Structural studies on proteins involved in human telomere maintenance

Ufuk Borucu

2017

Ufuk Borucu. (2017). Structural studies on proteins involved in human telomere maintenance. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/72778>

<https://doi.org/10.32657/10356/72778>

Structural Studies on Proteins

Student Name: Ufuk Borucu

Signature & Date: ()

Supervisor Name: Professor Daniela Rhodes

Signature & Date: ()

i

Acknowledgements

I would like to thank Prof. Daniela Rhodes for giving me the opportunity to work under her supervision in the telomere and telomerase field, for which I followed her from MRC Laboratory of Molecular Biology in Cambridge to NTU School of Biological Sciences in Singapore. I'm thankful for the very interesting and challenging projects I was allowed to work on. Through many hurdles and difficulties, her support and guidance has been of great value and is very much appreciated.

I thank all of my friends and colleagues at NTU. I would like to thank Dr. Simon Lattmann and Dr. Clare Jelinska for sharing their experience and aiding me in my research, Asst. Prof. Sara Sandin and her group members for the access to the EM facility and invaluable assistance during my EM work and Yibo Jin for helping me with EM processing. I would also like to thank my dear friends and colleagues who, in our regular coffee sessions, have entertained me with insightful discussions about our research projects, science and other topics, and have never failed to give me emotional support and friendly advice. In particular, I would like to express my gratitude to Hariharan Sivaraman, Serap Beldar and Dr. Abbas El Sahili.

Finally, I would like to thank my mother and sister. They have supported me in my goals and ambitions in so many ways, and I am grateful for their love, respect, and of course their inexhaustible patience. And last, but not least, I want to express my tremendous gratitude to my beloved partner, Debbie, who I had the fortune to meet in Singapore and who has continuously shown me her deep love and unwavering support. She has been both a motivation and inspiration during my course work.

Abstract

In order to sustain genomic stability and cellular viability, the ends of linear eukaryotic chromosomes are capped and protected by telomeres. The mechanism by which the length of telomeric DNA is maintained involves a specialized reverse transcriptase, called telomerase. This thesis focuses on the structural investigation of two proteins that play essential roles in the regulation of different stages of telomerase activity: TCAB1 and the CST complex. TCAB1 participates in the maturation process of the telomerase RNA subunit and is critical for telomerase trafficking in vivo. Mutations in TCAB1 lead to defects in telomere maintenance and give rise to severe forms of dyskeratosis congenita (DC). The CST complex, consisting of Ctc1, Stn1 and Ten1, limits telomerase activity in the late stages of telomere elongation. CST also initiates the fill-in synthesis of the complementary strand by recruiting the Pol α /Primase complex. Mutations in CST are associated with Coat Plus, DC and related diseases. In this thesis, I present various strategies for the expression and purification of TCAB1 and the CST complex and structural analysis using negative stain electron microscopy (EM). Using primary sequence analysis and computational methods, I demonstrated that TCAB1 contains seven putative WD40 repeats within its central domain rather than the five published. Using various strategies for expression of TCAB1 in *E.coli*, that lead to aggregated protein, it was published that TCAB1 folding requires an elaborate machinery including the TCP-Ring Complex (TRiC) that is only present in higher eukaryotes. For the CST complex, I found that expression of full length Ctc1 in *E.coli* was not feasible and that it necessitated co-expression with Stn1 and Ten1. This was achieved by setting up MultiBac expression in insect cells. Results are presented for a purification strategy for obtaining a pure complex that contains stoichiometric amounts of the three proteins in mg quantities. As an initial stage to determining the cryo-EM structure, I present a low-resolution three-dimensional structure (25Å) of the CST complex obtained using negative stain EM and single-particle reconstruction. Finally, I describe strategies to obtain a higher resolution structure bound to telomeric DNA.

Table of Contents

Acknowledgements	ii
Abstract	iii
Table of Contents	iv
List of Tables	v
List of Figures	vii
Abbreviations	x
Chapter 1: Introduction	1
1.1) Initiation of DNA Replication	1
1.2) DNA Replication and the End-Replication Problem	3
1.3) Biological Role of Telomeres and Telomere Structure	6
1.4) The Telomerase Complex	11
1.5) Telomerase Recruitment and Telomere Elongation	16
1.6) The CST Complex	24
1.7) Structural and Functional Evolution of Telomeric Proteins	29
1.8) CST is an RPA-like Complex	37
1.9) TCAB1	46
1.10) TCAB1 is a WD40 Proteins	49
Chapter 2: Materials and Methods	54
2.1) General Materials and Methods	54
2.1.1) E.coli strains	54
2.1.2) Plasmids	55
2.1.3) Growth media	58
2.2) DNA Methods	59
2.2.1) Poly-chain reaction (PCR)	59
2.2.2) DNA detection	59
2.2.3) Ligation independent cloning (LIC)	59
2.2.4) DNA isolation	60
2.2.5) MultiBac cloning	60
2.2.6) DNA annealing	64

2.3)	Protein Methods	65
2.3.1)	E.coli culture	65
2.3.2)	Insect cell culture	67
2.3.3)	Tissue culture	70
2.3.4)	Protein purification	71
2.3.5)	X-ray crystallography	76
Chapter 3:	Structural Investigations on the CST complex	77
3.1)	Recombinant Expression of Ctc1 in E.coli	77
3.1.1)	Initial Ctc1 construct design and expression results	77
3.2)	Recombinant Expression of Ctc1 and the CST complex in Insect Cells	87
3.2.1)	Expression of Ctc1 alone is inadequate for protein preparation and structural studies	87
3.2.2)	Co-expression and purification of the CST complex: Problems and Solutions	90
3.3)	Purified CST complexes did not crystallize despite extensive screening	102
Chapter 4:	CST Structure Determination by Electron Microscopy	108
4.1)	Introduction	108
4.2)	25 Å Structure of the CST Complex following Negative Stain EM	112
Chapter 5:	Structural Investigations on TCAB1	120
5.1)	Recombinant Expression of TCAB1 in E.coli	120
5.1.1)	Initial TCAB1 construct design and expression results	120
5.1.2)	TCAB1 forms soluble aggregates that hamper purification after expression in E.coli	125
5.1.3)	Pairwise comparison of HMMs reveals seventh putative WD70 repeat	129
5.1.4)	Screening of various constructs and conditions does not solve aggregation of TCAB1	132
5.1.5)	TCAB1 requires the chaperonin TRiC for proper folding	139
Chapter 6:	Discussion	145
References		154

List of Tables

Table 2-1: Genotypes of used bacterial strains.	54
Table 2-2: Bacterial plasmid names with respective inserts, tags of inserts and resistance genes.	55
Table 2-3: Tissue culture plasmid names with respective inserts, tags of inserts and resistance genes.	57
Table 2-4: Composition of various growth media used.	58
Table 2-5: Supplements to growth media.	58
Table 2-6: Composition of buffers and components used for agarose gel electrophoresis.	59
Table 2-7: LIC annealing program.	60
Table 2-8: Composition of buffers used for bacmid isolation.	64
Table 2-9: Oligonucleotide annealing program.	64
Table 2-10: Composition of components used for auto-induction.	66
Table 2-11: Composition of buffers used in protein purification.	74
Table 3-1: Composition of crystallization plates.	103
Table 5-1: Sequence identity and similarity between various human and D.melanogster TRiC subunits.	141

List of Figures

Figure 1-1: Initiation of DNA replication at origins of replication	2
Figure 1-2: The end replication problem	5
Figure 1-3: Organization of the human shelterin complex.....	8
Figure 1-4: hTERT and hTER features and domain architecture	14
Figure 1-5: Schematic diagram of telomerase reaction cycle	19
Figure 1-6: Schematic illustration of telomere elongation after completion of DNA replication	21
Figure 1-7: Schematic illustration of telomere associated proteins from various species	30
Figure 1-8: Structural comparison between scRap1, TRF1 and TRF2 homeodomains and DNA recognition.....	32
Figure 1-9: Structures of Cdc13, POT1 and TEBP α / β OB-folds in complex DNA.....	35
Figure 1-10: Comparison of domain architecture between human Ctc1 and yeast Cdc13	36
Figure 1-11: Domain architecture of RPA and CST complexes	38
Figure 1-12: X-ray crystal structure of Stn1 and Ten1	39
Figure 1-13: X-ray crystal structure of RPA70 DNA binding OB-folds 2 and 3 in complex with octadeoxycytosine oligonucleotide and of the RPA trimerization core.....	42
Figure 1-14: Schematic diagram of RPA binding mechanism to single-stranded DNA	43
Figure 1-15: Proposed model of Ctc1 C-terminal OB-fold interactions with Stn1-Ten1 ...	45
Figure 1-16: WD40 repeats and domain structure.....	51
Figure 1-17: Side view of β -transducin WD40 domain	52
Figure 2-1: General MultiBac cloning strategy for three- to four-gene assemblies.....	62
Figure 3-1: Ctc1 sequence analysis using ELM resource and construct design.....	77
Figure 3-2: Total and soluble expression results of Ctc1 constructs submitted to PPP	78
Figure 3-3: GST-Ctc1 N3 as an example of N-terminal Ctc-1 construct expression and purification.....	80

Figure 3-4: GST-Ctc1 N1 as an example for refolding and purification of N-terminal Ctc-1 constructs.....	82
Figure 3-5: Expression test of C-terminal Ctc1 constructs together with Stn1 and Ten1 .	83
Figure 3-6: Expression and purification tests of C-terminal Ctc1 constructs in complex with $\Delta 18$ Stn1 and Ten1.....	85
Figure 3-7: Tissue culture expression screen of C-terminal Ctc1 constructs.....	88
Figure 3-8: Purification of GST-tagged Ctc1 C4 from insect cell culture	89
Figure 3-9: Optimization CST expression in insect cells	91
Figure 3-10: Issues with the purification of Ctc1 C3- $\Delta 18$ Stn1-Ten1 (C3 Δ ST) following co-expression in insect cells.....	93
Figure 3-11: Low TEV-cleavage efficiency of the Ctc1 GST-tag	94
Figure 3-12: Effects on Ctc1 C3- $\Delta 18$ Stn1-Ten1 (C3 Δ ST) purification following optimization of purification conditions.....	95
Figure 3-13: Increased TEV-cleavage efficiency of the Ctc1 GST-tag following optimization of purification conditions.....	97
Figure 3-14: Introduction of a Heparin purification step enables the purification of a stoichiometric CST complex.....	98
Figure 3-15: Purification strategy of CST complex from insect cells	99
Figure 3-16: Established purification strategy for Ctc1 C3- $\Delta 18$ Stn1-Ten1 (C3 Δ ST)	100
Figure 3-17: Established purification strategy for Ctc1 C1- $\Delta 18$ Stn1-Ten1 (C1 Δ ST)	101
Figure 3-18: Limited proteolysis of C1 Δ ST	104
Figure 3-19: Design of new Ctc1 constructs that comprise N-terminal OB-folds for crystallization trials	105
Figure 4-1: Negative stain EM micrographs before and after optimization of EM grid preparation	113
Figure 4-2: Picked particles and gallery of 2D class averages of cross-linked CST complex	114
Figure 4-3: Surface representations of CST 3D maps.....	115
Figure 4-4: Fourier Shell Correlation (FSC) plot for the refined EM density map	116

Figure 4-5: Fitting of Stn1-Ten1 OB-folds into CST surface map	118
Figure 5-1: TCAB1 sequence analysis using ELM resource	120
Figure 5-2: Sequence alignment of TCAB1 with orthologues from different organisms	121
Figure 5-3: Total expression results of TCAB1 constructs submitted to PPP	122
Figure 5-4: Soluble expression results of TCAB1 constructs submitted to PPP.....	124
Figure 5-5: Initial small-scale expression test of full-length TCAB1	125
Figure 5-6: Large-scale expression and purification of full-length TCAB1.....	127
Figure 5-7: Common degradation product of TCAB1 1-548, 1-501 and 1-450	128
Figure 5-8: TCAB1 140-510 tertiary structure as modeled by HHpred using templates 1VYH_C, 3JCR_L and 3DW8_B and comparison with 1VYH_C solved structure	130
Figure 5-9: Sequence alignment of TCAB1 with orthologues from different organisms and revised WD40 repeats.....	131
Figure 5-10: TCAB1 domain architecture according to ELM resource and HHpred model, and construct design.....	132
Figure 5-11: Two examples of GST-tagged TCAB1 construct expression and purification	133
Figure 5-12: Protein yields from 1 L culture of various GST-tagged TCAB1 constructs following GSTrap affinity purification and gel filtration	134
Figure 5-13: Third round of TCAB1 construct design and purification results	136
Figure 5-14: Mouse and Zebrafish TCAB1 sequence analysis and purification results...	138
Figure 5-15: Purification results of TCAB1 co-expression with GroEL/ES.....	140
Figure 5-16: Purification of His ₆ -GST-TCAB1 following expression in Sf9 cells	142
Figure 5-17: Screening and expression tests of new TCAB1 constructs in HEK293T and Sf9 cells	143

Abbreviations

μL	microliter
μm	micrometer
μM	micromolar
Amp	Ampicillin
AND-1	Acidic nucleoplasmic DNA-binding protein 1
ATM	Ataxia-Telangiectasia Mutated
ATR	ATM- and Rad3-related
BLM	Bloom syndrome protein
bp	base pair
BV	Baculovirus
CAB-box	Cajal Body-box
Cam	Chloramphenicol
CB	Cajal Body
Cdc13	Cell Division Cycle protein 13
CDC6	Cell Division Cycle 6
CST	Ctc1, Stn1, Ten1
Ctc1	Conserved Telomere maintenance Component 1
CTE	C-Terminal Extension
CTF	Contrast Transfer Function
d	day
DAT	Dissociates Activities of Telomerase
DBD	DNA-binding domain
DDR	DNA Damage Response
Dia	Dialyzed fraction
DNA	Deoxyribonucleic Acid
DNA-PK	DNA-dependent Protein Kinase
dNTP	Deoxynucleoside Triphosphate

DOM	Domain
e. g.	exemplī grātiā (for example)
EPDE	End Point Dilution Experiment
EDTA	Ethylenediaminetetraacetic Acid
ELM	Eukaryotic Linear Motif
EM	Electron Microscopy
En	Elution fraction n
EST	Ever-Shorter Telomeres
FBS	Fetal Bovine Serum
Fen-1	Flap Endonuclease 1
FISH	Fluorescence <i>In Situ</i> Hybridization
Fn	Fraction n
FSC	Fourier Shell Correlation
FT	Flow-Through
<i>g</i>	gravity of earth
Gen	Gentamycin
GFP	Green Fluorescent Protein
GST	Glutathione S-Transferase
h	hour
His6	Hexa-Histidine
HMM	Hidden Markov Model
HSP	Heat Shock Protein
HTH	Helix-turn-helix
In	Input
IPTG	Isopropyl β -D-1-Thiogalactopyranoside
Kan	Kanamycin
kb	kilobase
kDa	kilodalton
LC	Low Complexity

LIC	Ligation Independent Cloning
M	molar
MCM2-7	MiniChromosome Maintenance 2-7
MDC1	Mediator of DNA Damage Checkpoint protein 1
mg	milligram
min	minute
mL	milliliter
mM	millimolar
MOI	Multiplicity Of Infection
MRX	Mre11-Rad50-Xrs2
ms	millisecond
MS	Mass Spectroscopy
MWCO	Molecular Weight Cut-Off
NAF1	Nef-Associated Factor 1
ng	nanogram
NHP2	High mobility group-like Nuclear Protein 2
nM	nanomolar
NOP10	Nucleolar Protein 10
nt	nucleotide
OB	Oligonucleotide/Oligosaccharide-Binding
OD ₆₀₀	Optical Density at 600 nm wavelength
ORC	Origin Replication Complex
OT	Other
PAGE	Polyacrylamide Gel Electrophoresis
PCNA	Proliferating Cell Nuclear Antigen
PCR	Poly-Chain Reaction
Pol	Polymerase
Pol α	DNA Polymerase α
POT1	Protection Of Telomeres 1

PPP	Protein Production Platform
RAP	Repeat Addition Processivity
RAP1	Repressor Activator Protein 1
RC	Replication Complex
RFC	Replication Factor C
RIF	Rap1-Interacting Factor
RNA	Ribonucleic Acid
RNF8	RING Finger protein 8
RNP	Ribonucleoprotein
RPA	Replication Protein A
rpm	round per minute
RS	Resuspended cell pellet
RT	Reverse Transcriptase
scaRNA	small cajal body specific RNA
SDS	Sodium Dodecyl Sulfate
sec	second
siRNA	small interfering RNA
SMA	Spinal Muscular Atrophy
SMN	Survival Motor Neuron
SN	Supernatant
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
snRNP	small nuclear RNP
Spc	Spectinomycin
STEM	Scanning Transmission Electron Microscopy
Stn1	Suppressor of cdc Thirteen homolog 1
TAP	Tandem Affinity Purification
TCAB1	Telomerase Cajal Body protein 1
TCP-1	T-Complex Protein 1

TEN	Telomerase N-terminal
Ten1	Telomere length regulation protein homolog 1
TER	Telomerase RNA
TERRA	Telomeric Repeat containing RNA
TERT	Telomerase Reverse Transcriptase
Tet	Tetracycline
TEV	Tobacco Etch Virus
TIF	Telomer dysfunction Induced Foci
TIN2	TRF1-Interacting Nuclear factor 2
TMG	Trimethylguanosine
TPP1	TIN2-and POT1 interacting Protein 1
TRBD	TER-Binding Domain
TRE	Template-Recognition Element
TRF	Telomeric Repeat binding Factor
TRFH	TRF Homology
u	unit
Vol.	Volume
WCL	Whole Cell Lysate
wHTH	winged Helix-Turn-Helix
Wn	Wash fraction n
WRAP53	WD40 Repeat-containing protein Antisense to TP53 gene
WRN	Werner syndrome ATP-dependent
X-Gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside
YFP	Yellow Fluorescent Protein

Chapter 1: Introduction

This thesis focuses on the structural investigation of two proteins, TCAB1 and the CST complex, that are involved in the maintenance of extreme ends of linear DNA molecules called telomeres. In order to understand the biological significance of telomeres and their maintenance, it is important to review the context in which telomere biology operates. In the first part of this introduction, the mechanism by which DNA is replicated will be summarized. The intent of this section is not to give a detailed and extended review of DNA replication, but instead aims to provide a brief overview highlighting the problems that arise during the replication of telomeres and their implications. In the following sections of this introduction, I will discuss the structure of telomeres, their maintenance and the important roles that TCAB1 and the CST complex play during this process.

1.1) Initiation of DNA Replication

The size of eukaryotic genomes covers a wide spectrum ranging from approximately 12 million nucleotides in *Saccharomyces cerevisiae* to more than 3 billion nucleotides in a human cell. In order to rapidly duplicate such large genomes, eukaryotes comprise multiple chromosomes on which the genetic information is distributed. In addition, each chromosome contains many genomic locations at which the initiation of DNA replication is strictly coordinated. These locations are called origins of replication, and the initiation process is referred to as firing. Cells have to ensure that origins of replication fire in a spatially and temporally regulated manner and not more than once per cell cycle.

Owing to the fundamental importance of DNA replication, replication factors and the mechanism of DNA replication are highly conserved. However, the regulatory complexity increases significantly not only from prokaryotes to eukaryotes but also from lower to higher eukaryotes. One such difference lies in the nature of origins of replication, which are recognized by origin recognition complexes (ORC).¹ Whereas *Saccharomyces cerevisiae* origins of replication comprise a well-defined consensus sequence, *Saccharomyces pombe* ORCs preferentially bind to AT-stretches and *Drosophila melanogaster* ORCs to AT rich sequences, human ORCs appear to lack similar sequence preferences.²⁻⁷ Instead, human origins of replication are particularly rich in G-tetrads, a

Figure 1-1: Initiation of DNA replication at origins of replication. The figure illustrates sequential events that lead to initiation of replication and formation of replication forks. For the sake of simplicity, many protein factors that constitute the pre-RC or the DNA replication machinery are not depicted. 1) After the origin of replication is recognized by ORC, recruitment of MCM2-7, CDC6 and CDT1 in early G1-phase gives license for the formation of the pre-RC. 2) In late G1-phase, a spike in kinase activity triggers the assembly of DNA polymerase and a multitude of other replication factors at the licensed origin of replication. 3) Kinase mediated phosphorylation of MCM2-7 induces helicase activity and DNA unwinding. In addition, it triggers the recruitment and assembly of other protein factors that form the DNA replication machinery (not depicted). Thus, the pre-RC is converted into two replication forks that travel in opposite directions.

feature that is conserved among metazoans, which increase in density near origins that are utilized more frequently.⁸⁻¹⁰ Other factors contribute to the recognition of origins of replication, such as distal DNA elements,¹¹⁻¹⁴ chromatin structure,¹⁵⁻²³ transcription²⁴⁻²⁸ and DNA topology.²⁹

After origins of replications are recognized by ORCs, they are “licensed” by the sequential assembly of pre-replication complex (pre-RC) components CDC6, CDT1 and replicative helicase MCM2-7. The tightly regulated formation of pre-RCs occurs during M to G1-phase and not all origins of replication need to be utilized. When certain environmental and cellular conditions are met, a rapid spike in kinase activity signals the transition from G1 into S-phase, inducing the full assembly of pre-RCs, conversion into active replication complexes (RC) (called “firing”) and preventing the formation of further pre-RCs.³⁰⁻³² Upon firing, replicative helicases present in RCs begin to unwind parental DNA and together with other factors recruit components of the replication machinery. Thus, they form replication forks that propagate along parental DNA in both directions until the entire genome has been replicated (Figure 1-1).

1.2) DNA Replication and the End-Replication Problem

DNA replication is a semi-conservative process, in which parental DNA is separated into two single strands that each serve as a template for the synthesis of its counterpart. However, due to the antiparallel nature of double-stranded DNA, replication of each strand is fundamentally different. DNA synthesis proceeds in a 5′ to 3′ direction and requires an oligonucleotide primer with a free 3′-OH group to initiate. As the progressing replication fork unwinds parental DNA, the template of one strand (leading strand) is presented in a 3′ to 5′ direction, allowing the synthesis of its counterpart to occur continuously from a single initial primer. The template of the opposite strand (lagging strand), however, emerges in a 5′ to 3′ direction. The manner in which cells deal with this

complication is by multiple priming events followed by the synthesis of short DNA stretches (Okazaki fragments). Eventually, the DNA fragments that run complementary to the lagging strand are processed and ligated. This process raises the necessity to cycle between various polymerases and replication factors and adds a profound level of complexity to the entire process.

The replication machinery itself is composed of a multitude of proteins that form three essential elements: a DNA polymerase, a “sliding clamp” and a “brace”. First, RNA primers complementary to the parental DNA template are synthesized. In eukaryotes, this is carried out by the DNA Polymerase α /Primase complex, which also adds a stretch of 20-25 nucleotides of DNA with low fidelity.³³⁻³⁶ Replication factor C (RFC), the “brace”, then recognizes and binds to these primer-template junctions in an ATP-dependent manner, displacing Pol α /Primase in the process.³⁷⁻⁴¹ PCNA is loaded next to RFC and forms a ring shaped “sliding clamp”. It serves as a processivity factor which encircles the double-stranded DNA behind the primer-template junction and together with RFC mediates the assembly of high fidelity DNA polymerases Pol δ for lagging-strand synthesis or Pol ϵ for leading-strand synthesis.^{34,42-48} As the replication fork progresses, lagging-strand synthesis necessitates a frequent exchange between Pol δ and Pol α /Primase activity, which is coordinated by the trimeric protein Ctf4 in *Saccharomyces cerevisiae* and its ortholog AND-1 in humans.⁴⁹⁻⁵¹ Finally, Okazaki fragments are processed by Fen1 and RNaseHII and joined together via DNA ligase 1.^{33,52}

It is clear that DNA replication is a very complex, yet efficient process. However, during the duplication of telomeres, the very ends of DNA molecules, a fundamental problem arises. This problem, called the end-replication problem, was first described in studies of linear phage DNA replication.^{53,54} Assuming blunt ended DNA, it was predicted that newly synthesized lagging-strands would lack a portion of DNA minimally equivalent to the length of the extreme RNA primer due to its removal after Okazaki fragment synthesis (Figure 1-2 a).^{53,54} Later, it was found that most eukaryotes possess telomeres

that contain 3' overhangs.^{55,56} As a consequence, newly synthesized lagging-strands would undergo 5' trimming which would recover a 3' overhang identical to that of the parental telomeres. However, the leading-strand synthesis would resume all the way to the 5' end of the parental DNA, yielding a blunt ended and significantly shortened telomere that would then further undergo 5' to 3' resection of the parental strand for the recovery of a 3' overhang (Figure 1-2 b).⁵⁵⁻⁵⁷ As a result, telomeres undergo progressive shortening as cells proliferate, which can lead to deleterious effects such as loss of genetic material and genomic instability.

Figure 1-2: The end replication problem. Owing to the nature of DNA replication, extreme ends of linear chromosomes fail to be duplicated faithfully. As a result, they undergo progressive shortening with every cell division. Solid lines represent parental DNA strands, whereas dashed lines represent newly synthesized strands. RNA primer is annotated in green. Grey lines indicate original length of parental DNA. **a)** DNA end shortening of blunt ended DNA as postulated for linear phage replication. **b)** Telomere shortening as evidenced in eukaryotic cells. Image was adapted from source material from Liew *et al.*⁵⁸

In most eukaryotes, the common solution to the end-replication problem is the synthesis of telomeric repeats at the 3' overhang, followed by the fill-in synthesis of the complementary strand. This process occurs after replication in early S-phase and is carried out by telomerase and other telomere associated factors, which will be described in the following sections.

1.3) Biological Role of Telomeres and Telomere Structure

Telomeres are extreme ends of linear DNA molecules that serve as a buffer against DNA loss. They consist of repetitive guanine-rich sequences that extend beyond their complementary strand to form a 3' overhang.⁵⁹ In accordance with this sequence bias, the protruding strand is generally referred to as the G-strand whereas its complementary strand is referred to as the C-strand. In humans, telomeres contain 5-15 kb of double-stranded telomeric TTAGGG repeats followed by a 50-300 nucleotide 3' overhang.⁵⁹⁻⁶¹ Both single- and double-stranded regions of telomeric repeats provide binding sites for specific proteins that form a capping structure called shelterin. These capping structures are essential for protecting telomeres from DNA damage repair and aberrant end-to-end fusion events by providing a distinctive feature which helps distinguish DNA double-strand breaks from naturally occurring chromosome ends.⁶²⁻⁶⁹ Another essential function of shelterin is the regulation and maintenance of telomere length, mainly through the recruitment of telomerase, an RNA-containing reverse transcriptase that catalyses the step-wise addition of telomeric repeats to the 3' overhang.⁷⁰⁻⁷³ Failure to maintain telomere length disrupts the integrity of telomeric DNA and triggers cell cycle arrest (cellular senescence) or programmed cell death. As such, it has direct implications in aging and provides a critical check-point for the prevention of run-away proliferation, which is a hallmark of cancer.

Proteins that are identified as components of the shelterin complex need to fulfil at least two of three distinguishing criteria: They must only accumulate at telomeres, retain telomeric localization throughout stages of the cell cycle and perform functions that are limited to telomeres. Currently, there are six proteins known to be part of the shelterin complex. These include telomeric repeat binding factors (TRF) 1 and 2, TRF1-interacting nuclear factor 2 (TIN2), repressor activator protein 1 (RAP1), protection of telomeres 1 (POT1), and the POT1- and TIN2-organizing protein (TPP1) (Figure 1-3). However, there

are many other proteins that associate with telomeres which serve important roles, but are not considered to be part of the shelterin complex. Paradoxically, these include a number of factors which are components of DNA repair and recombination machineries.⁷⁴ Why they fail to stimulate DNA damage repair in a telomeric context is mostly unknown.

The specificity of the shelterin complex to telomeres is mediated by three of its components that show high affinities for telomeric repeats. These are double-strand binding proteins TRF1, TRF2 and single-strand binding protein POT1. TRF1 and TRF2 share a highly analogous domain architecture, comprising a TRF homology (TRFH) domain, a flexible hinge domain and a nearly identical SANT/Myb domain.⁷⁵⁻⁸³ Only the very N-terminus of TRF1 and TRF2 constitutes a noticeable exception to their structural similarity, featuring a negative charge in TRF1 and a positive charge in TRF2. TRF1 and TRF2 bind double-stranded telomeric DNA as homodimers. The specificity of TRF1 and TRF2 towards their substrate is mediated by the C-terminal SANT/Myb domain,^{81,84} whereas homodimerization is enabled by the TRFH domain.^{75,83} In addition to functioning as a homodimerization domain, TRFH domain also provides a binding site for F/YxLxP motif proteins and plays a critical role in recruiting the rest of shelterin components and other factors to telomeres.^{76,77} Interestingly, even though their TRFH domains share a high degree of homology, TRF1 and TRF2 do not form heterodimers and recruit different protein targets.

Rap1 associates with TRF2, which is mediated by interactions between TRF2 hinge domain and a C-terminal domain of RAP1. The formation of a complex with TRF2 is essential for RAP1 localization to telomeres and its stability.^{85,86} Otherwise, RAP1 is poorly characterized. Metazoan RAP1 contains a Myb-like domain which unlike its yeast counterpart is insufficient for DNA binding and may instead be involved in protein-protein interactions with yet unidentified factors.^{85,87-89} In addition, RAP1 contains an N-terminal BRCT domain, which is predominantly found in DNA-damage sensitive cell cycle

checkpoint proteins and is predicted to be involved in the recruitment of phosphorylated proteins.⁸⁵

Creating a bridge between shelterin components that recognize double-stranded and single-stranded telomeric DNA, respectively, TIN2 provides binding sites for TRF1, TRF2 and the POT1-TPP1 complex.⁹⁰⁻⁹⁵ TIN2 interactions with its binding partners are mediated by the association of its C-terminal FxLxP motif with TRF1 TRFH domain, its N-terminal region with TRF2 hinge domain and another N-terminal region with TPP1.⁷⁶ These interactions can occur simultaneously since they are not mutually exclusive.⁹¹ Thus, TIN2 serves as a scaffolding protein that exerts a stabilizing effect on the shelterin complex.

The final component of shelterin is the POT1-TPP1 complex. TPP1 association with TIN2 is thought to be the primary means by which POT1 is recruited to telomeres. Depletion of TPP1 or mutation of its interaction site with POT1 abolishes POT1 localization to telomeres *in vivo* and gives rise to phenotypes that are consistent with POT1 loss.^{73,96,97} Interestingly, even though POT1 is known to bind single-stranded telomeric repeats with nanomolar affinity, it appears to be insufficient for independent localization to telomeres.⁹⁸⁻¹⁰¹ One partial explanation for the necessity of TPP1 for telomeric localization of POT1 was provided by biochemical *in vitro* studies which have revealed that TPP1 stimulates POT1 DNA binding activity and increases its affinity for single-stranded telomeric DNA even further.¹⁰² Moreover, depletion studies of TPP1 have underlined the reliance of POT1 on active recruitment mechanisms for its telomeric localization *in vivo*.^{90,103}

Figure 1-3: Organization of the human shelterin complex. The figure gives examples of a number of different configuration in which shelterin components can arrange on telomeric DNA.

Naturally, the association of the shelterin complex to telomeres influences the shape and structure of telomeric DNA. It has been suggested that the telomeric G-strand is able to fold back and invade downstream double-stranded telomeric repeats, displacing the complementary strand and forming a lariat structure called t-loops.¹⁰⁴ T-loops were first visualized by electron microscopy of protein-free telomeric restriction fragments that were cross-linked through psoralen treatment *in vivo*.¹⁰⁴ Subsequently, native telomere chromatin fibers were shown to form t-loops, as well.¹⁰⁵ The formation of t-loops is reported to be facilitated by TRF2.^{104,106} However, TRF2 comprises neither ATPase nor helicase activity and therefore cannot be considered to drive the conformational change. Instead, this role is presumably assumed by other proteins that are involved in DNA recombination and accumulate at telomeres.¹⁰⁷ Overall, it is unclear whether t-loop formation is the primary architectural organization of endogenous telomeres, if only a subset of cellular telomeres adopt that conformation or if other configurations are more prevalent.

TRF1 has been reported to have DNA remodeling activity which is further enhanced by TIN2.^{75,84,108,109} The ability of TRF1 to bend and loop telomeric repeats *in vitro* might possibly contribute to the formation of t-loops. Alternatively, it might simply allow telomeres to fold back onto themselves,¹¹⁰ encouraging interactions between upstream POT1-TPP1 and downstream TIN2. Thus, the architectural organization of telomeres *in vivo* remains mostly speculative.

As mentioned earlier, an essential role of the shelterin complex is the protection of telomeric DNA from DNA damage response mechanisms. Shelterin dysfunction leads to the accumulation of DNA damage response factors at telomeres, which in turn causes the formation of Telomere Dysfunction Induced Foci (TIFs).^{93,111-113} For a very long time, it was assumed that shelterin protects telomeres from DNA damage repair by physically hiding it from DNA damage surveillance. However, it has become increasingly clear that shelterin components actively suppress DNA damage signaling pathways. Two shelterin

components that are known to regulate ATM (ataxia-telangiectasia mutated) and ATR (ATM- and Rad3-Related) kinase induced DNA damage response pathways, respectively, are TRF2 and POT1. In the absence of TRF2, ATM kinase becomes activated which leads to an upregulation of p53 and p21 mediated cell cycle arrest.^{86,114,115} POT1-TPP1, on the other hand, protects single-stranded telomeric DNA, presumably by coating its length, preventing the binding of RPA and thereby suppressing ATR mediated DNA damage signaling.^{96,97,116,117} The exact mechanistic pathways by which TRF2 and POT1 suppress DNA damage signaling, however, remains unclear and requires further research.

Finally, the shelterin complex also plays a regulatory role in the formation of G-quadruplex structures at telomeric and subtelomeric regions. G-rich DNA has the propensity to arrange into planar G-quartets via cyclic Hoogsteen hydrogen-bonds, which then pile into hydrophobic stacks that give rise to four-stranded helical structures, called G-quadruplexes.¹¹⁸ In general, the formation of G-quadruplexes is driven by the chelation of sodium and potassium cations within each G-quartet, whereas their stability depends on the number of G-quartets present in the structure and the length and composition of interconnecting loops.¹¹⁹ The telomeric G-strand readily arranges into G-quadruplexes within ~55 ms *in vitro*, which is in the approximate range of DNA replication (~20 ms/nucleotide),¹²⁰ suggesting a high potential for the formation of G-quadruplexes during cellular events that transiently generate single-stranded repeats, such as transcription and DNA-replication, and within the 3' overhang. Numerous studies, including fluorescence and radio-labeling with specific antibodies and ligands, have provided indirect evidence for the presence of G-quadruplexes at a subset of human telomeres *in vivo*.¹²¹⁻¹²³ However, since G-quadruplex specific antibodies and ligands can shift the equilibrium towards their formation, it is important to treat findings derived from such experiments with caution. On the other hand, multiple helicases that have been shown to unwind G-quadruplexes, including WRN (Werner syndrome ATP-dependent) and BLM (Bloom syndrome protein) helicase, accumulate at telomeres and are necessary for telomere integrity,¹²⁴⁻¹²⁷ further substantiating the existence of G-quadruplexes at telomeres *in*

vivo. The involvement of shelterin in the regulation of G-quadruplexes arises from its components TRF2 and POT1, which are implicated in the recruitment of WRN and BLM to telomeres through direct interactions and stimulation of their activity.^{128,129} In addition, the DNA binding activity of POT1, TRF1 and TRF2 passively affect G-quadruplex formation by shifting the equilibrium towards single- and double-stranded DNA, respectively.

1.4) The Telomerase Complex

The shelterin complex is essential for the regulation and maintenance of telomere length. In most eukaryotes, this is performed through the recruitment and action of telomerase, a unique reverse transcriptase (RT) that forms a complex with its own RNA template.¹³⁰⁻¹³³ Telomerase specializes in the addition of telomeric repeats to the 3' overhang of telomeres which serves to circumvent the loss of telomeric DNA due to inadequate chromosome end-replication and telomere end-processing. Cells that are highly proliferative, such as germ line cells, stem cells but also most cancer cells require the activity of telomerase to maintain telomere length and genomic integrity. As such, telomerase can be considered to be an oncogene that is down-regulated in somatic cells in order to limit cancer formation. This mechanism of tumor suppression, however, also paves the way for progressive telomere shortening in telomerase-inactive somatic cells, which limits the potential number of cell divisions, causes replicative senescence and contributes to aging of the organism.¹³⁴

The minimal telomerase holoenzyme consists of a protein (TERT) and an RNA subunit (TER). Additional accessory factors such as Dyskerin, NOP10, NHP2 and GAR1 are required for telomerase activity *in vivo* but not *in vitro*.¹³⁵⁻¹³⁷ In metazoans, TERT (telomerase reverse transcriptase) can be divided into four functional elements: telomerase N-terminal (TEN) domain, TER-binding domain (TRBD), a central RT domain and a C-terminal extension (CTE) (Figure 1-4 a). The TEN domain plays an important role

in telomerase recruitment to telomeres which contributes to telomerase repeat addition processivity.¹³⁸⁻¹⁴⁴ In addition, it has been shown to participate in telomerase enzymatic activity.¹⁴⁵⁻¹⁴⁷ A flexible linker connects the TEN domain to TRBD. Together with the RT domain, TRBD is necessary for TER binding and thereby essential for telomerase activity.¹⁴⁸⁻¹⁵⁵ The central RT domain is the catalytic core of TERT. It comprises seven characteristic sequence motifs (RT motifs) that are otherwise found in retrotransposons and retroviral RTs. These motifs are well conserved among TERTs from different organisms, but correspond less accurately with those of canonical RTs.¹⁵⁶ However, the conservation of structurally and functionally important amino acid residues between TERTs and RTs strongly indicates an architectural and phylogenetic relation.¹⁵⁷⁻¹⁶⁰ Excluding the N-terminal extension, TERTs organize into a structure that is common among RTs as well as RNA and DNA polymerases. Reminiscent of a right-hand shape, TERT RT domain organizes into proto-typical finger and palm domains whereas its CTE forms the putative thumb domain.^{130,161-163} In contrast to the RT domain, however, CTE has very little sequence conservation and is therefore thought to have species-specific functions. Interestingly, this assumption is compounded by the observation that while CTE plays a critical role in telomerase activity in humans, it has been shown to be dispensable or even missing in other organisms.¹⁶⁴⁻¹⁶⁸

The RNA component of telomerase is highly divergent between species both in length and in sequence, covering a wide spectrum that ranges from only 150 nucleotides in ciliates to more than 2000 nucleotides in some fungi. However, despite their divergence in primary sequence, TRs share functionally conserved secondary structure elements, which can be divided into three common domains: a core domain, CR4/CR5 domain and a box H/ACA domain (Figure 1-4 b).¹⁶⁹⁻¹⁷¹ The TER core domain is essential for telomerase activity.¹⁷² Mutations in this region can have a negative effect on fidelity, processivity and substrate recognition.^{173,174} The core domain consists of the RNA template required for telomeric repeat synthesis, 5'- and 3'-boundary elements that prevent the addition of non-telomeric sequences and a conserved and catalytically important pseudoknot

structure.^{171,175-181} The pseudoknot structure, together with the CR4/CR5 domain, mediates TER-TERT interactions. CR4/CR5 domain of TER associates with TRBD, whereas the pseudoknot region, with a significantly weaker affinity, interacts with the RT domain of TERT.^{146,182} These interactions are essential for telomerase activity.^{166,183,184} In fact, telomerase activity can be reconstituted *in vitro* by assembling TERT with a minimal TER fragment consisting of the core and CR4/CR5 alone.¹⁸⁵ Finally, the H/ACA domain located at the 3'-end of TER provides a binding site for accessory factors Dyskerin, NHP2, NOP10 and the molecular chaperon NAF1 as well as other protein factors involved in TER maturation and telomerase bio-assembly. The H/ACA domain is therefore essential for telomerase activity *in vivo*.¹⁸⁶ Interestingly, H/ACA elements are normally found in small Cajal body specific (sca-) and small nucleolar (sno-) RNAs where they guide and specify targets for the pseudouridylation activity of Dyskerin.¹⁸⁷ However, TER H/ACA domain has not been shown to have any such function, but instead plays a role in TER maturation, stability and subcellular localization.¹⁸⁸⁻¹⁹⁰

As mentioned earlier, the oncogenic potential of telomerase necessitates a tight regulation of its activity, which begins with the expression and bio-assembly of its subunits. Unfortunately, the study of hTERT spatiotemporal regulation has been hampered by the fact that its endogenous levels are very low, making it difficult to detect by conventional methods.¹⁹¹ Moreover, hTERT overexpression has been shown to disrupt its subcellular localization, rendering any observation under such circumstances artificial.⁷¹ According to our current understanding, hTERT associates with several proteins which, when knocked-down, reduce telomerase activity. These proteins include molecular chaperones HSP90 and p23 as well as AAA⁺ ATPases Reptin and Pontin.^{192,193} Immunopurification of HSP90 and p23 respectively co-purifies telomerase activity, suggesting that both chaperones form a complex with mature telomerase.¹⁹² Further studies on the role of HSP90 and p23 in telomerase maturation, however, are complicated by the fact that the inhibition of key chaperones is likely to induce wide ranging pleiotropic

Figure 1-4: hTERT and hTER features and domain architecture. **a)** Schematic illustration of hTERT domains. **b)** Depiction of hTER secondary structure. The core domain carries the template region for telomere repeat synthesis. hTERT TRBD binds to both core and CR4/CR5 domains. Dyskerin, NHP2, NOP10 and NAF1 assemble on the H/ACA domain, which is stimulated by the BIO-box located in the terminal loop of the H/ACA domain. The adjacent CAB-box serves as a binding site for TCAB1. Image was adapted from source material from Schmidt *et al.*¹⁹⁴

effects that are difficult to predict. Reptin and Pontin, on the other hand, also pull down hTERT, but these fractions are catalytically inactive.¹⁹³ This suggests that Reptin and Pontin might play a role during the biogenesis of telomerase but dissociate once the telomerase complex has matured.

