPT, and doxorubicin (DOX), but are not significantly sensitive to UV, IR, and cisplatin

damage. Since HU, CPT, and DOX are primarily harmful during S-phase, this further strengthens the suggestion that RecQL4 plays a functional role in replication (Figure 1.4) [108] . A role for RecQL4 in telomere maintenance has also been proposed, since interaction with TRF2, a protein in this process, has been confirmed [109]. The authors of this study also found that RecQL4 associates with TRF1 foci at telomeres during S-phase and deficient cell lines display greater telomere fragility than wildtype cells. What role RecQL4 actively plays at the telomeres, however, has yet to be discovered.

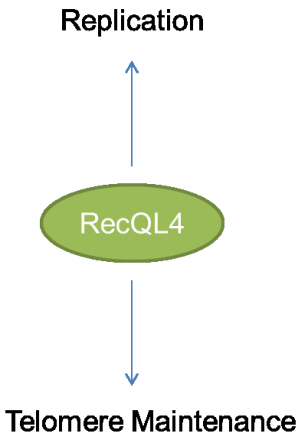


Figure 1.4: Proposed roles for RecQL4.

RecQL4 is predominantly expressed in the thymus and testis, with low-level expression also detected in the heart, brain, placenta, pancreas, small intestine, and colon; peak expression is detected in G1/S, consistent with a role in early replication [110]. Given the pattern of expression, it is unsurprising that RecQL4 deficient mice models have been found to have hypoplasia of the thymus; this finding perhaps explains the frequency of infections some patients of RecQL4-deficient syndromes incur. The same mice are also prone to embryonic growth retardation and death, low

birth weight, hair loss, skin lesions, sclerotic tails and premature greying, symptoms of aging, growth and skin problems similar to those seen in some human patients [111]. Perhaps suggestive of the diversity of its suggested roles, humans deficient in RecQL4 suffer from one of three syndromes: Rothmund-Thomson Syndrome (RTS), RAPADILINO syndrome, or Baller-Gerold syndrome.

Rothmund-Thomson syndrome was first described in 1868, by an ophthalmologist who observed children with poikiloderma and cataracts [112]. Patients with RTS can exhibit any of a myriad of symptoms, including short stature, bone deformations, poikiloderma, early aging, photosensitivity and neoplasia predisposition, particularly lymphomas and osteosarcomas; variability of symptoms between patients is often great, as indicated in several patient case studies where no two patients showed the same phenotype and no patient showed all of the possible symptoms associated with the syndrome (Figure 1.5) [112-115]. The mutation responsible for the disease varies from patient to patient, but all known mutations are exclusive to the helicase core and C-terminus of the protein [115]. RAPADILINO syndrome is characterized by radial deformations, patellae malformation/lack of formation, diarrhea, skeletal abnormalities including high or cleft palates, and short stature [116]. There are fewer than 20 reported cases of the syndrome as of 2009, but they are overwhelmingly occurring in Finland, implying the presence of a founder mutation. Cancer status is as high as 40% in patients with RAPADILINO, with osteosarcomas and lymphomas being the most common cancer types in the populations studied, particularly in those with dual A420 and A463 deletion genotypes [115]. Baller-Gerold syndrome patients exhibit bone deformities similar to those seen in RAPADILINO, and have anomalies including cranial

malformations, imperforate anus, rectovaginal fistulas, poor prenatal growth, and mental retardation [117].

Figure 1.5: Rothmund-Thomson patients. The patient on the left shows the poikiloderma and bone malformations characteristic of the syndrome. The patient on the right shows a closer look at poikiloderma [118, 119].

WRN

The Werner's helicase is much like other RecQ homologs in that it easily unwinds a diverse spectrum of DNA structures, including G4 quadruplex, bubbled duplex, 3'-overhang, synthetic x-junction, and forked duplexes; long-range unwinding is stimulated by the presence of RPA, which physically interacts with WRN [120, 121]. However, it presents another original case with regards to human "long form" RecQs in that it contains a 3'-5' exonuclease domain in the far N-terminus of the protein that is active particularly on forked DNA [122, 123]. Roles for WRN are summarized below (Figure 1.6). Phenotypically, cells which lack a functional WRN protein have a growth deficit, decreased survival rate, and sensitivity to HU, MMS, CPT and crosslinking agents [124-128]. A study with WRN-deficient mice found that embryos have a

decreased rate of survival versus wildtype mice; mice that survive embryogenesis show early evidence of myocardial fibrosis and higher incidence of cancer than mice that are wildtype or heterozygous at the WRN locus [129]. Cell lines derived from patients lacking WRN exhibit an increase in mutation rates over normal cell lines, especially in deletions [130]. Cells have an increased number of chromosome aberrations, including chromatid breaks and gaps in general and at fragile sites, both spontaneously and when treated with DNA-damaging agents. These cells have an increase in DSBs as evidenced by an increased level in γH2AX phosphorylation, both in untreated cells and in cells arrested in replication by HU [131]; the authors of the study suggest that this may be due to increased PCNA dissociation from chromatin at stalled forks in these cell lines and subsequent fork collapse, as PCNA levels in chromatin fractions are reduced in HU-treated Werner's syndrome cell lines. Whether this mechanism is true, however, remains unknown. WRN-deficient cells also have an increase in Rad51 foci, suggesting an increase in HR in these cell lines, possibly as a result of increased DSBs [131, 132].

Wildtype WRN protein interacts with a bevy of proteins from different biological pathways, suggesting a diverse role for both of its catalytic domains. The C-terminus of WRN interacts with the C-terminus of the oncoprotein p53, and WRN-deficient cells have been found to have a decrease in p53-mediated apoptosis and increase in senescence, implying a regulatory role for p53 with regards to WRN function [133, 134]. Indeed, p53 has been found to inhibit WRN exonuclease function, as well as helicase function on X-shaped junctions [135, 136].

WRN co-localizes with Mre11 after replication fork arrest in mid- to late- S phase and syndromes that are deficient in members of the MRN (Mre11/Rad50/Nbs1)

complex, including Nijmegen breakage syndrome, have a decrease in WRN relocalization to nuclear foci, suggesting that WRN may be recruited to breaks via MRN and this recruitment is needed for proper DNA repair via HR [137]. Physical interaction with Nbs1 and WRN has been confirmed, and the two proteins co-localize in response to IR and mitomycin C [138]. In addition to possibly recruiting WRN to the site of breaks being initially processed by MRN, the complex has also been found to enhance the activity of WRN helicase function *in vitro* [138]. In support of WRN's role in DSB repair via HR, WRN has also been found to co-localize with Rad51, Rad54, and Rad54b at sites of mitomycin C-induced damage [139].

In addition to being needed in HR-mediated DSB repair, WRN has been found to interact with, and unwind DNA intermediates from base excision repair (BER) [140]. This unwinding activity is regulated by Ape1, a protein active in BER that WRN physically interacts with. Ape1 inhibits the helicase function of WRN unless the complex is acted on by polβ; the suggested model for WRN's role in BER is that Ape1 binds a lesion that requires BER, and recruits WRN to the lesion. Polβ "takes" the DNA from WRN-bound Ape1, and the now catalytically-active WRN helps stimulate polβ-driven DNA repair of the lesion. Though this model is yet unconfirmed, it is clear that WRN plays some role in efficient BER repair.

WRN may also play a role in nucleotide-excision repair (NER) and base-excision repair (BER) as co-IPs have shown that the RQC domain of WRN interacts with NEIL1, a protein that repairs bases damaged by oxidation, and this interaction is most strong under conditions of oxidative stress. NEIL1 interaction inhibits both the exonuclease and helicase function of WRN, while WRN stimulates the glycosylase activity of NEIL1. The

exact role of WRN in NER is still unknown, but there is a significant increase in oxidative damage-derived base modifications in cells lacking WRN, and no additive increase in modification in cells lacking both NEIL1 and WRN, implying that NEIL1 and WRN work in the same NER pathway to prevent the persistence of damaged bases [141].

WRN's role in non-homologous end-joining (NHEJ) has also been described, though no further investigation into this role has been conducted in the last decade. It had been found that the N-terminus of WRN interacts physically with NHEJ protein Ku70, and the C-terminus interacts with Ku80 [142]. This interaction stimulates WRN exonuclease, but not helicase activity [142-144]. WRN also interacts with DNA-PKcs, which phosphorylates WRN in an unclear role and inhibits WRN helicase function in a manner that can be blocked by the Ku complex. It is suggested that WRN and DNA-PKcs form a complex where WRN is inactive until it binds Ku-bound DNA, after which WRN's helicase activity becomes active and helps process broken DNA ends for NHEJ [145]. The authors of this study also found interaction between DNA-PKcs and WRN is important at telomeric D-loops, as DNA-PKcs actually stimulates WRN helicase activity at telomeric D-loops. *In vitro*, DNA-PKcs also stimulates WRN D-loop dissolution in non-telomeric regions, though it cannot stimulate WRN-mediated dissolution of forked DNA, HJs, or G-tailed substrates. In agreement with this role at telomeric D-loops, WRN deficient cells have shorter telomeric G-tails [146].

A study in mice deficient for both WRN and Terc, which codes for telomerase RNA, also supports a role for WRN in telomere maintenance. Mutant mice exhibited aging symptoms more readily than mice expressing WRN in a Terc deficient background. Affected mice appeared to be normal in young adulthood, but began to

rapidly age shortly thereafter; these symptoms became more severe as subsequent generations ended up with shorter telomeres, implying that symptoms are due at least in part to insufficient telomere maintenance [147].

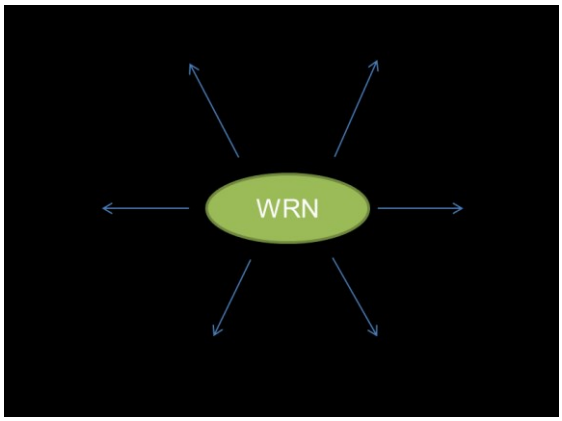


Figure 1.6: Proposed roles for WRN protein.

The WRN protein is prone to modification, the most-studied of which is phosphorylation. Phosphorylation of the protein occurs in response to replication fork blocks, and both ATR and ATM kinases have been implicated in modification of the protein in the presence of HU-stalled forks [148]. Interaction with the Abl kinase, which is active in response to genotoxic stress, has been detected in peptides containing the exonuclease and helicase domains. Abl phosphorylates tyrosine on WRN in response to bleomycin, and this phosphorylation inhibits exonuclease and helicase activities [149]. Phosphorylation by an unknown kinase also occurs at residues S440 and S467 in response to bleomycin. These residues are not in a catalytic region, and thus have not yet been found to influence the catalytic activity of the protein, but phosphomutants at

these sites are unable to relocalize WRN to the nucleoli post-damage, and repair DSBs more slowly following etoposide treatment [146].

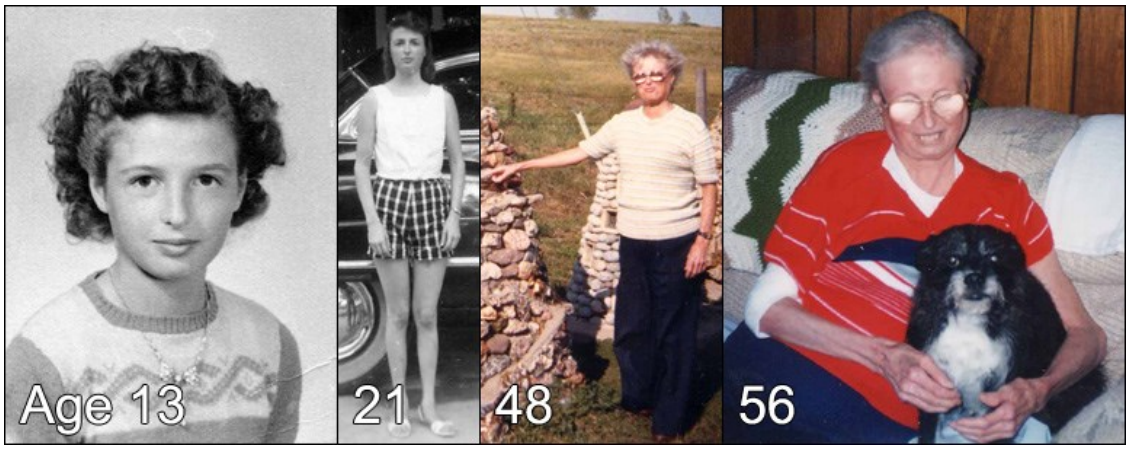


Figure 1.7: Werner's syndrome patient at various ages. Tracking of a Werner's patient show the normal aging to age 21, followed by rapid aging characteristic of the disease. Image from University of Washington.

Given the wide range of pathways and modifications of WRN, it is perhaps unsurprising those patients with Werner's Syndrome (WS) --who lack WRN-- have a range of symptoms that are the result of increased cell senescence and cancer-susceptibility. WS patients are different from sufferers of other RecQ-related syndromes in that they live relatively healthy lives for approximately the first 20 years of life, but age quickly thereafter and become more prone to age-related disease at a younger-than-expected ages, including bilateral cataracts, dermatological problems, premature greying/hair thinning, diabetes mellitus, hypogonadism, osteoporosis, premature atherosclerosis, and neoplasms (Figure 1.7) [150]. The mean age of first neoplasm in WS patients is 43 years, and patients have an excess in rare cancer types versus the

normal population, with an increase in soft-tissue sarcomas, osteosarcomas, and follicular thyroid carcinoma, amongst others [151]. These rare neoplasms are some of the most commonly encountered in the small patient pool, along with melanoma, meningioma and leukemia. The gross diversity of cancer types is perhaps reflective of the multiple pathways in which WRN is implicated.

BLM

BLM was first identified as a RecQ Helicase in 1995, when cDNA mapping verified that it contained the motifs common to the RecQ family [152]. Analysis of the gene product of this cDNA also confirmed its ATP and 3'-5' helicase activity present in the other family members [153]. Since then, many *in vitro* and *in vivo* studies of human BLM and other mammalian orthologs have tried to establish the role of BLM in maintaining genomic stability (Figure 1.8). Human and other mammalian cells lacking BLM have reduced survival in response to crosslinking agents, sensitivity to HU, MMS, etoposide, CPT, 4-NQO and UV-C irradiation, increased rates of homologous recombination resulting in non-crossovers, reduced NHEJ repair, and an increased numbers of SCEs [154-158]. Mice with mutant BLM have an increased rate in spontaneous tumor incidence, and this occurrence appears to be dosage-dependent, as homozygous mutants suffer they greatest occurrence, and heterozygous mice have a slightly elevated rate over wildtype mice [156]. Mitotic recombination in BLM-deficient mice was found to be increased in a separate study, and, as a result, an 18-fold increase in loss of heterozygosity (LOH) [159]. The effect of LOH became apparent in combination with $APC^{-/+}$ mice, which become prone to intestinal cancer if they become

$APC^{-/-}$ through aberrant crossing over between homologous chromosomes rather than sister chromatids. These mutants lose the tumor-suppressing effects of APC; mice lacking BLM in this background accumulated cancers at a greater rate than mice with WT BLM [159]. Embryogenesis of BLM-deficient mice is also impaired, as these embryos are developmentally delayed and small compared to their WT counterparts. Red cell production in these mutants is also impaired, and mutant fibroblasts incur an increased number of SCEs compared to fibroblasts from WT embryos [160].

Protein levels of BLM peak in S-phase and persist through G2/M before decreasing dramatically in G1, where repair by NHEJ takes over [161-163]. BLM protein can be found all over the nucleus, but form foci in response to damage and has been shown to co-localize with RPA, PMLs, and the nucleolus [163]. BLM catalyzes the branch migration of recombination intermediates, including dHJs, and has affinity for X-shaped DNA junctions, 3'-tailed duplex DNA, ssDNA, and G4 quadruplexes [153, 164-166]. It is also capable of resecting DNA in concert with Dna2, a function which requires helicase ability, is regulated by RPA, and enhanced by MRN, which acts upstream of BLM to resect short tracts of DNA at breaks [167, 168]. BLM catalysis is perhaps the most closely related to *S. cerevisiae* Sgs1 as it also harbors an interaction with Top3(α) and Rmi1 in the far N-terminus and is found to co-localize with these proteins *in vivo* [169, 170]. This interaction enhances BLM unwinding of dHJs, but does not affect the rate of ATP hydrolysis by BLM. Even in the presence of synthetic substrates that do not require the DNA-relaxing activity of Top3α, BLM unwinding activity of this DNA is enhanced, implying that the binding event between proteins itself may have some sort of effect on BLM processivity [170, 171].

Like WRN and other long-form RecQ helicases, BLM has been shown to interact with a bevy of different proteins acting in different pathways, implying a diverse cellular role. BLM has been shown to localize to the telomeres in cells lacking telomerase, and FRET analysis confirms interaction between BLM and telomeric protein Trf2 [172]. The same study found that overexpression of BLM in telomerase-deficient cells has been shown to increase telomere DNA synthesis, dampening the telomere-shortening effect seen in the telomerase-defective strains. This result, paired with the fact that telomere defects and shortened telomeres increase in BLM-deficient cells, suggests a role for BLM in the alternative lengthening of telomeres (ALT) pathway and general telomere maintenance [161, 172].

BLM has been detected in two large proteins complexes. The BRCA1-associated genome surveillance complex (BASC) contains Msh2, Msh6, Mlh1, ATM, BLM and the MRN complex proteins [173]. Co-localization between BRCA1 and BLM after HU exposure and partial co-localization after IR damage has been observed, implying this complex of proteins is needed to respond to DNA damage, but the functionality of this complex is yet unknown. BLM has also been found to work in complex with Top3α, RPA, Mlh1 and the Fanconi Anemia proteins FANCG, FANCC, FANCE, FANCF, and FANCA, and has been shown to co-localize and co-immunoprecipitate with FANCD2 [157]. Cells lacking the Fanconi core complex show a decrease in subnuclear relocalization of BLM to γH2AX foci following interstrand crosslink damage; this co-localization is driven by BLM phosphorylation, which is abolished when the FANC complex is absent. Cells derived from mice deficient in FANCD2 and both FANCD2 and BLM show equivalent sensitivity to crosslinking agents, suggesting a pathway overlap

for the proteins with regards to crosslink repair [174]. Deletion of both genes is not epistatic with regards to SCEs, however, suggesting that the overlap in FANC and BLM repair pathways are limited to certain types of regulation [174]. BLM has also been known to localize at ultrafine anaphase bridges and is proposed to be necessary for dissolving these bridges in anaphase to prevent chromosome fragmentation [175]. FANCD2 has also been detected on chromosomes during anaphase, and these ultrafine bridges were found to associate with FANCD2 spots, suggesting a role for the two proteins in faithful chromosome segregation [176]. While the exact relation between BLM and the FANC proteins has yet to be elucidated, it is clear that the interaction of these proteins is necessary for proper BLM response to DNA damage [157, 173].

Protein-protein interactions have also been detected with other singular proteins. SPIDR, a scaffolding protein that co-localizes and binds to BLM in response to DNA damage is required to formation of BLM foci [177]. The oncoprotein p53 physically interacts with BLM residues 1-431 between its own residues 155-393, and the localization of BLM to PML bodies in the nucleus has been found to be dependent on p53 [178, 179]. BLM and p53 have also been shown to co-localize at the site of stalled DNA replication forks and cells lacking BLM have a decrease in p53 recruited to these forks, implying that BLM is needed for p53 transport to these stalls. Thus, both play a role in regulating the location of the other in the cell [180]. Yeast-two-hybrid analysis has mapped an interaction between BLM and MLH1 to the BLM C-terminus, but cells lacking BLM have no difference in mismatch repair versus WT cells, suggesting that the interaction is not due to a role in mismatch repair, but rather a yet undescribed more specific repair [181]. Co-localization of BLM, Top3α, and Rmi1 with the PICH helicase

on ultrafine anaphase DNA bridges has been detected, and these bridges, indicative of defective sister chromatid separation, are increased in cell lines lacking BLM. The role of BLM at these bridges is unknown, but may reflect the need for the protein at unresolved DNA replication structures [175].

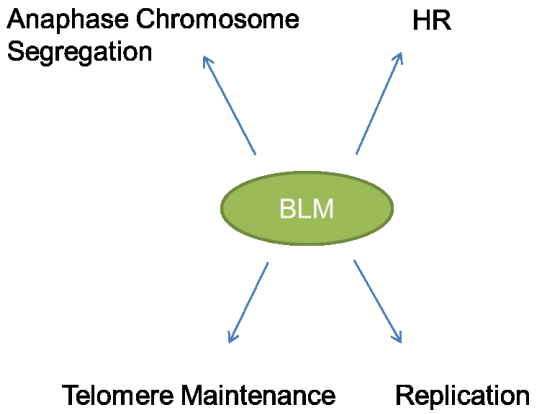


Figure 1.8: Proposed roles for BLM protein.

Like WRN, BLM is subject to extensive post-translational modification. BLM has been found to be phosphorylated by ATM in response to IR damage and by Chk1 and Chk2 in response to HU damage [154, 182]. Phosphorylation of residue T99 is increased in response to CPT and HU-induced damage, and occurs in a DNA replication-dependent manner by ATM and ATR. It is believed that this phosphorylation regulates the association of BLM with Top3α; when phosphorylated, BLM association with Top3α and PML bodies decreases, and association with γH2AX at breaks increases. This function suggests a potential role for BLM as a signaling molecule independent of its function with Top3α [183]. Residue S144 is phosphorylated by SAC

kinase MPS1; modification peaks in G2/M, suggesting a role for this modification in mitosis. Indeed, mutant S144A cells do not properly delay in mitosis in response to nocodazole and have an increase in multinucleated cells. S144A cells also have an increase in chromosome instability as illustrated by a broad distribution of chromosome number/cell vs. WT [184]. In S-phase, S338 is phosphorylated in an effort to enhance interaction with the protein TopBP1, which maintains BLM levels in the cell through G2/M before residue K3 is ubiquitinated by Mib1 in G1 to facilitate degradation of BLM in an effort to favor NHEJ DNA repair over HR [185]. A second study suggests that not S338, but S304 mediates this interaction [186]. In either case, a phosphorylation event is necessary to mediate at least one protein/protein interaction in BLM.

In humans, deficiency in BLM protein results in Bloom's syndrome, first described by dermatologist Dr. David Bloom when he described several patients with symptoms including telangiectatic erythema on the cheeks, nose, eyelid margins, lips, forehead, and ears, sun sensitivity, and stunted growth (Figure 1.9). Considering that some patients were related, he suggested that the disease has a genetic component [187, 188]. Case studies into patients listed in the Bloom's Syndrome Registry reveal that low birth weight at full term, café au lait spots, bone abnormalities, underdeveloped testes in males, irregular menses and early menopause in females, early onset diabetes (median age 24.8 years), high-pitched voices, and learning disabilities are also common symptoms [189, 190]. Of the 150 persons in the 1993 study, 118 had incidence of neoplasms; 86 of these were malignant, and mean age onset of these neoplasias was 24.4 years. Leukemias tended to predominate in children, and carcinomas in adults; lymphomas were prevalent in both age groups. Immune deficiency is also reported in

most patients [190]. In a survey of 134 persons with Bloom's Syndrome, 64 different mutations were found in 125 of them; 54 of these are premature stops, and 10 are missense mutations. Premature stops exist over the length of the protein from residue 104 to 1283. While there is no single mutation responsible for the disease, the most common mutation is the BLM^{Ash} mutation, a frame shift mutation at tyrosine 736 and prevalent in the Ashkenazi Jewish population [191]. Bloom's Syndrome is an autosomal recessive disease, suggesting a ¼ chance in inheritance from two carrier parents, but the observed rate in afflicted families is actually lower than this; the prevailing hypothesis is that BLM recessive homozygotes are prone to greater loss in embryogenesis, thus making the disease yet more rare and restricted to the 265 reported registered cases as of 2009 [189]. While heterozygotes do not have Bloom's Syndrome, it has been suggested that this genetic status results in an increased risk of cancer; a study of heterozygotes for BLM^{Ash} found that they were twice as prone to colorectal cancer than those without the mutant allele [192]. However, a human study of Jewish individuals with colorectal cancer found no significant relationship between cancer incidence and BLM^{Ash} heterozygosity [193]. A study in a functional BLM chimera in diploid yeast shows that non-Bloom's causing loss-of-function mutants in heterozygotic carriers have HU sensitivity intermediate to the wildtype and loss-of-function homozygotes [11]. This finding suggests that combination of a non-functional allele with a wildtype BLM allele could in fact impact cell survival in the presence of chronic DNA damage. This could have potential consequences for human patients with a heterozygotic loss-of-function BLM status, particularly with regards to long-term DNA damage (e.g. chemotherapy). A frameshift mutation in a polyadenine tract in BLM in

genetically unstable cancers prone to microsatellite instability was also discovered [194], and intronic SNPs in Rmi1, Top3α, and BLM have all been found to be associated with increased risk of acute myelogenous leukemia/myelodysplastic syndrome and melanoma (Rmi1 and Top3α) and bladder cancer (BLM) [194, 195]. This suggests an importance for BLM in maintaining genomic stability in the body, and that a degree of haploinsuffciency may exist with regards to certain kinds of cancer.

Figure 1.9: Patients with Bloom's syndrome. Patient on the left exhibits the classic facial telangiectasias of the disease. The middle and right patients exhibit the short stature of the disease. The middle patient grew to 138 cm in adulthood and the patient in the right photograph (147 cm) is next to his brother (183 cm), who does not have the disease. Photographs from Bloom's syndrome registry, Weill Cornell Medical College.

Intrinsically Disordered Proteins/Regions (IDP/Rs)

As the RecQ helicases participate in a wide range of functions, and, as shown in the case of Sgs1, many of the protein/protein interactions that regulate these functions are N-terminal, it is perhaps striking that little research has been done in defining the structure of the first half of these proteins as has been done for the catalytic regions. It

is well-known that the N-termini of many of the long-form RecQs contain acidic domains and that they are home to both protein interaction and modification sites, but the termini have been described as being “featureless [196].” Classically, shape and folding was believed to be the cornerstone in understanding protein function, and the so-called "lock-and-key" models were used to describe the biochemistry behind most protein-substrate interactions. While secondary and tertiary structures are still indicators of protein function for many proteins, this somewhat stringent view of structural science is moot for a rising class of proteins and protein domains. For many important proteins, including those used in cell cycle control, transcriptional and translational regulation, and protein phosphorylation, a low degree of even secondary structure is evident. These proteins are called "intrinsically disordered proteins (IDP)" or "natively unfolded" proteins [197].

Characteristics of IDPs

Many IDPs are characterized by a high degree of flexibility, low hydrophobicity, little sequence complexity, a low degree of compactness, and a large net charge. IDPs are generally rich in polar molecules (R, G, Q, E, S, P, K), two of which (P and G) have been shown to break secondary structures like alpha helices and beta sheets [197, 198]. In general, they tend to lack large hydrophobic or aromatic residues, as well as residues that promote tertiary structure (i.e. through disulfide bonds). These proteins therefore lack the sequence complexity of a more structured protein as they are depleted in several amino acids (I, L, V, W, Y, F, C, N). Because of this specialized composition, disordered proteins tend to be highly soluble in water and are generally

more heat-soluble than ordered proteins [199]. Even within the disordered proteins, there are proposed “flavors” of protein, and the composition of the different flavors (V, C, and S) vary slightly from one another [200]; type C are more enriched in H, M, and A than the others, type S has the least amount of H, and type V contain the greatest percentage of inflexible residues C, F, I, and Y versus the other types of disordered proteins. Interestingly, the role of these classes tend to differ as well, with C-type containing the most protein modification sites, S-type proteins being mostly protein-protein interactors, and V-type being mostly ribosomal proteins, thus implying a role for primary structure in protein function [200, 201]. Secondary structure in these proteins is typically confined to short segments, often helical and transient. As a whole, the “shape” of these proteins is classified as an ensemble of conformations which vary in their degree of transient secondary and tertiary structure. The protein will continuously shift between different conformations, and it is from these conformations that protein function(s) can arise [202-206]. In general, these disordered regions and proteins are believed to exist within a “structural continuum” along with folded proteins, and can generally be described as being “molten globules,” collapsed structure with no firm tertiary structure but developed secondary structure that is in a relatively fixed position; “multi-domain proteins,” which have ordered domains tied together by disordered linkers (also referred to as “beads on a string”), and “intrinsic coils,” which are mostly unfolded and have little secondary structure, if any [202, 207]. For disordered regions within proteins that have clusters of defined, folded structure, functionally disordered regions may be very short, such as 4-8 residue linkers and transient helices; they could also be large, being 10% of the protein in consecutive residues or longer [200]. Given depletion

in aromatic and hydrophobic residues that result in the lack of a folded structure, traditional techniques like X-ray crystallography are not available to IDPs. As such, alternative techniques, including NMR spectroscopy, small angle x-ray scattering (SAXS), circular dichroism, infared spectroscopy and atomic force microscopy are employed in order to determine the ensemble-averaged structures obtained from disordered proteins.

Prevalence of IDPs

A high percentage of eukaryotic proteins contain regions that are predicted to have a long stretch (greater than 30 residues) of disorder; in *S. cerevisiae*, 50-60% of the proteome is predicted to have these disordered domains, nearly three times as many disordered regions as predicted in *E. coli* [208]. Within the three kingdoms, these long stretches are predicted in 2% of Archea, 4.2% of Eubacteria, and 33% of Eukaryotic proteins. In mammals, up to 25% of the total number of proteins is predicted to be composed of completely unfolded proteins, such as Securin, Calpastatin, and Tau [209, 210], which makes study of IDP function and characteristics an important topic [200].

Thermodynamic Characteristics of IDPs

If one were to visualize the energy profile of a folded protein graphically, it could generally be described as an energy landscape ranging from conformations with high energy, to an energy minimum of the folded state in the pit of the graphical representation (Figure 1.10). Unfolded proteins represent a dense population of

conformational possibilities--transition from one disordered conformation to another incurs little energy penalty [211]. In ordered proteins, however, the transition from unfolded to folded protein is stepwise, and the energy landscape is “funnel-like.” As the protein reaches conformations of ever decreasing free energy, the conformational possibilities become less dense, as reversal from a lower energy state incurs an entropic penalty much greater than those seen in the disordered ensemble. Thus, the structural possibilities for a folded protein decrease as the folds approach the local energy minima. Upon binding, disordered proteins lose a degree of entropy, particularly if they undergo an induced-folding event; appropriate enthalpy compensation from modifications like phosphorylation can also counteract entropy loss from folding. It has also been inferred through experiments with the disordered interaction between pKID and KIX proteins that IDPs may have a lower energy barrier to overcome than an order/order interaction, as it is easier for disordered proteins to reach transitional states in the binding reaction; this also lowers the need for high entropy loss compensation [197, 212-214].

Functional Advantage of IDPs

IDPs are great for transient interactions, but given their high prevalence, particularly in higher eukaryotes, benefits beyond the transient nature likely exist. Indeed, there are several theories to the benefits of a fully- or largely-disordered protein. One potential scenario is that an ordered protein of importance in several protein pathways (i.e. kinases) has several downstream partners with disordered regions [213]; this allows the kinase to bind several partners with one single structure. On the other

hand, a disordered protein/region may have many ordered partners that it can support along the length of the disorder. The shape of the protein may change differently based on the partner being bound (or may even stay disordered, forming a “fuzzy complex [215]”). Beyond this, even these same proteins can bind different partners in the same region, using small molecular recognition motifs (MoRFs) as docking points to specifically recognize each different partner. By nature of its extended conformation, disordered proteins may also allow for a much larger interaction area with its partner than a folded partner affords; indeed, a disordered protein may well surround some or all of an ordered protein [213, 216, 217]. Beyond this variety of binding events, termed “promiscuous binding [210],” disordered proteins/regions also provide open access to post-translational modification of residues, can serve as flexible linkers that modulate the distance/position between two ordered domains, can be a stimulus for protein degradation (for transient protein expression), can contribute to molecular assemblies of proteins into large complexes, and can act as an area of rapid evolution that will not disrupt a protein’s catalytic or structural function (Figure 1.11) [200, 217-219]. A study into the putative roles and advantages of IDPs in yeast Chromatin Processes proteins can be found in chapter 4 of this dissertation.

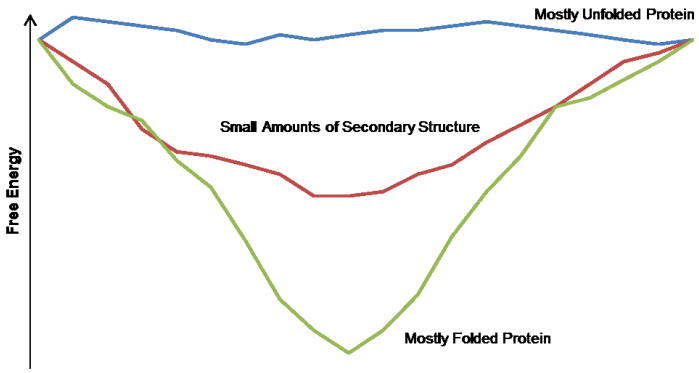


Figure 1.10: Free energy of proteins. Unfolded proteins tend to have high free energy that reaches no minimum energy conformation, while ordered proteins have a low-energy conformation termed "native." Proteins with small degrees of secondary structure vary between the energy extremes [220].

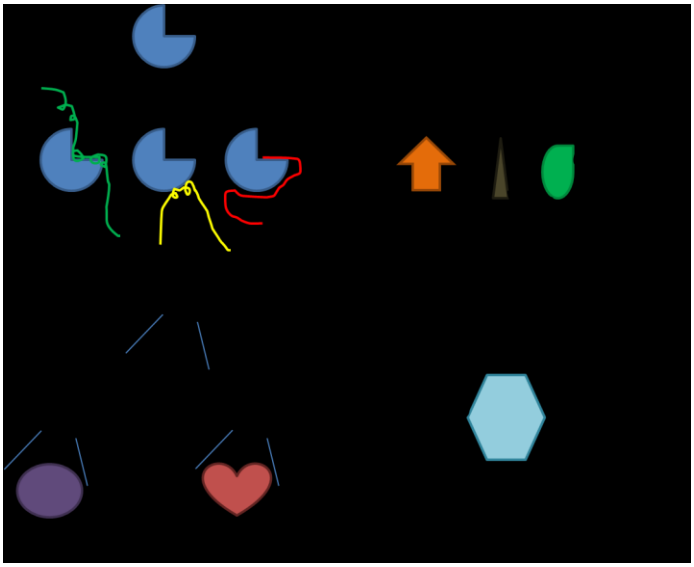


Figure 1.11: Potential roles of disorder in protein binding. Adapted from Dunker et al, 2002 [219].

Evolutionary Advantage of IDPs

IDPs present an interesting case in the study of the evolution of proteins. There are three definitions of structural conservation with regards to IDPs: flexible disorder, which conserves the disorder in orthologs but not the primary sequence, constrained disorder, which conserves both the disorder and the primary sequence, and non-conserved disorder [221]. Specific examples of flexible disorder evolution have been investigated where dynamic properties of disordered regions like linkers are conserved between orthologs, but primary sequence alignment is poor [222]. As such, it is important to consider evaluation beyond primary sequence alignments with regards to these regions; the different levels of disorder conservation are prevalent in different cell functions, with flexible disorder being most enriched in cell signaling and regulatory proteins and constrained disorder being most prevalent in proteins involving ribosome biogenesis, RNA binding, and protein folding [221]. Thus, the degree of primary sequence conservation may or may not be an important tool depending on the function of a group of orthologs. Also, while primary sequence in a great deal of IDPs tends to be poorly conserved in comparison to ordered proteins, a large percentage of them tend to conserve overall chemical composition; this allows a certain degree of flexibility in these regions with regards to acquisition of mutations and indels in these regions; function of these regions may still be retained if the mutagenesis does not affect overall chemistry [223, 224]. Indeed, long indels in homologous eukaryotic proteins have been confined mostly to the termini and disordered regions of the protein, and further disorder usually arises from disordered areas, which can more easily tolerate the modification of sequence [225, 226]. Conservation of disordered regions is particularly important with

regards to areas of post-translational modification, and the density of phosphosites in a protein highly correlates with disorder conservation (and less so with amino acid conservation) [221].

With regards to evolutionary time, it has been suggested that disorder tends to evolve more quickly than ordered regions [227]. An investigation into the evolution of disordered and ordered protein sets found that disordered proteins are generally more tolerant of evolution than ordered proteins, and evolution of individual residues tolerates a broader range of substitutions. Interestingly, this is not true of all residues in disorder; glutamine and asparagine are more prone to change in order than disorder, and tryptophan and tyrosine are often highly conserved in disordered domains [228]. Because protein/protein interaction events are also often driven by MoRFs that are not large, folded structures, interactions between partner proteins can be conserved so long as the MoRF remains, allowing for greater evolutionary change in the areas surrounding the MoRF [223]. It should be unsurprising, then, that disorder is more prevalent in eukaryotes than prokaryotes; over evolutionary time, disorder present in prokaryotes can add on more disorder to accommodate for further protein function in higher organisms. Indeed, length of disorder very closely correlates with the super-kingdoms, with eukaryota exhibiting much larger percentages of proteins with 80 or more consecutive residues in a disordered region than prokaryota and archea [229].

Examples of Intrinsic Disorder

The oncoprotein p53 has been shown to have disordered N- and C-termini that modulate several protein-protein interactions and house the majority of the protein's

PTMs, including acetylation, phosphorylation, and methylation events [230]. p53 is modulated by Mdm2 via its disordered transcription activation domain; binding regulates p53 interaction with transcriptional factors and targets p53 for destruction. As proper p53 function is important for suppressing cancer, the impact of mutated disordered regions on the ability of p53 to regulate downstream elements is in ongoing investigation [231]. Mutation of the disordered domain in the p53 transactivation domain showed that a simple P27A substitution increased binding to E3 ligase Mdm2, a result that lead to lowered expression of p53 target genes; this clearly exhibits the need for a degree of conserved disorder in proteins like p53 [232].

Tau protein in mammalian brain cells is a microtubule-associated protein that has been detected in neurofibrillary tangles in the brains of Alzheimer's Disease (AD) patients. In the normal brain, Tau protein is mostly a random coil; Tau has been found to be abnormally hyperphosphorylated in AD cases, and this modified protein is more prone to a stiffer, more partially folded conformation that tends to form the observed tangles [233-235]. Though it is currently unknown what mechanism this aggregation plays in AD onset and progression (if any), it remains a hallmark sign of the disease and the advancement of dementia in AD has been highly correlated with the degree of tangles present [236].

In yeast, ADR1, a regulator of alcohol dehydrogenase ADH2, contains a DNA-binding domain with an N-terminal disordered accessory domain and two zinc finger domains; overall, the unbound protein resembles a "beads on a string" model, where the two structured zinc finger domains lie on the disordered "string" that makes up the rest of the domain. Upon DNA binding, the region undergoes more extensive folding,

and the whole domain becomes structured. It is hypothesized that this increased folding around the fingers may increase the affinity of the region for DNA and is a good example of functional disorder in transcriptional regulation [237].

Disorder in *S. cerevisiae*

Studies in the *S. cerevisiae* proteome have estimated the percentage of proteins containing at least 30 consecutive disordered residues is between 31-54%, at least 50 consecutive disordered residues is between 19-30%, and wholly disordered proteins are up to 6% of the total proteome [209, 238]. With regards to the types of protein in the yeast genome that contain disorder, proteins are often found in the nucleus and are associated with transcription, protein modification (kinases), and binding activity, including DNA binding [238]. Disorder was also found to be a characteristic of “hub” proteins in yeast--proteins that act as a docker of several different partner proteins. Yeast hubs are enriched in wholly disordered proteins and depleted in ordered proteins, with hubs involved in protein binding being most strongly associated with predicted disorder, supporting the hypothesis that they are acting as “promiscuous binders” of several downstream protein elements; a study in promiscuous transcriptional hubs show that this specific class of protein reflects this model [239, 240].

Disorder in RecQ Helicases

Though there has been talk about the acidic domains of Sgs1 and BLM, implying consideration of the role of disorder in the RecQ family, the topic remains overall understudied. Using IUPRED prediction software [241], it is evident that the long-form

RecQs in humans and the yeast Sgs1 all contain a degree of long disorder in the N-terminus (Figure 1.12). Given that the RecQ helicases function in several different pathways as described above, and are subject to a wide array of protein/protein interactions and post-translational modifications in these N-termini, and, since these are major function of disordered proteins, disorder and other structural study in the RecQ helicases require more attention going forward. Thinking about the RecQs with regard to disorder, both flexible and non-conserved rather than traditional primary sequence alignments could potentially help elucidate some of the questions still surrounding these proteins. Chapter 3 reports a study we conducted on the N-terminal region of Sgs1 at the Top3 binding site, confirming predictions that this area is in fact, disordered. It is suggested, then, that disorder be taken into consideration when studying the RecQ proteins and their binding partner interactions. While NMR is the basis for study of disorder in that chapter, chapter 5 suggests a method for using site-directed mutagenesis as a means for targeting MoRFs in otherwise disordered regions. This can allow for the identification of important small/transient structure in these areas that may help mediate protein function and stability. Indeed, helical prediction of the first 250 residues of BLM using a bioinformatics tool for predicting alpha-helical structure, Agadir [242], reveals several putative helical regions that, based on our findings with Sgs1, may be transient helices facilitating the interaction of BLM with Top3α, Rmi1 and Rmi2 (Figure 1.13). Further study into this region may highlight areas of interest to look for SNPs in the general population that may contribute to an understanding of increased cancer risk.

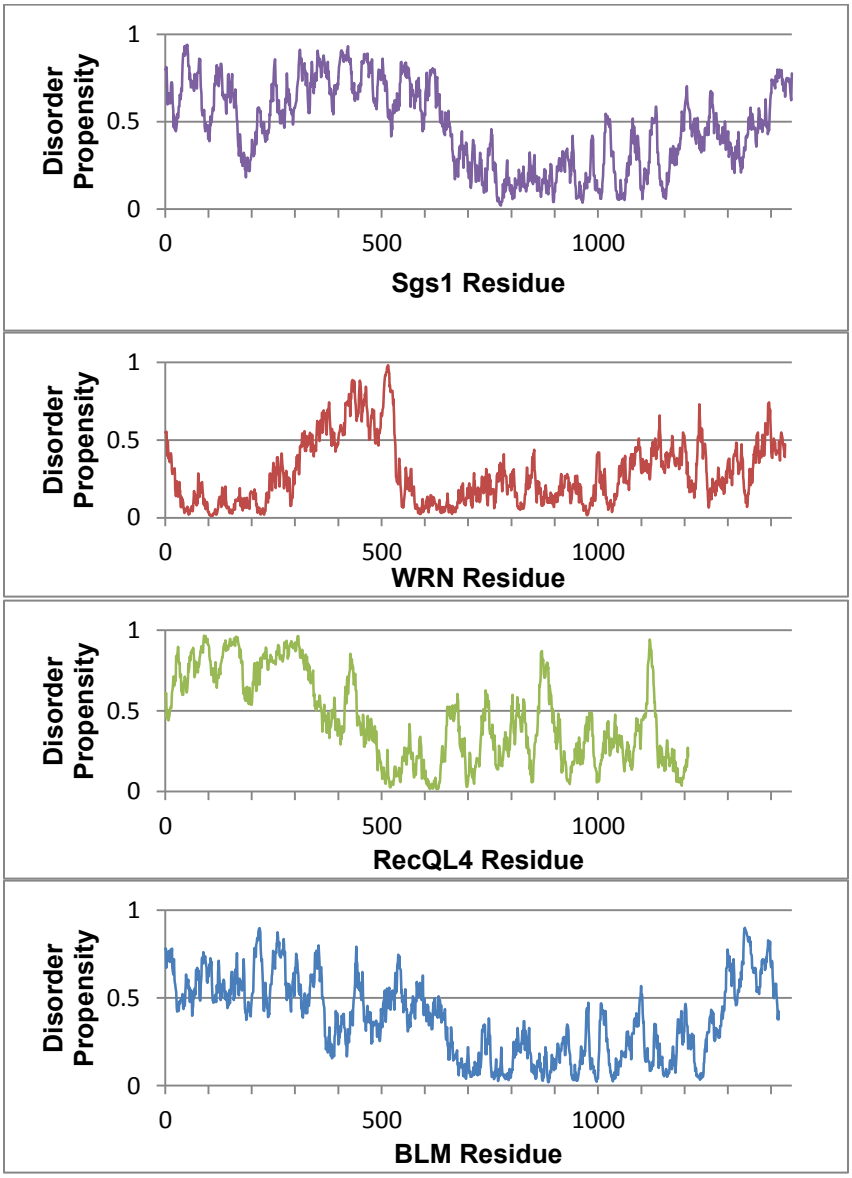


Figure 1.12: IUPRED predictions for BLM, WRN, RECQL4, and Sgs1. IUPRED [241] profiles show that long-form RecQs in humans have some degree of predicted disorder, particularly in protein/protein interaction domains.

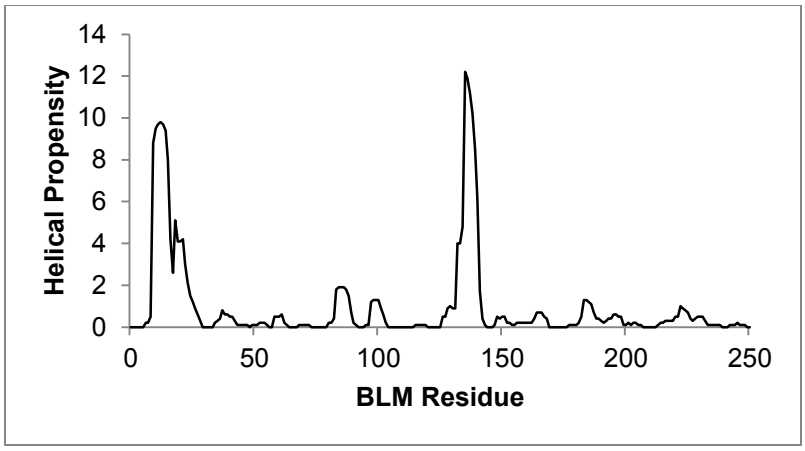


Figure 1.13: Agadir profile of BLM protein. The Agadir [242] profile of the first 250 residues of the BLM protein (Top3/Rmi1/Rmi2 interaction) reveals several putative areas of predicted alpha-helical structure for site-directed mutagenesis-based structural studies.

References

1. Bernstein, K.A., S. Gangloff, and R. Rothstein, *The RecQ DNA helicases in DNA Repair.* Annual review of genetics, 2010. **44**: p. 393-417.
2. Kadyk, L.C. and L.H. Hartwell, *Sister chromatids are preferred over homologs as substrates for recombinational repair in Saccharomyces cerevisiae.* Genetics, 1992. **132**(2): p. 387-402.
3. Hartung, F. and H. Puchta, *The RecQ gene family in plants.* Journal of Plant Physiology, 2006. **163**(3): p. 287-296.
4. Croteau, D.L., V. Popuri, P.L. Opresko, and V.A. Bohr, *Human RecQ Helicases* Biochemistry, 2014. **83**(1): p. 519-552.
5. Nakayama, K., N. Irino, and H. Nakayama, *The recQ gene of Escherichia coli K12: molecular cloning and isolation of insertion mutants.* Molecular and General Genetics MGG, 1985. **200**(2): p. 266-271.
6. Umezu, K., K. Nakayama, and H. Nakayama, *Escherichia coli RecQ protein is a DNA helicase.* Proceedings of the National Academy of Sciences, 1990. **87**(14): p. 5363-5367.
7. Harmon, F.G. and S.C. Kowalczykowski, *Biochemical Characterization of the DNA Helicase Activity of theEscherichia coli RecQ Helicase.* Journal of Biological Chemistry, 2001. **276**(1): p. 232-243.
8. Harmon, F.G. and S.C. Kowalczykowski, *RecQ helicase, in concert with RecA and SSB proteins, initiates and disrupts DNA recombination.* Genes & Development, 1998. **12**(8): p. 1134-1144.
9. Hickson, I.D., *RecQ helicases: caretakers of the genome.* Nature Reviews Cancer, 2003. **3**(3): p. 169-178.

10. Bernstein, D.A., M.C. Zittel, and J.L. Keck, *High-resolution structure of the E. coli RecQ helicase catalytic core.* The EMBO journal, 2003. **22**(19): p. 4910-4921.
11. Mirzaei, H. and K.H. Schmidt, *Non-Bloom syndrome-associated partial and total loss-of-function variants of BLM helicase.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(47): p. 19357-19362.
12. Bernstein, K.A., S. Gangloff, and R. Rothstein, *The RecQ DNA helicases in DNA repair.* Annual review of genetics, 2010. **44**: p. 393.
13. Gyimesi, M., G.M. Harami, K. Sarlós, E. Hazai, Z. Bikádi, and M. Kovács, *Complex activities of the human Bloom's syndrome helicase are encoded in a core region comprising the RecA and Zn-binding domains.* Nucleic Acids Research, 2012. **40**(9): p. 3952-3963.
14. Kitano, K., S.-Y. Kim, and T. Hakoshima, *Structural basis for DNA strand separation by the unconventional winged-helix domain of RecQ helicase WRN.* Structure, 2010. **18**(2): p. 177-187.
15. Park, C.-J., J. Ko, K.-S. Ryu, and B.-S. Choi, *Solution structure of the RecQ C-terminal domain of human Bloom syndrome protein.* Journal of Biomolecular NMR, 2014. **58**(2): p. 141-147.
16. von Kobbe, C., P. Karmakar, L. Dawut, P. Opresko, X. Zeng, R.M. Brosh, I.D. Hickson, and V.A. Bohr, *Colocalization, Physical, and Function and Interaction between Werner and Bloom Syndrome Proteins.* Journal of Biological Chemistry, 2002. **277**(24): p. 22035-22044.
17. Liu, Z., M.J. Macias, M.J. Bottomley, G. Stier, J.P. Linge, M. Nilges, P. Bork, and M. Sattler, *The three-dimensional structure of the HRDC domain and implications for the Werner and Bloom syndrome proteins.* Structure, 1999. **7**(12): p. 1557-1566.
18. Bernstein, D.A. and J.L. Keck, *Conferring substrate specificity to DNA helicases: role of the RecQ HRDC domain.* Structure, 2005. **13**(8): p. 1173-1182.
19. Wu, L., K.L. Chan, C. Ralf, D.A. Bernstein, P.L. Garcia, V.A. Bohr, A. Vindigni, P. Janscak, J.L. Keck, and I.D. Hickson, *The HRDC domain of BLM is required for the dissolution of double Holliday junctions.* The EMBO journal, 2005. **24**(14): p. 2679-2687.
20. Opresko, P.L., W.-H. Cheng, and V.A. Bohr, *Junction of RecQ helicase biochemistry and human disease.* Journal of Biological Chemistry, 2004. **279**(18): p. 18099-18102.
21. Bachrati, C. and I. Hickson, *RecQ helicases: suppressors of tumorigenesis and premature aging.* Biochem. J, 2003. **374**: p. 577-606.
22. Bennett, R.J., J.A. Sharp, and J.C. Wang, *Purification and characterization of the Sgs1 DNA helicase activity of Saccharomyces cerevisiae.* Journal of Biological Chemistry, 1998. **273**(16): p. 9644-9650.
23. Sun, H., R.J. Bennett, and N. Maizels, *The Saccharomyces cerevisiae Sgs1 helicase efficiently unwinds GG paired DNAs.* Nucleic acids research, 1999. **27**(9): p. 1978-1984.

24. Cejka, P., J.L. Plank, C.Z. Bachrati, I.D. Hickson, and S.C. Kowalczykowski, *Rmi1 stimulates decatenation of double Holliday junctions during dissolution by Sgs1-Top3.* Nat Struct Mol Biol, 2010. **17**(11): p. 1377-1382.
25. Watt, P.M., I.D. Hickson, R.H. Borts, and E.J. Louis, *SGS1, a homologue of the Bloom's and Werner's syndrome genes, is required for maintenance of genome stability in Saccharomyces cerevisiae.* Genetics, 1996. **144**(3): p. 935-945.
26. Yamagata, K., J.-i. Kato, A. Shimamoto, M. Goto, Y. Furuichi, and H. Ikeda, *Bloom's and Werner's syndrome genes suppress hyperrecombination in yeast sgs1 mutant: Implication for genomic instability in human diseases.* Proceedings of the National Academy of Sciences, 1998. **95**(15): p. 8733-8738.
27. Myung, K., A. Datta, C. Chen, and R.D. Kolodner, *SGS1, the Saccharomyces cerevisiae homologue of BLM and WRN, suppresses genome instability and homeologous recombination.* Nat Genet, 2001. **27**(1): p. 113-116.
28. Oh, S.D., J.P. Lao, P.Y.-H. Hwang, A.F. Taylor, G.R. Smith, and N. Hunter, *BLM Ortholog, Sgs1, Prevents Aberrant Crossing-over by Suppressing Formation of Multichromatid Joint Molecules.* Cell. **130**(2): p. 259-272.
29. Amin, A.D., A.B.H. Chaix, R.P. Mason, R.M. Badge, and R.H. Borts, *The Roles of the Saccharomyces cerevisiae RecQ Helicase SGS1 in Meiotic Genome Surveillance.* PLoS ONE, 2010. **5**(11): p. e15380.
30. Rockmill, B., J.C. Fung, S.S. Branda, and G.S. Roeder, *The Sgs1 helicase regulates chromosome synapsis and meiotic crossing over.* Current Biology, 2003. **13**(22): p. 1954-1962.
31. Gangloff, S., C. Soustelle, and F. Fabre, *Homologous recombination is responsible for cell death in the absence of the Sgs1 and Srs2 helicases.* Nature genetics, 2000. **25**(2): p. 192-194.
32. Parsons, A.B., R.L. Brost, H. Ding, Z. Li, C. Zhang, B. Sheikh, G.W. Brown, P.M. Kane, T.R. Hughes, and C. Boone, *Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways.* Nature biotechnology, 2004. **22**(1): p. 62-69.
33. Chang, M., M. Bellaoui, C. Boone, and G.W. Brown, *A genome-wide screen for methyl methanesulfonate-sensitive mutants reveals genes required for S phase progression in the presence of DNA damage.* Proceedings of the National Academy of Sciences, 2002. **99**(26): p. 16934-16939.
34. Sinclair, D.A., K. Mills, and L. Guarente, *Accelerated aging and nucleolar fragmentation in yeast sgs1 mutants.* Science, 1997. **277**(5330): p. 1313-1316.
35. Mimitou, E.P. and L.S. Symington, *Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing.* Nature, 2008. **455**(7214): p. 770-774.
36. Bennett, R.J., J.L. Keck, and J.C. Wang, *Binding specificity determines polarity of DNA unwinding by the Sgs1 protein of S. cerevisiae1.* Journal of Molecular Biology, 1999. **289**(2): p. 235-248.

