G-quadruplex recognition by the yeast telomere binding protein Rap1

Traczyk, Anna

2018

Traczyk, A. (2018). G-quadruplex recognition by the yeast telomere binding protein Rap1.
Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/82529>

<https://doi.org/10.32657/10220/47519>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**G-quadruplex recognition by the
yeast telomere binding protein Rap1**

Anna Traczyk

**SCHOOL OF BIOLOGICAL SCIENCES
NANYANG TECHNOLOGICAL
UNIVERSITY**

AUGUST 2018

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me except where otherwise stated in this thesis. The thesis work has not been submitted for a degree or professional qualification to any other university or institution. I declare that this thesis is written by myself and is free of plagiarism and of sufficient grammatical clarity to be examined. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

18/01/2013
.....
Date

Anna Treayle
.....
Signature

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it of sufficient grammatical clarity to be examined. To the best of my knowledge, the thesis is free of plagiarism and the research and writing are those of the candidate's except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

18th Jan 2019
.....
Date

B. Rhodes
.....
Signature

ACKNOWLEDGEMENTS

I would like to thank Prof. Daniela Rhodes for giving me the great opportunity to work in her group. It was an amazing learning experience through a challenging and interesting project. I would not have been able to complete my PhD project without her guidance and support.

I treasure the time I spend at Princeton University working with Prof. Virginia Zakian. It was a great opportunity to be part of her laboratory and learn so much more about *Saccharomyces cerevisiae*.

I want to say thank you to all my colleagues who were always there to support and motivated me on the bad days and listen to my complaints on even worse days. I am particularly grateful for the help and advice of Dr. Simon Lattmann, who taught me everything I know about G-quadruplexes and much else in science. It was always a pleasure to discuss with him any scientific doubts and without whom the learning curve would have been much steeper.

Last but not least I want to thank my parents. Without their support, I would not have embarked on a PhD. I am thankful for the way they taught me that it is important to be curious in life, open minded and not be afraid to explore.

ABSTRACT

G - Quadruplex recognition by the yeast telomere binding protein Rap1

Although biochemical and structural studies have shown that DNA can adopt various conformations other than the canonical right – handed double helix, such structures were thought to be *in vitro* artifacts. It has been known for decades, that single stranded guanine – rich DNA can fold into a structure called G – quadruplexes. Those structures arise from guanine self – association through Hoogsteen hydrogen-bonds into planar quartets that are further stabilized by stacking on top of each other. Remarkably, recently these non-canonical secondary DNA (or RNA) structures have drawn much attention, as there is accumulating evidence in literature pointing towards the existence of G-quadruplex structures *in vivo*, suggesting their function as nucleic acid based regulatory mechanism and possible drug targets.

Together with more evidence of the existence of G – quadruplex structure in cells, the proteins that might be involved in their regulation has been discovered. Over 20 years ago, the major telomere binding protein in budding yeast called Rap1 was shown to be able to bind to the single stranded telomeric G-overhangs and promote the formation of the G- quadruplex structures *in vitro*. Even though it has been more than two decades since this discovery, understanding of its consequence *in vivo* remains limited. The results presented in this dissertation help to better understand this interaction through determination of three – dimensional structure of the Rap1 DNA binding domain in complex with G – quadruplex DNA. Furthermore, I explore the possible function of this interaction *in vivo* as well as the involvement of Rap1 in the regulation of the TElomeric Repeat-containing RNA (TERRA) transcript.

TABLE OF CONTENTS

STATEMENT OF ORIGINALITY	II
SUPERVISOR DECLARATION STATEMENT	III
ACKNOWLEDGEMENTS	IV
ABSTRACT	V
TABLE OF CONTENTS	VI
ABBREVIATIONS	X
LIST OF FIGURES	XIII
LIST OF TABLES	XVII
CHAPTER ONE: INTRODUCTION	1
1.1. G – QUADRUPLLEXES AND THEIR ROLES IN BIOLOGY	1
1.1.1. <i>G – quadruplex structures</i>	1
1.1.2. <i>Potential G – quadruplex location in the genome</i>	3
1.1.3. <i>G – quadruplexes and their biological functions</i>	6
1.1.4. <i>Proteins involved in the regulation of G - quadruplexes</i>	7
1.1.5. <i>G-quadruplexes in RNA biology</i>	10
1.1.6. <i>Structural basis for G – quadruplex recognition</i>	13
1.1.7. <i>G – quadruplex structures at telomeres</i>	16
1.2. SOLVING THE END PROTECTION PROBLEM AT TELOMERES	19
1.3. THE ROLE OF BUDDING YEAST MULTIFUNCTIONAL DNA BINDING	
PROTEIN RAP1	21
1.3.1. <i>ScRap1 role in transcriptional regulation</i>	22
1.3.2. <i>Mechanism of scRap1 DBD recognition of double-stranded DNA</i> ...	24
1.3.3. <i>Rap1 and its interaction with G4 in vitro</i>	26

1.4. TERRA: TELOMERIC REPEAT – CONTAINING RNA	29
1.4.1. TERRA transcription and regulation.....	29
1.4.2. TERRA higher order structures.....	30
1.4.3. TERRA functions.....	32
RESEARCH OBJECTIVES.....	35
CHAPTER TWO: MATERIAL AND METHODS.....	36
2.1. EXPRESSION AND PURIFICATION OF RECOMBINANT ScRap1	
CONSTRUCTS	36
2.1.1. Expression and purification of the DNA Binding Domain (DBD) of <i>ScRap1</i>	36
2.1.2. Expression and purification of the Myb domains of <i>ScRap1</i>	37
2.1.3. Expression and purification of full length <i>scRap1</i>	37
2.1.4. Expression and purification of N-terminal BRCT domain of <i>scRap1</i>	39
2.1.5. Expression and purification of RCT domain of <i>scRap1</i>	39
2.2. G-QUADRUPLIX DNA PREPARATION	40
2.3. CIRCULAR DICHROISM.....	40
2.4. ³² P-LABELLING	40
2.5. ELECTROPHORETIC MOBILITY SHIFT ASSAY AND NATIVE GEL	
ELECTROPHORESIS	40
2.6. CRYSTALLIZATION AND STRUCTURE DETERMINATION	41
2.7. STRAINS AND PLASMID CONSTRUCTION	42
2.8. WESTERN BLOT	43
2.9. SOUTHERN BLOT	43
2.10. RNA IMMUNOPRECIPITATION (RNA–CO IP).....	44
2.11. CHROMATIN IMMUNOPRECIPITATION (CHIP)	46

CHAPTER THREE: INSIGHT INTO G-QUADRUPLEX

RECOGNITION BY SCRAP148

3.1. PROTEIN EXPRESSION AND PURIFICATION.....49

3.1.1. *Expression and purification of full length scRap1*..... 49

3.1.2. *DNA-binding domain (DBD) expression and purification* 52

3.1.3. *Purification of the BRCT domain* 55

3.1.4. *Purification of RCT domain*..... 58

3.2. OLIGONUCLEOTIDES DERIVED FROM YEAST TELOMERIC SEQUENCE

CAN FORM PARALLEL G – QUADRUPLEX STRUCTURES IN SOLUTION .60

3.3. IDENTIFICATION OF THE DOMAIN OF SCRAP1 RESPONSIBLE FOR

INTERACTION WITH G-QUADRUPLEX DNA..... 62

3.4. CHARACTERIZATION OF SCRAP1 – DBD G-QUADRUPLEX PROMOTION

ACTIVITY 65

3.5. CRYSTALLIZATION AND STRUCTURE DETERMINATION OF SCRAP1 IN

COMPLEX WITH G – QUADRUPLEX DNA.....67

3.6. STRUCTURAL BASIS OF PARALLEL G-QUADRUPLEX RECOGNITION BY

scRap1-DBD..... 71

3.7. BINDING OF DNA-RECOGNIZING A-HELIX PEPTIDES OF MYB1 AND

MYB2 TO THE G-QUADRUPLEX 75

3.8. SCRAP1 INTERACTION WITH G - QUADRUPLEX..... 79

CHAPTER FOUR: SCRAP1 INTERACTION WITH TELOMERIC

REPEAT-CONTAINING RNA (TERRA)83

4.1. TELOMERIC RNA CAN FORM A G – QUADRUPLEX STRUCTURE AND

SCRAP1 DBD CAN INTERACT WITH IT *IN VITRO*.....84

4.2. CAN SCRAP1 DIRECTLY INTERACT WITH TERRA IN YEAST CELLS? .86

4.3. ScRAP1 REGULATION OF TERRA.....	90
CHAPTER FIVE: THE IMPORTANCE OF THE MINIMAL DBD OF SCRAP1.....	93
5.1. ScRAP1 Δ (583-596) HAS A LOWER AFFINITY TO DOUBLE-STRANDED DNA WITHOUT AFFECTING INTERACTION WITH G-QUADRUPLEX	96
5.2. ScRAP1 Δ (583-596 AA) IS NOT EFFICIENT IN SUSTAINING CELL VIABILITY, BUT SHOW THE SAME TOXIC EFFECT AS WILD TYPE WHEN OVEREXPRESSED.....	98
5.3. ScRAP1 Δ (583-596 AA) CAN BIND AND OVER COMPETE WILD TYPE ScRAP1 WHEN IT IS OVEREXPRESSED.....	102
5.4. THE IMPORTANCE OF THE C- TAIL OF scRAP1 DBD WRAPPING AROUND DNA FOR YEAST FUNCTIONALITY.....	105
CHAPTER SIX: DISCUSSION, CONCLUSION AND FUTURE DIRECTIONS	107
6.1. G-QUADRUPLEX RESEARCH A CENTURY LATER: FROM GUANINE GELS TO BIOLOGICAL FUNCTIONS	107
6.2. ScRAP1 INTERACTION WITH THE G-QUADRUPLEX STRUCTURES	108
6.3. A STRUCTURAL ANALYSIS OF G – QUADRUPLEX RECOGNITION.....	111
6.3. TELOMERE REGULATION THROUGH THE INTERACTION OF SCRAP1 WITH TERRA.....	114
6.4. THE EVOLVING WORLD OF G – QUADRUPLEX RESEARCH: FUTURE PERSPECTIVES AND CHALLENGES	116
APPENDIX.....	119
REFERENCE:.....	123

ABBREVIATIONS

aa	Amino acid
AID	Activation-induced cytidine deaminase
ALT	Alternative lengthening of telomeres
ATP	Adenosine triphosphate
ARS	Autonomously Replicating Sequence
bp	base pair
ChIP	Chromatin immunoprecipitation
CD	Circular dichroism
CEN	Centromere sequence
CRS	Class-switch recombination
CYC1	Cytochrome 1
DSBs	Double-stranded DNA breaks
dsDNA	Double-stranded DNA
DBD	DNA binding domain
<i>E. coli</i>	<i>Escherichia coli</i>
EMSA	Electrophoretic mobility shift assay
G4	G – quadruplex
GT strand	Guanine, thymine rich DNA
HDR	homology direct repair
IMAC	Immobilized metal ion affinity chromatography
IP	Immunoprecipitation
LEU	Leucine
lncRNA	Lon non-coding RNA

mRNA	messenger RNA
nt	nucleotide
PGs	Potential G – quadruplex structures
POT1	Protection of telomere 1
Rap1	Repressor Activator Protein 1
RBP	RNA binding proteins
RCT	Rap1 C terminus domain
RHAU	RNA helicase associated with AU- rich element
Rif 1/2	Rap1-interacting factor 1 and 2
RIP	RNA co-immunoprecipitation
RSM	RHAU-specific motif
RT-qPCR	Real Time–quantitative Polymerase Chain Reaction
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SDS	Sodium dodecyl sulfate
Sir 3/4	Silent information regulator 3 and 4
TAS	telomere associated sequences
TelRNA	Telomeric RNA
TER	ADH1-terminator
TERRA	Telomeric repeat-containing RNA
TERC	Telomerase RNA component
TetO7	Tet – operator boxes/sequences
TEV protease	Tobacco Etch Virus nuclear-inclusion-a endopeptidase
TF	Transcriptional factor
TIN2	TEF-interacting nuclear protein 2
tTA	Tetracycline transactivator

TRF1/2	Telomeric repeat binding factor 1 and 2
URA3	URAcil requiring 3

Codes for amino acids

Amino Acid	Single letter code	3-letter code	Amino Acid	Single letter code	3-letter code
Alanine	A	Ala	Leucine	L	Leu
Arginine	R	Arg	Lysine	K	Lys
Asparagine	N	Asn	Methionine	M	Met
Aspartic Acid	D	Asp	Phenylalanine	F	Phe
Cystein	C	Cys	Proline	P	Pro
Glutamine	Q	Gln	Serine	S	Ser
Glutamic Acid	E	Glu	Threonine	T	Thr
Glycine	G	Gly	Tryptophan	W	Trp
Histidine	H	His	Tyrosine	Y	Tyr
Isoleucine	I	Ile	Valine	V	Val

LIST OF FIGURES

Figure 1. Structure and polymorphism of G-quadruplex.....	3
Figure 2. Possible formation and role of G-quadruplex structures in cells.	5
Figure 3. Possible mechanisms of the regulation of RNA G-quadruplexes. ...	12
Figure 4. NMR structure of RHAU specific motif (RSM) with parallel G- quadruplex.	14
Figure 5. Crystal structure of bovine RHAU helicase in complex with G – quadruplex DNA.	16
Figure 6. Telomeric DNA structure and its elongation by telomerase.	18
Figure 7. Comparison of the shelterin complexes of (a) <i>Homo sapiens</i> and (b) <i>S. cerevisiae</i>	20
Figure 8. Domain organization of Rap1 in different species:	22
Figure 9. Dynamics of transcription factor binding.	23
Figure 10. Two views of the structure of scRap1 DBD in complex with telomeric DNA.....	25
Figure 11. ScRap1 binds and promotes G-quadruplex structures.....	28
Figure 13. Model of the involvement of TERRA in telomere lengthening.	33
Figure 14. Purification of full length scRap1 by His-Trap HP column.	49
Figure 15. Purification of full length scRap1 by Heparin HP column.....	50
Figure 16. Ion exchange chromatography for full length scRap1.	51
Figure 17. Purification of full length scRap1 by Ni-NTA agarose beads.....	51
Figure 18. Purification of full length scRap1 by gel-filtration chromatography.	52
Figure 19. Affinity tag purification of scRap1-DBD.....	53
Figure 20. Reverse His-trap purification of scRap1-DBD.....	54

Figure 21. Purification of scRap1-DBD by gel-filtration chromatography.....	54
Figure 22. Mass spectrometry analysis for BRCT domain.....	55
Figure 23. Affinity tag purification of scRap1-BRCT.....	56
Figure 24. Ion exchange purification of the BRCT.	57
Figure 25. Reverse His-trap purification of scRap1-BRCT.	57
Figure 26. Size-exclusion chromatography for scRap1-BRCT.	58
Figure 27. Affinity tag purification of RCT.....	59
Figure 28. Purification of scRap1-RCT by gel-filtration chromatography.....	59
Figure 29. Telomeric DNA from yeast forms parallel G- quadruplex.	61
Figure 30. The BRCT and RCT domains of scRap1 do not bind to G-quadruplex DNA.....	62
Figure 31. Purified scRap1 DBD is active and binds to double stranded telomeric sequence.....	63
Figure 32. ScRap1 DBD domain is responsible for protein interaction with different parallel G-quadruplexes.	65
Figure 33. ScRap1-DBD promotes G-quadruplex formation of telomeric DNA.	66
Figure 34. Attempts to crystalize scRap1 DBD with yeast telomeric sequence.	67
Figure 36. The unit cell of scRap1 DBD in complex with G-quadruplex DNA.	71
Figure 37. Global view of the structure of the complex between scRap1-DBD and the parallel G-quadruplex.....	73
Figure 38. Comparison of scRap1 DBD structures interacting with different DNA structures: dsDNA vs G-quadruplex.	74

Figure 39. Comparison of DNA-recognizing α -helix of Myb1 and Myb2.....	76
Figure 40. The importance of interaction with the backbone of G – quadruplex	77
Figure 41. Myb2 domain can interact with G-quadruplex DNA.....	78
Figure 42. Comparison of two Myb motifs binding to G – quadruplex DNA.	79
Figure 43. Model of scRap1 DBD interaction with DNA.	81
Figure 44. ScRap1 regulation of TERRA.	84
Figure 45. Telomeric RNA similarly to telomeric DNA sequence from yeast can form parallel G-quadruplex.....	85
Figure 46. ScRap1 binds to RNA G-quadruplex with the same affinity as to DNA G4 structures	86
Figure 47. Tagging scRap1 with 13xmyc in an inducible 1L Terra strain.	88
Figure 48. ScRap1 interaction with TERRA	89
Figure 49. Model for TERRA mediated elongation of critically short telomeres.	91
Figure 51. One of the examples of scRap1 DBD mutant in DNA recognition α -helix.	93
Figure 53. ScRap1 DBD (360-583 aa) has a lower affinity for dsDNA than minimal DBD (360-596 aa) without changing the affinity to G - quadruplex DNA.....	96
Figure 54. Full length scRap1 missing C-tail of DBD shows no changes in binding affinity to G –quadruplex, but lower affinity to dsDNA in comparison to wild type protein.	97
Figure 55. Developing heterozygote strain Δ rap1/ Rap1.....	99

Figure 56. ScRap1 Δ (583-596 aa) is not viable without wild type copy of Rap1 and it is not able to rescue the phenotype.	101
Figure 57. ScRap1 Δ (583-596 aa) overexpression is lethal for yeast, showing the same phenotype as overexpression of wild type scRap1.	102
Figure 58. ScRap1 Δ (583-596 aa) induced by GALL promoter can bind to DNA and compete with wild type copy of scRap1 for binding.	104
Figure 59. Comparison of scRap1 DBD – DNA contact in structure with dsDNA and G – quadruplex.....	110
Figure 61. Comparison of recognition of G – quadruplex structures by scRap1 and RHAU.	113
Figure 62. Schematic representation of possible role of G- quadruplexes as target in cancer therapies.	117

LIST OF TABLES

Table 1. Telomeric DNA sequence repeats in different species	17
Table 2. DNA oligonucleotides tested for co – crystallization with scRap1 DBD.	68
Table 3. Data collection and refinement for X-ray crystal structure determination.	70
Table 4. Peptides used in the studies:	120
Table 5. Oligonucleotides used in the studies:.....	121
Table 6. Strains used in the studies:.....	122

CHAPTER ONE: Introduction

1.1. G – quadruplexes and their roles in biology

The canonical structure of right – handed double helical DNA has been known since the Watson and Crick paper in 1953 [1]. However, it is becoming increasingly evident that DNA can adopt various alternative conformations depending on DNA sequence and interacting proteins [2]. The observation that guanine – rich nucleic acid sequences can self–associate into G– quadruplex (G4) structures dates back to the analysis of guanosine gels in early 1960 by Gellert and his co – workers [3]. Even though these unusual structures have been known for decades, it was mostly considered as a structural curiosity observed *in vitro*.

Further interest in G – quadruplexes was sparked by the discovery that oligonucleotides derived from biologically relevant regions of the genome such as telomeres can spontaneously form G – quadruplex structures in physiological conditions [4–6]. Over the last several years, technical advancement has significantly improved our understanding of G – quadruplex structures and helped to accumulate more direct evidence for the existence of these structures in cells. Furthermore, G-quadruplexes have been recognized as a possible mechanism for the regulation of biological process in living cells [7].

1.1.1. G – quadruplex structures

G-quadruplexes form *in vitro* from DNA or RNA sequences containing tracts of guanine nucleotides (G–tracts) separated by other bases. The basic units for G-quadruplex structures are four guanine nucleotides that associate

through Hoogsteen hydrogen bonding into a cyclic arrangement forming a G-tetrad. In this arrangement, each guanine base makes two hydrogen bonds with an adjacent G-nucleotide. Hoogsteen base pairing, in addition to the N¹, N² and O⁶ atoms involved in hydrogen bonding in Watson – Crick DNA base pairing, uses the additional N⁷ position of the purine base as an acceptor (Figure 1a,b). These planar G–tetrads are stabilized further by stacking on top of each other. The formation of these structures is driven by monovalent cations such as K⁺ and Na⁺, that neutralize the electrostatic repulsion of inwardly pointing O⁶ atoms in the G-tetrads by occupying the central cavity between the planar stacks [2].

Structural analysis using both X-ray crystallography and NMR spectroscopy have revealed that depending on how the guanine bases are arranged, G-quadruplexes can adopt different topologies [8]. The diversity is coming from different possible syn/anti conformations of the guanines, orientation of the DNA strand as well as the number of DNA strands involved in the structure formation. Intra–molecular folding requires only one single strand of DNA containing four or more G–tracts, whereas inter – molecular structures can arise from two to four strands. Depending on the strands orientation G-quadruplex can be further classified into three groups: parallel, antiparallel or hybrids (Figure 1c).

G-quadruplexes can be very stable, although this depends on several factors including the sequence composition of the G – rich strand: the number of guanines in G–tracks and the length of the loop between them, as well as the nature of the bound cation [9–11]. Usually, small loops and longer G-tracts result in more stable structures.

Figure 1. Structure and polymorphism of G-quadruplex. (a) Structure of guanine. Black arrows are indicating atoms involved in hydrogen bonds in Watson – Crick DNA base pairing. The red arrow indicates additional hydrogen acceptor used in Hoogsteen pairing. (b) The basic unit of G-quartet, (c) Schematic representation of intramolecular G4's: parallel and antiparallel, (d) Schematic diagram of the sub-grouped polymorphic structure of intermolecular G-quadruplexes. The arrows indicate the direction of the DNA strands.

1.1.2. Potential G – quadruplex location in the genome

Several approaches have been used to determine if and where G-quadruplex structures occur *in vivo*. The first evidence of the possible location and role of G-quadruplexes in cells was revealed by genome – wide sequence analyses. By searching the human genome sequence database with a general sequence motif $G_{3-5}N_L G_{3-5}N_L G_{3-5}N_L G_{3-5}$, where N_L stands for unknown loops with

the length limit of up to 7 nucleotides, it was shown in 2005 that the human genome sequence data contain over 300 000 sequences with the potential to form G-quadruplex structures (PG4s) [12,13]. More recently, the development of high throughput G4-seq method, based on a combination of next – generation sequencing together with G – quadruplex dependent polymerase stop assay, allowed the discovery of an additional 400 000 PG4 sites that were not predicted by bioinformatic tools alone [14]. It is also likely that such calculations are an underestimation as many repeated DNA sequences are missing from the sequence databases. Importantly, PG4s found in these studies are not located at random positions, but rather cluster at important functional regions of the genome. Furthermore, G-quadruplex motifs have been shown to be highly conserved between species implying a selection pressure to preserve such sequences and signifying their importance for cellular functions [15].

As expected, human telomeres consisting of thousands of tandemly arranged TTAGGG repeats showed the highest abundance of PG4s in the human genome. Outside of chromosome ends, PG4's have been shown to be rather depleted in most of the coding regions of genes but highly enriched in promoter regions [16]. Most interestingly, G-quadruplex motifs were found in promoters of important oncogenes such as RAS and MYC, which come under attention as possible drug target [17,18]. Moreover, recent development in the identification of origins of replication in the human genome using high throughput sequencing suggested that 67% of identified replication origins overlap with G-quadruplex motifs [19,20]. This number increased to more than

90% when using a less stringent calculations for PG4s search, allowing the loop length to extend up to 15 nucleotides.

Figure 2. Possible formation and role of G-quadruplex structures in cells. Promoters and telomeres are specially enriched in G4 motifs, therefore G-quadruplex formation could control transcription (a) and be important for telomeres capping mechanism (b). dsDNA is transiently separated during replication which could favor G- quadruplex formation (c). Messenger RNA is in a single stranded form, therefore G-quadruplexes can easily form and influence translation process (d). (Taken from Ref [3]).

Of course, it is highly unlikely that this enormous number of G-quadruplexes would exist at the same time *in vivo*. From the biological point of view, it is important to understand the dynamic equilibrium between G-quadruplex and other forms of DNA such as the duplex form. Genomic DNA is primarily in the double-helical form which is furthermore stabilized by being packaged by histones into chromatin. There are two exceptions when nucleic acids are in a single stranded form, and therefore could fold into G - quadruplex structures: G-rich overhangs at the end of eukaryotic chromosomes and transcribed RNA. Furthermore, opportunities to form G-quadruplexes from double stranded DNA arise in the biological processes when DNA become transiently single stranded such as during DNA replication, transcription or recombination (Figure 2).

1.1.3. G – quadruplexes and their biological functions

A breakthrough in proving the presence and function of G-quadruplexes structures *in vivo* came more than 10 years ago from studies on ciliate (*Stylonychia*) telomeres [21]. Back then it was well established that guanine rich telomeric sequences are able to form G-quadruplexes *in vitro*, but direct evidence of their role *in vivo* was lacking. The development of antibodies which specifically recognize the antiparallel conformation of G-quadruplex formed by the *Stylonychia* 3'overhang d(T₄G₄)₂, allowed the direct probing for G-quadruplexes directly in cells [22]. The result from these immunostaining studies showed that G-quadruplex structures are present at the macronuclear telomeres. Moreover, the observation that the structure is absent during DNA replication and that the G-quadruplex formation is cell – cycle regulated through the phosphorylation of a telomere end-binding protein (TEBP β) suggested that the G – quadruplex structure may act as capping mechanism for *Stylonychia* chromosome ends [21].

Development of a variety of experimental strategies following this research resulted in the accumulation of evidence for the occurrence and biological significance of G-quadruplexes in other organisms. Most of the *in vivo* research on G-quadruplex in humans comes from the development of ligands binding specifically to G-quadruplex, which allowed the testing of cellular responses to such molecules. When G-quadruplex – stabilizing agents like telomestatin or RHPS4 were added to mouse or human cells, they caused shelterin complex disruption resulting in telomere dysfunction. Specifically, the ligands cause delocalization of the single stranded G-overhang binding protein POT1, but not the double stranded DNA interacting partner

TRF2 [23,24]. Similarly, by adding another G-quadruplex stabilizing ligand TMPyP4, it has been shown that G-quadruplex formation can affect the expression of 1200 genes, 700 of which are known to contain PG4s [18,25,26]. This provides support for the hypothesis that G-quadruplexes can have a regulatory role in transcription. Although G-quadruplex ligands are highly selective for G-quadruplex structures, very often their affinity is in μM concentration range. This raises the question about their specificity in the cellular context. Therefore, it is possible that some of the observed defects arise from non-specific binding to double-stranded DNA [8].

