Biochemical and structural analysis of shelterin subcomplexes

Inian, Oviya

2018

Inian, O. (2018). Biochemical and structural analysis of shelterin subcomplexes. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/75683>

<https://doi.org/10.32657/10356/75683>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**BIOCHEMICAL AND STRUCTURAL ANALYSIS OF
SHELTERIN SUBCOMPLEXES**

OVIYA INIAN

SCHOOL OF BIOLOGICAL SCIENCES

2018

BIOCHEMICAL AND STRUCTURAL ANALYSIS OF SHELTERIN SUBCOMPLEXES

OVIYA INIAN

SCHOOL OF BIOLOGICAL SCIENCES
NANYANG TECHNOLOGICAL UNIVERSITY

2018

A thesis submitted to Nanyang Technological University
for the degree of Doctor of Philosophy

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except where explicitly stated.

Oviya Inian

Acknowledgements

I would like to express sincere gratitude to my supervisor Dr Sara Sandin for providing me support, freedom and opportunity to pursue this study. Her constructive feedback has made me a better researcher and a critical thinker.

I am profoundly thankful to my co-supervisor Dr Daniela Rhodes for allowing me to work in her laboratory, providing useful ideas and valuable suggestions.

I would like to convey my gratitude to Daniela's group and MRC laboratory of molecular biology in Cambridge for the protein expression vectors. I am particularly grateful to Lynda Chapman and Louise Fairall for their pioneer work on shelterin proteins, which is the foundation on which this study is built.

I am grateful to the NTU protein production platform and mass spectrometry facility for the assistance with cloning and mass spectrometry analysis respectively.

My warm thanks to Dr Vinod K. Vogirala for training me when I was a novice, for being a patient mentor and a good friend. I am grateful to Dr Vasuki R. Chellamuthu for guiding me in the final years of my PhD and for critically reviewing this thesis. Her optimism and diligence have been a great inspiration to me.

I am thankful to all the members of Sara's and Daniela's groups. I am especially grateful for the cheerful company of Barbara Hübner, Satabdi Mishra, Dr Eugene Chua, Eric von Otter, Dr Rahul Jaiswal and Harvey Chan. I will cherish the entertaining discussions, encouragement and their scientific insight. I would like to express heartfelt gratitude to my friends. Kritika Shankar for being a constant support. My fellow PhD pals Uthra and Aretha for giving me company during the long writing sessions.

Finally, I would like to thank my parents and sister for their love, encouragement and unwavering support to pursue my dream. I dedicate this thesis to my family.

Table of Contents

Acknowledgements.....	III
List of figures	X
List of tables	XII
List of abbreviations	XIII
Abstract.....	XV
Chapter 1: Introduction.....	1
1.1 Telomeres.....	1
1.1.1 Introduction and historical perspective.....	1
1.1.2 Telomeric chromatin	3
1.1.3 Shelterin	4
1.1.4 Functions of shelterin.....	5
1.1.4.1 Telomere length homeostasis and telomerase recruitment	6
1.1.4.2 Chromosome end protection and T-loop model.....	9
1.1.5 Subunits of shelterin	12
1.1.5.1 Double-stranded DNA binding shelterin proteins.....	12
1.1.5.2 Single-stranded DNA binding shelterin proteins	15
1.1.3.4 TIN2.....	17
1.1.6 Telomeres, human diseases, and cancer	18
1.1.7 Higher-order structure of telomeres and shelterin subcomplexes.....	20
1.2 Electron microscopy	26
1.2.1 Overview and recent developments.....	26
1.2.2 Single particle analysis	28
1.2.2.1 Sample preparation	28
1.2.2.2 Contrast in EM.....	30
1.2.2.3 Data collection and processing.....	30

1.2.3 Phase-plate.....	33
1.2.4.1 Volta phase plate	34
1.3 Aims of this dissertation	36
Chapter 2: Materials and Methods	37
2.1 Materials	37
2.2 Methods	37
2.2.1 Expression and purification of TRF1	39
2.2.1.1 Transformation	40
2.2.1.2 Bacterial protein expression and cell harvest	40
2.2.1.3 Purification of TRF1	40
2.2.2 Expression and purification of TIN2	42
2.2.2.1 MBP-TIN2L cloning	42
2.2.2.2 Transformation	42
2.2.2.3 Bacterial protein expression and cell harvest	43
2.2.2.4 Purification of TIN2L and MBP-TIN2L.....	43
2.2.2.5 Purification of GST-TIN2S	44
2.2.3 Expression and purification of TRF2-RAP1-TIN2S	44
2.2.3.1 Transformation	44
2.2.3.2 Bacterial expression of proteins and cell harvest.....	44
2.2.3.3 Purification of TRF2-RAP1-TIN2S	45
2.2.4 MultiBac expression and purification of POT1-TPP1-TIN2S.....	46
2.2.4.1 Insect cell culture	47
2.2.4.2 Bacmid production	47
2.2.4.3 Transfection and initial virus (V_0) generation	48
2.2.4.4 V_1 generation and test purification	48
2.2.4.5 Viral titre determination.....	49
2.2.4.6 Scale-up and expression	50

2.2.4.7 Purification of POT1-TPP1-TIN2S	50
2.2.5 Biochemical studies	51
2.2.5.1 DNA binding studies	51
2.2.5.2 Crosslinking studies.....	53
2.2.5.3 Stoichiometry analysis	54
2.2.6 Negative staining	55
2.2.6.1 Sample preparation	55
2.2.6.2 Grid preparation.....	55
2.2.6.3 Data Collection	55
2.2.7 Cryo-electron microscopy	56
2.2.7.1 Sample preparation	56
2.2.7.2 Grid preparation.....	60
2.2.8 Data Processing	62
2.2.8.1 Micrograph screening and particle picking.....	62
2.2.8.2 CTF determination and particle extraction	62
2.2.8.3 2D Classification.....	62
2.2.8.4 3D Classification and refinement	63
Chapter 3: Expression, purification and negative staining of TRF1	65
3.1 Overview.....	65
3.2 Expression and purification.....	66
3.3 DNA binding studies	68
3.4 Negative stain EM analysis of TRF1 and telomeric DNA complex.....	70
3.4.1 Sample preparation	70
3.4.2 Initial 2D classification	73
3.5 Summary	73
Chapter 4: Expression, purification and negative stain EM analysis of TRF2- RAP1-TIN2S	75

4.1 Overview.....	75
4.2 Bioinformatics analysis of TIN2.....	76
4.3 TIN2 constructs.....	78
4.4 Purification of TRF2-RAP1-TIN2S	81
4.5 Analytical size-exclusion study	86
4.6 TRF2-RAP1-TIN2S stoichiometry analysis	89
4.7 Negative stain EM analysis of TRF2-RAP1-TIN2S	91
4.7.1 Data collection and 2D classification	92
4.7.2 3D classification and refinement.....	93
4.8 Summary	94
Chapter 5: Purification and negative stain EM analysis of POT1-TPP1-TIN2 complex.....	96
5.1 Overview.....	96
5.2 Purification of POT1-TPP1-TIN2S	97
5.2.1 Expression of POT1-TPP1-TIN2S	97
5.2.2 Viral titre determination and large-scale expression	100
5.2.3 Purification of TPP1- POT1 -TIN2S (Construct 9).....	100
5.2.4 Purification of TPP1- POT1 -TIN2L (Construct 7).....	105
5.3 Biochemical studies of POT1-TPP1-TIN2S	106
5.3.1 DNA binding assay	106
5.3.2 Crosslinking study.....	107
5.4 POT1-TPP1-TIN2S stoichiometry analysis	108
5.5 Negative staining analysis of POT1-TPP1-TIN2S.....	110
5.5.1 Data collection and 2D classification	110
5.5.2 3D Classification and Refinement.....	112
5.5.3 Docking of EM density map using model derived from SAXS	114
5.6 Summary	116

Chapter 6: Cryo-electron microscopy and Volta phase plate imaging.....	118
6.1 Overview.....	118
6.2 Phase plate conditioning and imaging	119
6.3 Imaging DNA and NCP with Volta phase plate	120
6.4 POT1-TPP1-TIN2S screening	123
6.5 Challenges faced in cryo-EM imaging of shelterin subcomplexes	124
6.6 Summary	126
Chapter 7: Discussion and future directions	127
7.1 Negative staining of TRF1	127
7.2 TIN2S can be copurified with TRF2 RAP1	129
7.3 TRF2-RAP1-TIN2S can form a complex with TRF1	129
7.4 POT1-TPP1-TIN2S form a stable complex.....	130
7.5 Stoichiometry of shelterin subcomplexes.....	131
7.6 TRF2-RAP1-TIN2S negative staining structure	132
7.7 POT1-TPP1-TIN2S negative staining structure	133
7.8 Glutaraldehyde crosslinking affects the DNA binding activity of shelterin proteins.....	134
7.9 Cryo-electron microscopy of shelterin subcomplexes.....	137
7.10 Nucleic acid and Protein-DNA complex imaging enhanced by Volta phase plate.....	138
7.11 Conclusions and future directions.....	140
References	141
Appendix.....	169
A1: Media and gel recipes	169
A2: Protein constructs.....	170
A3: Bacmids and plasmids.....	172
A4: Mass spectrometry	173
A5: Publication	174

List of figures

Figure 1.1: Shelterin – the mammalian telomere capping complex.....	4
Figure 1.2: Role of shelterin in telomere maintenance	6
Figure 1.3: Structure of human telomerase	7
Figure 1.4: Shelterin safeguards telomeres.....	10
Figure 1.5: Telomeric T-loop model.....	12
Figure 1.6: Structure of conserved domains in TRF proteins	13
Figure 1.7: Structure of conserved domains in hPOT1	15
Figure 1.8: Structure of TPP1 OB domain.....	16
Figure 1.9: Mutation hotspot in TIN2	19
Figure 1.10: Electron microscopy examination of telomeric chromatin	21
Figure 1.11: Summary of shelterin domains and interactions.....	25
Figure 1.12: Structure determination by EM.....	32
Figure 1.13: Schematic diagram of Volta-phase plate (VPP).	35
Figure 2.1: Schematic for NiNTA purification of C9	51
Figure 3.1: Summary of TRF1 purification. TRF1.....	66
Figure 3.2: TRF1 size-exclusion analysis.....	67
Figure 3.3: TRF1 DNA binding analysis.	69
Figure 3.4: Negative staining EM of TRF1 alone.....	70
Figure 3.5: Negative stain EM of TRF1 in complex with 4 telomeric repeats	71
Figure 3.6: TRF1 DNA binding and GA crosslinking	72
Figure 3.7: 2D classification of TRF1 - 4 repeat telomeric DNA complex.....	73
Figure 4.1: Structure prediction of TIN2.	77
Figure 4.2: Unsuccessful purification attempts of TIN2 constructs.	79
Figure 4.3: Size-exclusion chromatography of MBP-TIN2L.....	80
Figure 4.4: Purification of GST-TIN2S.....	81
Figure 4.5: Test expression of TRF2/RAP1/TRF1/TIN2S.....	82
Figure 4.6: NiNTA purification and removal of His-tag in TRF2-RAP1-TIN2S.....	84
Figure 4.7: Size-exclusion purification of TRF2-RAP1-TIN2S	85
Figure 4.8: Analytical size-exclusion study using TRF2-RAP1-TIN2S complex	87
Figure 4.9: Stoichiometry of TRF2-RAP1-TIN2S.....	90
Figure 4.10: Negative staining of TRF2-RAP1-TIN2S	91

Figure 4.11: 2D classification of TRF2-RAP1-TIN2S.....	93
Figure 4.12: Negative stain EM structure of TRF2-RAP1-TIN2S.....	94
Figure 5.1: Monitoring infection cycle using YFP.	98
Figure 5.2: Test expression of baculoviral constructs.....	99
Figure 5.3 Initial purification of construct C9.	101
Figure 5.4: NiNTA purification of TPP1- POT1 -TIN2S.....	103
Figure 5.5: Size-exclusion of TPP1-POT1-TIN2S.	104
Figure 5.6: Purification of TPP1- POT1 -TIN2L.....	105
Figure 5.7: DNA binding activity of POT1-TPP1-TIN2S	106
Figure 5.8: Crosslinking study of POT1-TPP1-TIN2S bound to hT10 DNA.....	107
Figure 5.9: Stoichiometry of POT1-TPP1-TIN2S.....	109
Figure 5.10: Estimation of CTF by CTFFIND3.	111
Figure 5.11: 2D classification of POT1-TPP1-TIN2S.....	112
Figure 5.12: 3D classification of the POT1-TPP1-TIN2S complex using RELION.	113
Figure 5.13: 3D structure of POT1-TPP1-TIN2S by negative stain EM.....	114
Figure 5.14: Comparison between POT1-TPP1-TIN2S and POT1-TPP1.	116
Figure 6.1: Optimal conditioning of VPP.....	119
Figure 6.2: Imaging Widom 601 DNA sequence with VPP	120
Figure 6.3: Imaging Holliday junctions with VPP	121
Figure 6.4: Imaging Nucleosome core particle (NCP) with VPP.....	122
Figure 6.5: Phase plate activation and imaging.....	123
Figure 6.6: Imaging POT1-TPP1-TIN2S using VPP	124
Figure 6.7: Aggregation of shelterin subcomplexes in Cryo-EM.....	125
Figure 7.1 Comparison of structures of TRF1	128
Figure 7.2: Comparison between TRF2-RAP1 and TRF2-RAP1-TIN2S	133
Figure 7.3: Docking of the POT1-TPP1-TIN2S structure.....	134
Figure 7.4: Lysine residue is essential for TRF1-DNA binding	136
Figure 7.5: Cryo-EM structure of NCP solved to 3.9 Å utilizing VPP.....	139

List of tables

Table 1.1: Key mutations in Dyskeratosis Congenita	19
Table 1.2: Relative abundance of shelterin components per telomere.....	23
Table 2.1 List of constructs used for protein purification	38
Table 2.2 Composition of buffers used in protein purification.....	39
Table 2.3: Sequence of primers used for SLIC cloning of MBP-TIN2	42
Table 2.4: Protein expression/induction condition for TIN2 constructs.....	43
Table 2.5: Overview of POT1-TPP1/-TIN2 MultiBac constructs.	46
Table 2.6: DNA probes and sequences.....	52
Table 2.7: Conditions for EMSA	53
Table 2.8: Gluteraldehyde crosslinking reaction.....	53
Table 2.9: Concentration of samples used to prepare cryo grids	61
Table 2.10: Starting models for negative staining.	63
Table 4.1: List of all TIN2 constructs tested for purification	78
Table 4.2: Size-exclusion studies using TRF2-RAP1-TIN2S.....	86
Table 4.3: TRF2-RAP1-TIN2S negative stain-EM data collection and processing...	92
Table 5.1: Viral titre determination for V1, Construct 9.....	100
Table 5.2: POT1-TPP1-TIN2S negative stain-EM data collection and processing.	110

List of abbreviations

2R	Two-repeat telomeric DNA
4R	Four-repeat telomeric DNA
Anti-Anti	Antibiotic-Antimycotic
ATM	Ataxia-telangiectasia mutated
ATR	Ataxia-telangiectasia Rad3 related
bp	Base pair
C.V	Column volume
CMOS	Complementary metal oxide sensor
Cryo-EM	Cryo-electron microscopy
CTF	Contrast transfer function
DC	Dyskeratosis Congenita
DDR	DNA damage response
dRI	Differential refractive index
dsDNA	Double-stranded DNA
EM	Electron microscopy
EMSA	Electrophoretic mobility shift assay
FBS	Fetal bovine serum
FSC	Fourier shell correlation
FT	Flow through
GA	Gluteraldehyde
GST	Glutathione S-transferase
His	Histidine
IPTG	Isopropyl β -D-1-thiogalactopyranoside
kb	Kilobase
kDa	Kilodalton
MOI	Multiplicity of infection
MWCO	Molecular weight cut-off
NCP	Nucleosome core particle

OB	Oligonucleotide/oligosaccharide binding domain
OD	Optical density
PAGE	Poly acrylamide gel electrophoresis
PARP	Poly-ADP ribose polymerase
PCR	Polymerase chain reaction
POT1	Protector of telomeres 1
RAP1	Repressor/activator protein 1
RPA	Replication protein A
rpm	Revolutions per minute
RT	Reverse transcriptase
SAXS	Small-angle X-ray scattering
SDS	Sodium dodecyl sulphate
SEC-MALS	Size-exclusion chromatography & multi-angle scattering
SELEX	Systematic evolution of ligands by exponential enrichment
SLIC	Sequence and ligation independent cloning
SPA	Single particle analysis
ssDNA	Single-stranded DNA
TERC	Telomerase RNA component
TERT	Telomerase reverse transcriptase subunit
TEV	Tobacco Etch virus protease
TIFs	Telomere dysfunction induced foci
TIN2	TRF1 interacting nuclear factor 1
TPP1	TIN2 and POT1 interacting protein
TRF1/2	Telomeric repeat-binding factor 1/2
TRFH	TRF1/2 homodimerization domain
UA	Uranyl acetate
VPP	Volta phase plate
YFP	Yellow fluorescent protein

Abstract

Telomeric chromatin consists of tandem 'TTAGGG' DNA repeats that associate with a telomere-specific complex, known as shelterin. Shelterin has an indispensable role in telomere length homeostasis and protection against genomic instability arising from erroneous DNA damage responses.

Shelterin collectively consists of six proteins - Telomeric repeat binding factors 1, 2 (**TRF1** and **TRF2**), Protector of telomeres 1 (**POT1**), TRF1 interacting nuclear factor 1 (**TIN2**), TIN2 POT1 interacting protein (**TPP1**) and Repressor/activator protein 1 (**RAP1**). Specific mutations in shelterin proteins leading to telomere dysfunction have been reported in human diseases such as Dyskeratosis, Hoyeraal-Hreidarsson syndrome and cancer. Although structures of the domains of the individual proteins are well-studied, the intact shelterin complex and its subcomplexes remain largely uncharacterized.

Results presented in this thesis discuss the biochemical purification of the shelterin subcomplexes TRF2-RAP1-TIN2S and the telomerase recruitment complex POT1-TPP1-TIN2S. The subcomplexes are further characterized by analysis of the stoichiometry and interaction with DNA. Structures of TRF2-RAP1-TIN2S and POT1-TPP1-TIN2S obtained through negative stain electron microscopy (EM) are presented.

Cryo-EM imaging of DNA and nucleosome core particle (NCP) samples utilizing Volta phase plate is shown. Subsequent Volta phase plate imaging of POT1-TPP1-TIN2S is presented, showing promise for forthcoming structure determination using cryo-EM. Finally, approaches to obtain high-resolution structures of shelterin subcomplexes in the future are discussed.

Chapter 1: Introduction

1.1 Telomeres

1.1.1 Introduction and historical perspective

Genetic information in eukaryotes is stored in compact nucleoprotein structures called chromosomes, which are replicated and passed on during cell division. However, with each round of DNA replication, there is a loss of information at the 3' end of linear chromosomes [1, 2]. Additionally, linear chromosomes could be inherently recognized as double-strand DNA breaks and subsequently elicit erroneous DNA damage repair responses (DDR) [3]. Telomeres are tandem arrays found at the end of eukaryotic chromosomes that act as protection against progressive chromosomal attrition owing to replication and DDR to maintain chromosomal stability.

An early understanding of telomeres was derived from cytogenetics experiments performed in the 1920s by Muller and McClintock, which recognized that natural chromosome ends had specialized characteristics distinguishing them from induced chromosomal breaks [4-6]. Chromosomal breaks and subsequent reattachments or fusions were observed using light microscopy in fruit fly and maize chromosomes irradiated with X-ray. However, the 'originally free ends' or 'natural ends' did not fuse with other chromosomal breaks and were later termed as 'telomeres' [4].

Human fetal cells in culture were observed to divide 40-60 times before entering cellular senescence [7]. Telomeres were suggested to act as a buffering mechanism against chromosomal attrition, gradually prompting senescence once they reached a critical length called the 'Hayflick limit' [8]. Almost a decade later, sperms cells were observed to contain longer telomeres in comparison to mature adult cells that had

undergone multiple proliferative cycles [9]. Around the same time, a telomere reverse transcriptase enzyme now known as telomerase was discovered in *Tetrahymena* by Greider and Blackburn [10]. Telomerase was observed to elongate telomeres through the addition of precise tandem telomeric DNA repeats using a RNA template [10].

Subsequently, a telomere double-stranded DNA binding protein Rap1c was identified in the yeast *S.cerevisiae* [11], previously recognized as a transcription factor [12]. Rap1c was shown to negatively regulate the telomere length, providing an initial indication that telomere binding proteins are involved in telomere maintenance [13]. A single-stranded DNA binding protein called TEBP was identified next in the ciliate *Oxytricha* [14, 15] and was later found to have two subunits TEBP α and β with differential functions [16].

Mammalian telomeres are now known to contain a complex of at least six telomere-specific proteins, collectively called as shelterin. Identification of mammalian telomeric factors began with the detection of Telomeric repeat-binding factor 1 (TRF1) in HeLa extracts as a telomere-specific DNA binding protein [17]. A related telomere binding factor, TRF2 was identified subsequently, on the basis of sequence similarity to TRF1 [18, 19] and its binding partner Repressor/activator protein 1 (RAP1) was detected by its interaction with TRF2 [20, 21]. Single-strand DNA binding factor Protector of telomeres 1 (POT1) was identified based on sequence similarity search to its ciliate homologue TEBP α [22]. TRF1 interacting nuclear factor 1 (TIN2) [23] and TIN2 and POT1 interacting protein (TPP1) [24-26] were recognised around the same time, based on their interaction with other shelterin components and each other using pull-down assays and yeast two-hybrid screens. The following sections provide an overview of telomeric chromatin and more specifically shelterin, its components, their functions and implication in human diseases.

1.1.2 Telomeric chromatin

Telomeric DNA consists of tandem repeats that end in a G-rich 3' overhang [27, 28]. In vertebrates, the repeat sequence is TTAGGG [29]. The number of repeats varies with the organism and cell type. In humans, the typical length of the double-strand region at birth is 10 kb, and the 3' overhang ranges from 50-500 nucleotides [9, 30-32].

Due to the inherent 'G-rich' nature of telomeres, they have high propensity to form G-quadruplex structures (reviewed in [33]). G-quadruplexes are four-stranded helical RNA and DNA structures formed by Hoogsteen base pairing. These structures are formed by four or more continuous G-rich stretches, stabilized by large cations such as Na⁺ and K⁺. Evidence for such structures *in vivo* was provided by a study with G-quadruplex specific antibody in Ciliates [34].

Recently, the presence of G-quadruplexes in human chromatin was also visualized by immunofluorescence using specific antibodies. Only a fraction of telomeres contained G-quadruplex structures, suggesting it may not be the primary form of telomeric DNA *in vivo* [35]. A cell division dependent role for G-quadruplexes, particularly in the alignment of sister chromatids during meiosis has been proposed [36].

Two types of factors can associate with telomeric DNA: nucleosomes and a telomere-specific complex called shelterin. Micrococcal nuclease digestions suggests that telomeric chromatin is highly sensitive to nucleases and contain shorter nucleosome repeat length in comparison to bulk chromatin [37, 38].

Additionally, telomeric DNA is a suboptimal sequence for nucleosome formation, as it has been shown that the energy required to form nucleosomes on telomeric DNA is

highest among known natural DNA sequences [39]. Nucleosomes on telomeres are likely to be mobile due to the repetitive nature of telomeric DNA [39].

Telomeric double-strand DNA shelterin binding proteins, TRF1 and TRF2, share the same DNA substrate as telomeric nucleosomes. Two predominant models have been proposed to elucidate the co-existence of nucleosomes and shelterin at the telomeres. A concentration-dependent model where the nucleosome and telomeric proteins compete for telomeric DNA [40], and alternatively telomeric proteins binding in the context of nucleosomes and modulating the nucleosomal organisation [41].

Telomeres are estimated to contain thousands of shelterin and nucleosome molecules. How these complexes interact with one another, assemble and fold telomeric DNA inside the nucleus remains mostly unknown.

1.1.3 Shelterin

Figure 1.1: Shelterin – the mammalian telomere capping complex. Schematic representation of the six protein shelterin complex at mammalian telomeres. Related dimeric DNA binding proteins TRF1 (blue) and TRF2 (green) with its binding partner RAP1 (grey) and single strand DNA binding POT1-TPP1 heterodimer (purple) connected by TIN2 (yellow).

Shelterin is the telomere capping six-protein complex in mammals (Figure 1.1). It consists of two related double-strand DNA binding proteins – TRF1 and TRF2, TRF2's binding partner RAP1 and single-stranded binding protein – POT1. TIN2 and TPP1 are the 'bridging' proteins linking the double-stranded DNA binding factors to the single-stranded DNA binding protein through various protein interactions.

Shelterin proteins are ubiquitously expressed and associate with telomeres throughout the cell-cycle. With the exception of RAP1 [42], shelterin proteins have exclusive telomere localization and function. Shelterin proteins are essential, as deletion of TRF1, TRF2, POT1, TPP1 or TIN2 result in embryonic lethality [43-47].

The six-protein complex has been shown to exist through fractionated nuclear extracts or pull-down assays. The intact complex has been stably assembled *in vitro* with or without DNA [48]. Role of individual components of shelterin and domain architecture are described in the following sections.

1.1.4 Functions of shelterin

Linear ends of chromosome face two major problems (Figure 1.2). Due to the nature of DNA replication, a small portion of the 3' end of the DNA is lost during each round of replication. Over time, this results in chromosomal erosion leading to the so-called "end-replication" problem. The "end-protection" problem is the resultant of linear ends being erroneously recognized as DNA damage and activating the DNA damage repair pathways (DDR).

Shelterin is involved in averting both problems, by its concerted action with the recruitment of telomerase and repression of DNA damage response proteins (Figure 1.2).

Figure 1.2: Role of shelterin in telomere maintenance. Deprotected telomeres have end-replication and end-protection problems. Shelterin can prevent the telomeres from being recognized as DNA damage sites to maintain the chromosomal stability and recruit telomerase to prevent chromosomal erosion. Image adapted from [49]

1.1.4.1 Telomere length homeostasis and telomerase recruitment

Telomere attrition with each round of replication occurs mainly through two mechanisms. Firstly, due to the semi-discontinuous nature of DNA replication, the very end of the lagging strand is lost with each round of replication [50-53]. Additionally, post-replicative processing of double-stranded telomeric DNA to generate the 3' overhang results in further loss of DNA [54, 55].

In somatic cells, the length of telomere progressively decreases until it reaches a critical limit after which the cells enter senescence [8]. However, in cells that are actively proliferating, such as the germ cells and adult stem cells, telomere length is maintained by telomerase. Telomerase is a telomere-specific reverse transcriptase, that adds tandem DNA repeats and counterbalances progressive attrition. As

anticipated, human telomerase is active only in germ and stem cells, and its improper expression promotes cancer progression [56, 57].

The telomerase enzyme consists of a catalytic reverse transcriptase subunit (TERT) and an RNA component (TERC) that contains the template for synthesis of telomeric repeats (Figure 1.3A) [58, 59]. TERC and TERT together form the holoenzyme [60, 61] that requires additional factors, such as Dyskerin [62] and TCAB1 [63] for its complete activity in cells. TERT essentially has three domains: N-terminal extension (NTE), reverse transcriptase domain (RT) and C-terminal extension (CTE). RT domain of telomerase contains three conserved aspartate residues that are essential for its catalytic activity [64]. NTE contains a telomerase essential N-terminal (TEN) domain and telomerase RNA binding domain (TRDB) connected by a linker. TEN domain can mediate the recruitment of telomerase through its interaction with TPP1 [65, 66].

Figure 1.3: Structure of human telomerase. (A) Telomerase consists of catalytic reverse transcriptase subunit TERT and RNA component which contains the template known as TERC. **(B)** A 30 Å negative staining structure of human telomerase in its dimeric form with an overall bilobal shape [67]. Image rendered from EMDB code 2310.

Biochemical studies performed on telomerase were indicative of its dimeric form [68, 69] consistent with negative stain electron microscopy (EM) analysis on active human telomerase (Figure 1.3 B) [67].

TPP1 is an essential factor for the recruitment of telomerase. TERT-TPP1 interactions have been shown through co-immunoprecipitation experiments. Deletion of TPP1 N-terminal OB domain leads to the loss of this interaction [70]. TPP1 siRNA impairs the recruitment of telomerase to telomeres [71].

In addition, several other studies done both *in vivo* and *in vitro* suggest that TPP1 recruits telomerase to the 3' overhang, using its N-terminal OB domain (reviewed in [49]). A short amino acid region on TPP1 called the TEL patch has been identified to interact directly with telomerase [72]. An *in vitro* study has shown a 2-3 fold increase in processivity of the telomerase enzyme on ssDNA bound with TPP1-POT1 compared to free ssDNA [67].

Other shelterin proteins such as TRF1 are also involved in telomerase regulation. TRF1 has been proposed to maintain telomere length based on a 'counting' model. In this model, when the number of TRF1 molecules falls below a critical number telomere elongation occurs, and conversely telomere shortening occurs when a large number of TRF1 molecules coat the telomeres. TRF1 deletion experiments using a dominant negative mutant shows telomere elongation. By contrast, TRF1 overexpression experiments have shown progressively shorter telomeres with each generation of cell culture [73].

TRF2, POT1, and TIN2 have also been implicated as negative regulators of telomere length using either dominant negative mutants or RNAi experiments [26, 74-76]. However, interfering with the function of these shelterin proteins may trigger DNA

damage response or alter the localization of other shelterin components to the telomere. For example, deletion of TIN2 results in reduced levels of TPP1 at telomeres and activates ATM (Ataxia telangiectasia mutated) pathway [77]. Therefore, it is not clear if the reduction of telomere length observed in deletion experiments is due to the individual proteins or an overlapping effect through association with other shelterin components.

1.1.4.2 Chromosome end protection and T-loop model

Linear ends of chromosome if inappropriately recognized as DNA breaks could activate various DNA repair pathways giving rise to chromosome fusions and deletions. Genomic instability created by DNA damage can result in senescence, apoptosis or transformation into tumour cells [78]. ATM and ATR (Ataxia telangiectasia Rad3 related) are the central protein signalling cascades that respond to DNA damage at deprotected telomeres. ATM recognizes dsDNA breaks, whereas ATR recognizes ssDNA breaks. Both kinases recruit other DNA damage response proteins such as γ -H2AX, 53BP1, and Mre11 into DNA damage foci [79-81]. These foci are cytologically visible and are called as telomere-dysfunction induced foci (TIFs) [82].

As a downstream response to DNA breaks, NHEJ (Non-homologous end joining) and HDR (Homology-directed repair) DNA repair pathways are activated. As an effect of the NHEJ activation, chromosome end-to-end fusions occur. Activation of HDR results in variable and dysfunctional telomeres leading to genomic instability [3].

The role of shelterin in protection against the DNA damage pathways was first observed when ATM kinase was upregulated in cells as a response to deletion of TRF2 [83]. Expression of a dominant negative mutant of TRF2 lacking both the N-

terminal basic domain and the Myb-domain could also elicit DNA damage response leading to chromosome end to end fusions [84]. Recent observation suggests that a region in TRF2 hinge domain (407-431) has a vital role in ATM signalling suppression [85]. NHEJ response is also repressed by TRF2, through its interaction with DNA damage response factor Ku 70/80 heterodimer [86].

Figure 1.4: Shelterin safeguards telomeres. Deletion of shelterin results in activation of DNA damage response and repair pathways. TRF2 protects telomeres from ATM and NHEJ pathways while POT1 prevents ATR and HDR pathways. Image adapted from [87].

POT1 binds to single-stranded telomeric DNA and naturally provides protection against single-strand DNA damage response ATR pathway. The proposed mechanism of ATR repression is through the exclusion of RPA proteins, which initiate the ATR pathway in response to ssDNA. TIN2 has been shown to mediate this process by stabilizing POT1-TPP1 heterodimer at the telomeres [77].

POT1 has also been implicated in defence against HDR, and the mechanism is thought to be redundant and overlapping with other shelterin protective mechanisms [87]. RAP1 was implied to contribute to HDR protection, however, subsequent

knockout studies have suggested that RAP1 may not be involved in telomere protection [42]. In addition, double deletion of TRF1 and TRF2 results in telomeres devoid of shelterin and the resultant deprotected telomeres were shown to elicit various DNA damage responses (Figure 1.4) [87].

An additional level of end protection, through the formation of T-loop at the end of telomeres, has been suggested. In this model, the single-stranded DNA displaces a part of double-stranded DNA (called the D-loop or displacement loop) and folds the ends into a closed loop-like structure (Figure 1.5 A). T-loops can provide enhanced protection from DNA damage response pathways by effectively hiding the single-stranded DNA [88].

Although direct *in vivo* evidence for the existence of T-loops at telomeres is lacking, owing to insufficient decondensation and resolution of chromatin [88], evidence for this model has been obtained mostly from electron microscopy analysis of isolated telomeres. Photo-crosslinking of human telomeres showed 'lariat' like structures [89]. T-loop like structures could be reconstituted *in vitro* with purified TRF2 and long double-stranded DNA with single-stranded overhang (Figure 1.5 B) [90].

Additionally, imaging of telomeres isolated from chicken erythrocytes and human fibroblasts revealed similar T-loop structures [91]. Under low-salt conditions (5 mM NaCl) the telomeres adopt a decondensed 'beads-on-a-string' structure, and at higher salt (75 mM NaCl) condense into a 30-nm fibre-like structure. It must be noted that shelterin proteins were lost during the telomere isolation procedure in this study and purified TRF1 was independently added before EM imaging [91].

A super-resolution microscopy (STORM) study suggests that T-loop formation depends on TRF2, with deletion of TRF2 resulting in a 5 fold reduction of T-loops in

isolated mouse embryonic fibroblasts [88]. In the same study, the frequency of T-loop structures was found to be 25% of all telomeres, indicating that T-loops may not be strictly the only structure that telomeric ends can adopt.