Despite containing an H/ACA domain, the maturation process of hTER is strikingly unconventional. Whereas sca- and snoRNAs are excised from intronic regions of other mRNAs, hTER is transcribed by RNA polymerase II from its own promoter.¹³³ hTER lacks the typical 3' poly-A tail that is common for RNA polymerase II transcription products but instead undergoes 3' to 5' exonucleolytic cleavage.^{195,196} The boundary of the exonucleolytic step is defined by the H/ACA domain and associated factors, which include Dyskerin, NHP2, NOP10 and the molecular chaperone NAF1.¹⁹⁷ Their association is further stimulated by a specialized BIO-box motif located in the terminal loop of the H/ACA domain.¹⁹⁸ The formation of the H/ACA RNP occurs co-transcriptionally and plays an important role in hTER stability by protecting it from excessive nucleolytic cleavage.¹⁹⁹ Thus, it participates in the regulation of hTER accumulation and consequently telomerase activity.²⁰⁰ In addition to the BIO-box, the terminal loop of the H/ACA domain also contains a CAB-box motif. This region is specifically recognized by Telomerase Cajal body protein 1 (TCAB1) which drives the localization of hTER to Cajal bodies (CBs).^{201,202} There the m⁷G cap of hTER, which is a byproduct of RNA polymerase II dependent transcription, is modified into a 2,2,7-trimethylguanosine (TMG) cap. Furthermore, the molecular chaperone NAF1 of the H/ACA RNP is replaced by GAR1.^{187,203,204}

Importantly, the localization of hTER to CBs requires its prior association with hTERT.²⁰⁵ In addition, it has been shown that depletion of TCAB1 has no diminishing effect on telomerase activity *in vitro*.²⁰⁶ Together, these findings indicate that telomerase is assembled into a catalytically active RNP before it is targeted to CBs. TCAB1 depletion also doesn't seem to have any effect on hTER accumulation, negating any potential role TCAB1 might have in the stabilization of hTER.²⁰⁶ Yet, the localization of telomerase to CBs is a crucial prerequisite for telomerase activity *in vivo*, which is highlighted by the fact that mutations in TCAB1 are associated with telomerase deficiency.²⁰⁷ Depletion of coilin, an essential protein for CB formation, prevents telomerase recruitment to telomeres.²⁰⁸ This phenotype can be rescued by the overexpression of telomerase which is known to disrupt its subcellular localization. Depletion of TCAB1 exhibits the same phenotype, but cannot

be rescued by the same method.²⁰⁸ Collectively, these observations hint at a crucial role of CBs and especially TCAB1 for telomerase recruitment to telomeres *in vivo*.

1.5) Telomerase Recruitment and Telomere Elongation

In contrast to most enzymatic reactions in which enzymes encounter their substrates by diffusion, telomerase activity relies on active recruitment mechanisms since both telomerase and telomeres exist in a very low abundance. In HEK293T and HeLa cells, which comprise 184 telomeres, the number of telomerase molecules per cell has been estimated to be around 250 during S-phase at which hTERT expression peaks.¹⁹¹ In comparison, haploid *Saccharomyces cerevisiae* contains 64 telomeres and roughly 30 telomerase molecules.²⁰⁹

Human telomerase recruitment was first demonstrated through the identification of hTERT DAT (dissociates activities of telomerase) mutants that retained telomerase activity *in vitro* but failed to maintain telomere length *in vivo*.^{210,211} The TEN domain of hTERT, which contains a number of these DAT mutations, is required for telomerase localization to telomeres.^{144,210,211} On the telomeric side, shelterin component TPP1 enhances the affinity of telomerase for single-stranded telomeric DNA through telomerase interactions via its N-terminal OB-fold.⁷³ Mutational analysis more accurately described a patch of amino acid residues (TEL patch) which are crucial for the association of TPP1 and the TEN domain.^{71,72,212,213} In a compensatory charge-swap mutation study, a mutation in the TEN domain of hTERT (K78E) which disrupted telomerase interactions with TPP1 *in vivo* and *in vitro* was counteracted by a complementary mutation in TPP1 (E215K), proving that their association is direct and does not require a bridging factor.¹⁴⁴ In summary, telomerase recruitment to telomeres requires direct interactions between hTERT TEN domain and TPP1 N-terminal OB-fold, more specifically the TEL patch.

However, these interactions are not sufficient since, as described earlier, other factors such as TCAB1 also play an essential role.²¹²

Considering the close relationship between telomere length maintenance and DNA replication as well as the oncogenic potential of telomerase activity, it is not surprising that telomerase recruitment is strictly regulated and that this regulation is governed in a cell cycle-dependent manner. For most of the cell cycle, telomerase localizes to CBs and only accumulates at telomeres during S-phase.²¹⁴ Interestingly, this timing accurately coincides with the transient association of CBs and a subset of telomeres which, in contrast to the majority of telomeres, are highly accessible to G-strand specific FISH probes and able to co-localize with hTER during S-phase.²¹⁵ In addition, it has been documented that mammalian telomerases have a preference for short telomeres.²¹⁶⁻²¹⁹ Thus, it is reasonable to assume that the function of CBs is not only limited to hTER maturation and telomerase storage, but also includes trafficking of telomerase to a subset of structurally distinct, short telomeres. TCAB1 probably occupies a key role in this process, since it is required for telomerase recruitment to telomeres and *in vivo* activity, even when telomerase is overexpressed.²⁰⁸ However, it is unclear whether the role of TCAB1 is to mediate interactions between telomeres and telomerase or to facilitate post-translational modifications of relevant factors. Since the assembly and recruitment of telomerase are governed in a cell cycle-regulated manner, it is very likely for post-translational modifications to occur, especially phosphorylation. A recently identified phosphorylation site in the N-terminal OB-fold of TPP1 (S111) might represent a good candidate for cell cycle-dependent modulation of TPP1-telomerase interactions.²²⁰ However, experimental data that is currently available for HeLa cells and embryonic stem cells is rather ambiguous, necessitating more research in order to ascertain the significance of this residue.^{220,221}

Once recruited, telomerase needs to be situated at the very 3' terminal end of the G-strand before it can initiate telomeric repeat synthesis. Telomerase is a highly

processive enzyme and its reaction cycle consists of three elementary steps: primer recognition, repeat addition and translocation (Figure 1-5). Primer recognition and template positioning is facilitated by a single-stranded template-recognition element (TRE) immediately downstream of the TER template region.²²² Even though telomeric DNA is considered to be the natural substrate of telomerase, it was shown that telomerase can extend almost any G-rich single-stranded DNA primer *in vitro*.^{131,132,223,224} In fact, telomerase can extend virtually any single-stranded DNA as long as it contains telomeric repeats at its 5' end.^{223,225} Together, these findings indicate that the template region of telomerase RNA is not the sole requisite for primer recognition and that there are additional upstream "anchor sites" which participate in telomerase-substrate binding. Biochemical studies further revealed that primer-template hybridization has indeed minimal contribution to the stabilization of telomerase-substrate interactions. Instead, telomerase-substrate interactions are mainly stabilized by the protein subunit of telomerase.²²⁶ This is evident in the fact that hTERT can specifically recognize and interact with telomeric single-stranded DNA even in the absence of hTR.²²⁷ These interactions are mainly mediated by the first 300 amino acids of the TEN domain with bases approximately 20-22 nucleotides upstream of the primer 3' end, whereas residues 300-1132 comprising the RT domain were shown to confer both binding affinity but more importantly substrate specificity.²²⁷ The telomerase RNA subunit participates in the regulation of anchor site interactions by further enhancing telomerase-substrate binding affinity.²²⁸⁻²³⁴ In summary, telomerase recognizes its DNA substrate with both template-proximal and template-distal elements located on RNA and protein subunit.

Figure 1-5: Schematic diagram of telomerase reaction cycle. Telomere repeat synthesis is catalyzed by TERT, which contains a TEN domain (yellow), TRBD (green), RT domain (blue) and a CTE (purple). The TEN domain contains a telomere anchor site with single-stranded DNA binding activity (grey box). Only the template region of TER is illustrated within RT domain. After primer recognition, telomeric DNA which is now aligned with TER template, telomerase catalyzes the addition of nucleotides by reverse transcription. The telomeric primer is extended until it reaches the 5'-boundary element, whereupon telomerase translocates from its original position to the newly synthesized 3'-end.

Repeat addition entails the successive incorporation of dNTPs to the 3' end of telomeric DNA.²³⁵ As nucleotides are added at one end of the template, bonds are melted at the other, thus keeping the RNA-DNA hybrid at a constant length of seven to eight base pairs.²³⁶ Reaching the 5' template boundary element prompts the dissociation of the RNA-DNA hybrid and the relocation of telomerase to the newly synthesized 3' end of its DNA substrate, allowing a new cycle of telomeric repeat synthesis. The ability of telomerase to catalyze the synthesis of consecutive telomeric repeats without dissociating from its DNA substrate is commonly referred to as repeat addition processivity (RAP). Interactions that decrease telomerase dissociation rate and facilitate translocation have been reported to enhance RAP.^{232,233,237-243} These include interactions between anchor sites and telomeric DNA as well as interactions between the TEN domain and shelterin components POT1-

TPP1. Proteins and protein complexes that enhance RAP, such as POT1-TPP1, are considered to be processivity factors for telomerase activity.

In contrast to positive regulators, negative regulators have an inhibitory effect on telomerase activity. In a single round of telomere extension, human telomerase adds around 60 nucleotides to the telomeric 3' overhang.²⁴⁴ The mechanism for defining the upper limit of repeat synthesis was recently discovered to involve the CST complex, consisting of Ctc1, Stn1 and Ten1.²⁴⁵ Upon reaching a certain length, the telomeric primer is sequestered by the association of CST in late S-phase/early G2-phase. In addition, interactions between CST and POT1-TPP1 are thought to inhibit TPP1 mediated stimulation of RAP and telomeric localization of telomerase.²⁴⁵ CST then initiates C-strand fill-in by recruiting DNA Polymerase α /Primase complex to telomeres. Presumably, this step is then followed up by exonucleolytic processing of the newly generated telomeric 5' end (Figure 1-6). Another negative regulator of telomerase activity that does not involve a protein factor is TERRA (Telomeric repeat containing RNA), the transcription of which initiates in subtelomeric regions.²⁴⁶ TERRA resembles the natural telomeric primer as it carries UUAGGG repeats near its 3' end. Thus, it acts in *cis* as a competitive inhibitor of telomerase activity.²⁴⁷

The method by which telomere binding proteins cooperate to regulate telomerase activity has been explained through the “protein counting” model which was first hypothesized based on experimental observations in *S. cerevisiae*.²⁴⁸⁻²⁵¹ Mutations in *RAP1*, which in budding yeast binds to double-stranded telomeric repeats, as well as knock-out of two proteins that interact with *RAP1*, *RIF1* and *RIF2*, individually lead to abnormally long telomeres, indicating that these proteins were involved in telomerase inhibition.²⁴⁸⁻²⁵⁰ Coupled with the finding that in yeast, there is a preference to elongate short telomeres,²⁵² it was proposed that telomere binding proteins that were negative regulators of telomerase activity had an accumulative repressive effect that increased

Figure 1-6: Schematic illustration of telomere elongation after completion of DNA replication. **a-c)** POT1-TPP1 recruit telomerase to telomeres, where it recognizes the telomeric primer and initiates G-strand elongation. **d,e)** Association of CST to the telomeric overhang inhibits telomerase activity by sequestering its DNA substrate and inhibiting POT1-TPP1 mediated stimulation. In addition, CST recruits DNA Polymerase α /Primase complex to telomeres. **f,g)** Primase/Pol α complex localizes to telomeres and initiates C-strand fill-in. **h)** Exonucleolytic degradation of the newly synthesized C-strand regenerates the telomeric overhang.

with telomere length.²⁵¹ Similarly, it was found that in metazoans shelterin components TRF1, TRF2, TIN2 and POT1 also had an inhibitory effect on telomere elongation and that the preference for the elongation of short telomeres recurred.²⁵³⁻²⁵⁶ Thus, the “protein

counting” model was later adopted to explain metazoan telomere length regulation, as well. The long-distance inhibitory effect of distal telomere binding components was proposed to occur through looping events, be it t-loop formation or folding back of telomeres, the probability of which is higher for long telomeres than short telomeres.^{104,110,257,258}

An alternative model that was recently proposed by Carolyn Greider, called the “replication fork” model, suggests that telomerase travels with the replication fork for which there is evidence from ciliates.²⁵⁹ It was shown that in *Oxytricha*, telomerase associates with so called replication bands that coordinate replication initiation and progression.^{260,261} The “replication fork” model postulates that telomere binding factors increase the possibility of telomerase dissociating from the replication fork, and thus exercise an inhibitory effect. The longer the telomere, the more telomere binding proteins are encountered by the replication fork, increasing the probability of telomerase dissociation and thereby explaining the preference for the elongation of short telomeres over long telomeres. The “replication fork” model further takes into account the observation that RIF1, which has an evolutionarily conserved function in repression of origin firing,²⁶²⁻²⁶⁹ has an inhibitory effect on telomere elongation. Origins of replications that are located near telomeres often do not fire in S-phase and are instead passively replicated by replication forks from neighboring origins.^{270,271} In fact, the firing efficacy of subtelomeric origins of replication has closer correlation with the location of the origin than its DNA sequence.^{272,273} The “replicative fork” model suggests that RIF1 accumulates at telomeres and represses origin firing from proximal origins of replication. Hence, it increases the distance that the replication fork has to travel, and with it the probability of telomerase dissociation. In a proposed feed-back mechanism, long telomeres that contain a larger number of RIF1 are more likely to repress proximal origins of replications than short telomeres. The “replication fork” model is an interesting new approach to view how telomeric factors contribute to regulate telomere length elongation and accounts for replication related observations that were puzzling under the “protein counting” model.

However, it does not explain the requirement of CB dependent telomerase trafficking for telomerase activity *in vivo*, since there is currently no evidence showing a connection between CBs and replication forks. Nevertheless, the “replication fork” model provides an alternative view that helps broaden scientists’ perspective with which they regard telomere maintenance and interpret data. While it is important to investigate immediate interactions and effects between telomeric factors, regarding telomere length regulation in the context of DNA replication and DNA damage repair is crucial for a thorough understanding of telomere biology.

1.6) The CST Complex

The CST complex consisting of Cdc13, Stn1 and Ten1 was first identified in *Saccharomyces cerevisiae*, where it was found to play an essential role in telomere maintenance.²⁷⁴⁻²⁷⁶ Later, when homologues of Stn1 and Ten1 and their interaction partner Ctc1 (conserved telomere maintenance component 1) were identified in both plants and humans,²⁷⁷⁻²⁷⁹ it became evident that the CST complex had an evolutionarily conserved function in telomere length regulation. However, whereas CST components Stn1 and Ten1 are well conserved, major components Cdc13 and Ctc1 vary greatly in primary sequence and overall length.²⁸⁰ Consequently, even though both yeast and human CST play a role in the protection of telomeres, their precise telomeric functions partially vary from one another.

One of the earliest observations in CST biology stems from a strain of *Saccharomyces cerevisiae* carrying a temperature sensitive *cdc13-1* allele. The *cdc13-1* phenotype was reported to lead to a significant increase in chromosome loss and extensive resection of the telomeric C-strand.^{281,282} Concurrently, another class of *CDC13* alleles were found to reproduce an EST (ever-shorter telomeres) phenotype that characteristically disrupts telomere elongation while retaining telomerase activity *in vitro*.²⁸³ Accounting for both phenotypes, Cdc13 protein was shown to bind telomeric single-stranded DNA and engage in interactions with both DNA polymerase α (Pol α) and telomerase subunit Est1.²⁸⁴⁻²⁸⁶ Through its association with Est1, Cdc13 is directly involved in the recruitment of telomerase to telomeres and therefore essential for telomerase activity *in vivo*.²⁸⁷⁻²⁸⁹ In addition, Cdc13 prevents excessive resection of the telomeric C-strand by blocking exonucleolytic access and initiating Pol α mediated C-strand synthesis.²⁹⁰

Stn1 and Ten1 have been implicated in the regulation of Cdc13 and its functions at telomeres. The N-terminus of Stn1 comprises binding sites for both Ten1 and Cdc13, thus

providing a scaffold for CST complex assembly. Its C-terminus, on the other hand, is implicated in interactions with Pol α subunit Pol12.^{291,292} CST complex formation downregulates telomerase activity by competitive inhibition due to mutually exclusive Est1/Stn1 interaction sites on Cdc13.^{293,294} In addition, since Cdc13 binds to a different Pol α subunit (Pol1) than Stn1, formation of the CST complex further enhances the affinity for Pol α and promotes C-strand synthesis. Finally, Ten1 has been shown to have weak binding affinity for telomeric single-stranded DNA, which is proposed to enhance the DNA binding activity of the CST complex.²⁹⁵

In summary, budding yeast Cdc13 protects single-stranded telomeric DNA, inhibits exonucleolytic C-strand degradation and recruits telomerase via direct interactions with Est1. Binding of Stn1-Ten1 and formation of the CST complex shifts Cdc13 function to telomerase inhibition and promotion of C-strand fill-in. This shift is regulated in a cell cycle-dependent manner. During late S/G2 phase, phosphorylation of Cdc13 allows for interactions with Est1, whereas subsequent dephosphorylation promotes the formation of CST complex.^{291,296} Given that in *Saccharomyces cerevisiae*, there is a preference for the elongation of short telomeres,²⁹⁷ it stands to reason that there is a likely correlation between telomere length and CST complex formation. Indeed, it has been shown that MRX (Mre11-Rad50-Xrs2) preferentially binds to short telomeres and catalyzes 5' to 3' resection of the telomeric C-strand.^{298,299} This in turn allows for Cdc13 binding as well as recruitment of Tel1 kinase via association with Xrs2.³⁰⁰⁻³⁰² Subsequently, Tel1 phosphorylation of Cdc13 prevents CST complex formation and promotes Est1 recruitment to short telomeres.³⁰³

Since Cdc13 DNA binding domain contains an OB-fold that shares structural similarity with that of POT1, it had been assumed that the shelterin complex in higher eukaryotes has replaced the CST complex in budding yeast.^{99,304,305} However, genetic and structural studies later revealed that the CST complex is instead much more closely related to Replication Protein A (RPA) both structurally and functionally.^{306,307}

RPA is an evolutionarily conserved heterotrimeric complex that binds single-stranded DNA in a sequence unspecific manner and is involved in a wide spectrum of cellular processes, including DNA damage response and DNA replication.³⁰⁸ Its subunits RPA70, RPA32 and RPA14 contain multiple OB folds that both confer DNA binding activity and mediate protein-protein interactions.³⁰⁹ It was predicted that Stn1 and Ten1 comprised OB-folds that were similar to those contained in RPA32 and RPA14.³⁰⁶ Structural information later confirmed that in *Candida tropicalis* Stn1-Ten1 and RPA32-RPA14 homologues Rpa2-Rpa3 share substantial similarities in domain architecture and subunit interfaces, whereas the structure of loops in the OB-fold and relative orientation of subunits constitute the main differences between the two complexes.³⁰⁷ The most striking difference, however, is the presence of two wHTH motifs in Stn1 whereas Rpa2 contains only one, indicating a functional adaptation of Stn1.^{307,310,311}

Similarities between CST and RPA were of great value for the identification of human CST orthologues, which was significantly hampered due to their very low to absent sequence identity with known genes. Using bioinformatic approaches and data base searches, potential plant and human orthologues of scStn1 only showed 17.7% and 21.5% sequence identity and 54.4% and 59.5% sequence similarity, respectively.^{306,312} Structural analysis, however, showed remarkable conservation of a signature motif within the Stn1 OB-fold comprising 27 amino acid residues across several species as well as RPA32.³⁰⁶ Eventually, a combination of bioinformatic and genetic approaches uncovered the identity of Stn1 and Ten1 homologues in fission yeast, plants (*A.thaliana*) and humans and revealed their conserved roles in telomere maintenance.^{277-279,312} A genetic screen for mutations that give rise to telomere capping defects in plants identified Ctc1 for the first time, which was in turn used to find metazoan Ctc1.³¹³ And while the sequence identity and similarity proved to be low, secondary structure prediction produced a high resemblance, including a putative OB-fold with homology to RPA70. Subsequent studies then confirmed seemingly conserved telomeric functions.^{279,313}

The first information about human Ctc1 and Stn1 originates from research focusing on Pol α /primase complex, in which Ctc1 and Stn1 were characterized as Pol α accessory factors (AAFs) that increased Pol α processivity and affinity for its template.³¹⁴⁻³¹⁶ Depletion studies and immunofluorescence experiments later revealed CST to have both telomeric and extra-telomeric functions.^{279,313} Only 20% of CST localizes to telomeres, whereas the remaining 80% accumulates at nuclear foci with unclear identity. Furthermore, the subnuclear distribution of CST appears to be cell cycle-independent.²⁷⁹ And while CST has a strong affinity for single-stranded DNA that increases with the length of its binding site, its sequence specificity to telomeric repeats has been the subject of conflicting reports.^{245,279} Interestingly, CST is a functional complex, meaning that all three of its components are required for high affinity DNA binding, which resemble the DNA binding properties of RPA.^{308,311}

Knock-down of Ctc1 leads to accumulation of chromatin bridges, micronuclei and γ H2AX foci, indicating end-to-end fusion related defects in chromosome segregation that induce global DNA damage response and genome instability.³¹³ Furthermore, depletion of Ctc1 causes an increase in telomeric overhang length, compromises telomere integrity and gives rise to sporadic loss of telomeres.³¹³ When treated with Exonuclease I, which efficiently resects the telomeric G-overhang, Ctc1 depleted cells retain a portion of single-stranded G-strand within the internal regions of telomeric DNA.³¹³ These phenotypes clearly implicate CST in the maintenance of telomeres. Given that both human and budding yeast CST recruit Pol α to single-stranded DNA,^{274,290,314-316} it stands to reason that the initiation of C-strand fill-in is a conserved function of CST, as well. However, in contrast to the budding yeast Cdc13, human CST displays no clear sequence specificity or preference for telomeric localization.²⁷⁹ Therefore, the accumulation of non-telomeric γ H2AX foci and internal single-stranded repeats that occur upon Ctc1 knock-down are suggestive of a secondary role in cellular processes such as DNA replication, fork stalling

or DNA damage repair. The involvement of CST in these processes, however, remains inconclusive and requires further elucidation.

Telomerase activity ensues only after the completion of DNA replication.^{317,318} Since it is unlikely for canonical Pol α /primase recruitment factors (And1 and Mcm10) to remain localized at telomeres,³¹⁹⁻³²¹ it is necessary for CST to re-employ Pol α /primase following telomere elongation in late S/G2 phase. But the question remains: what are the determining factors that affect telomeric localization of CST itself? Recent studies have revealed an intriguing connection between CST and shelterin component POT1-TPP1. Immunoprecipitation experiments identified interactions between Stn1 and TPP1, implicating POT1-TPP1 in the recruitment of CST to the telomeric overhang.²⁷⁷ On the other hand, depletion of POT1-TPP1 increases telomeric association of CST, suggesting that these two complexes compete for G-overhang binding.²⁴⁵ Since CST binding affinity increases with the length of target DNA,²⁷⁹ newly elongated telomeric overhang may favor CST binding over that of POT1-TPP1. In addition, interactions between CST and telomeric POT1-TPP1 may further enhance its association. By sequestering both substrate and processivity factor, CST eventually limits telomerase activity and coordinates C-strand fill-in.²⁴⁵ Interestingly, this model draws a compelling parallel between human TPP1 and yeast Cdc13, in which both proteins first recruit and activate telomerase and subsequently engage a third protein to downregulate telomerase activity and initiate the synthesis of the complementary strand.

Mutations in *CTC1* were found to result in Coat Plus, DC and related bone marrow failure syndromes.³²²⁻³²⁴ These mutations were found to affect CST complex formation, DNA binding activity, interactions with Pol α or subcellular localization, all of which are crucial for CST function in telomere maintenance.³²⁵ In order to further our understanding of the biophysical mechanisms underlying CST function and test currently proposed models, it is necessary to elucidate its three dimensional structure, both in isolation and in complex with various interaction partners.

1.7) Structural and Functional Evolution of Telomeric Proteins

Even though telomere protection and length maintenance are essential processes that ensure viability and genomic integrity, it is surprising to find that telomeric protein factors have undergone rapid divergence and display very poor evolutionary conservation.³⁰⁴ Case in point, even among the select species that utilize chromosome end-capping complexes and telomerase mediated length maintenance, there are prominent discrepancies, such as telomeric DNA sequence, its length, the number and function of Shelterin protein components and the composition of telomerase (Figure 1-7). This thorough divergence has presented a significant impediment in attempts to identify a unifying theme of telomerase dependent telomere maintenance. However, with the increasing availability of structural information, it has become clear that certain structural elements are seeing a common usage across related, as well as divergent telomeric factors from various species. In this section, I will briefly discuss the structural elements of telomeric factors that specifically bind to single- and double-stranded DNA with an eventual inspection of the relationship between the human CST complex and other telomeric proteins.

Double-stranded telomeric DNA binding proteins are Rap1 in *Saccharomyces cerevisiae*, Taz1 in *Saccharomyces pombe*, and TRF1 and TRF2 in humans and metazoans.^{77-79,326,327} These proteins adopt multidomain structures, comprising a number of specific domains that are separately implicated in protein-DNA or protein-protein interactions. Apart from this commonality, however, they showcase several variances in their function, organization and roles at telomeres. scRap1 and TRF1 perform a negative regulatory role on telomerase activity,^{253,328,329} whereas the primary function of TRF2 and Taz1 appears to be end-capping and telomere protection.^{326,330} Furthermore, whereas TRF1 and TRF2 bind to double-stranded telomeric DNA as homo-dimers, scRap1 binds as a monomer. In the case of Taz1, high flexibility regions have hampered the elucidation of

Figure 1-7: Schematic illustration of telomere associated proteins from various species. Colored elements suggest homology and/or functional parallels (see text), whereas grey elements do not.

its full-length structure. However, based on proteolysis studies and biochemical characterization, it is very likely that Taz1, similar to TRFs, acts as a homodimer.^{83,108}

Structural information on the DNA-binding domains (DBDs) of scRap1, TRF1 and TRF2 in complex with their DNA substrates reveals a common feature that specifically recognizes double-stranded DNA.^{80,331-336} This feature comprises a bundle of three α -helices, the second and third of which form a helix-turn-helix (HTH) motif. The third α -helix inserts itself into the major groove of double-stranded DNA and engages in base-specific interaction. In addition, amino-terminal arms contained on all three helices contact specific bases in the minor groove. Originally, TRF1 and TRF2 DBDs were named as Myb-domains, since their identification was derived from high sequence similarities to

the Myb-DNA binding motif.^{78,337,338} However, the presence of the aforementioned amino-terminal arms clearly distinguishes these DBDs as homeodomains.

scRap1, TRF1 and TRF2 all bind to double-stranded DNA through two separate homeodomains (Figure 1-8). But whereas TRF1 and TRF2 each carry only one such domain and therefore function as homodimers, scRap1 DBD comprises a bipartite structure consisting of two tandem homeodomains.^{80,331,334} The two scRap1 homeodomains wrap around telomeric DNA in an orthogonal anti-clockwise orientation. They are spaced at 8 base pairs from each other and each contact a different stretch of 8 base pairs in a sequence-specific manner (Figure 1-8 a). The binding site of a TRF1 homodimer, on the other hand, comprises only 15 base pairs with each DBD contacting a specific stretch of 7 base pairs.^{80,84,335} Similar to scRap1, TRF1 and TRF2 DBDs bind to double-stranded DNA at a rotation. However, instead of a 90° angle, they are arranged at 205°, placing them at roughly opposite sides of the DNA strand (Figure 1-8 b). The rotational arrangement that is a common feature between scRap1, TRF1 and TRF2 DBDs allows tight packaging of multiple proteins along telomeric DNA with a spacing that is smaller than their respective binding sites. In summary, despite being non-homologous proteins, scRap1, TRF1 and TRF2 share significant commonalities in the type of their DBDs and their rotational arrangement around their respective DNA binding sites.

Shelterin components that specifically bind to single-stranded telomeric DNA are TEBP α and β in *Oxytricha nova*, Cdc13 in *Saccharomyces cerevisiae*, Pot1 in *Saccharomyces pombe* and POT1 in humans and metazoan. Except for budding yeast Pot1 and human POT1, these proteins show considerable variation in primary sequence as well as their function at their respective telomeres, indicating that profound divergence has taken place over the course of evolution. However, similar to double-strand binding shelterin components, these proteins share a DBD motif that is ubiquitous among single-stranded DNA binding proteins.

Figure 1-8: Structural comparison between scRap1, TRF1 and TRF2 homeodomains and DNA recognition. The crystal structures of **a)** scRap1, **b)** TRF1 and **c)** TRF2 homeodomains in complex with double-stranded telomeric DNA illustrates a conserved concept of DNA interaction, which involves two DBDs binding at a rotational arrangement. The third α -helix of the homeodomain (highlighted in red) is inserted into the major groove of double-stranded DNA and engages in base-specific hydrogen-bond interactions. scRap1 contains two homeodomains arranged in tandem and binds to telomeric DNA as a monomer. hTRF1 and hTRF2 each contain a single homeodomain and bind to their target DNA as homodimers.

In ciliates, TEBP α can dimerize or form a heterodimer with TEBP β . Both complexes are capable of capping the protruding G-strand by specifically binding to the telomeric overhang,^{339,340} but each play different roles in telomere maintenance. The TEBP α homodimer (TEBP α/α) is organized in a head-to-tail arrangement. Each subunit comprises telomeric DNA binding sites on opposite faces of the dimer. This then allows the TEBP α/α -mediated association of a pair of telomeres into a higher order structure.³⁴¹ In contrast, TEBP α/β only binds to a single telomeric G-strand.³⁴² Importantly, TEBP α/β capping of telomeric ends buries the 3'-terminal OH group and thereby renders it inaccessible to telomerase. Consequently, whereas both TEBP dimers restrict telomerase access to extreme telomere ends, TEBP α/β inhibits telomerase activity much more strongly than TEBP α/α , a duality that plays a crucial role in *Oxytricha nova* telomere maintenance.³⁴³

Unlike TEBP α / α and TEBP α / β , which only form capping structures at the very end of telomeres, budding yeast Cdc13 is thought to coat and protect the entire length of the G-overhang. In addition, Cdc13 inhibits exonucleolytic degradation of the complementary C-strand. While phosphorylated, Cdc13 recruits telomerase to telomeres through direct interactions with Est1. Upon dephosphorylation, Cdc13 associates with Stn1-Ten1 and forms the budding yeast CST complex. Since Est1/Stn1 interaction sites on Cdc13 are mutually exclusive, the formation of the CST complex effectively downregulates telomerase activity.^{293,294} Furthermore, CST complex formation leads to an enhanced affinity to Pol α , thereby promoting C-strand synthesis.

In a fashion, the function of POT1 resembles that of yeast Cdc13, but also exhibits evident discrepancies. POT1 binds single-stranded telomeric DNA with nanomolar affinity. However, its localization to the telomeric overhang is aided by its association with TPP1.^{73,96-101} Similar to Cdc13, POT1-TPP1 protects telomeric DNA by coating its length and suppressing ATR mediated DNA damage signaling.^{96,97,116,117} In addition, POT1-TPP1 is a processivity factor for telomerase activity by providing anchor sites for telomerase through direct interactions, thus enhancing telomerase repeat addition processivity.^{239,243}

Despite having different functions and not being homologous proteins, structural elucidation of TEBP α / β , Cdc13 and POT1 identified a seemingly conserved single-stranded DNA binding domain, the OB-fold.^{68,344,345} Superposition of their respective OB-folds reveals a central core consisting of a β -barrel fold with a C-terminal α -helical extension. However, regions residing outside of this core display considerable variation in structural architecture. In all structures, single-stranded DNA binds to a concave groove on the face of the β -barrel and is enveloped by finger-like protrusions that extend from loops between β -strands 1 and 2, as well as β -strands 4 and 5. Contacts with amino acid residues on the two flanking loops locks the DNA strand in place (Figure 1-9). The binding of single-stranded DNA to the OB-folds is stabilized by stacking interactions between bases and hydrophobic amino acid side-chains, as well as between bases themselves. Thus, the DNA

strand is buried within the OB-fold, leaving its phosphate backbone to form an extended solvent-exposed ridge. Despite the common core structure and concept of DNA-binding, however, the detailed mode of DNA-binding varies for each OB-fold structure.

The single DNA-binding OB-fold of Cdc13 binds to a relatively long stretch of single-stranded DNA (d(GTGTGGGTGTG)). This is achieved through an extended loop region between β -strands 2 and 3, which increases the DNA-binding surface on the OB fold (Figure 1-9 a-c).³⁴⁴ POT1, on the other hand, binds to its target DNA consisting of 9 nucleotides (d(TAGGGTTAG)) with two tandem OB-folds. The second OB-fold is tilted and rotated so that the single-stranded DNA traces a more or less straight line through both DNA binding motifs (Figure 1-9 d-f).⁹⁹ Comparison between human POT1 and fission yeast Pot1 reveals similarities in protein-DNA contacts and intramolecular base stackings. However, in contrast to fission yeast Pot1, the minimal binding site of human POT1 does not comprise two Guanine bases on the upstream end of the telomeric DNA stretch. POT1 contacts two additional Guanine residues between its two OB-folds. This is likely to be true for its fission yeast counterpart, as well. However, since the only available structural information of Pot1 binding to DNA is limited to a single OB-fold, this remains to be demonstrated conclusively. The mode in which TEBP α/β caps telomeric DNA is very different, still. The first two OB-folds of TEBP α (OB1 and OB2) together with the single OB-fold of TEBP β form a deep DNA-binding pocket (Figure 1-9 g).⁶⁸ In complex with its minimal binding site (d(GGGGTTTTGGGG)), the two tandem OB-folds of TEBP α contact G1-T5, T10, T12 and G10-G12, whereas TEBP β OB-fold interacts with the intermittent nucleotides T5-G10. In this structure, the DNA strand is folded back so that OB1 of TEBP α contacts both G1 and G12 (Figure 1-9 h-j). Compared to Cdc13 and POT1, the TEBP α/β -DNA complex displays a more saturated map of interactions between bases and hydrophobic side chains as well as interbase stackings. Thus, the TEBP α/β buries the 3'-end of the DNA strand deep within its DNA-binding cleft, rendering it inaccessible.

Figure 1-9: Structures of Cdc13, POT1 and TEBP α / β OB-folds in complex DNA. The gallery of structures depicts OB-folds of various single-stranded telomeric DNA binding proteins in complex with their target DNA. Loop regions between β -strands 1 and 2, as well as 4 and 5 (highlighted in red) form finger-like protrusions that extend and wrap around the DNA strand. **a-c)** Cdc13 DBD consists of a single OB-fold. The extended loop region between β -strands 2 and 3 (highlighted in yellow) increases the OB-fold surface area for DNA-binding surface. **d-f)** POT1 DBD comprises two tandem OB-folds. The second OB fold is tilted and rotated in relation to the first. **e and f)** A side view of OB-fold 1 illustrates the degree to which it envelops the DNA strand. **g-j)** OB-folds 1 and 2 of TEBP α and OB-fold 1 of TEBP β bind DNA in a deep protein cleft. G12 at the 3'-end of the DNA strand (highlighted in red) is buried within the DNA binding pocket. **h-j)** Views of TEBP α OB-fold 1 from different angles illustrates how the DNA strand bends around and contacts this domain both with G1 and G12.

Compared to the single-stranded DNA binding proteins described above, the human CST complex displays significant variations, but also some striking similarities. Among its three components, Stn1 and Ten1 are most conserved. They display 17.2% and 21.6% sequence identity with their fission yeast counterparts, respectively. In addition, the OB core structures of yeast and human Stn1 and Ten1, as well as their functional arrangement into a stable heterodimeric complex are very similar.²⁸⁰ In contrast to Stn1 and Ten1, the third component of the CST complex, Ctc1 and Cdc13, vary greatly in primary sequence, overall length and domain organization.²⁸⁰ Furthermore, whereas the DBD of Ctc1 is localized at its N-terminus and is comprised of two putative OB folds, Cdc13 binds single-stranded telomeric DNA through a single OB-fold close to its C-terminus (Figure 1-10).³⁴⁴ This discrepancy suggests that there may be significant differences in how human and yeast CST complexes engage in interactions with telomeric DNA. However, in the absence of structural information on either CST complex, this remains to be seen.

Figure 1-10: Comparison of domain architecture between human Ctc1 and yeast Cdc13. RD: Recruitment Domain. Both Ctc1 and Cdc13 utilize OB-folds for protein-DNA and protein-protein interactions. However, Ctc1 shares no homology with Cdc13. Ctc1 DBD comprises two OB-folds, whereas Cdc13 DBD only one. The proteins bind to single-stranded DNA and Stn1 at opposing ends of the primary sequence.

When it comes to their roles at telomeres, human and yeast CST complexes again display noticeable similarities and differences. Both CST complexes, when fully formed, inhibit telomerase activity and initiate synthesis of the complementary strand through the recruitment of Primase and Pol α .^{291,292,314-316} However, the nature of the regulatory switch from telomerase activation to inhibition varies between the two complexes. In yeast, Cdc13 is directly involved in the recruitment of telomerase to telomeres through its

association with Est1.²⁸⁷⁻²⁸⁹ Upon formation of the CST complex, telomerase activity is downregulated via competitive inhibition due to mutually exclusive Est1/Stn1 interaction sites on Cdc13.^{293,294} In humans, on the other hand, the recruitment of telomerase to telomeres is managed through direct interactions with TPP1. Following telomerase activity, when the telomeric overhang reaches a certain length, CST localizes to telomeres and interacts with POT1-TPP1.²⁴⁵ Thus, it limits telomerase activity through competitive inhibition. Finally, whereas Cdc13 associates with yeast telomeres throughout the cell cycle in a specific manner, the human CST complex forms a functional complex of which only 20% localizes to telomeres, and 80% to other genomic loci. This calls into question the sequence specificity of the human CST complex to telomeric repeats.^{243,311}

In summary, the homeodomain and OB-fold are two DNA binding motifs that see an evolutionarily conserved usage among double-stranded and single-stranded telomeric DNA binding proteins, respectively. The DBD of Ctc1 consists of two putative OB-folds. This functional organization is more similar to that of POT1 than that of Cdc13. Telomere binding proteins are highly divergent, both in function and structural organization. However, they appear to utilize similar principles in binding to telomeric DNA. With regards to Ctc1, it would be interesting to elucidate the molecular mechanism by which it binds to single-stranded DNA. Structural information on human CST interactions with DNA would allow us to draw parallels to other single-stranded DNA binding proteins, such as POT1 and Cdc13, and potentially delineate evolutionary relationships.

1.8) CST is an RPA-like Complex

Both in yeast and humans, the CST complex bears structural and compositional similarity to RPA, a heterotrimeric complex consisting of RPA70, RPA32 and RPA14, that binds to single-stranded DNA in a sequence unspecific manner and is involved in a variety of cellular processes, such as DNA replication and DNA damage response.³⁰⁸ Both CST and

RPA contain multiple OB-folds that mediate DNA binding and protein-protein interactions.³⁰⁹ Stn1 and Ten1 display significant similarity with RPA32 and RPA14, respectively, including structural and compositional organization as well as weak DNA binding activity.^{279,280,306,307} The most evident discrepancy between Stn1 and RPA32 is the presence of two WHTH domains on the C-terminus of Stn1 compared to only one in RPA32.³⁰⁷ Ctc1, on the other hand, is noticeably different from RPA70 in that it is a substantially larger protein and is predicted to comprise three putative OB-folds, whereas RPA70 contains four.²⁷⁹ Nevertheless, Ctc1 and RPA70 display functional parallels in their composition. In both proteins, the C-terminal OB-fold partakes in complex formation, while a pair of internal OB-folds engage in strong DNA binding activity (Figure 1-11).^{279,308,346}

Figure 1-11: Domain architecture of RPA and CST complexes. Interactions between subunits is indicated by shaded areas. OB-fold interactions with single-stranded DNA and protein interaction partners are highlighted. The image burrows from source material by Chen *et al.*³⁴⁷

As mentioned earlier, Stn1 and Ten1 are the most conserved components of the CST complex, exhibiting a 17.7% and 21.5% sequence identity and 54.4% and 59.5%

sequence similarity with their fission yeast counterparts, respectively. Their ability to form a stable heterodimeric complex *in vitro* was of great advantage in the recent elucidation of their molecular structure using X-ray crystallography, which revealed a strong structural similarity to yeast Stn1 and Ten1, as well as RPA32 and RPA14 (Figure 1-12).²⁸⁰ The assembly of Stn1 and Ten1 is mediated by extensive interactions between highly conserved α -helices located on the C-terminus of their respective OB-folds, and additional auxiliary contacts between the β -barrel bodies of these domains. The association of Stn1 and Ten1 results in their putative DNA binding pockets to be aligned side by side

Figure 1-12: X-ray crystal structure of Stn1 and Ten1. **a)** Crystal structure of Stn1 and Ten1 OB-folds in interaction with each other, mainly mediated by extensive contacts between the two highly conserved C-terminal helices. **b)** Rotated view of the Stn1-Ten1 complex that highlights the parallel alignment of putative DNA binding pockets of both Stn1 and Ten1 OB-folds. **c)** Crystal structure of C-terminal wHTH domains of Stn1. **d)** Alignment of human Stn1-Ten1, human RPA32-RPA14 and fission yeast Stn1-Ten1 OB-folds reveals strong structural and compositional conservation.

(Figure 1-12 b), thereby combining them into an elongated DNA binding cleft. However, unlike its yeast counterpart which binds DNA with low affinity, human Ten1 reportedly lacks any DNA binding activity, whereas Stn1 was shown to bind single-stranded DNA with micromolar affinity in a sequence unspecific manner.²⁸⁰ Interestingly, WTH motifs in yeast Stn1 have been implicated in interactions with Pol α and Cdc13.^{296,348,349} But in order to verify the likely possibility of a functional conservation in human Stn1, further structural and biochemical investigations are necessary.

