A biochemical analysis of human shelterin complexes

Yexin Xie

Divisions of Structural Biology and Cancer Biology

Chester Beatty Labs

The Institute of Cancer Research

University of London

This thesis is submitted for the degree of Doctor of Philosophy

March 2022

Declaration

The work described in this thesis was carried out at the Institute of Cancer Research, London, under the supervision of Dr Sebastian Guettler.

I, Yexin Xie, declare that the work presented in this thesis is my own, and, where information has been derived from other sources, I confirm that this has been clearly indicated in the thesis.

Signature

Date

2022/03/31

“学而不思则罔，思而不学则殆”

“Learning without thinking means wasted work;

thinking without learning is dangerous”

Acknowledgements

Firstly, I would like to thank Sebastian for giving me this opportunity to conduct the research detailed here. Over the past four and half years, Sebastian has been hugely supportive, not only with respect to the project, but also in areas of personal and career development. I am very grateful for his contagious enthusiasm and curiosity, which has been the nucleation of many interesting experiments and discussions.

Next, I would like to thank previous and present lab members. In particular, I would like to thank Mike and Mariola for sharing their scientific expertise and guidance and easing my transition from a fresh BSc graduate into the fields of biochemistry, biophysics, and structural biology. I would like to thank Pete for initiating my path onto the shelterin project, Iris for the many constructs and protocols her work provided, and Oviya for sharing our passion on shelterin and the many wonderful conversations, which gave birth to countless ideas and experiments. Many thanks to Nisha and Matt, for giving their time and knowledge on mammalian tissue culture and electron microscopy, respectively, and to the PhD students Ben, Giacomo, Iona, Saira, and Muhammed who brought much cheer and inspiration.

I would like to thank all the teams in the Division of Structural biology for their genuine interest when it comes to troubleshooting and discussing challenging projects, and for creating such a fun and friendly working environment. I am indebted to Fabienne for dedicating her time for EM training and help with data collection, and her invaluable knowledge on image analysis and data interpretation. I am also very grateful for the contributions from Dr Jyoti Choudhary, Dr Lu Yu, and Dr Theo Roumeliotis at the ICR's Proteomics Facility for their help in running and optimising the mass spectrometry experiments and in subsequent data analyses. I would also like to thank Ruth for maintaining the baculovirus lab, Chris and his 'magic fingers' for solving all IT challenges, and Jane for managing the labs.

A huge thank you to my friends, in particular Elizabeth, Alireza and Tara, Hristina, Boyana, Veronica, Madalena and Emma, who have made my past nine years in London a pure joy.

Last but not least, I would like to thank my parents, Mingyan and Jicheng, and Buri for their endless support and their overflowing positivity, for being there at times of hardships and happiness. I will be forever indebted to their generosity.

Abstract

Linear chromosomes present two major challenges, known as the end-replication and end-protection problems. The human shelterin complex is a dynamic assembly of six components, consisting of TRF1, TRF2, RAP1, TIN2, TPP1, and POT1. These proteins decorate the telomeres, protecting the ends by suppressing DNA-damage responses and regulating telomere length maintenance. Impairment in either of these functions can lead to multiple human diseases. For instance, mutations in components of shelterin and telomerase cause the severe, progressive genetic disorder dyskeratosis congenita (DC), where the signs and symptoms reflect multi-system stem cell failure. Moreover, mutations in multiple shelterin subunits have been implicated in a wide range of human malignancies.

Over the past three decades, the protein-protein and protein-DNA interactions within shelterin has been extensively characterised. High-resolution structures have been solved for domains and captured domain-peptide interactions. However, the overall architecture of human shelterin complexes is still unknown. Here, I report the expression and purification of the full shelterin complex and four subcomplexes. I describe biochemical and biophysical studies to explore the size and stoichiometry of shelterin complexes, and their affinity for telomeric and non-telomeric DNA substrates. Integrating data from crosslinking mass spectrometry and negative-stain electron microscopy, I illustrate that human shelterin subunits can assemble into dynamic complexes that are heterogenous in both composition and conformation.

Table of Contents

List of figures

List of tables

List of Abbreviations

AcNPV	Autographa californica nuclear polyhedrosis virus
ALT	alternative lengthening of telomeres
ARC	ankyrin repeat cluster
ART	ADP-ribosyltransferase
ATM	ataxia-telangiectasia mutated
ATR	ataxia-telangiectasia Rad3 related
BME	β-mercaptoethanol
BMF	bone marrow failure
BRCT	BRCA1 C-terminal
BSA	bovine serum albumin
CV	column volume
CCD	charge-coupled device
CE	collision energy
CID	collision-induced dissociation
CMOS	complementary metal oxide sensor
cryo-EM	cryo-electron microscopy
cryo-ET	cryo-electron tomography
CST	CTC1-STN1-TEN1 complex
CTE	C-terminal extension
CTF	contrast transfer function
dStrepII	double-StrepII
DBD	DNA-binding domain
DC	dyskeratosis congenita
DDR	DNA damage response
DLS	dynamic light scattering
DMSO	dimethyl sulfoxide
DQE	detective quantum efficiency
dRI	differential refractive index
DSB	double strand break
dsDNA	double-stranded DNA
DSSO	disuccinimidyl sulfoxide
EM	electron microscopy
EMSA	electrophoretic mobility shift assay
FBS	foetal bovine serum
FPS	fluorescence proximity sensing
FSC	Fourier shell correlation
FT	Fourier transform

FTMS	Fourier transform mass spectrometry
FWHM	full width at half-maximum
GEC	gene expression cassette
GraFix	gradient fixation
HDR	homology-directed repair
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HID	high-energy collisional dissociation
His_6	hexa-histidine
His_{10}	deca-histidine
HJRL	Holliday junction resolvase-like
HTH	helix-turn-helix
HSV-tk	herpes simplex virus type 1 thymidine kinase
HYE	histidine-tyrosine-glutamine
IAA	iodoacetamide
IEX	ion exchange chromatography
IMAC	ion metal affinity chromatography
ITMS	ion trap mass spectrometry
KDE	kernel density estimates
LS	light scattering
ML	maximum likelihood
MALS	multi-angle light scattering
MAPS	monolithic active pixel sensors
MS	mass spectrometry
MSA	multivariate statistical analysis
MW	molecular weight
MWCO	molecular weight cut-off
NAD+	nicotinamide adenine dinucleotide
NHEJ	Non-homologous end joining
NHS	N-hydroxysuccinimide
NMR	nuclear magnetic resonance
NOB	N-terminus of the OB-fold
OB-fold	oligonucleotide/oligosaccharide-binding fold
PAGE	poly-acrylamide gel electrophoresis
PAR	poly-ADP-ribose
PARP	poly(ADP-ribose) polymerase
PARylation	poly(ADP-ribosyl)ation
PBS	phosphate-buffered saline
PBST	phosphate-buffered saline-Tween
PCA	principal component analysis

PCR	polymerase chain reaction
PIM	POT1-interacting motif
Pk/t	pseudoknot/template
POT1	protector of telomeres 1
PTM	post-translational modification
R_g	radius of gyration
RAP1	repressor/activator protein 1
RCT	RAP1 C-terminal
RI	refractive index
RIM	RAP1-interacting motif
RMS	root mean squared
RNP	ribonucleoprotein
RPA	replication protein A
RT	reverse transcriptase
SAM	sterile alpha motif
SDS	sodium dodecyl sulphate
SEC	size exclusion chromatography
SV40	simian virus 40
SNR	signal-to-noise ratio
ssDNA	single-stranded DNA
T-loop	telomeric loop
TBM	TRF-binding motif
TCEP	tris(2-carboxyethyl)phosphine
TEAB	triethylammonium bicarbonate buffer
TEBP	telomere end-binding protein
TEL	TPP1 glutamate- and leucine-rich
TEMED	tetramethylethylenediamine
TEN	telomerase N-terminal
TERC	telomerase RNA component
TERT	telomerase reverse transcriptase
TEV	tobacco etch virus
TIF	telomere dysfunction-induced foci
TIM	TIN2-interacting motif
TIN2	TRF1-interacting nuclear protein 2
TNKS	tankyrase 1
TNKS2	tankyrase 2
TPP1	TIN2- and POT1-interacting protein
TR	telomerase RNA
TRBD	telomerase RNA-binding

TRF	telomere repeat factor
TRFH	TRF homodimerisation
TERT	telomerase reverse transcriptase
TRF1	telomeric repeat-binding factor 1
TRF2	telomeric repeat-binding factor 2
Tris	tris(hydroxymethyl)aminomethane
XL-MS	crosslinking mass spectrometry
XRC	X-ray crystallography
YFP	yellow fluorescent protein

1 Introduction

In this chapter, I will first provide a historical perspective of the discovery of telomeres, followed by an overview of linear chromosomes in eukaryotes and the they pose for cell survival. Then, I describe the protein components involved in solving these challenges, with a focus on the shelterin complex and its interacting partners, namely telomerase and tankyrase. Finally, I pose the research questions central to my PhD project.

1.1 Telomeres

All organisms face a lifetime challenge of DNA maintenance, and the consequences of failure are fatal. Genomic instability is a major hallmark of aging and cancer (López-Otín, et al., 2013; Hanahan & Weinberg, 2011). In eukaryotes, this task is complicated by the linear structure of DNA. Telomeres (derived from the Greek nouns *telos* for 'end' and *meros* for 'part') are specialised nucleoprotein complexes that cap the end of linear chromosomes. Despite the variations in the composition of telomeric DNA and its associated proteins across different species, they perform similar functions: (1) the telomeric repeats and its associated proteins provides a mechanism of maintaining the chromosome ends, preventing sequence loss due to incomplete replication by the conventional DNA replication machinery during cell division; (2) the telomere-associated proteins seals the chromosome ends and suppresses the DNA damage repair machinery, protecting authentic chromosome ends from illicit repair and chromosomal aberrations.

1.1.1 The nature of linear chromosome ends in eukaryotes

Protein-encoding genetic information is stored and transmitted through condensed nucleoprotein structures known as chromosomes. The ends of eukaryotic linear chromosomes are sealed by a stretch of non-coding DNA of variable sequences and lengths, known as telomeres. Over 40 years ago, the first telomere sequence was characterised by Elizabeth Blackburn and Joseph Gall in a ciliated protozoan *Tetrahymena thermophila*, which consisted of a variable number of GC-rich tandemly repeated sequence (Blackburn & Gall, 1978). Subsequent experiments with *Tetrahymena* and yeast telomeric sequences by Jack Szostak and colleagues demonstrated that yeast cells were able to recognise and extend telomeric DNA from *Tetrahymena*, despite the distant evolutionary relationship between the two species (Szostak & Blackburn, 1982). Soon after, different telomeric sequences have been identified across a variety of eukaryotic organisms (Table 1.1). Interestingly, they all share similar characteristics with those originally described in *Tetrahymena*. The conserved nature of telomeres and the telomere maintenance infrastructure (described below) indicated that telomeric function is also evolutionary conserved.

It is now known that most eukaryotic telomeres consist of well-conserved GC-rich repetitive sequences, with the G-rich and C-rich strands synthesised as the leading and lagging strands, respectively. In humans, the chromosome ends are sealed by 5 – 15 kbp of TTAGGG dsDNA repeats, followed by a short protrusion of the G-rich strand at the 3'-end that forms a ssDNA overhang (Table 1.1). Both the dsDNA and ssDNA regions are critical for telomeric function (Klobutcher, et al., 1981). Besides its linear sequence, evidence suggest that telomeric DNA can generate several different higher order structures to assist telomeric function. Formation of these structures depends on the physical properties of the GC-rich DNA sequence commonly found at telomeres, and the presence of specific telomeric proteins (see sections 1.3.1 and **Error! Reference source not found.**).

Organism	Sequence	dsDNA length	ssDNA overhang length
H. sapiens	TTAGGG	5 – 15 kb	30 – 500 nt
M. musculus	TTAGGG	20 – 100 kb	Not determined
S. pombe	$GGTTACA_{0-1}C_{0-1}G_{0-6}$	5 kb	< 50 nt
S. cerevisiae	TG_{1-3}	300 bp	14 nt
T. brucei	GGGTTA	2 – 26 kb	10 nt (~ 40 nt at few ends)
T. thermophilia	TTGGGG	120 – 420 kb	14 – 15 or 20 – 21
C. elegans	TTAGGC	4 – 9 kb	> 30 nt (G-rich and C-rich)
A. thaliana	TTTAGGG	2 – 5 kb	20 – 30 nt
Oxytricha	TTTTGGGG	20 kb	16 nt

Table 1.1. Telomere length and composition vary across different species.

Adapted from (Sfeir, 2012).

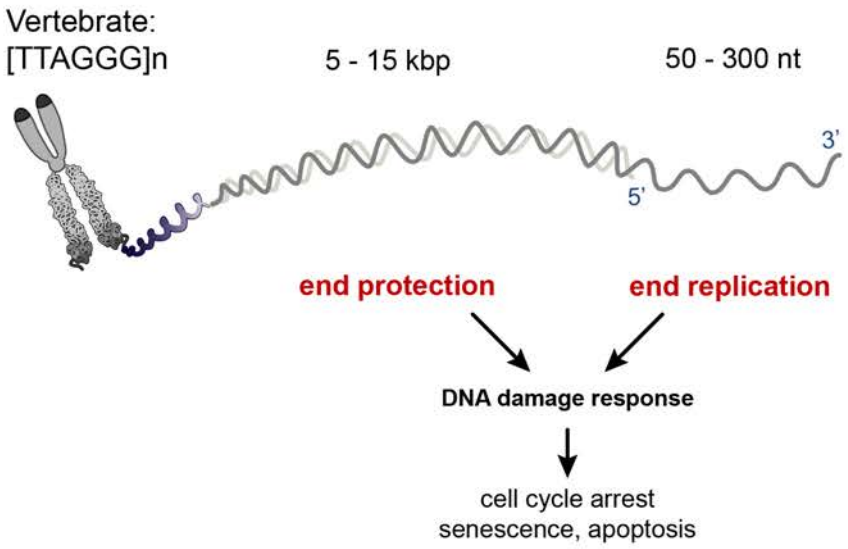

Figure 1.1. Telomeres and the problems of end protection and replication.

Vertebrates have the repetitive TTAGGG telomeric sequence at the end of the linear chromosomes. In humans, the telomeric dsDNA are ~ 5 – 15 kbp, followed by a short 50 – 300 nt of G-rich 3' overhang. This DNA structure resembles DNA double strand breaks and can elicit DNA damage responses, which results in cell cycle arrest, senescence, and apoptosis. The end replication problem results in gradual shortening of telomeres in somatic cells. Critically short telomeres can no longer be protected and become susceptible to DNA damage repair pathways.

1.1.1.1 **The end-replication problem**

As early as the 1860s, speculations were made by evolutionary biologists Alfred Russel Wallace and August Weismann that ‘cell-division is not ever-lasting’. They provided an explanation for the process of aging and death arises using theories of natural selection (Weismann, 1891). However, opposing views from the early 1900s argued against a cellular theory of aging. Among others, the Nobel-Prize winning French surgeon and biologist Alexis Carrel argued that the limit on cell replication was established by the culture conditions, and that under ideal conditions cells can be maintained indefinitely in vitro (Carrel & Ebeling, 1921; Carrel, 1912).

A century later, the idea of limited cell renewal resurfaced as Leonard Hayflick and Paul Moorhead described a series of changes in human cells in continuous culture (Hayflick & Moorhead, 1961; Hayflick, 1965). Signs of senescence include increased generation time, metabolic changes and reduced mitotic activity, and finally degeneration and cell death. The finite replicative capacity of non-transformed cells became known as the Hayflick limit and suggests a counting mechanism of cellular senescence. More clues to the underlying molecular mechanisms appeared soon after the elucidation of the structure of DNA and the mechanism of genetic material transmission through semiconservative DNA replication (Watson & Crick, 1953; Meselson & Stahl, 1958). DNA polymerase requires an RNA primer for initiating DNA synthesis, which proceeds in the 5’ to 3’ direction. Synthesis of the leading strand (in the 5’ to 3’ direction, towards the direction of travel of the replication fork) is straightforward and can continue until the end of the telomere, producing a blunt end. The lagging strand grows in a direction opposite to the replication fork and occurs in a series of steps, known as the Okazaki fragments, each requiring an RNA primer (Okazaki, et al., 1968). Removal of the final RNA primer at the extreme 3’ end of the template generates a gap that cannot be filled by DNA polymerase, resulting in a 3’ overhang in the lagging strand. Therefore, this mechanism of DNA replication implies that with each such cycle of genome duplication during cell replication, chromosome ends would become shorter (telomere attrition). This forms the basis of the ’end-replication problem’ (Figure

1.1) (Watson, 1972; Olovnikov, 1973). In one description, Alexey Olovnikov depicted the extreme terminal of linear chromosomes (’telogenes’) as ’buffers’, which are ’sacrificed’ in place of the ’vitally important genes’ located at the chromosome ends during cell division (Olovnikov, 1973). The solution to the end-replication problem came with the discovery of an alternative DNA lengthening mechanism that synthesised and elongated telomeres (see section 1.4.1).

1.1.1.2 **The end-protection problem**

Over 80 years ago, Hermann Muller and Barbara McClintock first observed in fruit flies and maize that authentic chromosome termini are resistant to a range of chromosomal aberrations resembling reactions to DNA breakage, which occurred when these natural ends were removed (Muller, 1938; McClintock, 1941). This problem is one of ’end protection’, which emerges from the resemblance of chromosome ends to DNA double strand breaks (DSBs), the most lethal type of DNA lesions (Figure 1.1). In cells, the exposure of DNA breakage induces cell cycle arrest to allow time for DNA damage response (DDR). DSB sensing is orchestrated by the ataxia-telangiectasia mutated (ATM) and ATM and RAD3-related (ATR) kinases (Awasthi, et al., 2015). Whilst ATM-CHK2 signalling is activated by dsDNA at the break site, ATR-CHK1 recruitment requires the presence of ssDNA (Blackford & Jackson, 2017). Optimal repair of DSBs is mainly mediated by homology-directed repair (HDR), and classical and alternative non-homologous end joining (c-NHEJ and alt-NHEJ, respectively) (Scully, et al., 2019). The pathway decisions are complex and depend on the exact circumstances, including the cell cycle stage and 5’ end resection (Hustedt & Durocher, 2016). NHEJ is primarily involved in DSB repair in G1, although it can occur throughout the cell cycle. On the other hand, HDR can only happen during S/G2 and is the dominant repair pathway in the presence of sister chromatids (Lieber, 2010). In addition, the extent of end resection is important in repair pathway choice. The process is tightly regulated, and its presence suppresses NHEJ whilst promoting RAD51-mediated HDR (Lieber, 2010).

Outside the telomeres, DDR is critical to allow cells to continue dividing with an intact genome. However, DDR activation at the telomeres induces cell cycle arrest, senescence, and genomic instability (Fumagalli, et al., 2012). Theoretically, human telomeres contain DNA structures can activate both ATM and ATR signalling and are susceptible to HDR and NHEJ repair. However, this is not observed in cells with functional telomeres, pointing towards the existence of inhibitory mechanism or mechanisms (described in section 1.3.3).

1.1.2 Discovery of proteins at the telomeres

The repetitive telomeric DNA sequences provide the basis for telomere maintenance. However, on its own does not solve either the end-protection or the end-replication problems. In fact, the 5' dsDNA coupled with the 3' ssDNA protrusion resembles structures at DSBs that has the potential to elicit DNA damage responses. Moreover, the end-replication problem arises from the incomplete replication of DNA at the 3' end during cell division. The nature of the telomeric DNA does not bypass this deficiency of the DNA-dependent DNA polymerase. Therefore, it soon became clear that telomeric DNA recruits various proteins to assist in telomere homeostasis.

The first telomere-associated proteins were identified in unicellular eukaryotes from the protozoan and fungal kingdoms (Forney, et al., 1987). They share features with proteins in the shelterin complex and 'shielded' chromosome ends and contributes to the chromosome end protection role attributed to telomeres. From early nuclease and chemical footprinting experiments, researchers noted that the terminal ~100 bp at the telomeres of *Oxytricha* and other ciliates are protected from cleavage (Gottschling & Cech, 1984; Gottschling & Zakian, 1986; Price, 1990). This indicated the presence of proteins at the telomeres and launched a search for telomere-associated proteins. In *Oxytricha nova*, this protection phenomenon was attributed to a two-subunit complex known

as telomere end binding protein (*On*TEBP) (Figure 1.2). This heterodimer was found to bind the ssDNA extension of *O. nova* telomeres with high specificity and affinity (Gottschling & Zakian, 1986). Next, a known transcription regulator in yeast, Rap1, were found to bind specifically to telomeric DNA (Buchman, et al., 1988). The first two telomere-binding proteins in human, TRF1 and TRF2, were identified also through their recognition of the human telomeric dsDNA sequence (Zhong, et al., 1992; Chong, et al., 1995; Bilaud, et al., 1997; Broccoli, et al., 1997). TIN2 and RAP1 emerged from yeast two-hybrid screens with TRF1 and TRF2, respectively (Kim, et al., 1999; Li, et al., 2000). POT1 was first identified from databases searches for homologs of fission yeast Pot1 (Baumann & Cech, 2001). Finally, TPP1 (previously known as TINT1/PTOP/PIP1) was discovered as a TIN2-interacting protein (Houghtaling, et al., 2004). Together these six proteins form the human shelterin complex.

1.2 The shelterin complex

Functional telomeres require a range of telomere-interacting proteins. One of the telomere-associated protein complexes is shelterin (de Lange, 2005). In humans, the shelterin complex consists of six proteins, namely TRF1, TRF2, RAP1, TIN2, TPP1, and POT1. All six components of shelterin are constitutively expressed in human cells (Uhlen, et al., 2015). Its expression levels are high and non-tissue specific due to its essential ‘full-time’ roles in protecting the telomeres, hence genome integrity (Uhlen, et al., 2015).

Different flavours of shelterin-like arrangements exist in other organisms (Figure 1.2). In the budding yeast *Saccharomyces cerevisiae* (*S. cerevisiae*), only Rap1 (*Sc*Rap1) is found at the dsDNA telomeric region (Figure 1.2) (Conrad, et al., 1990; Azad & Tomar, 2016), whereas the fission yeast *Schizosaccharomyces pombe* (*S. pombe*) shelterin complex is considerably more complex and consists of six proteins, namely Taz1, *Sp*Rap1, Poz1, Tpz1, Pot1 and Ccq1 (Figure 1.2)

(Moser & Nakamura, 2009). In the *S. pombe* shelterin, similar interactions are established between components analogous to the human shelterin complex, with several notable structural differences, including (1) the presence of only one TRF-like protein, Taz1, which localises to duplex telomeric DNA, (2) unlike human RAP1 that only interacts with TRF2, *Sp*Rap1 also forms part of the overarching bridge that connects dsDNA-binding subunits to the ssDNA-binding subunits, (3) Ccq1 is a unique component of *S. pombe* shelterin and is involved in regulating telomerase activity (Tomita & Cooper, 2008), similar to TPP1 functions in the human counterpart.

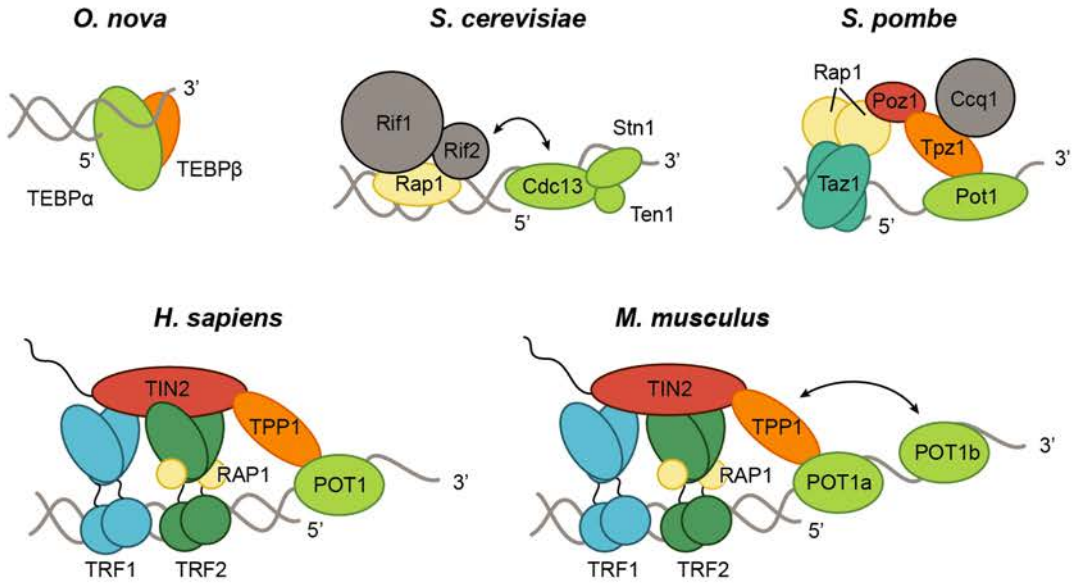

Figure 1.2. Shelterin-like arrangements are found in diverse organisms.

Shelterin-like telomeric proteins are found in a wide range of other eukaryotes. Orthologous proteins or proteins with similar domain architecture/function (e.g., TEBPα/TEBPβ, Tpz1/Pot1, and TPP1/POT1) are shown in the same colour.

Much of our knowledge of functions of shelterin proteins are derived from experiments with mice shelterin (de Lange, 2018). In rodents, although the same TTAGGG telomeric sequence are occupied by the shelterin complex as in humans (and other mammals), there are several notable differences (Figure 1.2).

Firstly, the telomere dynamics are significantly different. At birth, human telomeres have an average length of ~10–15 kbp of dsDNA that shortens at a rate of ~70 bp/year, whereas mice (of the common research C57BL6 strain) have telomeres of ~40–50 kbp that have an attrition rate 100 times faster than observed in humans (Canela, et al., 2007). Secondly, in mice there are two paralogues of POT1 (POT1a and POT1b, Figure 1.2), which share ~70% amino acid sequence identity with the single POT1 protein encoded in the human genome (Hockemeyer, et al., 2006). Deletion of POT1a, but not POT1b, results in embryonic lethality (Hockemeyer, et al., 2006). They also play non-redundant roles at the telomeres. Finally, the interactions between shelterin and its accessory factors are not well-conserved across species (see section 1.4.2.2). Therefore, cross-species generalisations should be made with caution.

1.2.1 Biological roles of the shelterin complex

The human shelterin complex plays critical roles in telomere protection and maintenance (Figure 1.3, Figure 1.4) (de Lange, 2018; Lim & Cech, 2021). TRF1 is primarily involved in (1) the semiconservative replication of telomeres through assisting the canonical DNA replication machinery (Figure 1.3) (Sfeir, et al., 2009; Zimmermann, et al., 2014), and (2) the regulation of telomere length maintenance by telomerase (Figure 1.4) (van Steensel & de Lange, 1997). During telomere replication in the S phase, TRF1 promotes efficient replication of the long TTAGGG repeat array through working in conjunction with helicases, including the BLM helicase, to remove secondary structures such as G4 quadruplexes (Sfeir, et al., 2009). In the absence of TRF1, telomeres activate ATR kinase signalling in S-phase, have fragile metaphase telomeres akin to common fragile sites, accompanied by sister telomere associations (Figure 1.3) (Sfeir, et al., 2009; Zimmermann, et al., 2014). In addition to its roles in telomere replication, TRF1 is also implicated in the regulation of telomerase-dependent telomere elongation (Figure 1.4). TRF1 overexpression in a telomerase-positive context *in vitro* leads to progressive telomere shortening (van Steensel & de Lange, 1997). Conversely, expression of a TRF1 mutant that removed endogenous TRF1 from telomeres

induced telomere elongation (van Steensel & de Lange, 1997). TRF1 also recruit other factors to regulate this process, including the poly(ADP-ribose) transferases tankyrase 1 and 2 (see section 1.4.2) (Smith, et al., 1998; Smith & de Lange, 2000; Cook, et al., 2002).

On the other hand, TRF2 plays critical roles in shaping the ends of telomeres (Griffith, et al., 1998; Doksani, et al., 2013; Benarroch-Popivker, et al., 2016; Lim & Cech, 2021) (see sections 1.3.1 and 1.3.2) and in telomere protection (see section 1.3.3) (Denchi & de Lange, 2007; Sfeir & de Lange, 2012; de Lange, 2018). TRF2 suppresses two DDR pathways, namely ATM signalling and NHEJ (Sfeir & de Lange, 2012). Several mechanisms have been proposed for how TRF2 achieves DDR repression, including (1) formation of T-loop structures that hide the ssDNA ends into the proximal telomeric dsDNA and hides the DSB-like structure from DSB sensors (see section 1.3.2) (Griffith, et al., 1999; Stansel, et al., 2001; Doksani, et al., 2013; Timashev & de Lange, 2020), (2) compaction of telomeric chromatin by TRF2 (and TRF1) excludes DDR factors from the telomere end (Bandaria, et al., 2016), and (3) interaction with several components of the ATM and NHEJ repair pathways (Song, et al., 2000; Karlseder, et al., 2004; Okamoto, et al., 2013). The TRF2 binding partner RAP1 is the only nonessential shelterin subunit, whereas removal of any other shelterin component is lethal on a cellular and organismal level (Karlseder, et al., 2003; Chiang, et al., 2004; Celli & de Lange, 2005). The role of RAP1 in telomere protection has been controversial. Some suggest that RAP1 does not contribute to repressing DDR (Kabir, et al., 2014), whist others provide evidence of its ability to suppress HDR (Sfeir, et al., 2010) and NHEJ (Bae & Baumann, 2007; Sarthy, et al., 2009), particularly in the absence of functional TRF2 (Benarroch-Popivker, et al., 2016). Besides protecting telomeres from illicit DDRs, TRF2 and RAP1 have also been shown to regulate telomere length maintenance. Overexpression of TRF2 in human primary fibroblast cells (IMR90) and a fibrosarcoma cell line (HTC75) result in telomere shortening (Smogorzewska, et al., 2000). RAP1 has also been shown to modulate telomere length maintenance (O'Connor, et al., 2004), although it does not affect telomerase processivity *in vitro* (Lim, et al., 2017).

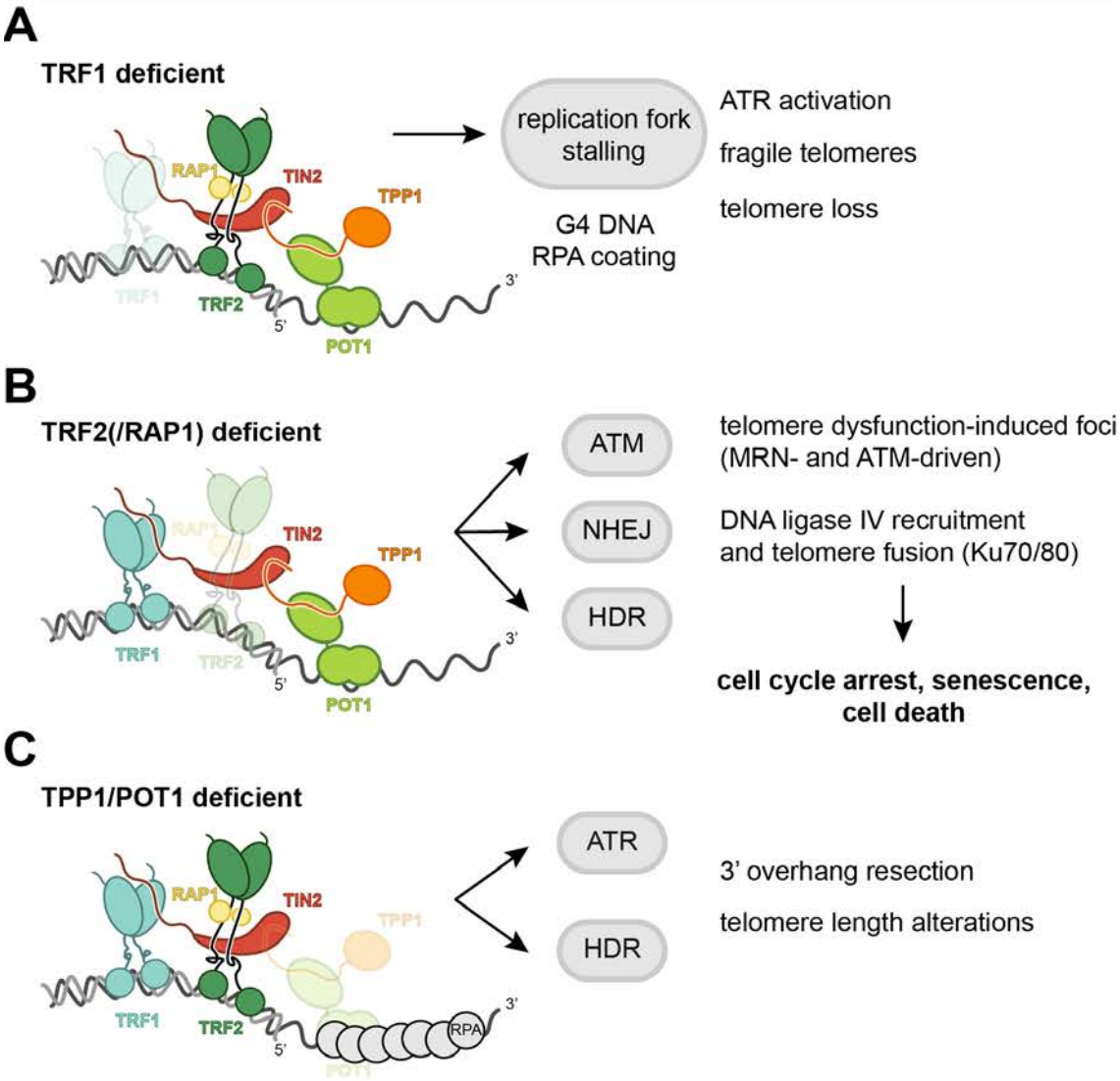

Figure 1.3. Shelterin components in telomere replication and protection.

(A) TRF1 promotes efficient replication of telomeres by preventing replication fork stalling, the removal of which leads to ATR activation, telomere fragility and telomere loss. **(B)** TRF2 represses ATM activation, non-homologous end joining (NHEJ), and homology-directed repair (HDR), the removal of which causes telomere dysfunction-induced foci (TIFs) and telomere fusion. **(C)** TPP1/POT1 represses ATR activation and HDR, preventing hyper-resection of the 3' overhang and telomere length alterations. Adapted from (Ruis & Boulton, 2021)

The TPP1/POT1 heterodimer play essential roles in both telomerase-dependent telomere length maintenance (Wang, et al., 2007; Zhong, et al., 2012; Sekne, et al., 2022) and telomere protection (Denchi & de Lange, 2007; Takai, et al., 2011; Kibe, et al., 2016). In telomerase-positive cells, TPP1/POT1 are positive telomere length regulators by recruiting telomerase to telomere 3' ends and enhancing telomerase processivity (Wang, et al., 2007; Nandakumar, et al., 2012; Zhong, et al., 2012; Lim, et al., 2017). On the telomere protection side, TPP1/POT1 prevent inappropriate activation of DDR throughout the cell cycle by repressing ATR signalling (see section 1.3.3). The TPP1/POT1 heterodimer is anchored at the telomeres by TIN2, which is essential for the heterodimer to perform roles in both telomere protection (Takai, et al., 2011; Frescas & de Lange, 2014) and telomerase-dependent telomere length maintenance (Abreu, et al., 2010; Pike, et al., 2019).

As mentioned above, TIN2 is the linchpin in the shelterin complex and is involved in stabilising multiple protein-protein and protein-DNA interactions that facilitate shelterin assembly and function at telomeres (Ye, et al., 2004a; Liu, et al., 2004b; Takai, et al., 2011; Zimmermann, et al., 2014). It also plays a role in telomere length regulation; TIN2 forms a ternary complex with tankyrase, a poly(ADP-ribose) polymerase (PARP) enzyme and a positive regulator of telomere length maintenance and TRF1 (see section 1.4.2.2) (Ye & de Lange, 2004c). Other telomeric functions of TIN2 can be difficult to ascertain given that disturbances of TIN2 likely destabilises many interactions within the shelterin complex. For instance, the frequent telomere fusions observed in TIN2 knockout cells can be attributed to an impact on TRF2 function (Takai, et al., 2011).

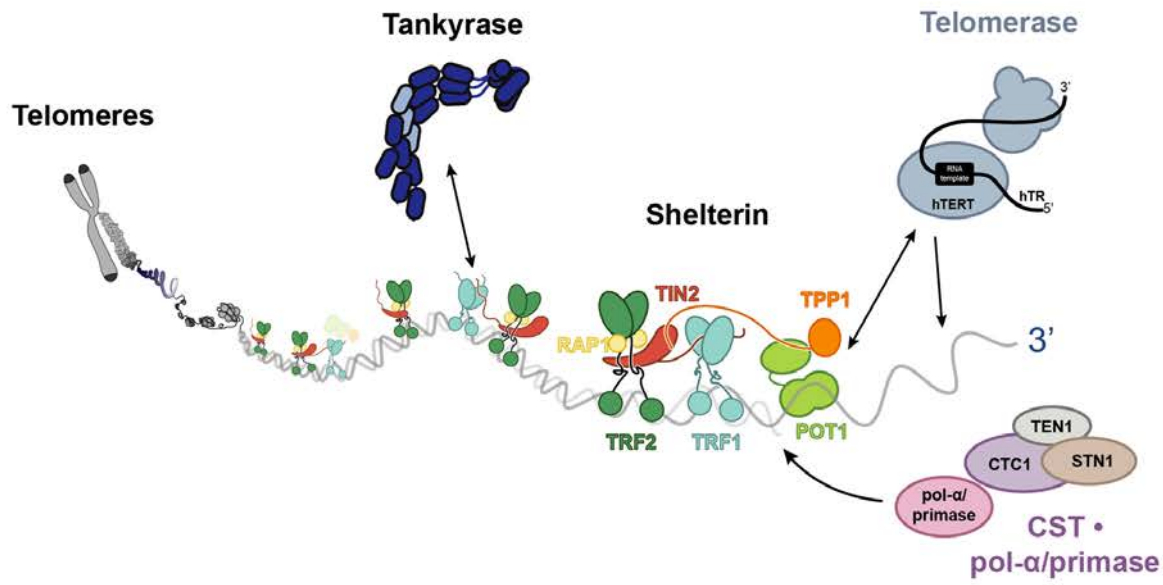

Figure 1.4. Shelterin components in telomere replication and length maintenance.

In the linear state, which presumably occurs during the S phase during telomere replication, the telomere is accessible to telomerase. Telomerase can bind to the 3′ ssDNA overhand and synthesis telomeric repeats de novo. The CST complex is recruited to the telomere ends and inhibits telomerase activity, thereby regulating telomere extension. CST also recruits the pol α-primase complex, which synthesises the complementary 5' strand (C-strand fill in).

1.2.2 Architecture of the human shelterin complex

1.2.2.1 TRF1 and TRF2/RAP1, the dsDNA-binding module

There are two dsDNA-binding subunits within the shelterin complex, namely the telomere-repeat binding factors (TRFs), TRF1 and TRF2. They share similar domain organisations (Figure 1.5). The TRFH domain at the N-terminus enables homodimerisation of TRF1 and TRF2 monomers (Fairall, et al., 2001), and is connected to the C-terminal Myb-type DNA-binding domain (DBD) by a long linker (~100 and ~150 amino acids in TRF1 and TRF2, respectively). One notable difference between the two TRF proteins is at the extreme N-terminus, where TRF1 has an acidic tail, as opposed to a basic one in TRF2.

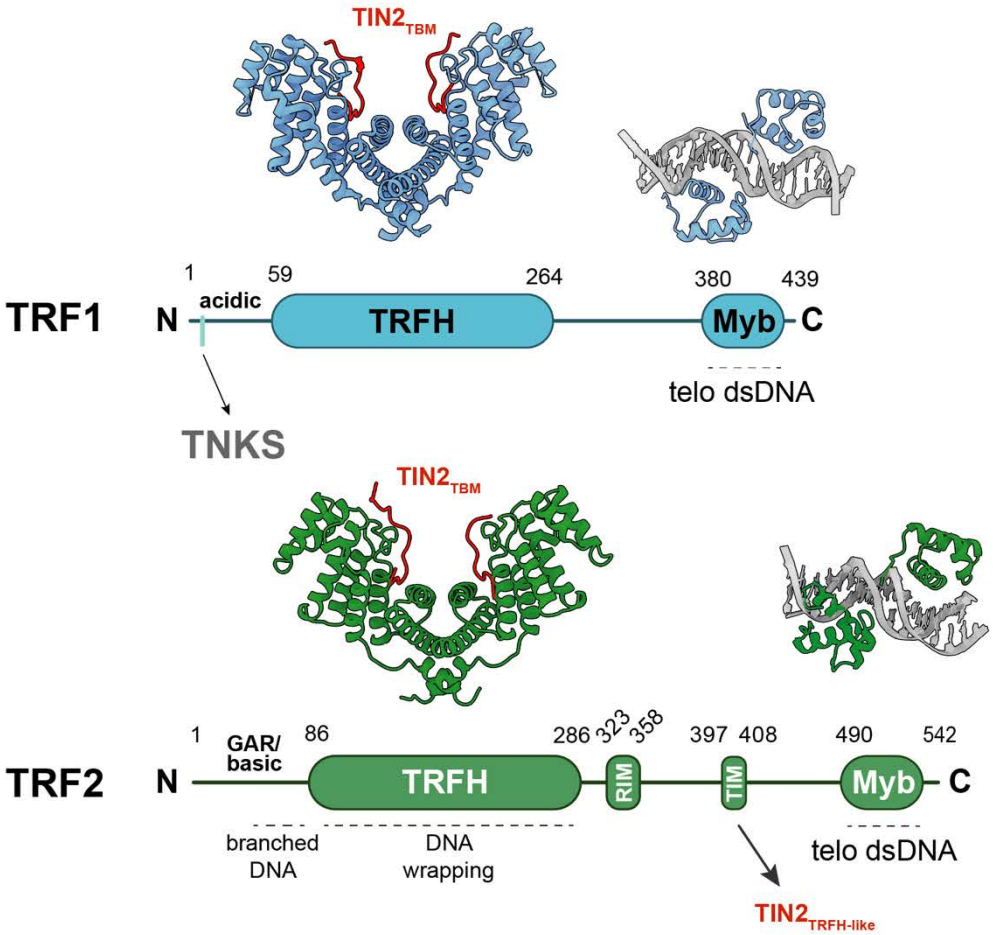

Figure 1.5. Domain organisations and structures of TRF1 and TRF2.

Domain organisation and crystal structures of TRF1 and TRF2 found at the telomeric dsDNA region. PDBs: TRF1, 3QBO and 1W0T; TRF2, 3BU8, 1W0U. (Chen, et al., 2008; Court, et al., 2005). RIM, RAP1-interacting motif; TIM, TIN2-interacting motif.

The TRFH domain in TRF1 and TRF2 is structurally conserved and form all-helix structures (Figure 1.6) (Broccoli, et al., 1997; Fairall, et al., 2001). Each TRFH domain consists of 10 α-helices and dimerisation is facilitated by helices 1, 2, and 10 from each monomer (Fairall, et al., 2001). The TRF1 and TRF2 TRFH domains recognise a conserved F/Y-x-L-x-P motif and serve as docking platforms that mediate key protein-protein interactions for shelterin assembly (TIN2 binding) and recruitment of factors that cooperate with shelterin to perform an array of functions at telomeres (Figure 1.6, see section 1.3) (Chen, et al., 2008). Subtle differences in this motif regulate the binding specificity for TRF1 and TRF2. $TRF1_{TRFH}$ prefers motifs with a positively charged tail and a hydrophobic residue (Phe) at beginning of this motif, whilst $TRF2_{TRFH}$ prefers a polar residue (Tyr) at the same position (Chen, et al., 2008).

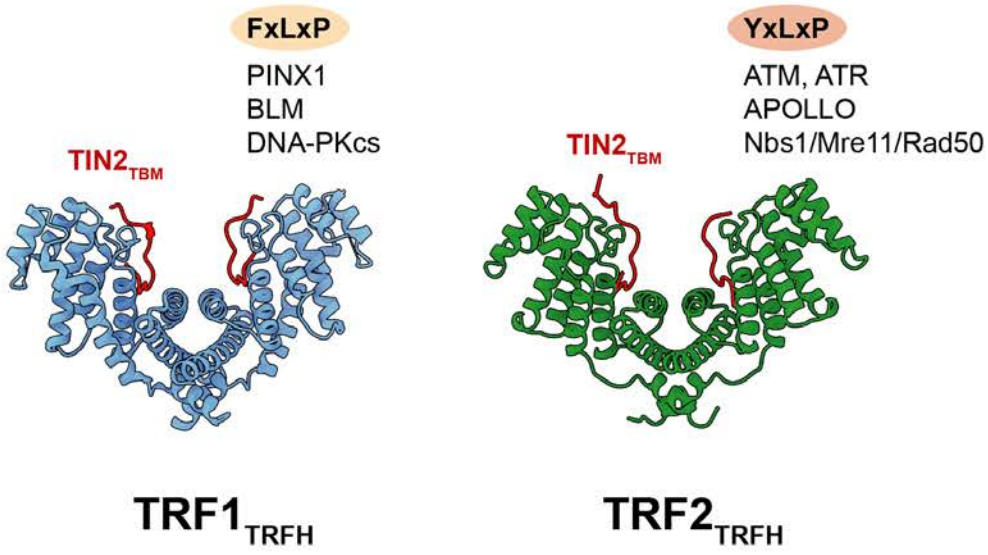

Figure 1.6. TRFH domains of TRF1 and TRF2 contain a peptide docking platform.

The docking site in the TRFH domain of TRF1 and TRF2 interact with proteins containing the TRFH-binding motif (TBM) FxLxP (e.g., PINX1, the BLM helices, and DNA-PKcs) and YxLxP (e.g., ATM and ATR kinases, APOLLO, and Nbs1 in the MRN complex), respectively (Palm & de Lange, 2008). Both TRF1 and TRF2 TRFH can bind to the TBM in TIN2, albeit the latter with lower affinity (Chen, et al., 2008).

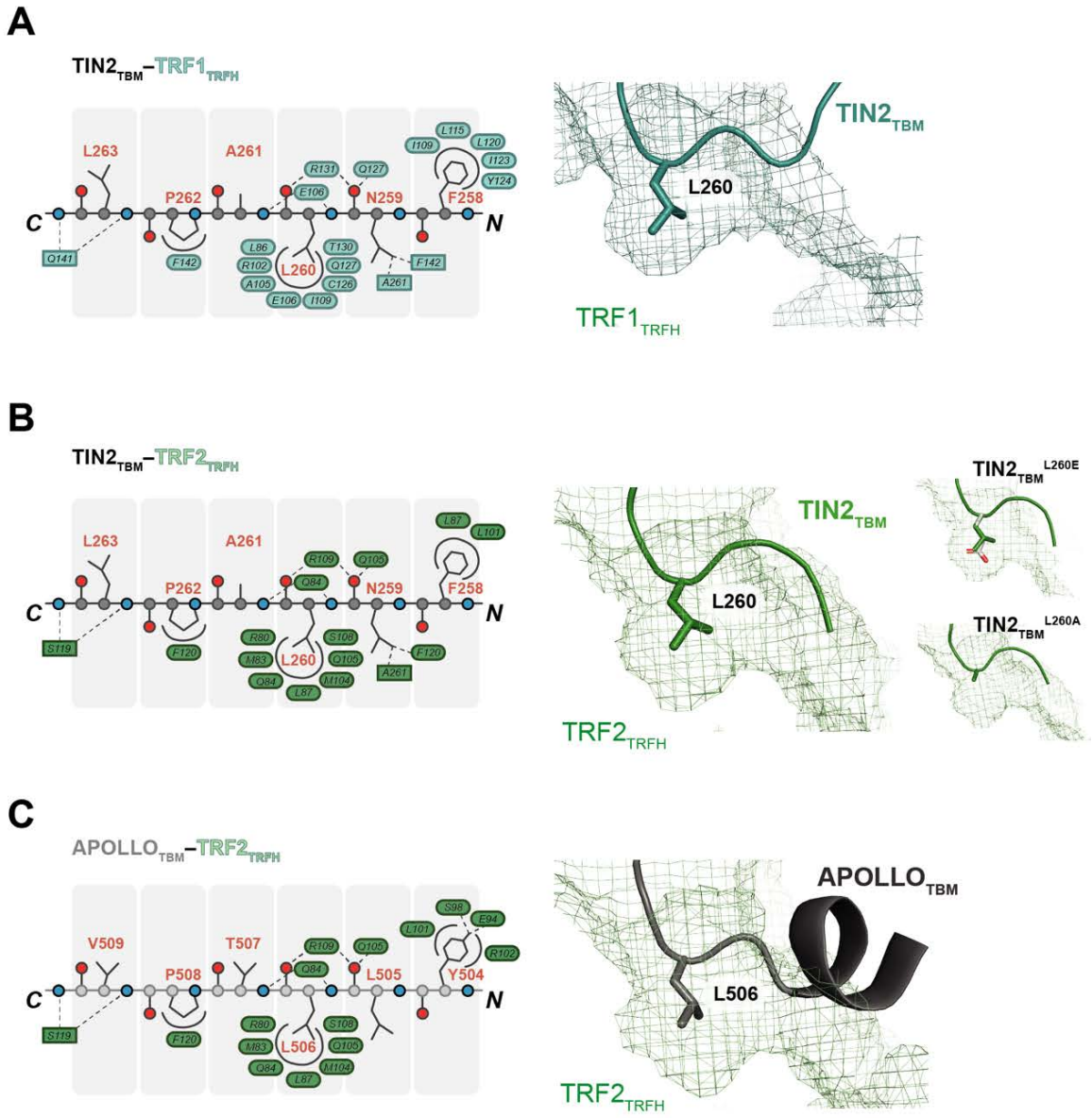

Figure 1.7. Interaction between TRFH domains from TRF1/TRF2 and TIN2$_{TBM}$.

Interactions between **(A)** TIN2$_{TBM}$ and TRF1$_{TRFH}$ (PDB: 3BQO), **(B)** TIN2$_{TBM}$ and TRF2$_{TRFH}$ (PDB: 3BU8), and **(C)** APOLLO$_{TBM}$ and TRF2$_{TRFH}$ (PDB: 3BUA). (Left) Cartoon representation of the interactions, with the backbone atoms of TIN2$_{TBM}$ and APOLLO$_{TBM}$ in circles (C in grey, O in red, N in blue). Residues of TRF1$_{TRFH}$ and TRF2$_{TRFH}$ are displayed as rods (side-chain interaction) and rectangle (backbone interaction) boxes and shaded blue and green, respectively. Hydrophilic and hydrophobic interactions are shown as straight dotted and curved solid lines, respectively. Adapted from Chen, et al., 2008. (Right) Crystal structures of the domain-peptide interactions (Chen, et al., 2008). The mesh represents the surface of TRF1 and TRF2 TRFH domains, and the cartoons show the TIN2 and APOLLO peptides, highlighting the conserved leucine residue (L260 in TIN2, L506 in APOLLO) and the TRFH domain. L260A and L260E mutations in TIN2$_{TBM}$ has been shown to interrupt this interaction. (Chen, et al., 2008).

The TRF1/2 Myb-type domain consists of a N-terminal arm and a 3-helix bundle, two of which form a helix-turn-helix (HTH) motif (Figure 1.8) (Court, et al., 2005). Crystal structures of TRF1 and TRF2 Myb domains show that they interact with telomeric DNA in a similar manner (Figure 1.8). A minimal two and a half duplex TTAGGG sequence is sufficient for binding of two Myb domains from one homodimer of TRF proteins (Court, et al., 2005). Each Myb domain contact a 7 bp TAGGGTT half site and makes several direct contacts with the ribose phosphate backbone and the bases via amino acid side chains, and indirect water-mediated contacts (Court, et al., 2005). Helix 2 and the N-terminal arm make direct contacts with the DNA backbone on both sides of the major groove, which docks the Myb domain onto the DNA. Helix 3 is the DNA-recognition helix. A lysine and arginine residue from each Myb domain (K421 and R425 in TRF1, and K530 and R534 in TRF2) form hydrogen bonds with two adjacent guanine (G) nucleotides in the major groove (Court, et al., 2005). Although a single Myb-type domain can bind telomeric dsDNA, high-affinity binding (in the 10 nM range) is mediated by the presence of two Myb domains in the context of the homodimer (Bianchi, et al., 1997; Bianchi, et al., 1999). As the Myb domains are tethered to the TRFH domains via long linkers with no apparent structure, the TRF proteins can also bind to half sites separated by up to 30 base pairs (Bianchi, et al., 1999). The basic Gly/Arg-rich (GAR) domain at the N-terminal of TRF2 has been shown to interact non-specifically with DNA through electrostatic interactions.

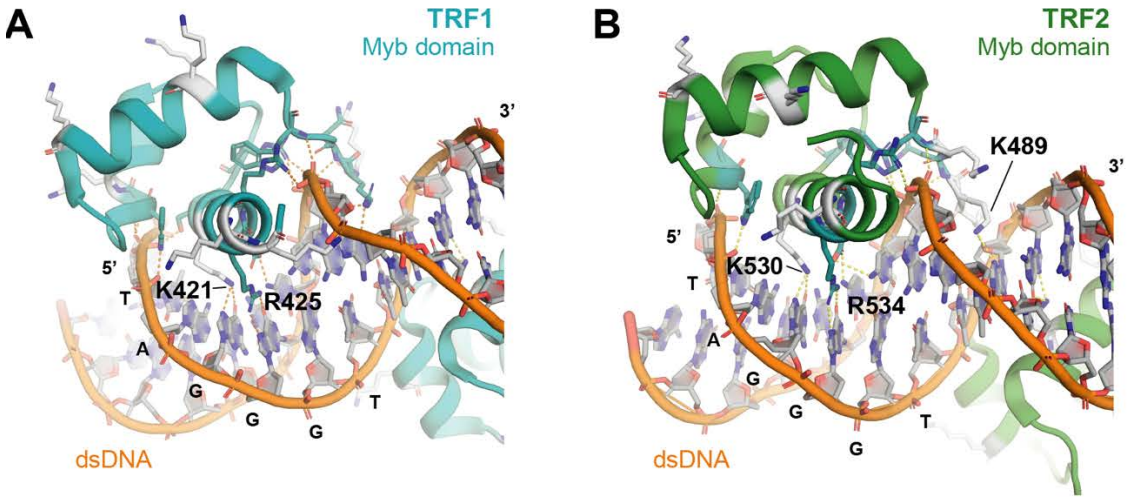

Figure 1.8. Key residues in the TRF1 and TRF2 Myb-dsDNA interactions.

Crystal structures of the Myb domains from **(A)** TRF1 (PDB: 1W0T) and **(B)** TRF2 (PDB: 1W0U) in complex with telomeric dsDNA. The lysine residue K421 from $TRF1_{Myb}$ is involved in recognising G3 in the major groove, whereas K530 and K489 in $TRF2_{Myb}$ are involved in binding to G3 and A8 in the major and minor grooves of DNA, respectively (Court, et al., 2005).

Finally, the shelterin subunit, RAP1, is recruited to the shelterin complex through its interaction with TRF2. RAP1 has a BRCA C-terminal (BRCT) domain, a single Myb domain, and a C-terminal domain (RCT) (Figure 1.9). The canonical BRCT domain, as in BRCA1, comprises of a central β-sheet stacked between two layers of α-helices (Gaiser, et al., 2004). The $RAP1_{BRCT}$ has only one layer of α-helices that cradles the β-sheet (Figure 1.10). BRCT domains are commonly implicated in phosphorylation-dependent interactions, and are also found to bind DNA and poly(ADP-ribose) (PAR) chains (Leung & Glover, 2011). Evidence suggests that the RAP1 BRCT and Myb domain are involved in regulating the length and heterogeneity of telomeres by recruiting other protein factor(s) to telomeres (Li & de Lange, 2003). However, other interacting partner(s) of RAP1 is still unknown. The Myb-type domain in RAP1 also possesses the canonical three-helix bundle (Figure 1.9), although its surface is less positively charged compared with the TRF1/2 Myb domains and does not bind directly to the telomeric DNA (Hanaoka, et al., 2001). Instead, it is recruited to the telomeres via binding to TRF2 (Li, et al., 2000). The RCT domain binds to a region in the TRF2

linker ($TRF2_{275–316}$) with a similar affinity as to the full-length TRF2 (K_D = 16.5 vs. 23.9 nM, respectively) (Chen, et al., 2011).

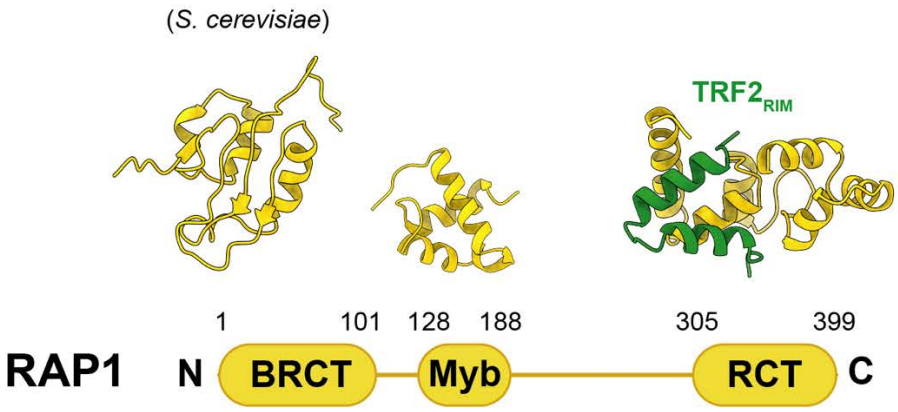

Figure 1.9. Domain organisation and structures of RAP1.

Domain organisation and crystal structures of RAP1, which is recruited by TRF2 to the telomeric dsDNA region. PDBs: 1FEX and 3K6G (Hanaoka, et al., 2001; Chen, et al., 2011). RIM, RAP1-interacting motif.

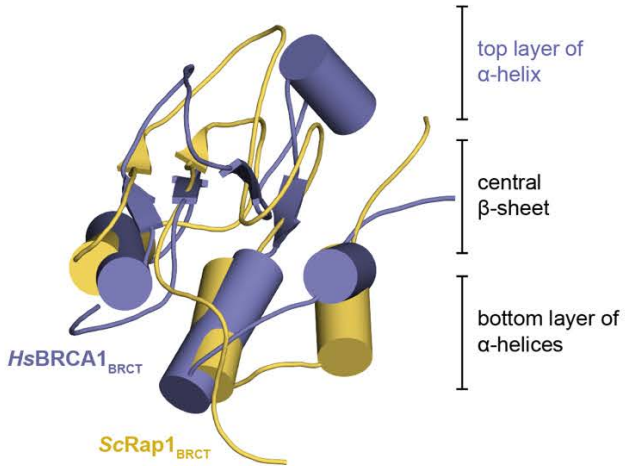

Figure 1.10. Solution structures of the BRCT domain from *Hs*BRCA1 and *Sc*Rap1.

Solution structures of the C-terminal BRCT domain from human BRCA1 (purple, PDB: 1OQA) (Gaiser, et al., 2004) and the BRCT domain from *S. cerevisiae* RAP1 (yellow, PDB: 2L42) (Zhang, et al., 2011).

1.2.2.2 ssDNA-binding activity and OB-folds in shelterin

The oligonucleotide/oligosaccharide-binding folds (OB-folds) are commonly found in ssDNA and RNA-binding proteins (Murzin, 1993). In the human shelterin complex, a total of four OB-folds are present in the TPP1/POT1 heterodimer. TPP1 and POT1 are human homologs of the ciliate *O. nova* telomere end-binding protein (TEBP) α and β subunits, respectively (Fang, et al., 1993; Lei, et al., 2004; Wang, et al., 2007). In contrast to *O. nova* where the ssDNA streams through a channel formed by both TEBP α and β subunits (Horvath, et al., 1998), only POT1 makes contacts with telomeric ssDNA in the TPP1/POT1 heterodimer (Wang, et al., 2007).

The domain organisation of TPP1 is similar to the *O. nova* TERBβ subunit (Wang, et al., 2007). The extreme N-terminus of TPP1 (residues 1-87) has no known structure and function, and are not conserved among TPP1 homologs from other organisms (Ye, et al., 2004b; Liu, et al., 2004a). The remainder of TPP1 comprises of an OB-fold towards the N-terminus, a central region forming interactions with the C-terminus of POT1, and a C-terminal region containing the TIN2-interacting motif (Figure 1.11) (Wang, et al., 2007; Chen, et al., 2017; Hu, et al., 2017). Although TPP1 also possess a single OB-fold, unlike its ciliate homolog TERBβ, it does not interact with telomeric ssDNA. Instead, the N-terminus of the OB-fold (NOB) and a patch rich in glutamate and leucine (TEL patch) on the OB-fold has been shown to be key in the mediating telomerase recruitment and processivity (Figure 1.11, see section 1.4.1) (Wang, et al., 2007; Zhong, et al., 2012; Nandakumar, et al., 2012; Grill, et al., 2018).

At the N-terminus, POT1 comprises of two adjoining OB-folds that recognises and bind telomeric ssDNA (OB1 and OB2), followed by a third OB-fold in the C-terminus (OB3) interrupted by a Holliday junction resolvase-like (HJRL) domain (Figure 1.11). OB1 and OB2 each consists of a five-stranded β-barrel, sealed by the N-terminal region and the loop connecting β3 and β4 at the top and bottom of the barrel, respectively (Lei, et al., 2004). The ssDNA binds in a continuous cleft formed at one side of the β-barrel with two flanking loops (Figure 1.12), with OB1 interacting with more of the telomeric ssDNA (5'-

TTAGGG) compared with OB2 (TTAG-3'). Binding of the ssDNA induces a kink in the DNA backbone by ~90^{o} at the phosphodiester group of thymidine at position 7 (Figure 1.12) (Lei, et al., 2004). Stacking interactions between the ssDNA and POT1 aromatic amino acid side chains contribute to the DNA-binding affinity of POT1, whereas several hydrogen bonding involving Watson-Crick donor-acceptor groups and surface residues of the DNA-binding groove defines the sequence specificity of the POT1 OB-folds (Lei, et al., 2004). POT1 has nanomolar affinity for a ssDNA with the minimal telomeric sequence 5'-TTAGGGTTAG-3' either at the 3' position (K_D = 8.3 nM) or at an internal position (K_D = 89 nM) (Lei, et al., 2004; Loayza, et al., 2004; Wang, et al., 2007). In addition, binding of TPP1 to POT1 enhances its DNA binding affinity by 10-fold (Wang, et al., 2007). OB1 and OB2 are separated by an eight amino acid short linker, which enables POT1 to bind ssDNA where the two half sites are separated by spacers of 6 bases albeit with drop in binding affinity (Smith, et al., 2022). The OB3 at the C-terminus does not bind ssDNA but contains a 150-residue HJRL domain in the loop connecting β1 and β2 that forms interactions with TPP1 (Figure 1.11) (Chen, et al., 2017; Rice, et al., 2017).

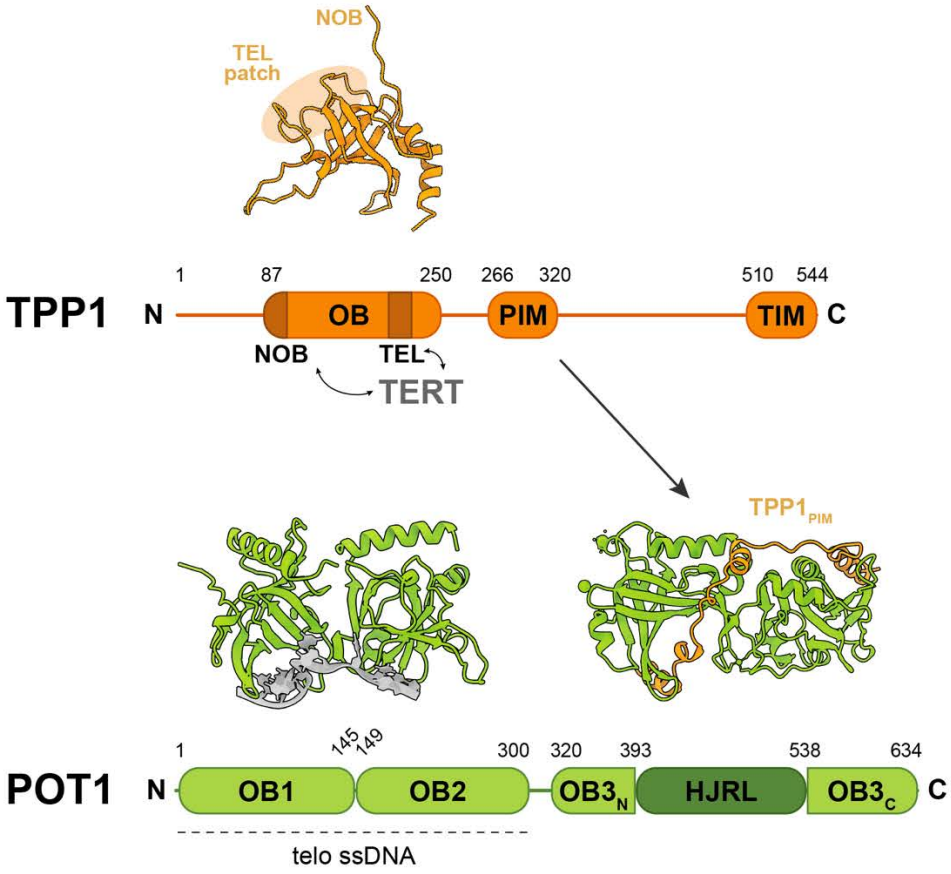

Figure 1.11. Domain organisation and structures of TPP1 and POT1.

Domain organisation and crystal structures of TPP1 (PDB: 2I46) (Wang, et al., 2007) and POT1 (PDB: 3KJP and 5UN7) (Nandakumar, et al., 2010; Rice, et al., 2017)

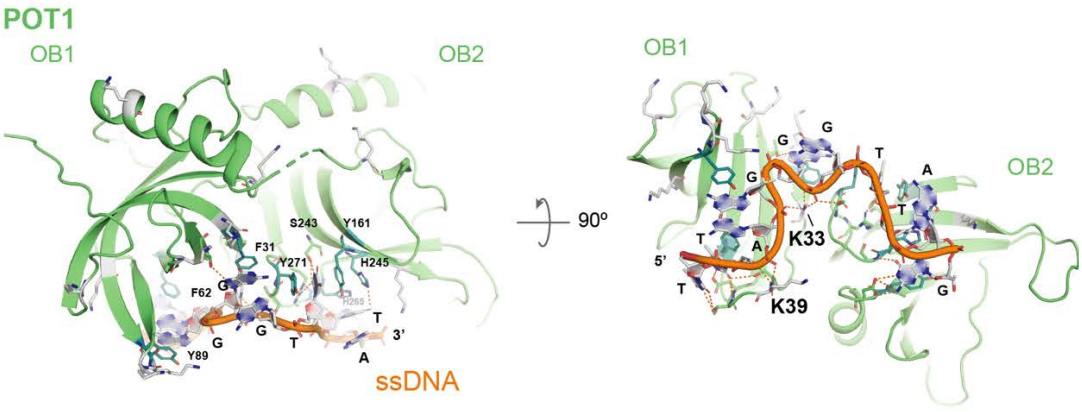

Figure 1.12. Key residues in the POT1 $OB_{1\text{-}2}$-ssDNA interactions.

Crystal structure of the POT1 OB1 and OB2 domains binding to telomeric ssDNA (PDB: 3KJP) (Nandakumar, et al., 2010). Residues K33 and K39 form hydrogen bonds with the DNA backbone, and several aromatic residues pack against the DNA bases.

1.2.2.3 TIN2, the linchpin of the shelterin complex

TIN2 is an essential scaffolding protein that arches between the dsDNA-binding TRF1 and TRF2 to the ssDNA-binding POT1, via TPP1. The only domain found in TIN2 is the N-terminal TRFH-like domain, which is structurally similar to the TRFH domain in TRF1/2 (Figure 1.13) (Hu, et al., 2017). Towards the C-terminus, TIN2 is largely disordered and contains a cluster of mutations identified in patients with Dyskeratosis Congenita (DC cluster) (Savage & Bertuch, 2010). Three isoforms of TIN2 have been reported (Kim, et al., 1999; Kaminker, et al., 2009; Pike, et al., 2019). The shortest isoform, TIN2S, lacks 97 amino acids at the extreme C-terminus, but retains all the regions essential for interactions with other shelterin subunits and the DC cluster (Figure 1.13).

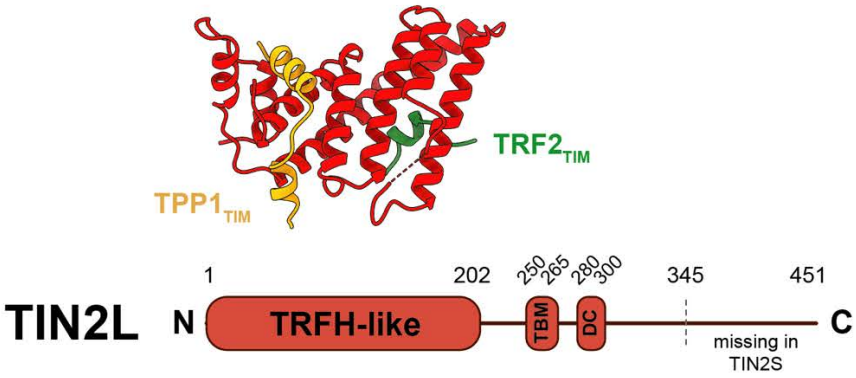

Figure 1.13. Domain organisation and structure of the bridging subunit TIN2.

The domain organisation of the long isoform of TIN2 (TIN2L). The short isoform, TIN2S, lacks 97 amino acids at the C-terminus of TIN2L. The only folded region in TIN2 is the TRFH-like domain, which interacts with TPP1 and TRF2 (PDB: 5XYF) (Hu, et al., 2017). The C-terminal portion of TIN2 is disordered and contains the TRFH-binding motif (TBM) and the DC mutation cluster.

TIN2 binds to TRF1 and TRF2 via two distinct interaction interfaces. The TRF-binding motif (TBM) at the TIN2 C-terminus can bind to the TRFH domains in TRF1 and TRF2 (Figure 1.14, Figure 1.7) (K_D = 0.3 and 6.5 μM, respectively), whilst the TRFH-like domain at the TIN2 N-terminus can bind to a short peptide in the linker region of TRF2 (Chen, et al., 2008; Hu, et al., 2017). TIN2 can bind to both TRF1 and TRF2 in the same shelterin complex, and this has been shown

to contribute to stabilisation of TRF2 on telomeres (Liu, et al., 2004b; Ye, et al., 2004a). In addition to the TRF proteins, TIN2 binds to a peptide motif at the extreme C-terminus of TPP1 (TIN2-interacting motif, $TPP1_{TIM}$) via the TRFH-like domain (Figure 1.13) (Hu, et al., 2017). Binding of TPP1 to TIN2 strengthens the interaction between TIN2 and TRF2, which has been shown to be key in allowing TIN2 to bind TRF1 and TRF2 simultaneously (Hu, et al., 2017; Janovič, et al., 2019).