Although T-loops offer a structural resolution to the end-protection problem, they remain to be characterized in depth. Further studies are required to answer critical questions such as the length variations in the loop and mechanism behind loop formation.

Figure 1.5: Telomeric T-loop model. Single-strand overhang displaces a part of the double strand DNA aided by TRF2 and forms a closed 'lariat' structure. T-loops can protect the telomeric end from DNA damage repair pathways by concealing the single-strand overhang. **(A)** Diagrammatic representation of T-loops. Adapted from [49] **(B)** Electron microscopy analysis of *in vitro* assembled T-loops [89]. Reused with permission from Elsevier ©.

1.1.5 Subunits of shelterin

1.1.5.1 Double-stranded DNA binding shelterin proteins

TRF1 and TRF2 are related double-strand telomeric repeat binding proteins with dimerization domain TRFH (TRF homology domain) and C-terminally located

SANT/Myb domain connected by a 100-200 amino acid long flexible hinge region [92]. TRF2 has N-terminal basic Gly/Arg-rich (GAR) domain and conversely, TRF1 contains an N-terminal acidic domain. This basic domain of TRF2 has been shown to interact with DNA non-specifically through electrostatic interaction [93]. A recent AFM study on TRF2 suggests that basic domain could contribute to wrapping of DNA around the TRFH domain [94].

Figure 1.6: Structure of conserved domains in TRF proteins. TRF1 and TRF2 are related proteins with a homodimerization domain and Myb domain. Crystal structure of TRFH dimerization domain at 2.9 Å showing a twisted horseshoe structure (left, monomers shown in different colours). Crystal structure of Myb domains at 2.0 Å shown interacting with the major groove of DNA (right). Structures of TRFH and Myb domain of TRF1 are shown. Crystal structures obtained from PDB accession code 1H6O for TRFH and 1W0T for Myb domain.

TRFH domains of TRF1 and TRF2 were shown to have only 27% amino acid sequence similarity, but exhibit similar 3D structure with 9 helices arranged in a twisted-horseshoe conformation [95]. Both proteins exist as pre-formed homodimers, independent of DNA binding. The TRFH of TRF1 and TRF2 do not heterodimerize with one another due to the distinct interactions found at the dimerization interface [95, 96]. TRFH has been shown as a region of interaction with other proteins containing a F/YxLxP peptide motif on the interacting proteins such as TIN2, Apollo, and PinX1 [97].

The DNA binding domains of TRF1 and TRF2 are strictly speaking Myb-like homeodomains, consisting of three-helix bundle arranged in helix-turn-helix motifs [98]. The homeodomains make extensive contact with the phosphodiester backbone and 'dock' on the DNA. Specificity to telomeric DNA is conferred by base-specific contacts in the major groove of the DNA by key lysine and arginine residues [98]. Additionally, the protein dimer binds to two half sites 5'YTAGGGTTR3' with high affinity in $\sim 10^{-9}$ nM range [98, 99]. Due to the flexible hinge region between the TRFH and Myb domains, TRF proteins have been shown to flexibly identify half sites spaced up to 30 base pairs apart from each other as determined through a Systematic evolution of ligands by exponential enrichment (SELEX) experiment [99].

Human RAP1, unlike its yeast homolog, lacks direct DNA binding activity and localizes to telomeres through its interaction with TRF2 [20]. It is known to contain an N-terminal BRCT domain, Myb domain, and a C-terminal TRF2 interacting domain. A gene-knockout study has reported a minimal role for RAP1 in telomere protection and maintenance. Transcriptome analysis suggests differential expression of several genes in human cells upon RAP1 knockout, suggesting a transcription regulation function [42]. Human RAP1 has lost all its telomeric functions including DNA binding but retained its function as a transcription factor during evolution and is the only known shelterin protein to have a non-telomeric function.

Structural information on the double-stranded DNA binding proteins is derived from domains in the individual proteins. Structures of Myb domains with DNA and TRFH in TRF1 and TRF2 have been solved to high-resolution using crystallography [95, 98]. Crystal structures of TRFH domain in complex with regions of other interacting proteins such as TIN2, Apollo and RAP1 have been solved [97, 100]. Structures of the

TRF2 interacting region and Myb-like domains of RAP1 have been solved using crystallography and solution NMR respectively [101, 102].

1.1.5.2 Single-stranded DNA binding shelterin proteins

Single-stranded DNA binding shelterin proteins contain characteristic oligosaccharide/oligonucleotide binding (OB) domains. OB domains contain a characteristic five-stranded β -barrel and are utilized extensively by single-strand shelterin protein for protein-protein interactions and DNA binding.

Figure 1.7: Structure of conserved domains in hPOT1. (A) Crystal structure of POT1 N-terminus at 1.73Å, displaying two OB folds (magenta and blue) binding to single-stranded telomeric DNA (grey) [103] **(B)** Crystal structure of POT1 C-terminus at 2.1 Å, containing a third OB domain (yellow) and a Holliday junction resolvase like domain (green) interacting with TPP1 (blue) [104]. Structures obtained from PDB accession code 1XJV for POT1 N-terminus and 5H65 for POT1 C-terminus

POT1 is the most well-conserved among shelterin proteins and was identified with a bioinformatics sequence similarity search to a ciliate single-stranded telomere binding protein, TEBP α [105]. It binds to telomeric single-strand DNA using two N-terminally located OB domains. POT1-DNA interactions predominantly involve aromatic amino acids such as tyrosine and phenylalanine stacking against the DNA bases, additionally enhanced by protein-DNA hydrogen bonds [103]. DNA makes contact with both the OB folds and has the backbone exposed to the solvent (Figure 1.7 A) [103]. POT1

heterodimerizes with TPP1 using a C-terminal region [106]. Recently, two independent studies have revealed the previously uncharacterized POT1 C-terminus to contain a third OB domain interspersed with a Holliday junction resolvase like (HJRL) domain [104, 107].

TPP1 heterodimerizes with POT1 and indirectly connects it to the double-strand binding proteins. TPP1 can recruit telomerase through an N-terminally located OB domain. Within the OB domain, a short sequence called the TEL patch has been identified which is responsible for both recruitment and increasing the processivity of telomerase [72].

Figure 1.8: Structure of TPP1 OB domain. Crystal structure of N-terminal OB domain of TPP1 at 2.7 Å, known to be involved in telomerase recruitment **[106]**. Structure obtained from PDB accession code 2I46

Phosphorylation of Ser 255 within the TEL patch is critical to the moderation of TPP1-Telomerase interaction [108]. Studies have shown that telomerase acts preferentially on the telomeric DNA coated with TPP1/POT1 *in vitro*. Additionally, a deletion mutant of TPP1 that lacks the OB domain shows impaired telomere elongation [70, 109]. In addition to the OB domain, TPP1 contains a centrally located POT1 binding domain [25] and a C-terminal TIN2 binding domain [110].

A recent study presents the small angle X-ray scattering (SAXS) density of full-length POT1 in complex with the N-terminus of TPP1. An overall V-shaped architecture was observed opposed to the 'heart' shaped architecture of the ciliate homolog TEBP α/β heterodimer [104].

1.1.3.4 TIN2

TIN2 is the bridging component of shelterin that can simultaneously interact with the single-stranded and the double-stranded DNA binding proteins. The TRFH domain of TRF1 interacts with TIN2 using a short FxLxP motif (also called TRFH binding motif, TBM) in the C-terminus [97]. The N-terminal 1-200 region of TIN2 interacts with TPP1 [110] and TRF2 [76].

In humans, TIN2 is expressed as two isoforms TIN2L (451 amino acids) and a shorter C-terminal truncated isoform called TIN2S (354 amino acids) [111]. Very little is known about the structure and domains of TIN2. The structural information available is limited to the crystal structure of the short peptide motif from 256-276 in complex with the TRF1 and TRF2 dimerization domains [97].

As mentioned earlier, deletion of TIN2 results in impaired telomere localizations of TRF1, TRF2, TPP1/ POT1 and up-regulation of DNA damage response pathways [77]. Mutations in TIN2 give rise to a premature ageing disorder Dyskeratosis congenita (DC) which is discussed in more detail in the following section.

Even though TIN2 has direct implication in telomere disorders and could be a potential target for drug design, no other domains/folds in TIN2 have been identified thus far, making it the least structurally characterized component of shelterin.

1.1.6 Telomeres, human diseases, and cancer

Specific mutations of proteins involved in telomere maintenance can result in telomere disorders, characterized by symptoms such as bone marrow failure, premature ageing and increased risk of cancer due to inherent genomic instability [112, 113]. The two most prevalent telomere disorder diseases in humans are Dyskeratosis congenita and Hoyeraal-Hreidarsson syndrome. Individuals affected by these diseases are born healthy, but symptoms start to appear within the first decade of their lives. The disorder manifests itself more severely in frequently dividing cells such as haematopoietic (bone marrow) cells [114, 115].

Dyskeratosis is a resultant of specific mutations of Dyskerin, TERT, TERC subunits of telomerase or shelterin protein TIN2. Dyskerin is the most frequently mutated factor, with several possible missense mutations [116]. Mutations in TERT occur in the reverse transcriptase (RT) domain [117, 118] and C-terminal region [119, 120]. These mutations, particularly in the RT domain, result in reduced telomerase catalytic activity [121, 122]. Several types of TERC mutations can lead to DC, including single and multiple base substitutions, deletions and insertions [123, 124].

TIN2 is so far the only known shelterin component associated to DC, with mutations clustered within the 214-298 amino acid region [125]. Significantly, the mutation hotspot also contains the TRF1 interaction motif (256–276) (Figure 1.9) [97]. Although the TIN2-TRF1 interactions were compromised in some mutants [126], not all mutants showed this impairment [127, 128], indicating it may not be the sole mechanism of disease progression. Examples of prominent DC mutations in each component are given (Table 1.1).

Figure 1.9: Mutation hotspot in TIN2. Interaction map of TRF1 and TIN2 is shown. Mutation hotspot region found mutated in Dyskeratosis Congenita overlaps with a TRF1 binding motif (TBM) on TIN2.

Table 1.1: Key mutations in Dyskeratosis Congenita

Mutated component	Selected examples of key mutations
Dyskerin	Arg353Val missense [129] Leu37del deletion [130, 131]
TERT	Pro704Ser substitution [132] Pro868Gly substitution [133]
TERC	108GC>AG [134] 408C>G [135]
TIN2	Lys280Glu substitution [136] Arg282His substitution [137]

Hoyeraal-Hreidarsson is a rare syndrome and a severe variant of DC. In addition to classic DC symptoms, it is characterized by growth retardation, cerebral encephaly and impaired motor skills [138]. Individuals affected by this syndrome have short lifespans, rarely surviving beyond a decade. Although the molecular basis of the disease is akin to DC, in addition to the characteristic DC mutations, certain other factors such as the helicase RTEL1 (regulator of telomere length) are frequently

mutated [139, 140]. In humans, telomerase is typically only active in stem cells and germ cells. However, in 85-90% of all cancers, telomerase is re-activated in order to maintain telomere length in proliferating cancer cells [141]. The remaining cancer cells (10-15%) utilize a telomerase-independent ALT pathway [142]. Alternative Lengthening of Telomeres (ALT) elongated telomeres by sister chromatid exchange similar to homologous recombination [143].

Shelterin component POT1 has been reported frequently mutated at both N-terminal and C-terminal OB-folds, in cases of melanoma [144, 145], glioma [146] and leukaemia [147]. Recent studies suggest that such mutations could lead to loss of POT1 from telomeres leading to genomic instability and in turn promoting tumorigenesis [104, 107].

Both cancer and telomere biology disorders are centred around maintenance of telomeres or lack thereof. Telomeric mutant proteins implicated in telomere disorders are not expressed in healthy cells. Additionally, telomerase is expressed in majority of cancer cells and not adult somatic cells. Therapeutic targeting of telomerase in cancer cells is in clinical trial phase. Imetelstat is a small molecule inhibitor of telomerase that has shown promising results in patients affected by different forms of cancer [148, 149]. Additionally, immunotherapy using novel peptide vaccines such as GV1001 and GRNVAC1 have shown positive results in patients with pancreatic cancer [150, 151]. High-resolution structural information on both telomerase and shelterin would be essential to further explore and develop targeted therapeutics.

1.1.7 Higher-order structure of telomeres and shelterin subcomplexes

Telomeres are packed into dense heterochromatin and contain characteristic features like heterochromatic protein 1 (HP1) [152] and epigenetic silencing marks such as

histone lysine trimethylation (H3K9me3) [153, 154]. Electron microscopy analysis of native telomeric chromatin from chicken erythrocytes revealed that increasing the salt concentration resulted in the compaction of the telomeric chromatin from a 10 nm (Figure 1.10) 'beads-on-a-string' form to the 30 nm chromatin fibre, indicating telomeric chromatin resembles canonical nucleosomal arrays [91].

As earlier mentioned, telomeric ends are proposed to end in closed-loop structures called T-loops. The possible implications of the T-loop model are two-fold. The primary effect is the effective protection of single-stranded DNA providing enhanced defence from DNA damage response pathways [88].

Additionally, T-loops are thought to mediate the telomeres into 'open' and 'closed' states. During DNA replication or telomere elongation, telomeres exist in an open state. Otherwise, the closed state would have inaccessible telomere ends in the T-loop formation [155]. It has been suggested that these states could be cell-cycle regulated, but further evidence is required to elucidate this model.

Figure 1.10: Electron microscopy examination of telomeric chromatin. Telomeric chromatin isolated from chicken erythrocytes, positively stained with phosphotungstic acid show 'beads-on-a-string' conformation at low salt (5 mM NaCl) [91]. Arrow mark indicates T-loop junction. Reused with permission from The Journal of Cell Biology (JBC).

TRF2 has N-terminal basic Gly/Arg-rich (GAR) domain, which can interact with DNA non-specifically through electrostatic interaction [93]. AFM study on TRF2 indicates that basic domain is involved in the wrapping of DNA around the TRFH domain, resulting in DNA compaction [94]. This compaction model is supported by *in vivo* evidence, through super-resolution STORM imaging of telomeres in human HeLa cells. Depletion of TRF1/2 has been shown to decompact the telomeres [156]. However, opposing studies using the same imaging technique have shown conflicting results as no significant decompaction upon complete knock-down of TRF1 and TRF2 [157, 158]. Additional investigation is required to validate the compaction model.

Experimental evidence indicates that shelterin can exist as an intact complex combining the six proteins. Biochemical experiments have shown that TIN2 can concurrently interact with TRF1 and TRF2 [97, 110]. Although RAP1 is only known to have direct interaction with TRF2, immunoprecipitation of RAP1 can pull-down the intact six-protein complex through an indirect interaction using TIN2 link [159]. Further, the intact complex has been stably purified using eukaryotic expression system in a recent *in vitro* biochemical study [48].

In addition to existing as a six-protein complex, shelterin has also been shown to exist as subcomplexes with one or two components missing. In a quantitative immunoblot study done in primary human fibroblast cells and HeLa cells, the six shelterin proteins did not exist in an equimolar ratio.

The relative abundance in the chromatin-bound fraction indicated that TRF2, TIN2 and RAP1 proteins are the most abundant, present in almost equimolar ratio. In comparison TRF1 was approximately 5-fold less in concentration. TPP1-POT1 heterodimer was the least abundant, found 15-fold less than TRF2 [160]. The relative

abundance of shelterin proteins bound to telomeres determined by this study is summarized (Table 1.2).

Table 1.2: Relative abundance of shelterin components per telomere, determined using quantitative immunoblot [160].

Shelterin component	No. of molecules per telomere
TRF1	180
TRF2	860
RAP1	1060
TIN2	840
TPP1	50
POT1	50

In several other studies, shelterin lacking either TRF1 or the TRF2-RAP1 complex have been shown to still localize to telomeres, indicating that formation of the six-protein complex is not necessary for the individual subunits to engage with the telomeres [110, 161, 162]. Nuclear extracts separated by size-exclusion shows fully-assembled shelterin and two other shelterin subcomplexes without either TRF1 or TRF2 [163]. Certain proteins are known to interact with specific shelterin components and perform distinct functions, for instance, PinX1 [164] & Tankyrase [165, 166] interact with TRF1 and Apollo [167] & SLX4 [168] interact with TRF2. The subcomplexes would also interact with a differential set of proteins conferring distinct functions by extension.

Photobleaching experiment to study binding kinetics of GFP-tagged individual shelterin DNA binding proteins showed the half-life of TRF1/2 on DNA was 24 seconds

which was approximately 5-fold faster than POT1's half-life of 128 seconds [169]. In comparison, the exchange rate of histones in the nucleosome was determined in the range of minutes using a similar GFP-tagging and FRAP approach in live cells. The 'fast fraction' of H2B had a half-life of 6 minutes, while the majority of H3 and H4 remained stably bound [170]. Binding kinetics of the telomere binding proteins and difference in binding half-life between POT1 and TRF proteins suggest that shelterin is dynamically associated with telomeres.

Early information on the shelterin domains derived from limited proteolysis and subsequent interaction mapping experiments using size-exclusion chromatography has shown that the individual proteins engage in mutually inclusive interactions to form the shelterin complex (Chapman and Fairall, unpublished data, Figure 1.11 A).

Telomeric chromatin contains nucleosomes and hundreds of dynamically binding shelterin proteins, but how these factors come together to form the compact telomeres is not understood. Although the binary interactions of shelterin proteins and their individual functions are well characterized biochemically (Figure 1.11 B), the subcomplexes of shelterin, their specific functions and relative abundance in the cell remain uncharacterized.

Figure 1.11: Summary of shelterin domains and interactions. (A) Interaction map of shelterin shows independent interactions between shelterin proteins. TIN2 interacts with TRF1 [76, 97, 110, 171], TRF2 [97, 110] and TPP1 [25, 26] and links the complex together. Other interactions within the complex comprise of POT1-TPP1 [25, 26] and TRF2-RAP1 [93, 102, 159] **(B)** Known domains and structure of the shelterin proteins rendered in grey (not in scale). Interacting proteins and DNA rendered in blue and orange respectively. PDB accession codes used in each row left to right are 1H6O [95], 1W0T [98], 1H6P [95], 1W0U [98], 1FEX [101], 3K6G [172], 3BQO [97], 2I46 [106], 1XJV [103] and 5H65 [104].

1.2 Electron microscopy

1.2.1 Overview and recent developments

In order to obtain mechanistic insights into the function and dynamics of biological processes, high-resolution structural information of the molecules and complexes involved is essential. In addition, near-atomic resolution structure of target molecules is an important element to drug discovery.

Traditionally NMR and X-ray crystallography were the techniques of choice in structure biology. However, NMR is limited to the study of proteins with a molecular weight of about 50 kDa. Crystallography has been and continues to be an important method for high-resolution structure determination [173]. However, obtaining crystals requires large amounts of highly homogeneous sample and crystallization of target protein remains a bottleneck for several challenging proteins such as membrane proteins or large multi-protein complexes.

In recent years, cryo-electron microscopy (Cryo-EM) has emerged as an alternative, with many structures solved routinely to near-atomic resolution (4 Å or higher). The major advantages of cryo-EM are the relative ease of sample preparation and tolerance to a certain extent of heterogeneity. Since the recent 'resolution revolution' [174], many biological questions have been answered with near-atomic resolution structures, such as *de novo* structures of proteins which were challenging to crystallize [175-178], conformational/transitional states of molecular machines [179-182], drug targets [183-185] and cell signalling molecules [186]. The highest resolution hitherto obtained with cryo-EM is the 2.2Å structure of β-Galactosidase bound to an inhibitor [187].

Direct detector technology [188, 189] and beam-induced motion correction [190, 191] are the two major advances that have led to an increase in the number of high-resolution structures determined by Cryo-EM in the recent years [192].

Detector quantum efficiency (DQE) is a parameter used to characterize the efficiency of the detectors in converting electrons into the image, based on the signal to noise ratio as a function of the spatial frequency [193]. Traditionally, cryo-EM imaging was performed using photographic films and charge coupled device (CCD) cameras. However, CCD cameras have poor DQE performance at high-resolution information range [194] and photographic films require significantly longer time for development and digitization. Direct detectors using complementary metal–oxide–semiconductor (CMOS) sensor [195, 196] were developed as an alternative. With significantly improved speed and DQE performance at both low-resolution and high-resolution range compared to the conventional detectors [192], direct electron detectors played an essential role in the ‘resolution revolution’.

Additionally, the ability to split the image into sub-frames or ‘movie mode’ has enabled correction of beam-induced motion that occurs while acquiring images [191, 197]. The motion vectors are determined globally and for individual particles, and then the frames are aligned computationally [198, 199]. High electron dose is unsuitable for biological samples as they are sensitive to radiation damage over time, subsequently resulting in a loss of information [200]. However, imaging with higher dose enables improved contrast, which is useful for identification of particles and determination of their orientations. With the advent of movie mode, images can be acquired with a high dose for particle selection, but during structure determination, the final frames of the image containing radiation damage can be discarded [201].

1.2.2 Single particle analysis

In a transmission electron microscope, a projection image of a molecule in random orientations is obtained. In order to obtain the 3D image from the 2D projection images, different views need to be combined in an iterative manner [202]. This process is collectively called single particle analysis (SPA) [203]. There are two ways to prepare biological samples for EM analysis called negative staining and vitrification. The following sections describe sample preparation and data processing approaches in SPA.

1.2.2.1 Sample preparation

Negative staining involves depositing a uniform layer of heavy metal stain (such as uranyl acetate) on a sample such that the background is evenly stained, and the particle of interest is excluded. The thickness of the stain is an important parameter to consider, as inadequate coverage may lead to only partial visualization of the macromolecule under observation.

Negative stain EM analysis has two major advantages. Firstly, due to strong electron scattering by the heavy metal atoms in the stain, high contrast is generated in comparison to unstained molecule [204]. Additionally, ease of negative stain grids preparation and utilization of low sample amount makes it a useful method for challenging samples that are restricted by quantity. The limitation of this method is that the stain itself limits the maximum achievable resolution of the reconstruction to $\sim 20\text{\AA}$, as air-drying the staining solution results in the formation of heavy metal micro-crystals [205]. Applications of negative stain EM include obtaining an initial model for samples [206] and analysis of small or highly heterogeneous samples [205].

Cryo-EM or vitrification is the current method of choice to obtain high-resolution structural information. Vitrification is a technique developed in 1984 by Dubochet *et al* [207]. It involves plunge-freezing specimen rapidly in liquid ethane, embedding molecules in amorphous vitreous ice and preserving the ultrastructure of molecules in near-native conditions [208]. The process of vitrification is accomplished by rapidly cooling the sample at the rate of 10^6 °C/s, to avoid crystal ice formation, which can damage the ultrastructure of biological samples [209]. In order to facilitate the high rate of heat transfer, the sample deposited on the grid must be sufficiently thin. This is usually achieved by loading 4-20 µl of the sample and blotting the grid prior to plunge-freezing.

Cryo-grids are typically made with carbon support with uniform holes distributed throughout. Several factors influence the quality of cryo-grid, one of which is ice thickness. The ice should be thick enough to completely embed the sample but thin enough to remain transparent to the electron beam [210]. Preparation of cryo-grids has been made more user-friendly and reproducible over the years with the development of semi-automatic plunge-freezing instruments for grid preparation. These instruments provide greater control by providing adjustable blotting parameters and controlling humidity.

The process of vitrification is not universally tolerated by all samples. Due to free diffusion in a thin layer of solution, protein molecules tend to interact with the air-water interface, which may result in preferential particle orientation, unfolding or even aggregation of proteins during grid preparation [211]. Affinity support grids such as NiNTA coated grids [212] have been suggested as a solution to prevent diffusion.

1.2.2.2 Contrast in EM

Biological samples are weak-phase objects and as a result the interaction between the electron beam and sample gives rise to variations in phase, but no significant amplitude changes. In negative stain EM, this problem is overcome with strong electron scattering by a heavy metal stain. However, in cryo-EM, this problem is exacerbated by embedding unstained low-contrast molecules in a layer of vitreous ice.

The image formed in a transmission electron microscope can be described by the Contrast transfer function (CTF) in Fourier space [213]. For a weak-phase object under parallel illumination, the CTF can be determined by the equation given below [214].

$$CTF(f) = 2 \sin 2\pi\gamma(f) \quad (\text{Equation 1})$$

Where f is the spatial frequency and γ is the wave aberration function. The wave aberration function can be further expanded with the equation below.

$$\gamma(f) = C_s \lambda^3 \frac{1}{4} f^4 - z \lambda \frac{1}{2} f^2 \quad (\text{Equation 2})$$

In this equation, C_s is the spherical aberration of the microscope, z is the defocus value and λ is the wavelength of the electron beam. From Equation 1 and 2, it can be inferred that CTF of the microscope is dependent on the spherical aberration and the defocus. Additionally, CTF is a sinusoidal function based on the spatial frequency [214]. The oscillations in the CTF results in loss of information at certain spatial frequencies. EM images can be collected with a range of defocus values in order to compensate [215].

1.2.2.3 Data collection and processing

EM data could be collected using automatic data acquisition software or manually. Approximately hundreds to millions of images of the particles in random orientations are recorded. Data processing converts individual 2D projection images to the 3D

structure in an iterative process. There are several processing software packages available with varying levels of complexity and user-friendliness, such as SPIDER [216], FREALIGN [217], EMAN [218], SPARX [219], XMIPP [220] and RELION [221].

In this section, the general steps involved in data processing are described, followed by a description of 2D and 3D classification techniques specific to RELION. The steps involved in data processing are summarized in a flowchart (Figure 1.12).

The initial step in processing EM data is correction of beam-induced motion, followed by CTF estimation and particle picking. Software packages either have an in-built CTF module or use an external application such as CTFFIND [222].

CTF correction is more applicable to Cryo-EM data than negative staining data, due to the resolution limitation and high contrast in negative staining. CTF is an oscillatory function that assumes negative values at certain frequencies. Phase flipping is a correction procedure, performed by changing the sign of the peaks but retaining the amplitude [223].

Particle picking can be the most labour-intensive step in data processing. It involves careful 'boxing' or picking of individual particles from the micrographs. Picking can be done manually or in a semi-automatic manner with template matching [224]. The box size used for picking particles is usually set to twice the longest particle dimension as a rule of thumb.

The purpose of 2D and 3D classification is to split the particles into different homogeneous subsets, in order to determine the true representative structure and additionally separate different transitional or heterogeneous structures that could exist within the sample.

Figure 1.12: Structure determination by EM. The work-flow for single particle structure determination. Adapted from [210]

Particle classification using RELION utilizes a regularized maximum likelihood approach with an application of Bayesian theorem [225]. In RELION a statistical approach rather than a heuristic approach is applied, where the probability of a map being correct based on the observed data is the function being optimized [221]. Particles from the homogenous and well-refined classes are iteratively selected for subsequent stages of processing.

2D classification does not require any initial references and the only critical parameter to be set by the user is a total number of classes (K) [226]. 3D classification works

using a similar basic algorithm as 2D, but additionally involves the use of an initial model, iterative back projections and weighted orientation assignment.

Initial models used can be derived from *ab initio* methods (EMAN [218]), from the crystal structure of known domains in the sample or using featureless spheres or discs that are low-pass filtered. Best 3D classes are selected and refined further to give the final map.

In RELION a ‘gold standard’ approach [227] is followed, where the data is divided into two random sets and kept independent throughout refinement. The cross-correlation values between the two maps are calculated and at the final step, the two independent maps are merged into one final map. This approach can generally prevent overfitting of noise and over-estimation of resolution in 3D reconstruction [228]. The final global resolution of the map is usually determined using the Fourier shell correlation (FSC) between the independent maps based on the frequency corresponding to either 0.143 or 0.5 FSC [227].

1.2.3 Phase-plate

Contrast in conventional TEM is created by applying defocus to the image. High-resolution information is attenuated at higher defocus values and the low-resolution contrast information is suppressed close to focus [229]. In particular, for molecules of size ≤ 200 kDa, a higher than average defocus needs to be applied to visualize them optimally, limiting the maximum resolution attainable [230].

Additionally, inaccuracies in alignments of the projection images during data processing increases with small-sized molecules [231]. The structure of 93 kDa Isocitrate dehydrogenase has been recently solved to 3.8 Å using EM. This was made possible by collecting large data set, extensive screening for very thin ice regions,

using small pixel size (counting mode) and sorting particles [232]. As an alternative approach, the newly developed phase-plate technology addresses the problems associated with small molecules and low-contrast samples.

Phase plate is a type of device that can enhance the contrast of the image by modulation of the phase contrast. In light microscopy, phase plates were developed in 1942 by Zernike [233]. Application of phase plate in electron microscopy was already suggested in theory during 1947 by Boersch [234]. Few models were since developed, such as the Zernike-phase plate (ZPP) [235] and spider thread phase plate [236, 237]. However, its practical implementation has only been possible in recent times with ZPP [238]. ZPP is essentially a thin carbon film with a central 1 μm hole, positioned in the back-focal plane such that the electron beam is aligned to the hole.

Under optimal conditions, ZPP can produce a $\pi/2$ phase shift resulting in significant enhancement of the contrast. Since its development, ZPP has been used in several applications [239, 240] including SPA [241]. There are a few practical limitations associated with ZPP such as a maximum achievable resolution of 10 Å, short usable lifetime, fringing artefacts and the prerequisite for precise alignment of the electron beam at the centre of a small hole [242].

1.2.4.1 Volta phase plate

The Volta phase plate (VPP) was recently developed as a successor to the ZPP [243]. VPP is essentially a hole-free, continuous carbon film placed in the back-focal plane of the microscope. Since the film is continuous, there is no requirement for precise alignment of the beam as in the case of ZPP. A negative potential is created upon irradiation of the film by the electron beam, resulting in a $\pi/2$ phase shift. Continuous

heating of VPP > 100°C is necessary to prevent beam induced contamination and charging. Schematic of the VPP in the microscope is shown (Figure 1.13).

Figure 1.13: Schematic diagram of Volta-phase plate (VPP). The VPP is a thin and continuous carbon film placed in the back focal plane of the microscope capable of producing phase-shift upon beam irradiation. Phase shift causes a contrast enhancement and helps visualize small-sized or low-contrast samples.

Initial activation or conditioning is performed by irradiating the phase plate for a short duration until the desired phase-shift and increase in contrast is observed. The phase-shift increases with the total dose applied over time and gradually becomes over-conditioned. Beam-induced spots are formed on the film and a 5-7 day recovery time is recommended for the VPP to return to its original state. The position of VPP is shifted to expose a new area on the film for subsequent data collection.

The exact working principle of VPP is not fully understood. The current hypothesis is that electron beam interacts with the surface of the carbon film resulting in local desorption of electrons and generation of Volta potential [243].

VPP has been used in several applications since its development, including cellular tomography [244-246] and single particle analysis [247-249]. Structure of several small-sized complexes such as 257 kDa PrX3 [250], 150 kDa GPCR-G protein complex [251] and 64 kDa haemoglobin have been solved. Volta phase plate, therefore, shows promising potential for structural analysis of small-sized complexes by cryo-EM.

1.3 Aims of this dissertation

Shelterin exists as a six-protein complex and subcomplexes at the telomeres. The individual shelterin proteins and interactions between these proteins have been well characterized. However, the *in vitro* biochemical and structural information on subcomplexes remain elusive. The main objectives of this thesis are as follows:

1. Biochemical purification of double-strand DNA binding subcomplex TRF2-RAP1-TIN2
2. Biochemical purification of the telomerase recruitment complex POT1-TPP1-TIN2
3. Investigation of stoichiometry and DNA binding activity of subcomplexes
4. Structural characterization of aforementioned complexes using single particle electron microscopy

Chapter 2: Materials and Methods

2.1 Materials

All reagents of suitable purity were obtained from Affymetrix, BioRad, BD, Sigma, VWR or Vivantis unless otherwise specified. Electron microscopy grade reagents and grids were obtained from Electron microscopy sciences. NiNTA resin used was either from Qiagen or Macherey-Nagel. GE columns were used for protein purification. BioRad standard protein mix was used for calibrating the gel filtration columns. cOmplete protease inhibitor from Roche was included in all purifications. Insect cell media and reagents were obtained from Thermo Fisher Scientific.

All bacterial and baculoviral constructs used in this study with the exception of MBP-TIN2 were graciously provided by Daniela Rhodes' lab (refer to Appendix A3 for details). Widom 601 145 DNA construct was graciously provided by Curt Davey. Lyophilized histones H2A, H2B, H3 and H4 were generously given by Lars Nordenskiöld. Bacterial media for cloning, protein production and protein and DNA gels were prepared in-house (refer to Appendix A1 for details).

2.2 Methods

Bacterial cells and insect cells were used for recombinant protein expression. Sonication was used to lyse the cells, with 10s ON and 20s OFF and 3s ON and 3s OFF pulses on ice, for bacterial and insect cells respectively. The maximum amplitudes used for bacterial cells and insect cells were 60% and 35% respectively. All the protein related work was performed either at 4°C or on ice unless mentioned. SDS-PAGE with final acrylamide concentration of 12% were used unless specified. Protein concentration was estimated using Thermo Scientific Nanodrop by measuring the absorbance at 280 nm and calculating the concentration using specific molar

extinction coefficients. Mass spectrometry was performed by the Proteomic Core Facility of the Biological Research Center, NTU. The constructs and composition of buffers used in following purification protocols are listed (Table 2.1 and 2.2 respectively).

Table 2.1 List of constructs used for protein purification. Bacterial and baculoviral constructs used in this study. All constructs of the full-length proteins unless otherwise mentioned.