Unlike Stn1-Ten1, there is currently no available structural data for the largest component of the CST complex, Ctc1, which of all CST subunits confers the strongest DNA binding affinity. Given the compositional and functional similarity between Ctc1 and RPA70 and the structural conservation of Stn1-Ten1, it is reasonable to assume that the CST complex is an RPA-like protein complex that has evolved telomere specific functions. However, only 20% of CST localizes to telomeres, and due to conflicting reports, it is currently unclear whether CST exhibits sequence specificity to telomeric or G-rich repeats.^{245,279} Instead, the only aspect of CST that confers telomere specificity appears to involve interactions between CST and POT1-TPP1, which may serve to favor its telomeric localization.²⁷⁷ Interestingly, depletion of POT1-TPP1 enhances CST accumulation at telomeres, presumably due to the loss of POT1-TPP1 coating of the G-overhang, and increases the accessibility of CST binding sites.²⁴⁵

Despite the lack of structural information on Ctc1, inferences from RPA70's mode of action in single-stranded DNA binding and interaction with RPA32 and RPA14 can provide some clues as to the structure and function of the CST complex. Biochemical and structural studies have identified a multistep mechanism by which RPA binds to single-stranded DNA that involves substantial conformational changes. First, OB-folds 2 and 3 of RPA70, which similarly to Ctc1 confer the strongest DNA binding activity of the trimeric complex, bind to 8-10 nucleotides of single-stranded DNA.³⁵⁰⁻³⁵² The formation of the RPA_{8nt} complex, which is relatively unstable, is followed by a significant conformational

change in which RPA aligns itself along the DNA, allowing direct interactions with 23-27 nucleotides. In this conformation, the subdomains of the RPA_{30nt} complex that mediate DNA binding include RPA70 OB-folds 2, 3 and 4, as well as RPA32 OB-fold 1.³⁵³⁻³⁵⁹ Using scanning transmission electron microscopy (STEM), RPA_{8nt} complexes appear globular in shape which allows them to be closely packed, whereas RPA_{30nt} complexes exhibit significant configurational flexibility depending on the length of the single-stranded DNA binding partner.³⁵⁸ Interestingly, the conformational change from RPA_{8nt} to RPA_{30nt} complex exposes a DNA-PK phosphorylation site, which may be of relevance in the intracellular regulation of RPA during double-stranded DNA damage repair and V(D)J recombination.

Co-crystallization of the two single-stranded DNA binding OB-folds of RPA70 with an octadeoxycytosine oligonucleotide revealed their parallel orientation, which creates a channel that allows DNA to be positioned in an approximately straight line with the sugar-backbone forming an outward-facing ridge (Figure 1-13 a and b).³⁶⁰ Each domain encompasses three nucleotides, whereas the gap between the OB-folds holds two. The three nucleotides that reside within each OB-fold display a very similar arrangement, in which their structure is maintained by extensive hydrogen bonds with RPA70 as well as stacking interactions with phenylalanine residues. An interesting feature of this structure entails the first β -strands of both OB-folds, which are broken into two by a short linker. In addition, the $\beta 1'$ - and $\beta 2$ -strands in each domain form a finger-like protrusion containing a β -hairpin that extends around the DNA and further stabilizes binding (Figure 1-13 a and b).

Structural information on the trimerization core of RPA, which is composed of RPA70 OB-fold 4, RPA32 OB-fold and RPA14, identifies the trimerization interface to consist of three conserved C-terminal α -helices that are arranged in parallel, engaging in extensive hydrophobic interactions (Figure 1-13 c).³⁰⁹ OB-fold 4 of RPA70 shares a similar structure with OB-folds 2 and 3, suggesting a comparable single-stranded DNA binding

mechanism. Interestingly, in the crystal, two trimeric complexes are packaged in a way that the three helices of each trimer form a complementary interface, giving rise to a hexa-helical arrangement (Figure 1-13 d). In accordance with electron microscope images that suggest a 6-fold symmetry of individual RPA complexes,³⁶¹ this particular crystal packaging

Figure 1-13: X-ray crystal structure of RPA70 DNA binding OB-folds 2 and 3 in complex with octadeoxycytosine oligonucleotide and of the RPA trimerization core. **a)** Co-crystal structure of RPA70 OB-folds 2 and 3 and an octadeoxycytosine oligonucleotide. In both OB-folds, a β -hairpin that connects $\beta 1'$ - and $\beta 2$ -strands protrudes and wraps around the oligonucleotide. Each OB-fold holds three nucleotides, whereas the gap between the OB-folds contains two. **b)** Side-view of the co-crystal structure that illustrates the extent of the channel created by the protein interface wrapping around the oligonucleotide. **c)** Crystal structure of the RPA trimerization core. The interface composed of three α -helices is annotated. **d)** Crystal packing of the RPA trimerization core features a C_2 symmetry axis. The red dashed circle highlights a hexa-helical arrangement consisting of two tri-helical bundles.

may indicate RPA70 OB-folds 1, 2 and 3 to contribute a three-helix bundle that interacts with that of the trimerization core. After single-stranded DNA binding by RPA70 OB-folds 2 and 3, the transition from the globular RPA_{8nt} into the elongated RPA_{30nt} complex may entail the disruption of the hexa-helical arrangement and an “unfolding” of the globular structure that brings the trimeric core in contact with DNA (Figure 1-14).

Figure 1-14: Schematic diagram of RPA binding mechanism to single-stranded DNA. **a)** In line with biochemical and structural data,^{309,361} RPA is likely to form a globular complex with a near 6-fold symmetry. The conserved C-terminal α -helices of OB folds in RPA70, RA32 and RPA14 are illustrated as protrusions facing inwards, whereas the wHTH domain of RPA32 is indicated with a protrusion facing outwards. **b)** Initial binding of RPA to single-stranded DNA involves OB-folds 2 and 3 of RPA70, thus forming a RPA_{8nt} complex. **c)** Substantial conformational changes result in the formation of an elongated RPA_{30nt} complex that aligns itself along the DNA and significantly enhances DNA binding affinity through the engagement of additional OB-folds.

Electromobility shift assays revealed that Ctc1, Ctc1-Stn1 and Stn1-Ten1 alone display weak single-stranded DNA binding affinity, whereas the affinity of the trimeric complex to single-stranded oligoes comprising three telomeric repeats is in the range of 1-20 nM.^{245,279} The increase of CST binding affinity with the length of its target DNA is suggestive of an RPA like binding mechanism, in which the accessibility of additional binding sites allow Stn1-Ten1 and possibly the C-terminal OB-fold of Ctc1 to engage in further interactions. However, this speculation remains to be tested, since there are currently no studies that have investigated this possibility, as was done for RPA.

Given the structural similarity between Stn1-Ten1 and RPA32-RPA14, including the extensive interactions between α -helices located on the C-terminus of their respective OB-folds, it is very likely that the Ctc1-Stn1-Ten1 interface resembles the RPA trimerization core. This assumption is further substantiated by a detailed analysis of disease associated mutations in Ctc1, which identified residues L1142H and 1196- Δ 7 in disrupting interactions with Stn1 and consequently the formation of the CST complex.³²⁵ These residues are located on the very C-terminus of the putative OB-fold of Ctc1. In a manner approximating the RPA tri-helical arrangement, these extreme residues may constitute a C-terminal α -helix which engages in interactions with those of Stn1 and Ten1 (Figure 1-15). The aforementioned study also characterized several mutations that fall within the uncharacterized region that stretches between the putative N- and C-terminal OB-folds of Ctc1 (V665G, R975G, C985 Δ , R987W).³²⁵ Interestingly, these mutations have been shown to negatively affect single-stranded DNA binding without preventing CST complex formation. However, in the absence of structural information, it remains unclear just how these mutations disrupt DNA binding.

Figure 1-15: Proposed model of Ctc1 C-terminal OB-fold interactions with Stn1-Ten1. **a)** Crystal structure of the RPA trimerization core. The interface composed of three α -helices is annotated. **b)** A proposed cartoon representation of Ctc1 C-terminus is superimposed over the crystal structures of Stn1-Ten1 OB folds in an approximation of the RPA trimerization core. Residues 1010-1100 constitute the putative C-terminal OB-fold of Ctc1 and are depicted as a globular domain. Deletion of residues 1196-1217 (highlighted in light pink) have been shown to disrupt Ctc1 interactions with Stn1-Ten1.³²⁵ It is possible that Ctc1 C-terminal residues form an α -helix that engages in interactions with Stn1-Ten1 similar to those observed in the RPA tri-helical interface.

The aim of the work presented in this thesis is to express and purify the CST complex and determine its structure, in order to gain insights into the molecular mechanism by which CST binds to single-stranded DNA and interacts with POT1-TPP1. Future research that builds on this work may elucidate how CST recruits Pol α /primase to telomeres. The elucidation of the structure of CST in isolation, and in complex with various lengths of single-stranded DNA would allow drawing direct comparisons between CST and RPA, and highlight similarities as well as functional differentiations. Finally, a high-resolution structure of CST in complex with a single-stranded DNA fragment should give a conclusive answer to whether or not CST possesses any sequence specificity towards telomeric repeats or G-rich sequences.

1.9) TCAB1

TCAB1, also called WRAP53 (WD40 repeat-containing protein antisense to TP53 gene), is a multifunctional adaptor protein that is involved in a wide range of cellular processes, including RNP biogenesis, DNA double strand break repair and telomere maintenance.^{201,202,206} Its main function is to guide subcellular localization of various factors by mediating their interactions at designated sites. TCAB1 itself is detectable in the cytoplasm, but appears highly enriched in CBs.²⁰²

CBs are evolutionarily conserved spherical objects with a diameter of 0.5-1 μm that are most prominent in cells with high transcriptional activity or that are proliferative.³⁶²⁻³⁶⁴ These subnuclear organelles contain a high concentration of various RNPs, including snRNPs, snoRNPs, scaRNPs and telomerase, and are implicated in their maturation and biogenesis.^{203,365-368} Importantly, CBs lack a phosphor-lipid membrane and instead adopt a hydrogel-like state that forms after polymerization of RNA and RNA-binding components into amyloid like fibers.^{369,370} As a result, CBs exist in a dynamic equilibrium in which their composition is subject to constant diffusional exchange with its environment.³⁷¹⁻³⁷³ Similar to coilin, TCAB1 plays an essential role in the formation and maintenance of CBs by providing a platform for the assembly of other CB components. In addition, TCAB1 also participates in the recruitment of the SMN complex, scaRNPs and telomerase.^{202,206}

The SMN complex is involved in the cytoplasmic assembly of spliceosomal snRNPs. After binding to the SMN complex in the cytoplasm, TCAB1 guides its nuclear import by mediating interactions between the SMN complex and nuclear pore receptor importin β .²⁰² Subsequently, interactions with coilin direct the localization of the SMN complex to CBs, where snRNPs resume their biogenesis.²⁰² Mutations in the *SMN1* gene results in spinal muscular atrophy (SMA) that is caused by impaired snRNP formation and splicing defects.³⁷⁴⁻³⁷⁶ Similarly, since depletion of TCAB1 disrupts nuclear localization of SMN and snRNP maturation, mutations in TCAB1 also give rise to SMA. It is important to

note, however, that SMA induced by TCAB1 dysfunction manifests in its most severe form, also referred to as type I SMA or Werdnig-Hoffmann disease.^{202,377-379} The increase in disease severity is likely to originate from the impairment of other TCAB1 regulated processes and their compounding impact on cellular homeostasis.

A probable candidate for a TCAB1 function that when impaired leads to an increase in disease severity is the involvement of TCAB1 in DNA double-strand break repair.³⁸⁰ After phosphorylation of histone variant H2AX and conversion into γ H2AX, direct interactions between γ H2AX and MDC1 BRCT domain recruit MDC1 to DNA double strand breakage sites.^{381,382} TCAB1 has been found to rapidly accumulate at these sites and facilitate interactions between MDC1 and ubiquitin ligase RNF8 by separately binding and tethering together both proteins.³⁸⁰ The RING domain of RNF8 then mediates the next step of the DDR (DNA damage response) recruitment cascade, which is the ubiquitylation of histones H2 and H2AX and recruitment of downstream factors.³⁸³⁻³⁸⁵ Depletion of TCAB1 has been reported to cause accumulation of spontaneous DNA double strand breaks and induce cell cycle arrest in G2/M phase.³⁸⁰ Consequently, loss of TCAB1 function can give rise to genomic instability.

Another function of TCAB1 is to target the two families of scaRNAs, C/D box scaRNAs and H/ACA box scaRNAs, to CBs, where they guide post-transcriptional modifications of snRNAs.²⁰¹ These modifications comprise C/D box scaRNA mediated methylation and H/ACA box scaRNA mediated pseudouridylation and are essential for proper spliceosomal assembly.^{386,387} scaRNA localization to CBs is driven by interactions between the CAB-box signal element and TCAB1.

Since telomerase RNA also contains an H/ACA domain, including a CAB-box motif, hTER is considered to be a scaRNA. However, hTER is exceptional in that its association with Dyskerin is not known to direct pseudouridylation of any target substrate. Instead, the main purpose of hTER H/ACA domain association with scaRNP proteins appears to be

hTER maturation, stability and intracellular accumulation.¹⁸⁸⁻¹⁹⁰ Like other scaRNAs, CB localization is driven by interactions between the CAB-box sequence and TCAB1. There, hTER undergoes further maturation, including modification of its m⁷G cap into a TMG cap and replacement of H/ACA RNP component NAF1 with GAR1.^{187,203,204} Even though TCAB1 is not required for telomerase assembly, hTER stability and telomerase activity *in vitro*, it plays a crucial role in telomerase recruitment to telomeres and activity *in vivo*.²⁰⁶⁻²⁰⁸ Consequently, mutations in TCAB1 result in defective telomere maintenance and give rise to dyskeratosis congenita (DC).²⁰⁷ Symptoms that are characteristic to DC include premature aging, bone marrow failure and predisposition for cancer.³⁸⁸⁻³⁹¹ Interestingly, rather than affecting interactions with the CAB-box or other factors, TCAB1 mutations that give rise to DC were found to impede proper folding of the protein, which was recently found to be guided by the chaperonin TCP-1 ring complex (TRiC).³⁹² In turn, improper folding impacts TCAB1 stability, prevents nuclear accumulation and disrupts all nuclear activity. Comparable to SMA, patients carrying dysfunctional TCAB1 exhibit more severe forms of DC, presumably due to the same compounding effect that is a combination between impaired DNA damage repair and genome instability.

The TCAB1 protein consists of three distinguishable domains: an N-terminal domain, a WD40 domain and a C-terminal extension. The N-terminal domain of TCAB1 contains a proline-rich stretch and is predicted to be highly disordered. Several putative phosphorylation sites are suggestive of cell cycle-regulated functions, such as telomerase recruitment.^{393,394} However, no disease related mutations have been reported in this portion of TCAB1. Instead, multiple mutations causing severe forms of DC have been identified in the adjacent WD40 domain.²⁰⁷ As mentioned earlier, these mutations disrupt TCAB1 activity by impeding TRiC aided protein folding, resulting in diminished TCAB1 levels and loss of nuclear activity.³⁹² The WD40 domain puts TCAB1 in a family of proteins that are very abundant in eukaryotes and serve a wide range of cellular functions by providing a rigid binding platform for various interaction partners.³⁹⁵ TCAB1 WD40 domain is most certainly involved in similar roles and critical for mediating protein interactions. Finally,

the WD40 domain is followed by a short, glycine-rich C-terminal extension of unknown function.

Currently, there is no structural information on TCAB1. Hence, available information on how TCAB1 participates in various cellular processes is exclusively derived from biochemical experiments, mutational analysis and depletion studies. In order to fully understand how TCAB1 realizes its various functions, it is important to obtain structural insights into the architecture of TCAB1 as well as its interactions with various factors. Thus, it would be possible to formulate the molecular basis by which TCAB1 promotes interactions between different factors, drives cellular localization and how these functions are influenced by cell cycle-dependent modifications.

1.10) TCAB1 is a WD40 Proteins

TCAB1 has not been shown to contain any catalytic activity. Instead, its function appears to be limited to mediating and facilitating protein-protein interactions as well as binding and trafficking specific RNA motifs (CAB-box). These activities are most likely attributable to the central WD40 domain, that in other proteins has been shown to fulfill similar functions. WD40 domain containing proteins are highly abundant in eukaryotes and participate in a wide range of cellular functions, such as signal transduction, cell division, vesicular trafficking, cytoskeletal changes, apoptosis, transcription regulation and RNA processing.³⁹⁶ In fact, yeast interactome datasets, which have been compiled using yeast two-hybrid systems and mass spectroscopy/tandem-affinity purification (MS/TAP), revealed WD40 domains to engage in more interactions than any other protein domain.^{397,398} Importantly, the number of protein components in WD40 domain protein complexes determined by MS/TAP significantly exceeds the number of binary interactions found in yeast two-hybrid experiments, suggesting that WD40 domains scaffold the assembly of larger complexes.^{397,398} WD40 domains have also been classified as one of the

most interacting domains in human interactome datasets.^{399,400} It is noteworthy, however, that both datasets are far from complete (20% for yeast interactome dataset and 1% for human interactome dataset), requiring any inference to be considered with caution.

WD40 domains were first identified in the β -subunit of the trimeric transducin complex G-protein.⁴⁰¹ The crystal structure of β -transducin revealed a seven bladed propeller shaped structure (β -propeller) with each of its blades forming a characteristic fold composed of four-stranded anti-parallel β -sheets (Figure 1-16 b).^{402,403} WD40 domains consist of repetitive sequence elements called WD40 repeats, the name of which was initially derived from its average length of about 40 amino acid residues that culminate in a conserved WD dipeptide. More recently, however, the definition of WD40 repeats has been revised as spanning 44 to 60 amino acid residues and comprising a conserved GH dipeptide 11 to 24 residues downstream from its N-terminus in addition to the C-terminal WD dipeptide (Figure 1-16 a).^{404,405} It is important to note that in most WD40 domain containing proteins, one WD40 does not constitute one β -sheet. Instead, each β -sheet is composed of three β -strands from one WD40 repeat (β -strands A, B and C) and one β -strand from the next (β -strand D) (Figure 1-16 c). Thus, each WD40 repeat forms a so-called “velcro”-closure which, together with extensive hydrophobic interactions between each pair of blades, contributes to the stabilization of the overall β -propeller structure.

Figure 1-16: WD40 repeats and domain structure. **a)** Hidden Markov model (HMM) of WD40 repeats from the protein families (Pfam) database (<http://pfam.xfam.org/family/WD40#tabview=tab4>).⁴⁰⁶ Stack height correlates with relative entropy of a certain sequence position whereas letter size correlates with relative frequency of the corresponding amino acid occupying that position. Underneath the chart, the first row gives values for occupancy, which is the likelihood of observing a gap character (blue color denotes low occupancy whereas white color denotes high occupancy). The second row signifies insert probability, whereas the third row specifies the length of the expected insert. Secondary structure elements are illustrated underneath the HMM chart. **b)** WD40 domain of β -transducin folds into a 7-bladed β -propeller structure. **c)** One WD40 repeat constitutes the outmost β -strand of one blade and the inner three β -strands of the next. Thus, it forms a “velcro”-closure that stabilizes the overall structure. WD40 repeat coloration is in accordance with that of secondary structure elements depicted in a).

In theory, WD40 domains can contain between 4 and 8 WD40 repeats.⁴⁰⁷ According to geometry modeling, however, 7-fold β -propellers are predicted to be energetically most favorable.⁴⁰⁸ This is in accordance with structural studies, which so far have only confirmed WD40 domain proteins with 7- and 8-bladed β -propellers. WD40 domains have three distinct surfaces, namely the top, the bottom and the circumference (Figure 1-17). The top surface of WD40 domains is defined as the site which harbors connecting loops between the first and last two β -strands of each WD40 repeat (β -strands D and A and β -strands B and C). Importantly, each surface can engage in separate interactions, allowing WD40 domains to act as protein scaffolds that can assemble large complexes and making them ideal hubs in cellular interaction networks.

Figure 1-17: Side view of β -transducin WD40 domain. Connecting loops between β -strand D and A, as well as β -strand B and C face the top surface.

Given these properties, it is conceivable that TCAB1 WD40 domain plays a central role in most of TCAB1 functions. The presence of individual binding sites is perfectly suitable for subcellular trafficking of associated factors and mediating interactions between various binding partners by bringing them in close proximity and conferring positive cooperativity. The functions of N- and C-terminal extensions, however, are less

evident. There are multiple precedents of WD40 domain proteins, in which N- and/or C-terminal extensions engage in stabilizing interactions with the WD40 domain. Furthermore, since the N-terminal region comprises several putative phosphorylation sites,^{393,394} intramolecular interactions between the N-terminal tail and the WD40 domain might be dependent on its phosphorylation state. Thus, there is an interesting possibility of cell cycle-regulated access to distinct binding sites, which could be of high relevance in telomerase trafficking.

The aim of the work described in this thesis is to express and purify TCAB1 and elucidate its structure, in order to reveal molecular mechanisms of its cellular functions. TCAB1 structure in isolation can convey important information regarding domain architecture, function of N- and C-terminal tails and molecular environment and effects of disease causing mutations. In complex with various binding partners, structural investigations can help elucidate the mechanism by which TCAB1 mediates binding between MDC1 and RNF8, interacts with H/ACA RNPs or tethers the SMN complex with importin β . In the context of the focus of our group, however, we are more interested in the involvement of TCAB1 in the recruitment of telomerase to telomeres, and potential cell cycle-dependent regulations. Also, it would be interesting to see the nature (e.g. direct or indirect) and architecture of interactions between TCAB1, TPP1 and telomerase.

Chapter 2: Materials and Methods

2.1) General Materials and Methods

2.1.1) *E.coli* strains

Table 2-1: Genotypes of used bacterial strains.

Strain	Genotype
BL21(DE3)pLysS	$F^- ompT gal dcm lon hsdS_B(r_B^- m_B^-) \lambda(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB^+]_{K-12}(\lambda^S) pLysS[T7p20 ori_{p15A}](CmR)$
C43(DE3)pLysS	$F^- ompT gal dcm lon hsdS_B(r_B^- m_B^-) \lambda(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB^+]_{K-12}(\lambda^S) pLysS[T7p20 ori_{p15A}](CmR)$
DH10 α pir	$F^- endA1 hsdR17 glnV44 (=supE44) thi-1 recA1 gyrA96 relA1 \Phi80lacZ\Delta M15 \Delta(lacZYA-argF) U169 zdg-232::Tn10 uidA::pir^+$
DH10Bac	$F^- mcrA \Delta(mrr-hsdRMS-mcrBC) \Phi80lacZ\Delta M15 \Delta lacX74 recA1 endA1 araD139 \Delta(ara leu) 7697 galU galk rpsL nupG \lambda-$
OmniMAX2	$F' (proAB+ lacIq lacZ\Delta M15 Tn10(TetR) \Delta(ccdAB)) mcrA \Delta(mrr-hsdRMS-mcrBC) \Phi80lacZ\Delta M15 \Delta(lacZYA-argF) U169 endA1 recA1 supE44 thi-1 gyrA96 relA1 tonA pand$
Rosetta(DE3)pLysS	$F^- ompT gal dcm lon hsdS_B(r_B^- m_B^-) \lambda(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB^+]_{K-12}(\lambda^S) pLysSRARE[T7p20 ileX argU thrU tyrU glyT thrT argW metT leuW proL ori_{p15A}](CmR)$

2.1.2) Plasmids

All plasmids listed below were prepared following the Ligation Independent Cloning protocol.

E.coli plasmids

Table 2-2: Bacterial plasmid names with respective inserts, tags of inserts and resistance genes.

Name	Insert	Tag	Resistance
pOPTG(Ctc1A1)	Ctc1	N' GST	Amp
pOPTG(Ctc1A2)	Ctc1 1-700	N' GST	Amp
pOPTG(Ctc1A3)	Ctc1 1-496	N' GST	Amp
pOPTG(Ctc1A4)	Ctc1 1-464	N' GST	Amp
pOPTG(Ctc1A5)	Ctc1 203-1217	N' GST	Amp
pOPTG(Ctc1A6)	Ctc1 203-700	N' GST	Amp
pOPTG(Ctc1A7)	Ctc1 203-496	N' GST	Amp
pOPTG(Ctc1A8)	Ctc1 203-464	N' GST	Amp
pOPTG(Ctc1A9)	Ctc1 521-1217	N' GST	Amp
pOPTG(Ctc1A10)	Ctc1 521-700	N' GST	Amp
pOPTG(Ctc1A11)	Ctc1 828-1217	N' GST	Amp
pOPTG(Ctc1A12)	Ctc1 958-1217	N' GST	Amp
pOPTG(TCAB1A1)	TCAB1	N' GST	Amp
pOPTG(TCAB1A2)	TCAB1 87-548	N' GST	Amp
pOPTG(TCAB1A3)	TCAB1 118-548	N' GST	Amp
pOPTG(TCAB1A4)	TCAB1 158-548	N' GST	Amp
pOPTG(TCAB1A5)	TCAB1 1-526	N' GST	Amp
pOPTG(TCAB1A6)	TCAB1 87-526	N' GST	Amp
pOPTG(TCAB1A7)	TCAB1 118-526	N' GST	Amp
pOPTG(TCAB1A8)	TCAB1 158-526	N' GST	Amp
pOPTG(TCAB1A9)	TCAB1 1-495	N' GST	Amp
pOPTG(TCAB1A10)	TCAB1 87-495	N' GST	Amp
pOPTG(TCAB1A11)	TCAB1 118-495	N' GST	Amp
pOPTG(TCAB1A12)	TCAB1 158-495	N' GST	Amp
pOPTG(TCAB1N1)	TCAB1 76-548	N' GST	Amp

pOPTG(TCAB1N2)	TCAB1 79-548	N' GST	Amp
pOPTG(TCAB1N3)	TCAB1 84-548	N' GST	Amp
pOPTG(TCAB1N4)	TCAB1 92-548	N' GST	Amp
pOPTG(TCAB1N5)	TCAB1 76-526	N' GST	Amp
pOPTG(TCAB1N6)	TCAB1 79-526	N' GST	Amp
pOPTG(TCAB1N7)	TCAB1 84-526	N' GST	Amp
pOPTG(TCAB1N8)	TCAB1 92-526	N' GST	Amp
pOPTG(mTCAB1)	mouse TCAB1	N' GST	Amp
pOPTG(zfTCAB1)	zebrafish TCAB1	N' GST	Amp
pNIC(ST)	Stn1 and Ten1	N' His ₆ (Stn1)	Kan
pNIC(Δ ST)	Stn1 19-368 and Ten1	N' His ₆ (Stn1)	Kan
pNIC(CSTA1)	Ctc1, Stn1 and Ten1	N' GST (Ctc1), N' His ₆ (Stn1)	Kan
pNIC(CSTA5)	Ctc1 203-1217, Stn1 and Ten1	N' GST (Ctc1), N' His ₆ (Stn1)	Kan
pNIC(CSTA9)	Ctc1 521-1217, Stn1 and Ten1	N' GST (Ctc1), N' His ₆ (Stn1)	Kan
pNIC(CSTA11)	Ctc1 828-1217, Stn1 and Ten1	N' GST (Ctc1), N' His ₆ (Stn1)	Kan
pNIC(CSTA1/ Δ)	Ctc1, Stn1 19-368 and Ten1	N' GST (Ctc1), N' His ₆ (Stn1)	Kan
pNIC(CSTA5/ Δ)	Ctc1 203-1217, Stn1 19-368 and Ten1	N' GST (Ctc1), N' His ₆ (Stn1)	Kan
pNIC(CSTA9/ Δ)	Ctc1 521-1217, Stn1 19-368 and Ten1	N' GST (Ctc1), N' His ₆ (Stn1)	Kan
pNIC(CSTA11/ Δ)	Ctc1 828-1217, Stn1 19-368 and Ten1	N' GST (Ctc1), N' His ₆ (Stn1)	Kan
pNIC(CSTA1')	Ctc1 1-1204, Stn1 and Ten1	N' GST (Ctc1), N' His ₆ (Stn1)	Kan
pNIC(CSTA5')	Ctc1 203-1204, Stn1 and Ten1	N' GST (Ctc1), N' His ₆ (Stn1)	Kan
pNIC(CSTA9')	Ctc1 521-1204, Stn1 and Ten1	N' GST (Ctc1), N' His ₆ (Stn1)	Kan
pNIC(CSTA11')	Ctc1 828-1204, Stn1 and Ten1	N' GST (Ctc1), N' His ₆ (Stn1)	Kan
pNIC(CSTA1'/ Δ)	Ctc1 1-1204, Stn1 19-368 and Ten1	N' GST (Ctc1), N' His ₆ (Stn1)	Kan

pNIC(CSTA5'/Δ)	Ctc1 203-1204, Stn1 19-368 and Ten1	N' GST (Ctc1), N' His ₆ (Stn1)	Kan
pNIC(CSTA9'/Δ)	Ctc1 521-1204, Stn1 19-368 and Ten1	N' GST (Ctc1), N' His ₆ (Stn1)	Kan
pNIC(CSTA11'/Δ)	Ctc1 828-1204, Stn1 19-368 and Ten1	N' GST (Ctc1), N' His ₆ (Stn1)	Kan
pFB(GFP-Ctc1A1)	GFP, Ctc1	N' GST (Ctc1)	Amp
pFB(GFP-Ctc1A11)	GFP, Ctc1 828-1204	N' GST (Ctc1)	Amp
pFL(Ctc1A1)	Ctc1	N' GST	Amp
pFL(Ctc1A5)	Ctc1 203-1204	N' GST	Amp
pFL(Ctc1A9)	Ctc1 521-1204	N' GST	Amp
pFL(Ctc1A11)	Ctc1 828-1204	N' GST	Amp
pFL(TCAB1B1)	TCAB1	N' GST, C' His ₆	Amp
pFL(TCAB1B2)	TCAB1 79-548	N' GST, C' His ₆	Amp
pFL(TCAB1B3)	TCAB1 84-548	N' GST, C' His ₆	Amp
pFL(TCAB1B1Δ)	TCAB1 1-526	N' GST, C' His ₆	Amp
pFL(TCAB1B2Δ)	TCAB1 79-526	N' GST, C' His ₆	Amp
pFL(TCAB1B3Δ)	TCAB1 84-526	N' GST, C' His ₆	Amp
pUCDM(ΔST)	Stn1 19-368 and Ten1	N' His ₆ (Stn1)	Cam

Tissue culture plasmids

Table 2-3: Tissue culture plasmid names with respective inserts, tags of inserts and resistance genes.

Name	Insert	Tag	Resistance
pUB(Ctc1A1)	Ctc1	N' 3xFLAG, C' GFP	Kan
pUB(Ctc1A5)	Ctc1 203-1204	N' 3xFLAG, C' GFP	Kan
pUB(Ctc1A9)	Ctc1 521-1204	N' 3xFLAG, C' GFP	Kan
pUB (Ctc1A11)	Ctc1 828-1204	N' 3xFLAG, C' GFP	Kan
pUB(TCAB1B1)	TCAB1	N' 3xFLAG, C' GFP	Kan
pUB(TCAB1B2)	TCAB1 79-548	N' 3xFLAG, C' GFP	Kan
pUB(TCAB1B3)	TCAB1 84-548	N' 3xFLAG, C' GFP	Kan
pUB(TCAB1B1Δ)	TCAB1 1-526	N' 3xFLAG, C' GFP	Kan
pUB(TCAB1B2Δ)	TCAB1 79-526	N' 3xFLAG, C' GFP	Kan
pUB(TCAB1B3Δ)	TCAB1 84-526	N' 3xFLAG, C' GFP	Kan

2.1.3) Growth media

Table 2-4: Composition of various growth media used.

Medium	Composition
LB	1% tryptone, 0.5% yeast extract, 10% NaCl, pH 7.5
2xTY	1.6% tryptone, 1% yeast extract, 0.5% NaCl, pH 7.4
SOC	2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl ₂ , 20 mM glucose
DMEM	see https://www.thermofisher.com
Sf900-III	see https://www.thermofisher.com
Grace's medium	see https://www.thermofisher.com

Table 2-5: Supplements to growth media.

Supplements	Stock solution	Medium concentration
Ampicillin	100 mg/mL	100 µg/mL
Kanamycin	50 mg/mL	50 µg/mL
Chloramphenicol	25 mg/mL	25 µg/mL
Gentamycin	7 mg/mL	7 µg/mL
Tetracycline	10 mg/mL	10 µg/mL
X-Gal	40 mg/mL	40 µg/mL
IPTG	1 M	0.1 – 0.5 mM
FBS	100%	2%
Anti-Anti	100x	1x

2.2) DNA Methods

2.2.1) Poly-chain reaction (PCR)

For PCR reactions, KOD and KOD Xtreme DNA Polymerase (*Millipore*) were used. Manufacturer's protocols were followed.

2.2.2) DNA detection

DNA probes were admixed with 1/5 Vol. of 6x sample buffer and loaded on an 1.5% agarose gel. Gel electrophoresis was performed with 1x TBE buffer and at 100 mV.

Table 2-6: Composition of buffers and components used for agarose gel electrophoresis.

Component	Composition
agarose (x%)	x% agarose in 1x TBE
1x TBE	89 mM tris-borate pH 8.3, 2 mM EDTA
6x sample buffer	0.25% bromophenol blue, 0.25% xylene cyanol FF, 30% glycerol

2.2.3) Ligation independent cloning (LIC)

Primers were designed to comprise a 20 nt complementary overlap with the extreme ends of the opened vector followed by a DNA sequence that is complementary to the extreme ends of the insert and has a melting temperature of 65°C. Inserts were generated by PCR with a pair of Forward and Reverse primers and purified by gel-extraction. 250 ng of open vector and insert were readied in a 1xNEB1 (*NEB*) buffered 13.5 µL volume solution and kept at 4°C. After the addition of 10 u/µL Exonuclease III (*NEB*), the reaction was incubated at 14°C for 1 min, quenched by the addition of 1.5 µL 500 mM EDTA, and plunged into liquid nitrogen. Subsequently, the reactions were subjected to a PCR annealing program.

Table 2-7: LIC annealing program.

Temperature	Time
75°C	20 min
75°C	6 sec
75 → 25°C	6 sec per 0.4°C step (750 s)
25°C	hold

2.2.4) DNA isolation

For **DNA isolation from gel slices** *GeneJET Gel Extraction Kit (Thermo Fischer Scientific)* was used and manufacturer's protocol was followed.

For **plasmid DNA isolation from *E.coli*** *GeneJET Plasmid Miniprep Kit (Thermo Fischer Scientific)* was used and manufacturer's protocol was followed.

2.2.5) MultiBac cloning

For the generation of bacmids encoding for more than two proteins, the MultiBac system, developed by Imre Berger and his group in EMBL Grenoble, was used and accompanying protocols were followed.⁴⁰⁹

In general, the Multibac system utilizes transfer vectors which are called acceptor and donor plasmids. Both plasmids contain identical p10 and polH cassettes, in which multiple genes of interest are inserted. Importantly, only acceptor plasmids contain Tn7 transposition elements, which are used to integrate gene cassettes into the MultiBac baculoviral genome. In order to generate larger gene assemblies, acceptor and donor plasmids are fused together by *in vitro* Cre-Lox recombination (Figure 2-1 a). The loxP site

on the acceptor plasmid is located in such a way that following recombination all recombinant DNA cargo is flanked by Tn7 transposition elements. Since acceptor plasmids contain pBR322-derived origins and donor plasmids contain conditional R6K γ replicons, successfully recombined transfer vectors can be selected by amplification in hosts that can recognize either one origin of replication, while at the same time utilizing antibiotic resistance markers on both plasmids. Finally, acceptor-donor fusions are integrated into the Multibac bacmid by Tn7 transposition in DH10MultiBac^{Cre} cells, which carry the baculoviral genome as an artificial bacterial chromosome and contain a helper plasmid that expresses the Tn7 transposon complex (Figure 2-1 b). Alternatively, one can use DH10EmBacY cells, which are the product of *in vivo* Cre-Lox recombination between the baculoviral genome contained in DH10MultiBac^{Cre} cells and a pUCDM vector encoding for YFP (yellow fluorescent protein). The fluorescent signal of YFP can then be used to evaluate infection efficiency and protein expression, as well as to determine the titer of viral stocks by means of end point dilution experiment (EPDE).

The transfer vectors used for the work presented in this thesis are pFL (donor) and pUCDM (acceptor). Ctc1 constructs were cloned into the polH cassette of pFL, whereas Stn1 and Ten1 constructs were cloned into the p10 and polH cassettes of pUCDM, respectively. pFL plasmids were amplified in OmniMax2 cells, whereas pUCDM plasmids were amplified in DH10 α pir cells. Plasmids were opened by enzymatic restriction (NcoI for pFL polH cassette, XhoI for pUCDM polH cassette and BamHI for pUCDM p10 cassette) and genes encoding for various protein constructs were sequentially inserted by ligation independent cloning. In order to assemble all three genes, various combinations were generated by *in vitro* Cre-Lox recombination of acceptor and donor plasmids (see Table 2-2). Successfully recombined transfer vectors were then selected and amplified in DH10 α pir cells in the presence of ampicillin (pFL) and chloramphenicol (pUCDM). Subsequently, transfer vectors containing all three genes were transformed into DH10EmBacY cells, where all recombinant DNA cargo was inserted into the LacZ α gene mini-attenuator sequence through *in vivo* Tn7 transposition. Productive recombination

events were then selected by blue-white screening. Finally, positive colonies were cultured and their bacmid DNA isolated.

Figure 2-1: General MultiBac cloning strategy for three- to four-gene assemblies. **a)** The figure illustrates the Cre-lox recombination of acceptor and donor plasmids. In the resulting transfer vector, Tn7 elements flank all four expression cassettes. **b)** The genetic cargo between Tn7 elements is integrated into the MultiBac baculoviral genome by Tn7 transposition. Integration into the Tn7 mini-attenuator sequence disrupts the LacZα gene, enabling blue-white screening for productive transposition events.

Cre-Lox recombination

For Cre-Lox recombination of acceptor and donor plasmid, Cre-recombinase (*NEB*) was used and manufacturer's protocol was followed.

Transformation and blue-white screening

50 μ L ultra-competent DH10EmBacY cells were admixed with 2 μ L fused transfer vectors and incubated at 4°C for 30 min. After a heat shock at 42°C for 1 min, 1 mL pre-warmed (37°C) SOC medium was added and the suspension was incubated at 37°C for 1 h. Subsequently, 200 μ L of 1:10, 1:100 and 1:1000 dilutions were streaked on LB agar plates containing 37 μ g/mL Cam, 100 μ g/mL Amp, 100 μ g/mL X-gal and 40 μ g/mL IPTG. After 3 days of incubation at 37°C, single white colonies were picked and restreaked for verification and incubated for another day at 37°C.

Bacmid isolation

Verified colonies were resuspended in 1 mL of 2xTY containing 7 μ g/mL gentamycin and 10 μ g/mL tetracycline and incubated overnight at 37°C and 700 rpm. The culture was then spun at 2,600x *g* for 20 min and the supernatant decanted. Residual supernatant was removed by tapping on filter paper. The cell pellet was frozen at -20°C for at least 30 min, thawed at room temperature for 15 min and resuspended in 250 μ L B1. In order to lyse the cells, 250 μ L B2 was added and the reaction was mixed by inverting, followed by a 10 min incubation at room temperature. The reaction was quenched by the addition of 300 μ L B3, mixed by inverting and incubated at 4°C for 20 min. Afterwards, it was spun at 2,600x *g*, 4°C for 30 min. The supernatant containing the bacmid was transferred into a fresh tube and spun again at 10,000x *g* and 4°C for 30 min. In order to precipitate the bacmid, 0.5 mL of the supernatant were transferred into a new tube with 0.5 mL isopropanol, mixed by inverting and incubated at 4°C for at

least 30 min. Then, it was spun at 10,000x g, 4°C for 30 min and the supernatant was discarded. The tube was sterilized with 70% EtOH and air dried under the laminar flow cabinet for 15-20 min. Finally, the bacmid DNA was resuspended in 50 µL sterile TE buffer.

Table 2-8: Composition of buffers used for bacmid isolation.

Component	Composition
B1	50 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0, 100 µg/ml RNaseA
B2	200 mM NaOH, 1% SDS
B3	3 M KOAc pH 5.5
TE buffer	10 mM Tris-HCl pH 7.5, 1 mM EDTA pH 8.0

2.2.6) DNA annealing

Complementary oligonucleotides were mixed in a stoichiometric ratio in a 50 µL volume to a final concentration of 20 µM each, and subjected to a PCR annealing program.

Table 2-9: Oligonucleotide annealing program.

Temperature	Time
99°C	5 min
99 → 25°C	1 min per 0.5°C step (140 min)
4°C	hold

2.3) Protein Methods

2.3.1) *E.coli* culture

Preparation of competent cells

For the preparation of chemo-competent *E.coli* cells, the Inoue method was followed.⁴¹⁰

Transformation

50 μ L of chemo-competent *E.coli* cells were added to 1 μ L of plasmid or 5 μ L LIC reaction, incubated at 4°C for 30 min, heat shocked at 42°C for 1 min and streaked on LB plates containing appropriate antibiotic supplements. Plates were then incubated at 37°C overnight until colonies of appropriate size were obtained.

IPTG induction

10 mL 2xTY was inoculated with a single colony of the desired bacterial *E.coli* strain and incubated overnight at 37°C. The pre-culture was then admixed with 1 L 2xTY and incubated at 37°C. For expression at 37°C, cells were grown to an OD₆₀₀ of 0.7-1.0, before IPTG was added (usually to a final concentration of 0.1 mM), followed by a 3 h incubation at 37°C. For expression at 18°C, cells were grown to an OD₆₀₀ of 0.5, before they were transferred to an 18°C incubator. By the time cells reached an OD₆₀₀ of 0.7-1.0, the temperature of the culture had dropped to 18°C, upon which expression was induced by the addition of IPTG to a final concentration of 0.1 mM.

Auto-induction

10 mL 2xTY was inoculated with a single colony of the desired bacterial *E.coli* strain and incubated overnight at 37°C. The pre-culture was then admixed with 1 L LB-medium, 50 mL NPS, 20 mL 5052, 1 mL Met-Mix and 1 mL 1M MgSO₄. After incubation for 2-3 days at 37°C, cells were harvested.

Table 2-10: Composition of components used for auto-induction.