Nevertheless, the recent development of G-quadruplex specific antibodies together with secondary and tertiary fluorochrome – labeled antibodies improved the signal sensitivity [27]. This technical progress permitted the visualization of the presence of G – quadruplexes in the genomic DNA of many human cell lines by immunostaining. Furthermore, this method allowed the demonstration that the occurrence of G-quadruplexes is cell – cycle regulated. As expected, the strongest signal from G-quadruplex foci were observed during S – phase when DNA replication takes place. Remarkably, only less than 20% of the G-quadruplex signal was localized to telomeres as judged by co – localization with TRF2. But the reason for this is not understood.

1.1.4. Proteins involved in the regulation of G - quadruplexes

For G-quadruplex formation to occur in the presence of the double helical DNA, the two strands of DNA have to be separated, and the single stranded form has to be stabilized which is an energetically unfavorable process. At the same time, the *in vitro* studies showed that once formed, G-quadruplexes can

be very stable, and that spontaneous unfolding would be very slow [7]. Given this, it is not a surprise that there is accumulating experimental evidence for proteins like chaperons or helicases to be involved in kinetic regulation of G -quadruplexes formation in cells and their role in biological processes.

Many studies are focused on helicases specializing in unwinding DNA or RNA G-quadruplexes. Helicases are a class of enzymes that play a major role in nucleic acid separation using energy from ATP hydrolyses. Several of these enzymes, that have been shown to unwind G-quadruplex *in vitro*, were discovered through mutations leading to genome instability and causing human pathologies, including BLM (associated with Bloom syndrome), FANCI (genetically linked to the bone marrow disease Fanconi anemia, and associated with breast cancer) or PIF1 (linked to increased cancer risk) [2,28,29]. Another interesting example of helicase involved in G-quadruplex resolving is protein discovered in *C. elegans* from genome integrity studies. DOG-1 which stands for “deletion of guanine rich DNA” was shown to be involved in G-quadruplex regulation after the observation that in cells lacking DOG-1, unusual phenotype of deletion of G-rich DNA throughout genome could be observed [30]. Because of similarity to DOG-1 as well as telomere fragility phenotype observed after adding G-quadruplex stabilizing reagent, the mammalian DNA helicase RTEL1 (regulator of telomere length 1) was implicated to be involved in G-quadruplex unwinding [31]. However, the direct biochemical conformation for this activity is still missing.

One of the best studied helicases involved in G-quadruplex unwinding is the PIF1 DNA helicase, which is conserved from bacteria to man. First, studies on *S. cerevisiae* scPif1 helicase showed that this enzyme is efficient

in *in vitro* unwinding of parallel intermolecular G-quadruplexes [32]. Subsequently, using chromatin immunoprecipitation, it was shown that scPif1 has more than 1000 binding sites in the genome of mitotic cells of which 10% overlap with G-quadruplex sequence motifs. This low number is consistent with the fact that Pif1 is a multifunctional helicase, but when the number of PG4s present in yeast genome are considered, this equals to 25% of the potential G-quadruplex motifs in the organism to be scPif1 associated [33]. Moreover, many genetic assays support the hypothesis that G-quadruplex structures form *in vivo* and their unwinding by scPif1 helicase is crucial for genome stability. First, scPif1– deficient cells show slower DNA replication and more double-stranded DNA breaks (DSBs). Furthermore, the G-quadruplex motifs in these cells show a high mutation rate, and the mutations remove the ability of the motif to fold into G-quadruplexes [33].

The possible role of proteins acting as chaperons to promote folding of G-quadruplex *in vivo* is much less understood. Most of the knowledge about G-quadruplex promotion by proteins comes from *in vitro* studies. Two decades ago, using a combination of biochemical and biophysical studies, the Rhodes group concluded that the main shelterin protein in *S. cerevisiae* – scRap1 is able to promote G - quadruplex formation *in vitro* [34,35]. But until now the consequences of this *in vivo* are not understood. Another recently discovered example is the yeast regulatory subunit of telomerase scEst1 [36,37]. It has been suggested that the G-quadruplex promotion by scEst1 in yeast cells is necessary for telomere length maintenance, but at the same time, scEst1 deficient cells do not show changes in telomerase recruitment.

The most convincing *in vivo* evidence for “G-quadruplex chaperons” was derived from studies on telomeres in ciliated protozoa. Telomeres in the macronucleus of *Stylonychia lemnae* are bound by the two proteins containing oligonucleotide binding folds (OB folds): TEBP α and TEBP β . TEBP α has been shown to directly bind specifically to telomeric DNA [38] and homodimerize [39]. However, for smaller TEBP β to be found at telomeric overhangs, it needs to heterodimerize with TEBP α in a DNA-dependent manner and it has been suggested to play important role in telomere capping [40]. TEBP β was one of the first protein shown to enhance the formation of G-quadruplexes by 10^5 to 10^6 fold *in vitro* [41]. More recent *in vivo* studies have shown that both TEBPs are necessary to cooperate to promote antiparallel G-quadruplex formation at telomeres [21,42]. Even though TEBP β alone is able to promote G-quadruplex formation *in vitro*, in the cells it requires TEBP α to be recruited to the correct sites. Furthermore, this interaction and G-quadruplex promotion activity is tightly regulated through phosphorylation of TEBP β in a cell cycle-dependent manner [21].

1.1.5. G-quadruplexes in RNA biology

Even though most of the reports in the literature have focused on DNA, the existence of RNA G-quadruplexes *in vivo* seems to be more inevitable. RNA molecules are single-stranded, so the formation of G-quadruplex does not need to compete with double-stranded DNA hybridization to the complementary strand. Furthermore, RNA G-quadruplexes have been implicated to be more stable than DNA G-quadruplexes [43]. Experimental studies in combination with bioinformatics approaches suggested that RNA G-quadruplexes

are important regulators of RNA post-transcriptional mechanisms as well as DNA processes [44].

Similarly, to their DNA counterparts, RNA G-quadruplex kinetic regulation in cells requires protein partners. However, up to date there is not much known about RBPs (RNA – binding proteins) that can modulate the G-quadruplex conformation, but it continues to be an active area of investigation. An example of a protein that was shown to interact with RNA G-quadruplex structures is activation-induced cytidine deaminase (AID). This enzyme, that plays a crucial role in immunoglobulin class-switch recombination (CSR), has been shown to interact with G-quadruplex RNA formed by germline switch RNAs. The binding to RNA is necessary for the recruitment of AID to the immunoglobulin switch (S) region of DNA in sequence – specific manner. Furthermore, it was shown that a single mutation in AID, that can cause defects in CSR and lead to diseases, impairs AID interaction with RNA and DNA G-quadruplex structures [45].

Evidence of RNA G-quadruplex unwinding in cells comes from studies on RNA helicase associated with AU-rich element (RHAU). Nucleocytoplasmic protein RHAU is a member of the DEAH-box helicase family and was shown to be able to bind and efficiently resolve G-quadruplex structures *in vitro* [46]. RNA co-immunoprecipitation experiment with RHAU has shown that the majority of RNA associated with this protein contains potential G-quadruplex motifs [47]. Furthermore, one of the most abundant target RNA was identified as TERC – human telomerase RNA subunit, and this interaction was dependent on G-quadruplex formation on 5'-region of TERC. These data together with the fact that knockdown of RHAU cause changes

in telomere length, strongly suggest that RNA G-quadruplex regulation is important for genome stability.

An increasing number of studies support the view that RNA G-quadruplexes are important regulators of biological processes. *Cis*-acting RNA G-quadruplexes have been shown to be critical in pre-mRNA maturation (splicing and polyadenylation), the transport of mature transcripts and the translation into proteins [48,49]. G-quadruplex motifs have been found not only in mRNA but in non – coding RNA's such as precursors of micro-RNA's [50] and lncRNAs [51]. This could suggest a potential role of RNA G-quadruplex structures in the regulation of gene expression *in trans*.

Figure 3. Possible mechanisms of the regulation of RNA G-quadruplexes. (a) RNA G-quadruplex recruit and interact with RBPs. (b) RNA G-rich sequence can exist in equilibrium between G4 and R-loop structures. (c) Interaction of protein with RNA G-quadruplex can influence the protein affinity to DNA structures.

The exact mechanism of the regulatory role of RNA G-quadruplexes remains elusive. It has been shown that in some cases protein interaction with

RNA G - quadruplex can influence the interaction with DNA [45]. In other cases, it seems that G-quadruplex hybrid structure can form by using guanines from both DNA and RNA molecules and influence telomere homeostasis [52] or transcription [53]. Moreover, it is possible that RNA structures *in vivo* exist in equilibrium between hairpins and G-quadruplex and influence gene expressions (Figure 3) [50].

1.1.6. Structural basis for G – quadruplex recognition

In recent years there has been a substantial effort by different researchers to understand the structural basis for specific G-quadruplex recognition by proteins. Of the few high-resolution structures solved, most did not show a direct interaction between the G-quadruplex structure and protein, but were rather dependent on the contact with duplex-quadruplex junction [54,55] or a loop binding [56].

The first information of G - quadruplex recognition by protein comes from NMR structure of RHAU specific motif (RSM) in the complex with parallel G-quadruplex determined recently and during the tie of this thesis writing [57]. As mentioned before RNA helicase associated with AU- rich element (RHAU) was shown to have ATP-dependent G-quadruplex unwinding activity with higher efficiency than resolving double-stranded DNA. While full length RHAU is necessary for G-quadruplex unwinding activity, the main interaction with a G-quadruplex RNA was mapped outside of helicase domain in the N-terminus region of the protein, known as RSM [58] (Figure 5a). The interaction with the G-quadruplex is via an α -helix formed by the RSM peptide with the face of G-tetrad through hydrophobic interactions further

stabilized by electrostatic interaction with the DNA backbone by positively charged amino acids (Figure 4d). It is interesting to point out here, that RSM helix forms only upon G-quadruplex binding, and its otherwise disorder as shown by proteolytic analysis [59]. Remarkably, the binding of the RSM domain resembles interaction of most known ligands targeting G-quadruplexes by binding to an exposed planar G-tetrad surface [57]. The mode of binding helps to explain the preference for parallel G-quadruplex binding, as in nonparallel G-quadruplexes the access to the guanine tetrads would be sterically hindered by the loops. It is worth noting that the G-quadruplex recognition is not necessarily associated with any specific G-quadruplex topology, but is rather depended on the accessibility of the G-tetrads surface.

Figure 4. NMR structure of RHAU specific motif (RSM) with parallel G-quadruplex. Side (a) and top (b) view of the structure (c) representation of residues in direct contact with planar guanines (d) details of three- anchor – points clamping G-quadruplex through electrostatic interaction (PDB ID: 2N21 [57]).

Even though, RSM domain of RHAU is essential for G-quadruplex interaction it is not sufficient for high affinity binding of the protein to G-quadruplex structures [59]. Furthermore, single – molecule experiments have shown that RSM presence is required for ATP-dependent G-quadruplex unwinding activity of RHAU, but the deletion of RSM does not abolish the binding to G-quadruplex DNA [60]. This result suggests the presence of another domain in RHAU protein responsible for binding to G-quadruplex DNA. Consistently, recent X-ray crystallography studies on bovine RHAU helicase have shown that additionally to RSM, an OB-fold – like subdomain in C-terminus part, folds back and helps to stabilize interaction with G-quadruplex, as well as interactions with single-stranded extension of nucleic acid [59]. Moreover, the interaction induces helical core rearrangements and cause nucleic acid binding channel formation, which is too narrow to accommodate dsDNA. This is consistent with previous biochemical studies showing the requirement of 3' single-stranded DNA/ RNA in front of G-quadruplex for helicase activity [61]. Interestingly, G – quadruplex observed in this co – crystal structure contains one non – canonical quartet consisting of A-T-G-G bases instead of the typical four guanines. This DNA reorganization seems to be caused by the association with the helicase domain of RHAU, which causes the pulling the last of guanine (G17) from the most bottom quartet into single-stranded region causing a shift of an A-T Watson Crick pair to the top quartet of G-quadruplex and from Hoogsteen bonds with remaining two guanines.

Figure 5. Crystal structure of bovine RHAU helicase in complex with G – quadruplex DNA. (a) Domain organization of bovine RHAU (b) top and side view of the color-coded representation of structure where pale green represents RSM fragment and additional OB-fold in contact with the G – quadruplex is in pink. (c) DNA sequence used for crystal structure. The underlined sequences represent the nucleotides forming G – quartets (PDB ID: 5VHE[59]).

1.1.7. G – quadruplex structures at telomeres

Telomeres are the specific protein/DNA complexes found at both physical ends of eukaryotic linear chromosomes. Due to the mechanism of DNA replication by DNA polymerase, the telomeric DNA ends cannot be fully replicated and therefore they gradually shorten with each cell division. Only the telomere dedicated reverse transcriptase, called telomerase, has the ability to elongate the 3' end of telomeric DNA by using its own RNA subunit as a template (Figure 6b) [62]. Telomerase activity is present in stem cells and during development, but it is inactive in most of normal human somatic

tissues. On the other hand, telomerase upregulation is observed in the majority of cancer cells, which results in immortalization and tumor growth [63].

Telomeric repeat	Group	Organisms
AGGGTT	Mammals	<i>Homo sapiens</i>
GGGTT	Ciliated protozoa	<i>Tetrahymena</i>
GGGGTTTT		<i>Oxytricha</i>
		<i>Stylonychia</i>
AGGGTTT	Higher plants	<i>Arabidopsis</i>
(A)G ₂₋₅ TTAC	Fission yeasts	<i>Saccharomyces pombe</i>
TG ₁₋₃	Budding yeast	<i>Saccharomyces cerevisiae</i>

Table 1. Telomeric DNA sequence repeats in different species. The sequence of the strand from 5' to 3' end. Adopted from [64].

Although the length of telomeric DNA in different species varies greatly, its sequence is highly conserved among species not only for their sequence composition (Table 1) but function as well. Typically, telomeres are double-stranded regions consisting of tandem repeats of guanine-rich sequences, where the G-rich strand extends in 3' direction forming a G-rich single stranded overhang.

In *S. cerevisiae*, telomeric DNA contains around 300 ± 75 bp of a somewhat heterogeneous sequence with the consensus C₁₋₃A/TG₁₋₃ [65]. Similarly, to other eukaryotes, *S. cerevisiae* telomeres end via a 3'- single stranded G-rich overhang. The length of this G-rich overhang change throughout the cell cycle: from around 12-15 nt through most of the cell cycle to 30– 100 nucleotides during late S/G2 phase [66]. G-overhang lengthening is not only controlled by telomerase, as this process is observed even in telomeres deficient cells, but also by C-strand degradation which is controlled by Cdk1p kinase activity [67].

yeast cells lacking of scCdc13 (*cdc13-1* mutant cells), G-quadruplex formation contributes to telomere maintenance [69].

1.2. Solving the end protection problem at telomeres

Telomeres containing exposed 3'- single stranded G-rich overhangs have the potential to be recognized as broken DNA and trigger DNA damage response mechanisms. Activating those signals could cause permanent cells arrest in order to “repair” chromosome ends by interchromosomal fusion or nucleolytic degradation, which would have a harmful effect on genome integrity [70].

The end – protection problem of chromosomes involving resistance to the DNA repair pathways is solved in humans and other eukaryotic species through the specific interaction of telomeric DNA with a complex of six proteins [71]. This complex which in humans was named shelterin, achieves its protective role through direct or indirect association with telomeric DNA and result in with the suppression of the action of checkpoint proteins and repair. In humans, the shelterin complex contains six proteins: three proteins which directly interact with DNA; TRF1 and TRF2 (Telomeric repeat binding factor 1 and 2) which bind to duplex DNA, and POT1 (Protection of telomere 1) interacting with the G – rich overhang (Figure 7). These three components are connected through two linker proteins: TIN2 (TEF-interacting nuclear protein 2) and TPP1. The additional 6th protein is Rap1 (Repressor activator protein 1) and it is part of the shelterin complex through its interaction with TRF2 [70].

Figure 7. Comparison of the shelterin complexes of (a) *Homo sapiens* and (b) *S. cerevisiae*. Rap1 in *S. cerevisiae* is main telomere interacting protein, whereas human homolog of Rap1 is on telomeres only through interaction with TRF2 and it does not directly interact with DNA.

S. cerevisiae together with divergent telomeric sequence contains a different composition of shelterin complex. In budding yeast, there is only one telomeric double-stranded DNA binding protein, scRap1, which binds with the very high affinity about every 18 bp [72]. Even though scRap1 shows little sequence similarity to the human telomeric DNA binding proteins TRF1 or TRF2 and has a different domain organization, DNA recognition is conserved. It binds to telomeric DNA using exactly the same DNA binding module called Myb, and it has the same DNA sequence recognition consensus [73]. ScRap1 protein is crucial for the recruitment of other components of the telomere capping complex like Rif1/Rif2 and Sir2/3/4 through interaction with the scRap1 C-terminal domain (RCT) [74,75]. Silent information regulator 2, 3 and 4 (Sir2/3/4) are important for telomere silencing whereas Rap1-interacting factor 1 and 2 (Rif1/Rif2) are required to prevent telomere elongation by telomerase

[76]. Instead of the single stranded binding protein POT1 observed in humans, in yeast Cdc13 protein is responsible for G-overhangs interaction. Although these proteins share very low sequence similarities, both use similar OB-fold scaffold for single stranded DNA recognition [77]. It is interesting to point out here that even though budding yeast are evolutionary very distinct from mammals and the telomere binding proteins are very different, the mechanism of both double and single stranded DNA recognition is highly conserved.

1.3. The role of budding yeast multifunctional DNA binding protein Rap1

ScRap1 (Repressor activator protein 1), originally known as TUF1, is the major telomere DNA repeats binding factor in *S. cerevisiae*, but it has remarkable functional versatility. This abundant nuclear protein was first discovered as a transcription factor regulating many metabolic genes [78]. Indeed only 10% of scRap1 is telomere associated [65], but telomeric sequences are among the highest affinity binding sites [79,80]. While bound to chromosome ends, scRap1 has been implicated in serving as “counting” molecule where the number of scRap1 proteins bound to telomeres functions as a readout for telomere elongation feedback loop dictating if a given telomere will be elongated in a given cell cycle [81]. Furthermore, besides controlling telomere length, scRap1 is involved in the silencing of genes near telomeres (TPE -telomere positioning effect), as well as protection from telomere-telomere fusion [82].

Figure 8. Domain organization of Rap1 in different species: (a) *Saccharomyces cerevisiae*, (b) *Homo sapiens*.

Consistent with its multi – functionality, scRap1 has a complex multidomain structure (Figure 8). As mentioned in the previous paragraph, its C-terminal domain (RCT) is essential for the recruitment of other proteins like Rif1/2 and Sir2/3/4 that are necessary for maintaining telomere structure and function. Located centrally the DNA binding domain, consisting of two Myb motifs, is responsible for interaction with DNA. This function is crucial for all the known roles of scRap1 and essential for cells viability [83]. Even though scRap1 protein has been extensively studied, the function of N-terminal BRCT domain, that is also conserved in other species like humans, is not understood and deletion of the first 340 amino acids of scRap1 does not result in a strong phenotype [83].

1.3.1. ScRap1 role in transcriptional regulation

ScRap1 is a transcription factor that has been shown to be able to either activate or repress transcription. This bimodal function, together with a wide range of target genes from glycolytic enzymes to ribosomal proteins, makes this transcriptional regulator essential for cell viability. Genome-wide studies have revealed that scRap1 targets 294 loci, that is about 5% of the yeast genome

in exponentially growing cells [84]. Despite the many years of studying scRap1 and advance in the field, an understanding of how this protein can efficiently accomplish both activation and repression of transcription remains elusive. One of the possible explanations could be through recruiting different interacting partners to different sites, which could be achieved by different post-transcriptional modification like phosphorylation.

Figure 9. Dynamics of transcription factor binding. (a) Domination of transcription factor – sites characterized by TF high affinity and low nucleosome occupancy resulting in high level of mRNA (b) Treadmilling. These loci have higher nucleosome occupancy and lower affinity for transcriptional factors, which cause their short residence time, but the occupancy observed by typical ChIP experiment is indistinguishable between a and b. (c) Domination of nucleosomes. Majority of the genome, where nucleosomes are deacetylated and there is high DNA-encoded nucleosome occupancy. Adopted from [85].

Recent studies have focused not only on occupancy of transcriptional factors on their theoretical binding sites across the genome, but understanding more the dynamics of these interactions [86]. It has been suggested that scRap1 has three different states of dynamic interaction with its binding sites and each of them controls the fate of downstream genes [87]. In this model, scRap1 long occupancy was linked to high transcription output, while treadmilling (fast binding turnover) was associated with the relatively low levels of mRNA (Figure 9 a,b). Furthermore, most of potential scRap1 sites were unoccupied as scRap1 has to compete for binding with nucleosome (Figure 9c).

Interestingly, telomeric and subtelomeric sites were showing fast turnover despite the high affinity for binding to those sites *in vitro*.

1.3.2. Mechanism of scRap1 DBD recognition of double-stranded DNA

ScRap1 was one of the first proteins discovered to specifically bind to double-stranded telomeric DNA repeats [88]. As described in previous paragraphs, scRap1 DBD contains two subdomains connected by a long flexible linker, and only after determining the crystal structure it was recognized that the two domains are structurally related to the homeodomain or Myb [89]. Interestingly, this tandem duplication of a domain is specific to Saccharomycotina, whereas only one copy is found in other yeasts and mammals (Figure 8), but in those organisms Rap1 lacks direct DNA binding activity and its localization on telomeres depends on interaction with other telomeric proteins. However, the same Myb domain is observed in mammalian telomeric DNA binding proteins TRF1 and TRF2, but the lack of tandem domains is replaced with TRFH domains which allow homodimerization [90]. Dimerization of TRF1 and TRF2, which guarantees the presence of two Myb binding motifs was shown to be necessary for high affinity binding to DNA and mutation in the dimerization domain causes delocalization of those proteins from telomeres [91].

Figure 10. Two views of the structure of scRap1 DBD in complex with telomeric DNA. The structure is color-coded, where Myb1 is shown in pale green, Myb2 – grey, two clusters of G•C base pairs which represent two telomeric repeats are shown in light blue (Adopted from ref [89]).

The minimal DBD of scRap1 was originally defined through a deletion analysis and is located between amino acid 360 and 596 [92]. The determination of crystal structure of this domain in complex with telomeric DNA revealed that it consists of two clearly defined homeodomains (Myb1 and Myb2) consisting of a helix–turn–helix motif (H–T–H) and a flexible C – terminal tail [89]. The homeodomains contain an N – terminal arms that specifically contact bases in minor groove followed by three – helical bundles, in which the last α -helix is the “DNA recognition” helix, making base specific contact in the major groove of DNA. Myb2 has an additional helix, that is likely important for maintaining the overall structure. The two Myb motifs bind to double stranded DNA in tandem orientations, where each structural unit contacts with an 8 base pair binding site making sequence specific interaction with G-steps. The relative positioning of those two domains is based on the interaction with DNA

(Figure 10), which is constrained by a long loop connecting C – terminal part of Myb1 with N – terminal arm of Myb2. The shortening of the loop by deletion of $\Delta 423$ -435 residues has been shown to be lethal for yeast. The importance of the loop was further confirmed by *in vitro* studies showing that scRap1 $\Delta 423$ -435 binds weaker to double stranded DNA than wild type [83].

Besides the interaction formed by the two Myb domains, a C – terminal tail of DBD wraps around the DNA helix increasing the surface area of the interaction by 25% [93]. Furthermore, this wrapping loop helps to stabilize the positioning of DBD on the DNA through interaction with Myb1 and locking it at a specific position [89,93].

1.3.3. Rap1 and its interaction with G4 *in vitro*

20 years ago, the ability of G-rich sequences to form G-quadruplex was considered as an interesting biochemical puzzle, rather than of real biological functional relevance. The interest was sparked by the demonstration that short oligonucleotides containing telomeric sequences can form G-quadruplex structures in physiological buffers, which in turn lead to the discovery of telomeric proteins being able to interact with this non – canonical DNA structures. This provided the first evidence for a possible biological function of G-quadruplexes [35,41].

The first research raising the possibility that scRap1 can interact with other forms of DNA than double helix, was inspired by an unexpected result from potassium permanganate sensitivity assay. This experiment have shown that double helical DNA acquires some single-stranded character upon scRap1

binding [72]. Even though this result was later explained through crystal structure showing that scRap1 caused local changes in sugar-phosphate backbone of dsDNA that allowed reaction with the potassium permanganate [89,94], it sparked the idea to test if scRap1 can bind to other forms of DNA. Subsequently, biochemical studies confirmed that indeed scRap1 can bind to single-stranded DNA, but the binding requires at least four GGG tracts [35]. The results presented in these studies showed that scRap1 binds to G-quadruplex DNA and that the full length scRap1 is required for the interaction.

Moreover, a methylation protection assay in combination with circular dichroism (CD) analysis showed that scRap1 not only interacts with G – rich DNA, but it promotes its G- quadruplex formation (Figure 11). The methylation protection assay is used to establish whether the guanine bases are involved in Hoogsteen hydrogen bonding, because dimethyl sulfate (DMS) reacts with the N⁷ groups of G-rich oligonucleotide. If the guanine base is involved in the G-tetrad formation, it is protected from the methylation reaction. The oligonucleotide annealed in 120mM KCl shows complete protection from methylation with or without the protein. However, in the lower salt concentration where G - quadruplex formation is not driven spontaneously, the guanines are not protected from methylation, but only after adding scRap1 the protection is detected again (Figure 6a). This observation strongly suggested that scRap1 promotes the formation of structures containing G-tetrads [34].

Furthermore, in the circular dichroism analysis where K⁺ ions were exchanged for NH₄⁺, that is considerably less efficient in driving G-quadruplex formation, the correct spectrum for parallel G-quadruplex was only observed

in the presence of scRap1. Interestingly, in contradiction to the biochemical studies, this biophysical study showed the G-quadruplex formation does not require full-length protein, but the DBD itself was sufficient (Figure 11b). This difference could be due to the fact that the DNA concentration required for CD analysis is higher than in the biochemical assays. It is also possible that DBD binds weakly to G – quadruplex DNA and dissociate in the electric field during the EMSA analysis.