Construct		Affinity tag(s)
Bacterial		
TRF1		His ₆ -S-TEV
TRF2		His ₆ -S-TEV
RAP1		His ₆ -S-TEV
TIN2S		GST-PreScission-His ₆ -TEV
TIN2L		His ₆ -S-TEV
MBP-TIN2L		His ₆ -S-TEV
Baculoviral		
C1	POT1 TPP1	His ₁₀ -SSG-TEV-POT1
		Strep-GS-TEV-TPP1
C2	POT1 TPP1 (87-544)	His ₁₀ -SSG-TEV-POT1
		Strep-GS-TEV-TPP1
C7	POT1 TPP1 TIN2L	His ₁₀ -SSG-TEV-POT1
		Strep-GS-TEV-TPP1
C8	POT1 TPP1 (87-544) TIN2L	His ₁₀ -SSG-TEV-POT1
		Strep-GS-TEV-TPP1
C9	POT1 TPP1 TIN2S	His ₁₀ -SSG-TEV-POT1
		Strep-GS-TEV-TPP1
C10	POT1 TPP1 (87-544) TIN2S	His ₁₀ -SSG-TEV-POT1
		Strep-GS-TEV-TPP1

Table 2.2 Composition of buffers used in protein purification.

Buffers	Composition
TRF1 lysis	50 mM HEPES, 500 mM NaCl, 20 mM Imidazole, 1% Triton, 2 mM PMSF, 0.25 mM EDTA, 1 mM DTT/5 μ M BME
TRF1 high salt wash	50 mM HEPES, 1 M NaCl, 20 mM Imidazole, 1 mM DTT/ 5 μ M BME
TRF1 wash	50 mM HEPES, 500 mM NaCl, 20 mM Imidazole, 1 mM DTT/ 5 μ M BME
TRF1 elution buffer	50 mM HEPES, 500 mM NaCl, 250 mM Imidazole, 1 mM DTT/ 5 μ M BME
TRF1 size-exclusion	50 mM HEPES, 500 mM NaCl, 10% Glycerol, 1 mM DTT/ 5 μ M BME
TIN2 lysis	50 mM HEPES, 300 mM NaCl, 20 mM Imidazole, 1% Triton, 2 mM PMSF, 0.25 mM EDTA, 1 mM DTT/5 μ M BME
TIN2 wash	50 mM HEPES, 300 mM NaCl, 20 mM Imidazole, 1 mM DTT/ 5 μ M BME
TIN2 elution	50 mM HEPES, 300 mM NaCl, 250 mM Imidazole, 1 mM DTT/ 5 μ M BME
TIN2 GST elution	50 mM HEPES, 300 mM NaCl, 10 mM reduced Glutathione, 1 mM DTT
TIN2 size-exclusion	50 mM HEPES, 300 mM NaCl, 10% Glycerol, 1 mM DTT/ 5 μ M BME
TRT lysis	50 mM HEPES, 300 mM NaCl, 5 mM MgCl ₂ , 10 mM Imidazole, 10% Glycerol, 1% Triton, 1 mM DTT/0.5 mM TCEP
TRT wash	50 mM HEPES, 300 mM NaCl, 20 mM Imidazole, 1 mM DTT/0.5 mM TCEP
TRT elution	50 mM HEPES, 300 mM NaCl, 300 mM Imidazole, 1 mM DTT/0.5 mM TCEP
TRT size-exclusion	50 mM HEPES, 300 mM NaCl, 10% Glycerol, 1 mM DTT/0.5 mM TCEP
C9 lysis	50 mM HEPES, 300 mM NaCl, 5 mM MgCl ₂ , 20 mM Imidazole, 10% Glycerol, 1% Triton, 1 mM DTT/0.5 mM TCEP
C9 wash	50 mM HEPES, 300 mM NaCl, 25 mM Imidazole, 1 mM DTT/0.5 mM TCEP
C9 elution	50 mM HEPES, 300 mM NaCl, 750 mM Imidazole, 1 mM DTT/0.5 mM TCEP
C9 size-exclusion	50 mM HEPES, 300 mM NaCl, 10% Glycerol, 1 mM DTT/0.5 mM TCEP

2.2.1 Expression and purification of TRF1

2.2.1.1 Transformation

Full length human TRF1 construct with a 6X Histidine (His) tag in a pET30a backbone was transformed into Rosetta DE3 bacterial cells by mixing 50 ng of the plasmid in 50 μ l cells and incubated on ice for 30 min. Subsequently, the cells were subjected to heat shock by incubating at 42°C for 30 seconds and immediately transferring the tube on ice for 2 min. 0.2 ml of LB media was added to the mix and incubated at 37°C for 1 hour. 0.1 ml of the transformation mix was plated on LB agar plate made with 50 μ g/ml Kanamycin (vector selection) and 35 μ g/ml Chloramphenicol (cell line selection).

2.2.1.2 Bacterial protein expression and cell harvest

Two flasks containing 25 ml LB media with 50 μ g/ml Kanamycin and 35 μ g/ml Chloramphenicol were inoculated with single colonies from the transformation plate. This starter culture was grown at 37°C overnight for 12-15 hours. Two 5 L flasks each containing 1 L of pre-autoclaved 2X-YT and the same antibiotics were inoculated each 2% v/v of starter culture and grown at 37°C and 220 rpm. Growth was closely monitored by checking the OD at 600nm.

Once the cells reached an OD of 0.6, they were induced with 0.4 mM of IPTG and allowed to grow further at 15°C for 15 hrs. Cells were harvested by spinning at 5000 rpm for 15 minutes. Cell pellet was transferred to a Falcon tube, flash frozen and stored at -80°C until further use.

2.2.1.3 Purification of TRF1

Purification protocol for TRF1 optimized by Lynda Chapman was used with minor variations. Pellets were dissolved in cold TRF1 lysis buffer, uniformly suspended and

sonicated. Cell lysate was spun at 20,000 rpm for 30 minutes. The supernatant was collected and filtered into a Falcon tube and 1 ml (bed volume) NiNTA resin pre-equilibrated with TRF1 lysis buffer was added and gently mixed. The tube was incubated at 4°C for 30 min and the slurry was subsequently transferred to a Bio-Rad glass column.

Flow through was collected and stored for SDS-PAGE analysis. The resin was then washed with 5 column volumes (C.V) of TRF1 high salt wash buffer, promptly washed with 2.5 C. V of TRF1 wash buffer and finally eluted in 3 C.V TRF1 Elution buffer. Wash and elution fractions were collected for SDS-PAGE analysis.

Based on the SDS-PAGE result, elution fractions containing the least contamination/degradation bands were pooled and concentrated using Vivaspin Turbo 4 concentrators with 30 kDa molecular weight cut-off (MWCO) to <0.5 ml volume. The concentrated protein was injected to a GE Superdex™ 200 10/300 equilibrated with TRF1 size-exclusion buffer. Fractions from the size-exclusion run were analysed using SDS-PAGE and pooled.

In order to cleave the 6X-His-S-Tev tag on the protein, in-house His-tagged TEV protease was used. TEV was added according to the yield of protein in the ratio 1:50, i.e. for every 50 mg of protein 1 mg of TEV protease was added. This reaction was incubated overnight 12-15 hr at room temperature. Protein sample before and after TEV digestion were saved for SDS-PAGE analysis.

After digestion, TEV and undigested TRF1 were removed by reverse NiNTA. 0.5 ml NiNTA was added post-digestion and incubated for 30 min at 4°C. The slurry was transferred to a Bio-Rad glass column and the flow-through which contained the digested protein was collected and concentrated to <1 ml.

As the final step in purification, the concentrated protein was again injected to Superdex™ 200 column. The peak fractions were analysed using SDS-PAGE and pooled for subsequent analysis.

2.2.2 Expression and purification of TIN2

Three different constructs were used for TIN2 namely: 6X His-TIN2L, 6X His MBP-TIN2L, and GST-TIN2S. MBP-TIN2 was cloned as described in the subsequent section.

2.2.2.1 MBP-TIN2L cloning

MBP in pMAL-c5X backbone and TIN2L in pET30a backbone were linearized by PCR using the primers mentioned below (Table 2.3).

Table 2.3: Sequence of primers used for SLIC cloning of MBP-TIN2

Primer	Sequence (5' → 3')
TIN2 Forward	CATTCTTCTGGTCTGGTGCCACGCGTTCTATGAAAATCGAAGAAGGTA
TIN2 Reverse	CATGGCGCCCTGAAAATACAGGTTTTTCATTACCTGCAGGGAATTCGG
MBP Forward	GAAAACCTGTATTTTCAGGGCGCCATGGCTACGCCC
MBP Reverse	AGAACCGCGTGGCACCAGACCAGAAGAATGATGAT

Sequence and ligation independent cloning (SLIC) was used to fuse Maltose Binding protein (MBP) to full-length TIN2L. The cloned constructs were verified with DNA sequencing before further use.

2.2.2.2 Transformation

The cloned construct was transformed as described in previous section (2.2.1.1). Cloning/sequencing and protein expression plasmid were transformed into NEB 10-β and Rosetta DE3 respectively.

2.2.2.3 Bacterial protein expression and cell harvest

All TIN2 constructs used were IPTG inducible. The cells were grown in 2YT to a cell density of 0.6-0.8 and subsequently induced. Based on test expression the following induction conditions were used for each construct (Table 2.4). The cells were spun at 5000 rpm for 15 min and the resultant pellets were flash frozen in liquid nitrogen and stored at -80°C until further use.

Table 2.4: Protein expression/induction condition for TIN2 constructs

Construct	Induction temperature(°C)	Duration(hours)	IPTG conc. (mM)
TIN2L	37	3-5	0.4
MBP-TIN2L	37	3	0.4
GST-TIN2S	15	12-15	0.5

2.2.2.4 Purification of TIN2L and MBP-TIN2L

The purification procedure for both TIN2L and MBP-TIN2L were identical, including the buffers. The frozen pellets were resuspended in TIN2 lysis buffer and every 50 ml of the cell suspension was sonicated on ice for a total time of 5 minutes. The cell lysate was spun at 20,000 rpm for 45 minutes at 4°C. The 1 ml bed volume of NiNTA was added to every 50 ml of supernatant and incubated at 4°C for 30 min. The slurry was added to BioRad glass column and allowed to settle down and the flow-through was collected.

The resin was then washed with 2 C.V of TIN2 wash buffer and subsequently eluted with 3 - 4 C.V TIN2 elution buffer. The elution fractions were analysed on SDS-PAGE and pooled. After concentrating the sample to <0.5 ml using a 10 kDa MWCO concentrator, it was injected into a GE Superdex 200 10/300 column which was pre-

equilibrated in TIN2 size-exclusion buffer. The fractions were finally analysed using SDS-PAGE.

2.2.2.5 Purification of GST-TIN2S

Cell pellet was resuspended in TIN2 lysis buffer and sonicated. The cell lysate was spun at 20,000 rpm for 45 min at 4°C. The lysate was then loaded on to a 5 ml GE GSTrap column. The column was washed with 2 C.V TIN2 binding buffer and eluted with 4 C.V of TIN2 GST elution buffer. The elution fractions were pooled, concentrated and injected to a GE Superdex 200 10/300 column which was pre-equilibrated with TIN2 size-exclusion buffer and the fractions were finally analysed using SDS-PAGE.

2.2.3 Expression and purification of TRF2-RAP1-TIN2S

2.2.3.1 Transformation

Full-length constructs of 6X His-tagged TRF2 with a, 6X His-tagged RAP1 and GST-tagged TIN2S, each in a pET30a were used for this complex. The constructs were separately transformed and grown in BL21 *pLysS* cells.

2.2.3.2 Bacterial expression of proteins and cell harvest

Single colonies from each transformation plate were inoculated in 10 ml of LB media containing 50 µg/ml Kanamycin (vector antibiotic) and 35 µg/ml Chloramphenicol (cell line antibiotic) and grown at 37°C, 200 rpm. Once the cultures reached an OD of 0.6, 1 ml of the culture was removed and spun at 5000 rpm for 5 minutes to pellet the cells. This pellet was dissolved in 50 µl of 1X Laemmli buffer and used to analyse the before induction sample. The culture was then promptly induced using 0.5 mM IPTG and the cells were further allowed to grow overnight at 15°C. Pellet from 1 ml was dissolved in

100 µl of Laemmli buffer to compensate for cell growth. 5 µl of the before and after induction samples were analysed on SDS-PAGE.

The volume of cultures were grown roughly in the ratio 2:1:6 for TRF2, RAP1 and TIN2 respectively. Overnight pre-cultures for each construct were grown sufficiently to inoculate 2% v/v of the large-scale cultures, in LB medium containing 50 µg/ml Kanamycin and 35 µg/ml Chloramphenicol. The large-scale cultures were grown in 2YT media with the same antibiotic selection at 37°C until the OD was 0.6-0.8 and subsequently induced with 0.5 mM IPTG and grown further overnight at 16°C. Cells were harvested by spinning at 5000 rpm and the resulting pellets were flash frozen using liquid nitrogen and stored at -80°C until further use.

2.2.3.3 Purification of TRF2-RAP1-TIN2S

Cell pellets were resuspended in TRT lysis buffer to a final volume of 150 ml. The cell suspension was sonicated in batches of 50 ml and paused in every 2 min intervals of sonication time to avoid overheating. The lysate was spun at 20,000 rpm for 1 hour. The supernatant was then filtered using a syringe filter with a cut-off of 0.2 µm.

GE 5 ml HisTrap column was equilibrated with TRT wash buffer. The filtered supernatant was applied to the column and washed with the wash buffer until the UV 280 absorbance dropped to zero and eluted as a gradient over 10 C.V. The elution fractions were analysed using SDS-PAGE and the fractions that had the least contaminations were pooled and the concentration was measured.

Affinity tags were removed by adding TEV protease to the final ratio of 1:50 of TEV to protein. The reaction was then transferred to a 10 kDa cut-off dialysis bag, and dialysed in two steps with 500 ml TRT Wash buffer each overnight at 4°C.

Reverse NiNTA was performed using 0.5 ml resin to remove TEV and undigested protein. The complex with the tags removed was in the flow through, and the complex with undigested tags and TEV protease were eluted later with TRT elution buffer. After the analysis using SDS-PAGE, the flow-through was concentrated to 1 ml and injected onto a Superdex 200 16/600 size-exclusion column. The peak fractions were pooled, concentrated and used for further analysis.

2.2.4 MultiBac expression and purification of POT1-TPP1-TIN2S

POT1-TPP1-TIN2S complex was expressed and purified from SF9 insect cells using MultiBac system. Six different EmBacY bacmid constructs with different combinations of POT1, TPP1, and TIN2 (Table 2.5) were cloned by Dr Vasuki and Dr Alessandra. Further construct details are given in the appendix section (A3). YFP reporter gene was also present in the system that helped to monitor of infection stage/level. Based on the initial screening, construct 9 or C9 which had full-length POT1, TPP1, and shorter isoform TIN2S was chosen for large-scale expression and purification. The following sections describe the bacmid and initial virus production for all the constructs and the scale-up, expression and purification for C9.

Table 2.5: Overview of POT1-TPP1/-TIN2 MultiBac constructs.

Construct 1	POT1 (1-634)	TPP1 (1-544)	-
Construct 2	POT1 (1-634)	TPP1 (87-544)	-
Construct 7	POT1 (1-634)	TPP1 (1-544)	TIN2L (1-451)
Construct 8	POT1 (1-634)	TPP1 (87-544)	TIN2L (1-451)
Construct 9	POT1 (1-634)	TPP1 (1-544)	TIN2S (1-354)
Construct 10	POT1 (1-634)	TPP1 (87-544)	TIN2S (1-354)

2.2.4.1 Insect cell culture

All insect cell culture work was carried out in Sf9 insect cells, derived from ovaries of *Spodoptera frugiperda*. For this study, the cells were grown in Sf900II SFM supplemented with 2% FBS and 1X Anti-Anti unless mentioned. Cells were maintained as a suspension culture at 27 °C, 120 rpm in Erlenmeyer flasks with vented caps. Cultures were grown until the cell density of 2 to 4 million cells per ml and subsequently sub-cultured to a cell density of 0.5 million cells per ml.

2.2.4.2 Bacmid production

DH10 EmBacY bacterial cells consisting of required bacmid was inoculated in 2 ml of LB medium containing 50 µg/ml Kanamycin, 10 µg/ml Gentamycin, 10 µg/ml Tetracycline and 30 µg/ml Chloramphenicol antibiotic selection and incubated overnight at 37°C.

Bacmid production was akin to standard plasmid mini-prep, except all spinning steps were done at 2,600 g for 30 min at 4°C to avoid shearing of the large bacmid DNA. Cells were spun, and the pellet was dissolved in 250 µl of resuspension buffer. Following that, 250 µl of the lysis buffer was added, inverted gently and incubated at room temperature for 5 min. Immediately 350 µl of neutralization buffer was added and inverted gently. The sample was kept on ice for 20 min and spun. The supernatant was then transferred to a new Eppendorf tube and spun again to ensure complete removal of the pellet.

The supernatant was carefully transferred to a fresh Eppendorf tube and an equal volume of isopropanol was added, inverted gently, incubated on ice for 30 min and spun. The tubes were then transferred into the laminar flow hood for subsequent steps to ensure sterility of Bacmid during transfection procedure. The supernatant was

discarded, and the pellet was washed with 70% ethanol. The final pellet was then carefully air dried to remove residual ethanol and dissolved in 20 μ l of pre-autoclaved TE buffer.

2.2.4.3 Transfection and initial virus (V_0) generation

Cells were dispensed in a 6 well plate to the cell density of 0.5 million cells/ml and incubated for 30 min at 27°C. Media supplemented with Anti-Anti but without FBS was aliquoted into two tubes, 100 μ l each. To the first tube, 10 μ l of bacmid DNA was added and to the other one 5 μ l of Cellfectin was added. Two solutions were incubated at room temperature for 5 min, mixed gently together and further incubated for 30 min at room temperature with addition of 1 ml of media.

The media from the 6 well plate was discarded carefully without disturbing the bottom of the plate after the initial 30 min incubation. DNA-Cellfectin mix was then added to the 6 well plate along with 2 ml of media and further incubated at 27°C for 5 hours. This media was then replaced with fresh media and incubated further for 60 hours at 27°C. The resulting supernatant from the well was collected and spun at 1500 rpm to derive V_0 virus.

2.2.4.4 V_1 generation and test purification

10 ml of cells grown to 2 million cells per ml were infected with 0.2 ml of V_0 virus. The infection cycle was carried on for 3 days at 27°C, 135 rpm and later spun at 1500 rpm for 10 min. The supernatant was saved as the next generation of virus or V_1 and the cell pellet was stored for a test purification.

A small-scale NiNTA purification was performed using 0.25 ml bed volume of NiNTA resin. The cells were suspended in 500 μ l C9 lysis buffer, transferred to a 2 ml Eppendorf tube and lysed using a sonicator with a microtip. For each construct, 0.25

ml pre-washed (with C9 wash buffer) NiNTA beads were added and incubated at 4°C for 30 min. The tubes were spun at 2000 rpm for 5 min and the supernatant was discarded.

The beads were then washed 3 times with 1 ml C9 Wash buffer by resuspending, spinning at 2000 rpm and discarding the supernatant. The complexes were finally eluted with 500 µl C9 elution buffer. The concentration of the elution fractions was measured with Nanodrop and 3µg of each fraction was loaded on a 12% SDS-PAGE.

To produce V2 for large-scale protein production, 50 ml of cells at 2 million cells per ml were infected with 1 ml of V₀₁ virus. The culture was grown for 3 days and spun at 2000 rpm. The resulting supernatant was stored in 50 ml Falcon tubes at 4°C, wrapped with aluminium foil to prevent exposure to light.

2.2.4.5 Viral titre determination

Viral titre for V1 was determined by endpoint dilution method. The original virus stock was diluted to 10⁻² with media. Ten-fold serial dilutions from 10⁻³ to 10⁻⁹ were done from the 10⁻² stock to 130 µl volume each and 1.3 ml of cells at 0.1 million/ml were added to every dilution.

The infected cells were dispensed in a 96 well plate, with each well containing 110 µl of the solution and each row containing one dilution (12 wells per dilution). The plate was incubated at 27°C for 72 hours.

GFP signal was then detected for each well, with a minimum of one fluorescing cell considered as GFP positive. The number of GFP positive wells per dilution was recorded. The viral titre was then determined by the following calculation.

$$\text{Viral titre} = \frac{\ln 2}{V(\text{virus}) \times 10^{\log(\text{dil50}) - \left(\frac{ab50 - 50\%}{ab50 - be50}\right)}} \frac{\text{pfu}}{\text{ml}}$$

V(virus) = Volume of virus added per well

dil50 = Dilution factor where more than 50% of the wells were positive

ab50 = Percentage of positive wells at dil50

be50 = Percentage of positive wells at dil50/10

2.2.4.6 Scale-up and expression

Cells were grown in either 2 L or 1 L vented Erlenmeyer flask to 40% volume of the flask to ensure adequate aeration. The cells were grown at 27°C, 135 rpm until a cell density of 2 million cells/ml was reached. The culture was infected with V₂ and incubated at 27°C. The amount of V₁ used was determined based on the following calculation.

$$\text{Volume of V2 virus required} = \frac{\text{MOI} \times \text{Vol (SF9)} \times \text{Cell density}}{\text{Viral titre}}$$

MOI = Multiplicity of infection/number of virus particle per cell

Vol(SF9) = Volume of culture to be infected

Infections were carried on 48-56 hours and the % GFP was recorded to keep track of the viral infectivity. The cells were then harvested by spinning at 1500 rpm and flash frozen using liquid N₂ and stored at -80°C.

2.2.4.7 Purification of POT1-TPP1-TIN2S

The frozen cell pellet was resuspended in C9 lysis buffer and sonicated. The lysate was spun at 20,000 rpm for 60 min and the resulting supernatant was filtered using a syringe-filter of pore size 0.2 µm. The filtered supernatant was loaded on a 1 ml GE

Histrap column at the flow rate of 0.8 ml/min. The column was then washed with 50 ml of C9 wash buffer to make sure all the unbound protein was washed away. The FPLC system was then programmed to wash the column with 1.3% of C9 elution buffer, which corresponds to 75 mM Imidazole. This wash step was added to further remove contaminants and was carried on for 2 C.V. The elution was performed as a gradient for 20 C.V and 5 C.V of 100% the C9 Elution buffer (Figure 2.1).

Figure 2.1: Schematic for NiNTA purification of C9; (C.V corresponds to column volume).

The fractions which contained the least contaminant were then pooled, concentrated with a 30 kDa MWCO concentrator. The sample was spun at the highest speed for 10 min at 4°C and injected in a GE Superose 6 10/300 column, pre-equilibrated with C9 size-exclusion buffer. The fractions which contained the complex were then collected and used for further analysis.

2.2.5 Biochemical studies

2.2.5.1 DNA binding studies

TRF1 and POT1 can bind to double-stranded and single-stranded telomeric DNA respectively. In this study, the DNA binding activity was analysed using the electrophoretic mobility shift assay (EMSA) to verify the purified protein/complex is

biologically active and to determine the optimal protein to DNA ratio for complex formation. The sequence of the probes used is given in the following table.

Table 2.6: DNA probes and sequences

Probe	Sequence (5' → 3')
hT10	Cy 5 TTAGGGTTAG
2R forward	GCG TGC <u>TTA GGG TTA GGG TTA</u> TAC
2R reverse	GTA <u>TAA CCC TAA CCC TAA</u> GCA CGC
4R forward	CTGGATGAATTCCTTCC <u>TTAGGGTTAGGGTTAGGGTTAGGGTTA</u> CCGAATTAACGATCTG
4R reverse	CAGATCGTTAATTCGGT <u>TAACCCTAACCCTAACCCTAACCCTAA</u> GGAAGGAATTCATCCAG

The forward and the reverse primers for 2R and 4R were mixed together to a equimolar ratio in TEN buffer (10 mM Tris pH 7.5, 0.1 mM EDTA and 100 mM NaCl) and heated to 95°C for 5min. The reaction was then gradually cooled down to room temperature and subsequently analysed by electrophoresis on a 1% agarose gel.

The amount of DNA was kept constant in binding reactions as 1 µM and 0.5 µM for 4R and hT10 respectively, and the amount of protein/complex was gradually increased.

The binding reactions were set up in 20 mM HEPES pH 7.5. The final NaCl concentration was adjusted to 150 mM and 50 mM for TRF1 and POT1-TPP1-TIN2 respectively. The reactions were analysed using 1% agarose gel or a 6% polyacrylamide gel. The conditions for EMSA for each sample is summarized (Table 2.7). EtBr gels were visualized using BioRad gel doc. Cy5 gels were visualized on a Typhoon7000 FLA.

Table 2.7: Conditions for EMSA

Sample	NaCl (mM)	Probe	Visualization
TRF1 + 4R	150	4R	EtBr
POT1-TPP1-TIN2 +hT10	50	hT10	Cy5

2.2.5.2 Crosslinking studies

Crosslinking studies were performed with POT1-TPP1-TIN2S to determine the optimal concentration of Glutaraldehyde (GA) required to retain DNA binding activity and completely crosslink the complex.

Based on the EMSA, the DNA binding master-mix reaction was set up such that the binding was completely saturated. The reaction was then aliquoted into seven tubes such that protein amount in each aliquot was 6 µg. GA was added to the aliquots in increasing amounts so that the final concentrations ranged from 0.0025% to 0.1% as summarized (Table 2.8). The reactions were then made up to 20 µl using a buffer containing 20 mM HEPES pH 7.5 and 50 mM NaCl.

Table 2.8: Glutaraldehyde crosslinking reaction

Component	No GA	0.0025%	0.005%	0.01%	0.025%	0.05%	0.1%
Complex+hT10	Final protein concentration 6 µg						
0.25% GA	-	-	-	-	2	4	6
0.025% GA	-	2	4	6	-	-	-
Buffer	Made up to 20 µl with buffer						

The reaction was incubated for 30 min on ice and then quenched with Tris-HCl pH 8 at a final concentration of 40mM. The reactions were then split into two, analysed

separately using 7% SDS-PAGE and 1% agarose gel to determine the optimal GA concentration.

2.2.5.3 Stoichiometry analysis

Stoichiometry of shelterin subcomplexes was determined using size-exclusion combined with multi-angle light scattering (SEC-MALS) and quantifying the relative intensities of subunits stained with SYPRO Ruby.

Concentration of the subcomplexes was adjusted to 1 mg/ml and spun at the highest speed before injecting 100 μ l of the sample to a Shimadzu HPLC attached with a Superdex 200 10/300 column. The column was pre-equilibrated with TRT/TPT size-exclusion buffer (Table 2.1) and the analysis was carried out at a flow rate of 0.25 ml/min at room temperature. For SEC-MALS analysis, the HPLC instrument was connected to a Wyatt DAWN 18 angle light scattering detector and Optilab refractive index detector. A commercial software called ASTRA® was used for data analysis and molecular weight determination.

For SYPRO ruby staining analysis, the subcomplexes were loaded on a 12% SDS-polyacrylamide gel with increasing total protein amounts of 1 μ g, 2 μ g, 3 μ g and 4 μ g. Gel electrophoresis was carried out until sufficient separation of the individual components was achieved.

The gel was subsequently fixed for 15 min x 2 with a fixation solution consisting of 50% methanol and 7% acetic acid. SYPRO Ruby stain was added after discarding the fixing solution. The stain was heated for 30 s thrice with resting intervals of 30 s, 5 min and 23 min on an orbital shaker for a total stain time of 30 min. Finally, the gel was washed with 10% methanol and 7% acetic acid solution for 30 min and imaged with a

BioRad ChemiDoc™ gel doc. Individual bands were detected and quantified using Image lab™ software for further analysis.

2.2.6 Negative staining

2.2.6.1 Sample preparation

The centre of the peak fraction from the final size-exclusion of TRF2-RAP1-TIN2S and TPP1-POT1-TIN2S were diluted to a concentration of 15 and 80 ng/μl respectively. TRF1-DNA sample was prepared as described (in section 2.2.5.1) and diluted to final concentration of 10 ng/μl in 20 mM HEPES pH 7.5 and 50 mM NaCl.

The sample was crosslinked by adding GA to the sample to a final concentration of 0.1%. The sample was spun at 15,000 rpm in a benchtop centrifuge for 10 minutes prior to grid preparation. The grids were made within 20 minutes of starting the crosslinking reaction.

2.2.6.2 Grid preparation

Homemade carbon coated copper rhodium grids (300 mesh) were glow discharged for 1 min and 4 μl of the protein sample was applied to the grid. After 1 min incubation time, the sample was blotted from the grid using Whatman filter paper. 2% uranyl acetate was immediately added to the grid, incubated for 1 min and carefully blotted without over drying the stain. The grid was further air dried for 10 minutes or longer before loading into the microscope.

2.2.6.3 Data Collection

Three different samples were analysed by negative staining. The images were collected on an FEI Technai T12 equipped with a 4K Eagle CCD camera at 49,000X

nominal magnification. The object pixel size was 2.1 Å/pixel. Micrographs were recorded at an electron dose of 20-25 e⁻/Å².

2.2.7 Cryo-electron microscopy

2.2.7.1 Sample preparation

Samples for cryo-EM were prepared with high purity and used immediately after purification procedures to avoid degradation or aggregation. POT1-TPP1-TIN2S and TRF2-RAP1-TIN2S samples were derived from the central peak fractions from gel-filtration and suitably diluted or buffer exchanged to keep the final glycerol content less than 5%. Additional sample preparation procedures for Holliday junction, Widom 601 sequence, nucleosome core particle (NCP) and crosslinking TRF2-RAP1-TIN2 are detailed in the following sections.

2.2.7.1.1 Holliday junction

Holliday junction (HJ) is a specialized DNA structure resembling a 'chicken foot' with four arms of complementary dsDNA. For this study, the four constituent sequences of the HJ were designed with equal length. The sequences with complementary regions are highlighted by the same colour as given below.

Sequence 1 5' **CCCCTGCGGCAACGTCGGAGCCCCTGGGTGCGGCCGGCTG** 3'

Sequence 2 5' **CAGCCGGCGCGACCCAGGGGGTCCCGCAGCGGACCGGGTC** 3'

Sequence 3 5' **GACCCGGTCCGCTGCGGGACGCTGCAGGAGTCTGCGCCCCG** 3'

Sequence 4 5' **CGGGCGCAGACTCCTGCAGCCTCCGACGTTGCCGCAGGGG** 3'

The four HJ sequences were mixed together in equimolar ratio to a final concentration of 10 mM each in TEN buffer (10 mM Tris pH 7.5, 0.1 mM EDTA and 100 mM NaCl) and heated to 95°C for 5min. The reaction was then gradually cooled down to room temperature and diluted to the required concentration before preparation of grids.

2.2.7.1.2 Widom 601 DNA

Widom 601 sequence is a chemically synthesized DNA sequence, identified by SELEX experiments to have high affinity for the histone octamer and strong nucleosome positioning ability [252]. It has been used widely for studying the biochemical and structural properties of NCP. In this study, Widom 601 DNA has been used directly for cryo-EM imaging with volta phase plate (VPP) and in reconstitution of NCP and subsequent EM analysis. The sequence of Widom DNA used is given below.

**5'ATCAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCA
CCGCTTAAACGCACGTACGCGCTGTCCCCGCGTTTTAACCGCCAAGGGGAT
TACTCCCTAGTCTCCAGGCACGTGTCAGATATATACATCGAT 3'**

Plasmid containing 145 bp of the Widom 601 DNA in pUC57 backbone was provided by Curt Alexander Davey, NTU. This plasmid was transformed as previously described (Section 2.2.1.1) into NEB 10- β cells with Ampicillin selection (100 μ g/ml).

A single colony was inoculated in 100 ml LB with Ampicillin and grown overnight at 37 °C and 220 rpm. The preculture was then transferred to a flask containing 1L of Terrific broth (TB) medium with Ampicillin and grown overnight at 37 °C, 220 rpm. The culture was spun down at 5000 rpm for 15 min and the resulting pellet was carefully resuspended in 50 ml of Alkaline lysis buffer 1 (25 mM Tris-HCl pH 8.0, 10 mM EDTA and 50 mM glucose). Cells were lysed upon the addition of 100 ml Alkaline lysis buffer 2 (0.2 M NaOH and 1% (w/v) SDS) and incubated on ice for 20 min with regular intervals of vigorous shaking to further assist the cell lysis.

Subsequently, 175 ml Neutralization buffer (2M acetic acid and 4M potassium acetate) was added to renature plasmid DNA and precipitate the genomic DNA. This solution was spun down at 10,000 g and the supernatant was further clarified through sterile

gauze filter. 100% isopropanol was added 0.52 times the volume of filtered supernatant to precipitate the plasmid DNA, which was pelleted by centrifugation at 16,000 g at 20°C for 30 min. Following that, the supernatant was discarded, and the pellet was dissolved in 40 ml of TE (10 mM Tris-HCl pH 8.0 and 50 mM EDTA). For the removal of RNA contamination, 250 µl of Rnase A was added, incubated overnight at 37°C and centrifuged at 10,000 g at 20°C for 30 min to remove insoluble particles.

In order to further purify the plasmid DNA, aqueous phase extractions using phenol and chloroform-isoamyl alcohol were performed. The supernatant after Rnase reaction was equally divided in two Teflon tubes, extracted with 10 ml of phenol and centrifuged at 27,000 g at 20°C for 20 min. The aqueous phase was transferred into a fresh Teflon tube and phenol extraction step was repeated two or three times until the interphase appears clear.

Chloroform – isoamyl alcohol mixture of 24:1 was added to the aqueous phase, spun down at 12,000 g at 20°C for 10 min and this process was repeated until the interphase appears clear. Finally, the aqueous phase was transferred to polycarbonate tubes and 0.2 volumes of 4 M NaCl and 0.4 volumes of 40% (w/v) PEG-6000 were added and incubated on ice for 30 min to precipitate the DNA. The mixture was subsequently spun at 27,000 g at 4°C for 10 min.

The pellet was dissolved in TE (10 mM Tris-HCl pH 8.0 and 0.1 mM EDTA) and the plasmid DNA was purified with a final round of ethanol precipitation by adding 0.1 volumes of 3 M sodium acetate and 2.5 volumes of 100% ethanol. This mixture was incubated on ice for 30 min and centrifuged at 15,000 g at 4°C for 10 min. The resulting pellet was air dried until excess ethanol was removed and dissolved in TE (10 mM Tris-HCl pH 8.0 and 0.1 mM EDTA).