Component	Composition
NPS	0.5 M (NH ₄) ₂ SO ₄ , 1 M KH ₂ PO ₄ , 1 M Na ₂ HPO ₄
5052	25% glycerol, 2.5% glucose, 10% α-lactose 50 mM FeCl ₃ , 20 mM CaCl ₂ , 2 mM Na ₂ SeO ₃ , 2 mM H ₃ BO ₃ , 10 mM
Met-Mix	MnCl ₂ , 10 mM ZnSO ₄ , 2 mM CoCl ₂ , 2 mM CuCl ₂ , 2 mM NiCl ₂ , 2 mM NaMoO ₂

Harvest of E.coli cells

Cell cultures were harvested at 4°C and 4000 rpm (*Avanti JXN-26, JLA-8.1 rotor, Beckman Coulter*) for 10 min and resuspended in lysis buffer. If the cell pellet was not immediately required for protein preparation, cells were harvested in plastic bags that were placed inside centrifuge bottles. After decanting the supernatant, the cell pellet was distributed into a thin layer within the plastic bag, submerged in liquid nitrogen and smashed into small flakes. The flakes were then transferred into a 50 mL falcon tube and stored at -80°C.

Disruption of E.coli cells

In order to obtain a homogenous solution that would not clog the machine in the following step, the cell suspension was first sonicated under mild conditions (4°C, 5 sec on, 5 sec off, 30% amplitude) (*Fisherbrand 700w Sonicator, Fisher Scientific*). Afterwards, it was homogenized at 4°C and 900 bar for 10 min (*PANDA 2K, GEA*).

2.3.2) Insect cell culture

In the following section, whenever Sf900-III medium is mentioned, it refers to Sf900-III medium supplemented with 2% FBS and 1x Anti-Anti.

Transfection of bacmid DNA and harvest of first-generation virus

Sf9 cells were diluted to 4×10^5 cell/mL in Sf900-III medium and 2 mL were dispensed into each well of a 6-well plate. For insect cells to attach to the bottom surface, plates were incubated at 27°C for 1 h. The medium in each well was aspirated and replaced with 2.5 mL Grace's insect medium. For transfection, Cellfectin II reagent (*Thermo Fischer Scientific*) was used and manufacturer's protocol followed. Following transfection, the Grace's medium was removed and replaced with 2 mL of Sf900-III medium. The plates were incubated for 4 days at 27°C. The supernatant, which at this time point contained the first-generation virus, was transferred into a 15 mL falcon tube and centrifuged at 4°C, 1,500x *g* for 10 min. In order to remove residual cells, the supernatant was transferred into a fresh 15 mL falcon tube and centrifuged again. Finally, the solution containing the first-generation viral stock was stored in the dark at 4°C.

Viral amplification

1x10⁶ cells/mL of Sf9 cells in Sf900-III were infected at an MOI (multiplicity of infection) of 1. The cells were incubated at 27°C and 150 rpm for 3 days. The cell culture was transferred into 50 mL falcon tubes and centrifuged at 4°C, 1,500x *g* for 10 min. In order to remove residual cells, the supernatant was transferred into a fresh 50 mL falcon tube and centrifuged again. Finally, the solution containing virus was stored in the dark at 4°C.

End point dilution experiment (EPDE)

In order to determine the titer of a viral stock, EPDE was performed. Importantly, EPDE can only be performed, if the baculoviral genome encodes for a fluorescent protein, such as YFP or GFP, since positive counts are determined by fluorescent signal.

The viral stock was diluted in a ten-fold series, ranging from 10⁻³ to 10⁻¹⁰ (depending on the expected titer of the examined viral stock), in Sf900-III medium. 130 µL of each dilution was added to 1.3 mL of 1x10⁵ cells/mL of Sf9 cells in Sf900-III medium, respectively, and mixed gently. Afterwards, for each dilution, 110 µL of the cell-virus mix was aliquoted into each well of a 96-well plate row. The 96-well plate was then incubated at 27°C for 3 days. At the end of the incubation period, each well was checked for fluorescence. If a well contained at least one cell that expressed a fluorescent protein, the well was counted as positive. Finally, the titer was calculated using following equation:

$$\text{Viral Titre} = \frac{\ln 2}{V(\text{virus}) \times 10^{\log(\text{dil}50) - \left(\frac{ab50 - 50\%}{ab50 - be50}\right)}} \frac{\text{pfu}}{\text{mL}}$$

where $V(\text{virus})$ is the volume of virus added per well (10 µL), $\text{dil}50$ is the dilution factor at which more than 50% of the wells in the corresponding row were infected, $ab50$ is the

percentage of wells infected at *dil50*, and *be50* the percentage of wells infected at *dil50/10*.

Optimization of expression conditions

In order to optimize expression conditions, three parameters were tested: cell count, MOI and incubation period. Tests were performed in 6-well deep-well blocks, sealed with air-permeable sealing films and incubated at 27°C and 450 rpm (*Titramax100*, *Heidolph*).

For the optimization of cell count at the time of infection, 3 mL of various concentrations of Sf9 cells in Sf900-III medium were placed in 6-well deep-well blocks and infected with virus at an MOI of 1. After 3 days of incubation at 27°C and 450 rpm, 100 µL samples were taken and analyzed by western blot.

For the optimization of MOI, 3 mL of 1×10^6 cells/mL of Sf9 cells in Sf900-III medium were placed in 6-well deep-well blocks and infected with virus at various MOIs. After 3 days of incubation at 27°C and 450 rpm, 100 µL samples were taken and analyzed by western blot.

For the optimization of the incubation period, 3 mL of 1×10^6 cells/mL of Sf9 cells in Sf900-III medium were placed in 6-well deep-well blocks and infected with virus at an MOI of 1. During incubation at 27°C and 450 rpm, 100 µL samples were taken at various time points and analyzed by western blot.

Protein expression in insect Cells

Protein expression in insect cells was performed at optimized conditions. But in general, for first time expression tests, 1×10^6 cells/mL of Sf9 cells in Sf900-III were infected at an MOI of 1. The cells were incubated at 27°C and 150 rpm for 3 days.

Harvest of insect cell cultures

Cell cultures were harvested at 4°C and 2,000 rpm (*Avanti JXN-26, JLA-8.1 rotor, Beckman Coulter*) for 10 min and resuspended in lysis buffer. If the cell pellet was not immediately required for protein preparation, cells flakes were prepared and stored away as described in “Harvest of *E.coli* cells”.

2.3.3) Tissue culture

Expression screen in HEK293T cells

For transfection of HEK293T cells in 6-well plates, Xfect polymer (*Clontech*) was used and manufacturer’s protocol was followed. Afterwards, plates were incubated at 37°C, 5% CO₂ for 48 h. The medium was decanted and cells were resuspended in 1 mL PBS (*Gibco*) per well. The resuspension was centrifuged at 4°C and 800x *g* for 5 min, and the supernatant aspirated. The cell pellets were analyzed by western blot.

2.3.4) Protein purification

Purification tests of GST-Ctc1 constructs from E.coli

Purification tests of GST-TCAB1 constructs from *E.coli* was performed in as described in “Purification tests of GST-TCAB1 constructs from *E.coli*”.

Co-purification tests of GST-Ctc1, His₆-Stn1, Ten1 constructs from E.coli

Co-purification tests of GST-Ctc1, His₆-Stn1, Ten1 constructs from *E.coli* was performed as described in “Purification tests of His₆-TCAB1 constructs from *E.coli*”.

Co-purification tests of GST-Ctc1, His₆-Stn1, Ten1 constructs from insect cells

Co-purification tests of GST-Ctc1, His₆-Stn1, Ten1 constructs from insect cells was performed as described in “Purification tests of His₆-GST-TCAB1 constructs from insect cells”.

Protocol for co-purification of GST-Ctc1, His₆-Stn1 19-368 and Ten1 from Insect Cells

Sf9 cells expressing GST-Ctc1, His₆-Stn1 19-368 and Ten1 were lysed in lysis buffer L8/150i and centrifuged at 4°C and 40,000 rpm for 30 min (*Avanti JXN-26, JLA-25.50 rotor, Beckman Coulter*). The supernatant was passed through a 0.45 µm pore filter, before it was applied on a HisTrap HP column (*GE Healthcare*) using a peristaltic pump at 3 mL/min. The column was then attached to an ÄKTAprime plus (*GE Healthcare*), and washed at 3 mL/min with 20 column volumes of NiW8/500, 20 column volumes NiW8/500+T and 5 column volumes NiW8/150. The protein was eluted at 1 mL/min with NiE8/150. Elution fractions were pooled and applied over a HiTrap Heparin HP column (*GE Healthcare*) at 1 mL/min using a peristaltic pump. The column was once again attached to an ÄKTAprime

plus (*GE Healthcare*), and washed at 1 mL/min with 5 column volumes of B8/150, B8/350, B8/450 and B8/1000, respectively. Protein fractions eluting with B8/450 were pooled and subjected to overnight TEV-cleavage at 4°C using a TEV to protein ratio of 1:100. The next morning, the protein solution was concentrated over a 100 kDa MWCO ultra-centrifugal filter unit (*Merck Millipore*) before it was applied on a superdex 200 pg column (*GE Healthcare*) of appropriate size. Gel filtration was run at 1 mL/min with B8/150T. Finally, fractions containing the purified protein were pooled and concentrated.

Protocol for co-purification of GST-Ctc1 521-1217, His₆-Stn1 19-368 and Ten1 from insect cells

Co-purification of GST-Ctc1 521-1217, His₆-Stn1 19-368 and Ten1 from insect cells was performed in a very similar manner to that of GST-Ctc1, His₆-Stn1 19-368 and Ten1 with only a few deviations. After the HisTrap HP eluate was loaded onto the HiTrap HP column, it was washed with 5 column volumes of B8/150, B8/250, B8/350 and B8/1000, instead, and protein fractions eluting with B8/250 were subjected to TEV-cleavage and gel filtration.

Purification tests of His₆-TCAB1 constructs from E.coli

His₆-TCAB1 expressing *E.coli* cells were lysed in lysis buffer L7.5/300i and centrifuged at 4°C and 40,000 rpm for 30 min (*Avanti JXN-26, JLA-25.50 rotor, Beckman Coulter*). The supernatant was passed through a 0.45 µm pore filter, before it was applied on a HisTrap HP column (*GE Healthcare*) using a peristaltic pump at 3 mL/min. The column was then attached to an ÄKTAprime plus (*GE Healthcare*), and washed at 3 mL/min with 10 column volumes of NiW7.5/500 and 5 column volumes NiW7.5/150. The protein was eluted at 2 mL/min with NiE7.5/150. Elution fractions were pooled and concentrated over a 10 kDa MWCO ultra-centrifugal filter unit (*Merck Millipore*) before they were applied on a superdex 200 pg column (*GE Healthcare*) of appropriate size. Gel filtration was run at 1

mL/min with B7.5/150T. Finally, fractions containing the purified protein were pooled and concentrated.

Purification tests of GST-TCAB1 constructs from E.coli

GST-TCAB1 expressing *E.coli* cells were lysed in lysis buffer L7.5/300 and centrifuged at 4°C and 40,000 rpm for 30 min (*Avanti JXN-26, JLA-25.50 rotor, Beckman Coulter*). The supernatant was passed through a 0.45 µm pore filter, before it was applied on a GSTrap HP column (*GE Healthcare*) using a peristaltic pump at 2 mL/min. The column was then attached to an ÄKTAprime plus (*GE Healthcare*), and washed at 2 mL/min with 10 column volumes of B7.5/500 and 5 column volumes B7.5/150. The protein was eluted at 1 mL/min with GSTE7.5/150. Elution fractions were pooled and concentrated over a 10 kDa MWCO ultra-centrifugal filter unit (*Merck Millipore*) before they were applied on a superdex 200 pg column (*GE Healthcare*) of appropriate size. Gel filtration was run at 1 mL/min with B7.5/150T. Finally, fractions containing the purified protein were pooled and concentrated.

Purification tests of His₆-GST-TCAB1 constructs from Insect Cells

His₆-GST-TCAB1 expressing Sf9 cells were lysed in lysis buffer L7.5/300i and centrifuged at 4°C and 40,000 rpm for 30 min (*Avanti JXN-26, JLA-25.50 rotor, Beckman Coulter*). The purification was performed as described in “Purification tests of His₆-TCAB1 constructs from *E.coli*”.

Table 2-11: Composition of buffers used in protein purification.

Component	Composition
L7.5/300	100 mM HEPES pH 7.5, 300 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol, 1x EDTA-free protease inhibitor cocktail (<i>Roche</i>)
L7.5/300i	100 mM HEPES pH 7.5, 300 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol, 25 mM imidazole, 1x EDTA-free protease inhibitor cocktail (<i>Roche</i>)
L8/150i	100 mM HEPES pH 8, 150 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol, 25 mM imidazole, 1x EDTA-free protease inhibitor cocktail (<i>Roche</i>)
NiW7.5/150	20 mM HEPES pH 7.5, 150 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol, 25 mM imidazole
NiW7.5/500	20 mM HEPES pH 7.5, 500 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol, 25 mM imidazole
NiE7.5/150	20 mM HEPES pH 7.5, 150 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol, 500 mM imidazole
NiW8/150	20 mM HEPES pH 8, 150 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol, 25 mM imidazole
NiW8/500	20 mM HEPES pH 8, 500 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol, 25 mM imidazole
NiW8/500+T	20 mM HEPES pH 8, 500 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol, 25 mM imidazole, 0.05% Triton X-100
NiE8/150	20 mM HEPES pH 8, 150 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol, 250 mM imidazole
GSTE7.5/150	20 mM HEPES pH 7.5, 150 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol, 10 mM reduced glutathione
B7.5/150	20 mM HEPES pH 7.5, 150 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol
B7.5/150T	20 mM HEPES pH 7.5, 500 mM NaCl, 2 mM TCEP, 10% glycerol
B7.5/500	20 mM HEPES pH 7.5, 500 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol
B8/150	20 mM HEPES pH 8, 150 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol
B8/250	20 mM HEPES pH 8, 250 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol
B8/350	20 mM HEPES pH 8, 350 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol
B8/450	20 mM HEPES pH 8, 450 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol
B8/1000	20 mM HEPES pH 8, 1000 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol

2.3.5) Electron microscopy (EM)

Sample preparation

During the final step of protein purification, namely gel filtration, glycerol was omitted in order to avoid dialysis of the protein sample. The sample for negative stain EM was collected off the superdex 200 pg column from the very apex of the corresponding protein peak. In doing so, the protein was obtained at high concentration and purity. Afterwards, the protein was diluted to working conditions (usually 0.01 mg/mL) using the same buffer as for gel filtration. 20 µL of the protein was admixed with 5 µL 0.5% glutaraldehyde solution (diluted from 25% EM grade glutaraldehyde) and incubated at 4°C for 10 min. Afterwards, the cross-linking reaction was centrifuged at 4°C and 12,000x *g* for 10 min, in order to sediment large aggregates, and immediately used for EM grid preparation.

Negative stain EM grid preparation

Carbon coated EM grids (copper/rhodium, 10-12 nm carbon layer, 400 square mesh) were obtained from Asst. Prof. Sara Sandin's group. For negative stain EM grid preparation, grids were first glow discharged for 60 sec (*PDC-32G, HARRICK*). Afterwards, 4 µL of protein sample was applied on top of the carbon surface for 1 min at room temperature and subsequently blotted with torn up whatman paper. Immediately after, 4 µL of aqueous 2% uranyl acetate solution was applied on the same grid for 1 min and blotted. The EM grid was then airdried for 10 min before it was stored away or used for data collection.

Negative stain EM imaging

For negative stain EM imaging, the T12 cryo-electron microscope (*FEI*) in Asst. Prof. Sara Sandin's laboratory was used. Manufacturer's operation manual was followed. Data collection was performed at defocus values of -1.5 to -1.0 μm .

Data processing

Particles were picked using Eman 2.0 and then exported to Relion 2.0 for 2D classification and averaging, 2D subclass selection, 3D reconstruction, masking and post-processing.^{411,412}

2.3.6) X-ray crystallography

Sample preparation

After purification, protein samples were concentrated using ultra-centrifugal concentration units (*Merck Millipore*), passed through a 0.2 μm pore filter and then centrifuged at 4°C and 12,000x *g* for 10 min.

Setting up of crystallization plates

For setting up sitting-drop crystallization plates, *Crystal Gryphon* crystallization robot (*ARI*) was used and manufacturer's operation manual followed. Afterwards, crystallization plates were sealed and incubated at 20°C in *Rock Imager 1000* (*Formulatrix*), in which droplet pictures were taken at a regular schedule up to 30 days.

Chapter 3: Structural Investigations on the CST complex

3.1) Recombinant Expression of Ctc1 in *E.coli*

3.1.1) Initial Ctc1 construct design and expression results

Ctc1 is 134.6 kDa protein that spans 1217 amino acid residues with a theoretical pI of 8.23. According to initial sequence analysis using ELM resource, Ctc1 is predicted to be fairly well ordered and contain multiple globular domains, including the putative OB-folds (Figure 3-1).⁴¹³ Based on this analysis, secondary structure predictions and a detailed inspection of primary sequence elements, constructs were designed to encompass regions of interest, namely the N-terminal single-stranded DNA binding domain and the C-terminal region involved in CST complex formation, and then submitted to PPP for cloning and expression screening.

Figure 3-1: Ctc1 sequence analysis using ELM resource and construct design. ELM resource identifies eukaryotic linear motifs and compiles relevant information from computational resources that provide algorithms for domain identification (SMART and Pfam),^{414,415} globular domain tendency plots (GlobPlot),⁴¹⁶ and disorder prediction (IUPred).⁴¹⁷ Putative OB folds were annotated according to literature reference.²⁷⁹ Ctc1 amino acid residues 221 to 464 are required for single-stranded DNA binding activity, whereas residues 700 to 1217 are sufficient for CST complex formation, but do not constitute the minimal Stn1-Ten1 binding site.

Figure 3-2: Total and soluble expression results of Ctc1 constructs submitted to PPP. DOM (domain): annotations correspond to GlobProt prediction of globular domains. LC (low complexity): disordered region as predicted by IUPred. Green boxes annotate putative OB-folds. **a)** Total expression results as can be conferred from whole cell lysates. **b)** Soluble expression results as can be conferred after NiNTA purification of the soluble fraction following cell lysis. **c)** Legend for PPP annotations. **d)** Coomassie stained SDS polyacrylamide gel after electrophoresis of whole cell lysates of various constructs. Red arrows highlight main protein bands.

According to PPP annotations, constructs spanning the two single-stranded DNA binding OB-folds (residues 221-464) were well-expressed but formed insoluble aggregates that were sequestered in inclusion bodies. Ctc1 constructs covering the C-terminal Stn1-Ten1 interaction domain, on the other hand, failed to express altogether (Figure 3-2). Consequently, no purified protein could be obtained from PPP.

The formation of inclusion bodies can have various reasons, such as improper folding, denaturation and low protein solubility. In order to address these issues, it is a common practice to add a globular, well-structured tag to the N-terminus of the protein. In doing so, the N-terminal tag may serve as a scaffold in protein folding, limit off-pathway folding events, and increase overall protein solubility. As a possible downside, cleavage of these tags can lead to precipitation and loss of purified protein.

In order to improve Ctc1 expression levels and prevent the formation of insoluble aggregates, all constructs were cloned as N-terminal GST-fusion proteins, which has the added benefit of providing an affinity tag for GSTrap purification. Following the introduction of GST-tags, Ctc1 constructs 1-700, 1-496, 1-464 and 203-700 (from here on referred to as Ctc1 N1, Ctc1 N2, Ctc1 N3 and Ctc1 N4, respectively), which had previously displayed high expression levels but low solubility, were tested first. Proteins were expressed in BL21(DE3)pLysS cells by IPTG induction and incubation at 18°C overnight and 37°C for three hours, respectively. Interestingly, in contrast to aforementioned observations, incubation at 18°C overnight did not lead to any discernable protein expression. Instead, following GSTrap purification, SDS-PAGE and Coomassie staining of elution fractions, a prominent band running at approximately the size of the GST-tag appeared, indicating extensive C-terminal degradation.

Figure 3-3: GST-Ctc1 N3 as an example of N-terminal Ctc-1 construct expression and purification. The protein was expressed at 37°C for 3 hours. Following cell lysis and centrifugation, the supernatant was subjected to GSTrap purification and gel filtration. WCL: Whole cell lysate, SN: Supernatant, FT: Flow-through, E: Pooled elution fractions, Fn: Fraction n. **a)** GSTrap chromatogram of GST-Ctc1 N3. **b)** Superdex 200 pg 10/300 GL chromatogram of GST-Ctc1 N3 following GSTrap purification. **c and d)** SDS-PAGE and Coomassie stain (c) and anti-GST western blot (d) of protein samples taken from indicated fractions. Despite high expression levels, large amounts of GST-Ctc1 N3 are missing from the supernatant following centrifugation of the whole cell lysate. Red boxes annotate degradation products that contain the GST-tag, which give a strong signal in anti-GST western blot. The band corresponding to GST-Ctc1 N3 (MW = 78.5 kDa) is highlighted with a red arrow, whereas the green arrow indicates the predominant 60 kDa contamination, presumably HSP60, which reappears in the purification of all N-terminal Ctc1 constructs, co-elutes with GST-Ctc1 N3 close to the void volume during gel filtration, and does not give a signal in anti-GST western blot.

Expression at 37°C, on the other hand, yielded high levels of protein, the majority of which, however, were sequestered in inclusion bodies. Similar to the expression at 18°C, GSTrap purification of the soluble fraction yielded extensive C-terminal degradation products containing the GST-tag, which was confirmed by anti-GST western blot. In addition to the degradation products, the GSTrap eluate also contained a protein which runs at approximately 60 kDa by SDS-PAGE analysis (Figure 3-3 c). Since the 60 kDa band reappears in the purification of various protein constructs of different lengths, and is not detectable in anti-GST western blots (Figure 3-3 d), it most likely corresponds to a prokaryotic molecular chaperone, such as HSP60. When GSTrap purified fractions were pooled, concentrated and subjected to gel filtration, Ctc1 constructs were shown to elute in a peak that was overlapping with the void volume. However, the volume at which this peak appears corresponds to a molecular weight of approximately 600,000 kDa. It is therefore unlikely to contain an ordered oligomer, but instead most probably consists of soluble aggregates. In addition, it displays significant contamination by the aforementioned molecular chaperone, which further implies aggregate formation.

Because expression tests did not provide a solution to prevent the formation of insoluble aggregates, an attempt was made to obtain soluble protein from inclusion bodies by refolding. This approach has been used successfully in the Rhodes group to refold small proteins. Despite the large size of the GST-Ctc1 constructs, which hampers refolding by broadening the free enthalpy landscape, the rigid and stable structure of the GST-tag could act as a scaffold to guide the folding process. For this reason, proteins contained in cell pellets were first resuspended under denaturing conditions in a concentrated urea solution, and then refolded using dialysis to reduce the concentration of the denaturant. Afterwards, the dialyzed solution was subjected to GSTrap purification and size exclusion chromatography. Unfortunately, this approach did not yield soluble N-terminal Ctc1 constructs. The previously observed degradation products containing the GST-tag reappeared as the dominant species, and only minute amounts of GST-Ctc1 could be recovered. And even though this approach eliminated chaperone

contamination, it did not prevent Ctc1 constructs from eluting close to the void volume during gel filtration, indicating the repeated formation of soluble aggregates. These observations indicate that the GST-tag was indeed refolded into its native structure,

Figure 3-4: GST-Ctc1 N1 as an example for refolding and purification of N-terminal Ctc-1 constructs. After lysis and centrifugation, cell pellets were resuspended, dialyzed and subjected to GSTrap purification and gel filtration. WCL: Whole cell lysate, SN: Supernatant, RS: Resuspended cell pellet, DIA: Dialyzed resuspension, FT: Flow-through, E: Pooled elution fractions, Fn: Fraction n. **a)** GSTrap chromatogram of refolded GST-Ctc1 N1. **b)** Superdex 200 pg 10/300 GL chromatogram of GST-Ctc1 N1 following GSTrap purification. **c)** SDS-PAGE and Coomassie stain of protein samples taken from indicated fractions. The contrast in the right half of the image was selectively increased in order to visualize weak bands. The band corresponding to GST-Ctc1 N1 (MW = 104.9 kDa) is highlighted with red arrows, whereas red boxes annotate degradation products that contain the GST-tag. A comparison between lanes WCL and RS demonstrates that resuspension of the cell pellet did not efficiently solubilize GST-Ctc1 N1.

allowing it to interact with the GSTrap column. A comparison between samples taken from the whole cell lysate and the resuspended cell pellet using SDS-PAGE analysis reveals that protein bands corresponding to Ctc1 constructs were not recovered efficiently following the resolubilization and purification step (Figure 3-4).

Having failed to obtain soluble Ctc1, an attempt to co-express Ctc1 constructs that comprised the C-terminal trimerization domain with Stn1 and Ten1 was made, in the anticipation that co-expression and complex formation with Stn1 and Ten1 would have a stabilizing effect on Ctc1, increase expression levels and also prevent the formation of soluble aggregates. GST-tagged Ctc1 constructs 1-1217, 203-1217, 521-1217, 828-1217 and 958-1217 (from here on referred to as Ctc1 C1, C2, C3, C4 and C5, respectively) and His₆-tagged full length Stn1 and Ten1, which were cloned into a separate plasmid, were co-expressed in BL21(DE3)pLysS cells at 37°C for three hours. Samples were taken at and three hours after IPTG induction and analyzed by SDS-PAGE and Coomassie staining. However, despite the emergence of two strongly expressed protein bands, which presumably belong to Stn1 and Ten1, there were no discernable bands corresponding to Ctc1 constructs (Figure 3-5).

Figure 3-5: Expression test of C-terminal Ctc1 constructs together with Stn1 and Ten1. The picture displays the result of SDS-PAGE and Coomassie stain of protein samples taken at the timepoint of induction (0 h) and after 3 h of incubation at 37°C. C1ST to C5ST indicate samples taken from cell cultures after 3 h of incubation that express CST complexes containing GST-tagged Ctc1 constructs C1 to C5. Protein bands corresponding to His₆-tagged Stn1 (MW = 44.6 kDa) are highlighted by a blue arrow, whereas the green arrow indicates protein bands corresponding to Ten1 (MW = 13.8 kDa).

Expression tests were also performed in C43(DE3)pLysS and Rosetta(DE3)pLysS cells. C43(DE3)pLysS is a BL21(DE3)pLysS derived strain that contains genetic mutations phenotypically selected for conferring tolerance to toxic proteins by preventing cell death associated with recombinant expression, and is generally used for difficult to express proteins. Rosetta(DE3)pLysS cells, on the other hand, express tRNAs for codons that are typically rare in *E.coli*, thereby enabling translation which is otherwise limited by *E.coli* codon usage. It is important to note that in these expression tests, an Stn1 construct was utilized, that, in accordance with the published X-ray crystal structure of Stn1-Ten1, was lacking 18 N-terminal amino acid residues (from here on referred to as $\Delta 18\text{Stn1}$).²⁸⁰ Unfortunately, using these strains also failed to produce satisfactory expression levels of Ctc1 constructs (Figure 3-6 a). Finally, in an attempt to slow folding kinetics, which has the potential to prevent off-pathway folding events and aggregate formation, the CST complex was co-expressed in Rosetta(DE3)pLysS cells at 18°C overnight. However, expression at lower temperatures also did not yield any visible amounts of Ctc1, while at the same time producing large quantities of $\Delta 18\text{Stn1-Ten1}$ (Figure 3-6 b and c). These results indicate that there was no problem with the method of expression, since both $\Delta 18\text{Stn1}$ and Ten1 were well-expressed. Instead, it appears that Ctc1 itself represents a difficult target for recombinant expression in *E.coli*.

In summary, all expression tests of Ctc1, including those performed by PPP, failed to express Ctc1 constructs that contained the C-terminal region (residues 958-1217), whereas constructs spanning N-terminal OB-folds (residues 221-464) resulted in significant amounts of recombinant protein, which, however, were sequestered in inclusion bodies. Since C-terminal constructs consistently failed to express, these observations were suggestive of C-terminal residues having a disruptive effect on protein stability and expression efficiency. In order to test this, C-terminal Ctc1 constructs were re-cloned using an alternative C-terminal boundary, which, however, would retain the ability to form a CST complex. In order to determine this boundary, the aforementioned

Figure 3-6: Expression and purification tests of C-terminal Ctc1 constructs in complex with Δ18Stn1 and Ten1. C1ΔST to C4ΔST refer to CST complexes comprising GST-tagged Ctc1 C1 to C4 and His₆-tagged Δ18Stn1. Blue arrows indicate the position of bands corresponding to His₆-tagged Δ18Stn1 (MW = 42.8 kDa), whereas green arrows highlight bands corresponding to Ten1 (MW = 13.8 kDa). WCL: Whole cell lysate, SN: Supernatant, FT: Flow-through, B: Beads. **a)** SDS-PAGE and Coomassie stain of protein samples taken at the timepoint of induction (0 h) and after 3 h of incubation at 37°C in C43(DE3)pLysS and Rosetta(DE3)pLysS cells, respectively, display high expression levels of both His₆-tagged Δ18Stn1 as well as Ten1, but not of Ctc1. **b)** SDS-PAGE and Coomassie stain of protein samples taken at the timepoint of induction (0 h) and after overnight incubation at 18°C in Rosetta(DE3)pLysS cells similarly displays prominent bands for His₆-tagged Δ18Stn1. There is no indication of Ctc1 expression for any of the tested Ctc1 constructs. **c)** SDS-PAGE and Coomassie stain of protein samples taken from indicated fractions. The supernatant was loaded onto NiNTA beads, which, after washing, were stripped with SDS-PAGE loading buffer. A comparison between all four constructs does not reveal any specific band that corresponds to any Ctc1 construct.

mutational study in which 1196-Δ7 had been shown to disrupt interactions with Stn1-Ten1,³²⁵ was taken into consideration. Consequently, the new C-terminal boundary was devised to terminate at S1204.

An additional concern was that in previous co-expressions, plasmids encoding for Ctc1 and Stn1-Ten1, respectively, were co-transfected. In order to exclude any possible artefacts, such as the favoured expression of the Stn1-Ten1 encoding plasmid, a second plasmid was produced that incorporated all three genes. In the following expression trials, which were performed in Rosetta(DE3)pLysS cells both at 18°C and 37°C, combinations between Stn1/Δ18Stn1-Ten1 and Ctc1 1-1217/1-1204 were tested.

Unfortunately, none of the condition and construct combinations resulted in the expression of Ctc1, whereas Stn1/Δ18Stn1 and Ten1 were expressed at the usual high levels. In conclusion, the attempt to express Ctc1 and Ctc1 complexes in *E.coli* was abandoned in favor of expression in insect cells.

3.2) Recombinant Expression of Ctc1 and the CST complex in Insect Cells

3.2.1) Expression of Ctc1 alone is inadequate for protein preparation and structural studies

Expression of proteins in insect cell expression has the benefit of providing a eukaryotic expression system with protein processing mechanisms, while also enabling expression levels that are usually higher than in tissue culture cells. It is a well-developed method that utilizes baculovirus as vectors to infect cells with genes of interest. In addition, well-established cloning strategies provide easy-to-follow protocols for the co-expression of multiple genes, which was of particular interest for this work. However, since the baculovirus-insect cell expression system needed to be established in our laboratory, first, and the means for co-expression of proteins in insect cells were not available at the time, Ctc1 constructs were first cloned and expressed in isolation while at the same time setting up the baculovirus system.

Because the production of viruses is a time-consuming process, as a preliminary test, Ctc1 constructs comprising C-terminal OB-folds were first screened for expression in HEK293T cells. Since in human derived cell lines, recombinant protein expression levels tend to be low, SDS-PAGE and Coomassie stain of whole cell lysates is usually unsuitable to visualize protein bands. Instead, in order to detect even low levels of recombinant protein, constructs were tagged with a 3xFLAG-tag on the N-terminus for western blot analysis, as well as a C-terminal GFP-tag, which serves to give an early indication of transfection efficiency, protein expression and stability. All C-terminal Ctc1 constructs that were screened resulted in low, but detectable expression levels, qualifying them for expression tests in insect cells (Figure 3-7).

Figure 3-7: Tissue culture expression screen of C-terminal Ctc1 constructs. The picture illustrates an anti-GST western blot of samples taken three days after transfection of HEK293T cells with plasmids encoding indicated proteins. C1 to C4 refer to respective C-terminal Ctc1 constructs. FLAG-GFP was expressed as a positive control. In comparison, expression levels of Ctc1 constructs are very low.

Initially, in order to provide a proof of concept, only GST-tagged Ctc1 constructs C1 and C4, which are the longest and shortest C-terminal Ctc1 constructs, respectively, were cloned into baculovirus vectors and amplified. Unfortunately, expression of full-length Ctc1 did not yield any detectable amounts of recombinant protein. GST-Ctc1 C4, on the other hand, was expressing, but displayed very low binding efficiency to the GSTrap affinity column, as evidenced by nearly equal protein band intensities between input and flow-through lanes following anti-GST western blot. In the subsequent gel filtration step, GST-Ctc1 C4 exclusively eluted in the void volume, indicating the formation of soluble aggregates. If so, the previously described low GSTrap purification efficiency may be attributable to the reduced accessibility of the GST-tag. The second peak eluting off the gel filtration column contained a protein approximately the size of GST. However, the corresponding protein band did not produce a signal in anti-GST western blot, suggesting that this contaminant is not a degradation product of GST-Ctc1 C4, but more likely an endogenous GST protein with an incompatible epitope (Figure 3-8).

In conclusion, expression of Ctc1 in insect cells lead to very low expression levels and the formation of soluble aggregates, which may possibly originate from protein instability or improper folding, as was seen for the bacterial expression of Ctc1. In order

to improve Ctc1 expression, Ctc1 was co-expressed with Stn1 and Ten1, anticipating interactions between Ctc1 and Stn1 to have a stabilizing effect.

Figure 3-8: Purification of GST-tagged Ctc1 C4 from insect cell culture. WCL: Whole cell lysate, SN: Supernatant, FT: Flow-through, E: Pooled elution fractions, Fn: Fraction n. **a)** GSTrap chromatogram of GST-Ctc1 C4. **b)** Superdex 200 pg 16/600 chromatogram of GST-Ctc1 C4 following GSTrap purification. **c and d)** SDS-PAGE and Coomassie stain (c) and anti-GST western blot (d) of protein samples taken from indicated fractions. The band corresponding to GST-Ctc1 C4 (MW = 69.9 kDa) is highlighted with red arrows. Red boxes annotate contaminants that specifically purify from GSTrap but do not give a signal in anti-GST western blot, indicating that these aren't degradation products, but more likely endogenous GST that has an incompatible epitope. Comparison of GST-Ctc1 C4 band intensities between SN and FT lanes on western blot reveals the efficiency of GSTrap binding to be very low. During the subsequent gel filtration step, GST-Ctc1 C4 exclusively elutes in the void volume.

3.2.2) Co-expression and purification of the CST complex: Problems and Solutions

As mentioned earlier, co-expression of a protein complex can have a stabilizing effect of its interacting subunits. When the MultiBac expression system became available, GST-tagged Ctc1 constructs C1, C2, C3 and C4 were cloned into one plasmid, His₆-tagged $\Delta 18\text{Stn1}$ and untagged Ten1 into another, and the two plasmids were fused via Cre-Lox recombination before the combined gene cassettes were inserted into an EmBacY bacmid. Western blots of expression tests during amplification of the baculovirus recurrently revealed that in comparison, expression levels of the shortest Ctc1 construct (Ctc1 C4) were the highest, followed by the second shortest (Ctc1 C3) and approximately the same for the two longer constructs which contained N-terminal single-stranded DNA binding OB-folds (Ctc1 C1 and Ctc1 C2), whereas $\Delta 18\text{Stn1}$ expression levels were virtually invariant (Figure 3-9 a and b). In order to optimize protein expression, the effects of various parameters were tested, including multiplicity of infection (MOI), cell count at the time of infection and incubation time. Whereas different cell counts had little to no impact on overall expression levels, increasing MOIs exerted a negative effect (Figure 3-9 c). When infected at an MOI of 1, it was found that protein expression peaked after three to four days (Figure 3-9 d). Hence, in order to save time and resources, subsequent protein expressions were performed by infecting 2-4 million cells/mL at an MOI of 1, followed by a three-day growth at 27°C.

During the purification of CST constructs, multiple problems were encountered. As evidenced by SDS-PAGE and Coomassie staining, a comparison between respective band intensities reveals that expression levels of the $\Delta 18\text{Stn1}$ -Ten1 subcomplex are much higher than that of the trimeric CST complex. Therefore, an additional purification step was required to separate the two complexes. However, since the eluate from the initial NiNTA column contains $\Delta 18\text{Stn1}$ -Ten1 in a much larger abundance than the CST complex, this also meant that final protein yields would be reduced drastically, which was a concern. Secondly, the GSTrap binding efficiency remained very low and almost all of the protein

appeared in the flow-through, making GSTrap affinity purification unsuitable for CST protein preparation. When NiNTA elution fractions were instead directly applied on gel filtration, CST eluted in or close to the void volume, once again implying the formation of

Figure 3-9: Optimization CST expression in insect cells. C1ΔST to C4ΔST refer to CST complexes comprising GST-tagged Ctc1 C1 to C4 and His₆-tagged Δ18Stn1. **a)** anti-GST western blot of samples taken three days after infection of Sf9 cells with baculovirus encoding for indicated protein complexes reveals higher expression levels of Ctc1 C3 and C4 compared to C1 and C2. **b)** anti-His₆ western blot of the same samples displays little to no difference in Δ18Stn1 expression levels. **c)** anti-GST western blot of samples taken three days after infection of Sf9 cells at indicated MOIs. Expression levels decrease with increasing MOI. **d)** anti-GST western blot of samples taken after the indicated time period post-infection of Sf9 cells. Ctc1 expression peaks after 3 to 4 days.

a)

b)

c)

d)

Figure 3-10: Issues with the purification of Ctc1 C3- Δ 18Stn1-Ten1 (C3 Δ ST) following co-expression in insect cells. The supernatant following centrifugation of the cell lysate was subjected to NiNTA affinity chromatography. Since C3 Δ ST did not bind to the GSTrap column, the GSTrap flow-through was concentrated through repeated application on NiNTA and subjected to gel filtration. Red, blue and green arrows annotate GST-Ctc1 C3 (MW = 104.2 kDa), His₆- Δ 18Stn1 (MW = 42.8 kDa) and Ten1 (MW = 13.8 kDa), respectively. WCL: Whole cell lysate, SN: Supernatant, FT: Flow-through, E: Pooled elution fractions, Fn: Fraction n. **a)** Superdex 200 pg 16/600 chromatogram of C3 Δ ST following concentration over NiNTA. **b,c,d)** SDS-PAGE and Coomassie stain (**b**), anti-GST western blot (**c**) and anti-His western blot (**d**) of protein samples taken from indicated fractions during the purification of C3 Δ ST. A comparison of band intensities reveals that the NiNTA eluate contains a significantly larger amount of His₆- Δ 18Stn1 and Ten1 than of GST-Ctc1 C3, indicating the enrichment of both CST and Δ 18Stn1-Ten1 complexes during NiNTA purification. As evidenced by anti-GST and anti-His western blots, the first two peaks (including the void volume peak) during gel filtration contain both GST-Ctc1 C3 and His₆- Δ 18Stn1, whereas the third prevalent peak consists of Δ 18Stn1-Ten1. Since the volume at which the second peak elutes roughly corresponds to a molecular weight of 300 kDa, the complex is either forming a higher-order oligomer or a soluble aggregate.

soluble aggregates or higher-order oligomers (Figure 3-10). Finally, TEV-cleavage of the Ctc1 GST-tag was very inefficient, likely due to the documented protein aggregation leading to inaccessibility of the cleavage site, and required an either very large amount of TEV protease or prolonged incubation at 4°C (Figure 3-11).

Figure 3-11: Low TEV-cleavage efficiency of the Ctc1 GST-tag. The picture illustrates an anti-GST western blot of samples taken after the indicated amount of incubation time at 4°C from reactions containing TEV in a TEV to protein ratio 1:100, 1:50 and 1:10. Even at a ratio of 1:50, GST-tag cleavage is still incomplete after 48 h. Only at a 1:10 ratio is the TEV-cleavage completed after 24 h.

The formation of soluble aggregates, and consequently the elution of the protein in the void volume during gel filtration, can have various causes. These include, but are not limited to, the co-purification of contaminants which drive the formation of higher-order complexes, as well as buffer conditions which may negate the net charge of the protein and thereby reduce its solubility. In order to address these possibilities, a broad screen was performed, testing various pH conditions, coupled with the use of 0.1% DDM during the entirety of the purification, or the introduction of an additional 0.05% Triton X-100 wash step during NiNTA affinity chromatography. Whereas the addition of the

Figure 3-12: Effects on Ctc1 C3-Δ18Stn1-Ten1 (C3ΔST) purification following optimization of purification conditions. The figure displays a pair-wise arrangement of superdex 200 pg 10/300 GL chromatograms and silver staining of samples taken from annotated fractions following SDS-PAGE. Red, blue and green arrows highlight protein bands corresponding to GST-Ctc1 C3 (104.2 kDa), His₆-Δ18Stn1 (MW = 42.8 kDa) and Ten1 (MW = 13.8 kDa), respectively. **a)** pH6, +0.05% Triton X-100 NiNTA wash step. **b)** pH6, +0.1% DDM. **c)** pH8, +0.05% Triton X-100 NiNTA wash step. **d)** pH8, +0.1% DDM. A comparison between chromatograms shows that the introduction of a 0.05% Triton X-100 wash step during NiNTA affinity purification drastically reduces the intensity of the void volume peak (a and c). Simultaneously, purification at pH 8 increases the peak containing the CST complex (c). Silver straining reveals that the obtained complex is not stoichiometric and is intermixed with the Δ18Stn1-Ten1 subcomplex.

detergent DDM (0.1%) to the entirety of the purification process, starting with the lysis buffer, increased overall protein yields, the purity was comparably low and the void volume peak intensity was reduced only slightly compared to previous experiments. The introduction of a 0.05% Triton X-100 wash step to NiNTA purification, on the other hand, lead to a significant increase in NiNTA elution purity and greatly reduced the amount of protein eluting in the void volume during gel filtration, while simultaneously increasing the intensity of the secondary peak containing the soluble CST complex. This effect appeared to be most pronounced when using buffers at pH 8.0. As a consequence, subsequent purifications were performed at pH 8.0, with the addition of a 0.05% Triton X-100 wash step during the initial NiNTA affinity purification (Figure 3-12).