37. Zhu, Z., W.-H. Chung, E.Y. Shim, S.E. Lee, and G. Ira, *Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double strand break ends.* Cell, 2008. **134**(6): p. 981-994.
38. Mitchel, K., K. Lehner, and S. Jinks-Robertson, *Heteroduplex DNA Position Defines the Roles of the Sgs1, Srs2, and Mph1 Helicases in Promoting Distinct Recombination Outcomes.* PLoS Genet, 2013. **9**(3): p. e1003340.
39. Ira, G., A. Malkova, G. Liberi, M. Foiani, and J.E. Haber, *Srs2 and Sgs1–Top3 suppress crossovers during double-strand break repair in yeast.* Cell, 2003. **115**(4): p. 401-411.
40. Spell, R.M. and S. Jinks-Robertson, *Examination of the Roles of Sgs1 and Srs2 Helicases in the Enforcement of Recombination Fidelity in Saccharomyces cerevisiae.* Genetics, 2004. **168**(4): p. 1855-1865.
41. Frei, C. and S.M. Gasser, *The yeast Sgs1p helicase acts upstream of Rad53p in the DNA replication checkpoint and colocalizes with Rad53p in S-phase-specific foci.* Genes & Development, 2000. **14**(1): p. 81-96.
42. Cobb, J.A., L. Bjergbaek, K. Shimada, C. Frei, and S.M. Gasser, *DNA polymerase stabilization at stalled replication forks requires Mec1 and the RecQ helicase Sgs1.* The EMBO journal, 2003. **22**(16): p. 4325-4336.
43. Cobb, J.A., T. Schleker, V. Rojas, L. Bjergbaek, J.A. Tercero, and S.M. Gasser, *Replisome instability, fork collapse, and gross chromosomal rearrangements arise synergistically from Mec1 kinase and RecQ helicase mutations.* Genes & development, 2005. **19**(24): p. 3055-3069.
44. Bjergbaek, L., J.A. Cobb, M. Tsai-Pflugfelder, and S.M. Gasser, *Mechanistically distinct roles for Sgs1p in checkpoint activation and replication fork maintenance.* The EMBO journal, 2005. **24**(2): p. 405-417.
45. Balogun, F.O., A.W. Truman, and S.J. Kron, *DNA resection proteins Sgs1 and Exo1 are required for G1 checkpoint activation in budding yeast.* DNA Repair, 2013. **12**(9): p. 751-760.
46. Hegnauer, A.M., N. Hustedt, K. Shimada, B.L. Pike, M. Vogel, P. Amsler, S.M. Rubin, F. van Leeuwen, A. Guénolé, H. van Attikum, N.H. Thomä, and S.M. Gasser, *An N-terminal acidic region of Sgs1 interacts with Rpa70 and recruits Rad53 kinase to stalled forks.* The EMBO Journal, 2012. **31**(18): p. 3768-3783.
47. Johnson, F.B., R.A. Marciniak, M. McVey, S.A. Stewart, W.C. Hahn, and L. Guarente, *The Saccharomyces cerevisiae WRN homolog Sgs1p participates in telomere maintenance in cells lacking telomerase.* The EMBO journal, 2001. **20**(4): p. 905-913.
48. Lundblad, V. and J.W. Szostak, *A mutant with a defect in telomere elongation leads to senescence in yeast.* Cell, 1989. **57**(4): p. 633-643.
49. Cohen, H. and D.A. Sinclair, *Recombination-mediated lengthening of terminal telomeric repeats requires the Sgs1 DNA helicase.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(6): p. 3174-3179.
50. Teng, S.-C. and V.A. Zakian, *Telomere-Telomere Recombination Is an Efficient Bypass Pathway for Telomere Maintenance in Saccharomyces cerevisiae.* Molecular and Cellular Biology, 1999. **19**(12): p. 8083-8093.

51. Lu, C.-Y., C.-H. Tsai, S.J. Brill, and S.-C. Teng, *Sumoylation of the BLM ortholog, Sgs1, promotes telomere–telomere recombination in budding yeast.* Nucleic acids research, 2009: p. gkp1008.
52. Chen, C.F. and S.J. Brill, *An essential DNA strand-exchange activity is conserved in the divergent N-termini of BLM orthologs.* The EMBO journal, 2010. **29**(10): p. 1713-1725.
53. Watt, P.M., E.J. Louis, R.H. Borts, and I.D. Hickson, *Sgs1: A eukaryotic homolog of E. coil RecQ that interacts with topoisomerase II in vivo and is required for faithful chromosome segregation.* Cell, 1995. **81**(2): p. 253-260.
54. Dunø, M., B. Thomsen, O. Westergaard, L. Krejci, and C. Bendixen, *Genetic analysis of the Saccharomyces cerevisiae Sgs1 helicase defines an essential function for the Sgs1-Top3 complex in the absence of SRS2 or TOP1.* Molecular and General Genetics MGG, 2000. **264**(1-2): p. 89-97.
55. Saffi, J., H. Feldmann, E.-L. Winnacker, and J.A. Henriques, *Interaction of the yeast Pso5/Rad16 and Sgs1 proteins: influences on DNA repair and aging.* Mutation Research/DNA Repair, 2001. **486**(3): p. 195-206.
56. Guzder, S.N., P. Sung, L. Prakash, and S. Prakash, *Yeast Rad7-Rad16 Complex, Specific for the Nucleotide Excision Repair of the Nontranscribed DNA Strand, Is an ATP-dependent DNA Damage Sensor.* Journal of Biological Chemistry, 1997. **272**(35): p. 21665-21668.
57. Aboussekhra, A., R. Chanet, Z. Zgaga, C. Cassier-Chauvat, M. Heude, and F. Fabre, *RADH, a gene of Saccharomyces cerevisiae encoding a putative DNA helicase involved in DNA repair. Characteristics of radH mutants and sequence of the gene.* Nucleic Acids Research, 1989. **17**(18): p. 7211-7219.
58. Chanet, R., M. Heude, A. Adjiri, L. Maloisel, and F. Fabre, *Semidominant mutations in the yeast Rad51 protein and their relationships with the Srs2 helicase.* Molecular and Cellular Biology, 1996. **16**(9): p. 4782-4789.
59. Chiolo, I., W. Carotenuto, G. Maffioletti, J.H. Petrini, M. Foiani, and G. Liberi, *Srs2 and Sgs1 DNA helicases associate with Mre11 in different subcomplexes following checkpoint activation and CDK1-mediated Srs2 phosphorylation.* Molecular and cellular biology, 2005. **25**(13): p. 5738-5751.
60. Hegnauer, A.M., N. Hustedt, K. Shimada, B.L. Pike, M. Vogel, P. Amsler, S.M. Rubin, F. van Leeuwen, A. Guénolé, H. van Attikum, N.H. Thomä, and S.M. Gasser, *An N-terminal acidic region of Sgs1 interacts with Rpa70 and recruits Rad53 kinase to stalled forks.* The EMBO Journal, 2012. **31**(18): p. 3768-3783.
61. Cejka, P., E. Cannavo, P. Polaczek, T. Masuda-Sasa, S. Pokharel, J.L. Campbell, and S.C. Kowalczykowski, *DNA end resection by Dna2-Sgs1-RPA and its stimulation by Top3-Rmi1 and Mre11-Rad50-Xrs2.* Nature, 2010. **467**(7311): p. 112-116.
62. Kim, R.A. and J.C. Wang, *Identification of the yeast TOP3 gene product as a single strand-specific DNA topoisomerase.* Journal of Biological Chemistry, 1992. **267**(24): p. 17178-17185.

63. Gangloff, S., J.P. McDonald, C. Bendixen, L. Arthur, and R. Rothstein, *The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase.* Molecular and Cellular Biology, 1994. **14**(12): p. 8391-8398.
64. Gangloff, S., B. de Massy, L. Arthur, R. Rothstein, and F. Fabre, *The essential role of yeast topoisomerase III in meiosis depends on recombination.* The EMBO Journal, 1999. **18**(6): p. 1701-1711.
65. Bocquet, N., A.H. Bizard, W. Abdulrahman, N.B. Larsen, M. Faty, S. Cavadini, R.D. Bunker, S.C. Kowalczykowski, P. Cejka, I.D. Hickson, and N.H. Thomä, *Structural and mechanistic insight into Holliday-junction dissolution by Topoisomerase IIIα and RMI1.* Nat Struct Mol Biol, 2014. **21**(3): p. 261-268.
66. Corbett, K.D. and J.M. Berger, *Structure, Molecular Mechanisms, and Evolutionary Relationships in DNA Topoisomerases.* Annual Review of Biophysics and Biomolecular Structure, 2004. **33**(1): p. 95-118.
67. Glineburg, M.R., A. Chavez, V. Agrawal, S.J. Brill, and F.B. Johnson, *Resolution by Unassisted Top3 Points to Template Switch Recombination Intermediates during DNA Replication.* Journal of Biological Chemistry, 2013. **288**(46): p. 33193-33204.
68. Harmon, F.G., R.J. DiGate, and S.C. Kowalczykowski, *RecQ Helicase and Topoisomerase III Comprise a Novel DNA Strand Passage Function: A Conserved Mechanism for Control of DNA Recombination.* Molecular Cell, 1999. **3**(5): p. 611-620.
69. Fricke, W.M., V. Kaliraman, and S.J. Brill, *Mapping the DNA topoisomerase III binding domain of the Sgs1 DNA helicase.* Journal of Biological Chemistry, 2001. **276**(12): p. 8848-8855.
70. Mullen, J.R., V. Kaliraman, and S.J. Brill, *Bipartite structure of the SGS1 DNA helicase in Saccharomyces cerevisiae.* Genetics, 2000. **154**(3): p. 1101-1114.
71. Bennett, R.J. and J.C. Wang, *Association of yeast DNA topoisomerase III and Sgs1 DNA helicase: studies of fusion proteins.* Proceedings of the National Academy of Sciences, 2001. **98**(20): p. 11108-11113.
72. Anderson, R.M. and D.A. Sinclair, *Yeast RecQ helicases: clues to DNA repair, genome stability and aging.* Molecular Mechanisms Of Werner's Syndrome, 2004: p. 78.
73. Mullen, J.R., F.S. Nallaseth, Y.Q. Lan, C.E. Slagle, and S.J. Brill, *Yeast Rmi1/Nce4 Controls Genome Stability as a Subunit of the Sgs1-Top3 Complex.* Molecular and Cellular Biology, 2005. **25**(11): p. 4476-4487.
74. Chang, M., M. Bellaoui, C. Zhang, R. Desai, P. Morozov, L. Delgado-Cruzata, R. Rothstein, G.A. Freyer, C. Boone, and G.W. Brown, *RMI1/NCE4, a suppressor of genome instability, encodes a member of the RecQ helicase/Topo III complex.* The EMBO Journal, 2005. **24**(11): p. 2024-2033.

75. Kennedy, J.A., G.W. Daughdrill, and K.H. Schmidt, *A transient α-helical molecular recognition element in the disordered N-terminus of the Sgs1 helicase is critical for chromosome stability and binding of Top3/Rmi1.* Nucleic Acids Research, 2013. **41**(22): p. 10215-10227.
76. Chen, C.-F. and S.J. Brill, *Binding and Activation of DNA Topoisomerase III by the Rmi1 Subunit.* Journal of Biological Chemistry, 2007. **282**(39): p. 28971-28979.
77. Yang, J., L. O'Donnell, D. Durocher, and G.W. Brown, *RMI1 Promotes DNA Replication Fork Progression and Recovery from Replication Fork Stress.* Molecular and Cellular Biology, 2012. **32**(15): p. 3054-3064.
78. Lai, M.S., M. Seki, A. Ui, and T. Enomoto, *Rmi1, a member of the Sgs1–Top3 complex in budding yeast, contributes to sister chromatid cohesion.* EMBO reports, 2007. **8**(7): p. 685-690.
79. Wang, F., Y. Yang, T.R. Singh, V. Busygina, R. Guo, K. Wan, W. Wang, P. Sung, A.R. Meetei, and M. Lei, *Crystal Structures of RMI1 and RMI2, Two OB-Fold Regulatory Subunits of the BLM Complex.* Structure, 2010. **18**(9): p. 1159-1170.
80. Xue, X., S. Raynard, V. Busygina, A.K. Singh, and P. Sung, *Role of Replication Protein A in Double Holliday Junction Dissolution Mediated by the BLM-Topo IIIα-RMI1-RMI2 Protein Complex.* Journal of Biological Chemistry, 2013. **288**(20): p. 14221-14227.
81. Guiraldelli, M.F., C. Eyster, and R.J. Pezza, *Genome instability and embryonic developmental defects in RMI1 deficient mice.* DNA Repair, 2013. **12**(10): p. 835-843.
82. Xu, D., R. Guo, A. Sobeck, C.Z. Bachrati, J. Yang, T. Enomoto, G.W. Brown, M.E. Hoatlin, I.D. Hickson, and W. Wang, *RMI, a new OB-fold complex essential for Bloom syndrome protein to maintain genome stability.* Genes & Development, 2008. **22**(20): p. 2843-2855.
83. Singh, T.R., A.M. Ali, V. Busygina, S. Raynard, Q. Fan, C.-h. Du, P.R. Andreassen, P. Sung, and A.R. Meetei, *BLAP18/RMI2, a novel OB-fold-containing protein, is an essential component of the Bloom helicase–double Holliday junction dissolvasome.* Genes & Development, 2008. **22**(20): p. 2856-2868.
84. Cui, S., R. Klima, A. Ochem, D. Arosio, A. Falaschi, and A. Vindigni, *Characterization of the DNA-unwinding Activity of Human RECQ1, a Helicase Specifically Stimulated by Human Replication Protein A.* Journal of Biological Chemistry, 2003. **278**(3): p. 1424-1432.
85. Puranam, K.L. and P.J. Blackshear, *Cloning and characterization of RECQL, a potential human homologue of the Escherichia coli DNA helicase RecQ.* Journal of Biological Chemistry, 1994. **269**(47): p. 29838-29845.
86. Muzzolini, L., F. Beuron, A. Patwardhan, V. Popuri, S. Cui, B. Niccolini, M. Rappas, P.S. Freemont, and A. Vindigni, *Different Quaternary Structures of Human RECQ1 Are Associated with Its Dual Enzymatic Activity.* PLoS Biol, 2007. **5**(2): p. e20.

87. Thangavel, S., R. Mendoza-Maldonado, E. Tissino, J.M. Sidorova, J. Yin, W. Wang, R.J. Monnat, A. Falaschi, and A. Vindigni, *Human RECQ1 and RECQ4 Helicases Play Distinct Roles in DNA Replication Initiation.* Molecular and Cellular Biology, 2010. **30**(6): p. 1382-1396.
88. Sharma, S., P. Phatak, A. Stortchevoi, M. Jasin, and J.R. LaRocque, *RECQ1 plays a distinct role in cellular response to oxidative DNA damage.* DNA Repair, 2012. **11**(6): p. 537-549.
89. Berti, M., A.R. Chaudhuri, S. Thangavel, S. Gomathinayagam, S. Kenig, M. Vujanovic, F. Odreman, T. Glatter, S. Graziano, R. Mendoza-Maldonado, F. Marino, B. Lucic, V. Biasin, M. Gstaiger, R. Aebersold, J.M. Sidorova, R.J. Monnat, M. Lopes, and A. Vindigni, *Human RECQ1 promotes restart of replication forks reversed by DNA topoisomerase I inhibition.* Nat Struct Mol Biol, 2013. **20**(3): p. 347-354.
90. Sharma, S. and R.M. Brosh, Jr., *Human RECQ1 Is a DNA Damage Responsive Protein Required for Genotoxic Stress Resistance and Suppression of Sister Chromatid Exchanges.* PLoS ONE, 2007. **2**(12): p. e1297.
91. Özsoy, A.Z., J.J. Sekelsky, and S.W. Matson, *Biochemical characterization of the small isoform of Drosophila melanogaster RECQ5 helicase.* Nucleic Acids Research, 2001. **29**(14): p. 2986-2993.
92. Sekelsky, J.J., M.H. Brodsky, G.M. Rubin, and R.S. Hawley, *Drosophila and human RecQ5 exist in different isoforms generated by alternative splicing.* Nucleic Acids Research, 1999. **27**(18): p. 3762-3769.
93. Shimamoto, A., K. Nishikawa, S. Kitao, and Y. Furuichi, *Human RecQ5β, a large isomer of RecQ5 DNA helicase, localizes in the nucleoplasm and interacts with topoisomerases 3α and 3β.* Nucleic Acids Research, 2000. **28**(7): p. 1647-1655.
94. Hu, Y., S. Raynard, M.G. Sehorn, X. Lu, W. Bussen, L. Zheng, J.M. Stark, E.L. Barnes, P. Chi, P. Janscak, M. Jasin, H. Vogel, P. Sung, and G. Luo, *RECQL5/Recql5 helicase regulates homologous recombination and suppresses tumor formation via disruption of Rad51 presynaptic filaments.* Genes & Development, 2007. **21**(23): p. 3073-3084.
95. Paliwal, S., R. Kanagaraj, A. Sturzenegger, K. Burdova, and P. Janscak, *Human RECQ5 helicase promotes repair of DNA double-strand breaks by synthesis-dependent strand annealing.* Nucleic Acids Research, 2014. **42**(4): p. 2380-2390.
96. Hu, Y., X. Lu, G. Zhou, E.L. Barnes, and G. Luo, *Recql5 Plays an Important Role in DNA Replication and Cell Survival after Camptothecin Treatment.* Molecular Biology of the Cell, 2009. **20**(1): p. 114-123.
97. Zheng, L., R. Kanagaraj, B. Mihaljevic, S. Schwendener, A.A. Sartori, B. Gerrits, I. Shevelev, and P. Janscak, *MRE11 complex links RECQ5 helicase to sites of DNA damage.* Nucleic Acids Research, 2009. **37**(8): p. 2645-2657.
98. Kanagaraj, R., D. Huehn, A. MacKellar, M. Menigatti, L. Zheng, V. Urban, I. Shevelev, A.L. Greenleaf, and P. Janscak, *RECQ5 helicase associates with the C-terminal repeat domain of RNA polymerase II during productive elongation phase of transcription.* Nucleic Acids Research, 2010.

99. Schwendener, S., S. Raynard, S. Paliwal, A. Cheng, R. Kanagaraj, I. Shevelev, J.M. Stark, P. Sung, and P. Janscak, *Physical Interaction of RECQ5 Helicase with RAD51 Facilitates Its Anti-recombinase Activity.* The Journal of Biological Chemistry, 2010. **285**(21): p. 15739-15745.
100. Lao, V.V., P. Welcsh, Y. Luo, K.T. Carter, S. Dzieciatkowski, S. Dintzis, J. Meza, N.E. Sarvetnick, R.J. Monnat, L.A. Loeb, and W.M. Grady, *Altered RECQ Helicase Expression in Sporadic Primary Colorectal Cancers.* Translational Oncology, 2013. **6**(4): p. 458-469.
101. Macris, M.A., L. Krejci, W. Bussen, A. Shimamoto, and P. Sung, *Biochemical characterization of the RECQ4 protein, mutated in Rothmund-Thomson syndrome.* DNA Repair, 2006. **5**(2): p. 172-180.
102. Capp, C., J. Wu, and T.-s. Hsieh, *Drosophila RecQ4 Has a 3′-5′ DNA Helicase Activity That Is Essential for Viability.* Journal of Biological Chemistry, 2009. **284**(45): p. 30845-30852.
103. Rossi, M.L., A.K. Ghosh, T. Kulikowicz, D.L. Croteau, and V.A. Bohr, *Conserved helicase domain of human RecQ4 is required for strand annealing-independent DNA unwinding.* DNA Repair, 2010. **9**(7): p. 796-804.
104. Sangrithi, M.N., J.A. Bernal, M. Madine, A. Philpott, J. Lee, W.G. Dunphy, and A.R. Venkitaraman, *Initiation of DNA Replication Requires the RECQL4 Protein Mutated in Rothmund-Thomson Syndrome.* Cell, 2005. **121**(6): p. 887-898.
105. Croteau, D.L., D.K. Singh, L. Hoh Ferrarelli, H. Lu, and V.A. Bohr, *RECQL4 in genomic instability and aging.* Trends in Genetics, 2012. **28**(12): p. 624-631.
106. Crevel, G., N. Vo, I. Crevel, S. Hamid, L. Hoa, S. Miyata, and S. Cotterill, *Drosophila RecQ4 Is Directly Involved in Both DNA Replication and the Response to UV Damage in S2 Cells.* PLoS ONE, 2012. **7**(11): p. e49505.
107. De, S., J. Kumari, R. Mudgal, P. Modi, S. Gupta, K. Futami, H. Goto, N.M. Lindor, Y. Furuichi, D. Mohanty, and S. Sengupta, *RECQL4 is essential for the transport of p53 to mitochondria in normal human cells in the absence of exogenous stress.* Journal of Cell Science, 2012. **125**(10): p. 2509-2522.
108. Jin, W., H. Liu, Y. Zhang, S. Otta, S. Plon, and L. Wang, *Sensitivity of RECQL4-deficient fibroblasts from Rothmund–Thomson syndrome patients to genotoxic agents.* Human Genetics, 2008. **123**(6): p. 643-653.
109. Ghosh, A.K., M.L. Rossi, D.K. Singh, C. Dunn, M. Ramamoorthy, D.L. Croteau, Y. Liu, and V.A. Bohr, *RECQL4, the Protein Mutated in Rothmund-Thomson Syndrome, Functions in Telomere Maintenance.* Journal of Biological Chemistry, 2012. **287**(1): p. 196-209.
110. Kitao, S., I. Ohsugi, K. Ichikawa, M. Goto, Y. Furuichi, and A. Shimamoto, *Cloning of Two New Human Helicase Genes of the RecQ Family: Biological Significance of Multiple Species in Higher Eukaryotes.* Genomics, 1998. **54**(3): p. 443-452.
111. Hoki, Y., R. Araki, A. Fujimori, T. Ohhata, H. Koseki, R. Fukumura, M. Nakamura, H. Takahashi, Y. Noda, S. Kito, and M. Abe, *Growth retardation and skin abnormalities of the Recql4-deficient mouse.* Human Molecular Genetics, 2003. **12**(18): p. 2293-2299.

112. Larizza, L., G. Roversi, and L. Volpi, *Rothmund-Thomson syndrome.* Orphanet Journal of Rare Diseases, 2010. **5**(1): p. 2.
113. Lindor, N.M., Y. Furuichi, S. Kitao, A. Shimamoto, C. Arndt, and S. Jalal, *Rothmund-Thomson syndrome due to RECQ4 helicase mutations: Report and clinical and molecular comparisons with Bloom syndrome and Werner syndrome.* American Journal of Medical Genetics, 2000. **90**(3): p. 223-228.
114. Pujol, L.A., R.P. Erickson, R.A. Heidenreich, and C. Cunniff, *Variable presentation of Rothmund-Thomson syndrome.* American Journal of Medical Genetics, 2000. **95**(3): p. 204-207.
115. Siitonen, H.A., J. Sotkasiira, M. Biervliet, A. Benmansour, Y. Capri, V. Cormier-Daire, B. Crandall, K. Hannula-Jouppi, R. Hennekam, D. Herzog, K. Keymolen, M. Lipsanen-Nyman, P. Miny, S.E. Plon, S. Riedl, A. Sarkar, F.R. Vargas, A. Verloes, L.L. Wang, H. Kaariainen, and M. Kestila, *The mutation spectrum in RECQL4 diseases.* Eur J Hum Genet, 2008. **17**(2): p. 151-158.
116. Jam, K., M. Fox, and B.F. Crandall, *RAPADILINO syndrome: A multiple malformation syndrome with radial and patellar aplasia.* Teratology, 1999. **60**(1): p. 37-38.
117. Dallapiccola, B.e.a., *Baller-Gerold syndrome: case report and clinical and radiological review.* Am J Med Genet, 1992. **42**(3): p. 365-8.
118. Bartyik, K., K.M. Gábor, B. Iványi, I. Németh, and E. Karg, *Rothmund-thomson syndrome and cutan T-cell lymphoma in childhood.* 2013.
119. Wang, L.L., M.L. Levy, R.A. Lewis, M.M. Chintagumpala, D. Lev, M. Rogers, and S.E. Plon, *Clinical manifestations in a cohort of 41 Rothmund-Thomson syndrome patients.* American Journal of Medical Genetics, 2001. **102**(1): p. 11-17.
120. Mohaghegh, P., J.K. Karow, R.M. Brosh Jr, V.A. Bohr, and I.D. Hickson, *The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases.* Nucleic Acids Research, 2001. **29**(13): p. 2843-2849.
121. Brosh, R.M., D.K. Orren, J.O. Nehlin, P.H. Ravn, M.K. Kenny, A. Machwe, and V.A. Bohr, *Functional and Physical Interaction between WRN Helicase and Human Replication Protein A.* Journal of Biological Chemistry, 1999. **274**(26): p. 18341-18350.
122. al, H.e., *The premature ageing syndrome protein, WRN, is a 3'right arrow5' exonuclease.* Nature Genetics, 1998. **20**: p. 114-116.
123. Opresko, P.L., J.-P. Laine, R.M. Brosh, M.M. Seidman, and V.A. Bohr, *Coordinate Action of the Helicase and 3' to 5' Exonuclease of Werner Syndrome Protein.* Journal of Biological Chemistry, 2001. **276**(48): p. 44677-44687.
124. Sidorova, J.M., K. Kehrli, F. Mao, and R. Monnat Jr, *Distinct functions of human RECQ helicases WRN and BLM in replication fork recovery and progression after hydroxyurea-induced stalling.* DNA Repair, 2013. **12**(2): p. 128-139.

125. Saintigny, Y., K. Makienko, C. Swanson, M.J. Emond, and J.R.J. Monnat, *Homologous Recombination Resolution Defect in Werner Syndrome.* Molecular and Cellular Biology, 2002. **22**(20): p. 6971-6978.
126. Pichierri, P., A. Franchitto, P. Mosesso, and F. Palitti, *Werner's Syndrome Protein Is Required for Correct Recovery after Replication Arrest and DNA Damage Induced in S-Phase of Cell Cycle.* Molecular Biology of the Cell, 2001. **12**(8): p. 2412-2421.
127. Machwe, A., E. Lozada, M.S. Wold, G.-M. Li, and D.K. Orren, *Molecular Cooperation between the Werner Syndrome Protein and Replication Protein A in Relation to Replication Fork Blockage.* The Journal of Biological Chemistry, 2011. **286**(5): p. 3497-3508.
128. Dhillon, K.K., J. Sidorova, Y. Saintigny, M. Poot, K. Gollahon, P.S. Rabinovitch, and R.J. Monnat, *Functional role of the Werner syndrome RecQ helicase in human fibroblasts.* Aging Cell, 2007. **6**(1): p. 53-61.
129. Lebel, M. and P. Leder, *A deletion within the murine Werner syndrome helicase induces sensitivity to inhibitors of topoisomerase and loss of cellular proliferative capacity.* Proceedings of the National Academy of Sciences, 1998. **95**(22): p. 13097-13102.
130. Fukuchi, K., G.M. Martin, and R.J. Monnat, *Mutator phenotype of Werner syndrome is characterized by extensive deletions.* Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(15): p. 5893-5897.
131. Franchitto, A., L.M. Pirzio, E. Prosperi, O. Sapora, M. Bignami, and P. Pichierri, *Replication fork stalling in WRN-deficient cells is overcome by prompt activation of a MUS81-dependent pathway.* The Journal of Cell Biology, 2008. **183**(2): p. 241-252.
132. Pirzio, L.M., P. Pichierri, M. Bignami, and A. Franchitto, *Werner syndrome helicase activity is essential in maintaining fragile site stability.* The Journal of Cell Biology, 2008. **180**(2): p. 305-314.
133. Spillare, E.A., A.I. Robles, X.W. Wang, J.-C. Shen, C.-E. Yu, G.D. Schellenberg, and C.C. Harris, *p53-Mediated apoptosis is attenuated in Werner syndrome cells.* Genes & Development, 1999. **13**(11): p. 1355-1360.
134. Blander, G., J. Kipnis, J.F.M. Leal, C.-E. Yu, G.D. Schellenberg, and M. Oren, *Physical and Functional Interaction between p53 and the Werner's Syndrome Protein.* Journal of Biological Chemistry, 1999. **274**(41): p. 29463-29469.
135. Brosh, R.M., P. Karmakar, J.A. Sommers, Q. Yang, X.W. Wang, E.A. Spillare, C.C. Harris, and V.A. Bohr, *p53 Modulates the Exonuclease Activity of Werner Syndrome Protein.* Journal of Biological Chemistry, 2001. **276**(37): p. 35093-35102.
136. Yang, Q., R. Zhang, X.W. Wang, E.A. Spillare, S.P. Linke, D. Subramanian, J.D. Griffith, J.L. Li, I.D. Hickson, J.C. Shen, L.A. Loeb, S.J. Mazur, E. Appella, R.M. Brosh, P. Karmakar, V.A. Bohr, and C.C. Harris, *The Processing of Holliday Junctions by BLM and WRN Helicases Is Regulated by p53.* Journal of Biological Chemistry, 2002. **277**(35): p. 31980-31987.

137. Franchitto, A. and P. Pichierri, *Werner Syndrome Protein and the MRE11 Complex are Involved in a Common Pathway of Replication Fork Recovery.* Cell Cycle, 2004. **3**(10): p. 1331-1339.
138. Cheng, W.-H., C. von Kobbe, P.L. Opresko, L.M. Arthur, K. Komatsu, M.M. Seidman, J.P. Carney, and V.A. Bohr, *Linkage between Werner Syndrome Protein and the Mre11 Complex via Nbs1.* Journal of Biological Chemistry, 2004. **279**(20): p. 21169-21176.
139. Otterlei, M., P. Bruheim, B. Ahn, W. Bussen, P. Karmakar, K. Baynton, and V.A. Bohr, *Werner syndrome protein participates in a complex with RAD51, RAD54, RAD54B and ATR in response to ICL-induced replication arrest.* Journal of Cell Science, 2006. **119**(24): p. 5137-5146.
140. Ahn, B., J.A. Harrigan, F.E. Indig, D.M. Wilson, and V.A. Bohr, *Regulation of WRN Helicase Activity in Human Base Excision Repair.* Journal of Biological Chemistry, 2004. **279**(51): p. 53465-53474.
141. Das, A., I. Boldogh, J.W. Lee, J.A. Harrigan, M.L. Hegde, J. Piotrowski, N. de Souza Pinto, W. Ramos, M.M. Greenberg, T.K. Hazra, S. Mitra, and V.A. Bohr, *The Human Werner Syndrome Protein Stimulates Repair of Oxidative DNA Base Damage by the DNA Glycosylase NEIL1.* Journal of Biological Chemistry, 2007. **282**(36): p. 26591-26602.
142. Cooper, M.P., A. Machwe, D.K. Orren, R.M. Brosh, D. Ramsden, and V.A. Bohr, *Ku complex interacts with and stimulates the Werner protein.* Genes & Development, 2000. **14**(8): p. 907-912.
143. Li, B. and L. Comai, *Functional Interaction between Ku and the Werner Syndrome Protein in DNA End Processing.* Journal of Biological Chemistry, 2000. **275**(37): p. 28349-28352.
144. Karmakar, P., C.M. Snowden, D.A. Ramsden, and V.A. Bohr, *Ku heterodimer binds to both ends of the Werner protein and functional interaction occurs at the Werner N-terminus.* Nucleic Acids Research, 2002. **30**(16): p. 3583-3591.
145. Yannone, S.M., S. Roy, D.W. Chan, M.B. Murphy, S. Huang, J. Campisi, and D.J. Chen, *Werner Syndrome Protein Is Regulated and Phosphorylated by DNA-dependent Protein Kinase.* Journal of Biological Chemistry, 2001. **276**(41): p. 38242-38248.
146. Kusumoto-Matsuo, R., P.L. Opresko, D. Ramsden, H. Tahara, and V.A. Bohr, *Cooperation of DNA-PKcs and WRN helicase in the maintenance of telomeric D-loops.* Aging (Albany NY), 2010. **2**(5): p. 274-284.
147. Chang, S., A.S. Multani, N.G. Cabrera, M.L. Naylor, P. Laud, D. Lombard, S. Pathak, L. Guarente, and R.A. DePinho, *Essential role of limiting telomeres in the pathogenesis of Werner syndrome.* Nature genetics, 2004. **36**(8): p. 877-882.
148. Pichierri, P., F. Rosselli, and A. Franchitto, *Werner's syndrome protein is phosphorylated in an ATR//ATM-dependent manner following replication arrest and DNA damage induced during the S phase of the cell cycle.* Oncogene, 0000. **22**(10): p. 1491-1500.

149. Cheng, W.-H., C.v. Kobbe, P.L. Opresko, K.M. Fields, J. Ren, D. Kufe, and V.A. Bohr, *Werner Syndrome Protein Phosphorylation by Abl Tyrosine Kinase Regulates Its Activity and Distribution.* Molecular and Cellular Biology, 2003. **23**(18): p. 6385-6395.
150. Lauper, J.M., A. Krause, T.L. Vaughan, and R.J. Monnat, Jr., *Spectrum and Risk of Neoplasia in Werner Syndrome: A Systematic Review.* PLoS ONE, 2013. **8**(4): p. e59709.
151. Goto, M., R.W. Miller, Y. Ishikawa, and H. Sugano, *Excess of rare cancers in Werner syndrome (adult progeria).* Cancer Epidemiology Biomarkers & Prevention, 1996. **5**(4): p. 239-246.
152. Ellis, N.A., J. Groden, T.-Z. Ye, J. Straughen, D.J. Lennon, S. Ciocci, M. Proytcheva, and J. German, *The Bloom's syndrome gene product is homologous to RecQ helicases.* Cell, 1995. **83**(4): p. 655-666.
153. Karow, J.K., R.K. Chakraverty, and I.D. Hickson, *The Bloom's Syndrome Gene Product Is a 3'-5' DNA Helicase.* Journal of Biological Chemistry, 1997. **272**(49): p. 30611-30614.
154. Ababou, M., S. Dutertre, Y. Lécluse, R. Onclercq, B. Chatton, and M. Amor-Guéret, *ATM-dependent phosphorylation and accumulation of endogenous BLM protein in response to ionizing radiation.* Oncogene, 2000. **19**(52): p. 5955-5963.
155. Wang, Y., K. Smith, B.C. Waldman, and A.S. Waldman, *Depletion of the bloom syndrome helicase stimulates homology-dependent repair at double-strand breaks in human chromosomes.* DNA Repair, 2011. **10**(4): p. 416-426.
156. McDaniel, L.D., N. Chester, M. Watson, A.D. Borowsky, P. Leder, and R.A. Schultz, *Chromosome instability and tumor predisposition inversely correlate with BLM protein levels.* DNA Repair, 2003. **2**(12): p. 1387-1404.
157. Pichierri, P., A. Franchitto, and F. Rosselli, *BLM and the FANC proteins collaborate in a common pathway in response to stalled replication forks.* The EMBO Journal, 2004. **23**(15): p. 3154-3163.
158. Imamura, O., K. Fujita, C. Itoh, S. Takeda, Y. Furuichi, and T. Matsumoto, *Werner and Bloom helicases are involved in DNA repair in a complementary fashion.* Oncogene, 2002. **21**(6): p. 954-963.
159. Luo, G., I.M. Santoro, L.D. McDaniel, I. Nishijima, M. Mills, H. Youssoufian, H. Vogel, R.A. Schultz, and A. Bradley, *Cancer predisposition caused by elevated mitotic recombination in Bloom mice.* Nat Genet, 2000. **26**(4): p. 424-429.
160. Chester, N., F. Kuo, C. Kozak, C.D. O'Hara, and P. Leder, *Stage-specific apoptosis, developmental delay, and embryonic lethality in mice homozygous for a targeted disruption in the murine Bloom's syndrome gene.* Genes & Development, 1998. **12**(21): p. 3382-3393.
161. Barefield, C. and J. Karlseder, *The BLM helicase contributes to telomere maintenance through processing of late-replicating intermediate structures.* Nucleic Acids Research, 2012. **40**(15): p. 7358-7367.

162. Dutertre, S., M. Ababou, R. Onclercq, J. Delic, B. Chatton, C. Jaulin, and M. Amor-Guéret, *Cell cycle regulation of the endogenous wild type Bloom's syndrome DNA helicase.* Oncogene, 2000. **19**(23): p. 2731-2738.
163. Sanz, M.M., M. Proytcheva, N.A. Ellis, W.K. Holloman, and J. German, *BLM, the Bloom's syndrome protein, varies during the cell cycle in its amount, distribution, and co-localization with other nuclear proteins.* Cytogenetics and cell genetics, 2000. **91**(1-4): p. 217-223.
164. Karow, J.K., R.H. Newman, P.S. Freemont, and I.D. Hickson, *Oligomeric ring structure of the Bloom's syndrome helicase.* Current Biology, 1999. **9**(11): p. 597-600.
165. Huber, M.D., D.C. Lee, and N. Maizels, *G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition.* Nucleic Acids Research, 2002. **30**(18): p. 3954-3961.
166. Bachrati, C.Z., R.H. Borts, and I.D. Hickson, *Mobile D-loops are a preferred substrate for the Bloom's syndrome helicase.* Nucleic acids research, 2006. **34**(8): p. 2269-2279.
167. Nimonkar, A.V., J. Genschel, E. Kinoshita, P. Polaczek, J.L. Campbell, C. Wyman, P. Modrich, and S.C. Kowalczykowski, *BLM–DNA2–RPA–MRN and EXO1–BLM–RPA–MRN constitute two DNA end resection machineries for human DNA break repair.* Genes & Development, 2011. **25**(4): p. 350-362.
168. Brosh, R.M., J.-L. Li, M.K. Kenny, J.K. Karow, M.P. Cooper, R.P. Kureekattil, I.D. Hickson, and V.A. Bohr, *Replication Protein A Physically Interacts with the Bloom's Syndrome Protein and Stimulates Its Helicase Activity.* Journal of Biological Chemistry, 2000. **275**(31): p. 23500-23508.
169. Johnson, F.B., D.B. Lombard, N.F. Neff, M.-A. Mastrangelo, W. Dewolf, N.A. Ellis, R.A. Marciniak, Y. Yin, R. Jaenisch, and L. Guarente, *Association of the Bloom Syndrome Protein with Topoisomerase IIIα in Somatic and Meiotic Cells.* Cancer Research, 2000. **60**(5): p. 1162-1167.
170. Raynard, S., W. Bussen, and P. Sung, *A Double Holliday Junction Dissolvasome Comprising BLM, Topoisomerase IIIα, and BLAP75.* Journal of Biological Chemistry, 2006. **281**(20): p. 13861-13864.
171. Bussen, W., S. Raynard, V. Busygina, A.K. Singh, and P. Sung, *Holliday Junction Processing Activity of the BLM-Topo IIIα-BLAP75 Complex.* Journal of Biological Chemistry, 2007. **282**(43): p. 31484-31492.
172. Stavropoulos, D.J., P.S. Bradshaw, X. Li, I. Pasic, K. Truong, M. Ikura, M. Ungrin, and M.S. Meyn, *The Bloom syndrome helicase BLM interacts with TRF2 in ALT cells and promotes telomeric DNA synthesis.* Human Molecular Genetics, 2002. **11**(25): p. 3135-3144.
173. Meetei, A.R., S. Sechi, M. Wallisch, D. Yang, M.K. Young, H. Joenje, M.E. Hoatlin, and W. Wang, *A Multiprotein Nuclear Complex Connects Fanconi Anemia and Bloom Syndrome.* Molecular and Cellular Biology, 2003. **23**(10): p. 3417-3426.

174. Hemphill, A.W., Y. Akkari, A.H. Newell, R.A. Schultz, M. Grompe, P.S. North, I.D. Hickson, P.M. Jakobs, S. Rennie, D. Pauw, J. Hejna, S.B. Olson, and R.E. Moses, *Topo IIIα and BLM Act within the Fanconi Anemia Pathway in Response to DNA-Crosslinking Agents.* Cytogenetic and Genome Research, 2009. **125**(3): p. 165-175.
175. Chan, K.-L., P.S. North, and I.D. Hickson, *BLM is required for faithful chromosome segregation and its localization defines a class of ultrafine anaphase bridges.* The EMBO Journal, 2007. **26**(14): p. 3397-3409.
176. Naim, V. and F. Rosselli, *The FANC pathway and BLM collaborate during mitosis to prevent micro-nucleation and chromosome abnormalities.* Nat Cell Biol, 2009. **11**(6): p. 761-768.
177. Wan, L., J. Han, T. Liu, S. Dong, F. Xie, H. Chen, and J. Huang, *Scaffolding protein SPIDR/KIAA0146 connects the Bloom syndrome helicase with homologous recombination repair.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(26): p. 10646-10651.
178. Garkavtsev, I.V., N. Kley, I.A. Grigorian, and A.V. Gudkov, *The Bloom syndrome protein interacts and cooperates with p53 in regulation of transcription and cell growth control.* Oncogene, 2001. **20**(57): p. 8276-8280.
179. Wang, X.W., A. Tseng, N.A. Ellis, E.A. Spillare, S.P. Linke, A.I. Robles, H. Seker, Q. Yang, P. Hu, S. Beresten, N.A. Bemmels, S. Garfield, and C.C. Harris, *Functional Interaction of p53 and BLM DNA Helicase in Apoptosis.* Journal of Biological Chemistry, 2001. **276**(35): p. 32948-32955.
180. Sengupta, S., S.P. Linke, R. Pedeux, Q. Yang, J. Farnsworth, S.H. Garfield, K. Valerie, J.W. Shay, N.A. Ellis, B. Wasylyk, and C.C. Harris, *BLM helicase-dependent transport of p53 to sites of stalled DNA replication forks modulates homologous recombination.* The EMBO Journal, 2003. **22**(5): p. 1210-1222.
181. Langland, G., J. Kordich, J. Creaney, K.H. Goss, K. Lillard-Wetherell, K. Bebenek, T.A. Kunkel, and J. Groden, *The Bloom's Syndrome Protein (BLM) Interacts with MLH1 but Is Not Required for DNA Mismatch Repair.* Journal of Biological Chemistry, 2001. **276**(32): p. 30031-30035.
182. Kaur, S., P. Modi, V. Srivastava, R. Mudgal, S. Tikoo, P. Arora, D. Mohanty, and S. Sengupta, *Chk1-Dependent Constitutive Phosphorylation of BLM Helicase at Serine 646 Decreases after DNA Damage.* Molecular Cancer Research, 2010. **8**(9): p. 1234-1247.
183. Rao, V.A., A.M. Fan, L. Meng, C.F. Doe, P.S. North, I.D. Hickson, and Y. Pommier, *Phosphorylation of BLM, Dissociation from Topoisomerase IIIα, and Colocalization with γ-H2AX after Topoisomerase I-Induced Replication Damage.* Molecular and Cellular Biology, 2005. **25**(20): p. 8925-8937.
184. Leng, M., D.W. Chan, H. Luo, C. Zhu, J. Qin, and Y. Wang, *MPS1-dependent mitotic BLM phosphorylation is important for chromosome stability.* Proceedings of the National Academy of Sciences, 2006. **103**(31): p. 11485-11490.

185. Wang, J., J. Chen, and Z. Gong, *TopBP1 Controls BLM Protein Level to Maintain Genome Stability.* Molecular Cell, 2013. **52**(5): p. 667-678.
186. Blackford, Andrew N., J. Nieminuszczy, Rebekka A. Schwab, Y. Galanty, Stephen P. Jackson, and W. Niedzwiedz, *TopBP1 Interacts with BLM to Maintain Genome Stability but Is Dispensable for Preventing BLM Degradation.* Molecular Cell, 2015. **57**(6): p. 1133-1141.
187. BLOOM, D., *Congenital telangiectatic erythema resembling lupus erythematosus in dwarfs: Probably a syndrome entity.* AMA American journal of diseases of children, 1954. **88**(6): p. 754-758.
188. Bloom, D., *The syndrome of congenital telangiectaticerythema and stunted growth.* The Journal of pediatrics, 1966. **68**(1): p. 103-113.
189. German, J., *Bloom's syndrome. I. Genetical and clinical observations in the first twenty-seven patients.* American Journal of Human Genetics, 1969. **21**(2): p. 196-227.
190. German, J., *Bloom Syndrome: A Mendelian Prototype of Somatic Mutational Disease.* Medicine, 1993. **72**(6): p. 393-406.
191. German, J., M.M. Sanz, S. Ciocci, T.Z. Ye, and N.A. Ellis, *Syndrome-causing mutations of the BLM gene in persons in the Bloom's Syndrome Registry.* Human Mutation, 2007. **28**(8): p. 743-753.
192. Gruber, S.B., N.A. Ellis, G. Rennert, K. Offit, K.K. Scott, R. Almog, P. Kolachana, J.D. Bonner, T. Kirchhoff, L.P. Tomsho, K. Nafa, H. Pierce, M. Low, J. Satagopan, H. Rennert, H. Huang, J.K. Greenson, J. Groden, B. Rapaport, J. Shia, S. Johnson, P.K. Gregersen, C.C. Harris, and J. Boyd, *BLM Heterozygosity and the Risk of Colorectal Cancer.* Science, 2002. **297**(5589): p. 2013-2013.
193. Cleary, S.P., W. Zhang, N. Di Nicola, M. Aronson, J. Aube, A. Steinman, R. Haddad, M. Redston, S. Gallinger, S.A. Narod, and R. Gryfe, *Heterozygosity for the BLMAsh Mutation and Cancer Risk.* Cancer Research, 2003. **63**(8): p. 1769-1771.
194. Broberg, K., E. Huynh, K.S. Engström, J. Björk, M. Albin, C. Ingvar, H. Olsson, and M. Höglund, *Association between polymorphisms in RMI1, TOP3A, and BLM and risk of cancer, a case-control study.* BMC cancer, 2009. **9**(1): p. 140.
195. Calin, G., V. Herlea, G. Barbanti-Brodano, and M. Negrini, *The Coding Region of the Bloom Syndrome BLM Gene and of the CBL Proto-Oncogene Is Mutated in Genetically Unstable Sporadic Gastrointestinal Tumors.* Cancer Research, 1998. **58**(17): p. 3777-3781.
196. Vindigni, A. and I.D. Hickson, *RecQ helicases: multiple structures for multiple functions?* HFSP Journal, 2009. **3**(3): p. 153-164.
197. Uversky, V.N., *Intrinsically disordered proteins from A to Z.* The International Journal of Biochemistry & Cell Biology, 2011. **43**(8): p. 1090-1103.
198. Richardson, J.S., *The anatomy and taxonomy of protein structure*. Vol. 34. 1981: Academic Press.

199. van der Lee, R., M. Buljan, B. Lang, R.J. Weatheritt, G.W. Daughdrill, A.K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D.T. Jones, P.M. Kim, R.W. Kriwacki, C.J. Oldfield, R.V. Pappu, P. Tompa, V.N. Uversky, P.E. Wright, and M.M. Babu, *Classification of Intrinsically Disordered Regions and Proteins.* Chemical Reviews, 2014. **114**(13): p. 6589-6631.
200. Dunker, A.K., I. Silman, V.N. Uversky, and J.L. Sussman, *Function and structure of inherently disordered proteins.* Current opinion in structural biology, 2008. **18**(6): p. 756-764.
201. Vucetic, S., C.J. Brown, A.K. Dunker, and Z. Obradovic, *Flavors of protein disorder.* Proteins: Structure, Function, and Bioinformatics, 2003. **52**(4): p. 573-584.
202. Dyson, H.J. and P.E. Wright, *Intrinsically unstructured proteins and their functions.* Nat Rev Mol Cell Biol, 2005. **6**(3): p. 197-208.
203. Uversky, V.N., J.R. Gillespie, and A.L. Fink, *Why are "natively unfolded" proteins unstructured under physiologic conditions?* Proteins: Structure, Function, and Bioinformatics, 2000. **41**(3): p. 415-427.
204. Daughdrill, G.W., *Determining structural ensembles for intrinsically disordered proteins.* Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure And Conformation, 2010: p. 107-129.
205. Tompa, P., M. Fuxreiter, C.J. Oldfield, I. Simon, A.K. Dunker, and V.N. Uversky, *Close encounters of the third kind: disordered domains and the nteractions of proteins.* BioEssays, 2009. **31**(3): p. 328-335.
206. Eliezer, D., *Biophysical characterization of intrinsically disordered proteins.* Current Opinion in Structural Biology, 2009. **19**(1): p. 23-30.
207. Uversky, V.N. and A.K. Dunker, *Understanding protein non-folding.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2010. **1804**(6): p. 1231-1264.
208. Tompa, P., Z. Dosztányi, and I. Simon, *Prevalent Structural Disorder in E. coli and S. cerevisiae Proteomes.* Journal of Proteome Research, 2006. **5**(8): p. 1996-2000.
209. Dunker, A.K., P. Romero, Z. Obradovic, E.C. Garner, and C.J. Brown, *Intrinsic Protein Disorder in Complete Genomes.* Genome Informatics, 2000. **11**: p. 161-171.
210. Tompa, P., *Intrinsically unstructured proteins.* Trends in Biochemical Sciences, 2002. **27**(10): p. 527-533.
211. Dill, K.A. and J.L. MacCallum, *The Protein-Folding Problem, 50 Years On.* Science, 2012. **338**(6110): p. 1042-1046.
212. Huang, Y. and Z. Liu, *Kinetic Advantage of Intrinsically Disordered Proteins in Coupled Folding–Binding Process: A Critical Assessment of the "Fly-Casting" Mechanism.* Journal of Molecular Biology, 2009. **393**(5): p. 1143-1159.
213. Dyson, H.J., *Expanding the proteome: disordered and alternatively-folded proteins.* Quarterly reviews of biophysics, 2011. **44**(4): p. 467-518.
214. Kovacs, D., B. Szabo, R. Pancsa, and P. Tompa, *Intrinsically disordered proteins undergo and assist folding transitions in the proteome.* Archives of Biochemistry and Biophysics, 2013. **531**(1–2): p. 80-89.

215. Tompa, P. and M. Fuxreiter, *Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions.* Trends in Biochemical Sciences, 2008. **33**(1): p. 2-8.
216. Mittag, T., L.E. Kay, and J.D. Forman-Kay, *Protein dynamics and conformational disorder in molecular recognition.* Journal of Molecular Recognition, 2010. **23**(2): p. 105-116.
217. Fuxreiter, M., P. Tompa, and I. Simon, *Local structural disorder imparts plasticity on linear motifs.* Bioinformatics, 2007. **23**(8): p. 950-956.
218. Teilum, K., J. Olsen, and B. Kragelund, *Functional aspects of protein flexibility.* Cellular and Molecular Life Sciences, 2009. **66**(14): p. 2231-2247.
219. Dunker, A.K., Brown, Celeste J. et al, *The Protein Trinity: Structure/Function Relationships That Include Intrinsic Disorder.* Proceedings of the Nature Biotechnology Winter Symposium, 2002. **2**(S2): p. 49-50.
220. Uversky, V.N., *Unusual biophysics of intrinsically disordered proteins.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2013. **1834**(5): p. 932-951.
221. Bellay, J., S. Han, M. Michaut, T. Kim, M. Costanzo, B.J. Andrews, C. Boone, G.D. Bader, C.L. Myers, and P.M. Kim, *Bringing order to protein disorder through comparative genomics and genetic interactions.* Genome Biol, 2011. **12**(2): p. R14.
222. Daughdrill, G.W., P. Narayanaswami, S.H. Gilmore, A. Belczyk, and C.J. Brown, *Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation.* Journal of molecular evolution, 2007. **65**(3): p. 277-288.
223. Forman-Kay, Julie D. and T. Mittag, *From Sequence and Forces to Structure, Function, and Evolution of Intrinsically Disordered Proteins.* Structure, 2013. **21**(9): p. 1492-1499.
224. Moesa, H.A., S. Wakabayashi, K. Nakai, and A. Patil, *Chemical composition is maintained in poorly conserved intrinsically disordered regions and suggests a means for their classification.* Molecular BioSystems, 2012. **8**(12): p. 3262-3273.
225. Light, S., R. Sagit, O. Sachenkova, D. Ekman, and A. Elofsson, *Protein Expansion Is Primarily due to Indels in Intrinsically Disordered Regions.* Molecular Biology and Evolution, 2013. **30**(12): p. 2645-2653.
226. Light, S., R. Sagit, D. Ekman, and A. Elofsson, *Long indels are disordered: A study of disorder and indels in homologous eukaryotic proteins.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2013. **1834**(5): p. 890-897.
227. Brown, C.J., S. Takayama, A.M. Campen, P. Vise, T.W. Marshall, C.J. Oldfield, C.J. Williams, and A. Keith Dunker, *Evolutionary rate heterogeneity in proteins with long disordered regions.* Journal of molecular evolution, 2002. **55**(1): p. 104-110.
228. Brown, C.J., A.K. Johnson, and G.W. Daughdrill, *Comparing Models of Evolution for Ordered and Disordered Proteins.* Molecular Biology and Evolution, 2010. **27**(3): p. 609-621.

229. Schlessinger, A., C. Schaefer, E. Vicedo, M. Schmidberger, M. Punta, and B. Rost, *Protein disorder ;a breakthrough invention of evolution?* Current Opinion in Structural Biology, 2011. **21**(3): p. 412-418.
230. Oldfield, C.J., J. Meng, J.Y. Yang, M.Q. Yang, V.N. Uversky, and A.K. Dunker, *Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners.* BMC genomics, 2008. **9**(Suppl 1): p. S1.
231. Uversky, V.N., C.J. Oldfield, and A.K. Dunker, *Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept.* Annual Review of Biophysics, 2008. **37**(1): p. 215-246.
232. Borcherds, W., F.-X. Theillet, A. Katzer, A. Finzel, K.M. Mishall, A.T. Powell, H. Wu, W. Manieri, C. Dieterich, P. Selenko, A. Loewer, and G.W. Daughdrill, *Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells.* Nat Chem Biol, 2014. **10**(12): p. 1000-1002.
233. Alonso, A.d.C., T. Zaidi, M. Novak, I. Grundke-Iqbal, and K. Iqbal, *Hyperphosphorylation induces self-assembly of τ into tangles of paired helical filaments/straight filaments.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(12): p. 6923-6928.
234. Hagestedt, T., B. Lichtenberg, H. Wille, E.M. Mandelkow, and E. Mandelkow, *Tau protein becomes long and stiff upon phosphorylation: correlation between paracrystalline structure and degree of phosphorylation.* The Journal of Cell Biology, 1989. **109**(4): p. 1643-1651.
235. Uversky, V.N., S. Winter, O.V. Galzitskaya, L. Kittler, and G. Lober, *Hyperphosphorylation induces structural modification of tau-protein.* FEBS letters, 1998. **439**(1): p. 21-25.
236. Gong, C.-X., I. Grundke-Iqbal, and K. Iqbal, *Targeting Tau Protein in Alzheimer's Disease.* Drugs & Aging, 2010. **27**(5): p. 351-365.
237. Hyre, D.E. and R.E. Klevit, *A disorder-to-order transition coupled to DNA binding in the essential zinc-finger DNA-binding domain of yeast ADR11.* Journal of Molecular Biology, 1998. **279**(4): p. 929-943.
238. Ward, J.J., J.S. Sodhi, L.J. McGuffin, B.F. Buxton, and D.T. Jones, *Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life.* Journal of Molecular Biology, 2004. **337**(3): p. 635-645.
239. Haynes, C., C.J. Oldfield, F. Ji, N. Klitgord, M.E. Cusick, P. Radivojac, V.N. Uversky, M. Vidal, and L.M. Iakoucheva, *Intrinsic Disorder Is a Common Feature of Hub Proteins from Four Eukaryotic Interactomes.* PLoS Comput Biol, 2006. **2**(8): p. e100.
240. Singh, G.P., M. Ganapathi, and D. Dash, *Role of intrinsic disorder in transient interactions of hub proteins.* Proteins: Structure, Function, and Bioinformatics, 2007. **66**(4): p. 761-765.
241. Dosztányi, Z., V. Csizmok, P. Tompa, and I. Simon, *IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content.* Bioinformatics, 2005. **21**(16): p. 3433-3434.
242. Muñoz, V. and L. Serrano, *Development of the multiple sequence approximation within the AGADIR model of α-helix formation: Comparison with Zimm-Bragg and Lifson-Roig formalisms.* Biopolymers, 1997. **41**(5): p. 495-509.

CHAPTER 2:

MATERIALS AND METHODS

Note to the reader: Experiment-specific methods are included in the chapter they apply to. Please refer to those chapters for these protocols. Below are detailed protocols for techniques used for several different experiments or for protocols only briefly outlined in publications.