The plasmid was digested with EcoRV restriction enzyme to release the Widom 601 145 fragment from the plasmid backbone. The final digestion reaction contained of 2 mg/ml plasmid DNA, 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 10 mM MgCl₂, 1 mM DTT and 300 U EcoRV per mg of plasmid and incubated at 37°C overnight. The completeness of the digestion was examined on a 1% agarose gel.

The plasmid backbone was separated from the insert by PEG precipitation. A final concentration of 6% (w/v) PEG-6000 and 0.5 M NaCl were added to the digestion reaction diluted to 0.5 mg/ml final concentration, incubated on ice for 30 min and Centrifuged at 27,000 g at 4°C for 20 min. The supernatant was subject to ethanol precipitation as described above. The final DNA pellet containing the Widom 601 DNA was sufficiently dried without over drying and resuspended in TE and used for preparing grids or NCP reconstitutions.

2.2.7.1.3 Nucleosome core particle

Purified lyophilized *Xenopus laevis* histones were provided by Lars Nordenskiöld, NTU. Histones were individually resuspended in unfolding buffer containing 7 M guanidine-HCl, 10 mM Tris-HCl (pH 7.5) and 10 mM DTT. The four histones H2A, H2B, H3 and H4 were then mixed with equal molar concentration of H3-H4 and approximately 25% excess of H2A and H2B as H2A-H2B dimer is relatively easy to separate from the octamer by size-exclusion compared to the H3-H4 tetramer. The histone mixture was then dialyzed overnight with three changes of refolding buffer consisting of 10 mM Tris-HCl pH 7.5, 2 M NaCl, 1 mM EDTA and 10 mM 2-mercaptoethanol. The refolded octamer was spun at highest speed to remove precipitates. Following concentration to a volume of 500 µl or less and another high-

speed centrifugation step, the histone octamer was purified using a Superdex 200 column to remove the excess H2A – H2B.

NCP was reconstituted using a decreasing salt gradient method described by Dyer *et al* [253]. For the NCP used in this study, the Widom 601 DNA and the histone octamer were mixed in equimolar ratio into a reaction mix with the final buffer condition of 20 mM HEPES pH 7.5, 1 mM EDTA, 1 mM DTT and 2 M KCl. The reconstitution reaction was dialyzed using buffers with 20 mM HEPES pH 7.5, 1 mM EDTA, 1 mM and step-wise decreasing salt concentrations of 0.85 M, 0.65 M, 0.45 M, 0.25 M and 0 M over a duration of 20 hours at room temperature. The reconstituted NCP was analysed on an 18% SDS-PAGE and 6% polyacrylamide gel prior to grid preparation.

2.2.7.1.4 Crosslinking and gel-filtration

This method of sample preparation was originally described by Chang *et al* [254] for preparation of APC/C complex for high-resolution cryo-EM. The advantages of this method over GraFix [255] include the reduced duration of exposure to the crosslinking agent and ability to monitor aggregates by checking the retention volume on the size-exclusion column. This method was utilized for screening TRF2-RAP1-TIN2S complex in cryo-EM. Purified complex of total amount 100 µg was crosslinked using 0.1% GA on ice for 20 min in a final volume of 100 µl. The reaction was quenched with 40 mM Tris pH 8.0. Quenched reaction was dialysed with TRT size exclusion buffer overnight to remove any traces of GA and finally purified with a Superose 6 column. The central peak fraction was used for grid preparation.

2.2.7.2 Grid preparation

Sample concentrations used for preparing cryo grids shown in this work are summarized below (Table 2.9). All samples were spun in a benchtop centrifuge at

15,000 rpm for 10 minutes prior to grid preparation. R2/2 Quantifoil grids were glow discharged for 1 min and carefully mounted on the Vitrobot tweezers. The Vitrobot was precooled to 4°C and set to 100% humidity. For most samples, a blot force of 1 and blot time of 2 s was used.

The grids were then plunged into liquid ethane container and carefully were transferred from liquid ethane to a grid box in liquid nitrogen compartment. Once prepared, the grids were continuously maintained at liquid nitrogen temperature.

Table 2.9: Concentration of samples used to prepare cryo grids

Sample	Concentration (mg/ml)
POT1-TPP1-TIN2S	0.8
POT1-TPP1-TIN2S + hT10 (DNA) [2:1]	0.8
TRF2-RAP1-TIN2S	0.6/0.2
601 Widom sequence	0.1
Holliday junction	0.1
NCP	1

2.2.7.3 Image acquisition

Images were acquired on a 200 kV FEI Arctica microscope with a Falcon2 direct electron detector and Volta phase plate. Images were acquired at a nominal magnification of 72000 X, with the object pixel size of 1.31 Å/pixel. Typically, a dose of 25-30 e⁻/Å was used to acquire images. Bright-field EM images were acquired in the defocus range of -2 to -3 µm. Phase plate images were acquired by first accurately focusing the beam on a carbon film, adjacent to the region of interest. Subsequently, an unexposed phase-plate position was activated by irradiating it with the electron

beam until sufficient phase shift (close to 90°) was achieved. Phase plate images have been band-pass filtered to remove information below 4 Å and above 500 Å with Fiji [256] for clarity.

2.2.8 Data Processing

Negative staining EM data was processed using EMAN2 [218], XMIPP [220] and RELION [221] software packages.

2.2.8.1 Micrograph screening and particle picking

Micrographs were imported in EMAN2 and screened based on the CTF rings and the quality of the image. Images with drift or astigmatism were removed. Particles were then picked using EMAN2 particle picker in SWARM mode. Particle coordinate files were saved for each micrograph as box files.

2.2.8.2 CTF determination and particle extraction

Micrographs converted to mrc format and the corresponding box files were transferred to a different directory for processing with RELION. CTF was then determined using CTFIND3, the output was checked to see if the calculated CTF matched the experimentally derived one. Individual particle images were extracted using the box files, with their metadata saved as 'particles.star' for subsequent processing.

2.2.8.3 2D Classification

For initial 2D runs, all particles were used as input and the number of classes was set such that each class contained 100-200 particles. Particle mask was applied based on the particle length roughly measured directly from the micrographs.

2D classification was repeated iteratively with different class numbers to see if the averaging technique yields similar results. The classes were then arranged based on

decreasing order of class distribution so that the classes with the most number of particles were at the top and classes with fewer particles or unrefined classes were at the bottom.

2.2.8.4 3D Classification and refinement

The initial model used for the three samples are given (Table 2.10). The reference map was modified such that it had the same dimensions as the box size of the particles. Particles selected from 2D were used as input and the 3D classification was set up such that there were at least 5000 particles per 3D class.

Based on the results of the 3D classification, particles from similar looking classes or individual class were saved as separate star files for 3D refinement.

Particles from 3D classification that had similar structure were combined and the 3D class density was used as a starting model along with the original starting models used for 3D classification.

Table 2.10: Starting models for negative staining.

Sample	Starting models (low pass filtered 60-80 Å)
TRF1+DNA	TRFH domain
TRF2-RAP1-TIN2S	TRFH domain and EM structure of TRF1 [257]
POT1-TPP1-TIN2S	POT1 OB domain [103] and Nucleosome core particle structure [258]

The default parameters for 3D classification and refinement are usually similar. Once the refinement was completed the FSC curve for the refinement was plotted to determine the resolution. Using the RELION printable command the Gold standard

FSC and the frequency ($1/\text{resolution}$) were listed and later plotted into a graph. The 0.143 and 0.5 FSC resolution criterion were recorded.

Chapter 3: Expression, purification and negative staining of TRF1

3.1 Overview

TRF1 is a negative regulator of telomere length, consisting of homodimerization domain (TRFH) in the N-terminus and Myb-like domain in the C-terminus connected by a flexible linker region. These domains were defined through limited proteolysis and sequence-based structure predictions [95]. Domains in TRF1 have been shown to interact with distinct proteins implicated in telomere dynamics. For instance, Tankyrase, a poly ADP ribose polymerase (PARP), has been proposed to indirectly control telomere length by interacting with and subsequently evicting TRF1 from telomeres through ubiquitin facilitated degradation [259].

Structures of the known domains in TRF1 were solved to high resolution using X-ray crystallography [95, 98]. However, the full-length protein was not structurally characterized due to the flexibility of the hinge region. Transitional or intermediate states can be classified out in electron microscopy (EM), as it is more tolerant to structural heterogeneity in the sample compared to other structure determination techniques [203]. Further, binding to DNA could stabilize the Myb domain and adjoining hinge regions, resulting in lower flexibility compared to free TRF1. Therefore, EM analysis of TRF1 bound to DNA was carried out using negative stain EM.

In order to obtain adequate purified sample for EM analysis, N-terminal 6X His-tagged human TRF1 was recombinantly purified from *E. coli*. Complex of TRF1 with DNA was assembled *in vitro* and subsequently analysed with negative stain EM. A recent study by Boskovic *et al* has utilized a similar approach and determined the negative staining structure of mouse TRF1 in apo and DNA bound states [257].

3.2 Expression and purification

Human TRF1 was recombinantly expressed and purified from bacteria based on the protocol established by Fairall *et al* [95] with alterations as described in methods (section 2.2.1). TRF1 purification essentially consists of His-tag affinity purification using NiNTA agarose, followed by an initial and a final size exclusion chromatography steps by Superdex 200 column and with a His-tag removal procedure in between the two size-exclusion steps. Purification summary and the final SDS-PAGE analysis of the purified protein is shown (Figure 3.1).

Figure 3.1: Summary of TRF1 purification. TRF1 was purified based on a protocol described in [95] as summarized in the flow chart (left). Recombinant TRF1 with His-tag was subject to affinity purification with NiNTA and size-exclusion chromatography using Superdex 200. TEV digestion was performed to cleave the affinity tag. The incompletely digested product (if any) and TEV protease were removed by 'reverse NiNTA'. The second and final round of size-exclusion chromatography using Superdex 200 yields purified TRF1 as shown on an SDS-PAGE (right).

Figure 3.2: TRF1 size-exclusion analysis. (A) Size-exclusion profile of purified TRF1 on Superdex 200 10/300 with peak fractions highlighted **(B)** SDS-PAGE analysis of corresponding peak fractions **(C)** Chromatographic separation of standard proteins on Superdex 200 10/300 and molecular weights of corresponding protein standards. Red arrow indicates the band corresponding to TRF1. Lanes are labelled with the corresponding fraction number.

Mass spectrometry analysis estimated the molecular weight of monomer as 50 kDa (Appendix 1). In the final size-exclusion chromatography step, elution volume of the protein was 11.5 ml on a Superdex 200 10/300 column (Figure 3.2). The apparent molecular weight of purified TRF1 was estimated at approximately 300 kDa based on

the protein standards. However, the molecular weight of dimeric TRF1 is 100 kDa, suggesting the elution volume of TRF1 does not correlate with the standards of globular proteins. This is a consequence of the elongated shape and larger radius of TRF1 due to the flexible hinge region. Dimensions of the purified TRF1 molecule under negative stain EM were later confirmed to be indicative of its 100 kDa dimeric form, as described in the following section (Section 3.4).

3.3 DNA binding studies

In order to determine the optimal protein to DNA ratio for EM sample preparation, DNA binding activity of purified TRF1 was analysed with electrophoretic mobility shift assay (EMSA). The practical limitation of sample size for EM is considered 100-200 kDa [260]. Therefore, a four telomeric repeat DNA probe (4R) was designed such that two individual TRF1 dimers could concomitantly bind and constitute a 200 kDa complex. The probe was also designed with flanking mixed DNA sequences of 16 base pairs (bp) on both ends to aid proper annealing of the telomeric repeat DNA.

The binding affinity of TRF1 to telomeric DNA previously determined from EMSA is 3.2 nM, and a non-specific DNA binding constant of 3.6 μ M [99]. In order to determine DNA binding in the working concentration for EM analysis, EMSA was performed with 1 μ M DNA. With the step-wise increase of TRF1, intensity of free DNA decreased, and two distinct band-shifts were observed in the EMSA (Figure 3.3).

Binding reactions were analysed using 6% polyacrylamide gel (as described in Appendix A1). First shift denotes the binding of the first TRF1 dimer to the DNA, and the second slower migrating band corresponds to two TRF1 dimers bound to DNA (Figure 3.3). The assay also indicated that TRF1 can simultaneously exist as single

and double dimer complexes with no preferential cooperative binding to DNA (Figure 3.3, lane 3, 4).

Figure 3.3: TRF1 DNA binding analysis. EMSA on a 6% polyacrylamide gel with 1 μ M of 4 telomeric repeat DNA and increasing molar concentrations of TRF1, where the final buffer condition of the binding reactions was 20 mM HEPES pH 7.5 and 150 mM NaCl. Protein to DNA ratio used for each reaction is labelled at the top of the gel and lanes at the bottom. Red and blue arrows indicate one TRF1 dimer-DNA complex and two TRF1 dimers-DNA complex respectively.

Complete saturation of DNA into the two-dimer complex was observed at five-fold molar excess of TRF1 to DNA (Figure 3.3, lane 6). Increasing the protein concentration beyond this ratio resulted in non-specific binding of TRF1 to the flanking mixed DNA of 16 bp on each side, as evidenced by an additional higher bandshift (Figure 3.3, lane 7). Therefore, molar ratio of TRF1 to 4R ratio of 5:1 was identified as the optimal ratio for sample preparation of EM grids.

3.4 Negative stain EM analysis of TRF1 and telomeric DNA complex

3.4.1 Sample preparation

Figure 3.4: Negative staining EM of TRF1 alone. Comparison between **(A)** Uncrosslinked complex and **(B)** crosslinked complex with representative particles in the panels below each image. Scale bars and the box size are 50 nm and 15 nm respectively.

Initial screening of purified TRF1 showed the sample was heterogeneous and dissociated (Figure 3.4 A). Crosslinking the sample with glutaraldehyde (GA), prevented dissociation of particles due to dehydration artefacts (Figure 3.4 B).

Dimensions of individual particles measured along the longest axes were 8 nm x 5.5 nm. Size of the particles agreed with dimensions of dimerization domain TRFH, indicating the sample consists of dimeric TRF1. TRF1 + 4R complex was reconstituted as described in the previous section and crosslinked with 0.1% GA prior to grid preparation. However, the particles did not exhibit the expected length as they were mostly in the size range of a single dimer rather than two dimers. Only a few particles with the dimensions of the two-dimer complex were observed (Figure 3.5).

Figure 3.5: Negative stain EM of TRF1 in complex with 4 telomeric repeats. (A) Selected area of crosslinked TRF1 – DNA complex stained with uranyl acetate is shown. The scale bar is 50 nm. Representative particles are shown for **(B)** two dimer complex **(C)** one dimer complex where the box size is 20 nm.

GA is a protein-protein crosslinking agent, known to react with primary amines of the lysine groups on the protein surface and does not crosslink protein to DNA [261]. In order to investigate the effect of GA crosslinking on DNA binding activity of the protein,

TRF1 + 4R DNA EMSA analysis was compared with and without 0.1% GA crosslinking for 20 min.

At the same protein concentration where the free DNA probe disappears in the native binding reaction, free DNA was still present in the crosslinked EMSA (Figure 3.6, lane 15). Interestingly, GA crosslinking could stabilize the non-specific bandshifts found to migrate slower than 2-dimer complex (Figure 3.6, lane 18 vs lane 9).

Figure 3.6: TRF1 DNA binding and GA crosslinking. EMSA for TRF1 binding to 1 μ M of 4 repeat telomeric DNA with and without GA. Increasing concentration of TRF1 from 1 to 8 times is indicated in the top. **(A)** Native EMSA with concentration of TRF1 sequentially increased in binding reactions **(B)** EMSA under crosslinking conditions with 0.1% GA added to the same reactions to check the effect of crosslinking on DNA binding shows partial loss of DNA binding activity. Blue and red arrows indicate 1:1 and 2:1 complex of TRF1 dimer to 4 repeat DNA respectively.

Furthermore, even an 8-fold excess of TRF1 was unable to completely bind the free DNA (Figure 3.6, lane 18). The detrimental effect of GA on the DNA binding activity of TRF1 could arise from inactivation of lysine in the active site of the protein.

3.4.2 Initial 2D classification

For initial 2D classification, particles were picked and sorted with a software called XMIPP [220]. Inspection of 2D class averages revealed only two out of 20 classes with the expected size of two TRF1 dimers bound to DNA (150-160 Å). Most classes had a maximum length of 80 Å. Hence, EM analysis confirms that GA crosslinking disrupts DNA binding of TRF1 as observed with EMSA analysis. Upon crosslinking, the saturated 2-dimer complex becomes a heterogeneous mixture of 2-dimer, 1-dimer DNA complexes and free TRF1 in varying amounts.

Figure 3.7: 2D classification of TRF1 - 4 repeat telomeric DNA complex (A) Negative stain electron micrograph with individual particles circled in black and blue. **(B)** Representative 2D class averages. Classes with similar dimensions are shown in **(C)**, **(D)**, **(E)**, **(F)**. Where **(E)** and **(F)** represent the one and two dimer complexes respectively with dimensions shown in Angstrom. Scale bar is 20 nm.

3.5 Summary

TRF1 was purified and its DNA binding activity with 4 telomeric repeat DNA was analysed with EMSA. The crosslinking procedure required to prevent dehydration artefacts and maintain the intactness of particles during negative staining had a

detrimental effect on DNA binding activity of the protein. 2D classification analysis of the complex indicated that it was highly heterogeneous, with only a fraction of the entire dataset containing particles of expected length and size. EMSA assay and EM analysis determined the final crosslinked sample used in this study was composed of four different components: TRF1 alone, one TRF1 dimer bound to the DNA, two TRF1 dimers bound to the DNA and the free DNA. Therefore, a complex of TRF1 with 4 telomeric repeat DNA assembled *in vitro* was determined unsuitable for EM analysis as the process of preparing grids and crosslinking made the sample heterogeneous.

Chapter 4: Expression, purification and negative stain EM analysis of TRF2-RAP1-TIN2S

4.1 Overview

TIN2 is known as the ‘bridging’ component of shelterin, as it can simultaneously mediate the interaction between single-strand DNA binding POT1-TPP1 heterodimer and double-stranded telomeric DNA binding proteins TRF1 and TRF2 [110, 262, 263]. Biochemical studies and deletion experiments have further indicated that TIN2 stabilises the binding of its interacting partners to telomeric DNA. Deletion of TIN2 results in either partial or complete loss of TRF1, TRF2 and POT1-TPP1 heterodimer from the telomeres and compromises their function [23, 76, 110, 262-265].

Interaction of TIN2 with TRF1 prevents its ADP-ribosylation by Tankyrase [171] and protects TRF1 from ubiquitin-mediated proteolysis [259]. Stabilization of the POT1-TPP1 heterodimer by TIN2 protects telomeres against ATR pathway by preventing RPA binding and subsequently initiating a signalling cascade [77]. TIN2 can also indirectly repress ATM signalling pathway by stabilising TRF2 [265]. Additionally, several known TIN2 disease mutants causing telomere dysfunction did not impair the expression and localisation of other shelterin proteins [127, 128] implying that stabilisation of its interacting proteins may not be the only mode of TIN2 function, and additional underlying mechanisms could exist.

Obtaining a complete understanding of the complex role played by TIN2 at telomeres has been challenging due to the lack of structural information. So far, the only available structure of this protein is a 20-amino acid peptide motif that interacts with TRF1 [97]. The following sections detail the attempts to purify TIN2 isoforms individually with

different tags, and subsequently co-purification of TIN2 short isoform with its interacting partners TRF2 and RAP1 as a complex. A negative stain EM structure of the TRF2-RAP1-TIN2S complex is also presented.

4.2 Bioinformatics analysis of TIN2

The amino acid sequence of full-length TIN2 was analysed using bioinformatics tools. Three different sequence based secondary structure prediction tools: PSIPRED [266], Predict protein [267], and J-Pred [268] gave a consensus prediction. N-terminal 1-200 amino acid region was predicted to contain 8-9 α -helices, and the rest of the protein was mostly unstructured with a few low probability helices (Figure 4.1 A).

Interestingly, the N-terminal 1-200 amino acid region has been reported to interact with TRF2 and TPP1 [26, 97, 110, 262, 263]. The 3D structure of TIN2 N-terminus (1-200) amino acid region was predicted using I-TASSER, an *ab initio* structure prediction software [269, 270]. The top five *ab initio* models sorted based on C-score were in agreement with the previous secondary structure predictions (Figure 4.1B), with each containing 6-9 α -helices.

4.3 TIN2 constructs

Purification of full-length TIN2 was not established in literature at the commencement of this study. Although a few studies described efforts to purify full-length TIN2, none of them had sufficient purity or quantity for structure analysis using electron microscopy (EM) or X-ray crystallography [23, 76, 171, 271]. Initial attempts to express and purify several different TIN2 constructs from bacteria proved challenging as most of the construct tested could not be satisfactorily purified. A shorter isoform of *Tinf2*, codon optimised for bacterial expression showed some promising results. Constructs tested as summarised below of which the highlighted TIN2L, MBP tag TIN2L and lastly GST-TIN2S are discussed in more detail in this section (Table 4.1).

Table 4.1: List of all TIN2 constructs tested for purification

Construct	Affinity tag
TIN2 long isoform (1-451)	6X Histidine
TIN2 long isoform (1-451)	6X Histidine with a longer linker
TIN2 long isoform (1-451)	6X Histidine with MBP fusion
TIN2 short isoform (1-354)	6X Histidine with GST fusion
TIN2 structured region (1-200)	6X Histidine
TIN2 structured region (1-200)	6X Histidine with MBP fusion

Both TIN2L and MBP-TIN2L were expressed with an N-terminal 6X His-tag. However, during NiNTA purification, there was an insufficient separation of the protein of interest from contaminating proteins or excessive degradation (Figure 4.2 A and B).

Figure 4.2: Unsuccessful purification attempts of TIN2 constructs. Purification of TIN2 was challenging as most of the constructs tested were not successful. Two constructs out of five that did not succeed are shown **(A)** Test purification of TIN2L showing insufficient separation of desired protein from the contaminants **(B)** Test purification of MBP fused TIN2L showing degradation.

Particularly in case of MBP-TIN2L, the intensity of the contaminant bands of lower molecular weight increased over time, indicative of degradation products. In an attempt to remove contaminants from MBP-TIN2L, a size-exclusion chromatography purification step was added which proved ineffective (Figure 4.3).

Figure 4.3: Size-exclusion chromatography of MBP-TIN2L. Elution fractions from the NiNTA purification concentrated and injected in Superdex 200 10/300 column. Two major peaks were obtained. The lower molecular weight peak contained an MBP degradation band. **(A)** Size-exclusion profile and **(B)** the corresponding fractions analysed using SDS-PAGE are shown. The red arrow and blue arrow indicate the band corresponding to MBP-TIN2L in the input lane and the band corresponding to MBP degradation respectively.

As an alternative purification strategy, the shorter isoform of TIN2 tagged with an N-terminal GST-tag was used. TIN2S (1-354) contained all the essential interaction sites with shelterin proteins and lacked approximately 100 amino acids of the predicted unstructured region found in C-terminus of TIN2L (1-451). GST affinity purification of the construct followed by size-exclusion showed promising results of higher purity than other constructs previously tested (Figure 4.4).

Figure 4.4: Purification of GST-TIN2S. Size-exclusion run of TIN2S after GST purification was better in comparison to the previous constructs tested. Some lower molecular weight contaminants were not separated during the purification. The induced and uninduced samples indicate the position of the GST-TIN2S band. Red arrows indicate the band corresponding to GST-TIN2.

4.4 Purification of TRF2-RAP1-TIN2S

Although GST-TIN2S was better than the other TIN2 constructs, the final purified protein was still not of sufficient purity, containing lower molecular weight

contaminants. These bands were suggestive of degradation as their intensity increased over time and their size corresponded to the GST-tag (roughly 25-26 kDa).

In order to aid the protein purification by providing separation from the contaminating degradation bands and to test the stability of protein interactions involved, TIN2S was co-purified from bacterial cells as a subcomplex with its interacting partner TRF2-RAP1. In addition to GST-tag on TIN2, each of the individual constructs were expressed with His-tags and TEV cleavage sites. The three constructs were grown and induced separately, and the pellets were suspended together for co-purification. Test expression for each construct was performed to identify the relative position of these proteins (Figure 4.5).

Figure 4.5: Test expression of TRF2/RAP1/TRF1/TIN2S. Before and after induction samples for each construct shows the relative position of bands corresponding to each protein. Note that the TIN2S construct was fused with GST affinity tag and therefore was close to 70 kDa in size.

After optimization, bacterial cells expressing TRF2, RAP1 and TIN2S were mixed roughly in a volume ratio of 2:1:6 to compensate for possible degradation of TIN2 and high expression levels of RAP1. Affinity purification using 5 ml HisTrap column was performed under an imidazole concentration gradient (Figure 4.6).

All elution fractions contained three proteins, which were pooled. As GST dimerization could interfere with structure determination, the affinity-tag was cleaved with overnight TEV digestion (Figure 4.6 B). Similar to TRF1 (Section 3.2) 'reverse' NiNTA step was used to remove the incompletely digested protein with affinity-tag still intact and the His-tagged TEV protease (Figure 4.6 C). Final size-exclusion chromatography step with a Superdex 200 column was used to separate the pure TRF2-RAP1-TIN2S complex from the degradation bands (Figure 4.7). Fractions D12 and D11 were collected for subsequent analysis.

Figure 4.6: NiNTA purification and subsequent removal of His-tag in TRF2-RAP1-TIN2S. NiNTA elution fractions contained TRF2, RAP1 and TIN2S **(A)** Run profile and **(B)** corresponding fractions analysed using SDS-PAGE are shown. **(C)** Reverse NiNTA purification of TRF2-RAP1-TIN2S complex after TEV cleavage. Complex with no affinity tags after TEV digestion is in the flow through fraction, while incompletely digested complex with intact His-tags and the TEV protease are found in elution fraction. The red, blue, white and green arrows indicate TRF2, RAP1, GST-TIN2 and TEV protease respectively.

Figure 4.7: Size-exclusion purification of TRF2-RAP1-TIN2S. The complex was separated from lower molecular weight contaminants using HiLoad Superdex 200 16/60. **(A)** Size-exclusion profile and **(B)** corresponding fractions analysed using SDS-PAGE are shown. The red, blue and white arrows indicate TRF2, RAP1 and TIN2S respectively.

4.5 Analytical size-exclusion study

A classic analytical size-exclusion study was performed with the following objectives:

(1) To test if TRF2-RAP1-TIN2S TRF1 can be stably assembled *in vitro* (2) To test the DNA binding ability of purified subcomplexes. A Superose 6 10/300 analytical size-exclusion column was used for this study. The samples used for this study are summarized below (Table 4.2). TRF2-RAP1-TIN2S and TRF1 were used as controls.

TRF1 used in this study was purified as discussed in the previous chapter

(Section 3.2). The individual components were mixed in the molar ratios given in the table and incubated on ice before injecting on to the column. The elution volumes of resulting peaks were recorded for each run. The peak fractions were analysed using SDS-PAGE and additionally agarose gels when DNA was added (Figure 4.8).

Table 4.2: Size-exclusion studies using TRF2-RAP1-TIN2S

Run no.	Components	Ratio	Peak Elution volume(ml)
1	TRF2-RAP1-TIN2S	control	11.56
2	TRF1	control	14.10
3	TRF2-RAP1-TIN2S + 2.5 repeat DNA	1:1.4	11.56
4	TRF2-RAP1-TIN2S + TRF1	1:1.2	11.27
5	TRF2-RAP1-TIN2S + TRF1 + 4 repeat DNA	1:1:1.2	10.95

Figure 4.8: Analytical size-exclusion study using TRF2-RAP1-TIN2S complex. (A) Overlay of five size-exclusion profiles showing peak migration upon addition of TRF1 to TRF2-RAP1-TIN2S complex with and without DNA to form higher-order complexes (B) Chromatographic separation of standard proteins on Superose 6 10/300 and molecular weights of corresponding protein standards. (C) Peak fractions from TRF2-RAP1-TIN2S and TRF2-RAP1-TIN2S TRF1 with two (left) or four (right) telomeric repeat DNA compared with free DNA controls, analysed on 1% agarose gels. (D) Comparison between peak fractions of TRF2-RAP1-TIN2S and TRF2-RAP1-TIN2S TRF1 using 10% SDS-PAGE. Arrow marks of the colour red, green, blue and white indicate TRF2, TRF1, RAP1 and TIN2S respectively.

Comparison of the peak fractions with molecular weight standards indicated a higher apparent molecular weight on the column as in the case of TRF1 (Chapter 3). The

complexes did not elute at void volume of the column but had a high apparent molecular weight of more than 660 kDa (Thyroglobulin). Comparison of elution volumes of TRF2-RAP1-TIN2S and TRF2-RAP1-TIN2S TRF1 indicates a peak shift and decrease in elution volume, with excess TRF1 eluting as a shoulder peak.

Analysis of the fractions from the two runs using SDS-PAGE affirms the observation with an additional band corresponding to TRF1 found in the four-protein complex, and excess TRF1 in the shoulder fraction (Figure 4.8 D, lanes 7-11). Excess TRF1 peak was also found to elute at the same volume as control TRF1.

DNA binding activity of TRF2-RAP1-TIN2S was examined by comparing free DNA with the peak fractions of the complex with the telomeric DNA. The peak fractions of complex bound to DNA were visualized as faint slow migrating bands compared to control DNA on an agarose gel (Figure 4.8 C, left).

To address if TRF2-RAP1-TIN2S-TRF1 complex could bind DNA in a stable manner, its interaction with four telomeric repeat DNA was tested. The reasoning behind using four telomeric repeat DNA was to allow sufficient binding sites for each TRF2 and TRF1 dimers in the complex. The four-protein complex produced a significant shift in elution volume compared to controls.

Peak fractions analysed on agarose gel suggests protein-DNA complex formation, with slower migrating bands compared to control DNA. Additionally, a band of lower size corresponding to the excess TRF1-DNA complex was also observed (Figure 4.8 C, right, fraction A14).

In conclusion, the results of analytical size-exclusion chromatography of TRF2-RAP1-TIN2S complex suggests that it can interact with DNA. Furthermore, it was shown to

interact with TRF1, with the ability to stably form a four-protein shelterin subcomplex in the presence or absence of DNA.

4.6 TRF2-RAP1-TIN2S stoichiometry analysis

The stoichiometry of the individual proteins in the TRF2-RAP1-TIN2S complex was determined using a combination of size-exclusion chromatography – multi-angle light scattering (SEC-MALS) and SYPRO Ruby staining (Figure 4.9).

Differential refraction (dRI) signal for the complex was observed as a single peak with minor shoulder peaks, which are indicative of degradation and therefore were excluded from analysis (Figure 4.9 A). The absolute molecular weight of the complex was calculated as 222 kDa using SEC-MALS, which is closest to the stoichiometry of 2:2:1 (TRF2: RAP1: TIN2) with an expected molecular weight of 235 kDa. According to this stoichiometry, each TRF2-RAP1 dimer would be associated with one TIN2 molecule.

Additionally, SYPRO Ruby and Coomassie blue R-250 staining were then used to determine the relative abundance of proteins in the complex (Figure 4.9 D). The intensities of the individual proteins were determined and normalized with the intensity of TIN2S as a reference. The ratio of TRF2:TIN2S and RAP1:TIN2S was calculated close to 1.5.

However, as the TRF2-RAP1 complex exists as a dimer of dimers in solution [93, 95, 98, 100], the ratio of 1.5:1.5:1 of TRF2-RAP1-TIN2 is not plausible. As SYPRO ruby and Coomassie blue can differentially stain proteins based on the size and the amino acid composition, a likely explanation for deviation from SEC-MALS calculation could be due to the staining differences between the individual proteins. The stoichiometry

determined by SEC-MALS, 2:2:1 of TRF2, RAP1 and TIN2S respectively, is therefore more likely to be the correct estimate.

Figure 4.9: Stoichiometry of TRF2-RAP1-TIN2S. (A) SEC-MALS to determine the absolute molecular mass of the complex. Differential refractive index signal & calculated molecular weight shown in blue and red respectively (B) Comparison of the measured and expected molecular weight (C) Purified TRF2-RAP1-TIN2S complex stained with SYPRO Ruby. Relative stoichiometry analysis of the complex using (D) SYPRO Ruby (E) Coomassie blue R-250 staining. Intensity of the bands were normalized with TIN2 signal and compared. Error bars represent the range of the data and $n = 4$.

4.7 Negative stain EM analysis of TRF2-RAP1-TIN2S

The TRF2-RAP1-TIN2S complex was analysed by negative staining EM. Native uncrosslinked sample was not intact through the negative staining procedure and hence 0.1% of GA was used to crosslink the complex for a total duration 20 min.

Figure 4.10: Negative staining of TRF2-RAP1-TIN2S. Glutaraldehyde crosslinking was required to keep the complex intact. **(A)** The uncrosslinked sample contained heterogeneous particles with 'fragmented' appearance **(B)** Crosslinked sample with uniform and intact particles. Examples of individual particles circled in black. Scale bars are 50nm.

Particles in the uncrosslinked sample had variable size and a dissociated appearance (Figure 4.10). By contrast, the crosslinked samples had a uniform size of approximately 15 nm X 11 nm and several repeating orientations or views. The EM

analysis of the TRF2-RAP1-TIN2S complex was carried out in collaboration with post-doctoral fellow Dr Vinod K. Vogirala.

4.7.1 Data collection and 2D classification

Images had uniform staining and well-distributed particles, with roughly 200-300 particles per micrograph. A total of 110 micrographs were recorded and down-sampled for 2D class averaging analysis. The following table summarises the data collection and processing details (Table 4.3).

Distinct views such as the inverted '8' view (Figure 4.11 C and D) and Y-shaped view (Figure 4.11 F) were observed in the 2D classes. Re-iteration of 2D classification by altering the total number of classes produced similar views. By visual inspection and particle sorting, the particles from well-refined classes were selectively saved for subsequent processing.