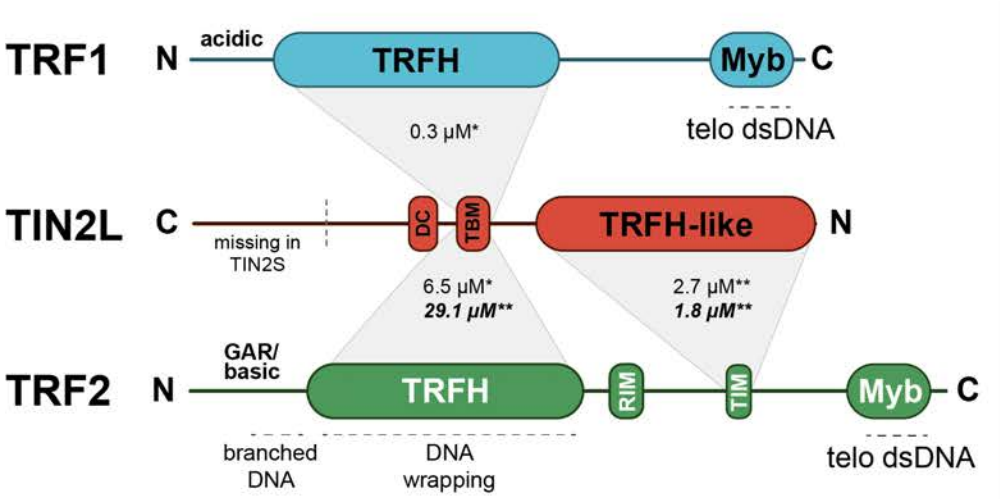

Figure 1.14. The TRF1-TIN2-TRF2 bridge.

Domain organisation of TRF1, TRF2 and TIN2L subunits. In the full shelterin complex, one homodimer of TRF1 occupies the $TIN2_{TBM}$ binding site via the $TRF1_{TRFH}$ domain, and one homodimer TRF2 occupies the $TIN2_{TRFH\text{-}like}$ domain via its linker. In the absence of TRF1, TRF2 has two potential binding sites in TIN2. * indicates K_D values determined by isothermal titration calorimetry (Chen, et al., 2008), and ** by fluorescence polarisation (Hu, et al., 2017). Values in bold represent the K_d for full length TRF2 and $TIN2_{TRFH}$ or $TIN2_{TBM}$.

1.3 Shelterin in telomere protection

1.3.1 Generating the 3' ssDNA overhang

After DNA replication, a series of steps take place at the telomeres to establish a 3' overhang. Formation of this structure at the 3' telomere end is essential for T-loop formation, which is a key mechanism of telomere protection (see below) (Griffith, et al., 1999). Apollo is an Artemis-Related Nuclease, also known as DNA cross-link repair 1B (DCLRE1B) protein, localises to telomeres via interactions with TRF2 and selectively initiates resection at the 3' overhang of leading-end telomeres (Chai, et al., 2006). RNAi knockdown of Apollo induces cellular senescence and indicators of DDR activation enriched in S-phase, including telomere-dysfunction-induced foci (TIFs), ATM activation and fusion of leading strand telomeres (Overbeek & de Lange, 2006). A splicing variant of the *DCLRE1B* gene producing a N-terminally truncated product lacking the TRF2-interacting motif has been reported in a patient with Hoyeraal-Hreidarsson syndrome (HHS) (Touzot, et al., 2010), a genetic disorder characterised by multisystem involvement after telomere dysfunction and premature cellular senescence. The resulting Apollo protein retains its DNA inter-strand cross-link repair function (Touzot, et al., 2010). However, patient-derived fibroblasts share a similar molecular and cellular phenotype as Apollo RNAi knockdown cells (Touzot, et al., 2010). This supports a role for Apollo in telomere maintenance through end processing during telomere replication. Another exonuclease, EXO1, also act on telomere ends in mice and humans, and generates transient extensive 3' overhangs on lagging-end telomeres during S phase (Overbeek & de Lange, 2006; Wu, et al., 2010). The length of the overhang then decreases during S/G2, likely from C-strand fill-in that requires DNA polymerase α (Polα) (Dai, et al., 2010; Lim & Cech, 2021).

Although resection of the 5' end of newly synthesised telomeres is necessary to generate the 3' overhang and re-establish the protective telomeric structure, the resection process must be tightly regulated. Excessive removal of DNA would accelerate telomere attrition and lead to premature cellular senescence. Hyper-resection of telomere ends is circumvented by TRF2 and

POT1. Inhibition of TRF2 or POT1 reduces the overall length of the 3' overhang by 30 – 50% (Van Steensel, et al., 1998; Hockemeyer, et al., 2005). In addition, nearly 80% of human chromosomes have 5' ends with sequence ATC-5′ (Hockemeyer, et al., 2005). Loss of POT1 introduces sequence heterogeneity to the 5' telomere termini, indicating a dysregulation of 5' resection (Hockemeyer, et al., 2005; Takai, et al., 2010).

1.3.2 T-loop formation

T-loop formation is dependent on the shelterin subunit TRF2 and not TRF1 (Griffith, et al., 1998; Stansel, et al., 2001; Timashev & de Lange, 2020). The TRFH domain in TRF2 have been implicated in the formation of T-loops (Figure 1.15) (Benarroch-Popivker, et al., 2016). The TRFH domain in TRF1 and TRF2 are known to facilitate protein-protein interactions within the shelterin complex and its accessory factors (Chen, et al., 2008). Recent evidence suggests that the TRF2 can also interact with DNA in a sequence-independent manner. Firstly, the extreme N-terminus of TRF2 is unstructured and basic in nature, and is rich in glycine (G), alanine (A), arginine (R) residues (GAR-rich) (Figure 1.5). This basic/GAR region has been shown to bind branched DNA, including three- and four-way junctions, which are structures that may be present at the base of T-loops (Schmutz, et al., 2017). Engaging the base of the T-loop potentially protects the structure from branch migration and prevent the formation of double Holliday junctions (Wang, et al., 2004; Saint-Léger, et al., 2014). Cleavage of the dHJs by dHJ resolvases can lead to telomere loss (Saint-Léger, et al., 2014). Secondly, the TRFH domain can bind to nonspecific DNA with low affinity and modify DNA topology (Benarroch-Popivker, et al., 2016). Each TRF2 homodimer can wrap ~90 bp of DNA (Benarroch-Popivker, et al., 2016). A series of surface lysine and arginine residues in the $TRF2_{TRFH}$ contribute to the DNA wrapping ability of TRF2, and mutations of these residues to alanines (in a TRF2 mutant known as 'Top-less') negatively impact the wrapping efficiency of TRF2 *in vitro* and its ability to generate T-loops *in vivo* (Benarroch-Popivker, et al., 2016; Timashev & de Lange, 2020). Expression of the wrapping-deficient TRF2 mutant in human HT1080

fibrosarcoma cells decreased the abundance of T-loops, triggered ATM activation, whilst maintaining the repression of NHEJ at telomeres (Benarroch-Popivker, et al., 2016). Given the low affinity of these sequence non-specific interactions, it is likely that the initial localisation of TRF2 at telomeres involves the high affinity and telomere-specific Myb domains. Subsequently, the structure-specific interactions localise TRF2, and its protein partners, to perform the plethora of roles in protecting and maintaining the telomeres.

1.3.3 Inhibiting DNA damage repair pathways

As mentioned, the ends of linear chromosomes resemble DSBs that has the potential to activate at least seven DDR pathways and checkpoints, including ATM and ATR signalling pathways, and HDR and NHEJ repair (de Lange, 2018). Several shelterin subunits have been implicated in averting DDR pathways at the telomeres.

Activation of ATM-dependent signalling at the telomeres has been shown to induce cell cycle arrest, senescence, and apoptosis (Karlseder, et al., 1999; Takai, et al., 2003). The ATM kinase is activated upon the interaction of the MRN complex, a DSB sensor, with DNA ends (Deng, et al., 2009). TRF2, but not TRF1, RAP1, TPP1, or POT1, is required to repress MRN-dependent activation of ATM at telomeres (Takai, et al., 2003) (Celli & de Lange, 2005; Denchi & de Lange, 2007; Sfeir, et al., 2009; Sfeir, et al., 2010). TIN2 deletion also led to ATM kinase activation, although this is partially due to destabilisation and loss of TRF2 from telomeres (Takai, et al., 2011). Several models have been proposed for inhibition of ATM activation at the telomeres. One model suggests that TRF2 may inhibit ATM activation by directly interacting with the ATM protein (Karlseder, et al., 2004). Finally, TRF2 (and TRF1) have been shown to compact DNA, which may restrict the access of DDR factors to the telomeres (Bandaria, et al., 2016). However, ATM is not found at detectable levels at telomeres, and does not explain how ATM activation can still occur at DSB sites further away from the extreme telomere termini (Doksani & de Lange, 2016). The current model is that

TRF2-dependent T-loop formation sequesters the telomere terminus in a structure that is no longer be recognised by the MRN complex (see section 1.3.2).

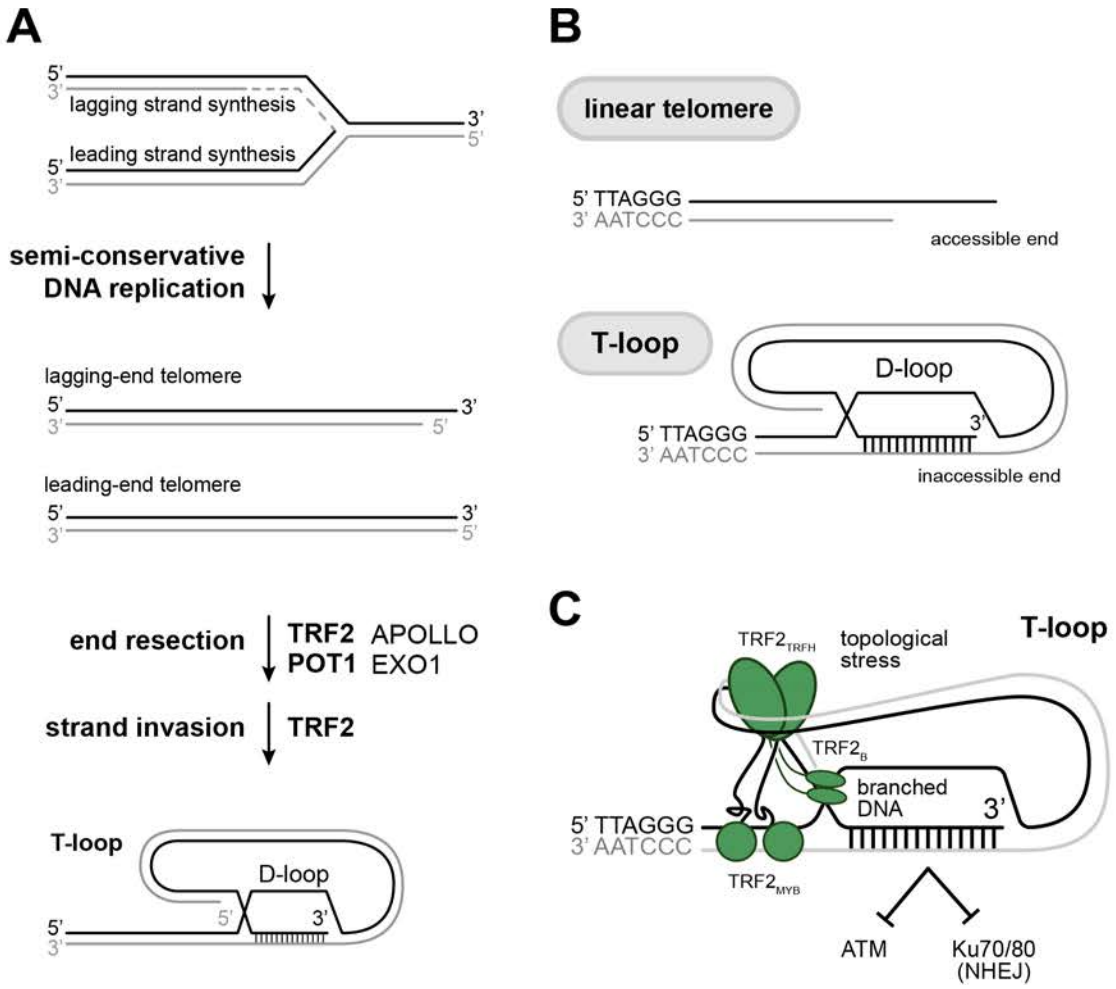

Figure 1.15. Shaping the DNA structures at the telomeric termini.

(A) Telomere replication and generation of the 3' overhang and T-loop structures. The leading-end telomere is blunt and requires additional processing to restore the 3' overhang. This is achieved by regulated resection of the 5' end by nucleases including APOLLO and EXO1, which are recruited by TRF2 and POT1, respectively. Adapted from (Palm & de Lange, 2008). **(B)** In the linear state, telomeres resemble DSBs with ends that are accessible for DSB sensors, whereas telomere ends in T-loop conformation are inaccessible. **(C)** TRF2 is involved in T-loop formation and stabilisation. The Myb domains ($TRF2_{MYB}$) bind to telomeric DNA with high affinity and specificity, recruiting TRF2 to telomeres dsDNA. The TRFH domain ($TRF2_{TRFH}$) wraps 90 bp of dsDNA with low affinity, generating topological stress that is suggested to promote the invasion of 3' ssDNA into the duplex telomeric region. The basic region of TRF2 ($TRF2_B$) can recognise branched DNA structures, including the junctions at the base of T-loops. This may stabilise T-loops by preventing branch migration and double Holliday junction (dHJ) formation, which can be cleaved by dHJ resolvases and lead to telomere loss. Adapted from (Ruis & Boulton, 2021)

Another potent threat to telomere integrity is the activation of ATR kinase, which be triggered at telomeres via two different mechanisms (Blackford & Jackson, 2017). One pathway requires the loading of the 9-1-1 clamp onto the adjacent 5' ds-ss junction, whereas the other does not involve 9-1-1 and does require the 5' ds-ss junction (Blackford & Jackson, 2017). Both pathways are initiated by binding of replication protein A (RPA) to ssDNA. The shelterin protein POT1 is involved in repressing ATR activation. A competition model has been proposed, whereby POT1 decorates the ssDNA at telomeres and excludes RPA from the ssDNA overhang (Hockemeyer, et al., 2005; Takai, et al., 2011). However, the DNA binding affinities and abundance of POT1 and RPA proteins suggest that POT1 alone is unlikely to block access of RPA (Wold, 1997; Lei, et al., 2004; Loayza, et al., 2004; Nandakumar, et al., 2010). TIN2 plays a critical role in mediating POT1-dependent ATR inhibition. Binding of TPP1/POT1 to TIN2 stabilises POT1 at telomeres through anchoring of TPP1/POT1 to the dsDNA binding shelterin proteins TRF1 and TRF2 (Takai, et al., 2011; Frescas & de Lange, 2014; Zimmermann, et al., 2014).

1.4 Shelterin in telomere length maintenance

The shelterin complex recruits and regulates several protein factors that play key roles in regulating telomere length maintenance. Here I discuss two shelterin interaction partners, namely telomerase and tankyrase.

1.4.1 Telomerase, a solution to the end-replication problem

Telomerase is a specialised ribonucleoprotein (RNP) complex that can compensate for the sequence loss that occurs at telomeres during genome duplication. The minimally active telomerase is composed of a protein catalytic subunit telomerase reverse transcriptase (TERT) and its RNA partner telomerase RNA (TR) (Weinrich, et al., 1997; Beattie, et al., 2001). The *TERT* gene encoding the human catalytic telomerase component (hTERT) is silenced in most healthy somatic cells. Spatiotemporal expression the hTERT protein is tightly controlled, and regulation occurs at many levels including transcription, mRNA splicing, folding and modifications, localisation, and the accessibility of the 3' ssDNA overhang (Wright, et al., 1996; Ulaner, et al., 2001; Venteicher, et al., 2009; Kim, et al., 2016; Lei, et al., 2005).

1.4.1.1 Domain architecture and structural studies of human telomerase

Human telomerase consists of two functional lobes bridged by one copy of hTR (Figure 1.16) (Nguyen, et al., 2018; Ghanim, et al., 2021). In the catalytic core, hTERT binds to the conserved regions 4 and 5 (CR4/5) and the pseudoknot/template (PK/t) domain of hTR, which contains the 3'-UCCCAAUC-5' template to direct telomeric repeat synthesis (Figure 1.16) (Lai, et al., 2001; Robart & Collins, 2010). The hTERT consists of four domains: the N-terminal (TEN) domain, the RNA-binding domain (TRBD), the reverse transcriptase (RT) domain, and the C-terminal extension (CTE) domain. The other lobe contains an H/ACA RNP complex formed by an array of telomerase-associated proteins key

for telomerase biogenesis and regulation (Figure 1.16) (Egan & Collins, 2012). Mutations in the H/ACA RNP are implicated in premature aging disorders known as telomeropathies (see section 1.5) (Opresko & Shay, 2017). A recent high-resolution cryo-EM structure of human telomerase also identified a histone H2A-H2B dimer that bound to C4/5 in hTR, suggesting a role for histones in stabilisation and function of hTR and possibly telomerase recruitment (Ghanim, et al., 2021).

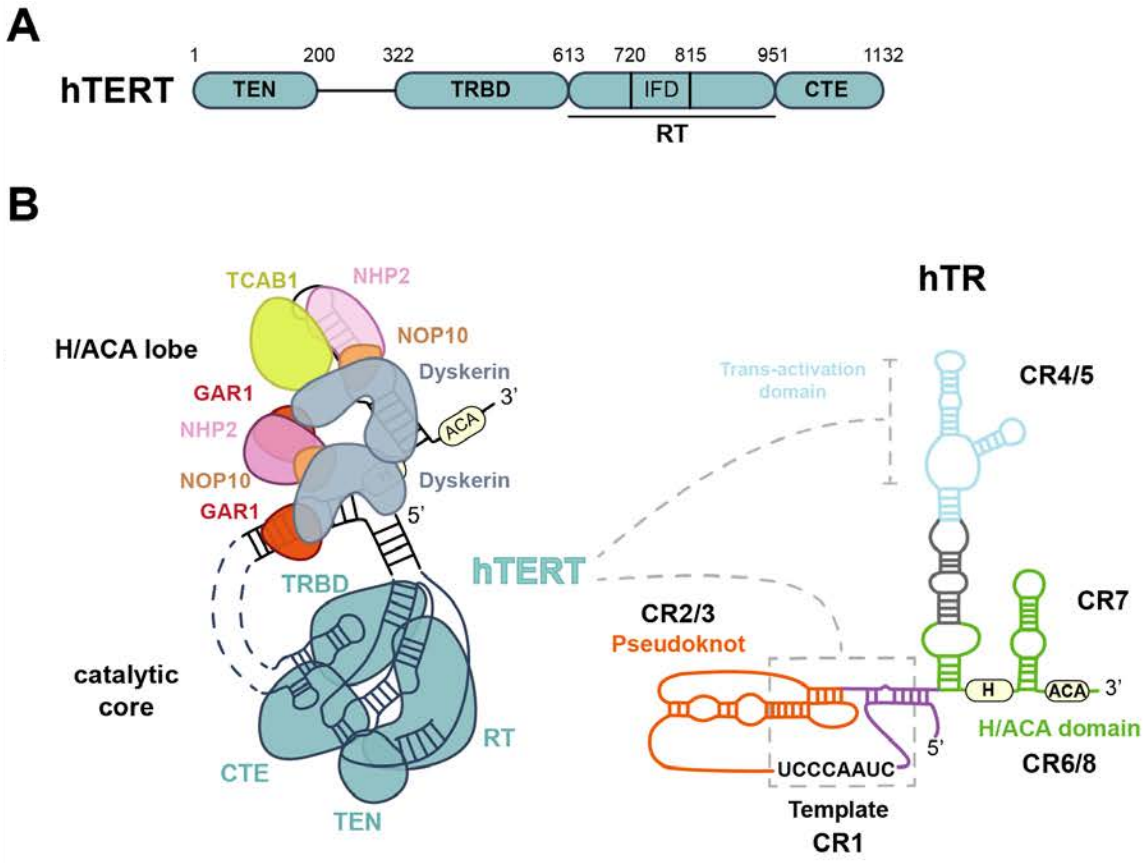

Figure 1.16. The telomerase ribonucleoprotein complex.

(A) Domain architectures of the human telomerase. TEN, telomerase essential N-terminal domain; TRBD, telomerase RNA-binding domain; RT, reverse transcriptase; IFD, insertion in the fingers subdomain; CTE, C-terminal extension; NTE, N-terminal extension. **(B)** Schematic of the human telomerase holoenzyme. The H/ACA lobe consists of proteins that binds to the H/ACA domain of the human telomerase RNA (hTR) that are key for the biogenesis and regulation of telomerase. The catalytic lobe is tethered to the H/ACA lobe by hTR. It binds to the template region in hTR and the telomeric ssDNA overhang, and catalyses the addition of telomeric repeats. Adapted from (Nguyen, et al., 2018)

1.4.1.2 Telomerase and telomere length homeostasis

Telomerase activity is essential to maintain populations of stem cells, such as cells in the reproductive systems, and the proliferative cells involved in haematopoiesis and in the renewal of the lining of the gastro-intestinal tract (Uhlen, et al., 2015). The lack of telomerase activity in somatic cells leads to cellular senescence and aging, as the regenerative ability of tissues are limited and functioning of the worn-out tissues deteriorates over time. In humans, telomerase dysregulation has been implicated in a range of diseases, most notably a family of rare conditions known as telomeropathies and the majority of cancers. Overexpression of telomerase imparts a replicative immortality to the majority (~85–90%) of human malignancies (Kim, et al., 1994; Shay & Bacchetti, 1997). On the other hand, mutations interrupting telomere extension by telomerase have been implicated in telomeropathies, where telomere shortening occurs rapidly and cells reach telomere crisis prematurely. Affected individuals often suffers from progressive bone marrow failure (BMF), pulmonary fibrosis, and predisposition to certain malignancies (see section 1.5).

The shelterin TPP1/POT1 heterodimer plays a crucial role in telomerase recruitment and stimulating telomerase processivity (Latrick & Cech, 2010). In 2022, two sub-4 Å structures of the human telomerase holocomplex bound to telomeric DNA and TPP1/POT1 was solved by cryo-EM, illustrating the molecular basis of interactions between shelterin components and telomerase (see section 1.6) (Sekne, et al., 2022; Liu, et al., 2022).

1.4.2 Tankyrase, a positive regulator of telomere length maintenance

Tankyrase 1 and 2 (TNKS and TNKS2) are members of the ADP-ribosyltransferases (ARTs) superfamily. ARTs are enzymes that catalyses the

transfer of ADP-ribose units from its cofactor nicotinamide adenine dinucleotide (NAD^+) onto a range of client substrates. This process is known as ADP-ribosylation. ADP-ribose modifications are reversible and dynamic. Several types 'reader' modules have evolved to recognise various structures of ADP-ribose. A delicate balance is established between ADP-ribose synthesis by 'writers' such as tankyrase, and its removal by specific hydrolases ('erasers') (Barkauskaite, et al., 2015).

Initially described as regulators of telomere maintenance (Smith, et al., 1998; Smith & de Lange, 2000; Cook, et al., 2002), tankyrase have been reported to partake in diverse cellular processes including resolution of sister chromatid cohesion during mitosis (Dynek & Smith, 2004; Canudas, et al., 2007), the Wnt/β-catenin signalling pathway (Huang, et al., 2009; Zhang, et al., 2011), mitotic spindle assembly (Chang, et al., 2005), centrosome maturation (Ozaki, et al., 2012), vesicle trafficking (Chi & Lodish, 2000), regulation of proteasome assembly and activity (Cho-Park & Steller, 2013), and in DDRs (Dregalla, et al., 2010; Nagy, et al., 2016). Whilst the double knockout is embryonically lethal in mice, germline inactivation of TNKS or TNKS2 alone produces developmentally normal mice with functional telomere maintenance and cell cycle control (Chiang, et al., 2008), reflecting a redundancy of TNKS/TNKS2 functions in mice. In humans, both tankyrases are ubiquitously expressed in the majority of adult and foetal tissues (Cook, et al., 2002).

1.4.2.1 Domain architecture and structural studies of tankyrases

Human tankyrases comprises five ankyrin repeat clusters (ARCs) in the N-terminus that are substrate recognition modules, a catalytic PARP domain at the C-terminus, connected by a sterile alpha motif (SAM) that mediates protein oligomerisation (Figure 1.17). The PARP domain in the tankyrases contains a structurally well-conserved ART fold, comprising of a split β-sheet with each half containing four antiparallel strands (Figure 1.18A). The PARP domain catalyses the addition of ADP-ribose groups onto its protein substrates using NAD^+ as a co-

substrate in a process known as poly(ADP-ribosyl)ation (PARylation). The key conserved amino acids involved in NAD^+ binding form the histidine-tyrosine-glutamine (HYE) triad (Figure 1.18A). Binding of the NAD^+ co-substrate to the catalytic site a kinked conformation, which strains the pyridinium *N*-glycosidic bond and reduces the activation energy required for glycosidic bond cleavage (Cohen & Chang, 2018). The small (~ 9 kDa) SAM domain is a five-helix bundle that facilitate oligomerisation of tankyrase and is key contributor to the catalysis-independent scaffolding functions of tankyrase (Figure 1.18B) (Mariotti, et al., 2016; Pollock, et al., 2017). An additional tail rich in histidine, proline and serine (referred to as the HPS region) is found only in TNKS and not TNKS2 (Figure 1.17).

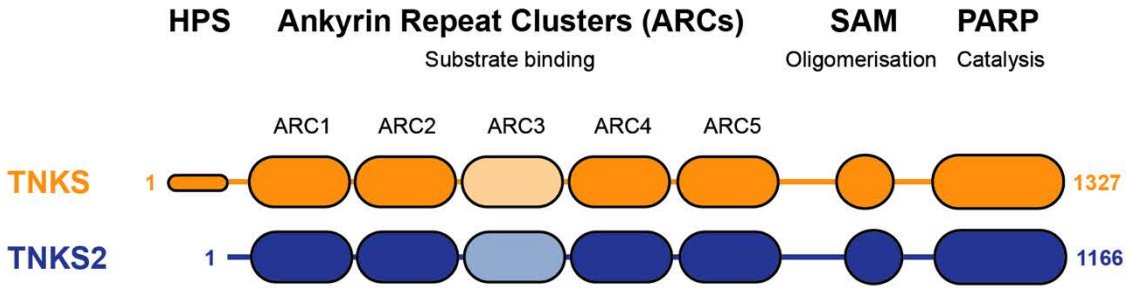

Figure 1.17. Domain architecture of human TNKS and TNKS2.

Human TNKS and TNKS2 contain five ARCs for substrate recognition and recruitment, an oligomerising SAM domain, and a catalytic PARP domain. TNKS also have a HPS region at the extreme N-terminus of the protein.

In tankyrases, substrate recognition is facilitated by the N-terminal ARCs (Figure 1.17). The architecture of ARCs is highly conserved, consisting of central ankyrin repeats that contains the substrate binding region, and two flanking ankyrin repeats (Figure 1.18C) (Guettler, et al., 2011). The majority of characterised tankyrase binders possesses the conserved but degenerate tankyrase-binding motif with the R-x-x-[small hydrophobic/G]-[D/E/I/P]-G-no P-[D/E] consensus sequence (Guettler, et al., 2011). Insertions in the tankyrase-binding motif, commonly between the most conserved arginine (position 1) and glycine (position 4) residues, can produce 'non-canonical' tankyrase-binding motifs more than eight amino acids in length (Morrone, et al., 2012; DaRosa, et

al., 2018). Four of the ARCs (ARC1/2/4/5) can bind to protein substrates with varying affinities. ARC2 and ARC5 generally have higher affinities for canonical tankyrase binding peptides compared with ARC1 (the weakest binder) and ARC4 (Guettler, et al., 2011).

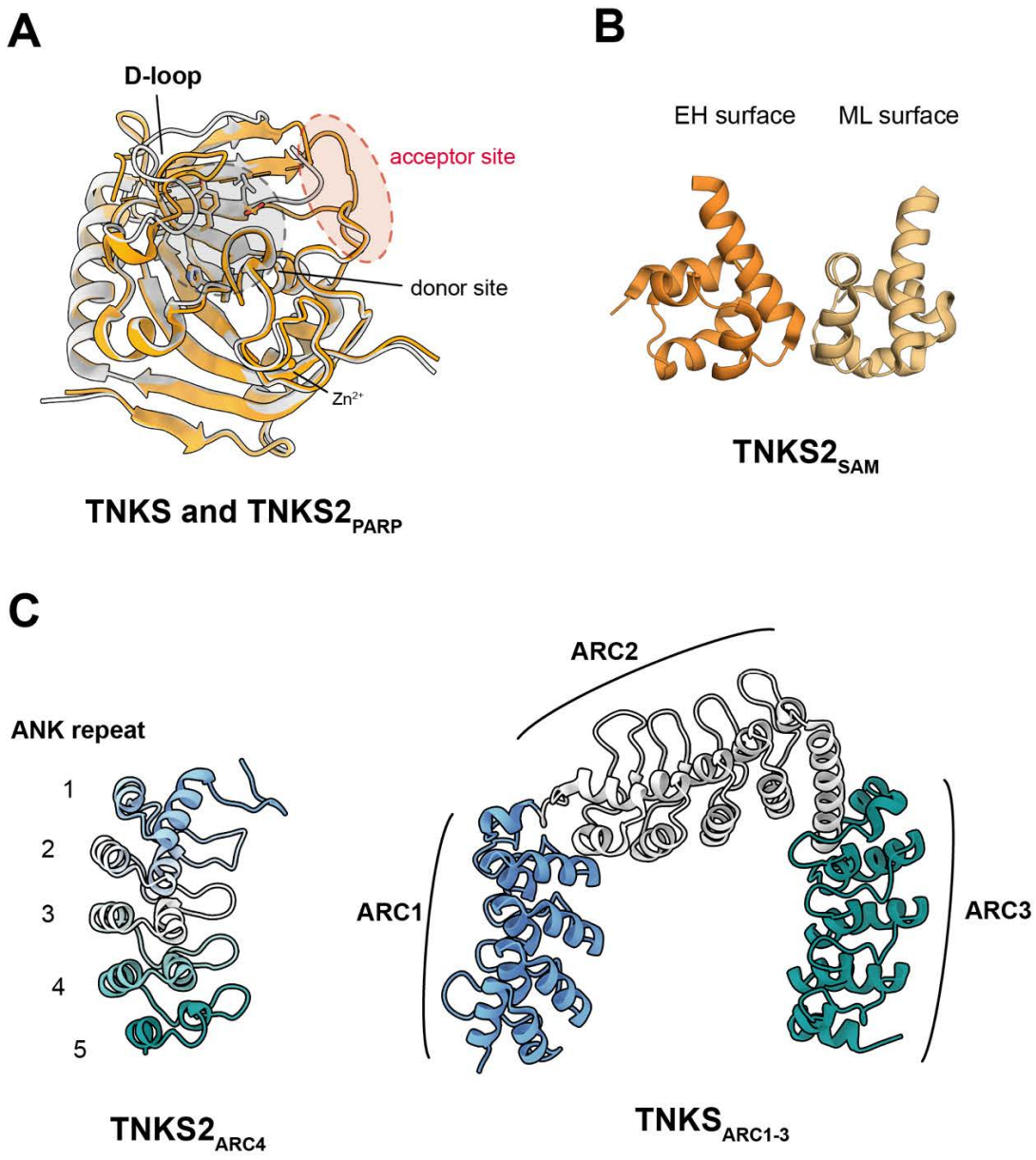

Figure 1.18. Crystal structures of human TNKS and TNKS2.

Crystal structures of **(A)** TNKS and TNKS2 PARP domains (PDBs: 2RF5 and 3KR7, respectively) (Lehtiö, et al., 2008; Karlberg, et al., 2010), (B) TNKS2 SAM domains and the two interfaces that mediates oligomerisation of TNKS (PDB: 5JRT) (Mariotti, et al., 2016), (C) TNKS2 ARC4 and the ANK repeats (1 – 5) are coloured from light blue (N-terminus) to teal (C-terminus) (PDB: 3TWQ, Guettler et al., 2011) and TNKS ARCs 1-3 (PDB: 5JHQ, Eisemann et al., 2016).

1.4.2.2 Tankyrase and telomere length homeostasis

In human cells, both tankyrases are involved in telomere length maintenance (Smith, et al., 1998; Smith & de Lange, 2000; Cook, et al., 2002). Tankyrases were first identified in yeast two-hybrid screens with the shelterin subunit TRF1 as the bait (Smith, et al., 1998; Kaminker, et al., 2001). TRF1 contains a canonical tankyrase-binding motif in the N-terminal acidic region and has been shown to interact with both TNKS and TNKS2 ARCs (Sbodio & Chi, 2002; Guettler, et al., 2011). PAR-dependent modification of TRF1 by TNKS reduce the DNA-binding affinity of TRF1 *in vitro* (Smith, et al., 1998). In telomerase-positive cells, overexpression of TNKS releases the negative regulatory effect of TRF1 and elongates the telomeres (Smith & de Lange, 2000). This is likely mediated by PARylation and removal of TRF1 from telomeres, followed by subsequent degradation via the proteasomal machinery (Chang, et al., 2005; Seimiya, et al., 2004; Cook, et al., 2002).

TNKS and its paralogue TNKS2 binds a short peptide motif, known as the tankyrase-binding motif, in the N-terminal acidic region of TRF1. *In vitro* reconstitution of the TNKS/TRF1/TIN2 ternary complex suggested that TIN2 inhibits TRF1 PARylation by TNKS, which may provide an additional level of control over telomere extension (Ye & de Lange, 2004c). The stoichiometry of this complex is still unclear. As TRF1 form constitutive homodimers, there are two tankyrase-binding motifs in shelterin complexes that can potentially interact with tankyrase. Cooperative binding of multiple ARCs to binding partners with more than one tankyrase-binding motif is less well characterised. Consecutive tankyrase-binding motifs can engage a single or multiple ARCs depending on the conformation restraints of the tankyrase ARCs and the spacing of the tankyrase-binding motifs in the interacting partner (Eisemann, et al., 2016). It is worth noting that mouse TNKS (mTNKS) can bind a TRF1 peptide containing the tankyrase-binding motif *in vitro* (Li, et al., 2016). However, mouse TRF1 lacks a tankyrase-binding motif and does not interact with mTNKS *in vivo* (Donigian & de Lange, 2007). In addition, mTNKS does not localise to telomeres in mouse embryonic fibroblasts (Donigian & de Lange, 2007). Overexpression of TNKS does not remove TRF1 from telomeres nor affect the abundance of other mouse shelterin

subunits (Donigian & de Lange, 2007). Therefore, tankyrase-dependent telomere extension is a regulatory mechanism unique to humans.

1.5 Premature aging and telomere spectrum diseases

To date, a spectrum of inherited disorders has been reported where patients suffer from conditions related to premature aging (Opresko & Shay, 2017). Telomeres from these individuals are usually shorter than age-matched healthy controls. These premature-aging syndromes, or telomeropathies, tend to be monogenetic and can be categorised into primary and secondary telomere diseases.

Primary telomeropathies are caused by dysregulations in the telomere maintenance processes, namely machineries directly involved in telomere end-protection and end-replication – the shelterin and CST complexes, and the telomerase RNP complex (Opresko & Shay, 2017). Dyskeratosis congenita (DC) was the first disorder discovered with a strong link to dysfunctional telomeres. The first case of DC was described in the 1920s, where a 20-year-old man presented with abnormal skin pigmentation, oral leukoplakia and nail dystrophy (Cole, 1930), the diagnostic triad for classic DC. A wide range of abnormalities have been documented in DC patients, including BMF, immune deficiency, enteropathy, premature hair loss, cerebellar hypoplasia with ataxia, and pulmonary disease, many of which can be attributed to the failure of the body to maintain the various stem cell compartments (Dokal, 2000). Increased prevalence of several malignancies has also been reported in patients with DC surpassing their third decade of life (Dokal, 2000; Baykal, et al., 2003; Alter, et al., 2009). The most prevalent solid tumours reported include those of the head and neck, skin, and increased risk of acute myeloid leukaemia and myelodysplastic syndrome (Alter, et al., 2009).

In comparison, secondary telomeropathies are typically inflicted by mutations in the DDR machinery and structural proteins that contribute to telomere protection and replication, telomere organisation and tethering to the nuclear envelope. Secondary telomere diseases and the implicated defective genes include ATM in Ataxia telangiectasia (Metcalfe, et al., 1996; Wood, et al., 2001; Wang, et al., 2007), LMNA in Hutchinson-Gilford progeria (Cao, et al., 2011; McCord, et al., 2012; van Steensel & Belmont, 2017), and the WRN in Werner syndrome (Du, et al., 2004; Crabbe, 2004; Edwards, et al., 2014). Studying these pathologies can expand our understanding on other potential molecular mechanisms involved in telomere maintenance. For instance, cells from patients with Werner syndrome display signs of telomere dysfunction, including accelerated telomere attrition and premature senescence that is characteristic of all telomeropathies (Opresko, et al., 2004). The defective gene in these patients is WRN, a 3' to 5' DNA helicase and exonuclease with the ability to unwind various DNA substrates including Holliday junctions and G quadruplexes (Mohaghegh, et al., 2001). Recent studies have suggested that WRN are involved in telomerase-independent recombination at telomeres, a process known as alternative lengthening of telomeres (ALT) (Opresko, et al., 2004) . It colocalises at nuclear foci formed at telomeres with the shelterin components TRF1 and TRF2, and several other proteins involved in DDR (Yeager, et al., 1999)), and may be involved in the resolution of telomeric D loops during DNA replication and recombination events (Opresko, et al., 2004).

1.6 The dynamic nature of shelterin complexes

Many of the pairwise protein-protein interactions involved in shelterin assembly have been established by biochemical, biophysical and structural studies (Figure 1.19). At the centre of action is TIN2. Theoretically, each TIN2 subunit can bind up to two homodimers of TRF proteins via the $TRF1_{TRFH}$-$TIN2_{TBM}$ (or $TRF2_{TRFH}$-$TIN2_{TBM}$) and the $TIN2_{TRFH\text{-}like}$-$TRF2_{linker}$ interaction (Chen, et al., 2008; Hu, et al., 2017). Simultaneously, TIN2 can bind to TPP1 via the $TIN2_{TRFH\text{-}like}$-$TPP1_{TIM}$ interaction (Hu, et al., 2017). TPP1 forms a heterodimer with POT1 in a 1:1 ratio (Lei, et al., 2004), whilst RAP1 binds only TRF2 also in a 1:1 stoichiometry (Chen, et al., 2011; Lim, et al., 2017), completing the six-subunit complex.

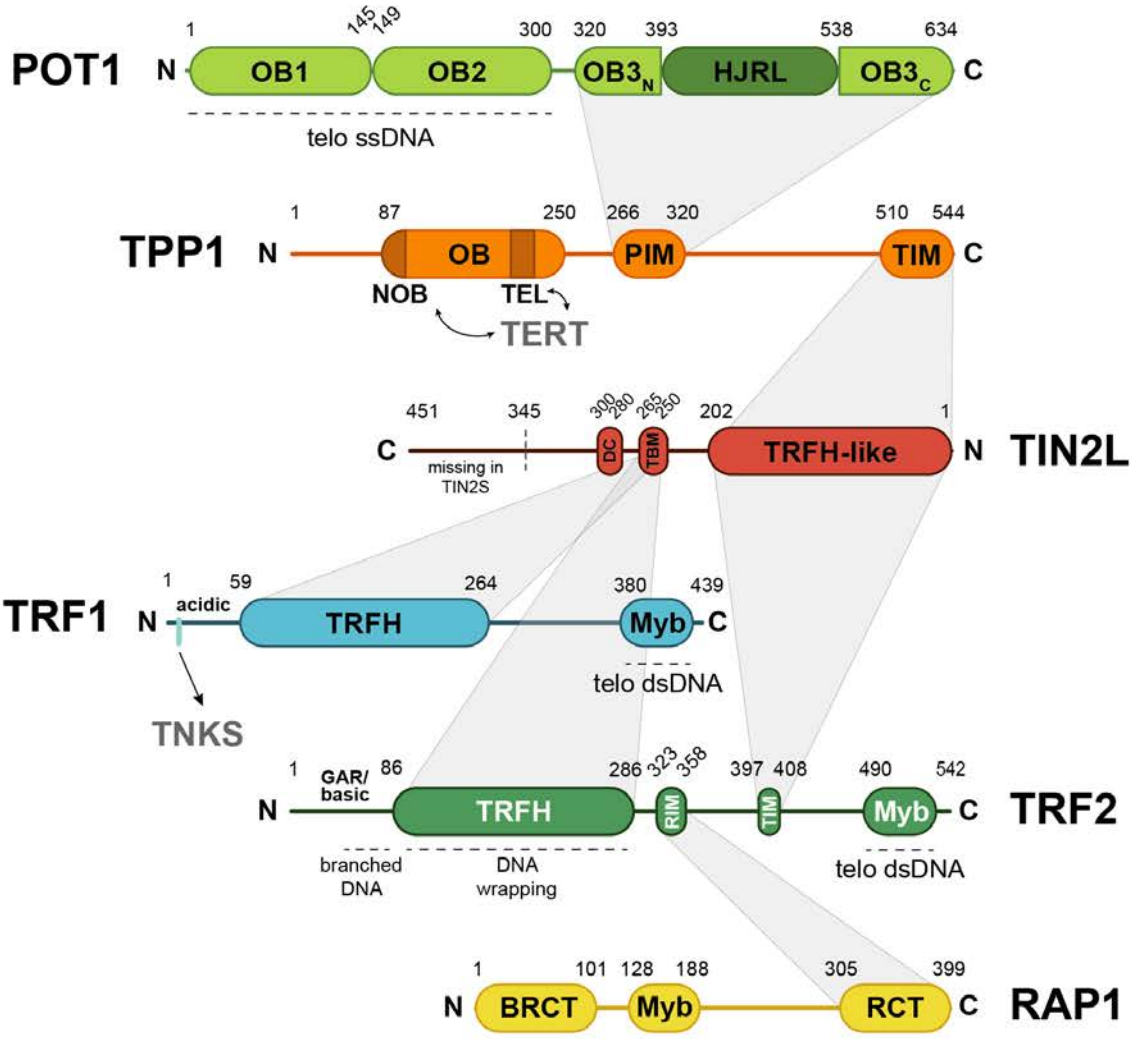

Figure 1.19. Interactions within the human shelterin complex.

1.6.1 Compositional heterogeneity

Shelterin has often been portrayed as a six-subunit complex. This was first observed in fractionated nuclear extracts (Liu, et al., 2004b). Immunoprecipitation studies in the presence of transiently overexpressed shelterin subunits in HEK293T cells also extracted all six subunits, corroborating the six-protein model (O'Connor, et al., 2006). On the other hand, experiments also suggested that the stoichiometry of shelterin *in vivo* may be varied and subcomplexes are likely to exist. Evidence shows that in mice and human cells, shelterin complexes lacking TRF1 or TRF2/RAP1 can still localise to telomeres, whereas simultaneous deletion of both TRF1 and TRF2 removes the entire shelterin complex from the telomeres (Celli & de Lange, 2005; Takai, et al., 2010). This indicated that the dsDNA binding subunits are key in establishing shelterin complexes at the telomeres. Photobleaching experiments have identified subcomplexes containing either TRF1 or TRF2 (Mattern, et al., 2004). A quantitative immunoblotting experiment suggests that several subcomplexes may be formed to explain the uneven stoichiometry of telomere-bound shelterin subunits (Takai, et al., 2010). TIN2 and TRF2/RAP1 appear to be the most abundant subunits at telomeres. TRF1 was found at a lower relative abundance compared to TRF2, and TPP1/POT1 were the least abundant (Takai, et al., 2010). This indicates that the most abundant complex minimally involves the TRF2/RAP1/TIN2 complex, followed by one with TRF1 (TRF2/RAP1/TIN2/TRF1). Finally, a smaller portion of the dsDNA-binding shelterin module also incorporates the TPP1/POT1 heterodimer. The proportion of several shelterin subunits found at telomeres also appeared to vary depending on telomere length (Takai, et al., 2010). TRF2, RAP1 and TIN2 coated shorter telomeres at a higher density, whist TRF1, TPP1 and POT1 subunits were found at a lower abundance on shorter telomeres (Takai, et al., 2010).

Furthermore, evidence shows that the shelterin subunits do not form static complexes. For instance, TRF1 has been shown to displace TRF2 from TIN2, and TPP1 binding can induce conformational changes in TIN2 to allow simultaneous association of TIN2 with TRF1 and TRF2 (O'Connor, et al., 2006; Chen, et al., 2008; Janovič, et al., 2019). The relatively weak affinity of

interactions between many shelterin subunits (in the high nanomolar to low micromolar range) also enables shelterin to form dynamic complexes (Chen, et al., 2008; Chen, et al., 2017; Hu, et al., 2017). It is conceivable that the composition of shelterin complexes formed may depend on spatiotemporal factors such as the cell cycle, the underlying telomere structure, and its interactions with other binding partners to perform its telomeric functions.

In 2017, an *in vitro* reconstitution of a five-subunit human shelterin complex lacking TRF1 identified a $(TRF2/RAP1)_2/(TIN2S/TPP1/POT1)_1$ complex, which supported the established protein-protein interactions within shelterin, except the surprising finding that only one TRF2 homodimer was associated with each TIN2S subunit (Lim, et al., 2017). Reconstitution of full six-subunit human complex has not been documented as I began this PhD project and was only reported by Zinder et al., (2022) as I wrote this thesis, which I will discuss in Chapter 5.

1.6.2 Conformational flexibility

All the protein-protein interactions between shelterin subunits are domain-peptide interactions (Chen, et al., 2008; Chen, et al., 2017; Rice, et al., 2017; Gaullier, et al., 2016). Besides POT1, the other five shelterin proteins have domains interspersed by long flexible linkers (Figure 1.5, Figure 1.11, Figure 1.13). Evidence from recent efforts in structural characterisation of shelterin using EM illustrates that these linkers contribute to the flexible conformations adopted by shelterin complexes (Sekne, et al., 2022; Smith, et al., 2022; Zinder, et al., 2022) (discussed below). These studies focused on solving the structure of the ssDNA-binding module TIN2/TPP1/POT1. This heterotrimer plays several essential roles in telomerase recruitment, regulation and telomere protection (Hockemeyer, et al., 2005; Denchi & de Lange, 2007; Takai, et al., 2011; Baumann & Cech, 2001; Ye, et al., 2004b; Pike, et al., 2019).

One of the EM studies reported only the most ordered subunit POT1 was resolved to 7.9 Å (Smith, et al., 2022). Even then, a 9.6 Å map of an alternative 'open' conformation of POT1 was obtained where the OB1 and OB2 were spaced further apart compared with that captured in the crystal structure of DNA-bound $POT1_{OB1-2}$ (Smith, et al., 2022; Lei, et al., 2004). The authors showed that the five-amino acid linker between OB1 and OB2 in POT1 (residues 145 – 149) allows some degree of flexibility in the orientation of these two domains (Smith, et al., 2022). This also permitted POT1 to recognise ssDNA with short spacers introduced between the optimal TTAGGGTTAG sequence without a drastic drop in binding affinity (Smith, et al., 2022). The authors proposed that the structural flexibility of the POT1 OB-folds may play a role in shaping the telomeric ssDNA region (Smith, et al., 2022). Previous studies have demonstrated that POT1 can unfold intramolecular G-quadruplexes, which are secondary structures formed by G-rich sequencing via Hoogsteen base pairing (Zaug, et al., 2005). This has been proposed as one level of telomerase regulation, as resolution of G-quadruplexes has been shown to restore telomerase processivity by providing the enzyme with linear DNA for telomere repeat synthesis (Zaug, et al., 2005).

Further evidence supporting the structural heterogeneity of shelterin was provided by negative-stain EM studies of the six-subunit shelterin and various subcomplexes (Zinder, et al., 2022). For the same heterotrimer as discussed above, the authors reported high conformational variability in the relative positioning between the domains in POT1, TPP1 and TIN2S (Zinder, et al., 2022). Attempts to reduce the extent of structural flexibility of this subcomplex by removing the unstructured linkers did not have a dramatic impact on conformational plasticity (Zinder, et al., 2022). Moreover, building up the complex with the dsDNA-binding subunits TRF1 and TRF2 did not alter the conformational variability of shelterin (Zinder, et al., 2022). The presence of telomeric DNA did not reduce structural heterogeneity of shelterin complexes (Smith, et al., 2022; Zinder, et al., 2022).

Another attempt in structural analysis of full-length shelterin subunits was in the context of the human telomerase (Sekne, et al., 2022). This study revealed the molecular basis behind how TPP1/POT1 channels and stabilises the hTERT-

DNA interaction to enhance telomerase processivity (Sekne, et al., 2022). The structures showed that upon binding to hTERT, the flexible loop forming the glutamate-rich region of the TEL patch in $TPP1_{OB}$ folds into an α-helix and interacts with the TEN domain in hTERT (Sekne, et al., 2022). Another surprising finding was that POT1 also bound to hTERT and formed a gate in front of the telomerase active site (Sekne, et al., 2022). This provided a mechanism for how TPP1/POT1 increases telomerase processivity, which entails cooperative effects resulting from the association of both TPP1 and POT1 to hTERT to reduce DNA dissociation during telomere repeat synthesis (Sekne, et al., 2022). Although TPP1 binding to hTERT induced conformation changes that reduced structural flexibility TEN domain, the same was not reciprocated for TPP1 nor POT1. This study also started with the TIN2/TPP1/POT1 heterotrimer as did the other two EM analyses (Smith, et al., 2022; Zinder, et al., 2022). However, only the OB-folds at the N-termini of TPP1 and POT1 were visible in the EM density maps (Sekne, et al., 2022). Both the $POT1_{OB3/HJRL}$ and $TIN2_{TRFH}$ domains in this heterotrimer interact with the unstructured C-terminal portion of TPP1, which were not stabilised upon TPP1/POT1 binding to telomerase (Sekne, et al., 2022). Together, these studies indicate that shelterin subunits and the assemblies they form are flexible and dynamic both in terms of composition and conformation, which contributes to the wide array of protein interactions engaged by shelterin and biological functions of shelterin at telomeres.

Project Aims

Studies thus far on the human shelterin complex *in vitro* mainly utilised individual shelterin subunits and subcomplexes, hence our knowledge of the behaviour of the full shelterin complex is still limited. Furthermore, current structural information on shelterin is limited to single motifs and domains of individual subunits, binary and ternary complexes solved by NMR and X-Ray crystallography. The overall architecture of the shelterin complex also remains uncharted.

In my PhD project, I aim to address the following questions:

1. What are the stoichiometries of shelterin complexes? Given the evidence supporting the formation of subcomplexes *in vivo*, which shelterin subcomplexes form stable protein assemblies?

2. What is the overall architecture of shelterin complexes? Are there any novel interfaces that regulate the assembly of the shelterin complex?

To start answering these questions, I expressed and purified the full shelterin complex and various subcomplexes. Extensive biochemical and biophysical assays were performed to characterise the stoichiometry and DNA-binding affinities of two recombinant shelterin complexes – the six-subunit shelterin complex and one lacking TRF1. Finally, I studied the architecture of the (–TRF1) complex and integrated the use of crosslinking mass spectrometry to explore protein-protein interactions in the context of the DNA-bound shelterin complex.

Statement of COVID impact

The COVID-19 pandemic has led to disruptions that affected my ability to complete the work I had originally planned. Due to the initial lockdown in 2020, I was unable to work in the lab from mid-March to June. Later waves of lockdowns also severely delayed work that involved internal and external collaborations. Whilst the work I document here revealed interesting aspects that contribute to our understanding of the compositional and conformation heterogeneity within shelterin complexes, there are several aspects that I would have explored further if the pandemic had not taken place.

Firstly, although the approach I took to reconstitute shelterin complexes gave us insights into the different stoichiometries of shelterin, these heterogeneous complexes were challenging to resolve using chromatographic methods. This complicated many biochemical and structural characterisations of shelterin. With more time, the next step I would have taken is to build shelterin bottom-up with individually expressed shelterin subunits. This would have allowed me to better understand what regulates complex assembly.

Secondly, the negative-stain EM study of the shelterin (–TRF1) subcomplex has shown that the absence of TRF1 does not dramatically alter the overall architecture of the shelterin complex. With more time, I would have taken this subcomplex forward for cryo-EM analysis to obtain a higher-resolution 3D reconstruction and better understand the protein-protein interactions and the positioning of the telomeric DNA within the shelterin complex.

The final aspect in my PhD was to examine the interaction of shelterin with tankyrase and telomerase to gain insights into how shelterin performs its roles in telomere maintenance. For this, I have purified tankyrase and telomerase, which sets the stage for future biochemical studies to characterise the effect of PARylation on the stability of shelterin and its ability to recruit and potentially enhance telomerase processivity.

Further details can be found in the discussions (Chapter 5).

Signature (Student: Yexin Xie)

Signature (Supervisor: Dr Sebastian Guettler)

Date

29 March 2022

2 Biochemical and biophysical characterisation of the full shelterin complex

Subunits that are recruited to the telomeres forming telomeric shelterin complex have been identified and well-characterised in the past two decades. Many protein-protein and protein-DNA interactions were uncovered by co-immunoprecipitation, yeast two-hybrid, and structural studies. Recent studies have documented reconstitutions of various subcomplexes of human shelterin (Lim, et al., 2017) and of the full mouse shelterin complex (Erdel, et al., 2017). The *in vitro* reconstitution of the full shelterin complex with all six subunits was undocumented [1].

In my PhD project, I aimed to study the structure of the shelterin complex, and its roles in telomere length maintenance. These pursuits require large amounts, and ideally highly pure, material. In this chapter, I begin by describing the expression and purification of the full shelterin complex with six full-length components, followed by biochemical and biophysical characterisations to determine its size and stoichiometry, and finally its affinity for telomeric and non-telomeric DNA substrates.

[1] In 2022, Zinder et al. reported *in vitro* reconstitutions of various shelterin complexes. There are notable differences between our systems that are interesting, which I will mention in the discussions (Chapter 5).

2.1 Expression and purification of the full shelterin complex

2.1.1 Overview

The full human shelterin complex was recombinantly expressed in Sf9 insect cells using a single baculovirus. All six full-length shelterin subunits were expressed together, namely TRF1, TRF2, RAP1, TPP1, POT1, and either dStrepII-tagged TIN2L or TIN2S (later referred to as shelterin^{TIN2L} and shelterin^{TIN2S}, respectively). All sequences were codon-optimised for expression in *E. coli*. See sections 6.1.1 and 6.2.1 for descriptions of the shelterin proteins and the expression constructs.

I adapted the two-step purification protocol (Figure 2.1A), which involves a Strep-tag affinity chromatography and size exclusion chromatography (SEC), to produce a range of shelterin complexes (Appendix Figure 7.1 and Figure 7.2). See section 6.3.3 for the detailed protocols. Briefly, all shelterin purifications began with Strep-tag affinity chromatography, where the recombinant shelterin complex was extracted from insect cell lysate through capturing dStrepII-tagged TIN2L/S. The lysis and the first wash step used a high-salt buffer (1 M NaCl) to remove contaminating DNA. The second wash step equilibrated the column with the SEC buffer and dropped the salt concentration to 0.5 M NaCl necessary for stabilising apo shelterin complexes. Both shelterin^{TIN2L} and shelterin^{TIN2S} complexes followed the same pattern of elution with similar yields during the affinity and SEC purifications (described below). Hence, I will focus on shelterin^{TIN2L} for the remaining chapter unless specified otherwise.

The protein concentration was quantified by spectroscopy using extinction coefficients calculated for shelterin, assuming the stoichiometry of the full shelterin complex containing (TRF1/TRF2/RAP1)$_2$/(dStrepII-TIN2/TPP1/POT1)$_1$ (see below, and sections 1.6 and 6.2.3). This stoichiometry has been established by a multitude of studies demonstrating the following:

- TRF1 and TRF2 form homodimers (Fairall, et al., 2001; Chen, et al., 2008).
- each TRF2 monomer can bind up to one RAP1 molecule (Chen, et al., 2011; Lim, et al., 2017).

- TIN2/TPP1/POT1 form a complex with a 1:1:1 stoichiometry (Lim, et al., 2017; Hu, et al., 2017; Chen, et al., 2017; Rice, et al., 2017).
- each TRF1 homodimer can bind up to two TIN2 subunits via the $TRF1_{TRFH}$-$TIN2_{TBM}$ interaction (Chen, et al., 2008).
- TRF2 can interact with TIN2 in two different binding modes, namely $TRF2_{linker}$-$TIN2_{TRFH\text{-}like}$ and $TRF2_{TRFH}$-$TIN2_{TBM}$, with the former being the higher-affinity interaction (Chen, et al., 2008; Hu, et al., 2017).

2.1.2 Purification of the full shelterin complex

The affinity purification step produced a high yield of shelterin complexes containing the dStrepII-tagged TIN2L/S subunit; the peak elution fraction from a StrepTrap 5 mL HP column contained 0.5 mL of up to 15 mg/mL of protein from 1L of insect cell culture (Figure 2.1B and Figure 2.3A). The affinity elution fractions were analysed by SDS-PAGE (Figure 2.1C), and the most concentrated fractions were further purified using a Superose 6 Increase 10/300 GL column. Frequently, two bands with lower intensities were often present in the affinity elution fractions (Figure 2.1C). The band below the 75 kDa marker and near 50 kDa corresponded to heat shock 70 kDa and tubulin α/β chains from Sf9 insect cells, as confirmed by mass spectrometry (Table 7.1). The latter was enriched in the void and mostly separated from the soluble shelterin complexes during SEC.

SEC purification of both shelterin^{TIN2L} and shelterin^{TIN2S} produced three visible species, with one main peak at 12 – 12.5 mL (species ***II***), and a shoulder eluting before and after this peak (species ***I*** and ***III***) (Figure 2.2A and Figure 2.3B). At this stage, the most concentrated fraction from the main peak (species ***II***) contained 1 – 2 mg/mL of protein. SDS-PAGE analysis followed by Coomassie staining and Western blotting showed that the six shelterin subunits were present in various abundances throughout the SEC elution (Figure 2.2B and C). Species ***I*** and ***II*** comprised of two variations of the full shelterin complex containing all six subunits (Figure 2.2B and C, Figure 2.3C). Notably, Western blotting analysis indicated that species ***I*** contained more TRF2 and its partner RAP1 compared

with species ***I***, and species ***III*** contained mainly the dStrepII-tagged TIN2 subunit and subcomplexes of shelterin (Figure 2.2C, Figure 2.3C). Species ***I*** and ***II*** contained robust shelterin complexes that remained stable over consecutive SEC purifications (Figure 2.4). In addition, the full shelterin complex with an expected MW of 490 kDa eluted earlier (~12 mL, species ***II***) from Superose 6 Increase 10/300 GL in comparison to the 669 kDa thyroglobulin (~13.5 mL). This may be explained by the disordered regions within many of the shelterin subunits, which would increase its hydrodynamic radius compared to another complex at the same molecular weight with a globular shape.

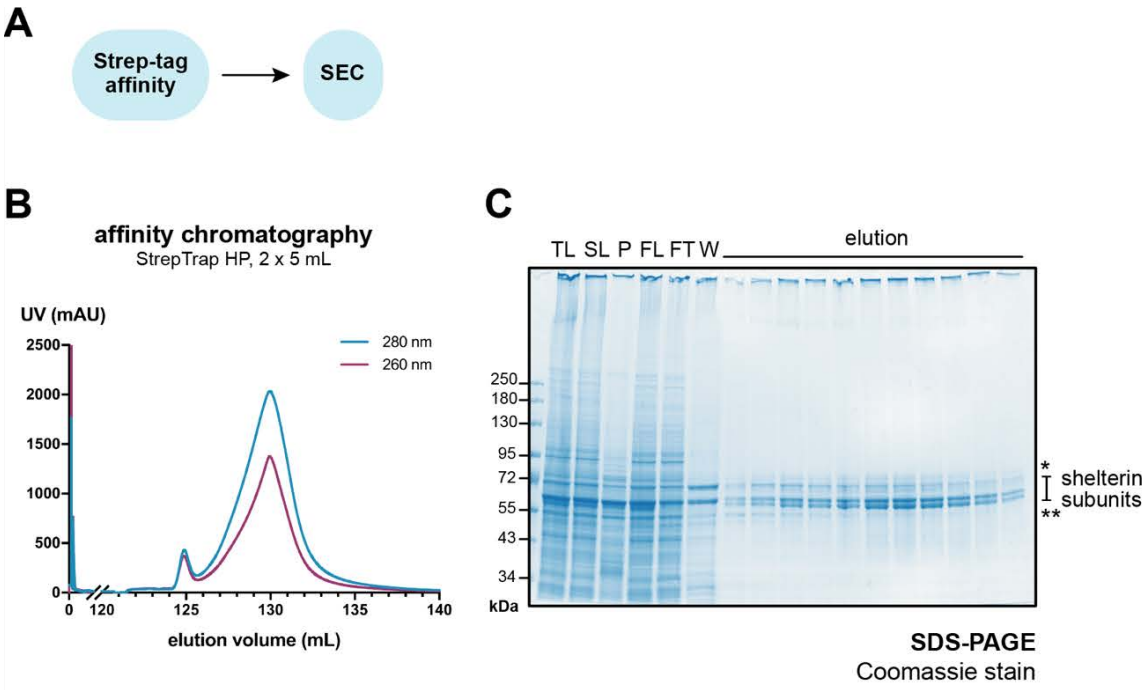

Figure 2.1. Affinity purification of the full shelterin^{TIN2L} complex.

(A) Schematic of the purification protocol. **(B)** Elution profiles from the Strep-tag affinity chromatograph. **(C)** SDS-PAGE analysis of the insect cell lysate and affinity elution fractions. TL – total lysate, SL – soluble lysate, P – insoluble pellet, FL – filtered lysate, FT – flowthrough, W – wash. * and ** indicate SF9 protein contaminants HSP70 and tubulin α/β chains, respectively.

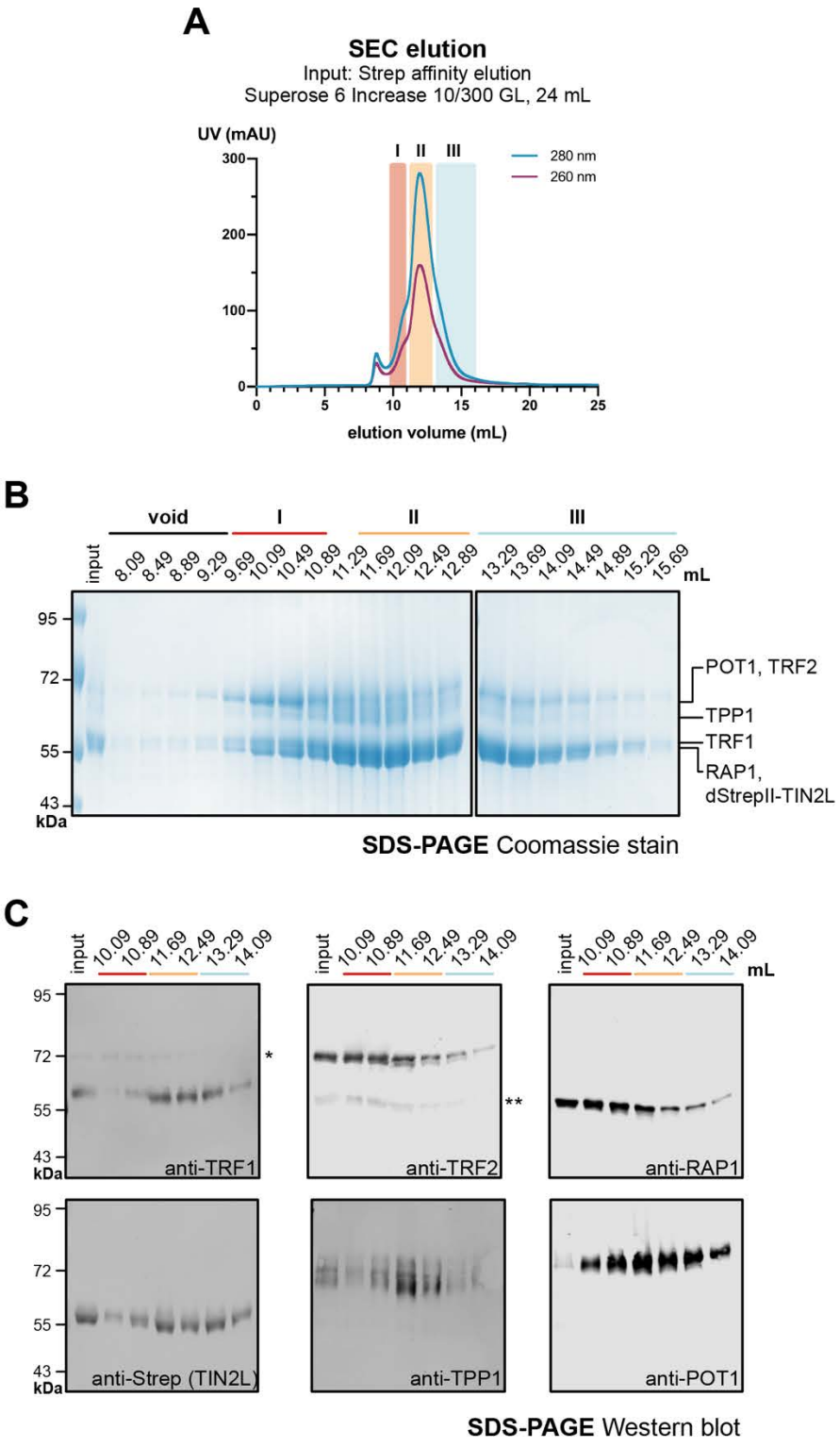

Figure 2.2. SEC purification of the full shelterin^{TIN2L} complex.

(A) Representative UV chromatogram from a SEC purification. 0.5 mL of the peak affinity elution fraction was injected and eluted at 0.3 mL/min in 200 µL fractions. **(B)** SDS-PAGE and **(C)** Western blot analysis of the SEC elution fractions as shown in (A). **(**The anti-TRF1 antibody used was custom generated by Thermo Fisher. * and ** indicate non-specific binding of the anti-TRF1 and anti-TRF2 antibodies to TRF2 and TRF1, respectively.

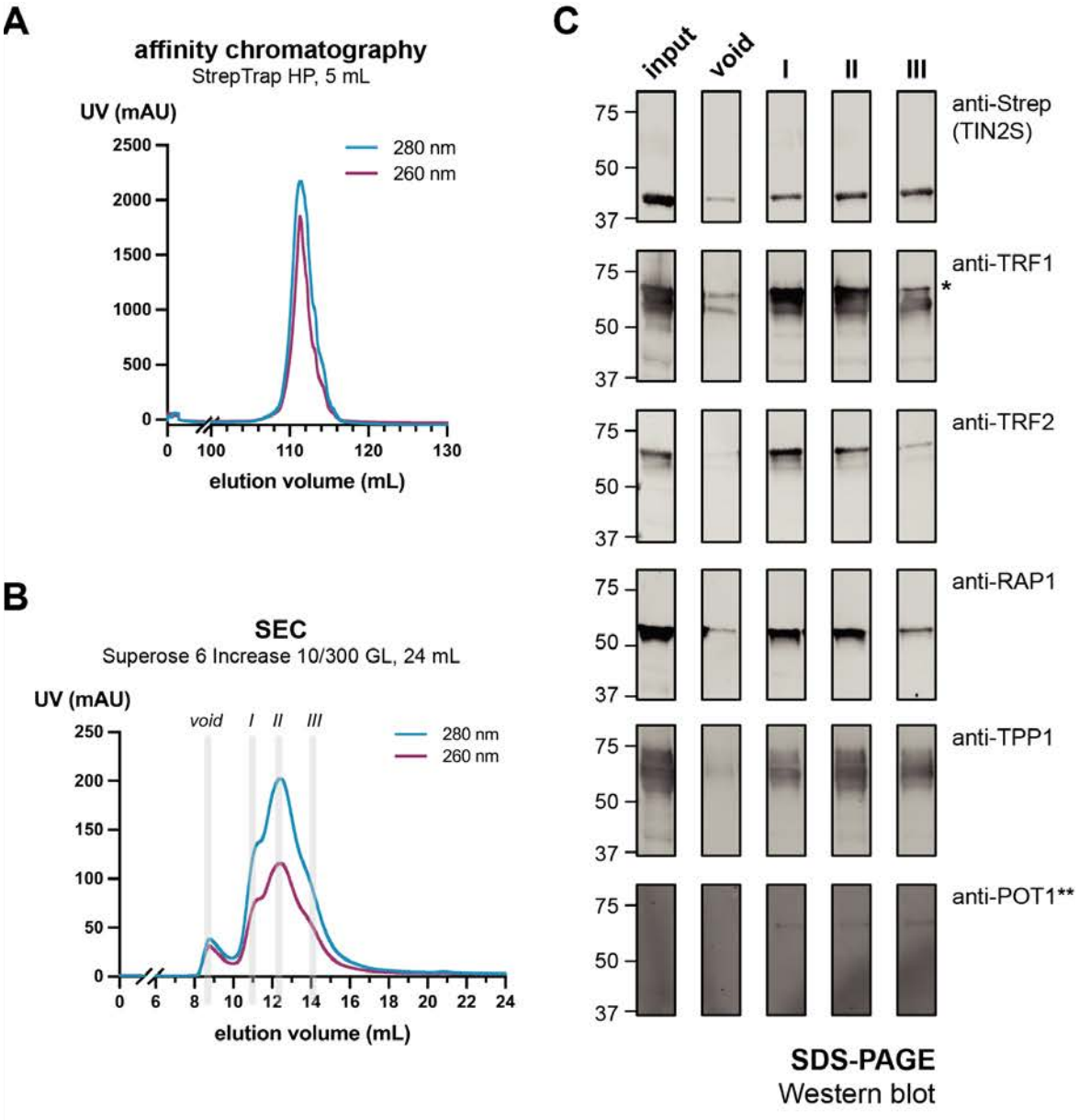

Figure 2.3. Two-step purification of the full apo shelterin^{TIN2S} complex

(A) Strep-tag affinity chromatography and **(B)** SEC purification of the apo shelterin^{TIN2S} complex. **(C)** The SEC input and SEC elution fractions from the void, species ***I***, ***II***, and ***III***, as shown in (B), were resolved by SDS-PAGE and analysed by Western blotting. All blots were scanned together with the same laser intensity (except POT1). The anti-TRF1 antibody (ab10579) recognises both TRF1 (lower band) and TRF2 (*). ** the quality of this batch of anti-POT1 antibody (ab124784) was suboptimal, and the blot was overexposed to visualise the bands for POT1 only.