Site-Directed Mutagenesis

Protocol was adapted from the Agilent Technologies protocol manual for Quick Change Mutagenesis. Primer numbers used in all site-directed mutagenesis experiments in the table 2.1 below and were designed as 35-mers with the mutation of interest in the center of the primer; the codon choice was determined by the least amount of nucleotide changes that could be made from original sequence to mutant, with consideration for codon commonality. All mutagenesis was conducted with a twelve and a half minute extension time at 68°C

Table 2.1: Site-Directed Mutagenesis Primer List

Primer Name	Gene	Mutation	Sequence
1877-sgs1-F30P-F	SGS1	F30P	AGACAAAGATTTCGTACCCCA GGCTATCCAAAAGC
1878-sgs1-F30P-R	SGS1	F30P	GCTTTTGGATAGCCTGGGGTA CGAAATCTTTGTCT
1879-sgs1-N80STOP-F	SGS1	Stop at residue 80	TGCTACGAAACAACATTAAGT CATGCAAACTTTGT
1880-sgs1-N80STOP-R	SGS1	Stop at residue 80	ACAAAGTTTGCATGACTTAATG TTGTTTCGTAGCA
1881-sgs1-L93P-F	SGS1	L93P	GAACGATACAGAATGGCTCTC GTACACTGCCACAT
1882-sgs1-L93P-R	SGS1	L93P	ATGTGGCAGTGTACGAGGGC CATTCTGTATCGTTC
1923-W92PF	SGS1	W92P	GTCGAACGATACAGAACCGCC CTCGTACACTGCCA
1924-W92PR	SGS1	W92P	TGGCAGTGTACGAGGGCGGT TCTGTATCGTTCGAC
1987-K26P-F	SGS1	K26P	GACTTTACAGGAAGACCCAGA TTTCGTATTCCAGG
1988-K26P-R	SGS1	K26P	CCTGGAATACGAAATCTGGGT CTTCCTGTAAAGTC
1989-Q34P-F	SGS1	Q34P	CGTATTCCAGGCTATCCCAAA GCACATCGCGAACA
1990-Q34P-R	SGS1	Q34P	TGTTCGCGATGTGCTTTGGGA TAGCCTGGAATACG
2014-Sgs1V29PF	SGS1	V29P	GGAAGACAAAGATTTCCCATT CCAGGCTATCCAAA
2015-Sgs1V29PR	SGS1	V29P	TTTGGATAGCCTGGAATGGGA AATCTTTGTCTTCC
2016-Sgs1I33PF	SGS1	I33P	TTTCGTATTCCAGGCTCCCCA AAAGCACATCGCGA
2017-Sgs1I33PR	SGS1	I33P	TCGCGATGTGCTTTTGGGGAG CCTGGAATACGAAA
2061- Sgs1K17P F	SGS1	K17P	GGAGCACAAATGGTTAccGGA AACGGCGACTTTAC
2062- Sgs1K17P R	SGS1	K17P	GTAAAGTCGCCGTTTCCggTAA CCATTTGTGCTCC
2063- Sgs1 T21P F	SGS1	T21P	GTTAAAGGAAACGGCGcCTTT ACAGGAAGACAAAG
2064-Sgs1 T21P R	SGS1	T21P	CTTTGTCTTCCTGTAAAGgCGC CGTTTCCTTTAAC
2065-Sgs1 D25P F	SGS1	D25P	GGCGACTTTACAGGAAccCAA AGATTTCGTATTCC

Table 2.1: Site-Directed Mutagenesis Primer List (continued)

Primer Name	Gene	Mutation	Sequence
2066-Sgs1D25P R	SGS1	D25P	GGAATACGAAATCTTTGggTTC CTGTAAAGTCGCC
2067-Sgs1 I37P F	SGS1	I37P	GGCTATCCAAAAGCACccCGC GAACAAAAGGCCTA
2068-Sgs1 I37P R	SGS1	I37P	TAGGCCTTTTGTTCGCGggGT GCTTTTGGATAGCC
2069-Sgs1T61P F	SGS1	T61P	ATGTGGACCAGGAACAcCAAA CTTTATAACCAGCA
2070-Sgs1T61P R	SGS1	T61P	TGCTGGTTATAAAGTTTGgTGT TCCTGGTCCACAT
2098-Sgs1D25A F	SGS1	D25A	GGCGACTTTACAGGAAGcCAA AGATTTCGTATTCC
2099-Sgs1D25A R	SGS1	D25A	GGAATACGAAATCTTTGgCTTC CTGTAAAGTCGCC
2100-Sgs1D25K F	SGS1	D25K	GGCGACTTTACAGGAAaAgAA AGATTTCGTATTCC
2101-Sgs1D25K R	SGS1	D25K	GGAATACGAAATCTTTcTtTTCC TGTAAAGTCGCC
2102-Sgs1I33A F	SGS1	I33A	TTTCGTATTCCAGGCTgcCCAA AAGCACATCGCGA
2103-Sgs1I33A R	SGS1	I33A	TCGCGATGTGCTTTTGGgcAG CCTGGAATACGAAA
2104-Sgs1I33K F	SGS1	I33K	TTTCGTATTCCAGGCTAAACAA AAGCACATCGCGA
2105-Sgs1I33K R	SGS1	I33K	TCGCGATGTGCTTTTGTTTAG CCTGGAATACGAAA
2273-Sgs1 L9P F	SGS1	L9P	GAAGCCGTCACATAACccAAG AAGGGAGCACAAAT
2274-Sgs1 L9P R	SGS1	L9P	ATTTGTGCTCCCTTCTTggGTT ATGTGACGGCTTC
2275-Sgs1H13P F	SGS1	H13P	TAACTTAAGAAGGGAGCcCAA ATGGTTAAAGGAAA
2276-Sgs1H13P R	SGS1	H13P	TTTCCTTTAACCATTTGgGCTC CCTTCTTAAGTTA
2307-W15P F Sgs1	SGS1	W15P	AAGAAGGGAGCACAAAccgTTA AAGGAAACGGCGA
2308-W15P R Sgs1	SGS1	W15P	TCGCCGTTTCCTTTAAcggTTT GTGCTCCCTTCTT
2607-Rmi1 F63P F	RMI1	F63P	AGACAGAGAACTGTTGccCCA GGTGTTGATGGTAG
2608-Rmi1 F63P R	RMI1	F63P	CTACCATCAACACCTGGggCA ACAGTTCTCTGTCT

Table 2.1: Site-Directed Mutagenesis Primer List (continued)

Primer Name	Gene	Mutation	Sequence
2610-Rmi1 P88A R	RMI1	P88A	CCACTTTCTGTTTTTTCGcATC CAGCTTGGTTTTT
2611-Rmi1 E220P F	RMI1	E220P	ATTTTGTGATTATTTGccATCTA AATTACAACGTG
2612-Rmi1 E220P R	RMI1	E220P	CACGTTGTAATTTAGATggCAA ATAATCACAAAAT
2694-Rmi1 A128P F	RMI1	A128P	GGCGGATAACAACTGCcCCAA GGAAAATAATAGCA
2695-Rmi1 A128P R	RMI1	A128P	TGCTATTATTTTCCTTGGgGCA GTTGTTATCCGCC
2712-rmi1A139P F	RMI1	A139P	CAACAATAATAGCAGTcCCGC CAAGAATAAAGCAG
2713-rmi1A139P R	RMI1	A139P	CTGCTTTATTCTTGGCGGgACT GCTATTATTGTTG
2716-rmi1Y218P F	RMI1	Y218P	TCAGAAATTTTGTGATccTTTG GAATCTAAATTAC
2717-rmi1Y218P R	RMI1	Y218P	GTAATTTAGATTCCAAAggATC ACAAAATTTCTGA
3109-Rmi1F63K F	RMI1	F63K	AGACAGAGAACTGTTGaaaCA GGTGTTGATGGTAG
3110-Rmi1F63K R	RMI1	F63K	CTACCATCAACACCTGtttCAAC AGTTCTCTGTCT
3111-Rmi1Y218K F	RMI1	Y218K	TCAGAAATTTTGTGATaAaTTG GAATCTAAATTAC
3112-Rmi1Y218K R	RMI1	Y218K	GTAATTTAGATTCCAAtTtATCA CAAAATTTCTGA
3113-Rmi1L7P F	RMI1	L7P	ATGTCTTTTTCATCTATCccAaa ACAGGATATCACAGATG
3114-Rmi1L7P R	RMI1	L7P	CATCTGTGATATCCTGTttTggG ATAGATGAAAAAGACAT
3115-Rmi1 Y35P F	RMI1	Y35P	GATTGTTTTCAGAGCTccCCAA AATGAACCTTGGT
3116-Rmi1 Y35P R	RMI1	Y35P	ACCAAGGTTCATTTTGGggAGC TCTGAAAACAATC

All mutagenic PCRs were transformed into *E. coli* 5α cells via electroporation and plated on selective media (all mutations except Sgs1 N^{1-80} STOP were selected on LB-Ampicillin plates; $Sgs1N^{1-80}$ was selected on LB-Kanamycin). Colonies were selected from the overnight growth and plasmid was prepped according to the protocol below.

Mutants were sequenced via MWG operon and results were evaluated via Sequencher software for integration of the mutation of interest.

Plasmid Isolation from Bacteria

All plasmid minipreps were prepared using the QIAprep miniprep kits from QIAGEN. Briefly, 5 ml of overnight bacterial culture was spun down at 3400 rpm for 10 min; the pellet was resuspended in QIAGEN buffer P1, followed by lysis with buffer P2, and protein precipitation with buffer N3. The solution was then spun at 14 000 rpm for 10 min, and the supernatant was harvested. The plasmid was purified on the kit-provided column and eluted in 10 mM Tris, pH 8.

Gross Chromosomal Rearrangement (GCR) Assay

All GCRs were done in strain KHSY1338 (*ura3-52*, *leu2Δ1*, *trp1Δ63*, *his3Δ200*, *lys2ΔBgl*, *hom3-10*, *ade2Δ1*, *ade8*, *YEL069C::URA3*, *sgs1::HIS3*). The assay surveys for the loss of both URA3 and CAN1 genes and is indicative of a high level of genomic instability. Five to ten colonies of KHSY1338 transformed with a mutant plasmid of choice were grown in synthetic media lacking leucine (SC-LEU) . From these cultures, a dilution was plated on SC-LEU and grown to obtain a viable cell count. The remaining culture was pelleted, resuspended in water, and plated on GCR plates containing canavanine and 5-FOA as previously described. [1].

Trichloroacetic Acid (TCA) Extraction

Protocol was adapted from the protocol first described in Wright et al, 1989 [2]. The yeast strain of interest was grown to saturation in liquid media overnight; from this culture, a culture is set up at OD_{600}= 0.2 and grown to desired OD (varied from 0.5 to 1.0, depending on protein being probed). Five ODs of the culture was then spun and the supernatant removed. The pellet was washed with water before being spun down again. The pellet was then resuspended in 20% TCA and acid-washed glass beads and beaten in a bead beater. The lysate was drawn off, and the beads were washed with TCA; this wash was pooled with the lysate. The sample was then spun down and the pellet was resuspended in laemmli buffer and pH-adjusted with 2M Tris pH 8. The sample was boiled, and then used in SDS-PAGE.

SDS-PAGE and Western Blot Analysis

Protocol was adapted from protocols in Schagger et al, 1987 and Towbin et al [3, 4]. Samples were loaded into an SDS-PAGE gel with 10%-18% polyacrylamide and run in tris-glycine buffer (250 mM Tris, 1.9 M glycine, 1% SDS) at 170 V until good resolution of a protein band can be expected as determined by the protein marker. The proteins were then transferred to a PVDF membrane in 1x transfer buffer (250 mM Tris, 1.9 M glycine, 20% methanol) for 1 hour and 15 minutes at 300 mV in a Hoefer semi-dry transfer machine. The membrane was then blocked overnight in 5% milk plus 1x TBST (150 mM NaCl, 10 mM Tris, 0.1% Triton-X100, pH 7.6). Primary antibody was added to 5 ml of the same milk solution and the blot was incubated for 1 hour at room

temperature. Following incubation, the blot was washed three times with TBST for 30 min. Secondary antibody (conjugated to horseradish peroxidase) was added to 5 ml milk solution and incubated for 1 hr at 4°C. The blot was then washed three times with TBST at 4°C for 30 min. The blot was then probed with 1:1 Amersham ECL prime kit reagents (GE Healthcare) for 2 min and visualized with film.

Hydroxyurea Sensitivity Assay

Cultures were grown in SC-LEU overnight to saturation. A culture was prepared from each of these saturations at an OD_{600} = 0.2 and grown to OD_{600}= 0.5. A volume equal to 0.5/OD times 50 µl of each culture was spun down at 14,000 rpm for 10 min and resuspended in 50 µl water. A series of 10-fold dilutions was prepared and 2 µl of each dilution was plated on YPD and YPD supplemented with 100 mM, 150 mM, or 200 mM HU and grown at 30°C. Plates were monitored for 4-5 days.

Yeast Mating for Diploids

Mating protocols were adapted from [5]. The two haploid strains of interest KHSY 2494 (MATa, *ura3Δ0*, *leu2Δ0*, *his3Δ1*, *lys2Δ0*, *RAD51.V5.6xHIS.KANMX6*, Open Biosystems) and KHSY 2497 (MATα, *ura3Δ0*, *leu2Δ0*, *his3Δ1*, *lys2Δ0*, *TOP3.V5.VSV.KANMX6*), plus mating strains KHSY1435 (MATa, Δthr4) and 1436 (MATα, Δthr4) were streaked on YPD and grown for 2 days at 30°C. A single colony from both strains of interest were mixed in 50 µl of water, spotted on YPD, and grown

overnight at 30°C. From this spot, a small scoop was taken and spread for singles on YPD and grown for 2 days at 30°C. A colony of tester strain 1435 was then suspended in 50 µl of water and spread on a warm YPD plate; this was repeated for a separate plate for tester strain 1436. Sixteen colonies from the mating plate were then suspended in 30 µl of water, and 2 µl was spotted on a YPD plate, and both tester lawn plates. Plates were incubated overnight at 30°C. The tester plates were replica plated onto minimal media (SC-Min) and incubated overnight at 30°C. Diploids were identified as those that did not grow on either tester plate as neither the tester nor strains of interest can survive on minimal media without mating with one another. Diploids were frozen and stored.

Lithium Acetate (LiAc) Transformation

The protocol used was previously described in Gietz and Woods, 2006 [6]. The strain used for transformation was grown overnight in YPD to saturation. This culture was then used to inoculate a 25 ml YPD culture to an OD_{600}= 0.2 and grown to OD_{600}=0.8± 0.04. The culture was then pelleted at 2000 rpm for 2 min, washed with 25 ml water, and spun down again. The pellet was then resuspended in 1 ml 100 mM LiAc and spun down at 14,000 rpm for 30 seconds. This pellet was then resuspended in 240 µl 100 mM LiAc and distributed into 4 equal aliquots before being spun down at 14,000 rpm for 1 min. These pellets were then treated with 50% PEG, 36 µl of 1 M LiAc, 75 µl of plasmid or PCR product plus water, and 10 µl boiled and snap-cooled salmon sperm DNA and incubated for 30 min at 30 °C. Post-incubation, the cells were heat-shocked at

42°C for 15 min. The cells were then spun down at 7000 rpm for 1 min, and the pellet was resuspended in 100 µl water and plated on selective media for the target gene or mutation (SC-LEU for point mutants, SC-HIS for Rmi1.myc integrants) and grown for 2 days at 30 °C. Single colonies were restreaked, incubated for 2 days at 30°C, and single colonies from this streak were frozen and stored post-verification.

Media Types

Media recipes were derived from Sherman, 2002 [5]. YPD liquid media was made by autoclaving 10 g/l Yeast Extract (US Biological), 20 g/l Bacto-peptone (US Biological), and 20 g/l glucose (Fisher Scientific). YPD agar uses the same ratios, but with 20g/L Agar (US Biological). Selective media agar was made by autoclaving 20 g/l agar and combining it with 2% glucose and a filter-sterilized solution of 6.7 g/l Yeast Nitrogen base (US Biological) and 2 g/l of a dropout mix lacking the amino acid being selected for (-LEU, -HIS; US Biological). Liquid selective media was made the same, but without agar. SC-Minimal media contains only 20 g/l agar and 2% glucose. LB media consisted of 10 g/l Tryptone (Fisher Scientific), 5 g/l Yeast extract, and 5 g/l NaCl (Fisher Scientific).

References

1. Schmidt, K.H., V. Pennaneach, C.D. Putnam, and R.D. Kolodner, *Analysis of Gross-Chromosomal Rearrangements in Saccharomyces cerevisiae*, in *Methods in Enzymology*, L.C. Judith and M. Paul, Editors. 2006, Academic Press. p. 462-476.
2. Wright, A.P., M. Bruns, and B.S. Hartley, *Extraction and rapid inactivation of proteins from Saccharomyces cerevisiae by trichloroacetic acid precipitation.* Yeast, 1989. **5**(1): p. 51-53.
3. Schägger, H. and G. von Jagow, *Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa.* Analytical Biochemistry, 1987. **166**(2): p. 368-379.
4. Towbin, H., T. Staehelin, and J. Gordon, *Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications.* Proceedings of the National Academy of Sciences, 1979. **76**(9): p. 4350-4354.
5. Sherman, F., *Getting started with yeast.* Methods in enzymology, 2002. **350**: p. 3-41.
6. Gietz, R.D. and R. Woods, *Yeast Transformation by the LiAc/SS Carrier DNA/PEG Method*, in *Yeast Protocol*, W. Xiao, Editor. 2006, Humana Press. p. 107-120.

CHAPTER 3:

A TRANSIENT α-HELICAL MOLECULAR RECOGNITION ELEMENT IN THE DISORDERED N-TERMINUS OF THE SGS1 HELICASE IS CRITICAL FOR CHROMOSOME STABILITY AND BINDING OF TOP3/RMI1.

Note to the reader: This chapter has been previously published with permission from the publisher as Kennedy, JA, Daughdrill, GW, and Schmidt, KH (2013). "A transient α-helical molecular recognition element in the disordered N-terminus of the Sgs1 helicase is critical for chromosome stability and binding of Top3/Rmi1." Nucleic Acids Res., 41(22) 10215-27.

Research was designed by all authors. All experiments in publication were performed by Jessica Kennedy. Corresponding author: Kristina Schmidt, Department of Cell Biology, Microbiology and Molecular Biology, University of South Florida, 4202 E. Fowler Avenue, ISA2015, Tampa, FL 33620. Phone: (813) 974-1592. Fax: (813) 974-1614.; E-mail: kschmidt@usf.edu

Abstract

The RecQ-like DNA helicase family is essential for the maintenance of genome stability in all organisms. Sgs1, a member of this family in *Saccharomyces cerevisiae*, regulates early and late steps of double-strand break repair by homologous recombination. Using nuclear magnetic resonance spectroscopy, we show that the N-terminal 125 residues of Sgs1 are disordered and contain a transient α-helix that

extends from residue 25 to 38. Based on the residue-specific knowledge of transient secondary structure, we designed proline mutations to disrupt this α-helix and observed hypersensitivity to DNA damaging agents and increased frequency of genome rearrangements. *In vitro* binding assays show that the defects of the proline mutants are the result of impaired binding of Top3 and Rmi1 to Sgs1. Extending mutagenesis N-terminally revealed a second functionally critical region that spans residues 9–17. Depending on the position of the proline substitution in the helix functional impairment of Sgs1 function varied, gradually increasing from the C- to the N-terminus. The multiscale approach we used to interrogate structure/function relationships in the long disordered N-terminal segment of Sgs1 allowed us to precisely define a functionally critical region and should be generally applicable to other disordered proteins.

Introduction

The maintenance of genome stability is essential for organismal survival. A complex and diverse system of proteins has evolved to accomplish this function. Sgs1 of *Saccharomyces cerevisiae* is a 3–5′ DNA helicase that belongs to the evolutionarily conserved RecQ helicase family whose members function in the maintenance of genome stability. Named after the RecQ helicase of *Escherichia coli*, members of this helicase family have been identified in all organisms, including five homologs in humans (RecQ1, BLM, WRN, RecQL4, RecQL5) [1]. Mutations in BLM, WRN and RecQL4 are associated with Bloom syndrome, Werner syndrome and Rothmund–Thompson syndrome, respectively, which are characterized by elevated levels of aberrant

recombination events, chromosome instability and extraordinary predisposition to cancer development early in life [1].

Saccharomyces cerevisiae cells that lack Sgs1 exhibit several phenotypes that are similar to those of cells from persons with Bloom syndrome, most notably dysregulated homologous recombination, hypersensitivity to DNA-damaging agents, meiotic defects and cell cycle delay [2,3]. These defects are caused when the helicase activity of Sgs1 is inactivated by mutations in the ATPase domain or the RecQ C-terminal domain, which together make up the helicase core. Also located in the C-terminal half of Sgs1 is the Helicase and RNAase D C-terminal (HRDC) domain thought to be involved in DNA substrate binding and protein–protein interactions. These domains are conserved in most RecQ homologs; they are structurally ordered and crystal structures of this region have been reported for *E. coli* RecQ and human RecQ1 [4,5]. In contrast, the N-terminal half of Sgs1 is devoid of conserved catalytic domains and provides binding sites for proteins with roles in DNA metabolism, including the topoisomerases Top2 and Top3, replication protein Rpa70, Rad16 and Srs2 [2,6–8]. Interaction with the Top3 homologs has also been shown for human BLM, RecQ1 and RecQ5, and the RecQ homolog of *Schizosaccharomyces pombe*, Rqh1 [9–12]. Superhelical relaxation activity and Holliday-junction dissolution activity of these topoisomerase/helicase complexes is greatly enhanced by interaction with the RecQ-mediated genome instability 1 (Rmi1) protein [13–15].

One of the most important functions of the Sgs1 N-terminus is the interaction with the Top3/Rmi1 complex (BLM/Topo IIIα/Rmi1/Rmi2 in humans, Rqh1/Top3/Rmi1 in *S. pombe*) [13–16]. The Top3 binding site is within the first 100–158 residues of Sgs1 [17–

19]. The loss of this region produces more severe phenotypes that exhibit slower growth and higher sensitivity to DNA damage than those produced by loss of Sgs1 alone [3]. This may be due to toxic intermediates produced by Sgs1 that accumulate during homologous recombination and require Top3 decatenation for resolution. Despite the fact that Sgs1 and BLM bind Top3 and its human homolog Topo IIIα, respectively, there is little primary sequence similarity between the N-terminal regions where these interactions are predicted to occur. Both N-termini are predicted to be intrinsically disordered [20], which may help explain their level of sequence divergence [21,22]. Such intrinsically disordered proteins/regions (IDPs/IDRs) are widespread in eukaryotes and function arises from an ensemble of conformations that contain varying degrees of secondary structure and rarely form transient tertiary contacts [21, 23–28]. A high percentage of eukaryotic proteins are predicted to contain significant stretches (>30 residues) of disorder; in *S. cerevisiae*, 50–60% of the total proteome are IDPs/IDRs, and a survey of cancer-associated human proteins found that ~79% of the proteins in the database are IDPs/IDRs [29,30].

Using multidimensional heteronuclear nuclear magnetic resonance (NMR) spectroscopy, we have identified a short segment within the first 125 residues of the intrinsically disordered N-terminus of unbound Sgs1 that has transient α-helical structure whose integrity is essential for Sgs1 function *in vivo*. We have rationally designed single amino acid substitutions that disrupt transient α-helices. Some of these mutations eliminate Top3 binding to Sgs1, cause DNA damage hypersensitivity and induce spontaneous chromosomal rearrangements.

Materials and Methods

Expression and Purification of Peptides for NMR Spectroscopy

Methods were based on a previously described procedure for the expression of an IDP [31]. Plasmid pKHS443, expressing $Sgs1^{1-125}$, was constructed by inserting the first 375 bp of *SGS1* into pET28a (Novagen) using *Nde*I and *Bam*HI sites. Plasmid pKHS463, expressing $Sgs1^{1-80}$, was constructed by introducing a stop codon after 240 bp in pKHS443. pKHS443 or pKHS463 was transformed into *E. coli* BL21 (DE) cells and grown at 37°C in 2 l of M9 media (42 mM Na_2HPO_4, 22 mM KH_2PO_4, 8 mM NaCl, 2 mM $MgSO_4$, 11 mM d-glucose, 0.1 mM $CaCl_2$, 10 µM $FeCl_3$, 1 mg of Vitamin B1/L, pH 7.3) plus 200 mg of ampicillin, supplemented with N^{15} ammonium chloride and C^{13} glucose. Protein expression was induced at OD_{600} = 0.6 for 3 h with 1 mM Isopropyl-beta-D-thiogalactopyranoside (IPTG) at 37°C. Cells were harvested via centrifugation at 8000 rpm before being resuspended in buffer A1 (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed at 19 000 psi via French press. The lysate was cleared via centrifugation (18 000 rpm, 1 h, 4°C) and the supernatant was loaded onto a 30 ml Ni-NTA column on an AKTA FPLC. The column was washed with 5 column volumes of buffer A2 (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0), and the peptide was eluted in buffer B (50 mM NaH_2PO_4, 300 mM NaCl, 300 mM imidazole, pH 8.0). Fractions containing the eluted protein were pooled and dialyzed into 50 mM Tris (pH 8.0) and 100 mM NaCl. The fractions were treated with 1 ml CleanCleave thrombin beads (Sigma) at room temperature for 8 h to remove the N-terminal $(HIS)_6$ tag. Cleaved proteins were dialyzed into gel filtration buffer (50 mM NaH_2PO_4, 300 mM NaCl, 1 mM EDTA, 0.02% NaN_3, 4 mM DL-Dithiothreitol (DTT), pH 7), then

concentrated to a volume of 10 ml and loaded onto a 120-ml GE Hiload 16/60 Superdex 70 column via fast protein liquid chromatography (FPLC) and harvested over four 2.5-ml runs. Fractions containing the peptide were pooled and dialyzed into NMR buffer (50 mM NaH_2PO_4, 100 mM NaCl, 1 mM EDTA, 0.02% NaN_3, 4 mM DTT, pH 6.8) before being concentrated to 600 µl (150 µM for $Sgs1^{1-125}$; 690 µM for $Sgs1^{1-80}$, 160 µM for $Sgs1^{1-80}$-F30P).

NMR Analysis

NMR data for $Sgs1^{1-80}$ and $Sgs1^{1-80}$-F30P were collected at 25°C on a Varian VNMRS 800 MHz spectrometer equipped with a triple resonance pulse field *Z*-axis gradient cold probe. To make the amide ^{1}H and ^{15}N as well as ^{13}C, $^{13}C_{\beta}$ and ^{13}CO resonance assignments, sensitivity enhanced ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC) and three-dimensional HNCACB and HNCO experiments were performed on a uniformly ^{15}N- and ^{13}C-labeled sample of $Sgs1^{1-80}$ at 470 µM (or $Sgs1^{1-80}$-F30P at 160 µM) in 90% H_2O/10% D2O, phosphate buffered saline (PBS) buffer, at a pH of 6.8 (32–34). For the HNCACB experiment, data were acquired in ^{1}H, ^{13}C and ^{15}N dimensions using 9615.3846 (t_3) × 16 086.4648 (t_2) × 2000 (t_1) Hz sweep widths, and 512 (t_3) × 128 (t_2) × 32 (t_1) complex data points. For the HNCO, the sweep widths were 9615.3846 (t_3) × 2000 (t_2) × 2000 (t_1) Hz, complex data points were identical to the HNCACB. The sweep widths and complex data points of the HSQC were 9615.3846 (t_2) × 2100 (t_1) Hz and 1024 (t_2) × 128 (t_1), respectively. Processing and analysis of the HNCACB data resulted in 66 nonproline amide ^{1}H, ^{15}N, $^{13}C\alpha$ and ^{13}Cb resonance assignments plus 8 proline $^{13}C\alpha$ and $^{13}C\beta$ resonance assignments. ^{1}H-^{15}N

steady-state nuclear Overhauser effect (NOE) experiments were recorded at 25°C on a Varian VNMRS 600 MHz spectrometer equipped with a triple resonance pulse field *Z*-axis gradient cold probe in the presence and absence of a 120 off-resonance ^{1}H saturation pulse every 5 ms for 3 s. A total of 512 (t_2) × 128 (t_1) complex points were recorded with 128 scans per increment with the sweep widths set to 7225.4335 (t_2) × 1700 (t_1) Hz. The ^{1}H-^{15}N heteronuclear Overhauser effect (NHNOE) values were determined by taking the quotient of the intensity for resolved resonances in the presence and absence of proton saturation. Three measurements were made on each protein and the values were averaged. Resonance assignments for Sgs1$^{1-125}$ were carried out at 25°C on a Varian VNMRS 600 MHz spectrometer equipped with a triple resonance pulse field *Z*-axis gradient cold probe. To make the amide ^{1}H and ^{15}N as well as $^{13}C_{\alpha}$, $^{13}C_{\beta}$ and ^{13}CO resonance assignments, sensitivity-enhanced ^{1}H-^{15}N HSQC and three-dimensional HNCACB and HNCO experiments were performed on a uniformly ^{15}N- and ^{13}C-labeled sample at 150 μM in 90% H_2O/10% D_2O, PBS buffer, at pH 6.8. For the HNCACB experiment, data were acquired in ^{1}H, ^{13}C and ^{15}N dimensions using 7225.4335 (t_3) × 12 064.1295 (t_2) × 1499.9813 (t_1) Hz sweep widths, and 512 (t_3) × 108 (t_2) × 32 (t_1) complex data points. For the HNCO, the sweep widths were 7225.4335 (t_3) × 1500 (t_2) × 1499.9813 (t_1) Hz, and 512 (t_3) × 74 (t_2) × 32 (t_1) complex data points. For the HNCACO, the sweep widths were 7225.4335 (t_3) × 12 000 (t_2) × 1499.9813 (t_1) Hz, and 512 (t_3) × 70 (t_2) × 28 (t_1) complex data points. Processing and analysis of the data resulted in 87 nonproline amide ^{1}H, ^{15}N, $^{13}C_{\alpha}$ and ^{13}Cb resonance assignments plus 12 proline $^{13}C\alpha$ and $^{13}C\beta$ resonance assignments. All NMR spectra were processed with nmrPipe and analyzed using nmrView software (31,35,36). Apodization was achieved in

the ^{1}H, ^{13}C and ^{15}N dimensions using a squared sine bell function shifted by 70°. Apodization was followed by zero filling to twice the number of real data points and linear prediction was used in the ^{15}N dimension of the HNCACB.

Hydroxyurea Hypersensitivity Assay

Yeast strain KHSY1338 (*ura3-52*, *leu2Δ1*, *trp1Δ63*, *his3Δ200*, *lys2ΔBgl*, *hom3-10*, *ade2Δ1*, *ade8*, *YEL069C::URA3*, *sgs1::HIS3*) was transformed with derivatives of plasmid pRS415-SGS1 (Supplementary Table S1) by standard lithium-acetate transformation [37] and selected on synthetic complete media lacking leucine (SC-Leu). Transformants were grown in liquid SC-Leu to OD_{600} = 0.5, then plated in 10-fold dilutions on YPD (yeast extract/peptone/dextrose) and on YPD supplemented with 100 mM hydroxyurea (HU). Colony growth at 30°C was documented after 3–5 days.

Top3 and Rmi1 Binding Assay

Plasmid pKHS462, expressing GST-Sgs1$^{1-250}$, was constructed by inserting the first 750 bp of *SGS1* into pGEX-6p-2 (GE Healthcare) using *Bam*HI and *Xho*I sites. The Sgs1 fragment was expressed in *E. coli* BL21 (DE) cells in LB media (10 g/l tryptone, 5 g/l NaCl, 5 g/l Yeast extract) supplemented with 1.5 mg ampicillin for 3 h in the presence of 1 mM IPTG. The cell pellet was resuspended in 100 µl GST buffer (125 mM Tris, 150 mM NaCl, pH 8.0) plus HALT protease inhibitors (Pierce) and sonicated for 10 × 3 pulses. Lysate was cleared by centrifugation at 14 000 rpm for 10 min at 4°C. Glutathione magnetic beads (Pierce) were then incubated with 625 µg of cleared lysate for 1 h at 4°C, and washed three times with GST buffer. Native yeast whole-cell extract

containing endogenous levels of Top3 and/or Rmi1 was prepared from a culture of KHSY2497 (MATα, *ura3Δ0*, *leu2Δ0*, *his3Δ1*, *lys2Δ0*, *TOP3.V5.VSV.KANMX6*, Open Biosystems), KHSY4695 (MATα, *ura3Δ0*, *leu2Δ0*, *his3Δ1*, *lys2Δ0*, *rmi1::HIS3*, *TOP3.V5.VSV.KANMX6*) or KHSY4696 (MATα, *ura3Δ0*, *leu2Δ0*, *his3Δ1*, *lys2Δ0*, *TOP3.V5.VSV.KANMX6*, *RMI1.myc.HIS3MX6*) grown at 30°C in YPD overnight. To construct a *top3Δ* yeast strain that expresses myc-epitope tagged Rmi1, a diploid generated by mating RDKY3837 (*MATa*, *ura3-52*, *trp1Δ63*, *his3Δ200*, *leu2Δ1*, *lys2Bgl*, *hom3-10*, *ade2Δ1*, *ade8*, *top3::TRP1*) and KHSY4696 (*MATα*, *ura3Δ0*, *leu2Δ0*, *his3Δ1*, *lys2Δ0, TOP3.V5.VSV.KANMX6*, *RMI1.myc.HIS3MX6*) was sporulated (38) to isolate a *top3::TRP1*, *RMI1.myc.HIS3MX6* haploid (KHSY4741) by genotyping on selective media. The presence of the *top3::TRP1* and *RMI1.myc.HIS3MX6* alleles was also confirmed by polymerase chain reaction. Yeast cells were collected by centrifugation at 2000 rpm for 4 min, washed and resuspended in Top3/Rmi1 buffer (50 mM Tris, pH 7.5, 0.01% NP-40, 5 mM β-glycerol phosphate, 2 mM magnesium acetate, 120 mM NaCl) plus HALT protease inhibitors (Pierce). The suspension was lysed via French press at 19 000 psi or in a BeadBeater (Biospec Products, Inc.) by beating three times for 1 min. Lysates were cleared by centrifugation at 14 000 rpm for 15 min at 4°C. Cleared yeast lysate of 20 (KHSY2497, KHSY4695) or 10 mg (KHSY4696, KHSY4741) was incubated with Sgs1-bound magnetic beads for 90 min at room temperature on a nutator. Beads were washed four times with Top3/Rmi1 buffer plus HALT protease inhibitors (Pierce) and boiled for 10 min in Laemmli buffer (BioRad). Beads were collected by centrifugation and eluted protein complexes were separated by 10% sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). Presence of Sgs1

fragments, Top3 and Rmi1 and was determined by western blotting using monoclonal antibodies against GST (Covance), VSV (Sigma) and myc (Covance) epitopes, respectively.

Gross-Chromosomal Rearrangement Assay

Accumulation of cells that had undergone simultaneous inactivation of the *URA3* and *CAN1* genes on chromosome V was determined as previously described [39] except that cells were grown in the absence of leucine to select for the presence of the pRS415-derived plasmids expressing the desired *sgs1* mutants. Briefly, yeast strain KHSY1338 was transformed with derivatives of plasmid pRS415-SGS1 containing proline mutations (Supplementary Table S3.I) and grown to saturation at 30°C in 10 ml of SC-Leu. Cells were washed in water and plated on selective media containing canavanine (can) and 5-fluoro-orotic acid (5-FOA) to select for cells with inactive *CAN1* and *URA3* genes. Cells were also plated on SC-Leu media to obtain a viable cell count. After incubation at 30°C, viable cell count was determined after 3 days, and colonies on 5-FOA/can were counted after 5 days. Mutation rates and 95% confidence intervals were calculated from 6 to 16 cultures as previously described [39,40].

Preparation of Yeast Whole-Cell Extracts by Trichloroacetic Acid Extraction

To assess expression levels of Top3 and Rmi1 in *rmi1::HIS3* and *top3::TRP1* strains, respectively, yeast cultures were grown in YPD with vigorous shaking and 10 ODs were harvested by centrifugation at 2000 rpm for 2 min. To assess expression levels of sgs1-F30P and sgs1-H13P, the 3′-end of *SGS1* in pKHS481 was fused to the

myc-epitope amplified from pFA6a-13Myc-HIS3MX6 [41] by gap repair of *Sac*I-linearized pKHS481 to generate pKHS596. *F30P* and *H13P* mutations were introduced into pKHS596 by QuikChange mutagenesis (Agilent Technologies) to generate pKHS598 and pKHS600, respectively. Cell pellets were washed in water and resuspended in ice-cold 20% trichloroacetic acid (TCA) and vortexed in a cell disruptor (USA Scientific) with acid-washed glass beads for 4 min at maximum speed. Cell lysate was cleared at 14 000 rpm for 3 min. The pellet was resuspended in Laemmli buffer, adjusted to neutral pH and boiled for 2 min before separation by 10% SDS-PAGE. Presence of Top3.VSV, Rmi1.myc, Sgs1.myc and GAPDH was determined by western blotting using monoclonal antibodies against VSV (Sigma) and myc (Covance) epitopes, and against GAPDH (Pierce), respectively.

Results

The First 125 Residues of the Structurally Disordered N-terminus of Sgs1 Contain Two Transient α-helices

Sgs1 is a modular protein containing both ordered and disordered domains. The ATPase domain, zinc-binding domain, winged-helix domain and the HRDC domain make up the structurally ordered C-terminal half of Sgs1. In contrast, most of the N-terminal half of Sgs1 (residues 1–654) is predicted to be disordered [20,42]. This is also the case for other members of the RecQ helicase family, most notably *S. pombe* Rqh1 and human BLM.

A previous study has shown that the first 158 residues of Sgs1 are sufficient for binding to the topoisomerase Top3 [18]. It is well established that short segments within

longer disordered regions will undergo coupled folding and binding in the presence of protein binding partners [43–45]. Disorder predictors like IUPred [46] will frequently display short dips into the ordered region (disorder tendency < 0.5) that correspond to these protein binding sites, and it is expected that these regions will contain some degree of transient secondary structure. The lowest dips in the IUPred plot of the first 158 residues of Sgs1 correspond to residues E24 and Y102 (Figure 3.1). To determine whether these small segments within the disordered N-terminus of Sgs1 could adopt functionally significant secondary structures, we characterized the solution structure of the first 125 residues of Sgs1 using NMR spectroscopy. Single (^{15}N)- and double (^{15}N/^{13}C)-labeled samples of Sgs1$^{1-125}$ were overexpressed in *E. coli* and purified to apparent homogeneity. The double-labeled sample was used to measure the HSQC spectrum (Figure 3.2) as well as the triple resonance spectra that were used to make resonance assignments. The HSQC spectrum shows narrow chemical shift dispersion in the ^{1}H dimension (7.85–8.5 ppm), consistent with a disordered peptide [47–49]. The ^{15}N-labeled sample was used to measure the NHNOE. NHNOE values are sensitive to the rotational correlation time for the residue of interest. In disordered regions, small positive NHNOE values indicate regions that are less dynamic and typically correlate with the presence of transient secondary structure, and negative NHNOE values indicate highly dynamic regions. The NHNOE values observed for Sgs1$^{1-125}$ are consistent with a mostly disordered protein that contains two transiently ordered regions centered on residues F30 and E92 (Figure 3.3A). Alpha carbon secondary chemical shifts (CAΔδ) were calculated for every residue by subtracting the amino acid–specific random coil chemical shift values for CA from the measured values [50]. This is a

reliable method for identifying the presence of transient secondary structure in IDPs [51–53]. The presence of transient α-helical secondary structure in Sgs1$^{1-125}$ was indicated by consecutive positive CAΔδ values for residues 23–34 and 88–97 (Figure 3.3B).

Several clusters of overlapping resonances in the HSQC and HNCACB spectra, and repeating amino acid motifs (e.g. Thr-Ala-Thr) limited resonance assignments to 77% of the nonproline residues for the Sgs1$^{1-125}$ fragment. Several of the residues that could not be assigned were in or near the two transient α-helical segments preventing an identification of the helix boundaries. To develop a more complete picture of the first helical region, NMR analysis of a shorter Sgs1 fragment containing residues 1–80 (Sgs1$^{1-80}$) was performed. Using this fragment, we were able to assign 93% of the nonproline resonances in the HSQC spectrum (Figure 3.2) and to fill in the gaps in the secondary $^{13}C_{\alpha}$ chemical shift analysis (Figure 3.3C and D). The overlap between the HSQC spectra of the Sgs1$^{1-80}$ and the Sgs1$^{1-125}$ peptides indicates that elimination of 36% of the residues of the Sgs1$^{1-125}$ peptide (45 residues) did not affect the solution structure of the first 80 residues of Sgs1, consistent with this being a disordered region. Secondary $^{13}C_{\alpha}$ chemical shift analysis indicates the presence of α-helical secondary structure for residues 25–38 and residues 88–97 within this disordered region (Figure 3B and D). However, as mentioned above, helical states for both regions are transient because secondary shift values of > 2.6 δppm would be expected for $^{13}C_{\alpha}$ in a persistent α-helix [51].

Functional Mapping of α-helices by Proline Mutagenesis

To determine if the transient α-helical structures for residues 25–38 and 88–97 are important for Sgs1 function, residues with the highest NHNOE and CAΔδ values in each helical region were replaced with prolines—a known helix breaker. V29 and F30 in the first helical region and W92 and L93 in the second helical region were changed to proline in the context of full-length Sgs1. Cells expressing the mutant helicases were plated on media containing 100 mM of the DNA-damaging agent HU (Figure 3.4). While the *sgs1-V29P* and *sgs1-F30P* mutants were as sensitive to HU as the *sgs1Δ* mutant, neither the W92P nor the L93P mutation caused increased sensitivity (Figure 3.4A and B), indicating that the α-helical structure centered on V29 and F30 contributes to Sgs1's role in DNA damage repair, whereas that centered on W92 and L93 does not.

According to the NHNOE and CAΔδ values, the strongest helical region in the first 125 residues of Sgs1 extends from residues 25 to 38. To determine the functional distance that this helical region extends on both sides of V29 and F30, residues were mutated according to the expected *i*, *i* + 4 intramolecular hydrogen-bonding pattern of a typical α-helix. The mutants using V29 as a starting point, therefore, were D25P, I33P and I37P and those based on F30 were K26P and Q34P. Because proline substitutions of disordered residues near the ordered region would not be predicted to affect Sgs1 function, a T61P mutation (IUPred disorder score: 0.73) was included as a negative control. The *sgs1-K26P*, *sgs1-Q34P*, *sgs1-I37P* and *sgs1-T61P* mutants exhibited wild-type levels of HU sensitivity, whereas the *sgs1-D25P* and *sgs1-I33P* mutants were hypersensitive, with a gradual decrease in functional impairment of Sgs1 being observed between proline substitutions near the N-terminus of the helix and those near

the C-terminus (Figure 3.4A). These observations are consistent with the functional α-helix extending from residues 25 to 33.

Whereas the lack of an effect of proline in position 26 argues against K26 being an internal residue of α-helix, our findings are consistent with K26 being in the first helical turn, more specifically in the N1 position, where proline is tolerated [54,55], whereas D25—as the N-cap residue [56]—defines the N-terminal helix boundary. Indeed, the AGADIR algorithm [57] identified a prominent peak of helical propensity centering on residue I33, and D25 received the highest N-cap score (Figure 3.5A). Consistent with the results of the DNA-damage-sensitivity assay, AGADIR predicted reduced helical content for the D25P mutant, but not for the K26P mutant (Figure 3.5B). Removing the N-cap by replacing the aspartic acid residue at position 25 with basic (D25K) or neutral (D25A) residues, which have excellent helical propensity, but are poor N-cap residues [58], leads to N-terminal extension of the helix in AGADIR (Figure 3.5C). This increase in helical content in the *sgs1-D25K* and *sgs1-D25A* mutants did not impair Sgs1 function *in vivo* (Figure 3.4C).

Further extending the proline mutagenesis starting from V29 toward the N-terminus revealed wild-type levels of HU sensitivity for *sgs1-T21P*, consistent with D25 defining the N-terminal end of the α-helix. In contrast, the *sgs1-K17P*, *sgs1-W15P*, *sgs1-H13P* and *sgs1-L9P* mutants were more sensitive to HU than cells expressing wild-type Sgs1, indicating that this region is also critical for Sgs1 function (Figure 3.4B and D). The stretch of consecutive positive CAΔδ values for N8 to R11 is consistent with α-helical propensity and the HU hypersensitivity assay suggests that it extends C-terminally to residue H17. At first sight, the negative CAΔδ value for W15 seems to

indicate that W15 is not in a transient helical structure (Figure 3.3D). If this is the case then it suggests that any helical structure in the bound state is not contiguous from residue 8 to 17. However, the inconsistent CAΔδ value for W15 could be owing to the inaccuracies associated with the random coil chemical shift library used for calculating the secondary chemical shifts [50] or related to an anomalous effect on the CA shift that results from the partial charge of the H13 and H17 side chains. Consistent with W15 being an α-helix, substituting the tryptophan with other residues with good helical propensity, such as alanine or arginine, did not affect Sgs1 function in the DNA-damage hypersensitivity assay (W15A, Figure 3.4B) or its ability to induce slow growth in the *sgs1Δ top3Δ* strain [59]. However, α-helical content in this region could not be further assessed as assignments, and therefore NHNOE and CAΔδ values for residues S6, E12, H13 and K14 were not available owing to overlapping resonances in the HSQC spectra of both $Sgs1^{1-80}$ and $Sgs1^{1-125}$. That W15 and W92 could be changed to nonaromatic residues without increasing sensitivity of cells to DNA damaging agents (Figure 3.4B) also shows that these two residues are not involved in stacking interactions with each other, with other aromatic residues in the region, or with DNA [60,61], or at least that such stacking interactions are not important for the role of Sgs1 in suppressing HU hypersensitivity.

To verify that proline mutations that cause HU hypersensitivity indeed disrupt the α-helix between residues D25 and A38, we analyzed the solution structure of the $sgs1^{1-80}$-F30P mutant by NMR (Figure 3.6). We found that the resonances that shifted notably in the HSQC spectrum of the F30P mutant compared with the wild type were limited to residues F28–A38 (Figure 3.6A, Merged), suggesting that changes induced by

the F30P mutation are probably localized to the α-helix. Indeed, the consecutive positive secondary alpha carbon chemical shifts (CAΔδ) between residues D25 and A38 in wild-type Sgs1, which indicate the presence of α-helical secondary structure, were markedly reduced in the F30P mutant (Figure 3.6B), demonstrating that a proline at position 30 is sufficient to prevent the formation of the α-helix between residues 25 and 38. We also confirmed that proline mutations that disrupt α-helical content in the N8–H17 region or the D25–A38 region and cause the highest HU sensitivity (H13P, F30P) do not affect Sgs1 expression levels and stability (Figure 3.9).

Disruption of Transient α-helices Impairs Complex Formation Between Sgs1, Top3 and Rmi1

The disordered region of Sgs1 where the transient α-helices were identified binds to the Type-1A topoisomerase Top3 [18]. To test if HU hypersensitivity caused by proline mutations in this region is owing to the disruption of transient helices that are required for the interaction between Sgs1 and Top3, the ability of various sgs1 mutants to form a complex with Top3 was assessed *in vitro*. Because overexpression of full-length Sgs1 leads to insolubility [62,63], we chose the N-terminal 250 residues of Sgs1 and expressed them as an N-terminal GST fusion in *E. coli*. This $Sgs1^{1-250}$ fragment pulled down endogenous Top3 from native yeast whole-cell extract in an Rmi1-dependent manner (Figure 3.7A). Similarly, binding of Rmi1 to $Sgs1^{1-250}$ was reduced in the absence of Top3 (Figure 3.7B), suggesting that Top3 and Rmi1 depend on each other for binding to the N-terminal 250 residues. Despite the effect on Sgs1 binding, expression levels of Top3 and Rmi1 were not affected by the absence of Rmi1 and

Top3, respectively (Figure 3.7C and D). $Sgs1^{1-250}$ binds to Top3 more strongly than $Sgs1^{1-160}$ and, similar to what has been reported previously for an Sgs1 fragment comprising residues 107–283 [19], $Sgs1^{125-250}$ did not bind to Top3 (Figure 3.7E and F). When we introduced L9P, H13P, K17P, D25P, V29P and F30P mutations into the $Sgs1^{1-250}$ fragment, its ability to pull down Top3 from cell extracts was diminished, whereas the T21P and K26P mutants were still able to bind Top3 (Figure 3.7G). Mutations of Sgs1 that disrupted binding to Top3 also disrupted binding to Rmi1 (Figure 3.7H).

Integrity of Transient α-helices is Critical for Maintaining Chromosomal Stability

Lack of Sgs1 or disruption of its conserved C-terminal helicase core domain leads to mitotic hyperrecombination and a moderate increase in the accumulation of gross-chromosomal rearrangements (GCRs), including translocations between nonallelic sites [39,64,65]. To determine if the inability of Sgs1 to interact with Top3 and Rmi1 also leads to increased genome instability, we tested the ability of D25P, K26P, V29P, F30P and I33P mutants of full-length Sgs1 expressed from a CEN/ARS plasmid to suppress the elevated GCR rate of an *sgs1Δ* mutant. Mirroring the results of the HU hypersensitivity assay, D25P, V29P, F30P and I33P were unable to complement the defects of *sgs1Δ* cells, whereas cells expressing the K26P mutant accumulated GCRs at a similar rate as cells expressing wild-type Sgs1 (Table 3.1).

Discussion

In the prokaryote-to-eukaryote transition, some members of the RecQ helicase family acquired long N-terminal regions that precede the ATPase domain of the helicase core. In Sgs1, the only RecQ homolog in *S. cerevisiae*, this N-terminal region is ~650 amino acids long, making up ~45% of the 1447-residue long protein. This entire region is predicted to be intrinsically disordered and to contain several short segments of transient secondary structure (Figure 3.1). Using NMR spectroscopy, we have demonstrated that the first 125 residues of this N-terminal region of Sgs1 are intrinsically disordered in the unbound solution state with two short segments, between residues 25–38 and 88–97, that adopt transient α-helical structure. Transient α-helices in disordered regions of proteins are often stabilized by interactions with a binding partner [43–45,66,67]. This principle was used to rationally design single residue substitutions that disrupted the transient α-helical structures of residues 25–38 and 88–97, and the effects of these mutations on Sgs1 function were tested *in vitro* and *in vivo*. Substitution of residues D25, V29, F30 and I33 with the α-helix breaker proline impaired Sgs1 function *in vivo*, as evidenced by increased sensitivity to DNA damage and increased chromosome instability, and reduced binding of Top3 and Rmi1 to Sgs1 *in vitro*. Additional proline mutagenesis following the *i*, *i* − *4* α-helix pattern revealed that L9, H13 and K17 were critical for the same Sgs1 functions as the D25–A38 α-helix.

Our work demonstrates that the integrity of a transient α-helix is required for the *in vivo* function of Sgs1 and the binding of Sgs1 to Top3 and Rmi1. This helps explain why previous attempts to identify functionally critical single residues through alanine scanning of the region were unsuccessful (K4A, P5A, L9A) [68]. Alanine scanning is

often useful for identifying residues important for catalytic function, such as the ATPase activity of Sgs1 (K706A in the Walker A motif). However, the effectiveness of this approach to detect functionally important structural motifs in disordered segments, such as transient α-helices, is hampered by the high helical propensity of alanine and will depend on whether the substitution occurs at a residue that forms part of the binding interface [58]. Substitution with lysine and valine residues, which also have excellent helical propensity, also had no effect on Sgs1 function (D25K, Figure 3.4C; D25V [59], whereas a proline substitution at this same residue disrupted function (D25P, Figure 3.4A). Even amino acid residues that have lower helical propensity and are therefore not commonly found in α-helices, such as glycine and serine, are not necessarily successful at disrupting transient α-helices when introduced as single-residue substitutions. For example, the E12G and H13S mutations by themselves were insufficient to disrupt the interaction between Sgs1 and Top3, but were effective when combined [68]. Rationally designing mutations based on residue-specific knowledge of transient secondary structure provided a direct test of structure/activity relationships for Sgs1 (and presumably other IDPs) that could only be realized by combining a high-resolution structural approach, like NMR, with the *in vivo* and *in vitro* functional tests that can be performed in a model organism like *S. cerevisiae*. While this type of multiscale approach has commonly been used to interrogate structure/activity relationships for ordered proteins, the widespread application of this approach to IDPs/IDRs has been hampered by a lack of understanding of the general rules that connect their dynamic structures to their function. We believe our study helps clarify an approach that can be consistently applied to identify the functionally critical regions of IDPs/IDRs.

What functional advantages might the long, intrinsically disordered N-terminal tail provide to Sgs1? One possibility is that it contains multiple protein interaction sites, in addition to Top3/Rmi1. This hypothesis is supported by multiple dips below the 0.5 threshold in the IUPred plot (Figure 3.1) and the fact that Sgs1 binds Top2, Rad16, Rpa70, Dna2 and Mre11 at sites that map to the disordered N-terminus, although the discrete binding sites have not been identified [6–8,13]. Sgs1 may need to bind several of these proteins, sequentially or concurrently, in the same process. For example, the Sgs1/Top3/Rmi1 complex is instrumental in DNA resection during double-strand break (DSB) repair in a reaction analogous to that performed by the RecBCD complex in bacteria. In this model, which was recently proposed by Cejka *et al.* [13], the Sgs1/Top3/Rmi1 complex is first recruited to the DSB by physically interacting with the Mre11 subunit of the Mre11/Rad50/Xrs2 complex. Subsequently, the Sgs1/Top3/Rmi1 complex physically interacts with Dna2 to stimulate preferential degradation of the 5′-end and with replication protein A (RPA) to protect the 3′-end. Still other physical interactions at the N-terminal tail, including those with Rad16 and Top2, are likely to be important for roles of the Sgs1/Top3/Rmi1 complex in DNA repair and chromosome segregation. Conformational flexibility may also be crucial to accommodating the various structures and sizes of DNA substrates that the Sgs1/Top3/Rmi1 complex acts on, which range from simple double-stranded or splayed ends to hairpins, quadruplexes, Holliday junctions and telomeres.

In *E. coli*, RecQ and Top3 interact functionally, but not physically. One advantage of gaining physical contact between Sgs1 and Top3 would be the ability of one subunit in the complex to regulate another subunit's enzymatic activity. Tight coordination

between a Type-IA topoisomerase activity, such as exhibited by Top3, and DNA-dependent ATPase activity, such as exhibited by the helicase core of Sgs1, can be seen in the reverse gyrases of thermophile and hyperthermophile bacteria and archaea, where the two activities are either contained in a single polypeptide [69,70] or are encoded by two separate genes [71]. In these enzymes, the topoisomerase domain has been found to reduce the activity of the helicase-like ATPase domain [72] and, conversely, the ATPase domain has been shown to inhibit the supercoil relaxation activity of the topoisomerase subunit to induce positive supercoiling [71]. Inhibition of the helicase activity of the human Werner syndrome helicase WRN by its associated Type-1B topoisomerase Topo I hints at the possibility of coordination between the two activities also in RecQ-like helicases. Similarly, in Sgs1, deletion of the Top3 contact site (*sgs1Δ1-158*) causes a more severe phenotype than that caused by the absence of Sgs1 [3], which could be explained by Top3 binding having an inhibitory effect on the ATPase activity of Sgs1.

The interaction with a Type-1 topoisomerase has been preserved in at least four of the five human RecQ-like helicases: BLM, WRN, RecQL1 and the long isoform of RecQL5. Like Sgs1, BLM and WRN interact with Topo IIIα (Type IA) and Topo I (Type IB), respectively, at the far end of a long N-terminal tail [12]. Human RecQL1 was also found to interact with Topo IIIα, whereas the long isoform of RecQL5 (RecQL5β) co-immunoprecipitated with Topo IIIα and Topo IIIβ [10,11]. The predicted helical content of the N-terminus of BLM does not resemble that of the Top3/Rmi1 contact site between residues 25 and 38 in Sgs1, which appears to be the result of a proline substitution in BLM at position 30 (Figure 3.8A). Instead, the helAcal content in the segment starting

with residue L9, which is weak in Sgs1, is predicted to be dominant in BLM. Thus, although both BLM and Sgs1 interact with topoisomerase 3 at the N-terminus, the structural elements in the two proteins that mediate this interaction may not be conserved. This is also supported by the finding that the C-terminal 156 residues of BLM also bind to Topo IIIα [12], whereas only the N-terminus of Sgs1 interacts with Top3. Strikingly, the predicted helical content for residues N23 to R36 in WRN is nearly a perfect match to that of the confirmed α-helix in Sgs1 (Figure 3.8B). However, WRN has not been shown to interact with Topo IIIα [10], possibly owing to the insertion of the exonuclease domain just downstream of this site, and it will be interesting to test if residues N23–R36 of WRN can provide a contact site for the Top3/Rmi1 complex when placed in Sgs1. In the case of *S. pombe* Rqh1, the first 322 N-terminal residues are required for interaction with Top3 [9]. Although helical content is not predicted for the first 100 residues of this region, noticeable helical content is evident for the 27-residue region between residues H264 and R291 and the 15-residue region between residues D112 and Q127, which could be investigated as putative Top3 binding sites (Figure 3.10). Although Topo IIIα also binds full-length RecQL1 and RecQL5 [10, 11], the binding regions in these two human RecQ homologs have not yet been narrowed down.

Applying the same NMR-based structure–function analysis to the remaining 525 residues of the disordered N-terminal tail of Sgs1 (and the tails of the other long RecQ-like helicases) will help to identify additional structural elements, either transient or persistent, that serve as molecular recognition elements for protein partners or DNA, and allow for the rational design of new separation of function alleles that encode

mutants of RecQ-like helicases with single residue substitutions that are defective in discrete cellular functions.