Table 4.3: TRF2-RAP1-TIN2S negative stain-EM data collection and processing

Molecular weight(kDa)	235	Total Dose (e/Å ²)	30
Sample concentration	15 µg/ml	Defocus range (µm)	-2 to -3
Microscope	Technai T12	Total images recorded	110
Detector	Eagle, CCD	Total Particles picked	23065
Pixel size(Å)	2.1	Box size(pixels)	168
Grid type	Homemade, 10 nm C	Resolution (Å)	28

Figure 4.11: 2D classification of TRF2-RAP1-TIN2S. (A) Field image of TRF2-RAP1-TIN2S showing uniform particle distribution. Selected individual particles circled in white and black. (B) Representative classes with highest particle distribution. Classes with similar views are shown in (C), (D), (E), (F) and (G). Scale bar and box size are 50 nm and 35 nm respectively.

4.7.2 3D classification and refinement

Particles were divided into three 3D-classes, and the previously reported full-length TRF1 negative staining structure or TRF2 dimerization domain low-pass filtered to 80 Å were used as the starting model [257]. One of the well refined classes with recognizable features was then used for 3D refinement with no applied symmetry to obtain the final map.

Global resolution of 28 Å was estimated from the 0.143 Fourier Shell Correlation (FSC) criterion. Recognizable features such as the ‘8 view’ and the ‘inverted-Y’ view,

previously seen in the 2D classes, were observed once again as different orientations in the refinement. Dimensions of the refinement were consistent with the observed particle size (Figure 4.12).

Figure 4.12: Negative stain EM structure of TRF2-RAP1-TIN2S. (A) Three views of a refined EM density map after ML3D classification in RELION. (B) FSC plot for the EM density map (C) Different views and dimensions of the refined map and the corresponding 2D classes.

4.8 Summary

Purification of different constructs of TIN2 was attempted in order to gain better understanding of the ‘bridging’ protein. Of the constructs and strategies tested, co-purification of GST-TIN2S along with its interacting partner TRF2-RAP1 was successful. The TRF2-RAP1-TIN2S complex was isolated and its interactions with TRF1 and DNA were analysed using size-exclusion chromatography. The complex (TRF2-RAP1-TIN2S) as expected was found to interact DNA. Additionally, it could

interact with TRF1 in the presence and absence of DNA to stably form the TRF2-RAP1-TIN2S-TRF1 complex. This four protein subcomplex of shelterin is expected to coat long stretches of double-stranded telomeric tracts [160, 272]. SEC-MALS analysis of TRF2-RAP1-TIN2 indicated a 2:2:1 stoichiometry. Negative stain EM analysis of the TRF2-RAP1-TIN2S complex gave a 3D refinement map of 28 Å resolution. The reconstruction had views that were identifiable in the 2D classes as well. The dimensions of the TRF2 RAP1 TIN2S were 15 nm x 11 nm x 7.5 nm without any recognizable symmetry.

Chapter 5: Purification and negative stain EM analysis of POT1-TPP1-TIN2 complex

5.1 Overview

Human telomeres contain conserved single-stranded G-rich extension at 3' terminus [55, 273, 274], which serves as a substrate for extension by telomerase [275]. This single-stranded telomeric DNA is bound in a sequence-specific manner by the shelterin protein POT1 [22, 105]. POT1 prevents the overhang from being incorrectly recognised as DNA damage by repression of ATR pathway [276]. TPP1 forms a heterodimer with POT1, increasing the DNA binding affinity of POT1 and the processivity of telomerase [106]. TPP1 has a short stretch of amino acids called the TEL patch which has been shown to interact with telomerase and aid its recruitment [72, 277, 278].

As mentioned previously, deletion of TIN2 results in the loss of POT1 and TPP1 at the telomeres and subsequent activation of DNA damage pathways [77]. Telomerase recruitment is also TIN2-mediated through its link with TPP1 [111]. Therefore, the shelterin subcomplex TPP1-POT1-TIN2 is collectively the telomerase recruitment complex.

Study of TPP1- POT1 -TIN2 complex was therefore undertaken to test if the complex can be stably assembled and to understand the molecular architecture of the complex. Previous purification attempts of TPP1 and POT1 using bacterial expression system in Rhodes' lab were observed to be sub-optimal. Additionally, post-translational modification by phosphorylation of Ser 255 residue on TPP1 is essential for its interaction with telomerase [108]. Therefore, insect cells were used for eukaryotic expression of the recruitment complex.

Multi-Bac system is a baculovirus based technique used to express subunits and purify large multi-protein complexes from a single virus [279]. This system also consists of a yellow fluorescent protein (YFP) reporter gene that allows for real-time monitoring of the protein expression and stage of viral infection.

MultiBac bacmids consisting of 10X His-tagged *Pot1* with different combinations and lengths of *Tinf2* and insect cell codon optimized *Tpp1* were cloned and provided by post-doctoral fellows' Dr Vasuki Chellamuthu and Dr Alessandra Beltrami from Daniela Rhodes' lab. The following sections detail the purification of these constructs expressing POT1-TPP1-TIN2 in insect cells and the subsequent EM analysis of the complex in negative staining.

5.2 Purification of POT1-TPP1-TIN2S

5.2.1 Expression of POT1-TPP1-TIN2S

Bacmids containing desired constructs were amplified and isolated from bacteria and transfected into SF9 cells using the protocol described in methods (section 2.2.4). The V0 virus generated for each construct was obtained from the transfected wells and was used to infect 10 ml of cells at a density of 2 million. The infection cycle was carried on for 3 days and the % YFP was determined (Figure 5.1).

A total of six constructs were screened, details are given in methods (section 2.2.4). Percentage YFP denotes the number of fluorescing cells out of the total number of cells. The % YFP was monitored over the course of 3 days for all the constructs (Figure 5.2 B). At day two, constructs 7 and 9 had very low YFP signal. However, by the end of day three, all constructs had greater than or equal to 80% YFP signal (Figure 5.2 B).

Figure 5.1: Monitoring infection cycle using YFP. Constructs were designed with a YFP marker that allows the infection cycle and protein expression to be monitored real-time. Cells imaged in bright-field mode (left) and fluorescent mode (right).

At the end of the viral infection cycle, cultures of each construct were spun down to extract the supernatant consisting of the V1 virus and cell pellet for test purification (Figure 5.2 A). Constructs C7 and C9 had distinct bands for POT1, TPP1 and TIN2L or TIN2S respectively as shown in lane 3 and 5. These constructs were therefore chosen for large-scale expression and purification. POT1 was the only subunit of the complex that was designed with a 10X His-tag, therefore co-purification of TPP1 and TIN2S along with POT1 by NiNTA affinity was an indication of stable complex formation.

A **B**

Figure 5.2: Test expression of baculoviral constructs. (A) Test NiNTA purification of all constructs with elution fractions analysed using SDS-PAGE. The red arrow indicates POT1, the green arrow indicates TPP1L, Blue and white arrows indicate TIN2L and TIN2S respectively **(B)** Percentage YFP for each construct plotted 48 and 72 hours post-infection.

5.2.2 Viral titre determination and large-scale expression

Viral titre is an important factor as it can be used to determine the optimal amount of virus required for infecting large-scale cultures. End-point dilution method (section 2.2.4.5) was used to determine the viral titre. Since the viral infection can be monitored real-time through YFP production, the YFP signals were recorded for each dilution. Table below summarizes the viral titre calculation for V1 generation virus of Construct 9 (Table 5.1)

Table 5.1: Viral titre determination for V1, Construct 9.

Dilution	% Cells infected
10^{-5}	75
10^{-6}	58
10^{-7}	58
10^{-8}	33
10^{-9}	25

The viral titre was determined to be 1.4×10^9 plaque forming units (Pfu). Based on the titre, the volume of virus required to infect 800 ml of culture at cell density 2 million/ml with MOI 1 was calculated as 1.142 ml. MOI parameter denotes the multiplicity of infection or the number of virus particles per cell. The infected cultures were incubated for 65-72 hours post-infection before harvest.

5.2.3 Purification of TPP1- POT1 -TIN2S (Construct 9)

A two-step protocol with affinity purification using a single 10X His-tag on POT1 and subsequent size-exclusion was applied to purify the three-protein complex. However,

purity of the complex was affected by higher molecular weight contaminants. (Figure 5.3).

Figure 5.3 Initial purification of construct C9. Size-exclusion fractions from the test purification analysed using SDS-PAGE. The complex did not get sufficiently separated from higher molecular weight contaminants. Red, green and white arrows indicate POT1, TPP1 and TIN2S respectively.

The purity of the complex was improved by the following measures: Harvesting the cells at 58-60 hours instead of 72 hours, infecting cells which have not exceeded 10-15 passages from the original stock and including a high concentration imidazole wash step.

10X His-tagged POT1 bound strongly to the column. Therefore, during gradient NiNTA elution, the complex started eluting at 100 mM Imidazole and extended up to 750 mM Imidazole wherein a single POT1 band was observed. An imidazole wash step of 75 mM concentration was added to remove any non-specific proteins bound to the NiNTA column. Contaminants that were removed during the wash (Figure 5.4 B, lane 3,) and

the small peak preceding the main protein peak during NiNTA purification (Figure 5.4 A) are shown.

Subsequently, the complex was further purified on a size-exclusion chromatography Superose 6 10/300 column (Figure 5.5 A and B). The central peak fraction was utilized for EM analysis and the remaining fractions were used for biochemical studies. Peptide mass fingerprinting (Appendix 1) and subsequent western blot analysis with antibodies specific to each protein of this complex performed by Dr Vasuki confirmed that all three components are present in the purified sample.

Figure 5.4: NiNTA purification of TPP1- POT1 -TIN2S. NiNTA purification strategy with 1 ml HisTrap gave the best results along with 75 mM imidazole wash (labelled as wash). A few fractions contained complex with high purity (A) HisTrap profile with fractions (B) corresponding fractions analysed using SDS-PAGE are shown. Red, green and white arrows indicate POT1, TPP1 and TIN2S respectively.

Figure 5.5: Size-exclusion of TPP1-POT1-TIN2S. Selectively pooled NiNTA fractions were further purified by size-exclusion. The purity of the complex was further improved from the previous purification trials **(A)** Size-exclusion profile with fractions **(B)** corresponding fractions analysed with SDS-PAGE. Red, green and white arrows indicate POT1, TPP1 and TIN2S respectively.

5.2.4 Purification of TPP1- POT1 -TIN2L (Construct 7)

As mentioned earlier, C9 and C7 had distinct bands for POT1, TPP1 and TIN2 L/S after test purification and hence were chosen. However, during large-scale purification, it was observed that C7 had the tendency to aggregate and therefore was not used for further analysis (Figure 5.6).

Figure 5.6: Purification of TPP1- POT1 -TIN2L. Size-exclusion fractions from the test purification analysed using SDS-PAGE. The complex has higher molecular weight contaminants and has a tendency to aggregate. **(A)** Size-exclusion profile indicating the aggregation peak in blue with fractions. **(B)** SDS-PAGE analysis of corresponding fractions. Red, green and blue arrows indicate POT1, TPP1 and TIN2L respectively.

5.3 Biochemical studies of POT1-TPP1-TIN2S

5.3.1 DNA binding assay

DNA binding activity of POT1-TPP1-TIN2S was studied with two objectives: to verify the DNA binding activity of purified complex and for cryo-EM grids preparation. A minimal POT1 binding telomeric DNA probe with 10 single-stranded bases called hT10, reported previously with a binding affinity of 9.2 nM was used for this study [103]. The probe was designed with an N-terminal Cy5 label to help visualize the reaction as ethidium bromide or SyBr stains do not effectively stain single-stranded DNA.

Figure 5.7: DNA binding activity of POT1-TPP1-TIN2S. Electrophoretic mobility shift assay (EMSA) for POT1-TPP1-TIN2S binding to hT10 single-stranded telomeric DNA analysed on a 1% agarose gel. DNA concentration of 0.5 μ M was used in binding reactions with increasing concentrations of the complex from 1 to 5 times molar excess. Complete binding of the complex to DNA was observed at 2 times molar excess of protein to DNA.

EMSA was performed with 0.5 μ M hT10 and step-wise increasing molar excess of the complex (Figure 5.7). The complex starts to bind to hT10 DNA at 1:1 molar ratio and completely shifts at 2:1 (Figure 5.7).

5.3.2 Crosslinking study

POT1-TPP1-TIN2S complex requires crosslinking to keep the complex intact for negative staining, like other shelterin proteins. Therefore, a titration with glutaraldehyde (GA), was performed to determine minimal concentration required for crosslinking the complex and to study the effect of GA on the DNA binding activity. The complex bound to hT10 DNA was crosslinked with GA amounts ranging from 0.0025% to 0.1%(v/v) (Figure 5.8).

Figure 5.8: Crosslinking study of POT1-TPP1-TIN2S bound to hT10 DNA. Increasing amounts of glutaraldehyde were added to 0.5 μ M of POT1-TPP1-TIN2S and DNA complex to determine the optimal concentration of fixing agent for EM. Glutaraldehyde concentrations used are labelled at the top of the gel and the DNA concentration. Analysis of corresponding reactions using **(A)** 8% SDS-PAGE and **(B)** 1% agarose gel are shown. Uncrosslinked and free DNA probes were also loaded as controls on the gel as labelled.

The crosslinking reaction was incubated with GA for 30 minutes and subsequently quenched with 40 mM Tris-HCl pH 8. The reactions were then examined simultaneously using SDS-PAGE and an agarose gel (Figure 5.8). The uncrosslinked sample and the probe without any complex bound were used as controls for the study. The amount of protein and DNA used in all the reactions were kept constant so that the lanes can be compared with each other. At a GA concentration of 0.1% for a duration of 30 minutes, it was observed that the protein gets completely crosslinked and can still retain its DNA binding activity. Although a very faint unbound DNA band was observed, most of the DNA was still tightly bound to the protein complex (Figure 5.8 B). Glutaraldehyde concentration of 0.1% (v/v) with 30-minute incubation was then used during sample preparation for negative staining.

5.4 POT1-TPP1-TIN2S stoichiometry analysis

Stoichiometry of POT1-TPP1-TIN2S was determined using a combination of SEC-MALS and SYPRO Ruby staining, similar to the previous chapter (Section 4.6). Differential refraction (dRI) showed a single peak followed by a sudden drop in the signal (Figure 5.9 A). This strong absorption of signal at 568 nm wavelength could be due to the presence of a Zinc atom coordinating with POT1 [104]. Light scattering peak was found in the same position as the dRI peak (Figure 5.9 B). An additional aggregate peak was also observed at the void volume of the column. SEC-MALS determined the total molecular weight of the complex as 157.1 ($\pm 1.7\%$) kDa, which indicates a stoichiometry of 1: 1: 1 of POT1, TPP1 and TIN2S (Figure 5.9 C). To further confirm the stoichiometry, the complex was stained with SYPRO Ruby and Coomassie R-250 and the relative intensities of the individual proteins were compared.

Figure 5.9: Stoichiometry of POT1-TPP1-TIN2S. (A) SEC-MALS to determine the absolute molecular mass of the complex. Differential refractive index signal & calculated molecular weight shown in blue and red respectively (B) Comparison of the measured and expected molecular weight (C) Overlay of differential refraction and light scattering, showing signal at the same position for the complex. Differential refractive index signal & light scattering shown in blue and red respectively (D) Purified POT1-TPP1-TIN2S complex stained with SYPRO Ruby. Relative stoichiometry analysis of the complex using (E) SYPRO Ruby and (F) Coomassie blue R-250 staining. Intensity of the bands were normalized with TIN2 signal and compared. Error bars represent the range of the data, $n = 4$

Using Coomassie staining, a stoichiometry of only 0.6 TPP1 per TIN2 was observed, likely due to unequal staining of the protein. SYPRO ruby staining showed an almost equal stoichiometry of the subunits, in agreement with the SEC-MALS. Therefore, the staining and light scattering analysis of the POT1-TPP1-TIN2S complex indicate that one molecule of TIN2S binds to every POT1, TPP1 heterodimer in a 1:1:1 stoichiometry.

5.5 Negative staining analysis of POT1-TPP1-TIN2S

5.5.1 Data collection and 2D classification

Negative staining of POT1-TPP1-TIN2S with 2% Uranyl acetate was performed. In order to obtain a more homogeneous sample for EM, a single fraction corresponding to the peak fraction from size-exclusion was crosslinked with 0.1% GA and subsequently used for grid preparation. It was observed that in general subjecting the sample to high-speed centrifugation just before grid preparation significantly reduced aggregates in the EM images. A concentration of 0.08 mg/ml was sufficient for obtaining a good distribution of particles on the grid. The dimensions of the particles were approximately 15 nm x 10 nm. Data collection and processing details are summarized in the following table (Table 5.2)

Table 5.2: POT1-TPP1-TIN2S negative stain-EM data collection and processing

Molecular weight (kDa)	167	Total Dose (e/Å ²)	25
Sample concentration	80 µg/ml	Defocus range (µm)	-1.25 to -2
Microscope	Technai T12	Total images recorded	100
Detector	Eagle, CCD	Total Particles picked	21907
Pixel size(Å)	2.1	Box size(pixels)	132
Grid type	Homemade, 10 nm C	Resolution (Å)	23

The micrographs had uniform staining, providing good contrast between the background and the particles of interest. Each micrograph had around 200-300 well-distributed particles and a total of 100 images were collected. Data was predominantly processed on RELION 1.4 [221], with the exception of EMAN2 [218] particle picker. CTF was estimated using CTFFIND3 [222]. The output files provided by this program displayed the calculated CTF and experimental CTF from the recorded micrograph side by side (Figure 5.10). At this stage, the micrographs with astigmatism or drift were removed. A total of 21907 particles were selected for subsequent processing.

Figure 5.10: Estimation of CTF by CTFFIND3. Example of micrograph selected for processing with estimated and the experimental CTF value side by side, showing the overlap of zeros with no astigmatism or drift.

Reference-free 2D class averages were calculated with the total number of classes varied between 100, 150 and 200. Several distinct features were observed in the 2D classes. Most of the classes had a tri-lobbed/triangular view, with some of the classes having a more elongated view (Figure 5.11 C-G). These views recurred in different runs of 2D classification and had higher particle distribution. The particles from the best 2D classes were then selected for subsequent 3D classification.

Figure 5.11: 2D classification of POT1-TPP1-TIN2S. (A) Field view image of POT1-TPP1-TIN2S with uniform particle distribution. Front and side views circled in black and white respectively. Scale bar is 50 nm. (B) Representative classes with highest particle distribution. Classes with similar views are shown in (C), (D), (E), (F) and (G). Box size is 28 nm.

5.5.2 3D Classification and Refinement

Particles sorted by 2D classification were divided into three 3D classes. As there are no previous homologous structures available that could be used as an initial model, three different low-pass filtered starting models were used in order to reduce model bias. (1) The POT1 OB crystal structure [103] (2) Class 3 from the previous run (Rendered in blue, Panel 1, Figure 5.12), (3) NCP structure low-pass filtered to

resemble a disk [258]. Different orientations of the final electron densities from each 3D classification are shown (Figure 5.12).

Figure 5.12: 3D classification of the POT1-TPP1-TIN2S complex using RELION. Run 1,2 and 3 are shown in the image. Different starting models were used for each run as follows **(1)** POT1 OB low-pass filtered to 60 Å **(2)** Class 3 from Run1 (rendered in blue, panel 1) low-pass filtered to 60 Å **(3)** NCP crystal structure low-pass filtered to 80 Å. Scale bars at the top of each panel are 15 nm each.

As there were no homologous or related starting models for the 3D classification and refinement, NCP low pass filtered to appear like a disc with no features was used as a control in 3D reconstructions to avoid model bias. Dimensions and molecular weight of NCP were comparable to POT1-TPP1-TIN2S complex. 3D classes generated with NCP and other starting models had similar overall triangular/tri-lobbed structure (Figure 5.12, panel 2 and 3).

Upon visual inspection, some classes appeared better refined than others. The particle distribution among the classes in Run 2 and 3 was uniform, with approximately 6500 particles in each class. Two classes from run 2 was selected for further refinement (Figure 5.12, rendered in green and blue respectively, panel 2).

The final refinement contained a total of 16,300 particles and different starting models were used such as the previous 3D classes and low-pass filtered NCP model, which produced similar looking densities. The model had an overall tri-lobbed structure. Global resolution of 23 Å was estimated from the 0.143 FSC criterion. Different views of the final 3D refinement with dimensions and the corresponding 2D classes are shown (Figure 5.13 C).

Figure 5.13: 3D structure of POT1-TPP1-TIN2S by negative stain EM (A) Three views of the refined EM map after ML3D classification **(B)** Fourier Shell Correlation (FSC) plot for the EM density map **(C)** Three views of the EM density and corresponding 2D class averages.

5.5.3 Docking of EM density map using model derived from SAXS

The structure of full-length POT1 with N-terminus of TPP1 has been solved previously using small angle x-ray scattering (SAXS) [104]. The low-resolution SAXS envelope

shows an overall V-shaped structure. Known structural domains in the complex, namely the N-terminal DNA binding OB domains of POT1 [103], the C-terminal OB domains of POT1 interspersed by the Holliday junction resolvase like domain bound to small peptide of TPP1 (POT1 binding motif) [104, 107] and the N-terminal OB domain of TPP1 [106] have been modelled in the envelope. Docking of these domain structures in the SAXS model has been validated by a control SAXS model lacking the TPP1 OB domain, showing loss of density in the expected region [104].

Both models were V-shaped, but the negative stain EM structure had larger dimensions and was of 23 Å resolution (Figure 5.14 B). The additional electron density in the negative stain EM model would, therefore, correspond to the C-terminal TPP1 region and TIN2S which were not included in the SAXS model (Figure 5.14 B, side view and top view). The only exception is the maximum length of the complex, which is 207 Å in the SAXS model but only 152 Å in the negative staining model. Possible reasons for this reduction in length could be a rearrangement of the TPP1 OB-fold due to the presence of TIN2S or compaction of the complex due to GA crosslinking. The dimensions of the POT1 OB1 and OB2 density region in the EM map and SAXS envelope were similar (rendered in magenta and deep blue, Figure 5.14 B).

At the final resolution of 23 Å, the negative stain EM structure cannot be unequivocally docked. Additionally, structure of the protein TIN2S in both negative staining structures discussed in this study, TRF2-RAP1-TIN2S and POT1-TPP1-TIN2S is not very clear owing to low-resolution of these models and unknown domain structure of TIN2. Nevertheless, the similarities between the SAXS envelope and the EM structure help deduct the position for known domains in POT1 and TPP1 within the negative staining structure.

Figure 5.14: Comparison between POT1-TPP1-TIN2S and POT1-TPP1. Both structures show a similar V-shaped architecture. **(A)** Side by side comparison of POT1-TPP1-TIN2S negative staining structure and the POT1-TPP1N SAXS envelope. **(B)** Comparison of different views and dimensions of the two models and preliminary docking of POT1-TPP1-TIN2S based on the SAXS docking shown by Chen Cong et al, Nature 2017 [104]. SAXS model of POT1 TPP1N image adapted with permission from Chen Cong et al, Nature 2017.

5.6 Summary

Constructs of different lengths of POT1, TPP1, and TIN2 were expressed by the eukaryotic MultiBac system in insect cells, of which POT1-TPP1-TIN2S complex was chosen for further analysis. The expression and purification of the POT1-TPP1-TIN2S subcomplex of shelterin was successfully performed using MultiBac system. The

complex was found to be stable and the DNA binding activity of the complex was verified.

Crosslinker titration experiment indicates 0.1% glutaraldehyde is required to carry out optimal crosslinking of the complex for negative staining. Although the DNA binding activity was affected to some extent upon glutaraldehyde crosslinking, the complex could still bind to DNA. Stoichiometry of 1:1:1 in the POT1-TPP1-TIN2S complex was shown by SEC-MALS and SYPRO Ruby staining analysis.

Negative stain EM structure of the complex was refined to 23 Å with an overall triangular/V shape and dimensions of 15 nm x 10.4 nm x 7.4 nm. Positions of the known domains in the EM map were deducted based on the recent low-resolution SAXS model of the POT1-TPP1N structure.

Analysis of POT1-TPP1-TIN2S indicates that it is a suitable candidate for future high-resolution structure determination. POT1-TPP1-TIN2S has been imaged using cryo-EM as discussed in the following chapter (Sections 6.4 and 6.5). Structural analysis of the complex using cryo-EM in order to unequivocally assign the domain positions and determine TIN2 structure is currently ongoing.

Chapter 6: Cryo-electron microscopy and Volta phase plate imaging

6.1 Overview

Phase contrast in brightfield cryo-EM is produced by interference between the scattered and unscattered electrons. However, biological samples are weak-phase objects and therefore intrinsically produce low contrast images [210]. This results in practical problems such as identifying and aligning small particles (below 200 kDa). In general, single particle analysis of large molecular weight complexes is more straightforward due to the increased signal to noise ratios.

Volta phase plate (VPP) is recently developed device to increase phase contrast in cryo-EM. The VPP is a thin layer (10 nm) of amorphous carbon kept continuously heated in the back-focal plane of the microscope. Irradiation of the VPP results in a beam-induced negative potential which can produce a phase-shift between the scattered and unscattered rays [243].

The mechanism by which the potential is created is still unclear. Development of VPP has facilitated near atomic structure determination of many proteins [250, 280, 281] including GPCR [282] and 64 kDa Haemoglobin [283]. The following sections detail the use of applying phase plate imaging to visualize DNA and the nucleosome core, which constitutes a part of a recent publication titled '3.9 Å structure of the nucleosome core particle determined by phase-plate cryo-EM' [249]. Cryo-EM imaging of POT1-TPP1-TIN2S using phase plate is presented and the challenges with two other shelterin subcomplex samples POT1-TPP1-TIN2S with DNA and TRF2-RAP1-TIN2S are also discussed. Nucleosome core particle (NCP) and 601 Widom DNA samples

were prepared by post-doctoral fellow Dr Vinod K. Vogirala. VPP and Technai Arctica TEM were operated with the help and guidance of Dr. Sara Sandin.

6.2 Phase plate conditioning and imaging

The Volta phase plate (VPP) is activated by exposing it to the electron beam to generate sufficient phase contrast, in a process called conditioning. An unexposed area of VPP is pre-irradiated with the electron beam for a fixed amount of time, causing a negative potential to develop on-the-fly [243]. To achieve ideal contrast using the Volta phase plate, a phase shift of 90° must be attained through conditioning [238], which can be monitored by inspection of the power spectra before and after conditioning.

In order to determine the activation or conditioning time, the microscope was set at the magnification required for phase plate image acquisition. Defocus was applied such that thorn rings could be clearly visualized in the power spectra. A new position on the VPP was exposed to the beam for varying amounts of time. Thorn rings and contrast of the micrographs were compared before and after activation (Figure 6.1).

Figure 6.1: Optimal conditioning of VPP. Volta potential is generated on-the-fly when VPP is irradiated with the electron beam. Ideal conditioning results in 90° phase shift as observed in the zeros of CTF rings before (left) and after (right) VPP activation.

It was observed that for most phase plate positions used in this study, an activation time of 7 seconds equivalent to a combined dose of $210 \text{ e}^-/\text{\AA}$ was sufficient to generate the required contrast and phase shift. Exposing for longer than optimal duration can cause 'over-conditioning' with undesirable 'halo' artefacts around the particles. After the ideal conditioning time was determined, new positions on the phase plate were activated for the same duration. Imaging of sample with activated phase plate was acquired in-focus conditions.

6.3 Imaging DNA and NCP with Volta phase plate

In order to test phase plate imaging of DNA, two different samples were utilized, a reconstituted Holliday junction and 601 Widom DNA [252]. Use of phase plate shows a marked increase in contrast for both the samples. 601 DNA was not easy to identify under bright field mode with applied defocus but was clearly visible with VPP (Figure 6.2).

Figure 6.2: Imaging Widom 601 DNA sequence with VPP. (A) Phase plate image of 145 bp linear DNA **(B)** Bright field image of the same area acquired with a defocus of $-2.5 \mu\text{m}$. Representative bent and straight DNA strands are indicated with white and black arrows respectively. Scale bars in each image are 50 nm

Figure 6.3: Imaging Holliday junctions with VPP. (A) Phase plate cryo-EM image of a Holliday junction recorded in focus **(B)** bright field EM image at -3 μm defocus. Arrows indicate the same DNA molecule in both images for comparison. Scale bars are 25 nm **(C)** Schematic representation of Holliday junction sample. Each arm contains 20 bp of DNA.

In case of the Holliday junction sample, particles could easily be detected as compared to the conventional bright-field image with an applied defocus of $-3\ \mu\text{m}$ (Figure 6.3A). Further, the arms in Holliday junction were visible in the phase plate image but not under bright field imaging. (Figure 6.3 A and B).

In order to test protein-DNA complexes and particularly a chromatin complex using VPP, NCP reconstituted with 601 Widom sequence was used. While there was an overall increase in contrast, detection of certain orientations of NCP such as the 'disk view' [284] (Figure 6.4) were particularly improved with the use of VPP.

Figure 6.4: Imaging Nucleosome core particle (NCP) with VPP. (A) Phase plate image of NCP showing representative disc views (in white circle) **(B)** Bright field image of the same area imaged at $-3\ \mu\text{m}$ defocus. Scale bars equivalent to 50 nm **(C)** Crystal structure of NCP indicating the disk and the side view of NCP.

Disk view of NCP is difficult to identify in cryo-EM owing to the low-contrast (Figure 6.4 B). However, disk view is important during data processing as it can influence alignment during 3D structure determination. Further, all views of the molecule need to be included to obtain isotropic resolution.

In conclusion, Volta phase plate is a versatile tool that can be used to augment imaging of DNA and protein-DNA complexes. Its application would be most beneficial for samples in the low molecular weight range that inherently generate low contrast under cryo-EM conditions. Imaging of 601L DNA and reconstituted NCP, along with phase-plate conditioning are summarized (Figure 6.5).

Figure 6.5: Phase plate activation and imaging. (A) Nucleosome core particle (NCP) imaged with VPP. (B) Bright field image recorded at $-3\ \mu\text{m}$ defocus. The red and white circles represent the ‘side view’ and ‘disk’ view of NCP respectively (C) DNA alone imaged with VPP. (D) Bright field image of the same region at $-2.5\ \mu\text{m}$ defocus. Arrows indicate the same DNA strand for comparison (E) Power spectra before and after VPP activation showing 90° phase shift. (F) Power spectra in focus. Scale bars are 50nm. Image reprinted with permissions from NAR [249].

6.4 POT1-TPP1-TIN2S screening

The POT1-TPP1-TIN2S complex was initially screened under bright field conditions, but the particles were difficult to identify (Figure 6.6). The complex is only 167 kDa in

size with no known symmetry, making it a challenging sample for bright field imaging. Thus, the sample was imaged in-focus conditions with VPP. Inspection of phase plate images of POT1-TPP1-TIN2S showed a clear distribution of particles which were not detectable in bright field mode. The particles size was in the range of 11-13 nm. The particles were of optimal concentration, well dispersed in ice and with no apparent aggregates. Therefore, POT1-TPP1-TIN2S was determined as a suitable sample for cryo-EM analysis. Further optimization and complete dataset collection utilizing VPP are currently ongoing.

Figure 6.6: Imaging POT1-TPP1-TIN2S using VPP. Comparison of bright-field and VPP images of POT1-TPP1-TIN2S indicates the particles are visible only with VPP **(A)** Imaging of POT1-TPP1-TIN2S under brightfield condition with -2.5 μm defocus **(B)** Phase plate image of POT1-TPP1-TIN2S at focus. Scale bar is 50 nm.

6.5 Challenges faced in cryo-EM imaging of shelterin subcomplexes

Two shelterin subcomplexes were imaged using cryo-EM, namely (1) Reconstituted POT1-TPP1-TIN2S with hT10 DNA complex (assembled as described earlier in section 5.3.1) (2) Purified TRF2-RAP1-TIN2S complex. Inspection of cryo-EM images revealed that both samples had the tendency to aggregate (Figure 6.7).

Cryo-EM grids of POT1-TPP1-TIN2S with DNA complex was made with the same concentration and purification batch as the protein alone grids imaged in the previous section (6.4). As the protein complex by itself showed no indication of aggregates unlike the protein-DNA complex, it was inferred that protein-DNA interaction somehow leads to aggregation. Protein-DNA complex was screened with increased salt concentration from 50, 100, 150 to 300 mM in an attempt to reduce intermolecular interactions and subsequently the aggregates. However, no clear improvement has been observed with the conditions screened so far.

Figure 6.7: Aggregation of shelterin subcomplexes in Cryo-EM. Shelterin subcomplexes had the tendency to aggregate in Cryo-EM. Representative cryo-EM images of **(A)** POT1-TPP1-TIN2S with hT10 DNA **(B)** TRF2-RAP1-TIN2S. Scale bars are 50 nm.

TRF2-RAP1-TIN2S sample aggregation was unexpected, as the sample did not appear to have aggregates on size-exclusion column or during negative staining. Several salt conditions were screened - 14, 50, 100, 150 and 300 mM. Additionally,

sample preparation by crosslinking combined with gel-filtration purification (as described in methods section 2.2.7.1.4) did not reduce aggregation.

Cryo-EM imaging of both complexes described in this section has not been straightforward so far. Extensive screening is to be employed in the future to determine optimal buffer conditions to reduce aggregation. Screening conditions would include additives such as detergents.

6.6 Summary

Volta phase plate (VPP) is a recent development in Cryo-EM field that can dramatically increase the contrast of sample in otherwise intrinsically low contrast biological samples. In this study, VPP was used to image various molecules such as long DNA strands, Holliday junction DNA, NCP and the telomerase recruitment complex POT1-TPP1-TIN2S. In each case, it was observed that the signal to noise ratio was drastically improved, which made it easier to identify the particles. Particularly, the low contrast 'disk' view of NCP stood out when imaged with VPP compared to the defocus image. POT1-TPP1-TIN2S sample which was virtually undifferentiable from the background in the conventional Cryo-EM, was clearly identified with VPP. Further data collection and processing is required to gain more information about the structure of this complex. Challenges faced in Cryo-EM imaging of shelterin subcomplex samples TRF2-RAP1-TIN2S and POT1-TPP1-TIN2S with DNA due to aggregation were also described.