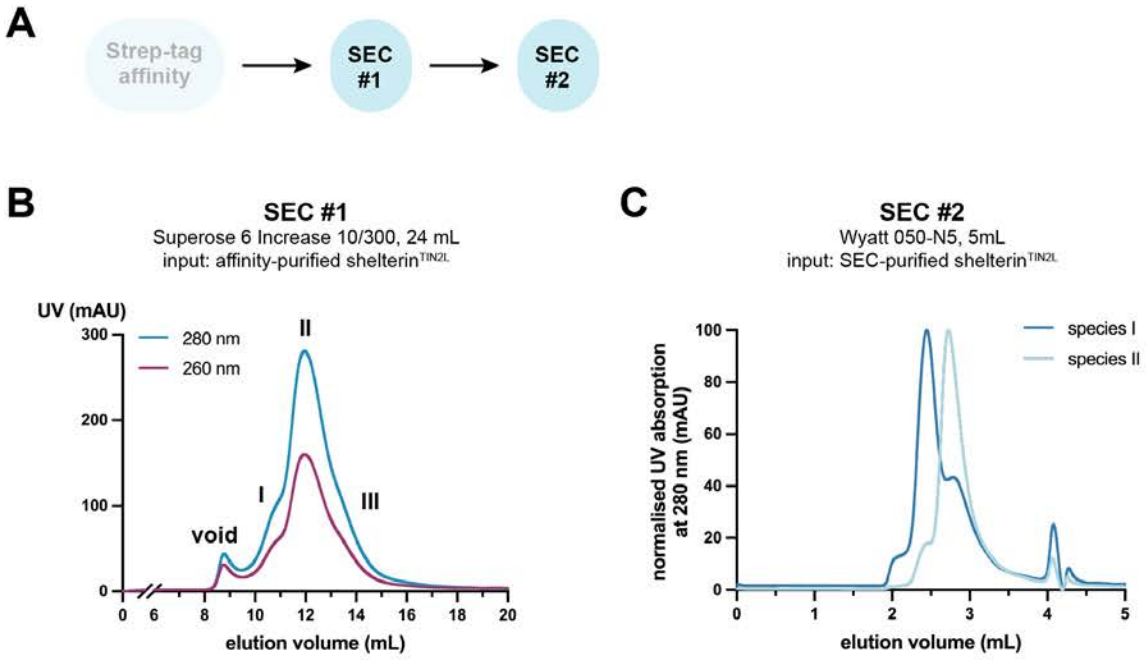

Figure 2.4. Two stable species of the full shelterin complex.

(A) Purification protocol for the full shelterin^{TIN2L} complex. **(B)** SEC purification #1. Affinity-purified protein was further subjected to SEC using Superose 6 Increase 10/300 GL, eluting at 0.3 mL/min. **(C)** SEC purification #2. Fractions from peak ***I*** and ***II*** were pooled separately from SEC #1 and re-injected into a Wyatt 050-N5 column at 0.2 mL/min.

The highest resolution was achieved using the Superose 6 Increase 10/300 GL column and eluting at a slow flow rate (0.1 – 0.3 ml/min). The Superose 6 XK 16/70 PG (120 mL) SEC column was tested for scaling up the protein production (Appendix Figure 7.3). However, this column had lower resolving power at the higher MW range than the Superose 6 Increase 10/300 GL column and was unable to separate the different species (Appendix Figure 7.3). For subsequent experiments where large amounts of shelterin complex were necessary, several affinity elution fractions were pooled and purified consecutively using the Superose 6 Increase 10/300 GL column. Since there was a substantial overlap of the three different peaks, the central peak fraction with up to two fractions on either side were usually pooled for downstream experiments to reduce contamination from other species.

2.2 Recombinant shelterin complexes have heterogeneous compositions

To further characterise the different species of shelterin complexes observed during protein purification, I determined the absolute MW using SEC-MALS and mass photometry, and I measured the size of affinity- and SEC-purified shelterin complexes using dynamic light scattering (DLS). For the biophysical characterisations described in this section, the TIN2L- and TIN2S-containing versions of the full shelterin complexes were expressed and purified as described above (section 2.1.2). The Wyatt column WTC-050-N5 (5 mL) was used to resolve the shelterin species for MALS analysis because it produced similar elution profiles for affinity-purified shelterin compared with Superose 6 Increase 10/300 GL column, whilst using less sample and drastically shortening the elution time due to its smaller bed volume.

2.2.1 Determining the MW and size of the full shelterin complex using SEC-MALS and DLS

Recombinant shelterin^{TIN2L} complex in each SEC species was monodisperse, as indicated by polydispersity indices of close to 1.00 (Figure 2.5). The MW of the full apo shelterin^{TIN2L} complex determined by SEC-MALS was 461.8 (± 26.3) kDa across the centre of species ***II***, which is near the expected MW of 491.3 kDa for a $(TRF1/TRF2/RAP1)_2/(dStrepII\text{-}TIN2L/TPP1/POT1)_1$ complex (Figure 2.5). The front (species ***I***) and trailing (species ***III***) shoulders were not resolved by SEC, so their MWs could not be determined by SEC-MALS. Moreover, the light scattering (LS) signal from aggregates eluting in the void buried the signal from soluble shelterin complexes in the front shoulder. Species ***II*** had a weight-averaged mean square radius (r_w) of 23.1 (± 8.59) nm. The large error was due to the size of shelterin approaching the detection limit of the MALS instrument.

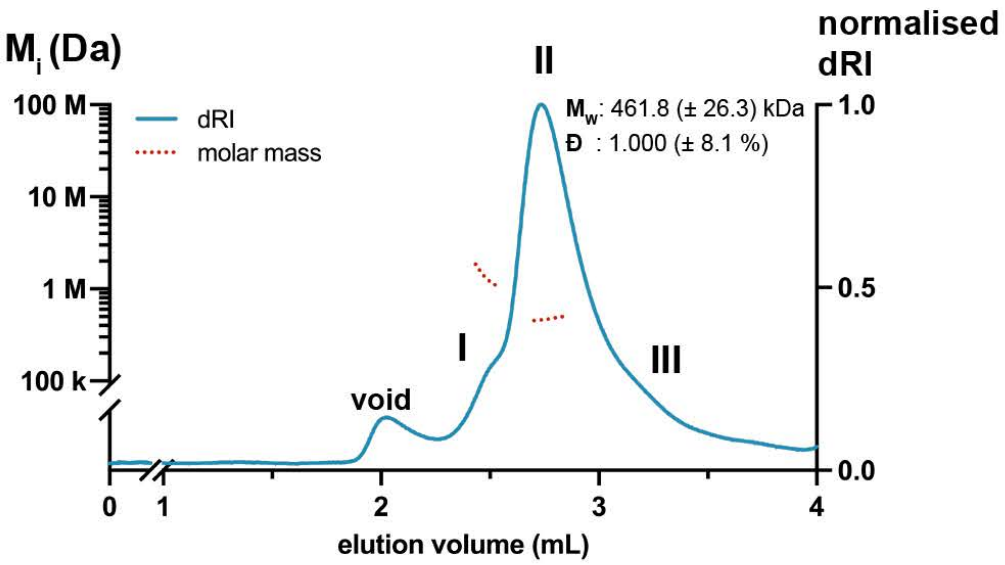

Figure 2.5. SEC-MALS analyses of the apo shelterin complex.

Affinity-purified (20 µL of 5 µM) shelterin^{TIN2L} was injected into a WTC-050N5 SEC column (5 mL) and eluted at 0.2 mL/min. Normalised dRI (blue) and the M_W distribution of proteins eluting at the centre of the peaks (red dotted line) are plotted. The mean average molecular weights (M_W) and dispersities (Đ) ± uncertainty are provided for species ***II***.

In addition, I used DLS to measure the hydrodynamic radii (R_H) of apo shelterin complexes, in order to gauge the size and homogeneity of complexes from different steps of the protein purification (Figure 2.6 and Table 2.1). Affinity-purified shelterin had the largest R_H values (~25 – 55 nm, weight-averaged). With additional SEC purification, shelterin complexes had a narrower range of R_H values and more symmetrical peaks in the frequency distribution plots, indicating an improvement in sample homogeneity (Figure 2.6). SEC-purified shelterin complexes from the front shoulder had the largest R_H values (species ***I***, ~20 – 31 nm), whereas shelterin from the centre of the main SEC peak had more uniform R_H values (species ***II***, ~17 – 19 nm). In conclusion, the recombinant six-member shelterin complexes can have various sizes. Compared with the full shelterin complex in the main SEC peak, the early eluting shelterin species have different conformations, different stoichiometry, or both.

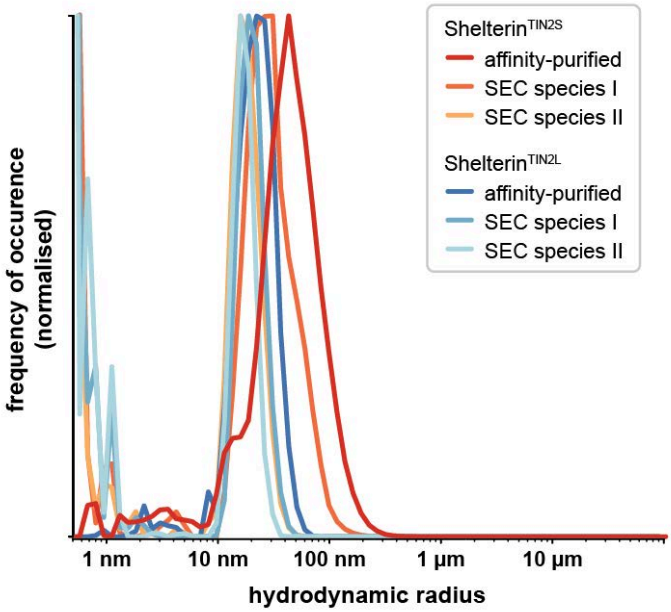

Figure 2.6. Hydrodynamic radii of shelterin^{TIN2L} and shelterin^{TIN2S} complexes.

Affinity- and SEC-purified shelterin complexes were subjected to DLS analysis. The hydrodynamic radii (log scale, weight-averaged) recorded for each sample were plotted against the frequency of occurrence, and then normalised.

Shelterin complex	Sample	Hydrodynamic radius (nm)
shelterin^{TIN2S}	affinity-purified	55.45 (50.72 – 60.18)
	SEC species ***I***	30.56 (26.80 – 34.32)
	SEC species ***II***	19.28 (16.44 – 22.12)
shelterin^{TIN2L}	affinity-purified	24.55 (21.37 – 27.73)
	SEC species ***I***	20.33 (19.34 – 21.32)
	SEC species ***II***	17.50 (17.02 – 17.98)

Table 2.1. Hydrodynamic radii of shelterin^{TIN2L} and shelterin^{TIN2S} complexes.

Brackets next to the hydrodynamic radius measurements indicate the error range in nm.

2.2.2 Determining the size of the full shelterin complex using mass photometry

Mass photometry is a label-free interferometric scattering microscopy technique for determining MW of molecules with high accuracy (Young & Kukura, 2019; Sonn-Segev, et al., 2020). See section 6.5.3.2 for details on mass photometry data collection and analysis. The contrast generated by a selection of protein standards was measured before each experiment to calibrate the mass photometer. Samples were diluted to ~100 nM prior to analysis to enable accurate determination of MW. For shelterin complexes, the weakest interactions (on the low micromolar scale) were difficult to capture at such low concentrations; high abundances of smaller species were observed due to dissociation of the full complex. Therefore, to stabilise larger protein complexes, the samples were crosslinked using a low concentration of glutaraldehyde (0.075% or 0.1%) and quenched prior to data acquisition. Many samples were measured before (native) and after crosslinking in order to obtain the most accurate MW whilst capturing any larger MW species present in the sample.

Mass photometry measurements of SEC-purified full shelterin complex showed two major species for shelterin^{TIN2L} at ~490 kDa and ~690 kDa (Figure 2.7A). Guided by our current knowledge of the protein-protein interactions within the shelterin complex (see section 2.1), I listed possible combinations that may explain the observed MW of detected species (Figure 2.7B and Table 2.2). The 490 kDa species was more abundant in SEC species ***II*** (Figure 2.8) and matched the expected MW of a $(TRF1/TRF2/RAP1)_2/(dStrepII\text{-}TIN2L/TPP1/POT1)_1$ complex, hereon referred to as the minimal shelterin complex. The 690 kDa complex was more abundant in earlier SEC fractions (Figure 2.8). The difference of ~200 kDa could be contributed by an extra $(TRF2/RAP1)_2$ subcomplex (MW: 207.7 kDa), giving a $(TRF1)_2/(TRF2/RAP1)_4/(dStrepII\text{-}TIN2L/TPP1/POT1)_1$ complex (expected MW: 699.1 kDa, Table 2.2). This is consistent with the observation that species ***I*** contains more TRF2/RAP1 (Figure 2.2B). Another possible explanation for the higher-MW complex is $(TRF1/TRF2/RAP1/dStrepII\text{-}TIN2L/TPP1/POT1)_2$ (expected MW: 674.6 kDa). Theoretically, this is possible as there are two distinct modes of interactions between the TRF1/2 subunits and

TIN2, and each monomer of TRF proteins has the potential to bind one copy of TIN2 (Chen, et al., 2008; Hu, et al., 2017). However, the stoichiometry of this larger MW species could not be analysed conclusively due to insufficient separation of the two complexes using the two-step purification protocol.

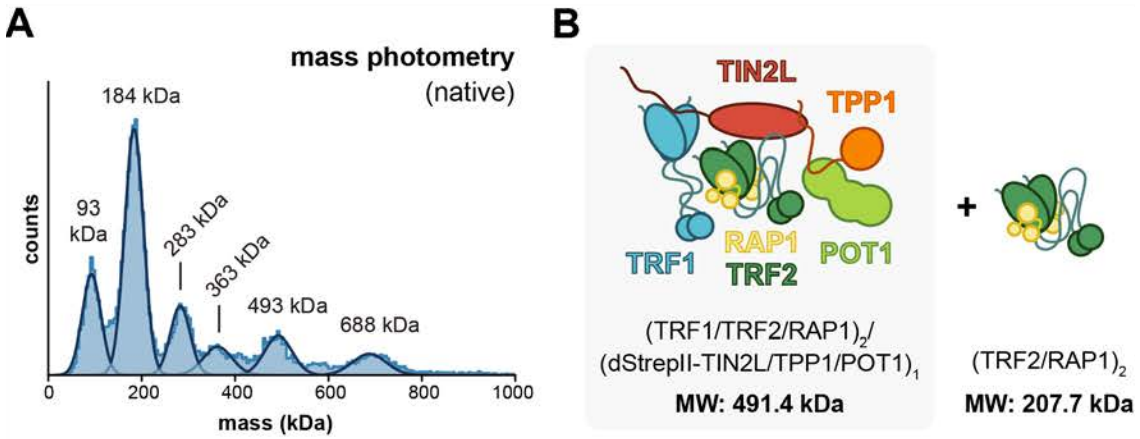

Figure 2.7. MW determination of the apo shelterin complex by mass photometry.

(A) MW distribution of a native (non-crosslinked) sample of SEC-purified apo shelterin^{TIN2L} complex. **(B)** Potential stoichiometry of the smaller (~490 kDa) and larger (~690 kDa) species of shelterin. The difference of ~200 kDa could be explained by an excess of TRF2 dimer plus its binding partner RAP1, as indicated by Western blot analyses.

Shelterin complex/subcomplex	Predicted mass (kDa)	Measured mass (kDa)
$(TRF1)_2$	100.50	93
dStrepII-TIN2L/TPP1/POT1	183.20	184
$(TRF1)_2$/dStrepII-TIN2L/TPP1/POT1	283.70	283
$(TRF1/TRF2/RAP1)_2$/dStrepII-TIN2L	362.23	363
$(TRF1/TRF2/RAP1)_2$/dStrepII-TIN2L/TPP1/POT1	491.40	493
$(TRF1/TRF2/RAP1/dStrepII\text{-}TIN2L/TPP1/POT1)_2$	674.60	688
$(TRF1)_2/(TRF2/RAP1)_4$/ dStrepII-TIN2L/TPP1/POT1	699.10	

Table 2.2. Potential stoichiometries of shelterin detected by mass photometry.

MW of species observed in mass photometry of SEC-purified apo shelterin^{TIN2L} complex, as shown in Figure 2.7A. Potential stoichiometries were calculated based on known protein-protein interactions (see section 2.1.1). The maximum mass error for this analysis is 6%.

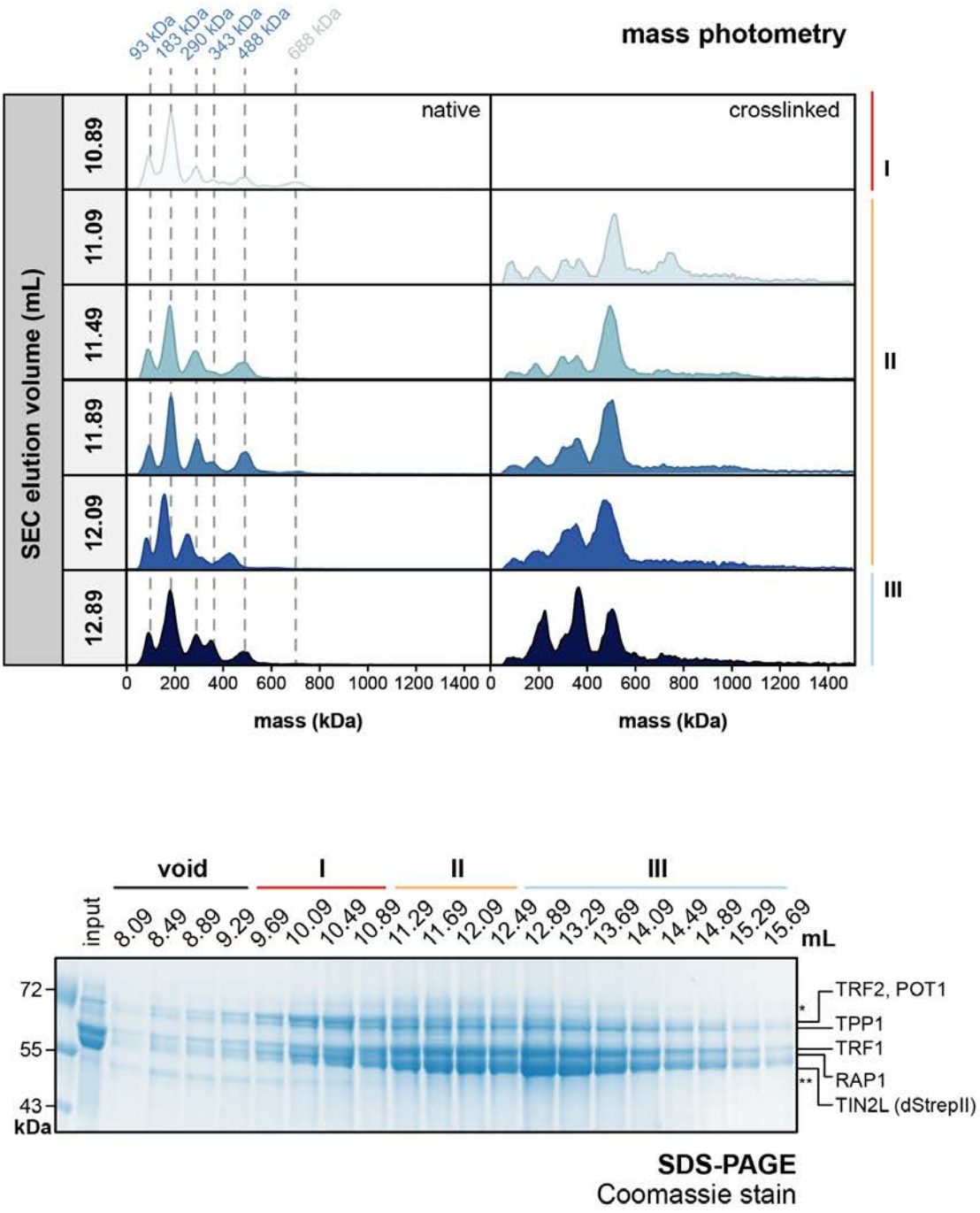

Figure 2.8. Mass photometry analysis of shelterin species across SEC elution.

The mean MW of shelterin^{TIN2L} species from SEC purification of the apo complex (as depicted in Figure 2.2) and the corresponding SDS-PAGE analysis.

2.3 Optimising the purification of the full shelterin complex

The two-step purification protocol described above produced heterogeneous populations of the full shelterin complex. I sought to adapt the purification strategy, in order to resolve the different species whilst maximising the yield of both the 490 kDa and the higher-order shelterin complex for downstream experiments.

2.3.1 Tandem SEC runs improved purity at the expense of yield

As SEC-purified species ***I*** and ***II*** were stable over additional SEC purifications. I repeated tandem SEC to improve the purity and yield of the 490 kDa species at the centre of SEC peak ***II*** (Figure 2.9A). After Strep-tag affinity chromatography, the most concentrated affinity fraction was injected into a Superose 6 Increase 10/300 SEC column (Figure 2.9B). Fractions from the middle of the SEC peak ***II*** were pooled and concentrated to 150 µL for a second round of SEC purification (Figure 2.9C). This additional SEC step improved sample homogeneity, judged by the lower proportions of the peak ***I*** and ***III*** (Figure 2.9D). However, as the various species overlapped during SEC elution, many of the elution fractions were discarded to ensure minimal carryover of the other species during the final SEC run. Therefore, the yield of the desired shelterin species at the end of two consecutive SEC runs is very low. Each SEC run also diluted the sample by ~ 5-fold, which required the elution fractions to be pooled and concentrated. Therefore, I explored other chromatographic methods to improve sample homogeneity after the Strep-tag affinity purification, and reserved SEC as the final polishing step.

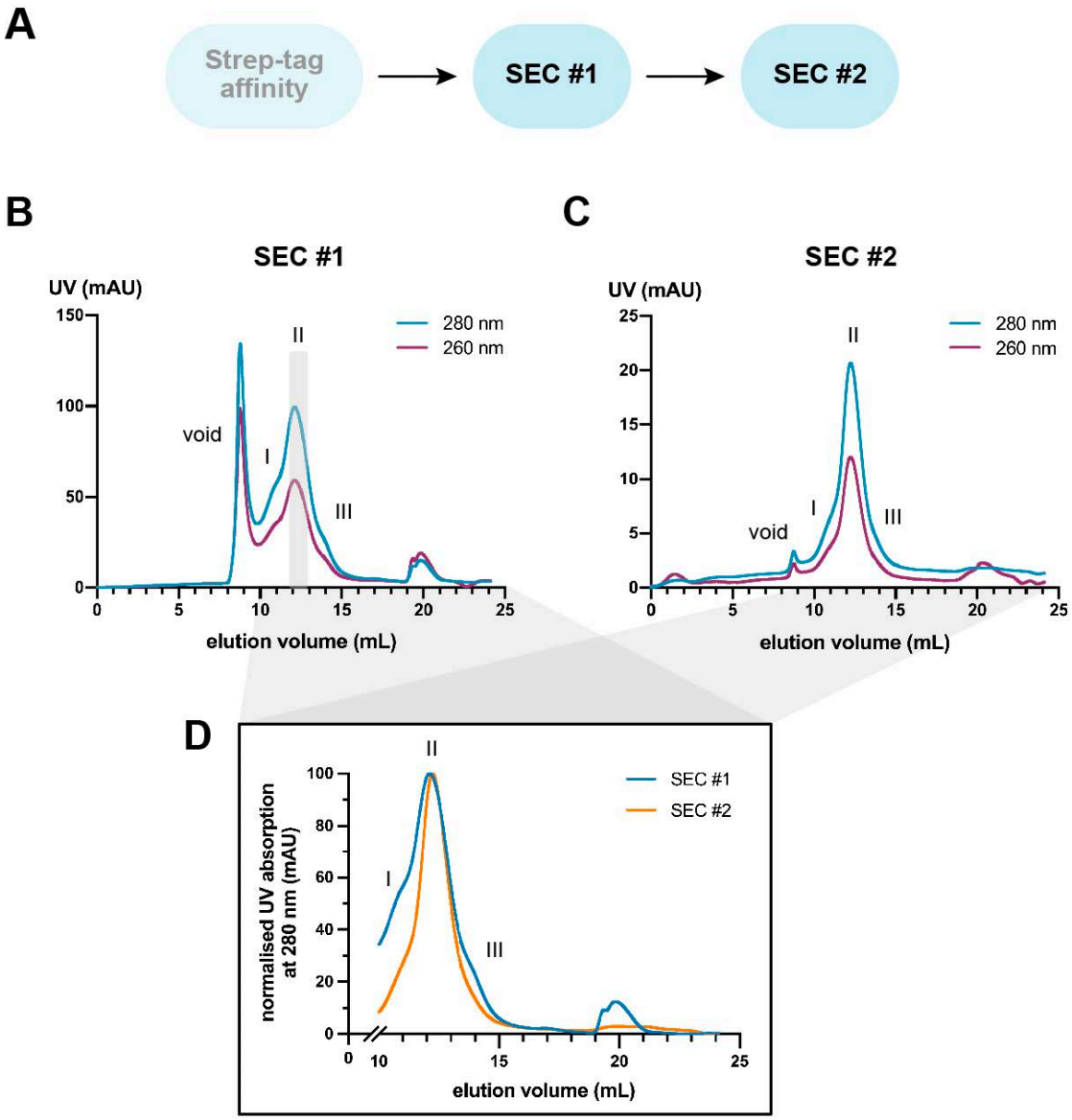

Figure 2.9. Sequential SEC improved enriched the 490 kDa shelterin complex.

(A) Schematic of the purification strategy. **(B)** SEC #1. 500 µL of affinity elution fraction of shelterin^{TIN2L} was injected into Superose 6 Increase 10/300 GL, eluting 0.3 mL/min in 200 µL fractions. Highlighted fractions were pool and concentrated to 150 µL. **(C)** SEC #2. Concentrated protein from the SEC #1 into the same column for a second round of SEC purification. **(D)** Overlay of the UV chromatograms from SEC #1 and #2, normalised to peak II to show the relative abundance of the three species.

2.3.2 Heparin chromatography resolved subcomplexes from the full shelterin complexes

Next, I explored the use of heparin chromatography, which is often used to purify DNA-binding proteins as heparin mimics the polyanionic structure of the DNA backbone. Elution fractions from Strep affinity were pooled and dialysed to drop the NaCl concentration to 300 mM to enable binding to the heparin HP column. This was the lowest salt concentration at which the recombinant apo shelterin complexes remained soluble over the course of the purification.

To begin with, a linear increase in the percentage of a high-salt buffer (2 M NaCl) was used to determine the salt concentrations at which the various shelterin species eluted from the column (Figure 2.10A and B). For shelterin^{TIN2L}, the flowthrough fractions consisted of dStrepII-TIN2L, TPP1, and POT1 (Figure 2.10C), and likely DNA contaminants due to the high 260 nm/280 nm UV absorbance ratio (A260/A280 = 1.29 at ~5 mL elution volume in Figure 2.10B). Affinity-purified shelterin^{TIN2L} and shelterin^{TIN2S} complexes eluted from heparin columns at comparable NaCl concentrations (Figure 2.10B and Figure 2.11A, respectively). Three species of shelterin complexes eluted during shallow linear [NaCl] gradients from 300 mM to ~ 1 M, over 20 CV (Figure 2.10B and Figure 2.11A). Step increases in NaCl concentration during elution was able to concentrate the various species during heparin elution (Figure 2.11B). Notably, species ***A*** and ***B*** contained all six shelterin subunits, whereas species ***C*** consisted of mainly TRF1 and dStrepII-TIN2 (Figure 2.10C, and Figure 2.11C and D). Fractions spanning the elution of species ***B*** were pooled, concentrated, and injected into a Superose 6 Increase 10/300 column to analyse the number and abundance of shelterin complexes present (Figure 2.10D). The SEC elution profile showed three species were present at the same elution volume as previously observed. However, the abundance of the trailing shoulder decreased markedly as the excess TIN2L/TPP1/POT1 and TRF1/TIN2L subcomplexes were removed during heparin chromatography in the flowthrough and in peak ***C*** (Figure 2.10D and E). The two other predominant species were the same as in previous SEC runs; both contained all the shelterin subunits, with the front shoulder being enriched in TRF2 and RAP1 (Figure 2.10E).

(figure on the next page)

Figure 2.10. Purification of apo shelterin^{TIN2L} using heparin chromatography.

(A) Purification strategy for the full shelterin^{TIN2L} complex. **(B)** UV chromatogram from the heparin chromatography step. Fractions from the Strep affinity elution were dialysed into the low-salt buffer (300 mM NaCl) prior to loading onto a Heparin HP 5 mL column, and eluted with a linear gradient from 300 mM to 2 M NaCl. **(C)** SDS-PAGE analysis of the flowthrough and elution fractions from heparin chromatography. **(D)** SEC analysis of peak ***B*** from heparin elution and **I** the corresponding SDS-PAGE analysis showed two predominant species (***I*** and ***II***).

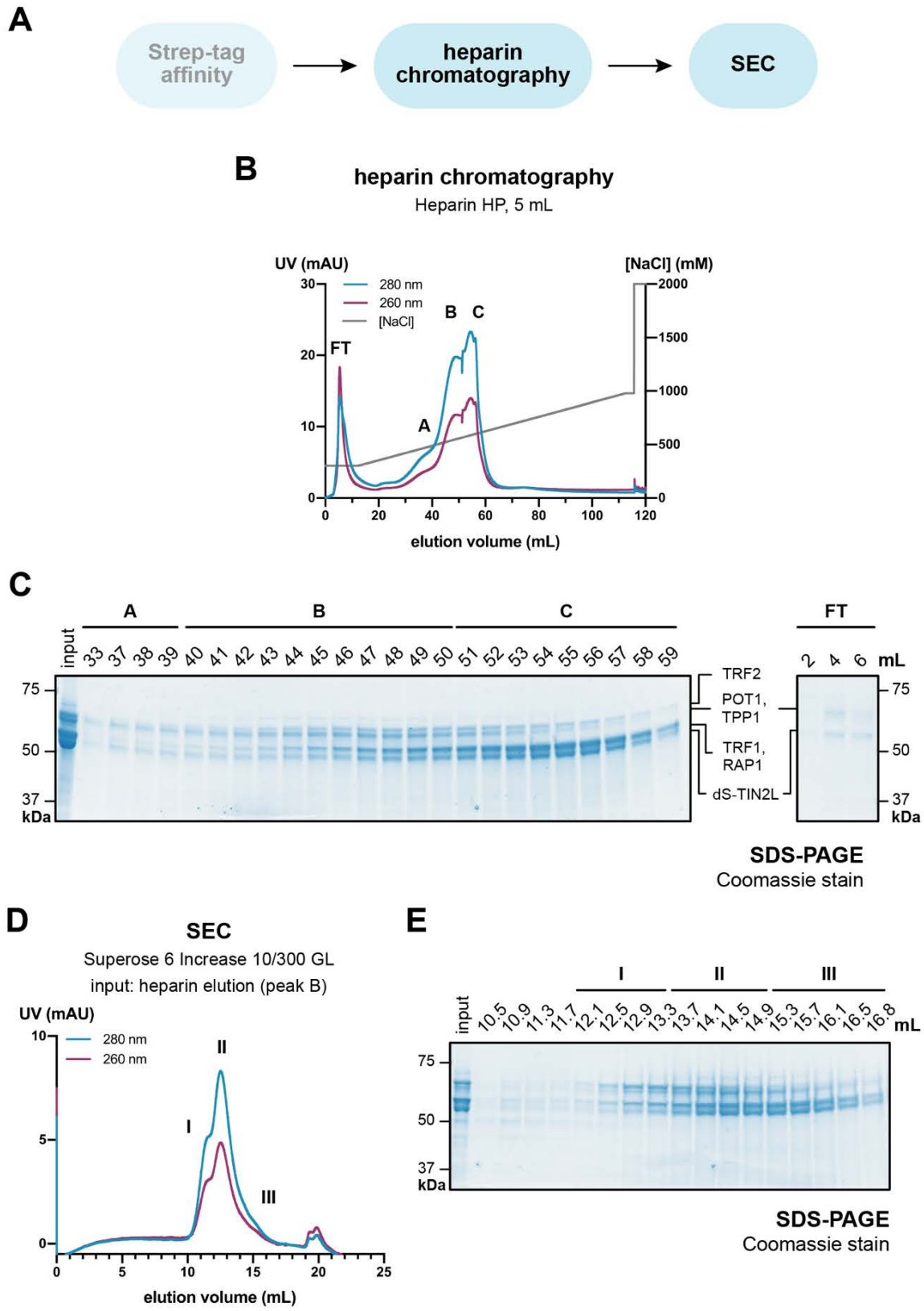

Figure 2.10. Purification of apo shelterin^{TIN2L} using heparin chromatography.

(legend on the previous page)

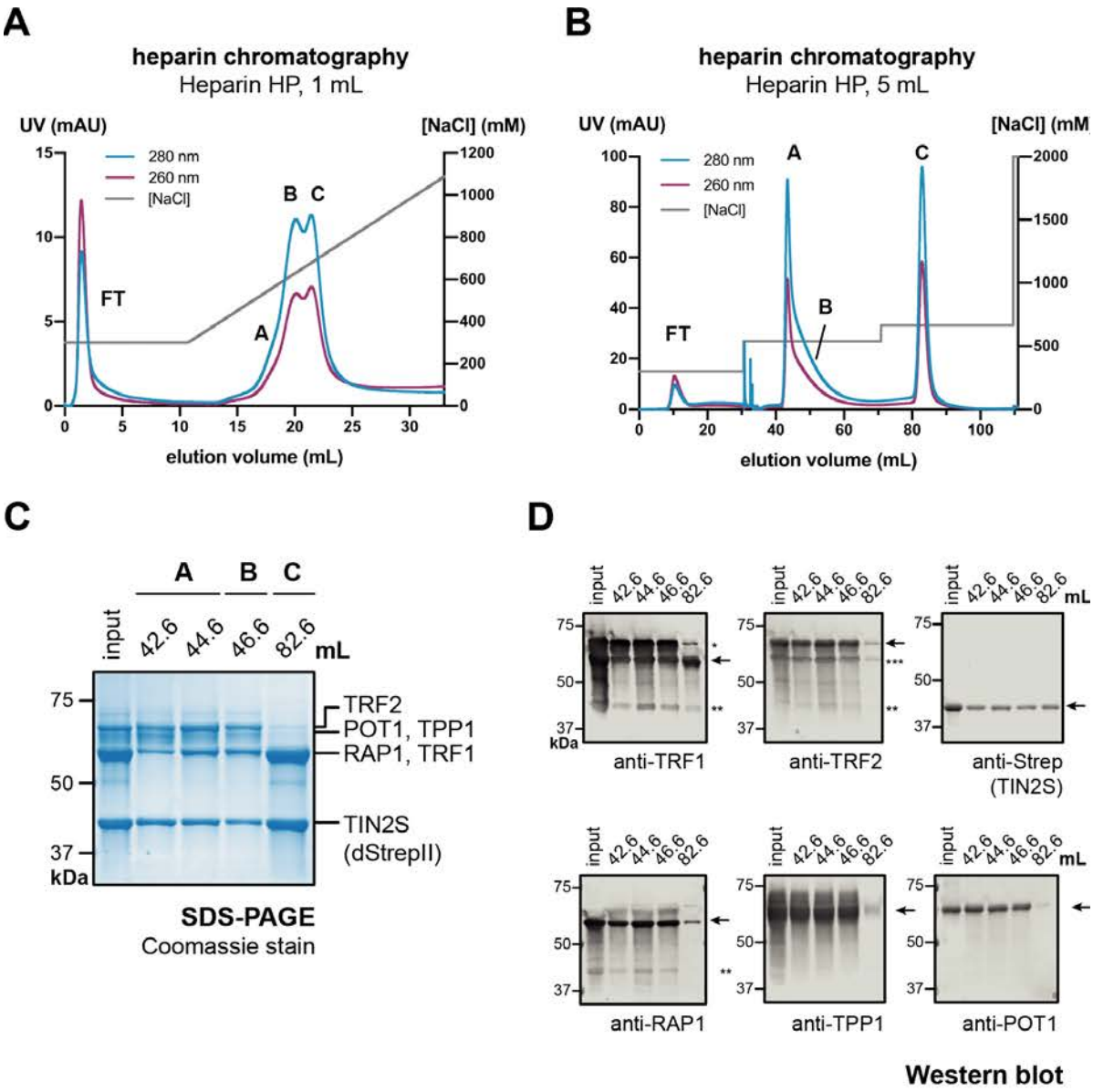

Figure 2.11. Purification of apo shelterin^{TIN2S} using heparin chromatography.

UV chromatograms from the heparin chromatography step eluted with **(A)** a linear [NaCl] gradient, or **(B)** a step increase in [NaCl]. Inputs were fractions from a Strep affinity purification of shelterin^{TIN2S} were pooled and dialysed into the low-salt buffer A (300 mM NaCl) prior to loading onto a 1- or 5-mL Heparin HP column. **(C)** SDS-PAGE and **(D)** Western blot analysis of the input and elution fractions from the step elution shown in (B). Arrows point to the intended target of the antibody when multiple bands were detected, and * indicates the non-specific binding of the anti-TRF1 antibody to its paralogue TRF2.

2.3.3 Resolving shelterin complexes using tandem affinity chromatography

Thus far, adding a heparin chromatography step in between the affinity and the final polishing SEC reduced the heterogeneity of the full shelterin complex by removing the majority of subcomplexes (Figure 2.10E). However, the larger-MW species remained. To explore whether different species of the full shelterin complex differ by the number of TRF2 as suggested by previous Western blotting analyses (Figure 2.2B), a TEV-cleavable deca-histidine (His_{10}) tag was cloned onto the N-terminus of TRF2 by Dr Oviya Inian. I tested the purification of dual-tagged shelterin constructs using tandem affinity chromatography (Figure 2.12A). The expected MW of the minimal dual-tagged shelterin^{TIN2L} complex is 495.9 kDa (Figure 2.12B). If the different shelterin species have different number of His-tagged TRF2 subunits, they will have different affinities for the Ni-NTA resin and may be eluted with different concentrations of imidazole.

For the detailed purification protocol, see section 6.3.3. Briefly, the shelterin^{TIN2L} complexes were first extracted from insect cell lysate using Strep-tag affinity purification as described previously (see section 2.1.2). Bound proteins were washed and eluted using a buffer containing 7 mM desthiobiotin and no imidazole. Next, the elution fractions were pooled for further purification by ion metal chromatography (IMAC) using a HisTrap HP 5 mL column. Two separate peaks were observed during a linear gradient elution at ~15% (peak ***A***) and 38% B (peak ***B***) (0 – 200 mM imidazole gradient, Figure 2.12C). Mass photometry analysis showed that peak ***A*** mainly contained the 490 kDa complex corresponding to the minimal (TRF1/His_{10}-TRF2/RAP1)$_2$/(dStrepII-TIN2L/ TPP1/ POT1)$_1$ complex, and a larger species at ~640 kDa appeared in later elution fractions (peak ***B***, Figure 2.12E).

Next, I translated the optimised conditions to a larger scale nickel affinity purification and used step increases in imidazole concentration to elute bound proteins over small volumes, thereby concentrating the sample (Figure 2.13A). SDS-PAGE analysis indicated the two peaks contained all six shelterin subunits at different stoichiometries (Figure 2.13B). The most concentrated fraction from

each peak were further purified by SEC. The SEC elution profiles confirmed that complexes eluted from the two different imidazole steps contained distinct shelterin^{TIN2L} species (Figure 2.14A). The species of shelterin^{TIN2L} eluted at lower imidazole appeared to contain very little amount of TRF2 and RAP1 (Figure 2.14B and C, left panels), whereas the species eluted at higher imidazole appeared to have a sub-stoichiometric amount of TIN2, TPP1 and POT1 (Figure 2.14B and C, right panels). These findings indicate that recombinant shelterin^{TIN2L} complexes can exist with TRF2 and RAP1 subunits in different stoichiometries, which can be further optimised to produce more homogeneous shelterin^{TIN2L} species at large scale.

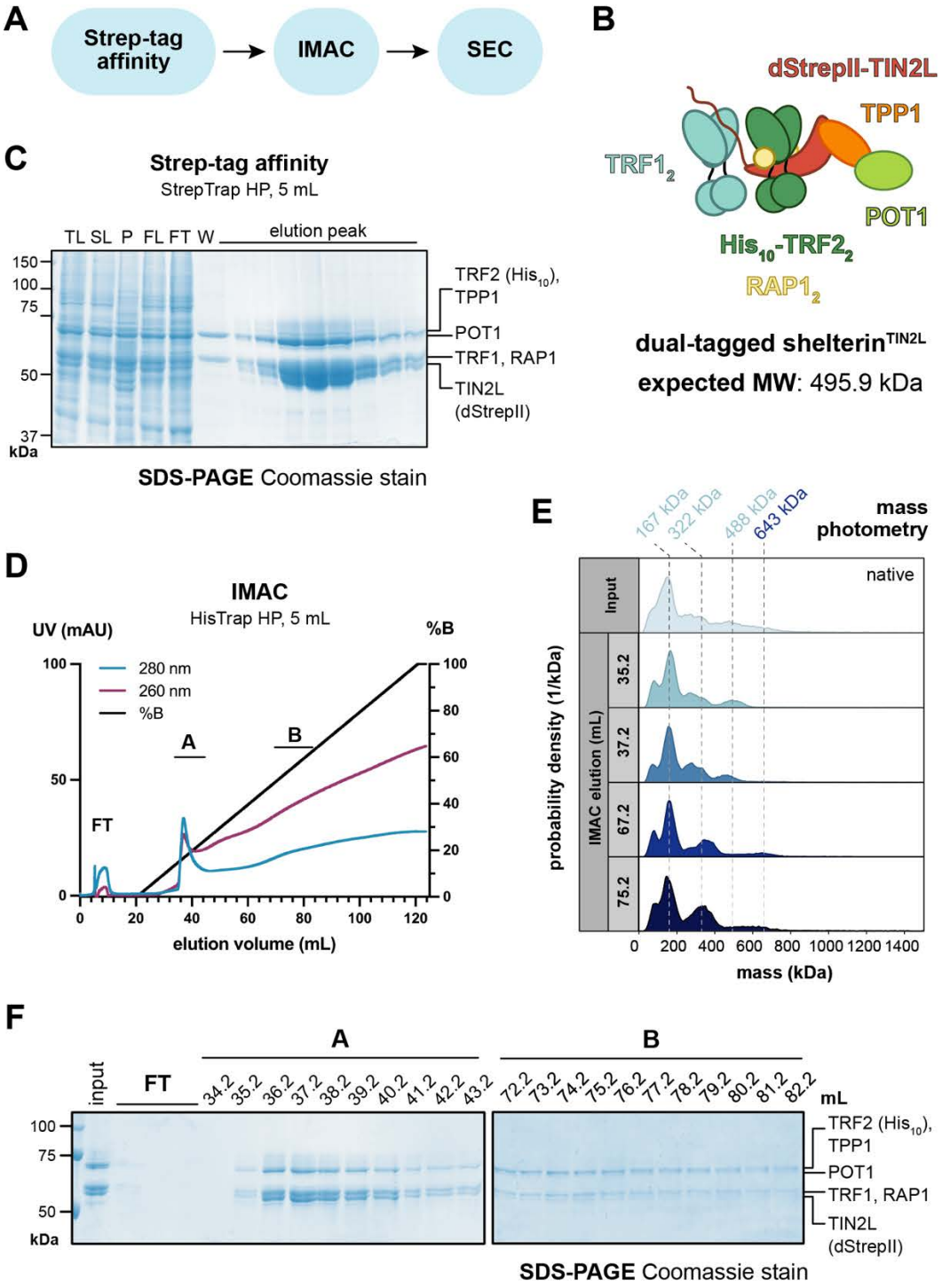

Figure 2.12. Purification of dual-tagged full shelterin complex.

(A) Purification strategy for the full shelterin^{TIN2L} complex. **(B)** Schematic of the minimal dual-tagged shelterin^{TIN2L} complex and its expected MW. **(C)** SDS-PAGE analysis of samples from the Strep-tag affinity purification. **(D)** UV chromatogram from the IMAC step. Protein from the strep elution was loaded into the HisTrap HP 5 mL column, washed, and eluted using a linear increase in imidazole concentration from 0 mM (0% B) to 200 mM (100% B) over 20 CV. **I** Mass photometry analysis of the input and elution fractions from the IMAC purification, as depicted in (D). **(F)** SDS-PAGE analysis of input, flowthrough (FT) elution fractions from IMAC.

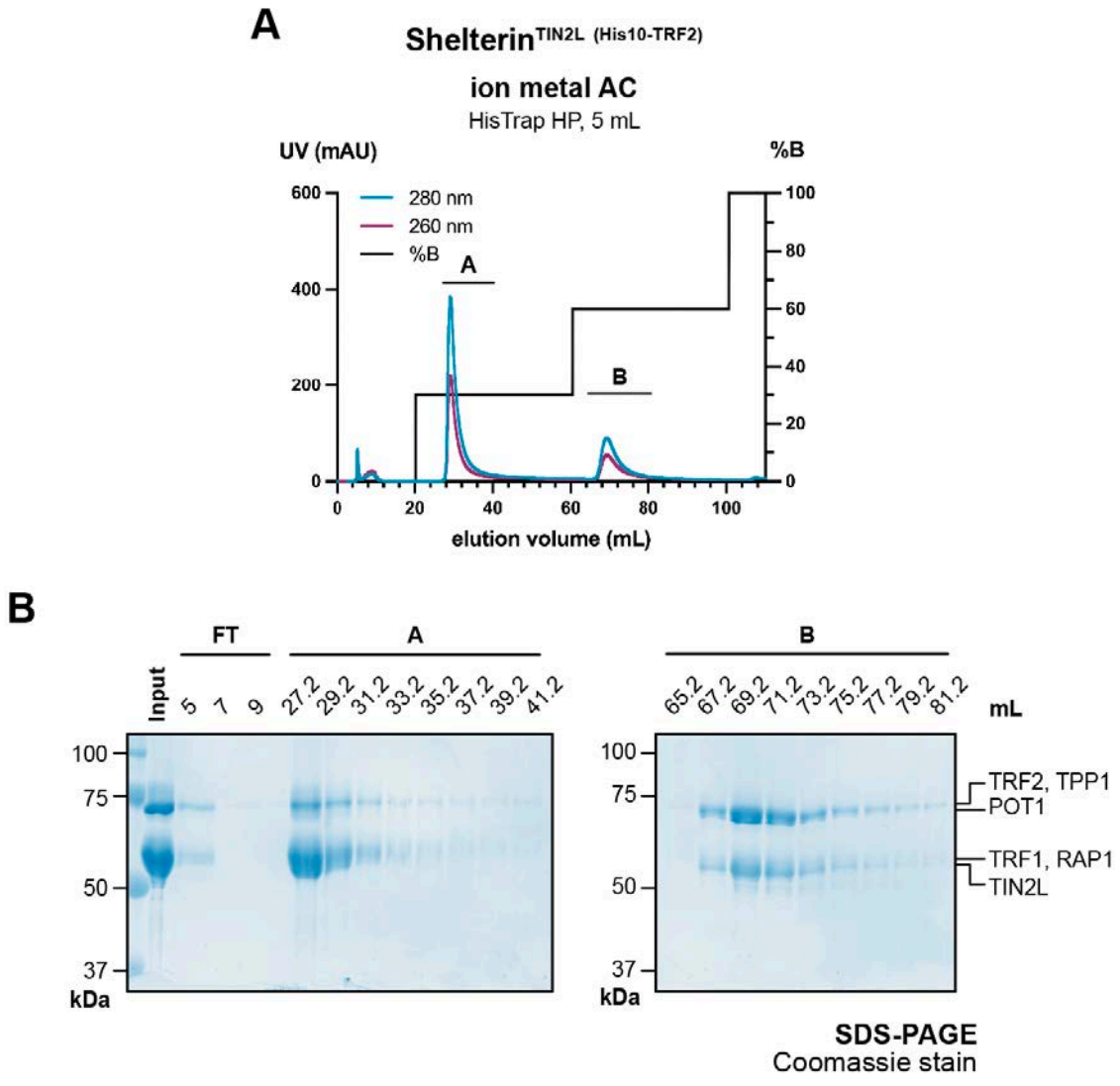

Figure 2.13. Step elution of dual-tagged shelterin from IMAC.

(A) Elution profile of dual-tagged shelterin^{TIN2L} from IMAC. Strep-tag AC elution fractions (8.84 mg, from Figure 2.12A) were loaded into a HisTrap HP 5 mL column, and step increases in imidazole were used to elute species in a small volume, thereby concentrating the samples. **(B)** SDS-PAGE analysis of the input, flowthrough (FT) and elution fractions from peak A and peak B, as depicted in (A).

(figure on the next page)

Figure 2.14. SEC analysis of purified dual-tagged shelterin complexes.

(A) SEC elution profiles of recombinant shelterin^{TIN2L} complexes purified by tandem affinity chromatography. Fractions from the HisTrap elution peaks (peak A and B in Figure 2.13) were centrifuged at 15,000 x g for 5 min, and 500 µL of the supernatant were injected onto a Superose 6 Increase 10/300 column and eluted at 0.2 mL/min. **(B)** SDS-PAGE and **(C)** Western blot analysis accompanying the elution profiles from (A).

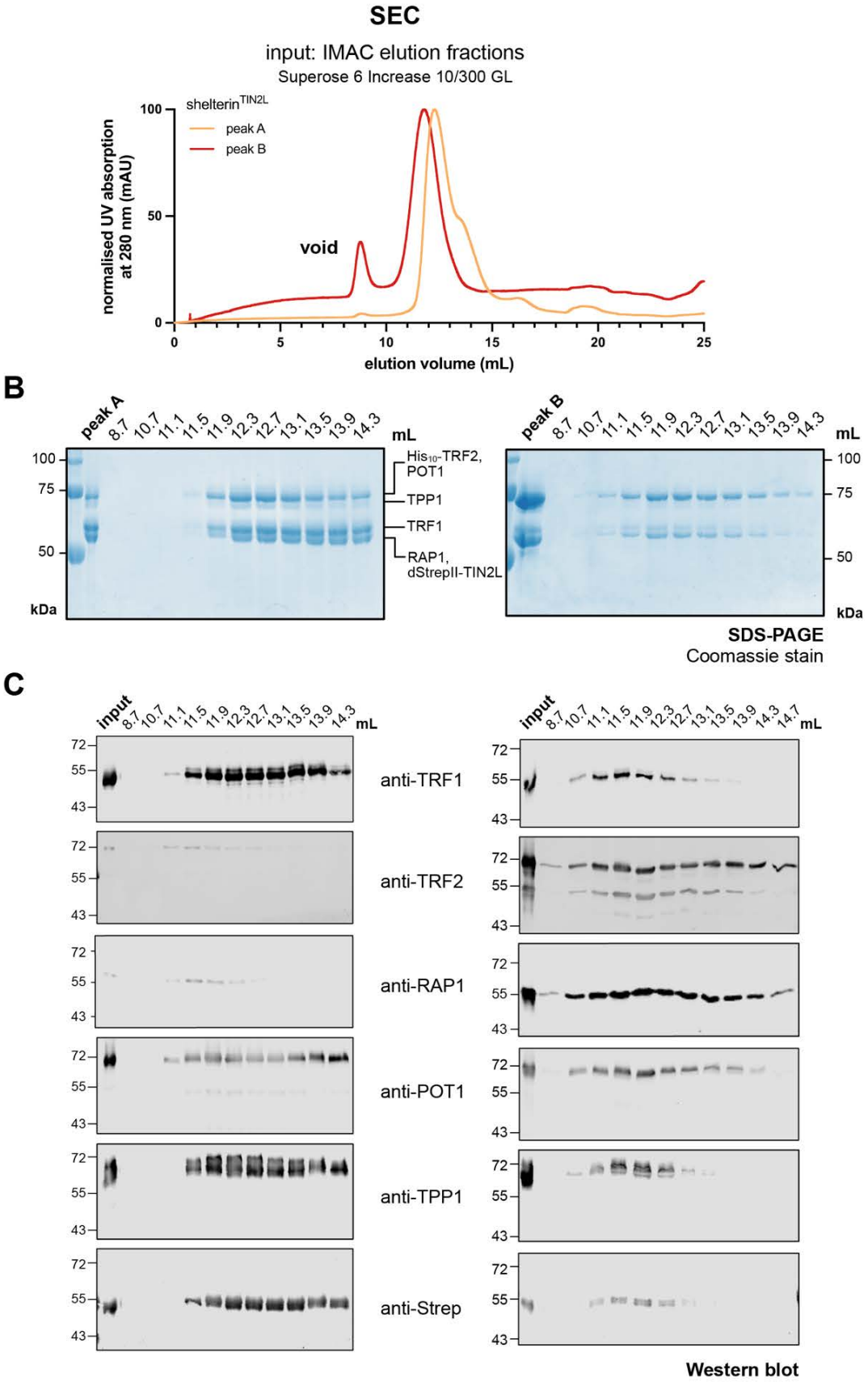

Figure 2.14. SEC analysis of purified dual-tagged shelterin complexes.

(legend on the previous page)

2.4 Shelterin binds telomeric DNA with nanomolar affinity

To check the ability of recombinant shelterin complexes to bind telomeric DNA, the affinity-purified full shelterin^{TIN2L} complex was incubated with 0.5 μM of a minimal telomeric sequence and analysed by electrophoretic mobility shift assay (EMSA) (Figure 2.15). The minimal telomeric sequence used for the full shelterin complex (referred to as teloDNA1) consists of 4.5 dsDNA TAGGGT repeats for the Myb domains from one dimer of TRF1 and one dimer of TRF2 (Court, et al., 2005), followed by ATC-5', which has been reported to form the ends of ~80% of C-rich strand at human telomeres (Sfeir, et al., 2005; Palm, et al., 2009). The ss 3'-overhang formed by the $(GGTTAG)_2$ sequence provides the optimal binding site to engage the OB-folds from one POT1 subunit (Lei, et al., 2004).

Increasing concentrations of affinity-purified shelterin^{TIN2L} bound to teloDNA1 to form a stable shelterin^{TIN2L}-DNA complex (Figure 2.15B). Traces of a shelterin-DNA complex first appeared at 0.25:1 molar ratio of shelterin:DNA (band ***III***, Figure 2.15B). This band likely consisted of shelterin subcomplexes present in the affinity-purified sample, which was enriched in the DNA-binding subunits TRF1 and POT1 (Figure 2.2B), and due to dissociation of the shelterin complex at low protein concentrations. All the DNA was engaged with equimolar amounts of shelterin and teloDNA1 and forming two species (bands ***II*** and ***III***, Figure 2.15B). Band ***II*** was less distinct than band ***III***, and likely contained both the minimal and higher-order shelterin^{TIN2L} complexes, which could be observed until it was in 5-fold molar excess over the telomeric DNA (Figure 2.15B). Finally, another species appeared at higher concentrations of shelterin^{TIN2L} (> 4-fold molar excess of protein over DNA), until it was the only band observed when shelterin^{TIN2L} was 10-fold molar excess over the DNA (band ***I***, Figure 2.15B). This is possibly due to non-specific binding of shelterin subunits to the non-telomeric region at the 5' end of the DNA oligonucleotide (Figure 2.15A), the purpose of which was to ensure correct annealing of the telomeric repeats.

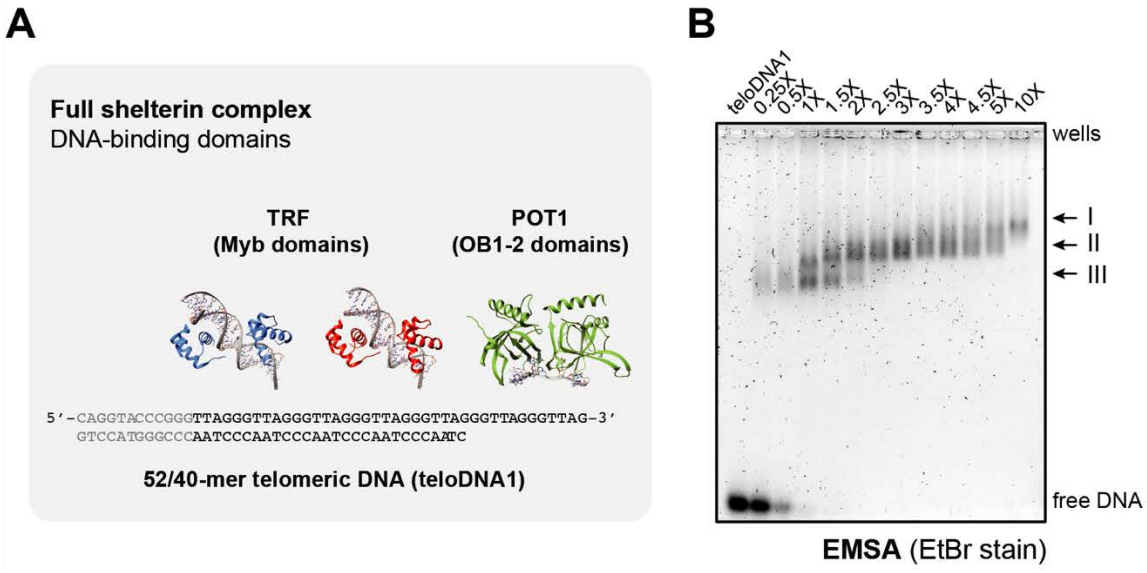

Figure 2.15. Recombinant full shelterin complexes bind to telomeric DNA.

(A) Schematic of the minimal telomeric DNA, teloDNA1, and DNA binding domains in the shelterin complex. TeloDNA1 consists of 4.5 ds telomeric repeats and 2 telomeric repeats in the 3' ss overhang. The two Myb domains from each TRF protein (TRF1 and TRF2; PDBs 1W0T and 1W0U, respectively) bind to the dsDNA, and the OB-folds 1 and 2 from POT1 (PDB: 1XJV) engages the 3' ssDNA overhang. **(B)** EMSA analysis of the DNA binding activity of affinity-purified shelterin. A serial dilution of apo shelterin^{TIN2L} was incubated with 0.5 μM teloDNA1 for 30 min on ice and subjected to electrophoresis in an 1.2% agarose 0.5X TBE gel.

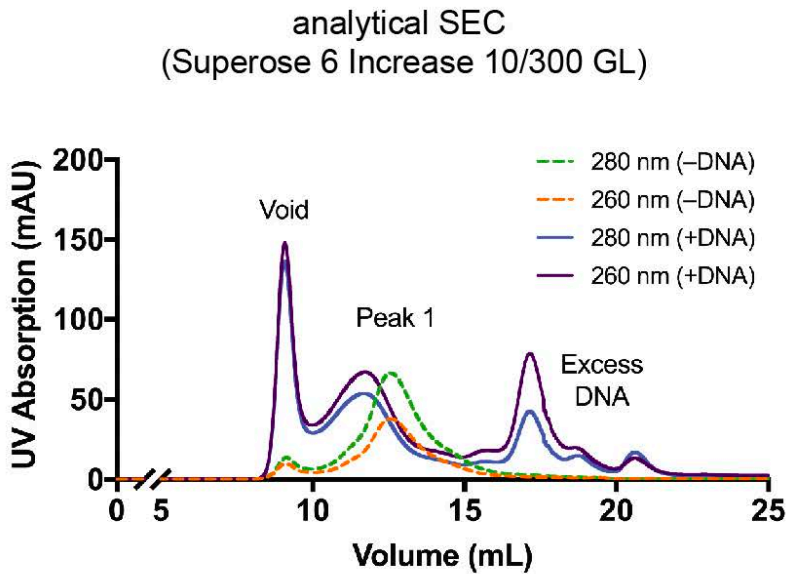

Figure 2.16. DNA-bound full shelterin complexes remain stable over SEC.

SEC elution profile of full shelterin^{TIN2L} incubated with buffer (dotted lines) or 1.2-fold molar excess of a minimal model telomeric sequence teloDNA1 (solid lines). Samples were subjected to analytical SEC and resolved using a Superose 6 Increase 10/300 GL column eluting at 0.5 mL/min.

Next, I determined the binding kinetics of recombinant shelterin^{TIN2L} complex to telomeric DNA using switchSENSE® technology. SwitchSENSE is a novel biophysical approach that uses biosensor technology to measure binding events on a chip (Rant, et al., 2004; Kaiser & Rant, 2010; Knezevic, et al., 2012; Langer, et al., 2015). See section 6.5.4.2 for details on the DNA sequences, data acquisition and analysis. In brief, the biochip contained a gold surface functionalised with 48-mer ssDNA nanolevers carrying either a red or a green fluorophore. The minimal teloDNA1 and the non-telomeric equivalent (mutDNA1) were hybridised to red and green fluorophore-labelled DNA nanolevers, respectively (Figure 2.17). A static electric potential was applied to the gold surface to keep the DNA nanolevers upright, away from the fluorescence-quenching gold surface. Binding of molecules was detected in real-time through changes on fluorescence emission. A range of concentrations of SEC-purified minimal shelterin^{TIN2L} complex was passed over the DNA nanolevers during the association phase (63 to 1000 nM), and buffer was injected during the dissociation phase (only conducted at 1 µM of shelterin).

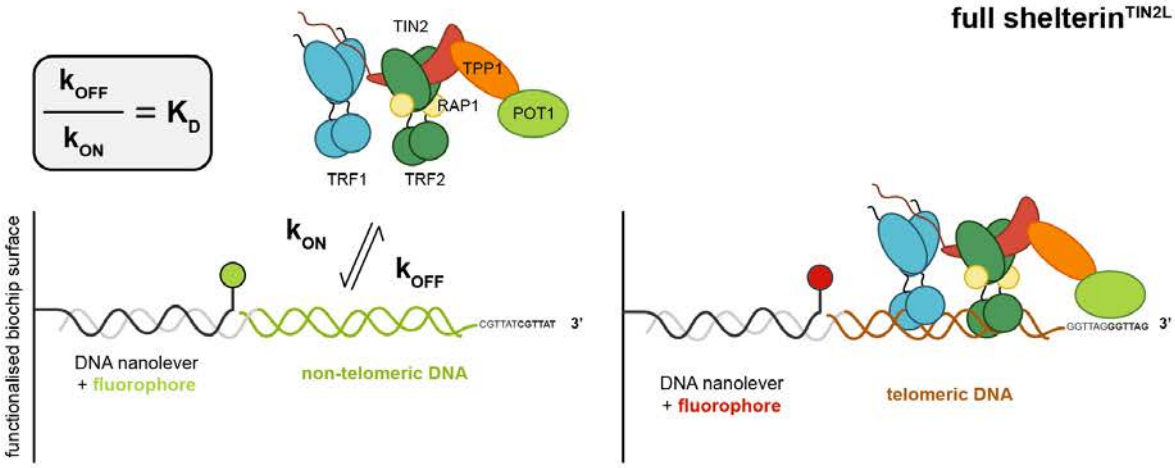

Figure 2.17. SwitchSENSE analysis of the affinity of the full shelterin complex for telomeric and non-telomeric DNA.

Schematic of the switchSENSE experimental setup to measure the binding kinetics (association and dissociation constants, k_{ON} and k_{OFF}, respectively) and to determine the affinity (equilibrium dissociation constant, K_D) of the full shelterin^{TIN2L} complex for a minimal telomeric target sequence (teloDNA1, red) and a non-telomeric DNA control (green).

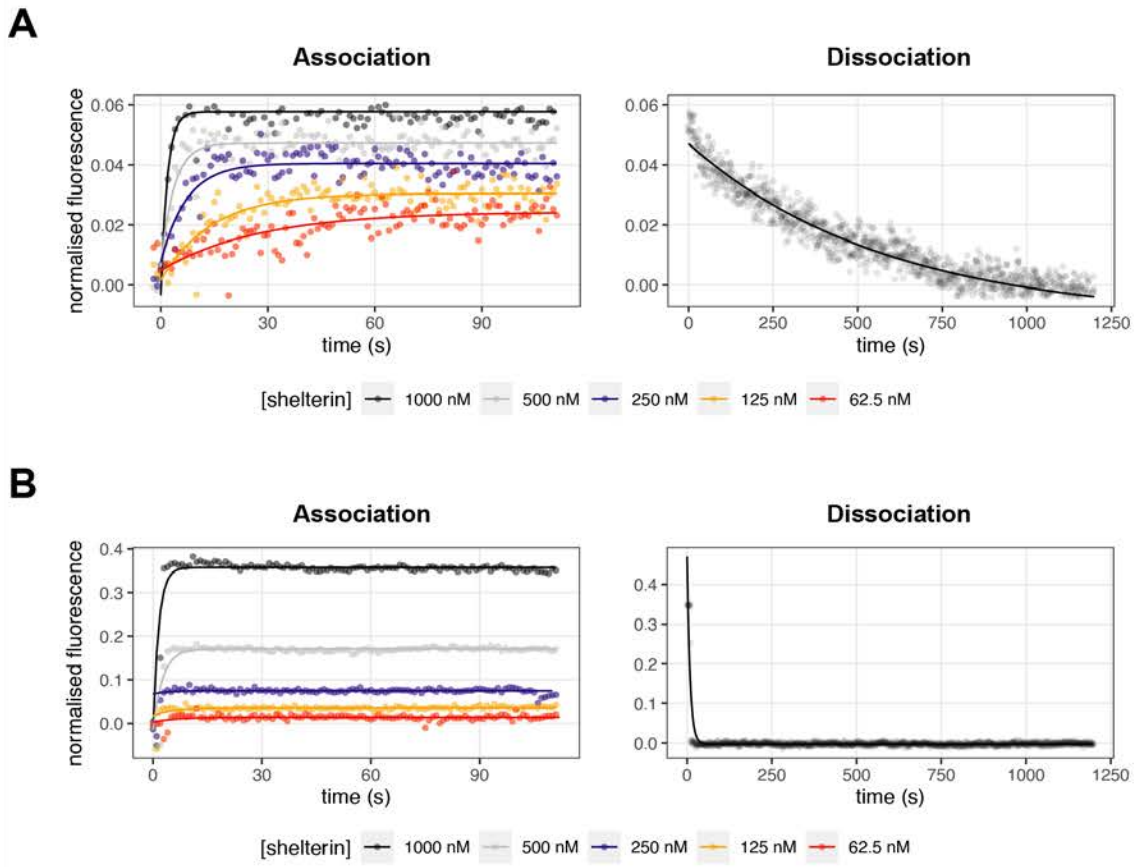

Figure 2.18. Time-resolved measurements of association and dissociation of shelterin from telomeric and non-telomeric DNA.

Association and dissociation curves from shelterin binding to **(A)** telomeric and **(B)** non-telomeric DNA from binding kinetic studies using switchSENSE. Various concentrations of SEC-purified shelterin^{TIN2L} (0, 62.5, 125, 250, 500 and 1000 nM) were passed over an electrode with both telomeric DNA and non-telomeric DNA. The change in fluorescence intensity was monitored over time. All six protein concentrations were used to determine k_{ON} (unit: M^{-1} s^{-1}), and one protein concentration (1000 nM) was used to determine k_{OFF} (unit: s^{-1}). The points were with single-exponential functions simultaneously (global fit) using the Langmuir binding model (1:1 model).

The shelterin^{TIN2L} complex bound to the telomeric and non-telomeric DNA tethered to the nanolever with nanomolar affinity (K_D = 3.11 ± 0.21 nM vs. 211 ± 12 nM, respectively Figure 2.18). These K_D values represent the upper limits of the affinity values, given the NaCl concentration of 300 mM in the buffer likely affected the rate of association and increased the rate of dissociation of shelterin from DNA. However, the high salt concentration was necessary for the apo shelterin^{TIN2L} complex to remain soluble during the assay. The binding of shelterin to both DNA sequences enhanced the fluorescence intensity in a concentration-

dependent manner, and the association curves showed single-exponential behaviour expected for one-to-one interactions (Figure 2.18A). Global fitting of associations yielded similar k_{ON} values for telomeric ($5.53 \pm 0.34 \times 10^5$ M^{-1} s^{-1}) and non-telomeric DNA ($5.25 \pm 0.25 \times 10^5$ M^{-1} s^{-1}). During the dissociation phase, sample buffer was flushed through the channel and fluorescence emission decreased in both the green and red channels, indicating that shelterin complexes disengaged from both types of DNA (Figure 2.18A and B). Notably, the fluorescence signal fell back to baseline within 30 s for the non-telomeric control, resulting in a high k_{OFF} value of $1.16 \pm 0.03 \times 10^{-1}$ s^{-1}. On the other hand, shelterin disengaged from the telomeric DNA in a gradual manner that lasted throughout the dissociation phase ($k_{OFF} = 1.72 \pm 0.04 \times 10^{-3}$ s^{-1}). This is due to the sequence-specific recognition of the Myb domains of TRF1 and TRF2 and the OB-folds of POT1, which allowed shelterin to bind stably to the telomeric sequence compared to the non-telomeric control. This difference of 67-fold in k_{OFF} explains the large difference observed for the binding affinity of shelterin to telomeric and non-telomeric DNA (Figure 2.19).

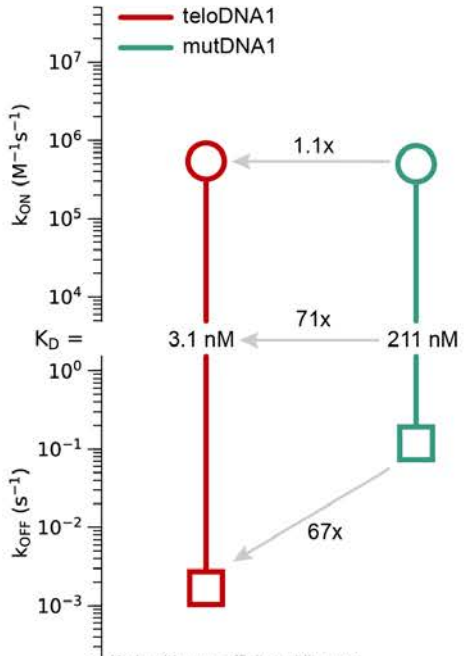

Figure 2.19. Rate map of shelterin binding to telomeric vs. non-telomeric DNA.

The rate map plots the rate constants for association (k_{ON}) and dissociation (k_{OFF}) events and the equilibrium constant for dissociation (K_D) (see Figure 2.18). Made with www.affinity-avidity.com.

2.5 Conclusions

In this chapter, I described the expression and purification of the full shelterin complex (with either the long or short isoform of TIN2) with a high yield and purity, making characterisations and functional studies more tractable. Expression of all six shelterin subunits in insect cells using a single baculovirus produced recombinant shelterin complexes that were heterogeneous in composition. The various shelterin species were partially resolved with a two-step protocol using strep-tag affinity and SEC. Adding heparin chromatography before the final SEC polishing step improved sample homogeneity, and when coupled with rapid screening using mass photometry, elution fractions containing the highest abundance of desired shelterin species could be selected for downstream assays.