References

1. Karow,J.K., Wu,L. and Hickson,I.D. (2000) *RecQ family helicases: roles in cancer and aging*. Curr. Opin. Genet. Dev., 10, 32–38.
2. Gangloff,S., McDonald,J.P., Bendixen,C., Arthur,L. and Rothstein,R. (1994) *The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase*. Mol. Cell. Biol., 14, 8391–8398.
3. Mullen,J.R., Kaliraman,V. and Brill,S.J. (2000) *Bipartite structure of the SGS1 DNA helicase in Saccharomyces cerevisiae*. Genetics, 154, 1101–1114.
4. Bernstein,D.A., Zittel,M.C. and Keck,J.L. (2003) *High-resolution structure of the E. coli RecQ helicase catalytic core*. EMBO J., 22, 4910–4921.
5. Pike,A.C., Shrestha,B., Popuri,V., Burgess-Brown,N., Muzzolini,L., Costantini,S., Vindigni,A. and Gileadi,O. (2009) *Structure of the human RECQ1 helicase reveals a putative strand-separation pin*. Proc. Natl Acad .Sci. USA, 106, 1039–1044.
6. Chiolo,I., Carotenuto,W., Maffioletti,G., Petrini,J.H., Foiani,M. and Liberi,G. (2005) *Srs2 and Sgs1 DNA helicases associate with Mre11 in different subcomplexes following checkpoint activation and CDK1-mediated Srs2 phosphorylation*. Mol. Cell. Biol., 25, 5738–5751.
7. Saffi,J., Feldmann,H., Winnacker,E.L. and Henriques,J.A. (2001) *Interaction of the yeast Pso5/Rad16 and Sgs1 proteins: influences on DNA repair and aging*. Mutat. Res., 486, 195–206.
8. Watt,P.M., Louis,E.J., Borts,R.H. and Hickson,I.D. (1995) *Sgs1: a eukaryotic homolog of E. coli RecQ that interacts with topoisomerase II in vivo and is required for faithful chromosome segregation*. Cell, 81, 253–260.
9. Ahmad,F. and Stewart,E. (2005) *The N-terminal region of the Schizosaccharomyces pombe RecQ helicase, Rqh1p, physically interacts with Topoisomerase III and is required for Rqh1p function. Mol. Genet. Genomics*, 273, 102–114.
10. Johnson,F.B., Lombard,D.B., Neff,N.F., Mastrangelo,M.A., Dewolf,W., Ellis,N.A., Marciniak,R.A., Yin,Y., Jaenisch,R. and Guarente,L. (2000) *Association of the Bloom syndrome protein with topoisomerase III alpha in somatic and meiotic cells*. Cancer Res., 60, 1162–1167.
11. Shimamoto,A., Nishikawa,K., Kitao,S. and Furuichi,Y. (2000) *Human RecQ5beta, a large isomer of RecQ5 DNA helicase localizes in the nucleoplasm and interacts with topoisomerases 3alpha and 3beta*. Nucleic Acids Res., 28, 1647–1655.
12. Wu,L., Davies,S.L., North,P.S., Goulaouic,H., Riou,J.F., Turley,H., Gatter,K.C. and Hickson,I.D. (2000) *The Bloom's syndrome gene product interacts with topoisomerase III*. J. Biol. Chem., 275, 9636–9644.

13. Cejka,P., Cannavo,E., Polaczek,P., Masuda-Sasa,T., Pokharel,S., Campbell,J.L. and Kowalczykowski,S.C. (2010) *DNA end resection by Dna2-Sgs1-RPA and its stimulation by Top3-Rmi1 and Mre11-Rad50-Xrs2.* Nature, 467, 112–116.
14. Mullen,J.R., Nallaseth,F.S., Lan,Y.Q., Slagle,C.E. and Brill,S.J. (2005) *Yeast Rmi1/Nce4 controls genome stability as a subunit of the Sgs1-Top3 complex. Mol. Cell. Biol*., 25, 4476–4487.
15. Raynard,S., Bussen,W. and Sung,P. (2006) *A double Holliday junction dissolvasome comprising BLM, topoisomerase III alpha, and BLAP75*. J. Biol. Chem., 281, 13861–13864.
16. Chang,M., Bellaoui,M., Zhang,C., Desai,R., Morozov,P., Delgado-Cruzata,L., Rothstein,R., Freyer,G.A., Boone,C. and Brown,G.W. (2005) *RMI1/NCE4, a suppressor of genome instability, encodes a member of the RecQ helicase/Topo III complex*. EMBO J., 24, 2024–2033.
17. Duno,M., Thomsen,B., Westergaard,O., Krejci,L. and Bendixen,C. (2000) *Genetic analysis of the Saccharomyces cerevisiae Sgs1 helicase defines an essential function for the Sgs1-Top3 complex in the absence of SRS2 or TOP1*. Mol. Gen. Genet., 264, 89–97.
18. Fricke,W.M., Kaliraman,V. and Brill,S.J. (2001*) Mapping the DNA topoisomerase III binding domain of the Sgs1 DNA helicase*. J. Biol. Chem., 276, 8848–8855.
19. Bennett,R.J., Noirot-Gros,M.F. and Wang,J.C. (2000) *Interaction between yeast sgs1 helicase and DNA topoisomerase III*. J. Biol. Chem., 275, 26898–26905.
20. Mirzaei,H., Syed,S., Kennedy,J. and Schmidt,K.H. (2011) *Sgs1 truncations induce genome rearrangements but suppress detrimental effects of BLM overexpression in Saccharomyces cerevisiae*. J. Mol. Biol., 405, 877–891.
21. Brown,C.J., Johnson,A.K. and Daughdrill,G.W. (2010) *Comparing models of evolution for ordered and disordered proteins.* Mol. Biol. Evol., 27, 609–621.
22. Brown,C.J., Johnson,A.K., Dunker,A.K. and Daughdrill,G.W. (2011) *Evolution and disorder*. Curr. Opin. Struct. Biol., 21, 441–446.
23. Dyson,H.J. and Wright,P.E. (2005) *Intrinsically unstructured proteins and their functions*. Nat. Rev. Mol. Cell Biol., 6, 197–208.
24. Uversky,V.N., Gillespie,J.R. and Fink,A.L. (2000) *Why are "natively unfolded" proteins unstructured under physiologic conditions?* Proteins, 41, 415–427.
25. Tompa,P., Fuxreiter,M., Oldfield,C.J., Simon,I., Dunker,A.K. and Uversky,V.N. (2009) *Close encounters of the third kind: disordered domains and the interactions of proteins*. Bioessays, 31, 328–335.
26. Dunker,A.K., Brown,C.J., Lawson,J.D., Iakoucheva,L.M. and Obradovic,Z. (2002) *Intrinsic disorder and protein function*. Biochemistry, 41, 6573–6582.
27. Wright,P.E. and Dyson,H.J. (1999) *Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm*. J. Mol. Biol., 293, 321–331.

28. Dunker,A.K., Lawson,J.D., Brown,C.J., Williams,R.M., Romero,P., Oh,J.S., Oldfield,C.J., Campen,A.M., Ratliff,C.M., Hipps,K.W. et al. (2001) I*ntrinsically disordered protein*. J. Mol. Graph Model., 19, 26–59.
29. Iakoucheva,L.M., Brown,C.J., Lawson,J.D., Obradovic,Z. and Dunker,A.K. (2002) *Intrinsic disorder in cell-signaling and cancer associated proteins*. J. Mol. Biol., 323, 573–584.
30. Tompa,P., Dosztanyi,Z. and Simon,I. (2006) *Prevalent structural disorder in E. coli and S. cerevisiae proteomes.* J. Proteome Res., 5, 1996–2000. 12 Nucleic Acids Research, 2013
31. Vise,P.D., Baral,B., Latos,A.J. and Daughdrill,G.W. (2005) *NMR chemical shift and relaxation measurements provide evidence for the coupled folding and binding of the p53 transactivation domain.* Nucleic Acids Res., 33, 2061–2077.
32. Kay,L.E., Keifer,P. and Saarinen,T. (1992) *Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity.* J. Am. Chem. Soc., 114, 10663–10665.
33. Kay,L.E., Xu,G.Y. and Yamazaki,T. (1994) *Enhanced-sensitivity triple-resonance spectroscopy with minimal H2O Saturation*. J. Magn. Reson. Ser. A, 129–133.
34. Wittekind,M. and Mueller,L. (1993) *HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins.* J. Magn. Reson. Ser. B, 201–205.
35. Vise,P.D., Baral,B., Latos,A.J. and Daughdrill,G.W. (2005) *NMR chemical shift and relaxation measurements provide evidence for the coupled folding and binding of the p53 transactivation domain*. Nucleic Acids Res., 33, 2061–2077.
36. Johnson,B.A. and Blevins,R.A. (1994) *NMR view - a computerprogram for the visualization and analysis of Nmr data.* J. Biomol. NMR, 4, 603–614.
37. Gietz,R.D. and Woods,R.A. (2006) *Yeast transformation by the LiAc/SS carrier DNA/PEG method. Methods Mol. Biol*., 313, 107–120.
38. Rockmill,B., Lambie,E.J. and Roeder,G.S. (1991) *Spore enrichment. Methods* Enzymol., 194, 146–149.
39. Schmidt,K.H., Pennaneach,V., Putnam,C.D. and Kolodner,R.D. (2006) *Analysis of gross-chromosomal rearrangements in Saccharomyces cerevisiae*. Methods Enzymol., 409, 462–476.
40. Nair,K.R. (1940) *Table of confidence intervals for the median in samples from any continuous population*. Sankhya, 4, 551–558.
41. Longtine,M.S., McKenzie,A. III, Demarini,D.J., Shah,N.G., Wach,A., Brachat,A., Philippsen,P. and Pringle,J.R. (1998) *Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae.* Yeast, 14, 953–961.
42. Mirzaei,H. and Schmidt,K. (2012) *Non-bloom-syndrome associated partial and total loss-of-function variants of BLM helicase.* Proc. Natl Acad. Sci. USA, 109, 19357–19362.
43. Dyson,H.J. and Wright,P.E. (2002) *Coupling of folding and binding for unstructured proteins*. Curr. Opin. Struct. Biol., 12, 54–60.

44. Wright,P.E. and Dyson,H.J. (2009) *Linking folding and binding*. Curr. Opin. Struct. Biol., 19, 31–38.
45. Oldfield,C.J., Cheng,Y., Cortese,M.S., Romero,P., Uversky,V.N. and Dunker,A.K. (2005) *Coupled folding and binding with alpha helix- forming molecular recognition elements*. Biochemistry, 44, 12454–12470.
46. Dosztanyi,Z., Csizmok,V., Tompa,P. and Simon,I. (2005) *IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content*. Bioinformatics, 21, 3433–3434.
47. Daughdrill,G.W., Pielak,G.J., Uversky,V.N., Cortese,M.S. and Dunker,A.K. (2005) *Natively disordered proteins*. In: Buchner,J. and Kiefhaber,T. (eds), Protein Folding Handbook. WILEY-VCH, Darmstadt, pp. 275–357.
48. Dyson,H.J. and Wright,P.E. (2001) *Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states*. Methods Enzymol., 339, 258–270.
49. Eliezer,D. (2007) *Characterizing residual structure in disordered protein States using nuclear magnetic resonance*. Methods Mol. Biol., 350, 49–67.
50. Tamiola,K., Acar,B. and Mulder,F.A. (2010) *Sequence-specific random coil chemical shifts of intrinsically disordered proteins.* J. Am. Chem. Soc., 132, 18000–18003.
51. Wishart,D.S. and Sykes,B.D. (1994) *Chemical shifts as a tool for structure determination*. Methods Enzymol., 239, 363–392.
52. Dyson,H.J. and Wright,P.E. (2002) *Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance*. Adv. Protein Chem., 62, 311–340.
53. Wishart,D.S., Sykes,B.D. and Richards,F.M. (1992) *The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy*. Biochemistry, 31, 1647–1651.
54. Kim,M.K. and Kang,Y.K. (1999) *Positional preference of proline in alpha-helices.* Protein Sci., 8, 1492–1499.
55. Richardson,J.S. and Richardson,D.C. (1988) *Amino acid preferences for specific locations at the ends of alpha helices*. Science, 240, 1648–1652.
56. Aurora,R. and Rose,G.D. (1998) *Helix capping*. Protein Sci., 7, 21–38.
57. Munoz,V. and Serrano,L. (1994) *Elucidating the folding problem of helical peptides using empirical parameters.* Nat. Struct. Biol., 1, 399–409.
58. Serrano,L. (2000) *The relationship between sequence and structure in elementary folding units.* Adv. Protein. Chem., 53, 49–85.
59. Bennett,R.J. and Wang,J.C. (2001) *Association of yeast DNA topoisomerase III and Sgs1 DNA helicase: studies of fusion proteins*. Proc. Natl Acad. Sci. USA, 98, 11108–11113.
60. Khamis,M.I., Casas-Finet,J.R. and Maki,A.H. (1988) *Binding of recA protein to single- and double-stranded polynucleotides occurs without involvement of its aromatic residues in stacking interactions with nucleotide bases*. Biochim. Biophys. Acta, 950, 132–137.

61. Khamis,M.I., Casas-Finet,J.R. and Maki,A.H. (1987) *Stacking interactions of tryptophan residues and nucleotide bases in complexes formed between Escherichia coli single-stranded DNA binding protein and heavy atom-modified poly(uridylic) acid. A study by optically detected magnetic resonance spectroscopy*. J. Biol. Chem., 262, 1725–1733.
62. Bennett,R.J., Keck,J.L. and Wang,J.C. (1999) *Binding specificity determines polarity of DNA unwinding by the Sgs1 protein of S. cerevisiae*. J. Mol. Biol., 289, 235–248.
63. Bennett,R.J., Sharp,J.A. and Wang,J.C. (1998) *Purification and characterization of the Sgs1 DNA helicase activity of Saccharomyces cerevisiae*. J. Biol. Chem., 273, 9644–9650.
64. Myung,K., Datta,A., Chen,C. and Kolodner,R.D. (2001) *SGS1, the Saccharomyces cerevisiae homologue of BLM and WRN, suppresses genome instability and homeologous recombination*. Nat. Genet., 27, 113–116.
65. Schmidt,K.H., Wu,J. and Kolodner,R.D. (2006) *Control of translocations between highly diverged genes by Sgs1, the Saccharomyces cerevisiae homolog of the Bloom's syndrome protein*. Mol. Cell. Biol., 26, 5406–5420.
66. Sugase,K., Dyson,H.J. and Wright,P.E. (2007) *Mechanism of coupled folding and binding of an intrinsically disordered protein*. Nature, 447, 1021–1025.
67. Boehr,D.D., Nussinov,R. and Wright,P.E. (2009) *The role of dynamic conformational ensembles in biomolecular recognition*. Nat. Chem. Biol., 5, 789–796.
68. Onodera,R., Seki,M., Ui,A., Satoh,Y., Miyajima,A., Onoda,F. and Enomoto,T. (2002) *Functional and physical interaction between Sgs1 and Top3 and Sgs1-independent function of Top3 in DNA recombination repair*. Genes Genet. Syst., 77, 11–21.
69. del Toro Duany,Y., Klostermeier,D. and Rudolph,M.G. (2011) *The conformational flexibility of the helicase-like domain from Thermotoga maritima reverse gyrase is restricted by the topoisomerase domain.* Biochemistry, 50, 5816–5823.
70. del Toro Duany,Y. and Klostermeier,D. (2011) *Nucleotide-driven conformational changes in the reverse gyrase helicase-like domain couple the nucleotide cycle to DNA processing.* Phys. Chem. Chem. Phys., 13, 10009–10019.
71. Capp,C., Qian,Y., Sage,H., Huber,H. and Hsieh,T.S. (2010) *Separate and combined biochemical activities of the subunits of a naturally split reverse gyrase. J. Biol. Chem.*, 285, 39637–39645.
72. del Toro Duany,Y., Jungblut,S.P., Schmidt,A.S. and Klostermeier,D. (2008) *The reverse gyrase helicase-like domain is a nucleotide-dependent switch that is attenuated by the topoisomerase domain*. Nucleic Acids Res., 36, 5882

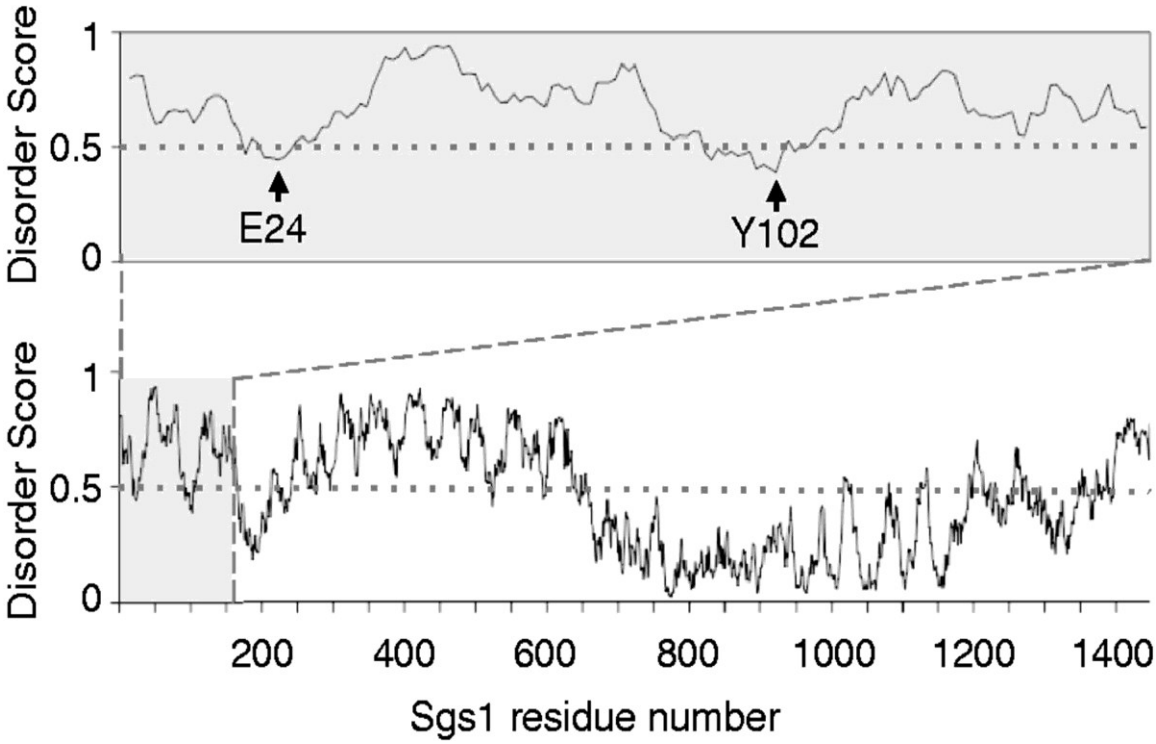


Figure 3.1: Prediction of intrinsically unstructured regions in Sgs1. Disorder scores are from IUPred (46) with scores >0.5 predicting disordered residues and scores <0.5 predicting ordered residues. Residues 1–158 (upper panel) are predicted to be mostly disordered with two short segments around residues E24 and Y102 dipping into the ordered region.

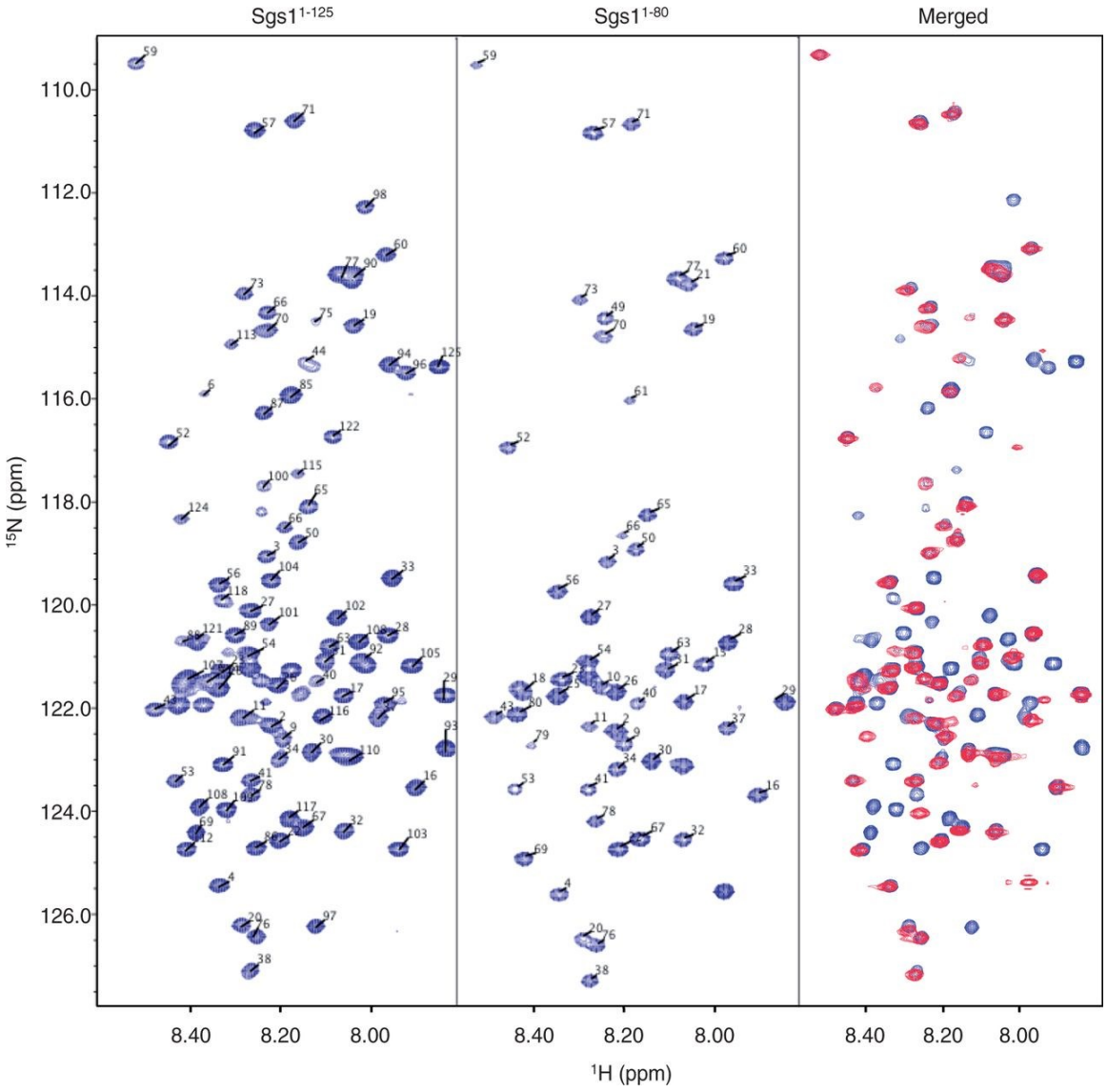


Figure 3.2: HSQC spectra of the first 125 residues of Sgs1 (Sgs1$^{1–125}$) and the first 80 residues of Sgs1 (Sgs1$^{1–80}$). Narrow chemical shift dispersion in the ^{1}H dimension in both the HSQC spectra of the long (Sgs1$^{1–125}$) and short (Sgs1$^{1–80}$) peptide are consistent with a disordered peptide. The overlay of the long and short peptide (Merged) shows little discrepancy in the peak assignments between the two proteins, implying conservation of structural elements, even with the truncation.

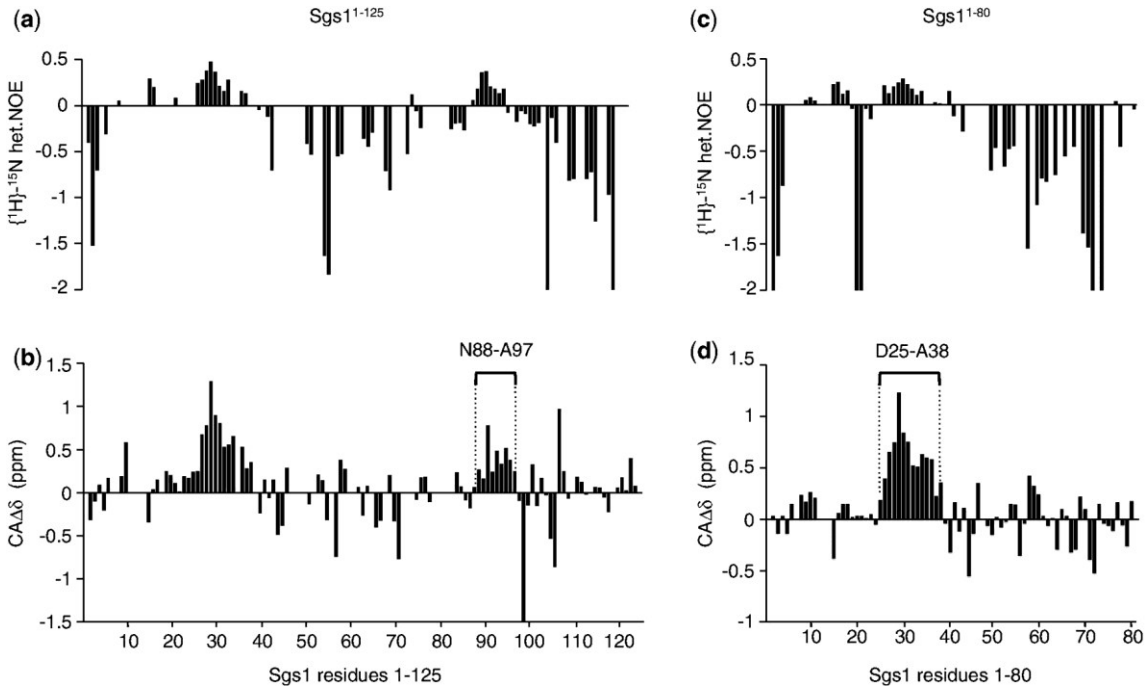


Figure 3.3: Measurement of NHNOE and secondary alpha carbon shifts (CAΔδ) of the Sgs1$^{1–125}$ peptide and the Sgs1$^{1–80}$ peptide. Consecutive positive values in the NHNOE plot for the Sgs1$^{1–125}$ peptide (**a**) and the Sgs1$^{1–80}$ peptide (**c**) indicate regions with a slower rotational correlation time that may adopt secondary structure. Consecutive positive secondary alpha carbon chemical shifts (CAΔδ) between residues 88 and 97 (**b**) and between residues 25 and 38 (**d**) indicate the presence of α-helical secondary structure in the unbound Sgs1 peptide as compared with standard chemical shifts in a random coil library (48,50).

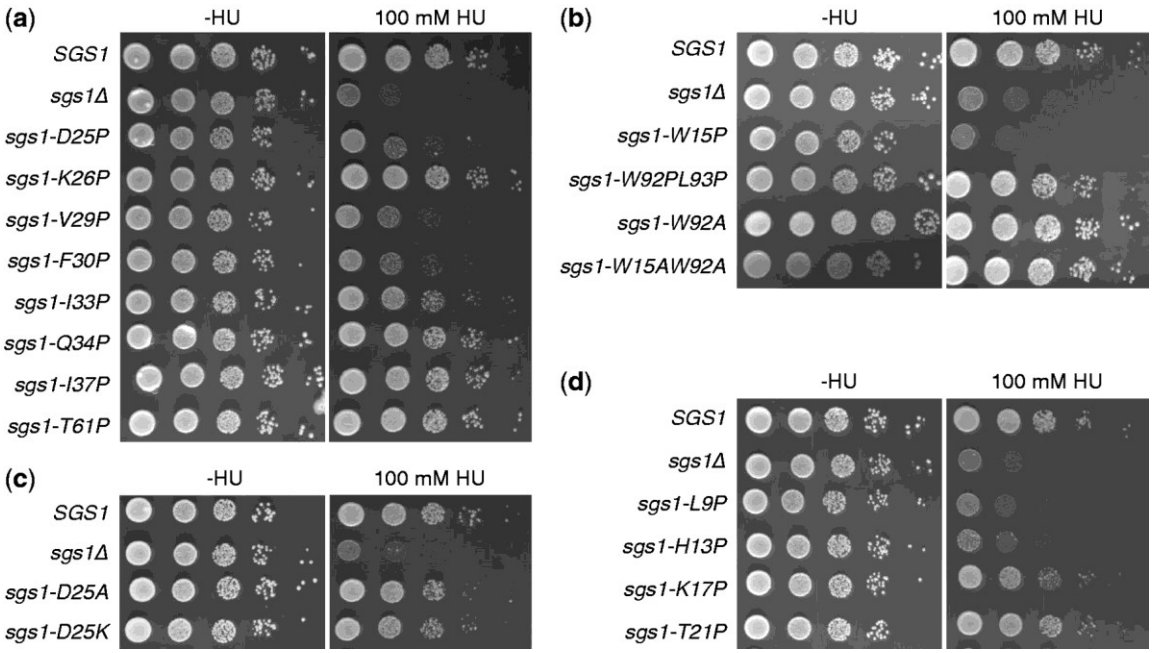


Figure 3.4: HU hypersensitivity of cells expressing *sgs1* alleles with mutations in (a) the α-helical region spanning residues 25–38 and (b) the α-helical region spanning residues 88–97. The wild-type phenotype exhibited by cells expressing W15A and W92A mutants of Sgs1 also demonstrates that these aromatic residues are not involved in stacking. T61P was included as a control for a disordered residue. (**c**) Replacing the N-cap residue D25 with basic (D25K) or neutral (D25A) residues that are poor N-caps, but have strong α-helical propensity, does not affect Sgs1 function. (**d**) Extending proline mutagenesis N-terminally of the first α-helical region reveals additional functional residues (L9, H13, H17).

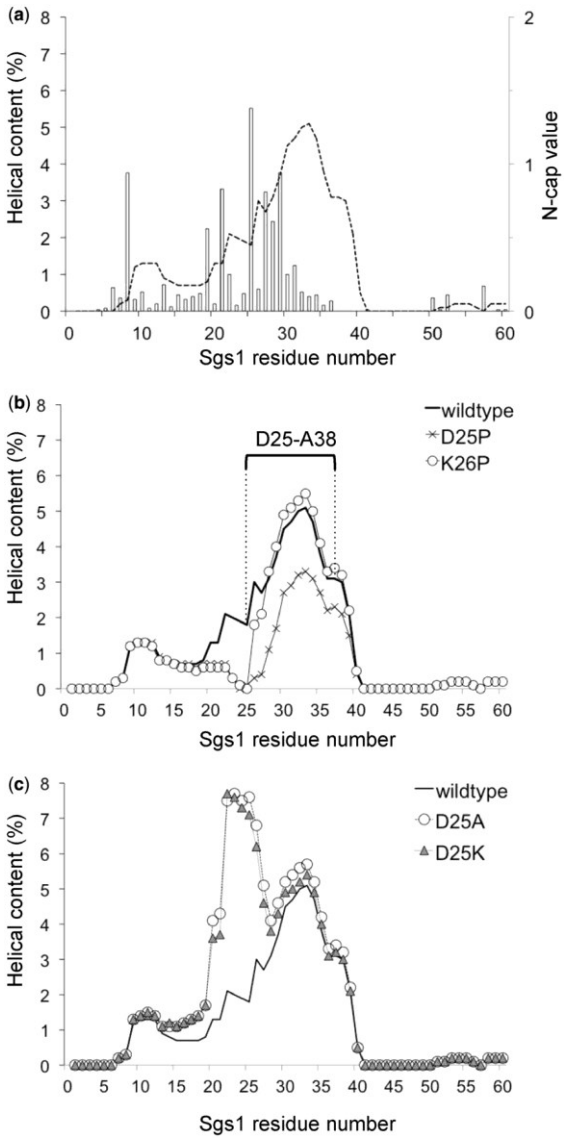


Figure 3.5: AGADIR (57) prediction of the helical content of the N-terminus of Sgs1. (**a**) In wild-type Sgs1 (dotted line), a prominent peak of helical propensity is predicted at residue I33 and a smaller peak at residues R10–E12. Residue D25 received the highest N-cap score (open columns). (**b**) The deleterious D25P mutation is predicted to reduce the helical content of the D25–A38 region, whereas the nondeleterious K26P mutation is not. (**c**) Replacing the N-cap residue D25 with residues that have excellent helical propensity, but are poor N-cap residues (lysine, alanine), is not predicted to reduce helical content of the D25–A38 region, but predicts an N-terminal extension of the helical region with a new peak of helical content at residue Q23 in the D25A mutant and at residue L22 in the D25K mutant.

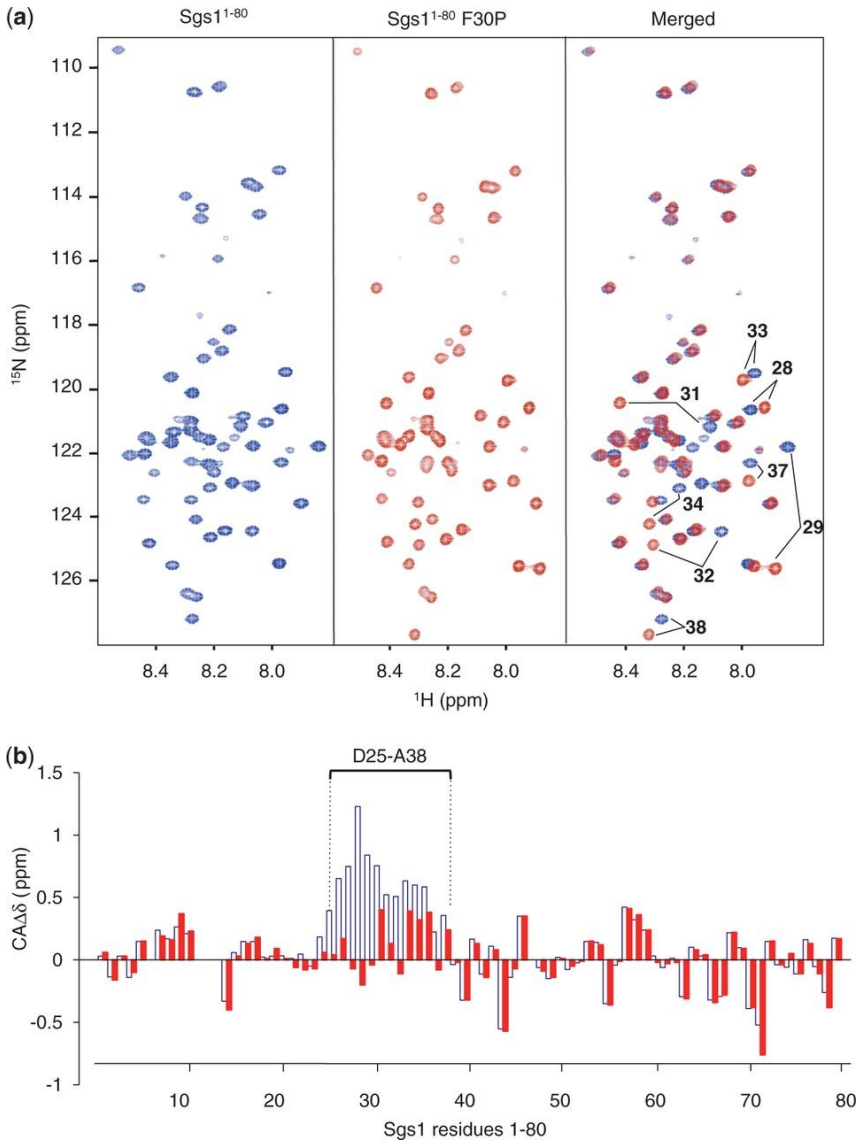


Figure 3.6: HSQC spectra and secondary chemical shift (CAΔδ) analysis of the first 80 residues of Sgs1 with a proline substitution at residue 30 (sgs1$^{1–80}$-F30P). (**a**) The overlay (Merged) of the HSQC spectra of wild-type Sgs1 (blue) and the sgs1-F30P mutant (red) reveals shifts in the peak assignments for residues F28, V29, Q31, A32, I33, Q34, I37 and A38, which form a transient α-helix in wild-type Sgs1. (**b**) Consecutive positive secondary alpha carbon chemical shifts (CAΔδ) between residues D25–A38 in wild-type Sgs1 (open blue columns), which indicate the presence of α-helical secondary structure in the unbound Sgs1 peptide, are markedly reduced in the Sgs1-F30P mutant (red filled columns).

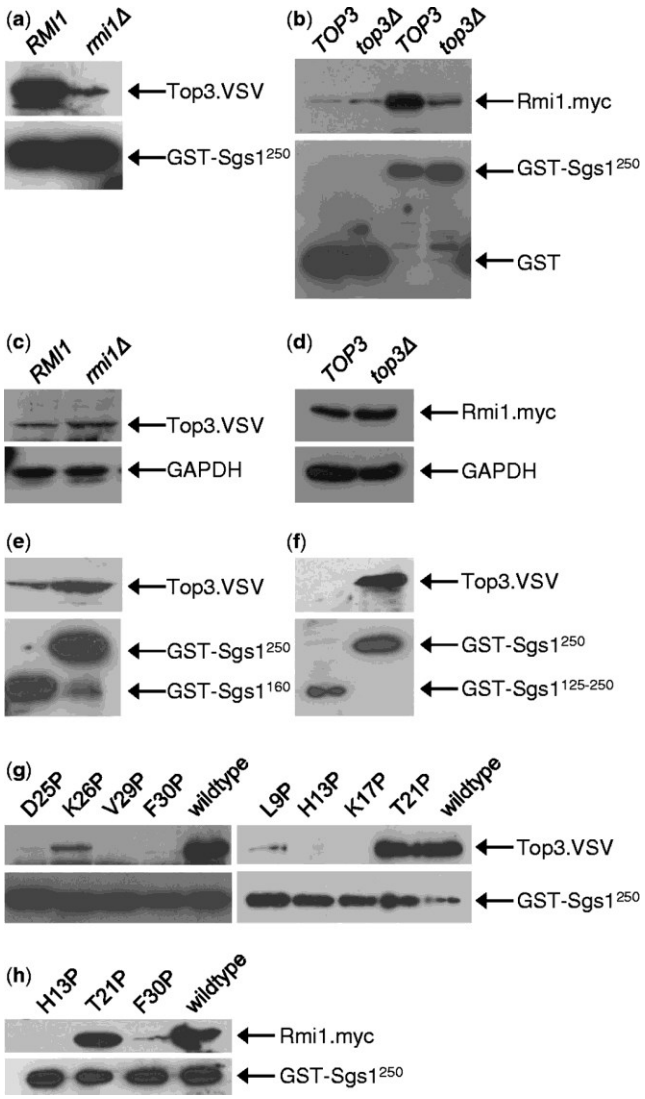


Figure 3.7: Loss of function of Sgs1 proline mutants is due to loss of Top3 and Rmi1 binding. Sgs1 proline mutants were expressed as N-terminal GST fusions in *E. coli* and purified by binding to glutathione beads. Top3 and Rmi1 were obtained from native whole-cell extracts of yeast strains KHSY2497 (*RMI1*), KHSY4695 (*rmi1Δ*) or KHSY4696 (*RMI1.MYC*), KHSY4741 (*top3Δ)*, which express epitope-tagged Top3 and/or Rmi1 from their chromosomal loci under their native promoters. (**a**) Binding of Top3 to the $Sgs1^{1–250}$ peptide is Rmi1-dependent. (**b**) Binding of Rmi1 to the $Sgs1^{1–250}$ peptide is Top3-dependent. (**c**) Deletion of *RMI1* does not lead to loss of Top3 expression. (**d**) Deletion of *TOP3* does not lead to loss of Rmi1 expression. (**e**) The $Sgs1^{1–250}$ peptide binds Top3 more strongly than the shorter $Sgs1^{1–160}$ peptide. (**f**) The $Sgs1^{125–250}$ peptide does not bind Top3, indicating that critical residues for Top3 binding are located in the first 125 residues of Sgs1. (**g**) Proline mutations at L9, H13, K17, D25, V29 and F30, but not at T21 and K26, reduce binding of $Sgs1^{1–250}$ to Top3. (**h**) Proline mutations at H13 and F30, which reduce binding of $Sgs1^{1–250}$ to Top3, also reduce binding to Rmi1.

Table 3.1. Effect of proline substitutions in the transient α-helix between residues D25 and A38 of Sgs1 on the rate of accumulating GCRs

Relevant genotype	Plasmid	GCR rate (Can^r 5–FOA^r × 10^{-8})	95% CI[a] (Can^r 5-FOA^r × 10^{-8})	Increase over wild type (*SGS1*)
SGS1	pKHS481	58	34–73	1
sgs1-D25P	pKHS494	334	260–789	6
sgs1-K26P	pKHS500	71	39–132	1
sgs1-V29P	pKHS492	320	189–352	6
sgs1-F30P	pKHS482	704	194–996	12
sgs1-I33P	pKHS496	211	165–255	4

- [a]95% confidence intervals were calculated according to Nair (40), with nonoverlapping confidence intervals indicating statistically significant differences ($\alpha < 0.05$) between median GCR rates.

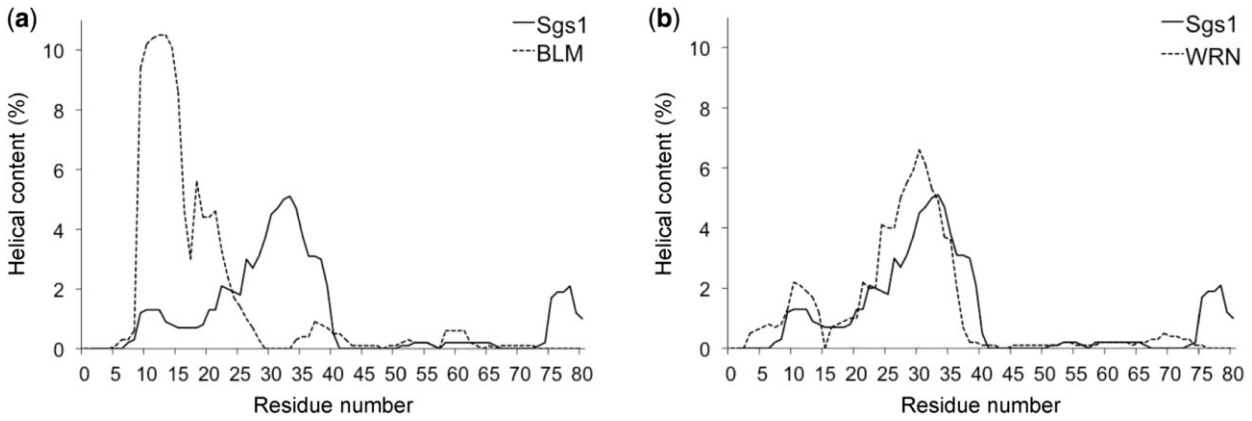


Figure 3.8. Helical content prediction for the N-termini of Sgs1, WRN and BLM (57). (**a**) In human BLM, which binds to the human Top3 homologue Topo IIIα, a prominent peak of helical content is predicted at residues Q12 and L13, which corresponds to the small R10–E12 peak in Sgs1. A peak corresponding to that at residue I33 in Sgs1 is not predicted in BLM, in part because of a proline residue at position 30. (**b**) The distribution of predicted helical content for the N-terminus of human WRN, which binds to Topo I, but has not been shown to bind to Topo IIIα, is similar to Sgs1, with two prominent peaks at residues E10 and A30, corresponding to similar peaks at R10–E12 and I33 in Sgs1.

Table 3.2: Plasmids used in this study

Plasmid	Description
pRS415	*CEN/ARS, LEU2*
pKHS443	*pET28a-Sgs1N1-125*
pKHS463	*pET28a-Sgs1N1-80*
pKHS481[i]	*pRS415-SGS1*
pKHS482	*pRS415-SGS1-F30P*
pKHS484	*pRS415-SGS1-W92PL93P*
pKHS485	*pRS415-SGS1-L93P*
pKHS489	*pRS415 -SGS1-K26P*
pKHS492	*pRS415-SGS1-V29P*
pKHS494	*pRS415-SGS1-D25P*
pKHS496	*pRS415-SGS1-I33P*
pKHS497	*pRS415-SGS1-Q34P*
pKHS546	*pGEX-6p-2-SGS1N1-250-K26P*
pKHS547	*pGEX-6p-2-SGS1N1-250-V29P*
pKHS548	*pGEX-6p-SGS1N1-250-F30P*
pKHS582	*pRS415-SGS1-L9P*
pKHS583	*pRS415-SGS1-H13P*
pKHS584	*pGEX-6p-2-SGS1N1-250-L9P*
pKHS585	*pGEX-6p-2-SGS1-250-H13P*
pKHS586	*pGEX-6p-2-SGS1N1-250-T21P*
pKHS587	*pGEX-6p-2-SGS1N1-250-K17P*

Table 3.2: Plasmids used in this study (Continued)

Plasmid	Description
pKHS588	*pRS415-SGS1-L181P*
pKHS589	*pRS415-SGS1-L215P*
pKHS590	*pRS415-SGS1-T61P*
pKHS591	*pRS415-SGS1-L176P*
pKHS592	*pRS415-SGS1-W15P*
pKHS594	*pRS415-SGS1-I37P*
pKHS595	*pGEX-6p-2-SGS1N1-250-D25P*
pKHS596	*pRS415-SGS1.MYC*
pKHS598	*pRS415-SGS1.MYC-F30P*
pKHS600	*pRS415-SGS1.MYC-H13P*

pRS415 is a gift from Dr. Steven Brill (Rutgers University)

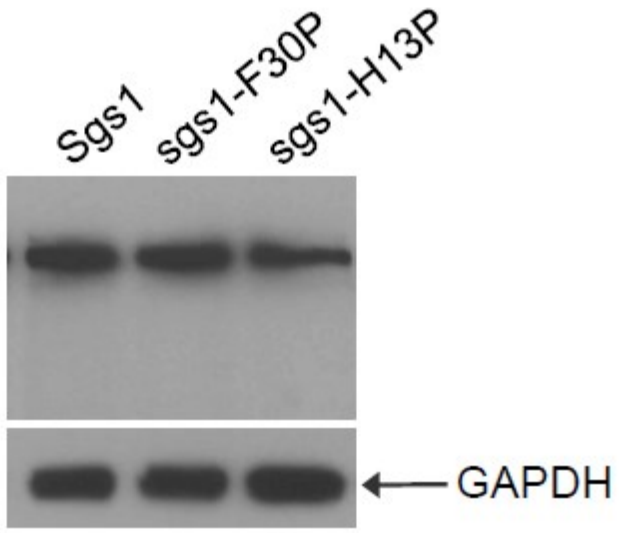


Figure 3.9: Proline substitutions in Sgs1 that cause hypersensitivity to the DNA damaging agent hydroxyurea do not affect expression levels of the sgs1 mutant proteins. The *sgs1Δ* mutant KHSY1338 was transformed with plasmids pKHS596 (Sgs1.myc), pKHS598 (sgs1-F30P.myc) or pKHS600 (sgs1-H13P.myc). Whole cell extracts were prepared by TCA extraction from mid-log phase cultures and separated by 10% SDSPAGE. Wildtype Sgs1 and the sgs1-F30P, sgs1-H13P mutants were detected by Western blotting with monoclonal antibody against the C-terminal myc-epitope. GAPDH was detected by a monoclonal antibody against GAPDH.

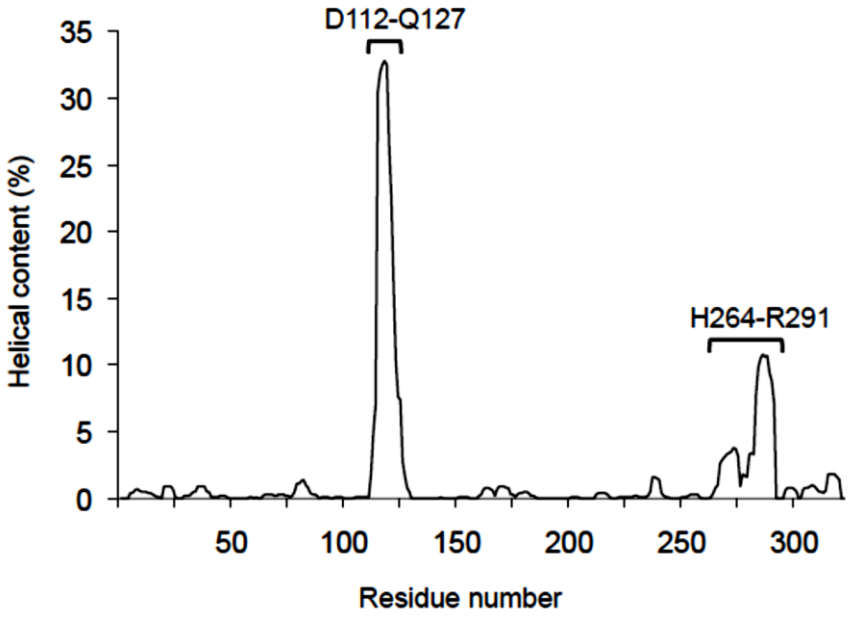


Figure 3.10: Helical content prediction for the N-terminus of *S. pombe* Rqh1 by AGADIR (57). The distribution of helical content for the first 322 residues, which are required for interaction with Top3 (9), are shown. The first prominent peak is predicted at residue M117 within a 15-residue segment spanning from D112 to Q127, and a second peak at I286 within a 27-residue segment spanning from H264 to R291.

CHAPTER FOUR:

THE ROLE OF PROTEIN DISORDER IN CHROMATIN PROCESSES OF *S. CEREVISIAE*

Introduction

Classically, proteins have been defined by three dimensional structure—function is dictated by the shape a protein takes. Recent developments, however, suggest that Intrinsically Disordered Proteins/Regions (IDP/Rs) play a large role in the sometimes complex functions of many proteins. Disordered proteins are found in the regulation of transcription/translation, cell signaling, phosphorylation, and small molecule storage [1-6]; given this wide range of duties, it should be unsurprising that this class of proteins is well-represented in many organisms. Studies using disorder prediction software have suggested that bacterial proteomes can be expected to be between 7-33% disordered, and Archaea 9-37% [7]. This percentage increases with organism complexity. In *Saccharomyces cerevisiae*, up to 60% of the total proteome is predicted to have some degree of disorder spanning at least 30 consecutive residues, the minimum definition for an IDR [8]. Approximately 6% of the proteome estimated to be wholly disordered [9]. Biologically, these proteins appear to be enriched in functions including the stress response, cell cycle, and electron transport [9]. Up to 25% of the total human proteome is predicted to be disordered, and include proteins with a wide range of functions, including Securin, Calpastatin, Caldesmon, and Tau [8-10]. As a class, IDRs are characterized by a large degree of polar and charged residues, a deficiency in large,

bulky, hydrophobic residues, a lack of persistent secondary and tertiary structure, little sequence complexity, a low degree of compactness, and a high degree of open, random-coil “shape,” existing as an ensemble of dynamic conformations [1, 11-13]. A lack of hydrophobic residues may prevent the protein from adopting structure as it prevents the formation of a hydrophobic core, but there are known instances of “induced-folding” models, whereupon disordered proteins assume structure once they are bound by a specific ligand [14-19]. Such disordered regions may have small pockets of persistent or transient secondary structure that may facilitate initial binding [20]. Because the proteins do not commit to one particular structure, a common theory of “binding promiscuity” in disordered proteins means that open conformation becomes a boon to proteins needed in several biological pathways, as the same area could serve as a binding site to several different proteins. A well-studied example is the Transactivation Domain of p53, which mediates interactions between several proteins, including p300/CBP and MDM2/MDM4 [21]. In addition to this benefit, IDP/Rs are also good sites for post-translational modifications, since the protein’s flexible nature allows for easy access to modification sites by a wide range of effectors [22]. Flexibility in the protein can also favor faster interaction rates between protein and ligand, as the open conformation’s larger target size may eliminate the restrictions of orientation present in the rigid “lock-and-key” models of structured proteins. IDRs provide easy access for kinases and phosphatases, thus making them ideal with regards to biological processes that have time constraints and the need for transient modification events [7]

IDP/Rs can exist as either entropic chains—which provide elasticity or act as “linker” regions between structured domains—or a second major functional class that is

comprised of “display sites” (sites for post-translational modification by modifying enzymes), “chaperones,” “effectors” (modifiers of binding partners), “assemblers” (responsible for binding several proteins, possibly in large complexes) and “scavengers” that store small molecules [23]. As proteins can have multiple disordered regions, it is not impossible that one protein can serve several of these functions. A look at the 356 wholly disordered proteins of *S. cerevisiae* reveals this complexity—142 are unidentified reading frames with no known function; truly, these proteins may serve in many capacities, and this level of complexity may make elucidation of some functions difficult [8]. Difficulty assigning a function to the protein products of these genes may also arise from the lack of structure that the proteins are hypothesized to assume; with no known folds and domains, it becomes difficult to make comparisons to proteins with similar structure and known function. Disordered regions can also adopt different structures upon binding to different ligands which further complicates the identification of relevant biochemical pathways and protein/protein interaction in that the same disordered area can serve as a functionally different hub for several different partners. For instance, the tumor suppressor p53 C-terminus has been found to adopt different types of secondary structure when bound to different partners, and the HIF-1α activation domain adopts either alpha helical or extended structures depending on whether it is binding co-activator p300 or transcriptional mediator FIH [1]. This caveat is important when considering conservation of certain structural motifs between homologs; since disordered proteins can undergo induced folding upon partner binding, the use of X-ray crystallography as a means to study these regions coupled with a partner may be misleading. In addition to perhaps underestimating the degree of disorder in a

proteome based on tandem purification and crystallization [8], the conformation of the protein that is disordered in apo formation may vary substantially depending on which partner it is coupled with in the 3d analysis. Hence, there are many other tools in use to study disordered proteins in both complexed and unbound form, including nuclear magnetic resonance, small-angle x-ray scattering, atomic force microscopy, and paramagnetic spin labelling [24].

When DNA damage in *S. cerevisiae* occurs, DNA repair proteins in several different pathways receive signals from DNA damage sensor proteins via mediator proteins and kinases in an effort to modify the chromatin in such a way that facilitates repair such as homologous recombination, non-homologous end joining, and excision repair. Depending on the type of damage, proteins involved in detecting and properly repairing damage can include several checkpoint proteins and kinases [25-27] (e.g. Mec1, Rad53), which require modification from and interaction with upstream and downstream protein elements, and enzymes for functions ranging from break resection [28-30] (e.g. Exo1, Sgs1, MRX proteins), base excision [31] (e.g. Rad27) nucleotide excision [32] (e.g.. Rad16), and ligation of DNA breaks [33] (e.g. Lig4), which require the ability to themselves be modified, as well as interact with other proteins that may be found at DNA breaks, unusual DNA structures, or stalled replication forks. Given this complex network of interaction and modification, we hypothesized that *S. cerevisiae* DNA repair proteins may contain an enrichment of functional disorder, since this type of non-structure facilitates these protein functions well. Study into some yeast proteins have indeed found that these proteins have predicted or confirmed regions of disorder important for proper protein function [34, 35]. Given that the proteins of DNA repair

and related chromatin processes are, overall, highly prone to interaction and modification, and since one of the major roles of disordered regions is to facilitate these, we asked whether or not disorder is overrepresented in the *S. cerevisiae* proteins, and, if so, which of these points seems to be a driving force in the occurrence of functional disorder. To this end, we have conducted a survey of 160 proteins classified in the protein database, Uniprot, as DNA repair proteins and analyzed them using PONDR-VLXT[36] and PONDRFIT [37], software that can predict both the presence of disorder and putative protein binding sites. Analysis established the degree to which disorder plays a role in budding yeast DNA repair, and how this disorder may be a function of protein/protein interaction and phosphorylation. Our analysis gives some insight into the role of functional disorder in this protein subset, and can be used as a source for experimental design beyond classical approaches like sequence alignments.

Methods

Formulation of protein data set

Our protein data set was manually constructed by searching the *S. cerevisiae* Uniprot (European Bioinformatics Institute) database for the term “DNA repair.” The top 160 non-repetitive hits were selected and their FASTA sequence files were obtained from the Saccharomyces Genome Database (SGD) at http://www.yeastgenome.org/.

Disorder prediction and interaction site prediction

Disorder prediction was provided by PONDR-VLXT [36], and interaction site prediction was analyzed using PONDRFIT [37] software, provided at http://www.disprot.org/pondr-fit.php and http://www.pondr.com/index.

Repair classification and phosphorylation

Repair classifications for all protein in the data set were manually determined by using the GO terms provided in the Saccharomyces Genome Database (SGD). Phosphorylation sites were manually evaluated by using the SGD Phosphogrid database on the "protein" section of each entry. Phosphogrid can be found at www.phosphogrid.org.