Figure 11. ScRap1 binds and promotes G-quadruplex structures. (a) Methylation protection patterns in the presence (+) or absence (–) of scRap1 in 20 or 120mM KCl concentration. TE indicates the pattern observed for single stranded DNA. Box 1 to 4 represents the positions of G-tracts. (Ref [35])). (b) CD spectra of the effects of scRap1 and DBA on the structure of Ytelo oligonucleotides: — TE, – – TE / 50 mM (NH₄)₂S₀₄, – - – TE / 50 mM (NH₄)₂S₀₄/ 3.5 μM scRap1, . . . TE / 50 mM (NH₄)₂S₀₄/ 3.5 μM DBD (Ref [34]).

Despite the clear evidence that scRap1 can bind and promote G-quadruplex formation, the possible biological role of this activity is not well understood. It has been suggested that scRap1 could act directly on the G – rich overhangs of telomeres in late S – phase, when they are getting extended [95]. Another

possibility is that quadruplex formation could inhibit the activity of the telomerase enzyme and be crucial for negative telomere length regulation [35].

1.4. TERRA: telomeric repeat – containing RNA

Evidence in the field suggested that the ends of linear eukaryotic chromosomes – the telomeres are constitutive heterochromatin. Indeed, their low gene density together with the nature of epigenetic marks such as histone H3 trimethylation at lysine 9 (H3K9me3) or histone 4 trimethylation at lysine 20 (H4K20me3) support this notion [96]. However, recent studies with both yeast and mammals have demonstrated that telomeres are not as transcriptionally silent as was believed. Telomeric DNA sequence is transcribed into TELomeric Repeat containing RNA (TERRA) [97]. This non-coding RNA appear to be an integral component of telomeric chromatin, adding an additional level of complexity of how telomeres are maintained [98].

1.4.1. TERRA transcription and regulation

Long non-coding RNAs transcribed from telomeres are heterogeneous in length and sequence consisting of a combination of telomeric and subtelomeric sequences. The transcription starts from the centromere towards the end of the chromosome direction using the C-rich strand. The length of TERRA transcript can range from 100 to up to more than 9000 nucleotides in mammals, which is caused by transcription termination in multiple places within telomeric DNA [97]. The majority of TERRA is transcribed by RNA polymerase II (RNAPII). However, there is evidence of other polymerases such as RNAPI and RNAPIII being involved as well. As expected for most

of RNAPII transcripts, most of TERRA molecules in yeast and around 7% in human are polyadenylated [96,99].

Even though the regulation of TERRA transcription is not well understood, there are indications that telomere binding proteins are involved in this process. For example in budding yeast, the major telomere binding protein scRap1 uses different protein partners to control TERRA transcription and degradation [100]. At telomeres containing so-called Y-elements, scRap1 negatively regulates the levels of TERRA via Rif1 and Rif2 and through degradation with 5'-3' Ribonucleic acid trafficking exonuclease (Rat1). However, at telomere lacking these elements, scRap1 mediated repression of TERRA transcription requires recruitment of another interacting partner such as Sir 2/3 [100].

Similarly, in mammalian cells, TRF1 and TRF2 have been shown to be important for the regulation of TERRA abundance. Overexpression of TRF2 downregulates the level of TERRA transcripts. At the same time, this overexpression caused more open chromatin formation through an increased nucleosome spacing [101]. Furthermore, TRF1 si-RNA knockdown in human cells resulted in a decrease of TERRA transcription [102]. However, depletion of TRF1 in mouse fibroblast cells did not affect the level of TERRA transcript [103]. This conflicting observation could reflect the differences in the regulation of TERRA expression between the two species.

1.4.2. TERRA higher order structures

As already mentioned, TERRA is transcribed from the telomeric C-strand, therefore generating a G-rich RNA species. X-ray crystallography studies combined with solution techniques, strongly suggest that at least the 3' part

of human TERRA that consists of repeats of UUAGG sequence, similarly to their DNA counterparts, can form G-quadruplex structures [104,105]. Furthermore, such G - quadruplexes were found to be able to stack on top of each other possibly forming a higher – order structures. One of the models predicts that TERRA is arranged in “beads-on-a-string” manner, where each bead can contain either single or double unit of three – layered G-quadruplexes (Figure 12a) [106]. Formation of this kind of structure could give an additional level of specificity for protein recognition and help to recruit other components to telomeres.

Figure 12. Higher order structures of TERRA. (a) “Beads-on-a-string model”, where long non-coding RNA forms G – quadruplex structures along the molecule. (b) R – loop structure, where TERRA invade duplex forming RNA – DNA hybrid.

Moreover, recent studies have shown that endogenous TERRA can also form RNA – DNA hybrids at telomeres [107,108]. This so – called R-loops consisting three-stranded nucleic acid (Figure 12b), can cause replication fork stalling, triggering the formation of double – stranded breaks and hence pose a threat to genome stability. However, RNA transcribed from telomeres

in budding yeast, have been shown to cause accumulation of R-loops on critically short telomeres in order to prevent premature senescence [107]. Persistence of the RNA – DNA hybrids cause a DNA damage response, which activates homology-directed repair and allows to re-elongate short telomeres.

1.4.3. TERRA functions

Even though it has been almost a decade since TERRA was discovered, our understanding of its role and mechanism of action is very limited. First implications that transcription from telomeres might affect telomere length itself came from studies, where inducible galactose promoter was integrated into the end of one of the chromosomes in budding yeast [109]. Induction of transcription at telomeres causing telomere shortening was further confirmed by a similar system, where transcription of native TERRA was controlled by doxycycline regulatable promoter [110].

It has been proposed that this in *cis*-regulation of telomere shortening is caused by inhibition of the telomerase. Since the nucleotide sequence of TERRA is identical to the template region of the telomerase RNA component, it can act as effect of an inhibitor through competitive base – pairing. Indeed, *in vitro* studies confirmed that the addition of short RNA oligonucleotides mimicking TERRA sequence can inhibit telomerase activity[102]. However, in cells, other RNA interacting proteins may cover telomeric tract of TERRA and interfere with the competition in the telomerase nucleic acid binding site. For example, hnRNAP1 has been shown to be associated with TERRA in cells, and addition of this protein to *in vitro* assays prevented TERRA inhibition of telomerase activity [111].

Figure 13. Model of the involvement of TERRA in telomere lengthening. In telomerase-positive cells, TERRA associated with telomeres to guide it to short telomeres. When cells are lacking the reverse transcriptase enzyme, TERRA – telomeres hybrids are accumulating and triggering HDR response.

On the other hand, TERRA's ability to form RNA – DNA hybrids and triggering homology direct repair in some cases, led to speculation that this long non-coding RNA could be involved in alternative lengthening of telomeres (ALT). ALT is homologues recombination mechanism that allows cells to extend telomeres in the absence of the telomerase enzyme. The use of this mechanism for maintaining telomeres length has been reported for approximately 15% of human cancer cells [98]. Moreover, ALT mechanism has been shown to help overcome senescence in other organisms including yeast. Cancer cells, as well as yeast survivors using ALT mechanism, demonstrated upregulation of TERRA. Recent studies in *S. cerevisiae* suggest a model where TERRA in the presence of telomerase assist in guiding

the enzyme to short telomeres where TERRA was originally transcribed to promote their elongation. On the other hand, when cells are lacking telomerase, TERRA forms R-loops on short telomeres and activates HDR response in order to prevent premature senescence[107].

RESEARCH OBJECTIVES

Although scRap1 has been known to be able to promote G – quadruplex formation *in vitro* for last quarter of a century, an understanding of the consequences of this *in vivo* is lacking.

The aim of the work presented in this thesis is to understand the role of scRap1 interaction with G – quadruplexes structures formed by both DNA and RNA. Determining three-dimensional structure of scRap1 in complex with G – quadruplex DNA is key to reveal the molecular mechanism of this interaction and paves the way to address the role of G -quadruplexes in yeast. Moreover, such knowledge will help to understand how the same protein can adapt to interact with different DNA conformations.

Since scRap1 has been shown to be involved in telomeric RNA regulation in this thesis, I explore the possibility of scRap1 interacting directly with TERRA via G – quadruplex recognition.

CHAPTER TWO: Material and Methods

2.1. Expression and purification of recombinant ScRap1 constructs

2.1.1. Expression and purification of the DNA Binding Domain (DBD) of ScRap1

The gene fragment of ScRap1 (360–598aa) encompassing the DBD of Rap1 was cloned into pET30a TEV vector using the exonuclease III-based ligation-independent cloning. The construct, which contains an N-terminus His₆-tag followed by TEV protease cleavage site, was transformed and expressed in *E. coli* BL21-CodonPlus (DE3)-RIPL competent cells. Cells were allowed to grow at 37°C until the OD₆₀₀ reached 0.6–0.8, then cooled down and induced with 0.5 mM IPTG at 18°C for overnight expression. Cells from a 6l culture were harvested by centrifugation for 30 mins/ 4000 rpm at 4°C. The cell pellet was then re-suspended in lysis buffer (100 mM HEPES–NaOH pH=7.5, 0.5 M KCl, 10% glycerol, 5 mM 2–mercaptoethanol, 5 mM MgCl₂, 5 µg/ml DNase I, 1 mM PMSF, 10 mM Imidazole), sonicated for 5 mins at 60% amplitude with 10 sec on/off intervals. The lysate was cleared by centrifugation for 30 mins/18000 rpm at 4°C to remove cell debris. The first purification step was done by loading the supernatant onto a 5ml His-Trap HP column (GE-Healthcare) which was equilibrated in advance with lysis buffer. Next, the column was washed with washing buffer (100 mM HEPES–NaOH pH=7.5, 0.5 M KCl, 10% glycerol, 5 mM 2–mercaptoethanol, 20 mM Imidazole) until the UV absorbance reached a steady baseline. Then, the protein was eluted by increasing the Imidazole concentration up to 400 mM. The fractions were analyzed for protein purity by 12% SDS-PAGE. Fractions containing the DBD domain were pooled and incubated overnight with TEV protease (in ratio 1:50)

at 4°C to cleave off the His₆-Tag. Next day, digested DBD was loaded on a 5ml His –Trap HP column (GE –Healthcare) to remove the His-tag together with the protease, and the flow-through fractions were collected and concentrated to a final volume of 10ml in a Vivaspin MWCO 10kDa centrifugal concentrator. The protein was further fractionated by gel filtration chromatography. The column (Superdex 75 prep grade 16/60), was pre-equilibrated in gel filtration buffer (20 mM MES-KOH pH=6.0, 400 mM KCl). Fractions containing the DBD protein were pooled, concentrated, snap frozen in liquid nitrogen and stored at –80°C.

2.1.2. Expression and purification of the Myb domains of ScRap1

The fragments of DBD DNA encompassing the Myb1 (354-422aa) or Myb2 (450-598aa) were cloned into pET30a TEV vector. The induction of the proteins were done in 37°C for 6h with 0.5 mM IPTG. The first steps of purification for both constructs were conducted using the same protocol as for full DBD only omitting TEV cleavage site step. However, after purification by gel filtration with gel filtration buffer: 20 mM MES-KOH pH=6.0, 400 mM KCl, both constructs, but especially Myb1 were DNA contaminated judging by 260/280 ratio measured by Nanodrop. In order to remove DNA second round of gel filtration with Superdex 75 prep grade 16/60 was conducted but with buffer containing 2M KCl. Small amount of Myb1 that was recovered from this purification was still DNA contaminated. Fractions containing Myb2 were pooled, concentrated, snap frozen in liquid nitrogen and stored at –80°C.

2.1.3. Expression and purification of full length scRap1

The full length scRap1 gene was cloned into pET30a TEV vector, similarly to the DBD. The first steps of purification were conducted using the same

protocol as for the shorter scRap1 construct. As the purity of the protein after the affinity purification on the His-Trap HP column was unsatisfactory, the fractions containing full length scRap1 were combined and diluted in Heparin column binding buffer (100 mM HEPES–NaOH pH=7.5, 10% glycerol, 5 mM 2–mercaptoethanol) to a final concentration of 100 mM NaCl. Then the sample was loaded on 5ml HiTrap Heparin HP (GE-Healthcare) and the protein was eluted by increasing the concentration of NaCl up to 1 M. For further purification eluted fractions containing the scRap1 protein were pooled and diluted to 100 mM KCl concentration with dilution buffer (20 mM MES–KOH pH=6, 10% glycerol, 1 mM EDTA, 5 mM 2–mercaptoethanol). The sample was then loaded onto an ion exchange column (1ml HiTrap SP FF, GE-Healthcare) and the protein eluted by increasing concentration of elution buffer (20 mM MES–KOH pH=6, 1 M KCl, 10% glycerol, 1 mM EDTA, 5 mM 2–mercaptoethanol). The fractions containing scRap1 were pooled and re-loaded on 1ml His-Trap HP column and eluted as describe before with increasing concentration of imidazole. The fractions containing protein were collected and concentrated down to a final volume 500µl by Vivaspin MWCO 50kDa centrifugal concentrator and then further purified by gel filtration chromatography on a Superdex 200 prep grade 10/30 column pre-equilibrated with gel filtration buffer (20 mM Tris–HCl pH8.0, 300 mM NaCl, 5% glycerol, 5 mM 2–mercaptoethanol). Fractions containing scRap1 protein were pooled together, concentrated, and snap frozen in liquid nitrogen and finally stored at –80°C.

2.1.4. Expression and purification of N-terminal BRCT domain of scRap1

The first steps of protein expression and purification including the fractionation on the 5ml His-Trap HP column were carried out as described above (section 2.1.1). Next, the pooled fractions from the His-Trap column were diluted to a final concentration of 100 mM of NaCl and the buffer exchanged to 20 mM Tris-HCl pH=8.0. The protein was then further purified using a 1ml HiTrap Q HP (GE-Healthcare). The protein was eluted by increasing concentration of NaCl up to 1 M (20 mM Tris-HCl pH=8, 1M NaCl, 5% glycerol, 1 mM EDTA, 5 mM 2-mercaptoethanol). Eluted fractions were analyzed by 12% SDS-PAGE, and the fractions containing the BRCT protein combined and incubated with TEV protease (in 1:50 ratio) at 4°C overnight. Next, the protein was reloaded on the His-Trap HP column and the flow-through (FT) containing the cleaved protein was collected. The rest of the impurities were eluted with increasing concentration of Imidazole up to 500 mM. FT fraction was concentrated by Vivaspin MWCO 30kDa centrifugal concentrator, loaded on size exclusion chromatography (Superdex 75 16/60) and eluted with storage buffer: 20 mM Tris-HCl pH=8, 300 mM NaCl, 5% glycerol, 5 mM 2-mercaptoethanol.

2.1.5. Expression and purification of RCT domain of scRap1

The first steps of expression and purification of the RCT domain were conducted as described in the previous paragraph (2.1.1). After affinity purification on the His-Trap column, the protein purity was satisfactory, so fractions from the elution peak were pooled and concentrated (Vivaspin MWCO 10kDa). Size based fractionation was carried on a Superdex 75 16/60

gel-filtration column equilibrated in 100 mM HEPES–HCl pH=7.5, 300 mM NaCl, 5% glycerol, 5 mM 2-mercaptoethanol.

2.2. G-quadruplex DNA preparation

Desalted DNA pellets (from IDT) were dissolved in water at a concentration of 200 μ M. Concentrations were confirmed by measuring the absorbance on NanoDrop using the extinction coefficients provided by the company. Then DNA samples were diluted to 50 μ M in 10 mM Tris–HCl pH=7.5 and 150 mM KCl. G – quadruplexes were formed by annealing the DNA for 5 min at 95 °C followed by slow cooling to room temperature overnight.

2.3. Circular dichroism

Circular dichroism spectra were recorded at room temperature using JASCO-815 spectropolarimeter with 1 cm path length cuvette. Spectra from 220 to 320 nm were taken for 500 μ l solutions containing 5 μ M DNA prepared as described in 2.2. Each spectrum was derived from the average of three scans. The spectral contribution of the buffer was subtracted.

2.4. 32 P-labelling

5 pmol of each oligonucleotide was labeled using 1,5 μ l of [γ - 32 P]ATP (3000 Ci/mmol) and 4 U of T4 PNK (New England BioLabs) in 5 μ l reaction. The samples were incubated for 1h in 37°C.

2.5. Electrophoretic mobility shift assay and native gel electrophoresis

Binding reactions were performed at 25°C 30 minutes in binding buffer containing: 20 mM Tris–HCl pH=7.5, 1 mM EDTA, 1 mM DTT, 6% glycerol, 100 μ g/ml BSA and 100 mM KCl in final volume of 12 μ l. After the incubation, 10 μ l of the samples were loaded on a pre-run 7% polyacrylamide native gels

(acrylamide and bisacrylamide in ratio 37,5:1). Gel and electrophoresis buffer was 0.5×TB (Tris-Borate acid buffer). Electrophoresis was performed at 4°C for 1 hour with constant voltage (200V).

2.6. Crystallization and structure determination

An aliquot of DNA prepared as described in 2.2 was mixed with scRap1 DBD protein in 1:1 or 1:2 ratio in binding buffer (20 mM MES–KOH pH=6.0, 400 mM KCl) and incubated at 25°C for 1 hour. The complex was concentrated up to 15 mg/ml using a Vivaspin 6 with 10kDA cutoff. The Crystal Phoenix (Art Robbins Instruments) was employed to set up high-throughput sitting drops vapor-diffusion crystallization plates using commercially available screens (Qiagen, Hampton Research and Molecular Dimensions). The initial crystal hit was obtained in condition from PEGRx: 18 % v/v 2-Propanol, 0.1 M Sodium citrate tribasic dihydrate pH=5.5, 20 % w/v Polyethylene glycol 4,000. The conditions were further optimized in 24 well plates using an additive screen and detergent screen (Hampton Research). The final, best-diffracting crystals were obtained in crystallization buffer containing: 0.1M Sodium citrate tribasic dihydrate pH 5.6, 10% isopropanol, 14% PEG4000 and 10% of glycerol as cryoprotectant.

X-ray diffraction data was collected on 5.0.2 beamline at The Advanced Light Source synchrotron in Berkeley. The structure was solved by molecular replacement with PHASER [112], using scRap1 structure (PDB ID: 1IGN [89]) entry as the protein model probe and NMR solved structure of G-quadruplex with the same sequences (PDB ID: 2lk7 [113]) for DNA model. The model was further improved by several cycles of refinement in Refmac5 [114].

2.7. Strains and plasmid construction

LVPS-151 Rap1-13xmyc: By4741 MATa his3 Δ 1 leu2 Δ 0 ura3 Δ 0 met15 Δ 0 Est2-8G-myc18::TRP1MX6trp1::NATade2::tTAURA_ADH1 TetO7_CYC1_1L_TERRA Ku80-8G-3HA::HIS3MX6 strain was provided by J. Lingner [110]. scRap1 was then tagged using 13xmyc epitope from pFA6a-13Myc-kanMX6 plasmid and PCR ligated with leucine (LEU) as a marker from pRS315 plasmid. The correct integration was confirmed by PCR, followed by Southern blot with a probe detecting unique site in LEU marker.

Rap1-Flag,Gall-Rap1 Δ (583-596)-myc: By4741 MATa Rap1-3xFlag::KanMX, GallRap1-9xmyc:: HIS [87] was provided by M.E. Boeck. Haploid strains were mated with BY4742 strain, sporulated, dissected and genotyped to obtain spore clones of By4741 MATa Rap1-3xFlag::KanMX. Gall_Rap1-9xmyc was PCR from genome DNA and cloned into pRS403 plasmid. Rap1-9xmyc was mutated to Rap1 Δ (583 – 596) and then integrated by homologous recombination into the HIS3 locus. The correct strains were confirmed by PCR and Southern blot with a probe detecting unique site from pRS403 plasmid.

Δ rap1/Rap1: KanMX was PCR from pFA6a-3Ha-kanMX and flanked by Rap1 sequence. The YPH501 was transformed with PCR fragment and integrated by homologous recombination. The strains were confirmed by resistance to KanMX and PCR. Strains were further used to transform with different plasmids.

2.8. Western blot

Yeast strains were grown overnight in 5ml YEPD liquid at 30°C. Next day 1ml of the culture was collected by centrifuging the culture for 5 min at 4000 rpm. The cells were washed with water, resuspended in 200µl of 0.1 M NaOH and incubated at room temperature for 10 min. The supernatant was removed by spinning down the samples for 1 min at 13200 rpm and then the pellet was resuspended in 100µl of Laemmli buffer (50mM Tris pH 6.8, 2 %SDS, 10% glycerol, 0.01% bromophenol blue dye, 2.5% β-mercaptoethanol). The mixed samples were boiled for 10 mins in 95°C and 10µl of each sample was analysed on a 7.8 % acrylamide SDS-PAGE gel. In order to transfer gels to a Hybond ECL membrane (GE Healthcare, #RPN303D), gels were blotted for 2h at 0.21 Amp at 4°C in transfer buffer (25 mM Tris base, 190 mM glycine and 20% ethanol). Membranes were blocked overnight in 5% skim milk at 4°C, followed by incubation for 1h in room temperature with c-Myc mouse monoclonal antibody (Clontech, #631206, 1:400). After three 10 mins washes with TBST, secondary goat-anti-mouse IgG (H&L) horseradish-peroxidase (BioRad, 1:3000) was added and incubated for 1h at RT, followed by another 3 washes with TBST. Western blots were developed using the ECL membrane kit (GE Healthcare).

2.9. Southern blot

Yeast strain were grown overnight in 5ml YEPD media at 30°C. Next day 1ml of the culture was collected by centrifugation for 5 min at 4000 rpm. Genomic DNA was isolated using MasterPure Yeast DNA Purification kit (Epicentre Illumina company). Purified DNA was then digested overnight with appropriate restriction enzymes. Digested DNA was analyzed on 1% agarose

gel with EtBr in TBE buffer for 10 hours at 40V. The DNA in the gel was nicked using Strata-linker, followed by 20 min incubation in denaturing solution (1.5M NaCl, 0.5N NaOH) and 20 min incubation in blotting solution (1.5M NaCl, 0.25N NaOH). The gel was then transferred to Hybond-XL membrane (GE Healthcare, #RPN303S) for 24 hours. After transfer, the membrane was cross-linked again with Strata-linker and incubated for 5 minutes at 65°C with pre-heated Church's buffer (0.5 M phosphate buffer, 1mM EDTA, 7% SDS). Next, the membrane was incubated for 4 h at 65 °C with the radioactive probe that was labeled with Rediprime II or Ready to Go (Both GE Healthcare) random prime labeling kit and α -32P dCTP (Perkin Elmer, BLU513H). The blot was washed 3 times for 5 min with Blot Wash I solution (150mM NaCl and 15mM trisodium citrate pH 7.0, 0.1% SDS), followed by 2x10 min washes with Blot Wash II solution (15mM NaCl and 1.5mM trisodium citrate pH 7.0, 0.1% SDS). The membrane was placed on Phosphor Screen (Molecular Dynamics) for at least 24 hours and then visualized with Typhoon 9401 (GE Healthcare).

2.10. RNA immunoprecipitation (RNA-co IP)

Yeast strains were grown in 500ml YEPD media without doxycycline until the OD₆₀₀ reached 0.6 to 0.7 at 30°C. Cells were harvested by centrifugation for 5 min at 3500 rpm at 4°C, washed with ice –cold water, followed by wash with cold TMG buffer (10mM Tris-HCl, pH 8.0, 1mM MgCl₂, 10% (v/v) glycerol, 100mM NaCl; 0.1 mM DTT, 0.1 mM EDTA). Pellets were resuspended in 600µl of TMG buffer containing cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail and mixed with 500µL of glass beads. Cells were lysed using a Fastprep bead breaker (3x 1 min at level 6.5m/s). In order to separate cell lysate from beads two holes were made with 21 ½ gauge needles

and spin into another 2ml Eppendorf tube (1 min, 1000 rpm, 4°C). The extract was diluted in 1ml of TMG buffer containing 20U of RNasin and 20U of Suprase In and rotated at 4°C for 10min. Next the extract was cleared by centrifugation for 10 min at 5600 g at 4°C, followed by 4-5 rounds of centrifugation for 10 min at 16000g at 4°C. For immunoprecipitation, the same amount (as measured by Bradford) of the lysate was incubated with 10 µl of c-Myc mouse monoclonal antibody (Clontech, #631206, 1:400) and 0.5% TWEEN 20 overnight at 4°C. After addition of 80 µl Protein G agarose Dynabeads the incubation was extended for another 4 h. A mock IP lacking the c-Myc antibody, but with Rabbit IgG antibody (Bethyl Laboratories P120-101) instead and a strain lacking the myc-tag at scRap1 gene served as controls. 20 µl of the extract was taken as input sample. Immunoprecipitants were washed 3 times with TMG containing 0.5% Tween 20, followed by one wash with TMG itself. The beads were resuspended in 25µL TMG from which 7µL was taken for Western blot. Rest of dynabeads and input RNA were resuspended in 175 µL of Proteinase K solution (0.4 mg/ml proteinase K, 10mM Tris-HCl, pH 8.0, 10 mm EDTA, 1% (w/v) SDS and 100 mM NaCl) and incubated at 37°C for 30 min. The RNA after DNase treatment was extracted using phenol/chloroform method. RNA was then reverse transcribe iScript™ Select cDNA Synthesis Kit with 0.5 mM CA oligonucleotide (oBL207; TERRA), 0.2 mM ACT1 R (oNI212; actin) [110]. The cDNA was then quantified by real-time PCR using iQ™ SYBR Green Supermix in CFX96 Touch™ Real-Time PCR Detection System.