Immobilising dStrepII-TIN2 during affinity chromatography extracted two shelterin complexes that contained all six shelterin subunits, with the predominant species having a MW of ~490 kDa, as determined by SEC-MALS and mass photometry. This matches the expected MW of 480 and 490 kDa for shelterin^{TIN2S} and shelterin^{TIN2L}, respectively for the minimal complex containing (TRF1/TRF2/RAP1)$_2$/(dStrepII-TIN2/TPP/POT1)$_1$. Interestingly, other species with larger MW were observed during mass photometry analyses. Shelterin complexes at ~640 – 690 kDa were consistently observed under native conditions. SDS-PAGE and Western blotting showed both the minimal and higher-MW species contained six shelterin subunits, with a greater abundance of TRF2 and RAP1 in the higher-order species. Therefore, it is possible that the larger 690 kDa species contains another (TRF2/RAP1)$_2$ module. However, it is yet unclear how another TRF2 dimer may be recruited to the complex, given the two TRF binding sites on TIN2 would have been occupied by one TRF1 and one TRF2 homodimer. To further explore this avenue, I expressed and purified shelterin^{TIN2L} with dStrepII-TIN2L and a His$_{10}$-tagged TRF2. Preliminary tests indicated the presence of shelterin species with higher and lower abundances of TRF2. Further characterisations will be necessary to confirm this observation.

Finally, I studied the interaction between the full shelterin^{TIN2L} complex and a minimal telomeric sequence using EMSA and determined the K_d values for

shelterin binding to telomeric and non-telomeric DNA using switchSENSE. Under the conditions tested, shelterin can bind to both types of DNA with nanomolar affinity (K_D = 3.11 vs. 211 nM for telomeric and non-telomeric DNA, respectively). However, under the conditions tested, shelterin had a 71-fold higher affinity for telomeric DNA compared with the non-telomeric control. This difference can be mainly attributed to the 67-fold difference between the dissociation rate constants for these two DNA constructs (k_{OFF} = 1.16 x 10^{-1} s^{-1} vs 1.72 x 10^{-3} s^{-1} for telomeric and non-telomeric DNA, respectively). These results suggested that shelterin can bind to DNA in a non-sequence specific manner. However, once it has found the target telomeric sequence, it had a much longer residing time on telomeric DNA compared with non-telomeric DNA.

3 Biochemical and biophysical characterisation of shelterin subcomplexes

Previous reports have provided evidence for various subcomplexes of shelterin *in vivo*, which likely perform different functions including efficient and persistent telomere protection and regulating telomere length maintenance (Ye, et al., 2004a; Takai, et al., 2010). Therefore, I generated a range of shelterin subcomplexes, with the aim to probe the stoichiometry and functional differences between the full and subcomplexes of shelterin using biochemical and biophysical assays, and eventually to study their architecture using electron microscopy (EM).

In this chapter, I will focus on a subcomplex lacking the dsDNA-binding component TRF1, thereby referred to as shelterin (–TRF1) or simply (–TRF1). This subcomplex is of particular interest in this project because in a telomerase-positive context, TRF1 is (1) a negative regulator of telomere length maintenance, and (2) a substrate of a PARP enzyme, tankyrase, which is itself a positive regulator of telomere length (Smogorzewska, et al., 2000; Smith & de Lange, 2000). Another use for this subcomplex was to complement ongoing efforts to characterise the structure of the human shelterin complex, described in section 4.1.

Later in this chapter, I will also mention the expression and purification of other shelterin subcomplexes, including one lacking the dsDNA binder TRF2 and its interactor RAP1 (–TRF2/RAP1), and shelterin lacking the ssDNA-binding protein POT1 and its binding partner TPP1 (–TPP1/POT1). These subunits have also been shown to regulate telomere length maintenance, particularly the TPP1 and POT1 heterodimer that recruits and enhances telomerase processivity (see section 1.4.1). Therefore, these variations of shelterin are also of interest to tease apart the contributions of shelterin subunits to telomere length maintenance.

3.1 Overview of protein expression and purification

The various human shelterin subcomplexes were recombinantly expressed in Sf9 insect cells using a single baculovirus that produced the desired combination of full-length shelterin subunits. As with the full complex, the subcomplexes contain one of the two TIN2 isoforms, TIN2L or TIN2S. All shelterin subcomplexes were purified using the same two-step purification protocol as for the full shelterin complex (see section 2.1.2). Molar concentrations of Strep affinity and SEC-purified shelterin subcomplexes were quantified by spectrophotometry (described below; also see sections 6.2.3). The stoichiometries of the recombinant subcomplexes were inferred from the stoichiometry used for the full six-subunit shelterin complex $(TRF1/TRF2/RAP1)_2/(TIN2/TPP1/POT1)_1$ (see sections 1.6.1 and section 2.1.1)

3.2 Purification of the (–TRF1) subcomplex

The shelterin $(–TRF1)^{TIN2L}$ and $(–TRF1)^{TIN2S}$ subcomplexes contained a TEV-cleavable dStrepII tag on TIN2L or TIN2S, respectively. The affinity purification extracted all five shelterin subunits with a high yield of up to 17.5 mg per litre of insect cell culture (Figure 3.1A and B). The absence of TRF1 was confirmed by mass spectrometry (Appendix Table 7.2). In the two-step protocol, the peak and adjacent affinity elution fractions were further purified by SEC (Figure 3.1C). Both shelterin $(–TRF1)^{TIN2L}$ and $(–TRF1)^{TIN2S}$ subcomplexes produced similar SEC elution profiles, which typically featured 3 peaks after the void volume (~8.5 mL for Superose 6 Increase 10/300 GL, Figure 3.1C). Peaks ***I*** and ***II*** were not fully resolved (Figure 3.1C and D). Western blot analyses showed that both peak ***I*** and ***II*** contained the five subunits of the shelterin (–TRF1) subcomplex at varying abundances (Figure 3.1E). Notably, peak ***I*** was enriched in TRF2 and RAP1 (lane ***I***), whilst the peak ***II*** contained more TIN2L, TPP1, and POT1 subunits (lane ***II***).

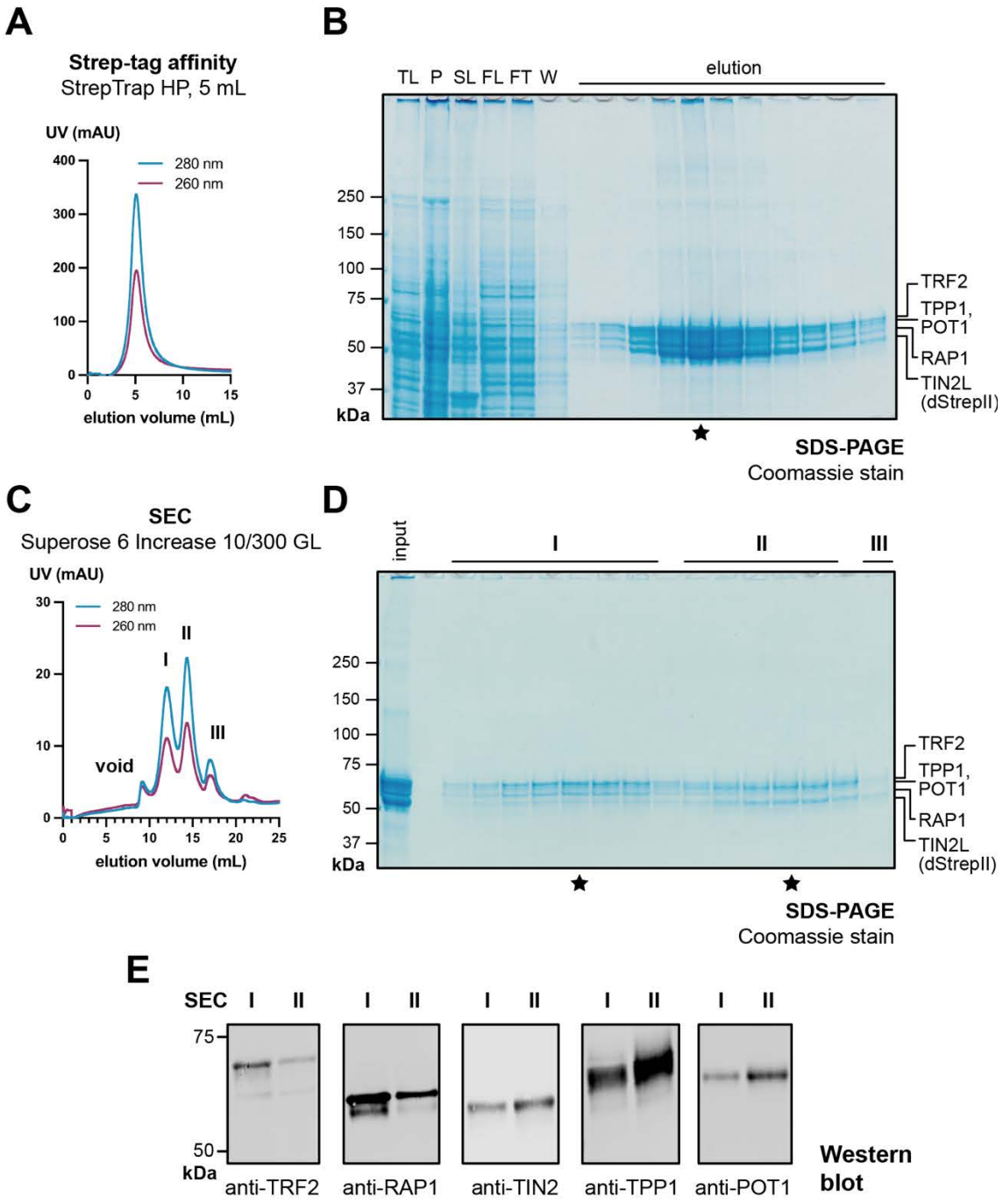

Figure 3.1. Two-step purification of the shelterin (–TRF1) subcomplex.

(A) Representative UV chromatogram from a Strep affinity purification of $(–TRF1)^{TIN2L}$. The fraction subjected to SEC purification is marked (★). **(B)** SDS-PAGE analyses of the lysates and elution fractions shown in (B). TL – total lysate, P – pellet, SL – soluble lysate, FL – filtered lysate, FT – flowthrough, W – wash. **(C)** Representative UV chromatogram from a SEC purification of $(–TRF1)^{TIN2L}$ using Superose 6 Increase 10/300 GL, eluting in 200 µL fractions. **(D)** SDS-PAGE and **I** Western blot analysis of eluates in SEC peak ***I***, ***II***, and ***III***, as shown in (C).

3.2.1 Determining the MW of recombinant shelterin (–TRF1)

To further characterise the different species of shelterin (–TRF1) obtained from affinity and SEC purifications, I determined the size and MW using SEC-MALS (section 3.2.1.1) and mass photometry (section 3.2.1.2).

3.2.1.1 SEC-MALS analysis of shelterin (–TRF1) complexes

In the absence of DNA (apo state), the recombinant shelterin $(–TRF1)^{TIN2L}$ complexes eluting in SEC peaks ***I*** and ***II*** were monodisperse, as indicated by polydispersity indices of close to 1.00 (Figure 3.2 and Table 3.1). SEC-MALS analysis indicated that species in SEC peak ***I*** had a MW of 374.2 ± 5.6 kDa (Table 3.1). This closely matched the expected MW of 390.8 kDa for the $(TRF2/RAP1)_2/(dStrepII\text{-}TIN2L/TPP1/POT1)_1$ complex. The species eluting across peak ***II*** had a MW of 150.6 ± 4.1 kDa (Table 3.1). Previous western blot analysis indicated that SEC peak ***II*** contained mostly dStrepII-TIN2L, TPP1 and POT1 subunits (Figure 3.1F), indicating that this species was a dStrepII-TIN2L/TPP1/POT1 subcomplex (1:1:1 ratio, expected MW: 183.2 kDa). The weight-averaged mean square radius for species in peaks ***I*** and ***II*** were 16.8 ± 3.7 nm and 14.6 ± 6.9 nm, respectively (Table 3.1). These sizes were close to the detection limit of the MALS detector, resulting in the large % errors.

To study the stability of recombinant apo (–TRF1) complexes, affinity-purified $(–TRF1)^{TIN2L}$ was diluted to 5, 2.5, 1.25 and 0.625 µM and subjected to an analytical SEC analysis (Figure 3.3). Notably, the elution volume of peak ***I*** shifted to the right with decreasing protein concentration, which suggests that species in this peak exist in a concentration-dependent equilibrium (Figure 3.3, Table 3.2). On the other hand, the elution volume of peak ***II*** remained constant upon dilution of the affinity-purified sample, indicating this as a stable module within the shelterin (–TRF1) subcomplex (Figure 3.3, Table 3.2).

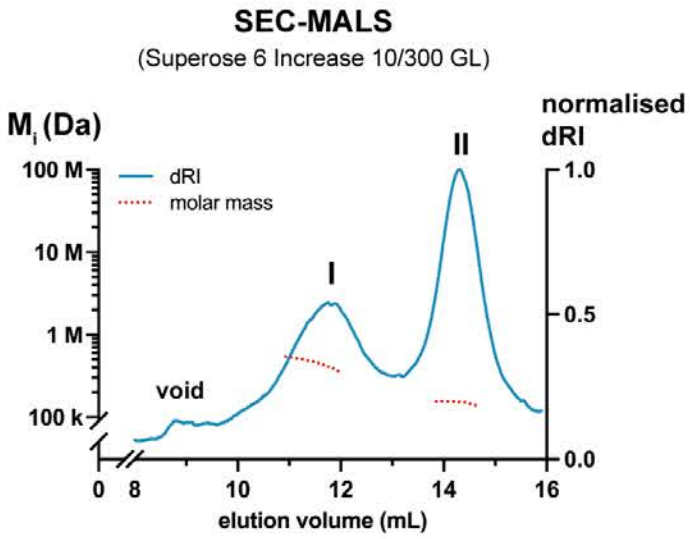

Figure 3.2. SEC-MALS analysis of the recombinant apo (–TRF1) complex.

MW distribution (dotted line) and normalised differential refractive index (dRI) were plotted for the SEC-MALS analysis of apo shelterin (–TRF1)^{TIN2L}. Affinity-purified (–TRF1)^{TIN2L} was further resolved using a Superose 6 Increase 10/300 SEC column connected to a MALS and RI detector (Wyatt).

	Shelterin (–TRF1)^{TIN2L}	
	Peak I	**Peak II**
M_W (kDa)	374.2 (± 5.60)	150.6 (± 4.70)
Polydispersity index	1.03 (± 0.02)	1.00 (± 0.04)
Radius, r_W (nm)	16.8 (± 3.70)	14.6 (± 6.90)

Table 3.1. SEC-MALS analysis of the recombinant apo (–TRF1) complex.

The weight average molar mass (M_W), polydispersity index, and weight-averaged mean square radius (r_w) from a SEC-MALS analysis of the shelterin (–TRF1) complexes as depicted in Figure 3.2. The polydispersity index is a ratio of M_W to M_n, where M_n is number average molar mass. The error values are indicated in brackets.

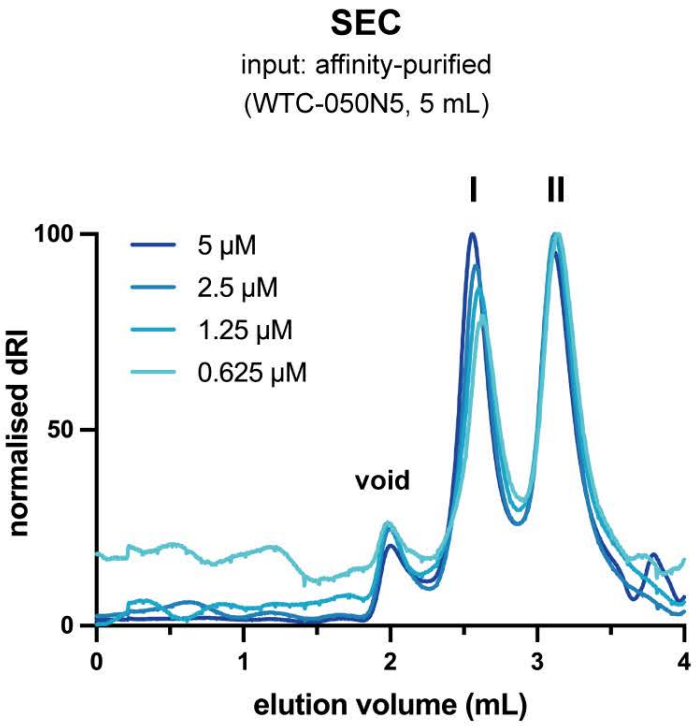

Figure 3.3. Concentration-dependent formation of apo (–TRF1) complexes.

Overlay of the analytical SEC profiles of a dilution series of shelterin (–TRF1)^{TIN2L}, highlighting the protein peaks. Affinity-purified shelterin (–TRF1)^{TIN2L} was diluted to 5, 2.5, 1.25, and 0.625 µM before each SEC run and resolved using a WTC-050N5 (5 mL) SEC column eluting at 0.2 mL/min. Data points were normalised to peak ***II,*** which had a constant elution volume across the dilution series.

(–TRF1)^{TIN2L} (µM)	elution volume (mL)	
	Peak I	Peak II
5	2.62	3.15
2.5	2.60	3.13
1.25	2.58	3.12
0.625	2.56	3.13

Table 3.2. Elution volumes of (–TRF1)^{TIN2L} species in SEC-MALS analysis.

Elution volumes of peaks ***I*** and ***II*** from analytical SEC analysis of (–TRF1)^{TIN2L}, as shown in the UV chromatograms in Figure 3.3.

3.2.1.2 MW determination using mass photometry

Mass photometry experiments were conducted in the same manner as previously described for the full complex (see sections 2.2.2 and 6.5.3.2). Most samples were measured both before (native) and after crosslinking with 0.075% or 0.1% glutaraldehyde in order to obtain the most accurate MW values and to capture less stable higher-order protein complexes within the sample. The conditions used for deriving possible stoichiometries of the observed MW of detected species are listed in section 2.1.1. Briefly, TRF2 forms stable homodimers, with each dimer recruiting two RAP1 proteins ($TRF2_2$:$RAP1_2$). Each TIN2 binds one TRF2 homodimer and up to one copy of the TPP1/POT1 heterodimer.

Five species of the shelterin (–TRF1) subcomplex were detected in SEC peak ***I*** by mass photometry analyses (Figure 3.4A). The species at ~390 kDa, referred to hereon as the minimal (–TRF1) complex, is consistent with a 2:2:1:1:1 stoichiometry of TRF2/RAP1/dStrepII-TIN2L/TPP1/POT1 (expected MW: 390.8 kDa, Figure 3.4A). The two larger species at ~550 and 600 kDa were typically present at a lower abundance when measured under native conditions (Figure 3.4A), possibly due to dissociation of the complexes during sample dilution. The extra ~ 160 to 200 kDa in the higher-order (–TRF1) complexes can be explained by different combinations of TRF2/RAP1 (Table 3.3). For example, an extra $(TRF2/RAP1)_2$ to the minimal (–TRF1) would give rise to a 598 kDa complex with $(TRF2/RAP1)_4$/$(dStrepII\text{-}TIN2L/TPP1/POT1)_1$, or an additional $TRF2_2/RAP1_1$ would produce a 553 kDa complex with $TRF2_4/RAP1_3$/$(dStrepII\text{-}TIN2L/TPP1/POT1)_1$. The MW of proteins in SEC peak ***II*** was determined to be 193 kDa (Figure 3.4B), which is close to the expected MW of 183.2 kDa for the dStrepII-TIN2L/TPP1/POT1 heterotrimer in a 1:1:1 ratio. There was a small amount of a 373 kDa complex also present due to the overlapping of species from peaks ***I*** and ***II*** during SEC purification (see Figure 3.1D).

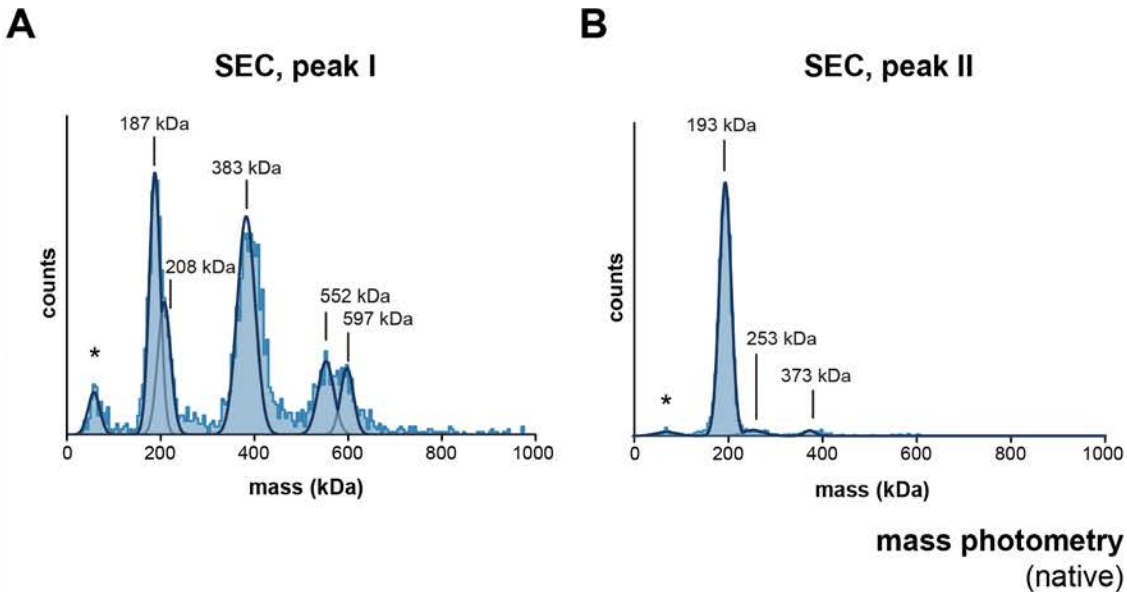

Figure 3.4. MW determination of the (–TRF1)^{TIN2L} complex by mass photometry.

Mass photometry measurement of native (non-crosslinked) samples of apo shelterin (–TRF1)^{TIN2L} subcomplex from **(A)** SEC peak ***I*** and **(B)** peak ***II***. The maximum mass error for this analysis is 1.1%.

Shelterin complex/subcomplex	Predicted mass (kDa)	Measured mass (kDa)
dStrepII-TIN2L/TPP1/POT1	183.20	187
$(TRF2/RAP1)_2$	207.70	208
$(TRF2/RAP1)_2$/dStrepII-TIN2L/TPP1/POT1	390.90	383
$(TRF2)_4/(RAP1)_3$/dStrepII-TIN2L/TPP1/POT1	554.34	552
$(TRF2/RAP1)_4$/dStrepII-TIN2L/TPP1/POT1	598.60	597

Table 3.3. Potential stoichiometries of (–TRF1)^{TIN2L} species from SEC peak *I*.

MW of species observed in mass photometry of shelterin (–TRF1)^{TIN2L} complex from SEC peak ***I***, as shown in Figure 3.4A). Potential stoichiometries were calculated based on known protein-protein interactions (see section 2.1.1). The error range was calculated based on the maximum mass error of 1.1%, determined from a mass calibration performed during this experiment.

To check whether the higher-order and minimal shelterin (–TRF1) in SEC peak ***I*** co-eluted across the entire peak, I pooled and concentrated the front and middle of peak ***I*** fractions from a SEC purification of (–TRF1)^{TIN2S} for SEC-MALS

and mass photometry analyses (Figure 3.5A). Shelterin (–TRF1)^{TIN2S} from middle of SEC peak ***I*** eluted as a single peak during SEC-MALS and had a MW of 380.0 kDa ± 17.0 kDa and polydispersity of 1.000 ± 0.06 (Figure 3.5B). This matched the 380.3 kDa expected of the minimal (–TRF1)^{TIN2S} complex with (TRF2/RAP1)$_2$/(dStrepII-TIN2S/TPP1/POT1)$_1$ (Figure 3.4C). When the same sample was analysed by mass photometry, two species of (–TRF1)^{TIN2S} were detected at 383 kDa and 563 kDa (Figure 3.5C). Two species of (–TRF1)^{TIN2S} of the same size were also found in similar proportions at the front of the peak, indicating that their elution profile overlapped during SEC purification (Figure 3.5D). This complicated the accurate mass determination by SEC-MALS, which relied on different species being resolved by SEC purification and explains the discrepancy between the number of species detected in SEC-MALS compared with mass photometry.

In this case, the maximum mass error associated with this set of mass photometry measurements was 5%, which led to an error range large enough to cover three potential stoichiometries for the 563 kDa complex (Table 3.4). One possible explanation may be the TRF2$_4$/RAP1$_3$/(dStrepII-TIN2/TPP1/POT1)$_1$ complex, which has been observed previously (Table 3.3). However, a complex with a tetramer of TRF2/RAP1 was outside the error range for this experiment. The remaining possible explanation for the 563 kDa complex could be a dimeric (–TRF1) complex with two copies of each subunit. Note that when the samples were crosslinked prior to data collection, the abundance of the smaller 174 kDa species decreased dramatically, suggesting that mild crosslinking stabilised the 383 kDa and 563 kDa complexes (Figure 3.5D). This 174 kDa subcomplex matches the expected MW of a dStrepII-TIN2S/TPP1/POT1 heterotrimer (expected MW: 173 kDa), and the sum of 174 kDa and 383 kDa gives the 563 kDa complex. Therefore, it is possible that the dStrepII-TIN2S/TPP1/POT1 heterotrimer and the 383 kDa minimal (–TRF1)^{TIN2S} complex formed the higher-MW complex, which dissociated upon sample dilution during mass photometry analysis.

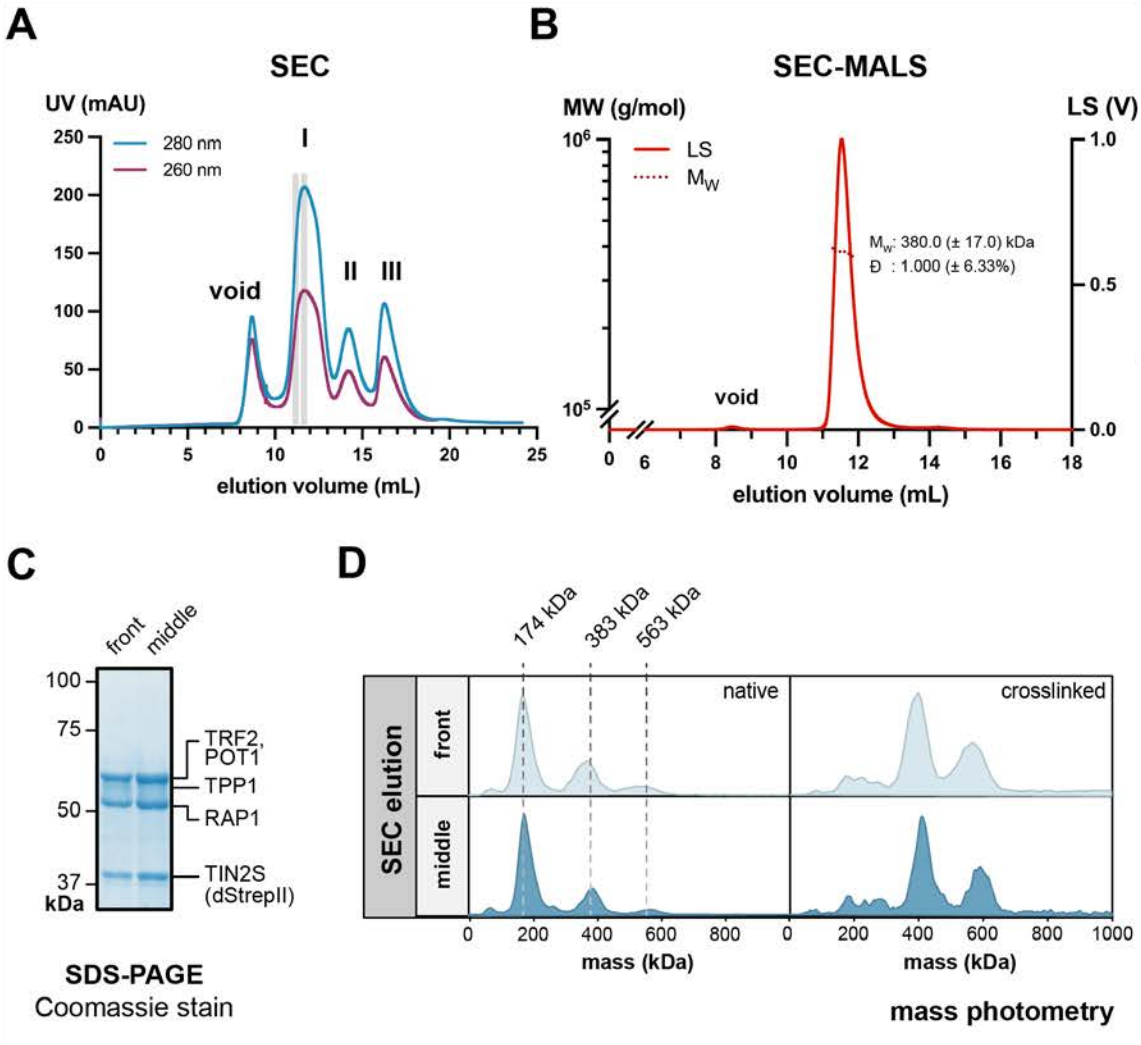

Figure 3.5. MW of (–TRF1)^{TIN2S} determined by SEC-MALS and mass photometry.

(A) SEC elution profile of the (–TRF1)^{TIN2S} complex. Fractions from the middle of SEC peak ***I*** were pooled and concentrated for further analysis. **(B)** SEC-MALS analysis of fractions from the centre of peak ***I***, showing the LS signal and the MW distribution across the peak. **(C)** SDS-PAGE analysis of shelterin (–TRF1)^{TIN2S}, as highlighted in (A). (**D)** Mass photometry analysis of species in SEC peak ***I*** as highlighted in (A). Gaussian curves were fitted to the mass distribution to give the mean MW values of each species under native conditions, with a maximum mass error of 5%. Schematics of shelterin complexes that would satisfy the native MW values are displayed on the top right.

Shelterin complex/subcomplex	Predicted mass (kDa)	Measured mass (kDa)
dStrepII-TIN2S/TPP1/POT1	172.62	179
$(TRF2/RAP1)_2$/dStrepII-TIN2S/TPP1/POT1	380.32	383
$(TRF2/RAP1/dStrepII-TIN2S/TPP1/POT1)_2$	552.94	563
$(TRF2)_4/(RAP1)_3$/dStrepII-TIN2S/TPP1/POT1	543.76	
$(TRF2/RAP1)_4$/dStrepII-TIN2S/TPP1/POT1	588.02	

Table 3.4. Potential stoichiometries of shelterin $(–TRF1)^{TIN2S}$ species from the middle of SEC peak *I*.

The top row contains the species detected in the MP analysis of shelterin $(–TRF1)^{TIN2S}$ from the middle of SEC peak ***I*** (Figure 3.5A), with a maximum mass error of 5%. The bottom row sums the predicted MW of the complex using the stoichiometry described in each column.

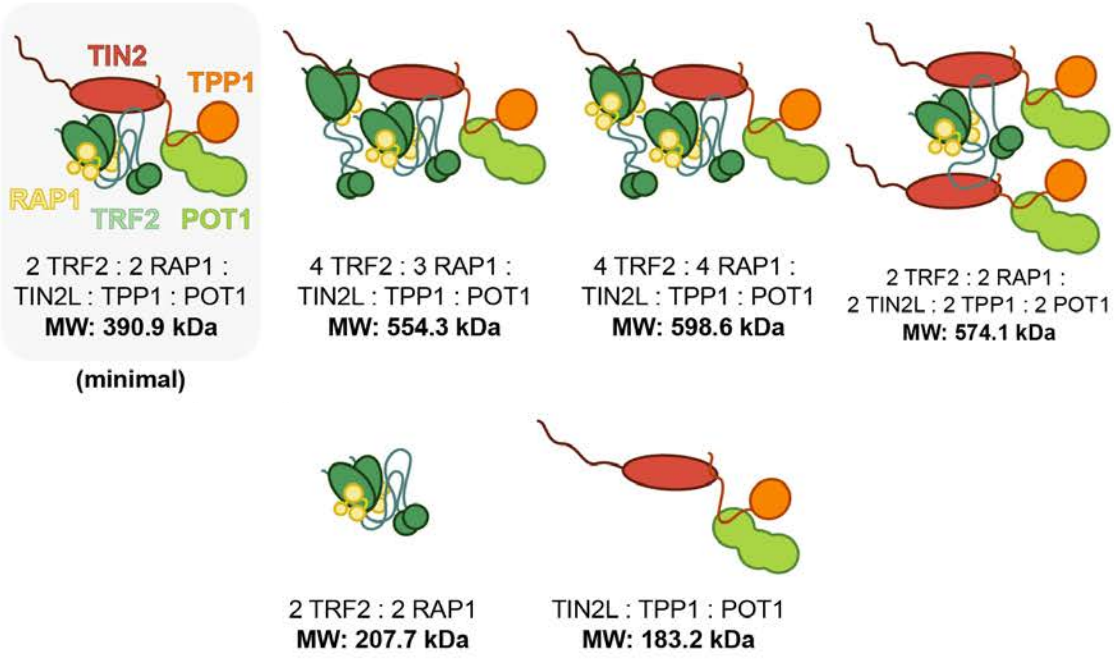

Figure 3.6. Potential stoichiometries of the (–TRF1) subcomplex.

Potential stoichiometries of the smaller (~390 kDa) and larger (~600 kDa) species of the shelterin $(–TRF1)^{TIN2L}$ subcomplex,. The difference of ~200 kDa could be explained by an excess of $(TRF2/RAP1)_2$, or by a subcomplex of $(dStrepII-TIN2/TPP1/POT1)_1$.

3.2.2 Optimising the purification of shelterin (–TRF1)

The two-step purification protocol described above produced heterogeneous populations of shelterin (–TRF1) subcomplex, as indicated by mass photometry analyses (see Figure 3.4, Figure 3.7). Therefore, I explored other strategies to isolate homogeneous shelterin (–TRF1) for downstream experiments.

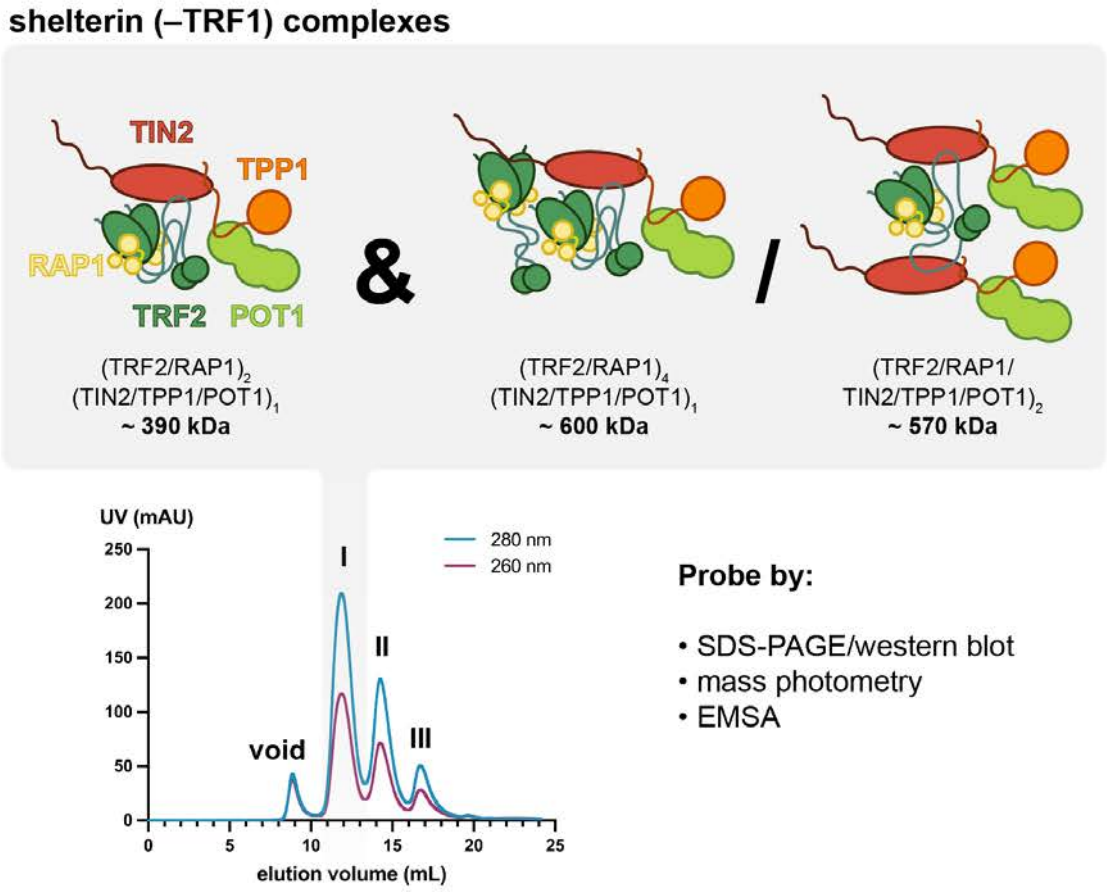

Figure 3.7. Working hypothesis of the stoichiometry of shelterin (–TRF1).

Species of (–TRF1) from SEC purification. After SEC, peak **I** contained two (–TRF1) complexes, as indicated by mass photometry analyses.

3.2.2.1 Resolving (–TRF1) species by ion exchange chromatography

First, I tried resolving the two different (–TRF1) species by ion exchange chromatography (IEX) using heparin, cation exchange (strong cation exchangers SP and Mono S) and anion exchange (strong anion exchanger Q) columns. Shelterin (–TRF1)^{TIN2L} from SEC peak ***I*** were dialysed into a low-salt buffer (200 mM or 300 mM) at two different pH values (pH 7.0 or pH 8.0) and loaded onto the cation and anion exchange columns (Figure 3.8), or the heparin column (Figure 3.9). The NaCl concentration was increased linearly to elute any bound proteins. Shelterin (–TRF1)^{TIN2L} complexes bound to the SP and Q columns at pH 7.0 with either 200 mM or 300 mM of NaCl, respectively (Figure 3.8A and B). However, the two species of (–TRF1) complexes were not resolved as all proteins eluted in one peak. No interactions were observed using the Mono S 5/50 column under the conditions tested (Figure 3.8B).

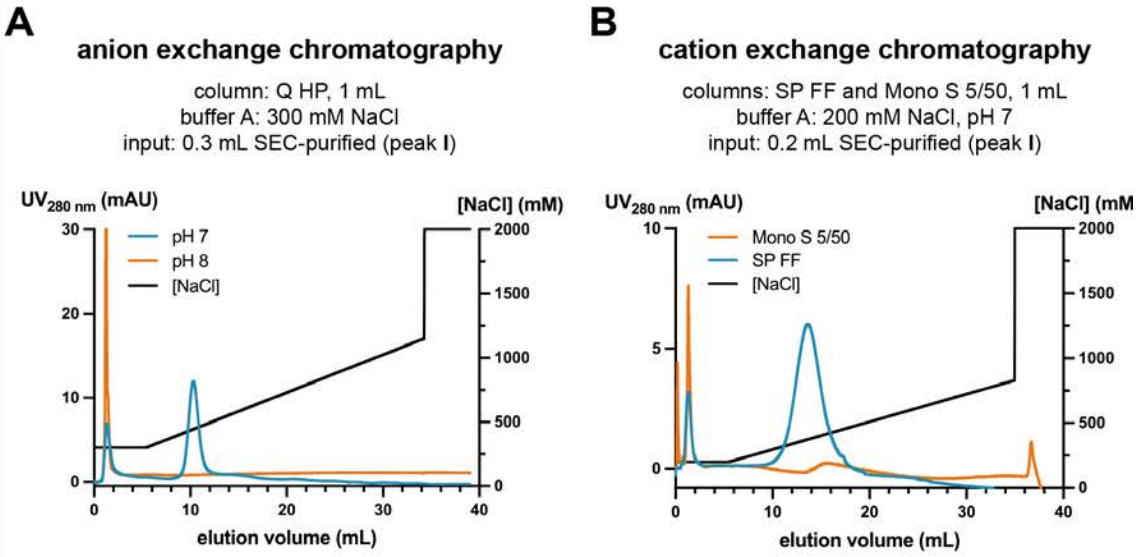

Figure 3.8. Purification of shelterin (–TRF1)^{TIN2L} complexes using ion exchange chromatography.

Elution profile of (–TRF1)^{TIN2L} from **(A)** an anion exchange column (HiTrap Q HP 1 mL) and **(B)** cation exchange columns (HiTrap SP FF and Mono S 5/50 GL, 1 mL). SEC-purified (–TRF1)^{TIN2L} complexes from peak ***I*** were pooled and dialysed into the starting buffer (200 or 300 mM NaCl in 20 mM of HEPES at pH 7.0 or 8.0) and loaded onto the IEX columns. The NaCl concentration increased linearly from buffer A (200 mM or 300 mM) to 35 or 50% of buffer B (2 M) over 30 CV.

SEC-purified shelterin (–TRF1)^{TIN2L} complexes eluted from the heparin column in one peak preceded by a small shoulder (Figure 3.9A and B). Mass photometry analysis showed that the shoulder contained a smaller subcomplex of 146 kDa (13.7 mL, Figure 3.9B and C), which likely originated from SEC peak ***II*** (see Figure 3.9A). The two species of (–TRF1) from SEC peak ***I*** co-eluted in the main peak, with most of the higher-order species (~600 kDa) found in earlier fractions (17.2 – 19.2 mL) whilst the minimal complex (~390 kDa) eluted across the entire peak (17.2 – 20.2 mL, Figure 3.9B and C).

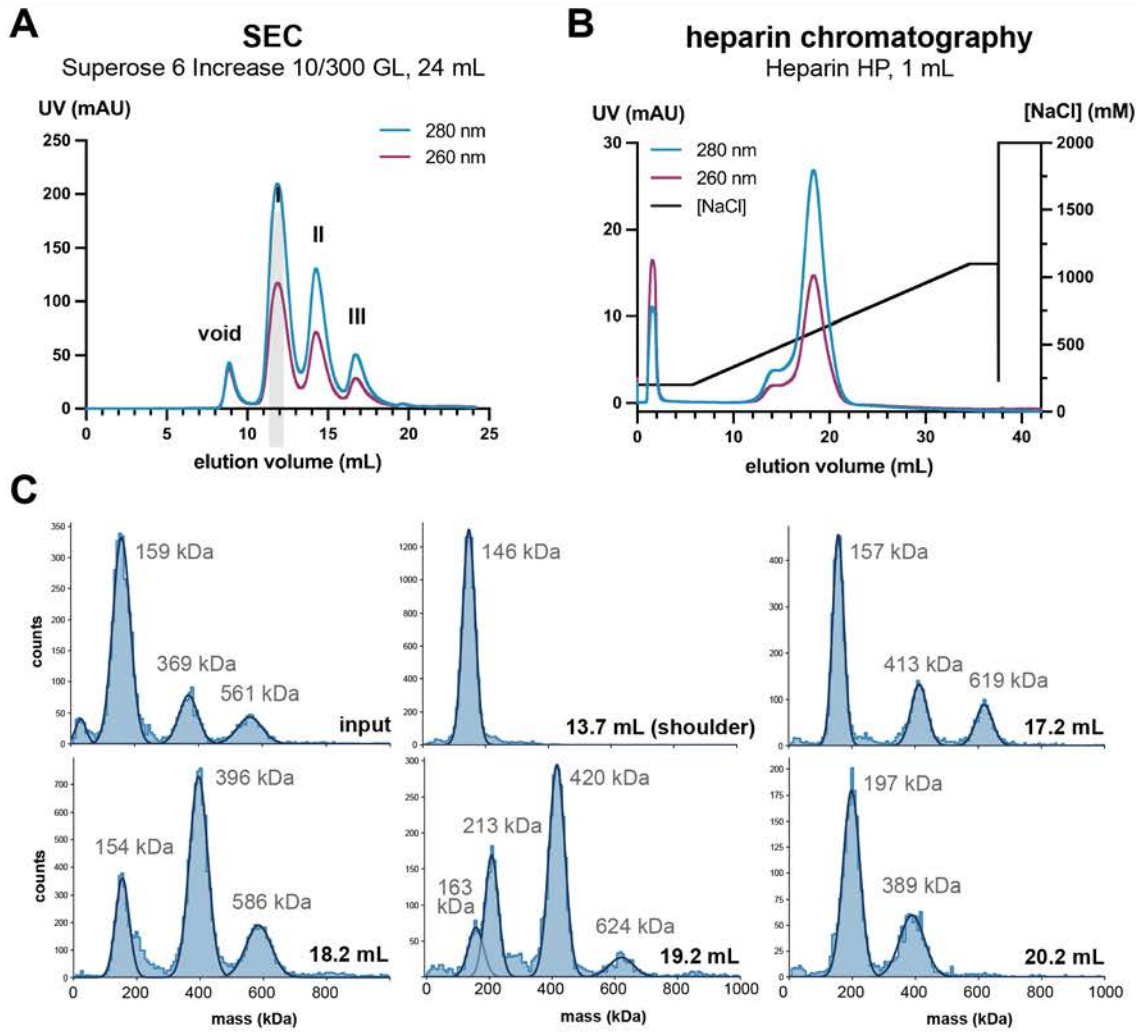

Figure 3.9. Purification of shelterin (–TRF1)^{TIN2S} by heparin chromatography.

(A) SEC elution profile of shelterin (–TRF1)^{TIN2S}. Fractions from peak ***I*** were pooled and dialysed into starting buffer A (200 mM NaCl, pH 8). **(B)** Elution profile from a Heparin HP 1 mL column, eluted using a linear salt gradient from 0% B (200 mM NaCl) to 50% B (2 M NaCl) over 30 CV. **(C)** Mass photometry measurements of the input and elution fractions as shown in (B).

3.2.2.2 Supplementing (–TRF1) with excess TRF2/RAP1 did not enrich the higher-MW species

As shown previously, mass photometry analysis showed (–TRF1) complexes formed a higher-order species at ~600 kDa, which is larger than the minimal (–TRF1) complex with $(TRF2/RAP1)_2$/(dStrepII-TIN2/TPP1/POT1)$_1$ by ~200 kDa. Given the known protein-protein interaction interfaces within shelterin, the two possible combinations are $(TRF2/RAP1)_2$ and (dStrepII-TIN2L/TPP1/POT1)$_1$. In the absence of TRF1, the $TIN2_{TBM}$ is unoccupied and has the potential to bind to $TRF2_{TRFH}$ (see section 1.2.2.3). Western blotting analysis of the full shelterin complex indicated that TRF2 and RAP1 subunits were present in greater amounts relative to TIN2 in the higher-order species compared to the minimal shelterin complex (Figure 2.2B). Therefore, I first hypothesised that the size difference between the larger and minimal (–TRF1) complexes may be due to additional copies of TRF2 and RAP1 (Figure 3.10). One potential approach to test this is to provide excess TRF2 and RAP1 subunits to (–TRF1) complexes and measure the abundance of the two different species. For this experiment, I reconstituted the TRF2/RAP1 subcomplex from recombinant dStrepII-TEV-tagged TRF2 and RAP1 proteins (Figure 3.11), see sections 6.2.3 and 6.3.4 for details of protein expression and purification). Fractions of shelterin (–TRF1)^{TIN2S} from SEC peak ***I*** were pooled and incubated with either 1.2- or 6-fold molar excess of TRF2/RAP1. The mixtures were subjected to analytical SEC and analysed by SDS-PAGE and mass photometry.

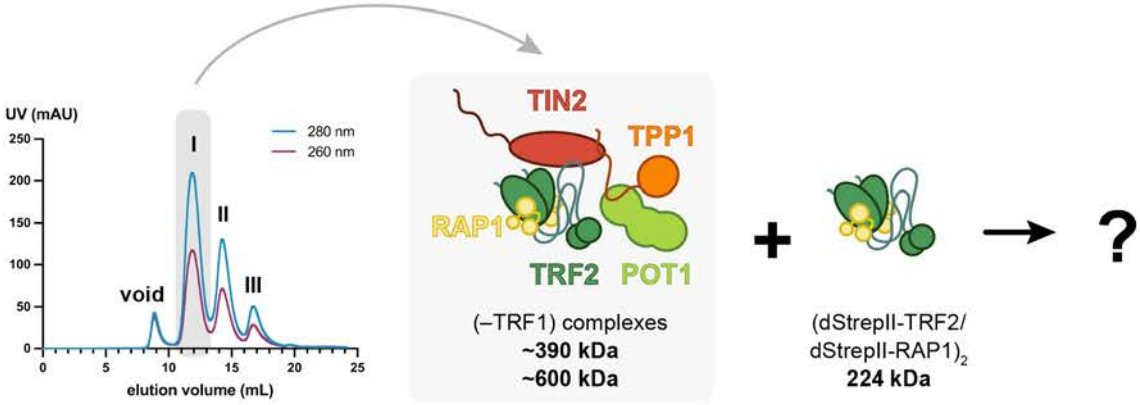

Figure 3.10. Overview of the rationale and experimental setup for enriching the higher order (–TRF1) complex.

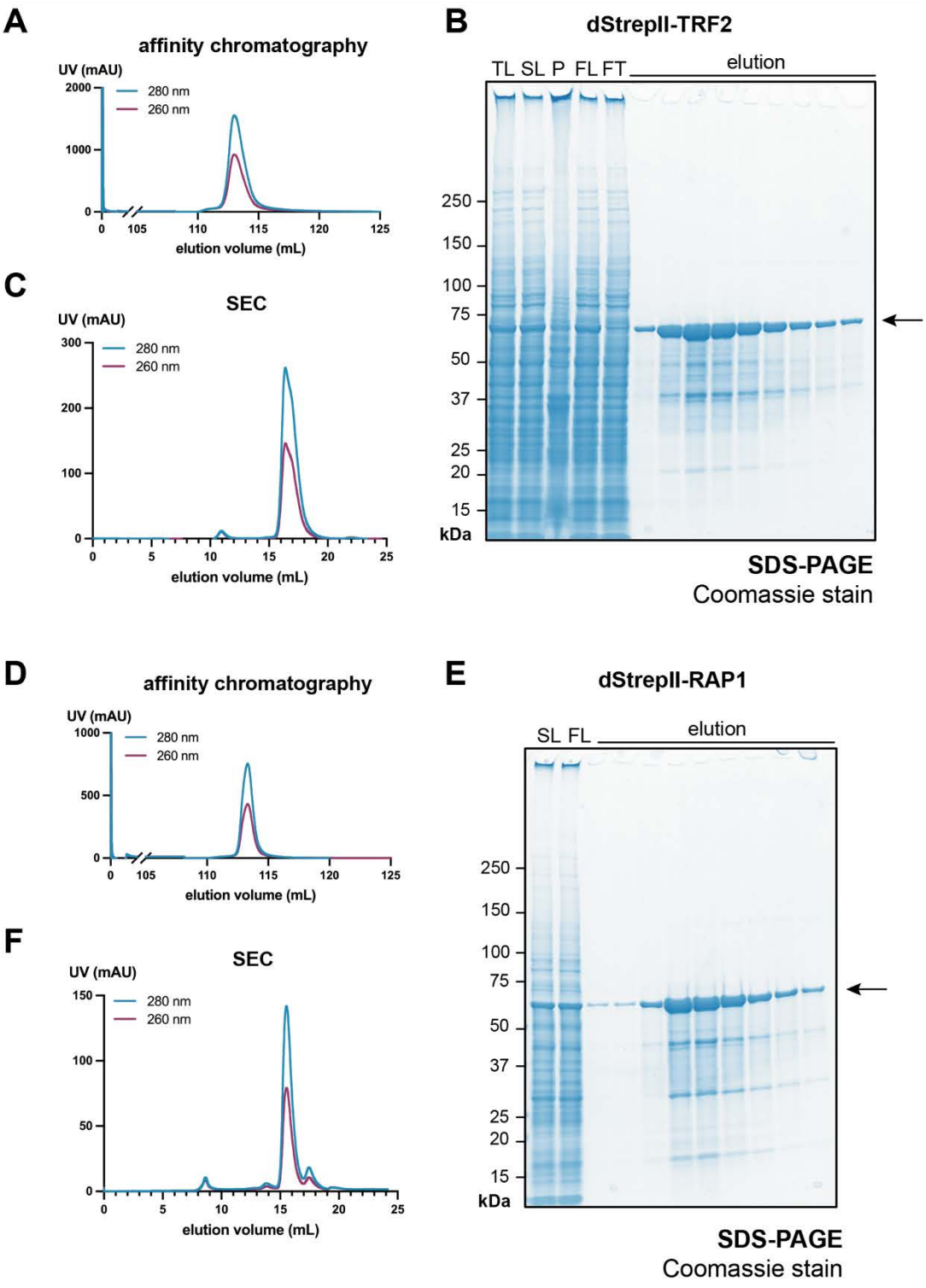

Figure 3.11. Purification of dStrepII-TRF2 and dStrepII- RAP1.

(A – C) Purification of dStrepII-TRF2. (A) Elution profile from the Strep affinity purification. **(B)** SDS-PAGE analysis of the affinity elution fractions. **(C)** Elution profile from the SEC purification. **(D – E) Purification of dStrepII-RAP1. (D)** Elution profile from the Strep affinity purification. **I** SDS-PAGE analysis of the affinity elution fractions. **(F)** Elution profile from the SEC purification. Arrows indicate full length dStrepII-tagged protein. TL – total lysate, SL – soluble lysate, P – pellet (insoluble lysate), FL – filtered lysate, FT – flowthrough.

During analytical SEC, dStrepII-RAP1 eluted because its small size and monomeric nature (15.5 mL, Figure 3.12A). The dStrepII-TRF2 subunit eluted earlier due to its larger size as a constitutive homodimer (14.1 mL, Figure 3.12A). The reconstituted dStrepII-tagged TRF2/RAP1 subcomplex and SEC-purified (–TRF1)^{TIN2S} eluted at 12.6 mL and 11.8 mL, respectively (Figure 3.12A). SDS-PAGE analysis showed dStrepII-RAP1 and dStrepII-TRF2 migrated as a single band (starting from lanes 22 and 15, respectively, Figure 3.12B), and as a stoichiometric complex when mixed in a 1:1 ratio (Figure 3.12B). With the addition of excess TRF2/RAP1 subcomplex, the (–TRF1)^{TIN2S} peak shifted to the right with a higher elution volume (11.8 to 12.5 mL, Figure 3.12A). This was confirmed by SDS-PAGE analysis, which showed that (–TRF1)^{TIN2S} supplemented with either 1.2 or 6-fold molar excess TRF2/RAP1 eluted later compared to (–TRF1)^{TIN2S} alone (panels 4 to 6, Figure 3.12B). This indicated that the added dStrepII-tagged TRF2/RAP1 was incorporated into the (–TRF1)^{TIN2S} complex and altered its shape, or size, or both.

The dStrepII-tagged TRF2 and RAP1 migrated higher on SDS-PAGE gels than their untagged counterparts within the SEC-purified (–TRF1)^{TIN2S} complex, which enabled detection of their incorporation into the complex. SDS-PAGE analysis showed that excess TRF2/RAP1 co-eluted with the (–TRF1)^{TIN2S} complex, which occurred earlier during SEC elution compared with the TRF2/RAP1 subcomplex alone (Figure 3.12B). At 1.2-fold molar excess, the majority of the TRF2/RAP1 subcomplex was incorporated into the (–TRF1) complex, whereas at 6-fold molar excess some of the TRF2/RAP1 subcomplex was seen eluting at the volume corresponding to isolated TRF1/RAP1 (Figure 3.12B). Moreover, untagged TRF2 and RAP1 appeared in elution fractions lacking TIN2S, i.e., later than the (–TRF1) complex alone (up until fraction 19 vs 15, respectively). Together, this indicates that some untagged TRF2 and RAP1 subunits were displaced by the dStrepII-tagged versions (Figure 3.12B).

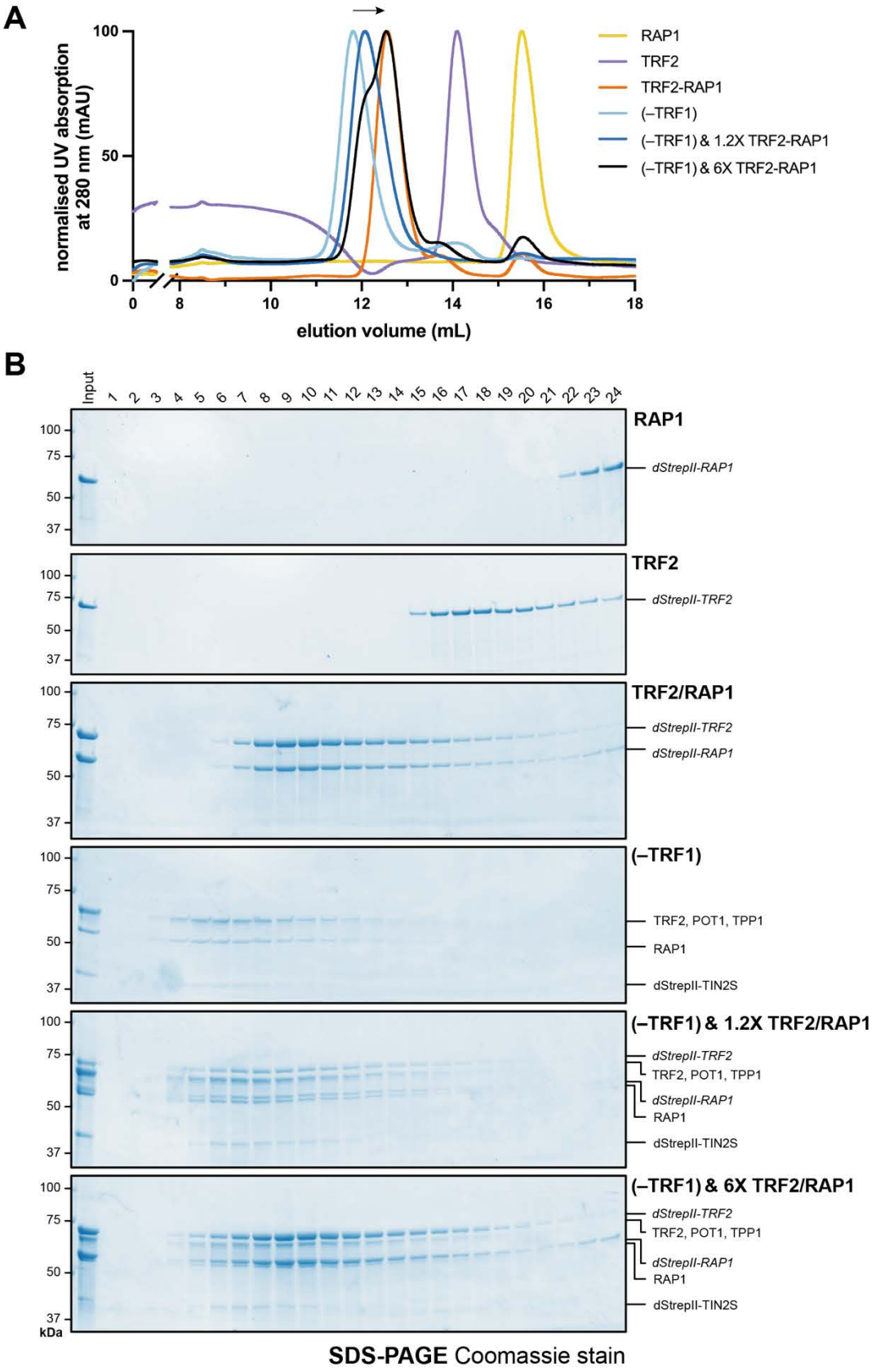

Figure 3.12. Analytical SEC of (–TRF1)^{TIN2S} supplemented with 1.2- or 6-fold molar excess TRF2-RAP1.

(legend on the next page)

(figure on the previous page)

Figure 3.12. Analytical SEC of (–TRF1)^{TIN2S} supplemented with 1.2- or 6-fold molar excess TRF2-RAP1.

(A) Normalised UV chromatograms showing the elution profiles of dStrepII-TRF2, dStrepII-RAP1, the dStrepII-tagged TRF2-RAP1 subcomplex, the (–TRF1) complex (previously SEC-purified), and the (–TRF1) complex in the presence of 1.2- or 6-times molar excess of TRF2-RAP1. All samples were resolved using a Superose 6 Increase 10/300 GL column at 0.1 mL/min in 200 µL fractions. **(B)** SDS-PAGE analysis of the elution fractions from analytical SEC depicted in (A).

Next, I checked whether the incorporation of dStrepII-tagged TRF2/RAP1 enriched the higher-order (–TRF1) species, or simply displaced their untagged counterparts. Using mass photometry, I measured the sizes of species from each condition before SEC purification (SEC input) and from the peak fraction from SEC purifications (SEC peak, Figure 3.13). The samples were measured directly (native) or after crosslinking with 0.1% of glutaraldehyde to capture the MW as accurately as possible, whilst detecting any larger-MW species that may have dissociated during the analysis. Mass photometry analysis showed that SEC-purified dStrepII-TRF2 had a MW of 125 kDa, which equates to a TRF2 homodimer (expected MW: 127.4 kDa, Figure 3.13). The calculated MW of one dStrepII-RAP1 subunit is 44.3 kDa, which approached the detection limit of the Refeyn One^{MP} instrument. This explains the large observed MW of 70 kDa for dStrepII-RAP1 (Figure 3.13). The reconstituted TRF2/RAP1 subcomplex contained three populations under native conditions (Figure 3.13). The most abundant species was 207 kDa in the SEC input (231 kDa in the SEC peak), close to the expected size of 224.2 kDa for the (dStrepII-TRF2/dStrepII-RAP1)$_2$ subcomplex (Figure 3.13).

The mass distribution of (–TRF1)^{TIN2S} alone was consistent with previous mass photometry measurements (see Figure 3.5D). The smallest species near 195 kDa were subcomplexes that formed during sample dilution, and their abundance decreased substantially when the sample was crosslinked prior to dilution (Figure 3.13). The species in the SEC input at 403 kDa (384 kDa in the SEC peak) is consistent with a minimal (–TRF1) complex with a (TRF2/RAP1)$_2$/

(dStrepII-TIN2S/TPP1/POT1)$_1$ stoichiometry. Another, larger species was detected in the input at 597 kDa (543 kDa in the SEC peak). Incubation of the (–TRF1) complexes with 1.2- or 6-fold molar excess of TRF2/RAP1 increased the MW of the minimal (–TRF1) complex, further suggesting that the dStrepII-tagged TRF2/RAP1 subcomplex displaced the untagged TRF2/RAP1 in the (–TRF1) input (SEC peak, Figure 3.13). However, it did not alter the relative abundances of the ~390 kDa and 600 kDa complexes (Figure 3.13), indicating that incorporation of the dStrepII-tagged TRF2/RAP1 subcomplex displaced the untagged counterparts in the (–TRF1)^{TIN2S} complex.

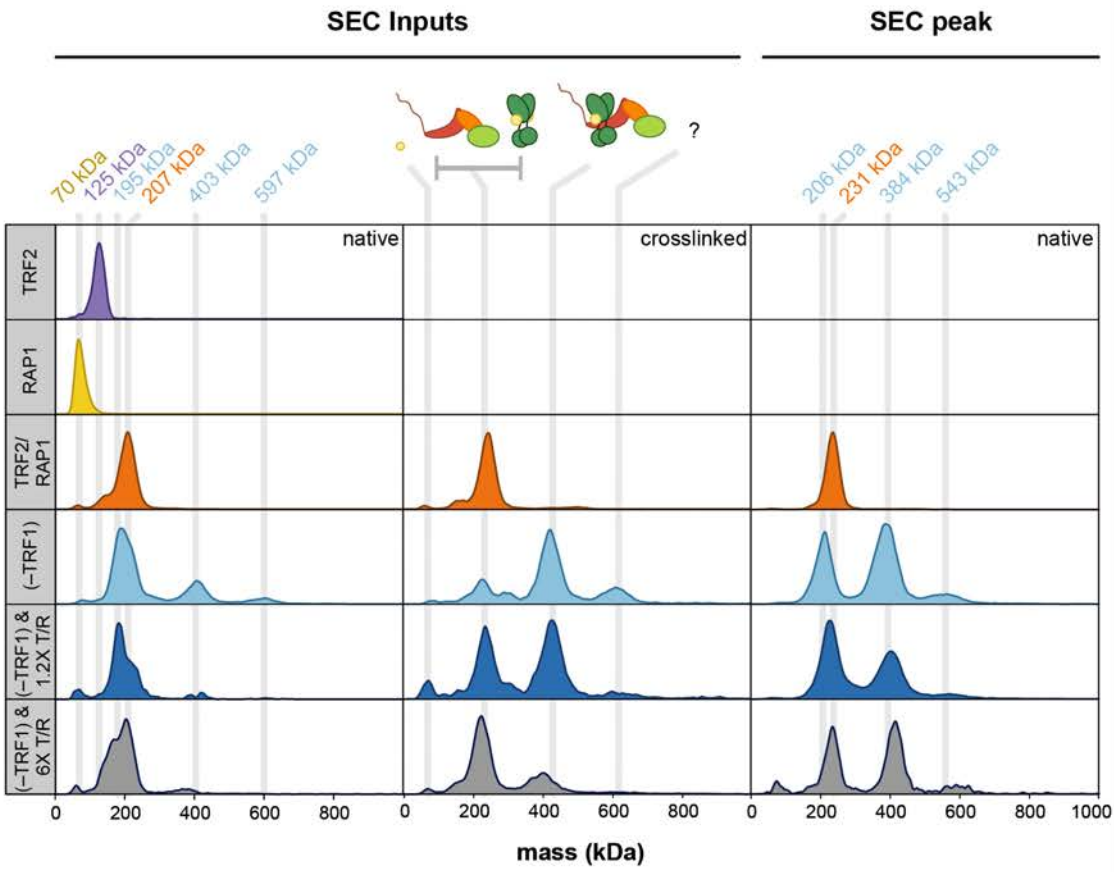

Figure 3.13. Mass photometry analysis of (–TRF1)^{TIN2S} supplemented with 1.2- and 6-fold molar excess of (TRF2/RAP1)$_2$.

The inputs to the analytical SEC experiment (SEC inputs) and the peak elution fraction from each SEC run (SEC peak) were analysed using mass photometry. For the SEC inputs, sample were measured under native conditions and after crosslinking with 0.01% of glutaraldehyde. Samples from the SEC peaks were analysed only under native conditions.

3.2.2.3 Disrupting the $TIN2_{TBM}$–$TRF2_{TRFH}$ interaction did not abolish the formation of higher-order (–TRF1) species

The scaffolding protein TIN2 offers a platform for the assembly of shelterin subunits. It has two known sites that interact with TRF1 and TRF2. The TRFH-like domain in TIN2 binds to the linker region in TRF2 ($TRF2_{linker}$), and the TIN2 TRF-binding motif ($TIN2_{TBM}$) can bind to the TRFH domain in both TRF1 and TRF2 (K_D = 6.5 μM vs 0.3 μM, respectively, Figure 1.14) (Chen, et al., 2008, Hu, et al., 2017). Therefore, I hypothesised that in the absence of TRF1, TIN2 can accommodate two TRF2 dimers, one through the $TIN2_{TRFH\text{-}like}$-$TRF2_{linker}$ interaction, and another through the $TIN2_{TBM}$-$TRF2_{TRFH}$ interaction (Figure 1.14 and Figure 1.7). Although the latter is weaker than the $TIN2_{TBM}$-$TRF1_{TRFH}$ interaction, it may facilitate the recruitment of another TRF2 dimer when there is no competition from TRF1. To test this hypothesis, I generated (–TRF1) complexes with mutations in $TIN2_{TBM}$ at a conserved leucine at position 260 in $TIN2_{TBM}$ to either a smaller alanine residue ($TIN2^{L260A}$) or a negatively charged glutamic acid ($TIN2^{L260E}$) residue. This has been reported to reduce binding of $TIN2_{TBM}$ to $TRF2_{TRFH}$ by disrupting the hydrophobic side-chain interactions between $TIN2_{TBM}$ and several residues in $TRF2_{TRFH}$ (Figure 1.7) (Chen, et al., 2008).

The affinity and SEC purification of the $(–TRF1)^{TIN2S(L260E)}$ complex is shown as a representative example (Figure 3.14). The SEC elution profile suggests that peak ***I*** still contained a mixture of two different species. Mass photometry data showed the same species were present in the SEC input and SEC elution fractions compared with the WT shelterin $(–TRF1)^{TIN2S}$ complex, including the higher-order complex at ~600 kDa (Figure 3.15B). Therefore, this suggests that the $TIN2_{TBM}$ site is not involved in recruiting additional TRF2 subunits in the absence of TRF1.

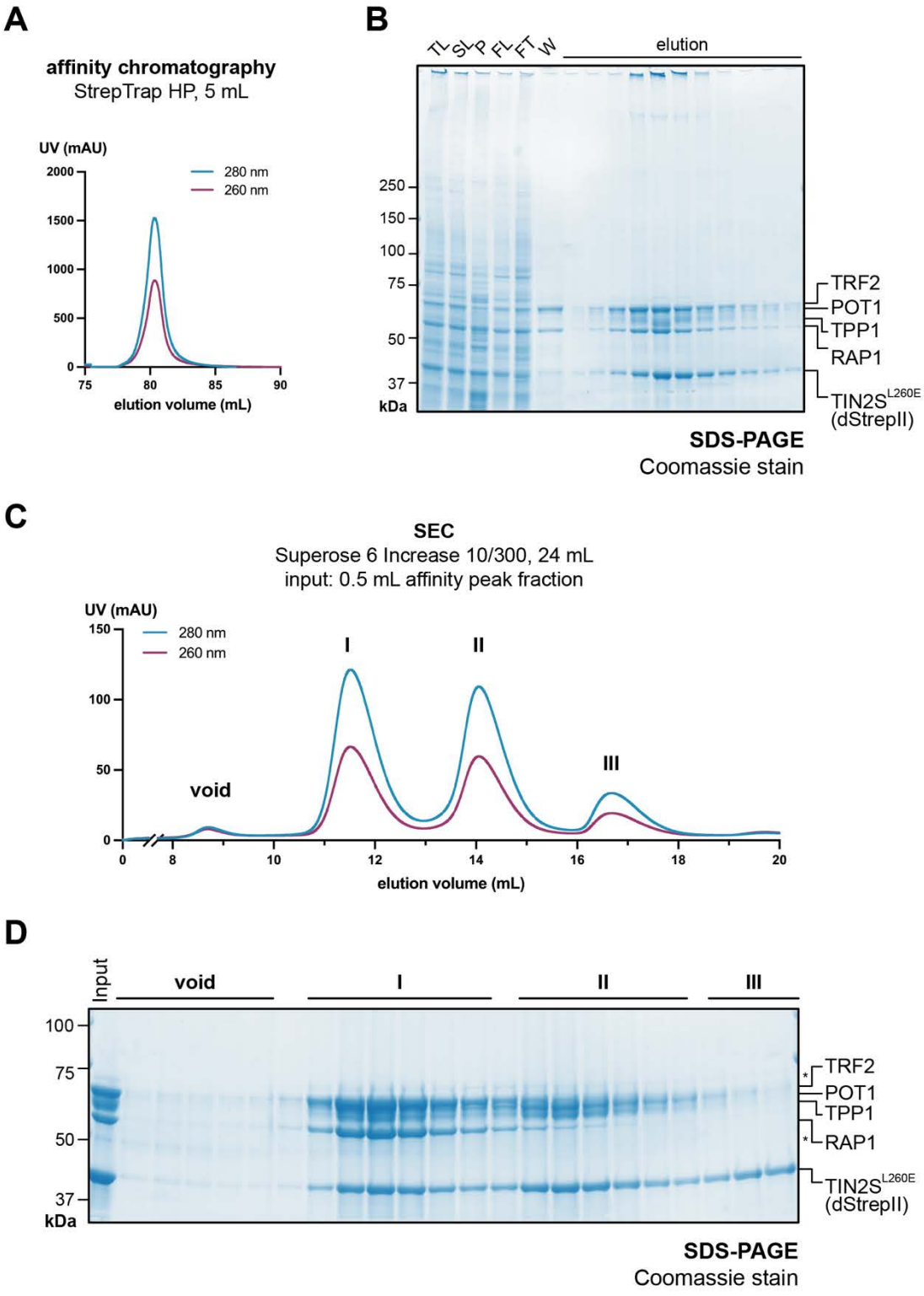

Figure 3.14. Purification of shelterin (–TRF1) with TIN2S^{L260E}.