Results

PONDR-VLXT predicts disorder based on a protein's likelihood to form a folded core. Analysis using PONDR-VLXT in human cancer-associated proteins found that these proteins, as well as regulatory and cytoskeletal proteins were two-fold as likely to contain disordered regions >30 residues than proteins involved in metabolism, biosynthesis and degradation [3]. Hub proteins, those with >10 binding partners, were found to be significantly more disordered than non-hubs in protein sets from *C. elegans*, *S. cerevisiae*, *D. melanogaster*, and *H. sapiens* [35]. A study of the yeast transcriptional network showed that within yeast transcriptional proteins, transcriptional hubs contained the greatest amount of disorder [38]. In addition to binding, functional disorder can be used to modulate post-translational modifications, as flexible linker regions, as a

stimulus for protein degradation, and as scaffolding regions for protein complexes [7, 39, 40]. Given the prevalence of IDRs, and given that proteins in the chromatin processes are often subject to roles that disorder facilitates, we asked whether or not *S. cerevisiae* chromatin processes proteins were more prone to disorder than the whole *S. cerevisiae* proteome. We evaluated proteins in the chromatin processes data set for the presence of at least one IDR as predicted by PONDR-VLXT; in the data set, 72.5% of the proteins contained disorder (Figure 4.1A), versus the predicted 60% in the total proteome [9]. In order to further investigate the function of this overrepresentation, we first considered the possibility of increased number of protein/protein interaction sites in chromatin processes proteins, since IDRs are prone to modulate several interactions. PONDRFIT has the capability to predict interaction sites on a protein by evaluating whether a stretch of amino acids will be chemically fit for burial into a partner protein. Of the 160 proteins in our set, PONDRFIT predicted at least one interaction site in 110 of the proteins, with an expected decrease in the number of proteins found with an increasing number of predicted sites (Figure 4.1B). To further investigate the role of disorder in protein/protein interaction of chromatin processes proteins, we asked what fraction of these interaction sites occurred in proteins with an IDR (Figure 4.1C). Interestingly, the proteins with no predicted interactions sites were overwhelmingly proteins without IDRs (20% ordered vs. 11% disordered), while proteins with one or more interactions were overwhelmingly proteins with at least one predicted IDR. Perhaps even more striking, proteins with 3 or more predicted interaction sites are exclusively proteins with IDRs, suggesting a primary role for protein/protein interaction with regards to disorder in the chromatin processes data set. When evaluating the data

for predicted interaction sites, we further focused on the proteins in the top 5% with regards to number of predicted interactions. These proteins contained a large number of interaction sites (8-13 sites). In order to establish a connection between number of interaction sites and biological function, we researched the putative physical binding partners for the proteins in this top 5%, including Mrc1, Rad9, Slx4, Sgs1, Zip1, Rad2, and Xrs2 (Figure 4.1D), using the SGD database, and found that all seven have been found to interact with a large number of protein partners (Table 4.1).

With these proteins in mind, we hypothesized that the high occurrence of disorder in the chromatin processes proteins was a result of disorder facilitating the predicted interaction sites. We then mapped the PONDRFIT data to the disorder prediction from PONDR-VLXT in an effort to ascertain to what degree disorder and interaction sites are connected. For the top 5%, there is a clear bias toward interaction sites located in disordered regions; for all proteins in this set, anywhere between 50 and 75% of the predicted interaction sites occurred in areas of disorder. There was also a large overall number of disordered regions in these proteins, with anywhere from 4 to 10 total regions ≥30 residues identified (Figure 4.2A). These regions vary in length between 30 residues in Rad2 and 304 residues in Sgs1; for the most part, however, there is little variation in average region length, with a mean range of 65-119 residues for the 7 proteins (Figure S4.2). Also of note was the location of the majority of those residues predicted by PONDR-VLXT to be disordered; for the top 5% of interactors, there was little difference between the percent of disordered residues found within IDRs and the overall percentage of disordered residues in the total protein, implying that the majority of disordered residues are in longer stretches of disorder, and suggesting that

the disorder is functional (Table 4.3). In an effort to understand the biological role that these interaction sites in disorder may play, we looked at the DNA damage signaling cascade, which contains 3 of the 7 proteins in the top 5%. The adaptors of the pathway, Rad9 and Mrc1, contain the greatest number of predicted interaction sites, and a large percentage (64% and 69%, respectively), are predicted in disordered regions. Because these proteins mediate several up and downstream partners, it is possible that the increased number of binding sites in disorder serves to accommodate all partners over the smallest amount of primary structure possible, as has been seen in other IDRs (e.g. p53) [41]. Another protein from the top 5%, Xrs2, is not a mediator of the signaling response, but is a member of an upstream complex whose partners contain small numbers of predicted interaction sites. Thus, Xrs2 presents another function for increased disorder and protein interaction in the chromatin processes set; Xrs2 can contain a large number of interaction sites, freeing the remaining complex members, Mre11 and Rad50, for other functions, particularly catalytic nuclease and ATPase activities; we discuss this function further below (Figure 4.2B).

Given that the top 5% had a large number of interactions predicted in IDRs, and given that a large number of predicted interactions seemed to connect well with a biological function requiring disorder modulating protein/protein interaction in the signaling cascade, we asked whether or not there were other traits of disorder in these proteins that correlated with an increased number of protein interaction sites. We first evaluated the data set for overall percentage of disordered residues per protein as predicted by PONDR-VLXT. No proteins in the set contained all residues with a disorder probability < 0.5, and one protein, Sml1, was 100% disordered. Sml1 is an inhibitor of

ribonucleotide reductase, and is degraded in response to DNA damage to increase dNTP pools needed for repair [42]. This degradation is regulated by post-translational modifications, which are easily accessed in IDRs. In addition, it has been suggested that there is a weak correlation between disorder and increased degradation rate; given Sml1's need to be overturned in the presence of DNA damage, the prevalence of disorder in this protein could have a functional role [43]. Overall, the proteins in the data set were skewed toward being 40% or less disordered, with 20% of the proteins exhibiting 1-10% total disorder, 26% exhibiting 11-20% total disorder, 24% exhibiting 21-30% disorder, and 14% exhibiting 31-40% disorder (Figure 4.3A). We predicted a relationship between the overall percentage of disordered residues and number of interactions predicted by PONDRFIT, but found little correlation between the two values (R^2= 0.3149) (Figure 4.3B). Extreme examples like Sml1 and Slx8 reveal why this correlation is so poor; the two proteins are 100% and 76% disordered, but contain two and zero predicted interaction sites, respectively. This could imply that the prevalence of disorder in these proteins is the result of an increased need for modification rather than protein/protein interaction. It is possible, too, that the PONDRFIT analysis is insufficient to detect binding if a protein is in a "fuzzy complex." These binding events are often more dynamic than those seen in induced-fit models [44], and may be overlooked in the algorithm, since they do not normally form a structural state that requires specific (likely hydrophobic) chemistry that the program accounts for when predicting binding sites. When we evaluate the checkpoint cascade, we saw little overall difference in percent disorder amongst most of the proteins in the cascade, despite differences in predicted interaction sites seen in Figure 2B (Figure 4.4). Thus, we suggest that while the high

degree of chromatin processes protein with IDRs is likely a result of need for protein/protein interaction, overall disorder is a poor indicator of increased protein interaction. We then reevaluated the distribution of disorder within the protein set by taking into account only the percentage of the protein residues that have both PONDR-VLXT scores ≥ 0.5 and are located in IDRs. This changed the protein profile from that seen in Figure 4.3A; almost 28% of the proteins contained no residues with PONDR-VLXT scores > 0.5 in IDRs (Figure 4.5A). This shift suggests that a considerable amount of the overall disorder in some proteins in the set are distributed in smaller regions, which may be less likely to be functionally protein-binding than a larger region exceeding 30 residues. We considered that not overall disorder, but rather the number of disordered regions a data set protein contains is a good predictor of protein/protein interaction propensity. When we compared this value to predicted interaction sites, a stronger correlation resulted (R^2= 0.6026) (Figure 4.5B). Looking back at the checkpoint cascade, we see that the proteins in the cascade with the highest number of disordered domains, Xrs2, Rad9, and Mrc1, are also the proteins identified as having a the largest number of putative interactions (Figure 4.6). Rad50 bucks this trend with 4 disordered domains (like Rad9 and Xrs2), but only 3 predicted interaction sites. One suggestion is that because Rad50 may play a role in binding DNA at double strand break ends, and because disorder has been found to be prevalent in proteins that bind DNA [45] . We propose that overall number of disordered domains, then, is a better indicator of increased protein/protein interaction, and that biological function, particularly as a binding hub or central binding protein in a complex, is highly connected to a greater number of disordered domains.

While evaluating the connectivity between biological function in the checkpoint cascade with protein interaction, we noticed Xrs2 as the sole member in its complex with Mre11 and Rad50 with a high number of predicted interaction sites versus the partners (10, versus 3 for both Mre11 and Rad50). We asked whether this pattern was also repeated in other complexes that had two or more members in our data set. When we compared the number of predicted interaction sites between complex protein members, we noticed that, in general, most complexes, including Mre11/Rad50/Xrs2, Slx4/Slx1, Rad55/Rad57, Slx5/Slx8, Lif1/Lig4 Sgs1/Rmi1, and Rad2/Tfb1/Tfb2/Rad3/Ssl1 contained a complex member that had 3 or more predicted interaction sites and a complex member with little or no predicted interaction (Figure 4.7A). This observation correlates well with the complex member with the highest number of predicted interactions also being the complex member with the greatest number of disordered domains, with Mre11/Rad50/Xrs2 and Rad55/Rad57 being the two notable exceptions (Figure 4.7B). Conservation of those proteins in complex with the most protein interaction and disorder is often the most poor of all complex members. Slx4 is so poorly conserved that BLAST analysis only returns hits for Saccharomyces species [46]. Xrs2 is poorly conserved when compared to Mre11 and Rad50, and is only found in eukaryotes [47]. The human homolog, Nbs1, only has 29% homology to Xrs2 in the structured N-terminal region, and no significant homology in the C-terminus, which is predicted to be disordered in both proteins [48]. As disordered regions tend to tolerate evolution better than ordered regions [49], it suggests an evolutionary advantage to these complex members with regards to chromatin processes. Evolution can occur in one protein more neutrally to account for increasing organismal complexity, while the

others, often with catalytic domains or specialized folds, can evolve more conservatively.

Another major role of functional disorder in proteins is post-translational modification. The “open” conformation of disordered regions provides a more accessible area for modifiers like kinases to apply modifications to residues than those seen in folded proteins, which bury a large percentage of its primary structure [50]. Post-translational modification like phosphorylation is also used as a means of enthalpy compensation for the entropy loss incurred when disordered regions undergo folding in response to protein/protein interaction; thus, modification and interaction at disordered sites are sometimes tied [51, 52]. Protein/protein interaction is clearly a driving force for increased need in disorder in the chromatin processes proteins over the yeast proteome at large, but since several of these proteins undergo post-translational modification in response to cellular signal for chromatin modification, we also considered that modification played a role in the disorder overrepresentation in our data set. Phosphorylation is one of the most closely studied modifications in *S. cerevisiae*, and high-throughput studies of the phosphoproteome have been conducted [53, 54]. Thus, we decided to use phosphorylation as an indicator of the role of disorder in modification of chromatin processes proteins. We used phosphogrid to acquire phosphorylation sites for all proteins and began by accessing whether phosphorylation tended to occur more often in proteins with or without disorder. In our set, 14% of the proteins have no IDRs or any phosphorylation sites; 11% were completely ordered, but contained at least one phosphorylation event. The most represented group in the set, however, were proteins both containing disorder and at least one phosphorylation site, which comprised 64% of

the total chromatin processes data set (Figure 4.8A). We decided to look at this particular group more closely to ascertain if this increase in phosphorylation was actually occurring in IDRs in these proteins. When we looked at proteins with at least a single phosphorylation event and mapped the site to disordered regions as predicted by PONDR-VLXT, we established that a small percentage (19%) had no known phosphorylation occurring in disorder. The overwhelming majority of proteins had at least one of the known phosphorylation events occurring in disorder, and 32% of the proteins had all known phosphorylation events occurring in disordered regions (Figure 4.8B). Given these results, it can be surmised that disorder in the chromatin processes proteins exists to accommodate both increased protein/protein interaction and phosphorylation events.

Discussion

In this study, we evaluated the role of disorder in the Chromatin Process proteome, and established that disorder is overrepresented in these proteins, likely as a result of the need for extensive protein/protein interactions and sites for modification via phosphorylation. We identified a subset of Chromatin Process proteins that are highly connected in the number of predicted interaction sites and the location of these sites in disorder. Based on other studies in disorder, the overrepresentation of disorder in DNA repair is consistent with other findings; it has been reported that biological function with a high need for regulation, signaling, and complex formation [55], all roles that are highly prevalent in DNA repair. While our analysis in the checkpoint cascade clearly shows that there is not a need for disorder at all steps, those proteins with the highest

orders of disordered domains are the mediators/adaptors of the cascade response, and have a larger number of substrates. A large number of disordered interaction areas allow the protein to accommodate the various proteins in downstream interaction without the need for the protein being too large or adding large, bulky domains over the course of evolutionary history as species and DNA damage response becomes more complex. Proteins in the cascade like the kinases, however, do not have a need to be extensively disordered if their downstream substrates are; indeed, the kinases in the cascade (Mec1, Tel1, Dun1, Chk1, Rad53), all display low orders of disorder. In the case of Mec1 and Tel1, the immediate substrates are rich in disordered domains, and in the complete data set, there is a high variability in disorder in the effector proteins; it is not beyond reason that a good deal of these will be disordered or in complex with a disordered protein, as approximately half of kinase substrates are highly unstructured, while only 19% are not, and 85% of kinases with 50% or more of substrates with a high level of disorder (>30% of the protein) are implicated in response to a particular stimulus or stress, such as DNA damage [56].

It is interesting that greater number of protein/protein interactions is not a function of total protein disorder, as a great deal of flexibility may imply a greater deal of binding promiscuity possible, allowing for several interactions over one large area via an induced-fit folding model. However, the role of linear motifs, short, sometimes well-conserved stretches often found in disordered regions can explain why we see a greater number of interaction in proteins with several smaller disordered domains; linear motifs are interaction areas that are a subset of induced-fit interactors, and are implicated in transient and low-affinity binding events, such as modification by kinases [57]. Using

the profiles of the high-number interactors as an example, one can see that there are several “dips” in the disorder prediction plots that cover short spans of residues (Figure 4.9). It is possible that these are linear motifs that break up what would otherwise be a long disordered domain, thereby increasing the number of domains, but decreasing the overall percentage of disorder, especially in shorter proteins. These motifs can work alone or in tandem with other motifs, and can acquire post-translational modifications that can change the target of the motif. In total, the presence of these motifs may serve as anchoring points for several protein/protein interactions and the disorder flanking them may serve to further facilitate binding. A large number of disordered domains with a large number of surrounding linear motifs are a likely explanation for the higher order interactions; in fact, we have already studied the presence of one motif in Sgs1 extensively using NMR [34]. The transient alpha-helix in the N-terminus of the protein facilitates the binding of the Top3/Rmi1 complex, and the downstream disorder has been shown to accommodate the binding as well, as the optimal binding domain extends well beyond the indispensable docking structure. There may, in fact, even be a second motif that facilitates binding as well, as we found that a longer 250 residue peptide bound Top3 and Rmi1 better than a shorter 160 residue peptide [34]. There is a notable dip in the disorder profile in Sgs1 in the region the shorter peptide lacks; while it is larger than one would expect from a linear motif, it is possible that there are several small structures that make up the region, skewing interpretation of the profile, but implying further binding events over the same length.

The finding that phosphorylation in DNA repair proteins has a strong bias in disorder is something that has been described in whole proteomes previously; a study

of the human proteome revealed that proteins described in GO terms as “phosphorylation” targets was one of the categories investigated most enriched in disordered proteins [58]. An intensive study of the residue composition flanking phosphorylated residues showed an enrichment of “surface exposed” residues around all 3 types of phosphorylated residues and the sequence complexity of these flanking regions were relatively low [59]; both qualities are used to define disordered proteins. p53 is a well-studied example of how phosphorylation provides versatility in disordered regions often not afforded in folded regions; the phosphorylation status of p53 changes the affinity of different protein interactions and some regions have been shown to adopt a different three-dimensional form depending on the partner bound to it, thus this important protein can regulate the protein/protein interactions in time and cover several interactions over a smaller distance needed than seen in a folded protein by utilizing disorder [16, 60, 61]. Because several of the proteins in chromatin processes participate in several different interactions and/or types of repair, the degree of disorder and phosphorylation overrepresented in the chromatin processes protein set is to be expected; in an effort to keep the proteins involved from becoming so large that they are metabolically draining on the cell, the addition of disorder in the transition from bacteria to yeast allows one protein to accommodate several binding events over one region and be modified by signaling proteins as a means to regulate when and where these interactions take place. The DNA damage checkpoint response and facets of DNA repair require interactions that are transient in nature, and disorder that is easily modified accommodates this need, as it tends to be high in specificity, low in affinity, and easily accessed by regulating enzymes. Biochemically, phosphorylation is likely

also biased to these areas in order to accommodate extended protein/protein interaction as the modification can sometimes be sufficient to drive the enthalpic penalty that is incurred when the disordered region undergoes a transition from disorder to order [61]. Since we hypothesize that many of these disordered regions in Chromatin Process proteins, especially in higher order interactors, contain recognition elements to facilitate binding and eventually undergo extensive folding, it is likely that, in some cases, modification of these proteins may be a function of providing the necessary energy for the folding of the element and its flanking disorder to occur.

Because we have shown that disorder is predicted to occur in a large percentage of chromatin processes proteins, and because there is a role for interaction and modification in these regions, we offer a warning to those studying these proteins using classical methods like primary sequence alignment. While it is possible that sequence conservation can exist in areas of conserved disorder, it is more likely that the sequence can vary substantially so long as disorder and the overall chemistry of the region does not change [62, 63]. It is possible, then, to miss areas of functional binding and modification if one relies solely on sequence identity as a means of studying homologs. We suggest that the presence of disordered regions in chromatin processes protein be considered in the future when trying to solve pertinent functional domains and when comparing regions of binding and modification between orthologs.

References

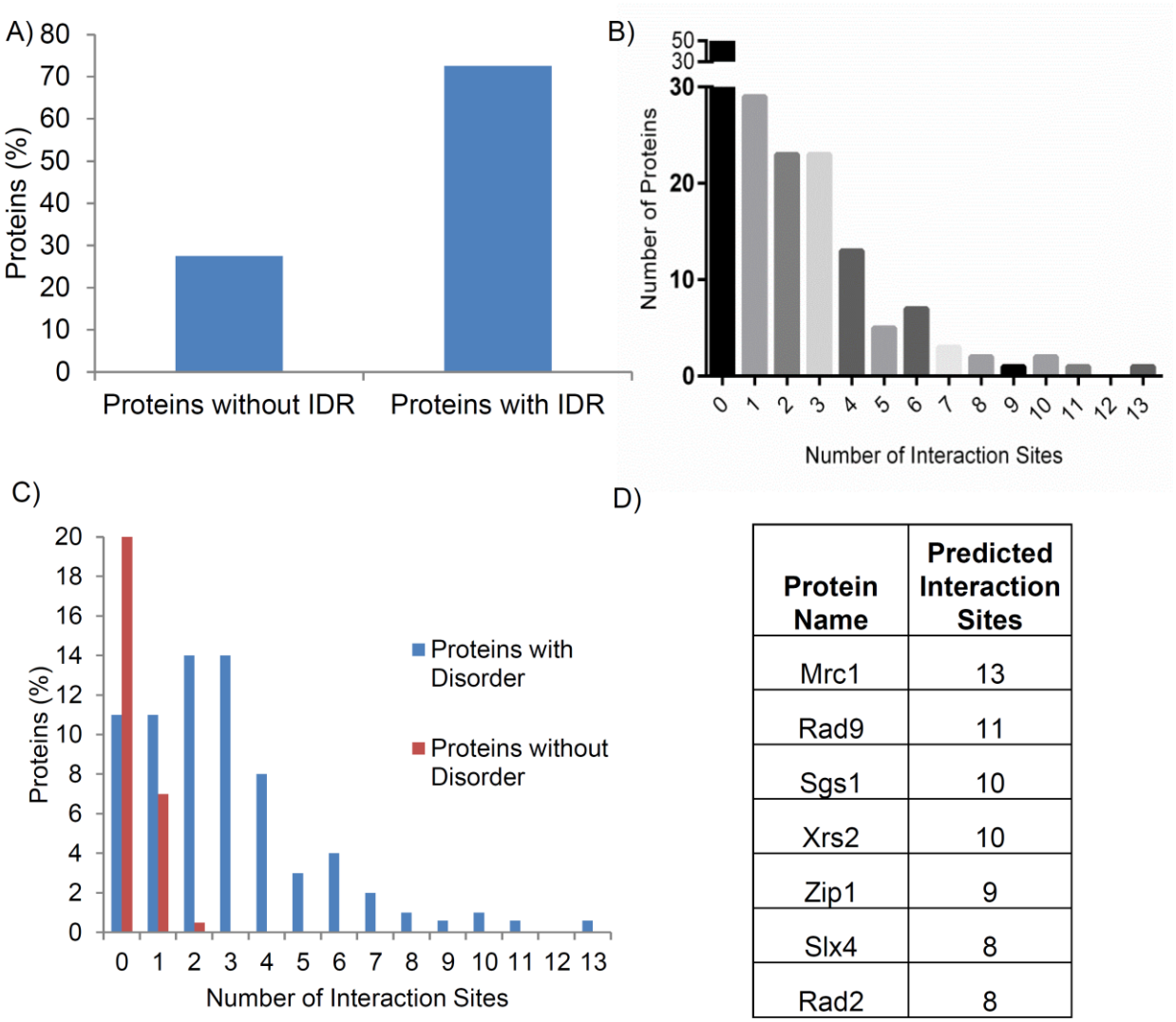
1. Dyson, H.J. and P.E. Wright, *Intrinsically unstructured proteins and their functions.* Nat Rev Mol Cell Biol, 2005. **6**(3): p. 197-208.
2. Liu, J., N.B. Perumal, C.J. Oldfield, E.W. Su, V.N. Uversky, and A.K. Dunker, *Intrinsic disorder in transcription factors.* Biochemistry, 2006. **45**(22): p. 6873-6888.
3. Iakoucheva, L.M., C.J. Brown, J.D. Lawson, Z. Obradović, and A.K. Dunker, *Intrinsic disorder in cell-signaling and cancer-associated proteins.* Journal of molecular biology, 2002. **323**(3): p. 573-584.
4. Uversky, V.N., C.J. Oldfield, and A.K. Dunker, *Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling.* Journal of Molecular Recognition, 2005. **18**(5): p. 343-384.
5. Tantos, A., K.-H. Han, and P. Tompa, *Intrinsic disorder in cell signaling and gene transcription.* Molecular and Cellular Endocrinology, 2012. **348**(2): p. 457-465.
6. Collins, M.O., L. Yu, I. Campuzano, S.G.N. Grant, and J.S. Choudhary, *Phosphoproteomic Analysis of the Mouse Brain Cytosol Reveals a Predominance of Protein Phosphorylation in Regions of Intrinsic Sequence Disorder.* Molecular & Cellular Proteomics, 2008. **7**(7): p. 1331-1348.
7. Dunker, A.K., Brown, Celeste J. et al, *The Protein Trinity: Structure/Function Relationships That Include Intrinsic Disorder.* Proceedings of the Nature Biotechnology Winter Symposium, 2002. **2**(S2): p. 49-50.
8. Dunker, A.K., P. Romero, Z. Obradovic, E.C. Garner, and C.J. Brown, *Intrinsic Protein Disorder in Complete Genomes.* Genome Informatics, 2000. **11**: p. 161-171.
9. Tompa, P., Z. Dosztányi, and I. Simon, *Prevalent Structural Disorder in E. coli and S. cerevisiae Proteomes.* Journal of Proteome Research, 2006. **5**(8): p. 1996-2000.
10. Tompa, P., *Intrinsically unstructured proteins.* Trends in Biochemical Sciences, 2002. **27**(10): p. 527-533.
11. Uversky, V.N., J.R. Gillespie, and A.L. Fink, *Why are "natively unfolded" proteins unstructured under physiologic conditions?* Proteins: Structure, Function, and Bioinformatics, 2000. **41**(3): p. 415-427.
12. Daughdrill, G.W., *Determining structural ensembles for intrinsically disordered proteins.* Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure And Conformation, 2010: p. 107-129.
13. Tompa, P., M. Fuxreiter, C.J. Oldfield, I. Simon, A.K. Dunker, and V.N. Uversky, *Close encounters of the third kind: disordered domains and the interactions of proteins.* BioEssays, 2009. **31**(3): p. 328-335.
14. Longhi, S., V. Receveur-Bréchot, D. Karlin, K. Johansson, H. Darbon, D. Bhella, R. Yeo, S. Finet, and B. Canard, *The C-terminal Domain of the Measles Virus Nucleoprotein Is Intrinsically Disordered and Folds upon Binding to the C-terminal Moiety of the Phosphoprotein.* Journal of Biological Chemistry, 2003. **278**(20): p. 18638-18648.

15. Belle, V., S. Rouger, S. Costanzo, E. Liquière, J. Strancar, B. Guigliarelli, A. Fournel, and S. Longhi, *Mapping α-helical induced folding within the intrinsically disordered C-terminal domain of the measles virus nucleoprotein by site-directed spin-labeling EPR spectroscopy.* Proteins: Structure, Function, and Bioinformatics, 2008. **73**(4): p. 973-988.
16. Oldfield, C.J., J. Meng, J.Y. Yang, M.Q. Yang, V.N. Uversky, and A.K. Dunker, *Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners.* BMC genomics, 2008. **9**(Suppl 1): p. S1.
17. Demarest, S.J., M. Martinez-Yamout, J. Chung, H. Chen, W. Xu, H.J. Dyson, R.M. Evans, and P.E. Wright, *Mutual synergistic folding in recruitment of CBP/p300 by p160 nuclear receptor coactivators.* Nature, 2002. **415**(6871): p. 549-553.
18. Demarest, S.J., S. Deechongkit, H. Jane Dyson, R.M. Evans, and P.E. Wright, *Packing, specificity, and mutability at the binding interface between the p160 coactivator and CREB-binding protein.* Protein Science, 2004. **13**(1): p. 203-210.
19. Mucsi, Z., F. Hudecz, M. Hollósi, P. Tompa, and P. Friedrich, *Binding-induced folding transitions in calpastatin subdomains A and C.* Protein Science, 2003. **12**(10): p. 2327-2336.
20. Fuxreiter, M., I. Simon, P. Friedrich, and P. Tompa, *Preformed Structural Elements Feature in Partner Recognition by Intrinsically Unstructured Proteins.* Journal of Molecular Biology, 2004. **338**(5): p. 1015-1026.
21. Joerger, A.C. and A.R. Fersht, *Structural Biology of the Tumor Suppressor p53.* Annual Review of Biochemistry, 2008. **77**(1): p. 557-582.
22. Gsponer, J. and M. Madan Babu, *The rules of disorder or why disorder rules.* Progress in Biophysics and Molecular Biology, 2009. **99**(2–3): p. 94-103.
23. Felli, I.C., R. Pierattelli, and P. Tompa, *Intrinsically Disordered Proteins*, in *NMR of Biomolecules*. 2012, Wiley-VCH Verlag GmbH & Co. KGaA. p. 136-152.
24. Wright, P.E. and H.J. Dyson, *Linking folding and binding.* Current Opinion in Structural Biology, 2009. **19**(1): p. 31-38.
25. Weinert, T.A., G.L. Kiser, and L.H. Hartwell, *Mitotic checkpoint genes in budding yeast and the dependence of mitosis on DNA replication and repair.* Genes & Development, 1994. **8**(6): p. 652-665.
26. Ma, J.-L., S.-J. Lee, J.K. Duong, and D.F. Stern, *Activation of the Checkpoint Kinase Rad53 by the Phosphatidyl Inositol Kinase-like Kinase Mec1.* Journal of Biological Chemistry, 2006. **281**(7): p. 3954-3963.
27. Hoch, N.C., E.S.W. Chen, R. Buckland, S.-C. Wang, A. Fazio, A. Hammet, A. Pellicioli, A. Chabes, M.-D. Tsai, and J. Heierhorst, *Molecular Basis of the Essential S Phase Function of the Rad53 Checkpoint Kinase.* Molecular and Cellular Biology, 2013. **33**(16): p. 3202-3213.
28. Nicolette, M.L., K. Lee, Z. Guo, M. Rani, J.M. Chow, S.E. Lee, and T.T. Paull, *Mre11–Rad50–Xrs2 and Sae2 promote 5' strand resection of DNA double-strand breaks.* Nature structural & molecular biology, 2010. **17**(12): p. 1478-1485.

29. Garcia, V., S.E.L. Phelps, S. Gray, and M.J. Neale, *Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1.* Nature, 2011. **479**(7372): p. 241-244.
30. Gravel, S., J.R. Chapman, C. Magill, and S.P. Jackson, *DNA helicases Sgs1 and BLM promote DNA double-strand break resection.* Genes & Development, 2008. **22**(20): p. 2767-2772.
31. Wu, X. and Z. Wang, *Relationships between yeast Rad27 and Apn1 in response to apurinic/apyrimidinic (AP) sites in DNA.* Nucleic Acids Research, 1999. **27**(4): p. 956-962.
32. Reed, S.H., Z. You, and E.C. Friedberg, *The Yeast RAD7 and RAD16 Genes Are Required for Postincision Events during Nucleotide Excision Repair.* Journal of Biological Chemistry, 1998. **273**(45): p. 29481-29488.
33. Wilson, T.E., U. Grawunder, and M.R. Lieber, *Yeast DNA ligase IV mediates non-homologous DNA end joining.* Nature, 1997. **388**(6641): p. 495-498.
34. Kennedy, J.A., G.W. Daughdrill, and K.H. Schmidt, *A transient α-helical molecular recognition element in the disordered N-terminus of the Sgs1 helicase is critical for chromosome stability and binding of Top3/Rmi1.* Nucleic Acids Research, 2013. **41**(22): p. 10215-10227.
35. Haynes, C., C.J. Oldfield, F. Ji, N. Klitgord, M.E. Cusick, P. Radivojac, V.N. Uversky, M. Vidal, and L.M. Iakoucheva, *Intrinsic Disorder Is a Common Feature of Hub Proteins from Four Eukaryotic Interactomes.* PLoS Comput Biol, 2006. **2**(8): p. e100.
36. Romero, P., Z. Obradovic, C.R. Kissinger, J.E. Villafranca, E. Garner, S. Guilliot, and A.K. Dunker. *Thousands of proteins likely to have long disordered regions*. in *Pac Symp Biocomput*. 1998.
37. Xue, B., R.L. Dunbrack, R.W. Williams, A.K. Dunker, and V.N. Uversky, *PONDR-FIT: A meta-predictor of intrinsically disordered amino acids.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2010. **1804**(4): p. 996-1010.
38. Singh, G.P. and D. Dash, *Intrinsic disorder in yeast transcriptional regulatory network.* Proteins: Structure, Function, and Bioinformatics, 2007. **68**(3): p. 602-605.
39. Fuxreiter, M., P. Tompa, and I. Simon, *Local structural disorder imparts plasticity on linear motifs.* Bioinformatics, 2007. **23**(8): p. 950-956.
40. Teilum, K., J. Olsen, and B. Kragelund, *Functional aspects of protein flexibility.* Cellular and Molecular Life Sciences, 2009. **66**(14): p. 2231-2247.
41. Uversky, V.N., C.J. Oldfield, and A.K. Dunker, *Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept.* Annual Review of Biophysics, 2008. **37**(1): p. 215-246.
42. Zhao, X., A. Chabes, V. Domkin, L. Thelander, and R. Rothstein, *The ribonucleotide reductase inhibitor Sml1 is a new target of the Mec1/Rad53 kinase cascade during growth and in response to DNA damage.* The EMBO Journal, 2001. **20**(13): p. 3544-3553.
43. Tompa, P., J. Prilusky, I. Silman, and J.L. Sussman, *Structural disorder serves as a weak signal for intracellular protein degradation.* Proteins: Structure, Function, and Bioinformatics, 2008. **71**(2): p. 903-909.

44. Tompa, P. and M. Fuxreiter, *Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions.* Trends in Biochemical Sciences, 2008. **33**(1): p. 2-8.
45. Guo, X.I.N., M.L. Bulyk, and A.J. Hartemink, *INTRINSIC DISORDER WITHIN AND FLANKING THE DNA-BINDING DOMAINS OF HUMAN TRANSCRIPTION FACTORS.* Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing, 2012: p. 104-115.
46. Fricke, W.M. and S.J. Brill, *Slx1—Slx4 is a second structure-specific endonuclease functionally redundant with Sgs1—Top3.* Genes & Development, 2003. **17**(14): p. 1768-1778.
47. Symington, L.S., *Role of RAD52 Epistasis Group Genes in Homologous Recombination and Double-Strand Break Repair.* Microbiology and Molecular Biology Reviews, 2002. **66**(4): p. 630-670.
48. Kobayashi, J., A. Antoccia, H. Tauchi, S. Matsuura, and K. Komatsu, *NBS1 and its functional role in the DNA damage response.* DNA Repair, 2004. **3**(8–9): p. 855-861.
49. Brown, C.J., S. Takayama, A.M. Campen, P. Vise, T.W. Marshall, C.J. Oldfield, C.J. Williams, and A. Keith Dunker, *Evolutionary rate heterogeneity in proteins with long disordered regions.* Journal of molecular evolution, 2002. **55**(1): p. 104-110.
50. Dunker, A.K., I. Silman, V.N. Uversky, and J.L. Sussman, *Function and structure of inherently disordered proteins.* Current opinion in structural biology, 2008. **18**(6): p. 756-764.
51. Huang, Y. and Z. Liu, *Kinetic Advantage of Intrinsically Disordered Proteins in Coupled Folding–Binding Process: A Critical Assessment of the "Fly-Casting" Mechanism.* Journal of Molecular Biology, 2009. **393**(5): p. 1143-1159.
52. Uversky, V.N., *Intrinsically disordered proteins from A to Z.* The International Journal of Biochemistry & Cell Biology, 2011. **43**(8): p. 1090-1103.
53. Fiedler, D., H. Braberg, M. Mehta, G. Chechik, G. Cagney, P. Mukherjee, A.C. Silva, M. Shales, S.R. Collins, S. van Wageningen, P. Kemmeren, F.C.P. Holstege, J.S. Weissman, M.-C. Keogh, D. Koller, K.M. Shokat, and N.J. Krogan, *Functional Organization of the S. cerevisiae Phosphorylation Network.* Cell, 2009. **136**(5): p. 952-963.
54. Ficarro, S.B., M.L. McCleland, P.T. Stukenberg, D.J. Burke, M.M. Ross, J. Shabanowitz, D.F. Hunt, and F.M. White, *Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae.* Nat Biotech, 2002. **20**(3): p. 301-305.
55. Bellay, J., S. Han, M. Michaut, T. Kim, M. Costanzo, B.J. Andrews, C. Boone, G.D. Bader, C.L. Myers, and P.M. Kim, *Bringing order to protein disorder through comparative genomics and genetic interactions.* Genome Biol, 2011. **12**(2): p. R14.
56. Gsponer, J., M.E. Futschik, S.A. Teichmann, and M.M. Babu, *Tight Regulation of Unstructured Proteins: From Transcript Synthesis to Protein Degradation.* Science, 2008. **322**(5906): p. 1365-1368.

57. Diella, F., N. Haslam, C. Chica, A. Budd, S. Michael, N.P. Brown, G. Travé, and T.J. Gibson, *Understanding eukaryotic linear motifs and their role in cell signaling and regulation.* Front Biosci, 2008. **13**: p. 6580-6603.
58. Lobley, A., M.B. Swindells, C.A. Orengo, and D.T. Jones, *Inferring Function Using Patterns of Native Disorder in Proteins.* PLoS Comput Biol, 2007. **3**(8): p. e162.
59. Iakoucheva, L.M., P. Radivojac, C.J. Brown, T.R. O'Connor, J.G. Sikes, Z. Obradovic, and A.K. Dunker, *The importance of intrinsic disorder for protein phosphorylation.* Nucleic acids research, 2004. **32**(3): p. 1037-1049.
60. Ferreon, J.C., C.W. Lee, M. Arai, M.A. Martinez-Yamout, H.J. Dyson, and P.E. Wright, *Cooperative regulation of p53 by modulation of ternary complex formation with CBP/p300 and HDM2.* Proceedings of the National Academy of Sciences, 2009. **106**(16): p. 6591-6596.
61. Dyson, H.J., *Expanding the proteome: disordered and alternatively-folded proteins.* Quarterly reviews of biophysics, 2011. **44**(4): p. 467-518.
62. Forman-Kay, Julie D. and T. Mittag, *From Sequence and Forces to Structure, Function, and Evolution of Intrinsically Disordered Proteins.* Structure, 2013. **21**(9): p. 1492-1499.
63. Moesa, H.A., S. Wakabayashi, K. Nakai, and A. Patil, *Chemical composition is maintained in poorly conserved intrinsically disordered regions and suggests a means for their classification.* Molecular BioSystems, 2012. **8**(12): p. 3262-3273.



Protein Name	Predicted Interaction Sites
Mrc1	13
Rad9	11
Sgs1	10
Xrs2	10
Zip1	9
Slx4	8
Rad2	8

Figure 4.1: Prevalence of disorder and predicted interaction sites in Chromatin Processes proteins. A) Proteins in the Chromatin Processes data set were evaluated using PONDR-VLXT for at least one stretch of disordered residues (VLXT value ≥ 0.5) greater than or equal to 30 residues (proteins with disorder). B) PONDRFIT was used to predict the protein/protein interaction sites of the proteins in the data set. C) Proteins were divided into proteins with and without at least one disordered region with regards to number of predicted interactions. Interaction prediction is most prevalent in proteins with at least one disordered region. D) Identity of the top 5% of proteins in the set with the highest number of predicted interaction sites.

A)

Protein Name	Total # of Predicted Interaction Sites	% interactions in Disorder	# of disordered regions ≥30 residues
Mrc1	13	69	10
Rad9	11	64	4
Sgs1	10	50	5
Xrs2	10	50	4
Zip1	9	55.5	7
Slx4	8	75	4
Rad2	8	50	5

B)

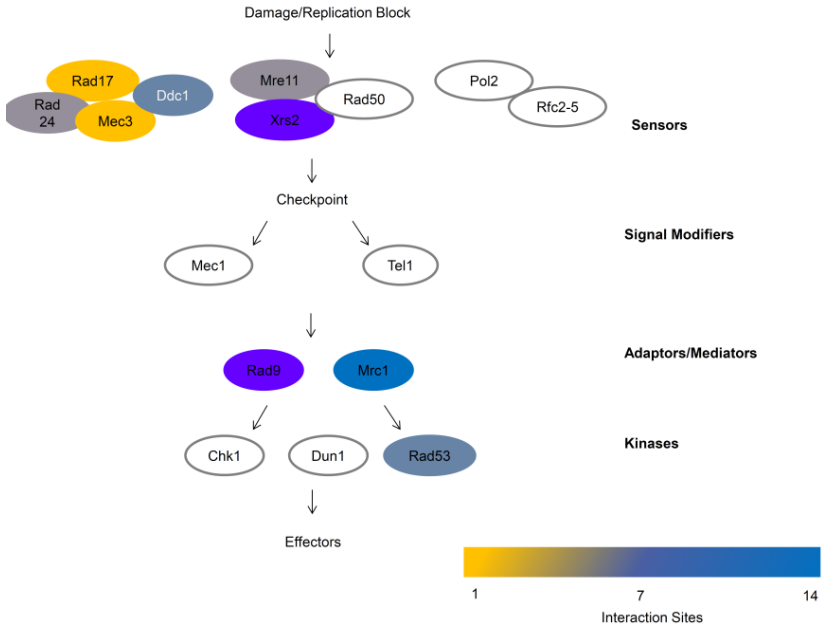


Figure 4.2: Predicted interaction in disorder for top 5% of highest-number interactors and role in checkpoint cascade. A) The top 5% of interacting proteins were evaluated by location of PONDRFIT-predicted interaction in disordered regions predicted by PONDR-VLXT ≥30 consecutive residues. B) The role of 3 of 7 of the top5% in a simplified model of the DNA damage checkpoint cascade. Mediators of the response have the highest number of predicted interaction sites, while the kinases up- and downstream have few or no interactions predicted.

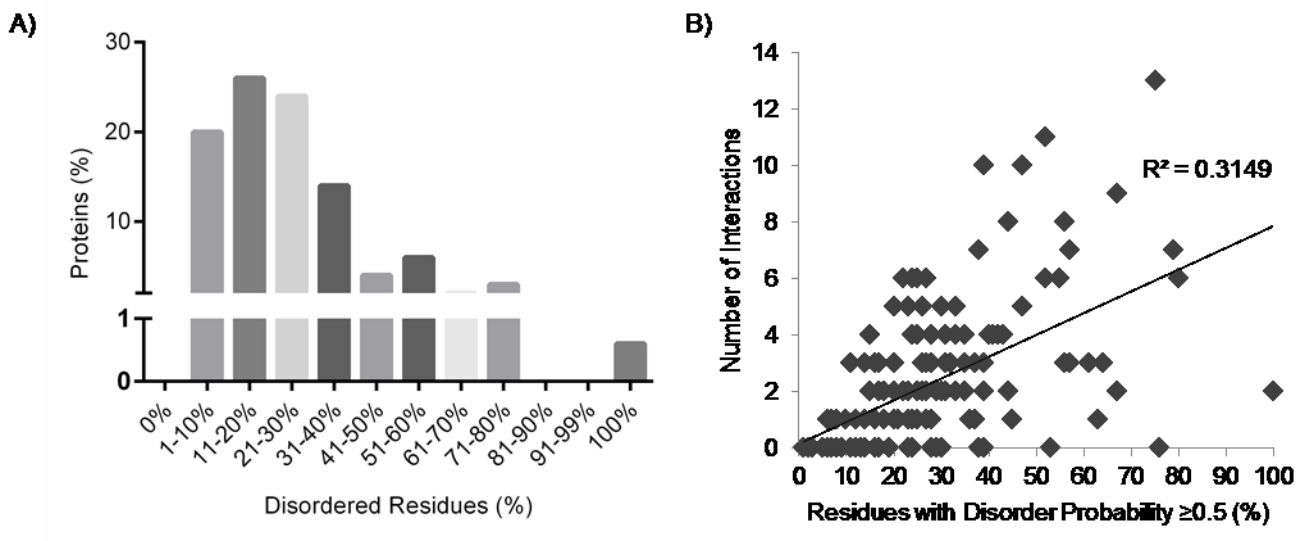


Figure 4.3: Percentage of disordered residues is not an indicator of a large number of predicted interactions. A) Overall protein disorder (residues with value ≥ 0.5) as predicted by PONDR-VLXT versus percent of the protein. B) Correlation between percentage of overall disordered residues and number of PONDFIT-predicted interaction sites.

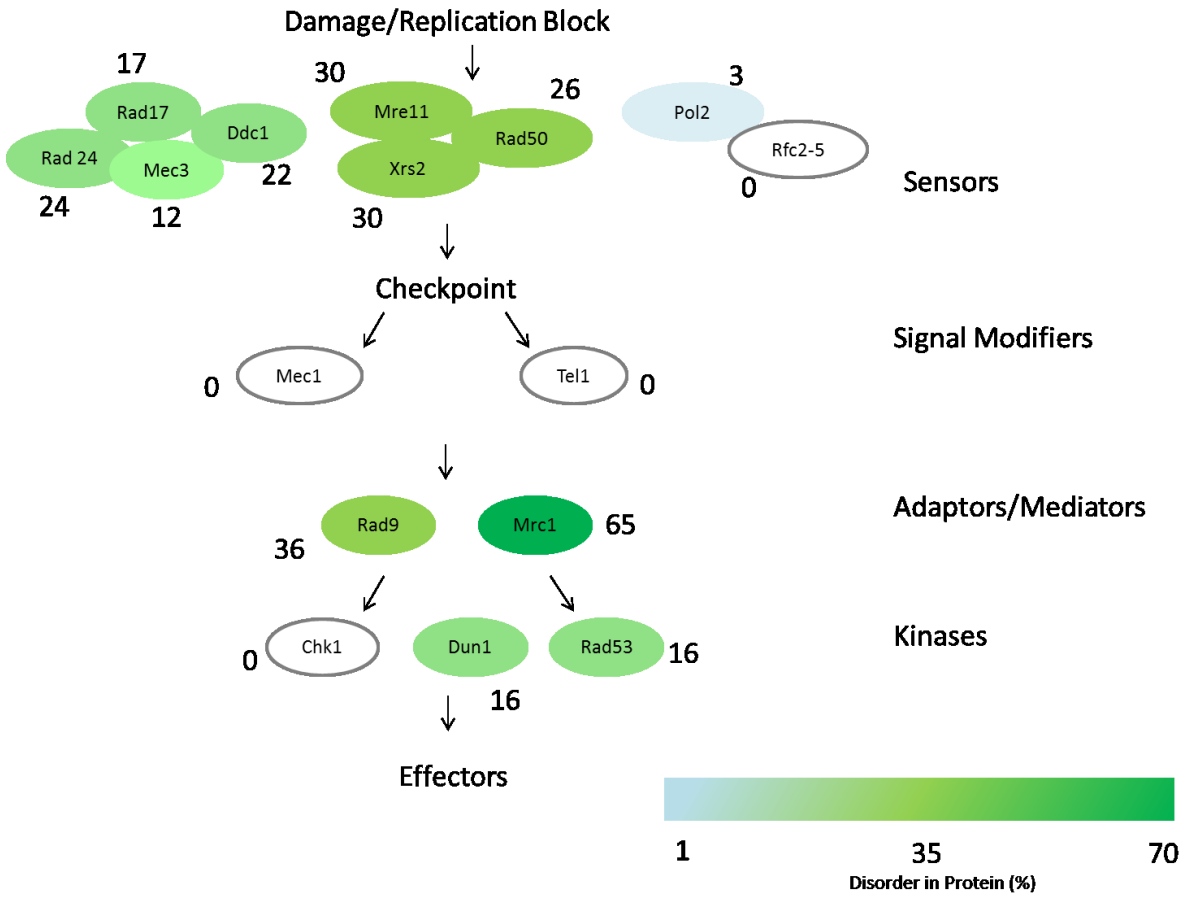


Figure 4.4: Evaluation of percent disordered residues in damage checkpoint response pathway. There is no correlation between role of a protein and/or the number of predicted interactions and percentage of overall disorder. Percent disorder is noted next to the protein.

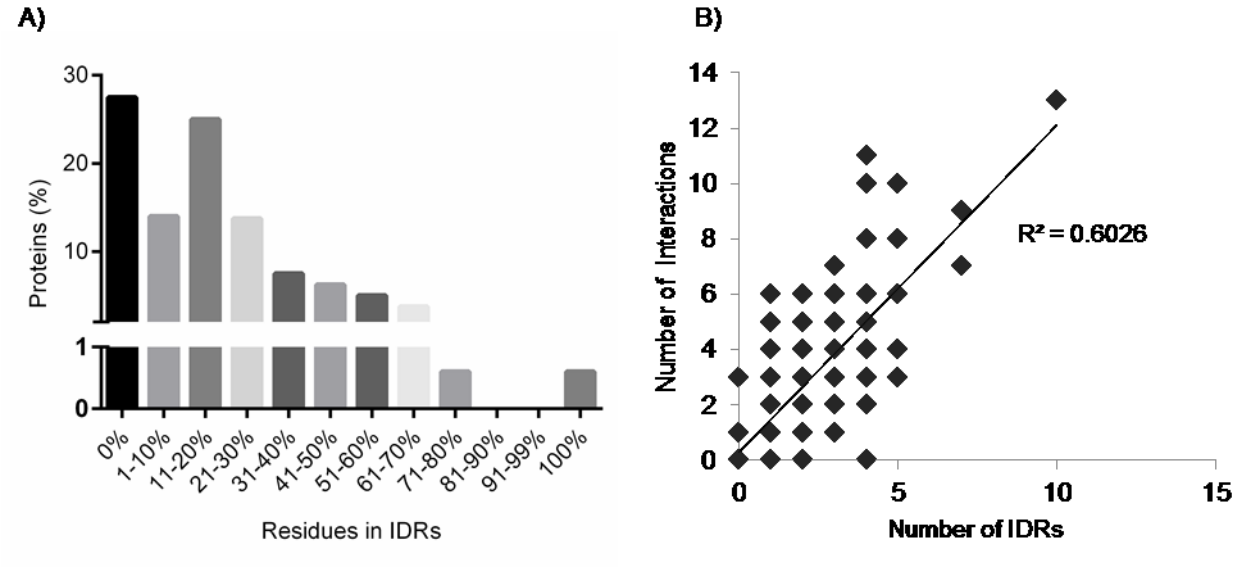


Figure 4.5: Number of disordered domains is the best indicator of a large number of predicted interactions. A) Percent of disordered residues in IDRs as predicted by PONDR-VLXT (residues with value ≥ 0.5). B) A moderate correlation exists between number of total IDRs in a Chromatin Processes protein and the number of predicted interaction sites.

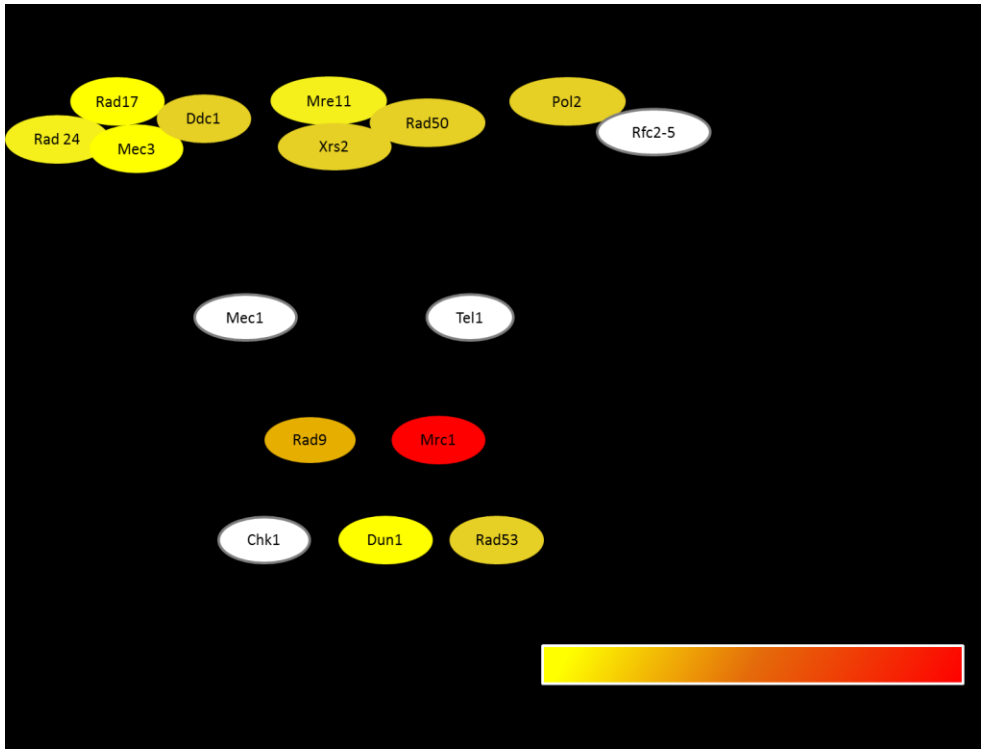


Figure 4.6: Evaluation of number of disordered domains in damage checkpoint response pathway. There is a correlation between role of a protein and/or the number of predicted interactions and number of IDRs as predicted by PONDR-VLXT. Mediators of the checkpoint appear to have the greatest number of IDRs; large numbers are also present in complexed proteins, usually in conjunction with one or more proteins with few IDRs. The number of domains is noted next to the protein.

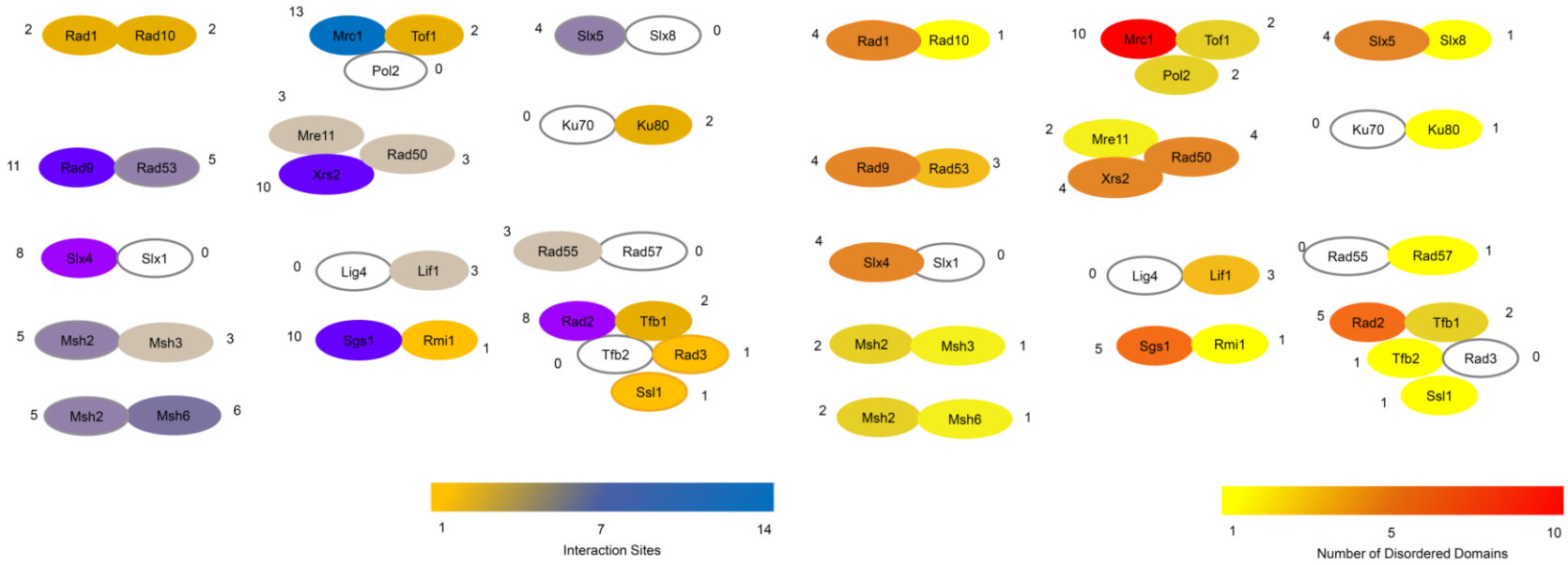


Figure 4.7: Protein complexes in Chromatin Processes proteins have one member with a large number of predicted interactions and disordered domains. PONDR-VLXT predicted disordered domains and PONDRFIT predicted interactions were compared between protein complex members who had at least two members in the data set. The number of predicted sites or domains is indicated next to the protein.

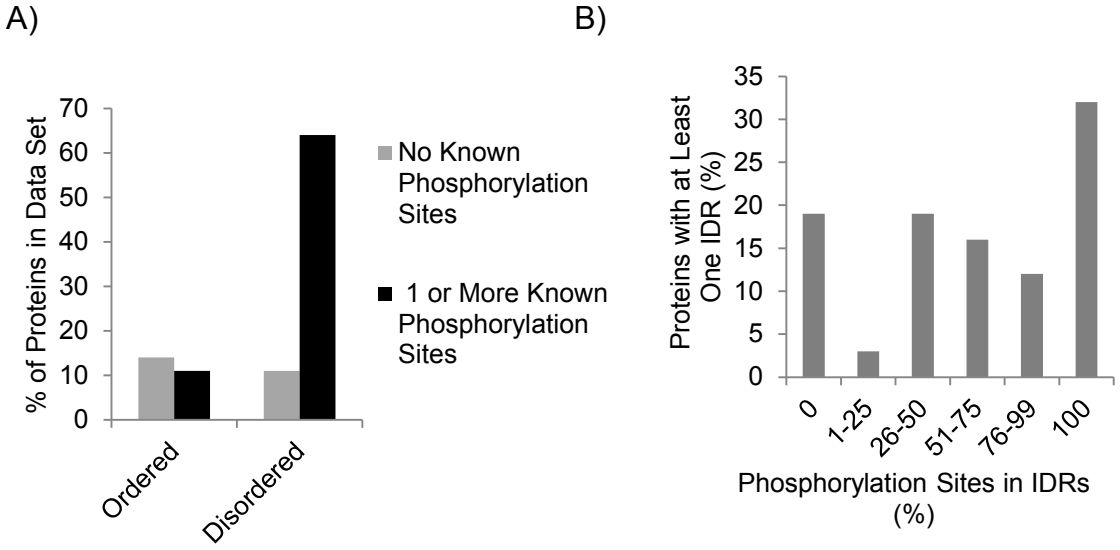


Figure 4.8: Phosphorylation in chromatin processes proteins occurs most frequently in disordered domains. A) Phosphorylation sites for the protein set were taken from PHOSPHOGRID; proteins without a disordered domain dominated the group of proteins in the set with no phosphorylation events, while almost 70% of the data set was both phosphorylated and had at least one disordered domain. B) Phosphorylation occurs completely in disorder in nearly 35% of all proteins with at least one disordered domain.