Coincidentally, as a consequence of preventing the formation of soluble aggregates and higher-order oligomers, the efficiency of TEV-cleavage also improved tremendously, presumably due to the GST-tag now being accessible for cleavage (Figure 3-13).

In order to isolate CST from the dimeric $\Delta 18\text{Stn1-Ten1}$ subcomplex, a Heparin purification step was introduced, making use of the large difference in DNA binding affinities between the two complexes. Heparin is a highly acidic polysaccharide that consists of alternating units of D-glucosamine and D-glucuronic acid, most of which are modified with sulfate and amidosulfate groups. As such, Heparin is a close analogue of DNA and is therefore commonly used for the purification of DNA binding proteins. The application of NiNTA purified fractions onto Heparin resulted in a clear separation between $\Delta 18\text{Stn1-Ten1}$ and CST complexes. Whereas $\Delta 18\text{Stn1-Ten1}$ did not bind to Heparin at 150 mM NaCl, (at which the sample was loaded) CST eluted as a complex containing stoichiometric amounts of Ctc1, Stn1 and Ten1 constructs, at around 250 mM to 450 mM NaCl, depending on the Ctc1 construct. In the following gel filtration step, only a small amount of protein eluted in the void volume, whereas the majority eluted as a single peak containing CST of very high purity (Figure 3-14).

Figure 3-13: Increased TEV-cleavage efficiency of the Ctc1 GST-tag following optimization of purification conditions. Red, light red, blue and green arrows annotate GST-Ctc1 C3 (104.2 kDa), Ctc1C3 (MW = 77 kDa), His6- Δ 18Stn1 (MW = 42.8 kDa) and Ten1 (MW = 13.8 kDa), respectively. The grey arrow highlights the protein band corresponding to TEV protease, whereas the orange arrow annotates the cleaved GST-tag. SDS-PAGE and Coomassie staining of samples taken after the indicated amount of incubation time at 20°C and 4°C from reactions containing TEV in a TEV to protein ratio 1:100, 1:50 and 1:10. TEV-cleavage at 20°C is complete in 2 h. At 4°C, TEV-cleavage at a TEV to protein ration 1:100 is already finished after 6 h. Comparing this to previous TEV-cleavage reactions, where TEV-cleavage wasn't complete after 48 h under similar conditions, it is clear that following optimization of purification conditions and the consequent prevention of soluble aggregate formation, TEV-cleavage efficiency has improved dramatically.

In conclusion, a purification strategy was established that included an additional 0.05% Triton X-100 wash during the initial NiNTA affinity purification and in which GSTrap was substituted for a Heparin purification step that efficiently separated trimeric CST from dimeric Stn1-Ten1 (Figure 3-15). However, between the four different Ctc1 constructs, only Ctc1 C1 and C3 could be purified following this strategy. The shortest construct (Ctc1 C4), which exhibited the highest levels of expression, did not have a higher binding affinity to the Heparin matrix when in complex with Δ 18Stn1-Ten1, and could therefore not be isolated from the dimeric subcomplex using this purification method. Interestingly, this points to a curious contrast between CST complexes comprising Ctc1 C3 (C3 Δ ST) and C4

Figure 3-14: Introduction of a Heparin purification step enables the purification of a stoichiometric CST complex. The figure illustrates the purification of Ctc1 C3-Δ18Stn1-Ten1 (C3ΔST) from insect cells. Red, blue and green arrows annotate GST-Ctc1 C3 (MW = 104.2 kDa), His6-Δ18Stn1 (MW = 42.8 kDa) and Ten1 (MW = 13.8 kDa), respectively. E: Pooled elution fractions, Fn: Fraction n. **a)** Heparin chromatogram of the concentrated eluate from the initial NiNTA purification step. Δ18Stn1-Ten1 elutes at 150 mM NaCl, at which the sample is loaded onto the column. C3ΔST elutes at a salt concentration roughly approximating 250-350 mM NaCl. **b)** Superdex 200 pg 10/300 GL chromatogram of Heparin fractions F8 to F11. The C3ΔST complex elutes at roughly the volume corresponding to a stoichiometric complex (approximately 180 kDa). **c and d)** SDS-PAGE and Coomassie stain of samples taken from indicated fractions. Heparin elution fractions F8-11 and F12-15 (c) were pooled and concentrated in order to better visualize their content. Whereas the first peak that elutes from the Heparin salt gradient contains C3ΔST in an already high purity, the second peak, which partially overlaps with the first, mostly consists of contaminations. C3ΔST elutes off the superdex 200 pg column as a highly pure and stoichiometric complex.

(C4ΔST). Whereas both Ctc1 constructs do not contain N-terminal single-stranded DNA binding OB-folds, only C3ΔST engaged in significantly stronger interactions with Heparin than the Δ18Stn1-Ten1 subcomplex. From the aforementioned analysis of disease associated Ctc1 mutations, it was reported that V665G negatively affects single-stranded DNA binding, which may provide an explanation for the difference in observations, since this residue is included in Ctc1 C3 but not in Ctc1 C4 constructs.³²⁵ However, further structural and biochemical information is required to propose a more definitive and detailed answer. In the case of Ctc1 C2, on the other hand, a large fraction of the protein underwent C-terminal degradation, which in turn disrupted CST complex formation. Consequently, it was not possible to purify Ctc1 C2 with His₆-tagged Δ18Stn1-Ten1 using NiNTA affinity chromatography.

Eventually, only CST complexes C1ΔST and C3ΔST, comprising Ctc1 C1 and C3, were purified in sufficiently large quantities for structural studies (Figure 3-16 and Figure 3-17). However, expression levels and consequently obtained protein yields between the two constructs varied significantly. Whereas for C3ΔST, from one liter of insect cell culture 1.6 mg of the purified protein complex could be obtained, the yield dropped to 0.3 mg per liter for C1ΔST. As a result, crystallization screens were predominantly carried out with the smaller protein complex, as crystallization requires larger amounts of purified protein, while the full-length construct was primarily used for electron microscopy (EM).

Figure 3-15: Purification strategy of CST complex from insect cells. See “Materials and Methods” for a detailed protocol of the purification strategy.

Figure 3-16: Established purification strategy for Ctc1 C3-Δ18Stn1-Ten1 (C3ΔST). After lysis of insect cells and centrifugation, the supernatant was subjected to NiNTA chromatography, Heparin purification, TEV-cleavage and gel filtration. Red, light red, blue and green arrows annotate GST-Ctc1 C3 (104.2 kDa), Ctc1C3 (MW = 77 kDa), His6-Δ18Stn1 (MW = 42.8 kDa) and Ten1 (MW = 13.8 kDa), respectively. E: Pooled elution fractions, TEV: TEV-cleavage product.

a) NiNTA affinity chromatogram. **b)** Heparin step-elution chromatogram of the concentrated eluate from the initial NiNTA purification step. **c)** Superdex 200 pg 16/600 chromatogram following TEV-cleavage of C3ΔST containing Heparin fractions. The C3ΔST complex elutes at roughly the volume corresponding to a stoichiometric complex (approximately 120 kDa). **d)** SDS-PAGE and Coomassie stain of samples taken from indicated fractions. Fractions of respective salt concentration during heparin step-elution were pooled and concentrated in order to better visualize their content. The already pure eluate at 250 mM NaCl was subjected to TEV-cleavage and gel filtration. C3ΔST elutes off the superdex 200 pg column as a highly pure and stoichiometric complex.

Figure 3-17: Established purification strategy for Ctc1 C1-Δ18Stn1-Ten1 (C1ΔST). This figure is analogous to Figure 3-16. Red, light red, blue and green arrows annotate GST-Ctc1 C1 (161.4 kDa), Ctc1C1 (MW = 134.3 kDa), His6-Δ18Stn1 (MW = 42.8 kDa) and Ten1 (MW = 13.8 kDa), respectively. E: Pooled elution fractions, TEV: TEV-cleavage product. **a)** NiNTA affinity chromatogram. **b)** Heparin step-elution chromatogram. **c)** Superdex 200 pg 16/600 chromatogram following TEV-cleavage of C1ΔST containing Heparin fractions. The C1ΔST complex elutes at roughly the volume corresponding to a stoichiometric complex (approximately 190 kDa). **d)** SDS-PAGE and Coomassie stain of samples taken from indicated fractions. Fractions of respective salt concentration during heparin step-elution were pooled and concentrated in order to better visualize their content. Heparin eluates at 400 mM and 450 mM NaCl were pooled and subjected to TEV-cleavage and gel filtration. C1ΔST elutes off the superdex 200 pg column as a highly pure and stoichiometric complex.

3.3) Purified CST complexes did not crystalize despite extensive screening

Since the C3ΔST, consisting of Ctc1 521-1217, Stn1 19-368 and full-length Ten1, yielded significantly higher protein yields per liter of insect cell culture compared to C1ΔST, crystallization trials were mainly performed with this CST complex. Crystallization plates were set up using a Gryphon LCP crystallization robot at various concentrations and buffer conditions (Table 3-1), and incubated at 20°C for 90 days.

The protein concentration with which to perform initial crystallization screens was determined based on how the protein solution behaved during centrifugal concentration. In theory, if during concentration of the protein solution, its volume does not change significantly, it is an indication of the solution reaching a point of saturation, at which the formation of protein crystals is more likely. Thus, crystallization plates were first set up using 2.5 mg/mL CST samples. However, since these plates contained many clear drops, higher protein concentrations were tested, as well. At 10 mg/mL, 50-80% of the crystallization plate featured drops with heavy precipitates, depending on the screen and buffer conditions.

Despite extensive screening of various conditions (see Table 3-1), no protein crystal could be obtained. Importantly, the probability of crystal formation is not only determined by protein concentration and buffer conditions, but also by construct boundaries, which may affect the nature and stability of protein-protein interactions within the crystal. In addition, it is desirable to exclude flexible regions, which may disrupt the formation of crystalline arrangements. It is therefore a common practice to also screen a wide range of protein constructs. However, insect cell expression puts a significant impediment on such an approach, since, in contrast to expression in *E.coli*, cloning, expression and purification of multiple constructs is a lengthy and time-consuming process. This is especially true, if the protein in question is difficult to express and requires the optimization of each purification step.

Table 3-1: Composition of crystallization plates.

Condition	Protein name	Protein concentration	Crystallization screen
#1	C3ΔST	2.5 mg/mL	Crystal Screen (<i>Hampton Research</i>)
#2	C3ΔST	2.5 mg/mL	PEG/Ion (<i>Hampton Research</i>)
#3	C3ΔST	2.5 mg/mL	Natrix (<i>Hampton Research</i>)
#4	C3ΔST	2.5 mg/mL	Midas (<i>Molecular Dimensions</i>)
#5	C3ΔST	2.5 mg/mL	Morpheus (<i>Molecular Dimensions</i>)
#6	C3ΔST	2.5 mg/mL	Nextal (<i>Hampton Research</i>)
#7	C3ΔST	2.5 mg/mL	Nucleix (<i>QIAGEN</i>)
#8	C3ΔST	2.5 mg/mL	SaltRx HT (<i>Hampton Research</i>)
#9	C3ΔST	2.5 mg/mL	Proplex (<i>Molecular Dimensions</i>)
#10	C3ΔST	5 mg/mL	Crystal Screen (<i>Hampton Research</i>)
#11	C3ΔST	5 mg/mL	PEG/Ion (<i>Hampton Research</i>)
#12	C3ΔST	5 mg/mL	Natrix (<i>Hampton Research</i>)
#13	C3ΔST	5 mg/mL	Midas (<i>Molecular Dimensions</i>)
#14	C3ΔST	5 mg/mL	Morpheus (<i>Molecular Dimensions</i>)
#15	C3ΔST	5 mg/mL	Nextal (<i>Hampton Research</i>)
#16	C3ΔST	5 mg/mL	Nucleix (<i>QIAGEN</i>)
#17	C3ΔST	5 mg/mL	SaltRx HT (<i>Hampton Research</i>)
#18	C3ΔST	5 mg/mL	Proplex (<i>Molecular Dimensions</i>)
#19	C3ΔST	10 mg/mL	Crystal Screen (<i>Hampton Research</i>)
#20	C3ΔST	10 mg/mL	PEG/Ion (<i>Hampton Research</i>)
#21	C3ΔST	10 mg/mL	Natrix (<i>Hampton Research</i>)
#22	C3ΔST	10 mg/mL	Midas (<i>Molecular Dimensions</i>)
#23	C3ΔST	10 mg/mL	Morpheus (<i>Molecular Dimensions</i>)
#24	C3ΔST	10 mg/mL	Nextal (<i>Hampton Research</i>)
#25	C3ΔST	10 mg/mL	Nucleix (<i>QIAGEN</i>)
#26	C3ΔST	10 mg/mL	SaltRx HT (<i>Hampton Research</i>)
#27	C3ΔST	10 mg/mL	Proplex (<i>Molecular Dimensions</i>)
#28	C1ΔST	10 mg/mL	Crystal Screen (<i>Hampton Research</i>)
#29	C1ΔST	10 mg/mL	PEG/Ion (<i>Hampton Research</i>)

In order to identify protein domains and hence better-defined construct boundaries, purified C1ΔST was subjected to limited proteolysis using trypsin and chymotrypsin (Figure 3-18). In turn, this information could then be used to design

constructs that omitted flexible regions on their termini and represented promising candidates for crystallization trials.

Figure 3-18: Limited proteolysis of C1ΔST. Red, blue and green arrows annotate, Ctc1C1 (MW = 134.3 kDa), His6-Δ18Stn1 (MW = 42.8 kDa) and Ten1 (MW = 13.8 kDa), respectively. Magenta and purple arrows highlight primary Ctc1 degradation products. **a)** SDS-PAGE and Coomassie stain of samples taken after 20 min incubation at 4°C from protease reactions at indicated protease to CST ratios. **b and c)** SDS-PAGE and Coomassie stain of samples taken at indicated time points after incubation at 4°C and a protease to CST ratio of 1 to 100.

At a protease (trypsin) to protein ratio of 1:100, two prominent bands corresponding to stable degradation products of Ctc1 emerged following trypsinization of C1ΔST for 1 hour (Figure 3-18), and persisted throughout a prolonged incubation time up to 20 hours. When these bands were sent for N-terminal Edman sequencing to IBS in Grenoble, the results which were returned were inconclusive and did not allow for determination of N-terminal boundaries. Consequently, in the absence of experimentally defined boundaries, multiple constructs were designed to include the N-terminal OB-fold, cloned as GST-fusion proteins and screened for expression in insect cells (Figure 3-19).

Figure 3-19: Design of new Ctc1 constructs that comprise N-terminal OB-folds for crystallization trials. a) Depiction of putative Ctc1 domains and designed constructs. Ctc1 amino acid residues 221-464 are required for single-stranded DNA binding, whereas residues 700-1217 are sufficient for the formation of a CST complex. Up to date, the minimal Stn1 binding site has not been determined. b) anti-GST western blot of samples taken three days after infection of Sf9 cells with baculovirus encoding for CST complexes comprising indicated Ctc1 constructs reveals higher expression levels of Ctc1 constructs spanning amino acid residues 9-1217 and 107-1217

The difference in molecular weight between the undigested protein and stable degradation products based on SDS-PAGE analysis is unreliable, since not only is the read-out very inaccurate, but also the migration behavior of the protein is affected by its pI value. As such, even though the molecular weight of Ctc1 following TEV-cleavage of the GST-tag is 134.8 kDa, the corresponding protein band runs at a molecular weight of 150 kDa. Following the expression test in insect cells, two new constructs stood out in that their expression levels were comparatively high (Figure 3-19 b). Between these two, the Ctc1 construct spanning amino acid residues 107-1217 was of particular interest, since its molecular weight after TEV-cleavage would be 123.3 kDa, which appears to coincide with the size of the degradation product highlighted with a magenta arrow in Figure 3-18 b. Since expression levels are, amongst others, an indication of protein stability, which in turn can affect the probability of crystal formation, it would be compelling to see how these two constructs behaved in crystallization trials.

In summary, the work outlined in this chapter demonstrates that the CST complex poses significant challenges for crystallographic studies. In part, these arise due to the requirement of expressing CST in insect cells. Because generating and amplifying baculoviral vectors is a lengthy process, screening for optimal protein constructs for crystallization trials becomes a very time-consuming endeavor. In addition, baculoviral expression in insect cells often provides low protein yields. This in turn necessitates the optimization of expression and purification for various protein constructs in order to achieve the protein amounts required for X-ray crystallography. Finally, CST poses an inherent challenge in that it does not readily form into protein crystals. Despite these challenges, I have managed to make significant progress. Through limited proteolysis studies and expression tests, I have identified protein boundaries and established new Ctc1 constructs that represent very promising candidates for future crystallization trials. The results presented in this chapter lay a solid foundation for further work on the structural elucidation of the CST complex through X-ray crystallography.

Our main challenge in the current chapter of the thesis was to satisfy requirements for X-ray crystallography, which were to obtain the protein of interest at high concentrations and to identify conditions at which it forms protein crystals. Sample requirements for electron microscopy (EM), on the other hand, are less stringent. Structural studies by EM are feasible with protein concentrations ranging from only 0.01 - 1 mg/mL and do not require the protein to pack into crystals. Another limitation of X-ray crystallography is that the formation of crystal lattices favors a preferred protein conformation and therefore fails to provide information on protein dynamics and alternative configurations. In contrast, EM has a strong potential to allow for the classification of distinct and alternative conformations. Therefore, in addition to crystallization screenings, I also conducted structural studies of CST using EM, which is discussed in the next chapter.

Chapter 4: CST Structure Determination by Electron Microscopy

4.1) Introduction

In electron microscopy (EM), structural information of a biological sample is obtained by collecting 2D images of single particles, which are subsequently used to reconstruct a three-dimensional volume. In contrast to conventional light microscopy, EM uses an accelerated high-voltage electron beam as a source of illumination. Since the wavelength of electrons is significantly shorter than that of visible light photons, electron microscopy has the potential to achieve atomic resolution. However, the performance of an electron microscope is primarily limited by the imaging system, including the spherical aberration of electromagnetic lenses, intensity of the electron beam and the resolution and sensitivity of the detector.

The imaging of protein samples is further complicated by the necessity to prepare a specimen that can withstand the vacuum inside an electron microscope and the exposure to high-energy electrons, which otherwise lead to the dehydration of the sample and radiation damage. The most common methods for EM sample preparation are negative staining and vitrification. In negative staining, the protein sample is applied onto a carbon coated grid and embedded in a layer of heavy metal salt crystals. Since the opaqueness to electrons correlates with the atomic number of an element, heavy metals scatter electrons much more strongly than biological material, thereby introducing a high contrast and allowing the visualization of even small proteins. The precipitation of heavy metal salts onto the sample also causes the protein to adopt preferred orientations while adsorbing to the carbon layer. This allows for a quick assessment of the homogeneity of the material, but on the other hand limits the number of angles at which the protein can be viewed. Consequently, it affects the number of 2D classes, which is relevant for the reconstruction of the three-dimensional structure (as will be discussed later). Due to the coating of the protein with heavy metal salts, negative staining also limits the achievable resolution to approximately 20 Å. Negative staining is therefore mostly used to evaluate

the quality of a protein sample prior to cryo-EM, or to image and collect structural data on small proteins for which cryo-EM is unsuitable.

In cryo-EM, the preparation of the protein sample is performed by vitrification. The protein solution is applied on a holey carbon grid, blotted and shock-frozen. Thus, the sample is solidified into a thin layer while at the same time avoiding the formation of ice crystals, which would otherwise damage the protein and generate high contrast in the acquired images. Since the protein solution is instantly frozen solid, the protein is preserved in a native environment and exists in various orientations and conformational states. As opposed to negative staining, vitrification introduces little to no artifacts during the preparation of the EM sample and does not impose a limit on the achievable resolution. However, since the scattering properties of the biological material and the surrounding buffer is fairly similar, the inherent contrast of vitrified samples is much lower in comparison to negative staining. Therefore, only proteins that are large enough to generate sufficient contrast (around 300 kDa) can be distinguished using this method, whereas the visualization and processing of smaller molecules may pose significant difficulties.

Because EM samples are very thin and predominantly consist of light elements that are rather opaque to electrons, in-focus images generally contain very little contrast. Therefore, image acquisition is usually performed in under-focus to induce phase shifts between scattered and unscattered electrons and thus generate phase contrast. Given a weakly scattering object, the contrast transfer function (CTF), a quasi-periodic sine function in reciprocal space, describes the linear correlation between the collected images and the imaged object, where amplitudes at various spatial frequencies indicate how much information is translated into an image. The point at which the CTF crosses the x-axis is called the point resolution. Since every time the CTF crosses the x-axis the phase-information is inversed, direct interpretation of images acquired at a defocus is not possible and requires corrective modulation using computational methods. In addition,

the amplitude of spatial frequencies is attenuated by the envelope function of the CTF, which declines at higher frequencies. Since the rate of the decline not only correlates with the spatial coherence of the electron beam, but also with the defocus value of the image, under-focusing the image reduces the amount of high-resolution information. Therefore, in order to obtain high-resolution structural data, it is important to collect images at very low defocus values that still generate sufficient contrast.

After EM micrographs have been acquired and CTF corrected, the reconstruction of a three-dimensional volume first requires the computational averaging of thousands single particle images, which are then divided into various sub-classes that contain different orientations and conformational states of the particle. Subsequently, these 2D projections are assigned different directions for the angular reconstruction of the 3D object using algorithms based on the central section theorem.⁴¹⁸ The thus obtained initial map is then refined by an iterative procedure called projection matching, in which 2D projections of the 3D template are generated and matched against single particle EM images, before a new 3D reconstruction is computed and used for the next iteration step.⁴¹⁹ The quality of the refinement is monitored by the Fourier shell correlation (FSC) curve, which plots the signal to noise ratio (SNR) as a function of the spatial frequency that is the inverse of the resolution of the map.⁴²⁰ The FSC curve is generated by computing correlation coefficients within resolution shells extracted from Fourier transforms of 3D volumes, that were generated from each of two random halves of the dataset. The resolution of the refined structure is then chosen as the spatial frequency at which either the FSC curve is equal to 0.5, or the SNR is 1.0 (which corresponds to an FSC curve value of 0.33).

Depending on the final resolution of the 3D volume, the EM density map can be validated by different means. For high-resolution structures ($<4 \text{ \AA}$), validation of the EM map is straightforward, since secondary structure details are clearly discernable and can be traced along the polypeptide backbone. Furthermore, this enables precise docking of

X-ray crystal structures into the obtained density map, as well as the definition of subclasses that exist in different conformational states. Intermediate-resolution maps (4-10 Å) primarily reveal the relative orientation of various modules and may resolve secondary structure elements. Therefore, at this resolution range, it is still feasible to fit X-ray crystal structures and define different conformational states. Low-resolution maps (>10 Å), on the other hand, only reveal the overall shape of the particle. Thus, construction of a molecular model by fitting X-ray crystal structures into the density map is highly ambiguous and should not be performed for verification purposes. Instead, 3D volumes obtained at a low resolution are usually validated by tilt-pair analysis, in which pairs of images (tilt-pairs) are acquired with and without a small angle tilt of the specimen. Subsequently, particles from the tilt-pair images are projection-matched to the 3D volume and the tilt angle is calculated. If the calculated value matches the tilt angle at which the images were collected, the EM map is considered to be valid.

At high-resolution, EM surface maps can be interpreted in a similar fashion to X-ray crystallographic maps to generate reliable models at an atomic scale. In addition, available information on local resolution and variation can facilitate the accurate interpretation of structural information and prevent over-interpretation. Finally, the ability to generate subclasses of different conformational states makes EM highly suitable for the structural study of protein dynamics. At lower resolution, the utility of EM maps is significantly reduced. Nevertheless, unique docking of known structures can offer valuable information, such as the architectural arrangement of the particle and the location of functional modules. Low-resolution EM maps can reveal the overall shape of the particle, which can then be used to design experiments that can further push the resolution of the obtained EM structure.

4.2) 25 Å Structure of the CST Complex following Negative Stain EM

Since the opacity of a vitrified biological sample to electrons is very similar to that of the surrounding environment, cryo-EM of such specimen usually results in low contrast images. Therefore, cryo-EM is generally more suitable for the analysis of larger proteins and protein assemblies that scatter electrons more strongly. Alternatively, high contrast can be achieved by negative staining with heavy metal salt crystals, enabling the visualization of comparatively small macromolecules. However, the detection limit of negative stain EM is still around 150-200 kDa, depending on the overall shape of the analyzed protein (globular or elongated) and the quality of the stain. In addition, during alignment and 2D classification of images, the accuracy of the reconstruction process decreases with the size of the single particles. Therefore, it was advisable to perform EM using the CST complex containing the full-length Ctc1 construct (C1ΔST), which has a combined molecular weight of 190.9 kDa (as opposed to C3ΔST, which has a molecular weight of 160.8 kDa).

In order to visualize the overall shape and identify different conformational states, the CST complex was first analyzed by negative stain EM. Screening of EM grids revealed a highly heterogeneous protein sample that contained various sizes of small particles as well as larger aggregates. This was surprising, since, as shown in Figure 3-17, C1ΔST elutes off the gel filtration column as a single peak at the appropriate volume and is obtained in very high purity, indicating that both the heterogeneity of the sample, as well as the appearance of protein aggregates occurred during the course of EM grid preparation.

Since the protein sample that was subjected to negative stain EM did not contain any low molecular weight contaminants, the emergence of small particles on acquired images was attributed to the CST complex falling apart during the preparation of negative stain EM grids. Presumably, this was due to the low pH of the 2% uranyl acetate solution (pH 4.2) disrupting electrostatic protein-protein interactions. In order to prevent the

complex from disassembling, purified material was cross-linked with glutaraldehyde at protein concentrations used for EM grid preparation. These concentrations were very low (0.005 – 0.02 mg/mL), thus favoring intramolecular cross-linking events over intermolecular ones. When analyzed by EM, the homogeneity of the generated sample was significantly improved, with the majority of particles being well-distributed and around 5x5 to 5x15 nm in size.

Figure 4-1: Negative stain EM micrographs before and after optimization of EM grid preparation. Micrographs **a-c** illustrate initially encountered issues like aggregation, positive staining and particle heterogeneity. After optimization of grid EM preparation (see text), micrographs with very high homogeneity, no aggregation and good negative staining were obtained (**d-f**).

The aggregation problem, on the other hand, was found to arise from the concentration step after pooling gel filtration elution fractions. In order to circumvent this issue, protein samples intended to be used for negative stain EM were collected at the very apex of their corresponding peak. Immediately after, the sample was diluted to working conditions, cross-linked and applied on the EM grid. This optimization greatly reduced the number of observed aggregates in micrographs while also improving overall quality of the obtained protein. A comparison of micrographs before and after optimization of grid preparation is illustrated in Figure 4-1.

For the reconstruction of the three-dimensional volume, more than 100 micrographs were acquired and a total of 24.400 particles were used to generate 50 2D class averages in 25 iteration steps, at the end of which estimated accuracy angles and offsets remained unchanged. From these averages, 36 images comprising a total of 22.061 particles were selected for 3D reconstruction of the complex (Figure 4-2).

Figure 4-2: Picked particles and gallery of 2D class averages of cross-linked CST complex. **a)** Field view of CST particles. **b)** 10 out of the 36 averages, to which 22.061 particles contributed. The side views of the complex are elongated and roughly J-shaped, whereas top and bottom views appear globular and compact in shape.

As previously mentioned in the introduction, the structural similarity between CST and RPA is likely to reflect parallels in their molecular mechanisms, such as single-stranded DNA binding, and architectural organization, such as globular and elongated

a)

b)

Figure 4-3: Surface representations of CST 3D maps. 3D reconstruction was scheduled using three (a), two (not shown) and one (a) class/-es. **a)** The distribution of particles among the classes is annotated underneath each volume is roughly equivalent. There are no significant configurational differences between individual classes, implying that the particles comprise only one class. **b)** The illustrated 3D structure of the CST complex, which according to Relion 2 software has a resolution of around 25 Å, is elongated in shape. Rotations along the x-axis display top and bottom views.

conformations.³⁵⁸ Thus, the possibility that the collected particles comprised multiple classes of CST that differed in their configuration needed to be taken into consideration. However, when the 3D reconstruction was first scheduled to produce three classes, the number of particles that were distributed between these were approximately equivalent. In addition, the resulting volumes were comparable in shape and did not display significant configurational differences (Figure 4-3). Hence, it was reasonable to assume that the collected particles either comprised less than three classes, or that the configurational differences were so small that they were virtually indistinguishable in the resolution of the final 3D volume. A scheduling of two classes yielded very similar results. Consequently, all particles were unified to reconstruct a single class of a CST 3D volume (Figure 4-3). Following refinement, the resolution of the structure was estimated to be at 25 Å based on an FSC curve value of 0.5 (Figure 4-4).

Figure 4-4: Fourier Shell Correlation (FSC) plot for the refined EM density map.

The low resolution at which the CST volume was obtained imposed significant limitations on the construction of its molecular model using available X-ray crystal structures of Stn1 and Ten1,²⁸⁰ since they were, compared to the full complex, too small to be docked without ambiguity ($MW_{\text{total}} = 65.8$ kDa). As a consequence, Stn1-Ten1 could be placed in multiple locations and orientations within the obtained surface map.

However, several restrictions and guidelines were of assistance in evaluating the plausibility of the fit. Since Stn1 mediates interactions between Ctc1 and Ten1, it had to be situated in a way that bridges Ten1 and the rest of the 3D volume, whereas Ten1 needed to occupy a peripheral region of the map. Furthermore, based on the structure of the RPA trimerization core, in which C-terminal α -helices of RPA70, RPA32 and RPA14 OB-folds are aligned in parallel to form a tri-helical bundle,³⁰⁹ orientations in which C-terminal α -helices of Stn1-Ten1 faced towards the rest of the volume were more tenable.

In accordance with these guidelines, two fits in which Stn1-Ten1 were placed within the top and bottom lobes of the elongated 3D volume, were of particular interest (Figure 4-5). However, when the X-ray crystal structure of the RPA trimerization core was superimposed on top of Stn1-Ten1 within these models, in both cases the RPA70 OB-fold protruded out of the surface map. Therefore, it is very likely that both fits were incorrect, since there is no space for interactions between Ctc1 and Stn1-Ten1 OB-folds to take place.

In summary, the overall shape of the CST complex was determined by negative stain EM. However, due to the low resolution at which the 3D volume was generated, molecular models using available X-ray crystal structures were highly ambiguous. For future experiments, the modular organisation of the CST complex can be defined by introducing large globular tags on either termini of Ctc1, Stn1 or Ten1. Subsequently, through the identification of additional densities, such tags can help pinpoint their location in the surface map. The purification protocol of CST offers an easy solution for such an approach. Omitting TEV-cleavage, the CST complex can be obtained with an N-terminal GST-tag on Ctc1 and an N-terminal His₆-tag on Stn1 (Figure 3-15). Whereas the GST-tag is likely to be sufficient to generate additional density in the EM map, the location of Stn1's N-terminus can be highlighted by the addition of a His₆-tag specific antibody to the purified sample prior to EM grid preparation.

It would also be interesting to also analyze CST in complex with DNA by EM, since this would not only reveal potential conformational changes upon DNA binding, but also pinpoint the Ctc1 N-terminal OB-folds. However, it has to be taken into account that cross-linking and negative staining can potentially disrupt protein-DNA interactions due to alterations in the molecular composition of the DNA binding site.

Figure 4-5: Fitting of Stn1-Ten1 OB-folds into CST surface map. a and b) Depiction of Stn1-Ten1 fits into top and bottom lobes of the CST volume obtained through negative stain EM at a resolution of 23.5 Å. The right panel illustrates a magnified view of the respective fits, including the superimposition of RPA trimerization core onto Stn1-Ten1. In both instances, RPA70 falls outside of the surface map. Red: Stn1, Green: Ten1, Orange-red: RPA32, Olive-green: RPA14, Blue: RPA70

The next step in elucidating the structure of CST and obtaining high-resolution maps involves the performance of cryo-EM. Potentially, the elongated conformation and small molecular weight of CST may complicate the detection of particles by generating low contrast images. Adjustable settings that can improve contrast during image acquisition include working at appropriate defocus, using objective apertures and increasing the electron beam intensity. Alternatively, this difficulty can be overcome by the use of the developing phase plate technology, which produces a cosine-type CTF with a widened region of information transfer at low frequencies and thereby increases contrast.^{421,422}

Chapter 5: Structural Investigations on TCAB1

5.1) Recombinant Expression of TCAB1 in *E.coli*

5.1.1) Initial TCAB1 construct design and expression results

TCAB1 is a 59.3 kDa protein that comprises 548 amino acid residues with a theoretical pI of 4.46. According to initial sequence analysis using Eukaryotic Linear Motif (ELM) resource,⁴¹³ TCAB1 WD40 domain is composed of five WD40 repeats that are flanked by highly disordered regions on both N- and C-terminus (Figure 5-1). It is important to note that prediction of WD40 repeats is significantly hampered by their functional diversity, low levels of sequence conservation and variation in the length of connecting loops between β -strands (Figure 1-16 a). Consequently, it is very common for sequence-based classification methods to underestimate the absolute number of WD40 repeats, which is likely to be the case for TCAB1, as well. In fact, the presence of characteristic WD40 sequence elements within the large gap between the first two annotated WD40 repeats is a strong indication of a sixth WD40 repeat that has escaped identification (Figure 5-2).

Figure 5-1: TCAB1 sequence analysis using ELM resource. ELM resource identifies eukaryotic linear motifs and compiles relevant information from computational resources that provide algorithms for domain identification (SMART and Pfam),^{414,415} globular domain tendency plots (Globplot),⁴¹⁶ and disorder prediction (IUPred).⁴¹⁷ TCAB1 is annotated to contain five WD40 repeats. Flanking regions are predicted to be highly disordered. Positions of putative phosphorylation sites and telomere dysfunction associated mutations are indicated with red and yellow arrow heads, respectively, and specified in the bottom panel.

Sequence alignment of TCAB1 with orthologues from different organisms show a high degree of conservation within the WD40 domain (Figure 5-2). N- and C-terminal regions, on the other hand, are markedly divergent, but appear to retain their particular compositional biases for being rich in proline and glycine residues, respectively. Secondary structure prediction places four β -strands into each identified WD40 repeat as well as the aforementioned possible sixth WD40 repeat. It furthermore is in agreement with IUPred disorder prediction, in that little to no secondary structure elements were annotated within the N-terminal region.

Figure 5-2: Sequence alignment of TCAB1 with orthologues from different organisms. Sequence alignment was carried out using Jalview and Clustal W.^{423,424} Secondary structure elements are predicted using Jpred and displayed underneath each row.⁴²⁵ WD40 repeats as annotated by ELM resource are depicted in blue boxes. The box with dashed black borders indicates the location of a sixth WD40 repeat that is very likely to be situated between the first annotated two.

Due to its biological significance and high likelihood of constituting a globular domain, initial TCAB1 constructs were designed to minimally span the WD40 domain, whereas N- and C- terminal borders were designated according to various parameters such as disorder propensity, secondary structure prediction, sequence conservation and amino acid type. TCAB1 constructs were then submitted to the Protein Production Platform (PPP) at SBS/NTU for cloning and expression screening. According to PPP annotations, all constructs except TCAB1 56-548 were well-expressed with respective bands being clearly visible after SDS-PAGE and Coomassie staining (Figure 5-3).

Figure 5-3: Total expression results of TCAB1 constructs submitted to PPP. DOM (domain): annotated domains correspond to predicted WD40 repeats. LC (low complexity): disordered region as predicted by IUPred. OT (other): annotations correspond to N-terminal proline rich patch, putative phosphorylation sites and disease associated mutations. **a)** Total expression results as can be conferred from whole cell lysates. **b)** Legend for PPP annotations. **c)** Coomassie stained SDS polyacrylamide gel after electrophoresis of whole cell lysates of various constructs. Red arrows highlight main protein bands, whereas red boxes feature main degradation products.

Careful inspection of the SDS-PAGE shows a recurring pattern when comparing lower molecular weight bands of constructs that begin with amino acid residues T144 and Q167 (highlighted with red boxes) to the background of other lanes (Figure 5-3 c). This is suggestive of degradation products that are construct specific. In fact, the size difference between these probable degradation products is small and may match that between their corresponding N-terminal amino acid residues T144 and Q167 (2.6 kDa). Together, these observations probably indicate C-terminal degradation which progresses deep into the WD40 domain. Interestingly, these bands do not appear for constructs that start at amino acid residues M1 or D56, suggesting a potential role of the N-terminal region in WD40 domain stability.

Despite positive total expression, most protein constructs turned out to be insoluble. Those constructs that were detectable in the supernatant following pelleting of the lysate, on the other hand, could not be purified (Figure 5-4). The emergence of distinct degradation patterns with ladder-like regularity, which previously were not visible in whole cell lysates, is suggestive of step-wise deletions of single WD40 repeats during protein purification (Figure 5-4 c). Interestingly, the purification of protein constructs with high level of total expression (144-548, 167-548, 144-450, 167-450, 144-501, 167-501) failed completely, whereas purification of protein constructs with low level of total expression (56-548, 1-450, 56-450, 1-511, 56-511) were successful in part. This observation further supports the notion that N-terminal residues might play a role in TCAB1 stability as well as solubility. In the case of full-length TCAB1, all protein was degraded over the course of the purification steps. In the end, no purified protein could be obtained from PPP.

Figure 5-4: Soluble expression results of TCAB1 constructs submitted to PPP. DOM (domain): annotated domains correspond to predicted WD40 repeats. LC (low complexity): disordered region as predicted by IUPred. OT (other): annotations correspond to N-terminal proline rich patch, putative phosphorylation sites and disease associated mutations. **a)** Soluble expression results as can be conferred after NiNTA purification of the soluble fraction following cell lysis. **b)** Legend for PPP annotations. **c)** Coomassie stained SDS polyacrylamide gel after electrophoresis of NiNTA purification peak fractions of various constructs. Red arrows highlight main protein bands, whereas red boxes feature main degradation products.

5.1.2) TCAB1 forms soluble aggregates that hamper purification after expression in *E.coli*

PPP uses a standardized protocol for the expression screen of all protein entries, which may not be appropriate for TCAB1 expression. Therefore, in order to improve the expression of TCAB1, alternative methods of TCAB1 expression and purification were investigated. After receiving PPP cloned plasmids encoding for various His₆-tagged TCAB1 constructs, small-scale test expression of full-length TCAB1 in Rosetta(DE3)pLysS, a cell line that supplies tRNAs for rare codons and thereby lifts limitations based on *E.coli* codon usage, at 37°C revealed little to no difference in total protein levels between auto-induction and induction at various IPTG concentrations (Figure 5-5). Subsequent NiNTA batch purification confirmed the very low efficiency of TCAB1 purification that was generally observed by PPP. In contrast, however, using this expression protocol, full-length TCAB1 did not degrade completely and could be obtained in small yields.

Figure 5-5: Initial small-scale expression test of full-length TCAB1. After electrophoresis of His₆-tagged TCAB1 1-548 whole cell lysate (WCL) and NiNTA elution samples, SDS polyacrylamide gel was stained with Coomassie Blue. Expression of TCAB1 1-548 induced at various IPTG concentrations or through auto-induction yields comparable amounts of protein in WCLs and after NiNTA batch purification. Red boxes annotate bands that correspond to TCAB1 1-548, as confirmed by mass spectroscopy and time course experiments (Figure 5-6).

Large-scale expression (Figure 5-6) later identified several problems that impede TCAB1 protein preparation from *E.coli*: most of TCAB1 protein accumulates in the

insoluble fraction of the cell lysate. The small fraction of TCAB1 that is soluble does not bind efficiently to NiNTA and elutes with the flow-through (Figure 5-6 a). Finally, TCAB1 elutes in the void volume of gel filtration columns (Figure 5-6 c). From this analysis, it was concluded that TCAB1 forms soluble aggregates which are inadequate for structural or biochemical studies. In conclusion, these results indicate a strong tendency of TCAB1 to form both soluble as well as insoluble aggregates that diminish protein yields and greatly challenge protein preparation. Interestingly, TEV mediated cleavage of the N-terminal TCAB1 His₆-tag following NiNTA purification proceeds relatively well (Figure 5-6 b), suggesting that for the majority of TCAB1 which did not bind to the NiNTA matrix, the affinity tag is inaccessible, presumably due to the formation of soluble aggregates, whereas for NiNTA purified fractions this is not the case.

Purification of TCAB1 1-501 and 1-450 constructs were similarly disappointing: most of the protein was lost during preparation, while the remainder formed soluble aggregates. Remarkably, purification of TCAB1 1-548 and 1-501 constructs produces a degradation band that migrates at 60 kDa, which roughly coincides with the electrophoretic shift of TCAB1 1-450. Given that K450 is in close proximity to the outermost WD40 repeat predicted by ELM resource and SMART (Figure 5-1), it may constitute the boundary of a global WD40 domain.

Figure 5-6: Large-scale expression and purification of full-length TCAB1. FT: Flow-through, Wn: Wash fraction n, En: Elution fraction n, Fn: Fraction n. **a)** SDS-PAGE and Coomassie stain of protein samples taken from expression time course and NiNTA purification. A lot of protein elutes in the flow-through and the first wash step. **b)** SDS-PAGE and Coomassie stain of protein samples taken from TEV-cleavage time course and reverse NiNTA purification. TEV-cleavage was performed at RT using a TEV to protein ratio of 1:20 and was mostly complete after 1 h. **c)** Superdex 200 pg 16/60 chromatogram of TCAB1 1-548 after TEV-cleavage and reverse NiNTA purification. Protein elutes in the void volume. **d)** SDS-PAGE and Coomassie stain of size exclusion fraction F13.

Figure 5-7: Common degradation product of TCAB1 1-548, 1-501 and 1-450. En: Elution fraction n. Protein samples were taken from NiNTA elution fractions of indicated His₆-tagged TCAB1 constructs and stained with Coomassie after SDS-PAGE. Red arrows highlight common degradation product in each gel.