(A) UV chromatogram from the affinity purification of (–TRF1)^{TIN2S} with a L260E point mutation in TIN2S. **(B)** SDS-PAGE analysis of samples from the affinity purification, as depicted in (A). **(C)** SEC elution profile from a Superose 6 Increase 10/300 GL column. The peak affinity elution fraction was injected into the SEC column at 0.3 mL/min. **(D)** SDS-PAGE analysis of the input to and elution from SEC as shown in (C).

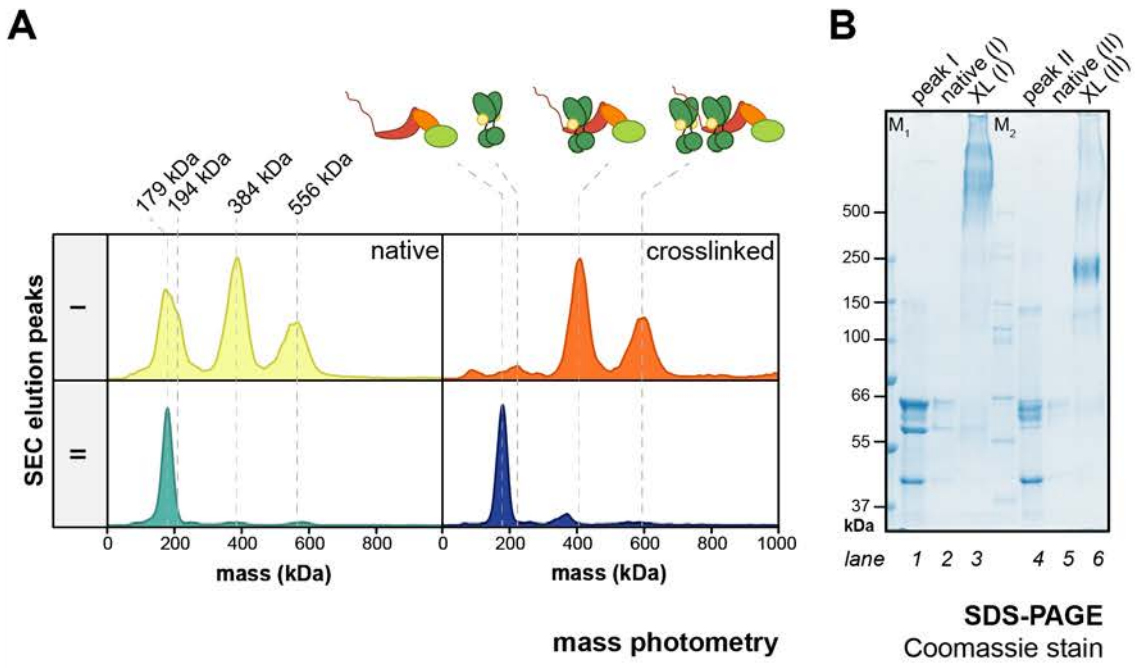

Figure 3.15. Mass photometry analysis of shelterin (–TRF1) with TIN2S^{L260E}.

(A) Mass photometry analysis of (–TRF1)^{TIN2S} from SEC peaks ***I*** and ***II***, as depicted in Figure 3.14C. Samples were measured under native (left) and crosslinked (right) conditions. The dotted lines mark the mean MW of each Gaussian peak. Schematics above the 'crosslinked' column show the predicted stoichiometry of shelterin subunits for each species (see Table 3.5). **(B)** SDS-PAGE analysis of the samples subjected to mass photometry analysis shown in (A): input (lanes 1 and 4), native samples after dilution (lanes 2 and 5), and crosslinked samples (lanes 3 and 6) prior to dilution. M_1 and M_2 are two protein markers ranging from 10 to 500 kDa.

Shelterin complex/subcomplex	Predicted mass (kDa)	Measured mass (kDa)
dStrepII-TIN2S/TPP1/POT1	172.62	179
$(TRF2/RAP1)_2$	207.70	194
$(TRF2/RAP1)_2$/dStrepII-TIN2S/TPP1/POT1	380.32	384
$(TRF2/RAP1/dStrepII\text{-}TIN2S/TPP1/POT1)_2$	552.94	556
$(TRF2)_4/(RAP1)_3$/dStrepII-TIN2S/TPP1/POT1	543.76	
$(TRF2/RAP1)_4$/dStrepII-TIN2S/TPP1/POT1	588.02	

Table 3.5. MW of shelterin (–TRF1) with TIN2S^{L260E} determined by mass photometry.

Potential stoichiometries of shelterin subunits in the mass photometry analysis of (–TRF1) complex with TIN2S^{L260E} (see Figure 3.15).

3.2.3 Shelterin lacking TRF1 binds telomeric DNA with similar affinity to the full complex

In the cellular context, shelterin complexes are recruited to the duplex and 3' ss-overhang regions of telomeric DNA via the TRF1, TRF2 and POT1 subunits. Therefore, I explored whether the absence of TRF1 affected the affinity of shelterin for telomeric DNA. First, I checked whether recombinant shelterin complexes lacking TRF1 can bind to a model telomeric DNA sequence (Figure 3.16A). The minimal telomeric DNA consisted of ds $(TAGGGT)_4$ repeats for one or two pairs Myb domains (i.e., from one or two homodimers) of TRF2 (Court, et al., 2005), followed by the ATC-5' sequence enriched at natural the ends of human telomeres (Sfeir, et al., 2005; Palm, et al., 2009). The ss 3'-overhang formed by the $(GGTTAG)_2$ sequence provided the optimal binding site to engage the OB-folds from one POT1 subunit (Lei, et al., 2004).

Various concentrations of affinity-purified shelterin $(–TRF1)^{TIN2L}$ were incubated with 0.5 μM of a minimal telomeric sequence, teloDNA1, and subjected to EMSA. Recombinant shelterin (–TRF1) complexes bound to telomeric DNA to form two stable protein-DNA complexes (Figure 3.16B). Traces of a (–TRF1)-DNA complex was first observed at 0.25:1 molar ratio of (–TRF1):DNA (band ***II***), and with a larger species forming at 1.5-times molar excess of DNA (band ***I***, Figure 3.16B). The smaller species is likely a subcomplex of shelterin (–TRF1) present in the affinity-purified sample and/or formed due to dissociation of the shelterin (–TRF1) complex upon dilution. All the DNA was engaged when the (–TRF1) complex was in equimolar concentration of the DNA (Figure 3.16B). The (–TRF1):DNA bands were smeared and less distinct compared with that observed for the full complex, which suggests that the shelterin subunits form a dynamic complex that varies in composition and/or conformation (Figure 2.15B). This is a likely explanation for why it is challenging to resolve the different (–TRF1) species during protein purification.

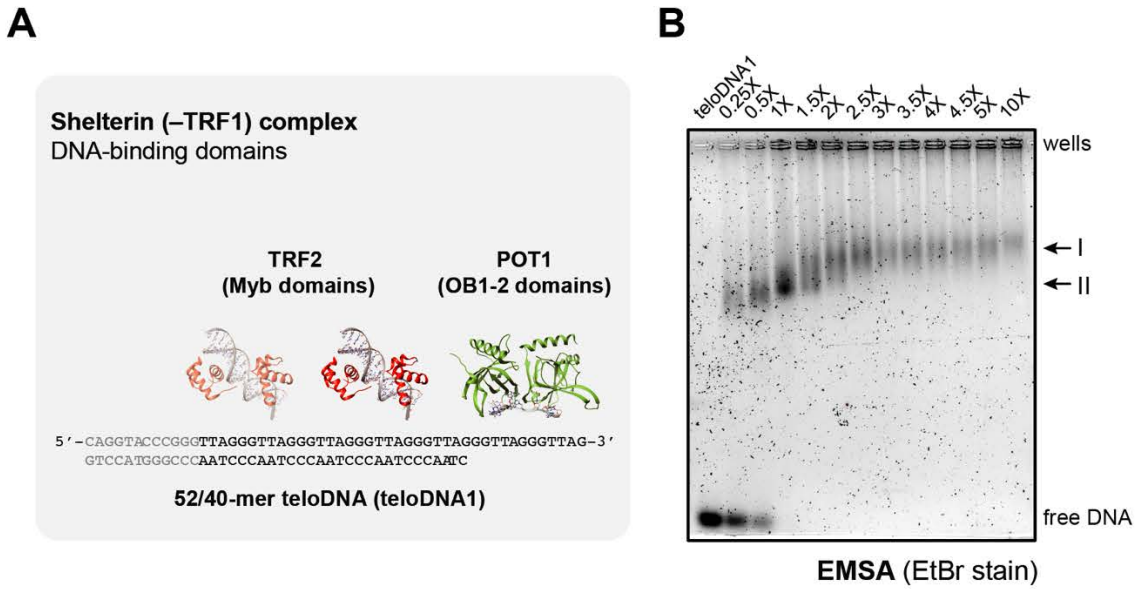

Figure 3.16. Recombinant shelterin (–TRF1) subcomplexes bind to a minimal telomeric DNA sequence.

(A) Schematic of the DNA-binding domains in the shelterin (–TRF1) subcomplexes and the model 52/40-mer telomeric DNA (teloDNA1). One or two pairs of TRF2 Myb domains can bind to the dsDNA and $OB_{1–2}$ from POT1 can bind to the 3' ssDNA. PDBs: 1W0U, 1XJV. **(B)** EMSA analysis of DNA-bound shelterin (–TRF1). Affinity-purified $(–TRF1)^{TIN2L}$ at various concentrations was incubated with 0.5 μM of teloDNA1 and resolved using a 0.6% agarose 0.5XTBE gel.

The DNA-bound $(–TRF1)^{TIN2L}$ complexes were stable during SEC purification (Figure 3.17A). In the presence of DNA, shelterin (–TRF1) eluted earlier compared with the apo complex, reflecting a change in hydrodynamic radius of the complex after DNA binding (Figure 3.17A and B). The increased 260/280 nm UV absorbance ratio of all the peaks after incubation with telomeric DNA is another indication that DNA was incorporated into the shelterin (–TRF1) complex (Figure 3.17A). The SEC elution fractions were further analysed by EMSA, which indicated the presence of one slow migrating species in peak ***I*** and one fast migrating species across peak ***II*** (Figure 3.17C). In summary, recombinant shelterin lacking TRF1 form stable complexes with DNA, which could be further purified by SEC. This renders this shelterin subcomplex an amenable target for structural investigation by EM to complement the cryo-EM studies of the full shelterin complex performed by Dr Oviya Inian.

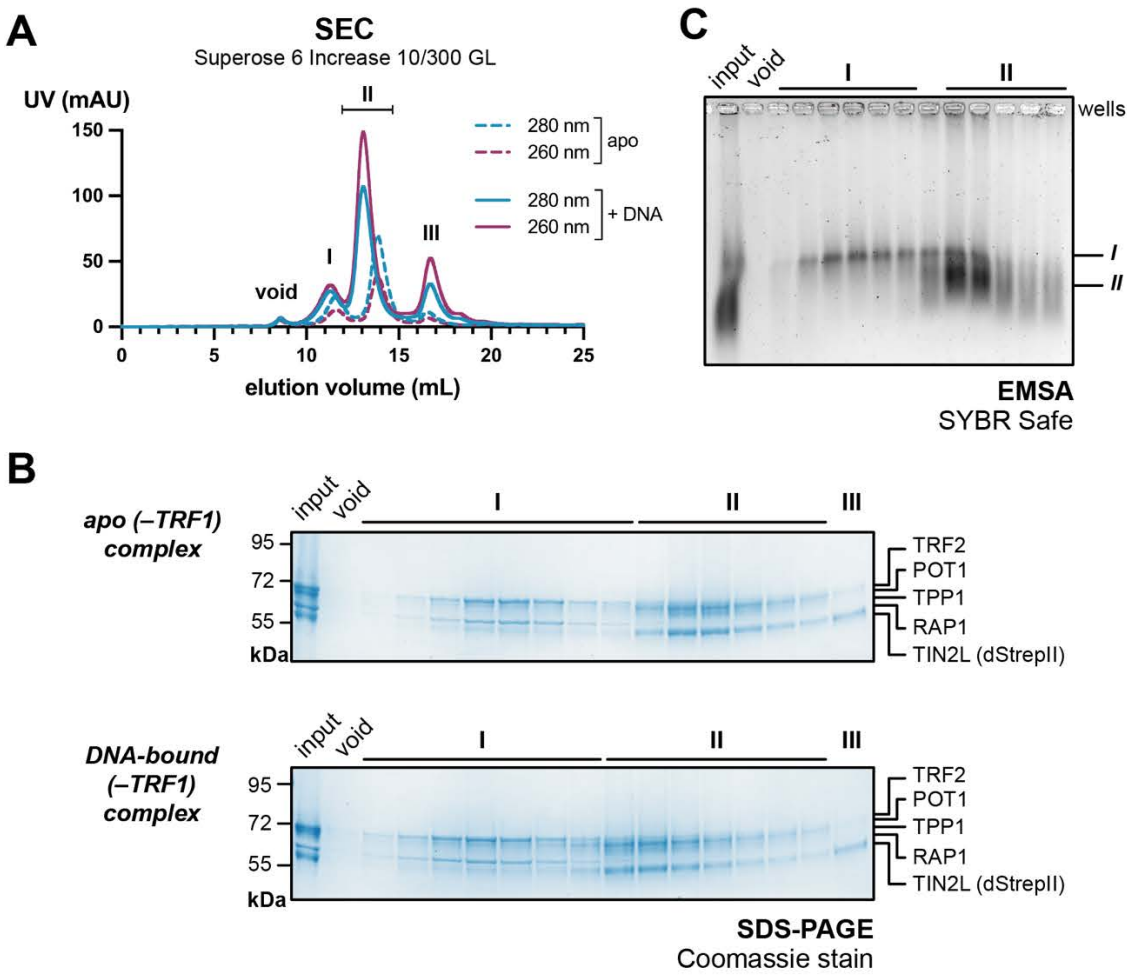

Figure 3.17. DNA-bound shelterin (–TRF1) complexes remain stable over SEC.

(A) SEC elution profile of shelterin (–TRF1)^{TIN2L} incubated with buffer (dotted lines) or 1.2-fold molar excess of teloDNA1 (solid lines). Samples were subjected to analytical SEC. **(B)** The SEC elution fractions from the reconstitution of the apo and DNA-bound (–TRF1) complexes were analysed by SDS-PAGE, and **(C)** the DNA-bound (–TRF1) complexes were analysed by EMSA.

Next, I determined the binding kinetics of recombinant (–TRF1)^{TIN2L} shelterin to a model telomeric and non-telomeric DNA using switchSENSE®. These measurements used the same DNA sequences, biochip and set-up as described for the full shelterin^{TIN2L} complex (see Figure 2.17). Similar to the full complex, the (–TRF1)^{TIN2L} shelterin complex bound to a minimal telomeric and non-telomeric DNA sequence in a 1:1 ratio with nanomolar affinity (K_D = 11.1 ± 0.7 nM vs. 96.7 ± 5.1 nM, respectively; Figure 3.18). Global fitting of associations yielded k_{ON} values that were different by nearly an order of magnitude (1.69 ±

0.09 x 10^5 M^{-1} s^{-1} vs 1.44 ± 0.07 x 10^6 M^{-1} s^{-1} for telomeric and non-telomeric DNA, respectively). In the absence of TRF1, the shelterin complex associated with non-telomeric DNA 8.5 times faster than telomeric DNA under the conditions tested (Figure 3.19). This is potentially due to the absence of Myb domains in the TRF1 homodimer, which mediates sequence-specific DNA binding. A greater difference was observed during the dissociation phase, where shelterin (–TRF1) disengaged from the non-telomeric control within seconds, giving a high k_{OFF} at 1.39 ± 0.04 x 10-1 s^{-1}. In contrast, the complex dissociated from telomeric DNA gradually, giving a k_{OFF} of 1.87 ± 0.04 x 10^{-3} s^{-1}. This 74-fold difference in k_{OFF} is largely responsible for the different binding affinity of the (–TRF1) complex to telomeric and non-telomeric DNA.

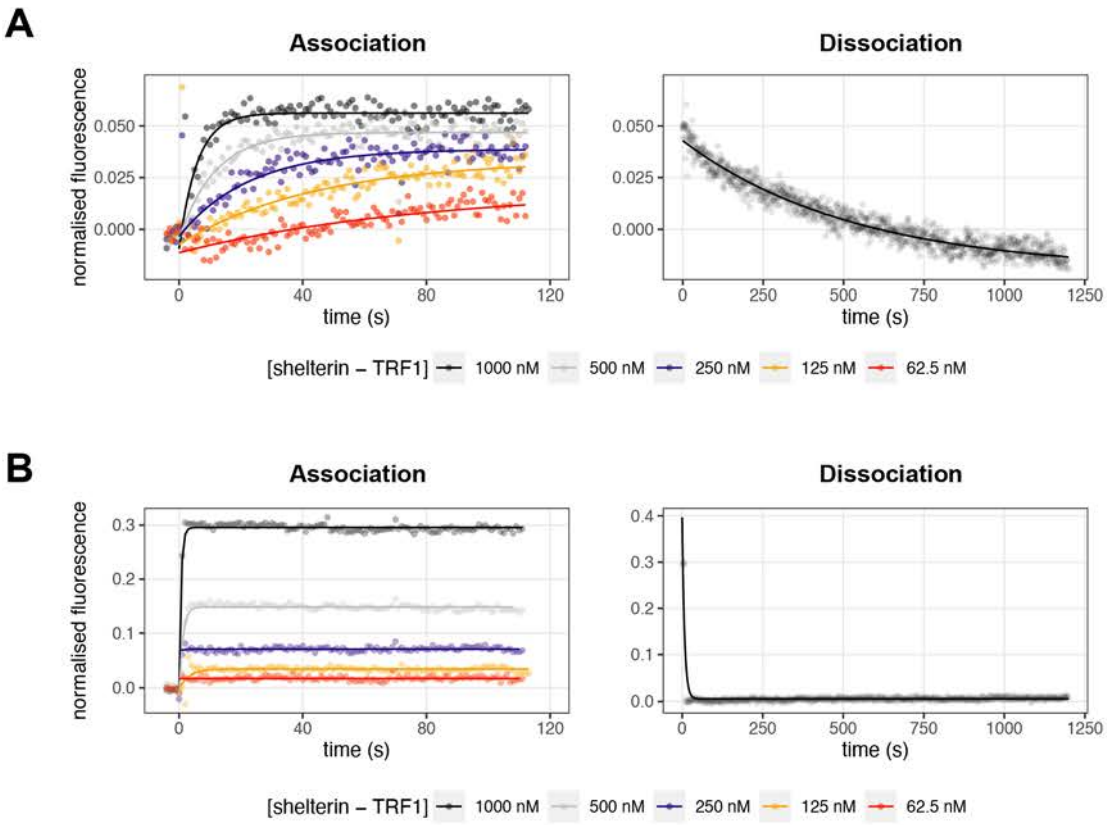

Figure 3.18. Time-resolved measurements of the association and dissociation of (–TRF1) from telomeric and non-telomeric DNA.

Association and dissociation of shelterin $(–TRF1)^{TIN2L}$ complexes from **(A)** telomeric DNA and **(B)** non-telomeric DNA. Six protein concentrations shelterin $(–TRF1)^{TIN2L}$ complexes (0, 62.5, 125, 250, 500 and 1000 nM) were used to determine k_{ON}, and one (1000 nM) to determine k_{OFF}.

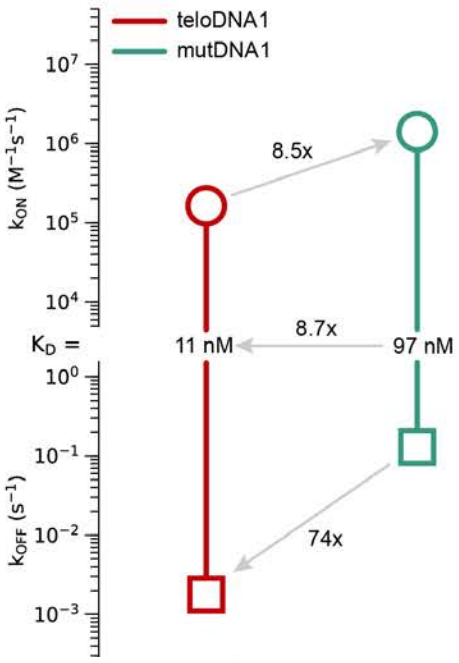

Figure 3.19. Rate map of (–TRF1) binding to telomeric vs. non-telomeric DNA.

The rate constants for association (k_{ON}) and dissociation (k_{OFF}) events, and the equilibrium constant for dissociation (K_D) (see Figure 3.18). Made with www.affinity-avidity.com.

Comparing with switchSENSE results between the full shelterin complex and the (–TRF1) subcomplex, both bound to the telomeric DNA with higher affinity compared with the non-telomeric DNA (Table 3.6). Interestingly, the full shelterin complex had a higher affinity for telomeric DNA compared to shelterin lacking TRF1 (K_D = 3.11 nM vs 11.1 nM, respectively), whereas the opposite was observed in the case of non-telomeric DNA where the (–TRF1) subcomplex had a lower K_D value compared with the full complex (K_D = 97 nM vs 211 nM, respectively) (Table 3.6). This can be explained by the varying rates of association; the full shelterin complex had a 3.3 times higher rate of association for telomeric DNA than the (–TRF1) subcomplex, and opposite was observed for the non-telomeric DNA where the (–TRF1) subcomplex had a 2.7 times higher rate of association compared with the full complex (Figure 3.20 and Table 3.6). Both complexes had similar rates of dissociation from telomeric and non-telomeric DNA. Together, these results indicated that the absence of TRF1 influenced the loading of shelterin onto DNA, but it did not significantly alter the stability of the resulting shelterin-DNA complexes.

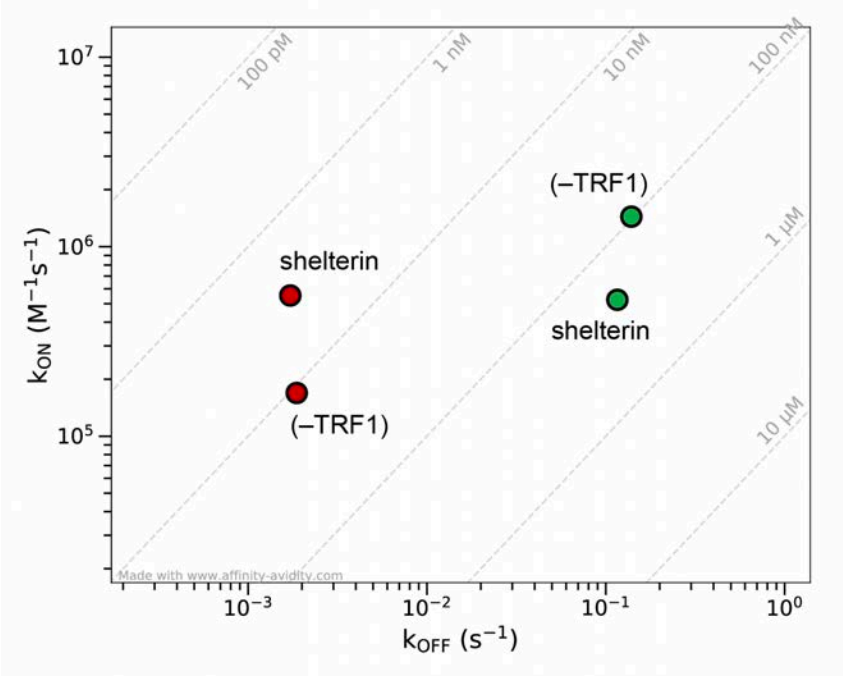

Figure 3.20. Binding kinetics of the full and (–TRF1) shelterin complexes for telomeric and non-telomeric DNA.

This plot maps the rate constant for association (k_{ON}) and dissociation (k_{OFF}) events were determined for the full and (–TRF1) shelterin^{TIN2L} complexes for telomeric DNA (red) and non-telomeric DNA (green) using switchSENSE (see Figure 2.18 and Figure 3.18).

Protein complex	DNA	k_{ON} [M^{-1} s^{-1}]	k_{OFF} [s^{-1}]	K_D [nM]
shelterin^{TIN2L}	teloDNA1	5.53 ± 0.34 x 10^5	1.72 ± 0.04 x 10^{-3}	3.11 ± 0.21
	mutDNA1	5.25 ± 0.25 x 10^5	1.16 ± 0.03 x 10^{-1}	211 ± 12
(–TRF1)^{TIN2L}	teloDNA1	1.69 ± 0.09 x 10^5	1.87 ± 0.04 x 10^{-3}	11.1 ± 0.7
	mutDNA1	1.44 ± 0.07 x 10^6	1.39 ± 0.04 x 10^{-1}	96.7 ± 5.1

Table 3.6. Kinetic rate and equilibrium dissociation constants determined for interactions between shelterin complexes and DNA.

The rate constant for association (k_{ON}) and dissociation (k_{OFF}) events, and the equilibrium dissociation constant (K_D) were determined for the full and (–TRF1) shelterin^{TIN2L} complexes binding to telomeric and non-telomeric DNA using switchSENSE (see Figure 2.18 and Figure 3.18).

3.3 Purification of other shelterin subcomplexes

Besides the full shelterin complex and the subcomplex lacking TRF1, I expressed and purified other variations of shelterin subcomplexes, including one lacking the dsDNA binding subunit TRF2 and its partner RAP1 (–TRF2/RAP1), and one lacking the TPP1/POT1 heterodimer (–TPP1/POT1). Together with the full complex and (–TRF1) subcomplex, these variations will serve as comparisons to further our understanding of the assembly and function of shelterin complexes. The same two-step purification protocol was used to purify these shelterin subcomplexes, starting with strep-tag affinity chromatography to extract shelterin complexes containing dStrepII-tagged TIN2L. Finally, the affinity purified proteins were resolved by SEC using the Superose 6 Increase 10/300 GL column.

3.3.1 Purification of the shelterin (–TRF2/RAP1) subcomplex

Expression of the 4-subunit shelterin (–TRF2/RAP1) subcomplex in insect cells, followed by the two-step purification reconstituted the apo (–TRF2/RAP1) subcomplexes (Figure 3.21). The minimal (–TRF2/RAP11)^{TIN2L} subcomplex was expected to have a $(TRF1)_2$/(dStrepII-TIN2L/TPP1/POT1)$_1$ stoichiometry, with a predicted MW of 283.7 kDa (Figure 3.21A). Affinity and SEC purification of the apo complex produced a stable (–TRF2/RAP1) subcomplex (Figure 3.21B – E). Initial mass photometry analysis of SEC-purified (–TRF2/RAP1)^{TIN2L} under native conditions showed five different peaks (Figure 3.22 and Table 3.7). The smaller species at 101 and 204 kDa likely resulted from dissociation of the (–TRF2/RAP1) subcomplex during sample dilution. The peak at 318 kDa likely corresponded with the minimal shelterin (–TRF2/RAP1) complex containing $TRF1_2$/(dStrepII-TIN2L/TPP1/ POT1)$_1$ at the expected MW of 283.7 kDa, or potentially with an additional dStrepII-TIN2L, giving a $(TRF1/dStrepII\text{-}TIN2L)_2$/TPP1/POT1 complex of 337.7 kDa. There were two higher order species detected by mass photometry (Figure 3.21D). The species at 389 kDa could be explained by two TRF1 homodimer, giving $(TRF1)_4$/dStrepII-TIN2L/TPP1/POT1 (predicted MW: 384.2 kDa), and the peak at 531 kDa may include extra dStrepII-TIN2L to give a

(TRF1/dStrepII-TIN2L)$_4$/TPP1/POT1 complex (predicted MW: 546.3 kDa). These results indicate that higher order shelterin complexes can form in the absence of TRF2/RAP1 and support the observation that shelterin subunits can self-assemble without telomeric DNA.

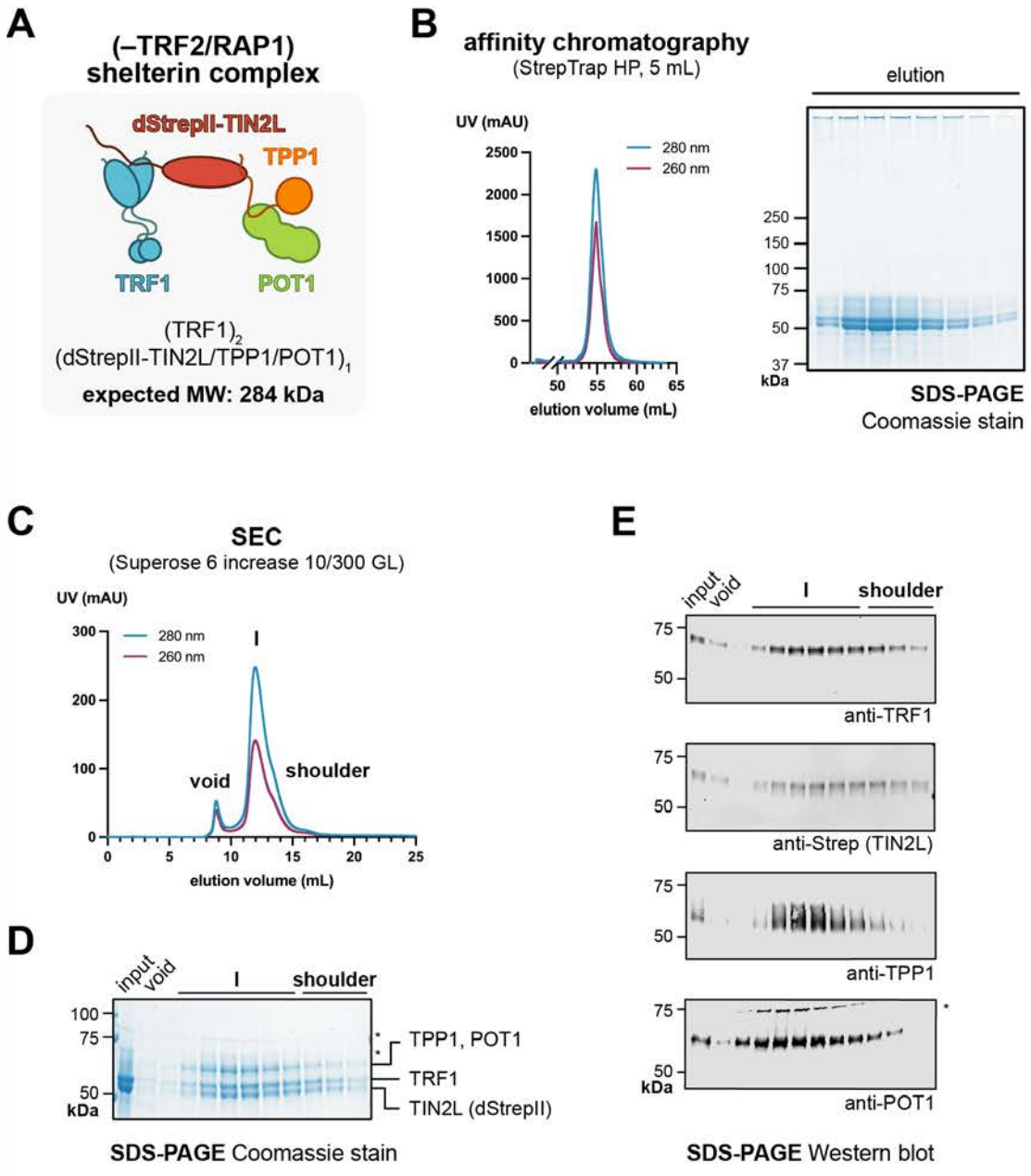

Figure 3.21. Purification of the shelterin (–TRF2/RAP1) complex.

(A) Schematic of the minimal (–TRF2/RAP1)^{TIN2L} shelterin complex. **(B)** UV chromatogram from the affinity purification of (–TRF2/RAP1)^{TIN2L}, and the corresponding SDS-PAGE analysis of the elution fractions. **(C)** UV chromatogram from the SEC purification, and the corresponding **(D)** SDS-PAGE and **I** Western blot analysis of the SEC input and SEC elution fractions. *Contaminant.

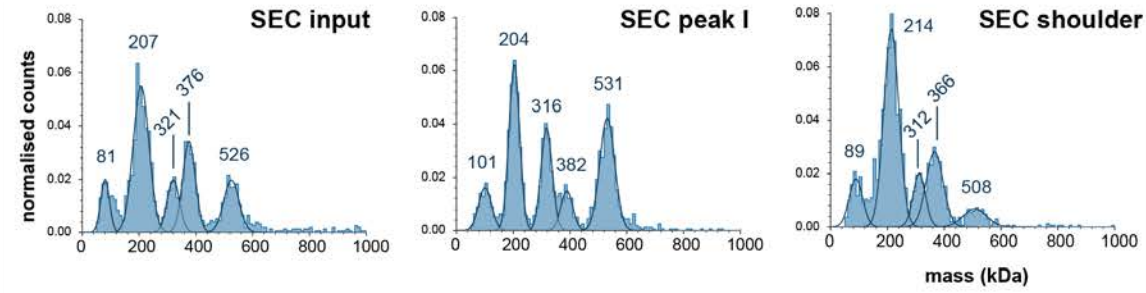

Figure 3.22. Mass photometry analysis of shelterin (–TRF2/RAP1).

Mass photometry measurements of SEC input and elution fractions from SEC peak ***I*** and shoulder, as shown in Figure 3.21C and D), under native conditions.

Shelterin complex/subcomplex	Predicted mass (kDa)	Measured mass (kDa)
$(TRF1)_2$	100.50	81 – 101
$(TRF1/dStrepII\text{-}TIN2L)_2$	208.56	204 – 214
$(TRF1)_2$/dStrepII-TIN2L/TPP1/POT1	283.70	312 – 321
$(TRF1/dStrepII\text{-}TIN2L)_2$/TPP1/POT1	337.73	
$(TRF1)_4$/dStrepII-TIN2L/TPP1/POT1	384.20	366 – 382
$(TRF1/dStrepII\text{-}TIN2L)_4$/TPP1/POT1	546.29	508 – 531

Table 3.7. Stoichiometries of shelterin (–TRF2/RAP1) subcomplexes.

The species detected in the mass photometry analysis of $(–TRF2/RAP1)^{TIN2L}$ from the SEC input, SEC peak ***I*** *and* SEC shoulder (see Figure 3.22) are listed, alongside potential stoichiometries that can explain the measured masses.

Recombinant affinity-purified shelterin $(–TRF2/RAP1)^{TIN2L}$ subcomplexes also bound to telomeric DNA (Figure 3.23). In EMSA studies, traces of a (–TRF2/RAP1)-DNA complex was first observed at 0.25:1 molar ratio of protein to DNA (Figure 3.23A). All the DNA was engaged when the shelterin (–TRF2/RAP1) complex was in 2-times molar excess of the DNA (Figure 3.23A). The DNA-bound $(–TRF2/RAP1)^{TIN2L}$ complex was also stable during SEC purification (Figure 3.23B). After incubation with telomeric DNA, the increased 260/280 nm UV absorbance ratio across the SEC elution indicated that DNA was incorporated into the shelterin (–TRF2/RAP1) complex (Figure 3.23B). In the presence of DNA, the (–TRF2/RAP1) subcomplex eluted earlier compared with the apo complex

(see Figure 3.21D and Figure 3.23D), reflecting a change in hydrodynamic radius of the complex after DNA binding. The EMSA analysis showed the presence of one DNA-bound complex in fractions across peak ***I*** (Figure 3.23D).

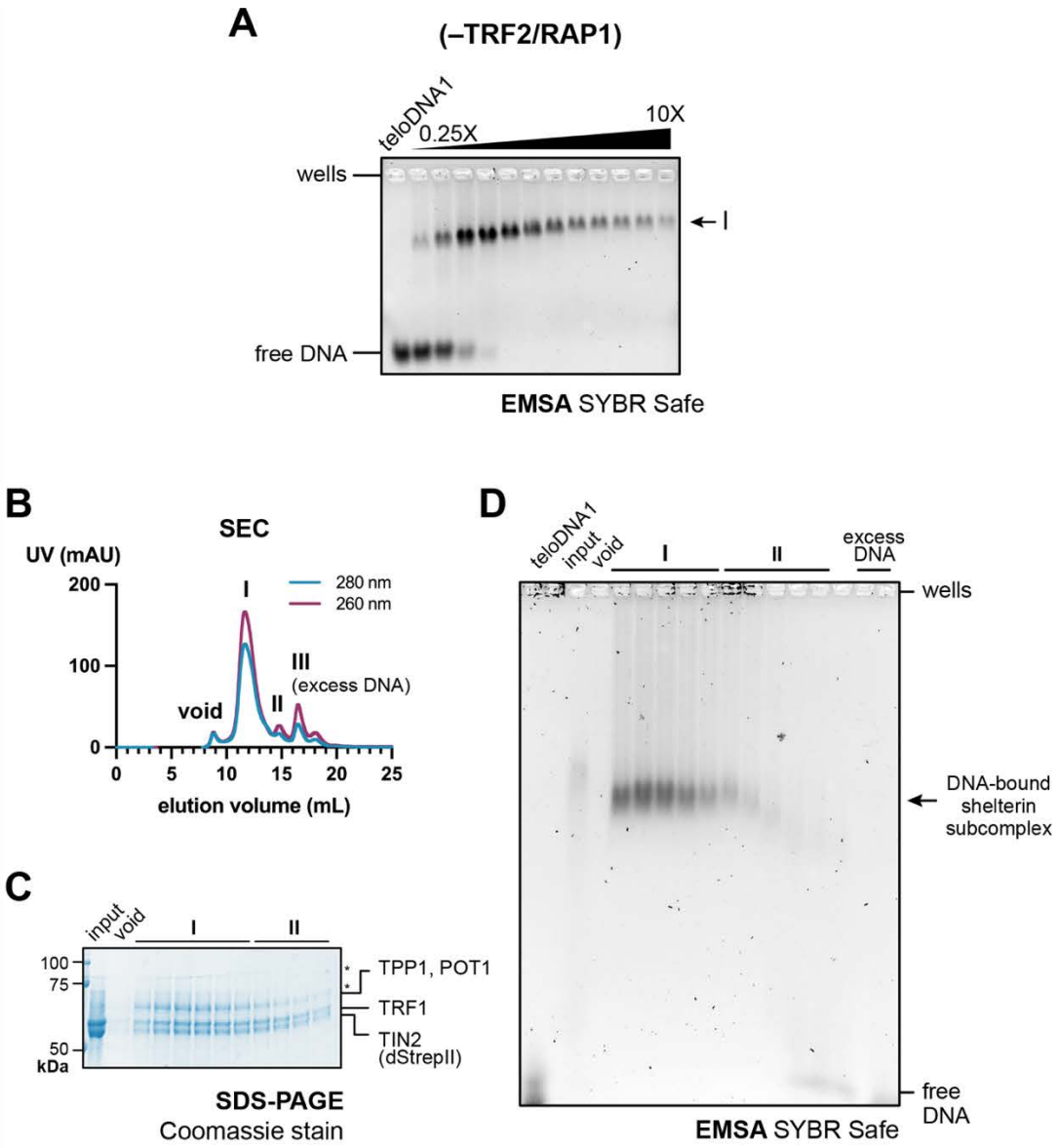

Figure 3.23. Recombinant shelterin (–TRF2/RAP1) binds to telomeric DNA.

(A) EMSA analysis of shelterin (–TRF2/RAP1)^{TIN2L} bound to teloDNA1. Various concentrations (0, 0.25, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 10-times molar excess) of affinity-purified (–TRF2/RAP1)^{TIN2L} subcomplex were incubated with 0.5 µM of telomeric DNA, teloDNA1, and resolved by a 0.6% agarose 0.5XTBE gel. **(B)** Analytical SEC analysis of DNA-bound (–TRF2/RAP1)^{TIN2L}. Affinity-purified subcomplex was incubated with 1.2-times molar excess of teloDNA1, and subjected to SEC. The SEC input and elution fractions were further analysed by **(C)** SDS-PAGE and **(D)** EMSA.

3.3.2 Purification of shelterin (–TPP1/POT1) subcomplexes

Expression of the 4-subunit shelterin subcomplex lacking TPP1/POT1 in Sf9 insect cells, followed by the two-step purification reconstituted the apo (–TPP1/POT1) subcomplex (Figure 3.24). The minimal $(–TPP1/POT1)^{TIN2L}$ subcomplex was expected to have a $(TRF1/TRF2/RAP1)_2$/dStrepII-TIN2L stoichiometry, with a predicted MW of 362.2 kDa (Figure 3.24A). During Strep-tag affinity chromatography, the elution fractions was enriched in dStrepII-TIN2L and TRF1 proteins (Figure 3.24B). These two subunits formed a stable complex (peak ***II***), which was resolved from the (–TPP1/POT1) subcomplex (peak ***I***) during SEC (Figure 3.24C). Mass photometry analysis of $(–TPP1/POT1)^{TIN2L}$ under native conditions showed up to four different peaks (Figure 3.25 and Table 3.8).

Peak ***I*** was consistently smaller compared with the peak ***II*** (Figure 3.24C). SDS-PAGE analyses showed that peak ***I*** contained all the expected shelterin subunits, whereas the peak ***II*** contained lower amounts of TRF2 and RAP1. The final peak consisted of mainly the dStrepII-tagged subunit (TRF1 or TIN2). Mass photometry analyses showed the presence of species near the expected MW for a minimal (–TPP1/POT1) complex in peak ***I*** SEC fractions (predicted MW for $TRF1_2/TRF2_2/RAP1_2$/dStrepII-$TIN2L_1$ is 362 kDa). Peak ***II*** consisted mainly of a complex at ~205 kDa for samples containing dStrepII-TIN2 (Figure 3.25). As this peak was enriched in TRF1 and TIN2, shown on SDS-PAGE, this MW value could only be explained by the $TRF1_2$/dStrepII-$TIN2L_2$ combination (expected MW: 209 kDa).

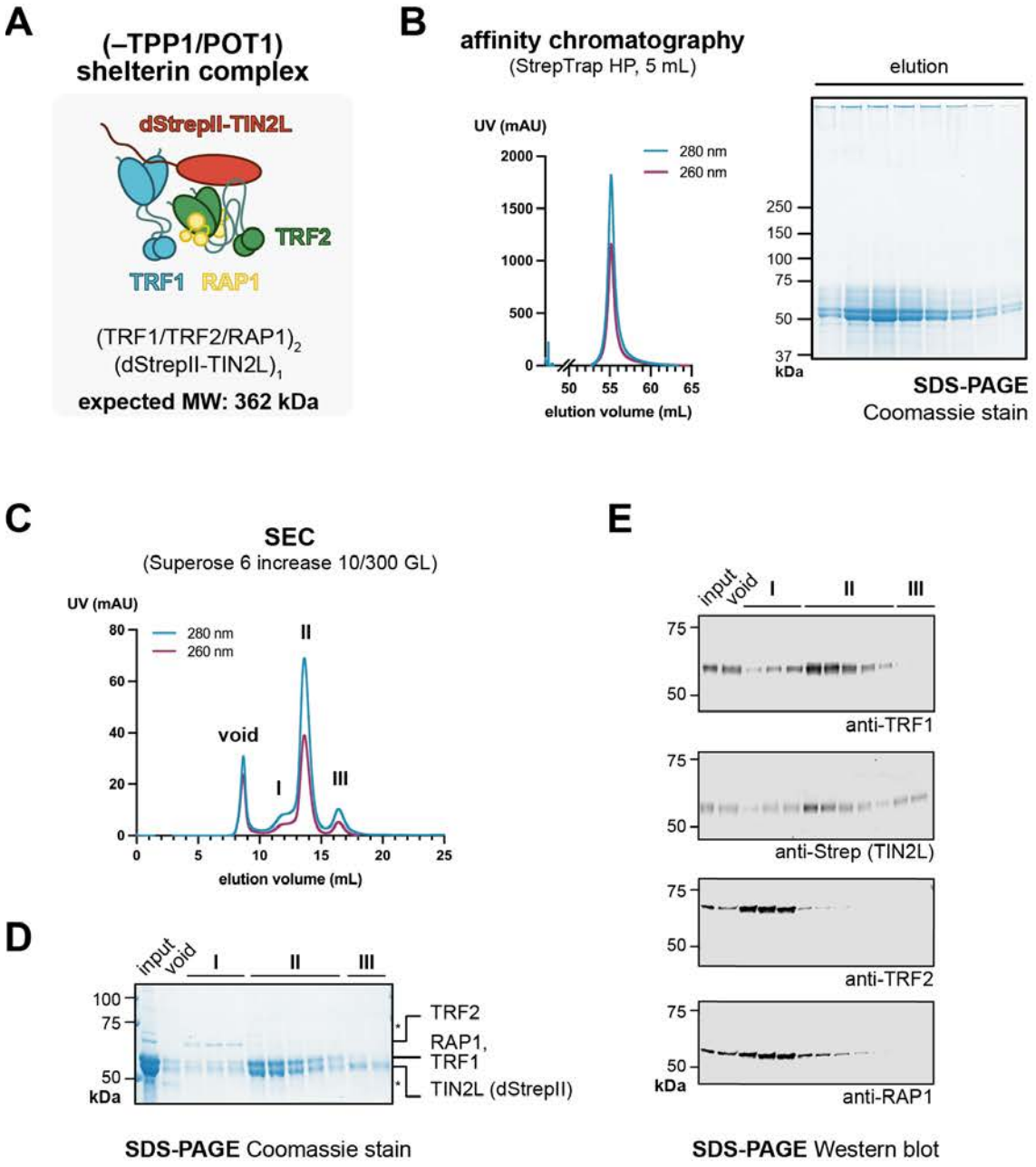

Figure 3.24. Purification of the shelterin (–TPP1/POT1) complex.

(4) Schematic of the minimal (–TPP1/POT1)^{TIN2L} shelterin complex. **(B)** UV chromatogram from the affinity purification of (–TPP1/POT1)^{TIN2L}, and the corresponding SDS-PAGE analysis of the elution fractions. **(C)** UV chromatogram from the SEC purification, and the corresponding **(D)** SDS-PAGE and **I** Western blot analysis of the SEC input and SEC elution fractions. *Contaminants and non-specific band.

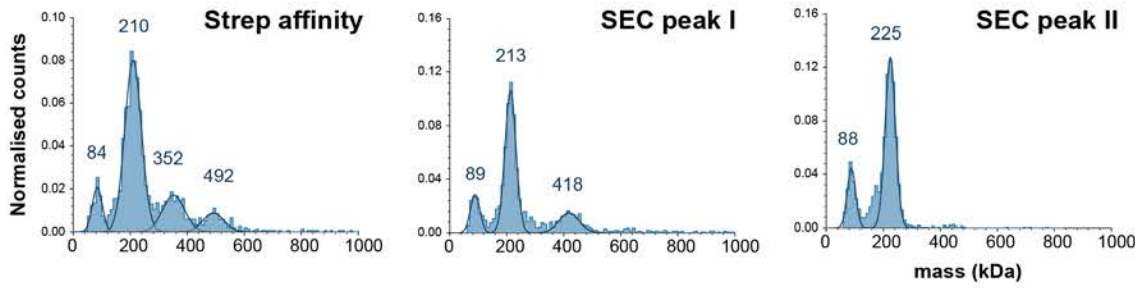

Figure 3.25. Mass photometry analysis of shelterin (–TPP1/POT1).

Mass photometry measurements of SEC input and elution fractions from SEC peak ***I*** and ***II***, as shown in Figure 3.24C and D), under native conditions.

Shelterin complex/subcomplex	Predicted mass (kDa)	Measured mass (kDa)
$(TRF1)_2$	100.50	84 – 88
$(TRF1/dStrepII\text{-}TIN2L)_2$	208.56	210 – 225
$(TRF1/TRF2/RAP1)_2/dStrepII\text{-}TIN2L$	362.23	352
$(TRF1/TRF2/RAP1/dStrepII\text{-}TIN2L)_2$	416.26	418
$(TRF1)_4/(TRF2/RAP1/dStrepII\text{-}TIN2L)_2$	516.76	492

Table 3.8. Stoichiometries of shelterin (–TPP1/POT1) subcomplexes.

The species detected in the mass photometry analysis of shelterin $(–TPP1/POT1)^{TIN2L}$ from the SEC input, SEC peak ***I*** *and* ***II*** (see Figure 3.25) are listed, alongside potential stoichiometries that can explain the measured masses.

The recombinant shelterin $(–TPP1/POT1)^{TIN2L}$ subcomplex bound to telomeric DNA (Figure 3.26). In EMSA studies, traces of a (–TPP1/POT1)-DNA complex was first observed at 0.25:1 molar ratio of protein to DNA (Figure 3.26A). All the DNA was engaged when the $(–TPP1/POT1)^{TIN2L}$ subcomplex was in 2-times molar excess of the DNA (Figure 3.26A). The DNA-bound $(–TPP1/POT1)^{TIN2L}$ complex was not stable during SEC purification under the conditions tested (Figure 3.26B). After incubation with DNA, the elution profile remained the same as the apo complex, and the 260/280 nm UV absorbance ratio of the SEC elution fractions did not change (Figure 3.26B and C). Moreover, the DNA eluted separately from the protein peaks (Figure 3.26B). These factors indicated that DNA was not stably bound to the shelterin (–TPP1/POT1) complex.

The EMSA analysis showed that DNA was bound to proteins within the SEC input, but this was lost during the SEC elution (Figure 3.26D).

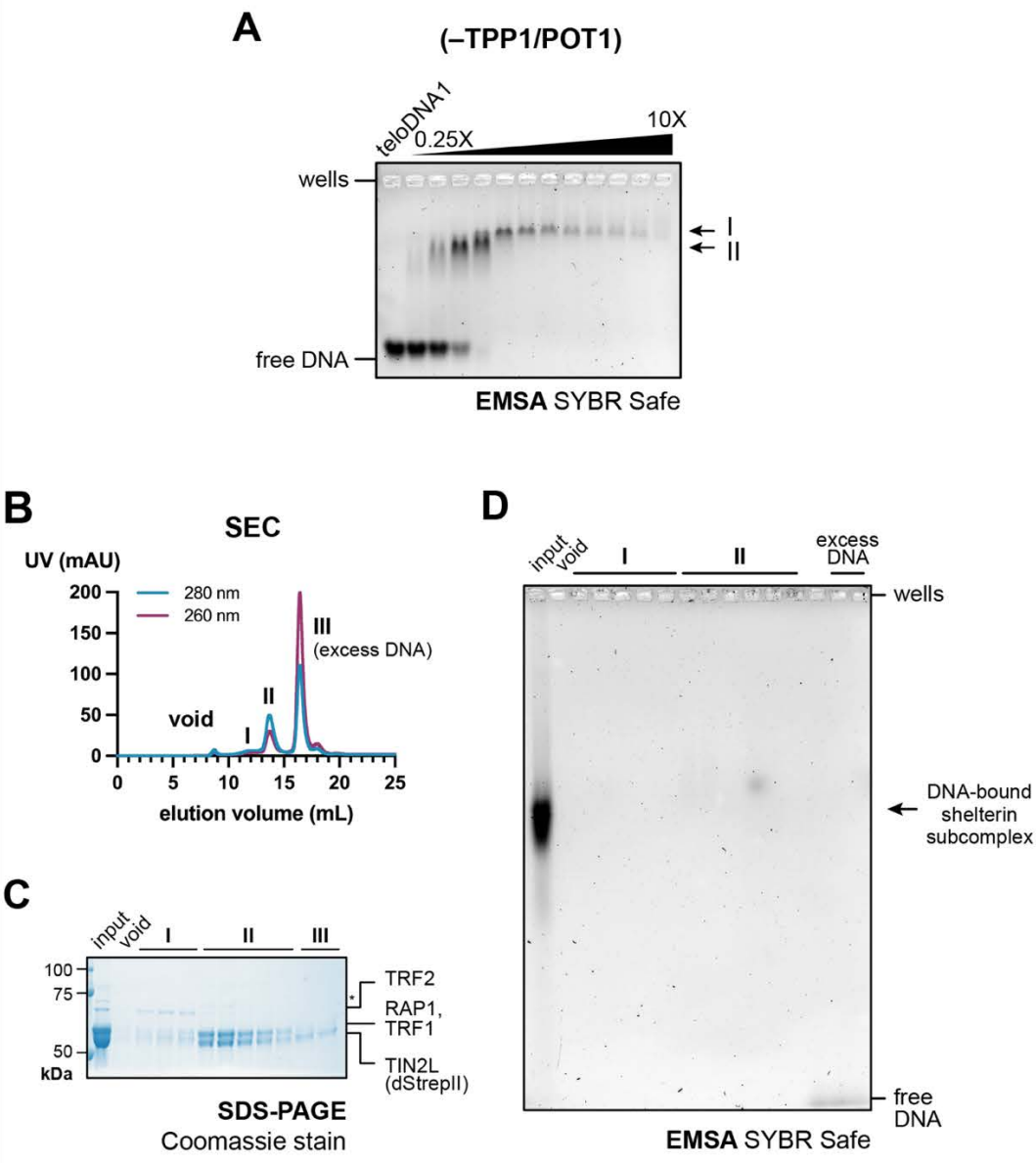

Figure 3.26. Recombinant shelterin (–TPP1/POT1) binds to telomeric DNA.

(4) EMSA analysis of shelterin (–TPP1/POT1)^{TIN2L}. Various concentrations (0, 0.25, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 10-times molar excess) of affinity-purified (–TPP1/POT1)^{TIN2L} subcomplex were incubated with 0.5 µM of telomeric DNA, teloDNA1. **(B)** UV chromatogram from the analytical SEC analysis of DNA-bound (–TPP1/POT1)^{TIN2L}. Affinity-purified subcomplex was incubated with 1.2-times molar excess of teloDNA1, and subjected to SEC. The SEC input and SEC elution fractions were further analysed by **(C)** SDS-PAGE and **(D)** EMSA. Protein-DNA complexes in EMSA analyses were resolved using 0.6% agarose 0.5X TBE gels.

3.4 Conclusion

In this chapter, I described the purification of three subcomplexes of human shelterin either lacking the dsDNA-binding modules TRF1 or TRF2/RAP1, or the ssDNA-binding module TPP1/POT1. Of these, subcomplexes lacking the dsDNA-binding modules could be expressed and purified from insect cells with higher yield compared with subcomplexes lacking TPP1/POT1. This supports previous studies showing that TPP1 is involved in forming stable shelterin complexes containing both TRF1 and TRF2 (O'Connor, et al., 2006; Janovič, et al., 2019).

Similar to the full complex, recombinantly expressed shelterin subunits formed multiple species of each subcomplex. In the case of the (–TRF1) subcomplex, two stable species were detected. The ~390 kDa species is consistent with the expected stoichiometry of one homodimer of TRF2, two RAP1 subunits, and a 1:1:1 ratio of TIN2L/TPP/POT1, as proposed previously (Lim, et al., 2017). Interestingly, I also observe another species at ~600 kDa. Given the known protein-protein interactions within shelterin, two combinations that may explain the additional MW of ~200 kDa are $(TRF2/RAP1)_2$ and $(dStrepII\text{-}TIN2L/TPP1/POT1)_1$. The former is possible given there are two TRF-binding sites in TIN2, namely $TIN2_{TBM}$ that interacts with the TRFH domains in TRF1 and TRF2, and the $TIN2_{TRFH\text{-}like}$ domain that binds to the linker region of TRF2. In the absence of competition from TRF1, $TIN2_{TBM}$ may recruit another TRF2 homodimer into the complex. However, supplementing the 390 kDa (–TRF1) complex with dStrepII-TRF2/dStrepII-RAP1 did not enrich the higher-MW species. Moreover, inhibiting the $TIN2_{TBM}$-$TRF2_{TRFH}$ interaction through point mutations in the conserved TBM in TIN2 (F-x-**L**-x-P, $TIN2^{L260A}$ and $TIN2^{L260E}$) did not abolish the formation of the higher-order species. These findings indicate that the size difference between the minimal and higher-MW (–TRF1) complex is unlikely to be mediated via the $TIN2_{TBM}$. Interestingly, higher-order species were observed for all three shelterin subcomplexes and the six-subunit full complex in the absence of DNA (sections 2.2.2). This suggests that other protein-protein interactions may mediate the recruitment of additional shelterin subunits. One

possibility is that TIN2, the central hub of shelterin and the subunit present in all the recombinant shelterin complexes studied, may bind to shelterin subunits via alternative binding sites. Splicing variants and post-translational modifications (PTMs), including the isoform and phosphorylation status of TIN2, has been suggested to play a role in regulating protein-protein interactions within shelterin and telomere length maintenance (Nelson, et al., 2018). The range of shelterin complexes with TIN2L or TIN2S generated and characterised in this work provides useful tools for further investigation into how different shelterin isoforms and PTMs contribute to shelterin assembly and function at the telomeres.

All three variations of shelterin subcomplexes were able to bind telomeric DNA. Among these, the (–TRF1) and (–TRF2/RAP1) shelterin subcomplexes formed stable DNA-bound complexes that could be further purified by SEC. These will be useful for further structural characterisation of shelterin complexes. In addition, I studied the interaction between the shelterin $(–TRF1)^{TIN2L}$ complex and DNA using switchSENSE. In the absence of TRF1, shelterin can bind to both telomeric and non-telomeric DNA in a 1:1 ratio with nanomolar affinity (K_D = 11.1 vs 96.7 nM, respectively). Under the conditions tested, shelterin (–TRF1) only had an 8.7-fold higher affinity for telomeric DNA compared with the non-telomeric DNA, compared with 71-fold difference observed for the full shelterin complex (see section 3.2.3). This can be mainly attributed to the difference between the foldchange in association rate constants for these two DNA constructs (1.1-fold vs. 8.5-fold for the full and (–TRF1) complex, respectively). This suggested that TRF1 contributes to determining the sequence specificity of shelterin. The next step would be to compare this with the (–TRF2/RAP1) and (–TPP1/POT1) subcomplexes, and other telomeric substrates including branched DNA and T-loops, to further investigate the contribution of the DNA-binding shelterin subunits to telomere association.

4 Characterising protein-protein interactions within human shelterin

After obtaining the holocomplex and subcomplexes of human shelterin, I aimed to further understand the architecture of shelterin using two approaches: negative-stain electron microscopy (EM) and crosslinking mass spectrometry (XL-MS). In the past two decades, many high-resolution structures of shelterin domains encompassing the known protein interaction interfaces and DNA-binding regions have been reported (Appendix Figure 7.4). The next step was to understand how the subunits are arranged to assemble the shelterin complex on telomeric DNA. Here, I describe structural studies of the shelterin (–TRF1) subcomplex using negative-stain EM. I chose XL-MS to complement our EM studies as it offers insight into the organisation of shelterin subunits in solution, relaying information on regions in spatial proximity in the 3D structure. See the Methods for more about each technique and details on sample preparation, data acquisition and analysis (section 6.6).

4.1 Structural characterisation of the shelterin (–TRF1) subcomplex using negative-stain EM

4.1.1 Overview

To study the architecture of the shelterin complex by EM, we decided to take a subtractive EM approach. This involved EM studies of the full shelterin complex, initiated by Dr Iris Gawarzewski and Dr Peter Saiu and continued by Dr Oviya Inian, accompanied by EM studies of the subcomplexes to help identify positions of shelterin subunits within the full complex. I chose to focus on the (–TRF1) subcomplex first for negative-stain EM as the two-step purification protocol produced more homogeneous complexes compared to the other subcomplexes (sections 3.2 and 3.3). Moreover, a shelterin complex lacking TRF1 may have

physiological relevance in the context of tankyrase-regulated telomere length maintenance (section 1.4.2), making it a particularly interesting structural target for EM.

4.1.2 Optimising the sample preparation using GraFix

For negative-stain EM analysis of the (–TRF1) subcomplex, I adapted the two-step purification protocol previously described for the reconstitution and further purification of the DNA-bound complex (Figure 4.1). Initial optimisations of sample preparation using ultracentrifugation were done using the $(–TRF1)^{TIN2L}$ complex. I later switched to the subcomplex containing TIN2S as the C-terminal region of TIN2L is predicted to be disordered, and likely will not be resolved by EM analysis. Moreover, the extra ~100 amino acids in the C-terminus appeared dispensable for the assembly of shelterin complexes, judged by the formation of similar TIN2S- and TIN2L-containing shelterin complexes in the absence and presence of telomeric DNA (see sections 2.1.2 and 3.2).

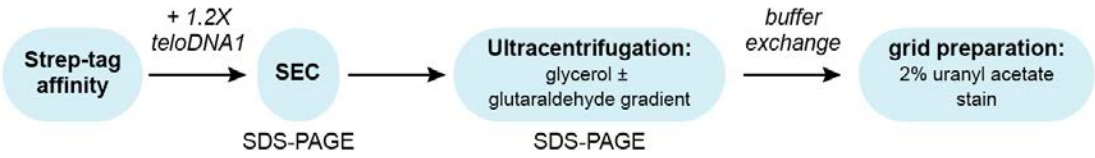

Figure 4.1. Schematic of the initial sample preparation using GraFix.

After the Strep-tag affinity purification, the peak fraction was incubated with teloDNA1 in 1.2-fold molar excess to saturate the $(–TRF1)^{TIN2L}$ complexes with DNA. This sample was further purified over SEC, which separated the excess DNA from the protein-DNA complexes (Figure 4.2A). The increase in the 260/280 nm UV absorption ratio compared with apo $(–TRF1)^{TIN2L}$ indicated that the DNA was incorporated into the $(–TRF1)^{TIN2L}$ subcomplexes. SDS-PAGE analysis showed the elution fractions from peak ***I*** contained the full (–TRF1)

subcomplex (Figure 4.2B). Next, fractions from the middle of SEC peak I were pooled and subjected to ultracentrifugation in a 15 – 50% glycerol gradient without or with gradient fixation (GraFix) with 0 – 0.2% glutaraldehyde (Figure 4.2C). GraFix was used to stabilise the complex by the gradual increase in crosslinker concentration whilst being purified by density gradient ultracentrifugation (Kastner, et al., 2008). The gradients were fractionated from the top and analysed by SDS-PAGE to select fractions for grid preparation (Figure 4.2D). During GraFix, $(–TRF1)^{TIN2L}$ complexes migrated further down the glycerol gradient by 2 fractions (~100 µL) and were exposed to <0.1% of glutaraldehyde (Figure 4.2D, right). Around this crosslinker concentration, all shelterin subunits were crosslinked as indicated by the disappearance of individual subunits running between 50 – 75 kDa and the appearance of bands at higher MW above the 250 kDa marker. After crosslinking, two species were observed on SDS-PAGE gel, and the smaller-MW species was resolved from the higher-MW species (compare fractions 5 – 8, Figure 4.2D). Fractions 5 and 8 were used to make negative-stain grids to see if a difference in size could be detected during image analysis.

Negative-stain grids were made using fractions from SEC peak ***I*** and ***II***, and the non-crosslinked complex from the glycerol gradient as comparisons (Figure 4.2E). Initial grid screening showed that the (–TRF1)-DNA complexes were unstable without crosslinking (Figure 4.2E). Few intact particles were found against a high background, presumably of dissociated shelterin components. On the other hand, in the presence of a low concentration of glutaraldehyde (<0.1%), prominent particles of similar sizes were observed (Figure 4.2E, GraFix fractions). In agreement with previous SDS-PAGE analysis (Figure 4.2D), GraFix fraction 5 contained smaller particles with dimensions of ~19.5 nm by 13.2 nm. Particles in GraFix fraction 8 appeared to be heterogenous in size, with the larger particles measuring ~23.0 nm by 17.7 nm (Figure 4.2E). In summary, this experiment showed that crosslinking was necessary to stabilise the DNA-bound shelterin $(–TRF1)^{TIN2L}$ subcomplex for negative-stain EM analysis.

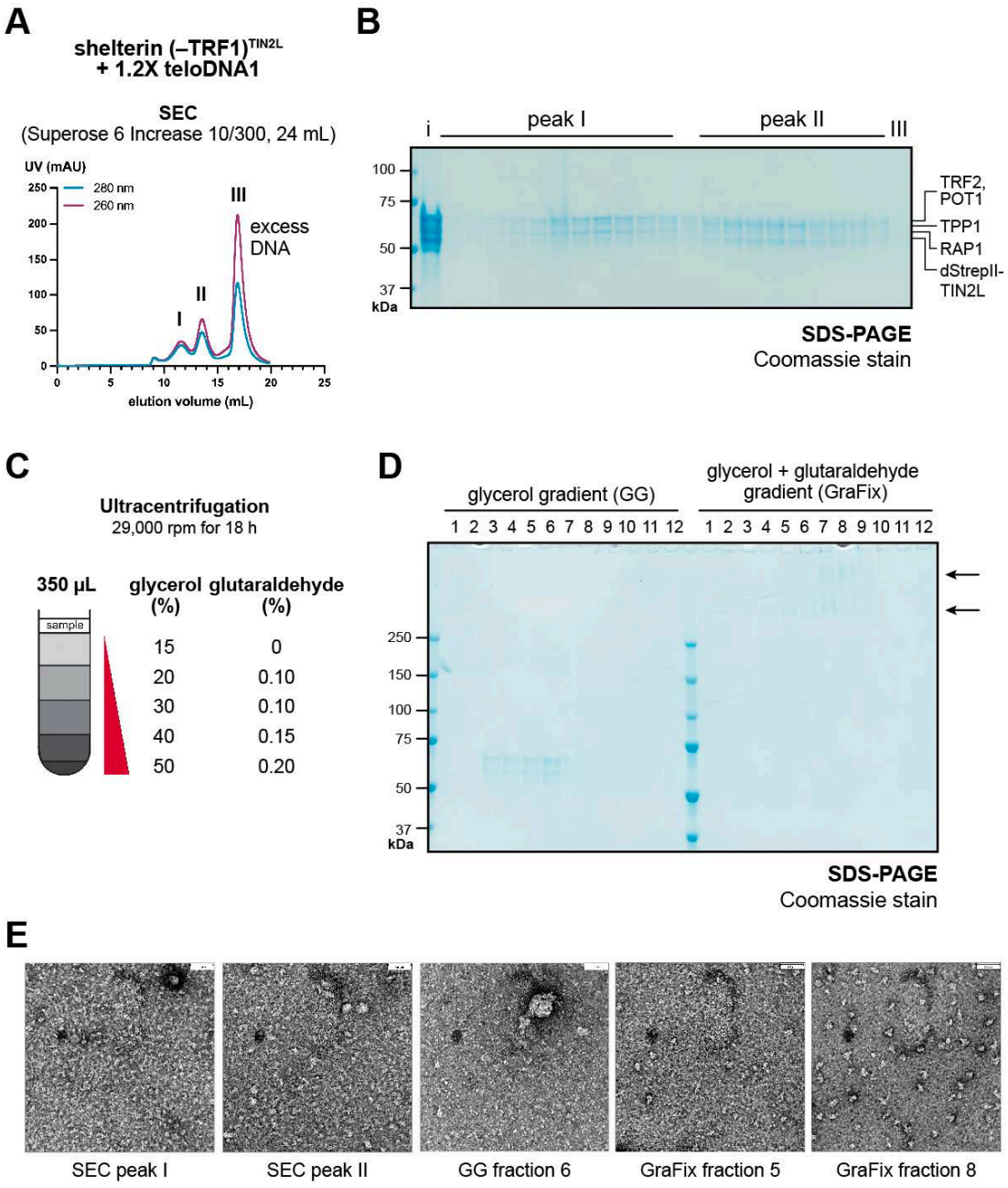

Figure 4.2. Preparing DNA-bound shelterin (–TRF1) for negative-stain EM analysis using GraFix.

(A) SEC elution profile of the (–TRF1)-DNA complex. 500 µL of affinity-purified $(–TRF1)^{TIN2L}$ was incubated with 1.2-fold molar excess of teloDNA1 on ice for 1 h and purified using SEC, eluting in 200-µL fractions. **(B)** SDS-PAGE analysis of every other SEC elution fraction from (A). I – input. Arrows indicate the two crosslinked (–TRF1) species. **(C)** Ultracentrifugation and GraFix. 350 µL of the (–TRF1)-DNA complex from SEC peak I was resolved using a 15 – 50% glycerol gradient (GG) with or without glutaraldehyde (0 – 0.2%). **(D)** SDS-PAGE analysis of fractions from ultracentrifugation. **I** Screening of negative-stain grids made with SEC peak I and II, GG fraction 6, and GraFix fractions 5 and 8. Scale bar = 50 nm.

4.1.3 Improving the homogeneity of DNA-bound shelterin (–TRF1) complexes using limited crosslinking

Whilst I performed the above analysis, my colleague Dr Peter Saiu showed that crosslinking using amine-reactive crosslinkers, such as glutaraldehyde, displaced DNA from the shelterin complex (data not shown). This can be explained by the presence of lysine residues in the Myb domains of TRF1 and TRF2, which form hydrogen bonds with telomeric dsDNA in the major and minor grooves (Figure 1.8A and B). Dr Saiu demonstrated that providing shelterin with a large excess DNA can overcome this problem, likely by increasing the probability of shelterin being bound to DNA at any time point during the crosslinking process. This is expected to protect the DNA-binding lysine residues in the TRF1/TRF2 Myb domains and in the POT1 OB-folds from glutaraldehyde. Another important factor that affected DNA binding was the crosslinker concentration; higher concentrations of glutaraldehyde displaced more telomeric DNA from the complex. Therefore, I performed a titration of the glutaraldehyde to the optimal crosslinker:protein ratio to stabilise shelterin (–TRF1) complexes whilst retaining the telomeric DNA (see section 6.6.1.7 for the detailed protocols).