2.11. Chromatin immunoprecipitation (ChIP)

Yeast strains were grown in 50ml YEPD media at 30°C to the OD₆₀₀ of 0.2. Then 2% of galactose was added to each sample and grown for another 2h. In order to crosslink the samples, 1.4 ml of 37% formaldehyde was added while shaking and incubated for 5 min at RT, followed by addition of 2.5 ml of 2.5 M glycine and continued shaking for another 5 mins. Cells were harvested by centrifugation for 5 min at 3000 rpm at 4°C, washed with HBS buffer (50mM HEPES pH 7.5, 140mM NaCl), followed by wash with cold ChIP lysis buffer (50mM HEPES pH 7.5, 140mM NaCl, 1mM EDTA, 0.1% Sodium deoxycholate, 1% IGEPAL CA-630, 1mM PMSF). The pellets were resuspended in 400µl of ChIP lysis buffer containing cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail. 10µl of 5 mg/ml zymolyase 100T in sorbitol was added to the cells and incubated for 5 min in 37°C (or till >80% of the cells were lysed). The lysed cells were diluted with 500µl of ChIP lysis buffer + Protease Inhibitor and centrifuged down (20,8000xg, 4°C, 30min). The pellets were resuspended with 500 µl of ChIP lysis buffer + Protease Inhibitor, and sonicated by two series of 8 pulses (setting 3, 50% duty cycle). In order to remove residual insolubilities, the samples were centrifuged twice for 5 min at 100000 g at 4 °C and 450 µl of supernatant was transferred to the fresh tube every time. The samples were incubated for 1h at 4°C with 10µl of c-Myc mouse monoclonal antibody (Clontech, #631206) or 10µl of anti-Flag (M2, Sigma). After addition of 80µl Protein G agarose Dynabeads, the incubation was prolonged for 2 h. immunoprecipitants were washed twice with SDS buffer (50mM HEPES pH 7.5, 140mM NaCl, 1mM EDTA, 0,025% SDS), followed by singles washes with High Salt buffer (50mM HEPES pH 7.5, 1M NaCl, 1mM

EDTA), T_L buffer (20mM Tris-Cl pH 7.5, 250mM LiCl, 1mM EDTA) and T_E buffer (20mM Tris-Cl pH 7.5, 1mM EDTA). The beads were resuspended in 145µL TE + 1% SDS and incubated at 65°C for 2 minutes. After that, the samples were centrifuged for 1 min and 120µl of ChIP elute samples were collected. In order to reverse-crosslink all the IP samples, as well as input samples (with addition of 115µl of TE+1%SDS), were incubated at 65°C for 12 hours. The samples were purified with Gel Purification Kit (Qiagen) and yield with the final volume of 60 µL. The DNA was then quantified by real-time PCR using iQTM SYBR Green Supermix in CFX96 TouchTM Real-Time PCR Detection System.

CHAPTER THREE: Insight into G-quadruplex recognition by scRap1

Significant experimental evidence to support the notion that nucleic acid sequences containing tandem repeats of guanine tracts can form G-quadruplexes has been known for decades. However, only recent studies are slowly uncovering the *in vivo* presence and function of such structures in genomes and RNA's. This information, taken together with previous extensive biophysical and biochemical characterization studies on the properties of G-quadruplexes is most consistent with the interpretation that G-quadruplex structures provide a nucleic acid based regulatory mechanism for the regulation of the crucial processes such as replication, transcription and translation.

Soon after the sequence of telomeric DNA had been revealed [115,116], three group independently demonstrated that telomeric DNA repeats can form G – quadruplex structures in physiological salt conditions [4–6]. This was followed by studies showing that scRap1, a major telomere binding protein in budding yeast [88,117], is able to promote and bind to DNA G-quadruplex structures *in vitro* [34,35].

The aim of the work described in this chapter was to identify the domain of scRap1 responsible for G-quadruplex interaction by biochemical studies and to elucidate the three-dimensional structure of the complex. The structure of the scRap1 protein in the complex with a G-quadruplex should provide the basis foundation for understanding the recognition of G- quadruplexes by proteins. Furthermore, this knowledge should allow us to address the question of the importance and mechanism of this interaction in yeast cells.

3.1. Protein expression and purification

As describe previously scRap1 is multidomain protein, therefore to re-examine and establish which regions of scRap1 are responsible for G-quadruplex binding and promotion, various scRap1 constructs were expressed in *E. coli*. In previous work done by the Rhodes group [34,118] full length scRap1 and scRap1 DBD domain were expressed and purified, but their solubility was low. I found that lowering the temperature of bacterial growth and using affinity tag purification improved the purity of these proteins.

3.1.1. Expression and purification of full length scRap1

Full length scRap1 (1-827aa, MW=92.6kDa) protein was expressed in *E. coli* at 18°C. The slow growth in lower temperature was necessary to obtain soluble protein (data not shown). The first step of purification was based on the His-affinity tag. Cell pellets were re-suspended in lysis buffer and loaded onto a His-Trap HP column (Figure 14).

Figure 14. Purification of full length scRap1 by His-Trap HP column. (a) Elution profile of the His-Trap HP 5ml column. The initial high absorbance value represents all untagged proteins from lysate (b) SDS-PAGE analysis of His-Trap HP purification. Lane 1 shows the MW marker (Precision Plus Protein™ All Blue Prestained Protein Standard), T-total, S-soluble, FT- flow through, W-wash, lanes E1 to E3 show the fractions representing the peak in (a).

In order to improve the purity, fractions E1 to E3 from the elution peak were combined and diluted to lower the salt concentration and loaded on a HiTrap

Heparin HP column, which mimics DNA. The protein was eluted by increasing concentration of sodium chloride (Figure 15).

Since the heparin column chromatography did not significantly improve the purity of the protein, scRap1 was further purified by ion exchange chromatography. The fractions containing scRap1 were pooled, diluted to lower the salt concentration and loaded onto a HiTrap S HP column. The binding to this ion exchange column is based on reversible binding of the charged protein to oppositely charged groups attached to the matrix. The protein was eluted by increasing concentration of potassium chloride (Figure 16).

Figure 15. Purification of full length scRap1 by Heparin HP column. (a) Elution profile of scRap1 from the Heparin HP 5ml column. (b) SDS – PAGE analysis of different fractions. Lane 1 shows the MW marker, input – sample after His-Trap purification, FT–flow through, W–wash, lanes B7 to C4 represent the peak containing full length scRap1.

Ion exchange chromatography improved the purity, but the sample still contained additional smaller protein fragments, which suggested degradation (Figure 16b). In the next step, the sample was mixed with Ni-NTA agarose beads. The rationale behind this atypical step was that if the degradation products do not contain the N-terminal His₆-tag, they cannot bind to the resin and hence will be removed. This step helped to remove some of the

contaminants but did not remove degradation products (Figure 17), which suggests that most of the degradation products are due to either C-terminal truncations or internal cut in the folded protein.

Figure 16. Ion exchange chromatography for full length scRap1. a) Elution profile of scRap1 from the HiTrap S HP 1ml column with increasing concentration of NaCl up to 1 M. (b) SDS-PAGE analysis of different fraction. Lane 1 shows the MW marker (Precision Plus Protein™ All Blue Prestained Protein Standard), input-sample after heparin purification, FT-flow through, W-wash, lane A7–B10 show different fractions of the peak.

Figure 17. Purification of full length scRap1 by Ni-NTA agarose beads. SDS – PAGE analysis of nickel beads purification. Lane 1 shows the MW marker (Precision Plus Protein™ All Blue Prestained Protein Standard), input – sample after ion exchange purification, FT-flow through, W-wash, lanes E1 to E6 represent the gradual elution of scRap1 by increasing the imidazole concentration from 100 to 500 mM.

The final purification step was by size exclusion chromatography. The scRap1 protein eluted much faster than would be expected by its size (Figure 18a, b). However, scRap1 has been shown to form a ringlike or C- shaped

structure with a radius of 6–6.6 nm, that does not behave like a globular protein [119], which can explain the atypical elution behavior. Fractions B1 to B5 were pooled and concentrated to 1 mg/ml. Fractions B6 to B10 were discarded as they contained substantial amounts of degradation products.

Figure 18. Purification of full length scRap1 by gel-filtration chromatography. a) Elution profile of scRap1 fractionation on a gel filtration Superdex 200 10/300 GL column. b) Elution profile of MW standards from Superdex 200 10/300 GL column: 1 – Thyroglobulin (MW 669 kDa), 2 – Ferritin (MW 440kDa), 3 – BSA (MW 67 kDa), 4 – β -lactoglobulin (MW 35 kDa), 5 – Ribonuclease A (MW 13.7 kDa), 6 – Cytochrome C (MW 13.6 kDa), 7 – Aprotinin (MW 6512 Da), 8– Vitamin B12 (MW 1 355Da) [GE Helthacare Superdex 75 10/300 GL and Superdex 200 10/300 GL Protocol] (c) SDS – PAGE analysis of different fractions. Lane 1 shows the MW marker (Precision Plus Protein™ All Blue Prestained Protein Standard), input–sample after Ni-NTA agarose beads. Peak fractions B1 to B5 were combined.

3.1.2. DNA-binding domain (DBD) expression and purification

Similarly, to the full length protein, the scRap1 DBD (MW=35.4kDa) was induced and expressed at 18°C. The first step was to load the re-suspended cell lysate on a 5ml His-Trap HP column. The sample was eluted by stepwise

increasing concentration of Imidazole up to 400mM (Figure 19). In order to remove the His₆-tag fractions A2–A4 were pooled and incubated with TEV protease and incubated overnight at 4°C. The cleaved DBD was then re-loaded on the His-Trap HP column and the flow-through was collected. Contaminants, as well as TEV protease, were eluted by increasing concentration of imidazole (Figure 20a). Fractions from those two purification steps were analyzed together by SDS–PAGE (Figure 20b).

Figure 19. Affinity tag purification of scRap1-DBD. The black line represents elution profile of the DBD from 5 ml His-Trap HP column. The green line indicates the stepwise increase in Imidazole concentration.

The final purification step was done by Superdex 75 16/60 gel-filtration chromatography. In the same way as full length protein, DBD domain elutes faster than it would be expected for a globular protein of that molecular weight (Figure 21a, b). The first small peak (fractions B2 to B12) are probably aggregates, therefore only fraction from the main peak (fractions C1 to D5) were collected and combined.

Figure 20. Reverse His-trap purification of scRap1-DBD. a) Elution profile of the DBD from 5 ml His-Trap HP column. The green line represents the increasing concentrations of Imidazole. b) SDS – PAGE analysis of different fraction. Lane 1 shows the MW marker (Precision Plus Protein™ All Blue Prestained Protein Standard), T– total, S– soluble, FT–flow through, W–wash, lanes 5 (HisTrap A2-A4) represent the combined fractions from the peak from the His-Trap HP column (Figure 19).

Figure 21. Purification of scRap1-DBD by gel-filtration chromatography. (a) Elution profile of DBD from the gel filtration chromatography on Superdex 75 16/60 GL. (b) Elution profile of MW standards for Superdex 75 16/60 GL column: 1– Myoglobin (MW 17kDa) 2 – Ovalbumin (MW 44 kDa), 3 – Albumin (MW 66 kDa), 4 – IgG (MW 158 kDa) 5 - Ferritin (MW 440 kDa) [GE Helthacare Superdex 75 16/60 GL and Superdex 200 16/60 GL Protocol] (c) SDS – PAGE analysis of different fraction. Lane 1 shows the MW marker (Precision Plus Protein™ All Blue Prestained Protein Standard), input–sample after His-Trap HP, fractions C1 to D6 represent the fractions that were combined.

3.1.3. Purification of the BRCT domain

During the test expression and small scale purification of the scRap1-BRCT domain it was observed that the BRCT domain, which has an expected MW=45.3kDa migrates much slower on the SDS page gel than would be expected from its molecular weight (data not shown). In order to confirm the size of the expressed protein, samples after small scale beads purification were sent for mass spectrometry analyses. The analyses confirm that the purified protein is around the expected molecular weight of 45.5 kDa and hence BRCT has an unexpected migration on SDS-PAGE (Figure 22).

Figure 22. Mass spectrometry analysis for BRCT domain. The main peak corresponds to a protein with a size of 45.5 kDa.

The N-terminal domain of scRap1 – BRCT domain, similarly to other scRap1 domains, was induced and expressed at 18°C. The first step of the purification was His-Trap HP column chromatography (Figure 23). The fractions A6 to B3 from the peak were pooled. To improve the purity ion exchange chromatography was carried out. For this, the sample was diluted to lower the salt concentration and additionally the pH of the buffer was increased to 8.0 to ensure that the protein is in an environment that causes the molecule to be negatively charged and bind to anion exchange resin. Then,

the sample was loaded onto a HiTrap Q HP column (Figure 24). Unfortunately, a substantial amount of protein eluted during the wash step, and the purity of the protein did not improve. Fractions A6 to B1 were pooled, incubated with TEV protease and incubated over – night at 4°C. The cleaved protein was then loaded again onto His-Trap HP column and the flow-through fraction was collected, and contaminants were eluted by increasing concentration of Imidazole (Figure 25).

Figure 23. Affinity tag purification of scRap1-BRCT. a) Elution profile of BRCT from His-Trap HP column. The green curve represents an increasing gradient of Imidazole concentration from 15mM up to 500mM. b) SDS – PAGE analysis of different fraction. Lane 1 shows the MW marker (Precision Plus Protein™ All Blue Prestained Protein Standard), T–total, S–soluble, FT–flow through, W–wash, A5–B7 represent different fractions from the main peak.

Even though there is no clear difference on the SDS – page gel in the migration of the protein before and after TEV cleavage (Figure 25b), it seems likely that removing of His–tag was successful as most of the protein was in the flow-through fraction and hence did not contain the His-tag anymore. Additionally, a band can be observed in the sample after TEV treatment, which is of expected size for the cleaved His–tag (6 kDa).

Figure 24. Ion exchange purification of the BRCT. a) Elution profile of BRCT from Hi Trap Q HP column. The green line represents gradual changes of sodium chloride concentration form 100 mM up to 1M. b) SDS – PAGE analysis of different fraction. Lane 1 shows the marker (Precision Plus Protein™ All Blue Prestained Protein Standard), FT – flow through, W – wash, A6-B1 represent fractions that were pooled together.

Figure 25. Reverse His-trap purification of scRap1-BRCT. a) Elution profile of BRCT from His-Trap HP column after TEV cleavage. The green line represents changes in Imidazole concentration. b) SDS – PAGE analysis of different fraction. Lane 1 shows the MW marker (Precision Plus Protein™ All Blue Prestained Protein Standard), input – sample after ion exchange purification, After TEV – sample after overnight incubation with TEV protease but before purification.

In the final step, the BRCT sample was further fractionized by size exclusion chromatography, which helped to remove a lot of small impurities (Figure 26b). Fractions A6 to A11 were pooled together, concentrated to 4 mg/ml, snap frozen in liquid nitrogen and stored at -80°C.

Figure 26. Size-exclusion chromatography for scRap1-BRCT. a) Elution profile of BRCT from Superdex 75 16/60 chromatography. b) SDS – PAGE analysis of different fraction. Lane 1 shows the MW marker (Precision Plus Protein™ All Blue Prestained Protein Standard), input – sample after reverse His-trap purification, lane A6-A11 indicates fraction that were pooled.

3.1.4. Purification of RCT domain

The C-terminus domain of scRap1 expression was induced with 0.5 mM IPTG at 18°C and the culture grown overnight. In the first step of purification, which was His-Trap affinity purification, the purity of the protein was satisfactory (Figure 27). The fractions A5 to B5 and C1 to C3 were combined and loaded on Superdex 75 16/60 column. The profile from gel filtration shows two peaks but the analysis of the peaks on SDS-PAGE gels shows that both peaks contain the same size protein (Figure 28b,c). The fractions A12 to B7 from the second peak were pooled and the fractions from the first peak were discarded as it probably contains aggregates.

Figure 27. Affinity tag purification of RCT. a) Elution profile of RCT from His-Trap HP column. The green line represents the increasing concentration of Imidazole up to 500mM. b) SDS – PAGE analysis of different fraction. Lane 1 shows the MW marker (Precision Plus Protein™ All Blue Prestained Protein Standard), T–total, S–soluble, FT–flow through, W–wash, lanes A5 to B5 represent main peak. Fraction C1 and C3 are representation of second peak.

Figure 28. Purification of scRap1-RCT by gel-filtration chromatography. (a) Elution profile of RCT from Superdex 75 16/60. SDS – PAGE analysis of different fraction from first (b) and second (c) peak, input–sample after His-trap purification.

3.2. Oligonucleotides derived from yeast telomeric sequence can form parallel G – quadruplex structures in solution

Guanine rich DNA sequences in presence of ions form G-quadruplex structures which can be highly polymorphic [120]. Although the sequence of *S. cerevisiae* telomeric DNA is more heterogeneous than that of other organism, there are naturally occurring sequences of G-tracts that have been shown to form G-quadruplex structures in the presence of monovalent cations [34,37].

To validate the structure of G-rich DNA oligonucleotides to be used in the binding studies with various scRap1 domains, circular dichroism (CD) spectroscopy was employed. CD provides a simple solution analytical method that discriminates between single stranded DNA and G-quadruplex structures, as well as distinguishing between different G-quadruplex topologies such as parallel and anti-parallel strand orientations. Typically, parallel G-quadruplex structures will result in a CD spectrum with negative maximum at around 240 nm and a positive maximum at around 260 nm. Whereas, an anti-parallel G- quadruplex will show a negative maximum at around 265 nm and positive one around 285 – 295nm [121] .

In order to analyse the structure of yeast telomeric DNA fragments, the oligonucleotide 4G3 (AATTCTGGG TGTGT GGG TGTGT GGG TGTGT GGGTGTGG) was used as in the previous studies [35]. In order to remove any other secondary structure that could be formed during oligonucleotide synthesis and pre-form G-quadruplexes, the 4G3 DNA pellet was re-suspended in water or in a solution containing 150mM KCl and boiled in 95°C followed by slow cooling. Comparison of the CD spectra shows that

there is a clear change in the spectrum of 4G3 depending on buffer conditions (Figure 29a). As expected, from the role of potassium ions in promoting G-quadruplex formation, the spectrum of this oligonucleotide in the presence of K^+ ions show a maximum at 260 nm and hence indicates the presence of a parallel G-quadruplex. An additional maximum around 270nm can be observed, which could suggest the presence of other topologies.

Figure 29. Telomeric DNA from yeast forms parallel G- quadruplex. Circular dichroism (CD) spectroscopy analysis of 4G3 (a) and 4G3 no tail (b) G-quadruplex conformation. The blue and red curves indicate the samples in the absence and presence of 150mM KCl, respectively.

In order to obtain a more homogeneous and stable G-quadruplex, the sequence of 4G3 DNA was modified and many other related oligonucleotides were analyzed (data not shown). It was found that removing the single stranded tails from both 3' and 5' end of the DNA oligonucleotide (4G3 no tail) resulted in a more structurally homogenous sample of the desired parallel G4 structure (Figure 29b).

3.3. Identification of the domain of scRap1 responsible for interaction with G-quadruplex DNA

Previous results from the Rhodes and co – workers showing that yeast Rap1 can promote and interact with G – quadruplex structures, but were somewhat inconclusive on which domain of scRap1 protein is sufficient for these functions [34,35]. The early biochemical binding assays were suggesting that the DNA binding domain alone is not sufficient for G – quadruplex interaction, whereas biophysical studies using CD showed that DBD is as efficient in G – quadruplex promotion as the full length protein.

Figure 30. The BRCT and RCT domains of scRap1 do not bind to G-quadruplex DNA. Binding of N- terminus domain BRCT. (a) and C – terminus RCT (b) to 4G3 oligonucleotide with telomeric yeast sequence pre-formed to G-quadruplex. Concentrations of proteins were increased from 0.015 – 16 μ M with a two-fold step and incubated with a constant DNA concentration (4nM). The analysis was carried out on native polyacrylamide gels and developed with Syber Gold. The sequence of G-quadruplex oligonucleotide 4G3: 5'-AATTCTGGGTGTGTGGGTGTGTGGGTGTGTGGGTGTGG-3'.

To establish how scRap1 binds to the G – quadruplex and which domain is responsible for this interaction EMSA analyses were carried out. The scRap1 protein contains 3 distinct domains (Figure 8), which were separately purified as described in the previous paragraph. As earlier experiments suggested that

the DBD domain of scRap1 is not sufficient for G-quadruplex binding [35], the two domains of scRap1 were analyzed first: the N-terminal BRCT and the C-terminal RCT. No interaction with G-quadruplex DNA was observed with neither domain individually even at very high protein concentrations (Figure 30).

Having excluded that the N- and C-termini of scRap1 are involved in the G – quadruplex interaction, the binding activity of the DBD was analyzed. First, to confirm that the purified DBD is active in DNA binding, EMSA was carried out using telomeric double stranded DNA containing one scRap1 binding site (Figure 31). The estimated apparent K_d of DBD binding is about 180pM and it is similar to previously reported values [89,119]. Moreover, the active DNA binding domain can bind to 4G3 G-quadruplex (Figure 32a), and observed apparent K_d is comparable to previously reported for the full length protein ($K_d=30\text{nM}$) [35].

Figure 31. Purified scRap1 DBD is active and binds to double stranded telomeric sequence. Concentrations of protein were increased from 0.05 to 3.9 nM incubated with a constant (^{32}P) labeled DNA concentration (1 pM). Analysis was carried out on 7% native polyacrylamide gels. dsDNA was made by annealing 1C3 = 5'-GATCCACACCCACACACCG-3' with 1G3 = 5'-AATTCGGTGTGTGGGTGTG-3' oligonucleotide.

To further analyze the scRap1 binding to the G-quadruplex DNA, several G-quadruplex forming oligonucleotides were tested. The first binding assays were carried out with the yeast telomeric oligonucleotide 4G3 in order to compare with results from previous studies [34]. But since it was shown above (Figure 29a) that this oligonucleotide does not form a pure species of G-quadruplex, the same sequence without the 3' and 5' overhangs (oligonucleotide 4G3–no tail) was also tested for binding. Interestingly, DBD shows a comparable binding affinity ($K_d = 22\text{nM}$) to the G-quadruplex formed by oligonucleotide 4G3–no tail as compared to the 4G3 (Figure 32c). Moreover, a similar apparent binding affinity (2-fold difference) is observed for the HIV integrase aptamer (Figure 32d, T-loops), which is one of the most well studied sequence that forms an unimolecular and highly stable parallel G-quadruplex [122,123]. Since the sequence of the tested G-quadruplexes is different this already indicates that the main interaction of DBD with G-quadruplexes must be through interaction with planar guanines. Whereas, stronger affinity for 4G3 and 4G3 – no tail could be explained by the additional interaction of the protein with the loops.

Interestingly, at the higher concentrations of scRap1 DBD, other larger molecular weight complexes appear than the expected 1:1 (Figure 32). In principle, both 5' and 3' ends of parallel G-quadruplex have exposed G- tetrads that scRap1 could interact with. This suggests that the larger complexes observed at higher protein concentration consist of two protein molecules bound to one DNA molecule. However, as planar G-tetrads in G-quadruplex DNA fragments can stack on top of each other, another

possibility is that the formation of higher complexity is through DNA dimerization resulting in super-shift observed in gels.

Figure 32. ScRap1 DBD domain is responsible for protein interaction with different parallel G-quadruplexes. EMSA analysis for 4G3 (a), 4G3 – no tail (b), and HIV integrase aptamer – T-loops (d) where DNA was labeled with ^{32}P and the concentration was kept constant – 100pM. (b) Calculation of the apparent affinity of DBD for 4G3 based on experiment (a). The sequence of oligonucleotides used for binding are as follow: 4G3 5'-AATTCTGGGTGTGTGGGTGTGTGGGTGTGTGGGTGTGG-3'; 4G3 no tail: 5'-TGGGTGTGTGGGTGTGTGGGTGTGTGGG-3'; T-loops: 5'-TTGGGTGGGTGGGTGGGT-3'.

3.4. Characterization of scRap1 – DBD G-quadruplex promotion activity

As mentioned in the previous chapter, there are many possible sites in genomes that have the potential to form G-quadruplexes, but it is rather unlikely that all form G-quadruplex structures at the same time. Their formation is likely tightly controlled by proteins that promote their formation as well as helicases that unwind such structures [8].

In previous studies, scRap1 was shown to be able to promote G-quadruplex formation in the buffer condition (with 50 mM NH_4^+) that are less efficient in supporting G-quadruplex structure than K^+ ions. To further characterize the involvement of scRap1 DBD in G-quadruplex promotion and to confirm that observed changes are not due to potassium ions present in protein buffer, CD analysis was employed.

Figure 33. ScRap1-DBD promotes G-quadruplex formation of telomeric DNA. CD spectra of the effects of buffer with 150mM KCl (a) or scRap1 DBD in 150mM KCl buffer in 1 to 1 molar ratio (b) on 4G3 no tail (5'-TGGGTGTGTGGG TGTGTGGGTGTGTGGG-3') DNA through time.

Similarly, to the previous observation, it can be seen from CD spectrum that scRap1 DBD promotes G – quadruplex formation (Figure 33b). Addition of potassium cations shifts the CD spectrum to some extent towards the characteristic parallel G-quadruplex spectrum, but the ellipticity of the spectrum does not change significantly with time. On the other hand, the addition of scRap1 DBD with K^+ buffer shows the same CD spectrum at the time 0, but with time the ellipticity of the spectrum increase resulting in the typical

characteristic parallel G-quadruplex spectrum. Therefore, it can be concluded that scRap1 DBD promotes G-quadruplex formation.

3.5. Crystallization and structure determination of scRap1 in complex with G – quadruplex DNA

In order to understand how scRap1 recognizes G – quadruplex structures and to be able to compare how different that interaction is from binding to canonical double-stranded DNA, I decided to determine the structure of the scRap1 DBD in complex with G-quadruplex. As the complex is relatively small and purification of a large amount of DBD is possible, I decided to use X-ray crystallography method to determine the three dimensional structure.

Figure 34. Attempts to crystallize scRap1 DBD with yeast telomeric sequence. (a) Morphology of crystals containing 4G3-5'T DNA. Testing crystal content by (b) Syber gold and (c) silver staining.

Many attempts to crystallize the DBD in complex with G-quadruplex forming oligonucleotides derived from yeast telomeric sequences were unsuccessful, probably due to long flexible loops between the GGG-tracts (Table 2). Crystals observed in the initial crystallization trials with yeast telomeric sequence contained only DNA (Figure 34). In order to obtain crystals

of the complex, it was necessary to change G-quadruplex sequence to a more stable and homogeneous one formed by the HIV integrase aptamer containing very short loops between the G-tracts, consisting of a single thymine (T-loop, Figure 35).