Chapter 7: Discussion and future directions

7.1 Negative staining of TRF1

An assembly of two TRF1 dimers bound to 4 telomeric repeat DNA was studied with negative stain electron microscopy (EM) to understand the architecture of full-length TRF1. In order to obtain a completely saturated complex with two dimers of TRF1 bound to DNA, optimal molar ratios were determined using electrophoretic mobility shift assay (EMSA). Crosslinking with glutaraldehyde (GA) was required to preserve TRF1 molecules during negative staining. However, crosslinking had a detrimental effect on the DNA binding activity of the protein.

2D class averaging of TRF1 DNA complex revealed most classes were of the TRF1 single dimer (around 8 nm) and only a few classes consistent with the dimensions of two dimers (15 nm) with bilobal structure were obtained. No distinct density for DNA or the Myb domain were visible in the 2D classes. Therefore, it was determined that TRF1 + four telomeric repeat DNA complex was not suitable for negative staining-EM. Using the minimal binding sequence of 2.5 telomeric repeats [98] for TRF1 dimer could improve the homogeneity of the sample compared to the 4 repeats. However, it is not a complete solution as crosslinking the sample would still result in a mixed population of DNA bound and apo-TRF1 particles.

The recent negative stain EM structure of mouse TRF1 by Boskovic *et al* describes the low-resolution structure of mTRF1 with and without DNA at 23 Å and 25 Å respectively [257]. Human and mouse TRF1 share 65% sequence similarity. The study does not mention the use of crosslinker to keep the complex intact for negative staining and the structure of mTRF1 bound to DNA does not show a clear nucleic acid density. Adjusting an equivalent threshold on the apo and DNA bound mTRF1 structures

makes them virtually the same and also comparable to the TRFH domain of TRF1 (Figure 7.1).

TRF1 domain architecture comprises of N-terminal dimerization and a small 50 amino acid C-terminal Myb domain interspersed by around 100 amino acids of disordered region. With flexible Myb domains on each dimer, TRF1 can bind to DNA with a high level of flexibility. A SELEX experiment has shown that TRF1 can recognize and bind to DNA half sites up to 40 base pairs apart [99].

Figure 7.1 Comparison between the negative staining EM structures of TRF1 [257]. Recent full length TRF1 negative staining structure published with the apo and DNA bound structures does not show clear DNA density upon comparison. Only the density corresponding to TRFH domain was clearly discernible in both models **(A)** Surface rendering of TRFH domain for comparison [95] **(B)** EM density of TRF1 DNA bound structure **(C)** EM density of apo-TRF1. Images rendered with the crystal structure of TRFH PDB code: 1H6O and TRF1 EM densities EMD codes 4105 and 4106. Vertical scale bar is 4 nm.

With respect to the TRF1 structure presented by Boskovic *et al* [257], the highly dynamic interaction between TRF1-DNA and the flexibility arising from the unstructured regions could have resulted in structural heterogeneity. Crosslinking may help to capture certain stable transitional states in the complex. However, the sample used was native and uncrosslinked material. This suggests that the nucleic acid and Myb domain densities may have been averaged out during single particle reconstruction.

It has been observed independently in the case of other human shelterin protein/complexes that GA is necessary to keep the structure intact during negative staining. Preliminary negative stain EM of uncrosslinked human TRF2 and TRF2 RAP1 published in 2012 shows characteristic dissociated particles [102], similar to TRF1. Due to the detrimental effect of GA on the DNA binding activity and paradoxically the necessity of GA crosslinking for sample stability, determination of negative staining EM structures of complexes containing TRF1 and TRF2 bound to DNA may not be feasible.

7.2 TIN2S can be copurified with TRF2 RAP1

In order to study TIN2, purification was tried with various constructs with different tags and isoforms. Most of the constructs proved difficult to express or purify owing to long stretches of disordered regions in TIN2. A codon optimized TIN2S construct with GST-tag and gave positive results. However, GST-TIN2S was sensitive to protease digestion and did not separate from its degradation products during purification.

As an alternative strategy, TIN2S was copurified with TRF2 and RAP1. Although the specific function of this complex *in vivo* remains undetermined, quantitative immunoblotting indicates TRF2, RAP1 and TIN2 are the three most abundant proteins present at telomeres in stoichiometric binding amounts [160]. Further, *in vivo* studies indicate that TIN2 binds to TRF2 and stabilizes it on the telomeres [23, 110]. TRF2-RAP1-TIN2S complex remained intact during size-exclusion chromatography and eluted at a higher apparent molecular weight similar to TRF1.

7.3 TRF2-RAP1-TIN2S can form a complex with TRF1

TIN2 can simultaneously bind to TRF1 and TRF2 through distinct regions of interaction [23, 97, 110]. Recently, the intact six-protein complex was isolated from co-transfected

human embryonic kidney cell line using tandem affinity purification [48]. The ability of TIN2S to concomitantly bind to TRF2 and TRF1 *in vitro* was tested using analytical size-exclusion. The purified TRF2-RAP1-TIN2S complex was incubated with TRF1 and its elution volume on an analytical size-exclusion column was determined and compared with controls.

A higher molecular weight complex was formed when TRF1 was added to TRF2-RAP1-TIN2S. As in the case of other shelterin proteins discussed so far, the complex with estimated molecular weight of 334 kDa, eluted at a higher apparent molecular weight in the range of 660-440 kDa. Further, the complex was also able to bind to telomeric DNA.

Previous interaction studies have predominantly utilized human or mouse cells. Although the proteins used in this study were human shelterin proteins, they were recombinantly expressed in bacteria and therefore lacked post-translational modifications. In conclusion, analytical size-exclusion study suggests that post-translational modifications may not be essential for TIN2 to simultaneously bind to TRF1 and TRF2 to form a stable complex.

7.4 POT1-TPP1-TIN2S form a stable complex

TPP1 and POT1 form heterodimers and are involved in recruiting telomerase [277, 285]. C-terminus of TPP1 is known to interact with N-terminal 1-200 region of TIN2. Deletion of TIN2 *in vivo* results in the loss of POT1/TPP1 and leads to deprotection of the telomeres [77]. Impairing interaction between TPP1 – TIN2 recapitulated classic deletion phenotype with upregulation of DNA response pathways [262]. Although TIN2 does not directly interact with telomerase, depletion of TIN2 using RNA interference results in decreased recruitment of telomerase [264]. Hence, TPP1-TIN2 link is

essential to maintain the two critical functions of the POT1/TPP1 heterodimer, namely protection of single-strand overhang and telomerase recruitment.

Although POT1-TPP1 heterodimer and the individual protein domains have been extensively studied, structural information on the intact POT1-TPP1-TIN2S complex is lacking. Six constructs with different lengths and combinations of POT1, TPP1 and TIN2 were expressed and purified from insect cells using MultiBac system. A complex containing full-length POT1, TPP1 and TIN2S was successfully isolated with adequate purity.

The purified complex was shown to bind to the minimal single-stranded DNA sequence using EMSA. Further, protein crosslinking experiment with a titration of glutaraldehyde indicated that 0.1% is sufficient to completely crosslink the complex and allow most of the DNA binding activity of the complex to remain intact. Crosslinked apo POT1-TPP1-TIN2S was analysed with negative staining. Native POT1-TPP1-TIN2S with or without DNA was imaged with Cryo-EM.

7.5 Stoichiometry of shelterin subcomplexes

Stoichiometry of the two shelterin subcomplexes in this study were analysed using a combination of SEC-MALS and SYPRO ruby staining. The stoichiometry of TRF2-RAP1-TIN2S and POT1-TPP1-TIN2S were determined as 2:2:1 and 1:1:1 respectively.

Previous *in vivo* study on the stoichiometry of shelterin reported the abundance of TIN2 was sufficient for equimolar binding to TRF2 inside the cells [160]. However, a more recent *in vitro* study on the reconstituted shelterin complex without RAP1 and TRF1, revealed a 2: 1 stoichiometry for TRF2: TIN2 [286].

Although TRF2 exists as a dimer in solution and contains potentially two binding sites for TIN2, it surprisingly binds to one TIN2 molecule. As this binding ratio is observed with or without RAP1, the resultant stoichiometry must be maintained exclusively through TRF2-TIN2 interaction. The significance of this unequal stoichiometry and the mechanism of regulation remain unclear.

The equal stoichiometry in the case of POT1-TPP1-TIN2 was expected as interaction studies indicate only one distinct binding site in each of the interacting subunits and all three intrinsically exist as monomers. Unequal stoichiometry was again reported with the larger complex of TRF2-TIN2-TPP1-POT1 (2:1:1:1) [286].

In summary, the results from this study and literature indicate that one TIN2 molecule is sufficient to connect the TRF2-RAP1 dimer of dimers to the POT1-TPP1 heterodimer. Further studies of the complex stoichiometry in the presence of TRF1 and determination of high-resolution structures could help us understand the how these telomeric proteins assemble into various complexes.

7.6 TRF2-RAP1-TIN2S negative staining structure

The structure of the TRF2-RAP1-TIN2S complex was determined to 28 Å using negative staining EM. Comparison of the TRF2-RAP1 structure previously determined in the laboratory (unpublished data) with the TRF2-RAP1-TIN2S reveals overall similarities in shape and dimensions (Figure 7.2).

A small additional density was observed in TRF2-RAP1-TIN2S structure, making it asymmetric in comparison to TRF2-RAP1 structure (Figure 7.2). This observation further reinforces the 2:2:1 stoichiometry determined by SEC-MALS as one TIN2S molecule could bind to one of the TRF2 monomers in the dimer, making the structure asymmetric.

Figure 7.2: Comparison of TRF2-RAP1 and TRF2-RAP1-TIN2S negative stain EM structures. Two views of **(A)** TRF2-RAP1-TIN2S and **(B)** TRF2-RAP1 are shown. Models are represented in the comparable scale and threshold. Position of the additional density in the three-protein complex is indicated with a black arrow. Additional density in TRF2-RAP1-TIN2S indicated by star. Scale bar is 15 nm.

7.7 POT1-TPP1-TIN2S negative staining structure

Negative staining structure of apo POT1-TPP1-TIN2S was obtained to 23 Å. The structure has a triangular shape, measuring 15 nm along the longest axis (Figure 7.3 B). Use of different starting models resulted in similar refinements. The known domains of TPP1 and POT1 have predominantly OB folds and therefore adopt the characteristic β -barrel architecture [287]. Final resolution of higher than 4Å is required for unambiguous β -sheet docking.

The position of the known domains in POT1 and TPP1 in the negative stain EM structure were assigned based on the recent small-angle X-ray scattering (SAXS) docking of POT1-TPP1 N-terminal OB [104]. The V-shaped feature described in the SAX envelope was also observed in the negative stain EM structure, however, the EM structure was 4 nm larger than the SAX model in the side view, consistent with the

inclusion of C-terminal TPP1 and TIN2S. Putative docking of known domains in POT1-TPP1-TIN2S based on the SAXS model is depicted (Figure 7.3).

Figure 7.3: Docking of the POT1-TPP1-TIN2S structure. (A) Domain architecture of POT1 and TPP1 **(B)** Comparison of different views and docking of POT1-TPP1-TIN2S with known domain structures of POT1 and TPP1 based on the SAXS docking of POT1-TPP1N by Chen Cong et al, Nature 2017 [104]

7.8 Glutaraldehyde crosslinking affects the DNA binding activity of shelterin proteins

Glutaraldehyde (GA) has widespread applications as a crosslinker, including immobilization of enzymes [288, 289], stabilization of protein crystals [290, 291] and crosslinking protein/cells for electron microscopy [292-294]. Combination of GA crosslinking with glycerol density gradient called GraFix has been shown to increase

homogeneity in samples and has been used successfully in case of certain heterogeneous samples [255].

Crosslinking was necessary during negative staining to preserve the structure of the shelterin proteins and subcomplexes discussed in this study. Crosslinking samples prior to negative staining kept the sample homogeneous and intact. However, crosslinking of protein-DNA complexes posed a challenge, as GA had a detrimental effect on DNA binding activity. DNA binding assays for TRF1 and to a lesser extent in POT1 indicated a correlation between protein crosslinking and reduction in DNA binding activity. Further, TRF2 was observed independently in the laboratory to undergo similar impairment in DNA binding upon GA treatment.

GA is known to exist in several forms in solution such as monomers, dimers and polymers, resulting in variable crosslinker lengths [295]. Mechanism of crosslinking by GA is complex since it can react in its various forms with several functional groups of amino acids present on the protein surface. The reaction of ϵ -amine groups in lysine with GA aldol group is essential for its crosslinking activity [261, 296].

DNA binding activity relies on the interaction between specific key lysine residues in the active site of these proteins with DNA (Figure 7.4). These polar lysine residues are usually found on the protein surface and are easily accessible in solution. Therefore, it is conceivable that reaction of GA with the primary amines to form a Schiff's base would strip its characteristic charge and steric properties which are essential to DNA binding.

The detrimental effect of GA is more pronounced in TRF1 as one key lysine residue in the Myb domain engaged in a water-mediated hydrogen bond with Guanine on the DNA is essential for its DNA binding activity (Figure 7.4) [98]. By contrast, POT1-DNA

interaction predominantly consists of stacking interaction between aromatic residues such as tyrosine and phenylalanine with the DNA bases [103]. In addition, 22 hydrogen bonds are found between amino acid residues on POT1 with DNA, only four of which are mediated by lysine [103]. This provides a possible explanation for the reduced effect of GA on the DNA binding activity of POT1 compared to TRF1.

Figure 7.4: Lysine residue is essential for TRF1-DNA binding. Base-specific contacts in DNA are made by lysine residues in the Myb domain. TRF1 Myb domain (grey) lysine residue K421 making contact with DNA (blue) is shown [98]. Image rendered using PDB code 1W0T for TRF1 Myb domain.

An alternative explanation for this observation could be that the protein is crosslinked in a conformation which causes steric hindrance to DNA binding. Although the DNA binding activity was reduced, it was never fully abolished even upon complete crosslinking. This suggests that the proteins were most likely not crosslinked into an inactive structure, but rather affected by a crosslinking artefact leading to a detrimental result on the DNA binding activity.

Surprisingly, the effect of GA on shelterin proteins binding to telomeric DNA has not been reported in the literature so far. Previously reported AFM studies of TRF2 have

in fact utilized GA [94, 297] and do not validate the DNA binding activity of the crosslinked protein using biochemical assays such as EMSA. Majority of these studies have used the same concentration and duration of GA crosslinking used in this study.

In conclusion, GA could interfere with shelterin DNA binding proteins TRF1 and POT1 as shown in this study. In the future, structure of native or uncrosslinked protein-DNA shelterin complexes sample should be used for structure determination to evade crosslinking artefacts

7.9 Cryo-electron microscopy of shelterin subcomplexes

Negative staining structures of the two shelterin subcomplexes: TRF2-RAP1-TIN2S and POT1-TPP1-TIN2S provided preliminary insights into domain organization in these complexes and can be used as suitable starting models for forthcoming EM studies. However, high-resolution cryo-EM structures will be required to dock the domain locations unequivocally and determine the structure of the previously uncharacterized TIN2.

POT1-TPP1-TIN2S was screened under native buffer conditions and varying concentrations. The complex was not clearly discernible under bright-field conditions even with high defocus values. This complex is challenging due to its lower molecular weight of 167 kDa and no known symmetry. As the sample did not produce enough contrast under conventional bright-field imaging, it was subsequently imaged with contrast-enhancing Volta phase plate (VPP). Distinct and uniform particles were observed with stark contrast when the same grids were screened with VPP. Collection of a complete VPP dataset on POT1-TPP1-TIN2S is currently ongoing.

However, cryo-EM imaging of the two other shelterin subcomplexes POT1-TPP1-TIN2s bound to DNA and TRF2-RAP1-TIN2S have revealed sample aggregation

tendency. The problem occurred even when the grids were made with fresh samples that were thoroughly spun before grid preparation as a precaution. Buffers with increasing salt concentrations were broadly screened, as higher salt concentration may help prevent weak intermolecular forces driving aggregation. However, none of the conditions screened were successful so far.

The driving factors for aggregation in EM are blotting and interaction of protein molecules with the air-water interface and subsequent unfolding during grid preparation [211]. Additives to prevent aggregation such as detergents, sugars or amino acid additives such as arginine need to be explored [298]. Alternatively, affinity grids could be utilized to capture the sample and prevent diffusion into the air-water interface [211, 212].

7.10 Nucleic acid and Protein-DNA complex imaging enhanced by Volta phase plate

VPP has recently emerged as a development in Cryo-EM for structure determination of progressively smaller proteins. Contrast enhancement of object without significantly increasing radiation damage has very well-suited applications for biological samples which are both weak-phase and radiation sensitive objects.

Two samples of DNA were imaged with VPP, namely 145 bp 601 Widom sequence and Holliday junction with 20 bp on each arm. VPP significantly enhanced the contrast compared to the conventional cryo-EM images and made the particles appear distinct. Kinks and bends in the DNA molecules were easily identifiable. While the concept of using TEM to study DNA has existed for several decades, with VPP the samples that were previously below the conventional size range can now be studied. In the context of telomeres, several interesting nucleic acid samples could be studied such as TERRA, TERC, T-loops and isolated telomeric chromatin.

200 kDa Nucleosome core particle (NCP) was used as a test object of known structure for VPP imaging of a small protein-DNA complex. With VPP, low contrast views of the molecule such as the ‘disk view’ were clearly visualized.

Figure 7.5: Cryo-EM structure of NCP solved to 3.9 Å utilizing VPP. (A) EM map of NCP (rendered in grey) with the crystal structure docked, rotated by 90° **(B)** Local resolution estimate using ResMap [299] showing better refinement of octamer core compared to DNA.

Structure of NCP was solved to 3.9 Å utilizing VPP imaging using only 26,060 particles with applied C2 symmetry (Figure 7.5) [249]. 2D classes and 3D reconstruction contained high resolution features such as the gyres of DNA and alpha-helices in the histone core. The side chains of amino acids were clearly defined in the final map. DNA wrapped around histone core had more flexibility than the core itself, resulting in lower resolution of DNA in comparison to histone octamer core. Furthermore, comparison to the previous NCP crystal structure showed small deviations in the DNA entry/exit points which could be due to crystal packing or DNA breathing on the octamer.

With the advent of VPP, 'resolution revolution' in Cryo-EM could be further driven towards more challenging proteins. The recent advances made with phase-plate imaging, for instance, the near-atomic structure of 64 kDa Haemoglobin [283] or the tomographic reconstruction of an entire HeLa nuclear periphery [244], indicates the size and contrast limitations previously faced in Cryo-EM could be overcome with VPP.

7.11 Conclusions and future directions

Biochemical purification, DNA binding activity, stoichiometry and low-resolution negative stain EM structures of two previously uncharacterized shelterin subcomplexes: TRF2-RAP1-TIN2S and POT1-TPP1-TIN2S have been addressed in this study. Further, TRF2-RAP1-TIN2S has been shown to interact with TRF1 with and without DNA. *In vivo* stoichiometry and distinct functions of these shelterin subcomplexes are yet to be established.

Determination of high-resolution cryo-EM structures of these subcomplexes is ongoing. However, aggregation of samples and low sample size remain as bottlenecks. Aggregation could be minimized with an extensive screening of buffer conditions and additives such as detergents.

Small sample size results in low contrast and subsequently practical problems with identification of particles. This issue could be overcome by utilizing phase plate imaging. Screening of POT1-TPP1-TIN2S sample with Volta phase plate (VPP) shows promising results. Additionally, VPP can enhance the contrast and imaging of test samples such as DNA and NCP. Future directions include collection and analysis of complete POT1-TPP1-TIN2S dataset with Volta phase plate.

References

1. Meselson, M. and F.W. Stahl, *THE REPLICATION OF DNA IN ESCHERICHIA COLI*. Proceedings of the National Academy of Sciences of the United States of America, 1958. **44**(7): p. 671-682.
2. Watson, J.D., *Origin of concatemeric T7 DNA*. Nat New Biol, 1972. **239**(94): p. 197-201.
3. Palm, W. and T. de Lange, *How shelterin protects mammalian telomeres*. Annual review of genetics, 2008. **42**: p. 301-334.
4. Muller, H.J., *The remaking of chromosomes*. Collecting net, 1938. **13**: p. 181-198.
5. McClintock, B., *The stability of broken ends of chromosomes in Zea mays*. Genetics, 1941. **26**(2): p. 234.
6. McClintock, B., *The association of non-homologous parts of chromosomes in the mid-prophase of meiosis in Zea mays*. Cell and Tissue Research, 1933. **19**(2): p. 191-237.
7. Hayflick, L., *A brief history of the mortality and immortality of cultured cells*. The Keio journal of medicine, 1998. **47**(3): p. 174-182.
8. 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, 1973. **41**(1): p. 181-190.
9. Cooke, H.J. and B. Smith. *Variability at the telomeres of the human X/Y pseudoautosomal region*. in *Cold Spring Harbor symposia on quantitative biology*. 1986. Cold Spring Harbor Laboratory Press.
10. Greider, C.W. and E.H. Blackburn, *Identification of a specific telomere terminal transferase activity in Tetrahymena extracts*. Cell, 1985. **43**(2): p. 405-413.
11. Berman, J., C.Y. Tachibana, and B.-K. Tye, *Identification of a telomere-binding activity from yeast*. Proceedings of the National Academy of Sciences, 1986. **83**(11): p. 3713-3717.
12. Buchman, A., W. Kimmerly, J. Rine, and R.D. Kornberg, *Two DNA-binding factors recognize specific sequences at silencers, upstream activating sequences, autonomously replicating sequences, and telomeres in*

- Saccharomyces cerevisiae*. Molecular and cellular biology, 1988. **8**(1): p. 210-225.
13. Conrad, M.N., J.H. Wright, A.J. Wolf, and V.A. Zakian, *RAP1 protein interacts with yeast telomeres in vivo: overproduction alters telomere structure and decreases chromosome stability*. Cell, 1990. **63**(4): p. 739-750.
 14. Gottschling, D.E. and V.A. Zakian, *Telomere proteins: specific recognition and protection of the natural termini of Oxytricha macronuclear DNA*. Cell, 1986. **47**(2): p. 195-205.
 15. Price, C.M. and T.R. Cech, *Telomeric DNA-protein interactions of Oxytricha macronuclear DNA*. Genes & development, 1987. **1**(8): p. 783-793.
 16. Gray, J.T., D.W. Celandier, C.M. Price, and T.R. Cech, *Cloning and expression of genes for the Oxytricha telomere-binding protein: specific subunit interactions in the telomeric complex*. Cell, 1991. **67**(4): p. 807-814.
 17. Zhong, Z., L. Shiue, S. Kaplan, and T. de Lange, *A mammalian factor that binds telomeric TTAGGG repeats in vitro*. Mol Cell Biol, 1992. **12**(11): p. 4834-43.
 18. Billaud, T., C. Brun, K. Ancelin, C.E. Koering, T. Laroche, and E. Gilson, *Telomeric localization of TRF2, a novel human telobox protein*. Nature genetics, 1997. **17**(2): p. 236-239.
 19. Broccoli, D., A. Smogorzewska, L. Chong, and T. de Lange, *Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2*. Nature genetics, 1997. **17**(2): p. 231-235.
 20. Li, B., S. Oestreich, and T. de Lange, *Identification of human Rap1: implications for telomere evolution*. Cell, 2000. **101**(5): p. 471-83.
 21. Xu-Dong, Z., B. Küster, M. Mann, J.H. Petrini, and T. de Lange, *Cell-cycle-regulated association of RAD50/MRE11/NBS1 with TRF2 and human telomeres*. Nature genetics, 2000. **25**(3): p. 347.
 22. Baumann, P., E. Podell, and T.R. Cech, *Human Pot1 (protection of telomeres) protein: cytolocalization, gene structure, and alternative splicing*. Mol Cell Biol, 2002. **22**(22): p. 8079-87.
 23. Kim, S.H., C. Beausejour, A.R. Davalos, P. Kaminker, S.J. Heo, and J. Campisi, *TIN2 mediates functions of TRF2 at human telomeres*. J Biol Chem, 2004. **279**(42): p. 43799-804.

24. Houghtaling, B.R., L. Cuttonaro, W. Chang, and S. Smith, *A dynamic molecular link between the telomere length regulator TRF1 and the chromosome end protector TRF2*. Current biology, 2004. **14**(18): p. 1621-1631.
25. Liu, D., A. Safari, M.S. O'Connor, D.W. Chan, A. Laegeler, J. Qin, and Z. Songyang, *PTOP interacts with POT1 and regulates its localization to telomeres*. Nat Cell Biol, 2004. **6**(7): p. 673-80.
26. Ye, J.Z.S., D. Hockemeyer, A.N. Krutchinsky, D. Loayza, S.M. Hooper, B.T. Chait, and T. de Lange, *POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex*. Genes Dev, 2004. **18**(14): p. 1649-54.
27. Henderson, E.R. and E.H. Blackburn, *An overhanging 3'terminus is a conserved feature of telomeres*. Molecular and Cellular Biology, 1989. **9**(1): p. 345-348.
28. Klobutcher, L.A., M.T. Swanton, P. Donini, and D.M. Prescott, *All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3'terminus*. Proceedings of the National Academy of Sciences, 1981. **78**(5): p. 3015-3019.
29. Moyzis, R.K., J.M. Buckingham, L.S. Cram, M. Dani, L.L. Deaven, M.D. Jones, J. Meyne, R.L. Ratliff, and J.-R. Wu, *A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes*. Proceedings of the National Academy of Sciences, 1988. **85**(18): p. 6622-6626.
30. Lejnine, S., V.L. Makarov, and J.P. Langmore, *Conserved nucleoprotein structure at the ends of vertebrate and invertebrate chromosomes*. Proceedings of the National Academy of Sciences, 1995. **92**(6): p. 2393-2397.
31. De Lange, T., L. Shiue, R. Myers, D. Cox, S. Naylor, A. Killery, and H. Varmus, *Structure and variability of human chromosome ends*. Molecular and cellular biology, 1990. **10**(2): p. 518-527.
32. Harley, C.B., A.B. Futcher, and C.W. Greider, *Telomeres shorten during ageing of human fibroblasts*. Nature, 1990. **345**(6274): p. 458.
33. Rhodes, D. and H.J. Lipps, *G-quadruplexes and their regulatory roles in biology*. Nucleic Acids Res, 2015. **43**(18): p. 8627-37.
34. Schaffitzel, C., J. Postberg, K. Paeschke, and H.J. Lipps, *Probing telomeric G-quadruplex DNA structures in cells with in vitro generated single-chain antibody fragments*. Methods Mol Biol, 2010. **608**: p. 159-81.

35. Biffi, G., D. Tannahill, J. McCafferty, and S. Balasubramanian, *Quantitative Visualization of DNA G-quadruplex Structures in Human Cells*. Nature chemistry, 2013. **5**(3): p. 182-186.
36. Moye, A.L., K.C. Porter, S.B. Cohen, T. Phan, K.G. Zyner, N. Sasaki, G.O. Lovrecz, J.L. Beck, and T.M. Bryan, *Telomeric G-quadruplexes are a substrate and site of localization for human telomerase*. 2015. **6**: p. 7643.
37. Tommerup, H., A. Dousmanis, and T. de Lange, *Unusual chromatin in human telomeres*. Molecular and Cellular Biology, 1994. **14**(9): p. 5777-5785.
38. Makarov, V.L., S. Lejnine, J. Bedoyan, and J.P. Langmore, *Nucleosomal organization of telomere-specific chromatin in rat*. Cell, 1993. **73**(4): p. 775-87.
39. Rossetti, L., S. Cacchione, M. Fua, and M. Savino, *Nucleosome assembly on telomeric sequences*. Biochemistry, 1998. **37**(19): p. 6727-37.
40. Rossetti, L., S. Cacchione, A. De Menna, L. Chapman, D. Rhodes, and M. Savino, *Specific interactions of the telomeric protein Rap1p with nucleosomal binding sites*. J Mol Biol, 2001. **306**(5): p. 903-13.
41. Galati, A., E. Micheli, C. Alicata, T. Ingegnere, A. Cicconi, M.C. Pusch, M.-J. Giraud-Panis, E. Gilson, and S. Cacchione, *TRF1 and TRF2 binding to telomeres is modulated by nucleosomal organization*. Nucleic Acids Research, 2015. **43**(12): p. 5824-5837.
42. Kabir, S., D. Hockemeyer, and T. de Lange, *TALEN Gene Knockouts Reveal No Requirement for the Conserved Human Shelterin Protein Rap1 in Telomere Protection and Length Regulation*. Cell Reports, 2014. **9**(4): p. 1273-1280.
43. Kibe, T., G.A. Osawa, C.E. Keegan, and T. de Lange, *Telomere Protection by TPP1 Is Mediated by POT1a and POT1b*. Molecular and Cellular Biology, 2010. **30**(4): p. 1059-1066.
44. Wu, L., A.S. Multani, H. He, W. Cosme-Blanco, Y. Deng, J.M. Deng, O. Bachilo, S. Pathak, H. Tahara, S.M. Bailey, Y. Deng, R.R. Behringer, and S. Chang, *Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres*. Cell, 2006. **126**(1): p. 49-62.
45. Hockemeyer, D., J.P. Daniels, H. Takai, and T. de Lange, *Recent expansion of the telomeric complex in rodents: Two distinct POT1 proteins protect mouse telomeres*. Cell, 2006. **126**(1): p. 63-77.

46. Martinez, P., I. Ferrara-Romeo, J.M. Flores, and M.A. Blasco, *Essential role for the TRF2 telomere protein in adult skin homeostasis*. Aging Cell, 2014. **13**(4): p. 656-68.
47. Chiang, Y.J., S.-H. Kim, L. Tessarollo, J. Campisi, and R.J. Hodes, *Telomere-Associated Protein TIN2 Is Essential for Early Embryonic Development through a Telomerase-Independent Pathway*. Molecular and Cellular Biology, 2004. **24**(15): p. 6631-6634.
48. Erdel, F., K. Kratz, S. Willcox, J.D. Griffith, E.C. Greene, and T. de Lange, *Telomere recognition and assembly mechanism of mammalian shelterin*. Cell reports, 2017. **18**(1): p. 41-53.
49. Nandakumar, J. and T.R. Cech, *Finding the end: recruitment of telomerase to the telomere*. Nature reviews. Molecular cell biology, 2013. **14**(2): p. 69.
50. Sugimoto, K., T. Okazaki, Y. Imae, and R. Okazaki, *MECHANISM OF DNA CHAIN GROWTH, III. EQUAL ANNEALING OF T4 NASCENT SHORT DNA CHAINS WITH THE SEPARATED COMPLEMENTARY STRANDS OF THE PHAGE DNA*. Proceedings of the National Academy of Sciences of the United States of America, 1969. **63**(4): p. 1343-1350.
51. Okazaki, R., T. Okazaki, K. Sakabe, K. Sugimoto, and A. Sugino, *Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains*. Proc Natl Acad Sci U S A, 1968. **59**(2): p. 598-605.
52. Sugimoto, K., T. Okazaki, and R. Okazaki, *Mechanism of DNA chain growth, II. Accumulation of newly synthesized short chains in E. coli infected with ligase-defective T4 phages*. Proc Natl Acad Sci U S A, 1968. **60**(4): p. 1356-62.
53. Okazaki, T. and R. Okazaki, *Mechanism of DNA chain growth. IV. Direction of synthesis of T4 short DNA chains as revealed by exonucleolytic degradation*. Proc Natl Acad Sci U S A, 1969. **64**(4): p. 1242-8.
54. Wellinger, R.J., K. Ethier, P. Labrecque, and V.A. Zakian, *Evidence for a new step in telomere maintenance*. Cell, 1996. **85**(3): p. 423-33.
55. Makarov, V.L., Y. Hirose, and J.P. Langmore, *Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening*. Cell, 1997. **88**(5): p. 657-66.
56. González-Suárez, E., E. Samper, A. Ramírez, J.M. Flores, J. Martín-Caballero, J.L. Jorcano, and M.A. Blasco, *Increased epidermal tumors and increased skin*

- wound healing in transgenic mice overexpressing the catalytic subunit of telomerase, *mTERT*, in basal keratinocytes. The EMBO journal, 2001. **20**(11): p. 2619-2630.
57. Gonzalez-Suarez, E., C. Geserick, J.M. Flores, and M.A. Blasco, *Antagonistic effects of telomerase on cancer and aging in K5-mTert transgenic mice*. Oncogene, 2005. **24**(13): p. 2256.
 58. Yu, G.-L., J.D. Bradley, L.D. Attardi, and E.H. Blackburn, *In vivo alteration of telomere sequences and senescence caused by mutated Tetrahymena telomerase RNAs*. Nature, 1990. **344**(6262): p. 126-132.
 59. Lingner, J., T.R. Hughes, A. Shevchenko, M. Mann, V. Lundblad, and T.R. Cech, *Reverse transcriptase motifs in the catalytic subunit of telomerase*. Science, 1997. **276**(5312): p. 561-567.
 60. Fu, D. and K. Collins, *Purification of human telomerase complexes identifies factors involved in telomerase biogenesis and telomere length regulation*. Molecular cell, 2007. **28**(5): p. 773-785.
 61. Cohen, S.B., M.E. Graham, G.O. Lovrecz, N. Bache, P.J. Robinson, and R.R. Reddel, *Protein composition of catalytically active human telomerase from immortal cells*. Science, 2007. **315**(5820): p. 1850-1853.
 62. Mitchell, J.R., E. Wood, and K. Collins, *A telomerase component is defective in the human disease dyskeratosis congenita*. Nature, 1999. **402**(6761): p. 551.
 63. Venteicher, A.S., E.B. Abreu, Z. Meng, K.E. McCann, R.M. Terns, T.D. Veenstra, M.P. Terns, and S.E. Artandi, *A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis*. Science, 2009. **323**(5914): p. 644-648.
 64. Lingner, J., T.R. Hughes, A. Shevchenko, M. Mann, V. Lundblad, and T.R. Cech, *Reverse transcriptase motifs in the catalytic subunit of telomerase*. Science, 1997. **276**(5312): p. 561-7.
 65. Schmidt, J.C., A.B. Dalby, and T.R. Cech, *Identification of human TERT elements necessary for telomerase recruitment to telomeres*. Elife, 2014. **3**.
 66. Zaug, A.J., E.R. Podell, J. Nandakumar, and T.R. Cech, *Functional interaction between telomere protein TPP1 and telomerase*. Genes Dev, 2010. **24**(6): p. 613-22.