Briefly, I incubated 1 µM of SEC-purified $(–TRF1)^{TIN2S}$ with 5-fold molar excess of telomeric DNA before treating the sample with various concentrations of glutaraldehyde (0.025 to 0.25%, Figure 4.3A). At a final glutaraldehyde concentration of 0.075%, all shelterin subunits were crosslinked and migrated as two smeared bands on the SDS-PAGE gel (Figure 4.3B). Although there were some proteins trapped in the wells for crosslinked samples, higher glutaraldehyde concentrations did not significantly drive the formation of higher-order species (Figure 4.3B). This was corroborated by the EMSA analysis, which also showed that glutaraldehyde crosslinking in the presence of excess DNA did not substantially reduce the amount of DNA retained by the $(–TRF1)^{TIN2S}$ subcomplex. However, crosslinking did induce compaction of the complex, thereby speeding up its migration through the agarose gel (Figure 4.3C, top). The strongest changes were observed between 0.025 and 0.075% of glutaraldehyde (lanes 10–12, Figure 4.3C). Beyond this range, further addition of the crosslinker did not impact the migration of protein-DNA complexes. Much of the non-crosslinked

DNA-bound complex was retained in the wells of the native polyacrylamide gel (Figure 4.3C, bottom). However, when crosslinked with a low glutaraldehyde concentration of 0.025%, the (–TRF1)^{TIN2S}-DNA complex entered the gel and was resolved into two different species. For future negative-stain EM sample preparations, I decided to proceed with using 0.075% glutaraldehyde as this was the minimum concentration required to stabilise the protein-protein interactions.

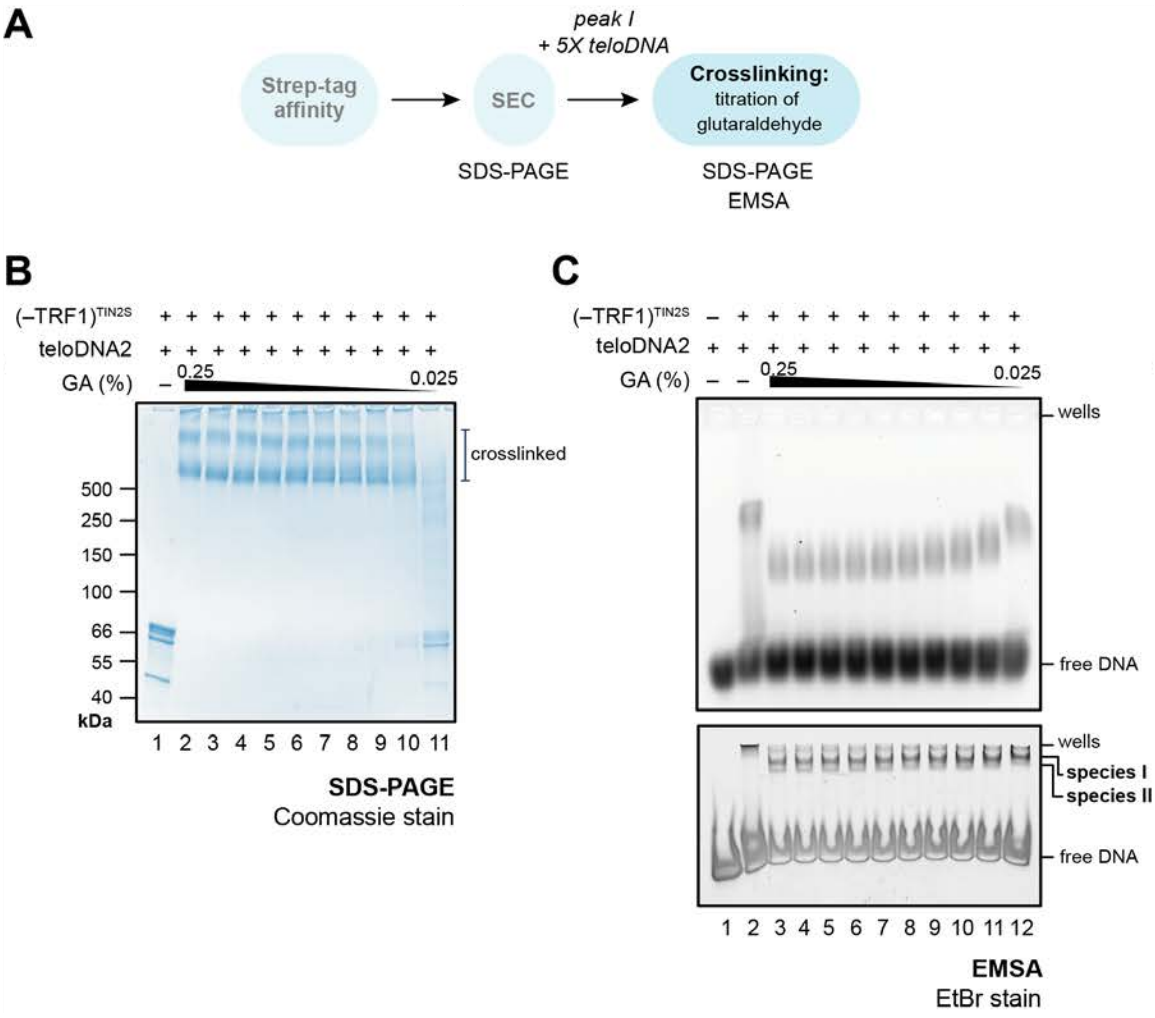

Figure 4.3. Limited crosslinking of DNA-bound shelterin (–TRF1).

(A) Schematic of the sample preparation process. 0.5 µM of SEC-purified (–TRF1)^{TIN2S} was incubated with various concentrations of glutaraldehyde (GA; 0, 0.025, 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.2, 0.225, 0.25%) for 30 min at 4 ºC. The reaction was quenched by adding Tris-HCl, pH 8.0 to a final concentration of 50 mM. **(B)** SDS-PAGE analysis of crosslinked samples using a 4 –15% Tris-glycine pre-cast gel. **(C)** EMSA analysis of crosslinked samples using a 0.5X TBE 1% agarose gel (top) and a 1X TBE 4 – 12% polyacrylamide gel (bottom).

4.1.4 Negative-stain EM analysis of DNA-bound (–TRF1)^{TIN2S}

After crosslinking, I subjected the sample to another round of SEC to remove the excess DNA. There was no protein aggregation, and the majority of protein-DNA complexes eluted in one symmetrical peak (Figure 4.4A). SDS-PAGE and EMSA analyses of the SEC fractions indicated the (–TRF1)^{TIN2S} complex was crosslinked and was bound to DNA (Figure 4.4B and C). Moreover, the EMSA analysis clearly showed the presence of two DNA-containing (–TRF1)^{TIN2S} complexes (Figure 4.4C). The earlier eluting species (starting from 11.9 mL) contained a complex that was larger in size and delayed the migration of DNA to a greater extent than the species from the main peak (starting from 12.7 mL) (Figure 4.4B). The fraction eluted at 12.7 mL appeared as one main band on EMSA and SDS-PAGE and was used to make negative-stain grids. Later, a mass photometry analysis of this fraction showed the presence of two populations of DNA-bound (–TRF1)^{TIN2S} complexes (Figure 4.4D). The 423 kDa species represented the minimal (–TRF1)^{TIN2S} complex with $(TRF2/RAP1)_2$/(dStrepII-TIN2S/TPP1/POT)$_1$, and a higher-order species was observed at 618 kDa. The micrographs showed crosslinked (–TRF1)-DNA complexes were uniformly distributed on the grid without any large aggregates (Figure 4.4E). However, the particles had different sizes that may represent complexes in different orientations and/or compositions, as indicated by the mass photometry analysis.

A dataset was collected using the FEI Tecnai F20 transmission electron microscope (TEM), operating at 200kV with a field emission gun (FEG) and a CMOS camera with a 1.732 Å pixel size. The raw micrographs were imported into cryoSPARC, and 55,342 particles were picked from 302 micrographs using the blob and template pickers (Figure 4.5A and B). Initial 2D classifications produced many classes of a triangular shape and a few rounder particles, which may represent different views and/or species of the DNA-bound (–TRF1)^{TIN2S} complex (Figure 4.5B). The 2D classes with larger and elongated particles measured at ~20.7 nm by 15.2 nm, which were similar to the dimensions observed during the screening stage (Figure 4.2E). Additionally, the classes with smaller rounded particles were ~11.7 nm by 13.7 nm (Figure 4.5B).

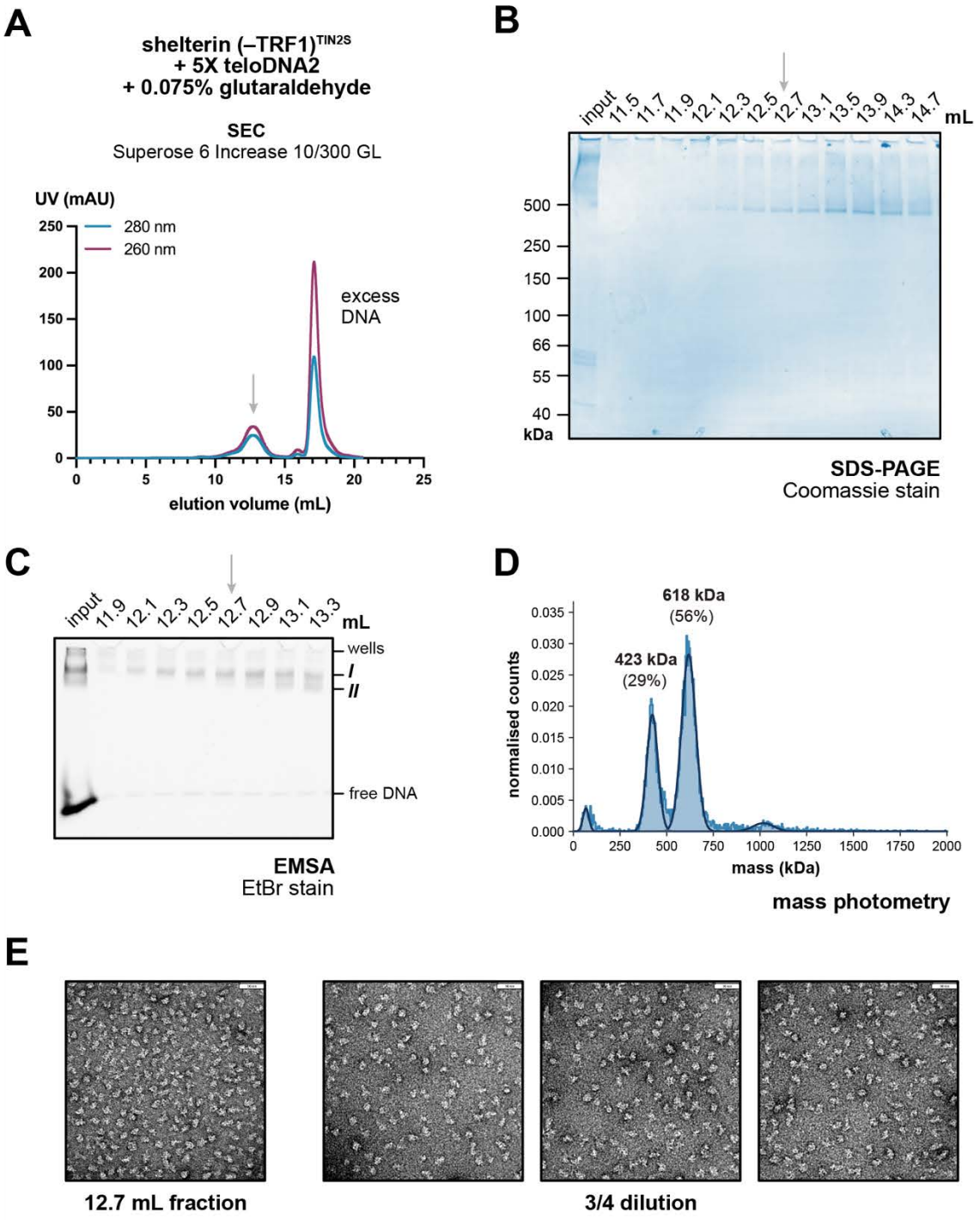

Figure 4.4. Screening negative-stain grids of crosslinked DNA-bound (–TRF1).

(A) SEC purification of crosslinked (–TRF1)^{TIN2S} bound to teloDNA2, which has 2.5 repeats of telomeric dsDNA and 2 repeats of ssDNA at the 3' overhang. **(B)** SDS-PAGE analysis of the SEC elution fractions. **(C)** EMSA analysis of the SEC elution fractions using a 1X TBE 4 – 12% polyacrylamide gel. **(D)** Mass photometry analysis of the SEC elution fraction at 12.7 mL. The maximum mass error is 6.1 %. **I** Screening of negative-stain EM grids made with the 12.7 mL fraction. Grids were made the sample before and after a ¾ dilution with buffer (20 mM HEPES pH 8.0). The grey arrows indicate the fraction used for negative-stain grid preparation.

Next, 2D classes that contained poorly aligned and/or few particles with low-resolution estimates (>40 Å), were removed. The remaining 36,553 particle images were used for 3D reconstruction (Figure 4.5C). As mass photometry analysis indicated the sample contained a mix of two species with different MWs (Figure 4.4D), I used the reference-free 3D *ab initio* reconstruction in cryoSPARC to generate one, two, or three low-resolution volumes. The majority of 3D volumes resembled the skull of a dinosaur, with a 'jaw' and 'nose' region that appeared consistently in multiple 3D *ab initio* reconstruction jobs (Figure 4.6A). The repeating features were apparent in multiple rounds of 2D classification and 3D *ab initio* reconstructions. There was a wide distribution of views along the azimuth axis (Figure 4.6B); most were within $\pm$ π/4 radians (45˚) in elevation, which represented side and oblique views of the (–TRF1) complex. Much fewer particle images were assigned as top view and bottom views ($\pm$ π/2 radians; 90˚). The number of particles assigned to each volume were approximately equal when two (46% and 54%) or three (33%, 28%, and 39%) initial volumes were generated (Figure 4.5C). While the overall shapes were similar, the detailed features varied between the initial reconstructions. This points to heterogeneity in DNA-bound shelterin (–TRF1), which can be due to variations in conformation and composition.

(figure on the next page)

Figure 4.5. Processing of negative-stain EM images of DNA-bound (–TRF1).

Representative examples of **(A)** raw micrographs and **(B)** 2D classes. Particles were initially picked using the blob picker to generate templates, which were used as input for the template picker. 55,342 particles were extracted and subjected to 2D classification (450 classes). **(C)** EM densities from three 3D *ab initio* reconstruction runs. 298 classes with 36,533 particles were selected and allocated to 1, 2, or 3 volumes for 3D *ab initio* reconstruction. The numbers below each 3D volume show the proportion and number of particles allocated to that 3D class. Volume 3 from the 3rd job (starred) was taken forward for further processing.

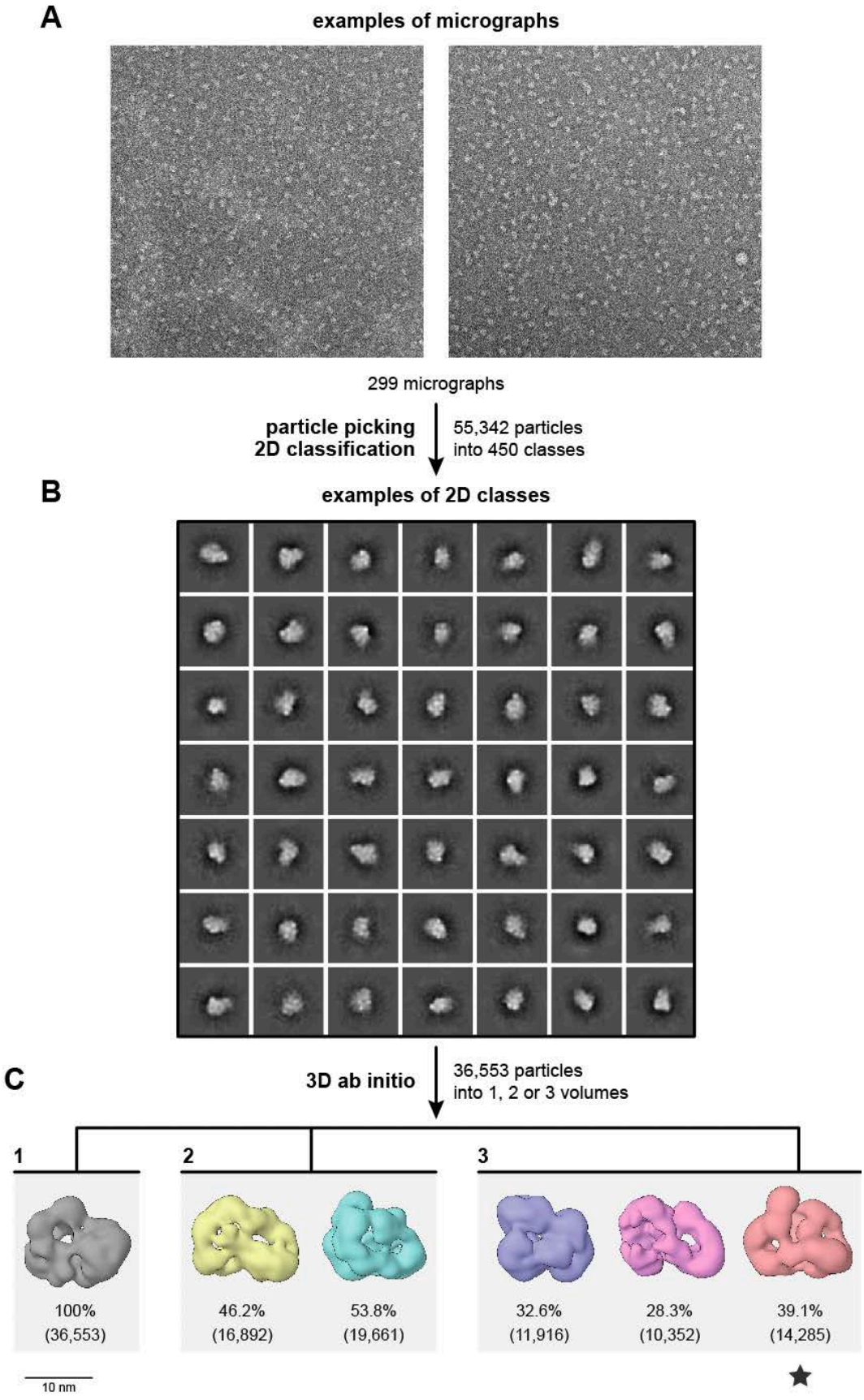

Figure 4.5. Processing of negative-stain EM images of DNA-bound (–TRF1).

(legend on the previous page)

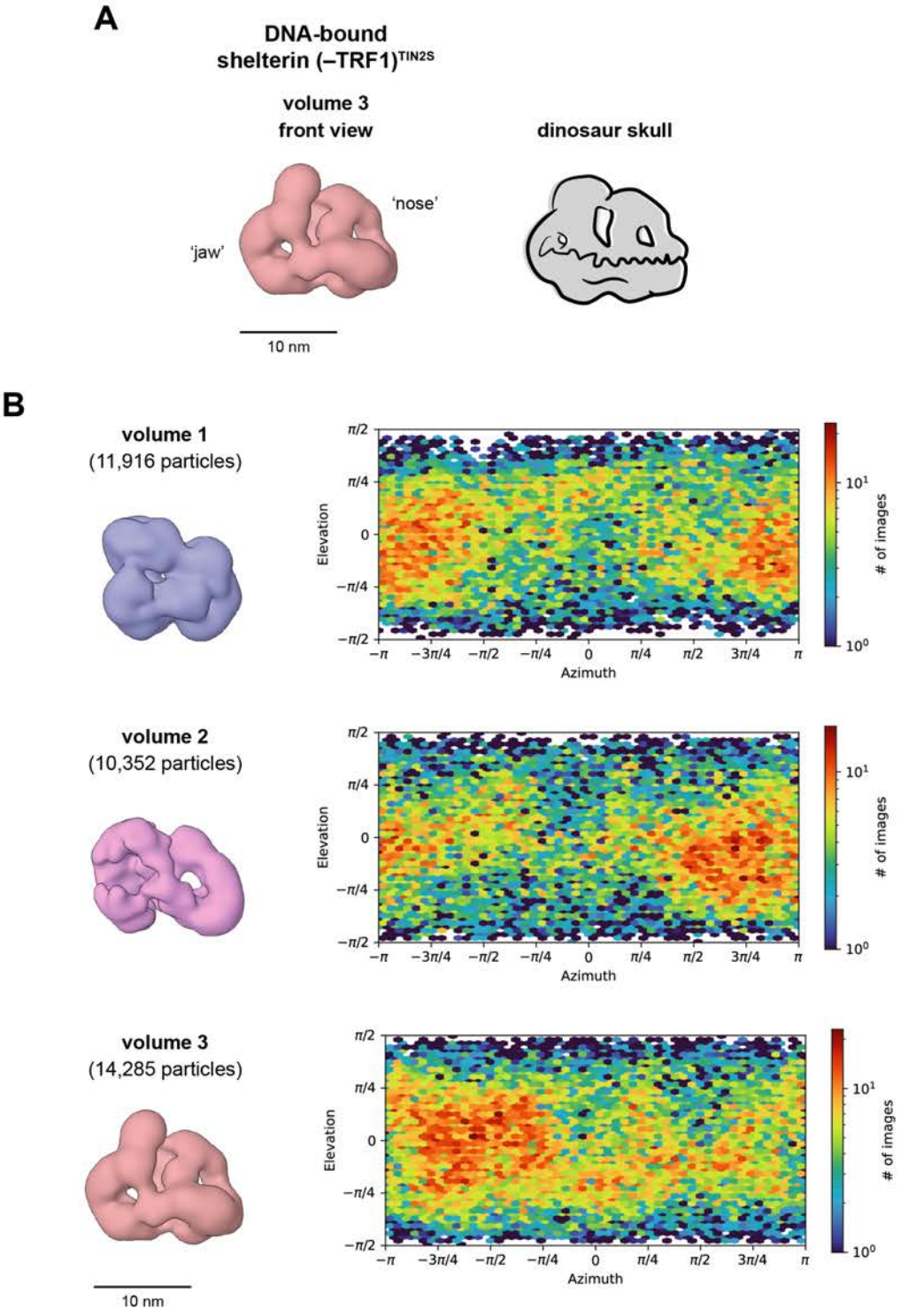

Figure 4.6. Distribution of particle orientations for 3D *ab initio* reconstruction of the DNA-bound (–TRF1) complex.

(A) 3D *ab initio* reconstruction of the DNA-bound shelterin (–TRF1)^{TIN2S} complex resembled the shape of a dinosaur skull, with a reproducible 'jaw' and 'nose' region. **(B)** Three 3D *ab initio* reconstructions with the number of particles contributing to each volume. The corresponding viewing direction distribution histograms (right) show the number of images were fitted to each viewing direction at each azimuth and elevation bin in radians.

Taking the ~14,000 particles contributing to the most populated 3D *ab initio* reconstruction (volume 3 in Figure 4.5C and Figure 4.7A), I repeated the 2D classification to remove poorly aligned particles (Figure 4.7B). The remaining 12,704 particles were used to generate one or two volumes during 3D *ab initio* reconstruction (Figure 4.7C). The starting density was reproducible when one volume was reconstructed from the similar set of particles (Figure 4.7C, left). When two volumes were generated from the selected particles, the particle images were designated equally, giving rise to densities with many varying but a few repeating features (Figure 4.7C, right). In particular, the 'jaw' region was observed in both negative-stain EM studies of the (–TRF1) subcomplex and the full shelterin complex (Figure 4.6A). A similar 3D *ab initio* reconstruction of the DNA-bound full shelterin^{TIN2S} complex was obtained by my colleague Dr Oviya Inian (Figure 4.8, grey). The EM map for the full shelterin complex was slightly larger than that for the (–TRF1) subcomplex. However, the extent of similarity between the two maps suggests that in the absence of TRF1, organisation of the remaining shelterin subunits was not significantly impacted under the conditions tested.

I was unable to identify any missing density to locate the 100 kDa TRF1 homodimer in the shelterin complex at this resolution (Figure 4.4D). This is possibly due to isolation of particles representing the 600 kDa species during data processing. The two species of (–TRF1) subcomplexes could not be separated despite efforts in further sample optimisation that I have previously described in Chapter 3 (see section 3.2.2). However, after crosslinking, it was possible to enrich the higher-MW (–TRF1) complex (Figure 4.4C and Figure 4.9B – E). After crosslinking the DNA-bound (–TRF1)^{TIN2S} complex, mass photometry analysis detected two predominant (–TRF1) complexes at 407 and 582 kDa (Figure 4.9A). After further purification by SEC to remove the excess DNA (Figure 4.9B), the elution fractions were analysed by SDS-PAGE, EMSA, and mass photometry (Figure 4.9C – E). Combining EMSA and mass photometry analysis, I identified an earlier SEC fraction that was enriched in the higher-MW species (11.3 mL, Figure 4.9C and D). This was a more homogenous sample and may be an interesting target for future EM analysis to study the architecture of the higher-order (–TRF1) species.

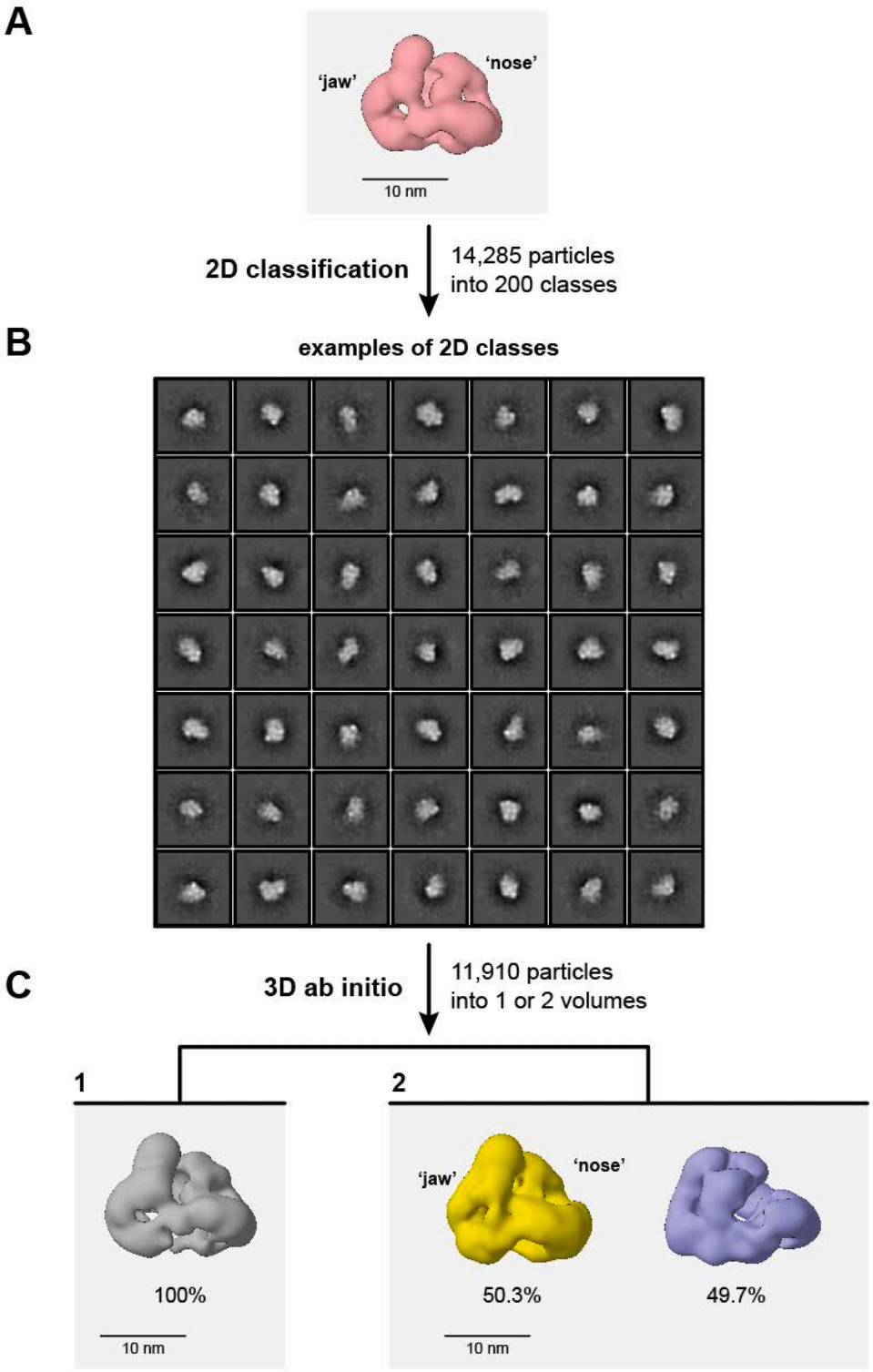

Figure 4.7. Further 2D classification and 3D *ab initio* reconstruction to explore conformational heterogeneity.

(A) 3D *ab initio* reconstructions of DNA-bound (–TRF1)^{TIN2S} shelterin complex. Particle images contributing to this reconstruction were selected and subjected to further 2D classification. **(B)** Examples of 2D classes. **(C)** One or two reference-free 3D *ab initio* reconstructions were generated using 12,704 particles from 121 2D classes.

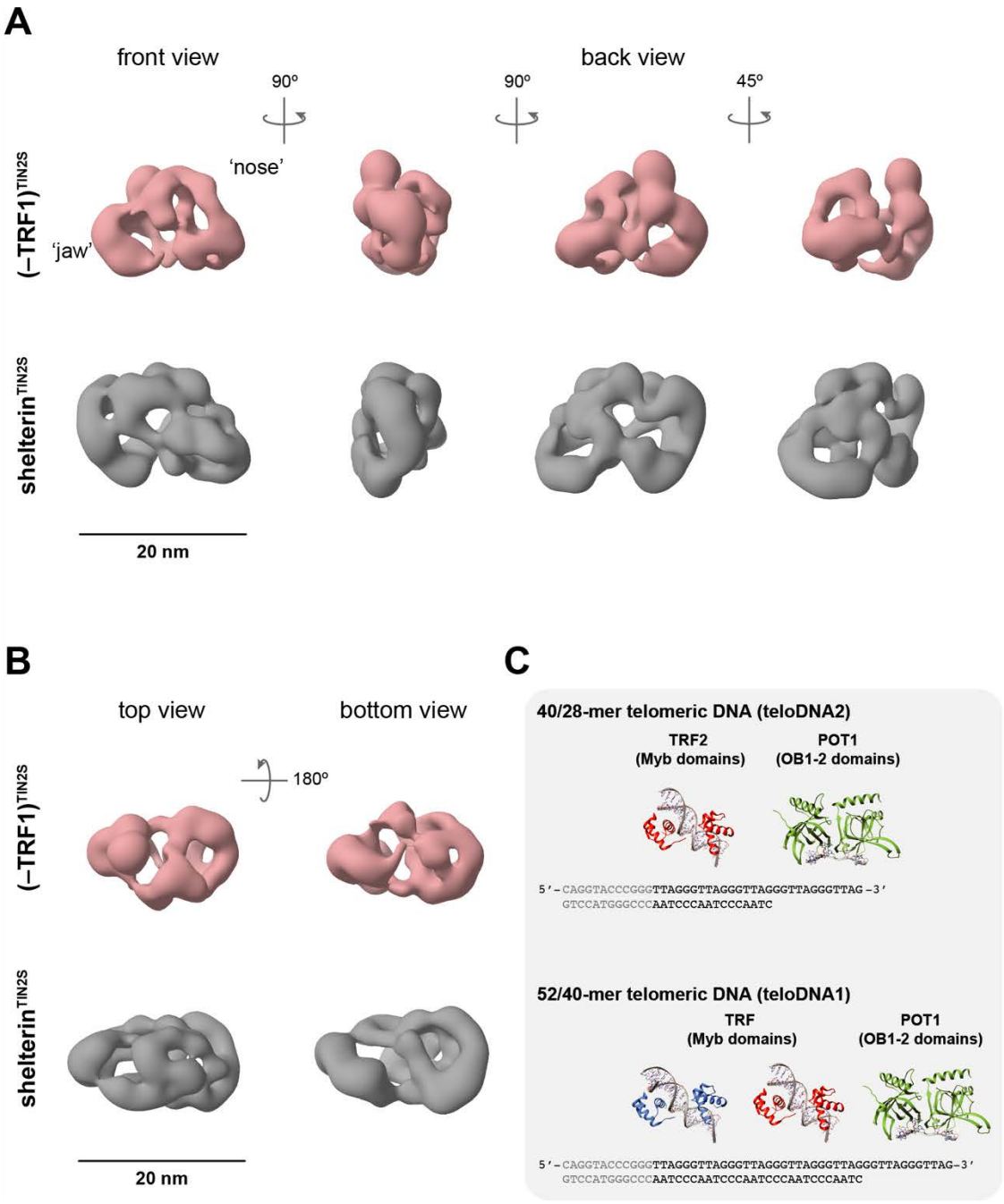

Figure 4.8. Comparisons of 3D *ab initio* reconstructions of the DNA-bound full and (–TRF1) shelterin^{TIN2S} complexes.

(A) Front and back views, and **(B)** top and bottom views of 3D *ab initio* reconstructions of the DNA-bound (–TRF1)^{TIN2S} subcomplex (pink), and the DNA-bound full shelterin^{TIN2S} complex (grey) from negative-stain EM analyses. Both samples were prepared using the same protocol (see section 6.6.1.7). **(C)** Schematics of the telomeric DNA used for reconstitution of DNA-bound shelterin complexes. The model telomeric DNA sequences, teloDNA1 and teloDNA2 were used to reconstitute DNA-bound full and (–TRF1)^{TIN2S} shelterin complexes, respectively.

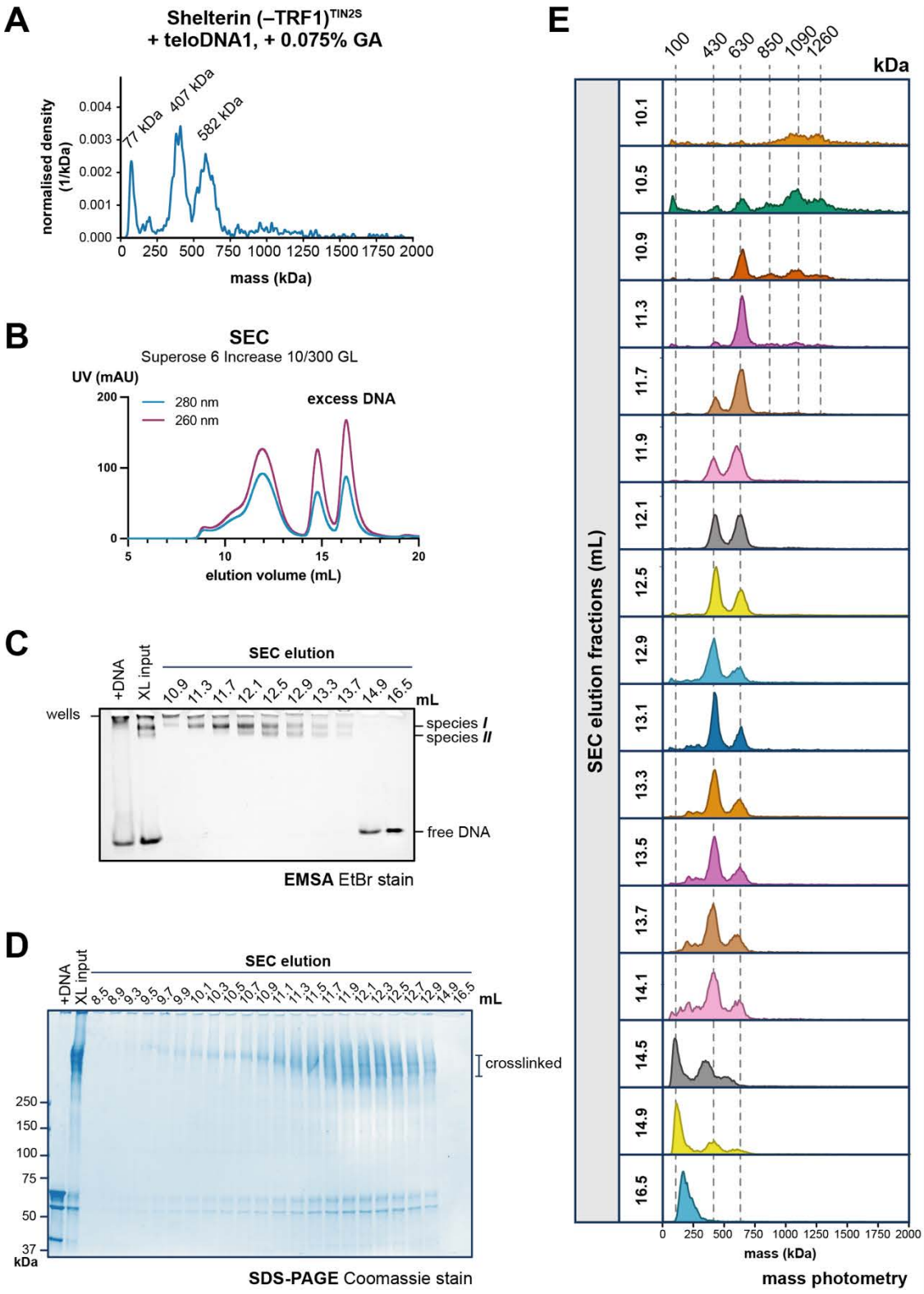

Figure 4.9. Purification of the higher-order DNA-bound (–TRF1) complex for EM analysis.

(A) Mass photometry analysis of the DNA-bound (–TRF1)^{TIN2S} after crosslinking with 0.075% of glutaraldehyde and quenched with 100 mM Tris-HCl pH 8. **(B)** SEC elution profile of crosslinked DNA-bound (–TRF1)^{TIN2S}. The elution fractions were analysed by **(C)** EMSA, **(D)** SDS-PAGE, and **I** mass photometry to identify homogeneous fractions for negative-stain EM analysis.

4.2 Probing protein-protein interaction interfaces within shelterin using XL-MS

4.2.1 Overview

In addition to the EM studies, XL-MS experiments were performed to explore protein-protein interactions between shelterin subunits in the full complex, and how DNA-binding changes the conformation of the shelterin complex. I chose to focus on the minimal shelterin complex $(TRF1/TRF2/RAP1)_2/(dStrepII\text{-}TIN2L/TPP1/POT1)_1$ to complement cryo-EM studies of the full shelterin complex performed by Dr Oviya Inian.

DSSO was used to perform the crosslinking reaction as it can be cleaved in the gas phase during tandem MS (MS^2), which enables MS^3 methods for accurate identification of the original crosslinked peptides. DSSO is a homo-bifunctional crosslinker that targets primary amines using NHS esters as the reactive moiety. The reactive moieties are connected by a spacer arm of 10.3 Å. Together, these properties enable identification of solvent-exposed lysine residues and provide a spatial restraint, with a theoretical upper boundary of ~26Å for the distance between the α carbon atoms (Cα) of crosslinked lysine residues (Kao, et al., 2010). The XL-MS experiments were conducted in collaboration with Dr Theo Roumeliotis from the Proteomics Facility at the ICR. See the Methods & Materials section for the detailed protocols and a schematic of the XL-MS workflow (Figure 6.8).

Crosslinks were filtered using a false discovery rate (FDR) of less than 1% and a XlinkX Score above 100 were visualised. I use the term ‘inter-link’ for inter-molecular crosslinks between two different shelterin subunits, and ‘self-link’ to refer to inter- and intra-molecular crosslinks between the same shelterin subunit. For instance, some of the self-links are likely inter-molecular crosslinks between two proximal TRF1, TRF2, or RAP1 subunits, as there are two copies of each within one minimal shelterin complex. The only case where it was possible to distinguish between intra- and inter-molecular self-links was when the same residue in one subunit was crosslinked by DSSO.

4.2.2 XL-MS analysis of the apo shelterin^{TIN2L} complex

To capture subunits in close spatial proximity in the full shelterin^{TIN2L} complex, I first optimised the protein:crosslinker ratio (Figure 4.10A). The shelterin^{TIN2L} complex was purified using the two-step purification as previously described (see section 2.1.2). SEC-purified apo shelterin^{TIN2L} was incubated with DMSO (control) or crosslinked using various DSSO concentrations at 20˚C for 1 hour and quenched with Tris-HCl buffer. Increasing DSSO concentration induced more extensive crosslinking of subunits in shelterin^{TIN2L} (Figure 4.10B and C). At 400 – 500 molar excess of DSSO, all shelterin subunits were crosslinked, judged by the disappearance of shelterin subunits at the lower-MW region and appearance of five bands above 250 kDa in the SDS-PAGE gel (Figure 4.10B and C). The most abundant species migrated near the 500 kDa marker (Figure 4.10B). Therefore, a 1:500 molar ratio of shelterin^{TIN2L}:DSSO was used to prepare samples for XL-MS experiments.

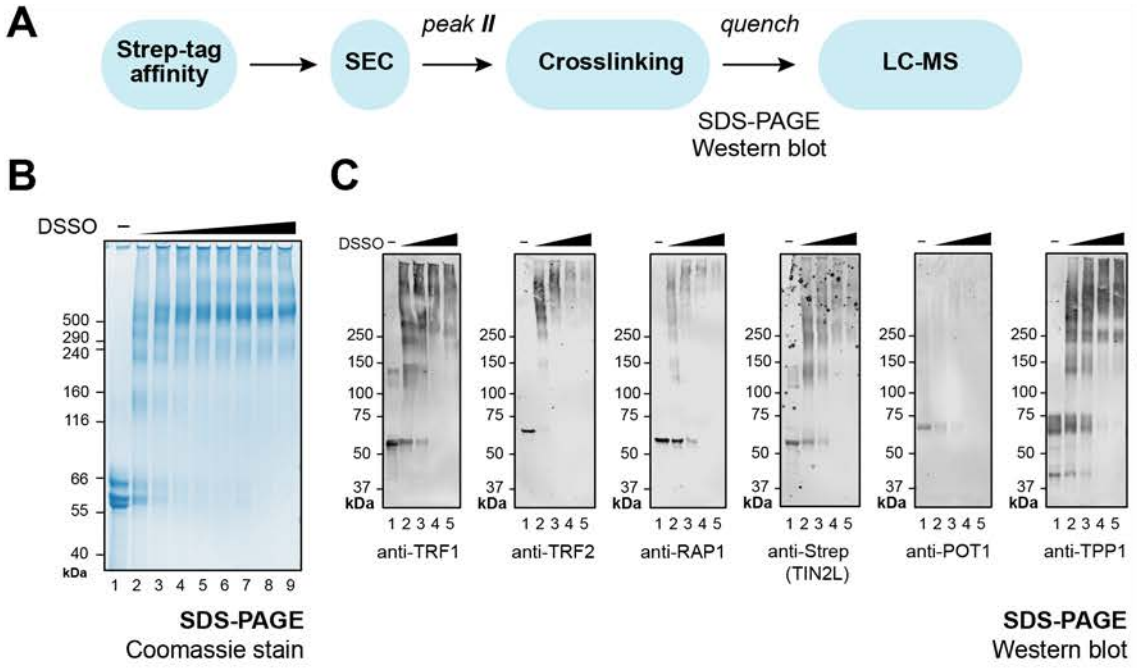

Figure 4.10. Optimising the shelterin:DSSO ratio for XL-MS analysis.

(A) Schematic of the sample preparation protocol for XL-MS. SEC-purified apo shelterin^{TIN2L} complex (1.5 µM) was incubated with 0, 100, 200, 300, 400, 500, 600, 800, or 1000-times molar excess of DSSO at 20 ºC for 1 hour. The crosslinked samples were quenched and analysed by **(B)** SDS-PAGE and **(C)** Western blot analysis (0, 100, 200, 500, 1000 molar excess of DSSO).

In a preliminary qualitative XL-MS experiment, the shelterin^{TIN2L} was crosslinked with 500-fold molar excess of DSSO (Figure 4.11A). The sequence coverage was above 57% for all subunits in the LC-MS analysis (Table 4.1). A total of 171 crosslinked peptide pairs were detected. Of these, 49% were inter-links and 51% were self-links. There were two surprising observations. Firstly, there were only two self-links and one inter-link detected from TRF2 (Figure 4.11B). One self-link was likely intra-molecular and found within the TRFH domain (Figure 4.11B), with a Cα–Cα distance of 20.5 Å between the lysine pair. The other self-link pair was in the linker region and may be either intra- or inter-molecular (Figure 4.11B). This lack of self-links was unexpected as TRF2 forms stable homodimers, similar to TRF1 from which many self-links were detected (Figure 4.11B). Both TRF1 and TRF2 contain similar percentage of lysines (8.2 vs. 8.3 %, respectively) distributed across the length of the amino acid sequence. Similar to the TRFH and Myb domains in TRF1, crystal structures of the TRF2 TRFH and Myb domains indicated that potentially there are many surface-exposed lysines (Figure 4.12A – D) (Benarroch-Popivker, et al., 2016). Therefore, the lack of crosslinks from TRF2 is unlikely to be due to the bias of lysine positions in the protein. The second striking observation was that no inter-links were identified between TRF2 and its binding partner RAP1 (Figure 4.11B). Although there are surface-exposed lysines in the interaction interface between TRF2$_{linker}$ and RAP1$_{RCT}$, some appear to be engaged in intra- and inter-molecular H-bonds, which reduces their likelihood of being crosslinked to other adjacent lysines by DSSO (Figure 4.12E) (Chen, et al., 2011).

To check whether the lysines in TRF2 and RAP1 are solvent-accessible and within the range of the DSSO spacer arm, the apo TRF2/RAP1 subcomplex was reconstituted as before (see section 3.2.2.2) and subjected to DSSO crosslinking under the same conditions as the full shelterin complex. This generated multiple inter- and self-links, indicating that in the TRF2/RAP1 subcomplex, there are surface lysines that can be crosslinked by DSSO (Figure 4.12F). This suggests that in the full shelterin complex, the lysines in TRF2 and RAP1 are less accessible to DSSO, either due to these subunits being buried inside the complex or are engaged in additional interactions formed in the context of other shelterin subunits.

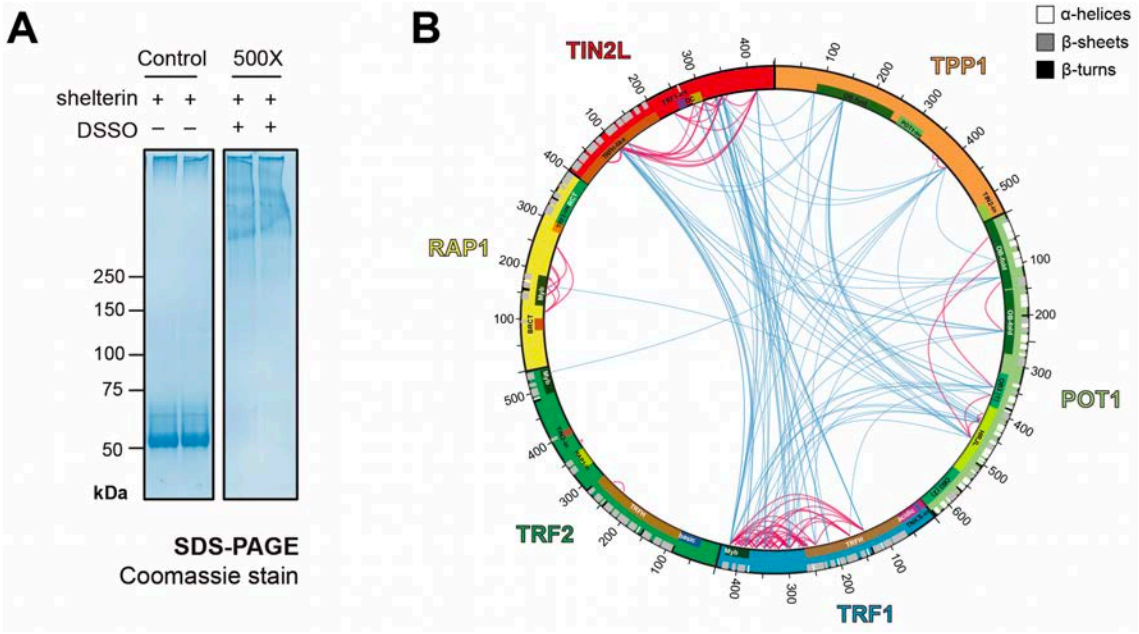

Figure 4.11. DSSO crosslinked peptides within the apo shelterin complex.

(A) SDS-PAGE analysis of native and crosslinked shelterin complex. SEC-purified apo shelterin^{TIN2L} at 1.64 μM was incubated with DMSO (control) or 500-fold molar excess of DSSO at room temperature shaking for 1 hour and quenched with Tris-HCl pH 8.0 (50 mM final concentration). **(B)** Circular plot displaying the inter-subunit crosslinks (blue) and self-links (red).

Shelterin subunit	sequence coverage (%)	# peptides	# crosslinks
TRF1	81	40	133
TIN2L	86	33	92
POT1	87	54	60
TPP1	78	32	29
RAP1	71	23	13
TRF2	57	31	3

Table 4.1. Sequence coverage of shelterin subunits in LC-MS.

The apo shelterin^{TIN2L} complex was crosslinked with 1:500 molar ratio of protein:DSSO and digested by trypsin to generate peptides for LC-MS. The number of unique peptides and total number of crosslinks found in each subunit are shown. Note that the sequence coverage may be underestimated due to the exclusion of peptides with charges <2+ (which are likely unmodified peptides and were filtered out to enrich crosslinked peptides for subsequent analysis).

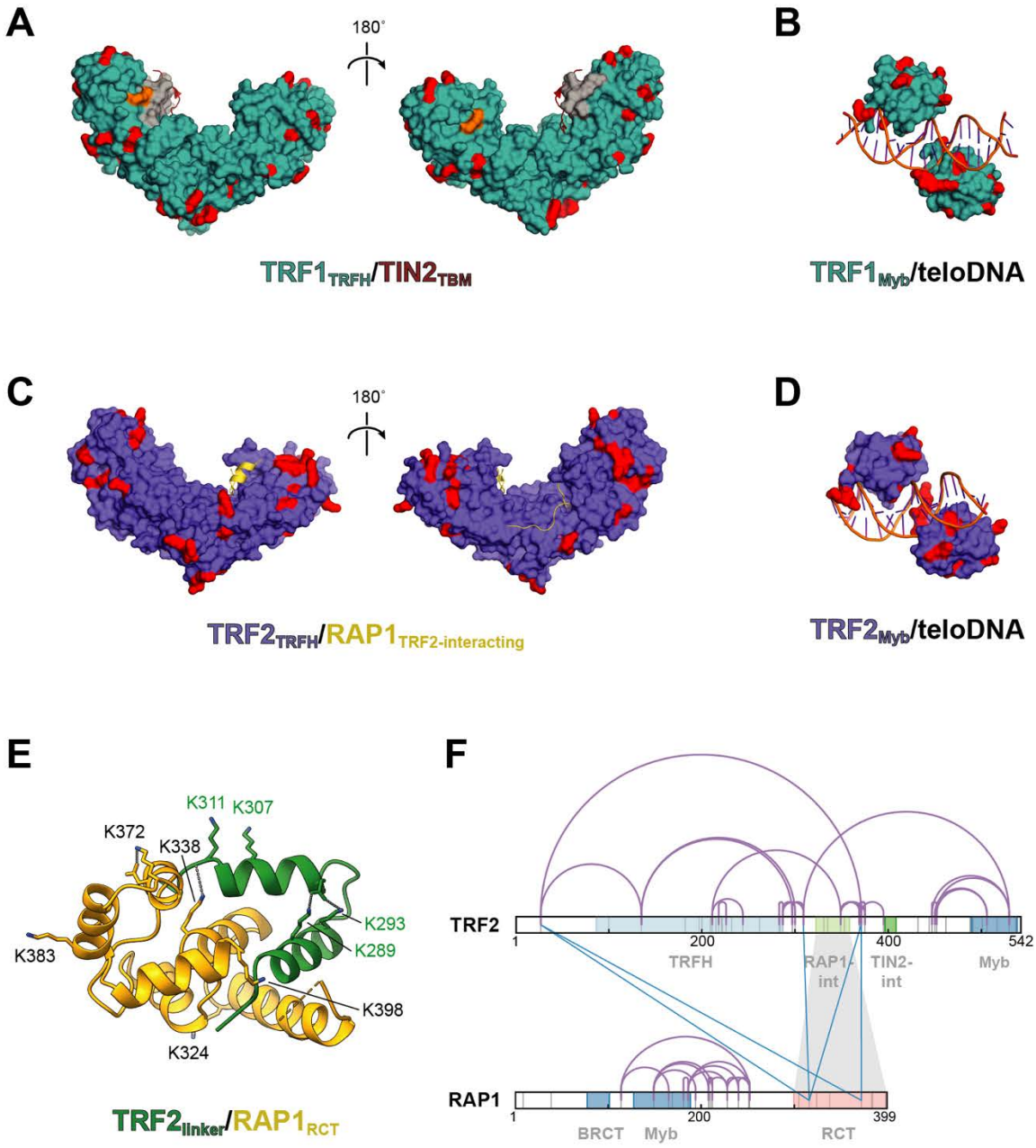

Figure 4.12. Potential surface-exposed lysines in TRF1 and TRF2 domains.

Surface lysines (in red) mapped onto crystal structures of **(A)** $TRF1_{TRFH}$ (with $TIN2_{TBM}$ peptide, red cartoon) (PDB: 3BQO); **(B)** TRF1 Myb domains in complex with telomeric dsDNA (PDB: 1W0T); **(C)** $TRF2_{TRFH}$, with the TRF2-interacting motif in RAP1 ($RAP1_{90-105}$, yellow cartoon) (PDB: 4RQI). **(D)** TRF2 Myb domains in complex with telomeric dsDNA (PDB: 1W0U). **I** Interacting interface between $TRF2_{linker}$ and $RAP1_{RCT}$ (PDB: 3K6G). The lysine side chains are represented as sticks and labelled (in black for RAP1 and green for TRF2), and H-bonds are represented as dashed grey lines. **(F)** Crosslinks detected in the TRF2/RAP1 subcomplex crosslinked with 500-times molar excess DSSO. The self-links and inter-links are mapped on the domain organisation of TRF2 and RAP1. The vertical grey lines mark the position of lysine residues. The shaded regions indicate known protein-protein interactions between TRF2 and RAP1.

Next, I tested whether the length of the crosslinking reaction affected the number and position of crosslinks. For this, SEC-purified apo shelterin^{TIN2L} was incubated with DMSO alone (control) or 500-fold molar excess of DSSO for 10 or 30 mins at 20 ˚C. The samples were efficiently digested by trypsin, judged by the appearance of protein bands below 15 kDa and a lack of species at higher MW (Appendix Figure 7.5), and TMT-labelled for quantitative LC-MS analysis. The positions of crosslinks were visualized using connectivity (Figure 4.13B) and network maps (Figure 4.14), and changes in the relative abundance of crosslink pairs in the 30-min compared to the 10-min DSSO crosslinking reaction were visualized using volcano plots (Figure 4.15).

In this experiment, 204 crosslink pairs were identified, with approximately the same number of self- and inter-links (49% and 51%, respectively). SDS-PAGE analysis indicated that nearly all shelterin subunits were crosslinked by DSSO, running as a prominent band near the 500 kDa protein marker (lanes 5–8, Figure 4.13A). There appeared to be more TRF1 and dStrepII-TIN2L subunits present in the input (lanes 1–4, Figure 4.13A), which explains the larger number of self- and inter-links from these subunits (Figure 4.13B). Once again, fewer inter-links were found to involve TRF2 and RAP1 compared to the other subunits (Figure 4.13B).

Many crosslinks were found in the flexible regions of shelterin subunits, in particular the linker in TRF1, the C-terminal portion of TIN2L, and the extreme N- and C-terminus of TPP1 (Figure 4.14A and B). Several crosslink hotspots were also found in domains, such as K430 and K433 in $POT1_{HJRL}$, K101 and K106 in $TIN2L_{TRFH\text{-}like}$, and K136 and K421 in $TRF1_{TRFH}$ and $TRF1_{Myb}$, respectively (Figure 4.14). These lysines were found in and near loops at the periphery of structured regions, both increasing their likelihood of being crosslinked by DSSO. The only shelterin subunit with structural information on the full-length protein is POT1, where a 7.9 Å EM density map enabled docking of crystal structures of OB1/2 and OB3/HJRL (Smith, et al., 2022) (Figure 4.16). Crosslink pairs satisfying the upper theoretical limit of ~26Å were nearby lysines or are found on the same face of POT1. For example, K121 in OB1 was crosslinked to K234 in OB2, which can be found on the same face of POT1 at a distance of 25.2 Å apart (Figure 4.16A).

Another example is K433, which was crosslinked to adjacent residues K422, K427, and K430 in POT1$_{HJRL,}$ and all within 10.3 Å of residue K433 (Figure 4.16B). Oppositely, crosslinks pairs separated by distances greater than ~26Å were found on opposite faces of POT1. For instance, K121 in OB1 was also crosslinked to K370 in part 1 of OB3 and K430 in the HJRL domain, both located at the opposite face of POT1 and were over 39 Å away from K121 (Figure 4.16A). These findings suggest that there may be other conformations of POT1 that were not captured by the cryoEM map (Smith, et al., 2022).

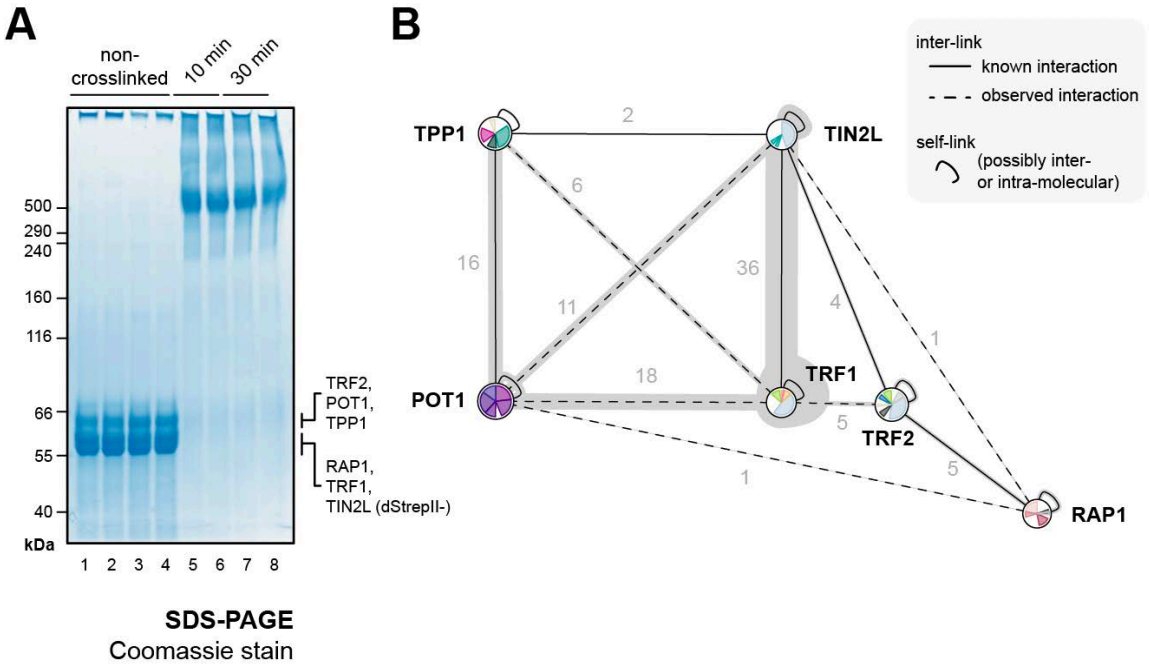

Figure 4.13. XL-MS analysis of apo shelterin after 10 or 30 mins of crosslinking.

(A) SDS-PAGE analysis of non-crosslinked samples of apo shelterin^{TIN2L} and samples crosslinked with DSSO for 10 or 30 mins. Two technical replicates of the control samples were analysed alongside two replicates of each crosslinking condition. **(B)** Connectivity map showing the crosslinks pairs found within (self-links) and between shelterin subunits (inter-links). The number of inter-links are labelled next to solid or dashed lines that indicate known and observed interactions, respectively. The number of crosslinks is indicated by the grey area surrounding a subunit or between two subunits.

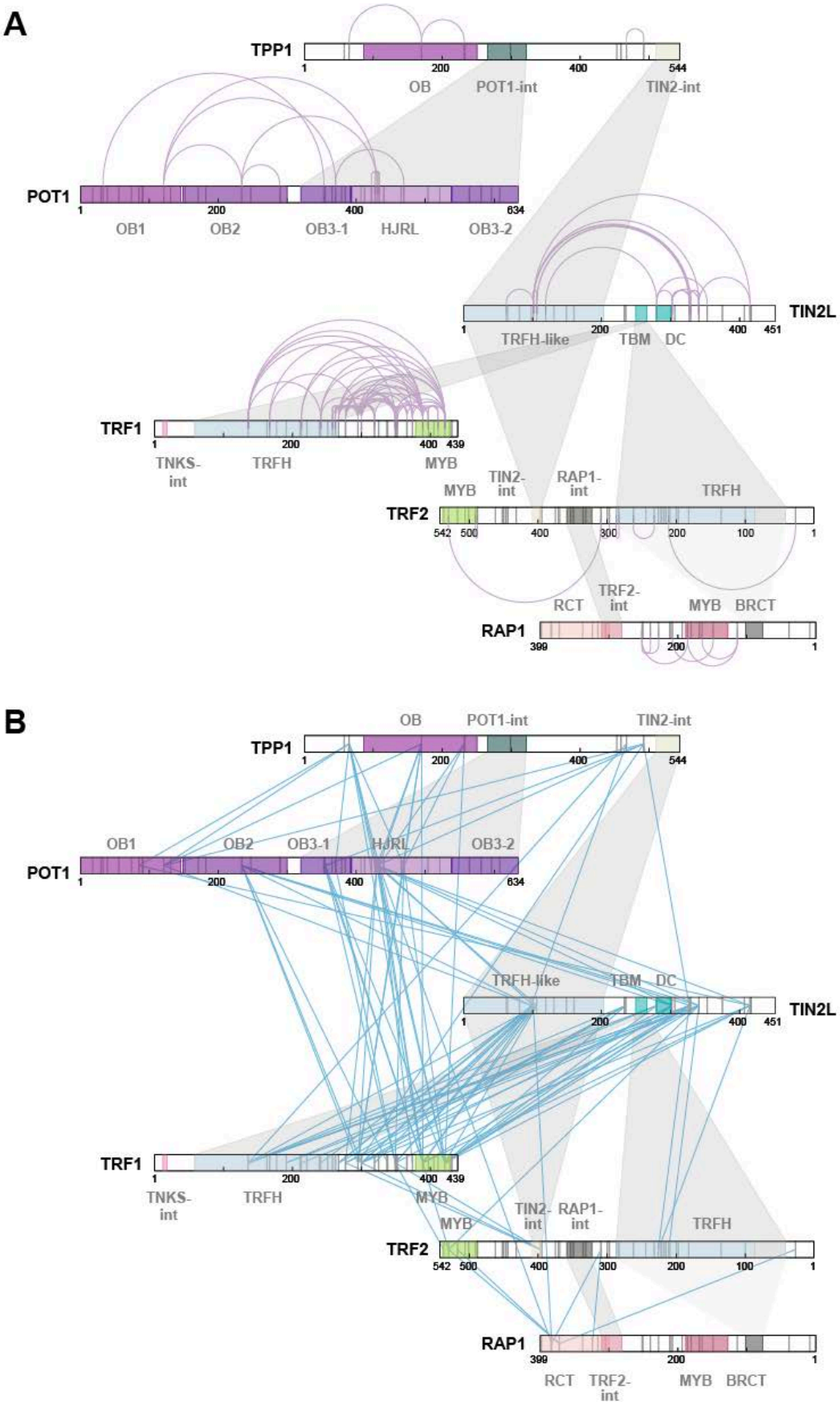

Figure 4.14. Crosslinks in the apo shelterin^{TIN2L} complex.

(legend on the next page)

(figure on the previous page)

Figure 4.14. Crosslinks in the apo shelterin^{TIN2L} complex.

Network maps of **(A)** the self-links and **(B)** inter-subunit crosslinks mapped on the domain organisation of each shelterin subunit. The vertical lines mark the position of lysine residues. The shaded regions indicate known protein-protein interactions between shelterin subunits.

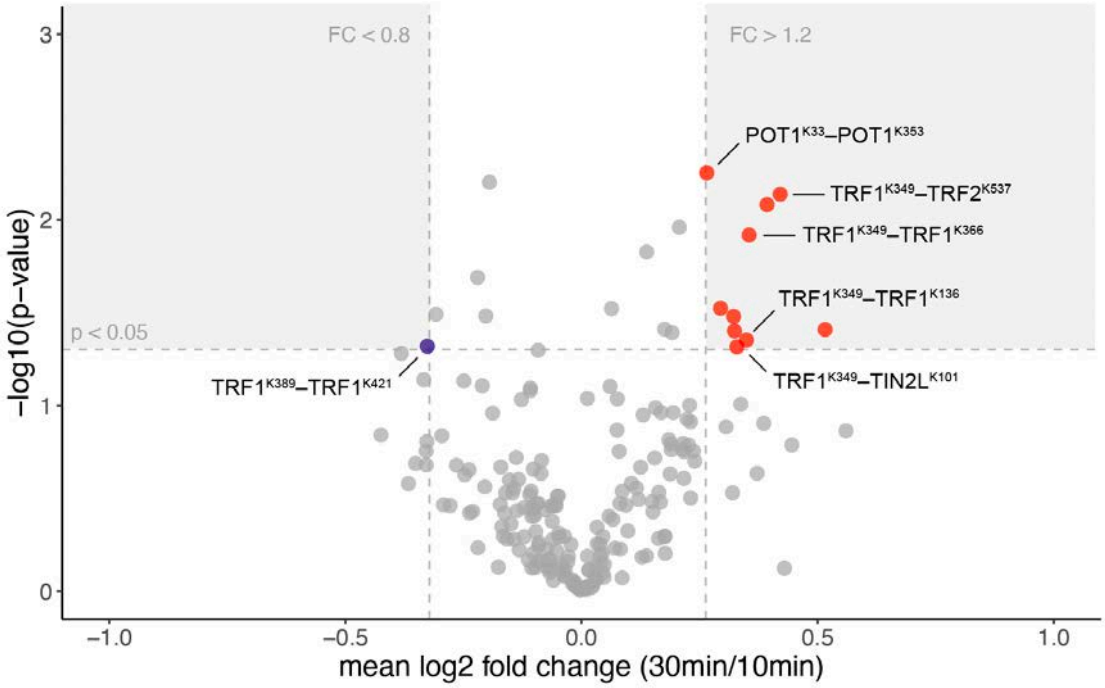

Figure 4.15. Relative abundance of crosslink pairs in the apo shelterin complex.

Volcano plot showing the mean log2 (30 min/10 min), which compares the abundance of crosslinks detected in the 30 mins vs. 10 mins crosslinking time (thresholds: fold-change of 0.8 and 1.2), and the -log10(p-value) (beyond threshold: $p < 0.05$).

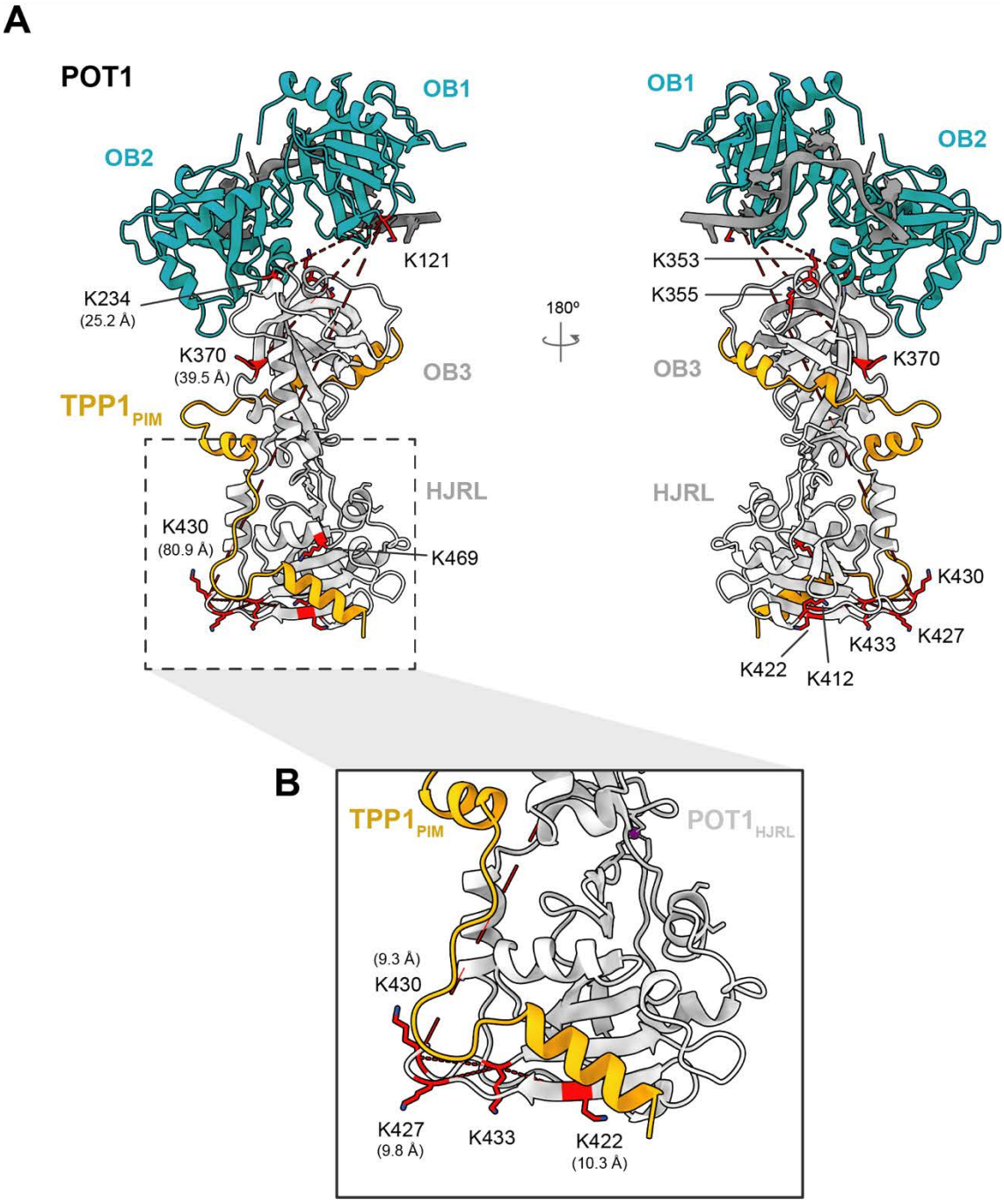

Figure 4.16. Self-links in POT1 identified by XL-MS analysis.