Name	Sequence
4 G3	AATTCT <u>GGG</u> TGTGT <u>GGG</u> TGTGT <u>GGG</u> TGTGT <u>GGG</u> TGTGG
Telo - 4G3	TGT <u>GGG</u> TGTGT <u>GGG</u> TGTGT <u>GGG</u> TGTGT <u>GGG</u> TGTGG
4G3- 5'T	T <u>GGG</u> TGTGT <u>GGG</u> TGTGT <u>GGG</u> TGTGT <u>GGG</u>
4G3- 5'T 3'T	T <u>GGG</u> TGTGT <u>GGG</u> TGTGT <u>GGG</u> TGTGT <u>GGG</u> T
TT loop	TT <u>GGG</u> TT <u>GGG</u> TT <u>GGG</u> TT <u>GGG</u> TT
3xT loop	TT <u>GGG</u> TTT <u>GGG</u> TTT <u>GGG</u> TTT <u>GGG</u> TT
TGT	TT <u>GGG</u> TGT <u>GGG</u> TGT <u>GGG</u> TGT <u>GGG</u> TT
3'5'T TGT loops	T <u>GGG</u> TGT <u>GGG</u> TGT <u>GGG</u> TGT <u>GGG</u> T
3'5'TGT loops	TGT <u>GGG</u> TGT <u>GGG</u> TGT <u>GGG</u> TGT <u>GGG</u> TGT
T - loop 6-TGT	TT <u>GGG</u> TGT <u>GGG</u> T GGG T GGG TT
T - loop 10-TGT	TT <u>GGG</u> T <u>GGG</u> TGT <u>GGG</u> T <u>GGG</u> TT
T - loop 14-TGT	TT <u>GGG</u> T <u>GGG</u> T <u>GGG</u> TGT <u>GGG</u> TT
T - loop 10,14-TGT	TT <u>GGG</u> T <u>GGG</u> TGT <u>GGG</u> TGT <u>GGG</u> TT
T - loop	TT <u>GGG</u> T <u>GGG</u> T <u>GGG</u> T <u>GGG</u> TT

Table 2. DNA oligonucleotides tested for co – crystallization with scRap1 DBD. G-tracts are underline. The name of each oligonucleotide is derived from the loops between G3 tracts.

The initial crystals of the scRap1 DBD/G-quadruplex complex diffracted to 2.8 Å. Further optimization of the crystals resulted in better diffraction and the final structure was determined to 2.4 Å resolution (Table 3). Native diffraction data were collected on the 5.0.2 beamline at The Advanced Light Source Synchrotron in Berkeley. The structure was solved by molecular replacement with PHASER [112], using the available scRap1 DBD structure

in complex with double-stranded DNA (PDB:1IGN [89]) for protein model probe and parallel G-quadruplex (PDB entry:2lk7 [113]) for DNA model.

Figure 35. Morphology of crystals containing scRap1 DBD in complex with T-loop DNA.

Initial PHASER runs resulted only in the solution for a single Myb domain. Therefore scRap1 DBD model was separated into two PDB files containing separated homeodomains without the flexible linker between them and molecular replacement calculations were repeated. This separation allowed to find a solution for both Myb domains. The model shows the relative orientation of the two Myb domains in the scRap1 DBD/G-quadruplex structure is different to the orientation of these domain in scRap1 DBD model from the structure with double-stranded DNA. The model was further improved by several cycles of refinement in Refmac5 [114].

The asymmetric unit of the crystal contains one protein molecule bound to one G-quadruplex and the final model consists residues 361-483 and 515-568 of scRap1 DBD. Due to poor electron density probably caused by flexibility,

the C- terminus (569-590 aa) region as well as the loop between 484 and 514 residues are missing from the model.

Data collection and refinement statistics	Value
Data collection	
Space group	C 1 2 1
Cell dimension	
a,b,c, Å	84.92, 70, 64.81
α,β,γ , °	90, 97.6, 90
Resolution, Å	64.16 – 2.4
R _{merge} (%)	0.051 (0.622)
$\langle I \rangle / \sigma \langle I \rangle$	40.5 (2.7)
Completeness (%)	98 (97.8)
Unique reflections	14498 (2078)
Refinement	
Resolution Å	64.16 – 2.4
Number of reflections	13702
R _{work} / R _{free} (%)	0.24/0.28

Table 3. Data collection and refinement for X-ray crystal structure determination.

Using the self-assembly properties of DNA to drive the packing and crystal formation have been shown to be a good strategy to obtain crystals for X-ray structures of many proteins in complex with double-stranded DNA. Suitably design dsDNA oligonucleotides tend to form a pseudo-continuous helices that helps with crystal formation through DNA ends packing against each other [124]. As mentioned before G-quadruplex structures can stack on top of each other and form structure with increasing complexity. Interestingly, the crystal packing of scRap1 DBD/G-quadruplex complex have shown similar continuous

packing of DNA molecules but through parallel G-quadruplex stacking on top of each other (Figure 36). This may explain the observed super-shift in EMSA experiments, where the appearance of the additional band seems to be caused by homodimerization of scRap1\G-quadruplex through DNA. Therefore the band of the slower mobility observed in EMSA assay likely contains two protein molecules with two G-quadruplex DNA oligonucleotides, rather than the predicted one DNA molecule.

Figure 36. The unit cell of scRap1 DBD in complex with G-quadruplex DNA. The structure is color-coded, where Myb 1 is shown in pale green with darker green representing DNA recognition α -helix, Myb 2 in grey, blue represents guanines and dark yellow represents thymine loops between GGG-tracts.

3.6. Structural basis of parallel G-quadruplex recognition by scRap1-DBD.

The crystal structure helps to understand how the scRap1-DBD recognizes parallel G-quadruplex structures. In the structure, scRap1 DBD folds into two separate homeodomains as observed in the structure of DBD in complex with dsDNA (Figure 37a,b). However, in the scRap1 DBD\G-quadruplex complex, only Myb1 is in contact with the bottom (3') G-tetrad

of bounded G-quadruplex. Myb1 domain contains three-helix bundle, in which second and third α -helices form 'helix-turn-helix' motif. The mode of binding to the G-quadruplex involves third helix, known as DNA-recognition α -helix, that covers almost entirely the exposed hydrophobic face of the bottom of the G-tetrad (Figure 37c). The position of the α -helix permits ring stacking of G-12 with His-405 (Figure 37e). This residue together with Ser-402 and Val-409 create hydrophobic patch that increase the specificity of interaction with G-quadruplex DNA. Besides hydrophobic interaction, DNA-recognition α -helix dock onto DNA through hydrogen bonds formation with the ribose phosphate backbone by residues Thr-399 and Asn-401 with T-14, as well as Arg-406 contact with sugar – phosphate backbone of G-9 and T-10 (Figure 3c). Furthermore, Ser-402 besides hydrophobic interaction is important for creating hydrogen bond with phosphate backbone of G-13. This mode of interaction reminds of most ligands selected for specific G-quadruplex binding as well as RHAU helicase, which occur through electrostatic interactions with the phosphate backbone and stacking on the G-tetrads. Moreover, Myb1 interacts with G-quadruplex DNA through short N-terminal arm, that similarly to what is observed for interaction with dsDNA, enters the groove of DNA and Ser-362 forms hydrogen bond with the DNA backbone of the G-quadruplex loop.

ScRap1 DBD interaction with G-quadruplex DNA buries 608Å of the protein, among which 323Å comes from Myb1 DNA-recognition α -helix interaction, corresponding to 53 % of the interaction surface. This three-time smaller surface of interaction than observed for dsDNA [93] caused by not equal involvement of Myb2 explains the difference in affinity to dsDNA versus

G-quadruplex. However, that suggest that interaction with these non-canonical structures may be important for more transient interaction in the cell.

Figure 37. Global view of the structure of the complex between scRap1-DBD and the parallel G-quadruplex. Side (a) and top view (b) with color-coded structure, where Myb1 is shown in pale green, Myb2 in light blue and DNA recognition α -helices are highlighted by darker shades of the domain's color. Grey represents guanines and dark yellow represents thymine bases. (c) Representation of DNA-recognition α -helix of Myb1 covering the exposed 3' face of the bottom of the G-tetrad (d) N-terminal arm of Myb1 interaction with the groove of G-quadruplex (e) Details of intermolecular interactions between DNA – recognition α -helix (S402, H405, V409) in Myb1 and DNA guanine bases. (f) Representation of the amino acids interacting with the phosphate backbone of G-quadruplex.

Figure 38. Comparison of scRap1 DBD structures interacting with different DNA structures: dsDNA vs G-quadruplex. Superposition of the Myb1 motifs of DBD in complex with dsDNA (pink, PDB: 1IGN [26]) and G-quadruplex (green and blue).

To further understand how scRap1 recognize different types of DNA, the two structure were superimposed through Myb1 (Figure 38). The comparison between available structures of DBD in complex with dsDNA and structure presented in this work of DBD with G – quadruplex shows that even though both homeodomains are well defined and there is no global difference in their architecture, their orientation in respect to each other changes substantially. Previous studies have shown that even though scRap1 DBD contains two separate homeodomains connected by long flexible region, binding to diverse sequences of dsDNA does not cause global changes in domain arrangement, but rather subtle side chain movements [118]. On the contrary, binding to G-quadruplex DNA requires them to adopt a different spatial arrangement. This rearrangement does not allow for both domains to form protein-protein interactions, thus their relative positioning on G-quadruplex DNA is determined only by the interaction with DNA. The flexible

region together with N-terminal arm of Myb2 forms a clamp over the minor groove in the structure with double stranded DNA. However, the conformational changes in the structure while bound to G-quadruplex DNA do not allowed for this region to contact DNA.

3.7. Binding of DNA-recognizing α -helix peptides of Myb1 and Myb2 to the G-quadruplex

As described in the previous paragraph, the mode of G-quadruplex recognition is through α -helix of Myb1 of the scRap1 DBD and it occurs by the helix sitting on the planar face of the top of G-tetrad. On the other hand, in the dsDNA binding each of the two Myb motifs contributes to DNA binding and both DNA recognition α -helixes makes specific hydrogen bonds to the guanines in the major groove of DNA (Figure 10) [89]. That raises the question why would only Myb1 be able to bind to G-quadruplex, but not Myb2. Comparison of the amino acid sequence of the DNA recognition α -helix of the two domains shows some crucial differences (Figure 39a). In place of Val-409 in Myb1 important for a hydrophobic interaction with the G-tetrad, Myb2 contains a positively charged lysine. Moreover, histidine-405 in domain 1 is swapped into a negatively charged aspartic acid and hence domain 2 would not be able to make the same hydrophobic interactions.

The two amino acids difference between Myb1 and Myb2 could provide an explanation for why only one of the two Myb motifs in the scRap1 DBD interacts with the G-quadruplex structures. On the other hand, the same amino acids that are important for interaction with phosphate backbone for Myb1 with G – quadruplex are in the same position in Myb2 DNA recognition α -helix

(Figure 39a,c). The role of these amino acids was further confirmed by a mutation analysis showing that mutation of arginine in position 406 to alanine, cause more than 2-fold difference in binding affinity to G-quadruplex DNA (Figure 40).

Figure 39. Comparison of DNA-recognizing α -helix of Myb1 and Myb2. (a) Alignment of the sequences (b) superposition of the structures of Myb1 (pale green, from the structure with G-quadruplex) and Myb2 (in grey, from the structure with double-stranded DNA). The view of DNA – recognizing α -helix of Myb1 interaction with G – quadruplex, with amino acids important for interaction with the DNA backbone. Blue represents guanines and dark yellow represents thymine bases.

In the crystal context, where there is one G-quadruplex structure available only Myb1 is interacting with the G-tetrads. However, this does not exclude the possibility that Myb2 could also interact with G-quadruplex in different settings, for example where there are two G-quadruplex structures in tandem. In order to test if Myb2 can bind to G-quadruplex DNA, both domains were expressed separately in *E.coli*. Initial expression and small-scale purification of Myb1 and Myb2 showed that even though both of the domains are getting

expressed, most of Myb1 is in the insoluble fraction (Figure 41a). A large-scale purification resulted in more protein, but the obtained Myb1 was contaminated with DNA, and it was not possible to remove the nucleic acid even with extensive washes. Nevertheless, Myb2 domain did not present any problems with purification and solubility, hence binding assays could be performed. Interestingly, as shown by EMSA, Myb2 can bind to G-quadruplex DNA on its own, but as expected the observed affinity is significantly lower (320nM) than for the full DNA binding domain of scRap1 (Figure 41b).

Figure 40. The importance of interaction with the backbone of G – quadruplex. EMSA analysis for scRap1 DBD mutated from arginine in position 406 to alanine binding to 4G3 oligonucleotide: AATTCT GGG TGTGT GGG TGTGT GGG TGTGT GGG TGTGG. Concentrations of proteins were increased from 8 to 2000 nM with three-fold steps and incubated with a constant DNA concentration (100pM) that was labeled with (32 P).

As expressing Myb1 was not possible, in order to compare the binding of two Myb motifs to a G-quadruplex, peptides containing the DNA recognition α -helix from both domains were used in EMSA experiment instead. Even though, many conditions were tested, the peptide/DNA complex did not enter the gel which could be caused by the positive charge of peptides used for the studies. However, if the apparent binding affinity is judged by the disappearance of free DNA, the EMSA assays shows that both Myb motifs can interact with

the G-quadruplex. As it one might expect, the binding affinity of peptides alone for the G-quadruplex DNA is 40-50 times lower than the affinity measured for the DBD, but the Myb1 peptide interaction (1.2 μ M) is stronger than Myb2 (5.4 μ M) (Figure 42). This suggests that in the crystal context where there is only one G-quadruplex structure available, Myb1 as the one with the higher affinity will interact with an exposed hydrophobic G-tetrads. However, in the context where there is more than one G-quadruplex structure, both Myb domains would be able to interact with such a form of DNA.

Figure 41. Myb2 domain can interact with G-quadruplex DNA. (a) Small scale protein expression and purification of Myb1 and Myb2. Lane 1 shows the MW marker, T– total, S – soluble fraction, FT–flow through, W–wash, B – beads, E – elution. (b) EMSA analysis for Myb2 binding to 4G3 oligonucleotide: AATTCT GGG TGTGT GGG TGTGT GGG TGTGT GGG TGTGG. Concentrations of proteins were increased from 8 to 6000 nM with three-fold steps and incubated with a constant DNA concentration (100pM) that was labeled with (32P).

Figure 42. Comparison of two Myb motifs binding to G – quadruplex DNA. EMSA analysis for Myb1 and Myb2 peptides binding to preformed G – quadruplex (4G3). Concentrations of peptides were increased from 0.5 to 32 μM with two-fold steps incubated with DNA that was labeled with radioactive ^{32}P and the concentration was kept constant – 100pM. The analysis was carried out on 7% native polyacrylamide gels.

3.8. ScRap1 interaction with G - quadruplex

The first reports in the literature on the possible interaction of yeast scRap1 with G-quadruplex structure were based entirely on biochemical and biophysical data. These studies were somewhat contradictory in that in one study G-quadruplex binding required full length protein [34], whereas in the other the DBD was sufficient to promote G-quadruplex formation and binding [35]. However, it is worth pointing out that these experiments were not carried out in the same experimental conditions, as CD analysis required much higher DNA concentrations.

The studies presented in this thesis establish that scRap1 DBD alone is sufficient to interact specifically with G-quadruplex structures. The differences to previous results could be explained by different experimental conditions and different types of gels for the EMSA. It should be remembered that electrophoretic mobility shift assay has several pitfalls as it is a non-

equilibrium method. Under the condition used in this study, the apparent binding affinity of the scRap1 DBD to G-quadruplex is similar to the one reported in previous study for the full length protein [34].

The crystal structure of the scRap1 DBD in complex with G-quadruplex DNA presented in this thesis is one of the first structures showing the mechanism of this interaction (Figure 37). In the three – dimensional structure only one of the two Myb domains of the DBD makes a specific interaction with G – quadruplex DNA, even though both domains are structurally similar, and both interact with double-stranded DNA in a similar manner. On the contrary, the binding assay shows that the DNA-recognition α -helix from both Myb motifs can interact with G-quadruplex structures although at significantly lower binding affinity than full DBD alone. It is interesting to note that this could explain lower affinity of DBD to G-quadruplex vs dsDNA (in nM versus pM range, respectively) as we observe in these studies. In the binding assay to double-stranded DNA, both Myb motifs contribute equally to the measured affinity, while the affinity to G-quadruplex measured by EMSA is only for one of two the Myb domains.

The mode of binding to the G-quadruplex involves the same DNA-recognition α -helix that makes specific contacts with G-steps in the major groove of double helical DNA [89]. In the structure with G-quadruplex, the DNA-recognition α -helix covers the exposed hydrophobic face of the top G-tetrad, similarly the structure of peptide of RNA helicase associated with AU-rich element (RHAU) in complex with a parallel G-quadruplex [57]. The position of the α -helix permits hydrophobic interaction with valine 409, and ring stacking of guanine 12 with histidine 405. Additional stabilization

of the binding as for any DNA protein interactions comes from hydrogen bonds with the backbone of the DNA. The mode of interaction suggests the specificity for parallel G - quadruplexes as guanine tetrads in nonparallel structures are usually flanked by different loops, which would sterically hinder protein binding. However, it should be noted that this kind of interaction is also possible with nonparallel structures if the G – tetrads are exposed.

Figure 43. Model of scRap1 DBD interaction with DNA. (a) scRap1 interaction with double-stranded DNA can be through an equal contribution from both Myb motifs (high affinity binding mode) or it can form higher stoichiometry complexes on binding site through interaction with only one Myb domain and results in space for more molecules to bind. (b) Model of scRap1 interaction with G-quadruplex DNA. (I) In the crystal structure where there is only one G-quadruplex structure available, Myb1 is the one interacting with G-quadruplex. In cells, the G - quadruplex DNA will not be that secluded. Myb2 that has lower affinity to G-quadruplex DNA can either stabilize the interaction with DNA by interaction with dsDNA on the junction (II) or by interaction with another G-quadruplex structure in tandem (III).

Although most of the studies of quadruplex-forming sequences are focusing on single G - quadruplex structures, it has been shown that sequences consisting of a number of PG4 repeats like telomeric DNA can form higher-order

G-quadruplex structures [125]. Furthermore, when a G-quadruplex is formed on genomic DNA there must be a junction between dsDNA and G-quadruplex. Even though scRap1 has been well characterized for its interaction with double-stranded DNA, recent studies suggested that scRap1 is able to interact with double-stranded DNA in multiple binding modes, besides the one observed in published structures. It has been suggested that at lower binding affinity sites, scRap1 DBD binds by only making contact with DNA through one of Myb domains allowing more molecules to bind (Figure 43)[126]. This led us to propose the model where scRap1 can also have different modes of interaction with G-quadruplex structures. In case of the crystal structure presented here where there is only one G – quadruplex molecule, scRap1 is interacting only through Myb1 domain as it is the one with higher affinity for G-quadruplex. However, if the G-quadruplex structure were at the junction with double-stranded DNA, Myb2 could help in stabilizing the interaction by binding to dsDNA. On the other hand, if there were G-quadruplex structures formed in tandem, each of the two Myb domains could interact with G-quadruplexes by the same mechanism as seen in the crystal structure (Figure 39) with the DNA-recognition helix covering the exposed hydrophobic face of the G-tetrads (Figure 43).

CHAPTER FOUR: ScRap1 interaction with Telomeric repeat-containing RNA (TERRA)

As described earlier, scRap1 is a negative regulator of TERRA levels through transcription and degradation that are regulated by different scRap1 interaction partners and sequence elements of telomeric DNA. In budding yeast, there are two types of telomere associated sequences (TAS): X and Y'. X - elements are present at all telomeres and they are heterogeneous in size and sequence, whereas Y'-elements come in two sizes: short 5.2 kb or long 6.7 kb and in any given strain half of telomeres contains these elements [65]. ScRap1-dependent TERRA regulation is controlled differently depending on whether telomeres contain Y'-element or not. At telomeres containing Y'-elements, scRap1 seems to repress transcription of TERRA via interacting factors Rif1 and Rif2 and promote degradation through the 5'-3'exonuclease Rat1. However, experimental data have shown that at X-element only telomeres, TERRA level additionally requires the recruitment of the silencing proteins - Sir2/3/4 [100].

Moreover, it has been demonstrated with a combination of different biochemical and biophysical methods that human telomeric RNA, like many other G-rich sequences can form G- quadruplexes structures in the presence of ions such as K^+ or Na^+ [104]. We hypothesized that the role of scRap1 in TERRA regulation could be through direct interaction with this long G-rich non-coding RNA, possibly by recognition of G - quadruplex structures. The aim of the project described in this chapter was to test this hypothesis.

Figure 44. ScRap1 regulation of TERRA. (a) Regulation on telomeres containing only X elements: TERRA level is regulated by transcriptional silencing through Sir2/3/4 complex, as well as exonuclease Rat1 (probably controlled by Rif1 and Rif2). (b) Telomeres containing Y' elements do not use Sir proteins for TERRA regulation instead only the presence of Rif1 and Rif2 is required.

4.1. Telomeric RNA can form a G – quadruplex structure and scRap1

DBD can interact with it *in vitro*

A number of studies of the 3' region of the human TERRA sequences have shown that indeed, as predicted by bioinformatic tools, those sequences can form G - quadruplex structures *in vitro*. In budding yeast, the telomeric DNA sequence, as well as the RNA that is transcribed from it, are more heterogeneous at the sequence level than other telomeric sequences. It has an irregular length of guanine tracts as well as the loops in between them (Table 1). To define the structural features of telomeric RNA oligonucleotide derived from *S. cerevisiae* TERRA sequence, CD spectroscopy was employed.

Figure 45. Telomeric RNA similarly to telomeric DNA sequence from yeast can form parallel G-quadruplex. Circular dichroism (CD) spectroscopy analysis of DNA(a) and RNA (rTelo) (b) G-quadruplex conformation. The blue and red curves indicate the samples in the absence and presence of 150 mM KCl, respectively.

As expected the rTelo oligonucleotide spectrum in the presence of potassium ions present the typical parallel G-quadruplex spectrum with a maximum at 260 nm and minimum 240 nm and hence very similar to that of corresponding G-rich DNA strand (Figure 45). Interestingly and in contrary to DNA oligonucleotide with the same sequence, TERRA oligonucleotides resuspended in miliQ water shows G-quadruplex signature minima and maxima peaks but with lower amplitude. This suggests some G -tetrad formation, even in water which is further stabilized by adding K^+ ions.

As described in chapter 3, the main interaction of scRap1 with a G-quadruplex is through the DNA recognition α -helix with the planar guanines in the parallel G-quadruplex structure. We hypothesized that in this case, the binding affinity of scRap1 for the RNA G-quadruplex with the same nucleotide sequence as the DNA one should be comparable. In order to analyse the binding electrophoretic mobility shift assay (EMSA) was carried out with either the scRap1 DBD (Figure 46a), or a peptide containing the

DNA-recognition helix of Myb1 of scRap1 (Figure 46b). In both cases, as predicted, the apparent binding constant for the RNA G-quadruplex is very similar to the one observed for DNA.

Figure 46. ScRap1 binds to RNA G-quadruplex with the same affinity as to DNA G4 structures. EMSA analysis for DBD of scRap1(a) and Myb1 peptide containing DNA-recognition helix (b) to telomeric DNA or RNA oligonucleotides. Both DNA and RNA were labeled with (³²P) and the concentration of nucleic acids was kept constant. The sequence of oligonucleotides used for binding are as follow DNA (4G3 no tail): 5'-TGGGTGTGTGGGTGTGTGGGTGTGTGGG-3', RNA: 5'-UGGGUGUGUGGGUGUGUGGGUGUGUGGG-3'.

4.2. Can scRap1 directly interact with TERRA in yeast cells?

As shown above, scRap1 can bind to RNA G-quadruplex oligonucleotides derived from TERRA sequence *in vitro* with a similar affinity (in nM range) as it binds to DNA G – quadruplexes. Therefore, we wonder if this direct interaction also occurs in cells. As scRap1 has been already implicated in the

regulation of TERRA levels, we speculated that it could be through a direct interaction.

Previous studies have shown that detection of the endogenous level of TERRA in wild type cells in *S. cerevisiae* requires very sensitive techniques as TERRA levels are very low [99]. This is a considerable limitation for using TERRA as a target for co - immunoprecipitation experiments to identify direct interacting factors. In order to improve the detection of TERRA by Real Time quantitative PCR (RT-qPCR), an inducible 1L TERRA expression yeast strain was used [110]. The strain contains the tetracycline\doxycycline controlled transcriptional activation system, where a modified tetracycline transactivator (tTa) was integrated into *ADE2* locus. Additionally, Tet – operator boxes (TetO7) followed by a *CYC1* start site (Figure 47a) were integrated into the telomeric sequence of chromosome 1. In the absence of doxycycline, tTa binds to TetO7 sequences resulting in overexpression of 1L-derived TERRA.

To enable pull-downs of scRap1, I further modified the strain by introducing 13xMyc epitope to the scRap1 gene. This was achieved by ligating a PCR fragment from the pFA6a-13Myc-kanMX6 plasmid with the leucine gene (*LEU*) to introduce a marker. The correct integration was confirmed by Southern blot analysis (Figure 47b) and scRap1 expression with the correct tag in different clones was further confirmed by Western blot analysis probing with a C- myc monoclonal antibody (Figure 47c).

Figure 47. Tagging scRap1 with 13xmyc in an inducible 1L Terra strain. (a) Schematic representation of chromosome 1 of strain with inducible 1L TERRA expression system. The strain was created by inserting seven TetO boxes and promoter of cytochrome 1 with a URA3 marker and the ADH1-terminator (Ter). Overexpression of 1L TERRA can be repressed by adding doxycycline to media [110]. (b) Southern blot analysis testing the insertion of 13xMyc-tag in the yeast genome in different clones. The probe was designed unique site in leucine marker. (c) Western Blot analysis with C-myc antibodies to confirm scRap1-13xMyc expression. For both analysis wild type and Tet O7 – 1L strains were used as negative controls.

In order to test if scRap1 can interact directly with TERRA, RNA co-immunoprecipitation (RIP) experiments were carried out using the 1L – TERRA 13x Myc tagged Rap1 strain. RIP using C-myc antibodies results in enrichment of scRap1 together with bound nucleic acids. In the final step, the associated RNAs were purified and quantified by quantitative real time PCR. In the RIP experiment presented here, an untagged 1L – TERRA strain as well as immunoprecipitation without antibody served as negative control.