In order to increase the solubility of TCAB1 and efficiency of its purification as well as to prevent the formation of soluble aggregates, multiple trials were performed with a focus on TCAB1 1-450, which appeared to represent a stable construct. The His₆-tag was swapped for a GST-tag, which besides serving as an affinity tag during protein purification was intended to increase TCAB1 solubility by acting as a prodomain. Proteins were expressed at low temperatures in order to slow folding kinetics and prevent possible denaturation. Finally, proteins were subjected to ammonium sulfate precipitation and on-column refolding. Unfortunately, none of these trials provided any noteworthy improvement or insight into the purification of TCAB1.

5.1.3) Pairwise comparison of HMMs reveals seventh putative WD70 repeat

Having failed to obtain soluble constructs, it became necessary to reinvestigate the primary sequence, secondary structure elements and probable domain architecture of TCAB1. Therefore, the HHpred method was utilized and compared with the analysis from aforementioned sequence based methods. HHpred is an online resource that detects remote protein homology by pairwise comparison of HMMs and uses available 3D information as a template to model tertiary structure.⁴²⁶ Due to significant sequential and functional diversity that exists among WD40 domain proteins, HMM based comparative searches for homologues using position specific gap penalties offers a more sensitive method for the identification of WD40 repeats than pairwise alignment of sequence profiles. In addition, since WD40 domain proteins share structural similarity, modeling TCAB1 to remote homologues with high scores but differing cellular functions represents a sensible strategy to derive useful information concerning domain architecture. Eventually, this analysis identified the likely presence of a seventh WD40 repeat.

TCAB1 amino acid residues 140-510 were submitted to HHpred and modeled according to tertiary structures of remote homologues, including Platelet-activating factor acetyl hydrolase (PAF-AH) α -subunit (1VYH_C), U4/U6 tri-snRNP subunit Prp4 (3JCR_L) and Protein Phosphatase 2A (PP2a) subunit B55 (3DW8_B) (Figure 5-8). TCAB1 was predicted to form a seven bladed β -propeller with the last WD40 repeat closely situated at the C-terminus of the sequence (amino acid residues 454-489). This finding makes structural sense and is compatible with sequence alignments and secondary structure predictions, since the predicted WD40 repeat covers four putative β -strands and accounts for a highly-conserved region located on the C-terminus of TCAB1 (Figure 5-2).

Figure 5-8: TCAB1 140-510 tertiary structure as modeled by HHpred using templates 1VYH_C, 3JCR_L and 3DW8_B and comparison with 1VYH_C solved structure. **a)** top and side views of TCAB1 140-510. **b)** top and side views of 1VYH_C. Encircled letters A-D annotate β -strands of N-terminal WD40 repeat. Cyan and green arrows point to large extensions between indicated WD40 repeats.

Figure 5-8 illustrates a comparison between top and side views of the predicted structure of TCAB1 140-510 and one of the three known protein structures that were used as templates for 3D modelling (1VYH_C). Importantly, neither HHpred nor Jpred predicted the N-terminal region of the first WD40 repeat to comprise a secondary structure element. However, both resources agree in the prediction of β -strands A, B and C which fall into the first WD40 repeat annotated by SMART and Pfam. It is therefore reasonable to assume

the presence of a β -strand D within this region. In addition, it is noteworthy that WD40 repeats 1 and 2 as well as 3 and 4 are connected by prominently long loops, which are likely to be susceptible to proteolytic cleavage and may be a factor in previously observed ladder-like degradation patterns. This analysis guided the design of new TCAB1 constructs that hopefully would fold properly and hence be soluble.

Figure 5-9: Sequence alignment of TCAB1 with orthologues from different organisms and revised WD40 repeats. Sequence alignment was carried out with Jalview and Clustal W.^{423,424} Secondary structure elements are predicted using Jpred and displayed underneath each row.⁴²⁵ WD40 repeats are derived from HHpred model (Figure 5-8) and annotated with blue boxes.

5.1.4) Screening of various constructs and conditions does not solve aggregation of TCAB1

A number of new GST-tagged TCAB1 constructs (12) were designed according to the revised domain architecture and expressed in Rosetta(DE3)pLysS cells at 37°C (Figure 5-10). Preparation of these protein constructs consistently reproduced the same observations: during GST affinity purification, only a small portion of GST-tagged TCAB1 constructs binds to the GSTrap HP column, while a large part remains in the flow-through. In addition, when subjected to gel filtration, GSTrap HP purified proteins elute in the void volume, suggesting the formation of soluble aggregates (Figure 5-11). Notably, protein constructs starting with amino acid residue S158 do not result in any expressed protein.

Figure 5-10: TCAB1 domain architecture according to ELM resource and HHpred model, and construct design. WD40 repeats are annotated with blue boxes. They vary in position between ELM resource determined domain architecture and that derived from HHpred model of TCAB1 and conserved primary sequence elements. Light blue boxes indicate WD40 repeats that ELM resource failed to identify.

Figure 5-11: Two examples of GST-tagged TCAB1 construct expression and purification. In: Input, FT: Flow-through, Fn: Fraction n. **a-c)** Expression and purification of GST-tagged TCAB1 1-548. **d-f)** Expression and purification of GST-tagged TCAB1 87-495. **a and d)** SDS-PAGE and Coomassie stain of protein samples taken from expression time course, GSTrap and Superdex 200 pg purification. Considerable amounts of GST-TCAB1 does not bind to GSTrap and remains in the flow-through. Red boxes annotate bands that correspond to GST-TCAB1 1-548 and GST-TCAB1 87-495, respectively. **b and e)** GSTrap HP chromatogram of TCAB1 1-548 (**b**) and TCAB1 87-495 (**c**). **c and f)** Superdex 200 pg 16/60 chromatogram of TCAB1 1-548 (**c**) and TCAB1 87-495 (**f**) after GSTrap purification. Proteins elute in the void volume.

These results replicate and reinforce initial findings from His₆-tagged protein expression and purification trials: protein constructs that begin at close proximity to the WD40 domain N-terminal boundary are unstable and fail in purification, whereas the rest is prone to form soluble aggregates, which hamper affinity purification, diminish protein yield and are unsuitable for use in structural studies.

A comparison between various constructs and their respective protein yield following purification reveals that certain N- and C-terminal boundaries give rise to higher protein levels whereas others diminish protein yield significantly (Figure 5-12). For instance, purification of TCAB1 constructs starting at amino acid residue S87 were significantly more successful. On the C-terminus, constructs terminating in I548 consistently yielded largest protein amounts. It is important to mention that for screening purposes these purifications were performed in parallel, but only once. Hence, no error bars are shown. Despite this caveat, however, there is a clear indication that certain boundaries produce much higher protein yields than others.

Figure 5-12: Protein yields from 1 L culture of various GST-tagged TCAB1 constructs following GSTrap affinity purification and gel filtration. The x-axis indicates N-terminal boundaries of protein constructs, whereas the C-terminal boundary is specified by color code according to the legend.

In general, protein yields can indicate a direct correlation with expression levels, as well as the tendency of a protein to remain in the soluble fraction. Since protein stability can have an effect on both, interpretations of these results were utilized to formulate a third round of TCAB1 constructs with the anticipation that high protein stability may help circumvent the formation of soluble aggregates (Figure 5-13 a). N-terminal boundaries were chosen around amino acid residue S87 based on primary sequence and secondary structure prediction, whereas C-terminal boundaries were set at S526 and I548. Unfortunately, expression and purification of these new construct did not improve TCAB1 production and failed to solve the problem of soluble aggregates (Figure 5-13 b and c).

a)

Constructs:

b)

GST-TCAB1(76-548) superdex 200 pg 16/60

GST-TCAB1(79-548) superdex 200 pg 16/60

GST-TCAB1(84-548) superdex 200 pg 16/60

GST-TCAB1(92-548) superdex 200 pg 16/60

c)

Figure 5-13: Third round of TCAB1 construct design and purification results. **a)** TCAB1 domain architecture is based on HHpred model. Construct design was informed by protein yields of previous purifications possibly being an indication for protein stability. **b and c)** Superdex 200 pg 16/60 chromatogram following GSTrap purification of GST-tagged TCAB1 constructs ending at I548 (**b**) and S526 (**c**), together with SDS-PAGE and Coomassie stain of protein samples taken from peak fraction. All protein constructs elute in the void volume. Fn: Fraction n.

After it became evident that expression of alternate TCAB1 constructs was not a feasible strategy to stabilize native protein structure and suppress aggregation events, the effects of additives during the purification of TCAB1 were tested. Selection of additives was broadly based on sufficiently distinct chemical classes, chemical properties and effects on protein folding. As such, trials were performed using urea (denaturant), ammonium sulfate (antichaotropic salt), glucose (saccharide) and tween 20 (non-ionic surfactant).

Unfortunately, none of these additives prevented the formation of soluble aggregates, which is evident by the elution of GSTrap purified proteins in the void volume of subsequent gel filtration steps. Negative results are not depicted here, since they are identical to previously displayed purification results.

Finally, TCAB1 orthologues from mouse and zebrafish, which share an 82.2% and 56.5% sequence similarity and a 77.1% and 43.4% sequence identity with human TCAB1, respectively, were cloned and expressed with the expectation that these orthologues may perform differently in protein purification. As a general trend, the compositional and structural complexity and flexibility of homologous proteins and their assemblies increases with the level of complexity of their corresponding organisms. In addition, due to their adaptation to distinct conditions and molecular environments, including temperature, homologous proteins may vary in stability and behave differently in purification and structural studies. When submitted to ELM resource, the overall domain architecture, globular domain tendency and disorder prediction are very similar for human, mouse and zebrafish TCAB1. Primary sequence alignment reveals a very close conservation of the WD40 domain and to a lesser extent the C-terminal tail, whereas the N-terminal region, which is predicted to be highly disordered, displays significant variation (Figure 5-2). Unfortunately, expression and purification of mouse and zebrafish TCAB1 only replicated previously described results and did not provide a solution to the aggregation problem (Figure 5-14).

Figure 5-14: Mouse and Zebrafish TCAB1 sequence analysis and purification results. ELM resource was used to plot globular domain tendency (GlobPlot) and protein disorder (IUPred) of mouse **(a)** and zebrafish TCAB1 **(b)**. WD40 repeats were derived from primary sequence alignment and manual comparison with putative WD40 repeats in human TCAB1. Underneath each chart, Superdex 200 pg 16/60 chromatogram following GSTrap purification of respective GST-tagged TCAB1 constructs together with SDS-PAGE and Coomassie stain of protein samples taken from peak fraction are depicted. Fn: Fraction n.

5.1.5) TCAB1 requires the chaperonin TRiC for proper folding

During the course of my work, in December 2014, Freund *et. al* published a study in which they discovered that folding of endogenous TCAB1 requires the participation of the chaperonin TCP-1 ring complex (TRiC).³⁹² Using a genome wide siRNA screen based on RNA fluorescence *in situ* hybridization (FISH), they identified TRiC as a regulator of telomerase trafficking. Depletion of TRiC leads to mislocalization of telomerase RNA as well as scaRNAs, and negatively affects telomerase recruitment to telomeres and *in vivo* activity. Importantly, depletion of TRiC does not impact TER stability or intracellular accumulation. The observed phenotype is very similar to that following depletion of TCAB1.^{193,207} In fact, knock-down of any TRiC subunit, but not of molecular chaperones HSPA8 and HSPA5, diminishes overall TCAB1 protein levels, revealing a specific requirement of TRiC for TCAB1 stability. Subsequent immunoprecipitation experiments of various TCAB1 N- and C-terminal deletion mutants demonstrated interactions between TRiC and TCAB1 to be mediated by the WD40 domain. They further showed that in HeLa S3 cells, both WD40 domain and C-terminal extension are required for TCAB1 stability and intracellular accumulation.³⁹²

TRiC is a eukaryotic group II chaperonin composed of eight paralogous subunits that arrange into two stacked rings.^{427,428} Each ring forms a folding chamber which encapsulates substrate proteins and promotes folding by isolating them inside a chemically controlled environment away from the crowded cytoplasm. Unlike prokaryotic group I chaperonins, of which GroEL/ES from *E.coli* is a prominent member, TRiC does not require a co-factor to form a “lid”.⁴²⁹ Instead, helical protrusions in the apical domain of each subunit form an aperture encircling the folding chamber that opens and closes in an ATP-dependent manner.^{427,430} Owing to its hetero-oligomeric nature, each folding chamber of TRiC presents protein substrates with a chemically and functionally asymmetric environment. In contrast, GroEL consists of identical subunits that arrange into heptameric rings with chemically symmetrical cavities. Approximately 10% of all

cytosolic proteins require TRiC to mediate folding.⁴³¹ And while there are no clear identifiers that confer TRiC dependency, some motifs are more frequently found to interact with TRiC, including the WD40 domain.^{432,433}

The finding that TRiC plays an essential role in TCAB1 folding and stability explains why recombinant expression and purification of various TCAB1 constructs from *E.coli* has consistently resulted in the formation of both soluble and insoluble aggregates. TRiC shares many similarities with GroEL in overall structure, subunit domain architecture and mechanistic principles. So to improve folding of TCAB1, it was co-expressed with GroEL/ES in BL21(DE3)pLysS cells. However, this approach remained insufficient to ensure proper folding of TCAB1 (Figure 5-15). In conclusion, except for co-expression of TCAB1 with all eight subunits of TRiC, these findings make it clear that expression in *E.coli* is inadequate for obtaining properly folded, and hence soluble, TCAB1.

Figure 5-15: Purification results of TCAB1 co-expression with GroEL/ES. Superdex 200 pg 16/60 chromatogram following GSTrap purification of respective GST-tagged TCAB1 constructs together with SDS-PAGE and Coomassie stain of protein samples taken from peak fraction are depicted. Fn: Fraction n.

Given that TRiC is an absolute requirement for the folding of a subset of proteins that are essential for cellular viability, it is unsurprising that its subunits are highly conserved among eukaryotes. Table 5-1 lists sequence identities and similarities between human and *D.melanogaster* TRiC subunits, which average at 69.3% and 81.5%, respectively. On account of the close evolutionary distance between *D.melanogaster* and *S.frugiperda*, it is reasonable to assume that TRiC is similarly well conserved between humans and Sf cell lines that are commonly used for protein expression, making it an interesting option to explore.

Table 5-1: Sequence identity and similarity between various human and *D.melanogaster* TRiC subunits.

	TCP-1 α	TCP-1 β	TCP-1 γ	TCP-1 δ	TCP-1 ϵ	TCP-1 ζ	TCP-1 η	TCP-1 θ
identity	72.8%	70.1%	68.6%	67.6%	70.3%	68.9%	74.9%	61.2%
similarity	82.8%	82.3%	79.6%	82.1%	81.0%	83.5%	83.1%	77.8%

While setting up an insect cell facility within our own lab, GST-tagged full-length TCAB1 was submitted to PPP, which also provides baculovirus (BV) expression services. Even though PPP expression and purification concluded negatively, we obtained an Sf9 cell pellet infected with His₆-GST-TCAB1 encoding BV. Subsequent protein preparation shows a clear difference to previous purification attempts from *E.coli*, in that the elution profile after size exclusion chromatography features multiple dominant peaks (Figure 5-16). Only a small portion of Sf9 purified His₆-GST-TCAB1 forms soluble aggregates that elute in the void volume, whereas the rest constitutes a prevalent peak that elutes shortly after. Sample analysis of the purification steps using SDS-PAGE and western blot reveals various well-defined C-terminal degradation products which extend all the way to the N-terminal His₆-GST-tag. This is reminiscent of previous protein preparations from *E.coli*, in which after SDS-PAGE distinct degradation patterns emerged, presumably due to proteolytic cleavage of connecting loops and consequent removal of single or multiple WD40 repeats. Following gel filtration, full-length TCAB1 containing fractions are mostly pure and elute in a single peak, suggesting that they purify as a homogenous species. Poor

resolution and overlap with peaks that contain degradation products remains a source of contamination. The degree of protein purity, however, presents a promising prospect for structural and biochemical studies.

Following these results, multiple TCAB1 constructs were screened for expression in HEK293T cells, since this cell line allows for a very quick evaluation of expression efficiency in higher eukaryotes. After their expression was confirmed, they were cloned into the Multibac system. In order to eliminate degradation products during the purification process, TCAB1 constructs were fused with an N-terminal GST-tag and a C-terminal His₆-tag. Viral stocks were created and amplified. Initial expression tests in Sf9 cells were all positive and produced comparatively strong bands after western blot (

Figure 5-17).

Figure 5-17: Screening and expression tests of new TCAB1 constructs in HEK293T and Sf9 cells. **a)** Representation of TCAB1 domain architecture and construct design. Constructs 79-548 and 79-526 failed in cloning. **b)** Anti-FLAG western

blot of TCAB1 constructs screened for expression in HEK293T cells following transient transfection. **c)** Anti-GST western blot of GST-TCAB1-His₆ constructs after expression test in Sf9 cells.

The readiness of viral stocks and feasibility of both expression and purification signify promising progress, allowing the project to be advanced to crystallization trials of TCAB1 or TCAB1 complexes after optimization of both expression and purification. Alternatively, sufficiently large assemblies (over 200 kDa) can be studied by negative stain and cryo-EM. Structural insights into TCAB1 complexes with telomerase and other telomere components can expand our understanding of telomerase trafficking, including its recruitment to telomeres, which is a prospect that is very worthwhile pursuing. The work presented in this chapter lays the foundation for further research that focuses on these questions.

Chapter 6: Discussion

The maintenance of telomere length is a crucial process that ensures the genomic integrity and viability of eukaryotic cells, and involves a large number of cellular factors participating in functions such as telomere capping, elongation and quality control. These processes are extensively regulated and necessarily interdependent. The work presented in this thesis focuses on two protein factors, TCAB1 and the CST complex.

TCAB1 is involved in telomerase trafficking between Cajal bodies (CBs) and telomeres. Its primary function is the mediation of protein-protein and protein-nucleic acid interactions, and is presumably the main mechanism by which TCAB1 guides its interaction partners to designated sub-cellular locations. However, in the absence of structural information on TCAB1 and TCAB1 assemblies, these inferences remain mostly speculative.

Establishing a strategy for the expression and purification of TCAB1 was greatly hampered by the fact that TCAB1 failed to be purified from *E.coli* cells without the formation of soluble aggregates. Following a conventional method, TCAB1 constructs were expressed as N-terminal GST-fusion proteins. Generally, N-terminal GST-tags often help to improve solubility and protein folding of their fusion partners. However, despite extensive screening of conditions and constructs, this approach did little to solve the aggregation problem of TCAB1. In total, three rounds of cloning and twenty-two constructs, including homologues from mouse and zebrafish, were tested under various conditions, before expression in *E.coli* was eventually abandoned. During the course of this work, two additional putative TCAB1 WD40 repeats were identified, characterizing TCAB1 as a WD40 protein with a total of seven repeats. The proper identification of putative protein domains is very important to inform the design of constructs for recombinant expression. At the same time, a newly published finding reported the requirement of the eukaryotic group II chaperonin TRiC for proper folding of TCAB1, and

identified disease associated mutations in TCAB1 to affect its association with TRiC. Since subunits of TRiC are highly conserved among higher eukaryotes, expression of TCAB1 was tested in Sf9 insect cells. Following expression in insect cells, it was possible to obtain TCAB1 as non-aggregated material, as evidence by its elution profile following gel filtration. The spontaneous change in TCAB1 purification behavior after switching to a different expression system verifies that TCAB1 requires a more complex expression machinery which is absent in *E.coli* but available in Sf9.

In conclusion, I was able to provide a solution for the problems encountered in TCAB1 expression as well as purification, and demonstrated a way towards a feasible purification strategy. Thus, it is now possible to obtain purified material in the foreseeable future, which can then be used for structural studies using X-ray crystallography. Since the main function of TCAB1 is the mediation of protein-protein interactions, it will be very interesting to test and characterize interactions between TCAB1 and other telomeric components, such as POT1-TPP1, hTER H/ACA domain and telomerase RNP, using *in vitro* biochemical studies. Larger assemblies of TCAB1 can then be analyzed using crystallographic methods, as well as negative stain and cryo-EM. These structural studies can provide a larger picture and substantial insight into the regulation and mechanism of telomere maintenance via telomerase activity. It is possible that TCAB1 is part of the telomerase holo-enzyme, serving as a factor that is involved in its maturation and cell-cycle dependent trafficking. It would then be interesting to identify stable intermediates starting from the bio-assembly of the telomerase holo-enzyme up until its cell-cycle dependent “activation” and recruitment to telomeres, and determine their compositional and structural properties. Consequently, this may reveal new avenues in cancer therapy by identifying novel drug targets for medication.

The second part of the work presented in this thesis focuses on the CST complex. The CST complex is the final player in telomere elongation. It inhibits telomerase activity by sequestering both its telomeric DNA substrate and binding partner (POT1-TPP1), and

initiates C-strand fill-in synthesis by recruiting Pol α /Primase to telomeres. The CST complex consists of Ctc1, Stn1 and Ten1. The structure of Stn1 and Ten1 have been elucidated by X-ray crystallography and strongly resemble that of RPA32 and RPA14 of the single stranded DNA binding RPA complex. The molecular structure of the largest component, Ctc1, however, remains unknown. Again, in order to elucidate its molecular mechanism and mode of action, it is imperative to obtain structural information on both CST alone and in complex with telomeric DNA and/or other interaction partners.

Expression of Ctc1 in *E.coli* resulted in the formation of soluble aggregates, as evidenced by gel filtration chromatograms. In order to circumvent the formation of soluble Ctc1 aggregates, sixteen constructs were tested under various conditions, including precipitation and refolding of aggregated material, as well as co-expression with the Stn1-Ten1 subcomplex. However, these attempts failed to provide non-aggregated material that is the sample requirement for structural and biochemical studies. Eventually, expression of Ctc1 in *E.coli* was abandoned in favor of expression in Sf9 insect cells. When examining Ctc1 constructs spanning the C-terminal OB-fold, it was shown that expression of Ctc1 alone in Sf9 cells lead to either no expression, or to the formation of soluble aggregates. Finally, in order to improve the stability of Ctc1 through protein-protein interactions, Ctc1 was co-expressed in insect cells with Stn1 and Ten1.

As anticipated, co-expression of Ctc1 with Stn1 and Ten1 in Sf9 cells had a stabilizing effect, greatly increasing Ctc1 expression levels and purification efficiency. However, purification of the CST complex still yielded a large amount of aggregated material. This impediment was solved by the introduction and modification of purification steps and conditions, including the optimization of salt and buffer conditions, and the introduction of detergent wash steps during the initial NiNTA purification step. The use of Heparin purification enabled the efficient separation of the dimeric Stn1-Ten1 subcomplex from the trimeric CST complex, while also removing nucleic acid contaminations through competition. Consequently, protein purity and homogeneity was significantly improved.

The described solution to the aggregation problem indicates that the primary source for the formation of soluble CST aggregates was the presence of contaminants, such as chaperones and other protein factors. Eventually, optimized purification strategies were established for a total of two CST constructs, namely C1 Δ ST and C3 Δ ST.

The failure to obtain C2 Δ ST and C4 Δ ST following the same general protocol nevertheless gave rise to some interesting observations and contemplations. C2 Δ ST, which comprises Ctc1 construct 203-1217 consistently displayed C-terminal degradation of Ctc1, as evidenced by anti-GST western blots and the failure to co-purify with His₆-tagged Δ 18Stn1. Intriguingly, this was not the case for Ctc1 constructs 1-1217 (Ctc1 C1) or 521-1217 (Ctc1 C3). This distinction in degradation behavior suggests that N-terminal residues may play a role in the stabilization of the C-terminus. This is an appealing assumption, since it also draws a parallel with the globular structure of the RPA_{8nt} complex, in which N-terminal OB folds are thought to contact C-terminal OB-folds through the formation of a hexa-helical bundle. However, this explanation is not sufficient to account for the observation that Ctc1 C3, which also lacks N-terminal OB-folds, is not only stable, but exhibits significantly higher expression levels than full-length Ctc1 C1. It would be interesting to revisit this phenomenon, once high-resolution structural data on CST is available.

In contrast to C1 Δ ST and C3 Δ ST, C4 Δ ST containing Ctc1 construct 828-1217 did not display enhanced binding affinity to Heparin. Interestingly, both C3 Δ ST and C4 Δ ST lack Ctc1 N-terminal single-stranded DNA binding OB-folds. The only discrepancy between these two Ctc1 constructs is the lack of amino acid residues 521-827 in Ctc1 C4. Thus, the observed difference in Heparin binding affinity suggests a significant contribution of these amino acid residues in Heparin binding, and consequently may also play an important role in DNA binding activity. This speculation is supported by the finding that disease associated mutation V665G negatively affects CST single-stranded DNA binding. Structural data on CST-DNA complexes are likely to provide further and more conclusive insights into

the role of these amino acid residues and molecular basis by which they affect DNA binding.

Crystallization trials of C1 Δ ST and C3 Δ ST did not produce any protein crystals, despite extensive screening of buffer conditions and protein concentrations. For future experiments, it is worthwhile testing different Ctc1 constructs, of which two promising candidates are provided in this thesis. Both constructs include N-terminal single-stranded DNA binding OB-fold domains, which is essential if DNA binding activity of the CST complex is to be examined. However, despite the relatively high expression levels based on western blot comparisons with other constructs, it will still be necessary to scale up expression volumes and further optimize purification strategies in order to obtain sufficient amounts of protein for X-ray crystallography.

Finally, the work presented in this thesis culminates in the elucidation of the low-resolution structure of C1 Δ ST using negative stain EM. It is important to point out that the main challenges of this project arise from the requirement to express CST in insect cell. The time-consuming process of generating viral vectors, low protein yields, and the necessity to optimize expression and purification protocols represent significant impediments for crystallographic studies. The reconstruction of a low-resolution electron density map following negative stain EM provides an important and necessary stepping stone for cryo-EM experiments. Therefore, for future structural studies, it would be more sensible to focus on EM methods instead of X-ray crystallography.

The reconstructed volume of C1 Δ ST exhibits a very low resolution, as is expected from negative stain EM, but gives a general sense of the shape and architecture of the complex. Surprisingly, 3D reconstruction of the CST complex only provides a single class of particles. Since CST is an RPA-like protein complex, it would have been reasonable to assume that like RPA, CST can assume various conformations that comprise globular and elongated states. It is important to remember, however, that prior to grid preparation, the

CST complex was fixed by glutaraldehyde cross-linking. Thus, the single class that was obtained may not be an accurate representation of CST conformational flexibility, but that of the favored configuration that emerges following the cross-linking reaction. The reconstructed volume of CST is elongated in shape with a height of 155 Å and a width of 55 Å at its “belt”. The globular lobes present at both extremes are noticeably different in size, with the “bottom” lobe being approximately twice as large and featuring armrest-like protrusions (Figure 4-3).

The construction of a molecular model for the CST complex gave rise to considerable ambiguity when trying to fit the small-sized Stn1 and Ten1 OB-fold crystal structures into the low-resolution CST volume (Figure 4-5). In order to appraise the plausibility of the two Stn1-Ten1 fits that were considered to be tenable, RPA32-RPA14 were superimposed onto Stn1-Ten1. Then, the orientation with which RPA70 formed the RPA trimerization core was viewed (Figure 4-5). In both molecular models, RPA70 protruded out of the surface map, suggesting that the trimerization interface in the obtained 3D volume of CST is either different from that of RPA, or that the proposed fits are inaccurate. Given the high ambiguity of the fits, the latter is more likely.

In conclusion, the current resolution of the obtained CST volume is not yet sufficient to build a definite molecular model of the CST complex. In order to pinpoint the location of various subunits and domains within the surface map, there are multiple strategies that should be considered for future work. One such strategy involves performing negative stain EM of CST complexes that contain additional density at designated sites, i.e. a large protein tag at the N-terminus of Ctc1. Most feasibly, the purification strategy can be adjusted to omit TEV-cleavage, so that following 3D reconstruction of negative stain EM data, additional density in the 3D volume would correspond to the Ctc1 N-terminal GST-tag. Alternatively, antibodies can be used to bind and point to smaller tags, such as the Stn1 N-terminal His₆-tag. Finally, negative stain EM

structures of CST in complex with substrate DNA can reveal the location of Ctc1 N-terminal OB-folds.

More importantly, however, improving resolution would reduce or eliminate the ambiguity during the generation of a molecular model and facilitate determining the position of Stn1 and Ten1 within the obtained 3D volume. Collecting a much larger negative stain EM dataset following the protocols laid out in this thesis may improve the resolution by another 5-10 Å. But even at a resolution of 15 Å, the construction of a molecular model using the generated EM density map would remain unreliable and provide limited information. In order to approach intermediate- (4-10 Å) to high-resolution (<4 Å) maps, it is necessary to perform cryo-EM using the high-quality material that is obtainable by following the established CST purification protocol. Potentially, the elongated shape of CST in combination with its comparatively small size ($MW_{\text{total}} = 190.9$ kDa) may cause complications in the detection of particles due to the generation of low contrast images. Measures to increase the contrast during cryo-EM include the adjustment of settings such as defocus, objective apertures and electron beam intensity. However, since ice-embedded biological material is sensitive to irradiation, the electron beam intensity can only be increased to a certain threshold. This limits the extent to which the signal-to-noise ratio can be improved and hampers the collection of high-contrast images. High contrast is essential for the accurate alignment of observed particles, which in turn has a direct effect on the resolution of the final reconstruction. Another complication arises due to the application of defocus, which leads to underrepresentation of certain low spatial frequencies. This can be circumvented by collecting and combining images of various defocus values that contain contrast information for different spatial frequencies. However, it is difficult to distinguish small protein particles at low defocus values, especially with low signal-to-noise ratios.

Alternatively, the use of a phase plate can enhance the transfer of information at low spatial frequencies and thereby increase contrast.^{421,422} Volta phase plates (VPPs)

enable high contrast transfer of low spatial frequencies while images are being collected close to focus. This increases the signal-to-noise ratio significantly and therefore allows for ready detection of weak-phase particles such as small molecular weight proteins.⁴³⁴ In addition, in contrast to other phase plates, images collected with VPPs retain high spatial frequency information which is essential for the reconstruction of high-resolution volumes.⁴³⁵⁻⁴³⁸ In a recent publication, the structure of haemoglobin (MW = 64 kDa) was solved at 3.2 Å resolution using a combination of VPP and the application of approx. 500 nm defocus, playing to the strengths of both phase plate technology and conventional transmission EM.⁴³⁹ A similar approach can be taken in solving the structure of CST at high resolution.

Future work on the structural elucidation of CST alone and in complex with interaction partners, such as single-stranded DNA and telomeric protein factors will give invaluable insights into the regulation of telomere maintenance. This in turn can provide new drug targets for medical treatment of cancer. It is important to keep in mind that only a small fraction of CST, approximately 20%, localizes to telomeres, whereas the remaining 80% accumulate at other nuclear foci where CST is likely to carry out other cellular functions.²⁷⁹ Targeting these mechanisms that are unspecific to telomere maintenance with inhibitory drugs may prove toxic for the organism, and should be avoided. The protein interface between POT1-TPP1 and CST, on the other hand, is of great interest, since this interaction is specific to the regulation of telomere length. When disrupted, targeted medication may prevent CST localization to telomeres and reproduce knock-down phenotypes, such as accumulation of chromatin bridges, micronuclei and γH2AX foci.³¹³ The inhibition of CST localization to telomeres has a potential of compromising telomere integrity and leading to sporadic loss of telomeres in proliferative, telomerase-active cells.³¹³ It remains to be seen, however, if the disruption of CST interaction with POT1-TPP1 would also affect healthy somatic cells.

From an academic perspective, elucidating the structure of CST in complex with single-stranded DNA would provide a clear understanding of its molecular basics. Currently, Ctc1 is thought to bind single-stranded DNA through two putative OB-folds located close to the N-terminus of the protein. However, there are conflicting reports on the specificity of CST towards telomeric repeats.^{243,277} High-resolution structural data would help define exact boundaries of the two putative OB-folds and conclusively determine whether or not CST exhibits DNA-binding specificity. Furthermore, similar to RPA, it is possible that CST binding to DNA can give rise to alternative conformational states. To address this possibility, structural studies by cryo-EM are especially well suited, since instant vitrification of the sample has the potential of capturing alternative conformations of particles in amorphous ice. The results presented in this thesis and technical considerations discussed in this chapter provide a solid foundation and guideline for further cryo-EM studies.

In conclusion, I have demonstrated a way towards the purification of TCAB1 for use in biochemical and structural studies, and established an expression and purification protocol for the CST complex. Furthermore, I have generated a low-resolution EM density map of the full-length CST complex and elaborated on experimental strategies to further improve the resolution of the structure.

References

- 1 Bell, S. P. & Stillman, B. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357**, 128-134, doi:10.1038/357128a0 (1992).
- 2 Vashee, S. *et al.* Sequence-independent DNA binding and replication initiation by the human origin recognition complex. *Genes & development* **17**, 1894-1908, doi:10.1101/gad.1084203 (2003).
- 3 Theis, J. F. & Newlon, C. S. The ARS309 chromosomal replicator of *Saccharomyces cerevisiae* depends on an exceptional ARS consensus sequence. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 10786-10791 (1997).
- 4 Schaarschmidt, D., Baltin, J., Stehle, I. M., Lipps, H. J. & Knippers, R. An episomal mammalian replicon: sequence-independent binding of the origin recognition complex. *The EMBO journal* **23**, 191-201, doi:10.1038/sj.emboj.7600029 (2004).
- 5 Palzkill, T. G. & Newlon, C. S. A yeast replication origin consists of multiple copies of a small conserved sequence. *Cell* **53**, 441-450 (1988).
- 6 Marahrens, Y. & Stillman, B. A yeast chromosomal origin of DNA replication defined by multiple functional elements. *Science* **255**, 817-823 (1992).
- 7 Chuang, R. Y. & Kelly, T. J. The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 2656-2661 (1999).
- 8 Cayrou, C., Gregoire, D., Coulombe, P., Danis, E. & Mechali, M. Genome-scale identification of active DNA replication origins. *Methods* **57**, 158-164, doi:10.1016/j.ymeth.2012.06.015 (2012).
- 9 Cayrou, C. *et al.* Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome research* **21**, 1438-1449, doi:10.1101/gr.121830.111 (2011).
- 10 Besnard, E. *et al.* Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nature structural & molecular biology* **19**, 837-844, doi:10.1038/nsmb.2339 (2012).
- 11 Mesner, L. D. & Hamlin, J. L. Specific signals at the 3' end of the DHFR gene define one boundary of the downstream origin of replication. *Genes & development* **19**, 1053-1066, doi:10.1101/gad.1307105 (2005).
- 12 Kalejta, R. F. *et al.* Distal sequences, but not ori-beta/OBR-1, are essential for initiation of DNA replication in the Chinese hamster DHFR origin. *Molecular cell* **2**, 797-806 (1998).
- 13 Hayashida, T. *et al.* Replication initiation from a novel origin identified in the Th2 cytokine cluster locus requires a distant conserved noncoding sequence. *Journal of immunology* **176**, 5446-5454 (2006).
- 14 Aladjem, M. I. *et al.* Participation of the human beta-globin locus control region in initiation of DNA replication. *Science* **270**, 815-819 (1995).
- 15 Prioleau, M. N., Gendron, M. C. & Hyrien, O. Replication of the chicken beta-globin locus: early-firing origins at the 5' HS4 insulator and the rho- and betaA-globin genes show opposite epigenetic modifications. *Molecular and cellular biology* **23**, 3536-3549 (2003).
- 16 Pasero, P., Bensimon, A. & Schwob, E. Single-molecule analysis reveals clustering and epigenetic regulation of replication origins at the yeast rDNA locus. *Genes & development* **16**, 2479-2484, doi:10.1101/gad.232902 (2002).

- 17 Pappas, D. L., Jr., Frisch, R. & Weinreich, M. The NAD(+)-dependent Sir2p histone deacetylase is a negative regulator of chromosomal DNA replication. *Genes & development* **18**, 769-781, doi:10.1101/gad.1173204 (2004).
- 18 Miotto, B. & Struhl, K. HBO1 histone acetylase is a coactivator of the replication licensing factor Cdt1. *Genes & development* **22**, 2633-2638, doi:10.1101/gad.1674108 (2008).
- 19 Iizuka, M., Matsui, T., Takisawa, H. & Smith, M. M. Regulation of replication licensing by acetyltransferase Hbo1. *Molecular and cellular biology* **26**, 1098-1108, doi:10.1128/MCB.26.3.1098-1108.2006 (2006).
- 20 Danis, E. *et al.* Specification of a DNA replication origin by a transcription complex. *Nature cell biology* **6**, 721-730, doi:10.1038/ncb1149 (2004).
- 21 Crampton, A., Chang, F., Pappas, D. L., Jr., Frisch, R. L. & Weinreich, M. An ARS element inhibits DNA replication through a SIR2-dependent mechanism. *Molecular cell* **30**, 156-166, doi:10.1016/j.molcel.2008.02.019 (2008).
- 22 Burke, T. W., Cook, J. G., Asano, M. & Nevins, J. R. Replication factors MCM2 and ORC1 interact with the histone acetyltransferase HBO1. *The Journal of biological chemistry* **276**, 15397-15408, doi:10.1074/jbc.M011556200 (2001).
- 23 Aggarwal, B. D. & Calvi, B. R. Chromatin regulates origin activity in Drosophila follicle cells. *Nature* **430**, 372-376, doi:10.1038/nature02694 (2004).
- 24 Snyder, M., Sapolsky, R. J. & Davis, R. W. Transcription interferes with elements important for chromosome maintenance in *Saccharomyces cerevisiae*. *Molecular and cellular biology* **8**, 2184-2194 (1988).
- 25 Saha, S., Shan, Y., Mesner, L. D. & Hamlin, J. L. The promoter of the Chinese hamster ovary dihydrofolate reductase gene regulates the activity of the local origin and helps define its boundaries. *Genes & development* **18**, 397-410, doi:10.1101/gad.1171404 (2004).
- 26 Nieduszynski, C. A., Blow, J. J. & Donaldson, A. D. The requirement of yeast replication origins for pre-replication complex proteins is modulated by transcription. *Nucleic acids research* **33**, 2410-2420, doi:10.1093/nar/gki539 (2005).
- 27 Haase, S. B., Heinzl, S. S. & Calos, M. P. Transcription inhibits the replication of autonomously replicating plasmids in human cells. *Molecular and cellular biology* **14**, 2516-2524 (1994).
- 28 Donato, J. J., Chung, S. C. & Tye, B. K. Genome-wide hierarchy of replication origin usage in *Saccharomyces cerevisiae*. *PLoS genetics* **2**, e141, doi:10.1371/journal.pgen.0020141 (2006).
- 29 Dayn, A., Malkhosyan, S. & Mirkin, S. M. Transcriptionally driven cruciform formation in vivo. *Nucleic acids research* **20**, 5991-5997 (1992).
- 30 Tada, S., Li, A., Maiorano, D., Mechali, M. & Blow, J. J. Repression of origin assembly in metaphase depends on inhibition of RLF-B/Cdt1 by geminin. *Nature cell biology* **3**, 107-113, doi:10.1038/35055000 (2001).
- 31 Wohlschlegel, J. A. *et al.* Inhibition of eukaryotic DNA replication by geminin binding to Cdt1. *Science* **290**, 2309-2312, doi:10.1126/science.290.5500.2309 (2000).
- 32 Noton, E. & Diffley, J. F. CDK inactivation is the only essential function of the APC/C and the mitotic exit network proteins for origin resetting during mitosis. *Molecular cell* **5**, 85-95 (2000).
- 33 Waga, S. & Stillman, B. Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication in vitro. *Nature* **369**, 207-212, doi:10.1038/369207a0 (1994).
- 34 Tsurimoto, T., Melendy, T. & Stillman, B. Sequential initiation of lagging and leading strand synthesis by two different polymerase complexes at the SV40 DNA replication origin. *Nature* **346**, 534-539, doi:10.1038/346534a0 (1990).

- 35 Nethanel, T., Reisfeld, S., Dinter-Gottlieb, G. & Kaufmann, G. An Okazaki piece of simian virus 40 may be synthesized by ligation of shorter precursor chains. *Journal of virology* **62**, 2867-2873 (1988).
- 36 Frick, D. N. & Richardson, C. C. DNA primases. *Annual review of biochemistry* **70**, 39-80, doi:10.1146/annurev.biochem.70.1.39 (2001).
- 37 Stukenberg, P. T., Studwell-Vaughan, P. S. & O'Donnell, M. Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. *The Journal of biological chemistry* **266**, 11328-11334 (1991).
- 38 Munn, M. M. & Alberts, B. M. DNA footprinting studies of the complex formed by the T4 DNA polymerase holoenzyme at a primer-template junction. *The Journal of biological chemistry* **266**, 20034-20044 (1991).
- 39 Kong, X. P., Onrust, R., O'Donnell, M. & Kuriyan, J. Three-dimensional structure of the beta subunit of E. coli DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell* **69**, 425-437 (1992).
- 40 Gogol, E. P., Young, M. C., Kubasek, W. L., Jarvis, T. C. & von Hippel, P. H. Cryoelectron microscopic visualization of functional subassemblies of the bacteriophage T4 DNA replication complex. *Journal of molecular biology* **224**, 395-412 (1992).
- 41 Capson, T. L., Benkovic, S. J. & Nossal, N. G. Protein-DNA cross-linking demonstrates stepwise ATP-dependent assembly of T4 DNA polymerase and its accessory proteins on the primer-template. *Cell* **65**, 249-258 (1991).
- 42 Podust, V. N. & Hubscher, U. Lagging strand DNA synthesis by calf thymus DNA polymerases alpha, beta, delta and epsilon in the presence of auxiliary proteins. *Nucleic acids research* **21**, 841-846 (1993).
- 43 Lee, S. H., Kwong, A. D., Pan, Z. Q. & Hurwitz, J. Studies on the activator 1 protein complex, an accessory factor for proliferating cell nuclear antigen-dependent DNA polymerase delta. *The Journal of biological chemistry* **266**, 594-602 (1991).
- 44 Pursell, Z. F., Isoz, I., Lundstrom, E. B., Johansson, E. & Kunkel, T. A. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science* **317**, 127-130, doi:10.1126/science.1144067 (2007).
- 45 Nick McElhinny, S. A., Gordenin, D. A., Stith, C. M., Burgers, P. M. & Kunkel, T. A. Division of labor at the eukaryotic replication fork. *Molecular cell* **30**, 137-144, doi:10.1016/j.molcel.2008.02.022 (2008).
- 46 Miyabe, I., Kunkel, T. A. & Carr, A. M. The major roles of DNA polymerases epsilon and delta at the eukaryotic replication fork are evolutionarily conserved. *PLoS genetics* **7**, e1002407, doi:10.1371/journal.pgen.1002407 (2011).
- 47 Georgescu, R. E. *et al.* Mechanism of asymmetric polymerase assembly at the eukaryotic replication fork. *Nature structural & molecular biology* **21**, 664-670, doi:10.1038/nsmb.2851 (2014).
- 48 Clausen, A. R. *et al.* Tracking replication enzymology in vivo by genome-wide mapping of ribonucleotide incorporation. *Nature structural & molecular biology* **22**, 185-191, doi:10.1038/nsmb.2957 (2015).
- 49 Kouprina, N. *et al.* CTF4 (CHL15) mutants exhibit defective DNA metabolism in the yeast *Saccharomyces cerevisiae*. *Molecular and cellular biology* **12**, 5736-5747 (1992).
- 50 Villa, F. *et al.* Ctf4 Is a Hub in the Eukaryotic Replisome that Links Multiple CIP-Box Proteins to the CMG Helicase. *Molecular cell* **63**, 385-396, doi:10.1016/j.molcel.2016.06.009 (2016).