Self-links in POT1 involving the lysine residues **(A)** K121, and **(B)** K433. Crystal structures of POT1 OB1-2 (PDB: 3KJP) and OB3/HJRL with $TPP1_{PIM}$ (PDB: 5UN7) were docked into the 7.9Å EM density map of POT1 (EMD: 30596) (Smith, et al., 2022). Self-links identified from XL-MS was mapped onto the crystal structures to determine the distance between the crosslinked lysine pairs. Dotted lines are shown between the paired lysines. In brackets are the Cα–Cα distances between lysines crosslinked to either K121 or K433 in (A) and (B), respectively.

Approximately half of the crosslink pairs identified were inter-subunit crosslinks. Inter-links were found between all shelterin subunits known to participate in direct protein-protein interactions (solid lines, Figure 4.13B). Many crosslinks were observed adjacent to the established interaction interfaces. For example, 5 out of 16 crosslink pairs from the TPP1/POT1 heterodimer connected $POT1_{HJRL}$ to $TPP1_{OB}$, which is immediately upstream of the extended coil in TPP1 known to make extensive contacts with $POT1_{HJRL}$ and $POT1_{OB3}$ (Figure 4.13B and Figure 4.14B). In addition, many crosslinks suggest that several shelterin subunits exist in close spatial proximity in the apo $shelterin^{TIN2L}$ complex (dashed lines, Figure 4.13B). For instance, 11 crosslinks were found between TIN2L and POT1 (Figure 4.13B), which likely formed when these subunits were brought close together by their interactions with TPP1. Similarly, crosslink pairs from TPP1 and POT1 to TRF1 may have been established as TIN2L bridged these proteins together.

Finally, TMT-labelling of the crosslinked peptides enabled quantitative comparisons of crosslinks detected at the two different time points. All crosslink pairs in the 30-min crosslinking reaction were within 0.7 to 1.5-fold change from crosslinking for 10 mins. In the absence of telomeric DNA, lysines that are key for DNA-binding were also crosslinked by DSSO (Figure 4.14). For example, lysines at positions 421 and 530 in the Myb domains of TRF1 and TRF2 were found in several crosslink pairs (Figure 4.14). However, only one crosslink pair, K33 in $POT1_{OB1}$ and K353 in $POT1_{OB3\text{-}}$, was found at a higher abundance with longer DSSO crosslinking (1.2-fold change, Figure 4.15). Several crosslink pairs were significantly more or less abundant at 30 mins compared with the 10-min crosslinking reaction (< 0.8-fold or > 1.2-fold, $p < 0.05$, Figure 4.15). Within these, only one crosslink was less abundant at 30 mins. The remaining crosslinks were more abundant with the longer DSSO incubation time (Figure 4.15). Several of these crosslink pairs also involved lysines in the TRF1 linker region. In particular, residue K349 was crosslinked to K101 and K329 in the TRFH-like domain and the C-terminus of TIN2L, respectively, to $POT1_{HJRL}$ at K469, and finally to itself at multiple lysine positions (Figure 4.14 and Figure 4.15). This indicated that increasing the crosslinking duration raises the likelihood of capturing disordered regions of shelterin subunits in proximity.

4.2.3 Comparing crosslinks from apo and DNA-bound shelterin complexes

In a cellular context, shelterin complexes are recruited to telomeric DNA where they perform essential functions in telomere protection and maintenance. To explore the conformational changes induced upon binding of the full shelterin complex to telomeric DNA, I subjected both the apo and DNA-bound complex to quantitative XL-MS analysis.

First, I confirmed whether the conditions optimised thus far were suitable for crosslinking the shelterin-DNA complex. I reconstituted the shelterin^{TIN2L} complex bound to the minimal telomeric DNA, teloDNA1, and titrated in DSSO ranging from 0 to 1000-fold molar excess. In the presence of DNA, the majority of shelterin subunits became crosslinked at 500-fold molar excess of DSSO (Figure 4.17A, lane 3). Western blotting confirmed that all six subunits were present in this higher MW species (Figure 4.17B). Mass photometry analysis of the native complex prior to crosslinking identified a species at 515 kDa, which is approximately the 490 kDa shelterin^{TIN2L} species plus the 29 kDa teloDNA1 (Figure 4.16C, lane 0X). The presence of three predominant subcomplexes was mostly due to dissociation of the complex during sample dilution. Crosslinking by DSSO stabilised the full shelterin complex at ~600 kDa and significantly reduced the abundance of the two smaller species at ~100 and 200 kDa (Figure 4.17C, lanes 250X – 1000X). In addition, the crosslinking process did not produce species larger than the minimal shelterin complex or large aggregates (Figure 4.17C). As the DSSO concentration increased from 250- to 1000-fold molar in excess of the protein complex, the MW of the crosslinked shelterin^{TIN2L} complex increased from 552 kDa to 615 kDa, indicating the complex was getting increasingly modified by DSSO and the quenching agent Tris (Figure 4.17C). EMSA analysis showed that shelterin complex crosslinked by DSSO retained DNA at all DSSO concentrations (Figure 4.17D). However, beyond 500-fold molar excess of DSSO, the abundance of the shelterin-DNA complex decreased substantially, indicating that the DNA-binding capability of shelterin was becoming impaired at these crosslinker concentrations (Figure 4.16B, lanes 5–

6). Therefore, 1:500 molar ratio of shelterin:DSSO was suitable for crosslinking both the apo and DNA-bound shelterin^{TIN2L} complexes.

In the previous XL-MS experiment, there was an excess of TRF1 and dStrepII-TIN2L subunits, which was likely carried over from an adjacent species during the final SEC purification (see section 1.1.2). To improve sample homogeneity, I introduced an additional heparin chromatography step after the Strep-tag affinity purification (Figure 4.18A). A species of shelterin containing dStrepII-TIN2L, TPP1 and POT1 did not bind to the heparin column and was removed in the flowthrough (Figure 4.18B and C). Although both peak ***A*** and ***B*** appeared to contain the same proteins on SDS-PAGE, the shelterin species in peak ***B*** also contained an additional band near the 55 kDa marker (Figure 4.18C). The higher 260/280 nm ratio seen from the UV chromatogram also indicated contaminating DNA co-eluting with proteins in the peak ***B*** compared with peak ***A*** (0.65 vs. 0.56, respectively; Figure 4.18B). Therefore, fractions in peak ***A*** from the heparin elution were pooled and concentrated for the final SEC polishing step (Figure 4.18D). Mass photometry analysis revealed that early SEC fractions contained the minimal 490 kDa apo shelterin^{TIN2L} complex (11.29 mL to 11.69 mL, Figure 4.18D and E). These fractions were used for reconstituting the DNA-bound shelterin complex and as comparisons during XL-MS analysis. Finally, the native (control) and crosslinked shelterin-DNA samples were analysed by EMSA, which confirmed the presence of telomeric DNA (band ***I***, Figure 4.18F). All samples were digested by trypsin and TMT-labelled for quantitative MS analysis.

From this XL-MS experiment, 179 crosslink pairs were identified (47% and 53% of inter-links and self-links, respectively). The positions of crosslinks were isualized using connectivity (Figure 4.20) and network maps (Figure 4.21). Significant changes in the relative abundance of crosslink pairs in the DNA-bound shelterin complex compared with the apo complex were isualized using volcano plots (Figure 4.22).

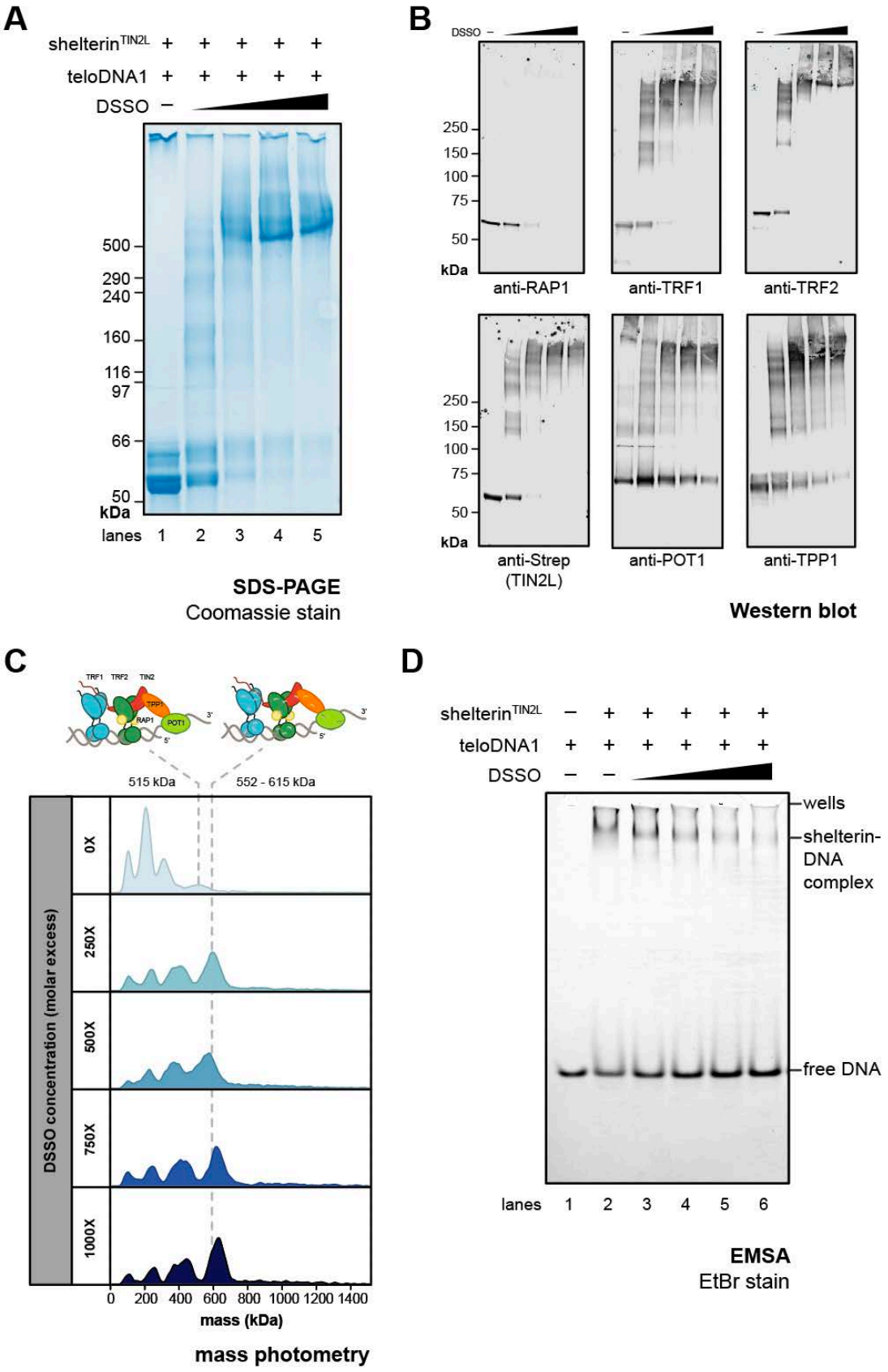

Figure 4.17. Titration of DSSO to optimise crosslinking of the DNA-bound shelterin complex.

(legend on the next page)

(figure on the previous page)

Figure 4.17. Titration of DSSO to optimise crosslinking of the DNA-bound shelterin complex.

Titration of the DSSO crosslinker in 0, 250, 500, 750, 1000-fold molar excess of the shelterin^{TIN2L} complex at 1.5 µM. **(A)** SDS-PAGE and **(B)** Western blot analysis of the native and crosslinked shelterin^{TIN2L} complexes. **(C)** Mass photometry analysis of the native (0X) and crosslinked DNA-bound shelterin^{TIN2L} complex treated with 250, 500, and 1000-fold molar excess of DSSO. The expected MW of the complex with telomeric DNA is 519 kDa. **(D)** EMSA analysis of the native and crosslinked DNA-bound shelterin^{TIN2L} complexes.

Although the majority of crosslinks was still detected in the TRF1 and dStrepII-TIN2L subunits, the additional heparin purification and quality control steps reduced the number of crosslinks from these two subunits (compare Figure 4.20 with Figure 4.13B). As seen previously, there were many crosslinks arising from flexible regions within shelterin subunits (Figure 4.21). Notably, there was a reduction in the abundance of crosslinks from the DNA-binding domains of shelterin subunits in the presence of telomeric DNA (Figure 4.22 and Appendix Figure 7.6). The abundance of the $POT1_{OB1}{}^{K33}$–$POT1_{OB3\text{-}1}{}^{K353}$ crosslink pair decreased to 4% of that identified in the apo complex (Figure 4.22). Furthermore, 16 crosslinks involving K421 from $TRF1_{Myb}$, a key residue contributing to specificity and affinity of DNA-binding, were significantly less abundant in the crosslinked DNA-bound shelterin^{TIN2L} complex ($p < 0.05$, Appendix Figure 7.6). The equivalent lysine in $TRF2_{Myb}$ (K530) was found in a self-link with K211 in $TRF2_{TRFH}$ and an inter-link with K372 in $RAP1_{RCT}$, with the former being 0.78-fold less abundant in the DNA-bound complex ($p < 0.05$). These findings indicated that with the addition of the minimal telomeric DNA, lysine residues in the DNA-binding surfaces were engaged in DNA binding and became less accessible to DSSO. This gave confidence that differences between the abundance of crosslinks in the apo and DNA-bound state were induced by conformational changes upon DNA-binding.

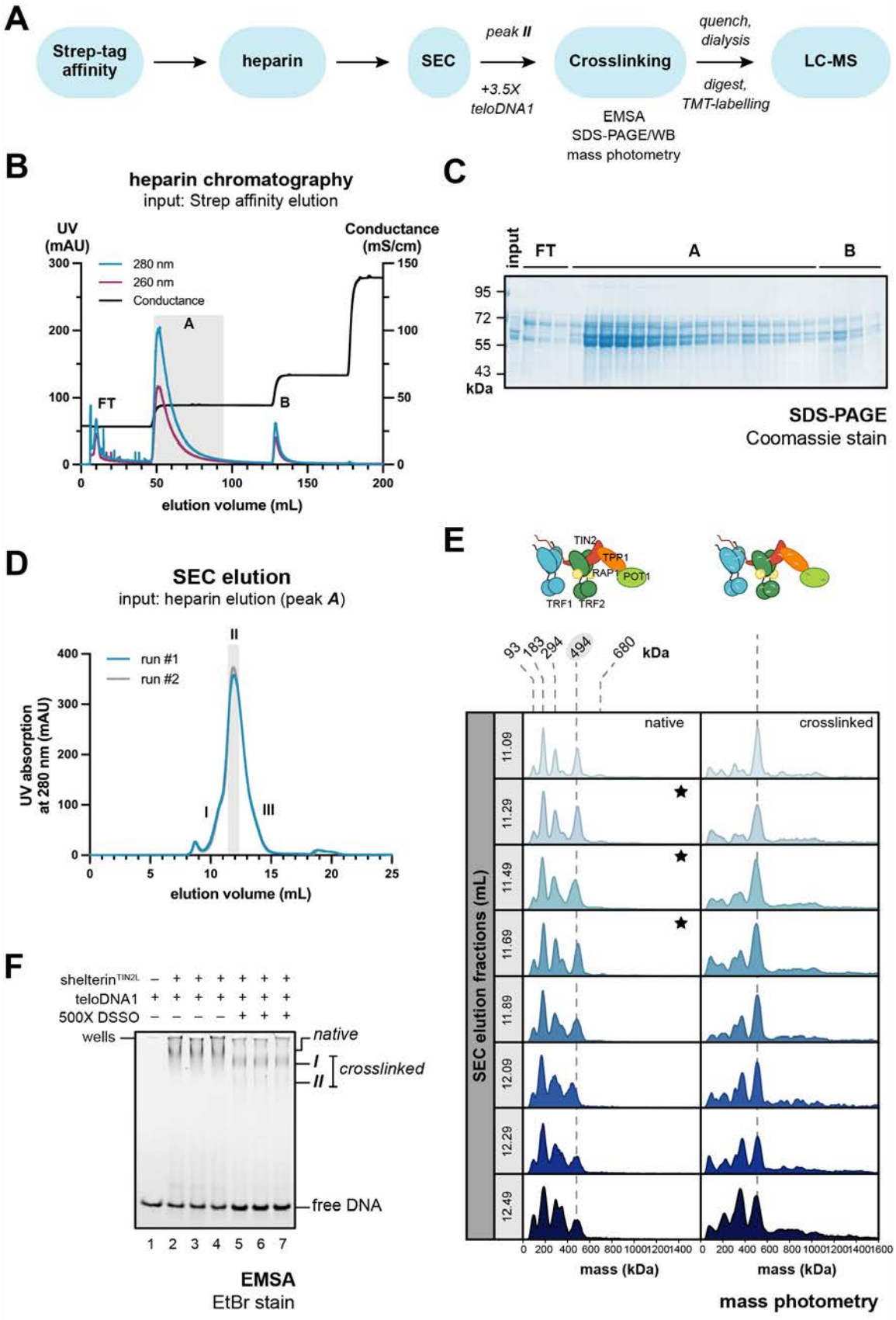

Figure 4.18. Reconstitution of apo and DNA-bound shelterin for XL-MS.

(legend on the next page)

(figure on the previous page)

Figure 4.18. Reconstitution of apo and DNA-bound shelterin for XL-MS.

(A) Schematic of the optimised sample preparation procedure. **(B)** Elution profile from heparin chromatography. Affinity-purified apo shelterin^{TIN2L} was dialysed to a low-salt buffer (300 mM NaCl), loaded onto a Heparin HP column followed by step elution to separate the full complex from subcomplexes. **(C)** SDS-PAGE analysis of heparin elution. **(D)** Final SEC purification. Fractions from the heparin elution peak ***I*** were pooled, concentrated, split into two and purified over 2 consecutive runs. **(E)** Mass photometry analysis of SEC elution fractions. Fractions across species ***II*** were analysed under native and crosslinked conditions to identify homogeneous samples for XL-MS analysis. Dotted lines indicate the minimal 490 kDa shelterin^{TIN2L} complex. The stars indicate the SEC fractions taken forward for DSSO crosslinking. **(F)** EMSA analysis of DNA-bound shelterin^{TIN2L} samples.

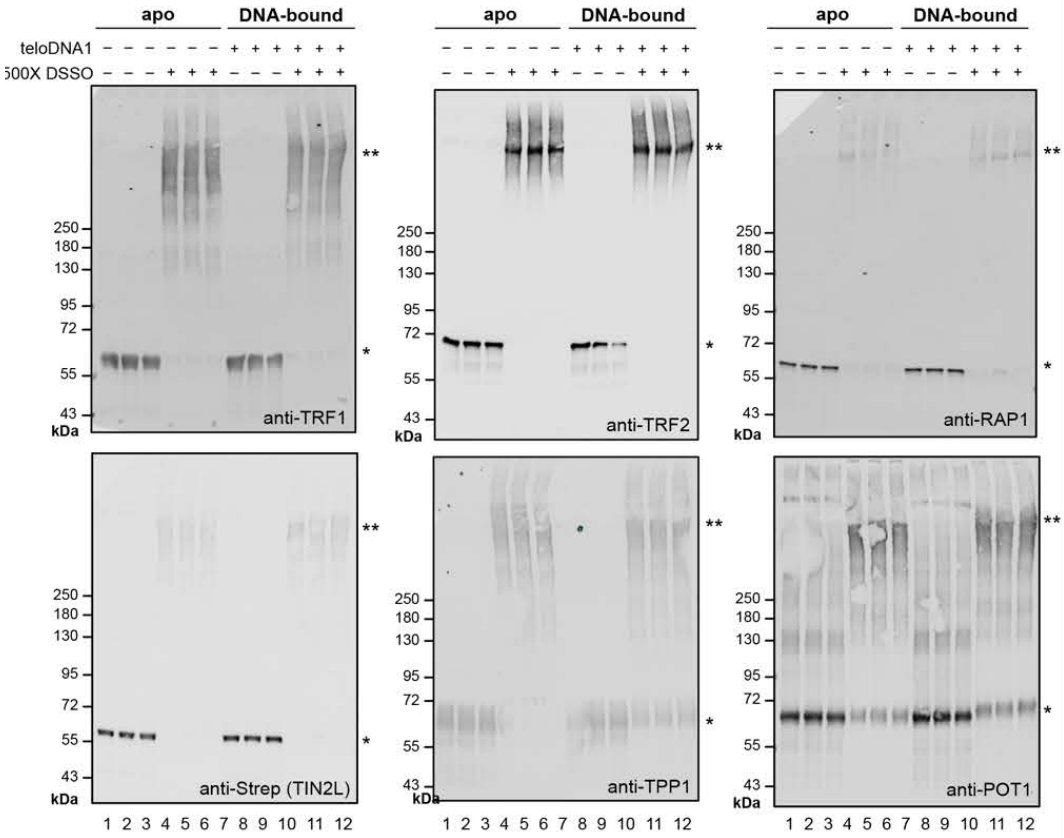

Figure 4.19. Western blotting analysis of apo and DNA-bound shelterin complexes for XL-MS.

SEC-purified shelterin^{TIN2L} complexes in the apo (lanes 1-6) and DNA-bound (7-12) states were incubated with DMSO or DSSO for 20 mins at room temperature. The quenched samples were resolved by SDS-PAGE and probed with antibodies against all six shelterin subunits. * and ** indicate the presence of native and crosslinked species.

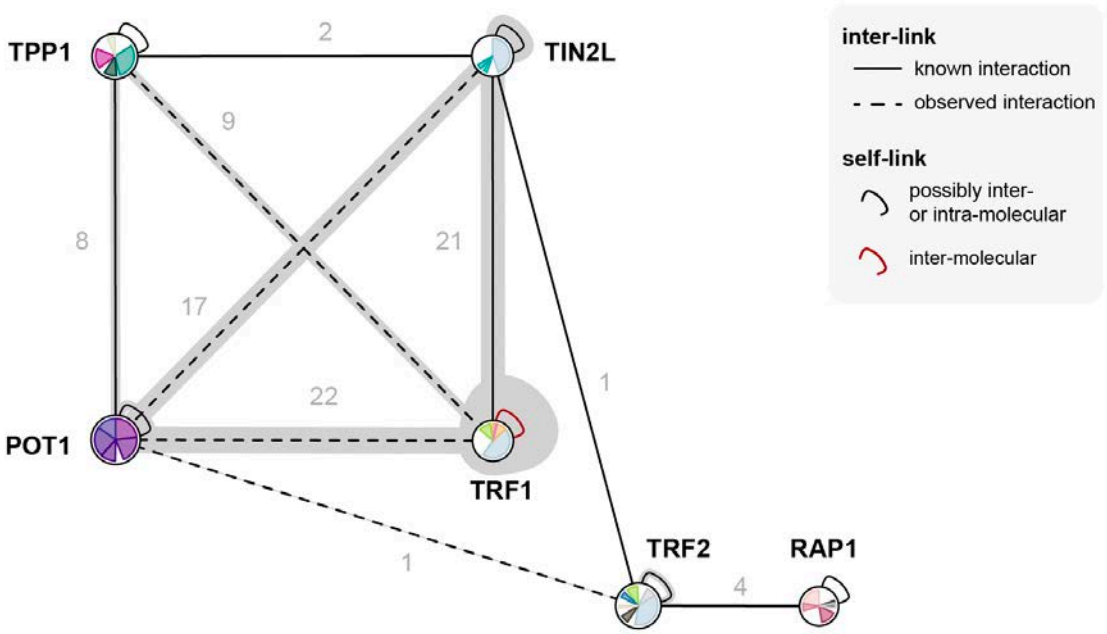

Figure 4.20. XL-MS analysis of the apo and DNA-bound shelterin complex.

Connectivity map showing the crosslinks pairs found within (self-links) and between shelterin subunits (inter-links). The number of interlinks are labelled adjacent to solid or dashed lines that indicate known and observed interactions, respectively. The number of protein crosslinks correlates with the size of the grey area surrounding the subunit or between two subunits.

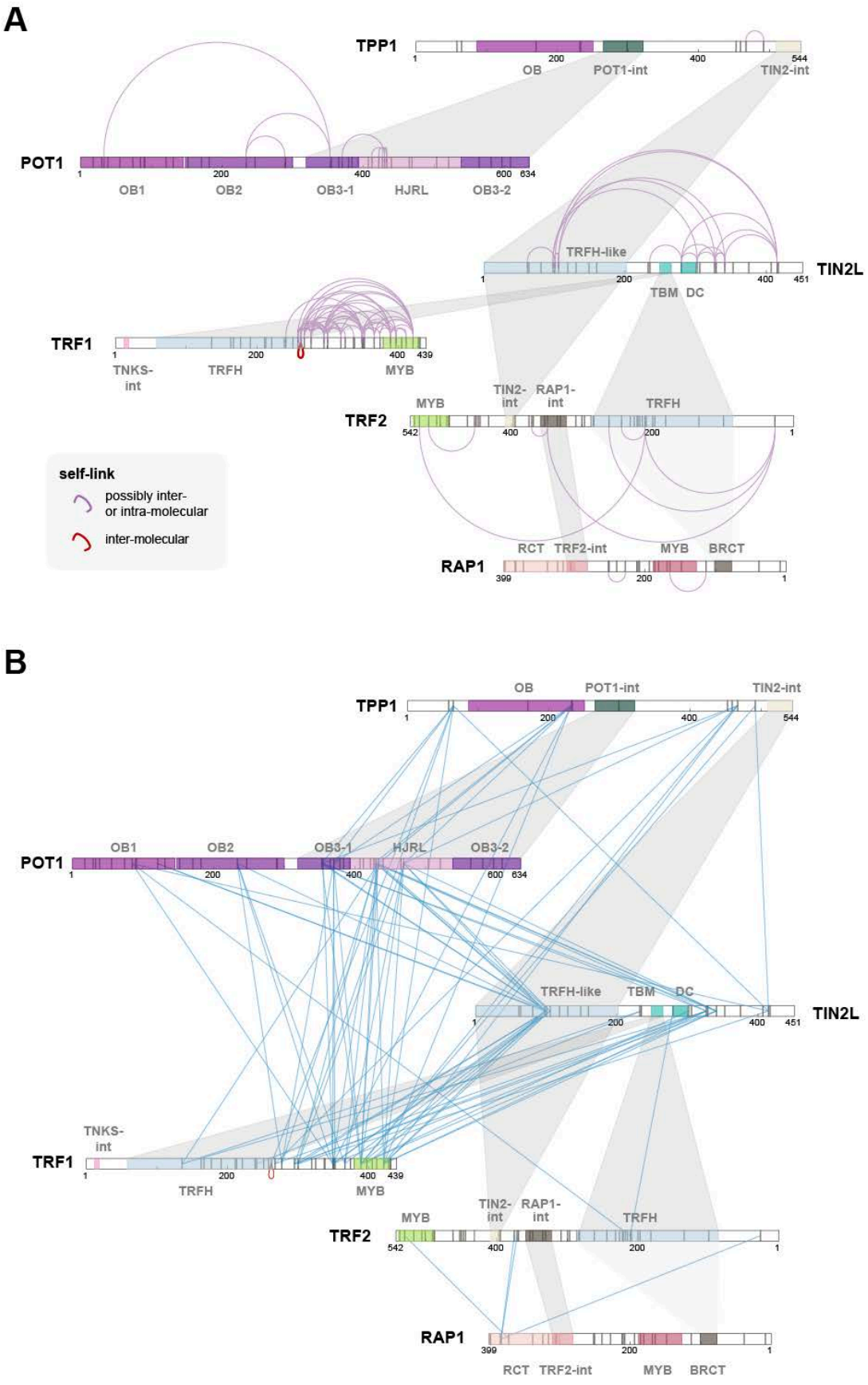

Figure 4.21. Crosslinks in the apo and DNA-bound shelterin complex.

(legend on the next page)

(figure on the previous page)

Figure 4.21. Crosslinks in the apo and DNA-bound shelterin complex.

(A) The self-links and **(B)** inter-subunit crosslinks mapped on the domain organisation of each shelterin subunit. The vertical lines mark the position of lysine residues. The shaded regions indicate known protein-protein interactions between shelterin subunits.

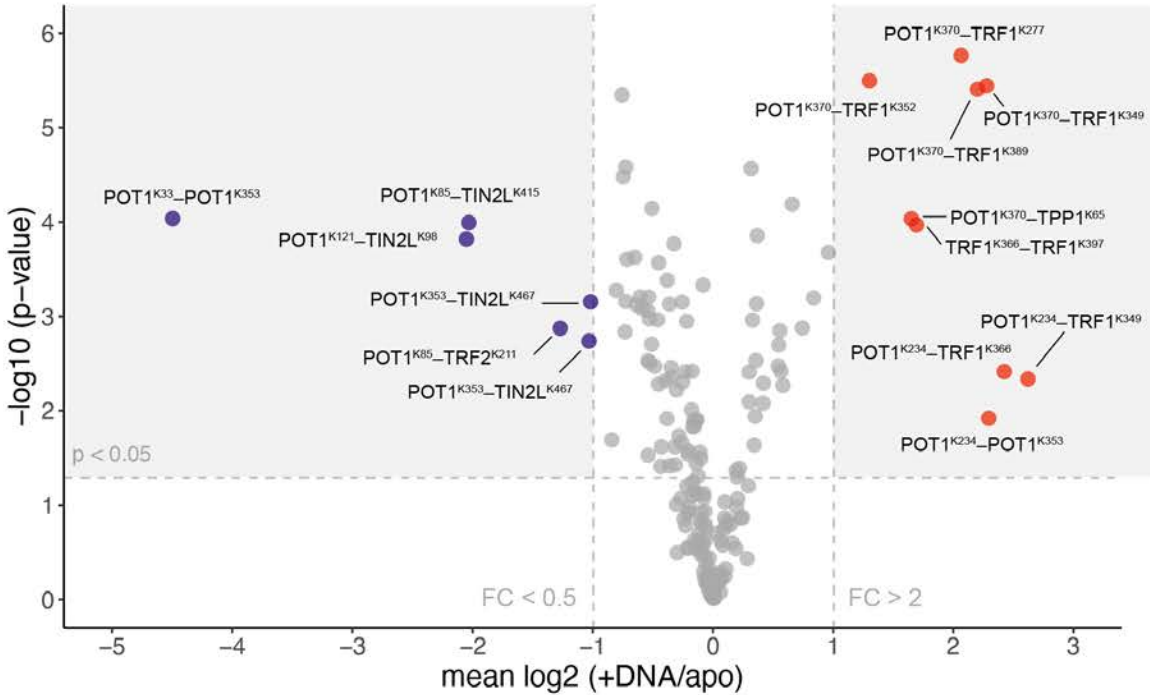

Figure 4.22. Changes in abundance of crosslink pairs in the DNA-bound vs. apo shelterin complex.

Volcano plot showing the mean log2 (+DNA/apo), which compares the relative abundance of crosslink pairs detected in the DNA-bound vs. the apo shelterin^{TIN2L} complex. Thresholds: vertical lines represent fold changes of < 0.5 and > 2 (mean log2 < –1 and > 1, respectively), and $p < 0.05$ beyond the horizontal line crossing the -log10(p-value) axis.

4.2.3.1 Crosslinks from the dsDNA-binding module of shelterin

Focusing on the TRF1/TIN2L/TRF2/RAP1 portion of the shelterin complex, inter-links between TRF1 and TIN2 were most common (24%, Figure 4.20). Many crosslinks involved the linker region in TRF1 and the unstructured C-terminal portion of TIN2L (Figure 4.23A). Crosslinks were also found between lysines in $TRF1_{TRFH}$ and near the $TIN2_{TBM}$, such as the $TRF1^{K136}$-$TIN2L^{K235}$ and $TRF1^{K136}$-$TIN2L^{K329}$ pairs (orange lines in Figure 4.23A). The TRFH-like domain in TIN2L contained a crosslink hotspot consisting of 3 lysine residues, K98, K101 and K106, in a loop connecting helices α4 and α5, and in helix α5 (Figure 4.23C). These lysine residues formed extensive crosslinks with the C-terminal half of TRF1 (Figure 4.23A). The only crosslink between TRF2 and TIN2L was found in the TRFH domain of TRF2 and a flexible region between the TBM and DC cluster in TIN2L (orange line in Figure 4.23A, Figure 4.23B). Interestingly, in the crystal structure of TRFH-like domain TIN2 bound to TRF2 and TPP1 peptides, K397 in TRF2 was found within 26 Å of the crosslink hotspot in TIN2 and was not engaged in H-bonding (Figure 4.23C). However, crosslinks near this region were not observed, which may be due to this region of the TRF2 linker being further apart from $TIN2_{TRFH\text{-}like}$ or inaccessible in solution. Four crosslinks were observed between the TRF2 and RAP1, two of which involved K383 in the $RAP1_{RCT}$ and a region in the TRF2 linker close to the known RAP1-interacting motif (Figure 4.23A and D).

Upon DNA-binding, significant changes in crosslink abundance occurred in TRF1 self-links, notably between the linker region and the DNA-binding Myb domain (Figure 4.22 and Appendix Figure 7.6). In the DNA-bound complex, the abundance of crosslinks between several lysines in linker (including K277, K295, K300, K349, and K366) and K421 in helix 3 of the 'helix-turn-helix' motif in $TRF1_{Myb}$ decreased significantly (Figure 4.23E). Conversely, the abundance of self-links between K349, K366 and other lysine residues in helices 1 and 2, and the turn preceding helix 3 in the Myb domain (K389, K397, and K411) increased significantly ($p < 0.05$, Figure 4.22 and Figure 4.23E). Together, this indicated that the TRF1 linker lay proximal to helix 3 in the Myb domain in the apo shelterin complex, which was displaced in the presence of DNA.

(figure on the next page)

Figure 4.23. Crosslinks in the dsDNA-binding module of the shelterin complex.

(A) Network plot mapping the DSSO crosslinks in the dsDNA module of the shelterin^{TIN2L} complex. Grey vertical lines mark the lysine distributions. Shaded areas map the known protein-protein interactions. The orange lines highlight crosslinks from TRF1/2 TRFH domains to regions near the TBM in TIN2. **(B)** Crystal structures of TRFH domains from TRF1 (dark blue, PDB: 3BQO) and TRF2 (grey, PDB: 3BU8) with the TIN2$_{TBM}$ (yellow and orange in complex with TRF1$_{TRFH}$ and TRF2 $_{TRFH}$, respectively). **(C)** Crystal structure of TRFH-like domain in TIN2, and the TIN2-interacting motif in TRF2$_{linker}$. (PDB: 5XYF). **(D)** Crystal structure of the C-terminal domain in RAP1 (RAP1$_{RCT}$) and the RAP1-interacting motif in the linker region of TRF2 (TRF2$_{linker}$) (PDB: 3K6G). **(E)** Crystal structure of the TRF1 Myb domain bound to telomeric dsDNA (PDB: 1W0T). Lysine residues are shown in stick representation, with crosslinked residues coloured in red. The yellow dashed lines indicate H-bonds and black dashed lines the Cα–Cα distance between TIN2^{K106} and TRF2^{K355}.

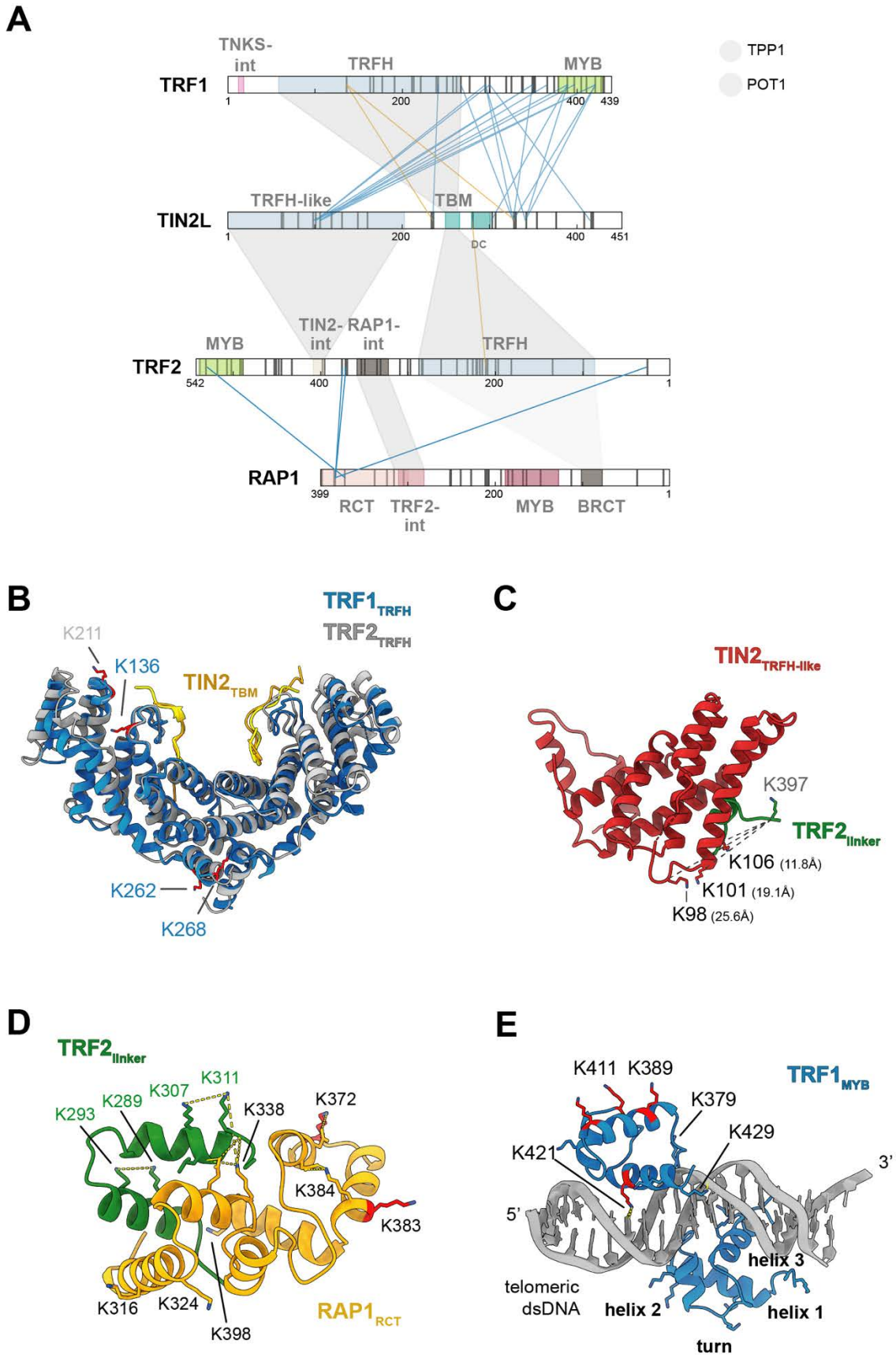

Figure 4.23. Crosslinks in the dsDNA-binding module of the shelterin complex.

(legend on the previous page)

4.2.3.2 Crosslinks from the ssDNA-binding module of shelterin

Interlinks were found between all three proteins in the ssDNA-binding module of shelterin (TIN2L/TPP1/POT1). All three subunits contained crosslink hotspots, namely $TPP1_{OB}$ (K232), several lysines in the C-terminal half of POT1 (K353/K355, K430/433, K469) and the TRFH-like domain in TIN2L (Figure 4.24). These lysines are located at the periphery of domains and in short loops (Figure 4.24C – E). As they were not part of interaction interfaces, they have a higher likelihood of being crosslinked to other proximal lysines by DSSO. For instance, K232 in $TPP1_{OB}$ is located in helix α3 that caps one end of the central β-barrel and was found in multiple crosslink pairs (Figure 4.24D). On the other hand, the adjacent K233 is involved in H-bonding with helix αA, which protected the ε-amine from the reactive NHS ester group in DSSO (Figure 4.24D).

Several crosslinks were also observed connecting $TIN2L_{TRFH\text{-}like}$ to several domains in POT1 in both the apo and DNA-bound states (Figure 4.24A). Notably, when bound to DNA, the crosslink between K98 in $TIN2L_{TRFH\text{-}like}$ and K121 in POT_{OB1} was significantly less abundant compared to the apo complex (0.24-fold, $p < 0.05$, Figure 4.22). On the other hand, lysines in the same region (K98, K101, K106) formed 1.23- to 1.79-fold more crosslinks with a cluster of lysines in $POT1_{HJRL}$ (K355, K430, K433 and K469) in the DNA-bound complex ($p < 0.05$, Figure 4.22). This suggested that DNA binding induced a conformational change whereby $TIN2L_{TRFH\text{-}like}$ translocated from a position near OB1/2 of POT1 towards the HJRL domain and regions of the OB3 fold that are distal to the site of DNA interaction. A similar change was also observed between TPP1 and POT1. In both apo and DNA-bound states, $TPP1^{K65}$ in the flexible N-terminus preceding $TPP1_{OB}$ formed crosslinks with K353 in $POT1_{OB3}$ pointing towards the ssDNA-binding region of $POT1_{OB1/2}$, and with K370 in $POT1_{OB3}$ distal to the ssDNA-binding site (Figure 4.16A and B). Upon DNA binding, the abundance of the former pair dropped to 0.6-fold ($p < 0.05$), whilst the latter increased to 3.14-fold ($p < 0.05$, Figure 4.22). In addition, an inter-link between K232 in $TPP1_{OB}$ and $POT1^{K370}$ was 1.67-fold more abundant in the DNA-bound shelterin complex ($p < 0.05$). Together, these results indicate that the N-terminal portion of TPP1 shifted away from the ssDNA-binding region of POT1 upon DNA binding.

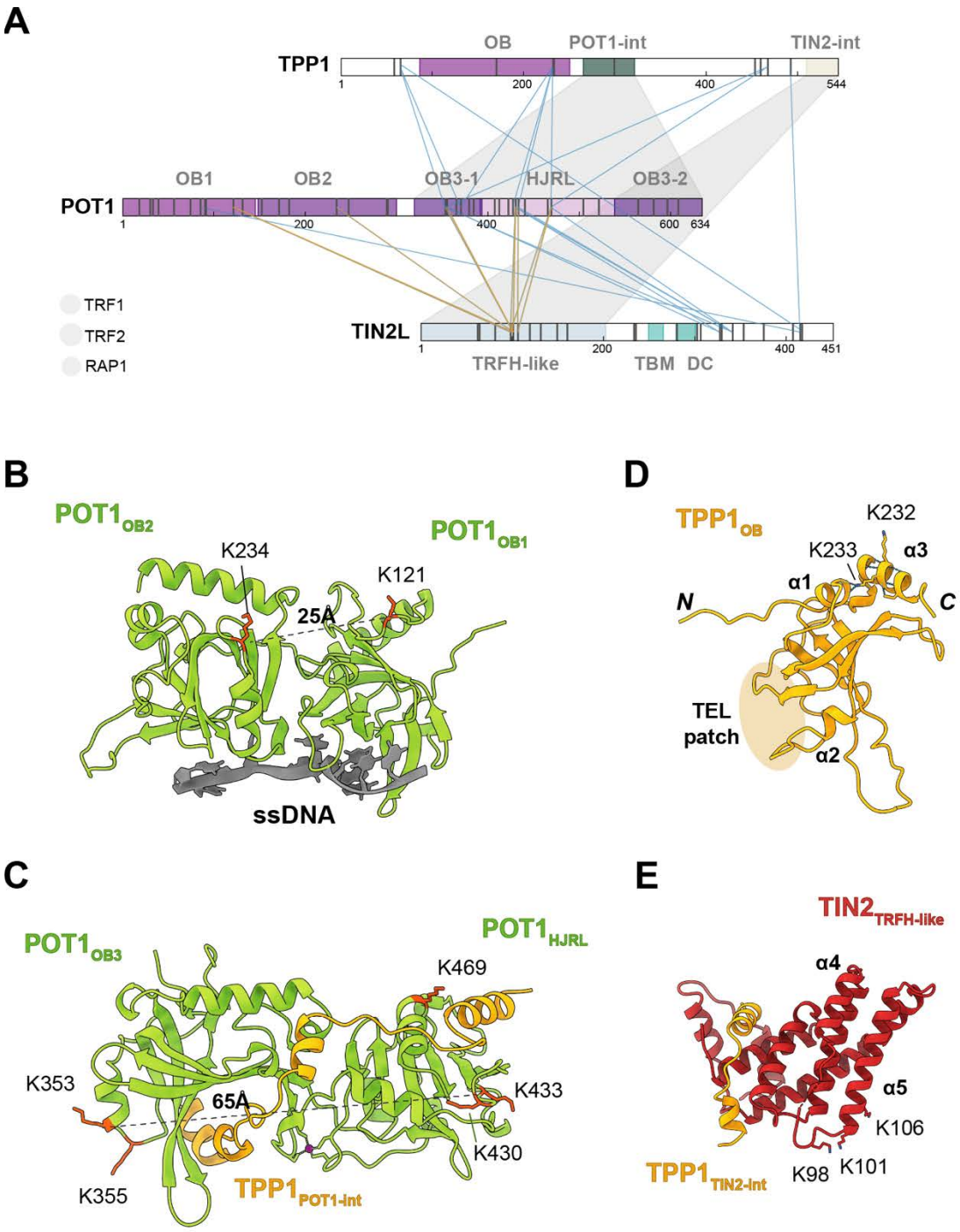

Figure 4.24. Crosslinks in the ssDNA-binding module of the shelterin complex.

(A) Network plot mapping the DSSO crosslinks in the ssDNA module of the shelterin^{TIN2L} complex. Grey vertical lines mark the lysine distributions. Shaded areas map the known protein-protein interactions. **(B)** Crystal structure of POT1 OB1 and OB2 domains (green) ssDNA (grey) (PDB: 3KJP), and **(C)** the POT1 split OB3/HJRL domains in complex with POT1-interacting motif in TPP1 (PDB: 5UN7). **(D)** Crystal structure of the OB-fold in TPP1 (PDB: 2I46). The TEL patch is highlighted in orange. **(E)** Crystal structure of TRFH-like domain in TIN2, and the TIN2-interacting motif in TPP1 (PDB: 5XYF). Lysine residues crosslinked by DSSO are coloured in red. The black dashed lines indicate the Cα–Cα distance between the indicated lysine pairs.

4.2.3.3 Crosslinks between TRF1 and TPP1/POT1

Many crosslinks were observed between the TRF1 and the TPP1/POT1 heterodimer (Figure 4.20). One crosslink pair was found between K136 in $TRF1_{TRFH}$ and K433 in $POT1_{HJRL}$ (Figure 4.21), the abundance of which increased by 1.23-fold in the DNA-bound complex ($p = 0.06$). Another lysine residue K262 in the TRFH domain of TRF1 was paired with K65 in the unstructured N-terminus of TPP1 (Figure 4.21). The abundance of this lysine pair was unchanged by the presence of DNA in the complex (1.01-fold change, $p = 0.85$). The remaining crosslinks were found in the linker region and Myb domain in TRF1, which were bridged by DSSO to lysines in the flexible extreme N- and C-terminal regions in TPP1 and across the entire POT1 protein (Figure 4.21). Notably, DNA-binding led to striking differences in crosslink abundance between $TRF1_{linker}$ and POT1 (Figure 4.25). For instance, the abundance of 6 crosslink pairs between $TRF1_{linker}$ and $POT1_{OB2}$ / $POT1_{OB3\text{-}N}$ increased by more than 2-fold in the DNA-bound complex (Figure 4.25). Additionally, the same lysine in the TRF1 linker formed several crosslink pairs with different lysines in POT1, and vice versa. For example, $TRF1^{K349}$ was crosslinked to K234, K355, and K370 located in the OB2 and $OB3_N$ of POT1 (Figure 4.25). Conversely, K370 in POT1 was found in crosslink pairs with K277, K349, K352, and K389 residues in the TRF1 linker and Myb domain (Figure 4.25). It is important to note that the TRF1 proteins exist as homodimer, and hence the crosslinks could have involved lysines in either one of the TRF1 subunits. Interestingly, K232 in $TPP1_{OB}$ was found crosslinked to both the C-terminal region of TRF1, and to $POT1_{HJRL}$ and $POT1_{OB3\text{-}N}$. K232 in TPP1 is immediately upstream of $TPP1_{PIM}$ that interacts with the C-terminal half of POT1. In the DNA-bound shelterin complex, the crosslink pair $TPP1^{K232}$–$TRF1^{K421}$ was 0.79-fold less abundant ($p<0.05$), whilst $TPP1^{K232}$–$POT1^{K370}$ was 1.67-fold more abundant ($p<0.05$). The abundance of crosslinks between $TPP1^{K232}$ to lysines in $POT1_{HJRL}$ did not change significantly between the two samples ($p<0.05$). These findings suggest that DNA binding induced changes in the conformation of TRF1, bringing the Myb domain away from the TPP1 OB-fold. In the absence of $TRF1_{Myb}$, the $POT1_{OB3\text{-}N}$ and $TPP1_{OB}$ were found to be closer in proximity. In summary, although TRF1 was found in close proximity to the TPP1/POT1 heterodimer in both the apo and the DNA-bound $shelterin^{TIN2L}$ complex, DNA binding induced

major conformational changes including reorganisation of the linker region of TRF1 and movements at the from the TPP1/POT1 interacting surfaces.

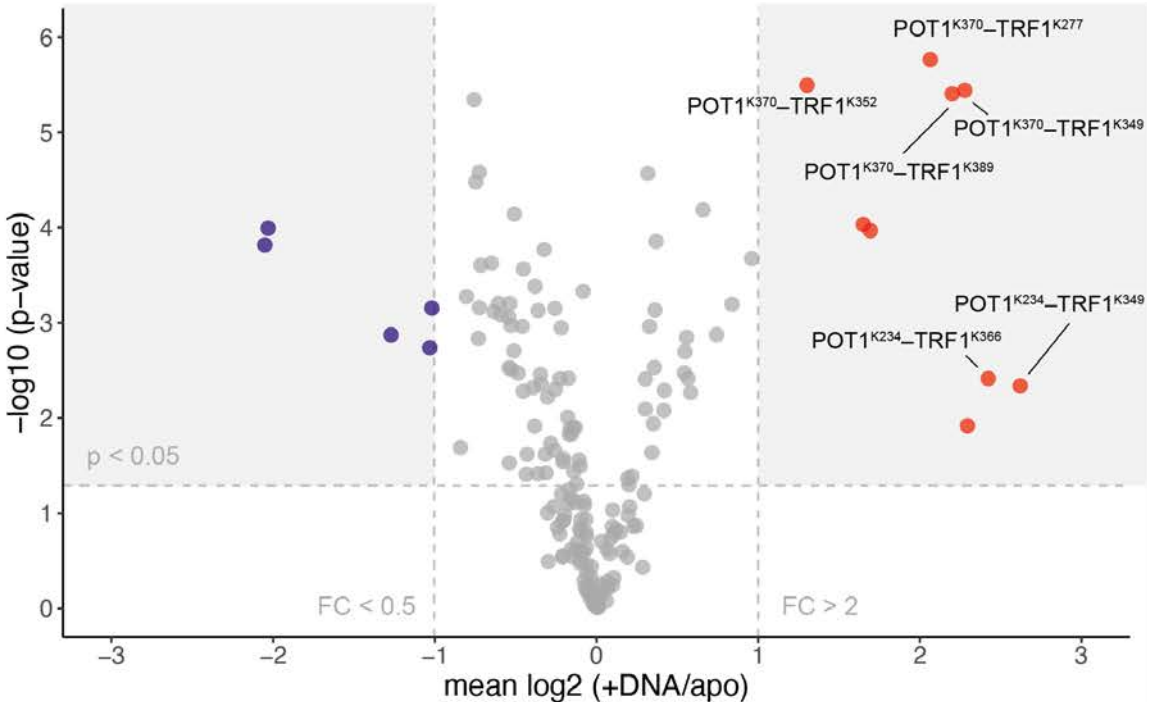

Figure 4.25. Changes in abundance of crosslink pairs between TRF1 and POT1.

Volcano plot showing the mean log2 (+DNA/apo), which compares the relative abundance of crosslink pairs detected in the DNA-bound vs. the apo shelterin^{TIN2L} complex. Thresholds: vertical lines represent fold-changes < 0.5 and > 2, and p < 0.05 beyond the horizontal line crossing the -log10(p-value) axis.

4.3 Conclusions

In this chapter, I presented negative-stain EM work revealing that a subcomplex of human shelterin lacking TRF1 assembled on a minimal telomeric DNA had similar dimensions and overall architecture to the full shelterin complex. After optimising the purification protocols, I was able to enrich two species of DNA-bound shelterin $(–TRF1)^{TIN2S}$ complexes. One contained the minimal 380 kDa $(–TRF1)^{TIN2S}$ complex with 2:2:1:1:1 stoichiometry of TRF2:RAP1:dStrepII-TIN2S:TPP1:POT1. The other was a higher-order species at ~580 kDa, the stoichiometry of which is yet to be confirmed. Many rounds of 2D classification provided a range of particle views for generating initial models using 3D *ab initio* reconstruction. The resulting volumes had repeating features and resembled a dinosaur skull, with two poles as the 'jaw' and 'nose' regions. At this stage, I was unable to unambiguously dock the shelterin domains. Interestingly, a comparison with the negative-stain map of the DNA-bound full $shelterin^{TIN2S}$ complex showed a density that was slightly larger than that of $(–TRF1)^{TIN2S}$ .The maps had a very similar overall shape and I could not identify the density that would correspond to TRF1 in the full $shelterin^{TIN2S}$ map. The extent of similarity between the two maps suggests that the removal of TRF1 did not significantly impact the organisation of the remaining shelterin subunits.

To complement our EM pursuits, I conducted quantitative XL-MS experiments to study the topology of the apo and DNA-bound $shelterin^{TIN2L}$ complexes. Careful quality-control steps guided the selection of the minimal $shelterin^{TIN2L}$ complex during protein purification and confirmed the presence of all subunits after DSSO crosslinking. One unexpected finding was the lack (or very low abundance) of crosslinks from TRF2 and RAP1. The lysine content and distribution throughout these two proteins were similar to TRF1, which were detected in many inter- and self-links. The samples were efficiently digested by trypsin, resulting in good coverage of all shelterin subunits (Appendix Figure 7.5, Table 4.1). The limited representation of TRF2 and RAP1 is unlikely due to the lack of solvent-accessible lysines, particularly in TRF2, as there are many lysines in the flexible linker connecting the TRFH and Myb domains. The equivalent region in its paralog

TRF1 was consistently a crosslink hotspot (Figure 4.21A). Other possibilities include the spacer arm between the crosslinker, which may not be long enough to capture the lysines on the surface of shelterin.

Interestingly, many crosslinks were detected between two DNA-binding proteins TRF1 and POT1 in both states. Although there are no direct protein-protein interactions reported between these two subunits, the abundant lysine pairs formed between them by DSSO indicated they are positioned close to each other (within ~30Å) within the shelterin complex. Changes in the abundance of crosslink pairs indicated that DNA binding induced many conformational changes within the shelterin^{TIN2L} complex. For instance, binding of ssDNA by POT1$_{OB1/2}$ led to local rearrangements of TIN2L$_{TRFH}$ and the N-terminal half of TPP1.

Finally, there are many indications of conformational heterogeneity in the shelterin^{TIN2L} complex. For instance, the same lysine in the flexible region of TRF1 was found crosslinked to several different regions in multiple proteins in the same sample. Several shelterin domains were also identified as crosslink hotspots, including three lysines in TIN2L$_{TRFH\text{-}like}$ that formed crosslink pairs with lysines across all POT1 domains. Through mapping the self-links onto the crystal structures of POT1, I was able to confirm several crosslinks as intra-molecular self-links that were within the theoretical upper limit imposed by DSSO (~26 Å, Figure 4.16). However, it was evident that several self-links were improbable to represent intramolecular interactions given that the lysine pairs were located on opposite faces of POT1 (Figure 4.16). Possible explanations could be that POT1 may exist in other conformations besides the one captured in the cryoEM analysis of POT1 (Smith, et al., 2022), or there may be other POT1 subunit in the close vicinity during DSSO crosslinking. This is unlikely to have formed between two shelterin complexes as the formation of large aggregates were minimised through careful optimisation of protein and DSSO concentrations. The other scenario could be that POT1 was present in more than one copy in the full shelterin complex, although this is less likely as the minimal shelterin complex was enriched in the sample for crosslinking. The work on determining the stoichiometry and structure of this shelterin complex is ongoing.

5 Discussion

5.1 Reconstitution of human shelterin complexes

Since the discovery of the first shelterin subunit, TRF1, in the early 1990s (Zhong, et al., 1992), there have been many cellular, biochemical, and structural studies that identified all the components of the complex and extensively characterised the protein-protein and protein-DNA interaction surfaces involved in shelterin assembly (de Lange, 2018; Chen, 2019). In recent years, various recombinantly purified human shelterin subcomplexes (Lim, et al., 2017) and the full mouse shelterin complex (Erdel, et al., 2017) have been reported; these provided further insight into the stoichiometry of the shelterin complex and mechanisms of shelterin function. More recent studies have demonstrated extensive compositional and structural heterogeneity in shelterin complexes (Smith, et al., 2022; Sekne, et al., 2022; Zinder, et al., 2022). The dynamic nature of shelterin is consistent with the relatively weak (high nanomolar to low micromolar) domain-peptide interactions between shelterin subunits (Chen, et al., 2008; Chen, et al., 2017; Rice, et al., 2017; Gaullier, et al., 2016). This property of shelterin is also likely to contribute to the many roles shelterin subunits play at the telomeres, where the composition and conformation of shelterin may be determined by factors including the length and structure of telomeric DNA, interactions with other protein partners and stages of the cell cycle (Takai, et al., 2010; Vannier, et al., 2012; Sarek, et al., 2019; Zhu, et al., 2000). These considerations highlight the importance of studying the compositional dynamics of shelterin in the context of complexes, which will help shed light on how shelterin performs a myriad of functions in telomere protection and maintenance.

Here, I report the expression and purification of the full six-subunit human shelterin complex (shelterin^{TIN2L} and shelterin^{TIN2S}) and various subcomplexes, with the aim to further our understanding of the structural differences and the dynamic nature of human shelterin. The small TEV-cleavable dStrepII tag at the N-terminus of TIN2L/S allowed us to capture remarkable amounts of shelterin from baculovirus-infected insect cells, which were further purified to obtain

recombinant shelterin complexes with high yield and purity. This makes downstream biochemical, biophysical, and structural characterisations feasible.

I focused my efforts on purification and characterisation of the full shelterin complex and a subcomplex lacking TRF1 because they will be key in our quest to understand the interplay between TRF1 and the PARP enzyme tankyrase, and how tankyrase function modulates telomerase-dependent telomere extension (see section 1.4.2). In accordance with previous studies, the formation of human shelterin complexes does not require telomeric DNA (Lim, et al., 2017). I purified both TIN2S- and TIN2L-containg shelterin complexes to avoid missing any potential interactions that depend on the ~200 amino acids at the TIN2 C-terminus, which has been absent in most *in vitro* reconstitutions of shelterin reported thus far. Notably, both the full and subcomplexes of shelterin^{TIN2L} and shelterin^{TIN2S} purifications consistently produced similar complexes (see sections 2.1.2, 3.2, and 3.3), indicating that the extreme C-terminus of TIN2L does not contribute to the assembly of the shelterin complex.

5.1.1 Recombinant shelterin form higher-order complexes

Recombinant full and (–TRF1) shelterin complexes both formed two stable species (see sections 2.1.2 and 3.2). This was unexpected given the result from a previous reconstitution of shelterin without TRF1 that consisted of a single species with the minimal $(TRF2/RAP1)_2/(TIN2S/TPP1/POT1)_1$ stoichiometry (Lim, et al., 2017). For the full shelterin complex, SEC-MALS and mass photometry identified a species with a MW of ~490 kDa, which was consistent with a minimal six-subunit shelterin complex with a 2:2:2:1:1:1 stoichiometry of TRF1:TRF2:RAP1:dStrepII-TIN2L/S:TPP1:POT1. Moreover, a higher-order species of shelterin was visible during early SEC elution fractions and was determined to be ~690 kDa by mass photometry. These two species were stable over consecutive SEC purifications in the absence of DNA (Figure 2.4). Western blotting analysis indicated that TRF2 and RAP1 were present at a higher abundance in the larger species compared to SEC fractions enriched in the 490

kDa complex. Therefore, the additional 200 kDa may be explained by an extra module of $(TRF2/RAP1)_2$, which has an expected MW of 207.7 kDa.

Similarly, the (–TRF1) shelterin complex also formed two species of different sizes. One was determined to be ~390 kDa that was consistent with a $(TRF2/RAP1)_2/(dStrepII\text{-}TIN2/TPP1/POT1)_1$ complex. Mass photometry also detected another (–TRF1) complex with a MW of ~600 kDa (Figure 3.8). However, separation of the two species of (–TRF1) proved more challenging by chromatographic methods compared to the full shelterin complex. Therefore, I tried other approaches to enrich and characterise the stoichiometries of these different (–TRF1) complexes. I started with the hypothesis that the higher-order (–TRF1) complex contained an extra $(TRF2/RAP1)_2$ compared with the smaller 390 kDa species. However, providing excess insect-cell expressed recombinant TRF2/RAP1 to a mixed population of the two (–TRF1) complexes did not drive the formation of the higher-MW (–TRF1) complex (section 3.2.2.2). It is worth noting that another possible regulatory switch that may affect protein-protein interactions is PTMs. For instance, phosphorylation of TIN2L has been suggested to enhance the strength of the TIN2-TRF2 interaction, whilst TRF2 phosphorylation have been shown to regulate cell cycle-dependent unwinding of T-loops to mediate telomere replication (Nelson, et al., 2018; Sarek, et al., 2019). Although numerous PTMs have been reported for shelterin subunits, the roles they play in regulating protein-protein interactions within shelterin and how they impact shelterin-mediated functions are still poorly understood. Furthermore, mutations in the conserved leucine 260 in $TIN2_{TBM}$, previously shown to abolish the $TIN2_{TBM}$-$TRF2_{TRFH}$ interaction (Chen, et al., 2008), did not markedly reduce the abundance of the higher-order (–TRF1) complex. Together, these observations suggest that under the conditions tested, $TIN2_{TBM}$ is unlikely to bind to TRF2 via the TBM-TRFH interaction interface in the absence of TRF1. In summary, it is unclear whether the higher-MW complex contains more TRF2/RAP1, TIN2/TPP1/POT1, or a mix of both, compared to the minimal full shelterin complex.

As I wrote this thesis, a reconstitution of the full human shelterin complex was reported (Zinder, et al., 2022). The authors observed a dimeric shelterin

complex at 780 kDa by mass photometry composed of two copies of each shelterin subunit (expected MW: 759 kDa) (Zinder, et al., 2022). As each TRF1 and TRF2 homodimer have two TIN2 binding sites, the authors proposed that the TRF1 and TRF2 together recruit two TIN2/TPP1/POT1 heterotrimers (Zinder, et al., 2022). There are two important differences between the shelterin reconstitution reported by Zinder et al. (2022) and the one I described. Firstly, the expression constructs were different. The shelterin complexes I produced had a dStrepII tag (~4 kDa) on TIN2L or TIN2S. In the Zinder et al. (2022) study, two subunits retained their affinity tags throughout the purification and for mass photometry analyses: the TIN2S subunit carried an eGFP affinity tag (~27 kDa), and each RAP1 subunit had a His_6-MBP tag (~44 kDa) (Zinder, et al., 2022). Although the authors demonstrate that the eGFP tag does not induce dimerisation of TIN2S in the TIN2S/TPP1/POT1 subcomplex (Zinder, et al., 2022), it is uncertain whether these larger affinity tags impact the assembly of higher-order shelterin complexes. Secondly, the reconstitution approach was different. I expressed all six shelterin subunits in Sf9 insect cells using a single baculovirus and extracted all shelterin species that contained the dStrepII-TIN2L/S subunit. Zinder et al. (2022) built their systems bottom-up; they first generated their TRF1 complex ($TRF1_2$/eGFP-$TIN2S_1$/$TPP1_1$/$POT1_1$) complex by mixing $(TRF1)_2$ and eGFP-TIN2S/TPP1/POT1 in a 1:2 ratio. Next, the $(TRF1)_2$/eGFP-TIN2S/TPP1/POT1 complex was incubated with (TRF2/His_6MBP-Rap1)$_2$ in a 1.33:1 ratio and further purified by ultracentrifugation and enrichment prior to mass photometry analysis. Due to these differences, it is difficult to make direct comparisons between the two. In addition, the shelterin subunits have similar MWs, which makes mass photometry data difficult to interpret given the wider error ranges associated with this technique compared with others such as native mass spectrometry. Therefore, further investigation with more accurate methods of determining MW will help elucidate the stoichiometry of shelterin subunits.

5.1.2 TPP1 plays an important role in shelterin assembly

The shelterin subcomplex lacking the TPP1/POT1 heterodimer is of interest as a control for studying how the full shelterin complex is involved in recruiting and enhancing the enzymatic processivity of telomerase (Wang, et al., 2007; Lim, et al., 2017; Sekne, et al., 2022). Purifications of the (–TPP1/POT1) subcomplex only produced a small proportion of the 4-subunit subcomplex, as determined by SDS-PAGE/Western blotting and mass photometry analyses (see Figure 3.25). Instead, the most stable and abundant complex consisted of the TRF1/dStrepII-TIN2L subcomplex in a 2:2 ratio (see section 5.1.3). This is in accordance with previous studies, which have shown that TPP1 is key in driving higher-order assemblies of shelterin (O'Connor, et al., 2006). Binding of TPP1 to TIN2 induces allosteric changes in the TRFH-like domain of TIN2, enabling TIN2 to accommodate both TRF1 and TRF2 (Hu, et al., 2017; Janovič, et al., 2019). Given that TPP1 and POT1 are the least abundant shelterin subunits on the telomeres (Takai, et al., 2010), it would be interesting to investigate the proportion of TRF2 and TRF1 that are bridged by TIN2 on telomeres and the functions of the (–TPP1/POT1) subcomplex *in vivo*.

5.1.3 TRF1 forms a stable complex with TIN2L/S in a 2:2 stoichiometry

Crystal structures of the TRFH domain of TRF1 and TRF2 have shown that each homodimer of TRF protein can potentially bind two $TIN2_{TBM}$ peptides via the TRFH domain (Chen, et al., 2008). However, there has only been a report of full-length TRF2 and TIN2 existing in a 2:1 stoichiometry in the (–TRF1) subcomplex (Lim, et al., 2017), and no evidence thus far has been provided for a $(TRF1:TIN2)_2$ complex. I obtained the TIN2/TRF1 subcomplex as a by-product in several protein purifications, which was often found present with a 2:2 stoichiometry. During SEC purifications of the (–TPP1/POT1) subcomplex, the predominant and stable species consisted of mainly TRF1 and dStrepII-TIN2L/S, as judged by Western blotting (Figure 3.24). Mass photometry analysis identified an abundant

species with a mean mass of ~210 kDa, which corresponded to a TRF1/dStrepII-TIN2 subcomplex with a 2:2 stoichiometry (expected MW: 209 kDa). The $(TRF1/dStrepII\text{-}TIN2S)_2$ subcomplex was also observed during purifications of the full shelterin complex (Figure 2.11C and D). Recently, a study of the full shelterin complex have also proposed that both TRF1 and TRF2 bind to TIN2 in a 2:2 stoichiometry, resulting in the formation of a dimeric shelterin complex (discussed in section 5.1.1) (Zinder, et al., 2022). Further investigation is required to confirm whether the 2:2 ratio of TRF1:TIN2 can form in the presence of the other four shelterin subunits (see section 5.3).

5.1.4 Recombinant shelterin complexes bind to telomeric and non-telomeric DNA with nanomolar affinities

Previous studies have shown that the three individual DNA-binding subunits, TRF1, TRF2 and POT1, have high affinities for telomeric ds and ssDNA (K_D < 10 nM), respectively (Lei, et al., 2004; Court, et al., 2005). In the TRF1 and TRF2 Myb domains, helices 1 and 2 make extensive contacts with the DNA backbone that dock the Myb domains onto dsDNA (Court, et al., 2005), while base-specific contacts between key lysine and arginine residues in helix 3 and the DNA major groove, and several indirect water-mediated contacts, confer specificity for the highly conserved 5'-TAGGGTT-3' telomeric sequence (Bianchi, et al., 1999; Court, et al., 2005). The flexible linker of 100 – 200 amino acids between the TRFH and Myb domains in TRF1/2 enables the TRF homodimers to bind half-sites separated by up to 30 base pairs (Bianchi, et al., 1999). On the other hand, the two OB folds in the N-terminus of POT1 recognise the 5'-TAGGGTTAG-3' sequence in telomeric ssDNA (Baumann and Cech, 2001; Lei et al., 2004; Loayza et al., 2004). OB1 binds to the first six nucleotides (TTAGGG) and the latter four is bound by OB2. A short linker of 5 amino acids between POT1 OB1 and OB2 has been shown to tolerate a separation between these two parts of the minimal sequence with reduced affinities still within the low nanomolar range (Smith, et al., 2022). The OB-fold in TPP1 does not make direct contacts with DNA, but TPP1 binding to POT1 enhances the $POT1_{OB1\text{-}2}$-ssDNA interaction (Wang et al.,

2007; Xin et al., 2007; Taylor et al., 2011). Previously, a study of the mouse full shelterin complex showed that removal of POT1a reduced the affinity of the complex for a 64/28-mer telomeric DNA by 16-fold (Erdel, et al., 2017). Similarly, an *in vitro* reconstitution of the human shelterin complex lacking TRF1 showed that removal of POT1 reduced its affinity for a model telomeric ds/ss DNA by 14-fold (Lim, et al., 2017). A larger effect was observed when TRF2 was absent in the (–TRF1) complex, where affinity for the model telomeric DNA reduced by 50-fold (Lim, et al., 2017). This shows that TRF2 has a greater contribution to the high affinity binding shelterin complexes to telomeric DNA. This is in agreement with *in vivo* experiments where evidence show that removal of TRF2 from telomeres negatively impacts the localisation of POT1 at telomeres.

Both studies of the mouse and human shelterin mentioned above used the agarose-EMSA method to assess the binding affinities of shelterin complexes to DNA (Erdel, et al., 2017; Lim, et al., 2017). The advantages of this approach include that smaller quantities of reagents are required, particularly with ^{32}P-radiolabeling of the nucleic acid, for measurement of high-affinity interactions (low picomolar apparent K_D, $K_{D,app}$, values). However, an important limitation of this technique is that the samples are not at chemical equilibrium during electrophoresis. Using the switchSENSE technology (Rant, et al., 2004; Langer, et al., 2013), I have determined the $K_{D,app}$ of the full shelterin^{TIN2L} and (–TRF1)^{TIN2L} subcomplex to the minimal model telomeric DNA and a non-telomeric DNA control under equilibrium conditions (sections 2.4 and 3.5). Both complexes were able to bind the telomeric sequence in a 1:1 ratio with nanomolar affinity. The full shelterin complex had a 3.6-fold higher affinity for telomeric DNA compared with the subcomplex lacking TRF1 ($K_{D,\ app}$ = 3.11 vs. 11.1 nM, respectively). This difference in affinity can be explained by the lower rate of association of the (–TRF1) subcomplex compared with the full complex (k_{ON} = 1.69 x 10^5 vs. 5.53 x 10^5 M^{-1} s^{-1}, respectively). This indicated that TRF1 contributes to the overall binding affinity of shelterin to telomeric DNA by offering another pair of Myb domains to dock the shelterin complex onto telomeric DNA. Interestingly, this trend is reversed in the context of non-telomeric DNA. Albeit at reduced affinities compared with the telomeric substrate, the (–TRF1) subcomplex had a 2.2-fold higher affinity for non-telomeric DNA than the full shelterin complex ($K_{D,\ app}$ = 96.7

vs. 211 nM, respectively). This was due to a higher rate of association of the (–TRF1) complex to non-telomeric DNA compared with the full complex. Notably, the (–TRF1) was 8.5-times faster in binding to non-telomeric compared with telomeric DNA (k_{ON} = 1.44 vs. 0.17 x 10^6 M^{-1} s^{-1}, respectively), whereas no such difference was observed for the full complex. Finally, for both complexes, the k_{OFF} values were 2 orders of magnitudes higher for the non-telomeric DNA (10^{-1} vs 10^{-3} s^{-1} for non-telomeric and telomeric DNA, respectively).