Figure 48. ScRap1 interaction with TERRA: Co- immunoprecipitation experiments where TERRA and actin were quantified by RT-qPCR (a) TERRA level normalized to the enrichment of immunoprecipitants (IP) over input. (b) TERRA level normalized using actin as a reference gene. Error bars represent standard deviations from measurements of three independent biological replicates. The significance of the data was determined using Holm-Sidak one-way Anova.

The normalization of RT-qPCR data can be carried out in various ways. One of the method for quantifying pull-down experiments is to normalize the enrichment of the transcript of interest pulled down with protein against the total cellular content of particular RNA transcript. Another way is to normalize the level of the RNA of interest against the reference gene transcript (Figure 48b). Using an endogenous and unrelated RNA transcript as an internal control can improve the reliability of RT-qPCR results. However, the correct selection of reference gene is crucial as any changes in “the normalizer” will produce artificial changes and errors in quantification of the transcript of interest. I decided to use actin as gene of reference, as the same gene was used in similar studies using the same yeast to show that TERRA is interacting with Ku proteins [110]. Furthermore, actin is one of the most common “housekeeping gene” that is used as a reference. Unfortunately, actin level was changing significantly between the samples, especially much larger amount of actin can be detected

in the negative control when immunoprecipitation was performed without antibodies (Figure 48b). That could be caused by the fact that actin transcript as one of the highly expressed in yeast cells, stick to unblocked beads and cause much higher background. Thus, actin transcript does not seem to be the correct reference gene for this experiment. Furthermore, using a mock antibody instead of empty beads can be better negative control as it is possible that the absence of antibodies cause higher background as beads are not blocked and hence more prone to bind unspecific RNA's.

Interestingly, if actin is not used for normalization, but rather if the enrichment of TERRA purified from immunoprecipitation (IP) over the total amount of TERRA (input) is calculated directly, the scRap1 – 13xMyc tag strain shows significant enrichment of TERRA transcript in comparison to both negatives controls: untagged strain and no antibody control (Figure 48a). Further experiments to confirm this result are necessary, but if this experiment can be reproduced it would mean that scRap1 binds directly to TERRA in yeast. This gives an interesting possibility of the additional mechanism of scRap1 involvement in telomeres maintenance.

4.3. ScRap1 regulation of TERRA

Telomere length homeostasis results from the appropriate balance between telomeres elongation and shortening processes. The interpretation from recent studies in budding yeast suggests a model in which TERRA is primarily involved in telomere elongation of critically short telomeres [107]. TERRA transcribed from normal or long telomeres is kept at a low level and further decays during the DNA replication process. On the other hand, at critically short

telomeres, TERRA transcription is increased which seems to be caused by partial telomere uncapping due to loss of Rif2 together with its interacting partner RNase H2. Upregulation of TERRA together with lack of RNase H2 which is responsible for R-loops removal, leads to replication machinery staling. Replication stress caused by an increase in R-loops formation of TERRA with telomeric DNA sources in DNA damage responses and triggers homology direct repair (HDR). This promotes re-elongation of short telomeres, and at the same time protecting cells from inducing senescence (Figure 49).

Figure 49. Model for TERRA mediated elongation of critically short telomeres. On long telomeres (top part of the panel) Rif2 recruits RNase H2 and Rat1 causing R-loop removal and TERRA degradation. On short telomeres (bottom part of the panel) TERRA accumulates and through the formation of R – loops promotes homology – directed repair causing telomeres elongation and delays of replicative senescence (based on [107])

Taken together, these observation with experimental data showing the involvement of scRap1 in TERRA level regulation and with studies presented in this chapter, it is thought-provoking to speculate that, when TERRA is bound to DNA on short telomeres and form R – loops, further part of this long non – coding RNA can directly interact and be protected from

degradation by one of the Myb domains of scRap1 DBD, whereas the other Myb domain could still remain in contact with double-stranded DNA (Figure 50). As mentioned before, scRap1 has been proposed to have multiple binding modes of interaction with DNA [126] (Figure 43), where on some binding sites, only a single Myb motif of the DBD makes contact with nucleic acid leaving the second Myb domain free for interaction with other DNA or RNA molecules. Importantly, the two Myb domains in scRap1 are connected by a long flexible linker (40aa) and their relative positioning is determined by binding to nucleic acids. Previous studies have suggested that this reorganization is important for the orientation of other scRap1 domains (Figure 8) in order to make them accessible to interacting partners [93]. ScRap1 interaction with TERRA could hinder the RCT domain and prevent the recruiting of Rif2, which consequently would cause TERRA upregulation from these telomeres. Moreover, this model is in agreement with previous studies where scRap1 has been implicated to serve as “counting” molecule in a feedback loop of telomere length regulations [81] and helps cells to distinguish between extendible and non-extendible telomeres by recruiting an appropriate number of its interacting partners Rif1 and Rif2 [127].

Figure 50. Cartoon model of a possible interaction of scRap1 with TERRA.

CHAPTER FIVE: The importance of the minimal DBD of scRap1

ScRap1 is an essential protein involved in the regulation of several important processes in the cells such as telomere maintenance and transcription. Most of its functionality is connected to its DNA binding domain. Using the information from the three-dimensional structures of scRap1 DBD in the complex with different DNA forms we wanted to create mutants that would allow separating the function: binding to double-stranded DNA versus binding to the G-quadruplex. The hope was that mutational analysis would enable us to test and understand the importance of the interaction of this protein with non-canonical DNA structures in yeast cells.

Figure 51. One of the examples of scRap1 DBD mutant in DNA recognition α -helix. EMSA analysis of mutation of hydrophobic amino acids in Myb1 that stack on top of G- quadruplex in the structure to larger and charged residues. The mutation does not show significant difference in binding to G-quadruplex DNA. Concentrations of proteins were increased from 8 to 2000 nM with three-fold steps and incubated with a constant DNA concentration (100pM) that was labeled with (32P). 4G3 oligonucleotide: AATTCT GGG TGTGT GGG TGTGT GGG TGTGT GGG TGTGG.

The strategy was to make point mutations in the scRap1 DBD that would reduce the binding affinity to G-quadruplex DNA without affecting the double-stranded DNA binding. In the first experiment the two essential amino acids making hydrophobic interaction with G-tetrad were mutated into larger and

negatively charged residues (H405E, V409E). However, this experimental approach failed as these point mutation did not show significant difference in binding affinity to G-quadruplex (Figure 51). In other experiments with other mutants a lower affinity to G-quadruplex structures resulted in reducing the affinity to double-stranded DNA as well (Figure 40, data not shown). It is important to note that it is possible that the EMSA analysis is not sufficiently sensitive to detect subtle differences between mutants and wild type DBD binding to G-quadruplex DNA.

Subsequently, I took the opposite approach to designing the mutants that would have a lower binding affinity to double-stranded DNA without affecting binding to G-quadruplex DNA. Previous deletion analysis have shown that the minimal DNA binding domain of scRap1 is centrally located, 235aa fragment spanning from residue 360 to 596 aa (Figure 52a). Furthermore, shortening of the domain by the last 13aa residues resulted in a loss of binding as judged by EMSA experiment [92]. The importance of this C-terminal tail is evident when examining the crystal structure of scRap1 DBD (360-598 aa construct) in complex with telomeric dsDNA. The structure contains electron density for the amino acid 587 to 594 residues, even though the density is missing for residues 579 to 586 that are connecting C-terminus tail with structured Myb2 and hence are disorder. This “C-tail” region of DBD has been shown to interact with DNA by entering the major groove, where it is involved in extensive interaction with cysteines, as well as forming a hydrophobic pocket around methyl group of thymine-9 of the DNA binding site. Moreover, amino acids on this loop fold back towards Myb1. This wrapping around DNA and interaction with Myb1 seems to clamp the protein in a specific position

on double-stranded DNA and at the same time influences the orientation of other domains of scRap1 that are important for its functionality [93]. On the contrary, this wrapping loop is not visible in the electron density map of the structure of DBD in complex with G-quadruplex. This suggests that this C-tail is disorder and not involved in interaction with the non-canonical DNA structures.

Figure 52. The importance of C- tail in the interaction with double-stranded DNA. (a) Combination of C- and N- terminal deletion of scRap1 to characterize minimal DNA binding domain of scRap1. Adopted from [92] (b) Color coded structure of scRap1 in complex with double-stranded DNA where Myb1 is shown pale green, Myb2 in grey, and the DNA is yellow with orange phosphate backbone. Pink represents C-terminal tail. Adopted from [89].

The aim of work presented in this chapter was to further characterize the interaction of scRap1 with DNA. Constructing a *S. cerevisiae* strain lacking “C- tail” of scRap1 DBD might give some insight into the role of the interaction of scRap1 with G-quadruplex *in vivo*. Moreover, the mutant strain will allow addressing the importance of a minimal DNA binding domain for yeast viability and proper functionality.

5.1. ScRap1 $\Delta(583-596)$ has a lower affinity to double-stranded DNA without affecting interaction with G-quadruplex

In order to test the importance of the C-tail of scRap1 DBD in DNA binding, scRap1 DBD $\Delta(583-596)$ and full length scRap1 $\Delta(583-596)$ proteins were expressed and purified from bacteria. Similarly, to other scRap1 constructs, lowering the temperature to 18°C during induction and expression was essential to obtain soluble proteins (data not shown).

Figure 53. ScRap1 DBD (360-583 aa) has a lower affinity for dsDNA than minimal DBD (360-596 aa) without changing the affinity to G - quadruplex DNA. EMSA analysis for dsDNA (a), and HIV integrase aptamer – T -loops: 5'-TTGGGTGGGTGGGTGGGTGTT (b) where DNA was labeled with (32 P) and the concentration was kept constant. dsDNA was made by annealing 1C3 = 5'-GATCCACACCCACACACCG-3' with 1G3 = 5'-AATTCGGTGTGTGGGTGTG-3' oligonucleotide.

Binding affinities of the two scRap1 DBD constructs were analyzed by EMSA. Contrary to what was defined as minimal DNA binding domain

in previous studies [92], the DBD $\Delta(583-596)$ missing the C- tail is able to bind to double-stranded DNA (Figure 53a), but the binding affinity is 5 fold lower than for the minimal DBD (360-596aa). In contrast, removing the last 13 amino acids of this domain does not influence the binding affinity to G-quadruplex DNA (Figure 53b). This is consistent with what was predicted from the crystal structures of the scRap1 DBD\G-quadruplex complex. Furthermore, these findings were confirmed for full length scRap1 (Figure 54).

Figure 54. Full length scRap1 missing C-tail of DBD shows no changes in binding affinity to G –quadruplex, but lower affinity to dsDNA in comparison to wild type protein. EMSA analysis for dsDNA (a), and HIV integrase aptamer – T-loops (b) where DNA was labeled with (^{32}P) and the concentration was kept constant.

5.2. ScRap1 Δ (583-596 aa) is not efficient in sustaining cell viability, but show the same toxic effect as wild type when overexpressed

Due to its multi-functionality, scRap1 is essential for normal growth and functioning of yeast. Previous studies have shown that both depletion or overexpression of this protein has lethal consequences for budding yeast [128]. Furthermore, it was previously shown that deletion of only 6 amino acid in the C-terminal region of DNA binding domain influences cell viability [129].

In order to test the viability of yeast with scRap1 Δ (583-596aa) protein in which the C-tail of DBD was deleted, diploid heterozygote strain was developed. In the heterozygote strain, one copy of scRap1 was exchange with KanMX module (Figure 55a,c) giving yeast resistance to gentamycin. Deletion of one copy of scRap1 did not influence the growth of yeast as shown by spotting assay (Figure 55b). The diploid yeast strain Δ rap1/ Rap1 was then allowed to sporulated on nutrient-poor medium resulting in meiotic division and formation of four haploid cells. Similarly, to mammalian cells, meiosis in yeast results in four genetically non-identical cells that can be separated under a dissection microscope [130]. Simple Mendelian inheritance predicts that alleles of each gene segregate from each other during meiosis and haploid cells carry only one allele of each gene. By this law, in the case of Δ rap1/ Rap1 strain, only two of the four haploids will contain scRap1 gene and another two will contain the KanMx module instead. As scRap1 is an essential gene, only these two spores can be observed growing on the plates, and they are not resistant to gentamicin which is in agreement with the Mendelian segregation (Figure 55c).

Subsequently, the yeast diploid heterozygote strain $\Delta rap1/Rap1$ was transformed with yeast centromere plasmid (pRS315) that contains scRap1 endogenous promoter followed by either the wild type copy of scRap1 or scRap1 $\Delta(583-596aa)$ mutant gene. These type of yeast vectors are based on the backbone of multipurpose pBLUESCRIPT plasmid and consist of Autonomously Replicating Sequence (ARS) together with yeast centromere DNA (CEN) [131]. In consequence, the plasmid does not get integrated into the genome, but it is replicated and treated by yeast cell as a small but independent chromosome. CEN vectors are considered to be stable without integration but they are lost during cell division with much higher rate than natural yeast chromosomes [132], so growing cells on selective media with appropriate marker helps to ensure plasmid preservation.

Figure 55. Developing heterozygote strain $\Delta rap1/ Rap1$. (a) One copy of scRap1 was exchange by homologous recombination with KanMX module. (b) Spot assay with wild type strain (YPH 501) and 3 different clones of heterozygote construct showing no difference in growth. (c) Dissection of $\Delta rap1/ Rap1$ on YEPD plate and replica plate 2 days later on gentamicin, confirming that Rap1 is an essential protein and only 2 out for tetrads are surviving, and none of them carrying resistance to gentamycin.

The strains $\Delta rap1/ Rap1$ transformed with appropriate vectors were then sporulated and dissected on selective media. After 2 days plates were replica plated on gentamicin. Similarly, to what was observed before for $\Delta rap1/ Rap1$, spores will not be able to grow on gentamycin, unless the protein transcribed from the plasmid is able to rescue the lack of scRap1 in the genome. Haploid cells from dissection of heterozygote strain $\Delta rap1/ Rap1$ were growing on plates with gentamicin, only if the haploids were transformed with the plasmid containing a wild type copy of scRap1, whereas the scRap1 $\Delta(583-596aa)$ mutant was not able to rescue viability of yeast lacking the endogenous copy of scRap1. Therefore, even though deletion of this 13aa C – tail resulted in a 5-fold lower binding affinity to double-stranded DNA *in vitro*, it seems to be required for correct scRap1 functioning *in vivo*. It is worth noting, that plasmid inheritance does not follow Mendel's law, so even in strains transformed with wild type scRap1, not all four sporulate cells were growing on selective media.

One of the intriguing features of scRap1 is that over-expression of the protein is toxic and leads to growth inhibition. The toxic domain has been characterized to 34aa starting immediately downstream of the DBD domain (597-630aa) and overexpression of this small fragment together with scRap1 DBD is sufficient to cause growth changes [128]. Even though the mechanism underlying this toxicity is not known, one of the most convincing hypotheses is that excess of scRap1 can lead to binding to lower affinity binding sites and inappropriate activation or silencing of transcription.

Figure 56. ScRap1 $\Delta(583-596)$ aa is not viable without wild type copy of Rap1 and it is not able to rescue the phenotype. Heterozygote strain $\Delta rap1/Rap1$ was transformed with the plasmid containing C-terminus HA-tag WT copy of scRap1 (a) or scRap1 $\Delta(583-596)$ aa (b) and dissected on –LEU plate (plasmid resistance) and replica plate on gentamicin plates after 2 days (resistance of heterozygote).

As shown in the previous paragraph, deletion of residues 583-596aa of scRap1 is lethal for yeast. This in combination with our *in vitro* study suggests that the mutant is not able to bind to dsDNA in cells or because of its lower affinity to dsDNA, cells would need more scRap1 to be viable. In order to test if scRap1 $\Delta(583-596aa)$ mutant overexpression can rescue cells by compensation for the amount of scRap1, or whether it is toxic to yeast cells as observed for overexpression of the endogenous copy, mutant and a wild type scRap1 were cloned into p424 plasmid. This episomal plasmid that was developed from natural yeast plasmid (2-micron plasmid) allows stable high copies propagation in cells. Additionally, it contains a GAL-1 inducible promoter, which can be activated by galactose in media.

All strains transformed with p424 plasmid display normal growth on glucose. However, induction of overexpression of either wild type or mutant scRap1 by galactose presents the same toxicity phenotype (Figure 57). Remarkably, that seems to suggest that scRap1 $\Delta(583-596\text{aa})$ is able to bind to DNA in the cells, but possibly overexpression together with lower affinity cause its binding to inappropriate sites and activation or silencing of incorrect target genes.

Figure 57. ScRap1 $\Delta(583-596\text{ aa})$ overexpression is lethal for yeast, showing the same phenotype as overexpression of wild type scRap1. Serial dilutions of cells were made for a control strain (By4741) transformed with empty plasmid or plasmid consisting of WT-scRap1 or scRap1 $\Delta(583-596\text{ aa})$ under the GAL-1 promoter. Cells were spotted onto plates under the indicated conditions and incubated for 2 to 4 days.

5.3. ScRap1 $\Delta(583-596\text{ aa})$ can bind and over compete wild type scRap1 when it is overexpressed

My analysis shows that scRap1 lacking the wrapping loop at the C-terminus of the DBD is binding with five fold lower affinity to double-stranded DNA *in vitro* and it is not able to keep yeast viable when endogenous scRap1 is missing. Taken together, these observations suggest that scRap1 $\Delta(583-596\text{ aa})$ is not able to bind to dsDNA in the cell and because of this, it cannot function properly. On the contrary, overexpression of scRap1 $\Delta(583-596\text{aa})$ presents the same toxicity phenotype as observed in the case of overexpression of wild type scRap1.

To test if scRap1 $\Delta(583-596 \text{ aa})$ can bind to DNA *in vivo*, a competition ChiP strain was developed using the same concept as presented in dynamic studies of scRap1 binding [87]. In this strain, the endogenous copy of scRap1 was tagged with 3xFlag tag by homologous recombination. The additional copy of scRap1 $\Delta(583-596 \text{ aa})$ mutant with 9x Myc epitope was integrated into the HIS locus with weakened galactose - inducible promoter GALL. In this setting wild type copy of scRap1 is expressed constitutively while scRap1 $\Delta(583-596 \text{ aa})$ can be induced as a competitor by adding galactose to the media. The two proteins can be easily distinguished by the two different tags (Figure 58a). As shown by Western Blot analysis, the level of scRap1 $\Delta(583-596 \text{ aa})$ increase with time after galactose induction, whereas there is no observed difference in endogenous scRap1 abundance through time (Figure 58b).

In order to test if scRap1 $\Delta(583-596 \text{ aa})$ can bind to genomic DNA *in vivo*, yeast strains were collected after 2h of galactose induction and chromatin immunoprecipitations were performed (ChiP-qPCR). ChIP experiments using antibodies against different epitopes was were used to separately purified and enriched wild type scRap1 as well as scRap1 $\Delta(583-596 \text{ aa})$ mutant protein with DNA associated with them. In the next step, DNA was purified and quantified. An untagged strain, as well as immunoprecipitation with mock antibody, served as negative control, and the level of DNA enrichment in the different sample was measured by quantitative real time PCR and normalized using ARO1 as a reference gene.

Interestingly, when the level of scRap1 $\Delta(583-596 \text{ aa})$ is much higher than the level of wild type protein as judge by Western Blot analysis, the mutant

protein not only can bind to telomeric sites tested in this experiment, but it is also able to over compete endogenous protein (Figure 58c). Furthermore, the binding was tested for scRap1 binding sites with the potential to form G-quadruplex structures that are known to be targets for Pif1 helicase (Zakian at all, unpublished). Similarly, to what was observed for telomeric sites, scRap1 $\Delta(583-596$ aa) can bind to such sites and compete with wild type scRap1 for the binding.

Figure 58. ScRap1 $\Delta(583-596$ aa) induced by GALL promoter can bind to DNA and compete with wild type copy of scRap1 for binding. (a) Schematic representation of ChIP competition strain. (b) Western blot analysis of scRap1 $\Delta(583-596$ aa)- myc induction after adding 2% galactose to the media. (c) Binding of scRap1 $\Delta(583-596$ aa) and Wt-scRap1 to telomeres by ChIP-qPCR experiment after 120 mins of induction with galactose (d) Binding to potential G – quadruplex site in the genome, that has been also shown to be bound by scPif1 helicase.

5.4. The importance of the C- tail of scRap1 DBD wrapping around DNA for yeast functionality

The DNA binding domain of scRap1 was demonstrated to be crucial for the functioning of scRap1 protein and even point mutation in it can cause cells death [79]. As discussed earlier, the crystal structure shows that the C-tail of the DBD contributes to scRap1 interaction with DNA [89,93], as it increases the surface area of interaction with nucleic acid by 25% [93]. Remarkably, deletion of the last 13 amino acids from C – terminal part of the DBD domain (C-tail), which results in only 5-fold difference in apparent binding affinity to double-stranded DNA *in vitro*, fails to compensate for the endogenous scRap1 loss in cells. Furthermore, even though the mutant yeast strain is not viable, scRap1 $\Delta(583-596 \text{ aa})$ protein can interact with DNA in cells as shown by ChiP-qPCR.

Taking all these results together, it is surprising that although the scRap1 $\Delta(583-596 \text{ aa})$ mutation does not abolish interaction with DNA *in vitro* as well as *in vivo*, it is not sufficient for rescuing yeast cell viability. Molecular dynamic stimulation studies together with the X-ray structures of scRap1 DBD in complex with different DNA's shows that the conformation of scRap1 is imposed by DNA binding and this adjustment is crucial for correct orientation of the N- and C- terminal domains of the protein [93]. Deletion of the C-tail of DBD which helps to position scRap1 correctly on DNA could influence that and sterically hinder the RCT domain, and at the same time blocking binding of interaction partners. Moreover, scRap1 $\Delta(583-596 \text{ aa})$ overexpression has shown the same toxicity phenotype that can be observed for wild type scRap1 overexpression. Since the toxicity phenotype can be observed

by overexpression of only scRap1 DBD with the following 34aa, this suggests that the toxicity phenotype is not due to scRap1 interacting partners [133].

On the other hand, even though deletion of full endogenous scRap1 was shown to be lethal for yeast, the phenotype could be compensated by expression of scRap1-DBD alone or even DBD from closely related species [83]. These results suggest that the DNA binding domain itself is critical for the function of scRap1, and interaction with other proteins is not necessary to restore viability. A possible interpretation is that the lack of the wrapping loop influences the dynamics of the interaction with DNA, rather than the interactions with other partners. Conventional ChIP experiment presented here give only the information on the occupancy of the protein on DNA sites. Recent studies have shown that measuring the dynamics of the interaction between proteins and nucleic acid can be more extrapolative to the biological function [86]. As described in the previous chapter, transcriptional factors can have three different modes of interaction with chromatin: more stable interactions that result in higher gene transcription, whereas “treadmilling” is observed on sites that resulted in lower transcription level. Depletion of C-tail of DBD influences the DNA clamping, which likely result in changes of the kinetics of the interaction and resulting in an incorrect level of transcription of target genes.

CHAPTER SIX: Discussion, conclusion and future directions

6.1. G-quadruplex research a century later: from guanine gels to biological functions

The first indication that guanine rich DNA can somehow self – assemble and form gels was shown at the beginning of the last century [134]. It took another 50 years to reveal the chemistry of guanines interactions and the formation of structures now known as G – quadruplexes. Since Gellert and his co-workers proposed G-quadruplex formation [3], this non-canonical structure of DNA has been the focus of much interest but mostly from biophysicists and chemists. This structural curiosity emerged as possible nucleic acid based mechanism in cells in 80s and 90s, when several biologically important genomic regions like telomeres and promoters were shown to form G-quadruplex structures in physiologically relevant conditions [4–6], as well as the discovery of proteins that are able to unwind, or promote G-quadruplex formation [35,135,136].

Although in the past decade, the field of G – quadruplexes has been blossoming, their heterogeneity together with possible transient existence in cells, research into the presence and function of G – quadruplexes *in vivo* continue to be a great challenge. However, as pointed out by Aron Klug over 30 years ago: “If G – quadruplex structures form so easily *in vitro*, Nature will have a way to use them *in vivo*”. Indeed, looking at the accumulating biophysical studies on the properties of G – quadruplexes, it is difficult to imagine that cells would not have found mechanisms to exploit such structures on guanine rich sequences for functions, or even at least find

the mechanism to protect from G- quadruplex interference with normal cellular processes.

Bioinformatic analysis in combination with experimental data have found at least 700 000 DNA sequences with the potential to form G – quadruplex structures in the human genome. Initial experiments of mapping DNA and RNA G-quadruplexes in cells are drawing a picture in which only a minority of the predicted PG4's structure are detected by chromatin immunoprecipitation [137] or RNA G4-seq experiments [138]. Even though those numbers are relatively small, they are consistent with the prediction that careful control of G-quadruplex dynamics between folded and unfolded state would be crucial for the involvement of G-quadruplexes in multiple regulatory functions. A better understanding of proteins controlling the formation and resolution of G-quadruplexes structures is necessary for the advancement of the field.

6.2. ScRap1 interaction with the G-quadruplex structures

ScRap1 is a multifunctional protein that was first discovered through its role as a transcription factor. Only later it was connected to telomere maintenance and it was one of the first protein to be shown to bind specifically to telomeric dsDNA repeats [139]. All the key roles of this essential protein appear to be controlled through its interaction with DNA. The current model for how scRap1 recognizes DNA is largely based on available three-dimensional crystal structures, where scRap1 DBD consisting of two Myb motifs binds to telomeric dsDNA site containing two G-steps. Besides using two DNA binding motifs, DNA binding is further stabilizes through a flexible linker between the two structured domains that enters the minor groove of DNA

and with unstructured C- terminal tail that wraps around DNA and folds back to the Myb1 domain.