Table 4.1: Known interacting partners of top 5% highest-number interactors in the Chromatin Processes data set

Protein Name	Total # of Predicted Interaction Sites	Known Interacting Proteins
Mrc1	13	Cdc45, Cmk2, Cse4, Csm3, Ctf4, Dia2, Dpb2, Gwt1, Hht1, Hht2, Hog1, Mcm2, Mcm3, Mcm4, Mcm6, Mec1, Pol2, Prp9, Psf2, Rad53, Rpn3, Sld5, Smt3, Spt7, Stb2, Taf9, Tof1, Tye7, Ubi4
Rad9	11	Cdc14, Cdc28, Cdc5, Chk1, Clb2, Dpb11, Dun1, Hta1, Hta2, Mec1, Nab2, Rad17, Rad52, Rad53, Rad9, Smt3, Ubi4
Sgs1	10	Bud27, Cdc28, Cmr1, Dna2, Frk1, Gis1, Ksp1, Mec1, Mlh1, Mlh3, Mre11, Msh6, Prp45, Rad16, Rad51, Rad53, Rfa1, Rmi1, Rtt107, Smt3, Srs2, Stu2, Swd3, Top3, Ubc9, Yck2
Xrs2	10	Ahp1, Cdc16, Cmr1, End3, Erg20, Hmt1, Htb1, Htb2, Ino4, Lif1, Mre11, Nab2, Pch2, Pst2, Rad50, Rec107, Rif2, Smt3, Tek1, Xrs2, YBR063C
Zip1	9	Apl1, Atg1, Bli1, Cdc28, Cmr1, Cnl1, Cst9, Did2, Eto1, Far10, Gyp5, Hhf1, Hhf2, Htb1, Htb2, Kar3, Kin3, Lhs1, Mad1, Met4, Mps3, Myo2, Nam7, Nnf1, Nnf2, Not3, Nuf2, Pan1, Red1, Sec9, Smt3, Smy1, Spra2, Spc24, Spc72, Spr3, Tae2, Tpk3, Tpm2, Ubi4, Yaf9, YHR080C, Yta6
Slx4	8	Ahp1, Bmh2, Cdc27, Cpr1, Dpb11, Gus1, Hxk1, Rad1, Rad10, Rtt107, Slx1, Smc5, Smt3, Ssz1, Tdh1, Yef3
Rad2	8	Cmk2, Dhh1, Frk1, Gal4, Msh2, Nab2, Pat1, Pex15, Puf3, Rad3, Rpo21, Sat4, Sen1, Slf1, Smt3, Srb4, Srs2, Ssl1, Ssl2, Tfb1

Table 4.2: Characterization of disordered domains in top 5% highest-number interactors in the Chromatin Processes data set

Protein Name	Length of Disordered Regions (# of residues)	Average Region Length
Mrc1	38, 41, 43, 51, 56, 63, 67, 81, 120, 155	72
Zip1	41, 42, 44, 49, 59, 69, 182	69
Slx4	36, 62, 88, 102	72
Rad9	61, 73, 134, 201	117
Sgs1	48, 56, 59, 129, 304	119
Rad2	30, 44, 76, 85, 105	68
Xrs2	44, 66, 70, 79	65

Table 4.3: Characterization of disordered residues in top 5% highest-number Interactors in the Chromatin Processes data set

Protein Name	Total Protein Length	% disordered residues in domains ≥30 residues	% disordered residues in protein
Mrc1	1096	65	75
Zip1	875	55.5	67
Sgs1	1447	41	47
Slx4	748	38.5	39
Rad9	1309	35.8	56
Rad2	1031	33	44
Xrs2	854	30	39

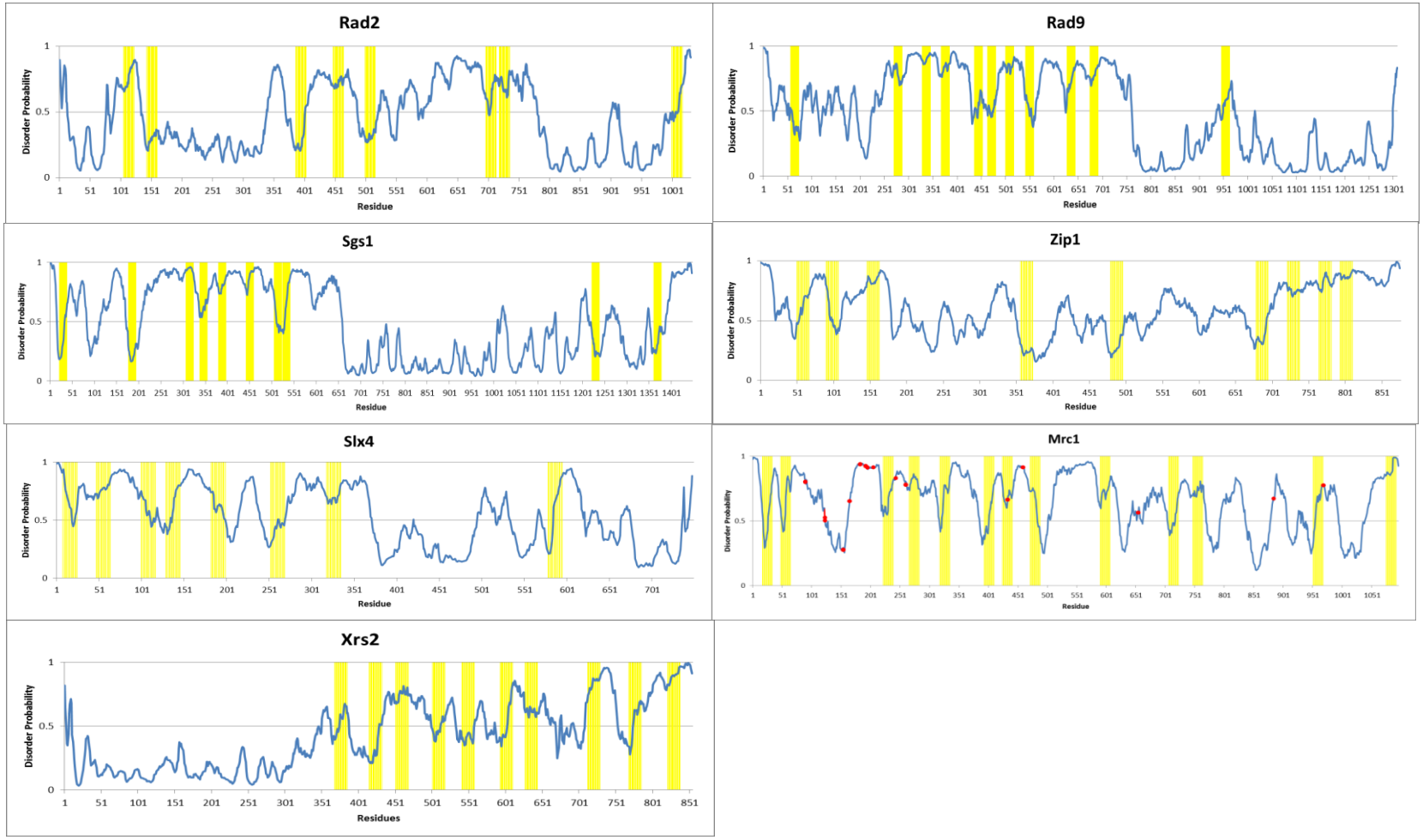


Figure 4.1: IUPRED plots of high-level interactors with putative binding sites. Binding sites predicted by PONDRFIT are highlighted in yellow. A disorder probability > 0.5 suggests a disordered section of the protein. Mrc1 phosphorylation are mapped in red as described .

Table 4.4: Proteins in the Chromatin Processes data set

Protein	% Residues ≥.5	% in Disordered Domains ≥30 Residues	Interactions	# Disordered Domains	# Phosphorylation sites	Pathways (GO terms)
Apn1	24	19	1	1	3	BER
Apn2	24	9	0	1	0	BER
Bre1	40	31	4	3	2	HR, Checkpoint
Cdc25	38	34	7	7	29	Checkpoint
Cdc28	7	0	0	0	6	Checkpoint
Cdc45	23	16	1	2	2	HR, Checkpoint
Cdc7	16	6	3	1	5	HR, Checkpoint, Replication
Cdc9	24	19	1	2	5	HR, NER, BER
Chk1	6	0	0	0	8	Checkpoint
Chl1	16	8	0	1	3	SCC
Csm1	23	0	1	0	0	HR
Csm3	35	10	2	1	1	Replication
Ctf18	20	0	1	0	4	HR, SCC
Ctf4	17	11	2	3	7	HR, SCC
Ddc1	30	22	5	3	7	HR, Checkpoint
Din7	12	0	1	0	0	Replication
Dmc1	8	0	1	0	0	HR
Dna2	23	18	5	2	9	HR
Dot1	33	25	4	2	3	HR, NER, Checkpoint
Dpb11	33	19	2	2	5	MMR, Replication
Dun1	20	16	1	2	25	Checkpoint
Eco1	25	13	1	1	1	SCC
Esc2	61	53	3	2	9	HR, Checkpoint
Est1	12	7	0	1	2	Misc
Exo1	39	31	2	2	5	HR, MMR
Hmi1	5	0	0	0	2	Misc
Hnt3	14	0	0	0	0	Misc
Hsm3	9	0	0	0	1	MMR
Htb1	53	43	0	1	1	Misc

Table 4.4: Proteins in the Chromatin Processes data set (Continued)						
Protein	**% Residues ≥.5**	**% in Disordered Domains ≥30 Residues**	**Interactions**	**# Disordered Domains**	**# Phosphorylation sites**	**Pathways (GO terms)**
Irc20	17	12	3	4	1	HR, NHEJ
Ku70	13	0	0	0	3	HR, NHEJ
Ku80	17	5	2	1	0	HR, NHEJ
Lif1	57	49	3	3	1	NHEJ
Lig4	7	0	0	0	3	NHEJ
Mag1	10	0	1	0	2	BER
Mcd1	43	29	4	2	8	SCC
Mcm10	42	37	4	2	9	Replication
Mcm2	32	28	3	2	8	HR, Replication
Mcm3	26	20	3	2	20	HR, Replication
Mcm4	28	25	3	2	24	HR, Replication
Mcm5	19	4	0	1	4	Replication
Mcm6	30	22	2	3	16	Replication
Mcm7	16	9	1	2	9	HR, Replication
Mec1	2	0	0	0	4	Checkpoint, Replication
Mec3	36	12	1	1	2	Checkpoint
Mgm101	29	20	0	1	1	Misc
Mgs1	18	11	2	1	0	Replication
Mgt1	13	0	0	0	0	Misc
Mhr1	17	0	0	0	0	HR
Mlh1	20	15	2	2	3	MMR
Mlh2	20	13	5	2	0	MMR
Mlh3	5	0	0	0	0	HR, MMR
Mms1	5	0	0	0	2	Replication
Mms2	27	0	1	0	1	Replication
Mms21	14	0	1	0	1	Replication
Mms22	22	14	6	3	0	Replication
Mms4	29	15	2	2	8	HR
Mph1	20	16	3	3	2	HR, Replication
Mrc1	75	65	13	10	31	Checkpoint, Replication
Mre11	35	30	3	2	2	HR, Checkpoint
Msh1	11	5	0	1	2	HR, MMR

Table 4.4: Proteins in the Chromatin Processes data set (Continued)						
Protein	% Residues ≥.5	% in Disordered Domains ≥30 Residues	Interactions	# Disordered Domains	# Phosphorylation sites	Pathways (GO terms)
Msh2	3	0	0	0	1	HR, MMR
Msh3	14	11	3	1	0	HR, MMR
Msh4	14	11	3	1	0	HR
Msh5	7	0	1	0	0	HR
Msh6	25	24	6	1	16	MMR
Mus81	26	13	2	2	2	HR
Nej1	27	23	2	1	3	NHEJ
Nhp10	56	42	3	2	4	Misc
Nse1	7	0	0	0	0	Replication
Nse4	45	37	1	3	2	Replication
Ntg1	13	9	0	1	0	BER
Ntg2	8	0	1	0	0	BER
Ogg1	11	0	0	0	0	BER
Pan2	6	0	0	0	1	Replication
Pan3	32	23	3	1	8	Replication
Pap2	35	34	4	3	3	Misc
Pds1	79	60	7	3	19	Misc
Phr1	6	0	0	0	4	Misc
Pif1	31	26	4	3	5	HR
Pms1	26	23	3	4	6	MMR
Pol1	24	18	6	4	21	Replication
Pol2	6	3	0	2	3	Replication
Pol30	9	0	0	0	0	Replication
Pol31	13	0	0	0	2	Replication
Pol32	67	64	2	1	9	Replication
Pol4	14	9	0	1	0	NHEJ
Pph3	7	0	0	0	0	HR, Checkpoint
Pri1	9	0	0	0	2	Replication
Psf1	9	0	0	0	0	Replication
Pso2	28	19	1	2	2	Misc
Rad1	22	19	2	4	5	HR, NER, MMR
Rad10	44	42	2	1	0	HR, NER, MMR
Rad14	56	42	3	2	2	NER
Rad16	22	20	2	1	6	NER

Table 4.4: Proteins in the Chromatin Processes data set (Continued)						
Protein	**% Residues ≥.5**	**% in Disordered Domains ≥30 Residues**	**Interactions**	**# Disordered Domains**	**# Phosphorylation sites**	**Pathways (GO terms)**
Rad17	24	17	1	1	4	HR, Checkpoint
Rad18	64	53	3	4	9	Replication
Rad2	44	33	8	5	7	NER
Rad23	63	55	1	3	15	NER
Rad24	28	24	4	2	4	HR, NER, Checkpoint
Rad27	24	8	1	1	2	BER, NHEJ, Replication
Rad3	8	0	1	0	0	NER
Rad30	11	6	0	1	0	Replication
Rad34	17	13	2	1	0	NER
Rad4	27	25	3	2	0	NER
Rad5	25	18	4	4	7	Replication
Rad50	37	26	3	4	3	HR, Checkpoint
Rad51	23	19	2	1	0	HR
Rad52	55	49	6	2	6	HR
Rad53	26	16	5	3	36	Checkpoint
Rad54	18	8	2	1	2	HR
Rad55	20	0	3	0	5	HR
Rad57	17	7	0	1	1	HR
Rad59	12	0	0	0	0	HR
Rad6	37	37	1	2	1	Checkpoint, Replication
Rad7	31	25	3	1	7	NER
Rad9	52	36	11	4	45	NER, Checkpoint
Rdh54	25	14	2	2	3	HR
Rev1	11	4	3	1	1	Replication
Rfa1	12	9	1	1	2	Replication
Rfa2	28	15	1	1	7	Replication
Rfa3	23	0	0	0	1	Replication
Rfc1	31	23	3	2	7	Replication
Rfc2	6	0	1	0	0	Replication
Rfc3	7	0	0	0	0	Replication
Rfc4	7	0	0	0	0	Replication
Rmi1	30	14	0	1	1	HR, Checkpoint

Table 4.4: Proteins in the Chromatin Processes data set (Continued)						
Protein	**% Residues ≥.5**	**% in Disordered Domains ≥30 Residues**	**Interactions**	**# Disordered Domains**	**# Phosphorylation sites**	**Pathways (GO terms)**
Rrm3	33	32	5	1	1	Replication
Rsc2	41	28	4	3	15	Misc
Sae2	35	21	4	2	7	HR
Sae3	53	45	0	1	1	HR
Sgs1	47	41	10	5	2	HR, Checkpoint
Shu2	8	0	0	0	0	HR, Replication
Sir2	20	13	2	1	4	Replication
Sld2	80	62	6	5	8	HR, Replication
Sld3	47	41	5	4	60	HR, Replication
Sld5	28	14	0	1	0	Replication
Slx1	9	0	0	0	0	HR
Slx4	56	39	8	4	13	HR
Slx5	52	47	6	4	2	Misc
Slx8	76	72	0	1	4	Misc
Smc1	27	7	6	2	2	Misc
Smc5	18	9	2	1	0	Misc
Sml1	100	100	2	1	2	Misc
Snf5	57	52	7	3	3	Misc
Srs2	33	24	4	5	9	HR, NHEJ
Ssl1	21	18	1	1	2	NER, Replication
Ssl2	24	16	4	2	1	NER, Replication
Tah11	18	12	1	1	4	Replication
Tdp1	14	9	1	1	0	Misc
Tel1	1	0	0	0	2	Checkpoint
Tfb1	31	20	2	2	2	NER
Tfb2	14	8	0	1	2	NER
Tfb5	38	0	0	0	1	NER
Tof1	15	13	2	2	12	SCC, Checkpoint, Replication
Ubc13	16	0	0	0	0	Replication
Ung1	16	0	0	0	0	Misc
Xrs2	39	30	10	4	9	HR, Checkpoint

Table 4.4: Proteins in the Chromatin Processes data set (Continued)						
Protein	**% Residues ≥.5**	**% in Disordered Domains ≥30 Residues**	**Interactions**	**# Disordered Domains**	**# Phosphorylation sites**	**Pathways (GO terms)**
Yen1	39	32	3	5	4	HR
Zip1	67	56	9	7	6	HR

CHAPTER 5:

BIOINFORMATICALLY GUIDED MUTAGENESIS IN RMI1, THE NONCATALYTIC SUBUNIT OF THE *S. CEREVISIAE* SGS1/TOP3/RMI1 COMPLEX, REVEALS TWO FUNCTIONAL MOTIFS

Introduction

The RecQ-like DNA helicase family is evolutionarily conserved and necessary for genomic stability from bacteria to humans. In yeast the RecQ-like DNA helicase Sgs1 forms a complex with Top3/Rmi1 (STR) and facilitates both early and late stage DNA break repair [1]. Early in double strand break (DSB) repair, STR resects the ends of the DSB to facilitate the formation of a single-strand 3' overhang to which Rad51 binds [2-5]. This Rad51 filament is then able to initiate a genome-wide search for sequence homology (strand invasion), eventually leading to the formation of Holliday Junctions (HJs) that need to be resolved prior to cell division. Resolution can be achieved by the HJ-specific endonuclease Yen1, randomly leading to crossover and noncrossover products, or HJs can be dissolved by STR in a process involving HJ migration and decatenation of the single strands that yields noncrossover products [6]. STR has also been implicated in the reversal of strand invasion after extension of the invading 3'end by DNA synthesis to promote DSB repair by synthesis-dependent strand annealing, as well as reversal of strand invasion prior to 3'end extension (D-loop reversal) [7]. Through these functions, STR promotes noncrossover outcomes of HR and regulates

HR levels. Hence, yeast cells that lack the helicase activity of the STR complex (*sgs1Δ*) are prone to hyperrecombination, increased chromosomal instability, gross chromosomal rearrangement (GCR) formation, hypersensitivity to DNA-damaging agents and, in the case of *top3Δ* and *rmi1Δ* single mutants, very poor growth [8-11].

Despite the severe growth phenotype of the *rmi1Δ* mutant, the functional contribution of Rmi1 to the STR complex is still poorly understood. Rmi1, was first discovered in *S. cerevisiae* in a screen for components of the Sgs1/Top3 pathway [9]. Yeast cells lacking Rmi1 were found to be hypersensitive to hydroxyurea (HU) and methylmethanesulfonate (MMS), have an increased rate of spontaneous DNA damage as indicated by an increase in Rad52 foci, an increase in GCRs, and deficiency in Rad53 phosphorylation [9, 11]. Diploids lacking Rmi1 are defective in meiosis, and deletion of genes with roles in the checkpoint response to replication stress, such as Mrc1, Tof1, Csm3, lead to synthetic lethality, implying a diverse role for Rmi1 in several chromatin processes [9, 11]. Despite the severity of *rmi1Δ* phenotypes, Rmi1 has no known catalytic function. It has been shown to stimulate the catalytic functions of Sgs1/Top3, particularly the decatenation of HJs [12-14]. This function is conserved in the BLM/TopoIIIα/Rmi1/Rmi2 (BTR) complex, the human variant of STR, and studies in human cell lines imply a role for Rmi1 in TopoIIIα stability and expression [14-16].

The N-terminus of human Rmi1 has been crystallized, providing some clues to its role in catalytic enhancement and BTR complex stability [17, 18]. The N-terminus of human Rmi1 is most closely related to the *S. cerevisiae* Rmi1, is capable of binding BLM and TopoIIIα, and contains a central oligonucleotide-binding (OB) fold that is similar in structure to that of the replication protein A subunit RPA70, though it is

suggested that it is incapable of binding DNA like RPA [18, 19]. Human Rmi1 contains a disordered loop needed for dHJ dissolution enhancement of TopoIIIα [17]. Co-crystallization of the Rmi1 N-terminal lobe peptide with TopoIIIα reveals that the OB-fold of Rmi1 lies opposite of the ssDNA-binding domain of TopoIIIα, and the loop of Rmi1 physically interacts with the topoisomerase by inserting itself into the topoisomerase gate. It has been hypothesized that this loop may be what facilitates the catalytic enhancement of TopoIIIα by regulating opening and closing of the topoisomerase gate [17, 18].

In an effort to better understand the molecular basis of Rmi1 function, we have combined bioinformatics tools with an *in vivo* mutational analysis of *RMI1* function in yeast. This approach has identified short, N- and C-terminal structural motifs that are essential for Rmi1 function and are conserved in human Rmi1. We propose hypotheses for how these motifs contribute to Rmi1's role in maintaining the functional integrity of the STR complex.

Experimental Procedures

Bioinformatics analysis

The 241 residues of *S. cerevisiae* Rmi1 and the N-terminal 240 residues of the 625-residue human Rmi1 were analyzed for helical propensity, structural disorder, and amino acid sequence similarity [20-24].

Plasmids

The open reading frame of *RMI1* plus 500 bp up- and downstream was amplified by PCR from the endogenous *RMI1* locus of KHSY1338 (*ura3-52, leu2Δ1, trp1Δ63, his3Δ20,0 lys2-Bgl, hom3-10, ade2Δ1ade8, YEL069C::URA3, sgs1::HIS3*). The fragment was inserted into *Xba*I-digested pRS415 by gap-repair cloning using the non-homologous-endjoining deficient yeast strain KHSY2331 (*ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, lig4::loxP-G418-loxP*) and standard lithium-acetate transformation [25]. The integrity of *RMI1* and the promoter region in the resulting pRS415-RMI1 plasmid (pKHS 621) was verified by sequencing. Point mutations were introduced into pKHS 621 by QuikChange site-directed mutagenesis (Agilent Technologies). The list of plasmids used in this study is provided in Table 5.1.

Yeast strains

To construct yeast strains with chromosomally integrated *rmi1* mutants that are expressed as C-terminally myc-epitope-tagged proteins, pKHS 621 was linearized with *Box*I and the *HIS3*-linked myc-coding sequence from pFA6a-13MYC-HIS3MX6 [26] was inserted by gap-repair cloning. Point mutations were introduced into the resulting plasmid (pKHS630) using the QuikChange protocol (Agilent Technologies). Fragments spanning MYC-tagged *RMI1* and *rmi1* mutant alleles were amplified from pKHS 630 and its derivatives by PCR and used to replace the endogenous *RMI1* locus in the BY4711-derived yeast strain *MATa, ura3Δ0, leu2Δ0, his3Δ1, lys2Δ0, TOP3.V5.VSV.KANMX6*. Integrity of *RMI1* and the mutant alleles was confirmed by

sequencing and expression verified by Western blot analysis with c-myc monoclonal antibody (Covance).

Hydroxyurea hypersensitivity assay

Derivatives of pKHS621 were transformed into KHSY4695 (*MAT*α*, ura3Δ0, leu2Δ0, his3Δ1, lys2Δ0, rmi1::HIS3, TOP3.V5.VSV.KANMX6*), grown to OD_{600} = 0.5 in synthetic complete media lacking leucine (SC-Leu), and spotted in 10-fold dilutions on yeast extract/peptone/dextrose (YPD) and on YPD supplemented with 150 mM HU. Growth was documented after 3 to 5 days of incubation at 30°C.

Cycloheximide chase

Yeast cultures were grown overnight to OD_{600} = 1.0. Cells were synchronized by addition of 2 µg/µl alpha factor for 1 hour, followed by addition of 1 µg/µl alpha factor for an additional 1 hour. Cells were washed twice with warm YPD and resuspended in YPD to reach OD_{600} = 1.0. Cycloheximide was added to the culture at a final concentration of 50 µg/ml and the culture was incubated with vigorous shaking at 30 °C. Aliquots equivalent to 2 ODs were removed at the indicated intervals over a 24-hour time course. Whole cell extracts were prepared as previously described [27]. Briefly, washed cell pellets were resuspended in 20% trichloroacetic acid and vortexed in a cell disruptor with glass beads for 10 minutes at maximum speed. Lysate was cleared by centrifugation at 14000 rpm for 2 min. The pellet was resuspended in Laemmli buffer, the pH adjusted with 2 M Tris, pH 8.3, and boiled. Extracted proteins were separated by

SDS-PAGE and transferred to PVDF membrane before being probed for Rmi1 and rmi1 mutants using a c-myc monoclonal antibody (Covance).

Results

Determining functionally important residues in *S. cerevisiae* Rmi1 has been challenging as there, are no known catalytic domains, no crystal structure, and only minimal conservation of primary sequence (~35% between yeast genera, 18% between *S.cerevisiae* and human Rmi1) and length, ranging from 241 residues in *S. cerevisiae* to 625 in humans. We reasoned that combining structural prediction tools [23, 24, 28] in such a way that they detect 'order within disorder' could reveal functionally important motifs in Rmi1. We focused our analysis on the N- and C-terminal tails flanking the predicted OB-fold, and an apparent disordered loop emerging from the OB-fold (Figure 5.1A). We identified two regions of increased helical propensity, spanning residues 58-69 and residues 212-228, as well as two regions of lesser helical propensity between residues 126-131 and 138-145 (Figure 5.1B). We had previously determined that disruption of an α-helix was most effective when a residue with high helical propensity near the peak or in the N-terminal half of the helix was replaced with the helix breaker proline. Therefore, we constructed F63P, A128P, A139P and E220P mutations (Fig. 5.1A) to determine the importance of these regions for Rmi1 function *in vivo*. These proline substitutions led to marked decreases in helical propensity in these regions (Figure 5.1C-F). We also noted that the proline at position 88 seemed to disrupt what might otherwise be a region with high helical propensity, and hypothesized that this native break was helping to maintain a degree of flexibility in what would otherwise be a

persistent, structured region. We considered that replacing P88 with a residue with high helical propensity that was otherwise benign, such as alanine, would restore helicity to this region. Indeed, the P88A mutation is predicted to lead to an extraordinary increase in helical propensity not seen in any region of the wildtype forms of yeast or human Rmi1 (Figure 5.1G). We exploited the HU hypersensitivity of yeast cells lacking Rmi1[11] to assess the functional impact of these proline substitutions *in vivo*. We found that *rmi1Δ* cells expressing rmi1-A128P and rmi1-A139P exhibited the same HU sensitivity as the *rmi1Δ* mutant complemented with wildtype *RMI1*, whereas rmi1-P88A was able to partially suppress the HU hypersensitivity of *rmi1Δ* (Figure 5.1H). The *rmi1-F63P* mutant caused the same degree of HU hypersensitivity as a deletion of *RMI1*, indicating that it was a null allele (Figure 5.1H). We considered the possibility that the phenotype of the F63P mutation could also be due to the loss of a strong hydrophobic interaction via the aromatic residue, and decided to evaluate the importance of the chemical composition of this predicted helical region. We chose to replace F63 with a hydrophilic residue with high helical propensity, such as lysine, that would be predicted to maintain the structural integrity of the motif, but change its chemical characteristics. We found that the F63K mutation caused the same hypersensitivity to HU as the F63P mutations (Figure 5.1G), implicating that this residue maps to an α-helical structure that must conserve both its shape and hydrophobic character in order to maintain wildtype function of Rmi1. Similarly to F63, substitution of E220 with proline abolished Rmi1 function (Figure 5.1H).

Next we analyzed primary sequence alignments of *S.c.* Rmi1 to identify conserved residues and regions of conserved chemical character that could also be

indicative of a functional role. We found that the 241-residue long *S.c.* Rmi1 is ~85% identical to Rmi1 of other *Saccharomyces* species, but identity markedly decreased to ~30% when compared to yeast species outside of the genus (e.g., *K. lactis*, *C. glabrata*), and to ~18% when compared to the N-terminal 241 residues of human Rmi1. Because of this low level of sequence conservation we decided to analyze an alignment of twelve closely related Rmi1 sequences from fully sequenced *Saccharomyces* and non-*Saccharomyces* yeast species in PhylomeDB v4 [22] (Figure 5.2C, Supplemental Figure 5.1). This alignment revealed that the chemical characteristics of the α-helical region centered on residue E220 were conserved, with a short stretch of hydrophobic residues surrounded by charged residues. Whereas neither E220 nor the acidic or hydrophilic character of the residue was conserved outside of the *Saccharomyces* genus, the hydrophobic residues were, including a tyrosine at position 218 (Figure 5.2C). We hypothesized that this residue was not only part of the functional α-helical structure we had inferred from the E220P mutant, but was also a key residue for binding in an otherwise fairly charged α-helix. Indeed we found that either breaking the helix (rmi1-Y218P) or increasing its hydrophilicity (rmi1-Y218K) abolished Rmi1 function (Figure 5.2D). Although *S.c.* Rmi1 and human Rmi1 are only 18% identical, we found that they share regions of similar helical propensity, including the region that surrounds Y218 in yeast and Y201 in human Rmi1 (Figure 5.2A,B).

When we extended the computational analysis to the N-terminus of human Rmi1, we identified three regions of increased helical propensity (Figure 3B), which have been shown to form three α-helices in the crystal structure [17, 18]. Comparisons of helical propensity and primary sequences of *S.c.* Rmi1 and the N-terminus of human Rmi1

suggest a structural equivalence between the predicted sole α-helix in yeast and α3 in human Rmi1, with a potential equivalent of F63 at residue F50 in human Rmi1 (Figure 5.3A,B; Figure 5.5). The lack of helical propensity in the first 57 residues of *S.c.* Rmi1 suggests that this region does not form α-helices in the apo form as human Rmi1 does. However, we noticed two discrete regions in the yeast sequence alignment (residues S2-T16 and E28-E38) that contain hydrophobic residues in the *i*+4 pattern typical of an α-helix and are separated from each other by residues with the lowest helical propensity, proline and glycine (Figure 5.3C). To test the possibility that these two mildly hydrophobic regions could become helical upon binding to another protein, possibly Sgs1, or could be analogous to α1 and α2 in human Rmi1 we replaced L7 and Y35 with proline (Figure 5.3C). Expression of either mutant, however, was sufficient to fully restore wildtype growth to the *rmi1Δ* mutant on HU (Figure 5.3D), suggesting either that, unlike in human Rmi1, this region in *S.c.* Rmi1 does not adopt α-helical structures or that any helical structure or binding-induced folding in this region is not required for Rmi1's role in tolerating HU-induced DNA-damage.

Discussion

In this study, we have used three bioninformatic tools – disorder prediction, helical propensity prediction and phylome alignments – to elucidate structure/function relationships in the N- and C-terminal regions of *S.c.* Rmi1 that surround it's postulated, central OB-fold.

Short structured motifs are common in areas of disorder, and are often Molecular Recognition Features (MoRFs), which are short sequences of marginal order used for

protein binding that induces further disorder-to-order transition [29-31]. PONDR-VLXT [32] and IUPRED [33] predict order for residues 134-227, the majority of which likely form the OB-fold observed in the crystal structure of human Rmi1. This ordered region contains a disordered loop (residues 120-136) that seems to be equivalent to the disordered insertion loop in the OB-fold of human Rmi1 [17]. Deletion of this loop in human Rmi1 eliminates complex formation with BLM and TopoIIIα [18] whereas replacement of the equivalent loop in *S.c.* Rmi1, as estimated by sequence alignment with a scrambled version of equal chemistry, showed that Rmi1 was able to bind Sgs1 and Top3, but was unable to stimulate Top3 catalytic activity and dHJ dissolution [17]. Our proline mutagenesis (A128P, A139P) suggests that the adoption of helical structure in this disordered loop is not required for Rmi1's function in tolerating HU-induced DNA damage.

The N- and C-terminal regions flanking the OB-fold contain many disordered residues and short ordered segments. In these regions, we have identified three mutations (F63P/K, Y218P/K, E220P) that display *Δrmi1*-like defects during chronic exposure to HU, and one mutation, P88A, that exhibits an intermediate growth defect on HU. To date, only one other point mutant in *S.c.* Rmi1 has been reported, E69K [34]. At the permissive temperature of 25°C the sensitivity of this mutant to DNA-damaging agents resembles that of wildtype, but at 35°C it exhibits the defect of an *Δrmi1*. That E69K appears to be better tolerated than F63P/K can be explained by the location of E69 at the C-terminus of the predicted helical region (residues 58-69), where mutations appear to have less functional impact [27], whereas F63 maps closer to the center and, in contrast to E69, is conserved in other yeast species (Figure 5.3C).

At the C-terminus, E220P and Y218P/K mutations caused the same HU hypersensitivity as F63P/K mutations at the N-terminus, leading us to propose that residues 58-69 and 212-228 of *S.c.* Rmi1 form functionally critical α-helices that must retain a certain degree of hydrophobic character for full functionality. Based on the hydrophobicity of the α-helices we hypothesize that they function as interaction sites, with the F63 and Y218 residues part of the binding interfaces. This type of small binding motif paired with disorder has been seen in other proteins, including the yeast protein Adr1, which contains two small zinc finger domains in a disordered domain [35]; interestingly, the disordered components of this domain undergo extensive folding when contact is made between the zinc fingers and DNA. We tested if binding-induced helix formation was also a function of the extended unstructured N-terminus of yeast Rmi1, but found that introducing proline residues at positions (L7, Y35), where prospective helices might form, did not impair Rmi1 function as assessed by growth on HU.

Regarding the binding events that the F63 helix and Y218 helix may be engaged in, we present two possibilities: First, one or both of the helices may be stabilizing the OB-fold of Rmi1 as seen in other proteins containing this fold type [36]. Crystallography of human Rmi1 suggests that the three alpha helices in the N-terminus (α1-3) mediate protein stabilization via L57 and T59 in the third helix and K166 in the OB fold via hydrogen bonding [18]. If the predicted α1 (58-69) in *S.c.* Rmi1 is analogous to α3 in human Rmi1, it is possible that F63P eliminates the structural element needed for stabilizing stacking interactions, and F63K induces an electrostatic repulsion strong enough to prevent the packing between the strands of the central fold and the N-terminal α-helix.

The functional impairment of *S.c* Rmi1 by the P88A mutation is consistent with the role of the putative N-terminal α-helix in stabilizing the OB-fold. We propose that the helix-breaking property of the proline at position 88 (Figure 5.1G) contributes to the flexibility of a linker between the α-helix and the first β-sheet of the OB-fold. Replacement of this proline with alanine, which has excellent helical propensity, is predicted to make this linker helical. The resulting rigidity may weaken the interactions between the N-terminal α-helix and the first β-sheet of the OB-fold, leading to a destabilization of the OB-fold, but not as severe as that resulting from disrupting the integrity of the α-helix itself by F63P/K. Crystallography of human Rmi1 supports this role for the N-terminal α-helix and, by extension, the flexible linker, with the corresponding residues being in close proximity to the central OB-fold. F50 in α3 of human Rmi1, which we propose to be equivalent to F63 in the predicted α1 of *S.c.* Rmi1, is in close proximity to β1 and β4 of the OB-fold (Figure 5.4). Similarly, at the C-terminus, Y201, the equivalent of Y218 of *S.c.* Rmi1, appears close to the rear of the OB-fold, also in proximity to β1 (Figure 5.4) [17].

Second, one or both of the α-helices in Rmi1 may mediate physical interactions between the Rmi1/Top3 complex and Sgs1. Crystallography of the human Rmi1/TopoIIIα complex shows that the central OB-fold of Rmi1 is the primary interactor with TopoIIIα [17]. This suggests that aside from stabilizing the OB-fold, it is unlikely that the putative helices at F63 and Y218 play a role in Top3 binding. Where the RecQ helicase interacts with the Rmi1/topoisomerase complex is still unknown, but the location of the α helices at the N- and C-terminal ends of *S.c* Rmi1, their hydrophobic

faces, and their conservation in human Rmi1 may make the α-helices at F63 and Y218 candidates for Sgs1 binding.

Future Directions

In an effort to test the hypothesis that the structures destroyed by proline mutation are providing stability to the protein at large, a cycloheximide chase will be employed to compare the stability of the wildtype Rmi1 protein to the mutant proteins. Upon introduction of cycloheximide, all protein synthesis ceases, and the degradation time of a protein can be probed. To date, the wildtype protein has been evaluated via chase as described in the method section; preliminary results suggest that protein levels reach near zero around 7-8 hours following cycloheximide exposure. It is hypothesized that both F63P and Y218P are contributing to protein instability that yields the sick phenotypes seen on HU; thus, it is expected that the chase results will show that the proteins degrade much more quickly than the wildtype protein. A128P is being used as a positive control, and is expected to have a similar stability profile to wildtype Rmi1. Future directions also include P88A protein stability with the cycloheximide chase. Considering that the phenotype on HU is intermediate to the wildtype and F63P/Y218P phenotypes, it is possible that the protein will be intermediate in stability and time to complete degradation. It is also possible that P88A plays no role in stability of Rmi1 itself, but rather prevents some other sort of binding event or function that requires the flexible linker that P88 maintains. To test this possibility, we may test the ability for the mutant protein to bind its only known partners, Sgs1 and Top3. This could be accomplished via *in vitro* pulldown, as previously described [27]. In this case, it is

hypothesized that Sgs1 binding is more likely to be mediated by this flexible loop, as previous crystallography of the human orthologs puts the binding of Top3 out of reach of the region [17].

References

1. Chu, W.K. and I.D. Hickson, *RecQ helicases: multifunctional genome caretakers.* Nat Rev Cancer, 2009. **9**(9): p. 644-654.
2. Mimitou, E.P. and L.S. Symington, *Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing.* Nature, 2008. **455**(7214): p. 770-774.
3. Bennett, R.J., J.L. Keck, and J.C. Wang, *Binding specificity determines polarity of DNA unwinding by the Sgs1 protein of S. cerevisiae1.* Journal of Molecular Biology, 1999. **289**(2): p. 235-248.
4. Daley, J.M., T. Chiba, X. Xue, H. Niu, and P. Sung, *Multifaceted role of the Topo IIIα–RMI1-RMI2 complex and DNA2 in the BLM-dependent pathway of DNA break end resection.* Nucleic Acids Research, 2014. **42**(17): p. 11083-11091.
5. Zhu, Z., W.-H. Chung, E.Y. Shim, S.E. Lee, and G. Ira, *Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double strand break ends.* Cell, 2008. **134**(6): p. 981-994.
6. Mitchel, K., K. Lehner, and S. Jinks-Robertson, *Heteroduplex DNA Position Defines the Roles of the Sgs1, Srs2, and Mph1 Helicases in Promoting Distinct Recombination Outcomes.* PLoS Genet, 2013. **9**(3): p. e1003340.
7. Fasching, C.L., P. Cejka, S.C. Kowalczykowski, and W.-D. Heyer, *Top3-rmi1 dissolve rad51-mediated D loops by a topoisomerase-based mechanism.* Molecular cell, 2015. **57**(4): p. 595-606.
8. Myung, K., A. Datta, C. Chen, and R.D. Kolodner, *SGS1, the Saccharomyces cerevisiae homologue of BLM and WRN, suppresses genome instability and homeologous recombination.* Nat Genet, 2001. **27**(1): p. 113-116.
9. Mullen, J.R., F.S. Nallaseth, Y.Q. Lan, C.E. Slagle, and S.J. Brill, *Yeast Rmi1/Nce4 Controls Genome Stability as a Subunit of the Sgs1-Top3 Complex.* Molecular and Cellular Biology, 2005. **25**(11): p. 4476-4487.
10. Gangloff, S., J.P. McDonald, C. Bendixen, L. Arthur, and R. Rothstein, *The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase.* Molecular and Cellular Biology, 1994. **14**(12): p. 8391-8398.
11. Chang, M., M. Bellaoui, C. Zhang, R. Desai, P. Morozov, L. Delgado-Cruzata, R. Rothstein, G.A. Freyer, C. Boone, and G.W. Brown, *RMI1/NCE4, a suppressor of genome instability, encodes a member of the RecQ helicase/Topo III complex.* The EMBO Journal, 2005. **24**(11): p. 2024-2033.

12. Chen, C.-F. and S.J. Brill, *Binding and Activation of DNA Topoisomerase III by the Rmi1 Subunit.* Journal of Biological Chemistry, 2007. **282**(39): p. 28971-28979.
13. Cejka, P., J.L. Plank, C.Z. Bachrati, I.D. Hickson, and S.C. Kowalczykowski, *Rmi1 stimulates decatenation of double Holliday junctions during dissolution by Sgs1-Top3.* Nat Struct Mol Biol, 2010. **17**(11): p. 1377-1382.
14. Yang, J., L. O'Donnell, D. Durocher, and G.W. Brown, *RMI1 Promotes DNA Replication Fork Progression and Recovery from Replication Fork Stress.* Molecular and Cellular Biology, 2012. **32**(15): p. 3054-3064.
15. Guiraldelli, M.F., C. Eyster, and R.J. Pezza, *Genome instability and embryonic developmental defects in RMI1 deficient mice.* DNA Repair, 2013. **12**(10): p. 835-843.
16. Wu, L., C.Z. Bachrati, J. Ou, C. Xu, J. Yin, M. Chang, W. Wang, L. Li, G.W. Brown, and I.D. Hickson, *BLAP75/RMI1 promotes the BLM-dependent dissolution of homologous recombination intermediates.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(11): p. 4068-4073.
17. Bocquet, N., A.H. Bizard, W. Abdulrahman, N.B. Larsen, M. Faty, S. Cavadini, R.D. Bunker, S.C. Kowalczykowski, P. Cejka, I.D. Hickson, and N.H. Thomä, *Structural and mechanistic insight into Holliday-junction dissolution by Topoisomerase IIIα and RMI1.* Nat Struct Mol Biol, 2014. **21**(3): p. 261-268.
18. Wang, F., Y. Yang, T.R. Singh, V. Busygina, R. Guo, K. Wan, W. Wang, P. Sung, A.R. Meetei, and M. Lei, *Crystal Structures of RMI1 and RMI2, Two OB-Fold Regulatory Subunits of the BLM Complex.* Structure, 2010. **18**(9): p. 1159-1170.
19. Raynard, S., W. Bussen, and P. Sung, *A Double Holliday Junction Dissolvasome Comprising BLM, Topoisomerase IIIα, and BLAP75.* Journal of Biological Chemistry, 2006. **281**(20): p. 13861-13864.
20. Muñoz, V. and L. Serrano, *Development of the multiple sequence approximation within the AGADIR model of α-helix formation: Comparison with Zimm-Bragg and Lifson-Roig formalisms.* Biopolymers, 1997. **41**(5): p. 495-509.
21. Dosztanyi, Z., V. Csizmok, P. Tompa, and I. Simon, *The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins.* J Mol Biol, 2005. **347**(4): p. 827-39.
22. Huerta-Cepas, J., S. Capella-Gutiérrez, L.P. Pryszcz, M. Marcet-Houben, and T. Gabaldón, *PhylomeDB v4: zooming into the plurality of evolutionary histories of a genome.* Nucleic Acids Research, 2014. **42**(Database issue): p. D897-D902.
23. Lacroix, E., A.R. Viguera, and L. Serrano, *Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters.* J Mol Biol, 1998. **284**(1): p. 173-91.

24. Peng, K., S. Vucetic, P. Radivojac, C.J. Brown, A.K. Dunker, and Z. Obradovic, *Optimizing long intrinsic disorder predictors with protein evolutionary information.* J Bioinform Comput Biol, 2005. **3**(1): p. 35-60.
25. Gietz, R.D. and R. Woods, *Yeast Transformation by the LiAc/SS Carrier DNA/PEG Method*, in *Yeast Protocol*, W. Xiao, Editor. 2006, Humana Press. p. 107-120.
26. Longtine, M.S., A. McKenzie III, D.J. Demarini, N.G. Shah, A. Wach, A. Brachat, P. Philippsen, and J.R. Pringle, *Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae.* Yeast, 1998. **14**(10): p. 953-961.
27. Kennedy, J.A., G.W. Daughdrill, and K.H. Schmidt, *A transient α-helical molecular recognition element in the disordered N-terminus of the Sgs1 helicase is critical for chromosome stability and binding of Top3/Rmi1.* Nucleic Acids Research, 2013. **41**(22): p. 10215-10227.
28. Munoz, V. and L. Serrano, *Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: comparison with Zimm-Bragg and Lifson-Roig formalisms.* Biopolymers, 1997. **41**(5): p. 495-509.
29. Dyson, H.J. and P.E. Wright, *Coupling of folding and binding for unstructured proteins.* Current Opinion in Structural Biology, 2002. **12**(1): p. 54-60.
30. Fuxreiter, M., I. Simon, P. Friedrich, and P. Tompa, *Preformed Structural Elements Feature in Partner Recognition by Intrinsically Unstructured Proteins.* Journal of Molecular Biology, 2004. **338**(5): p. 1015-1026.
31. Mittag, T., L.E. Kay, and J.D. Forman-Kay, *Protein dynamics and conformational disorder in molecular recognition.* Journal of Molecular Recognition, 2010. **23**(2): p. 105-116.
32. Romero, P., Z. Obradovic, X. Li, E.C. Garner, C.J. Brown, and A.K. Dunker, *Sequence complexity of disordered protein.* Proteins: Structure, Function, and Bioinformatics, 2001. **42**(1): p. 38-48.
33. Dosztányi, Z., V. Csizmok, P. Tompa, and I. Simon, *IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content.* Bioinformatics, 2005. **21**(16): p. 3433-3434.
34. Ashton, T.M., H.W. Mankouri, A. Heidenblut, P.J. McHugh, and I.D. Hickson, *Pathways for Holliday Junction Processing during Homologous Recombination in Saccharomyces cerevisiae.* Molecular and Cellular Biology, 2011. **31**(9): p. 1921-1933.
35. Hyre, D.E. and R.E. Klevit, *A disorder-to-order transition coupled to DNA binding in the essential zinc-finger DNA-binding domain of yeast ADR11.* Journal of Molecular Biology, 1998. **279**(4): p. 929-943.
36. Theobald, D.L., R.M. Mitton-Fry, and D.S. Wuttke, *NUCLEIC ACID RECOGNITION BY OB-FOLD PROTEINS.* Annual review of biophysics and biomolecular structure, 2003. **32**: p. 115-133.

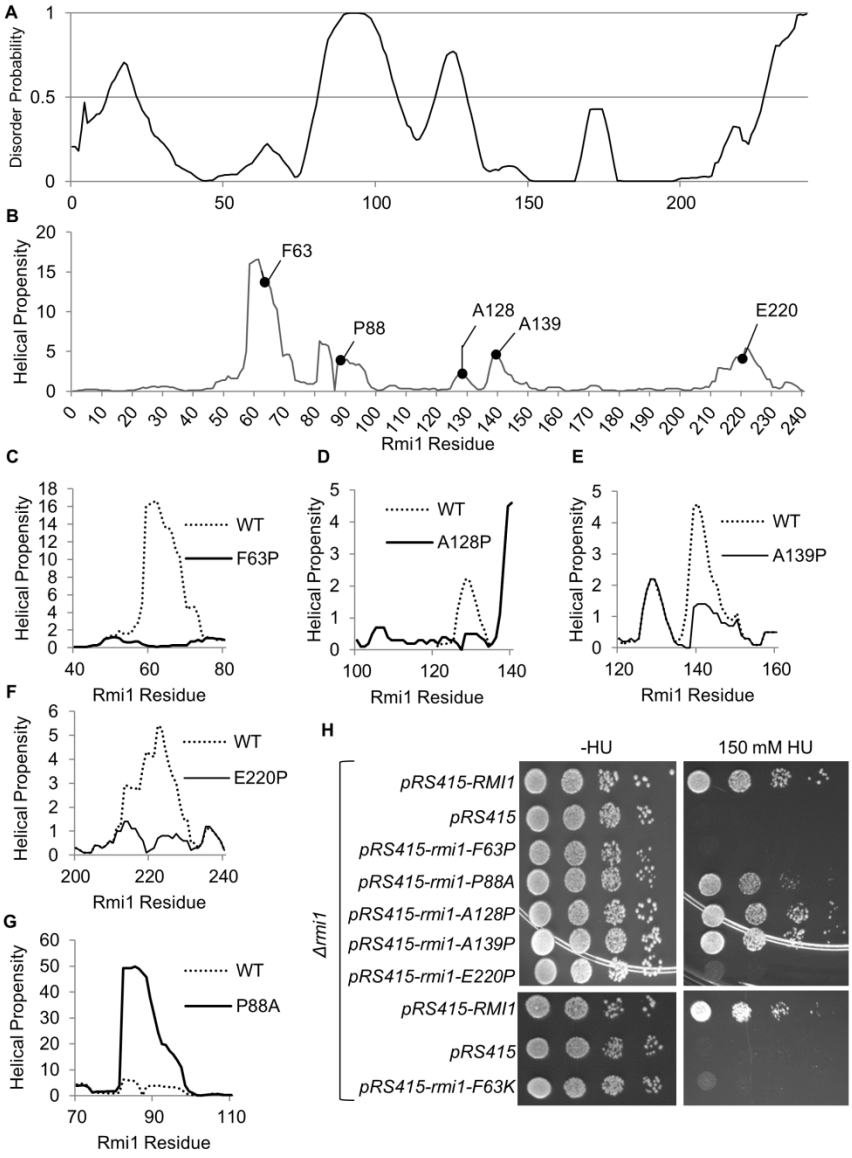


FIGURE 5.1: Sites of directed mutagenesis predicted by Agadir software. A) Residues F63, A128, A139, and E220 are residues near the peak of the helical prediction. P88 is predicted to be responsible for the sudden loss in helical prediction in the area just N-terminal of it. B-F) Wildtype and Mutation Agadir profiles show dissolution of predicted structure when proline mutants are introduced to these areas or induction of structure when proline is removed. G) Plasmids containing the ORF of RMI1 and rmi1 mutants were transformed into a Δrmi1 strain and challenged with HU. F63P and F63K are phenotypically like the empty vector negative control; E220P and P88A have intermediate phenotypes.

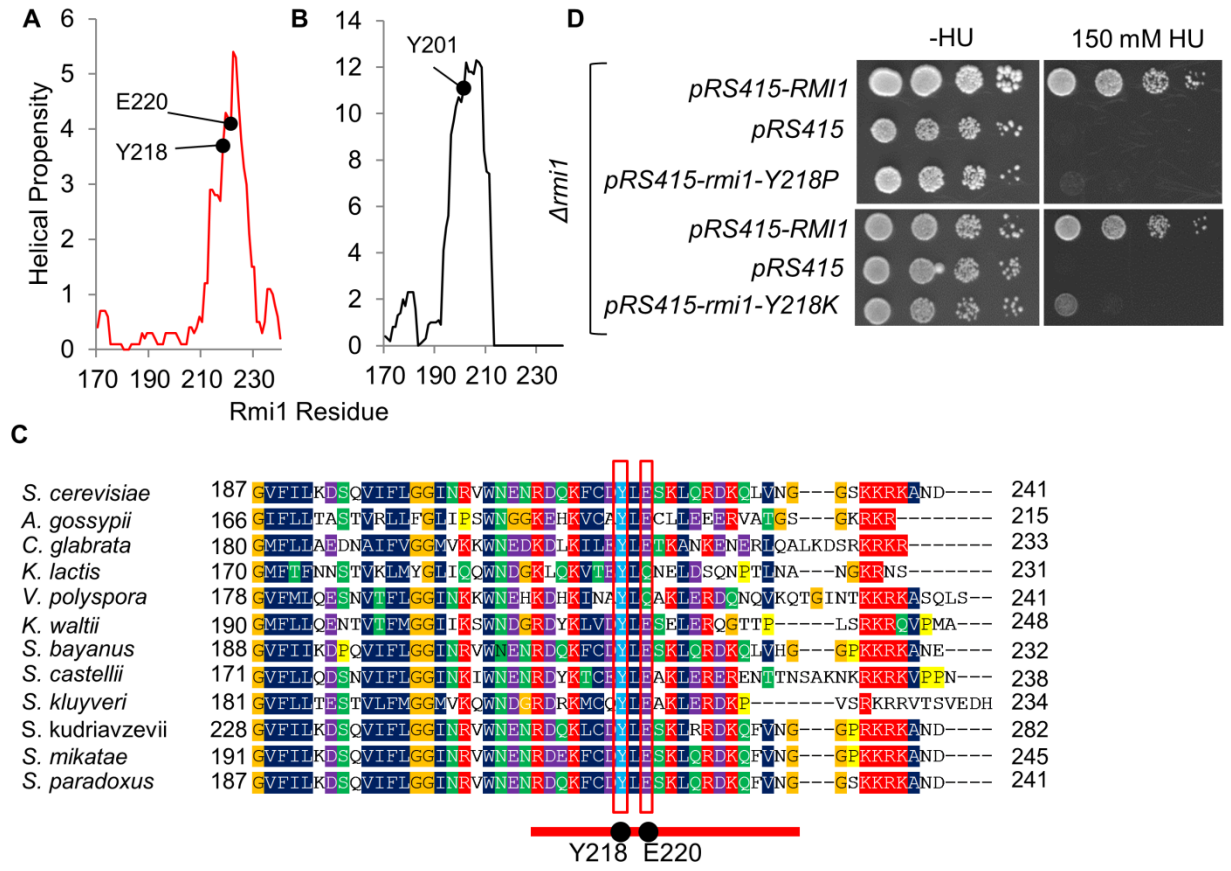


FIGURE 5.2: Yeast Rmi1 and human Rmi1-N have comparable predicted structure in the far C-terminus. A) The Agadir profile of yeast Rmi1 predicts an alpha helix at the end of the protein. B) An equivalent helix is predicted at the end of human Rmi1-N, which is the only region of the human protein with sequence similarity to the yeast ortholog. C) Phylome DB alignment of yeast species reveals a highly conserved tyrosine at position 218. This residue is in the same predicted helix as E220 and has a potential equivalent in Y201 in humans. D) Y218P and Y218K rmi1 plasmids cannot complement Δrmi1 strains on HU.

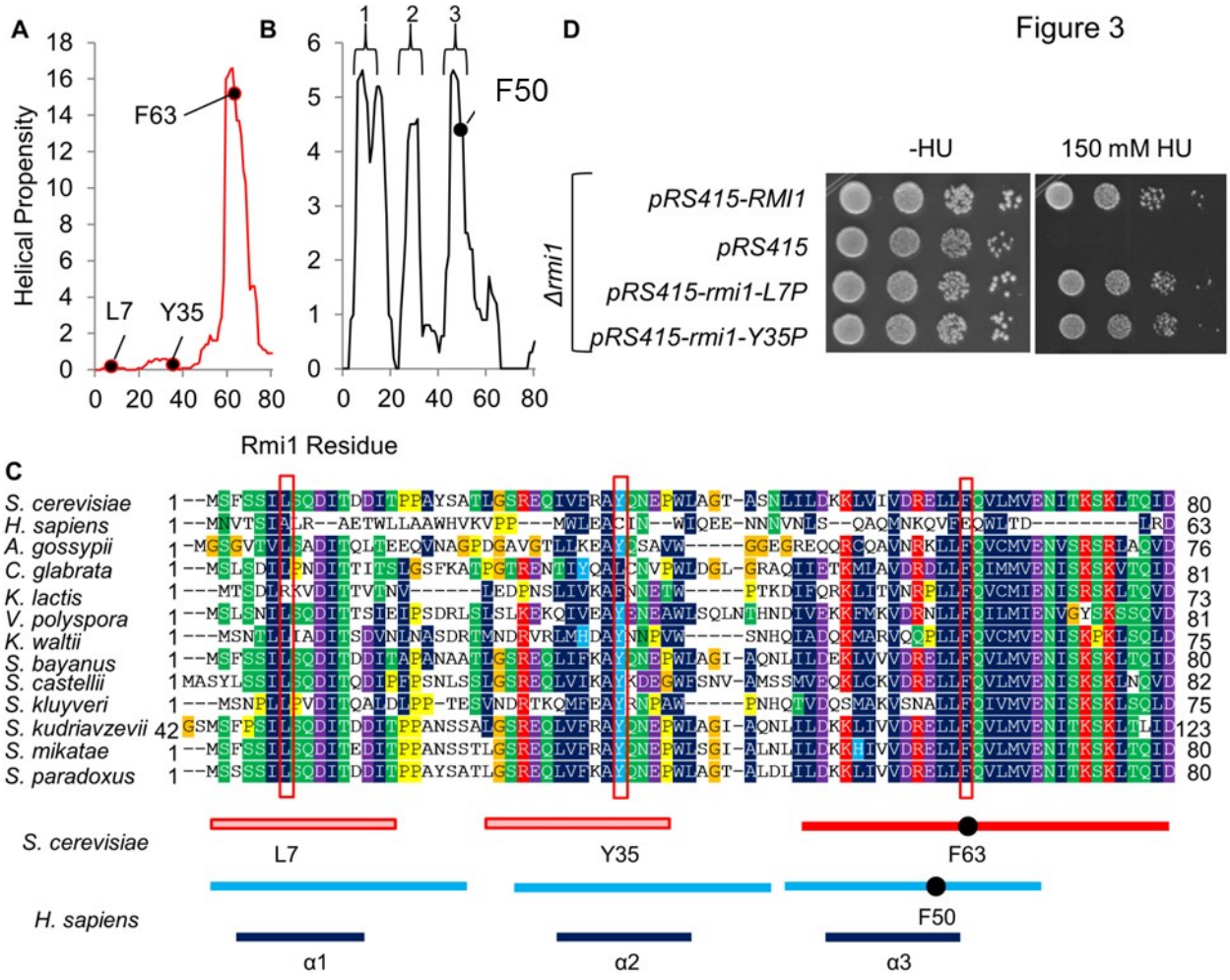


FIGURE 5.3: Yeast and human Rmi1 have potentially comparable predicted structure and "proto-structure" in the far N-terminus. A) Agadir profile of *S. cerevisiae* Rmi1 N1-80. B) Agadir profile of *H. sapiens* Rmi1 N1-80. The profile predicts three helices, as labeled. The estimated human equivalent to yeast F63, F50, is highlighted. C) Phylome DB alignment of yeast species illustrates clusters of conserved residue chemistry about L7 and Y35, which are themselves highly conserved. The predicted threshold of the yeast Rmi1 helix is plotted in the red box; the thresholds of two hypothesized proto-helices, which may assume structure upon a binding event, are plotted in the red hashed boxes. The human Agadir prediction (light blue boxes) and actual crystal structure helical thresholds [79] (dark blue boxes) are included for comparison. D) Mutagenesis in L7 and Y35 meant to prevent the induction of alpha-helical structure does not affect growth on HU.