67. Sauerwald, A., S. Sandin, G. Cristofari, S.H. Scheres, J. Lingner, and D. Rhodes, *Structure of active dimeric human telomerase*. Nature structural & molecular biology, 2013. **20**(4): p. 454-460.
68. Arai, K., K. Masutomi, S. Khurts, S. Kaneko, K. Kobayashi, and S. Murakami, *Two independent regions of human telomerase reverse transcriptase are important for its oligomerization and telomerase activity*. J Biol Chem, 2002. **277**(10): p. 8538-44.
69. Wenz, C., B. Enenkel, M. Amacker, C. Kelleher, K. Damm, and J. Lingner, *Human telomerase contains two cooperating telomerase RNA molecules*. Embo j, 2001. **20**(13): p. 3526-34.
70. Xin, H., D. Liu, M. Wan, A. Safari, H. Kim, W. Sun, M.S. O'Connor, and Z. Songyang, *TPP1 is a homologue of ciliate TEBP-[bgr] and interacts with POT1 to recruit telomerase*. Nature, 2007. **445**(7127): p. 559-562.
71. Tejera, A.M., M.S. d'Alcontres, M. Thanasoula, R.M. Marion, P. Martinez, C. Liao, J.M. Flores, M. Tarsounas, and M.A. Blasco, *TPP1 is required for TERT recruitment, telomere elongation during nuclear reprogramming, and normal skin development in mice*. Developmental cell, 2010. **18**(5): p. 775-789.
72. Nandakumar, J., C.F. Bell, I. Weidenfeld, A.J. Zaug, L.A. Leinwand, and T.R. Cech, *The TEL patch of telomere protein TPP1 mediates telomerase recruitment and processivity*. Nature, 2012. **492**(7428): p. 285-9.
73. Van Steensel, B. and T. De Lange, *Control of telomere length by the human telomeric protein TRF1*. Nature, 1997. **385**(6618): p. 740.
74. Smogorzewska, A., B. van Steensel, A. Bianchi, S. Oelmann, M.R. Schaefer, G. Schnapp, and T. de Lange, *Control of human telomere length by TRF1 and TRF2*. Mol Cell Biol, 2000. **20**(5): p. 1659-68.
75. Kelleher, C., I. Kurth, and J. Lingner, *Human protection of telomeres 1 (POT1) is a negative regulator of telomerase activity in vitro*. Molecular and cellular biology, 2005. **25**(2): p. 808-818.
76. Kim, S.-h., P. Kaminker, and J. Campisi, *TIN2, a new regulator of telomere length in human cells*. Nature genetics, 1999. **23**(4): p. 405.
77. Takai, K.K., T. Kibe, J.R. Donigian, D. Frescas, and T. de Lange, *Telomere protection by TPP1/POT1 requires tethering to TIN2*. Mol Cell, 2011. **44**(4): p. 647-59.

78. Karlseder, J., D. Broccoli, Y. Dai, S. Hardy, and T. de Lange, *p53-and ATM-dependent apoptosis induced by telomeres lacking TRF2*. Science, 1999. **283**(5406): p. 1321-1325.
79. DiTullio, R.A., Jr., T.A. Mochan, M. Venere, J. Bartkova, M. Sehested, J. Bartek, and T.D. Halazonetis, *53BP1 functions in an ATM-dependent checkpoint pathway that is constitutively activated in human cancer*. Nat Cell Biol, 2002. **4**(12): p. 998-1002.
80. Paull, T.T., E.P. Rogakou, V. Yamazaki, C.U. Kirchgessner, M. Gellert, and W.M. Bonner, *A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage*. Curr Biol, 2000. **10**(15): p. 886-95.
81. Mirzoeva, O.K. and J.H. Petrini, *DNA damage-dependent nuclear dynamics of the Mre11 complex*. Mol Cell Biol, 2001. **21**(1): p. 281-8.
82. Takai, H., A. Smogorzewska, and T. de Lange, *DNA damage foci at dysfunctional telomeres*. Current Biology, 2003. **13**(17): p. 1549-1556.
83. Smogorzewska, A., J. Karlseder, H. Holtgreve-Grez, A. Jauch, and T. de Lange, *DNA ligase IV-dependent NHEJ of deprotected mammalian telomeres in G1 and G2*. Current Biology, 2002. **12**(19): p. 1635-1644.
84. van Steensel, B., A. Smogorzewska, and T. de Lange, *TRF2 protects human telomeres from end-to-end fusions*. Cell, 1998. **92**(3): p. 401-13.
85. Okamoto, K., C. Bartocci, I. Ouzounov, J.K. Diedrich, J.R. Yates, 3rd, and E.L. Denchi, *A two-step mechanism for TRF2-mediated chromosome-end protection*. Nature, 2013. **494**(7438): p. 502-5.
86. Ribes-Zamora, A., S.M. Indiviglio, I. Mihalek, C.L. Williams, and A.A. Bertuch, *TRF2 Interaction with Ku Heterotetramerization Interface Gives Insight into c-NHEJ Prevention at Human Telomeres*. Cell Rep, 2013. **5**(1): p. 194-206.
87. Sfeir, A. and T. de Lange, *Removal of shelterin reveals the telomere end-protection problem*. Science, 2012. **336**(6081): p. 593-597.
88. Doksani, Y., J.Y. Wu, T. de Lange, and X. Zhuang, *Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent t-loop formation*. Cell, 2013. **155**(2): p. 345-356.
89. Griffith, J.D., L. Comeau, S. Rosenfield, R.M. Stansel, A. Bianchi, H. Moss, and T. de Lange, *Mammalian telomeres end in a large duplex loop*. Cell, 1999. **97**(4): p. 503-14.

90. Stansel, R.M., T. de Lange, and J.D. Griffith, *T-loop assembly in vitro involves binding of TRF2 near the 3' telomeric overhang*. The EMBO Journal, 2001. **20**(19): p. 5532-5540.
91. Nikitina, T. and C.L. Woodcock, *Closed chromatin loops at the ends of chromosomes*. The Journal of Cell Biology, 2004. **166**(2): p. 161-165.
92. Bianchi, A., S. Smith, L. Chong, P. Elias, and T. de Lange, *TRF1 is a dimer and bends telomeric DNA*. The EMBO journal, 1997. **16**(7): p. 1785-1794.
93. Janouskova, E., I. Necasova, J. Pavlouskova, M. Zimmermann, M. Hluchy, V. Marini, M. Novakova, and C. Hofr, *Human Rap1 modulates TRF2 attraction to telomeric DNA*. Nucleic Acids Res, 2015. **43**(5): p. 2691-700.
94. Benarroch-Popivker, D., S. Pisano, A. Mendez-Bermudez, L. Lototska, P. Kaur, S. Bauwens, N. Djerbi, C.M. Latrick, V. Fraissier, B. Pei, *et al.*, *TRF2-Mediated Control of Telomere DNA Topology as a Mechanism for Chromosome-End Protection*. Mol Cell, 2016. **61**(2): p. 274-86.
95. Fairall, L., L. Chapman, H. Moss, T. de Lange, and D. Rhodes, *Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2*. Mol Cell, 2001. **8**(2): p. 351-61.
96. Broccoli, D., A. Smogorzewska, L. Chong, and T. de Lange, *Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2*. Nat Genet, 1997. **17**(2): p. 231-5.
97. Chen, Y., Y. Yang, M. van Overbeek, J.R. Donigian, P. Baci, T. de Lange, and M. Lei, *A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins*. Science, 2008. **319**(5866): p. 1092-6.
98. Court, R., L. Chapman, L. Fairall, and D. Rhodes, *How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures*. EMBO Rep, 2005. **6**(1): p. 39-45.
99. Bianchi, A., R.M. Stansel, L. Fairall, J.D. Griffith, D. Rhodes, and T. de Lange, *TRF1 binds a bipartite telomeric site with extreme spatial flexibility*. EMBO J, 1999. **18**(20): p. 5735-44.
100. Gaullier, G., S. Miron, S. Pisano, R. Buisson, Y.-V. Le Bihan, C. Tellier-Lebègue, W. Messaoud, P. Roblin, B.G. Guimarães, R. Thai, M.-J. Giraud-Panis, E. Gilson, and M.-H. Le Du, *A higher-order entity formed by the flexible assembly of RAP1 with TRF2*. Nucleic Acids Research, 2016. **44**(4): p. 1962-1976.

101. Hanaoka, S., A. Nagadoi, S. Yoshimura, S. Aimoto, B. Li, T. de Lange, and Y. Nishimura, *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, 2001. **312**(1): p. 167-175.
102. Arat, N.Ö. and J.D. Griffith, *Human Rap1 interacts directly with telomeric DNA and regulates TRF2 localization at the telomere*. Journal of Biological Chemistry, 2012. **287**(50): p. 41583-41594.
103. Lei, M., E.R. Podell, and T.R. Cech, *Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection*. Nat Struct Mol Biol, 2004. **11**(12): p. 1223-9.
104. Chen, C., P. Gu, J. Wu, X. Chen, S. Niu, H. Sun, L. Wu, N. Li, J. Peng, S. Shi, *et al.*, *Structural insights into POT1-TPP1 interaction and POT1 C-terminal mutations in human cancer*. Nature Communications, 2017. **8**: p. 14929.
105. Baumann, P. and T.R. Cech, *Pot1, the putative telomere end-binding protein in fission yeast and humans*. Science, 2001. **292**(5519): p. 1171-5.
106. Wang, F., E.R. Podell, A.J. Zaug, Y. Yang, P. Baciú, T.R. Cech, and M. Lei, *The POT1-TPP1 telomere complex is a telomerase processivity factor*. Nature, 2007. **445**(7127): p. 506-10.
107. Rice, C., P.K. Shastrula, A.V. Kossenkova, R. Hills, D.M. Baird, L.C. Showe, T. Doukov, S. Janicki, and E. Skordalakes, *Structural and functional analysis of the human POT1-TPP1 telomeric complex*. Nat Commun, 2017. **8**: p. 14928.
108. Hirai, Y., M. Tamura, J. Otani, and F. Ishikawa, *NEK6-mediated phosphorylation of human TPP1 regulates telomere length through telomerase recruitment*. Genes to Cells, 2016. **21**(8): p. 874-889.
109. Latrick, C.M. and T.R. Cech, *POT1-TPP1 enhances telomerase processivity by slowing primer dissociation and aiding translocation*. The EMBO journal, 2010. **29**(5): p. 924-933.
110. Ye, J.Z., J.R. Donigian, M. van Overbeek, D. Loayza, Y. Luo, A.N. Krutchinsky, B.T. Chait, and T. de Lange, *TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres*. J Biol Chem, 2004. **279**(45): p. 47264-71.
111. Frank, A.K., D.C. Tran, R.W. Qu, B.A. Stohr, D.J. Segal, and L. Xu, *The Shelterin TIN2 Subunit Mediates Recruitment of Telomerase to Telomeres*. PLOS Genetics, 2015. **11**(7): p. e1005410.

112. Alter, B.P., N. Giri, S.A. Savage, and P.S. Rosenberg, *Cancer in dyskeratosis congenita*. Blood, 2009. **113**(26): p. 6549-57.
113. Alter, B.P., N. Giri, S.A. Savage, J.A. Peters, J.T. Loud, L. Leathwood, A.G. Carr, M.H. Greene, and P.S. Rosenberg, *Malignancies and survival patterns in the National Cancer Institute inherited bone marrow failure syndromes cohort study*. Br J Haematol, 2010. **150**(2): p. 179-88.
114. Mason, P.J. and M. Bessler, *The genetics of dyskeratosis congenita*. Cancer Genet, 2011. **204**(12): p. 635-45.
115. Gu, B.W., M. Bessler, and P.J. Mason, *A pathogenic dyskerin mutation impairs proliferation and activates a DNA damage response independent of telomere length in mice*. Proc Natl Acad Sci U S A, 2008. **105**(29): p. 10173-8.
116. Heiss, N.S., S.W. Knight, T.J. Vulliamy, S.M. Klauck, S. Wiemann, P.J. Mason, A. Poustka, and I. Dokal, *X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions*. Nat Genet, 1998. **19**(1): p. 32-8.
117. Yamaguchi, H., R.T. Calado, H. Ly, S. Kajigaya, G.M. Baerlocher, S.J. Chanock, P.M. Lansdorp, and N.S. Young, *Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia*. N Engl J Med, 2005. **352**(14): p. 1413-24.
118. Vulliamy, T.J., A. Walne, A. Baskaradas, P.J. Mason, A. Marrone, and I. Dokal, *Mutations in the reverse transcriptase component of telomerase (TERT) in patients with bone marrow failure*. Blood Cells Mol Dis, 2005. **34**(3): p. 257-63.
119. Xin, Z.-T., A.D. Beauchamp, R.T. Calado, J.W. Bradford, J.A. Regal, A. Shenoy, Y. Liang, P.M. Lansdorp, N.S. Young, and H. Ly, *Functional characterization of natural telomerase mutations found in patients with hematologic disorders*. Blood, 2007. **109**(2): p. 524-532.
120. Vulliamy, T.J., M.J. Kirwan, R. Beswick, U. Hossain, C. Baqai, A. Ratcliffe, J. Marsh, A. Walne, and I. Dokal, *Differences in disease severity but similar telomere lengths in genetic subgroups of patients with telomerase and shelterin mutations*. PloS one, 2011. **6**(9): p. e24383.
121. Armanios, M., J.-L. Chen, Y.-P.C. Chang, R.A. Brodsky, A. Hawkins, C.A. Griffin, J.R. Eshleman, A.R. Cohen, A. Chakravarti, and A. Hamosh, *Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita*. Proceedings of the National

- Academy of Sciences of the United States of America, 2005. **102**(44): p. 15960-15964.
122. Marrone, A., A. Walne, H. Tamary, Y. Masunari, M. Kirwan, R. Beswick, T. Vulliamy, and I. Dokal, *Telomerase reverse-transcriptase homozygous mutations in autosomal recessive dyskeratosis congenita and Hoyeraal-Hreidarsson syndrome*. Blood, 2007. **110**(13): p. 4198-4205.
 123. Vulliamy, T., A. Marrone, R. Szydlo, A. Walne, P.J. Mason, and I. Dokal, *Disease anticipation is associated with progressive telomere shortening in families with dyskeratosis congenita due to mutations in TERC*. Nat Genet, 2004. **36**(5): p. 447-9.
 124. Vulliamy, T., A. Marrone, F. Goldman, A. Dearlove, M. Bessler, P.J. Mason, and I. Dokal, *The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita*. Nature, 2001. **413**(6854): p. 432-5.
 125. Savage, S.A., N. Giri, G.M. Baerlocher, N. Orr, P.M. Lansdorp, and B.P. Alter, *TINF2, a component of the shelterin telomere protection complex, is mutated in dyskeratosis congenita*. The American Journal of Human Genetics, 2008. **82**(2): p. 501-509.
 126. Sasa, G.S., A. Ribes-Zamora, N.D. Nelson, and A.A. Bertuch, *Three novel truncating TINF2 mutations causing severe dyskeratosis congenita in early childhood*. Clinical genetics, 2012. **81**(5): p. 10.1111/j.1399-0004.2011.01658.x.
 127. Yang, D., Q. He, H. Kim, W. Ma, and Z. Songyang, *TIN2 Protein Dyskeratosis Congenita Missense Mutants Are Defective in Association with Telomerase*. The Journal of Biological Chemistry, 2011. **286**(26): p. 23022-23030.
 128. Xin, Z.T. and H. Ly, *Characterization of interactions between naturally mutated forms of the TIN2 protein and its known protein partners of the shelterin complex*. Clin Genet, 2012. **81**(3): p. 301-2.
 129. Trahan, C., C. Martel, and F. Dragon, *Effects of dyskeratosis congenita mutations in dyskerin, NHP2 and NOP10 on assembly of H/ACA pre-RNPs*. Hum Mol Genet, 2010. **19**(5): p. 825-36.
 130. Cossu, F., T.J. Vulliamy, A. Marrone, M. Badiali, A. Cao, and I. Dokal, *A novel DKC1 mutation, severe combined immunodeficiency (T+ B-NK-SCID) and bone marrow transplantation in an infant with Hoyeraal-Hreidarsson syndrome*. British journal of haematology, 2002. **119**(3): p. 765-768.

131. Knight, S., N. Heiss, T. Vulliamy, S. Greschner, G. Stavrides, G. Pai, G. Lestringant, N. Varma, P. Mason, and I. Dokal, *X-linked dyskeratosis congenita is predominantly caused by missense mutations in the DKC1 gene*. The American Journal of Human Genetics, 1999. **65**(1): p. 50-58.
132. Du, H.-Y., E. Pumbo, P. Manley, J.J. Field, S.J. Bayliss, D.B. Wilson, P.J. Mason, and M. Bessler, *Complex inheritance pattern of dyskeratosis congenita in two families with 2 different mutations in the telomerase reverse transcriptase gene*. Blood, 2008. **111**(3): p. 1128-1130.
133. Sharma, A., K. Myers, Z. Ye, and J. D'orazio, *Dyskeratosis congenita caused by a novel TERT point mutation in siblings with pancytopenia and exudative retinopathy*. Pediatric blood & cancer, 2014. **61**(12): p. 2302-2304.
134. Comolli, L.R., I. Smirnov, L. Xu, E.H. Blackburn, and T.L. James, *A molecular switch underlies a human telomerase disease*. Proc Natl Acad Sci U S A, 2002. **99**(26): p. 16998-7003.
135. Theimer, C.A., L.D. Finger, L. Trantirek, and J. Feigon, *Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase RNA*. Proceedings of the National Academy of Sciences, 2003. **100**(2): p. 449-454.
136. Frescas, D. and T. de Lange, *A TIN2 dyskeratosis congenita mutation causes telomerase-independent telomere shortening in mice*. Genes & Development, 2014. **28**(2): p. 153-166.
137. Sasa, G.S., A. Ribes-Zamora, N.D. Nelson, and A.A. Bertuch, *Three novel truncating TINF2 mutations causing severe dyskeratosis congenita in early childhood*. Clinical genetics, 2012. **81**(5): p. 470-478.
138. Knight, S., N. Heiss, T. Vulliamy, C. Aalfs, C. McMahon, P. Richmond, A. Jones, R. Hennekam, A. Poustka, and P. Mason, *Unexplained aplastic anaemia, immunodeficiency, and cerebellar hypoplasia (Hoyeraal-Hreidarsson syndrome) due to mutations in the dyskeratosis congenita gene, DKC1*. British journal of haematology, 1999. **107**(2): p. 335-339.
139. Ballew, B.J., M. Yeager, K. Jacobs, N. Giri, J. Boland, L. Burdett, B.P. Alter, and S.A. Savage, *Germline mutations of regulator of telomere elongation helicase 1, RTEL1, in dyskeratosis congenita*. Human genetics, 2013. **132**(4): p. 473-480.

140. Le Guen, T., L. Jullien, F. Touzot, M. Schertzer, L. Gaillard, M. Perderiset, W. Carpentier, P. Nitschke, C. Picard, and G. Couillault, *Human RTEL1 deficiency causes Hoyeraaal–Hreidarsson syndrome with short telomeres and genome instability*. Human molecular genetics, 2013. **22**(16): p. 3239-3249.
141. Shay, J. and S. Bacchetti, *A survey of telomerase activity in human cancer*. European journal of cancer, 1997. **33**(5): p. 787-791.
142. Bryan, T.M., A. Englezou, L. Dalla-Pozza, M.A. Dunham, and R.R. Reddel, *Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines*. Nat Med, 1997. **3**(11): p. 1271-4.
143. Bryan, T.M., A. Englezou, L. Dalla-Pozza, M.A. Dunham, and R.R. Reddel, *Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines*. Nature medicine, 1997. **3**(11): p. 1271-1274.
144. Shi, J., X.R. Yang, B. Ballew, M. Rotunno, D. Calista, M.C. Fargnoli, P. Ghiorzo, B. Bressac-de Paillerets, E. Nagore, M.F. Avril, *et al.*, *Rare missense variants in POT1 predispose to familial cutaneous malignant melanoma*. 2014. **46**(5): p. 482-6.
145. Robles-Espinoza, C.D., M. Harland, A.J. Ramsay, L.G. Aoude, V. Quesada, Z. Ding, K.A. Pooley, A.L. Pritchard, J.C. Tiffen, M. Petljak, *et al.*, *POT1 loss-of-function variants predispose to familial melanoma*. Nat Genet, 2014. **46**(5): p. 478-481.
146. Bainbridge, M.N., G.N. Armstrong, M.M. Gramatges, A.A. Bertuch, S.N. Jhangiani, H. Doddapaneni, L. Lewis, J. Tombrello, S. Tsavachidis, Y. Liu, *et al.*, *Germline mutations in shelterin complex genes are associated with familial glioma*. J Natl Cancer Inst, 2015. **107**(1): p. 384.
147. Ramsay, A.J., V. Quesada, M. Foronda, L. Conde, A. Martinez-Trillos, N. Villamor, D. Rodriguez, A. Kwarciak, C. Garabaya, M. Gallardo, *et al.*, *POT1 mutations cause telomere dysfunction in chronic lymphocytic leukemia*. Nat Genet, 2013. **45**(5): p. 526-30.
148. Dikmen, Z.G., G.C. Gellert, S. Jackson, S. Gryaznov, R. Tressler, P. Dogan, W.E. Wright, and J.W. Shay, *In vivo inhibition of lung cancer by GRN163L: a novel human telomerase inhibitor*. Cancer Res, 2005. **65**(17): p. 7866-73.

149. Ouellette, M.M., W.E. Wright, and J.W. Shay, *Targeting telomerase-expressing cancer cells*. Journal of Cellular and Molecular Medicine, 2011. **15**(7): p. 1433-1442.
150. Harley, C.B., *Telomerase and cancer therapeutics*. Nat Rev Cancer, 2008. **8**(3): p. 167-179.
151. Brunsvig, P.F., S. Aamdal, M.K. Gjertsen, G. Kvalheim, C.J. Markowski-Grimsrud, I. Sve, M. Dyrhaug, S. Trachsel, M. Møller, J.A. Eriksen, and G. Gaudernack, *Telomerase peptide vaccination: a phase I/II study in patients with non-small cell lung cancer*. Cancer Immunology, Immunotherapy, 2006. **55**(12): p. 1553-1564.
152. Canudas, S., B.R. Houghtaling, M. Bhanot, G. Sasa, S.A. Savage, A.A. Bertuch, and S. Smith, *A role for heterochromatin protein 1γ at human telomeres*. Genes & development, 2011. **25**(17): p. 1807-1819.
153. Benetti, R., M. García-Cao, and M.A. Blasco, *Telomere length regulates the epigenetic status of mammalian telomeres and subtelomeres*. Nature genetics, 2007. **39**(2): p. 243.
154. García-Cao, M., R. O'Sullivan, A.H. Peters, T. Jenuwein, and M.A. Blasco, *Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases*. Nature genetics, 2004. **36**(1): p. 94.
155. Rhodes, D., L. Fairall, T. Simonsson, R. Court, and L. Chapman, *Telomere architecture*. EMBO Reports, 2002. **3**(12): p. 1139-1145.
156. Bandaria, J.N., P. Qin, V. Berk, S. Chu, and A. Yildiz, *Shelterin Protects Chromosome Ends by Compacting Telomeric Chromatin*. Cell, 2016. **164**(4): p. 735-746.
157. Timashev, L.A., H. Babcock, X. Zhuang, and T. de Lange, *The DDR at telomeres lacking intact shelterin does not require substantial chromatin decompaction*. Genes Dev, 2017. **31**(6): p. 578-589.
158. Vancevska, A., K.M. Douglass, V. Pfeiffer, S. Manley, and J. Lingner, *The telomeric DNA damage response occurs in the absence of chromatin decompaction*. Genes & Development, 2017.
159. Li, B., S. Oestreich, and T. de Lange, *Identification of human Rap1: implications for telomere evolution*. Cell, 2000. **101**(5): p. 471-483.

160. Takai, K.K., S. Hooper, S. Blackwood, R. Gandhi, and T. de Lange, *In vivo stoichiometry of shelterin components*. Journal of Biological Chemistry, 2010. **285**(2): p. 1457-1467.
161. Sfeir, A., S.T. Kosiyatrakul, D. Hockemeyer, S.L. MacRae, J. Karlseder, C.L. Schildkraut, and T. de Lange, *Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication*. cell, 2009. **138**(1): p. 90-103.
162. Celli, G.B. and T. de Lange, *DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion*. Nat Cell Biol, 2005. **7**(7): p. 712-8.
163. Kim, S.-h., A.R. Davalos, S.-J. Heo, F. Rodier, Y. Zou, C. Beausejour, P. Kaminker, S.M. Yannoni, and J. Campisi, *Telomere dysfunction and cell survival: roles for distinct TIN2-containing complexes*. The Journal of cell biology, 2008. **181**(3): p. 447-460.
164. Zhou, X.Z. and K.P. Lu, *The Pin2/TRF1-interacting protein PinX1 is a potent telomerase inhibitor*. Cell, 2001. **107**(3): p. 347-59.
165. Kaminker, P.G., S.H. Kim, R.D. Taylor, Y. Zebardjian, W.D. Funk, G.B. Morin, P. Yaswen, and J. Campisi, *TANK2, a new TRF1-associated poly(ADP-ribose) polymerase, causes rapid induction of cell death upon overexpression*. J Biol Chem, 2001. **276**(38): p. 35891-9.
166. Smith, S., I. Gariat, A. Schmitt, and T. de Lange, *Tankyrase, a poly(ADP-ribose) polymerase at human telomeres*. Science, 1998. **282**(5393): p. 1484-7.
167. van Overbeek, M. and T. de Lange, *Apollo, an Artemis-related nuclease, interacts with TRF2 and protects human telomeres in S phase*. Curr Biol, 2006. **16**(13): p. 1295-302.
168. Svendsen, J.M., A. Smogorzewska, M.E. Sowa, B.C. O'Connell, S.P. Gygi, S.J. Elledge, and J.W. Harper, *Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair*. Cell, 2009. **138**(1): p. 63-77.
169. Mattern, K.A., S.J. Swiggers, A.L. Nigg, B. Löwenberg, A.B. Houtsmuller, and J.M.J. Zijlmans, *Dynamics of protein binding to telomeres in living cells: implications for telomere structure and function*. Molecular and cellular biology, 2004. **24**(12): p. 5587-5594.
170. Kimura, H. and P.R. Cook, *Kinetics of Core Histones in Living Human Cells: Little Exchange of H3 and H4 and Some Rapid Exchange of H2b*. The Journal of Cell Biology, 2001. **153**(7): p. 1341-1354.

171. Ye, J.Z. and T. de Lange, *TIN2 is a tankyrase 1 PARP modulator in the TRF1 telomere length control complex*. Nat Genet, 2004. **36**(6): p. 618-23.
172. Chen, Y., R. Rai, Z.-R. Zhou, J. Kanoh, C. Ribeyre, Y. Yang, H. Zheng, P. Damay, F. Wang, and H. Tsujii, *A conserved motif within RAP1 has diversified roles in telomere protection and regulation in different organisms*. Nature structural & molecular biology, 2011. **18**(2): p. 213-221.
173. Shi, Y., *A glimpse of structural biology through X-ray crystallography*. Cell, 2014. **159**(5): p. 995-1014.
174. Kuhlbrandt, W., *Biochemistry. The resolution revolution*. Science, 2014. **343**(6178): p. 1443-4.
175. Liao, M., E. Cao, D. Julius, and Y. Cheng, *Structure of the TRPV1 ion channel determined by electron cryo-microscopy*. Nature, 2013. **504**(7478): p. 107-12.
176. Ludtke, S.J., M.L. Baker, D.H. Chen, J.L. Song, D.T. Chuang, and W. Chiu, *De novo backbone trace of GroEL from single particle electron cryomicroscopy*. Structure, 2008. **16**(3): p. 441-8.
177. Baker, M.L., C.F. Hryc, Q. Zhang, W. Wu, J. Jakana, C. Haase-Pettingell, P.V. Afonine, P.D. Adams, J.A. King, W. Jiang, and W. Chiu, *Validated near-atomic resolution structure of bacteriophage epsilon15 derived from cryo-EM and modeling*. Proc Natl Acad Sci U S A, 2013. **110**(30): p. 12301-6.
178. Yu, X., L. Jin, and Z.H. Zhou, *3.88 Å structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy*. Nature, 2008. **453**(7193): p. 415-9.
179. Frauenfeld, J., J. Gumbart, E.O.v.d. Sluis, S. Funes, M. Gartmann, B. Beatrix, T. Mielke, O. Berninghausen, T. Becker, K. Schulten, and R. Beckmann, *Cryo-EM structure of the ribosome–SecYE complex in the membrane environment*. Nat Struct Mol Biol, 2011. **18**(5): p. 614-621.
180. Frank, J. and R.L. Gonzalez, Jr., *Structure and dynamics of a processive Brownian motor: the translating ribosome*. Annu Rev Biochem, 2010. **79**: p. 381-412.
181. Tsai, K.-L., X. Yu, S. Gopalan, T.-C. Chao, Y. Zhang, L. Florens, M.P. Washburn, K. Murakami, R.C. Conaway, J.W. Conaway, and F.J. Asturias, *Mediator structure and rearrangements required for holoenzyme formation*. Nature, 2017. **544**(7649): p. 196-201.

182. Sengupta, J., J. Nilsson, R. Gursky, M. Kjeldgaard, P. Nissen, and J. Frank, *Visualization of the eEF2-80S ribosome transition-state complex by cryo-electron microscopy*. J Mol Biol, 2008. **382**(1): p. 179-87.
183. Merk, A., A. Bartesaghi, S. Banerjee, V. Falconieri, P. Rao, M.I. Davis, R. Pragani, M.B. Boxer, L.A. Earl, J.L. Milne, and S. Subramaniam, *Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery*. Cell, 2016. **165**(7): p. 1698-707.
184. Kostyuchenko, V.A., E.X.Y. Lim, S. Zhang, G. Fibriansah, T.-S. Ng, J.S.G. Ooi, J. Shi, and S.-M. Lok, *Structure of the thermally stable Zika virus*. Nature, 2016. **advance online publication**.
185. Banerjee, S., A. Bartesaghi, A. Merk, P. Rao, S.L. Bulfer, Y. Yan, N. Green, B. Mroczkowski, R.J. Neitz, P. Wipf, *et al.*, *2.3 Å resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition*. Science, 2016. **351**(6275): p. 871-5.
186. Baretić, D., H.K. Pollard, D.I. Fisher, C.M. Johnson, B. Santhanam, C.M. Truman, T. Kouba, A.R. Fersht, C. Phillips, and R.L. Williams, *Structures of closed and open conformations of dimeric human ATM*. Science Advances, 2017. **3**(5).
187. Bartesaghi, A., A. Merk, S. Banerjee, D. Matthies, X. Wu, J.L. Milne, and S. Subramaniam, *2.2 Å resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor*. Science, 2015. **348**(6239): p. 1147-51.
188. McMullan, G., A.T. Clark, R. Turchetta, and A.R. Faruqi, *Enhanced imaging in low dose electron microscopy using electron counting*. Ultramicroscopy, 2009. **109**(12): p. 1411-6.
189. McMullan, G., A.R. Faruqi, R. Henderson, N. Guerrini, R. Turchetta, A. Jacobs, and G. van Hooten, *Experimental observation of the improvement in MTF from backthinning a CMOS direct electron detector*. Ultramicroscopy, 2009. **109**(9-3): p. 1144-1147.
190. Brilot, A.F., J.Z. Chen, A. Cheng, J. Pan, S.C. Harrison, C.S. Potter, B. Carragher, R. Henderson, and N. Grigorieff, *Beam-induced motion of vitrified specimen on holey carbon film*. Journal of Structural Biology, 2012. **177**(3): p. 630-637.
191. Campbell, M.G., A. Cheng, A.F. Brilot, A. Moeller, D. Lyumkis, D. Veasler, J. Pan, S.C. Harrison, C.S. Potter, B. Carragher, and N. Grigorieff, *Movies of ice-*

- embedded particles enhance resolution in electron cryo-microscopy*. Structure, 2012. **20**(11): p. 1823-8.
192. Li, X., P. Mooney, S. Zheng, C.R. Booth, M.B. Braunfeld, S. Gubbens, D.A. Agard, and Y. Cheng, *Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM*. Nat Methods, 2013. **10**(6): p. 584-90.
 193. McMullan, G., S. Chen, R. Henderson, and A.R. Faruqi, *Detective quantum efficiency of electron area detectors in electron microscopy*. Ultramicroscopy, 2009. **109**(9): p. 1126-43.
 194. Meyer, R.R., A.I. Kirkland, R.E. Dunin-Borkowski, and J.L. Hutchison, *Experimental characterisation of CCD cameras for HREM at 300kV*. Ultramicroscopy, 2000. **85**(1): p. 9-13.
 195. Faruqi, A.R., D.M. Cattermole, and C. Raeburn, *Direct electron detection methods in electron microscopy*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2003. **513**(1): p. 317-321.
 196. Deptuch, G., A. Besson, P. Rehak, M. Szelezniak, J. Wall, M. Winter, and Y. Zhu, *Direct electron imaging in electron microscopy with monolithic active pixel sensors*. Ultramicroscopy, 2007. **107**(8): p. 674-84.
 197. Veessler, D., M.G. Campbell, A. Cheng, C.Y. Fu, Z. Murez, J.E. Johnson, C.S. Potter, and B. Carragher, *Maximizing the potential of electron cryomicroscopy data collected using direct detectors*. J Struct Biol, 2013. **184**(2): p. 193-202.
 198. Li, X., P. Mooney, S. Zheng, C.R. Booth, M.B. Braunfeld, S. Gubbens, D.A. Agard, and Y. Cheng, *Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM*. Nat Meth, 2013. **10**(6): p. 584-590.
 199. Bai, X.-c., I.S. Fernandez, G. McMullan, and S.H.W. Scheres, *Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles*. eLife, 2013. **2**: p. e00461.
 200. Baker, L.A. and J.L. Rubinstein, *Radiation damage in electron cryomicroscopy*. Methods Enzymol, 2010. **481**: p. 371-88.
 201. Scheres, S.H., *Beam-induced motion correction for sub-megadalton cryo-EM particles*. Elife, 2014. **3**: p. e03665.