Almost a quarter-century ago, it was proposed from biochemical and biophysical studies that scRap1 not only binds to duplex DNA, but also to G-quadruplex structures that can form from guanine rich DNA sequences. An understanding of how scRap1 recognizes G-quadruplex structures is presented in this thesis through structural studies. The comparison between available structures of DBD in complex with dsDNA and structure presented in this thesis of DBD with G – quadruplex shows how the protein can adapt its conformation to recognize different DNA structures using the same domains. Although the same DNA recognition α -helix of Myb1 is used to recognize and bind to double stranded and G-quadruplex DNA, the detailed nature of the interaction shows some crucial differences. First of all, the term “recognition helix” in H-T-H motifs describes the helix that interacts with DNA by entering the major groove of DNA where is able to make specific hydrogen bonds with DNA bases. Therefore, this interaction in the major groove is crucial for protein sequence specific recognition [140]. On the other hand, the same helix in complex with G-quadruplex DNA stack on the top of planar G-tetrads through hydrophobic interactions, and the recognition is mainly based on G-quadruplex structure itself rather than the nucleotide sequence.

A detailed analysis of the DBD interactions with different DNA structures shows that in both cases residues Thr-399 and Asn-401 contact the phosphate group of the same nucleotide base as well as Arg-406 contact sugar – phosphate backbone. Furthermore, His-405 increase the specificity to double-stranded DNA through hydrophobic interaction with an exposed cysteine, whereas

in interaction with G – quadruplex besides hydrophobic interaction, it helps to stabilize binding through ring stacking with guanine-12. This hydrophobic interaction with planar G – quarter is additionally increased by Val- 409 that is not involved in duplex DNA interaction and Ser-402 (Figure 59).

Figure 59. Comparison of scRap1 DBD – DNA contact in structure with dsDNA and G – quadruplex. (a) Representation of Myb1 interaction with double stranded DNA. (b) Schematic representation of G – quadruplex DNA bound to Rap1 – DBD in crystal structure. (c) Representation of Myb1 interaction with G-quadruplex DNA. In both representations arrows are indicating direct hydrogen bonds between protein and DNA. The arrows with round heads are indicating hydrophobic interactions. Broken arrows show water mediate interactions. The broken line represents ring stacking. The green color amino acid showing amino acid involved in interaction with both DNA structures, black one represents residues involved in recognition of dsDNA, whereas blue residues represent amino acid involved in G-quadruplex interaction. Histidin-405 was marked blue in G-quadruplex recognition because of different nature of interaction with DNA than observed with dsDNA through ring stacking rather than interaction through water.

Even though crystal structures provide a lot of information about the complexes and possible interactions between atoms, it is only a “snap shot” of a single conformation of the protein, and do not give information on dynamics of the interactions. It is interesting to speculate that beside structures of scRap1 DBD with DNA observed by crystallography, other possible complexes may exist in the cell. As mentioned in the previous paragraph, these complexes could exist at a junction between duplex with G-quadruplex DNA or on nucleic acid forming G-quadruplexes in tandems (Figure 43b). This idea is consistent with recent biochemical studies showing that scRap1 in addition to forming the anticipated 1:1 complex with DNA, can also form complexes of higher stoichiometries, where possibly only one of the Myb-like domain is involved in double stranded DNA interactions [126].

It is interesting to note, that human telomeric proteins: TRF1 and TRF2 show the same mechanism of telomeric dsDNA recognition and the Myb domains between those proteins are highly conserved [141]. Furthermore, TRF2 have been implied to be able to interact with G-quadruplex structures [142,143]. Taken this together with the fact that telomeric DNA is guanine rich in different species, it seems highly possible that the role of G-quadruplex structures in telomere maintenance and their interaction with telomeric protein is conserved between species.

6.3. A structural analysis of G – quadruplex recognition

Prior to protein\G-quadruplex structures, an understanding of the mechanism of recognition of G-quadruplex structures was mostly built on the research on small molecules binding to such structures. Most of these

compounds are based on planar aromatic polycyclic structures providing specificity through π - π stacking with guanine rings [144]. The binding of such ligands to the large aromatic surface formed by a G-quartet has been used to provide specificity and enable distinguishing between G-quadruplex and duplex DNA. Furthermore, the interactions are usually stabilized by direct or water mediated electrostatic interactions of positively charged groups of the compounds with the phosphate backbone of DNA structure [145].

Figure 60. The basics of G-quadruplex recognition by small molecules. Solution structure of parallel G-quadruplex in complex with one of the best characterized ligands – PhenDC (PDB ID: 2MGN [146]) Top (a) and side (b) view showing planar aromatic structure stacking on top of G – quadruplex. Orange represents guanines, silver represents other nucleotides.

One of the best structurally characterized interaction of a protein in complex with G-quadruplex DNA is with the RHAU helicase. First, the structure of RHAU specific motif (RSM) peptide in complex with G- quadruplex DNA was published[57]. But during this thesis writing a follow up study of the structure of almost the full length RHAU protein from bovine in complex with G-quadruplex DNA was published [59]. The mode of RSM binding resembles G-quadruplex recognition by ligands specifically selected to recognize such structures. Remarkably, even though the scRap1 DNA recognition α -helix only

shares low sequence similarities with RSM, the structural basis of G-quadruplex recognition shares some similarities (Figure 61). In both cases an α -helix stack on top of the planar face of the G – quarter through hydrophobic interactions. However, the longer scaffold of scRap1 Myb1 stabilize the binding by direct hydrogen bonding with the sugar-phosphate backbone of DNA bases, whereas RHAU use three charged amino acids to clamp the G – quadruplex structure by electrostatic interaction.

Figure 61. Comparison of recognition of G – quadruplex structures by scRap1 and RHAU. (a) Sequence alignment of DNA recognizing α -helix of Myb1 from scRap1 DBD with RHAU specific motif. (b) top view of Myb1 α -helix interacting with G -quadruplex DNA. (c) top view of RSM in complex with G – quadruplex (PDB ID: 2N21 [57]). Orange represents guanines, silver represents other nucleotides.

The three-dimensional structure of scRap1 in complex with the G-quadruplex presented in this thesis provide an understanding of structural basis for protein G – quadruplex recognition. Unfortunately, because scRap1 is a multifunctional protein and G-quadruplex structures have been proposed to have diverse activities, it is difficult to pinpoint the critical role of this interaction *in vivo*.

Interestingly, we found that deletion of 13 amino acid forming the wrapping loop in DNA binding domain that changes the binding affinity to duplex DNA but does not change that to G-quadruplex DNA *in vitro* was not able to restore yeast viability in cells lacking an endogenous copy of scRap1. This mutational analysis confirm the importance of scRap1 interaction with dsDNA for yeast survival. However, this fact does not exclude involvement of scRap1 interaction with G-quadruplex for cells viability, especially in the case of critically short telomeres. Moreover, in the scRap1 – depleted cells through temperature degradation, adding PhenDC3 which stabilize G-quadruplex structure shows growth defects (Lim and Zakian, unpublished). Whereas adding the same PhenDC3 concentration to wild type cells does not influence cells growth. This seem to indicate that scRap1 G-quadruplex recognition is important for controlling the balance between unfolded and folded states of G-quadruplexes possibly through recruiting other proteins. Addition of the drug that stabilize the G-quadruplex structures cause disruption of this homeostasis and shifting dynamic toward G-quadruplex folded state and causing phenotypical changes in cells lacking scRap1.

6.3. Telomere regulation through the interaction of scRap1 with TERRA

The conventional view of the flow of genetic information that was centered around protein, where RNA is only a messenger, the middle product between DNA and protein that it codes for, has been challenged in the past decade. It has become increasingly clear that not only proteins, but also RNAs that are not translated can have very important regulatory roles in the cells. Indeed, only

small portion of genome in complex organisms is coding for proteins, whereas many non-coding regions are being transcribed [147]. The implication of the importance of non-coding regions of DNA has received a lot of attention in recent years, but we are still far from understanding the importance of the vast majority of such regions. One of the proposed long noncoding RNA mediated mechanism of action is targeting the proteins to correct sites in the genome through complementary regions in DNA in sequence specific manner [148].

One of particularly unique long non-coding RNA is TERRA (telomere repeat – containing RNA) that is transcribed from telomeric DNA. Even though the endogenous levels of TERRA are kept very low through most of the cell cycle[99], the regulation of this RNA has been proposed to play an important role for cells to be able to distinguish between non-extendible and extendible telomeres [107]. In mammalian cells, TERRA has been shown to be one of the components of the telomere nucleoprotein complex through direct contact with the main double stranded binding telomeric protein – TRF2 [149]. This interaction, which is important for heterochromatin formation and telomere maintenance, was further implied to be depended more on RNA G – quadruplex structures than the RNA sequence itself. This has led us to the interesting hypothesis that a similar mechanism for the interaction of TERRA with telomeric DNA will be observed in yeast, as the main telomere protein in *S. cerevisiae* (Rap1) does interact with G – quadruplex structures.

The data presented in this thesis supports the hypothesis that scRap1 directly interacts with TERRA in budding yeast. More experiments will be required to fully understand the role of scRap1 in TERRA regulation. Nevertheless,

the hypothesis presented in this chapter may explain the importance of scRap1 interaction with non – canonical nucleic acids structures.

6.4. The evolving world of G – quadruplex research: future perspectives and challenges

Overall the growing field of G – quadruplex research is getting more and more exciting as development of techniques has allowed us understand these structures and their biological significantly more. In the future the combination, of the knowledge from biophysical studies together with biological experiments should help to unravel novel nucleic acid based regulatory ways of controlling cellular pathways.

G – quadruplex forming sequences with their implication in genome instability and telomere maintenance present a great opportunity for novel anti -cancer treatment. However, given their abundance and possible involvement in multiple biological processes, the known small molecules recognizing G-quadruplexes are expected to cause side effect. The selectivity of ligands binding preferentially to G- quadruplex DNA versus duplex can be achieved by recognition of planar surface of G-quartet, but this feature by itself does not permit distinguishing between different sequences or topologies of G-quadruplexes. Better selectivity for therapeutics can be perhaps achieved by developing ligands that can interact with G-quadruplex not only by recognizing guanine platform but by additional sequence specific interaction with loops and grooves.

On the other hand, the limitation of G-quadruplex recognizing molecules in therapies could be possibly overcome by a consideration of the targeting

of protein- quadruplex complexes. In theory, designed compound could influence the stability of the binary complex and impaired G-quadruplex related biological functions. While there is accumulating three-dimensional data on different topologies of G-quadruplexes, there is still very little information on G-quadruplex recognition by proteins. Further exploration of this fertile area will be crucial for better understanding of the fundamental roles of G-quadruplexes in biological processes as well as progress towards drug development.

Figure 62. Schematic representation of possible role of G- quadruplexes as target in cancer therapies. Compounds binding to G – quadruplex structures are considered as possible anticancer drug target as they can influence duplex – quadruplex equilibrium (a) as well as cause changes in the binding and stability of its complex with resolving helicases, promoting chaperons or the complexes with telomerase (b).

The biological significance of interaction of scRap1 with G – quadruplexes remains unclear. The research presented in this thesis provides further insight

into the interaction of scRap1 with G – quadruplexes *in vitro* and can pave the way for a better understanding of the importance of this interaction *in vivo*. Given that biochemical assays did not show significant difference in binding to RNA versus DNA G – quadruplexes, both models presented in chapter three and four could be correct (Figure 43, Figure 49).

Besides the usual challenges of testing G- quadruplex role *in vivo*, scRap1 project presented with additional challenges. Although the comparison of both structure of scRap1 in complex with DNA's revealed that the mode of binding to double-stranded DNA versus G-quadruplex is different, the same DNA recognition α -helix is used. Despite different recognition mechanism the mutational analysis could not separate binding to duplex versus quadruplex DNA. Moreover, additional challenges come from the fact that DBD is essential for yeast survival and testing many mutations *in vivo* is not possible as they are lethal for yeast.

Appendix

His₆-Rap1 (1-827aa)

HHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTGSENLYF
QGAMADIGSEFELRRQAYLNMSSPDDFETAPAEYVDALDPSMVVVD
GSAAVTAPSDSAAEVKANQNEENTGATAAETSEKVDQTEVEKKDDD
DTTEVGVTTPSIADTAATANIASTSGASVTEPTTDDTADEKKEQVS
GPPLSNMKFYLNRDADAHDSLNDIDQLARLIRANGGEVLDSKPREKE
NVFIVSPYNHTNLPTVTPTYIKACCQSNLLNMENYLVYPD NFREVVD
SRLQEESHSGVDNSNSNSDNKDSIRPKTEIISTNTNGATEDSTSEKVM
VDAEQQARLQEQALLRQHVSSTASITSGGHNDLVQIEQPQKDTSEN
NNSNVNDEDNDLLTQDNNPQTADEGNASFQAQRSMISRGALPSHNKA
SFTDEEDEFILDVVRKNPTRRTTHTLYDEISHYVPNHTGNSIRHRFRVY
LSKRLEYVVEVDKFGKLVRRDDGNLIKTKVLPPSIKRKFSADEDYTLA
IAVKKQFYRDLFQIDPDTGRSLITDEDTPTAIARRNMTMDPNHVPGSEP
NFAAYRTQSRRGPIAREFFKHFAEEHAAHTENAWRDRFRKFLLAYGID
DYISYYEAEKAQNREPEPMKNLTNRPKRPGVPTPGNYNSAAKRARNY
SSQRNVQPTANAASANAAAAAASNSYAIPENELLDEDTMNFISSL
KNDLSNISNSLPFEYPHEIAEAIRSDFSNEDIYDNIDPDTISFPPKIATDL
FLPLFFHFGSTRQFMDKLHEVISGDYEPSQAEKLVQDLCDETGIRKNFS
TSILTCLSGDLMVFPRYFLNMFKDNVNPPPNVPGIWITHDDDESLSKND
QEQIRKLVKKHGTGRMEMRKRFFEKDLL

His₆-DBD (359-827aa)

HHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTGSENLYF
QGGALPSHNKASFTDEEDEFILDVVRKNPTRRTTHTLYDEISHYVPNH
TGNSIRHRFRVYLSKRLEYVVEVDKFGKLVRRDDGNLIKTKVLPPSIK
RKFSADEDYTLAIAVKKQFYRDLFQIDPDTGRSLITDEDTPTAIARRNM
TMDPNHVPGSEPNFAAYRTQSRRGPIAREFFKHFAEEHAAHTENAWR
DRFRKFLLAYGIDDYISYYEAEKAQNREPEPMKNLTNRPKRPGVPTPG
NYNSAAKR

His₆-BRCT (1-353 aa)

MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTGSENLY
FQGAMADIGSEFELRRQAYLNMSSPDDFETAPAEYVDALDPSMVVVD
SGSAAVTAPSDSAAEVKANQNEENTGATAAETSEKVDQTEVEKKDDD
DTTEVGVTTPSIADTAATANIASTSGASVTEPTTDDTADEKKEQVS
GPPLSNMKFYLNRDADAHDSLNDIDQLARLIRANGGEVLDSKPREKE
NVFIVSPYNHTNLPTVTPTYIKACCQSNLLNMENYLVYPD NFREVVD
SRLQEESHSGVDNSNSNSDNKDSIRPKTEIISTNTNGATEDSTSEKVM

VDAEQQARLQEQAQLLRQHVSSTASITSGGHNDLVQIEQPQKDTSNN
NNSNVNDEDNDLLTQDNNPQTADEGNASFQAQRSMISRGA

His₆-RCT (599-827 aa)

MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTGSENLY
FQGAMADIGSEFELRRQAWARNYSSQRNVQPTANAASANAAAAA
AASNSYAIPENELLEDTMNFISSLKNDLSNISNSLPFEYPHEIAEAI
RSDFSNEDIYDNIDPDTISFPPKIATTDLFLPLFFHFGSTRQFMDKL
HEVISGDYEPSQAEKLVQDLCDETGIRKNFSTSILTCLSGDLMVFPRY
FLNMFKDNVNPPPNVPGIWITHDDDESLSKNDQEQIRKLVKKHGTGR
MEMRKRF EKDLL

His₆-Myb1 (359-411aa)

MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTGSENLY
FQGGALPSHNKASFTDEEDEFILDVVRKNPTRRTTHTLYDEISHYV
PNHTGNSIRHRFRVYLS

His₆-Myb2 (449-598aa)

MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTGSENLY
FQGGALPSHNKASFTDEEDEFILDVVRKNPTRRTTHTLYDEISHYV
PNHTGNAWRDRFRKFLSKRLEYVVEVDKFGKLVRRDDG NLIKT
KVLPPSIKRKFSADEDYTLAIAVKKQFYRDLFQIDPDTGRSLIT
DEDTPTAIARNMTMDPNHVPGSEPNFAAYRTQSRRGPIAREFFKH
FAEEHAAHTENAWRDRFRKFLLAYGIDDYISYYEAEKAQNREPE
PMKNLTNRPKRPGVPTP GNYNSAAKR

Table 4. Peptides used in the studies:

Name	Sequence
Myb1	HTGNSIRHRFRVYLSKR
Myb2	HTENAWRDRFRKFLLAY

Table 5. Oligonucleotides used in the studies:

Name		Sequence
4G3		AATTCTGGGTGTGTGGGTGTGTGGGTGTGTGGGTG TGG
4G3-no tail		TGGGTGTGTGGGTGTGTGGGTGTGTGGG
rTelo		UGGGUGUGUGGGUGUGUGGGUGUGUGGG
T-loops(HIV integrase aptamer)		TTGGGTGGGTGGGTGGGTT
dsDNA	1G3	AATTCGGTGTGTGGGTGTG
	1C3	GATCCACACCCACACACCG
1L F (oNI54)		CGGTGGGTGAGTGGTAGTAAGTAGA
1L R (oBL1181)		CAGCCCTAATCTAACCCTGG
CA (oBL207)		CACCACACCCACACACCACCCACA
RT VI-R (+):		ATC ATT GAG GAT CTA TAA TC
RT VI-R (-):		CTT CAC TCC ATT GCG
RT XV-L (+):		TAA CCC TGT CCA ACC TGT CT
RT XV-L (-):		ATA CTA TAG CAT CCG TGG GC
Actin F		GTAACATCGTTATGTCCGGTGGTAC
Actin R		CCAAGATAGAACCACCAATCCAGAC

Table 6. Strains used in the studies:

Name	Sequence	Source
LVPS-151 Ku80-8G- 3HA/TetO7-1L	By4741 MATa his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 Est2-8G- myc18::TRP1MX6 trp1::NAT ade2::tTA URA_ADH1_TetO7_CYC1_1L_TER RA Ku80-8G-3HA::HIS3MX6	Pfeiffer at all 2012 [110]
LVPS-151 Rap1-13xmyc	By4741MATa his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 Est2-8G- myc18::TRP1MX6 trp1::NAT ade2::tTA URA_ADH1_TetO7_CYC1_1L_TER RA Ku80-8G-3HA::HIS3MX6 Rap1- 13myc::Leu	This study
Δrap1/Rap1	YPH501 rap1Δ::KanMX4/Rap1	This study
Rap1- Flag,Gall-Rap1- myc	By4741 MATa Rap1- 3xFlag::KanMX, Gall_Rap1-9xmyc:: HIS3, Δbar1::LEU	Lickwar at all 2012 [87]
Rap1- Flag,Gall- Rap1Δ(583- 596)-myc	By4741 MATa Rap1- 3xFlag::KanMX, Gall_Rap1Δ (583- 596)-13xmyc:: HIS	This study

Reference:

1. Watson JD, Crick FHC. A structure for deoxyribose nucleic acid. *Nature*. 1953. pp. 737–738. doi:10.1016/S0735-1097(03)00800-3
2. Bochman ML, Paeschke K, Zakian V a. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet*. Nature Publishing Group; 2012;13: 770–780. doi:10.1038/nrg3296
3. Gellert M, Lipsett MN, Davies DR. Helix Formation By Guanylic Acid. *Proc Natl Acad Sci*. 1962;48: 2013–2018. doi:10.1073/pnas.48.12.2013
4. Sundquist WI, Klug a. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*. 1989;342: 825–829. doi:10.1038/342825a0
5. Williamson JR, Raghuraman MK, Cech TR. Monovalent cation-induced structure of telomeric DNA: The G-quartet model. *Cell*. 1989;59: 871–880. doi:10.1016/0092-8674(89)90610-7
6. Sen D, Gilbert W. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature*. 1990. pp. 410–414. doi:10.1038/344410a0
7. Rhodes D, Lipps HJ. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res*. 2015; gkv862. doi:10.1093/nar/gkv862
8. Lipps HJ, Rhodes D. G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol*. 2009;19: 414–22. doi:10.1016/j.tcb.2009.05.002
9. Huppert JL. Structure, location and interactions of G-quadruplexes. *FEBS J*. 2010;277: 3452–3458. doi:10.1111/j.1742-4658.2010.07758.x
10. Burge S, Parkinson GN, Hazel P, Todd a. K, Neidle S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res*. 2006;34: 5402–5415. doi:10.1093/nar/gkl655
11. Guedin a., Gros J, Alberti P, Mergny J-L. How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res*. 2010;38: 7858–7868. doi:10.1093/nar/gkq639
12. Todd AK. Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res*. 2005;33: 2901–2907. doi:10.1093/nar/gki553
13. Huppert JL. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res*. 2005;33: 2908–2916. doi:10.1093/nar/gki609
14. Chambers VS, Marsico G, Boutell JM, Di Antonio M, Smith GP, Balasubramanian S. High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol*. Nature Publishing Group; 2015;33: 877–881. doi:10.1038/nbt.3295
15. Nakken S, Rognes T, Hovig E. The disruptive positions in human G-quadruplex motifs are less polymorphic and more conserved than their neutral counterparts. *Nucleic Acids Res*. 2009;37: 5749–56. doi:10.1093/nar/gkp590
16. Eddy J, Maizels N. Gene function correlates with potential for G4 DNA

- formation in the human genome. *Nucleic Acids Res.* 2006;34: 3887–3896. doi:10.1093/nar/gkl529
17. Cogoi S, Xodo LE. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* 2006;34: 2536–2549. doi:10.1093/nar/gkl286
18. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc Natl Acad Sci U S A.* 2002;99: 11593–8. doi:10.1073/pnas.182256799
19. Besnard E, Babled A, Lapasset L, Milhavet O, Parrinello H, Dantec C, et al. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat Struct Mol Biol.* Nature Publishing Group; 2012;19: 837–844. doi:10.1038/nsmb.2339
20. Cayrou C, Grégoire D, Coulombe P, Danis E, Méchali M. Genome-scale identification of active DNA replication origins. *Methods.* Elsevier Inc.; 2012;57: 158–164. doi:10.1016/j.ymeth.2012.06.015
21. Paeschke K, Simonsson T, Postberg J, Rhodes D, Lipps HJ. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat Struct & Mol Biol.* 2005;12: 847–854. doi:10.1038/nsmb982
22. Schaffitzel C, Berger I, Postberg J, Hanes J, Lipps HJ, Plückthun a. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Stylonychia lemnae* macronuclei. *Proc Natl Acad Sci U S A.* 2001;98: 8572–7. doi:10.1073/pnas.141229498
23. Salvati E, Leonetti C, Rizzo A, Scarsella M, Mottotese M, Galati R, et al. Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. *J Clin Invest.* 2007;117: 3236–3247. doi:10.1172/JCI32461
24. Gomez D, Wenner T, Brassart B, Douarre C, O'Donohue MF, El Khoury V, et al. Telomestatin-induced telomere uncapping is modulated by POT1 through G-overhang extension in HT1080 human tumor cells. *J Biol Chem.* 2006;281: 38721–38729. doi:10.1074/jbc.M605828200
25. Verma A, Yadav VK, Basundra R, Kumar A, Chowdhury S. Evidence of genome-wide G4 DNA-mediated gene expression in human cancer cells. *Nucleic Acids Res.* 2009;37: 4194–4204. doi:10.1093/nar/gkn1076
26. Grand CL, Han H, Muñoz RM, Weitman S, Von Hoff DD, Hurley LH, et al. The Cationic Porphyrin TMPyP4 Down-Regulates c- MYC and Human Telomerase Reverse Transcriptase Expression and Inhibits Tumor Growth in Vivo. *Mol Cancer Ther.* 2002;1: 565–573.
27. Biffi G, Tannahill D, McCafferty J, Balasubramanian S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem.* Nature Publishing Group; 2013;5: 182–186. doi:10.1038/nchem.1548
28. Sun H, Karow JK, Hickson ID, Maizels N. The Bloom's syndrome helicase unwinds G4 DNA. *J Biol Chem.* 1998;273: 27587–27592. doi:10.1074/jbc.273.42.27587

29. Suhasini AN, Brosh RM. Disease-causing missense mutations in human DNA helicase disorders. *Mutat Res - Rev Mutat Res.* Elsevier B.V.; 2013;752: 138–152. doi:10.1016/j.mrrev.2012.12.004
30. Cheung I, Schertzer M, Rose A, Lansdorp PM. Disruption of dog-1 in *Caenorhabditis elegans* triggers deletions upstream of guanine-rich DNA. *Nat Genet.* 2002;31: 405–409. doi:10.1038/ng928
31. Vannier J-B, Pavicic-Kaltenbrunner V, Petalcorin MIR, Ding H, Boulton SJ. RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. *Cell.* Elsevier Inc.; 2012;149: 795–806. doi:10.1016/j.cell.2012.03.030
32. Ribeyre C, Lopes J, Boulé J-B, Piazza A, Guédin A, Zakian V a., et al. The Yeast Pif1 Helicase Prevents Genomic Instability Caused by G-Quadruplex-Forming CEB1 Sequences In Vivo. *PLoS Genet.* 2009;5: e1000475. doi:10.1371/journal.pgen.1000475
33. Paeschke K, Capra J a, Zakian V a. DNA replication through G-quadruplex motifs is promoted by the *Saccharomyces cerevisiae* Pif1 DNA helicase. *Cell.* Elsevier Inc.; 2011;145: 678–91. doi:10.1016/j.cell.2011.04.015
34. Giraldo R, Suzuki M, Chapman L, Rhodes D. Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: a circular dichroism study. *Proc Natl Acad Sci U S A.* 1994;91: 7658–62. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=44461&tool=pmcentrez&rendertype=abstract>
35. Giraldo R, Rhodes D. The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *EMBO J.* 1994;13: 2411–20. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=395107&tool=pmcentrez&rendertype=abstract>
36. Zhang M-L, Tong X-J, Fu X-H, Zhou BO, Wang J, Liao X-H, et al. Yeast telomerase subunit Est1p has guanine quadruplex-promoting activity that is required for telomere elongation. *Nat Struct Mol Biol.* Nature Publishing Group; 2010;17: 202–9. doi:10.1038/nsmb.1760
37. Li QJ, Tong XJ, Duan YM, Zhou JQ. Characterization of the intramolecular G-quadruplex promoting activity of Est1. *FEBS Lett.* 2013;587: 659–665. doi:10.1016/j.febslet.2013.01.024
38. Gottschling DE, Zakian VA. Telomere proteins: Specific recognition and protection of the natural termini of *Oxytricha* macronuclear DNA. *Cell.* 1986;47: 195–205. doi:10.1016/0092-8674(86)90442-3
39. Peersen OB, Ruggles JA, Schultz SC. Dimeric structure of the *Oxytricha* nova telomere end-binding protein α -subunit bound to ssDNA. *Nat Struct Biol.* 2002;9: 182–187. doi:10.1038/nsb761
40. Fang G, Cech TR. *Oxytricha* telomere-binding protein: DNA-dependent dimerization of the alpha and beta subunits. *Proc Natl Acad Sci U S A.* 1993;90: 6056–6060. doi:10.1126/JB.01125-10
41. Fang G, Cech TR. The beta subunit of *Oxytricha* telomere-binding protein promotes G-quartet formation by telomeric DNA. *Cell.* 1993;74:

- 875–85. doi:10.1016/0092-8674(93)90467-5
42. Paeschke K, Juranek S, Rhodes D, Lipps HJ. Cell cycle-dependent regulation of telomere tethering in the nucleus. *Chromosom Res.* 2008;16: 721–728. doi:10.1007/s10577-008-1222-x
 43. Saccà B, Lacroix L, Mergny JL. The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Res.* 2005;33: 1182–1192. doi:10.1093/nar/gki257
 44. Cammas A, Millevoi S. RNA G-quadruplexes: emerging mechanisms in disease. *Nucleic Acids Res.* 2017;45: 1584–1595. doi:10.1093/nar/gkw1280
 45. Zheng S, Vuong BQ, Vaidyanathan B, Lin J-Y, Huang F-T, Chaudhuri J. Non-coding RNA Generated following Lariat Debranching Mediates Targeting of AID to DNA. *Cell.* 2015;161: 762–773. doi:10.1016/j.cell.2015.03.020
 46. Vaughn JP, Creacy SD, Routh ED, Joyner-Butt C, Jenkins GS, Pauli S, et al. The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J Biol Chem.* 2005;280: 38117–38120. doi:10.1074/jbc.C500348200
 47. Lattmann S, Stadler MB, Vaughn JP, Akman S a, Nagamine Y. The DEAH-box RNA helicase RHAU binds an intramolecular RNA G-quadruplex in TERC and associates with telomerase holoenzyme. *Nucleic Acids Res.* 2011;39: 9390–404. doi:10.1093/nar/gkr630
 48. Kumari S, Bugaut A, Huppert JL, Balasubramanian S. Europe PMC Funders Group An RNA G-quadruplex in the 5 ' UTR of the NRAS proto- oncogene modulates translation. 2008;3: 218–221. doi:10.1038/nchembio864.An
 49. Millevoi S, Moine H, Vagner S. G-quadruplexes in RNA biology. *Wiley Interdiscip Rev RNA.* 2012;3: 495–507. doi:10.1002/wrna.1113
 50. Mirihana Arachchilage G, Dassanayake AC, Basu S. A potassium ion-dependent RNA structural switch regulates human pre-miRNA 92b maturation. *Chem Biol. Elsevier Ltd;* 2015;22: 262–272. doi:10.1016/j.chembiol.2014.12.013
 51. Jayaraj GG, Pandey S, Scaria V, Maiti S. Potential G-quadruplexes in the human long non-coding transcriptome. *RNA Biol.* 2012;9: 81–86. doi:10.4161/rna.9.1.18047
 52. Xu Y, Ishizuka T, Yang J, Ito K, Katada H, Komiyama M, et al. Oligonucleotide models of telomeric DNA and RNA form a hybrid G-quadruplex structure as a potential component of telomeres. *J Biol Chem.* 2012;287: 41787–41796. doi:10.1074/jbc.M112.342030
 53. Zheng KW, Wu RY, He Y De, Xiao S, Zhang JY, Liu JQ, et al. A competitive formation of DNA:RNA hybrid G-quadruplex is responsible to the mitochondrial transcription termination at the DNA replication priming site. *Nucleic Acids Res.* 2014;42: 10832–10844. doi:10.1093/nar/gku764

54. Phan AT, Kuryavyi V, Darnell JC, Serganov A, Majumdar A, Ilin S, et al. Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat Struct Mol Biol.* 2011;18: 796–804. doi:10.1038/nsmb.2064
55. Krauss IR, Pica A, Merlino A, Mazzarella L, Sica F. Duplex-quadruplex motifs in a peculiar structural organization cooperatively contribute to thrombin binding of a DNA aptamer. *Acta Crystallogr Sect D Biol Crystallogr. International Union of Crystallography;* 2013;69: 2403–2411. doi:10.1107/S0907444913022269
56. Horvath MP, Schultz SC. DNA G-quartets in a 1.86 Å resolution structure of an *Oxytricha nova* telomeric protein-DNA complex. *J Mol Biol.* 2001;310: 367–77. doi:10.1006/jmbi.2001.4766
57. Heddi B, Cheong VV, Martadinata H, Phan AT. Insights into G-quadruplex specific recognition by the DEAH-box helicase RHAU: Solution structure of a peptide-quadruplex complex. *Proc Natl Acad Sci.* 2015;637371: 201422605. doi:10.1073/pnas.1422605112
58. Lattmann S, Giri B, Vaughn JP, Akman S a, Nagamine Y. Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU. *Nucleic Acids Res.* 2010;38: 6219–33. doi:10.1093/nar/gkq372
59. Chen MC, Tippa R, Demeshkina NA, Murat P, Balasubramanian S, Myong S, et al. Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature. Springer US;* 2018; doi:10.1038/s41586-018-0209-9
60. You H, Lattmann S, Rhodes D, Yan J. RHAU helicase stabilizes G4 in its nucleotide-free state and destabilizes G4 upon ATP hydrolysis. *Nucleic Acids Res.* 2017;45: 206–214. doi:10.1093/nar/gkw881
61. Yangyuoru PM, Bradburn DA, Liu Z, Xiao TS, Russell R. The G-quadruplex (G4) resolvase DHX36 efficiently and specifically disrupts DNA G4s via a translocation-based helicase mechanism. *J Biol Chem.* 2018;293: 1924–1932. doi:10.1074/jbc.M117.815076
62. Feng J, Funk W, Wang S, Weinrich S, Avilion A, Chiu C, et al. The RNA component of human telomerase. *Science (80-).* 1995;269: 1236–1241. doi:10.1126/science.7544491
63. Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PLC, et al. Specific association of human telomerase activity with immortal cells and cancer. *Science.* 1994. pp. 2011–2015.
64. Blackburn EH. *Telomeres.* Elsevier Science. 1991;16: 378–381. doi:/10.1016/0968-0004(91)90155-O
65. Wellinger RJ, Zakian V a. Everything you ever wanted to know about *Saccharomyces cerevisiae* telomeres: beginning to end. *Genetics.* 2012;191: 1073–105. doi:10.1534/genetics.111.137851
66. Chang M, Arneric M, Lingner J. Telomerase repeat addition processivity is increased at critically short telomeres in a Tel1-dependent manner in *Saccharomyces cerevisiae*. *Genes Dev.* 2007;21: 2485–94. doi:10.1101/gad.1588807

67. Larrivée M, LeBel C, Wellinger RJ. The generation of proper constitutive G-tails on yeast telomeres is dependent on the MRX complex. *Genes Dev.* 2004;18: 1391–1396. doi:10.1101/gad.1199404
68. Zahler a M, Williamson JR, Cech TR, Prescott DM. Inhibition of telomerase by G-quartet DNA structures. *Nature.* 1991;350: 718–720. doi:10.1038/350718a0
69. Smith JS, Chen Q, Yatsunyk L a, Nicoludis JM, Garcia MS, Kranaster R, et al. Rudimentary G-quadruplex-based telomere capping in *Saccharomyces cerevisiae*. *Nat Struct Mol Biol.* Nature Publishing Group; 2011;18: 478–85. doi:10.1038/nsmb.2033
70. de Lange T. How Telomeres Solve the End-Protection Problem. *Science* (80-). 2009;326: 948–952. doi:10.1126/science.1170633
71. De Lange T. Shelterin: The protein complex that shapes and safeguards human telomeres. *Genes Dev.* 2005;19: 2100–2110. doi:10.1101/gad.1346005
72. Gilson E, Roberge M, Giraldo R, Rhodes D, Gasser SM. Distortion of the DNA Double Helix by RAP1 at Silencers and Multiple Telomeric Binding Sites. *J Mol Biol.* 1993;231: 293–310. doi:10.1006/jmbi.1993.1283
73. Court R, Chapman L, Fairall L, Rhodes D. How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures. *EMBO Rep.* 2005;6: 39–45. doi:10.1038/sj.embor.7400314
74. Chen Y, Rai R, Zhou Z-R, Kanoh J, Ribeyre C, Yang Y, et al. A conserved motif within RAP1 has diversified roles in telomere protection and regulation in different organisms. *Nat Struct Mol Biol.* 2011;18: 213–221. doi:10.1038/nsmb.1974
75. Shi T, Bunker RD, Mattarocci S, Ribeyre C, Faty M, Gut H, et al. Rif1 and Rif2 shape telomere function and architecture through multivalent Rap1 interactions. *Cell.* 2013;153: 1340–53. doi:10.1016/j.cell.2013.05.007
76. Lue NF. Plasticity of telomere maintenance mechanisms in yeast. *Trends Biochem Sci.* 2010;35: 8–17. doi:10.1016/j.tibs.2009.08.006
77. Steinberg-Neifach O, Lue NF. Telomere DNA recognition in *Saccharomycotina* yeast: potential lessons for the co-evolution of ssDNA and dsDNA-binding proteins and their target sites. *Front Genet.* 2015;6: 162. doi:10.3389/fgene.2015.00162
78. Huet J, Cottrelle P, Cool M, Vignais ML, Thiele D, Marck C, et al. A general upstream binding factor for genes of the yeast translational apparatus. *EMBO J.* 1985;4: 3539–47.
79. Buchman AR, Lue NF, Kornberg RD. Connections between transcriptional activators, silencers, and telomeres as revealed by functional analysis of a yeast DNA-binding protein. *Mol Cell Biol.* 1988;8: 5086–99. doi:10.1128/MCB.8.12.5086.Updated
80. Longtine MS, Wilson NM, Petracek ME, Berman J. A yeast telomere

- binding activity binds to two related telomere sequence motifs and is indistinguishable from RAP1. *Curr Genet.* 1989;16: 225–239. doi:10.1007/BF00422108
81. Marcand S, Gilson E, Shore D. A Protein-Counting Mechanism for Telomere Length Regulation in Yeast A Protein-Counting Mechanism for Telomere Length Regulation in Yeast *Ste. Science* (80-). 2010;986. doi:10.1126/science.275.5302.986
82. Lue NF. Plasticity of telomere maintenance mechanisms in yeast. *Trends Biochem Sci.* 2010;35: 8–17. doi:10.1016/j.tibs.2009.08.006
83. Graham IR, Haw RA, Spink KG, Halden KA, Chambers A. In vivo analysis of functional regions within yeast Rap1p. *Mol Cell Biol.* 1999;19: 7481–90. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=84746&tool=pmcentrez&rendertype=abstract>
84. Lieb JD, Liu X, Botstein D, Brown PO. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nat Genet.* 2001;28: 327–334. doi:10.1038/ng569
85. Lickwar CR, Mueller F, Hanlon SE, McNally JG, Lieb JD. Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature.* 2012;484: 251–5. doi:10.1038/nature10985
86. Lickwar CR, Mueller F, Lieb JD. Genome-wide measurement of protein-DNA binding dynamics using competition ChIP. *Nat Protoc.* 2013;8: 1337–1353. doi:10.1038/nprot.2013.077
87. Lickwar CR, Mueller F, Hanlon SE, McNally JG, Lieb JD. Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature.* Nature Publishing Group; 2012;484: 251–5. doi:10.1038/nature10985
88. Lustig a J, Kurtz S, Shore D. Involvement of the silencer and UAS binding protein RAP1 in regulation of telomere length. *Science.* 1990;250: 549–553. doi:10.1126/science.2237406
89. Konig P, Giraldo R, Chapman L, Rhodes D. The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. *Cell.* 1996;85: 125–36. Available: <http://www.ncbi.nlm.nih.gov/pubmed/8620531>
90. Bianchi A, Smith S, Chong L, Elias P, De Lange T. TRF1 is a dimer and bends telomeric DNA. *EMBO J.* 1997;16: 1785–1794. doi:10.1093/emboj/16.7.1785
91. Fairall L, Chapman L, Moss H, De Lange T, Rhodes D. Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2. *Mol Cell.* 2001;8: 351–361. doi:10.1016/S1097-2765(01)00321-5
92. Henry Y a, Chambers a, Tsang JS, Kingsman a J, Kingsman SM. Characterisation of the DNA binding domain of the yeast RAP1 protein. *Nucleic Acids Res.* 1990;18: 2617–23. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=330744&to>

ol=pmcentrez&rendertype=abstract

93. Matot B, Le Bihan Y-V, Lescasse R, Pérez J, Miron S, David G, et al. The orientation of the C-terminal domain of the *Saccharomyces cerevisiae* Rap1 protein is determined by its binding to DNA. *Nucleic Acids Res.* 2012;40: 3197–207. doi:10.1093/nar/gkr1166
94. Le Bihan YV, Matot B, Pietrement O, Giraud-Panis MJ, Gasparini S, Le Cam E, et al. Effect of Rap1 binding on DNA distortion and potassium permanganate hypersensitivity. *Acta Crystallogr Sect D Biol Crystallogr.* 2013;69: 409–419. doi:10.1107/S0907444912049311
95. Wellinger RJ, Wolf AJ, Zakian VA. *Saccharomyces* telomeres acquire single-strand TG1-3 tails late in S phase. *Cell.* 1993;72: 51–60. doi:10.1016/0092-8674(93)90049-V
96. Azzalin CM, Lingner J. Telomeres: The silence is broken. *Cell Cycle.* 2008;7: 1161–1165. doi:10.4161/cc.7.9.5836
97. Azzalin CM, Reichenbach P, Khorauli L, Giulotto E, Lingner J. and RNA Surveillance Factors at. *Science.* 2007;318: 798–801. doi:10.1126/science.1147182
98. Azzalin CM, Lingner J. Telomere functions grounding on TERRA firma. *Trends Cell Biol. Elsevier Ltd;* 2015;25: 29–36. doi:10.1016/j.tcb.2014.08.007
99. Luke B, Panza A, Redon S, Iglesias N, Li Z, Lingner J. The Rat1p 5' to 3' exonuclease degrades telomeric repeat-containing RNA and promotes telomere elongation in *Saccharomyces cerevisiae*. *Mol Cell. Elsevier Inc.;* 2008;32: 465–77. doi:10.1016/j.molcel.2008.10.019
100. Iglesias N, Redon S, Pfeiffer V, Dees M, Lingner J, Luke B. Subtelomeric repetitive elements determine TERRA regulation by Rap1/Rif and Rap1/Sir complexes in yeast. *EMBO Rep. Nature Publishing Group;* 2011;12: 587–593. doi:10.1038/embor.2011.73
101. Benetti R, Schoeftner S, Muñoz P, Blasco MA. Role of TRF2 in the assembly of telomeric chromatin. *Cell Cycle.* 2008;7: 3461–3468. doi:10.4161/cc.7.21.7013
102. Schoeftner S, Blasco MA. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol.* 2008;10: 228–236. doi:10.1038/ncb1685
103. Sfeir A, Kosiyatrakul ST, Hockemeyer D, MacRae SL, Karlseder J, Schildkraut CL, et al. Mammalian Telomeres Resemble Fragile Sites and Require TRF1 for Efficient Replication. *Cell. Elsevier Ltd;* 2009;138: 90–103. doi:10.1016/j.cell.2009.06.021
104. Collie GW, Haider SM, Neidle S, Parkinson GN. A crystallographic and modelling study of a human telomeric RNA (TERRA) quadruplex. *Nucleic Acids Res.* 2010;38: 5569–80. doi:10.1093/nar/gkq259
105. Martadinata H, Tua A. Structure of Propeller-Type Parallel-Stranded RNA G-Quadruplexes , Formed by Human Telomeric RNA Sequences in K Solution Structure of Propeller-Type Parallel-Stranded RNA G-Quadruplexes , Formed by Human Telomeric RNA Sequences in K +

- Solution. *Structure*. 2009; 2570–2578. doi:10.1021/ja806592z
106. Martadinata H, Heddi B, Lim KW, Phan AT. Structure of long human telomeric RNA (TERRA): G-quadruplexes formed by four and eight UUAGGG repeats are stable building blocks. *Biochemistry*. 2011;50: 6455–6461. doi:10.1021/bi200569f
 107. Graf M, Bonetti D, Lockhart A, Serhal K, Kellner V, Maicher A, et al. Telomere Length Determines TERRA and R-Loop Regulation through the Cell Cycle. *Cell*. 2017;170: 72–85.e14. doi:10.1016/j.cell.2017.06.006
 108. Arora R, Lee Y, Wischnewski H, Brun CM, Schwarz T, Azzalin CM. RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat Commun. Nature Publishing Group*; 2014;5: 1–11. doi:10.1038/ncomms6220
 109. Sandell LL, Gottschling DE, Zakian VA. Transcription of a yeast telomere alleviates telomere position effect without affecting chromosome stability. *Proc Natl Acad Sci U S A*. 1994;91: 12061–12065. doi:10.1073/pnas.91.25.12061
 110. Pfeiffer V, Lingner J. TERRA promotes telomere shortening through exonuclease 1-mediated resection of chromosome ends. *PLoS Genet*. 2012;8. doi:10.1371/journal.pgen.1002747
 111. Flynn RL, Centore RC, O’Sullivan RJ, Rai R, Tse A, Songyang Z, et al. TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. *Nature. Nature Publishing Group*; 2011;471: 532–538. doi:10.1038/nature09772
 112. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr. International Union of Crystallography*; 2007;40: 658–674. doi:10.1107/S0021889807021206
 113. Do NQ, Phan AT. Monomer-dimer equilibrium for the 5’-5’ stacking of propeller-type parallel-stranded g-quadruplexes: NMR structural study. *Chem - A Eur J*. 2012;18: 14752–14759. doi:10.1002/chem.201103295
 114. Murshudov GN, Skubák P, Lebedev A a., Pannu NS, Steiner R a., Nicholls R a., et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr Sect D Biol Crystallogr*. 2011;67: 355–367. doi:10.1107/S0907444911001314
 115. Blackburn EH, Gall JG. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in *Tetrahymena*. *J Mol Biol*. 1978;120: 33–53. doi:10.1016/0022-2836(78)90294-2
 116. Blackburn EH, Budarf ML, Challoner PB, Cherry JM, Howard EA, Katzen AL, et al. DNA termini in ciliate macronuclei. *Cold Spring Harb Symp Quant Biol*. 1982;47: 1195–1207. doi:10.1101/SQB.1983.047.01.135
 117. Conrad MN, Wright JH, Wolf AJ, Zakian V a. RAP1 protein interacts with yeast telomeres in vivo: Overproduction alters telomere structure and decreases chromosome stability. *Cell*. 1990;63: 739–750. doi:10.1016/0092-8674(90)90140-A

118. Taylor HO, O'Reilly M, Leslie a G, Rhodes D. How the multifunctional yeast Rap1p discriminates between DNA target sites: a crystallographic analysis. *J Mol Biol.* 2000;303: 693–707. doi:10.1006/jmbi.2000.4161
119. Williams TL, Levy DL, Maki-Yonekura S, Yonekura K, Blackburn EH. Characterization of the yeast telomere nucleoprotein core: Rap1 binds independently to each recognition site. *J Biol Chem.* 2010;285: 35814–24. doi:10.1074/jbc.M110.170167
120. König SLB, Evans AC, Huppert JL. Seven essential questions on G-quadruplexes. *Biomol Concepts.* 2010;1: 197–213. doi:10.1515/bmc.2010.011
121. Masiero S, Trotta R, Pieraccini S, De Tito S, Perone R, Randazzo A, et al. A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. *Org Biomol Chem.* 2010;8: 2683–2692. doi:10.1039/c003428b
122. Kelley S, Boroda S, Musier-Forsyth K, Kankia BI. HIV-integrase aptamer folds into a parallel quadruplex: A thermodynamic study. *Biophys Chem. Elsevier B.V.;* 2011;155: 82–88. doi:10.1016/j.bpc.2011.03.004
123. Arora A, Maiti S. Stability and molecular recognition of quadruplexes with different loop length in the absence and presence of molecular crowding agents. *J Phys Chem B.* 2009;113: 8784–8792. doi:10.1021/jp809486g
124. Hollis T. Crystallization of protein-DNA complexes. *Methods Mol Biol.* 2007;363: 225–237. doi:10.1385/1-59745-209-2:225
125. Bauer L, Tlučková K, Tóthová P, Viglaský V. G-quadruplex motifs arranged in tandem occurring in telomeric repeats and the insulin-linked polymorphic region. *Biochemistry.* 2011;50: 7484–7492. doi:10.1021/bi2003235
126. Feldmann EA, Galletto R. The DNA-Binding Domain of Yeast Rap1 Interacts with dsDNA in Multiple Binding Modes. *Biochemistry.* 2014; doi:10.1021/bi501049b
127. Levy DL, Blackburn EH. Counting of Rif1p and Rif2p on. *Society.* 2004;24: 10857–10867. doi:10.1128/MCB.24.24.10857
128. Freeman K, Gwadz M, Shore D. Molecular and genetic analysis of the toxic effect of RAP1 overexpression in yeast. *Genetics.* 1995;141: 1253–1262.
129. Gallo A, Lo Sterzo C, Mori M, Di Matteo A, Bertini I, Banci L, et al. Structure of nucleophosmin DNA-binding domain and analysis of its complex with a G-quadruplex sequence from the c-MYC promoter. *J Biol Chem.* 2012;287: 26539–48. doi:10.1074/jbc.M112.371013
130. Morin A, Moores AW, Sacher M. Dissection of *Saccharomyces cerevisiae* asci. *J Vis Exp.* 2009; 6–8. doi:10.3791/1146
131. Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics.* 1989;122: 19–27. doi:0378111995000377 [pii]

132. Dani GM, Zakian VA. Mitotic and meiotic stability of linear plasmids in yeast. *Proc Natl Acad Sci U S A*. 1983;80: 3406–10. doi:10.1073/pnas.80.11.3406
133. Chambers A. Highly conserved toxicity of *Saccharomyces cerevisiae* Rap1p. *Mol Microbiol*. 1996;22: 449–458. doi:10.1046/j.1365-2958.1996.00134.x
134. Bang I. Untersuchungen über die Guanylsäure. *Biochem Z*. 1910;26: 293–311.
135. Fang G, Cech TR. The beta subunit of *Oxytricha* telomere-binding protein promotes G-quartet formation by telomeric DNA. *Cell*. 1993;74: 875–85. doi:10.1016/0092-8674(93)90467-5
136. Fang G, Cech TR. Characterization of a G-Quartet Formation Reaction Promoted by the β -Subunit of the *Oxytricha* Telomere-Binding Protein. *Biochemistry*. 1993;32: 11646–11657. doi:10.1021/bi00094a022
137. Hänsel-Hertsch R, Beraldi D, Stefanie L, Giovanni M, Di Antonio M, Tannahill D, et al. G-quadruplex structures mark human regulatory chromatin. *Nat Genet*. 2016;accepted. doi:10.1038/ng.3662
138. Kwok CK, Marsico G, Sahakyan AB, Chambers VS, Balasubramanian S. RG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat Methods*. 2016;13: 841–844. doi:10.1038/nmeth.3965
139. Lustig AJ, Kurtz S, Shore D. Involvement of the silencer and UAS binding protein RAP1 in regulation of telomere length. *Science* (80-). 1990;250: 549–553. doi:10.1126/science.2237406
140. Harrison SC, Aggarwal AK. DNA Recognition by Proteins with the Helix-Turn-Helix Motif. *Annu Rev Biochem*. 1990;59: 933–969. doi:10.1146/annurev.bi.59.070190.004441
141. König P, Rhodes D. Recognition of telomeric DNA. *Trends Biochem Sci*. 1997;22: 43–47. doi:10.1016/S0968-0004(97)01008-6
142. Biffi G, Tannahill D, Balasubramanian S. An intramolecular G-quadruplex structure is required for binding of telomeric repeat-containing RNA to the telomeric protein TRF2. *J Am Chem Soc*. 2012;134: 11974–11976.
143. Pedroso IM, Hayward W, Fletcher TM. The effect of the TRF2 N-terminal and TRFH regions on telomeric G-quadruplex structures. *Nucleic Acids Res*. 2009;37: 1541–54. doi:10.1093/nar/gkn1081
144. Monchaud D, Teulade-Fichou MP. A hitchhiker's guide to G-quadruplex ligands. *Org Biomol Chem*. 2008;6: 627–636. doi:10.1039/b714772b
145. Balasubramanian S, Neidle S. G-quadruplex nucleic acids as therapeutic targets. *Curr Opin Chem Biol*. 2009;13: 345–53. doi:10.1016/j.cbpa.2009.04.637
146. Chung WJ, Heddi B, Hamon F, Teulade-Fichou MP, Phan AT. Solution structure of a G-quadruplex bound to the bisquinolinium compound phen-DC3. *Angew Chemie - Int Ed*. 2014;53: 999–1002. doi:10.1002/anie.201308063

147. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet.* Nature Publishing Group; 2016;17: 47–62. doi:10.1038/nrg.2015.10
148. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci.* 2009;106: 11667–11672. doi:10.1073/pnas.0904715106
149. Deng Z, Norseen J, Wiedmer A, Riethman H, Lieberman PM. TERRA RNA Binding to TRF2 Facilitates Heterochromatin Formation and ORC Recruitment at Telomeres. *Mol Cell.* Elsevier Ltd; 2009;35: 403–413. doi:10.1016/j.molcel.2009.06.025