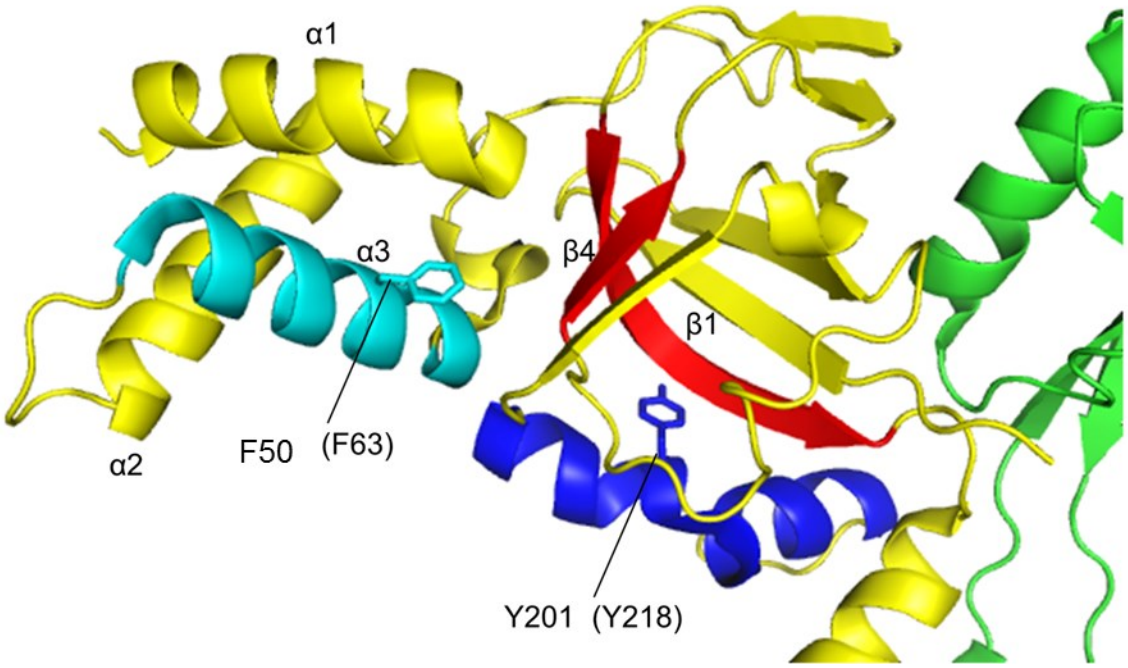


FIGURE 5.4: Structural hypotheses for proline mutant phenotypes based on disorder prediction and crystal structure. Mapping the equivalent human residues to F63 (light blue helix) and Y218 (dark blue helix) in the human Rmi1-N crystal structure (PDB database: 4CGY) [65] shows a close proximity of these residues and their accompanying structures to the central barrel of the OB fold , suggesting a role for these structures in protein stabilization. Both residues are removed in space from Top3α (green), suggesting that they are not important for direct Top3α/Rmi1 interaction.

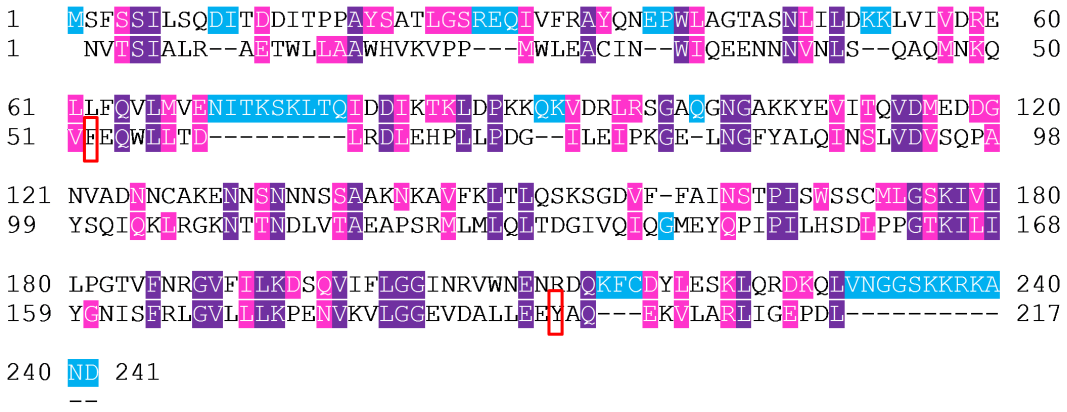


FIGURE 5.5: Alignment between S.c. Rmi1 and H.s. Rmi1N. Human equivalents to yeast F63 and Y218 are highlighted in red boxes.

Table 5.1: Plasmids used in this study

Plasmid	Description
pKHS 621	*pRS415-Rmi1*
pKHS 622	*pRS415-Rmi1-F63P*
pKHS 623	*pRS415-Rmi1-E220P*
pKHS 624	*pRS415-Rmi1-A128P*
pKHS 625	*pRS415-Rmi1-A139P*
pKHS 626	*pRS415-Rmi1-P88A*
pKHS 627	*pRS415-Rmi1-Y218P*
pKHS 628	*pRS415-Rmi1-L7P*
pKHS 629	*pRS415-Rmi1-Y35P*
pKHS 630	*pRS415-Rmi1.myc*
pKHS 631	*pRS415-Rmi1.myc- F63P*
pKHS 632	*pRS415-Rmi1.myc-A128P*
pKHS 633	*pRS415-Rmi1.myc-Y218P*

pRS415 is a gift from Dr. Steven Brill (Rutgers University)

APPENDIX A:
SGS1 TRUNCATIONS INDUCE GENOME REARRANGEMENTS BUT SUPPRESS DETRIMENTAL EFFECTS OF BLM OVEREXPRESSION IN *SACCHAROMYCES CEREVISIAE*

Note to the reader: This chapter has been previously published with permission from the publisher as Mirzaei, H, Syed, S, Kennedy, JA, and Schmidt KH (2011). "Sgs1 Truncations Induce Genome Rearrangements but Suppress Detrimental Effects of BL Overexpression in Saccharomyces cerevisiae." J Mol Biol., 405(4); 877-891.

Research was designed by K. Schmidt. Sgs1 Truncations and experiments were performed by S. Syed. BLM diploid experiments and chimera protein construction was done by H. Mirzaei. Point mutations and experiments on the zinc-binding domain were done by H. Mirzaei and J. Kennedy. Corresponding author: Kristina Schmidt, Department of Cell Biology, Microbiology and Molecular Biology, University of South Florida, 4202 E. Fowler Avenue, ISA2015, Tampa, FL 33620. Phone: (813) 974-1592. Fax: (813) 974-1614.; E-mail: kschmidt@usf.edu

Abstract

RecQ-like DNA helicases are conserved from bacteria to humans. They perform functions in the maintenance of genome stability, and their mutation is associated with cancer predisposition and premature aging syndromes in humans. Here, a series of C-terminal deletions and point mutations of Sgs1, the only RecQ-like helicase in yeast,

show that the HRDC and Rad51 interaction domain are dispensable for Sgs1's role in suppressing genome instability, whereas the zinc-binding domain and the helicase domain are required. BLM expression from the native *SGS1* promoter had no adverse effects on cell growth, but also was unable to complement any *sgs1Δ* defects. BLM overexpression, however, significantly increased the rate of accumulating GCRs in a dosage dependent manner and greatly exacerbated sensitivity to DNA-damaging agents. Co-expressing *sgs1* truncations of up to 900 residues, lacking all known functional domains of Sgs1, suppressed HU sensitivity of BLM overexpressing cells, suggesting a functional relationship between Sgs1 and BLM. Indeed, protein disorder prediction analysis of Sgs1 and BLM was used to produce a functional Sgs1-BLM chimera by replacing the N-terminus of BLM with the disordered N-terminus of Sgs1. The functionality of this chimera suggests that it is the disordered N-terminus, a site of protein binding and post-translational modification, that confers species-specificity to these two RecQ-like proteins.

Introduction

RecQ-like DNA helicases, named after the DNA repair protein RecQ of *E. coli* [1-3] are evolutionarily highly conserved. These 3'- to 5'-helicases function at the interface between DNA replication and recombination to maintain genome integrity. Sgs1 is the only known member of this helicase family in *Saccharomyces cerevisiae* [4]. Sgs1-deficient cells show increased sensitivity to the DNA-damaging agents hydroxyurea (HU) and methylmethane sulfonate (MMS), missegregate chromosomes, accumulate gross-chromosomal rearrangements (GCRs) and have a shortened lifespan [5-8]. In

contrast, five RecQ-like helicases (RecQL1, BLM, WRN, RecQL4 and RecQL5) are known in humans, and mutations in the *BLM*, *WRN* and *RECQL4* genes are associated with the rare, cancer-prone Blooms syndrome, Werner syndrome and Rothmund Thompson syndrome, respectively [9-13]. All RecQ-like helicases share a seven-motif helicase domain with Walker A and DEAH motifs. The RQC (RecQ-helicase-conserved) domain, located C-terminal to the helicase domain, is thought to be involved in DNA binding and conferring specificity of binding to DNA structures, such as G4-tetrads 14; [15-17]. The HRDC (Helicase and RNaseD C-terminal) domain is the most C-terminal of the conserved domains and resembles domains in other proteins that are involved in nucleic acid metabolism, such as RNase D and UvrD; but, like the RQC domain, it is not found in all RecQ-like helicases [18; 19]. The HRDC domain has been implicated in binding and resolving DNA structures, such as Holliday junctions, and in mediating protein-protein interactions [18; 20-23]. Two acidic regions have also been identified N-terminal of the helicase domain and may be involved in mediating protein-protein interactions [10; 24; 25]. Sgs1 is found in a complex with Top3 and Rmi1, and there is also evidence of physical interactions of the N-terminal half of Sgs1 with Top2, Srs2 and Rad16, and interactions of the C-terminus with Mlh1 and Rad51 [26; 27] [28-31].

Defects in BLM, the human RecQ-helicase considered to be most closely related to Sgs1 cause Bloom's syndrome (BS), an autosomal recessive disorder characterized by chromosome gaps and breaks, elevated sister chromatid exchange, mitotic hyperrecombination, and aberrant DNA replication events [32-34]. Affected individuals suffer from a high incidence and wide variety of cancers, infertility and dwarfism

(reviewed in reference 33). BLM catalyses ATP-dependent 3' to 5' DNA unwinding with a preference for DNA structures that may arise spontaneously during DNA replication or as a result of homologous recombination (HR) [35]. For example, by unwinding unusual secondary DNA structures, BLM may aid replication fork progression, prevent illegitimate recombination during replication and assist in restarting stalled forks [36-39]. Evidence supporting a role of BLM in maintaining genome integrity has been accumulating. For example, BLM-defective cells exhibit a retarded rate of strand elongation during DNA replication 40, accumulate abnormal replication intermediates [41] and are hypersensitive to agents that impair DNA replication [42]. BLM physically interacts with several proteins that play important roles during DNA replication and repair, such as replication protein A (RPA), the flap-endonuclease FEN-1, chromatin assembly factor CAF-1, the mismatch repair protein Mlh1, HR factor Rad51 and topoisomerase III α [43-49]. BLM peaks in S phase and it localizes to replication foci, most likely through its physical interaction with a subunit of DNA polymerase δ [50-54].

Here we have determined the role of C-terminal domains and protein interaction sites of Sgs1 in suppressing GCR accumulation by expressing point mutants and truncations of Sgs1 lacking as few as 20 and as many as 1428 residues. To investigate BLM's ability to complement *sgs1Δ* defects, such as increased genome instability and sensitivity to HU and MMS, human *BLM* cDNA was expressed under control of the native *SGS1* promoter and overexpressed from a galactose-inducible promoter, revealing that BLM could suppress *sgs1Δ* defects neither in haploid nor in diploid cells. However, using computational protein disorder prediction tools, we have designed a yeast/human chimera that consists of two nonfunctional segments of BLM and Sgs1.

The ability of this chimera to suppress all *sgs1Δ* defects that we tested suggests a functional relationship between BLM and Sgs1, which is also supported by our finding that short N-terminal fragments of Sgs1, which are devoid of all known functional domains for helicase activity and DNA binding, suppress severely detrimental effects of BLM overexpression in yeast.

Materials and Methods

Yeast Strains and Media

All strains are derived from KHSY802, a derivative of S288C. Yeast strains expressing truncations of Sgs1 helicase were constructed by homologous-recombination-mediated integration of PCR products, replacing the desired 3'-segment of *SGS1* on chromosome VIII with a myc-epitope coding sequence (from pFA6a-13Myc.His3MX655, gift from Mark Longtine, University of Washington) in frame with the *SGS1* coding sequence. Expression of all truncation alleles and the myc-epitope-tagged wildtype allele of *SGS1* was confirmed by western blot analysis. All gene replacements, insertions and truncations were performed by the standard LiAc protocol [56], using PCR products with at least 50-nucleotides on each end that matched the chromosomal target locus. To express BLM from the native *SGS1* promoter (PSGS1), a PCR fragment containing *BLM* cDNA (Open Biosystems) and a *HIS3* cassette was amplified by PCR from plasmid pKHS293 using primers that include 50-nt homology to the chromosomal *SGS1* locus. This PCR product was fused to the native chromosomal SGS1 promoter by homologous-recombination-mediated integration [56]. A PCR fragment coding for a 13Myc epitope tag was amplified from pFA6a-13Myc-kanMX6 55

and integrated in-frame at the 3'end of cDNAs or *sgs1* alleles for detection of protein expression by western blot analysis. In strain KHSY3350 and KHSY3218, galactose-inducible promoters amplified from plasmids pFA6a-kanMX6-PGAL1 or pFA6a-TRP1-PGAL155, respectively, were used to replace the native *SGS1* promoter. To construct KHSY3355, the 3'-terminal 2313 bp of BLM cDNA linked to a *HIS3* cassette were amplified by PCR from plasmid pKHS293 and used to replace the 3'-terminal 2400 bp of *SGS1* in KHSY802. The accuracy of PCR-derived *SGS1* or *BLM* integrations was confirmed by sequencing. Amino acid changes C1047F and F1056A in Sgs1 were made by sitedirected mutagenesis (QuikChange, Stratagene) of pKHS360 and integrated at the *sgs1::HIS3* locus in KHSY1338. All yeast strains used in this study are listed in Table S1 (Supplementary Information). Cells were grown in YPD consisting of 10g/l yeast extract (Fisher Scientific), 20 g/l Bacto-peptone (BD Diagnostic Systems), 2% glucose (Fisher Scientific), unless indicated otherwise. For plates, agar (BD Diagnostic Systems) was added at a concentration of 20 g/l.

Western blot analysis

To confirm expression of myc-epitope tagged *BLM* and *SGS1* alleles, cells were grown to OD600 = 0.5 in YPD and whole cell extracts were prepared from 5 ml of culture (~ 3.5 × 107 cells) by standard trichloroacetic acid (TCA, Fisher Scientific) extraction [57]. Five microliters of TCA extract were separated on 10% polyacrylamide gels, transferred to a PVDF membrane (BioRad), probed with anti-c-myc monoclonal antibody (9E10, Covance Research Products) and visualized by chemiluminescence (ECL Plus, GE Healthcare). To confirm expression of *SGS1* and *BLM* from the *GAL1* promoter the same western blot procedure was used, but cells were grown overnight in

YP ((10g/l yeast extract (Fisher Scientific), 20 g/l Bacto-peptone (BD Diagnostic Systems)) supplemented with 2% sucrose (Fisher Scientific), then diluted to OD_{600} = 0.2 either in YP supplemented with 2% sucrose (uninduced sample) or 2% galactose (induced sample) and harvested for TCA extraction when cultures reached OD_{600} = 0.5. Molecular weight marker (Broad Range) was from BioRad.

Sensitivity to DNA damaging agents HU and MMS

Cell cultures were grown in YPD to OD_{600} = 0.5 and 10-fold serial dilutions were spotted on YPD supplemented with 0.05% methyl-methanesulfonate (MMS, Sigma Aldrich) or hydroxyurea (HU, US Biologicals) at 50 mM or 100 mM, as indicated. For experiments that included strains expressing BLM or Sgs1 from the *GAL1* promoter (Figure 5), cultures were grown in YP-2% sucrose instead of YPD, and spotted on YP-1 % sucrose + 1% galactose (to induce gene expression) supplemented with 100 mM HU, or without HU as the growth control.

GCR rate measurements

Rates of accumulating gross-chromosomal rearrangements (GCRs) in YPD were determined as previously described [58]. For GCR rate measurements of yeast strains expressing BLM or Sgs1 from the *GAL1* promoter the same procedure was followed, except that media was supplemented with 2% galactose to induce gene expression. Briefly, 10 ml of YP-2% galactose were inoculated with a single colony, which had been grown on YPD agar for 3 days. After 3 days of growth in liquid media at 30° C with vigorous shaking, cells were plated on GCR plates 58 supplemented with 2% galactose

instead of 2% glucose, and 10−6 dilutions were plated on YPD to obtain the viable cell count. Colonies on GCR plates were counted after 5 days of incubation at 30° C. For GCR rate measurements in the presence of varying BLM expression levels (Table 2), 0.1% or 0.5% galactose was added to liquid YP media and to GCR plates instead of 2% galactose, and sucrose was supplemented to reach a total of 2% sugar in the media. 95% confidence intervals were calculated according to Nair [59].

Random Spore Analysis

Diploids heterozygous for the desired mutant alleles were grown overnight at 30°C in YPD, washed, transferred to 0.1% potassium acetate (Fisher Scientific) and incubated for 5 days at 30°C with vigorous shaking. Asci were incubated in the presence of zymolase (MP Biomedicals, 500 µg/ml) in 1 M sorbitol (Fisher Scientific) for 20 min at 30°C and enriched for haploid spores as previously described [60]. Spores were plated on YPD, incubated at 30°C and genotyped by spotting on synthetic drop-out media (US Biologicals) to detect the presence of *TRP1* and *HIS3* marker cassettes linked to the mutant alleles. Presence of mutant alleles linked to the kanMX6 cassette was detected by the ability of haploids to grow on YPD supplemented with 200 µg/ml G418 (Axxora LLC, San Diego, CA).

Results

Requirement of the RQC domain of Sgs1, but not the HRDC domain, for GCR suppression

Sgs1 contains a conserved DEAH helicase domain, a conserved helicase and RNaseD Cterminal (HRDC) domain, two acid regions (AR1, AR2) and a RecQ-conserved (RQC) domain composed of zinc-binding and winged-helix domains. Several protein interaction sites have also been located in the 1447-amino-acid long protein (Figure A.1A). To determine the role of these domains in the maintenance of genome stability, systematic deletions to the 3' end of the chromosomal *SGS1* gene were generated, such that truncations of the C-terminus of Sgs1, ranging from 20 to 1428 amino acids, were expressed as fusions to a myc epitope. Truncations of up to 80 amino acids were constructed to not affect any known functional domain of Sgs1 while *ΔC100* and *ΔC200* deletions partially or completely, respectively, removed the HRDC domain and *ΔC300* and *ΔC400* deletions partially or completely removed the RQC domain. The largest deletions (*ΔC700, ΔC800, ΔC900, ΔC1000* and *ΔC1100*) eliminate the entire helicase domain, including the Walker A motif (803–812 aa), with the *ΔC800-ΔC1100* deletions also affecting the part of the N-terminal half of Sgs1 that contains protein interaction sites (e.g., Rad16, residues 421–792; Top2, residues 432–724; Srs2, residues 422–722) and two acid regions (AR1, residues 321–447; AR2, residues 502–648), whereas *ΔC500* and *ΔC600* deletions partially remove the helicase domain while leaving the Walker A motif intact (Figure 1A). All truncation alleles were stably expressed from the chromosomal *SGS1* locus under control of the native *SGS1* promoter (Figure A.1B). C-terminal fusion to the myc-epitope did not adversely affect

Sgs1 function, as indicated by equal sensitivity to HU and MMS of strains expressing tagged and untagged Sgs1 (wildtype) (Figure 2A). The largest deletion, leaving intact only the 19 N-terminal amino acids of Sgs1 (*sgs1ΔC1428*), was as sensitive to HU and MMS as a complete *SGS1* deletion (*sgs1Δ*), thus behaving like a null allele (Figure A.2A). Loss of up to 200 C-terminal amino acids did not increase sensitivity to HU or MMS, whereas loss of 300 or more amino acids led to sensitivity similar to that of the *sgs1ΔC142*8 and *sgs1Δ* mutants (Figure A.2A). The construction of additional 20-amino-acid truncations extended the C-terminal region that is dispensable for HU/MMS resistance to 240 amino acids (Figure A.2B).

It was previously shown that cells lacking the DNA helicase Srs2 (*srs2Δ*) depend on functional Sgs1 for their viability [61]. To assess the ability of *sgs1* truncation alleles to support growth of the *srs2Δ* mutant, we constructed diploid strains heterozygous for the *srs2Δ* deletion and heterozygous for the *sgs1ΔC200*, *sgs1ΔC260* or *sgs1ΔC300* alleles. The meiotic products of the sporulated diploids were spread on nonselective, rich media (YPD), allowing all spores to grow (Figure A.2C). Diploids heterozygous for the *srs2Δ* deletion and the *sgs1ΔC200* truncation yielded spores that grew into colonies of the same size, suggesting that the C-terminal 200 amino acid residues of Sgs1, which harbor the HRDC domain and an interaction site with the homologous recombination factor Rad51, are not required for the viability of the *srs2Δ* mutant. In contrast, sporulation of diploids heterozygous for the *srs2Δ* deletion and *sgs1ΔC260* or *sgs1ΔC300* alleles yielded mixtures of normal-sized and small colonies. Genotyping revealed that the small colonies were *srs2Δ sgs1ΔC260* or *srs2Δsgs1ΔC300* mutants whereas the normal-sized colonies corresponded to wildtype spores or single mutants.

Thus, Sgs1 that lacks 260 or more C-terminal residues and therefore does not contain a complete RQC domain cannot support normal growth of cells lacking Srs2. When we tested the effect of the C-terminal deletions on the accumulation of GCRs, we found that the C-terminal 240 amino acids were dispensable for maintaining genome integrity, whereas deleting as little as an additional 20 amino acids (*sgs1ΔC260*) caused the GCR rate to increase to that exhibited by the null mutant without a discernable intermediate phenotype (Table A.1). Combining the *sgs1ΔC300* truncation allele with a deletion of the DNA-damage checkpoint sensor *MEC3* led to a synergistic GCR rate increase, while, as expected, combining the *sgs1ΔC200* allele with a *mec3Δ* mutation did not. Thus, these findings show that the HRDC domain and the previously reported C-terminal interaction with Rad51 are not required for Sgs1's role in preventing the accumulation of GCRs and supporting normal growth of the *srs2Δ* mutant, whereas the integrity of the RQC domain, which has been suggested to span amino acids 1075 to 1207 based on the alignment of three-dimensional structures 62, is essential.

Bloom's Syndrome Associated RQC Domain Mutations Cause Loss of Sgs1 Function *in vivo*

Of the 32 exonic base substitutions that are causative of Bloom's syndrome, thirteen are missense mutations [9; 13; 63; 64], with six of these mutations affecting conserved residues that have been shown *in vitro* to participate in zinc binding and G-tetrad DNA binding activity (Figure A.3A). Studies, however, have been limited to biochemical and biophysical analyses of mutant proteins and were hampered by the inability to purify some mutant BLM proteins [16; 17; 65]. Since the cysteine residues

are highly conserved between RecQ-like helicases, including Sgs1, we replaced the corresponding cysteine residue in Sgs1 with the BS-associated mutation (*sgs1*-C1047F). Unlike BLM with mutations in any of the three conserved cysteine residues C1036, C1063 or C1066, which degraded upon purification and could therefore not be characterized [65], the *sgs1-C1047F* mutant allele was stably expressed *in vivo* from the native *SGS1* locus (Figure A.3B). The *sgs1-C1047F* mutant showed increased HU and MMS sensitivity, which, however, did not reach the level of the *sgs1Δ* allele, and exhibited levels of GCR accumulation comparable to the *sgs1Δ* mutant, demonstrating that the C1047F mutation severely impairs Sgs1 function (Figure A.3C, Table A.1). In addition to conserved cysteine residues and immediately adjoining arginine (R1037) and aspartic acid (D1064) residues, ClustalW2 alignments showed F1056 to be the only other fully conserved amino acid residue in the zinc-binding domain of Sgs1 (Figure 3A). Although the corresponding residue in BLM (F1045) is not associated with a BS mutation, the BLMF1045A mutation has been shown to cause a severe helicase defect and ssDNA binding deficiency *in vitro* [65]. When we introduced the corresponding mutation into Sgs1 (F1056A), however, the mutant was no more sensitive to HU and MMS than wildtype cells (Figure A.3C), but instead appeared fully functional with a wildtype GCR rate (Table A.1).

Expression of human BLM cDNA from the endogenous *SGS1* promoter does not complement *Δsgs1* defects

RecQ-like DNA helicases are evolutionarily conserved from bacteria to humans. Since cells from BS patients share defects seen in *sgs1Δ* cells, including increased sensitivity to DNA-damaging agents, increased levels of aberrant genetic exchange and reduced life-span, it has been suggested that RecQ-like DNA helicases from different phyla or even kingdoms might complement each other, thus allowing the development of simple model organisms for the functional and mutational characterization of disease-associated human RecQ-like helicases, such as BLM and WRN [66]. Thus, to assess the ability of BLM to suppress genome instability in the *sgs1Δ* mutant, BLM cDNA was inserted in-frame with the start codon of *SGS1* at its chromosomal locus (*PSGS1BLM*). We reasoned that insertion at the wildtype *SGS1* locus would promote cell-cycle-dependent regulation of BLM expression and expression levels similar to those previously shown for Sgs1 [67]. Stable expression of BLM was confirmed by western blot analysis, using a yeast strain expressing myc-tagged BLM (Figure A.4A); however, all subsequent experiments were carried out with untagged BLM. Expression of a single copy of BLM (*PSGS1 BLM*) did not lead to a statistically significant difference in the GCR rate compared to the *sgs1Δ* mutant (Table A.1, Table A.2), or alleviate HU sensitivity (Figure A.4B), demonstrating that *BLM* can be successfully expressed in yeast under control of the native *SGS1* promoter without detrimental effects on cell growth, but is unable to complement the tested *sgs1Δ* defects to any extent.

Overexpression of BLM leads to increased sensitivity to DNA damaging agents and rapid accumulation of GCRs

Since a single copy of *BLM (PSGS1BLM)* did not complement *Δsgs1* defects, we examined the effect of increasing BLM expression levels on *sgs1Δ* mutant phenotypes. For this purpose, the native *SGS1* promoter was replaced with a *GAL1* promoter and galactose-dependent expression of BLM was verified by fusing *BLM* to a myc-epitope tag (Figure A.4A). Overexpression of BLM did not compensate for the lack of Sgs1 when cells were exposed to HU, but instead led to a further increase in sensitivity to HU compared to the *sgs1ΔC1428* cells or cells expressing BLM under the *SGS1* promoter (Figure A.4B). We found that maximum induction of *BLM* expression led to a 1665-fold increase in the GCR rate compared to wildtype and a 34-fold increase compared to the *sgs1Δ* mutant assayed under the same conditions (Table A.2). In contrast, overexpression of Sgs1 from the *GAL1* promoter did not lead to GCR accumulation (Table A.2). The GCR rate increase upon BLM overexpression was dependent on induction levels, with the GCR rate gradually decreasing to that of the *sgs1Δ* mutant as the galactose concentration in the media decreased (Table A.2). Thus, *sgs1Δ* defects cannot be complemented by any level of BLM expression; in fact, increasing BLM expression levels induce higher sensitivity to DNA damaging agents and significantly higher genome instability compared to the *sgs1Δ* mutant.

N-terminus of Sgs1 suppresses detrimental effects of BLM overexpression

Since Sgs1 is important for the suppression of illegitimate recombination between identical sequences, such as those found in related genes, on homologous chromosomes and sister chromatids, we tested HU sensitivity of diploid strains expressing truncated *sgs1* alleles in the presence or absence of the *SGS1* wildtype allele (Figure A.5). HU sensitivity was fully suppressed for all alleles if a single copy of wildtype *SGS1* was expressed from the other allele (Figure A.5A), demonstrating that the *sgs1* truncation alleles did not have a dominant effect. As in haploid cells, only the *sgs1ΔC200* allele complemented HU sensitivity of the *sgs1Δ* diploid completely (Figure A.5B); however, cells expressing the *sgs1ΔC300* to *sgs1ΔC900* alleles were less sensitive than diploids that expressed larger truncations or the *sgs1ΔC1428* null allele (Figure A.5B). This ability of *sgs1Δ300* to *sgs1ΔC900* truncation alleles to at least partially suppress HU sensitivity indicates that there may be N-terminal segments in Sgs1 that contribute to HU resistance.

Diploids expressing *BLM* from native *SGS1* promoters on both alleles were as sensitive to HU as diploids not expressing Sgs1, whereas diploids *over*expressing BLM from one allele or from both alleles were severely HU-sensitive, with the highest expression level lacking any growth on 100 mM HU (Figure A.5C), reflecting the severe HU sensitivity of haploid cells expressing the *PGALBLM* allele (Figure A.4B). Diploids overexpressing BLM also appeared to grow more slowly than any other diploid tested here (Figure A.5C). Remarkably, expression of a single copy of *SGS1* from its endogenous promoter (*SGS1*/*PGALBLM*) completely eliminated the severe HU sensitivity conferred by overexpression of BLM. To determine if full-length

Sgs1 was required for this suppression, we crossed the haploid strain overexpressing BLM with haploids expressing various Sgs1 truncations. We found that a single copy of the *sgs1ΔC200* allele was as sufficient as wildtype Sgs1 in suppressing HU sensitivity and slow growth of the BLM overexpressing strain, and as few as the N-terminal 547 amino acids remaining in the *sgs1ΔC900* allele were sufficient for significant suppression of HU sensitivity and slow growth caused by BLM overexpression (Figure A.5C). These findings suggest that none of the known enzymatic activities or functional and conserved domains are required for suppressing the HU sensitivity of the BLM overexpressing diploids, but that the N-terminal 547 amino acids are sufficient for suppressing the detrimental effects of BLM overexpression in a diploid. That the *sgs1ΔC1000* and *sgs1ΔC1100* alleles were clearly less effective at suppressing HU sensitivity shows that the N-terminal 447 amino acids, which contain the Top3 interaction site, are necessary but not sufficient for complementation.

Design of a functional Sgs1-BLM chimera

Sgs1 and BLM share about 21% of their amino acid residues in a pair-wise alignment of the full-length proteins (ClustalW2), with most of the identical residues in the helicase domain. In fact, the N-terminal segment of Sgs1 expressed by the *sgs1ΔC800* allele, which is able to suppress the HU sensitivity of BLM-overexpressing diploids, shares only 11% with the corresponding N-terminal segment of BLM. Devoid of conserved domains and known enzymatic activities, the N-terminus of Sgs1 has been shown to be required for physical interactions with Top3, Top2, Srs2 and Rad16 [6; 26-30]. Using IUPred, an algorithm for the prediction of intrinsically disordered proteins, we

found that the N-terminal 650 residues contain a similar distribution of ordered and intrinsically disordered segments (Figure A.6A, B). In disorder prediction algorithms, such as IUPred [68; 69], a score of > 0.5 predicts a disordered amino acid residue and a score of < 0.5 predicts an ordered residue, with 30 consecutive disordered amino acids commonly being used as a lower limit for detecting disorder in whole proteome searches [68-71]. The helicase domains of Sgs1 and BLM coincide with the predicted ordered regions in both proteins, starting at around residue 648, and are surrounded by a long N-terminal and a short C-terminal segment, which contain mostly disordered residues. In fact, using the IUPred output scores, 83% of the 648 N-terminal residues of Sgs1 (538/648) are disordered, with 70% of all 648 residues being located in segments of more than 30 consecutive disordered residues, whereas only 16% of the C-terminal 800 residues of Sgs1 are predicted to be disordered, with only a single disordered segment that is longer than 30 residues (residues 1396–1447).

Based on the IUPred prediction, BLM can also be divided into a disordered N-terminus and an ordered C-terminus (Figure A.6A, B). For BLM, 52% of the N-terminal 648 residues are predicted to be disordered but only 15% of these residues are found in stretches of more than 30 disordered residues. The difference in the pattern of disorder predicted for the N-terminal segments of Sgs1 and BLM led us to hypothesize that this region may be involved in conferring species-specificity to BLM and Sgs1 function and, thus, prevent BLM from functioning in yeast. This hypothesis is supported by the fact that the N-terminus of Sgs1 is sufficient for complementation of the HU sensitivity induced by overexpression of BLM. To test this hypothesis, we constructed a yeast-human chimera in which the N-terminal 647 residues of BLM were replaced by the N-

terminal 647 residues of Sgs1 (*sgs1ΔC800- blmΔN647*) (Figure A.6C). To express this chimera from the native *SGS1* promoter we replaced nucleotides 1941 to 4344 of the endogenous *SGS1* gene with nucleotides 1941 to 4254 of *BLM* cDNA (Figure A.6E). Remarkably, the chimera was nearly as effective as wildtype *SGS1* in conferring resistance to HU, whereas the N-terminal segment of Sgs1 by itself was ineffective (Figure A.6D). Moreover, when we combined the chimeric allele with a *mec3Δ* mutation, GCRs accumulated at a significantly lower rate than in the *mec3Δ* mutant carrying the GCR-deficient *sgs1ΔC300* or *sgs1ΔC800* alleles, albeit not at the low rate of the *mec3Δ* mutant carrying the GCR-proficient *sgs1ΔC200* allele, signifying partial functionality of the chimerical protein in the suppression of chromosomal rearrangements (Table A.1). Finally, besides Srs2, the *sgs1Δ* mutant also requires the DNA helicase Rrm3 for viability. Synthetic lethality between *sgs1Δ* and *rrm3Δ* mutations is suppressed by disrupting HR factors such as Rad51 and Rad55, suggesting that the lethality is due accumulation of aberrant HR intermediates [72-74]. To assess if the Sgs1-BLM chimera was capable of preventing the accumulation of lethal levels of aberrant recombination intermediates we constructed a diploid heterozygous for the *rrm3Δ* mutation and heterozygous for the *sgs1ΔC800-blmΔN647* allele, expressing the Sgs1-BLM chimera. Spreading of spores from this diploid on YPD, which allows all spores to grow, showed that the *rrm3Δ* mutant expressing the chimera grows normally with the diameter of double mutant colonies measuring approximately 90% of that of the single mutants (Figure A.6F). These findings indicate that the Sgs1-BLM chimera is functional and, while not capable of fully suppressing chromosomal rearrangements, prevents the

accumulation of lethal levels of aberrant recombination intermediates when Rrm3 helicase is absent.

Discussion

Yeast cells that lack Sgs1 exhibit upregulated and aberrant recombination in mitosis, increased sensitivity to DNA damaging agents, accumulation of GCRs, synthetic lethality with mutations in other DNA metabolic genes, such as the *SRS2* and *RRM3* helicase genes, and meiotic defects that lead to poor spore viability [8; 30; 61; 67; 73; 75-79]. Sgs1 contains several conserved domains (DEAD-helicase, RQC, HRDC, AR1 and AR2) and protein interaction sites (Top2, Top3, Srs2, Rad16, Rad51, Mlh1) have been identified by two-hybrid screens [27; 29; 30; 46; 80]. How the integrity of these conserved motifs and protein-protein interaction sites affects the role of Sgs1 in suppression of aberrant genome rearrangements has not been determined. The requirement of some domains and/or protein interaction sites, but not others, may shed light on the poorly understood mechanism(s) by which Sgs1 contributes to the maintenance of genome integrity in yeast. Here, we find that the C-terminal 240 amino-acid segment, which contains Rad51 and Mlh1 interaction sites as well as the conserved HRDC domain thought to be involved in DNA binding and in recognition and processing of double Holliday junctions [20; 23], is dispensable for Sgs1's role in suppressing GCRs. The integrity of the RQC domain, however, is essential for GCR suppression. That zinc-binding is crucial for Sgs1 activity, and loss of function of the Cterminal truncation allele was not due to disruption of protein structure/function because of such a large deletion, was further confirmed by the finding that the point

mutation of a conserved zinc-coordinating cysteine, which has also been observed in BS patients 64, led to loss of Sgs1's ability to suppress HU sensitivity and GCR accumulation. This loss of function was not due to degradation of the mutant protein as had been previously observed for some cysteine mutants of BLM during attempts at overexpression and purification from *E. coli*. However, we cannot exclude the possibility that the loss of function resulted from intracellular mislocalization of the mutant protein. Previously, modeling of the zinc-binding domain of BLM and instability of purified mutant proteins had indicated that hydrogen bonds between three conserved residues, Y1029 (Y1040 in Sgs1), R1037 (R1048 in Sgs1) and D1064 (D1070 in Sgs1), are required for folding of the zinc-binding domain and overall protein stability 17. Although F1056 of Sgs1 does not appear to be involved in this zinc-domain stabilization and the Sgs1-F1056A mutant protein appears stable in this study, F1056 is the only other fully conserved residue in the zinc-binding domain of RecQ-like helicases, suggesting functional significance. However, introduction of the F1056A mutation had no effect on Sgs1 function *in vivo* when we assessed HU sensitivity, consistent with a previous study 81, or GCR accumulation. That in a previous *in vitro* study 65 the corresponding BLM mutation (F1045A) had severely impaired helicase and ssDNA binding activities could either be due to differences in the importance of this residue for enzymatic activity of BLM and Sgs1 or, more likely, be due to the fact that only the helicase-core segment of BLM, lacking 769 residues of N- and C-termini, was purified. The *in vitro* function of this isolated domain could be more strongly affected by a mutation than the *in vivo* function of the full-length Sgs1 mutant protein assessed here. Although nearly half of all BLM alleles that are associated with single-amino-acid changes (7 of 17 alleles) in BS

patients are located in the RQC domain [9; 13; 63; 64] none affect F1045, consistent with our finding that mutation of this conserved residue may not be associated with significant loss of function *in vivo*.

We find that Sgs1 retains partial functionality even when it lacks the HRDC, RQC and DEAH helicase domains, as demonstrated by the greater HU resistance of diploids that only express the N-terminal 547 amino acid residues compared to those alleles expressing fewer than 447 residues of Sgs1. One explanation for this finding could be that protein-protein interactions conferred by the N-terminus could contribute to the structural stability of multi-protein complexes, such as the Sgs1/Top3/Rmi1 31 complex or, even more relevant to HU resistance, DNA-damage-specific complexes with Srs2 and Mre11 [29]. In these multi-protein complexes, enzymatic activity of Sgs1 may be dispensable. Indeed, *sgs1* alleles with point mutations in the helicase domain have been shown to be capable of performing some functions of the wildtype allele, including those carried out during meiosis and checkpoint activation [79; 82].

In contrast to two previous reports [66; 83], which both used the same yeast strain that constitutively expressed *BLM* from a *GAPDH* promoter and showed partial suppression of some *sgs1Δ* defects, including HU sensitivity, we found that neither *BLM* expression under control of the natural *SGS1* promoter nor varying levels of *BLM* expression under control of a galactose-inducible promoter had any positive effect on the *sgs1Δ* mutant. That a single copy of BLM, when expressed under control of the native *SGS1* promoter, cannot alleviate *sgs1Δ* defects initially suggested to us that BLM had no functionality in yeast. In fact, the strong increase in genome instability, accompanied by severe HU sensitivity and some growth retardation upon

overexpression of BLM, indicated that BLM expression is detrimental to yeast cells. The absence of any GCR accumulation upon Sgs1 overexpression suggests that increased accumulation of GCRs in BLM overexpressing cells is not simply due to increased unwinding. Rather, we propose that BLM may possess helicase activity in yeast, leading to increased unwinding upon overexpression, but fails to elicit proper downstream responses, for example due to lack of proper N-terminal protein-protein interactions, which ultimately leads to an overabundance of aberrantly repaired lesions. That endogenous levels of N-terminal segments of Sgs1 as short as 547 residues suppressed the slow growth phenotype and the severe HU sensitivity of BLM-overexpressing cells argues in favor of a functional relationship between Sgs1 and BLM. For example, co-expression of Sgs1 and BLM could alleviate HU sensitivity in BLM overexpressing cells by acting as a bridge between BLM and Top3 (and/or other protein complexes interacting with the Sgs1 N-terminus), thereby linking enzymatic activity to appropriate upstream and downstream events. Remarkably, even relatively short N-terminal fragments of Sgs1 are sufficient for the suppression of the increased HU sensitivity of BLM-overexpressing cells, further supporting the importance of the Sgs1 N-terminus with its role in mediating interaction with other DNA metabolic factors. HU resistance comparable to wildtype cells and significantly reduced GCR accumulation of cells expressing a chimeric fusion of the Sgs1 N-terminus, which is devoid of enzymatic function and dispensable for helicase activity and ssDNA binding *in vitro*, and the BLM C-terminus, which contains helicase/RQC and HRDC domains, is consistent with helicase activity of BLM in yeast and a biologically significant, functional

interaction between BLM and Sgs1. That not only fusion of the Sgs1 and BLM segments provides HU resistance, but also co-expression of BLM and Sgs1 polypeptides from separate alleles in the same cell may indicate that the N-terminus of Sgs1 can physically interact with BLM. Our findings also suggest that it is the inability of the N-terminus of BLM to interact with or be modified by yeast proteins that leads to the inability of BLM to function in yeast. A previous report that BLM expression in yeast alleviates several *sgs1Δ* phenotypes, including partial suppression of HU sensitivity [66;83], could be explained by the fact that in the earlier study BLM was expressed from a *GAPDH* promoter, whereas here it was expressed either from the native *SGS1* promoter or from a galactose-inducible promoter. However, in light of the findings presented here, there could also be an alternative explanation. Since the *GAPDH*-promoter-*BLM* construct appears to have been inserted into the middle of the wildtype *SGS1* gene, an N-terminal segment of Sgs1 could have been expressed from the native *SGS1* promoter in addition to BLM being expressed from the *GAPDH* promoter. As shown here for haploids expressing the chimera and for diploids coexpressing the N-terminus of Sgs1 and full length BLM, such co-expression of an Sgs1-N-terminal segment from the native *SGS1* promoter and *BLM* from the *GAPDH* promoter could be the an explanation for the reported increase in HU resistance of BLM-expressing cells compared to *sgs1Δ* cells.

Of the five human RecQ-like DNA helicases, BLM is considered to be most closely related to Sgs1. Even though we show here that BLM cannot suppress any defects of the *sgs1Δ* mutant, the functional chimera does provide evidence for a functional relationship between the two RecQ-like helicases and provides a model

system for the further characterization of BLM functional domains in yeast. In fact, all BS-associated missense mutations and numerous polymorphisms are located within the 770-residue C-terminal fragment of BLM that is part of the chimera, so that they are now accessible to further functional and mutational characterization in yeast. The *in vivo* functionality of the Sgs1-BLM chimera also demonstrates the remarkable utility of protein disorder prediction as a tool for the construction of functional mutants. It will be interesting to see whether domains of any of the other human RecQ-like helicases will, like BLM, be able to form functional chimeras with the Sgs1 N-terminus.

References

1. Hegde SP, Qin MH, Li XH, Atkinson MA, Clark AJ, Rajagopalan M, Madiraju MV. *Interactions of protein with RecO, RecR, and single-stranded DNA binding proteins reveal roles for the RecF-RecO-RecR complex in DNA repair and recombination*. Proc Natl Acad Sci U S A. 1996; 93:14468–73.
2. Morimatsu K, Kowalczykowski SC. *RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair*. Mol Cell. 2003; 11:1337–47.
3. Ivancic-Bace I, Salaj-Smic E, Brcic-Kostic K. *Effects of recJ, recQ, and recFOR mutations on recombination in nuclease-deficient recB recD double mutants of Escherichia coli*. J Bacteriol. 2005; 187:1350–6.
4. Gangloff S, McDonald JP, Bendixen C, Arthur L, Rothstein R. *The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase*. Mol Cell Biol. 1994; 14:8391–8.
5. Fricke WM, Brill SJ. *Slx1-Slx4 is a second structure-specific endonuclease functionally redundant with Sgs1-Top3*. Genes Dev. 2003; 17:1768–78.
6. Mullen JR, Kaliraman V, Brill SJ. *Bipartite structure of the SGS1 DNA helicase in Saccharomyces cerevisiae*. Genetics. 2000; 154:1101–14.
7. Sinclair DA, Mills K, Guarente L. *Accelerated aging and nucleolar fragmentation in yeast sgs1 mutants*. Science. 1997; 277:1313–6.
8. Schmidt KH, Wu J, Kolodner RD. *Control of Translocations between Highly Diverged Genes by Sgs1, the Saccharomyces cerevisiae Homolog of the ome Protein*. Mol Cell Biol. 2006; 26:5406–20.
9. Ellis NA, Groden J, Ye TZ, Straughen J, Lennon DJ, Ciocci S, Proytcheva M, German J. *The Bloom's syndrome gene product is homologous to RecQ helicases*. Cell. 1995; 83:655–66.

10. Kitao S, Ohsugi I, Ichikawa K, Goto M, Furuichi Y, Shimamoto A. *Cloning of two new human helicase genes of the RecQ family: biological significance of multiple species in higher eukaryotes.* Genomics. 1998; 54:443–52.
11. Kitao S, Shimamoto A, Goto M, Miller RW, Smithson WA, Lindor NM, Furuichi Y. *Mutations in RECQL4 cause a subset of cases of Rothmund-Thomson syndrome.* Nat Genet. 1999; 22:82–4.
12. Yu CE, Oshima J, Fu YH, Wijsman EM, Hisama F, Alisch R, Matthews S, Nakura J, Miki T, Ouais S, Martin GM, Mulligan J, Schellenberg GD. *Positional cloning of the Werner's syndrome gene*. Science. 1996; 272:258–62.
13. German J, Sanz MM, Ciocci S, Ye TZ, Ellis NA. *Syndrome-causing mutations of the BLM gene in persons in the Bloom's syndrome registry*. Hum Mutat. 2007
14. Lee JW, Kusumoto R, Doherty KM, Lin GX, Zeng W, Cheng WH, von Kobbe C, Brosh RM Jr, Hu JS, Bohr VA. *Modulation of Werner syndrome protein function by a single mutation in the conserved RecQ domain*. J Biol Chem. 2005; 280:39627–36.
15. von Kobbe C, Thoma NH, Czyzewski BK, Pavletich NP, Bohr VA. *Werner syndrome protein contains three structure-specific DNA binding domains*. J Biol Chem. 2003; 278:52997–3006.
16. Huber MD, Duquette ML, Shiels JC, Maizels N. *A conserved G4 DNA binding domain in RecQ family helicases*. J Mol Biol. 2006; 358:1071–80.
17. Guo RB, Rigolet P, Zargarian L, Fermandjian S, Xi XG. *Structural and functional characterizations reveal the importance of a zinc binding domain in Bloom's syndrome helicase*. Nucleic Acids Res. 2005; 33:3109–24.
18. Kitano K, Yoshihara N, Hakoshima T. *Crystal structure of the HRDC domain of human Werner syndrome protein, WRN*. J Biol Chem. 2007; 282:2717–28.
19. Morozov V, Mushegian AR, Koonin EV, Bork P. *A putative nucleic acid-binding domain in Bloom's and Werner's syndrome helicases*. Trends Biochem Sci. 1997; 22:417–8.
20. Wu L, Chan KL, Ralf C, Bernstein DA, Garcia PL, Bohr VA, Vindigni A, Janscak P, Keck JL, Hickson ID. *The HRDC domain of BLM is required for the dissolution of double Holliday junctions*. Embo J. 2005; 24:2679–87.
21. Killoran MP, Keck JL. *Structure and function of the regulatory C-terminal HRDC domain from Deinococcus radiodurans* RecQ. Nucleic Acids Res. 2008; 36:3139–49.
22. Bernstein DA, Keck JL. *Conferring substrate specificity to DNA helicases: role of the RecQ HRDC domain*. Structure. 2005; 13:1173–82.
23. Liu Z, Macias MJ, Bottomley MJ, Stier G, Linge JP, Nilges M, Bork P, Sattler M. *The three-dimensional structure of the HRDC domain and implications for the Werner and Bloom syndrome proteins. Structure*. 1999; 7:1557–66.
24. Miyajima A, Seki M, Onoda F, Ui A, Satoh Y, Ohno Y, Enomoto T. *Different domains of Sgs1 are required for mitotic and meiotic functions*. Genes Genet Syst. 2000; 75:319–26.

25. Bernstein KA, Shor E, Sunjevaric I, Fumasoni M, Burgess RC, Foiani M, Branzei D, Rothstein R. *Sgs1 function in the repair of DNA replication intermediates is separable from its role in homologous recombinational repair*. Embo J. 2009; 28:915–25.
26. Bennett RJ, Noirot-Gros MF, Wang JC. *Interaction between yeast sgs1 helicase and DNA topoisomerase III*. J Biol Chem. 2000; 275:26898–905.
27. Duno M, Thomsen B, Westergaard O, Krejci L, Bendixen C. *Genetic analysis of the Saccharomyces cerevisiae Sgs1 helicase defines an essential function for the Sgs1-Top3 complex in the absence of SRS2 or TOP1*. Mol Gen Genet. 2000; 264:89–97.
28. Fricke WM, Kaliraman V, Brill SJ. *Mapping the DNA topoisomerase III binding domain of the Sgs1 DNA helicase*. J Biol Chem. 2001; 276:8848–55.
29. Chiolo I, Carotenuto W, Maffioletti G, Petrini JH, Foiani M, Liberi G. *Srs2 and Sgs1 DNA helicases associate with Mre11 in different subcomplexes following checkpoint activation and CDK1-mediated Srs2 phosphorylation*. Mol Cell Biol. 2005; 25:5738–51.
30. Watt PM, Louis EJ, Borts RH, Hickson ID. *Sgs1: a eukaryotic homolog of E. coli RecQ that interacts with topoisomerase II in vivo and is required for faithful chromosome segregation*. Cell. 1995; 81:253–60.
31. Chang M, Bellaoui M, Zhang C, Desai R, Morozov P, Delgado-Cruzata L, Rothstein R, Freyer GA, Boone C, Brown GW. *RMI1/NCE4, a suppressor of genome instability, encodes a member of the RecQ helicase/Topo III complex*. Embo J. 2005; 24:2024–33.
32. Chaganti RS, Schonberg S, German J. *A manyfold increase in sister chromatid exchanges in Bloom's syndrome lymphocytes*. Proc Natl Acad Sci U S A. 1974; 71:4508–12.
33. Bachrati CZ, Hickson ID. *RecQ helicases: suppressors of tumorigenesis and premature aging*. Biochem J. 2003; 374:577–606.
34. Hojo ET, van Diemen PC, Darroudi F, Natarajan AT. *Spontaneous chromosomal aberrations in Fanconi anaemia, ataxia telangiectasia fibroblast and Bloom's syndrome lymphoblastoid cell lines as detected by conventional cytogenetic analysis and fluorescence in situ hybridisation (FISH) technique*. Mutat Res. 1995; 334:59–69.
35. Mohaghegh P, Karow JK, Brosh RM Jr, Bohr VA, Hickson ID. *The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases.* Nucleic Acids Res. 2001; 29:2843–9.
36. Wu L, Hickson ID. *DNA helicases required for homologous recombination and repair of damaged replication forks*. Annu Rev Genet. 2006; 40:279–306.
37. Ralf C, Hickson ID, Wu L. *The Bloom's syndrome helicase can promote the regression of a model replication fork*. J Biol Chem. 2006; 281:22839–46.
38. Hanada K, Hickson ID. *Molecular genetics of RecQ helicase disorders*. Cell Mol Life Sci. 2007; 64:2306–22.
39. Bachrati CZ, Hickson ID. *RecQ helicases: guardian angels of the DNA replication fork*. Chromosoma. 2008; 117:219–33.
40. Hand R, German J. *A retarded rate of DNA chain growth in Bloom's syndrome*. Proc Natl Acad Sci U S A. 1975; 72:758–62.

41. Lonn U, Lonn S, Nylen U, Winblad G, German J. *An abnormal profile of DNA replication intermediates in Bloom's syndrome*. Cancer Res. 1990; 50:3141–5.
42. Davies SL, North PS, Dart A, Lakin ND, Hickson ID. *Phosphorylation of the Bloom's syndrome helicase and its role in recovery from S-phase arrest*. Molecular and Cellular Biology. 2004; 24:1279–1291.
43. Jiao RJ, Bachrati CZ, Pedrazzi G, Kuster P, Petkovic M, Li JL, Egli D, Hickson ID, Stagljar I. *Physical and functional interaction between the Bloom's syndrome gene product and the largest subunit of chromatin assembly factor 1*. Molecular and Cellular Biology. 2004; 24:4710–4719.
44. Johnson FB, Lombard DB, Neff NF, Mastrangelo MA, Dewolf W, Ellis NA, Marciniak RA, Yin Y, Jaenisch R, Guarente L. *Association of the Bloom syndrome protein with topoisomerase III alpha in somatic and meiotic cells*. Cancer Res. 2000; 60:1162–7.
45. Wu L, Bachrati CZ, Ou J, Xu C, Yin J, Chang M, Wang W, Li L, Brown GW, Hickson ID. *BLAP75/RMI1 promotes the BLM-dependent dissolution of homologous recombination intermediates*. Proc Natl Acad Sci U S A. 2006; 103:4068–73.
46. Wu L, Davies SL, Levitt NC, Hickson ID. *Potential role for the BLM helicase in recombinational repair via a conserved interaction with RAD51*. J Biol Chem. 2001; 276:19375–81.
47. Brosh RM Jr, Li JL, Kenny MK, Karow JK, Cooper MP, Kureekattil RP, Hickson ID, Bohr VA. *Replication protein A physically interacts with the Bloom's syndrome protein and stimulates its helicase activity*. J Biol Chem. 2000; 275:23500–8.
48. Sharma S, Sommers JA, Wu L, Bohr VA, Hickson ID, Brosh RM Jr. *Stimulation of flap endonuclease-1 by the Bloom's syndrome protein*. J Biol Chem. 2004; 279:9847–56.
49. Langland G, Kordich J, Creaney J, Goss KH, Lillard-Wetherell K, Bebenek K, Kunkel TA, Groden J. *The Bloom's syndrome protein (BLM) interacts with MLH1 but is not required for DNA mismatch repair.* J Biol Chem. 2001; 276:30031–5.
50. Selak N, Bachrati CZ, Shevelev I, Dietschy T, van Loon B, Jacob A, Hubscher U, Hoheisel JD, Hickson ID, Stagljar I. *The Bloom's syndrome helicase (BLM) interacts physically and functionally with p12, the smallest subunit of human DNA polymerase delta*. Nucleic Acids Res. 2008; 36:5166–79.
51. Bischof O, Kim SH, Irving J, Beresten S, Ellis NA, Campisi J. *Regulation and localization of the Bloom syndrome protein in response to DNA damage*. J Cell Biol. 2001; 153:367–80.
52. Dutertre S, Ababou M, Onclercq R, Delic J, Chatton B, Jaulin C, Amor-Gueret M. *Cell cycle regulation of the endogenous wild type Bloom's syndrome DNA helicase*. Oncogene. 2000; 19:2731–8.
53. Sanz MM, Proytcheva M, Ellis NA, Holloman WK, German J. *BLM, the Bloom's syndrome protein, varies during the cell cycle in its amount, distribution, and co-localization with other nuclear proteins*. Cytogenet Cell Genet. 2000; 91:217–23.