202. De Rosier, D.J. and A. Klug, *Reconstruction of three dimensional structures from electron micrographs*. Nature, 1968. **217**(5124): p. 130-4.
203. Frank, J., *Three-dimensional electron microscopy of macromolecular assemblies: visualization of biological molecules in their native state*. 2006: Oxford University Press.
204. De Carlo, S. and J.R. Harris, *Negative staining and Cryo-negative Staining of Macromolecules and Viruses for TEM*. Micron (Oxford, England : 1993), 2011. **42**(2): p. 117-131.
205. Ohi, M., Y. Li, Y. Cheng, and T. Walz, *Negative Staining and Image Classification – Powerful Tools in Modern Electron Microscopy*. Biological Procedures Online, 2004. **6**: p. 23-34.
206. Thompson, R.F., M. Walker, C.A. Siebert, S.P. Muench, and N.A. Ranson, *An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology*. Methods (San Diego, Calif.), 2016. **100**: p. 3-15.
207. Adrian, M., J. Dubochet, J. Lepault, and A.W. McDowell, *Cryo-electron microscopy of viruses*. Nature, 1984. **308**(5954): p. 32-6.
208. Taylor, K.A. and R.M. Glaeser, *Electron diffraction of frozen, hydrated protein crystals*. Science, 1974. **186**(4168): p. 1036-7.
209. Dubochet, J., M. Adrian, J.J. Chang, J.C. Homo, J. Lepault, A.W. McDowell, and P. Schultz, *Cryo-electron microscopy of vitrified specimens*. Q Rev Biophys, 1988. **21**(2): p. 129-228.
210. Cheng, Y., N. Grigorieff, Pawel A. Penczek, and T. Walz, *A Primer to Single-Particle Cryo-Electron Microscopy*. Cell, 2015. **161**(3): p. 438-449.
211. Glaeser, R.M. and B.-G. Han, *Opinion: hazards faced by macromolecules when confined to thin aqueous films*. Biophysics Reports, 2017. **3**(1): p. 1-7.
212. Kelly, D.F., P.D. Abeyrathne, D. Dukovski, and T. Walz, *The Affinity Grid: a pre-fabricated EM grid for monolayer purification*. Journal of molecular biology, 2008. **382**(2): p. 423-433.
213. Erickson, H.P. and A. Klug, *Measurement and Compensation of Defocusing and Aberrations by Fourier Processing of Electron Micrographs*. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 1971. **261**(837): p. 105-118.
214. Wade, R.H., *The phase contrast characteristics in bright field electron microscopy*. Ultramicroscopy, 1978. **3**: p. 329-334.

215. Coene, W., G. Janssen, M. Op de Beeck, and D. Van Dyck, *Phase retrieval through focus variation for ultra-resolution in field-emission transmission electron microscopy*. Physical Review Letters, 1992. **69**(26): p. 3743-3746.
216. Shaikh, T.R., H. Gao, W.T. Baxter, F.J. Asturias, N. Boisset, A. Leith, and J. Frank, *SPIDER image processing for single-particle reconstruction of biological macromolecules from electron micrographs*. Nature protocols, 2008. **3**(12): p. 1941-1974.
217. Grigorieff, N., *FREALIGN: high-resolution refinement of single particle structures*. J Struct Biol, 2007. **157**(1): p. 117-25.
218. Tang, G., L. Peng, P.R. Baldwin, D.S. Mann, W. Jiang, I. Rees, and S.J. Ludtke, *EMAN2: an extensible image processing suite for electron microscopy*. J Struct Biol, 2007. **157**(1): p. 38-46.
219. Hohn, M., G. Tang, G. Goodyear, P.R. Baldwin, Z. Huang, P.A. Penczek, C. Yang, R.M. Glaeser, P.D. Adams, and S.J. Ludtke, *SPARX, a new environment for Cryo-EM image processing*. J Struct Biol, 2007. **157**(1): p. 47-55.
220. Sorzano, C.O., R. Marabini, J. Velazquez-Muriel, J.R. Bilbao-Castro, S.H. Scheres, J.M. Carazo, and A. Pascual-Montano, *XMIPP: a new generation of an open-source image processing package for electron microscopy*. J Struct Biol, 2004. **148**(2): p. 194-204.
221. Scheres, S.H., *RELION: implementation of a Bayesian approach to cryo-EM structure determination*. J Struct Biol, 2012. **180**(3): p. 519-30.
222. Rohou, A. and N. Grigorieff, *CTFFIND4: Fast and accurate defocus estimation from electron micrographs*. J Struct Biol, 2015. **192**(2): p. 216-21.
223. Penczek, P.A., *Image restoration in cryo-electron microscopy*. Methods Enzymol, 2010. **482**: p. 35-72.
224. Scheres, S.H.W., *Semi-automated selection of cryo-EM particles in RELION-1.3*. J Struct Biol, 2015. **189**(2): p. 114-22.
225. Scheres, S.H.W., *A Bayesian View on Cryo-EM Structure Determination*. Journal of Molecular Biology, 2012. **415**(2): p. 406-418.
226. Scheres, S.H., *Processing of Structurally Heterogeneous Cryo-EM Data in RELION*. Methods Enzymol, 2016. **579**: p. 125-57.
227. Scheres, S.H.W. and S. Chen, *Prevention of overfitting in cryo-EM structure determination*. Nat Meth, 2012. **9**(9): p. 853-854.

228. Chen, S., G. McMullan, A.R. Faruqi, G.N. Murshudov, J.M. Short, S.H.W. Scheres, and R. Henderson, *High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy()*. Ultramicroscopy, 2013. **135**: p. 24-35.
229. Scherzer, O., *The Theoretical Resolution Limit of the Electron Microscope*. Journal of Applied Physics, 1949. **20**(1): p. 20-29.
230. Cheng, Y., *Single-particle cryo-EM at crystallographic resolution*. Cell, 2015. **161**(3): p. 450-7.
231. Henderson, R., S. Chen, J.Z. Chen, N. Grigorieff, L.A. Passmore, L. Ciccarelli, J.L. Rubinstein, R.A. Crowther, P.L. Stewart, and P.B. Rosenthal, *Tilt-Pair Analysis of Images from a Range of Different Specimens in Single-Particle Electron Cryomicroscopy*. Journal of Molecular Biology, 2011. **413**(5-3): p. 1028-1046.
232. Merk, A., A. Bartesaghi, S. Banerjee, V. Falconieri, P. Rao, M.I. Davis, R. Pragani, M.B. Boxer, Lesley A. Earl, Jacqueline L.S. Milne, and S. Subramaniam, *Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery*. Cell. **165**(7): p. 1698-1707.
233. Zernike, F., *Phase contrast, a new method for the microscopic observation of transparent objects*. Physica, 1942. **9**(7): p. 686-698.
234. Boersch, H., *Über die Kontraste von Atomen im Elektronenmikroskop*. Zeitschrift für Naturforschung A, 1947. **2**(11-12): p. 615-633.
235. Kanaya, K., H. Kawakatsu, K. Itō, and H. Yotsumoto, *Experiment on the electron phase microscope*. Journal of Applied Physics, 1958. **29**(7): p. 1046-1049.
236. Unwin, P., *Phase contrast and interference microscopy with the electron microscope*. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 1971: p. 95-104.
237. Unwin, P., *An electrostatic phase plate for the electron microscope*. Berichte der Bunsengesellschaft für physikalische Chemie, 1970. **74**(11): p. 1137-1141.
238. Danev, R. and K. Nagayama, *Transmission electron microscopy with Zernike phase plate*. Ultramicroscopy, 2001. **88**(4): p. 243-252.
239. Dai, W., C. Fu, D. Raytcheva, J. Flanagan, H.A. Khant, X. Liu, R.H. Rochat, C. Haase-Pettingell, J. Piret, S.J. Ludtke, K. Nagayama, M.F. Schmid, J.A. King,

- and W. Chiu, *Visualizing virus assembly intermediates inside marine cyanobacteria*. Nature, 2013. **502**(7473): p. 707-10.
240. Yamaguchi, M., R. Danev, K. Nishiyama, K. Sugawara, and K. Nagayama, *Zernike phase contrast electron microscopy of ice-embedded influenza A virus*. J Struct Biol, 2008. **162**(2): p. 271-6.
 241. Rochat, R.H., X. Liu, K. Murata, K. Nagayama, F.J. Rixon, and W. Chiu, *Seeing the Portal in Herpes Simplex Virus Type 1 B Capsids*. Journal of Virology, 2011. **85**(4): p. 1871-1874.
 242. Danev, R., R.M. Glaeser, and K. Nagayama, *Practical factors affecting the performance of a thin-film phase plate for transmission electron microscopy*. Ultramicroscopy, 2009. **109**(4): p. 312-25.
 243. Danev, R., B. Buijsse, M. Khoshouei, J.M. Plitzko, and W. Baumeister, *Volta potential phase plate for in-focus phase contrast transmission electron microscopy*. Proc Natl Acad Sci U S A, 2014. **111**(44): p. 15635-40.
 244. Mahamid, J., S. Pfeffer, M. Schaffer, E. Villa, R. Danev, L. Kuhn Cuellar, F. Förster, A.A. Hyman, J.M. Plitzko, and W. Baumeister, *Visualizing the molecular sociology at the HeLa cell nuclear periphery*. Science, 2016. **351**(6276): p. 969-972.
 245. Asano, S., Y. Fukuda, F. Beck, A. Aufderheide, F. Forster, R. Danev, and W. Baumeister, *Proteasomes. A molecular census of 26S proteasomes in intact neurons*. Science, 2015. **347**(6220): p. 439-42.
 246. Fukuda, Y., U. Laugks, V. Lučić, W. Baumeister, and R. Danev, *Electron cryotomography of vitrified cells with a Volta phase plate*. Journal of Structural Biology, 2015. **190**(2): p. 143-154.
 247. Danev, R., D. Tegunov, and W. Baumeister, *Using the Volta phase plate with defocus for cryo-EM single particle analysis*. eLife, 2017. **6**: p. e23006.
 248. Danev, R. and W. Baumeister, *Cryo-EM single particle analysis with the Volta phase plate*. eLife, 2016. **5**: p. e13046.
 249. Chua, E.Y.D., V.K. Vogirala, O. Inian, A.S.W. Wong, L. Nordenskiöld, J.M. Plitzko, R. Danev, and S. Sandin, *3.9 Å structure of the nucleosome core particle determined by phase-plate cryo-EM*. Nucleic Acids Research, 2016. **44**(17): p. 8013-8019.

250. Khoshouei, M., M. Radjainia, A.J. Phillips, J.A. Gerrard, A.K. Mitra, J.M. Plitzko, W. Baumeister, and R. Danev, *Volta phase plate cryo-EM of the small protein complex Prx3*. Nat Commun, 2016. **7**: p. 10534.
251. Liang, Y.-L., M. Khoshouei, M. Radjainia, Y. Zhang, A. Glukhova, J. Tarrasch, D.M. Thal, S.G.B. Furness, G. Christopoulos, T. Coudrat, *et al.*, *Phase-plate cryo-EM structure of a class B GPCR–G-protein complex*. Nature, 2017. **546**(7656): p. 118-123.
252. Lowary, P.T. and J. Widom, *New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning*. J Mol Biol, 1998. **276**(1): p. 19-42.
253. Dyer, P.N., R.S. Edayathumangalam, C.L. White, Y. Bao, S. Chakravarthy, U.M. Muthurajan, and K. Luger, *Reconstitution of nucleosome core particles from recombinant histones and DNA*, in *Methods in enzymology*. 2003, Elsevier. p. 23-44.
254. Chang, L., Z. Zhang, J. Yang, S.H. McLaughlin, and D. Barford, *Atomic structure of the APC/C and its mechanism of protein ubiquitination*. Nature, 2015. **522**(7557): p. 450-454.
255. Kastner, B., N. Fischer, M.M. Golas, B. Sander, P. Dube, D. Boehringer, K. Hartmuth, J. Deckert, F. Hauer, E. Wolf, *et al.*, *GraFix: sample preparation for single-particle electron cryomicroscopy*. Nat Methods, 2008. **5**(1): p. 53-5.
256. Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, and B. Schmid, *Fiji: an open-source platform for biological-image analysis*. Nature methods, 2012. **9**(7): p. 676-682.
257. Boskovic, J., J. Martinez-Gago, M. Mendez-Pertuz, A. Buscato, J.L. Martinez-Torrecedrada, and M.A. Blasco, *Molecular Architecture of Full-length TRF1 Favors Its Interaction with DNA*. J Biol Chem, 2016. **291**(41): p. 21829-21835.
258. Vasudevan, D., E.Y. Chua, and C.A. Davey, *Crystal structures of nucleosome core particles containing the '601' strong positioning sequence*. J Mol Biol, 2010. **403**(1): p. 1-10.
259. Chang, W., J.N. Dynek, and S. Smith, *TRF1 is degraded by ubiquitin-mediated proteolysis after release from*. Genes Dev, 2003. **17**(11): p. 1328-33.
260. Henderson, R., *The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules*. Q Rev Biophys, 1995. **28**(2): p. 171-93.

261. Richards, F.M. and J.R. Knowles, *Glutaraldehyde as a protein cross-linking reagent*. Journal of Molecular Biology, 1968. **37**(1): p. 231-233.
262. Frescas, D. and T. de Lange, *Binding of TPP1 protein to TIN2 protein is required for POT1a,b protein-mediated telomere protection*. J Biol Chem, 2014. **289**(35): p. 24180-7.
263. O'Connor, M.S., A. Safari, H. Xin, D. Liu, and Z. Songyang, *A critical role for TPP1 and TIN2 interaction in high-order telomeric complex assembly*. Proc Natl Acad Sci U S A, 2006. **103**(32): p. 11874-9.
264. Abreu, E., E. Artonovska, P. Reichenbach, G. Cristofari, B. Culp, R.M. Terns, J. Lingner, and M.P. Terns, *TIN2-tethered TPP1 recruits human telomerase to telomeres in vivo*. Mol Cell Biol, 2010. **30**(12): p. 2971-82.
265. Frescas, D. and T. de Lange, *TRF2-tethered TIN2 can mediate telomere protection by TPP1/POT1*. Molecular and cellular biology, 2014. **34**(7): p. 1349-1362.
266. McGuffin, L.J., K. Bryson, and D.T. Jones, *The PSIPRED protein structure prediction server*. Bioinformatics, 2000. **16**(4): p. 404-5.
267. Yachdav, G., E. Koppmann, L. Kajan, M. Hecht, T. Goldberg, T. Hamp, P. Honigschmid, A. Schafferhans, M. Roos, M. Bernhofer, et al., *PredictProtein--an open resource for online prediction of protein structural and functional features*. Nucleic Acids Res, 2014. **42**(Web Server issue): p. W337-43.
268. Cuff, J.A., M.E. Clamp, A.S. Siddiqui, M. Finlay, and G.J. Barton, *JPred: a consensus secondary structure prediction server*. Bioinformatics, 1998. **14**(10): p. 892-3.
269. Zhang, Y., *I-TASSER server for protein 3D structure prediction*. BMC Bioinformatics, 2008. **9**: p. 40.
270. Yang, J., R. Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang, *The I-TASSER Suite: protein structure and function prediction*. Nature methods, 2015. **12**(1): p. 7-8.
271. Kim, S.H., S. Han, Y.H. You, D.J. Chen, and J. Campisi, *The human telomere-associated protein TIN2 stimulates interactions between telomeric DNA tracts *in vitro**. EMBO reports, 2003. **4**(7): p. 685-691.
272. Greider, C.W., *Telomeres do D-loop-T-loop*. Cell, 1999. **97**(4): p. 419-22.

273. Wright, W.E., V.M. Tesmer, K.E. Huffman, S.D. Levene, and J.W. Shay, *Normal human chromosomes have long G-rich telomeric overhangs at one end*. Genes Dev, 1997. **11**(21): p. 2801-9.
274. McElligott, R. and R.J. Wellinger, *The terminal DNA structure of mammalian chromosomes*. Embo j, 1997. **16**(12): p. 3705-14.
275. Lingner, J. and T.R. Cech, *Purification of telomerase from Euplotes aediculatus: requirement of a primer 3' overhang*. Proc Natl Acad Sci U S A, 1996. **93**(20): p. 10712-7.
276. Denchi, E.L. and T. de Lange, *Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1*. Nature, 2007. **448**(7157): p. 1068-71.
277. Zhong, F.L., L.F. Batista, A. Freund, M.F. Pech, A.S. Venteicher, and S.E. Artandi, *TPP1 OB-fold domain controls telomere maintenance by recruiting telomerase to chromosome ends*. Cell, 2012. **150**(3): p. 481-94.
278. Sexton, A.N., D.T. Youmans, and K. Collins, *Specificity requirements for human telomere protein interaction with telomerase holoenzyme*. J Biol Chem, 2012. **287**(41): p. 34455-64.
279. Bieniossek, C., T.J. Richmond, and I. Berger, *MultiBac: multigene baculovirus-based eukaryotic protein complex production*. Curr Protoc Protein Sci, 2008. **Chapter 5**: p. Unit 5.20.
280. Danev, R. and D. Tegunov, *Using the Volta phase plate with defocus for cryo-EM single particle analysis*. 2017. **6**.
281. Danev, R. and W. Baumeister, *Cryo-EM single particle analysis with the Volta phase plate*. Elife, 2016. **5**.
282. Liang, Y.L., M. Khoshouei, M. Radjainia, Y. Zhang, A. Glukhova, J. Tarrasch, D.M. Thal, S.G.B. Furness, G. Christopoulos, T. Coudrat, *et al.*, *Phase-plate cryo-EM structure of a class B GPCR-G-protein complex*. Nature, 2017. **546**(7656): p. 118-123.
283. Khoshouei, M., M. Radjainia, W. Baumeister, and R. Danev, *Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate*. 2017. **8**: p. 16099.
284. Chua, E.Y.D. and S. Sandin, *Advances in phase plate cryo-EM imaging of DNA and nucleosomes*. Nucleus, 2017. **8**(3): p. 275-278.

285. Xin, H., D. Liu, M. Wan, A. Safari, H. Kim, W. Sun, M.S. O'Connor, and Z. Songyang, *TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase*. *Nature*, 2007. **445**(7127): p. 559-62.
286. Lim, C.J., A.J. Zaug, H.J. Kim, and T.R. Cech, *Reconstitution of human shelterin complexes reveals unexpected stoichiometry and dual pathways to enhance telomerase processivity*. *Nature Communications*, 2017. **8**(1): p. 1075.
287. Murzin, A.G., *OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences*. *The EMBO Journal*, 1993. **12**(3): p. 861-867.
288. Lopez-Gallego, F., J.M. Guisan, and L. Betancor, *Glutaraldehyde-mediated protein immobilization*. *Methods Mol Biol*, 2013. **1051**: p. 33-41.
289. Walt, D.R. and V.I. Agayn, *The chemistry of enzyme and protein immobilization with glutaraldehyde*. *TrAC Trends in Analytical Chemistry*, 1994. **13**(10): p. 425-430.
290. Lusty, C., *A gentle vapor-diffusion technique for cross-linking of protein crystals for cryocrystallography*. *Journal of Applied Crystallography*, 1999. **32**(1): p. 106-112.
291. Heras, B. and J.L. Martin, *Post-crystallization treatments for improving diffraction quality of protein crystals*. *Acta Crystallogr D Biol Crystallogr*, 2005. **61**(Pt 9): p. 1173-80.
292. Prento, P., *Glutaraldehyde for electron microscopy: a practical investigation of commercial glutaraldehydes and glutaraldehyde-storage conditions*. *Histochem J*, 1995. **27**(11): p. 906-13.
293. Stark, H., *GraFix: Stabilization of Fragile Macromolecular Complexes for Single Particle Cryo-EM*. *Methods in Enzymology*, 2010. **481**: p. 109-126.
294. Sabatini, D.D., K. Bensh, and R.J. Barnett, *Cytochemistry and electron microscopy*. *The Journal of Cell Biology*, 1963. **17**(1): p. 19-58.
295. Migneault, I., C. Dartiguenave, M.J. Bertrand, and K.C. Waldron, *Glutaraldehyde: behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking*. *Biotechniques*, 2004. **37**(5): p. 790-6, 798-802.
296. Habeeb, A.J. and R. Hiramoto, *Reaction of proteins with glutaraldehyde*. *Arch Biochem Biophys*, 1968. **126**(1): p. 16-26.

297. Yoshimura, S.H., H. Maruyama, F. Ishikawa, R. Ohki, and K. Takeyasu, *Molecular mechanisms of DNA end-loop formation by TRF2*. *Genes to Cells*, 2004. **9**(3): p. 205-218.
298. Baynes, B.M. and B.L. Trout, *Rational Design of Solution Additives for the Prevention of Protein Aggregation*. *Biophysical Journal*, 2004. **87**(3): p. 1631-1639.
299. Kucukelbir, A., F.J. Sigworth, and H.D. Tagare, *Quantifying the local resolution of cryo-EM density maps*. *Nat Methods*, 2014. **11**(1): p. 63-5.

Appendix

A1: Media and gel recipes

LB media	1.0% (w/v) Bacto Tryptone 0.5% (w/v) Yeast extract 5.0% (w/v) NaCl
2xTY media	1.6% (w/v) Bacto Tryptone 1.0% (w/v) Yeast extract 5.0% (w/v) NaCl
8 % SDS-PAGE denaturing gel	8% (v/v) Acrylamide 0.25% (v/v) Bis 375 mM Tris-HCl (pH 8.8) 0.1% (w/v) SDS 0.1% (v/v) APS 0.1% (v/v) TEMED
12 % SDS-PAGE denaturing gel	12% (v/v) Acrylamide 0.25% (v/v) Bis 375 mM Tris-HCl (pH 8.8) 0.1% (w/v) SDS 0.1% (v/v) APS 0.1% (v/v) TEMED
4% SDS-PAGE stacking gel	4% (v/v) Acrylamide 0.25% (v/v) Bis 125 mM Tris-HCl (pH 6.8) 0.1% (w/v) SDS 0.1% (v/v) APS 0.1% (v/v) TEMED
6% Native polyacrylamide gel	6% (v/v) Acrylamide 0.15% (v/v) Bis 0.25x TBE buffer 0.1% (v/v) APS 0.1% (v/v) TEMED
1% Agarose gel	1% (w/v) agarose in 0.5x TBE buffer
TBE	89 mM Tris-HCl (pH 7.5) 89 mM Boric acid 2 mM EDTA

A2: Protein constructs

1. Full-length human POT1 (634 amino acids, 71 kDa)

MSLVPATNYIYTPLNQLKGGTIVNVYGVVKFFKPPYLSKGTDYCSVVTIVDQTNVKTCLLFS
GNYEALPIIYKNGDIVRFHRLKIQVYKKETQGITSSGFASLTFFEGTLGAPIIPRTSSKYFNF
TTEDHKMVEALRVWASTHMSPSWTLLKLCADVQPMQYFDLTCQLLGKAEVDGASFLKVVWDGT
RTPFPSSWRVLIQDLVLEGDLSHIHRLQNLTIDILVYDNHVVHVARSLKVGSLFRIYSLHTKLQ
SMNSENQTMLSLEFHLHGGTSYGRGIRVLPESNSDQDLKKDLESANLTANQHSDVICQSEP
DDSFPSGSGSVSLYEVEERCQQLSATILTDHQYLERLTPLCAILKQKAPQQYRIRAKLSYKPRR
LFQSVKLHCPKCHLLQEVPHGEDLDIIFQDGATKTPDVKLQNTSLYDSKIWTTKNQKGRKVA
VHFVKNNGILPLSNECLLLIEGGTLSEICKLSNKFNSVIPVRSGHEDLELLDLSAPFLIQGT
IHYYGCKQCSSLRSIQNNSLVDKTSWIPSSVAEALGIVPLQYVFVMTFTLDDGTGVLEAYLM
DSDKFFQIPASEVLMDDDLQKSVDMMDMFCPPGIKIDAYPWLECFIKSYNVTNGTDNQICY
QIFDTTVAEDVI

2. Full-length human RAP1 (399 amino acids, 44 kDa)

MAEAMD LGKDPNGPTHSSSTLFVRDDGSSMSFYVRPSPAKRRLSTLILHGGGTVCRVQEPGAV
LLAQPG EALAEASGDFISTQYILDCVERNERLELEAYRLGPASAADTGSEAKPGALAEGAAE
PEPQRHAGRIAF TDADDVAILTYVKENARSPSSVTGNALWKAMEKSSLTQHSWQSLKDRYLK
HLRGQEHKYL LGDAPVSPSSQKLKRKAEDPEAADSGEPQNKRTPDLP EEEYVKEEIQENEE
AVKKMLVEATREFEEVVVDESPDFEIHITMCDDDPPTPEEDSETQPDEEEEEEEEEKVSQPE
VGAAIKIIRQLMEKFNLDLSTVTQAFLKNSGELEATSAFLASGQRADGYPIWSRQDDIDLQK
DDEDTREALVKKFGAQNVARRIEFRKK

3. Full-length human TIN2L (451 amino acids, 50 kDa)

MATPLVAGPAALRFAAAASWQVVRGRCVEHFPRVLEFLRSLRAVAPGLVRYRHHHERLCMGLK
AKVVVELILQGRPWAQVLKALNHHFPESGPIVRDPKATKQDLRKILEAQETFYQQVKQLSEA
PVDLASKLQELEQEYGEFPLAAMEKLLFEYLCQLEKALPTPQAQQQLQDVLSWMQPGVSITSS
LAWRQYGVDMGWLLPECSVTDSVNLAEPMEQNPPQQQRLALHNPLPKAKPGTHLPQGPSSRT
HPEPLAGRHFNLAPLGRRRVQSQWASTRGGHKERPTVMLFPFRNLGSPTQVISKPEESKEEHA
IYTADLAMGTRAASTGKSKSPCQTLGGRALKENPVDLPATEQKENCLDCYMDPLRLSLLPPR
ARKPVCPPSLCSSLVITIGDLVLDSDEEENGQGEKESLENYQKTKFDTLIPTLCEYLPPSGH
GAIPVSSCDCRDSSRPL

4. Full-length human TIN2S (354 amino acids, 39 kDa)

MATPLVAGPAALRFAAAASWQVVRGRCVEHFPRVLEFLRSLRAVAPGLVRYRHHHERLCMGLK
AKVVVELILQGRPWAQVLKALNHHFPESGPIVRDPKATKQDLRKILEAQETFYQQVKQLSEA
PVDLASKLQELEQEYGEFPLAAMEKLLFEYLCQLEKALPTPQAQQQLQDVLSWMQPGVSITSS
LAWRQYGVDMGWLLPECSVTDSVNLAEPMEQNPPQQQRLALHNPLPKAKPGTHLPQGPSSRT
HPEPLAGRHFNLAPLGRRRVQSQWASTRGGHKERPTVMLFPFRNLGSPTQVISKPEESKEEHA
IYTADLAMGTRAASTGKSKSPCQTLGGRALKENPVDLPATEQKE

5. Full-length human TPP1 (544 amino acids, 57 kDa)

MPGRCQSDAAMRVNGPASRAPAGWTSGLHTGPRAGRPRQAQARGVRGRGLLLRPRPAKELPL
PRKGGAWAPAGNPGPLHPLGVAVGMAGSGRLVLRPWIRELILGSETPSSPRAGQLLEVLQDA
EAAVAGPSHAPDTSVDGATLLVSDGTHSVRCLVTREALDTSWEEKEFGFRGTEGRLLLLQD
CGVHVQVAEGGAPAEFYLVDRFSLLPTEQPRLRVPGCNQDLVDVQKKLYDCLEEHLSESTSS
NAGLSLSQLLDEMREDQEHQGALVCLAESCLTLEGPCCTAPPVTHWAASRCKATGEAVYTVPS
SMLCISENDQLILSSLGPCQRTQGPELPPDPALQDLSLTLIASPPSSPSSSGTPALPGHMS
SEESGTSISLLPALSLAAPDPGQRSSSQPSAICSAPATLTSPRSPHASRTPSSPLQSCTPSL
SPRSHVPSPHQALVTRPQKPSLEFKEFVGLPCKNRPPFPRTGATRGAQEPSCSVWEPPKRHRD
GSAFQYEEYPPCTSLCARVQAVRLPPQLMAWALHFLMDAQPGSEPTPM

6. Full-length human TRF1 (439 amino acids, 50 kDa)

MAEDVSSAAPSPRGCADGRDADPTEEQMAETERNDEEQFECQELLECCQVQVGAPEEEEEEEE
DAGLVAAEAEAVAAGWMLDFLCLSLCRAFRDGRSEDFRTRNSAEAI IHGLSSLTACQLRTIY
ICQFLTRIAAGKTLDAQFENDERITPLESALMIWGSIEKEHDKLHEEIQNLIKIQAIAVCME
NGNFKEAAEEVFERIFGDPNSHMPFKSKLLMIISQKDTFHSFFQHFSYNHMMEKIKSYVNYVL
SEKSSTFLMKAAAKVVESKRTRTITSQDKPSGNDVEMETEANLDTRKSVSDKQSAVTESSEG
TVSLLRSHKNLFLSKLQHGTQQQDLNKKERRVGTTPQSTKKKKESRRATESRIPVSKSQPVTP
EKHRARKRQAWLWEEDKNLRSGVRKYGEGNWSKILLHYKFNNRTSVMLKDRWRTMKKLKLIS
SDSED

7. Full-length human TRF2 (500 amino acids, 55 kDa)

MAGGGGSSDGSGRAAGRASRSSGRARRGRHEPGLGGAERGAGEARLEEAVNRWVLKFYFH
EALRAFRGSRYGDFRQIRDIMQALLVRPLGKEHTVSRLLRVMQCLSRIEEENLDCSFDMEA
ELTPLESAINVLEMIKTEFTLTEAVVESRKLKVEAAVVICIKNKEFEKASKILKKHMSKDP
TTQKLRLNDLLNIIREKNLAHPVIQNFSYETFQQKMLRFLESHLDDAEPYLLTMAKKALKSES
AASSTGKEDKQPAPGPVEKPPREPARQLRNPPTTIGMMTLKAAFKTLSGAQDSEAAFAKLDQ
KDLVLPTQALPASPALKNRPRKDENESSAPADGEGGSELQPKNKRMTISRVLLEEDSQSTE
PSAGLNSSQEAAASAPPSKPTVLNQPLPGEKNPKVPKGKWNSSNGVEEKETWVEEDELQVQQA
APDEDSTTNITKKQKWTVEESEWVKAGVQKYGEGNWAAISKNYPFVNRTAVMIKDRWRTMKR
LGMN

A3: Bacmids and plasmids

Bacterial constructs	Plasmid	Source
His ₆ -S-TEV-TRF1	pET30a	Rhodes' lab/ MRC, Cambridge
His ₆ -S-TEV-TRF2	pET30a	Rhodes' lab/ MRC, Cambridge
His ₆ -S-TEV-RAP1	pET30a	Rhodes' lab/ MRC, Cambridge
His ₆ -S-TEV-TIN2L	pET30a	Rhodes' lab/ MRC, Cambridge
His ₆ -S-TEV-MBP-TIN2L	pET30a	Cloned
GST-PreScission-His ₆ -TEV-TIN2S	pNIC28-Bsa4	Rhodes' lab/ Protein production platform (PPP), NTU

Baculoviral constructs	Bacmid	Source
Construct 1 MG-His ₁₀ -SSG-TEV-POT1 MG-Strep-GS-TEV-TPP1 (1-544)	pFL + pUCDM	Rhodes' lab
Construct 2 MG-His ₁₀ -SSG-TEV-POT1 MG-Strep-GS-TEV-TPP1 (87-544)	pFL + pUCDM	Rhodes' lab
Construct 6 MG-His ₁₀ -SSG-TEV-POT1 MG-Strep-GS-TEV-TPP1 (1-544) TIN2L	pFL + pUCDM	Rhodes' lab
Construct 7 MG-His ₁₀ -SSG-TEV-POT1 MG-Strep-GS-TEV-TPP1 (87-544) TIN2L	pFL + pUCDM	Rhodes' lab
Construct 9 MG-His ₁₀ -SSG-TEV-POT1 MG-Strep-GS-TEV-TPP1 (1-544) TIN2S	pFL + pUCDM	Rhodes' lab
Construct 10 MG-His ₁₀ -SSG-TEV-POT1 MG-Strep-GS-TEV-TPP1 (87-544) TIN2S	pFL + pUCDM	Rhodes' lab

A5: Publication

Chua, Eugene YD, Vinod K. Vogirala, Oviya Inian, Andrew SW Wong, Lars Nordenskiöld, Juergen M. Plitzko, Radostin Danev, and Sara Sandin. **"3.9 Å structure of the nucleosome core particle determined by phase-plate cryo-EM."** *Nucleic acids research* 44, no. 17 (2016): 8013-8019.