- 51 Samora, C. P. *et al.* Ctf4 Links DNA Replication with Sister Chromatid Cohesion Establishment by Recruiting the Chl1 Helicase to the Replisome. *Molecular cell* **63**, 371-384, doi:10.1016/j.molcel.2016.05.036 (2016).
- 52 Turchi, J. J., Huang, L., Murante, R. S., Kim, Y. & Bambara, R. A. Enzymatic completion of mammalian lagging-strand DNA replication. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 9803-9807 (1994).
- 53 Watson, J. D. Origin of concatemeric T7 DNA. *Nature: New biology* **239**, 197-201 (1972).
- 54 Olovnikov, A. M. A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *Journal of theoretical biology* **41**, 181-190 (1973).
- 55 LeBel, C. & Wellinger, R. J. Telomeres: what's new at your end? *Journal of cell science* **118**, 2785-2788, doi:10.1242/jcs.02394 (2005).
- 56 Larrivee, M., LeBel, C. & Wellinger, R. J. The generation of proper constitutive G-tails on yeast telomeres is dependent on the MRX complex. *Genes & development* **18**, 1391-1396, doi:10.1101/gad.1199404 (2004).
- 57 Soudet, J., Jolivet, P. & Teixeira, M. T. Elucidation of the DNA end-replication problem in *Saccharomyces cerevisiae*. *Molecular cell* **53**, 954-964, doi:10.1016/j.molcel.2014.02.030 (2014).
- 58 Liew, L. P. & Norbury, C. J. Telomere maintenance: all's well that ends well. *Archives of toxicology* **83**, 407-416, doi:10.1007/s00204-009-0423-1 (2009).
- 59 Makarov, V. L., Hirose, Y. & Langmore, J. P. Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. *Cell* **88**, 657-666 (1997).
- 60 Moyzis, R. K. *et al.* A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 6622-6626 (1988).
- 61 Wright, W. E., Tesmer, V. M., Huffman, K. E., Levene, S. D. & Shay, J. W. Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes & development* **11**, 2801-2809 (1997).
- 62 de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes & development* **19**, 2100-2110, doi:10.1101/gad.1346005 (2005).
- 63 de Lange, T. Activation of telomerase in a human tumor. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 2882-2885 (1994).
- 64 Hockemeyer, D., Daniels, J. P., Takai, H. & de Lange, T. Recent expansion of the telomeric complex in rodents: Two distinct POT1 proteins protect mouse telomeres. *Cell* **126**, 63-77, doi:10.1016/j.cell.2006.04.044 (2006).
- 65 Konishi, A. & de Lange, T. Cell cycle control of telomere protection and NHEJ revealed by a ts mutation in the DNA-binding domain of TRF2. *Genes & development* **22**, 1221-1230, doi:10.1101/gad.1634008 (2008).
- 66 Wu, L. *et al.* Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres. *Cell* **126**, 49-62, doi:10.1016/j.cell.2006.05.037 (2006).
- 67 Gottschling, D. E. & Stoddard, B. Telomeres: structure of a chromosome's aglet. *Curr Biol* **9**, R164-167 (1999).
- 68 Horvath, M. P., Schweiker, V. L., Bevilacqua, J. M., Ruggles, J. A. & Schultz, S. C. Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single strand DNA. *Cell* **95**, 963-974 (1998).

- 69 Gray, J. T., Celandier, D. W., Price, C. M. & Cech, T. R. Cloning and expression of genes for the *Oxytricha* telomere-binding protein: specific subunit interactions in the telomeric complex. *Cell* **67**, 807-814 (1991).
- 70 Blackburn, E. H. & Collins, K. Telomerase: an RNP enzyme synthesizes DNA. *Cold Spring Harbor perspectives in biology* **3**, doi:10.1101/cshperspect.a003558 (2011).
- 71 Zhong, F. L. *et al.* TPP1 OB-fold domain controls telomere maintenance by recruiting telomerase to chromosome ends. *Cell* **150**, 481-494, doi:10.1016/j.cell.2012.07.012 (2012).
- 72 Sexton, A. N., Youmans, D. T. & Collins, K. Specificity requirements for human telomere protein interaction with telomerase holoenzyme. *The Journal of biological chemistry* **287**, 34455-34464, doi:10.1074/jbc.M112.394767 (2012).
- 73 Xin, H. *et al.* TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase. *Nature* **445**, 559-562, doi:10.1038/nature05469 (2007).
- 74 Weaver, D. T. Telomeres: moonlighting by DNA repair proteins. *Curr Biol* **8**, R492-494 (1998).
- 75 Bianchi, A., Smith, S., Chong, L., Elias, P. & de Lange, T. TRF1 is a dimer and bends telomeric DNA. *The EMBO journal* **16**, 1785-1794, doi:10.1093/emboj/16.7.1785 (1997).
- 76 Chen, Y. *et al.* A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins. *Science* **319**, 1092-1096, doi:10.1126/science.1151804 (2008).
- 77 Billaud, T. *et al.* Telomeric localization of TRF2, a novel human telobox protein. *Nature genetics* **17**, 236-239, doi:10.1038/ng1097-236 (1997).
- 78 Broccoli, D., Smogorzewska, A., Chong, L. & de Lange, T. Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nature genetics* **17**, 231-235, doi:10.1038/ng1097-231 (1997).
- 79 Chong, L. *et al.* A human telomeric protein. *Science* **270**, 1663-1667 (1995).
- 80 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 reports* **6**, 39-45, doi:10.1038/sj.embor.7400314 (2005).
- 81 Hanaoka, S., Nagadoi, A. & Nishimura, Y. Comparison between TRF2 and TRF1 of their telomeric DNA-bound structures and DNA-binding activities. *Protein science : a publication of the Protein Society* **14**, 119-130, doi:10.1110/ps.04983705 (2005).
- 82 Nishikawa, T. *et al.* Solution structure of a telomeric DNA complex of human TRF1. *Structure* **9**, 1237-1251 (2001).
- 83 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. *Molecular cell* **8**, 351-361 (2001).
- 84 Bianchi, A. *et al.* TRF1 binds a bipartite telomeric site with extreme spatial flexibility. *The EMBO journal* **18**, 5735-5744, doi:10.1093/emboj/18.20.5735 (1999).
- 85 Li, B., Oestreich, S. & de Lange, T. Identification of human Rap1: implications for telomere evolution. *Cell* **101**, 471-483 (2000).
- 86 Celli, G. B. & de Lange, T. DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion. *Nature cell biology* **7**, 712-718, doi:10.1038/ncb1275 (2005).
- 87 Zhu, X. D., Kuster, B., Mann, M., Petrini, J. H. & de Lange, T. Cell-cycle-regulated association of RAD50/MRE11/NBS1 with TRF2 and human telomeres. *Nature genetics* **25**, 347-352, doi:10.1038/77139 (2000).
- 88 Li, B. & de Lange, T. Rap1 affects the length and heterogeneity of human telomeres. *Molecular biology of the cell* **14**, 5060-5068, doi:10.1091/mbc.E03-06-0403 (2003).

- 89 Hanaoka, S. *et al.* NMR structure of the hRap1 Myb motif reveals a canonical three-helix bundle lacking the positive surface charge typical of Myb DNA-binding domains. *Journal of molecular biology* **312**, 167-175, doi:10.1006/jmbi.2001.4924 (2001).
- 90 Ye, J. Z. *et al.* POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex. *Genes & development* **18**, 1649-1654, doi:10.1101/gad.1215404 (2004).
- 91 Ye, J. Z. *et al.* TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres. *The Journal of biological chemistry* **279**, 47264-47271, doi:10.1074/jbc.M409047200 (2004).
- 92 Kim, S. H., Kaminker, P. & Campisi, J. TIN2, a new regulator of telomere length in human cells. *Nature genetics* **23**, 405-412, doi:10.1038/70508 (1999).
- 93 Kim, S. H. *et al.* TIN2 mediates functions of TRF2 at human telomeres. *The Journal of biological chemistry* **279**, 43799-43804, doi:10.1074/jbc.M408650200 (2004).
- 94 Houghtaling, B. R., Cuttonaro, L., Chang, W. & Smith, S. A dynamic molecular link between the telomere length regulator TRF1 and the chromosome end protector TRF2. *Curr Biol* **14**, 1621-1631, doi:10.1016/j.cub.2004.08.052 (2004).
- 95 Abreu, E. *et al.* TIN2-tethered TPP1 recruits human telomerase to telomeres in vivo. *Molecular and cellular biology* **30**, 2971-2982, doi:10.1128/MCB.00240-10 (2010).
- 96 Hockemeyer, D. *et al.* Telomere protection by mammalian Pot1 requires interaction with Tpp1. *Nature structural & molecular biology* **14**, 754-761, doi:10.1038/nsmb1270 (2007).
- 97 Denchi, E. L. & de Lange, T. Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. *Nature* **448**, 1068-1071, doi:10.1038/nature06065 (2007).
- 98 Loayza, D., Parsons, H., Donigian, J., Hoke, K. & de Lange, T. DNA binding features of human POT1: a nonamer 5'-TAGGGTTAG-3' minimal binding site, sequence specificity, and internal binding to multimeric sites. *The Journal of biological chemistry* **279**, 13241-13248, doi:10.1074/jbc.M312309200 (2004).
- 99 Lei, M., Podell, E. R. & Cech, T. R. Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nature structural & molecular biology* **11**, 1223-1229, doi:10.1038/nsmb867 (2004).
- 100 Kelleher, C., Kurth, I. & Lingner, J. Human protection of telomeres 1 (POT1) is a negative regulator of telomerase activity in vitro. *Molecular and cellular biology* **25**, 808-818, doi:10.1128/MCB.25.2.808-818.2005 (2005).
- 101 Baumann, P. & Cech, T. R. Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science* **292**, 1171-1175, doi:10.1126/science.1060036 (2001).
- 102 Wang, F. *et al.* The POT1-TPP1 telomere complex is a telomerase processivity factor. *Nature* **445**, 506-510, doi:10.1038/nature05454 (2007).
- 103 Liu, D. *et al.* PTOP interacts with POT1 and regulates its localization to telomeres. *Nature cell biology* **6**, 673-680, doi:10.1038/ncb1142 (2004).
- 104 Griffith, J. D. *et al.* Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503-514 (1999).
- 105 Nikitina, T. & Woodcock, C. L. Closed chromatin loops at the ends of chromosomes. *The Journal of cell biology* **166**, 161-165, doi:10.1083/jcb.200403118 (2004).
- 106 Stansel, R. M., de Lange, T. & Griffith, J. D. T-loop assembly in vitro involves binding of TRF2 near the 3' telomeric overhang. *The EMBO journal* **20**, 5532-5540, doi:10.1093/emboj/20.19.5532 (2001).
- 107 de Lange, T. & Petrini, J. H. A new connection at human telomeres: association of the Mre11 complex with TRF2. *Cold Spring Harbor symposia on quantitative biology* **65**, 265-273 (2000).

- 108 Kim, S. H., Han, S., You, Y. H., Chen, D. J. & Campisi, J. The human telomere-associated protein TIN2 stimulates interactions between telomeric DNA tracts in vitro. *EMBO reports* **4**, 685-691, doi:10.1038/sj.embor.embor872 (2003).
- 109 Griffith, J., Bianchi, A. & de Lange, T. TRF1 promotes parallel pairing of telomeric tracts in vitro. *Journal of molecular biology* **278**, 79-88, doi:10.1006/jmbi.1998.1686 (1998).
- 110 Strahl-Bolsinger, S., Hecht, A., Luo, K. & Grunstein, M. SIR2 and SIR4 interactions differ in core and extended telomeric heterochromatin in yeast. *Genes & development* **11**, 83-93 (1997).
- 111 Takai, H., Smogorzewska, A. & de Lange, T. DNA damage foci at dysfunctional telomeres. *Curr Biol* **13**, 1549-1556 (2003).
- 112 Hockemeyer, D., Sfeir, A. J., Shay, J. W., Wright, W. E. & de Lange, T. POT1 protects telomeres from a transient DNA damage response and determines how human chromosomes end. *The EMBO journal* **24**, 2667-2678, doi:10.1038/sj.emboj.7600733 (2005).
- 113 d'Adda di Fagagna, F. *et al.* A DNA damage checkpoint response in telomere-initiated senescence. *Nature* **426**, 194-198, doi:10.1038/nature02118 (2003).
- 114 van Steensel, B., Smogorzewska, A. & de Lange, T. TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**, 401-413 (1998).
- 115 Karlseder, J., Broccoli, D., Dai, Y., Hardy, S. & de Lange, T. p53- and ATM-dependent apoptosis induced by telomeres lacking TRF2. *Science* **283**, 1321-1325 (1999).
- 116 Else, T. *et al.* Tpp1/Acd maintains genomic stability through a complex role in telomere protection. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **15**, 1001-1013, doi:10.1007/s10577-007-1175-5 (2007).
- 117 Kibe, T., Osawa, G. A., Keegan, C. E. & de Lange, T. Telomere protection by TPP1 is mediated by POT1a and POT1b. *Molecular and cellular biology* **30**, 1059-1066, doi:10.1128/MCB.01498-09 (2010).
- 118 Gellert, M., Lipsett, M. N. & Davies, D. R. Helix formation by guanylic acid. *Proceedings of the National Academy of Sciences of the United States of America* **48**, 2013-2018 (1962).
- 119 Konig, S. L., Evans, A. C. & Huppert, J. L. Seven essential questions on G-quadruplexes. *Biomolecular concepts* **1**, 197-213, doi:10.1515/bmc.2010.011 (2010).
- 120 Zhang, A. Y. & Balasubramanian, S. The kinetics and folding pathways of intramolecular G-quadruplex nucleic acids. *Journal of the American Chemical Society* **134**, 19297-19308, doi:10.1021/ja309851t (2012).
- 121 Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nature chemistry* **5**, 182-186, doi:10.1038/nchem.1548 (2013).
- 122 Yang, Q. *et al.* Verification of specific G-quadruplex structure by using a novel cyanine dye supramolecular assembly: I. recognizing mixed G-quadruplex in human telomeres. *Chemical communications*, 1103-1105, doi:10.1039/b820101c (2009).
- 123 Granotier, C. *et al.* Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic acids research* **33**, 4182-4190, doi:10.1093/nar/gki722 (2005).
- 124 Lin, W. *et al.* Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity. *The EMBO journal* **32**, 1425-1439, doi:10.1038/emboj.2013.88 (2013).
- 125 Crabbe, L., Verdun, R. E., Haggblom, C. I. & Karlseder, J. Defective telomere lagging strand synthesis in cells lacking WRN helicase activity. *Science* **306**, 1951-1953, doi:10.1126/science.1103619 (2004).

- 126 Brosh, R. M., Jr. DNA helicases involved in DNA repair and their roles in cancer. *Nature reviews. Cancer* **13**, 542-558, doi:10.1038/nrc3560 (2013).
- 127 Ding, H. *et al.* Regulation of murine telomere length by Rtel: an essential gene encoding a helicase-like protein. *Cell* **117**, 873-886, doi:10.1016/j.cell.2004.05.026 (2004).
- 128 Opresko, P. L. *et al.* Telomere-binding protein TRF2 binds to and stimulates the Werner and Bloom syndrome helicases. *The Journal of biological chemistry* **277**, 41110-41119, doi:10.1074/jbc.M205396200 (2002).
- 129 Opresko, P. L. *et al.* POT1 stimulates RecQ helicases WRN and BLM to unwind telomeric DNA substrates. *The Journal of biological chemistry* **280**, 32069-32080, doi:10.1074/jbc.M505211200 (2005).
- 130 Lingner, J. *et al.* Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* **276**, 561-567 (1997).
- 131 Greider, C. W. & Blackburn, E. H. The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell* **51**, 887-898 (1987).
- 132 Greider, C. W. & Blackburn, E. H. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* **43**, 405-413 (1985).
- 133 Feng, J. *et al.* The RNA component of human telomerase. *Science* **269**, 1236-1241 (1995).
- 134 de Lange, T. & Jacks, T. For better or worse? Telomerase inhibition and cancer. *Cell* **98**, 273-275 (1999).
- 135 Weinrich, S. L. *et al.* Reconstitution of human telomerase with the template RNA component hTR and the catalytic protein subunit hTERT. *Nature genetics* **17**, 498-502, doi:10.1038/ng1297-498 (1997).
- 136 Nakayama, J. *et al.* Telomerase activation by hTERT in human normal fibroblasts and hepatocellular carcinomas. *Nature genetics* **18**, 65-68, doi:10.1038/ng0198-65 (1998).
- 137 Collins, K. & Gandhi, L. The reverse transcriptase component of the Tetrahymena telomerase ribonucleoprotein complex. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 8485-8490 (1998).
- 138 Moriarty, T. J., Marie-Egyptienne, D. T. & Autexier, C. Functional organization of repeat addition processivity and DNA synthesis determinants in the human telomerase multimer. *Molecular and cellular biology* **24**, 3720-3733 (2004).
- 139 Lue, N. F. A physical and functional constituent of telomerase anchor site. *The Journal of biological chemistry* **280**, 26586-26591, doi:10.1074/jbc.M503028200 (2005).
- 140 Moriarty, T. J., Ward, R. J., Taboski, M. A. & Autexier, C. An anchor site-type defect in human telomerase that disrupts telomere length maintenance and cellular immortalization. *Molecular biology of the cell* **16**, 3152-3161, doi:10.1091/mbc.E05-02-0148 (2005).
- 141 Lue, N. F. & Li, Z. Modeling and structure function analysis of the putative anchor site of yeast telomerase. *Nucleic acids research* **35**, 5213-5222, doi:10.1093/nar/gkm531 (2007).
- 142 Zaug, A. J., Podell, E. R. & Cech, T. R. Mutation in TERT separates processivity from anchor-site function. *Nature structural & molecular biology* **15**, 870-872, doi:10.1038/nsmb.1462 (2008).
- 143 Zaug, A. J., Podell, E. R., Nandakumar, J. & Cech, T. R. Functional interaction between telomere protein TPP1 and telomerase. *Genes & development* **24**, 613-622, doi:10.1101/gad.1881810 (2010).
- 144 Schmidt, J. C., Dalby, A. B. & Cech, T. R. Identification of human TERT elements necessary for telomerase recruitment to telomeres. *eLife* **3**, doi:10.7554/eLife.03563 (2014).

- 145 Wu, R. A. & Collins, K. Human telomerase specialization for repeat synthesis by unique handling of primer-template duplex. *The EMBO journal* **33**, 921-935, doi:10.1002/embj.201387205 (2014).
- 146 Robart, A. R. & Collins, K. Human telomerase domain interactions capture DNA for TEN domain-dependent processive elongation. *Molecular cell* **42**, 308-318, doi:10.1016/j.molcel.2011.03.012 (2011).
- 147 Jurczyk, J. *et al.* Direct involvement of the TEN domain at the active site of human telomerase. *Nucleic acids research* **39**, 1774-1788, doi:10.1093/nar/gkq1083 (2011).
- 148 Rouda, S. & Skordalakes, E. Structure of the RNA-binding domain of telomerase: implications for RNA recognition and binding. *Structure* **15**, 1403-1412, doi:10.1016/j.str.2007.09.007 (2007).
- 149 Miller, M. C., Liu, J. K. & Collins, K. Template definition by Tetrahymena telomerase reverse transcriptase. *The EMBO journal* **19**, 4412-4422, doi:10.1093/emboj/19.16.4412 (2000).
- 150 Bosoy, D., Peng, Y., Mian, I. S. & Lue, N. F. Conserved N-terminal motifs of telomerase reverse transcriptase required for ribonucleoprotein assembly in vivo. *The Journal of biological chemistry* **278**, 3882-3890, doi:10.1074/jbc.M210645200 (2003).
- 151 Bachand, F. & Autexier, C. Functional regions of human telomerase reverse transcriptase and human telomerase RNA required for telomerase activity and RNA-protein interactions. *Molecular and cellular biology* **21**, 1888-1897, doi:10.1128/MCB.21.5.1888-1897.2001 (2001).
- 152 Lai, C. K., Mitchell, J. R. & Collins, K. RNA binding domain of telomerase reverse transcriptase. *Molecular and cellular biology* **21**, 990-1000, doi:10.1128/MCB.21.4.990-1000.2001 (2001).
- 153 Bryan, T. M., Goodrich, K. J. & Cech, T. R. Telomerase RNA bound by protein motifs specific to telomerase reverse transcriptase. *Molecular cell* **6**, 493-499 (2000).
- 154 O'Connor, C. M., Lai, C. K. & Collins, K. Two purified domains of telomerase reverse transcriptase reconstitute sequence-specific interactions with RNA. *The Journal of biological chemistry* **280**, 17533-17539, doi:10.1074/jbc.M501211200 (2005).
- 155 Moriarty, T. J., Huard, S., Dupuis, S. & Autexier, C. Functional multimerization of human telomerase requires an RNA interaction domain in the N terminus of the catalytic subunit. *Molecular and cellular biology* **22**, 1253-1265 (2002).
- 156 Bryan, T. M., Sperger, J. M., Chapman, K. B. & Cech, T. R. Telomerase reverse transcriptase genes identified in Tetrahymena thermophila and Oxytricha trifallax. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 8479-8484 (1998).
- 157 Ren, J. *et al.* The structure of HIV-1 reverse transcriptase complexed with 9-chloro-TIBO: lessons for inhibitor design. *Structure* **3**, 915-926, doi:10.1016/S0969-2126(01)00226-X (1995).
- 158 Nakamura, T. M. & Cech, T. R. Reversing time: origin of telomerase. *Cell* **92**, 587-590 (1998).
- 159 Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**, 1783-1790 (1992).
- 160 Jacobo-Molina, A. *et al.* Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 6320-6324 (1993).
- 161 Nakamura, T. M. *et al.* Telomerase catalytic subunit homologs from fission yeast and human. *Science* **277**, 955-959 (1997).

- 162 Cote, M. L. & Roth, M. J. Murine leukemia virus reverse transcriptase: structural comparison with HIV-1 reverse transcriptase. *Virus research* **134**, 186-202, doi:10.1016/j.virusres.2008.01.001 (2008).
- 163 Gillis, A. J., Schuller, A. P. & Skordalakes, E. Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT. *Nature* **455**, 633-637, doi:10.1038/nature07283 (2008).
- 164 Bachand, F. & Autexier, C. Functional regions of human telomerase reverse transcriptase and human telomerase RNA required for telomerase activity and RNA-protein interactions. *Molecular and cellular biology* **21**, 1888-1897, doi:10.1128/MCB.21.5.1888-1897.2001 (2001).
- 165 Beattie, T. L., Zhou, W., Robinson, M. O. & Harrington, L. Polymerization defects within human telomerase are distinct from telomerase RNA and TEP1 binding. *Molecular biology of the cell* **11**, 3329-3340 (2000).
- 166 Lai, C. K., Mitchell, J. R. & Collins, K. RNA binding domain of telomerase reverse transcriptase. *Molecular and cellular biology* **21**, 990-1000, doi:10.1128/MCB.21.4.990-1000.2001 (2001).
- 167 Friedman, K. L. & Cech, T. R. Essential functions of amino-terminal domains in the yeast telomerase catalytic subunit revealed by selection for viable mutants. *Genes & development* **13**, 2863-2874 (1999).
- 168 Malik, H. S., Burke, W. D. & Eickbush, T. H. Putative telomerase catalytic subunits from *Giardia lamblia* and *Caenorhabditis elegans*. *Gene* **251**, 101-108 (2000).
- 169 Legassie, J. D. & Jarstfer, M. B. The unmasking of telomerase. *Structure* **14**, 1603-1609, doi:10.1016/j.str.2006.09.004 (2006).
- 170 Theimer, C. A. & Feigon, J. Structure and function of telomerase RNA. *Curr Opin Struct Biol* **16**, 307-318, doi:10.1016/j.sbi.2006.05.005 (2006).
- 171 Chen, J. L., Blasco, M. A. & Greider, C. W. Secondary structure of vertebrate telomerase RNA. *Cell* **100**, 503-514 (2000).
- 172 Ly, H., Xu, L., Rivera, M. A., Parslow, T. G. & Blackburn, E. H. A role for a novel 'trans-pseudoknot' RNA-RNA interaction in the functional dimerization of human telomerase. *Genes & development* **17**, 1078-1083, doi:10.1101/gad.1060803 (2003).
- 173 Gilley, D. & Blackburn, E. H. Specific RNA residue interactions required for enzymatic functions of *Tetrahymena* telomerase. *Molecular and cellular biology* **16**, 66-75 (1996).
- 174 Lin, J., Smith, D. L. & Blackburn, E. H. Mutant telomere sequences lead to impaired chromosome separation and a unique checkpoint response. *Molecular biology of the cell* **15**, 1623-1634, doi:10.1091/mbc.E03-10-0740 (2004).
- 175 Chen, J. L. & Greider, C. W. Template boundary definition in mammalian telomerase. *Genes & development* **17**, 2747-2752, doi:10.1101/gad.1140303 (2003).
- 176 Tzfati, Y., Fulton, T. B., Roy, J. & Blackburn, E. H. Template boundary in a yeast telomerase specified by RNA structure. *Science* **288**, 863-867 (2000).
- 177 Lai, C. K., Miller, M. C. & Collins, K. Template boundary definition in *Tetrahymena* telomerase. *Genes & development* **16**, 415-420, doi:10.1101/gad.962602 (2002).
- 178 Box, J. A., Bunch, J. T., Zappulla, D. C., Glynn, E. F. & Baumann, P. A flexible template boundary element in the RNA subunit of fission yeast telomerase. *The Journal of biological chemistry* **283**, 24224-24233, doi:10.1074/jbc.M802043200 (2008).
- 179 Theimer, C. A., Blois, C. A. & Feigon, J. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Molecular cell* **17**, 671-682, doi:10.1016/j.molcel.2005.01.017 (2005).

- 180 Kim, N. K. *et al.* Solution structure and dynamics of the wild-type pseudoknot of human telomerase RNA. *Journal of molecular biology* **384**, 1249-1261, doi:10.1016/j.jmb.2008.10.005 (2008).
- 181 Chen, J. L. & Greider, C. W. Functional analysis of the pseudoknot structure in human telomerase RNA. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8080-8085; discussion 8077-8089, doi:10.1073/pnas.0502259102 (2005).
- 182 Mitchell, J. R. & Collins, K. Human telomerase activation requires two independent interactions between telomerase RNA and telomerase reverse transcriptase. *Molecular cell* **6**, 361-371 (2000).
- 183 Autexier, C., Pruzan, R., Funk, W. D. & Greider, C. W. Reconstitution of human telomerase activity and identification of a minimal functional region of the human telomerase RNA. *The EMBO journal* **15**, 5928-5935 (1996).
- 184 Chen, J. L., Opperman, K. K. & Greider, C. W. A critical stem-loop structure in the CR4-CR5 domain of mammalian telomerase RNA. *Nucleic acids research* **30**, 592-597 (2002).
- 185 Tesmer, V. M. *et al.* Two inactive fragments of the integral RNA cooperate to assemble active telomerase with the human protein catalytic subunit (hTERT) in vitro. *Molecular and cellular biology* **19**, 6207-6216 (1999).
- 186 Collins, K. Physiological assembly and activity of human telomerase complexes. *Mech Ageing Dev* **129**, 91-98, doi:10.1016/j.mad.2007.10.008 (2008).
- 187 Kiss, T., Fayet-Lebaron, E. & Jady, B. E. Box H/ACA small ribonucleoproteins. *Molecular cell* **37**, 597-606, doi:10.1016/j.molcel.2010.01.032 (2010).
- 188 Mitchell, J. R., Cheng, J. & Collins, K. A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. *Molecular and cellular biology* **19**, 567-576 (1999).
- 189 Pogacic, V., Dragon, F. & Filipowicz, W. Human H/ACA small nucleolar RNPs and telomerase share evolutionarily conserved proteins NHP2 and NOP10. *Molecular and cellular biology* **20**, 9028-9040 (2000).
- 190 Fu, D. & Collins, K. Distinct biogenesis pathways for human telomerase RNA and H/ACA small nucleolar RNAs. *Molecular cell* **11**, 1361-1372 (2003).
- 191 Xi, L. & Cech, T. R. Inventory of telomerase components in human cells reveals multiple subpopulations of hTR and hTERT. *Nucleic acids research* **42**, 8565-8577, doi:10.1093/nar/gku560 (2014).
- 192 Holt, S. E. *et al.* Functional requirement of p23 and Hsp90 in telomerase complexes. *Genes & development* **13**, 817-826 (1999).
- 193 Venteicher, A. S., Meng, Z., Mason, P. J., Veenstra, T. D. & Artandi, S. E. Identification of ATPases pontin and reptin as telomerase components essential for holoenzyme assembly. *Cell* **132**, 945-957, doi:10.1016/j.cell.2008.01.019 (2008).
- 194 Schmidt, J. C. & Cech, T. R. Human telomerase: biogenesis, trafficking, recruitment, and activation. *Genes & development* **29**, 1095-1105, doi:10.1101/gad.263863.115 (2015).
- 195 Jady, B. E., Bertrand, E. & Kiss, T. Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. *The Journal of cell biology* **164**, 647-652, doi:10.1083/jcb.200310138 (2004).
- 196 Fu, D. & Collins, K. Human telomerase and Cajal body ribonucleoproteins share a unique specificity of Sm protein association. *Genes & development* **20**, 531-536, doi:10.1101/gad.1390306 (2006).
- 197 Fu, D. & Collins, K. Distinct biogenesis pathways for human telomerase RNA and H/ACA small nucleolar RNAs. *Molecular cell* **11**, 1361-1372 (2003).

- 198 Egan, E. D. & Collins, K. An enhanced H/ACA RNP assembly mechanism for human telomerase RNA. *Molecular and cellular biology* **32**, 2428-2439, doi:10.1128/MCB.00286-12 (2012).
- 199 Darzacq, X. *et al.* Stepwise RNP assembly at the site of H/ACA RNA transcription in human cells. *The Journal of cell biology* **173**, 207-218, doi:10.1083/jcb.200601105 (2006).
- 200 Armanios, M. & Blackburn, E. H. The telomere syndromes. *Nature reviews. Genetics* **13**, 693-704, doi:10.1038/nrg3246 (2012).
- 201 Tycowski, K. T., Shu, M. D., Kukoyi, A. & Steitz, J. A. A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles. *Molecular cell* **34**, 47-57, doi:10.1016/j.molcel.2009.02.020 (2009).
- 202 Mahmoudi, S. *et al.* WRAP53 is essential for Cajal body formation and for targeting the survival of motor neuron complex to Cajal bodies. *PLoS biology* **8**, e1000521, doi:10.1371/journal.pbio.1000521 (2010).
- 203 Jady, B. E., Bertrand, E. & Kiss, T. Human telomerase RNA and box H/ACA scaRNAs share a common Cajal body-specific localization signal. *The Journal of cell biology* **164**, 647-652, doi:10.1083/jcb.200310138 (2004).
- 204 Girard, C. *et al.* Characterization of a short isoform of human Tgs1 hypermethylase associating with small nucleolar ribonucleoprotein core proteins and produced by limited proteolytic processing. *The Journal of biological chemistry* **283**, 2060-2069, doi:10.1074/jbc.M704209200 (2008).
- 205 Tomlinson, R. L. *et al.* Telomerase reverse transcriptase is required for the localization of telomerase RNA to cajal bodies and telomeres in human cancer cells. *Molecular biology of the cell* **19**, 3793-3800, doi:10.1091/mbc.E08-02-0184 (2008).
- 206 Venteicher, A. S. *et al.* A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science* **323**, 644-648, doi:10.1126/science.1165357 (2009).
- 207 Zhong, F. *et al.* Disruption of telomerase trafficking by TCAB1 mutation causes dyskeratosis congenita. *Genes & development* **25**, 11-16, doi:10.1101/gad.2006411 (2011).
- 208 Stern, J. L., Zyner, K. G., Pickett, H. A., Cohen, S. B. & Bryan, T. M. Telomerase recruitment requires both TCAB1 and Cajal bodies independently. *Molecular and cellular biology* **32**, 2384-2395, doi:10.1128/MCB.00379-12 (2012).
- 209 Mozdy, A. D. & Cech, T. R. Low abundance of telomerase in yeast: implications for telomerase haploinsufficiency. *Rna* **12**, 1721-1737, doi:10.1261/rna.134706 (2006).
- 210 Armbruster, B. N. *et al.* Rescue of an hTERT mutant defective in telomere elongation by fusion with hPot1. *Molecular and cellular biology* **24**, 3552-3561 (2004).
- 211 Armbruster, B. N., Banik, S. S., Guo, C., Smith, A. C. & Counter, C. M. N-terminal domains of the human telomerase catalytic subunit required for enzyme activity in vivo. *Molecular and cellular biology* **21**, 7775-7786, doi:10.1128/MCB.21.22.7775-7786.2001 (2001).
- 212 Nandakumar, J. *et al.* The TEL patch of telomere protein TPP1 mediates telomerase recruitment and processivity. *Nature* **492**, 285-289, doi:10.1038/nature11648 (2012).
- 213 Kocak, H. *et al.* Hoyerlaal-Hreidarsson syndrome caused by a germline mutation in the TEL patch of the telomere protein TPP1. *Genes & development* **28**, 2090-2102, doi:10.1101/gad.248567.114 (2014).
- 214 Zhu, X. *et al.* Cell cycle-dependent modulation of telomerase activity in tumor cells. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 6091-6095 (1996).

- 215 Jady, B. E., Richard, P., Bertrand, E. & Kiss, T. Cell cycle-dependent recruitment of telomerase RNA and Cajal bodies to human telomeres. *Molecular biology of the cell* **17**, 944-954, doi:10.1091/mbc.E05-09-0904 (2006).
- 216 Zhu, L. *et al.* Telomere length regulation in mice is linked to a novel chromosome locus. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 8648-8653 (1998).
- 217 Samper, E., Flores, J. M. & Blasco, M. A. Restoration of telomerase activity rescues chromosomal instability and premature aging in Terc^{-/-} mice with short telomeres. *EMBO reports* **2**, 800-807, doi:10.1093/embo-reports/kve174 (2001).
- 218 Ouellette, M. M. *et al.* Subsenescent telomere lengths in fibroblasts immortalized by limiting amounts of telomerase. *The Journal of biological chemistry* **275**, 10072-10076 (2000).
- 219 Hemann, M. T., Strong, M. A., Hao, L. Y. & Greider, C. W. The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. *Cell* **107**, 67-77 (2001).
- 220 Zhang, Y. *et al.* Phosphorylation of TPP1 regulates cell cycle-dependent telomerase recruitment. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 5457-5462, doi:10.1073/pnas.1217733110 (2013).
- 221 Sexton, A. N. *et al.* Genetic and molecular identification of three human TPP1 functions in telomerase action: recruitment, activation, and homeostasis set point regulation. *Genes & development* **28**, 1885-1899, doi:10.1101/gad.246819.114 (2014).
- 222 Miller, M. C. & Collins, K. Telomerase recognizes its template by using an adjacent RNA motif. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 6585-6590, doi:10.1073/pnas.102024699 (2002).
- 223 Morin, G. B. Recognition of a chromosome truncation site associated with alpha-thalassaemia by human telomerase. *Nature* **353**, 454-456, doi:10.1038/353454a0 (1991).
- 224 Collins, K. & Greider, C. W. Tetrahymena telomerase catalyzes nucleolytic cleavage and nonprocessive elongation. *Genes & development* **7**, 1364-1376 (1993).
- 225 Harrington, L. A. & Greider, C. W. Telomerase primer specificity and chromosome healing. *Nature* **353**, 451-454, doi:10.1038/353451a0 (1991).
- 226 Wallweber, G., Gryaznov, S., Pongracz, K. & Pruzan, R. Interaction of human telomerase with its primer substrate. *Biochemistry* **42**, 589-600, doi:10.1021/bi026914a (2003).
- 227 Wyatt, H. D., Lobb, D. A. & Beattie, T. L. Characterization of physical and functional anchor site interactions in human telomerase. *Molecular and cellular biology* **27**, 3226-3240, doi:10.1128/MCB.02368-06 (2007).
- 228 Jacobs, S. A., Podell, E. R. & Cech, T. R. Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase. *Nature structural & molecular biology* **13**, 218-225, doi:10.1038/nsmb1054 (2006).
- 229 Romi, E. *et al.* High-resolution physical and functional mapping of the template adjacent DNA binding site in catalytically active telomerase. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 8791-8796, doi:10.1073/pnas.0703157104 (2007).
- 230 Sealey, D. C. *et al.* The N-terminus of hTERT contains a DNA-binding domain and is required for telomerase activity and cellular immortalization. *Nucleic acids research* **38**, 2019-2035, doi:10.1093/nar/gkp1160 (2010).
- 231 Finger, S. N. & Bryan, T. M. Multiple DNA-binding sites in Tetrahymena telomerase. *Nucleic acids research* **36**, 1260-1272, doi:10.1093/nar/gkm866 (2008).

- 232 Bryan, T. M., Goodrich, K. J. & Cech, T. R. A mutant of Tetrahymena telomerase reverse transcriptase with increased processivity. *The Journal of biological chemistry* **275**, 24199-24207, doi:10.1074/jbc.M003246200 (2000).
- 233 Lue, N. F. A physical and functional constituent of telomerase anchor site. *The Journal of biological chemistry* **280**, 26586-26591, doi:10.1074/jbc.M503028200 (2005).
- 234 Hammond, P. W., Lively, T. N. & Cech, T. R. The anchor site of telomerase from Euplotes aediculatus revealed by photo-cross-linking to single- and double-stranded DNA primers. *Molecular and cellular biology* **17**, 296-308 (1997).
- 235 Autexier, C. & Lue, N. F. The structure and function of telomerase reverse transcriptase. *Annual review of biochemistry* **75**, 493-517, doi:10.1146/annurev.biochem.75.103004.142412 (2006).
- 236 Forstemann, K. & Lingner, J. Telomerase limits the extent of base pairing between template RNA and telomeric DNA. *EMBO reports* **6**, 361-366, doi:10.1038/sj.embor.7400374 (2005).
- 237 Lue, N. F., Lin, Y. C. & Mian, I. S. A conserved telomerase motif within the catalytic domain of telomerase reverse transcriptase is specifically required for repeat addition processivity. *Molecular and cellular biology* **23**, 8440-8449 (2003).
- 238 Lue, N. F. & Li, Z. Modeling and structure function analysis of the putative anchor site of yeast telomerase. *Nucleic acids research* **35**, 5213-5222, doi:10.1093/nar/gkm531 (2007).
- 239 Zaug, A. J., Podell, E. R. & Cech, T. R. Mutation in TERT separates processivity from anchor-site function. *Nature structural & molecular biology* **15**, 870-872, doi:10.1038/nsmb.1462 (2008).
- 240 Xie, M., Podlevsky, J. D., Qi, X., Bley, C. J. & Chen, J. J. A novel motif in telomerase reverse transcriptase regulates telomere repeat addition rate and processivity. *Nucleic acids research* **38**, 1982-1996, doi:10.1093/nar/gkp1198 (2010).
- 241 Moriarty, T. J., Ward, R. J., Taboski, M. A. & Autexier, C. An anchor site-type defect in human telomerase that disrupts telomere length maintenance and cellular immortalization. *Molecular biology of the cell* **16**, 3152-3161, doi:10.1091/mbc.E05-02-0148 (2005).
- 242 Moriarty, T. J., Marie-Egyptienne, D. T. & Autexier, C. Functional organization of repeat addition processivity and DNA synthesis determinants in the human telomerase multimer. *Molecular and cellular biology* **24**, 3720-3733 (2004).
- 243 Latrick, C. M. & Cech, T. R. POT1-TPP1 enhances telomerase processivity by slowing primer dissociation and aiding translocation. *The EMBO journal* **29**, 924-933, doi:10.1038/emboj.2009.409 (2010).
- 244 Zhao, Y. *et al.* Processive and distributive extension of human telomeres by telomerase under homeostatic and nonequilibrium conditions. *Molecular cell* **42**, 297-307, doi:10.1016/j.molcel.2011.03.020 (2011).
- 245 Chen, L. Y., Redon, S. & Lingner, J. The human CST complex is a terminator of telomerase activity. *Nature* **488**, 540-544, doi:10.1038/nature11269 (2012).
- 246 Schoeftner, S. & Blasco, M. A. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nature cell biology* **10**, 228-236, doi:10.1038/ncb1685 (2008).
- 247 Redon, S., Reichenbach, P. & Lingner, J. The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic acids research* **38**, 5797-5806, doi:10.1093/nar/gkq296 (2010).
- 248 Sussel, L. & Shore, D. Separation of transcriptional activation and silencing functions of the RAP1-encoded repressor/activator protein 1: isolation of viable mutants affecting both

- silencing and telomere length. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 7749-7753 (1991).
- 249 Wotton, D. & Shore, D. A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to regulate telomere length in *Saccharomyces cerevisiae*. *Genes & development* **11**, 748-760 (1997).
- 250 Hardy, C. F., Sussel, L. & Shore, D. A RAP1-interacting protein involved in transcriptional silencing and telomere length regulation. *Genes & development* **6**, 801-814 (1992).
- 251 Marcand, S., Gilson, E. & Shore, D. A protein-counting mechanism for telomere length regulation in yeast. *Science* **275**, 986-990 (1997).
- 252 Marcand, S., Brevet, V. & Gilson, E. Progressive cis-inhibition of telomerase upon telomere elongation. *The EMBO journal* **18**, 3509-3519, doi:10.1093/emboj/18.12.3509 (1999).
- 253 van Steensel, B. & de Lange, T. Control of telomere length by the human telomeric protein TRF1. *Nature* **385**, 740-743, doi:10.1038/385740a0 (1997).
- 254 Loayza, D. & De Lange, T. POT1 as a terminal transducer of TRF1 telomere length control. *Nature* **423**, 1013-1018, doi:10.1038/nature01688 (2003).
- 255 Ye, J. Z. & de Lange, T. TIN2 is a tankyrase 1 PARP modulator in the TRF1 telomere length control complex. *Nature genetics* **36**, 618-623, doi:10.1038/ng1360 (2004).
- 256 Takai, K. K., Hooper, S., Blackwood, S., Gandhi, R. & de Lange, T. In vivo stoichiometry of shelterin components. *The Journal of biological chemistry* **285**, 1457-1467, doi:10.1074/jbc.M109.038026 (2010).
- 257 Teixeira, M. T., Arneric, M., Sperisen, P. & Lingner, J. Telomere length homeostasis is achieved via a switch between telomerase- extendible and -nonextendible states. *Cell* **117**, 323-335 (2004).
- 258 Grunstein, M. Molecular model for telomeric heterochromatin in yeast. *Current opinion in cell biology* **9**, 383-387 (1997).
- 259 Greider, C. W. Regulating telomere length from the inside out: the replication fork model. *Genes & development* **30**, 1483-1491, doi:10.1101/gad.280578.116 (2016).
- 260 Olins, D. E., Olins, A. L., Cacheiro, L. H. & Tan, E. M. Proliferating cell nuclear antigen/cyclin in the ciliate *Euplotes eurystomus*: localization in the replication band and in micronuclei. *The Journal of cell biology* **109**, 1399-1410 (1989).
- 261 Fang, G. & Cech, T. R. Telomerase RNA localized in the replication band and spherical subnuclear organelles in hypotrichous ciliates. *The Journal of cell biology* **130**, 243-253 (1995).
- 262 Yamazaki, S. *et al.* Rif1 regulates the replication timing domains on the human genome. *The EMBO journal* **31**, 3667-3677, doi:10.1038/emboj.2012.180 (2012).
- 263 Sreesankar, E., Bharathi, V., Mishra, R. K. & Mishra, K. *Drosophila* Rif1 is an essential gene and controls late developmental events by direct interaction with PP1-87B. *Scientific reports* **5**, 10679, doi:10.1038/srep10679 (2015).
- 264 Peace, J. M., Ter-Zakarian, A. & Aparicio, O. M. Rif1 regulates initiation timing of late replication origins throughout the *S. cerevisiae* genome. *PloS one* **9**, e98501, doi:10.1371/journal.pone.0098501 (2014).
- 265 Mattarocci, S. *et al.* Rif1 controls DNA replication timing in yeast through the PP1 phosphatase Glc7. *Cell reports* **7**, 62-69, doi:10.1016/j.celrep.2014.03.010 (2014).
- 266 Lian, H. Y. *et al.* The effect of Ku on telomere replication time is mediated by telomere length but is independent of histone tail acetylation. *Molecular biology of the cell* **22**, 1753-1765, doi:10.1091/mbc.E10-06-0549 (2011).
- 267 Hayano, M. *et al.* Rif1 is a global regulator of timing of replication origin firing in fission yeast. *Genes & development* **26**, 137-150, doi:10.1101/gad.178491.111 (2012).