54. Yankiwski V, Marciniak RA, Guarente L, Neff NF. *Nuclear structure in normal and Bloom syndrome cells*. Proc Natl Acad Sci U S A. 2000; 97:5214–9.
55. Longtine MS, McKenzie A 3rd, Demarini DJ, Shah NG, Wach A, Brachat A, Philippsen P, Pringle JR. *Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae*. Yeast. 1998; 14:953–61.
56. Gietz RD, Woods RA. *Yeast transformation by the LiAc/SS Carrier DNA/PEG method*. Methods Mol Biol. 2006; 313:107–20.
57. Foiani, M.; Liberi, G.; Piatti, S.; Plevani, P. *Saccharomyces cerevisiae as a model system to study DNA replication. In: Cotterill, S., editor. Eukaryotic DNA Replication. A Practical Approach*. Oxford University Press; Oxford, UK: 1999. p. 185-200.
58. Schmidt KH, Pennaneach V, Putnam CD, Kolodner RD. *Analysis of gross-chromosomal rearrangements in Saccharomyces cerevisiae.* Methods Enzymol. 2006; 409:462–76.
59. Nair KR. *Table of confidence intervals for the median in samples from any continuous population*. Sankhya. 1940; 4:551–558.
60. Rockmill B, Lambie EJ, Roeder GS. *Spore enrichment*. Methods Enzymol. 1991; 194:146–9.
61. Lee SK, Johnson RE, Yu SL, Prakash L, Prakash S. *Requirement of yeast SGS1 and SRS2 genes for replication and transcription*. Science. 1999; 286:2339–42.
62. Kitano K, Kim SY, Hakoshima T. *Structural basis for DNA strand separation by the unconventional winged-helix domain of RecQ helicase WRN.* Structure. 18:177– 87.
63. Barakat A, Ababou M, Onclercq R, Dutertre S, Chadli E, Hda N, Benslimane A, Amor-Gueret M. *Identification of a novel BLM missense mutation (2706T>C) in a Moroccan patient with Bloom's syndrome*. Hum Mutat. 2000; 15:584–5.
64. Foucault F, Vaury C, Barakat A, Thibout D, Planchon P, Jaulin C, Praz F, Amor-Gueret M. *Characterization of a new BLM mutation associated with a topoisomerase II alpha defect in a patient with Bloom's syndrome*. Hum Mol Genet. 1997; 6:1427–34.
65. Janscak P, Garcia PL, Hamburger F, Makuta Y, Shiraishi K, Imai Y, Ikeda H, Bickle TA. *Characterization and mutational analysis of the RecQ core of the bloom syndrome protein*. J Mol Biol. 2003; 330:29–42.
66. Yamagata K, Kato J, Shimamoto A, Goto M, Furuichi Y, Ikeda H. *Bloom's and Werner's syndrome genes suppress hyperrecombination in yeast sgs1 mutant: implication for genomic instability in human diseases*. Proc Natl Acad Sci U S A. 1998; 95:8733–8.
67. Frei C, Gasser SM. *The yeast Sgs1p helicase acts upstream of Rad53p in the DNA replication checkpoint and colocalizes with Rad53p in S-phase-specific foci.* Genes Dev. 2000; 14:81–96.
68. Dosztanyi Z, Csizmok V, Tompa P, Simon I. *The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins*. J Mol Biol. 2005; 347:827–39.

69. Dosztanyi Z, Csizmok V, Tompa P, Simon I. *IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content*. Bioinformatics. 2005; 21:3433–4.
70. Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z. *Optimizing long intrinsic disorder predictors with protein evolutionary information*. J Bioinform Comput Biol. 2005; 3:35– 60.
71. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. *Prediction and functional analysis of native disorder in proteins from the three kingdoms of life*. J Mol Biol. 2004; 337:635–45.
72. Ooi SL, Shoemaker DD, Boeke JD. *DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray*. Nat Genet. 2003; 35:277–86.
73. Schmidt KH, Kolodner RD. *Requirement of Rrm3 helicase for repair of spontaneous DNA lesions in cells lacking Srs2 or Sgs1 helicase*. Mol Cell Biol. 2004; 24:3213–26.
74. Torres JZ, Schnakenberg SL, Zakian VA. *Saccharomyces cerevisiae Rrm3p DNA helicase promotes genome integrity by preventing replication fork stalling: viability of rrm3 cells requires the intra-S–phase checkpoint and fork restart activities*. Mol Cell Biol. 2004; 24:3198–212.
75. Cobb JA, Bjergbaek L, Gasser SM. *RecQ helicases: at the heart of genetic stability*. FEBS Lett. 2002; 529:43–8.
76. Ira G, Malkova A, Liberi G, Foiani M, Haber JE. *Srs2 and Sgs1–Top3 suppress crossovers during double-strand break repair in yeast*. Cell. 2003; 115:401–11.
77. Myung K, Datta A, Chen C, Kolodner RD. *SGS1, the Saccharomyces cerevisiae homologue of BLM and WRN, suppresses genome instability and homeologous recombination*. Nat Genet. 2001; 27:113–6.
78. Versini G, Comet I, Wu M, Hoopes L, Schwob E, Pasero P. *The yeast Sgs1 helicase is differentially required for genomic and ribosomal DNA replication*. Embo J. 2003; 22:1939–49.
79. Miyajima A, Seki M, Onoda F, Shiratori M, Odagiri N, Ohta K, Kikuchi Y, Ohno Y, Enomoto T. *Sgs1 helicase activity is required for mitotic but apparently not for meiotic functions*. Mol Cell Biol. 2000; 20:6399–409.
80. Saffi J, Feldmann H, Winnacker EL, Henriques JA. *Interaction of the yeast Pso5/Rad16 and Sgs1 proteins: influences on DNA repair and aging.* Mutat Res. 2001; 486:195–206.
81. Ui A, Satoh Y, Onoda F, Miyajima A, Seki M, Enomoto T. *The N-terminal region of Sgs1, which interacts with Top3, is required for complementation of MMS sensitivity and suppression of hyper-recombination in sgs1 disruptants.* Mol Genet Genomics. 2001; 265:837–50.
82. Bjergbaek L, Cobb JA, Tsai-Pflugfelder M, Gasser SM. *Mechanistically distinct roles for Sgs1p in checkpoint activation and replication fork maintenance*. Embo J. 2005; 24:405–17.
83. Heo SJ, Tatebayashi K, Ohsugi I, Shimamoto A, Furuichi Y, Ikeda H. *Bloom's syndrome gene suppresses premature ageing caused by Sgs1 deficiency in yeast*. Genes Cells. 1999; 4:619–25.

84. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. *Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 2003*; 31:3497–500.

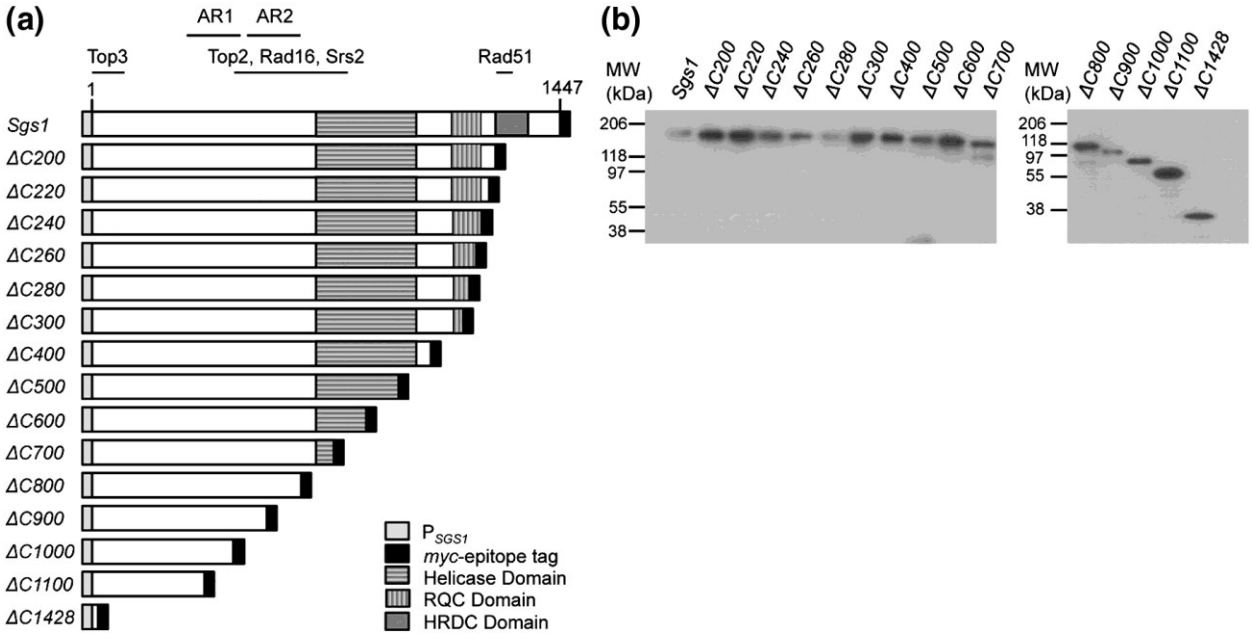


Figure A.1. C-terminal truncations of Sgs1 used in this study. (A) Full-length Sgs1 contains a DEAH helicase domain, an RQC domain and an HRDC domain in its C-terminal half and acidic regions AR1 and AR2 in its N-terminal half; interaction sites with Top3, Top2, Srs2, Rad51 and Rad16 are indicated. C-terminal truncations ranging in size from 200 residues to 1428 residues were constructed by fusion to a myc-epitope tag. All truncations were introduced at the endogenous *SGS1* locus on chromosome VIII. **(B)** Expression of wildtype Sgs1 and truncation alleles from the endogenous SGS1 promoter (*PSGS1*) was confirmed by western blotting, using a myc-antibody. Molecular weights (MW) are indicated on the left.

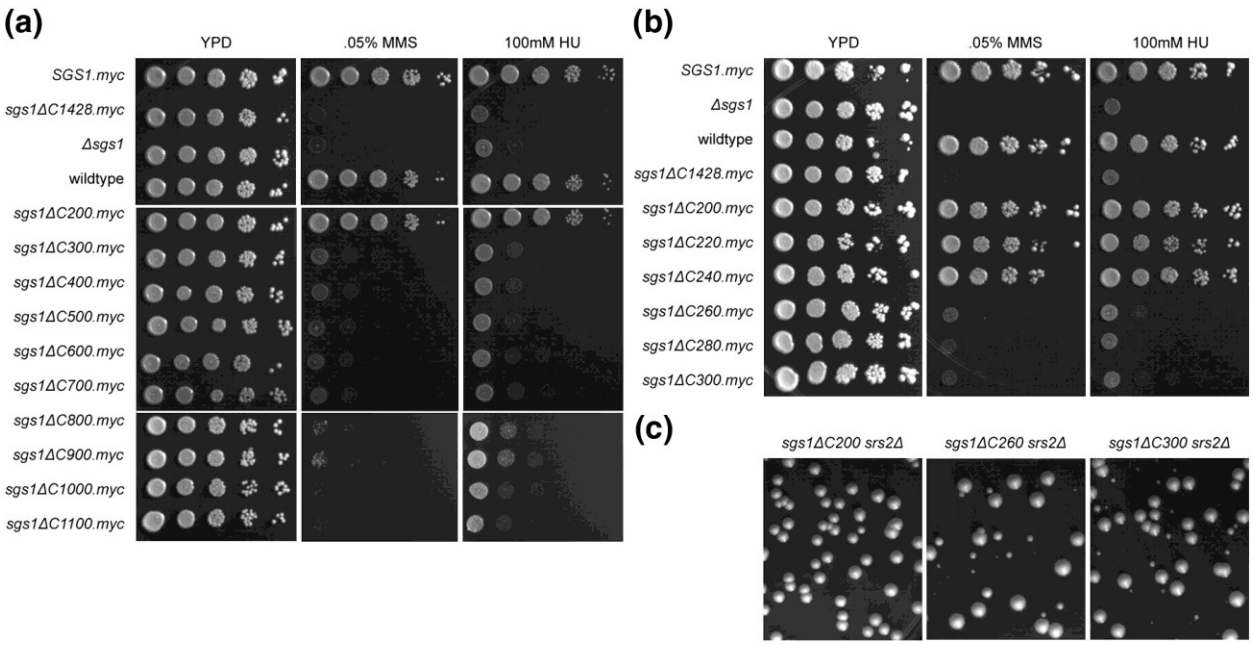


Figure A.2. Sensitivity of cells expressing Sgs1 truncation alleles to the DNA damaging agents HU and MMS. Ten-fold dilutions of exponentially growing cultures (OD600 = 0.5) were spotted on YPD for viable cell count and on YPD containing 100mM HU or 0.05% MMS, followed by incubation at 30° C. **(A)** Haploid cells expressing *sgs1* alleles lacking 300 or more residues from the C-terminus are as sensitive to HU and MMS as the null allele. **(B).** Additional incremental 20-amino-acid deletions reveal that cells expressing *sgs1* alleles lacking up to 240 residues are as resistant to HU and MMS as wildtype cells whereas those lacking 260 or more residues are as sensitive as the *sgs1Δ* mutant. **(C)** Spores from diploids heterozygous for an *srs2Δ* deletion and heterozygous either for the *sgs1-ΔC200, sgs1Δ-C260* or *sgs1-ΔC300* were spread on YPD to allow for growth of spores of all possible genotypes. Similar sized colonies obtained from the spores of the diploid heterozygous for *sgs1ΔC200* and *srs2Δ* mutations (left) indicate that the *sgs1ΔC200 srs2Δ* mutant grows as well as the single mutants, suggesting that deletion of the C-terminal 200 amino acid residues does not negatively affect growth of the *srs2Δ* mutant. In contrast, spores from diploids heterozygous for the *srs2Δ* mutation and the *sgs1ΔC260* allele (middle) or the *sgs1ΔC300* allele (right), grew into a mixture of normal-sized colonies (corresponding to single mutants and wildtype) and small-sized colonies (corresponding to *srs2Δ sgs1ΔC260* or *srs2Δ sgs1ΔC300* mutants as determined by genotyping), demonstrating that an intact RQC domain in Sgs1 is required for the viability of the *srs2Δ* mutant.

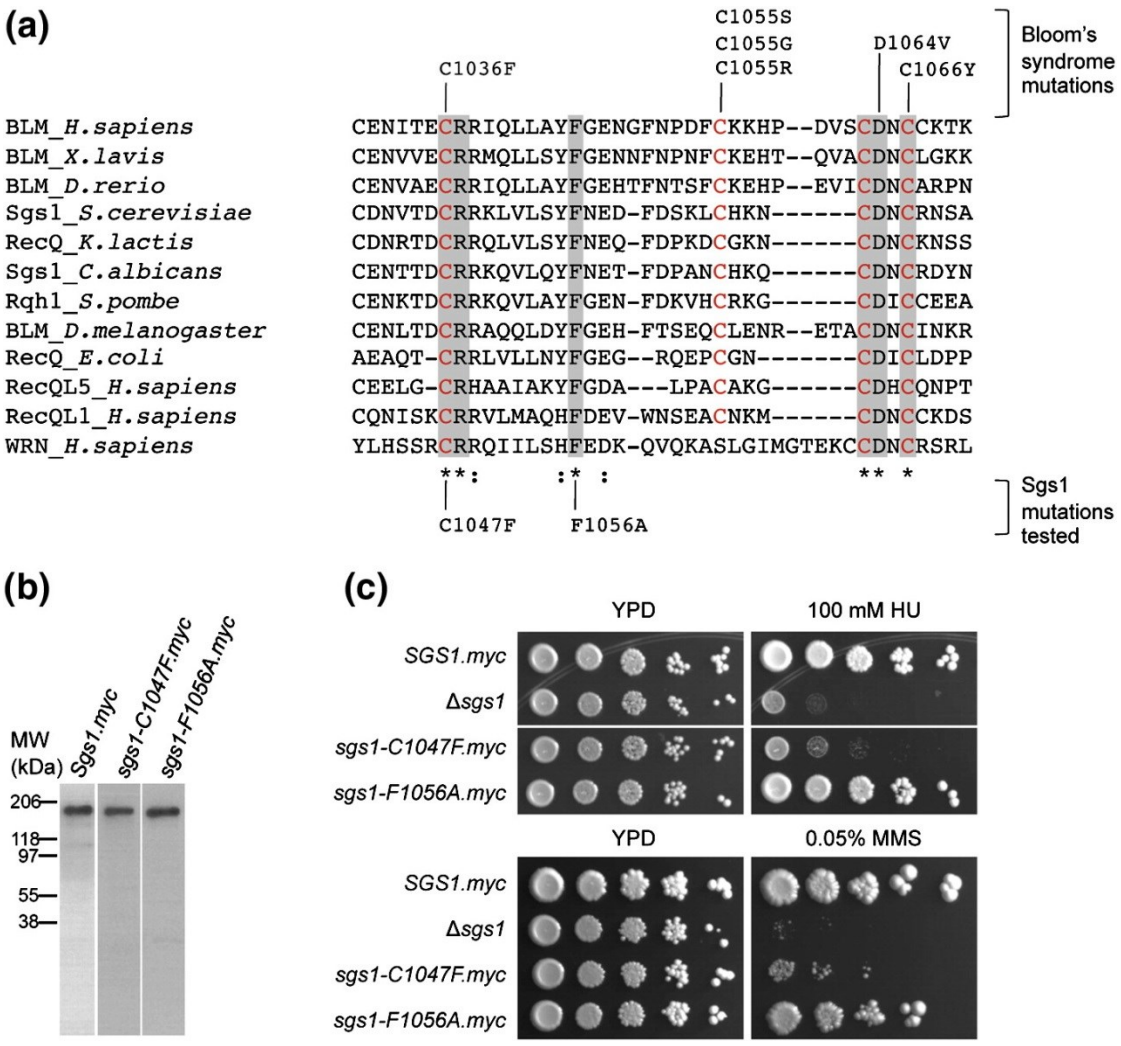


Figure A.3. Effect of zinc-binding domain mutations on Sgs1 function *in vivo*. (A) Zinc-binding domain is conserved from bacterial to human RecQ-like DNA helicases. Protein sequences were aligned with ClustalW2 84. The alignment of RecQL1 was manually adjusted. Amino acid residues identical in all sequences are highlighted in gray and indicated by '*' below the alignment, conserved substitutions are indicated by ':' below the alignment, and cysteine residues thought to be involved in zinc-binding are shown in red. At least six different missense mutations in the zinc-binding domain are associated with Bloom's syndrome. **(B)** C1047F and F1056A mutations were introduced into Sgs1 and expression was confirmed by western blot using antibody against the C-terminal myc-epitope. Molecular weights (MW) are indicated in kDa to the left. **(C)** Mutation of the highly conserved F1056 does not impair Sgs1 function whereas the C1047F mutation leads to an increase in sensitivity to HU and MMS, but not to the level seen in the *sgs1Δ* mutant.

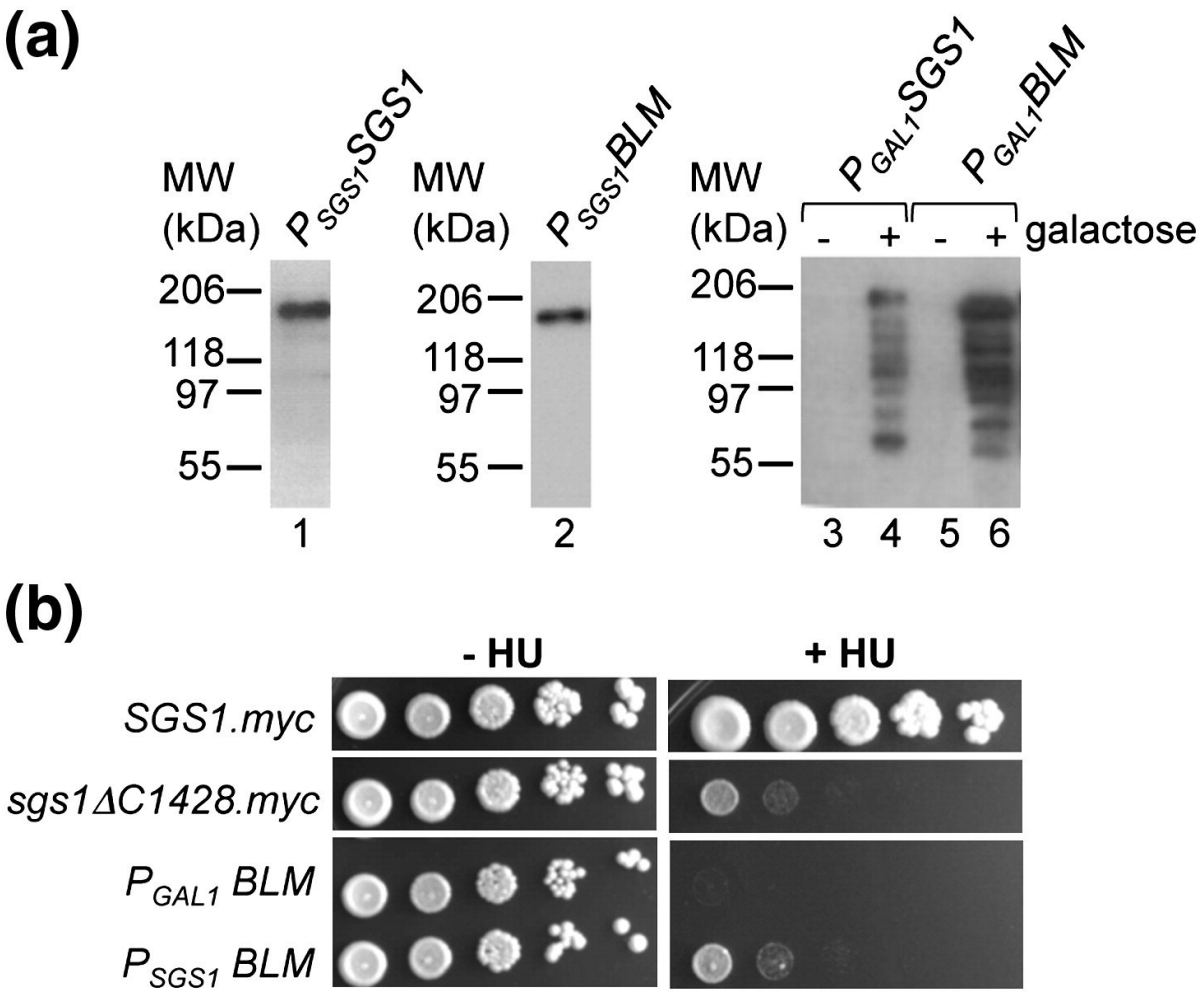


Figure A.4. BLM expression does not suppress *sgs1Δ* defects and BLM overexpression is detrimental to yeast cells. (A) Expression of myc-epitope tagged Sgs1 (lane 1) and BLM (lane 2) from the native chromosomal *SGS1* locus or galactose-inducible overexpression of myc-epitope tagged Sgs1 (lane 4) and BLM (lane 6) in yeast cells grown in YP supplemented with 1% sucrose and 1% galactose (to induce expression, lanes 4 and 6) or without galactose (lanes 3 and 5). Both BLM and Sgs1 show signs of degradation upon overexpression (lanes 4 and 6) whereas expression from the native *SGS1* promoter is stable (lanes 1 and 2). Molecular weights (MW) are indicated in kDa on the left. **(B)** Cells expressing BLM from the *SGS1* promoter on chromosome VIII are as sensitive to HU as cells lacking Sgs1 (*Δsgs1*). Replacement of the natural *SGS1* promoter with a galactose-inducible *GAL1* promoter induces BLM overexpression and leads to increased HU sensitivity. Ten-fold dilutions of cells were spotted on media containing 1% sucrose and 1% galactose (to induce BLM overexpression) with and without 100 mM HU.

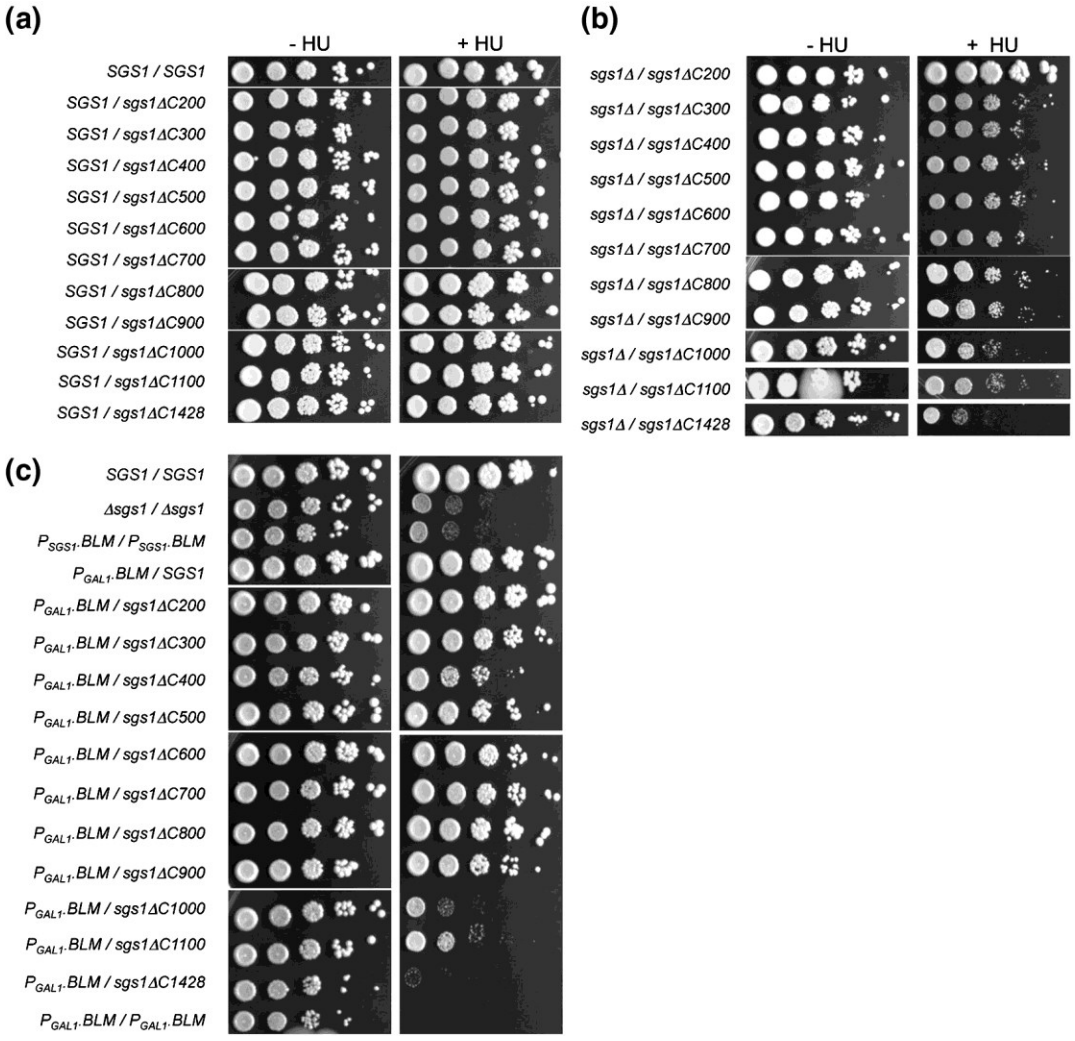


Figure A.5. HU sensitivity of diploid cells expressing *BLM* and mutant alleles of *SGS1*. (A) Ten-fold dilutions of exponentially growing diploids expressing truncation alleles of *SGS1* in the presence of a wildtype allele were spotted on YPD media with and without 100 mM HU. **(B)** Ten-fold dilutions of exponentially growing diploids expressing truncation alleles of *SGS1* in the absence of a wildtype allele were spotted on YPD media with and without 100 mM HU. **(C)** Ten-fold dilutions of exponentially growing diploids overexpressing BLM from a *GAL1* promoter inserted at the native *SGS1* locus and expressing truncation alleles of *SGS1* under control of the native *SGS1* promoter on the other allele were spotted on media containing 1% galactose (to induce gene expression) and 1% sucrose with or without 100 mM HU.

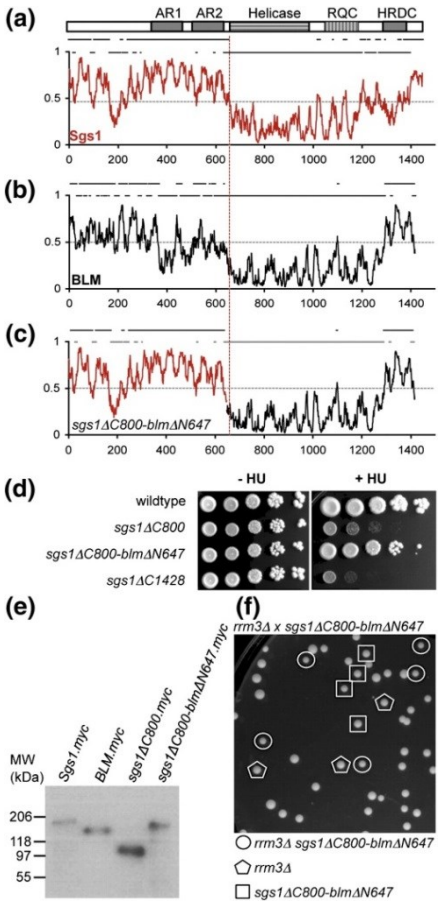


Figure A.6. Construction of a functional chimerical protein composed of the N-terminus of Sgs1 and the C-terminus of BLM. (A – B) Protein disorder prediction of Sgs1 (red) and BLM (black) using the IUPred algorithm. Values above 0.5 indicate a disordered residue whereas values below 0.5 indicate ordered residues; amino acid residue numbers (1–1447) are indicated on the abscissa. Black lines above the graph show a simplified order and disorder distribution along the length of the protein with values above 0.5 being assigned a "1" and values below 0.5 being assigned a "0". The vertical red line indicates the site in Sgs1, BLM and the chimera where the disordered N-terminal segment transitions into the ordered helicase domain at residue 647/648. This site was chosen as the fusion site for the chimera. The approximate location of Sgs1 domains is indicated above panel A. **(C)** Disorder prediction for the Sgs1-BLM chimera in which the N-terminal 647 residues of BLM (black) were replaced with the N-terminal 647 residues of Sgs1 (red). **(D)** Ten-fold dilutions of exponentially growing haploids were spotted on YPD with or without 100 mM HU. **(E)** The C-terminus of the Sgs1-BLM chimera was fused to a myc-epitope tag and expression was confirmed by western blotting. Molecular weight marker bands (kD) are indicated on the left **(F)** A diploid heterozygous for the *rrm3Δ* mutation and the *sgs1ΔC800-blmΔN647* allele expressing the chimera was sporulated and random spores were plated on YPD to allow all spores to grow. An open circle indicates the haploid double mutant, and the open square and pentagon indicate haploid *sgs1ΔC800-blmΔN647* and *rrm3Δ* single mutants, respectively.

Table A.1.

Accumulation of GCRs in cells expressing mutant alleles of SGS1

Relevant Genotype	GCR rate (Can^r 5-FOA^r × 10^{-10})	95% CI[b] (Canr 5-FOA^r × 10^{-10})
Wild type	1.1	< 1–6.2
sgs1Δ	251	80–310
sgs1ΔC200	7	< 6–23
sgs1ΔC220	31	5–41
sgs1ΔC240	10	< 6–27
sgs1ΔC260	159	85–362
sgs1ΔC280	244	166–387
sgs1ΔC300	145	76–204
sgs1ΔC400	106	60–180
sgs1ΔC500	102	53–252
sgs1ΔC600	152	26–283
sgs1ΔC700	189	49–271
sgs1ΔC800	133	71–225
sgs1ΔC1428	206	97–273
sgs1-C1047F	64	35–131
sgs1-F1056A	< 16	< 10–26
mec3Δsgs1ΔC200	11	< 7–22
mec3Δsgs1ΔC300	1003	691–1500
mec3Δsgs1ΔC800	758	645–895
mec3Δsgs1ΔC800-blmΔN647 c	361	330–419

i

All *sgs1* truncations (*sgs1ΔC*) are C-terminally fused to a myc-epitope tag.

ii

95% confidence intervals (CI) were calculated according to Nair [59].

c

The *sgs1ΔC800-blmΔN647* allele expresses a chimeric protein that consists of the N-terminal 647 residues of Sgs1 and the C-terminal 770 residues of human BLM.

Table A.2.
Effect of BLM expression on GCR accumulation in the sgs1Δ mutant

Relevant genotype	Galactose concentration in media (%)	GCR rate (Can^r 5-FOA^r × 10−10)	95% CI[b] (Can^r 5-FOA^r × 10−10)
Wild type	0	1.1	< 1–6.2
PSGS1BLM	0	70	56–151
PGALBLM	0	61	30–153
PGALBLM	0.1	335	233–576
PGALBLM	0.5	382	170–777
PGALBLM	2	1832	1090–2910
PGALSGS1	2	< 11	< 9–12
sgs1Δ	2	54	23–104

1
Human *BLM* cDNA was inserted at the endogenous *SGS1* locus, fused to the native *SGS1* promoter (*PSGS1*) or fused to a galactose-inducible promoter (*PGAL*). In *PGALSGS1*, the native *SGS1* promoter region was disrupted by fusing the *SGS1* ORF to a galactose-inducible promoter. If strains expressing *BLM* or *SGS1* genes from the galactose-inducible GAL1 promoter were grown in less than 2% galactose (to lower protein expression levels) media was supplemented with sucrose to reach a total sugar concentration of 2%.
2
95% confidence intervals (CI) were calculated according to Nair [59].

TABLE A.3: Yeast strains used in this study

Strain	Genotype
KHSY802	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3*
KHSY1338	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1::HIS3*
KHSY1705	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::ura3::TRP1, sgs1::BLM.HIS3*
KHSY2341	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, mec3::kanMX6, sgs1ΔC200.MYC.HIS3*
KHSY2347	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, mec3::kanMX6, sgs1ΔC300.MYC.HIS3*
KHSY2599	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3*
KHSY2602	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::ura3::TRP1/YEL069C::ura3::TRP1, sgs1::kanMX6/sgs1::kanMX6*
KHSY2726	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, SGS1/sgs1C200.MYC.HIS3*

Table A.3: Yeast strains used in this study (Continued)

KHSY2828	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::ura3::TRP1, sgs1::kanMX6/sgs1C200.MYC.HIS3*
KHSY2837	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::ura3::TRP1, sgs1::kanMX6/sgs1C300.MYC.HIS3*
KHSY2880	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, SGS1/sgs1ΔC300.MYC.HIS3*
KHSY2883	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, SGS1/sgs1ΔC400.MYC.HIS3*
KHSY2886	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, SGS1/sgs1ΔC500.MYC.HIS3*
KHSY2889	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, SGS1/sgs1ΔC600.MYC.HIS3*

Table A.3: Yeast strains used in this study (Continued)

KHSY2892	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, SGS1/sgs1ΔC700.MYC.HIS3*
KHSY2895	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, SGS1/sgs1ΔC800.MYC.HIS3*
KHSY2898	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, SGS1/sgs1ΔC1428.MYC.HIS3*
KHSY2928	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::ura3::TRP1, sgs1::kanMX6/sgs1ΔC400.MYC.HIS3*
KHSY2931	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::ura3::TRP1, sgs1::kanMX6/sgs1ΔC500.MYC.HIS3*
KHSY2934	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::ura3::TRP1, sgs1::kanMX6/sgs1ΔC600.MYC.HIS3*

Table A.3: Yeast strains used in this study (Continued)

KHSY2937	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::ura3::TRP1, sgs1::kanMX6/sgs1ΔC700.MYC.HIS3*
KHSY2940	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::ura3::TRP1, sgs1::kanMX6/sgs1ΔC800.MYC.HIS3*
KHSY2943	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::ura3::TRP1, sgs1::kanMX6/sgs1ΔC1428.MYC.HIS3*
KHSY2970	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1.MYC.HIS3*
KHSY2972	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1ΔC200.MYC.HIS3*
KHSY2973	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1ΔC300.MYC.HIS3*
KHSY2974	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1ΔC400.MYC.HIS3*
KHSY2975	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1ΔC500.MYC.HIS3*

Table A.3: Yeast strains used in this study (Continued)

KHSY2976	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1ΔC600.MYC.HIS3*
KHSY2977	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1ΔC700.MYC.HIS3*
KHSY2978	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1ΔC800.MYC.HIS3*
KHSY2979	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1ΔC1428.MYC.HIS3*
KHSY3181	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, SGS1/sgs1ΔC1100.MYC.HIS3*
KHSY3218	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, TRP1.PGAL1.SGS1*
KHSY3332	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1::BLM.HIS3*
KHSY3346	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::ura3::TRP1, BLM.MYC.kanMX6.HIS3*
KHSY3350	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, kanMX6.PGAL1.BLM.HIS3*
KHSY3353	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, kanMX6.PGAL1.BLM.HIS3*

Table A.3: Yeast strains used in this study (Continued)

KHSY3355	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1ΔC800-blmΔN647.HIS3*
KHSY3363	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1ΔC800-blmΔN647.HIS3, Δmec3::kanMX6*
KHSY3372	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, sgs1::BLM.HIS3/sgs1::BLM.HIS3*
KHSY3409	*ura3-52/ura3-52, leu2Δ/leu2Δ , trp1Δ63/ trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ys2ΔBgl, hom3-10/hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, kanMX6.PGAL1.BLM.HIS3/SGS1.MYC.HIS3*
KHSY3410	*ura3-52/ura3-52, leu2Δ/leu2Δ , trp1Δ63/ trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ys2ΔBgl, hom3-10/hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, kanMX6.PGAL1.BLM.HIS3/sgs1ΔC200.MYC.HIS3*
KHSY3412	*ura3-52/ura3-52, leu2Δ/leu2Δ , trp1Δ63/ trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ys2ΔBgl, hom3-10/hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, kanMX6.PGAL1.BLM.HIS3/sgs1ΔC300.MYC.HIS3*

Table A.3: Yeast strains used in this study (Continued)

KHSY3414	*ura3-52/ura3-52, leu2Δ/leu2Δ , trp1Δ63/ trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ys2ΔBgl, hom3-10/hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, kanMX6.PGAL1.BLM.HIS3/sgs1ΔC400.MYC.HIS3*
KHSY3416	*ura3-52/ura3-52, leu2Δ/leu2Δ , trp1Δ63/ trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ys2ΔBgl, hom3-10/hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, kanMX6.PGAL1.BLM.HIS3/sgs1ΔC500.MYC.HIS3*
KHSY3417	*ura3-52/ura3-52, leu2Δ/leu2Δ , trp1Δ63/ trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ys2ΔBgl, hom3-10/hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, kanMX6.PGAL1.BLM.HIS3/sgs1ΔC600.MYC.HIS3*
KHSY3419	*ura3-52/ura3-52, leu2Δ/leu2Δ , trp1Δ63/ trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ys2ΔBgl, hom3-10/hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, kanMX6.PGAL1.BLM.HIS3/sgs1ΔC700.MYC.HIS3*
KHSY3420	*ura3-52/ura3-52, leu2Δ/leu2Δ , trp1Δ63/ trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ys2ΔBgl, hom3-10/hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, kanMX6.PGAL1.BLM.HIS3/sgs1ΔC800.MYC.HIS3*

Table A.3: Yeast strains used in this study (Continued)

KHSY3422	*ura3-52/ura3-52, leu2Δ/leu2Δ , trp1Δ63/ trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ys2ΔBgl, hom3-10/hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, kanMX6.PGAL1.BLM.HIS3/sgs1ΔC900.MYC.HIS3*
KHSY3423	*ura3-52/ura3-52, leu2Δ/leu2Δ , trp1Δ63/ trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ys2ΔBgl, hom3-10/hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, kanMX6.PGAL1.BLM.HIS3/sgs1ΔC1000.MYC.HIS3*
KHSY3424	*ura3-52/ura3-52, leu2Δ/leu2Δ , trp1Δ63/ trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ys2ΔBgl, hom3-10/hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, kanMX6.PGAL1.BLM.HIS3/sgs1ΔC1100.MYC.HIS3*
KHSY3425	*ura3-52/ura3-52, leu2Δ/leu2Δ , trp1Δ63/ trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ys2ΔBgl, hom3-10/hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, kanMX6.PGAL1.BLM.HIS3/sgs1ΔC1428.MYC.HIS3*
KHSY3426	*ura3-52/ura3-52, leu2Δ/leu2Δ , trp1Δ63/ trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ys2ΔBgl, hom3-10/hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, kanMX6.PGAL1.BLM.HIS3/kanMX6.PGAL1.BLM.HIS3*

Table A.3: Yeast strains used in this study (Continued)

KHSY3429	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::ura3::TRP1, sgs1::kanMX6/sgs1ΔC1100.MYC.HIS3*
KHSY3470	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1ΔC220.MYC.HIS3*
KHSY3473	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1ΔC240.MYC.HIS3*
KHSY3476	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1ΔC260.MYC.HIS3*
KHSY3479	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1ΔC280.MYC.HIS3*
KHSY3500	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1ΔC900.MYC.HIS3*
KHSY3502	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1ΔC1000.MYC.HIS3*
KHSY3504	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1ΔC1100.MYC.HIS3*
KHSY3510	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, TRP1.PGAL1. SGS1.MYC.HIS3*
KHSY3512	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1-C1047F.TRP1*

Table A.3: Yeast strains used in this study (Continued)

KHSY3516	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1-F1056A.TRP1*
KHSY3517	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1-C1047F.MYC.HIS3*
KHSY3520	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, sgs1-F1056A.MYC.HIS3*
KHSY3523	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8, YEL069C::URA3, kanMX6.PGAL1.BLM.MYC.TRP1.HIS3*
KHSY3528	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::ura3::TRP1, sgs1::kanMX6/sgs1ΔC1000.MYC.HIS3*
KHSY3534	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::ura3::TRP1, sgs1::kanMX6/sgs1ΔC900.MYC.HIS3*
KHSY3536	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, SGS1/sgs1ΔC1000.MYC.HIS3*
KHSY3539	*ura3-52/ura3-52, leu2Δ1/ leu2Δ1, trp1Δ63/trp1Δ63, his3Δ200/his3Δ200, lys2ΔBgl/ lys2ΔBgl, hom3-10/ hom3-10, ade2Δ1/ade2Δ1, ade8/ade8, YEL069C::URA3/YEL069C::URA3, SGS1/sgs1ΔC900.MYC.HIS3*

Table A.3: Yeast strains used in this study (Continued)

KHSY3543	*MATa ura3-52, leu2Δ1, trp1Δ63, his3Δ200, lys2ΔBgl, hom3-10, ade2Δ1, ade8. YEL069C::URA3, sgs1ΔC800-blmΔN647.MYC.TRP1*

APPENDIX B:

NMR DATA

Table B1: Peak assignments for Sgs1 N^{1-125} peptide

Residue Number	Residue	H	HN	CA	CB
1	M				
2	V	8.21988	122.3445	62.3294	32.91862
3	T	8.22853	119.026	61.76092	69.8252
4	K	8.33585	125.4552	54.21737	32.62978
5	P	N/A	N/A	63.02	32.21
6	S	8.36899	115.9146	58.29224	63.82262
7	H				
8	N				
9	L	8.19468	122.5807	55.48658	42.32477
10	R	8.28251	122.1783	56.49396	30.52714
11	R				
12	E				
13	H				
14	K				
15	W				
16	L	7.90041	123.5679	55.17055	42.47335
17	K	8.06267	121.7686	56.57215	32.96658
18	E				
19	T	8.04	114.57	62.1205	69.6257
20	A	8.2848	126.2067	52.82452	19.27114
21	T	8.042 8.20142	113.5352	62.00811	69.79695
22	L	124.57117	124.5712	55.31136	42.28904
23	Q	8.329	121.3023	56.0418	29.3875
24	E			56.8542	30.5093
25	D	8.335	121.6091	54.5717	41.1865

Table B1: Peak assignments for Sgs1 N^{1-125} peptide (Continued)

26	K	8.2065	121.5207	57.0863	32.7918
27	D	8.26583	120.1133	54.91777	41.10337
28	F	7.96111	120.5838	58.7914	39.41052
29	V	7.83837	121.8131	63.23366	32.5935
30	F	8.13081	122.8441	58.80898	39.36015
31	Q	8.10104	121.0846	56.29585	29.31666
32	A	8.0602	124.3929	53.1189	18.90968
33	I	7.95223	119.4848	61.86297	38.55723
34	Q	8.20301	122.9763	56.26645	29.1325
35	K				
36	H			56.6091	30.5955
37	I	7.98315	122.1882	61.41187	38.63476
38	A	8.26634	127.0843	52.79229	19.19498
39	N				
40	K	8.12	121.49	56.2169	32.95
41	R	8.26615	123.4053	54.00874	
42	P	N/A	N/A	63.01	32.31
43	K	8.47766	122.0044	56.45931	33.05701
44	T	8.14235	115.244	61.56435	69.71656
45	N			52.83696	38.78605
46	S			56.50993	63.42659
47	P				
48	P				
49	T				
50	T	8.16	118.78	54.59	41.2
51	P	N/A	N/A	63.04	39.01
52	S	8.44744	116.8972	58.21897	63.95412
53	K	8.43197	123.4124	56.67553	32.972
54	D	8.2693	120.9317	54.6034	41.1861
55	E	8.2659	121.222	56.6222	30.2675
56	C	8.33404	119.5746	58.50417	28.30785
57	G	8.2578	110.8407	44.69437	
58	P	N/A	N/A	63.56	32.1
59	G	8.51983	109.4864	45.40379	
60	T	7.9646	113.204		
61	T				
62	N			53.1974	
63	F	8.0905	120.8059	57.8361	
64	I	8.0588	122.8569	61.0498	38.7329

Table B1: Peak assignments for Sgs1 N^{1-125} peptide (Continued)

65	T	8.13716	118.0419	61.6789	69.62651
66	S	8.19	118.5	57.96791	64.03301
67	I	8.14958	124.2996	58.65614	38.70634
68	P	N/A	N/A	63.17	32.11
69	A				
70	S	8.22905	114.6649	58.1819	64.03219
71	G	8.16959	110.6062	44.67143	
72	P	N/A	N/A		
73	T	8.28	113.97	61.86	69.7
74	N				
75	T	8.1234	114.4951		
76	A	8.26019	126.4146	52.79309	19.33326
77	T	8.06419	113.6282	62.13061	69.70393
78	K	8.26435	123.6846	56.3553	32.92902
79	Q				
80	H				
81	E			56.728	30.5256
82	V	8.17	121.25	62.6101	56.728
83	M				
84	Q				
85	T	8.17685	115.9563	61.94339	69.68748
86	L	8.25404	124.6848	55.16385	42.39013
87	S	8.23658	116.2874	58.19368	63.84549
88	N	8.41612	120.698	53.35475	38.95899
89	D	8.30018	120.5722	54.77021	41.1655
90	T	8.03624	113.6258	62.31408	69.52244
91	E	8.32668	123.086	57.37228	29.98405
92	W	8.01026	120.9964	57.7331	29.16892
93	L	7.83456	122.8286	55.60242	42.37008
94	S	7.95789	115.3333	58.62363	63.59823
95	Y	7.96941	121.879	58.30853	38.75819
96	T	7.92071	115.502	61.87001	69.80556
97	A	8.12105	126.2384	52.8683	19.27598
98	T	8.01092	112.2691	61.84889	69.65426
99	S				
100	N	8.23843	117.7062	53.1433	38.91206
101	Q	8.22509	120.3727	56.30673	29.22771
102	Y	8.07296	120.253	57.5453	38.5643
103	A	7.93554	124.7221	52.4258	19.67012
104	D	8.22067	119.5048	54.25103	41.13511

Table B1: Peak assignments for Sgs1 N^{1-125} peptide (Continued)

105	V	7.90708	121.1807	59.67993	32.80018
106	P	N/A	N/A	62.5757	
107	M	8.21988	122.3445	56.36728	32.91862
108	V	8.02629	120.714	61.93245	33.12821
109	D	8.31981	123.9739	54.21426	41.31996
110	I				
111	P	N/A	N/A	63.09	32.15
112	A	8.4098	124.753	52.59244	19.25044
113	S	8.30954	114.9381	58.36703	63.77139
114	T	8.1277	115.3548	61.7155	69.656
115	S	8.1623	117.4504	58.3601	63.7295
116	V	8.11231	122.2318	62.34681	32.73477
117	V	8.17994	124.1036	62.21606	32.7663
118	S	8.33011	119.8839	58.13893	63.73322
119	N				
120	P	N/A	N/A	63.27	32.14
121	R	8.38416	120.6465	56.16673	30.7691
122	T	8.08281	116.708	59.55571	69.56049
123	P	N/A	N/A	63.66	32.11
124	N	8.4193	118.3227	53.33006	39.16951
125	G	7.84815	115.3577	46.22267	

Table B2: Peak assignments for Sgs1 N^{1-80} peptide

Residue Number	Residue	H	HN	CA	CB
1	M			55.3785	32.8186
2	V	8.21	122.31	62.28	32.87
3	T	8.23151	119.0454	61.7235909	69.80896
4	K	8.33908	125.4415	54.1603394	32.6217995
5	P	N/A	N/A	63.0794182	83.009697
6	S	8.37255	115.8686	58.2680206	63.7881699
7	H				
8	N			53.3577805	38.7228394
9	L	8.19247	122.5594	55.4583092	42.2596207
10	R	8.25	121.46	56.1738	30.5
11	R	8.27	122.21	56.26	30.68
12	E				
13	H				
14	K				32.7890816
15	W	8.01492	121.0278	57.1287804	29.399929
16	L	7.89708	123.5431	55.1884117	42.4949112
17	K	8.05774	121.7429	56.566761	32.9739304
18	E	8.41845	121.5595	56.8657	30.219
19	T	8.03966	114.5798	61.9319115	69.7065277
20	A	8.28776	126.295	52.65	19.2876396
21	T	8.04672	113.706	61.9318314	69.7565918
22	L	8.20583	124.5754	55.31213	42.2676888
23	Q	8.33352	121.3175	55.8962097	29.42

Table B2: Peak assignments for Sgs1 N^{1-80} peptide (Continued)

24	E	8.43	121.98	56.63	30.45
25	D	8.34	121.64	54.5437508	41.1975098
26	K	8.20896	121.5487	57.0044098	32.8010292
27	D	8.26834	120.1106	54.9319	41.0997314
28	F	7.96073	120.5829	58.7588997	39.3948593
29	V	7.8332	121.7531	63.1692696	32.5655403
30	F	8.13168	122.9027	58.7493401	39.359169
31	Q	8.0996	121.1583	56.244709	29.3470402
32	A	8.06327	124.4081	53.1120911	18.0088692
33	I	7.94951	119.468	61.8188896	38.5292702
34	Q	8.21003	123.063	56.2424812	29.1556606
35	K	8.1676	121.789	56.89	32.9058
36	H			56.6144295	30.5993195
37	I	7.96672	122.2637	61.3540916	38.6244011
38	A	8.26951	127.1183	52.7955513	19.1597309
39	N	8.2482	117.7394	53.2123	38.8374
40	K	8.1294	121.5962	56.1281	32.9709015
41	R	8.27519	123.4202	54.0261688	30.1029
42	P	N/A	N/A	62.956459	32.2294
43	K	8.48149	122.0392	56.4200401	33.0219688
44	T	8.15418	115.3279	61.4978905	69.9117432
45	N	8.32	121.0177	53.08	38.94
46	S	8.24	117.72	56.57	63.14
47	P	N/A	N/A		
48	P	N/A	N/A	63.0189781	32.0674896

Table B2: Peak assignments for Sgs1 N^{1-80} peptide (Continued)

49	T	8.23617	114.3647	61.68	69.8253784
50	T	8.16683	118.8081	59.6694298	69.7886505
51	P	N/A	N/A	63.0924911	32.2289505
52	S	8.44991	116.8439	58.2057915	63.8908081
53	K	8.43316	123.4211	56.6177711	32.9634895
54	D	8.27	120.98	54.6	41.2
55	E	8.27	121.13	56.64	30.33
56	C	8.34	119.62	58.4585609	28.2946396
57	G	8.26367	110.7965	43.9230804	
58	P	N/A	N/A	63.5016594	32.0774689
59	G	8.52181	109.4896	45.3688393	
60	T	7.96866	113.2012	61.8311691	69.9331436
61	T	8.18	115.94	61.9390488	69.6863632
62	N			53.1432495	
63	F	8.09147	120.8253	57.8064499	39.4358597
64	I	8.064	122.95	61.06	38.63
65	T	8.13863	118.1441	61.7212601	69.6549683
66	S	8.2	118.54	58.05	63.7762413
67	I	8.15659	124.3697	58.6881409	38.7136688
68	P	N/A	N/A	63.1273804	32.1472511
69	A	8.41648	124.7613	52.5667801	19.2261391
70	S	8.2387	114.7023	58.1207008	64.0062027
71	G	8.1787	110.6324	83.009697	63.99
72	P	N/A	N/A	63.2068481	32.2116318
73	T	8.29179	114.0037	61.8410492	69.7118683

Table B2: Peak assignments for Sgs1 N^{1-80} peptide (Continued)

74	N	8.43	121.4487	53.1906	38.7594
75	T	8.12689	114.5378	61.9174118	69.5784836
76	A	8.29	126.4438	52.7818794	19.3586197
77	T	8.07365	113.5912	61.8951683	69.7438736
78	K	8.25609	124.0322	56.1994705	33.1184387
79	Q	8.39793	122.5607	55.9633102	29.68297
80	H	7.9755	125.3586	57.1394	30.5982

APPENDIX C:

PERMISSIONS

License Number	3633171257086
License date	May 20, 2015
Licensed content publisher	Oxford University Press
Licensed content publication	Nucleic Acids Research
Licensed content title	A transient α-helical molecular recognition element in the disordered N-terminus of the Sgs1 helicase is critical for chromosome stability and binding of Top3/Rmi1:
Licensed content author	Jessica A. Kennedy, Gary W. Daughdrill, Kristina H. Schmidt
Licensed content date	12/01/2013
Volume number	41
Issue number	22
Type of Use	Thesis/Dissertation
Requestor type	Academic/Educational institute
Format	Print and electronic
Portion	Text Extract
Number of pages requested	12
Will you be translating?	No
Author of this OUP article	Yes
Order reference number	None
Title of your thesis / dissertation	Structure-Function Analysis of the DNA Damage Repair Complex STR in Saccharomyces cerevisiae
Expected completion date	Jun 2015
Estimated size(pages)	200
Publisher VAT ID	GB 125 5067 30
Total	0.00 USD

Figure C3: Permissions for text content from chapter 3.

License Number	3633171362838
License date	May 20, 2015
Licensed content publisher	Oxford University Press
Licensed content publication	Nucleic Acids Research
Licensed content title	A transient α-helical molecular recognition element in the disordered N-terminus of the Sgs1 helicase is critical for chromosome stability and binding of Top3/Rmi1:
Licensed content author	Jessica A. Kennedy, Gary W. Daughdrill, Kristina H. Schmidt
Licensed content date	12/01/2013
Volume number	41
Issue number	22
Type of Use	Thesis/Dissertation
Requestor type	Academic/Educational institute
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	11
Will you be translating?	No
Author of this OUP article	Yes
Order reference number	None
Title of your thesis / dissertation	Structure-Function Analysis of the DNA Damage Repair Complex STR in Saccharomyces cerevisiae
Expected completion date	Jun 2015
Estimated size(pages)	200
Publisher VAT ID	GB 125 5067 30
Total	0.00 USD

Figure C4: Permission for figures/tables in chapter 3.

License Number	3633171462232
License date	May 20, 2015
Licensed content publisher	Elsevier
Licensed content publication	Journal of Molecular Biology
Licensed content title	Sgs1 Truncations Induce Genome Rearrangements but Suppress Detrimental Effects of BLM Overexpression in Saccharomyces cerevisiae
Licensed content author	Hamed Mirzaei,Salahuddin Syed,Jessica Kennedy,Kristina H. Schmidt
Licensed content date	28 January 2011
Licensed content volume number	405
Licensed content issue number	4
Number of pages	15
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Title of your thesis/dissertation	Structure-Function Analysis of the DNA Damage Repair Complex STR in Saccharomyces cerevisiae
Expected completion date	Jun 2015
Estimated size (number of pages)	200
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD

Figure C5: Permissions for content in appendix A.