- 268 Cornacchia, D. *et al.* Mouse Rif1 is a key regulator of the replication-timing programme in mammalian cells. *The EMBO journal* **31**, 3678-3690, doi:10.1038/emboj.2012.214 (2012).
- 269 Buonomo, S. B., Wu, Y., Ferguson, D. & de Lange, T. Mammalian Rif1 contributes to replication stress survival and homology-directed repair. *The Journal of cell biology* **187**, 385-398, doi:10.1083/jcb.200902039 (2009).
- 270 Raghuraman, M. K. *et al.* Replication dynamics of the yeast genome. *Science* **294**, 115-121, doi:10.1126/science.294.5540.115 (2001).
- 271 McCarroll, R. M. & Fangman, W. L. Time of replication of yeast centromeres and telomeres. *Cell* **54**, 505-513 (1988).
- 272 Ferguson, B. M. & Fangman, W. L. A position effect on the time of replication origin activation in yeast. *Cell* **68**, 333-339 (1992).
- 273 Smith, Z. E. & Higgs, D. R. The pattern of replication at a human telomeric region (16p13.3): its relationship to chromosome structure and gene expression. *Human molecular genetics* **8**, 1373-1386 (1999).
- 274 Grandin, N., Reed, S. I. & Charbonneau, M. Stn1, a new *Saccharomyces cerevisiae* protein, is implicated in telomere size regulation in association with Cdc13. *Genes & development* **11**, 512-527 (1997).
- 275 Grandin, N., Damon, C. & Charbonneau, M. Ten1 functions in telomere end protection and length regulation in association with Stn1 and Cdc13. *The EMBO journal* **20**, 1173-1183, doi:10.1093/emboj/20.5.1173 (2001).
- 276 Grandin, N., Damon, C. & Charbonneau, M. Cdc13 prevents telomere uncapping and Rad50-dependent homologous recombination. *The EMBO journal* **20**, 6127-6139, doi:10.1093/emboj/20.21.6127 (2001).
- 277 Wan, M., Qin, J., Songyang, Z. & Liu, D. OB fold-containing protein 1 (OBFC1), a human homolog of yeast Stn1, associates with TPP1 and is implicated in telomere length regulation. *The Journal of biological chemistry* **284**, 26725-26731, doi:10.1074/jbc.M109.021105 (2009).
- 278 Song, X. *et al.* STN1 protects chromosome ends in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 19815-19820, doi:10.1073/pnas.0807867105 (2008).
- 279 Miyake, Y. *et al.* RPA-like mammalian Ctc1-Stn1-Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway. *Molecular cell* **36**, 193-206, doi:10.1016/j.molcel.2009.08.009 (2009).
- 280 Bryan, C., Rice, C., Harkisheimer, M., Schultz, D. C. & Skordalakes, E. Structure of the human telomeric Stn1-Ten1 capping complex. *PloS one* **8**, e66756, doi:10.1371/journal.pone.0066756 (2013).
- 281 Hartwell, L. H. & Smith, D. Altered fidelity of mitotic chromosome transmission in cell cycle mutants of *S. cerevisiae*. *Genetics* **110**, 381-395 (1985).
- 282 Garvik, B., Carson, M. & Hartwell, L. Single-stranded DNA arising at telomeres in *cdc13* mutants may constitute a specific signal for the RAD9 checkpoint. *Molecular and cellular biology* **15**, 6128-6138 (1995).
- 283 Lendvay, T. S., Morris, D. K., Sah, J., Balasubramanian, B. & Lundblad, V. Senescence mutants of *Saccharomyces cerevisiae* with a defect in telomere replication identify three additional EST genes. *Genetics* **144**, 1399-1412 (1996).
- 284 Qi, H. & Zakian, V. A. The *Saccharomyces* telomere-binding protein Cdc13p interacts with both the catalytic subunit of DNA polymerase alpha and the telomerase-associated est1 protein. *Genes & development* **14**, 1777-1788 (2000).

- 285 Nugent, C. I., Hughes, T. R., Lue, N. F. & Lundblad, V. Cdc13p: a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance. *Science* **274**, 249-252 (1996).
- 286 Lin, J. J. & Zakian, V. A. The *Saccharomyces* CDC13 protein is a single-strand TG1-3 telomeric DNA-binding protein in vitro that affects telomere behavior in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 13760-13765 (1996).
- 287 Taggart, A. K., Teng, S. C. & Zakian, V. A. Est1p as a cell cycle-regulated activator of telomere-bound telomerase. *Science* **297**, 1023-1026, doi:10.1126/science.1074968 (2002).
- 288 Pennock, E., Buckley, K. & Lundblad, V. Cdc13 delivers separate complexes to the telomere for end protection and replication. *Cell* **104**, 387-396 (2001).
- 289 Bianchi, A., Negrini, S. & Shore, D. Delivery of yeast telomerase to a DNA break depends on the recruitment functions of Cdc13 and Est1. *Molecular cell* **16**, 139-146, doi:10.1016/j.molcel.2004.09.009 (2004).
- 290 Maringele, L. & Lydall, D. EXO1-dependent single-stranded DNA at telomeres activates subsets of DNA damage and spindle checkpoint pathways in budding yeast yku70Delta mutants. *Genes & development* **16**, 1919-1933, doi:10.1101/gad.225102 (2002).
- 291 Puglisi, A., Bianchi, A., Lemmens, L., Damay, P. & Shore, D. Distinct roles for yeast Stn1 in telomere capping and telomerase inhibition. *The EMBO journal* **27**, 2328-2339, doi:10.1038/emboj.2008.158 (2008).
- 292 Grossi, S., Puglisi, A., Dmitriev, P. V., Lopes, M. & Shore, D. Pol12, the B subunit of DNA polymerase alpha, functions in both telomere capping and length regulation. *Genes & development* **18**, 992-1006, doi:10.1101/gad.300004 (2004).
- 293 Grandin, N., Damon, C. & Charbonneau, M. Cdc13 cooperates with the yeast Ku proteins and Stn1 to regulate telomerase recruitment. *Molecular and cellular biology* **20**, 8397-8408 (2000).
- 294 Chandra, A., Hughes, T. R., Nugent, C. I. & Lundblad, V. Cdc13 both positively and negatively regulates telomere replication. *Genes & development* **15**, 404-414, doi:10.1101/gad.861001 (2001).
- 295 Qian, W. *et al.* Ten1p promotes the telomeric DNA-binding activity of Cdc13p: implication for its function in telomere length regulation. *Cell research* **19**, 849-863, doi:10.1038/cr.2009.67 (2009).
- 296 Li, S. *et al.* Cdk1-dependent phosphorylation of Cdc13 coordinates telomere elongation during cell-cycle progression. *Cell* **136**, 50-61, doi:10.1016/j.cell.2008.11.027 (2009).
- 297 Arneric, M. & Lingner, J. Tel1 kinase and subtelomere-bound Tbf1 mediate preferential elongation of short telomeres by telomerase in yeast. *EMBO reports* **8**, 1080-1085, doi:10.1038/sj.embor.7401082 (2007).
- 298 Viscardi, V., Bonetti, D., Cartagena-Lirola, H., Lucchini, G. & Longhese, M. P. MRX-dependent DNA damage response to short telomeres. *Molecular biology of the cell* **18**, 3047-3058, doi:10.1091/mbc.E07-03-0285 (2007).
- 299 Negrini, S., Ribaud, V., Bianchi, A. & Shore, D. DNA breaks are masked by multiple Rap1 binding in yeast: implications for telomere capping and telomerase regulation. *Genes & development* **21**, 292-302, doi:10.1101/gad.400907 (2007).
- 300 Shima, H., Suzuki, M. & Shinohara, M. Isolation and characterization of novel xrs2 mutations in *Saccharomyces cerevisiae*. *Genetics* **170**, 71-85, doi:10.1534/genetics.104.037580 (2005).

- 301 Sabourin, M., Tuzon, C. T. & Zakian, V. A. Telomerase and Tel1p preferentially associate
with short telomeres in *S. cerevisiae*. *Molecular cell* **27**, 550-561,
doi:10.1016/j.molcel.2007.07.016 (2007).
- 302 Goudsouzian, L. K., Tuzon, C. T. & Zakian, V. A. *S. cerevisiae* Tel1p and Mre11p are required
for normal levels of Est1p and Est2p telomere association. *Molecular cell* **24**, 603-610,
doi:10.1016/j.molcel.2006.10.005 (2006).
- 303 Tseng, S. F., Lin, J. J. & Teng, S. C. The telomerase-recruitment domain of the telomere
binding protein Cdc13 is regulated by Mec1p/Tel1p-dependent phosphorylation. *Nucleic
acids research* **34**, 6327-6336, doi:10.1093/nar/gkl786 (2006).
- 304 Linger, B. R. & Price, C. M. Conservation of telomere protein complexes: shuffling through
evolution. *Critical reviews in biochemistry and molecular biology* **44**, 434-446,
doi:10.3109/10409230903307329 (2009).
- 305 Mitton-Fry, R. M., Anderson, E. M., Hughes, T. R., Lundblad, V. & Wuttke, D. S. Conserved
structure for single-stranded telomeric DNA recognition. *Science* **296**, 145-147,
doi:10.1126/science.1068799 (2002).
- 306 Gao, H., Cervantes, R. B., Mandell, E. K., Otero, J. H. & Lundblad, V. RPA-like proteins
mediate yeast telomere function. *Nature structural & molecular biology* **14**, 208-214,
doi:10.1038/nsmb1205 (2007).
- 307 Sun, J. *et al.* Stn1-Ten1 is an Rpa2-Rpa3-like complex at telomeres. *Genes & development*
23, 2900-2914, doi:10.1101/gad.1851909 (2009).
- 308 Wold, M. S. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein
required for eukaryotic DNA metabolism. *Annual review of biochemistry* **66**, 61-92,
doi:10.1146/annurev.biochem.66.1.61 (1997).
- 309 Bochkareva, E., Korolev, S., Lees-Miller, S. P. & Bochkarev, A. Structure of the RPA
trimerization core and its role in the multistep DNA-binding mechanism of RPA. *The EMBO
journal* **21**, 1855-1863, doi:10.1093/emboj/21.7.1855 (2002).
- 310 Gelinas, A. D. *et al.* Telomere capping proteins are structurally related to RPA with an
additional telomere-specific domain. *Proceedings of the National Academy of Sciences of
the United States of America* **106**, 19298-19303, doi:10.1073/pnas.0909203106 (2009).
- 311 Fanning, E., Klimovich, V. & Nager, A. R. A dynamic model for replication protein A (RPA)
function in DNA processing pathways. *Nucleic acids research* **34**, 4126-4137,
doi:10.1093/nar/gkl550 (2006).
- 312 Martin, V., Du, L. L., Rozenzhak, S. & Russell, P. Protection of telomeres by a conserved
Stn1-Ten1 complex. *Proceedings of the National Academy of Sciences of the United States
of America* **104**, 14038-14043, doi:10.1073/pnas.0705497104 (2007).
- 313 Surovtseva, Y. V. *et al.* Conserved telomere maintenance component 1 interacts with STN1
and maintains chromosome ends in higher eukaryotes. *Molecular cell* **36**, 207-218,
doi:10.1016/j.molcel.2009.09.017 (2009).
- 314 Goulian, M., Heard, C. J. & Grimm, S. L. Purification and properties of an accessory protein
for DNA polymerase alpha/primase. *The Journal of biological chemistry* **265**, 13221-13230
(1990).
- 315 Goulian, M. & Heard, C. J. The mechanism of action of an accessory protein for DNA
polymerase alpha/primase. *The Journal of biological chemistry* **265**, 13231-13239 (1990).
- 316 Casteel, D. E. *et al.* A DNA polymerase- α primase cofactor with homology
to replication protein A-32 regulates DNA replication in mammalian cells. *The Journal of
biological chemistry* **284**, 5807-5818, doi:10.1074/jbc.M807593200 (2009).
- 317 Zhao, Y. *et al.* Telomere extension occurs at most chromosome ends and is uncoupled
from fill-in in human cancer cells. *Cell* **138**, 463-475, doi:10.1016/j.cell.2009.05.026 (2009).

- 318 Moser, B. A. *et al.* Differential arrival of leading and lagging strand DNA polymerases at
fission yeast telomeres. *The EMBO journal* **28**, 810-820, doi:10.1038/emboj.2009.31
(2009).
- 319 Zhu, W. *et al.* Mcm10 and And-1/CTF4 recruit DNA polymerase alpha to chromatin for
initiation of DNA replication. *Genes & development* **21**, 2288-2299,
doi:10.1101/gad.1585607 (2007).
- 320 Gambus, A. *et al.* A key role for Ctf4 in coupling the MCM2-7 helicase to DNA polymerase
alpha within the eukaryotic replisome. *The EMBO journal* **28**, 2992-3004,
doi:10.1038/emboj.2009.226 (2009).
- 321 Chattopadhyay, S. & Bielinsky, A. K. Human Mcm10 regulates the catalytic subunit of DNA
polymerase-alpha and prevents DNA damage during replication. *Molecular biology of the
cell* **18**, 4085-4095, doi:10.1091/mbc.E06-12-1148 (2007).
- 322 Walne, A. J. *et al.* Mutations in the telomere capping complex in bone marrow failure and
related syndromes. *Haematologica* **98**, 334-338, doi:10.3324/haematol.2012.071068
(2013).
- 323 Keller, R. B. *et al.* CTC1 Mutations in a patient with dyskeratosis congenita. *Pediatric blood
& cancer* **59**, 311-314, doi:10.1002/pbc.24193 (2012).
- 324 Anderson, B. H. *et al.* Mutations in CTC1, encoding conserved telomere maintenance
component 1, cause Coats plus. *Nature genetics* **44**, 338-342, doi:10.1038/ng.1084 (2012).
- 325 Chen, L. Y., Majerska, J. & Lingner, J. Molecular basis of telomere syndrome caused by
CTC1 mutations. *Genes & development* **27**, 2099-2108, doi:10.1101/gad.222893.113
(2013).
- 326 Cooper, J. P., Nimmo, E. R., Allshire, R. C. & Cech, T. R. Regulation of telomere length and
function by a Myb-domain protein in fission yeast. *Nature* **385**, 744-747,
doi:10.1038/385744a0 (1997).
- 327 Longtine, M. S., Wilson, N. M., Petracek, M. E. & Berman, J. A yeast telomere binding
activity binds to two related telomere sequence motifs and is indistinguishable from RAP1.
Current genetics **16**, 225-239 (1989).
- 328 Lustig, A. J., Kurtz, S. & Shore, D. Involvement of the silencer and UAS binding protein RAP1
in regulation of telomere length. *Science* **250**, 549-553 (1990).
- 329 Conrad, M. N., Wright, J. H., Wolf, A. J. & Zakian, V. A. RAP1 protein interacts with yeast
telomeres in vivo: overproduction alters telomere structure and decreases chromosome
stability. *Cell* **63**, 739-750 (1990).
- 330 Ancelin, K., Brun, C. & Gilson, E. Role of the telomeric DNA-binding protein TRF2 in the
stability of human chromosome ends. *BioEssays : news and reviews in molecular, cellular
and developmental biology* **20**, 879-883, doi:10.1002/(SICI)1521-
1878(199811)20:11<879::AID-BIES2>3.0.CO;2-I (1998).
- 331 Taylor, H. O., O'Reilly, M., Leslie, A. G. & Rhodes, D. How the multifunctional yeast Rap1p
discriminates between DNA target sites: a crystallographic analysis. *Journal of molecular
biology* **303**, 693-707, doi:10.1006/jmbi.2000.4161 (2000).
- 332 Nishikawa, T. *et al.* Structure of the DNA-binding domain of human telomeric protein, TRF1
and its interaction with telomeric DNA. *Nucleic acids research. Supplement*, 273-274
(2001).
- 333 Nishikawa, T., Nagadoi, A., Yoshimura, S., Aimoto, S. & Nishimura, Y. Solution structure of
the DNA-binding domain of human telomeric protein, hTRF1. *Structure* **6**, 1057-1065
(1998).
- 334 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* **85**, 125-136 (1996).

- 335 Konig, P., Fairall, L. & Rhodes, D. Sequence-specific DNA recognition by the myb-like domain of the human telomere binding protein TRF1: a model for the protein-DNA complex. *Nucleic acids research* **26**, 1731-1740 (1998).
- 336 Kissinger, C. R., Liu, B. S., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell* **63**, 579-590 (1990).
- 337 Konig, P. & Rhodes, D. Recognition of telomeric DNA. *Trends in biochemical sciences* **22**, 43-47 (1997).
- 338 Billaud, T. *et al.* The telobox, a Myb-related telomeric DNA binding motif found in proteins from yeast, plants and human. *Nucleic acids research* **24**, 1294-1303 (1996).
- 339 Price, C. M. & Cech, T. R. Telomeric DNA-protein interactions of Oxytricha macronuclear DNA. *Genes & development* **1**, 783-793 (1987).
- 340 Gottschling, D. E. & Zakian, V. A. Telomere proteins: specific recognition and protection of the natural termini of Oxytricha macronuclear DNA. *Cell* **47**, 195-205 (1986).
- 341 Peersen, O. B., Ruggles, J. A. & Schultz, S. C. Dimeric structure of the Oxytricha nova telomere end-binding protein alpha-subunit bound to ssDNA. *Nature structural biology* **9**, 182-187, doi:10.1038/nsb761 (2002).
- 342 Fang, G. & Cech, T. R. Oxytricha telomere-binding protein: DNA-dependent dimerization of the alpha and beta subunits. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 6056-6060 (1993).
- 343 Froelich-Ammon, S. J., Dickinson, B. A., Bevilacqua, J. M., Schultz, S. C. & Cech, T. R. Modulation of telomerase activity by telomere DNA-binding proteins in Oxytricha. *Genes & development* **12**, 1504-1514 (1998).
- 344 Mitton-Fry, R. M., Anderson, E. M., Theobald, D. L., Glustrom, L. W. & Wuttke, D. S. Structural basis for telomeric single-stranded DNA recognition by yeast Cdc13. *Journal of molecular biology* **338**, 241-255, doi:10.1016/j.jmb.2004.01.063 (2004).
- 345 Lei, M., Podell, E. R., Baumann, P. & Cech, T. R. DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA. *Nature* **426**, 198-203, doi:10.1038/nature02092 (2003).
- 346 Iftode, C., Daniely, Y. & Borowiec, J. A. Replication protein A (RPA): the eukaryotic SSB. *Critical reviews in biochemistry and molecular biology* **34**, 141-180, doi:10.1080/10409239991209255 (1999).
- 347 Chen, L. Y. & Lingner, J. CST for the grand finale of telomere replication. *Nucleus* **4**, 277-282, doi:10.4161/nucl.25701 (2013).
- 348 Lue, N. F., Chan, J., Wright, W. E. & Hurwitz, J. The CDC13-STN1-TEN1 complex stimulates Pol alpha activity by promoting RNA priming and primase-to-polymerase switch. *Nature communications* **5**, 5762, doi:10.1038/ncomms6762 (2014).
- 349 Huang, C., Dai, X. & Chai, W. Human Stn1 protects telomere integrity by promoting efficient lagging-strand synthesis at telomeres and mediating C-strand fill-in. *Cell research* **22**, 1681-1695, doi:10.1038/cr.2012.132 (2012).
- 350 Walther, A. P., Gomes, X. V., Lao, Y., Lee, C. G. & Wold, M. S. Replication protein A interactions with DNA. 1. Functions of the DNA-binding and zinc-finger domains of the 70-kDa subunit. *Biochemistry* **38**, 3963-3973, doi:10.1021/bi982370u (1999).
- 351 Pfuetzner, R. A., Bochkarev, A., Frappier, L. & Edwards, A. M. Replication protein A. Characterization and crystallization of the DNA binding domain. *The Journal of biological chemistry* **272**, 430-434 (1997).
- 352 Blackwell, L. J. & Borowiec, J. A. Human replication protein A binds single-stranded DNA in two distinct complexes. *Molecular and cellular biology* **14**, 3993-4001 (1994).

- 353 Philipova, D. *et al.* A hierarchy of SSB protomers in replication protein A. *Genes & development* **10**, 2222-2233 (1996).
- 354 Kim, C., Snyder, R. O. & Wold, M. S. Binding properties of replication protein A from human and yeast cells. *Molecular and cellular biology* **12**, 3050-3059 (1992).
- 355 Gomes, X. V., Henriksen, L. A. & Wold, M. S. Proteolytic mapping of human replication protein A: evidence for multiple structural domains and a conformational change upon interaction with single-stranded DNA. *Biochemistry* **35**, 5586-5595, doi:10.1021/bi9526995 (1996).
- 356 Bochkareva, E., Korolev, S. & Bochkarev, A. The role for zinc in replication protein A. *The Journal of biological chemistry* **275**, 27332-27338, doi:10.1074/jbc.M000620200 (2000).
- 357 Bochkareva, E., Frappier, L., Edwards, A. M. & Bochkarev, A. The RPA32 subunit of human replication protein A contains a single-stranded DNA-binding domain. *The Journal of biological chemistry* **273**, 3932-3936 (1998).
- 358 Blackwell, L. J., Borowiec, J. A. & Mastrangelo, I. A. Single-stranded-DNA binding alters human replication protein A structure and facilitates interaction with DNA-dependent protein kinase. *Molecular and cellular biology* **16**, 4798-4807 (1996).
- 359 Bastin-Shanower, S. A. & Brill, S. J. Functional analysis of the four DNA binding domains of replication protein A. The role of RPA2 in ssDNA binding. *The Journal of biological chemistry* **276**, 36446-36453, doi:10.1074/jbc.M104386200 (2001).
- 360 Bochkarev, A., Pfuetzner, R. A., Edwards, A. M. & Frappier, L. Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature* **385**, 176-181, doi:10.1038/385176a0 (1997).
- 361 Alani, E., Thresher, R., Griffith, J. D. & Kolodner, R. D. Characterization of DNA-binding and strand-exchange stimulation properties of γ -RPA, a yeast single-strand-DNA-binding protein. *Journal of molecular biology* **227**, 54-71 (1992).
- 362 Strzelecka, M., Oates, A. C. & Neugebauer, K. M. Dynamic control of Cajal body number during zebrafish embryogenesis. *Nucleus* **1**, 96-108, doi:10.4161/nucl.1.1.10680 (2010).
- 363 Pena, E., Berciano, M. T., Fernandez, R., Ojeda, J. L. & Lafarga, M. Neuronal body size correlates with the number of nucleoli and Cajal bodies, and with the organization of the splicing machinery in rat trigeminal ganglion neurons. *The Journal of comparative neurology* **430**, 250-263 (2001).
- 364 Cioce, M. & Lamond, A. I. Cajal bodies: a long history of discovery. *Annual review of cell and developmental biology* **21**, 105-131, doi:10.1146/annurev.cellbio.20.010403.103738 (2005).
- 365 Zhu, Y., Tomlinson, R. L., Lukowiak, A. A., Terns, R. M. & Terns, M. P. Telomerase RNA accumulates in Cajal bodies in human cancer cells. *Molecular biology of the cell* **15**, 81-90, doi:10.1091/mbc.E03-07-0525 (2004).
- 366 Richard, P. *et al.* A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *The EMBO journal* **22**, 4283-4293, doi:10.1093/emboj/cdg394 (2003).
- 367 Filipowicz, W. & Pogacic, V. Biogenesis of small nucleolar ribonucleoproteins. *Current opinion in cell biology* **14**, 319-327 (2002).
- 368 Carmo-Fonseca, M., Pepperkok, R., Carvalho, M. T. & Lamond, A. I. Transcription-dependent colocalization of the U1, U2, U4/U6, and U5 snRNPs in coiled bodies. *The Journal of cell biology* **117**, 1-14 (1992).
- 369 Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753-767, doi:10.1016/j.cell.2012.04.017 (2012).

- 370 Han, T. W. *et al.* Cell-free formation of RNA granules: bound RNAs identify features and
components of cellular assemblies. *Cell* **149**, 768-779, doi:10.1016/j.cell.2012.04.016
(2012).
- 371 Handwerger, K. E., Murphy, C. & Gall, J. G. Steady-state dynamics of Cajal body
components in the *Xenopus* germinal vesicle. *The Journal of cell biology* **160**, 495-504,
doi:10.1083/jcb.200212024 (2003).
- 372 Dundr, M. *et al.* In vivo kinetics of Cajal body components. *The Journal of cell biology* **164**,
831-842, doi:10.1083/jcb.200311121 (2004).
- 373 Deryusheva, S. & Gall, J. G. Dynamics of coilin in Cajal bodies of the *Xenopus* germinal
vesicle. *Proceedings of the National Academy of Sciences of the United States of America*
101, 4810-4814, doi:10.1073/pnas.0401106101 (2004).
- 374 Fischer, U., Liu, Q. & Dreyfuss, G. The SMN-SIP1 complex has an essential role in
spliceosomal snRNP biogenesis. *Cell* **90**, 1023-1029 (1997).
- 375 Pellizzoni, L., Yong, J. & Dreyfuss, G. Essential role for the SMN complex in the specificity
of snRNP assembly. *Science* **298**, 1775-1779, doi:10.1126/science.1074962 (2002).
- 376 Zhang, Z. *et al.* SMN deficiency causes tissue-specific perturbations in the repertoire of
snRNAs and widespread defects in splicing. *Cell* **133**, 585-600,
doi:10.1016/j.cell.2008.03.031 (2008).
- 377 Tapia, O. *et al.* Reorganization of Cajal bodies and nucleolar targeting of coilin in motor
neurons of type I spinal muscular atrophy. *Histochemistry and cell biology* **137**, 657-667,
doi:10.1007/s00418-012-0921-8 (2012).
- 378 Oskoui, M. *et al.* The changing natural history of spinal muscular atrophy type 1. *Neurology*
69, 1931-1936, doi:10.1212/01.wnl.0000290830.40544.b9 (2007).
- 379 Lefebvre, S. *et al.* Correlation between severity and SMN protein level in spinal muscular
atrophy. *Nature genetics* **16**, 265-269, doi:10.1038/ng0797-265 (1997).
- 380 Henriksson, S. *et al.* The scaffold protein WRAP53beta orchestrates the ubiquitin response
critical for DNA double-strand break repair. *Genes & development* **28**, 2726-2738,
doi:10.1101/gad.246546.114 (2014).
- 381 Stucki, M. *et al.* MDC1 directly binds phosphorylated histone H2AX to regulate cellular
responses to DNA double-strand breaks. *Cell* **123**, 1213-1226,
doi:10.1016/j.cell.2005.09.038 (2005).
- 382 Durocher, D. & Jackson, S. P. DNA-PK, ATM and ATR as sensors of DNA damage: variations
on a theme? *Current opinion in cell biology* **13**, 225-231 (2001).
- 383 Mailand, N. *et al.* RNF8 ubiquitylates histones at DNA double-strand breaks and promotes
assembly of repair proteins. *Cell* **131**, 887-900, doi:10.1016/j.cell.2007.09.040 (2007).
- 384 Kolas, N. K. *et al.* Orchestration of the DNA-damage response by the RNF8 ubiquitin ligase.
Science **318**, 1637-1640, doi:10.1126/science.1150034 (2007).
- 385 Huen, M. S. *et al.* RNF8 transduces the DNA-damage signal via histone ubiquitylation and
checkpoint protein assembly. *Cell* **131**, 901-914, doi:10.1016/j.cell.2007.09.041 (2007).
- 386 Jady, B. E. *et al.* Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal
body following import from the cytoplasm. *The EMBO journal* **22**, 1878-1888,
doi:10.1093/emboj/cdg187 (2003).
- 387 Darzacq, X. *et al.* Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation
and pseudouridylation guide RNAs. *The EMBO journal* **21**, 2746-2756,
doi:10.1093/emboj/21.11.2746 (2002).
- 388 Bellodi, C. *et al.* H/ACA small RNA dysfunctions in disease reveal key roles for noncoding
RNA modifications in hematopoietic stem cell differentiation. *Cell reports* **3**, 1493-1502,
doi:10.1016/j.celrep.2013.04.030 (2013).

389 Zhang, Y., Morimoto, K., Danilova, N., Zhang, B. & Lin, S. Zebrafish models for dyskeratosis
congenita reveal critical roles of p53 activation contributing to hematopoietic defects
through RNA processing. *PLoS one* **7**, e30188, doi:10.1371/journal.pone.0030188 (2012).

390 Mochizuki, Y., He, J., Kulkarni, S., Bessler, M. & Mason, P. J. Mouse dyskerin mutations
affect accumulation of telomerase RNA and small nucleolar RNA, telomerase activity, and
ribosomal RNA processing. *Proceedings of the National Academy of Sciences of the United
States of America* **101**, 10756-10761, doi:10.1073/pnas.0402560101 (2004).

391 Ruggero, D. *et al.* Dyskeratosis congenita and cancer in mice deficient in ribosomal RNA
modification. *Science* **299**, 259-262, doi:10.1126/science.1079447 (2003).

392 Freund, A. *et al.* Proteostatic control of telomerase function through TRiC-mediated
folding of TCAB1. *Cell* **159**, 1389-1403, doi:10.1016/j.cell.2014.10.059 (2014).

393 Olsen, J. V. *et al.* Quantitative phosphoproteomics reveals widespread full
phosphorylation site occupancy during mitosis. *Science signaling* **3**, ra3,
doi:10.1126/scisignal.2000475 (2010).

394 Olsen, J. V. *et al.* Global, in vivo, and site-specific phosphorylation dynamics in signaling
networks. *Cell* **127**, 635-648, doi:10.1016/j.cell.2006.09.026 (2006).

395 Stirnimann, C. U., Petsalaki, E., Russell, R. B. & Muller, C. W. WD40 proteins propel cellular
networks. *Trends in biochemical sciences* **35**, 565-574, doi:10.1016/j.tibs.2010.04.003
(2010).

396 Yu, L., Gaitatzes, C., Neer, E. & Smith, T. F. Thirty-plus functional families from a single
motif. *Protein science : a publication of the Protein Society* **9**, 2470-2476,
doi:10.1110/ps.9.12.2470 (2000).

397 Yu, H. *et al.* High-quality binary protein interaction map of the yeast interactome network.
Science **322**, 104-110, doi:10.1126/science.1158684 (2008).

398 Collins, S. R. *et al.* Toward a comprehensive atlas of the physical interactome of
Saccharomyces cerevisiae. *Molecular & cellular proteomics : MCP* **6**, 439-450,
doi:10.1074/mcp.M600381-MCP200 (2007).

399 Rual, J. F. *et al.* Towards a proteome-scale map of the human protein-protein interaction
network. *Nature* **437**, 1173-1178, doi:10.1038/nature04209 (2005).

400 Stelzl, U. *et al.* A human protein-protein interaction network: a resource for annotating
the proteome. *Cell* **122**, 957-968, doi:10.1016/j.cell.2005.08.029 (2005).

401 Fong, H. K. *et al.* Repetitive segmental structure of the transducin beta subunit: homology
with the CDC4 gene and identification of related mRNAs. *Proceedings of the National
Academy of Sciences of the United States of America* **83**, 2162-2166 (1986).

402 Sondek, J., Bohm, A., Lambright, D. G., Hamm, H. E. & Sigler, P. B. Crystal structure of a G-
protein beta gamma dimer at 2.1A resolution. *Nature* **379**, 369-374,
doi:10.1038/379369a0 (1996).

403 Wall, M. A. *et al.* The structure of the G protein heterotrimer Gi alpha 1 beta 1 gamma 2.
Cell **83**, 1047-1058 (1995).

404 Smith, T. F., Gaitatzes, C., Saxena, K. & Neer, E. J. The WD repeat: a common architecture
for diverse functions. *Trends in biochemical sciences* **24**, 181-185 (1999).

405 Neer, E. J., Schmidt, C. J., Nambudripad, R. & Smith, T. F. The ancient regulatory-protein
family of WD-repeat proteins. *Nature* **371**, 297-300, doi:10.1038/371297a0 (1994).

406 Schuster-Bockler, B., Schultz, J. & Rahmann, S. HMM Logos for visualization of protein
families. *BMC bioinformatics* **5**, 7, doi:10.1186/1471-2105-5-7 (2004).

407 Paoli, M. Protein folds propelled by diversity. *Progress in biophysics and molecular biology*
76, 103-130 (2001).

- 408 Murzin, A. G. Structural principles for the propeller assembly of beta-sheets: the
preference for seven-fold symmetry. *Proteins* **14**, 191-201, doi:10.1002/prot.340140206
(1992).
- 409 Bieniossek, C., Richmond, T. J. & Berger, I. MultiBac: multigene baculovirus-based
eukaryotic protein complex production. *Current protocols in protein science* **Chapter 5**,
Unit 5 20, doi:10.1002/0471140864.ps0520s51 (2008).
- 410 Sambrook, J. & Russell, D. W. The inoue method for preparation and transformation of
competent e. Coli: "ultra-competent" cells. *CSH protocols* **2006**,
doi:10.1101/pdb.prot3944 (2006).
- 411 Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy.
Journal of structural biology **157**, 38-46, doi:10.1016/j.jsb.2006.05.009 (2007).
- 412 Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure
determination with parallelisation using GPUs in RELION-2. *eLife* **5**,
doi:10.7554/eLife.18722 (2016).
- 413 Dinkel, H. *et al.* ELM--the database of eukaryotic linear motifs. *Nucleic acids research* **40**,
D242-251, doi:10.1093/nar/gkr1064 (2012).
- 414 Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P. & Bork, P. SMART: a web-based tool for
the study of genetically mobile domains. *Nucleic acids research* **28**, 231-234 (2000).
- 415 Bateman, A. *et al.* The Pfam protein families database. *Nucleic acids research* **28**, 263-266
(2000).
- 416 Linding, R., Russell, R. B., Neduva, V. & Gibson, T. J. GlobPlot: Exploring protein sequences
for globularity and disorder. *Nucleic acids research* **31**, 3701-3708 (2003).
- 417 Dosztanyi, Z., Csizmek, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of
intrinsically unstructured regions of proteins based on estimated energy content.
Bioinformatics **21**, 3433-3434, doi:10.1093/bioinformatics/bti541 (2005).
- 418 van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the
IMAGIC image processing system. *Journal of structural biology* **116**, 17-24,
doi:10.1006/jsbi.1996.0004 (1996).
- 419 Penczek, P., Radermacher, M. & Frank, J. Three-dimensional reconstruction of single
particles embedded in ice. *Ultramicroscopy* **40**, 33-53 (1992).
- 420 Penczek, P. A. Resolution measures in molecular electron microscopy. *Methods in
enzymology* **482**, 73-100, doi:10.1016/S0076-6879(10)82003-8 (2010).
- 421 Danev, R., Okawara, H., Usuda, N., Kametani, K. & Nagayama, K. A Novel Phase-contrast
Transmission Electron Microscopy Producing High-contrast Topographic Images of Weak
objects. *Journal of biological physics* **28**, 627-635, doi:10.1023/A:1021234621466 (2002).
- 422 Danev, R. & Nagayama, K. Transmission electron microscopy with Zernike phase plate.
Ultramicroscopy **88**, 243-252 (2001).
- 423 Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version
2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-
1191, doi:10.1093/bioinformatics/btp033 (2009).
- 424 Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948,
doi:10.1093/bioinformatics/btm404 (2007).
- 425 Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure
prediction server. *Nucleic acids research* **43**, W389-394, doi:10.1093/nar/gkv332 (2015).
- 426 Soding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology
detection and structure prediction. *Nucleic acids research* **33**, W244-248,
doi:10.1093/nar/gki408 (2005).

- 427 Frydman, J. *et al.* Function in protein folding of TRiC, a cytosolic ring complex containing
TCP-1 and structurally related subunits. *The EMBO journal* **11**, 4767-4778 (1992).
- 428 Kubota, H., Hynes, G. & Willison, K. The eighth Cct gene, Cctq, encoding the theta subunit
of the cytosolic chaperonin containing TCP-1. *Gene* **154**, 231-236 (1995).
- 429 Xu, Z., Horwich, A. L. & Sigler, P. B. The crystal structure of the asymmetric GroEL-GroES-
(ADP)7 chaperonin complex. *Nature* **388**, 741-750, doi:10.1038/41944 (1997).
- 430 Spiess, C., Meyer, A. S., Reissmann, S. & Frydman, J. Mechanism of the eukaryotic
chaperonin: protein folding in the chamber of secrets. *Trends in cell biology* **14**, 598-604,
doi:10.1016/j.tcb.2004.09.015 (2004).
- 431 Yam, A. Y. *et al.* Defining the TRiC/CCT interactome links chaperonin function to
stabilization of newly made proteins with complex topologies. *Nature structural &
molecular biology* **15**, 1255-1262, doi:10.1038/nsmb.1515 (2008).
- 432 Miyata, Y., Shibata, T., Aoshima, M., Tsubata, T. & Nishida, E. The molecular chaperone
TRiC/CCT binds to the Trp-Asp 40 (WD40) repeat protein WDR68 and promotes its folding,
protein kinase DYRK1A binding, and nuclear accumulation. *The Journal of biological
chemistry* **289**, 33320-33332, doi:10.1074/jbc.M114.586115 (2014).
- 433 Kubota, S., Kubota, H. & Nagata, K. Cytosolic chaperonin protects folding intermediates of
Gbeta from aggregation by recognizing hydrophobic beta-strands. *Proceedings of the
National Academy of Sciences of the United States of America* **103**, 8360-8365,
doi:10.1073/pnas.0600195103 (2006).
- 434 Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J. M. & Baumeister, W. Volta potential phase
plate for in-focus phase contrast transmission electron microscopy. *Proceedings of the
National Academy of Sciences of the United States of America* **111**, 15635-15640,
doi:10.1073/pnas.1418377111 (2014).
- 435 Khoshouei, M. *et al.* Volta phase plate cryo-EM of the small protein complex Prx3. *Nature
communications* **7**, 10534, doi:10.1038/ncomms10534 (2016).
- 436 Danev, R., Tegunov, D. & Baumeister, W. Using the Volta phase plate with defocus for cryo-
EM single particle analysis. *eLife* **6**, doi:10.7554/eLife.23006 (2017).
- 437 Danev, R. & Baumeister, W. Cryo-EM single particle analysis with the Volta phase plate.
eLife **5**, doi:10.7554/eLife.13046 (2016).
- 438 Chua, E. Y. *et al.* 3.9 Å structure of the nucleosome core particle determined by phase-
plate cryo-EM. *Nucleic acids research* **44**, 8013-8019, doi:10.1093/nar/gkw708 (2016).
- 439 Khoshouei, M., Radjainia, M., Baumeister, W. & Danev, R. Cryo-EM structure of
haemoglobin at 3.2 Å determined with the Volta phase plate. *Nature communications* **8**,
16099, doi:10.1038/ncomms16099 (2017).