Together, these observations suggest shelterin can bind to DNA in a non-sequence specific manner. This maybe partially explained by the non-sequence specific and low affinity interactions made by the TRFH domain and basic N-terminal region of TRF2. In non-telomeric regions, the high on- and off-rates enable shelterin to perform a rapid 3D search for its target telomeric sequence. Once it has found the target TTAGGG repeats, it had a much longer residing time on DNA compared with a non-telomeric sequence. Furthermore, the different rates of association between the two complexes for telomeric and non-telomeric DNA indicate that TRF1 contributed to the higher affinity and specificity of the shelterin complex for telomeric DNA. This supports the model for telomere recognition proposed by Erdel et al., (2017).

5.2 Towards understanding the architecture of shelterin complexes

Finally, I integrated structural information from electron microscopy and crosslinking mass spectrometry to characterise the protein-protein interactions within the human shelterin complex. To do this, I joined forces with Dr Oviya Inian in our lab, who is working on studying the full shelterin complex at high resolution using cryo-EM. My focus was two-fold:

(1) examining the structure of DNA-bound (–TRF1) subcomplex. This will help us identify the position of TRF1 in the full complex. This may also provide structural insights into how the interplay between TRF1 and tankyrase regulates telomere length maintenance, which will be complemented with extensive biochemical and biophysical assays (see Perspectives).

(2) Explore the protein-protein interactions within shelterin to complement EM studies of the full complex.

5.2.1 EM studies of the full and (–TRF1) shelterin complex reveal structural heterogeneity

For structural studies of the (–TRF1) subcomplex, I reconstituted the DNA-bound $(–TRF1)^{TIN2S}$ complex using limited crosslinking and proceeded with negative-stain EM to examine the structural integrity of the complex. Previous studies have shown that the extensive linkers in TRF1/2 and the short linker between OB1 and OB2 in POT1 mean that the DNA-binding domains are very flexible and tolerant of telomeric DNA substrates of various lengths. With the aim to reduce structural heterogeneity and to simplify the image processing downstream, I used a minimal telomeric DNA with sequence lengths optimised for binding of the minimal (–TRF1) complex to dock two Myb domains from a TRF2 homodimer and one POT1 subunit with OB1-2 domains (teloDNA2) (Court, et al., 2005; Palm, et al., 2009; Sfeir, et al., 2005; Lei, et al., 2004). As discussed previously, the (–TRF1)

subcomplex exists in two flavours (see section 1.1 above). The samples used for negative-stain EM analysis also contained both populations, which complicated the image analysis process. To explore the extent of compositional heterogeneity in the sample, I used 3D *ab initio* reconstruction in cryoSPARC to reconstruct several 3D initial models. This is a quick and computationally cheap way to identify whether 2D images of the (–TRF1) complex are best explained by one or more than one 3D volume. The stochastic gradient descent (SGD) algorithm used by this process also renders the process insensitive to initialisation. This means that with enough views representing all the 3D structures in the dataset, the SGD implementation increases the likelihood of finding accurate initial 3D volumes for later high-resolution refinement, even with an arbitrarily computed volume as the starting model (see Methods).

After multiple rounds of 2D and 3D classification (using the SGD algorithm in 3D *ab initio* reconstruction), I was able to obtain several low-resolution reconstructions of DNA-bound (–TRF1)^{TIN2S}. I then compared these maps with EM density maps obtained from negative-stain and cryo-EM studies of the full complex (performed by Dr Inian). The sample preparation for the full shelterin^{TIN2S} bound to a minimal telomeric DNA sequence (teloDNA1, instead of teloDNA2 for the subcomplex). The other difference was that the two species of crosslinked DNA-bound shelterin^{TIN2S} was better resolved during the final SEC step, and a sample enriched in the minimal shelterin complex was taken forward for EM experiments.

I identified one initial map of the DNA-bound (–TRF1)^{TIN2S} complex that closely resembled the overall shape and low-resolution features the full shelterin complex. This model was taken further for 3D homogenous refinement, which did not drastically change the 3D density. Comparison of these two densities showed the core of the density matched well, although there were several regions that were different, indicative of major conformational changes. It is possible that through several rounds of 2D and 3D classifications in cryoSPARC, I have enriched one of the two species of the (–TRF1) subcomplex. In summary, these EM indicated that the absence of TRF1 did not significantly alter the overall architecture of the full shelterin complex. This supports the concept that shelterin

can exist as various subcomplexes at telomeres as independent modules to offer different functions. In addition, conformational heterogeneity was expected as many shelterin subunits contain extensive disordered regions. This is the nature of the shelterin complex; many protein-protein interactions within shelterin and between shelterin and its accessory factors at the telomeres are mediated by domain-peptide interactions that likely facilitate its role in telomere protection and maintenance.

5.2.2 XL-MS experiments reveal spatial relationships of shelterin subunits within the full shelterin complex

In recent years, XL-MS has become an increasingly popular tool to complement structural studies of dynamic multi-subunit protein complexes. It offers lower-resolution structural information that is dependent on the crosslinker, the spacer arm length and the reactive moiety. DSSO is a popular crosslinker used for XL-MS studies for several reasons. Firstly, it is a homo-bifunctional crosslinker with a NHS ester moiety at either side of the spacer, which can react with ε-amines in lysines. This is favourable due to the relatively high prevalence of lysine residues in proteins (~6% of all residues) and lysines are often distributed across solvent-accessible areas (Yu & Huang, 2017). Secondly, it has a spacer arm length of 10.3 Å, which can theoretically crosslink two lysine residues with Cα-Cα distance of ~26Å. This covers a wide enough distance to capture lysines across regions within a protein and between proteins (intra- and inter-links). Finally, another key advantage of DSSO is that it is cleavable during MS, which allows for more accurate identification of crosslinked peptides.

The aims of my XL-MS analysis of the full shelterin complex are two-fold: (1) to explore the spatial arrangement of shelterin subunits in solution, and (2) to detect conformational changes that may occur upon DNA-binding. To address these questions, I performed two quantitative XL-MS experiments. Experiment 1 compares the crosslinks formed after a 10- or 30-min crosslinking time, and experiment 2 compares the crosslinks identified in the apo vs. the DNA-bound

shelterin^{TIN2L} complex. Two initial observations increased my confidence in the XL-MS results. Firstly, lysines near three of five of the known domain-peptide interactions interfaces were observed in both experiments, namely at the $TRF1_{TRFH}$-$TIN2_{TBM}$ and $TRF2_{TRFH}$-$TIN2_{TBM}$ interface, and the $TPP1_{PIM}$-$POT1_{HJRL}$ interface (Chen, et al., 2008; Rice, et al., 2017). This indicates that DSSO crosslinking was able to capture some of the surfaces that are known to be in spatial proximity to each other. Secondly, three key lysines residues involved in DNA binding ($TRF2^{K530}$, $TRF1^{K421}$, and $POT1^{K33}$) were found to be crosslinked in the apo state. In the presence of telomeric DNA, the abundance of the latter two significantly decreased, indicating that the DNA-binding domains are engaged in DNA binding and are protected from the reactive NHS ester in DSSO. This gave confidence that changes in abundance of crosslinks in experiment two are correlated with conformational changes upon DNA-binding.

5.2.2.1 TRF1, TPP1 and POT1 are in close spatial proximity

Several crosslink pairs were formed between the C-terminal portion of TRF1 including the linker and the Myb domain and the C-terminal half of POT1 covering the OB3-1 and HJRL domains, and between $TRF1_{Myb}$ and $TPP1_{OB}$. These were observed in both apo and DNA-bound states, indicating that these three surfaces were found adjacent to each other independently of telomeric DNA. However, there were indications of conformational changes that occurred upon DNA binding. For instance, crosslinks between the Myb domain of TRF1 and near the POT1-interacting surface of TPP1 reduced significantly when DNA was present in the shelterin complex, indicating rearrangement of the TRF1 Myb domain further away from this TPP1 region compared to the apo complex.

5.2.2.2 Additional TRF1-TIN2

Besides crosslinks between regions near the $TRF1_{TRFH}$-$TIN2_{TBM}$ interaction surface, a number of crosslinks was also found between the TRFH-like domain of TIN2 and the C-terminal portion of TRF1 covering the linker and the Myb domain. Several of these crosslinks were consistently observed in both XL-MS experiments comparing crosslinking time and the two different states of shelterin, indicating that the $TRF1_{Myb}$ may be found in close spatial proximity with $TIN2_{TRFH}$. Interestingly, the lysines in $TIN2_{TRFH}$ involved in the crosslink pairs were adjacent to the TRF2 binding site. Previously, it has been shown that the addition of TRF1 to the TIN2/TRF2 complex can disturb the TRF2-TIN2 interaction (Janovič, et al., 2019). Therefore, it is possible that the Myb domain of TRF1 is orientated close to the TRFH domain in the shelterin complex, and in a position that blocks TRF2 binding site in TIN2.

5.2.2.3 Evidence of sample heterogeneity

As mentioned in section 5.1.3, there is evidence from purifications of the full complex that TRF1 and TIN2 can exist in a 2:2 stoichiometry. Often, in the SDS-PAGE analyses of shelterin prior to crosslinking, it appeared that the lower bands corresponding to TRF1 and TIN2L were stained more intensely compared with the upper bands, which corresponded to TRF2, TPP1 and POT1 subunits. Although the additional heparin purification removes a substantial amount of excess TRF1/TIN2, there are still more crosslinks originating from these two subunits compared to any other shelterin component. After crosslinking, all the subunits shifted to a smear at the higher-MW region of the gel, indicating the subunits were crosslinked. Therefore, the excess TRF1/TIN2L observed in the non-crosslinked sample was incorporated in the full shelterin complex and was not merely a separate by-product of the purification. This possibly explains the high abundance of crosslinks between TRF1 and TIN2L subunits that was detected in both XL-MS experiments. In addition, both experiments found very few crosslinks from TRF2 and RAP1, although Western blot analyses indicated both subunits were present and crosslinked by DSSO. One possible explanation

for this is that there is a mix of different species of shelterin after the final SEC purification: one may consist of more TRF1 and TIN2L (such as in the 2:2 ratio), whilst the other may have incorporated TRF2/RAP1. Although mass photometry indicated that the predominant species in the sample used for XL-MS is the size expected for a $(TRF1/TRF2_2RAP1)_2/(TIN2L/TPP1/POT1)_1$ complex, the difference in size between this and the other species with more TRF1/TIN2L (and with maybe less TRF2/RAP1) may be too small to be detected by mass photometry. The dual-tagged shelterin appears to be a better construct to produce both the minimal and higher-order shelterin species, and can be further optimised to give more homogeneous shelterin complexes for biochemical and structural studies in the future.

5.3 Future directions

The biochemical, biophysical and structural characterisations of shelterin have shown that shelterin subunits can form dynamic and heterogeneous complexes. These findings also raised new questions that need to be addressed, including:

- What are the key interactions between shelterin subunits that regulate the formation of shelterin complexes? For instance, how does TPP1 alter the TRF1/TIN2 interaction interface in a way that enables TIN2 to bind both TRF1 and TRF2 simultaneously? Understanding how protein-protein and protein-DNA interactions contribute to the formation of shelterin complexes may help identify and purify homogenous shelterin complexes for structural and biochemical characterisation.

- Does the spatial proximity of TRF1 affect the ability of the TPP1/POT1 heterodimer to recruit telomerase? TRF1 is a negative regulator of telomere length maintenance. Understanding the spatial organisation of TRF1 and TPP1/POT1 within the shelterin complex may provide insight into factors that regulate telomerase recruitment and telomere length maintenance.

- How does tankyrase and PARylation change the stoichiometry of shelterin complexes? TRF1 is a substrate of the PARP enzyme tankyrase. *In vitro* biochemical studies have shown that PARylation of TRF1 reduces its ability to bind to telomeric DNA, and overexpression of tankyrase in telomerase-positive cells removes TRF1 from telomeres and promotes telomere elongation. It is yet unclear how PARylation of TRF1 displaces it from telomeres, whether it is due to loss in DNA binding alone or whether it also affects the protein-protein interactions that tether TRF1 to telomeres.

Answering these questions will provide insight into how shelterin complexes form and how the various complexes contribute to telomere length regulation and protection.

6 Methods and Materials

6.1 Plasmids and Cloning

6.1.1 Shelterin plasmids

The cDNA constructs of all six shelterin subunits were codon-optimised for *E. coli* expression by GenScript (see Table 6.1), and are under control of either the polyhedron (polH) or p10 gene promoter of the *Autographa californica* nuclear polyhedrosis virus (AcNPV) and ended in polyadenylation signals from the simian virus 40 (SV40) late gene or the herpes simplex virus type 1 thymidine kinase (HSV-tk) gene (see Table 6.2). Each gene expression cassette (GEC), spanning from promoter to terminator was flanked by two identical restriction sites to enable specific addition or removal of GECs with ease (see Table 6.2). The codon-optimised shelterin cDNAs were inserted into the pACEBac1 acceptor vector as follows:

RE site – promoter – Kozak sequence (AGCCGCCACC) – Start codon – [Tag & TEV] – insert – Stop codon – terminator – RE site

Affinity-tagged shelterin subunits were synthesised by extending the open reading frame (ORF) with a double-StrepII (dStrepII) tag or a polyhistidine tag followed by a TEV protease site. The amino acid sequences of the affinity tags and the TEV cleavage site were as follows, starting with the start codon:

dStrepII-TEV: M-SA-WSHPQFEK-$(GGGS)_2$-GGSA-WSHPQFEK-GA-ENLYFQG

His_6-TEV: M-HHHHHH-ENLYFQG

His_{10}-TEV: M-HHHHHHHHHH-ENLYFQG

The dStrepII-TEV-TRF1 and dStrepII-TEV-TIN2L expression constructs were synthesised and codon-optimised for *E. coli* expression by GenScript. The dStrepII-TEV-TIN2S construct was created by PCR amplification of dStrepII-TEV-TIN2L using the Kapa HiFi HotStart PCR kit (Kapa Biosystems), starting from the unique restriction site preceding the promoter and ending with a premature TGA stop codon to give a gene product lacking the residues 355 – 451 at the C terminus of TIN2L. The dStrepII-TEV-RAP1 and dStrepII-TEV-TRF2 constructs were cloned by stepwise PCR amplification to extend the ORF of the untagged versions with a dStrepII-TEV sequence by Anthony Marchand and Dr Oviya Inian, respectively. The His_6-TEV and His_{10}-TEV-TRF2 constructs were cloned by ligation-independent cloning using the codon-optimised untagged TRF2 with a His_6-TEV or His_{10}-TEV sequence by Dr Oviya Inian. The unique restriction sites flanking the genes were kept intact.

The plasmid containing all six components of shelterin, with a TEV-cleavable dStrepII tag preceding the TRF1 coding sequence, was assembled in the pACEBac1 acceptor vector by Dr Iris Gawarzewski. The protocol was a modified version of the MultiBac Turbo System, based on the nicking cloning strategy developed in David Barford's laboratory and later adapted by Dr David Rees in Dale Wigley's laboratory. I generated the other multi-protein shelterin constructs in this thesis by restriction cloning to switch dStrepII-TEV-TRF1 for untagged TRF1, and untagged TIN2L for dStrepII-TEV-TIN2L or dStrepII-TEV-TIN2S (Figure 6.1). From these two new shelterin constructs, shelterin^{TIN2L} and shelterin^{TIN2S}, the coding sequence for various subunits was removed to generate constructs of the shelterin subcomplexes. The dual-tagged shelterin constructs were generated by Dr Oviya Inian. See Table 6.5 for the list of expression constructs. Correct assembly was checked by restriction digest using the unique restriction sites and verified by sequencing of the ORFs.

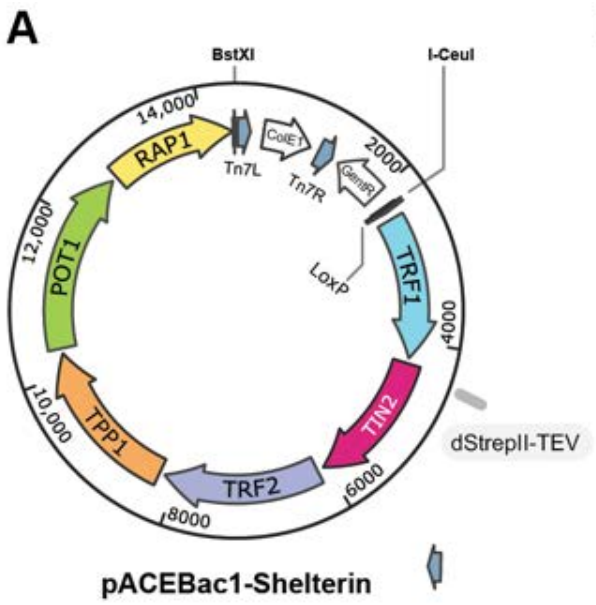

Figure 6.1. Schematic of the shelterin expression construct.

Gene	Protein	NCBI RefSeq	Mass (Da)	Length (a.a.)
TERF1	TRF1	NP_059523.2	50,246	439
TERF2	TRF2	NP_005643.2	59,594	542
TERF2IP	RAP1	NP_061848.2	44,260	399
TINF2	TIN2L	NP_001092744.1	50,023	451
TINF2	TIN2S	NP_036593.2	39,444	354
ACD	TPP1	AAH16904.1	57,733	544
POT1	POT1	NP_056265.2	71,442	634

Table 6.1. List of identifiers for human shelterin subunits.

Gene	Promoter	Terminator	Restriction enzyme
TERF1	polH	SV40	*MluI*
TERF2	polH	SV40	*XhoI*
TERF2IP	p10	HSVtk	*EcoRI*
TINF2	polH	SV40	*NheI*
ACD	polH	SV40	*XmaI*
POT1	p10	HSVtk	*BamHI*

Table 6.2. List of gene expression cassettes used for constructing donor and acceptor plasmids.

Abbreviations: HA, haemagglutinin; HSVtk, Herpes simplex virus thymidine kinase; polH, polyhedrin.

Primer	Sequence (5' → 3')
TIN2S forward	TATAATATTGCCGCCACCATGTCTG
TIN2S reverse	GCCAGCGACGGAACAAAAAGAATGATAATATTGTG

Table 6.3. Primers used for generating dStrepII-TEV-TIN2S.

6.2 Protein sequences and quantification

6.2.1 Sequences of human shelterin subunits

The full human shelterin complex consists of six different proteins, namely TRF1, TRF2, RAP1, TIN2L or TIN2S, TPP1 and POT1 (see Table 6.1 for the list of protein identifiers). The full shelterin complex and subcomplexes lacking TRF1, TRF2/RAP1, POT1, or TPP1/POT1 were expressed in Sf9 insect cells using a single baculovirus, all carrying an N-terminal TEV-cleavable dStrepII tag on TIN2L or TIN2S. The dual-tagged full and (–TRF1) shelterin complexes also have a N-terminal TEV-cleavable His_6 or His_{10} tag on TRF2 (see section 6.1 and Table 6.5).

6.2.2 SDS-PAGE and Western blotting

For SDS-PAGE analysis, protein samples were boiled at 100 °C for 5 mins in SDS sample buffer (4X SDS sample buffer contained 200 mM Tris-HCl pH 6.8, 8% SDS, 40% glycerol, 20% BME, 0.04% bromophenol blue), before loading onto SDS-PAGE gels alongside 3 µL of Precision Plus Protein Marker (Bio-Rad) or Color Prestained Protein Standard, Broad Range (New England Biolabs). The SDS-PAGE gels used include 4–15% and 4–20% Criterion™ TGX™ Precast gels (Bio-Rad), 10-20% and 16% Novex™ WedgeWell™ Tris-Glycine Mini Protein gels (Invitrogen), and homemade gels (10%, 13%, and 15%), which were run using either the Criterion system (for precast gels, Bio-Rad) or the Mini-PROTEAN system (for homemade gels, Bio-Rad) in Tris-Glycine running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3). SDS-PAGE gels were run at constant voltage of 180 V at 4 °C to prevent overheating. The protein bands were visualised using Coomassie staining (InstantBlue, Expedeon) and scanned using the ChemiDoc Imaging System (Bio-Rad).

For Western blotting, protein resolved using unstained SDS-PAGE gels were transferred onto nitrocellulose membranes (Amersham Protran Premium Nitrocellulose Membrane 0.45 µm, Cytiva) in a Tris/Glycine transfer buffer (25 mM Tris, 192 mM glycine, pH 8.3, Bio-Rad) with added 20% methanol, and run at a constant current of 300 mA for 1 hour at 4 °C. Protein transfer to the membrane was checked using reversible Ponceau S staining (Sigma-Aldrich). Then, membranes were blocked for 1 hour at 4 °C using a blocking solution consisting of 0.01% (v/v) Tween-20 and 5% (w/v) milk powder dissolved in phosphate-buffered saline (PBS). Next, membranes were incubated with primary antibodies (see Table 6.4) in fresh blocking solution for 2 hours at room temperature or 4 °C overnight, followed by three washes with 0.01% (v/v) of Tween-20 in PBS (PBST) for 15 mins each. The appropriate secondary antibodies (1:10,000 dilutions of IRDye® 800CW donkey anti-mouse and/or IRDye® 680RD donkey anti-rabbit antibodies, LI-COR) were incubated with the membrane for 1 hour at room temperature and rinsed three times with PBST, for 15 mins each. Finally, the membranes were scanned using the Odyssey CLx

Infrared Imaging System (LI-COR) and visualised using the Image Studio™ Lite software (LI-COR).

Antibody	Host	Dilution	Source
anti-TRF1 (TRF-78)	mouse mAb	1:10,000	#ab10579, abcam
anti-TRF1	rabbit pAb	1:2,000	custom generated, ThermoFisher
anti-TRF2 (D1Y5D)	rabbit mAb	1:2,000	#13136, Cell Signaling Technology
anti-RAP1 (D9H4)	rabbit mAb	1:2,000	5#433, Cell Signaling Technology
anti-TPP1 (D4E2R)	rabbit mAb	1:4,000	#14667, Cell Signaling Technology
anti-POT1 (EPR6319)	rabbit mAb	1:1,000	#ab124784, abcam
anti-Strep	mouse mAb	1:4,000	#71590, Merk

Table 6.4. List of primary antibodies.

6.2.3 Protein quantification by UV spectrophotometry

Protein concentration was measured using the NanoDrop UV-Vis spectrophotometer (ThermoFisher) with the theoretical molecular weights (MW, in kDa) and extinction coefficients ($M^{-1}\,cm^{-1}$) derived from ExPASy ProtParam (Gasteiger, et al., 2005). For single proteins, these parameters were derived from individual polypeptide sequences. For the six-subunit shelterin complex, two copies of the TRF1, TRF2, and RAP1 sequences, and one copy of TIN2, TPP1, and POT1 were used to reflect the 2:2:2:1:1:1 stoichiometry of TRF1:TRF2:RAP1: TIN2L/S:TPP1:POT1 (see section 1.6.1). Parameters for the shelterin (–TRF1) subcomplexes were calculated as for the full complex but lacking TRF1.

The MW is simply the addition of weights of all the amino acids in the protein or the protein complex. The extinction coefficient, ε, at a set wavelength collectively describes all the photon absorbing species present within the molecule at a defined wavelength. From a known protein sequence, the molar extinction coefficient of a protein (in water, assuming all Cys residues are reduced) at 280 nm, ε_{280}, can be calculated from first principles (Pace, et al., 1995; Edelhoch, 1967):

$$\varepsilon_{protein} = \sum n(A) \times \varepsilon(A)$$

where $n(A)$ is the number of a particular amino acid in the protein sequence, and $\varepsilon(A)$ is the extinction coefficient of the amino acid.

Constructs		Affinity tag(s)	MW (kDa)	Extinction coefficient * (M^{-1} cm^{-1})	Theoretical pI
Individual shelterin subunits	TRF2 ^	dStrepII-TEV	63.60	59,930	9.30
	RAP1 ^^	dStrepII-TEV	48.27	40,910	4.69
Full shelterin complex	TRF1, TRF2. RAP1, **TIN2L**, TPP1, POT1	dStrepII-TEV-TIN2L	491.26	414,710	5.95
	TRF1, TRF2. RAP1, **TIN2S**, TPP1, POT1	dStrepII-TEV-TIN2S	480.68	410,240	6.01
	TRF1, TRF2. RAP1, **TIN2L**, TPP1, POT1 ^	dStrepII-TEV-TIN2L, His_{10}-TEV-TRF2	495.71	417,690	6.04
shelterin (–TRF1)	TRF2, RAP1, **TIN2L**, TPP1, POT1	dStrepII-TEV-TIN2L	390.81	330,830	5.94
	TRF2, RAP1, **TIN2S**, TPP1, POT1	dStrepII-TEV-TIN2S	380.23	326,360	6.01
	TRF2, RAP1, **TIN2S**, TPP1, POT1 ^	dStrepII-TEV-TIN2S, His_6-TEV-TRF2	484.03	413,220	6.05

Table 6.5. List of expression constructs.

The theoretical parameters were derived from ExPASy ProtParam by inputting the amino acid sequence of individual proteins or the combined sequences for protein complexes (Gasteiger, et al., 2005).

* at 280 nm measured in water, and assuming all Cys residues are reduced.

^ generated by Dr Oviya Inian

^^ generated by Anthony Marchand

6.3 Protein expression and purification

6.3.1 Bacmid generation for protein expression in insect cells

For an overview of the generation of baculoviruses for protein expression in insect cells, see Figure 6.2. For the expression of individual genes or complexes using insect cells, the plasmids were integrated into the baculovirus genome harboured by DH10^{MultiBacTurbo} *E. coli* competent cells via Tn7 transposition (Bieniossek, et al., 2008). These cells carry the baculovirus genome and the transposase required for Tn7 transposition. In addition, the DH10^{EMBacY} cells contain a constitutively expressing YFP expression cassette in its LoxP site, distal to the Tn7 attachment site, to enable monitoring of viral replication and protein expression via fluorescence. Colonies containing the composite bacmids were identified by (1) blue/white screening (successful Tn7 transposition prevents the expression of the lacZ-peptide, and colonies remain white on agar plates supplemented with X-gal) and (2) antibiotic resistance (inherited from the acceptor and donor plasmids). The bacmid DNA was then extracted and purified using alkaline lysis and isopropanol precipitation and used to transfect insect cells.

Sf9 insect cells were grown in the Insect-XPRESS medium (Lonza) and were used for both baculovirus production and protein expression. Insect cell transfection was performed in 6-well tissue culture plates. Briefly, 2 mL of log-phase Sf9 insect cells at 0.5 x 10^6 cells/mL were seeded per well and transfected with purified bacmids using the Cellfectin™ II Reagent (Invitrogen). After 96 hours of incubation at 27 °C, the supernatant containing the baculovirus was harvested (low titre P1 virus) and amplified by infecting 25 mL of log-phase Sf9 cells (at 0.5 x 10^6 cells/mL). Samples were taken from the cultures after 48, 72 and 96 hours post-infection and examined using the LUNA-FL Automated Fluorescence Cell Counter (LabTech) to check for YFP fluorescence, cell count and viability. After 3 to 4 days of incubation at 27 ºC shaking at 130 rpm, the culture was centrifuged at 300 xg for 10 mins to harvest the supernatant containing the P2 virus. On occasion, another round of amplification (P3) was necessary to raise the viral titre and improve the protein yield. The harvested viruses were supplemented with 4% v/v FBS and stored at 4 °C. The process from the generation of bacmids to

scaled-up protein expression typically took around 15 days or more, depending on whether the viral titre was sufficiently high at the P2 stage. Western blot analyses and small-scale test expressions were also performed to check protein expression. For protein expression in insect cells, Sf9 cultures of 200 mL – 500 mL were infected with 1:250 or 1:500 dilutions of baculovirus at ~ 0.8 to 1 x 10^6 cells/mL. The cultures were incubated in shaking incubators at 27 °C, 130 rpm for 3 to 4 days. Cells were harvested when cell viability dropped to between 80 – 90% or when %YFP fluorescent cells reached over 80%. Cell pellets were flash-frozen and stored at −80°C until use.

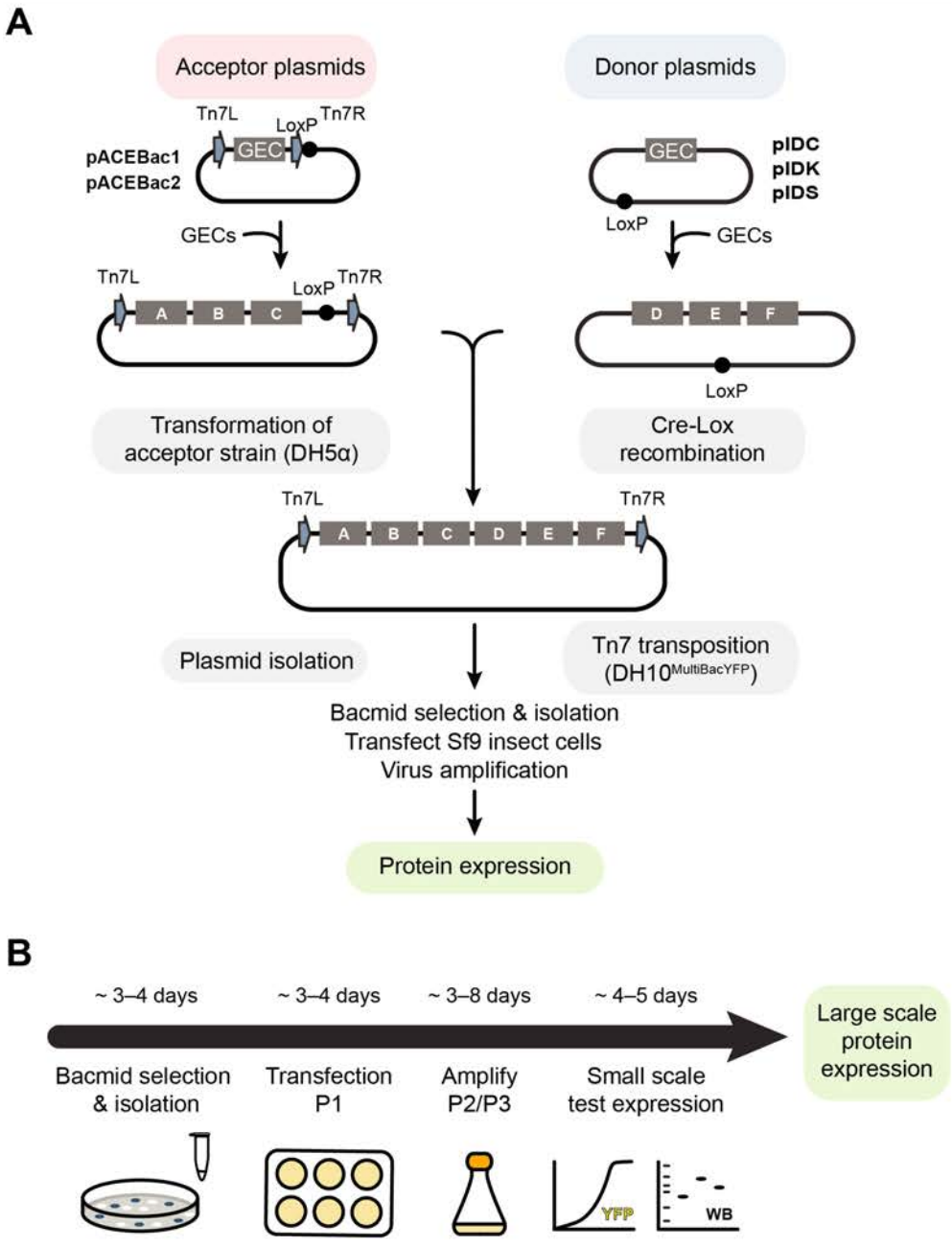

Figure 6.2. Overview of the baculovirus-insect cell expression system.

(A) GECs are inserted into acceptor and donor plasmids. Various combinations of acceptor-donor fusions can be generated via in vitro Cre-Lox recombination, which are identified using antibiotic selection of transformed *E. coli* cells. The combined plasmid is purified and integrated into baculovirus genome via Tn7 transposition in DH10^{MultiBacYFP} cells. Colonies containing the GEC-containing bacmids are selected by blue/white screening and antibiotic resistance (inherited from the acceptor, donor, or donor-acceptor combinations). The bacmid DNA is then extracted and used to transfect insect cells for protein production. Amplification of the baculovirus is usually required to improve the protein yield. **(B)** Typical timeline for protein expression using the baculovirus-insect cell system.

6.3.2 Overview of purification strategy

Generally, purifications of single proteins involved one affinity chromatography step, whilst complex protein assemblies required initial enrichment using affinity chromatography followed by further purification. Depending on the protein yield, the subsequent steps involved size exclusion chromatography (SEC) or further affinity-based purifications.

For all protein purifications, the first chromatography step was facilitated by a dStrepII affinity tag (see Table 6.5) appended to the N-terminus of the protein of interest. Affinity purifications were performed using 1-, 5-mL or two 5-mL columns in tandem. column or batch formats, where the surface-modified resin was prepacked into columns or used as loose resin, respectively.

Although single-step purifications involving affinity chromatography produces proteins with high purity, especially for single dStrepII-TEV fusion proteins, to produce homogeneous macromolecular protein complexes, the final purification step is commonly SEC. This further separates any subspecies that may have formed due to the presence of excess subunits, including the tagged protein. During SEC, separation of molecules depends on the interactions between the sample and the resin. This is affected by the hydrodynamic radii of the molecules, which in turn are dependent on size, shape, and other biophysical and biochemical properties. The smaller the hydrodynamic radii, the more readily the particles penetrate the pores in the resin, and therefore the greater is the retention time in the column. Different SEC columns are packed with various resins of defined pore size suitable for resolving a range of molecular weights (MWs).

The conventional calibration method uses a selection of protein standards across a MW range. The calibration curve of a particular SEC column shows the relationship between retention volume of a set of protein or polymer standards against their MW, which can be used to estimate MW of unknown samples. The exclusion limit for a particular SEC column gives an indication for the size of molecules that are excluded by the resin pores, represented by molecules that

elute in the void volume. On the opposite spectrum are molecules that are small enough to permeate all the pores, which can be used to determine the total permeation limit of the column. Separation of molecules with the desired molecular weight range should lie comfortably within these two limits, i.e., within the linear region of a calibration curve. An extended linear range indicating that a wider range of molecular weights can be resolved with a particular column.

However, not all macromolecules behave the same. A comparison between the calibration curves of protein with those of polymers (e.g., polysaccharides and PEG) reinforces the fact that separation during SEC, as stated in the name, depends on sizes of the molecules rather than their molecular weight, which is usually the only parameter provided for SEC calibration standards. Therefore, column calibration with SEC standards of compact globular proteins is inappropriate for MW measurements for elongated molecules, such as partially disordered proteins or multi-subunit protein complexes with extended tertiary structures. The absolute MW of proteins and protein complexes can be determined from other methods, including analytical SEC coupled to multi-angle light scattering (MALS) (see section 6.5.1).

6.3.3 Expression and purification of shelterin complexes

The full human shelterin^{TIN2L} complex and the shelterin –TRF1^{TIN2L} subcomplex were expressed in 400 to 500 mL Sf9 cultures in 2 L flasks, infected at 1 x 10^6 cells/ml with 1 in 250 or 1 in 500 dilution of a P3 baculovirus. All steps were performed at 4 ^oC, unless otherwise stated. The cell pellet was lysed with lysis buffer (50 mM HEPES pH 8.0, 1 M NaCl, 10 mM BME, 2 mM $MgCl_2$, benzonase (abcam), and EDTA-free protease inhibitors (Roche)) and sonicated using a Vibra-Cell sonicator (Sonics & Materials) with 3 second on/off cycles at 20% amplitude for 3 mins. The total lysate was cleared by centrifugation at 15,000 x g and filtered using a 0.45-μm filter.

The same two- or three-step purification protocol was used for both the full shelterin complex and subcomplexes. In both protocols, the first step was Strep-tag affinity purification. The filtered lysate was loaded onto a pre-equilibrated StrepTrap HP column (one or two 5-mL columns in tandem, Cytiva) at 1 mL/min using a peristaltic pump. Then, the column was transferred to an ÄKTA Pure protein purification system (Cytiva), and washed with 5 – 10 column volumes (CV) of wash buffer A (50 mM HEPES pH 8.0, 1 M NaCl, 5 mM BME) and 5 CV of wash buffer B (50 mM HEPES pH 8.0, 0.5 M NaCl, 1 mM TCEP, 10% glycerol) at 1 mL/min. The captured proteins were eluted with 5 CV of elution buffer (50 mM HEPES pH 8.0, 0.5 M NaCl, 1 mM TCEP, 10% glycerol, 5 mM desthiobiotin) at 0.3 mL/min. The elution was monitored by the change in UV absorption at 280 nm and 260 nm using the UNICORN 7 software (Cytiva). In the two-step protocol, the affinity-purified proteins were purified by SEC using a Superose 6 Increase 10/300 GL column (Cytiva) pre-equilibrated with SEC buffer (50 mM HEPES pH 8.0, 0.5 M NaCl, 2 mM TCEP, 10% glycerol). Up to 0.5 mL of affinity-purified protein were injected and eluted at 0.1 – 0.3 mL/min in 200 µL fractions. In the three-step protocol, the affinity-purified shelterin complexes were subjected to heparin chromatography prior to SEC purification. For the heparin purification, the affinity-purified protein was dialysed against 50 mM HEPES pH 8.0, 0.3 mM NaCl, 2 mM TCEP, and then applied to a pre-equilibrated HiTrap Heparin HP column (1 or 5 mL, Cytiva) at 0.5 mL/min. The column was washed with 3 CV of wash buffer (50 mM HEPES pH 8.0, 0.3 M NaCl, 2 mM TCEP) and eluted with a linear or step gradient from 0.3 to 2 M NaCl in a buffer also containing 50 mM HEPES pH and 2 mM TCEP. The fractions were analysed with SDS-PAGE, and fractions containing the proteins of interest were pooled, concentrated, and subjected to SEC purification as described above.

During protocol optimisations, proteins from Strep-tag affinity elution were also purified using ion exchange chromatography (IEX) and ion metal affinity chromatography (IMAC). For IEX, the affinity-purified protein was dialysed against 50 mM HEPES (pH 8.0 or pH 7.0), 0.2 or 0.3 mM NaCl, 2 mM TCEP, and then applied to pre-equilibrated columns (HiTrap SP FF, Mono S 5/50 GL, or HiTrap Q HP, Cytiva) at 0.5 mL/min. The column was washed with 3 CV of wash buffer (50 mM HEPES pH 8.0, 0.2 or 0.3 M NaCl, 2 mM TCEP) and eluted with a

linear or step gradient from 0.2 or 0.3 M to 2 M NaCl in a buffer also containing 50 mM HEPES pH and 2 mM TCEP. For IMAC, the affinity-purified protein was loaded directly onto a HisTrap HP column (Cytiva), washed with 3 CV of wash buffer (50 mM HEPES pH 8.0, 0.3 M NaCl, 2 mM TCEP) and eluted with a linear or step gradient using 20 to 350 mM imidazole (pH 8) in a buffer also containing 50 mM HEPES pH 8.0, 0.3 M NaCl and 2 mM TCEP.

6.3.4 Expression and purification of individual shelterin subunits

Recombinant TRF2 and RAP1 proteins were produced individually as dStrepII-TEV fusion proteins with the same two-step protocol, namely Strep-tag affinity purification and SEC, using the same buffers and columns as described above (section 6.3.3).

6.4 Biochemical assays

6.4.1 Electrophoretic mobility shift assay

The shelterin complexes used for electrophoretic mobility shift assay (EMSA) experiments were expressed and purified as previously described (see section 6.3.3). For the protein titration experiments, the gel shift reaction mixture contained 10 µL of affinity-purified proteins from 0.25 to 10-fold molar excess of protein relative to DNA, incubated with a final concentration of 0.5 µM of a telomeric oligonucleotide containing both a dsDNA region and a ssDNA 3' overhang (see Table 6.6 for the list of oligonucleotides). The DNA binding reactions were performed in 50 mM HEPES pH 8.0, 250 mM NaCl, 10% glycerol, 1 mM TCEP, and incubated for 30 mins on ice, and then loaded on 0.5X TBE 1.2 % agarose gels and run in 0.5X TBE buffer for 1 hour at 100 V at 4 ˚C. For

checking DNA binding in SEC-purified shelterin complexes after crosslinking, 10 µL of the SEC elution fractions were loaded onto either 0.5X TBE 1.2 % agarose gel or 1X TBE 4 – 12% precast polyacrylamide gels (Invitrogen) and run in 0.5X TBE buffer for 1 hour or 1.5 hours at 100 V at 4˚C. The gels were rinsed with water, stained with ethidium bromide for 15 mins, and washed with water twice before scanning using the ChemiDoc Imaging System (BioRad).

Name	Sequence
teloDNA1	5'-CAGGTACCCGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAG 3'-GTCCATGGGCCCAATCCCAATCCCAATCCCAATCCCAATC
teloDNA2	5'-CAGGTACCCGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAG 3'-GTCCATGGGCCCAATCCCAATCCCAATC

Table 6.6. List of telomeric oligonucleotides used for DNA-binding studies and EM.

6.5 Biophysical assays

6.5.1 Size exclusion chromatography coupled to multi-angle light scattering

6.5.1.1 Background

Analytical SEC can be used in tandem with MALS for absolute MW determination. The amount of light scattered by a molecule is directly proportional to mass, concentration, and the refractive index (RI) of the sample, and the RI of the solvent. Unlike SEC, SEC-MALS calculates MW from first principles (based on the Rayleigh-Debye-Gans light scattering (LS) model, see equation 1), thus is independent of elution volume. However, resolution of the various species in the injected sample is still important to allow individual species to enter the MALS and concentration detectors for accurate MW determination.

$$MW = \frac{R(0)}{Kc\left(\frac{dn}{dc}\right)^2} \qquad \text{equation 1}$$

where MW is the weight-average molecular weight (molar mass) of the particle, R(0) is the Rayleigh ratio R_θ (i.e., the ratio of scattered light to the incident laser intensity; see equation 2) measured by the MALS detector and extrapolated to angle, θ equals zero, *c* is the sample concentration determined by the UV or dRI detector (the latter is preferable, especially for molecules without a UV chromophore such as proteins lacking the aromatic residues tyrosine and tryptophan responsible for UV absorption at around 280 nm), *n* is the solvent refractive index, and *(dn/dc)* is the refractive index increment (i.e., the difference between the refractive index of the particle and the solvent; for pure proteins in water this value is very close to 0.185 mL/g), and *K* is a system constant.

$$R_\theta = \frac{I_\theta r^2}{I_0 V} \qquad \text{equation 2}$$

where I_θ is the intensity of the scattered light at angle θ; I_0 is the incident light intensity, and *r* is the distance from the scattering volume to the detector; V is the illuminated scattering volume from which the scattered light reaches the detector at angle θ.

The angles used in a MALS setup are fixed and LS is detected simultaneously at all angles. One of the main advantages of measuring LS as a function of scattering angle is the possibility of determining the root mean squared (RMS) radius, also referred to as the radius of gyration (R_g). This provides information on the dimensions of the molecules. However, most molecules subjected to SEC-MALS are relatively small, hence the intensity of the LS signal does not vary much for different angles (i.e., the LS profile is isotropic). In these instances, R_g cannot be measured. When LS varies with the scattering angle (i.e., there is an angular dependence of LS), the scattering profile is said to be anisotropic. Regardless of the scattering profile, MW can be calculated since it is

derived from extrapolation to zero angle from the multi-angle measurements (R(0) in equation 1).

Another useful calculation is the degree of polydispersity of the sample. Polymers and polymerising proteins are formed by repeating units of varying lengths, i.e., they are polydisperse. Therefore, the MW of polymers is an average value that is representative of the distribution of chain lengths and MWs existing in the sample. For protein macromolecules, commonly used MW averages are the number average and weight average MWs (M_n and M_w, respectively). The former is simply the statistical average MW of all the polymer chains (see equation 3). The latter is sensitive to the length of the polymer rather than the number alone; the larger the polymer chain, the more it contributes to the MW average (denoted by the squared variable M_i in equation 4). They are defined by:

$$\boldsymbol{M_n} = \frac{\sum \boldsymbol{N_i M_i}}{\sum \boldsymbol{N_i}} \qquad \text{equation 3}$$

$$\boldsymbol{M_w} = \frac{\sum \boldsymbol{N_i M_i^2}}{\sum \boldsymbol{N_i}} \qquad \text{equation 4}$$

where M_i is the MW of a particular chain and N_i is the number of chains with that particular MW.

The polydispersity index (Đ) is a ratio of M_w and M_n and is a measure of the MW distribution of a polymer (see equation 5). Monodisperse samples with identical chain lengths (such as non-polymerising proteins) have a polydispersity index of 1 (i.e., $M_n = M_w$). Deviation from a M_w/M_n value of 1 indicates differences in lengths in the sample, with larger M_w/M_n values representing a broader MW distribution within a sample.

$$Đ = \frac{\boldsymbol{M_w}}{\boldsymbol{M_n}} \qquad \text{equation 5}$$

6.5.1.2 SEC-MALS data acquisition and analysis

Shelterin complexes were resolved using SEC columns connected to an HPLC system (Agilent Infinity II LC System). LS and dRI of the samples were measured in-line using a MALS detector (DAWN®, Wyatt Technologies) and a dRI detector (Optilab®, Wyatt Technologies), which related information on sample concentration, solvent absolute RI and sample refractive increments (dn/dc), all of which are necessary for MW calculations.

For SEC-MALS analyses of the shelterin^{TIN2L} complex, 20 µL of 5 µM affinity-purified protein was injected onto a Wyatt 050-N5 SEC column (Wyatt, #WTC-050N5) pre-equilibrated with buffer (50 mM HEPES pH 8.0, 0.5 M NaCl, 1 mM TCEP) and eluted at 0.2 ml/min. For SEC-MALS analyses of the shelterin (–TRF1)^{TIN2L} complex, 100 µL of 5 µM affinity-purified protein was injected onto a Superose 6 Increase 10/300 GL column (Cytiva) pre-equilibrated with buffer (50 mM HEPES pH 8.0, 0.5 M NaCl, 1 mM TCEP) and eluted at 0.1 or 0.5 ml/min (see figure legends for the exact flow rate). For the analytical SEC analysis of shelterin (–TRF1)^{TIN2L} shown in Figure 3.3, 20 µL of serial dilutions of affinity-purified protein (0.625 µM, 1.25 µM, 2.5 µM, 5 µM) were injected sequentially onto a pre-equilibrated Wyatt 050-N5 SEC column (Wyatt Technologies) and eluted at 0.2 ml/min, starting from the lowest protein concentration. Injections were separated with a 3 CV SEC buffer run to re-equilibrate the column and prevent contamination of subsequent runs.

Data on light scattering and dRI were collated in and analysed using the ASTRA 7 software (Wyatt Technologies) to calculate weight average MW (M_w), number average MW (M_n) and the polydispersity index (Đ).

6.5.2 Dynamic light scattering

6.5.2.1 Background

Dynamic light scattering (DLS) is a technique for investigating how particles diffuse within a solution, from which the size of particles (i.e., the hydrodynamic radius) can be calculated. DLS records the Brownian motion of particles as they interact with surrounding solvent molecules. The larger the particle, the larger the hydrodynamic radius, the slower the magnitude of the Brownian motion. The velocity of the Brownian motion is defined by the translational diffusion coefficient, D, which is related to the hydrodynamic radius, R_H of the particles. This relationship is defined by the Stokes-Einstein equation,

$$D = \frac{k_B T}{6\pi\eta R_H} \qquad \text{equation 6}$$

where k_B is the Boltzmann constant, T is the temperature, and η is viscosity of the solution. Temperature, T, is related to the viscosity, η, of the solution in which the measurements are made. Unstable temperature control can lead to convection currents in the sample and induce motion that will overcast the Brownian motion required for size interpretation.

Two other important factors that affect the diffusion of particles are the surface structure of a particle and whether it is globular or has an extended conformation. A particle with a smooth surface will have a smaller hydrodynamic radius compared with one with the same molecular weight but has flexible protrusions. Importantly, the equation assumes a spherical particle, which is the only volume that can be described with a single hydrodynamic radius. Therefore, the hydrodynamic radius of an extended particle calculated as above is the radius of a spherical particle that has the same translational diffusion speed. Moreover, for extended particles, changes in the length dimension will have a larger impact on diffusion speed, and hence hydrodynamic radius calculations, whereas changes in the width (diameter) will have a smaller impact and are harder to detect.

6.5.2.2 DLS data acquisition and analysis

Dynamic light scattering measurements were acquired using a SpectroLight 600 (Xtal Concepts GmbH). The laser used an optical power of 100 mW and wavelength of 658 nm, and the detector was located at a scattering angle of 142º. The refractive index of water (n = 1.33) and a viscosity of 1.006 cP were used for all calculations. Two μL of each sample were pipetted in T well plate and immersed under oil. All measurements were taken at 18 ºC, and 15 repeats were performed for each sample to give precise results.

6.5.3 Mass photometry

6.5.3.1 Background

Mass photometry is a label-free interferometric scattering microscopy (iSCAT) technique to determine absolute molecular mass by quantifying light scattering from single particles in solution with high accuracy (Young, et al., 2018; Young & Kukura, 2019). It was built on the principles of interference reflection microscopy (Verschueren, 1985) and interferometric scattering microscopy (Ortega-Arroyo & Kukura, 2012).

In dark-field microscopy, the incident light is focused onto the sample on the glass slide by condenser lenses. The scattered light from the sample is further focused by the objective and ocular lenses before reaching a detector. To improve the signal-to-noise ratio, dark-field microscopy minimises the background noise by filtering out the unscattered incident light from the signal (Figure 6.3A). This produces an image with a dark background with bright scattering objects. In general, the intensity registered by the detector, I_{det}, is dependent on the amplitudes of the background electric field, E_b, the scattered electric field, E_s, and the phase difference between the reflected and scattered light field, φ,

$$I_{det} = |E_b + E_s|^2 = |E_i|^2 \left[b^2 + |s|^2 - 2b\,|s|\,\sin(\varphi)\right] \qquad \text{equation 7}$$

where E_i is the electric field of incident light used to illuminate the sample, and *s* is the amplitude of the scattered light.

Thus, the three different contributions to the detected intensity are (1) the background intensity, $|E_i|^2 b^2$, (2) the pure scattering signal, $|E_i|^2\,|s|^2$, and (3) the interference term $2b\,|s|\sin(\varphi)$ (Ortega-Arroyo & Kukura, 2012). Since dark-field microscopy minimises the background field, the pure scattering signal is the dominating signal at the detector (Ortega-Arroyo & Kukura, 2012).

A typical iSCAT setup involves a laser as the incident light source, which is focused onto the sample on a glass slide by a high-numerical-aperture objective (Figure 6.3B). Unlike dark-field microscopy, iSCAT does not exclude all background light, and instead collects the light reflected from the glass/solution interface as the reference field, E_r. The detected intensity, I_{det}, is now given by,

$$I_{det} = |E_r + E_s|^2 = |E_i|^2 \left[r^2 + |s|^2 - 2r\,|s|\,\sin(\varphi)\right] \qquad \text{equation 8}$$

where r^2 is the reflectivity of the glass-solution interface. At the glass/water interface, r ~ 0.065, which is much greater than the scattering signal $|s|$ for particles with D < 50 nm (Ortega-Arroyo & Kukura, 2012). Therefore, for samples with small scatterers, the reference contribution is the dominant signal at the detector, with a small contribution from the light scattered by the sample.

According to Mie theory (Mie, 1908), the scattering amplitude of small scatterers can be represented as

$$s \sim \varepsilon_m\,\pi\,\frac{D^3}{2}\frac{\varepsilon_p - \varepsilon_m}{\varepsilon_p + 2\varepsilon_m} \qquad \text{equation 9}$$

where ε_p and ε_m are the dielectric constants of the particle and its surrounding medium, respectively, and D is the diameter of a spherical particle. This indicates that the scattering signal scales with the polarisability of the scatterer and is

therefore a function of refractive index and particle diameter/volume. According to this, the pure scattering term $|s|^2$ scales as D^6, and the square of the particle volume. This means a particle with half its diameter will only generate $(1/2)^6$ = 1/64th of the signal (Ortega-Arroyo & Kukura, 2012). Therefore, for small scatterers, this scattering signal drops significantly and rapidly approaches zero. Moreover, most organic matter has a dielectric constant that is very similar to the surrounding aqueous solution, so the pure scattering term $|s|^2$ also becomes negligible for weak scatterers. This leads to one important advantage of iSCAT operated in reflection over other label-free dark-field microscopy approaches, which lies in its improved detection of weakly scattering objects (Ortega-Arroyo & Kukura, 2012). In iSCAT, the interference term that is used to generate contrast scales linearly with the scattering amplitude and thus the volume of the scattering object. Now, reducing the diameter of the scattered by half would only cause a 1/8th drop in the detected signal.

The interferometric scattering contrast, c, can be defined by separating the light intensity in the presence $I_{particle}$ and absence ($I_{background}$) of a scatterer,

$$c = \frac{I_{particle}}{I_{background}} = 1 + \frac{2\,|s|\cos(\varphi)}{r} \qquad \text{equation 10}$$

In the mass photometer, a partial reflector is placed in the centre of the beam path to increase the measured contrast. It does so by attenuating the reflected light travelling in the centre of the beam (thus reducing $I_{background}$), and allows most of the scattered light to pass through (that contribute to $I_{particle}$. Finally, a background subtraction step is performed during image analysis. Consecutive frames of the raw images can be used to remove background scattering from the glass-buffer interface (Soltermann, 2020). Now, the single binding events are registered as black circles that appear and disappear with time. Their intensity corresponding to their ratiometric contrast value, and their maximum contrast is later converted into molecular mass after instrument calibration with known protein standards.

In the mass photometer setup (Figure 6.4A), the incident laser light travels through the same objective lens as the reflected and scattered light used for imaging. The majority of the incident light travels past the transparent glass slide and interacts with the particles within the sample. Some of the incident light becomes scattered by the sample in all directions. A small portion of the incident light is reflected at the glass interface, which is merged with some of the light scattered by the sample as they travel back through the objective lens. The merged light sources constructively interfere, which is used to generate an image. Improvements in signal detection (hardware) combined with streamlined data analysis pipelines (software) has enabled high signal-to-noise ratio (SNR) imaging and accurate determination of the interferometric contrast generated by single binding events. This approach is used in mass photometry as an absolute measure of molecular mass with a high mass accuracy. By registering many single landing events, a distribution of contrasts could be constructed from the single counts.

Mass photometry does not require labelling of samples since the individual particles in the samples generate contrast as they land on the glass slide. This interferometric contrast is generated by the difference in refractive index between a particle and its surrounding medium. When measuring protein samples, assuming all proteins share similar optical properties and densities, the interferometric contrast is proportional to the molecular mass of the protein. Proteins with known masses can then be used to calibrate the instrument (Figure 6.4B). Resolution is defined by full width at half-maximum (FWHM), which is dependent on the width of the distribution arising from a single species. This is related to the precision with which the contrast of each landing particle can be measured. The higher the precision with which the centre of the distribution could be determined, the narrower the distribution, and the higher the resolution. This parameter is particularly important for multi-subunit complexes and more complicated biomolecular samples where clear separation of different species is essential for accurate measurement of molecular mass of the various species that may exist in the sample.

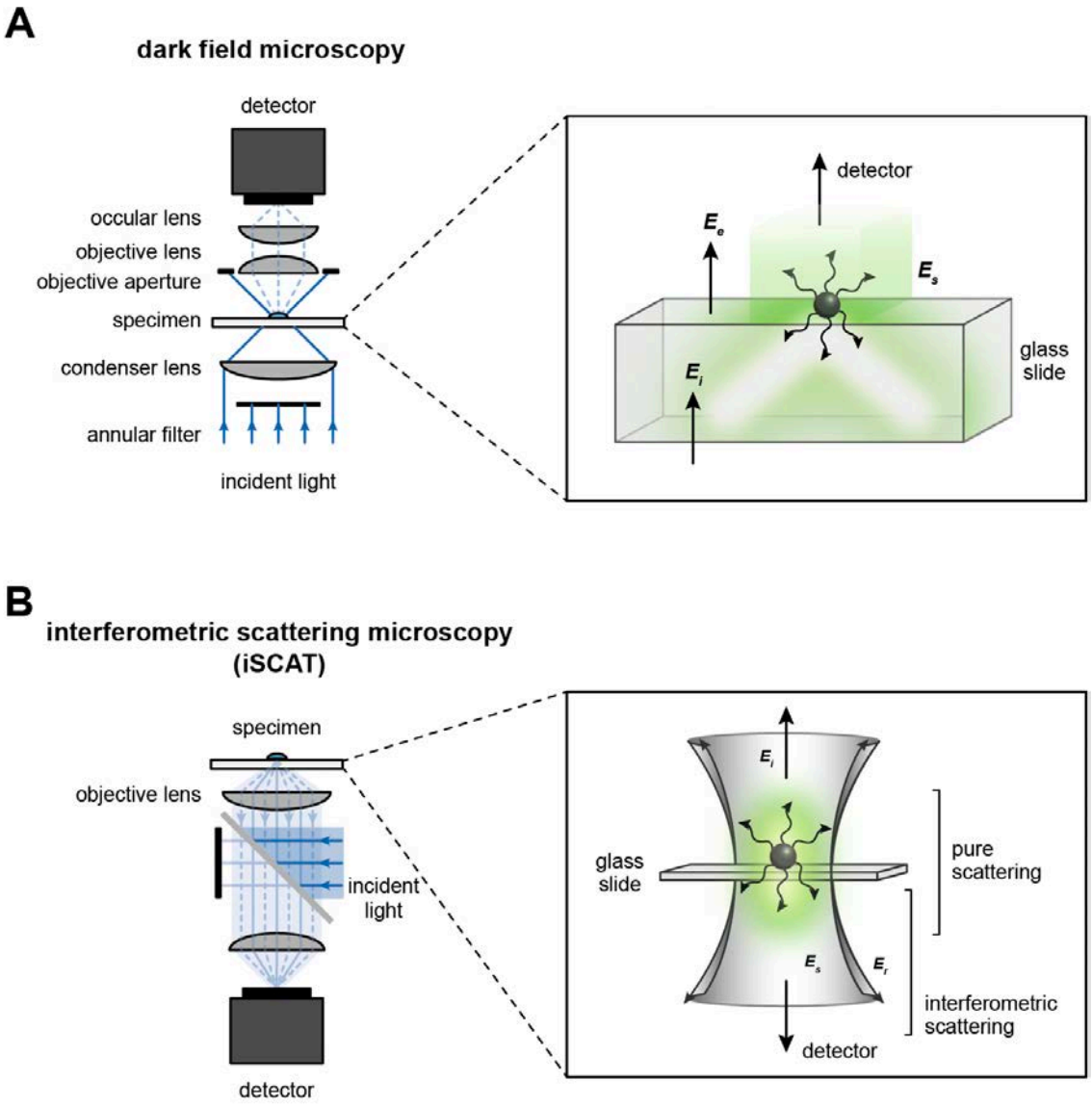

Figure 6.3. Detection geometries for dark-field and scattering microscopy.

(A) Schematic of a dark-field microscope and the principle of dark-field detection, where the incident light is separated from scattered light by total internal reflection. **(B)** Schematic of a dark-field microscope and the principle of iSCAT detection, where light reflected at the glass/solution interface and any scattered light from the sample travel back through the objective lens and focused onto the detector. E_i, incident electric field; E_s, scattered electric field; E_r, reflected or reference electric field. Adapted from (Ortega Arroyo, et al., 2016; Young & Kukura, 2019)

Being label- and immobilisation-free greatly simplifies the experimental procedure and reduces the amount of starting material required. Moreover, proteins samples can be measured directly in their native buffers, assuming the

native buffer does not generate significant background noise. The technique also requires very little sample (pmols per measurement) since it can detect individual binding events. Several parameters can have a major influence the relative abundances of the species counted during measurements. One of the main type of samples that remains challenging for mass photometry are complexes where weak interactions dominate. Weak interactions are difficult to characterise because of their fast off-rates, which can easily fall apart at low concentrations required for accurate mass measurements in mass photometry. In addition, when the molecular masses of individual components within the complex are small (less than the FWHM), they may not be well resolved and Gaussian fitting during image analysis may only produce an average of different overlapping species.

6.5.3.2 Mass photometry data acquisition and analysis

Mass photometry experiments were performed on a Refeyn One^{MP} instrument (Refeyn Ltd). Coverslips and gaskets were cleaned by washing with sterile water, 100% isopropanol, sterile water, and air dried with N_2 gas. A mass calibration was performed using a set of native protein standards (Invitrogen, #LC0725) diluted in filtered mass photometry buffer (20 mM HEPES pH 8.0, 0.3 mM NaCl) before each experiment. Shelterin complexes were diluted to 100 nM using the mass photometry buffer and added immediately to a 15-µL buffer drop and mixed by pipetting up and down. To crosslink shelterin complexes for mass photometry analysis, protein complexes were diluted to 1 µM and crosslinked with 0.075 – 0.1% glutaraldehyde on ice for 30 mins. The crosslinking reaction was quenched with 50 mM Tris pH 8.0, and samples were then diluted to 100 nM prior data acquisition.

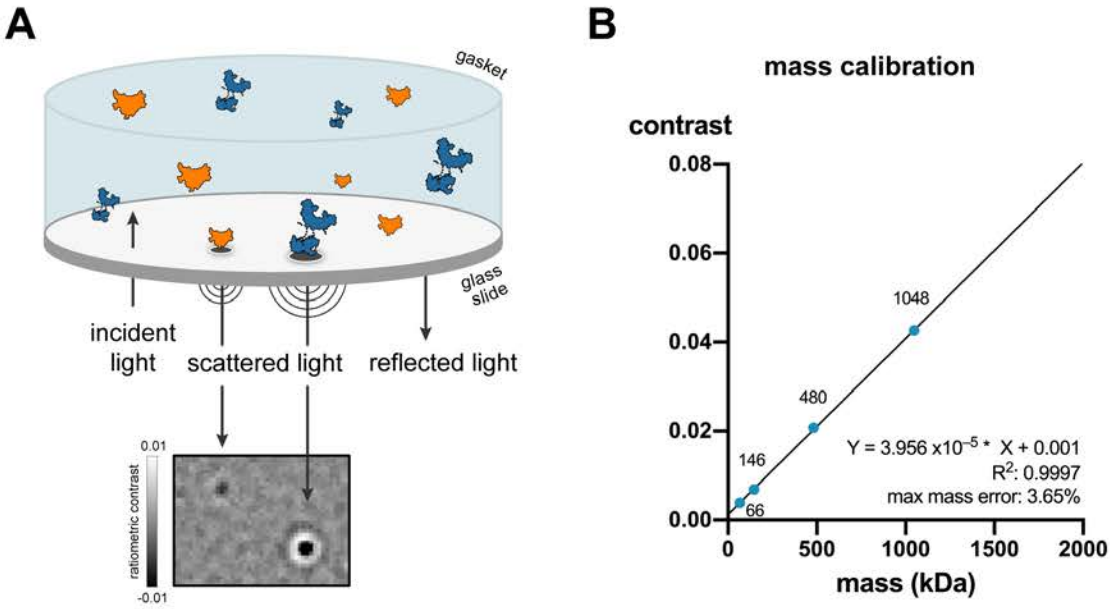

Figure 6.4. Principles of mass photometry.

(A) Schematic of the mass photometry experimental set-up. The incident light is focused onto the sample on the glass slide, and light reflected at the glass interface and the light scattered by the particles in the sample are collected. Figure inspired by (Soltermann, et al., 2020). **(B)** Calibration curve. The calibration curve plotting the measured peak contrast against the expected mass of the proteins in the calibration standard (blue dots). A linear line can be fitted to the data with a correlation coefficient R^2 of close to 1. The mass error (%) was determined from measuring the deviation of the predicted from the expected mass for each calibrant, taking the maximum error value as the maximum mass error (%).

Movies were collected for 10,000 or 20,000 frames in the largest detection area using the AcquireMP software (Refeyn Ltd) and analysed using the DiscoverMP software (Refeyn Ltd). The outputs include a list of measured contrast for individual particles, which were plotted as a histogram to view the distribution of species in the sample. The measured contrasts were converted to MW values using the calibration factor from the mass calibration. From the native protein markers, up to 4 separate proteins can be routinely identified (66, 146, 480 and 1048 kDa, Figure 6.4B). As the mass photometry signal scales linearly with mass, the measured contrast and corresponding molecular masses can be plotted and fitted to a straight line, y = mx, where y is the mean peak contrast, x is the molecular mass of the sample, and m is the calibration factor (Figure 6.4B). The errors were calculated from the residuals of measured contrast and

calibration curve values for each calibrant. The maximum error (%) is the maximum deviation of the measured molecular mass from the expected molecular mass. The maximum errors associated with each experiment are listed in the associated figure legend. The data were visualised as histograms with specified bandwidth to determine the mean mass for each peak using Gaussian fitting in the DiscoverMP software. In addition, kernel density estimates (KDE) were generated for each sample using a Gaussian kernel to visualise the data.

6.5.4 Real-time binding kinetics with switchSENSE

6.5.4.1 Background

SwitchSENSE® is a chip-based technology that uses one- or two-colour fluorescence sensing and molecular dynamics measurement to provide a wide range of biophysical information such as binding kinetics, binding-induced protein conformational changes and thermodynamics (Rant, et al., 2004; Langer, et al., 2013). SwitchSENSE experiments can be conducted in two measurement modes, a static and dynamic mode. In the static fluorescence proximity sensing (FPS) mode, the electrodes are kept at a constant negative potential to keep the DNA nanolever at a fixed angle to the surface, and only the fluorophore emission intensity is measured. The fluorescence intensity of the dye in the FPS mode is influenced by two variables, (1) the distance between the fluorophore and the gold surface (as mentioned above), and (2) quenching or enhancement of the fluorescence intensity of the dye due to the presence of other molecules bound in close proximity to the dye. The dynamic mode is where an alternating potential is applied to the gold electrodes to attract and repel the DNA nanolever, and thus switch its orientation.

DNA is a highly charged polyanion, and its structure in solution is strongly dependent on ionic conditions (including ion valency and concentration) and temperature. In a buffer with a low concentration of monovalent ions, such as Na^+,

DNA molecules can be approximated to a stiff rod as bending is electrostatically unfavourable because of the strong charge repulsion on the backbone. With increasing Na^+ concentration, the electrostatic repulsion is reduced and thus bending becomes more favourable.

At the vicinity of the electrically biased metal surface, the range of the effect of the electric force is dependent on several factors. Important parameters include temperature, and the ion valency and concentration in the buffer. Free ions in the salt-containing buffer reorganise around at the charged metal surface to create a balancing counter charge in the solution, which has a screening effect. At flat surfaces, the diffusive potential distribution Φ on the buffer side can be represented by the Gouy-Chapman equation,

$$\Phi(d) = \frac{2kT}{e} In\left(\frac{1+\gamma \exp\left(1d/l_D\right)}{1-\gamma \exp\left(1d/l_D\right)}\right), \gamma = tanh\left(\frac{e\Phi_0}{4k_BT}\right) \quad \text{equation 11}$$

where the Debye length, l_D is the screening length that varies with the salt concentration in the buffer, d is the distance to the charged surface, Φ_0 is the surface potential, and n is the ion density.

This screen of oppositely charged ions at the charged metal/solution interface becomes more disperse with increasing distance from the metal surface. The result of this screening effect is that the potential is short-ranged (within the first few nanometres from the surface), beyond which Brownian motion dominates. At low to moderate salt concentrations (10 – 100 mM NaCl), this screening layer emanates from the metal surface for a few (1-3) nanometres. This covers the first few base pairs of the DNA nanolever and allows for switching the nanolever orientation through alternating the metal surface potential. At higher salt concentrations (> 140 mM NaCl), the screening layer is sub-nanometre, and switching the surface potential is ineffective at attracting and repelling the DNA nanolever. Therefore, for measurements using the dynamic mode, it is recommended to use buffers with salt concentration of less than 300 mM, where fluorescence measurement modes are less affected and can tolerate up to 3M.

The binding theory that underlies the affinity and kinetic assays assumes two species, an analyte A and a ligand B interact in solution to form a product, AB. Association occurs when the analyte encounters the ligand with the correct orientation and sufficient energy, and dissociation occurs when the analyte and ligand separate from each other. This model also assumes that after dissociation, the two reactants are not modified by this interaction.

$$A + B \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} A\,B \qquad \text{equation 12}$$

At equilibrium, the rate at which new AB complexes are formed (given by the forward reaction rate constant, k_{on}) equals the rate at which they decay (represented by the reverse reaction rate constant, k_{off}).

$$[A][B] \times k_{on} = [A \cdot B] \times k_{off} \qquad \text{equation 13}$$

The above equation can be rearranged to give the dissociation constant at equilibrium, (K_D),

$$K_D = \frac{[A][B]}{[A \cdot B]} = \frac{k_{off}}{k_{on}} \qquad units: M = \frac{s^{-1}}{M^{-1}\,s^{-1}} \qquad \text{equation 14}$$

An important distinction can be made between the concentration of ligands in solution and surface-tethered ligands (n). In solution, $[A \cdot B]$ depends on both $[A]$ and $[B]$. However, for surface-tethered ligands the total number of ligands immobilised on the electrode surface, $n_{B,0}$, is constant, and hence only the concentration of the analyte is varied between runs.

$$n_{B,0} = n_B + n_{AB} = constant \qquad \text{equation 15}$$

where n_B is the number of free ligands on the surface, and n_{AB} is the number of ligands are bound to analyte molecules. Between these two scenarios, $[B]$ is comparable to n_B, and $[A \cdot B]$ is the equivalent of n_{AB}. The fraction of binding sites that are occupied by the analyte ($n_{AB}/n_{B,0}$) is proportional to the readout

signal. When $[A]$ is equal to K_D, approximately 50% of the binding sites are occupied by the analyte at equilibrium, $f_{eq}(c_A = K_D) = 0.5$. Although the fraction bound depends on $[A]$, the values of k_{on} and k_{off} are constant rate values that are independent of $[A]$. Taking association and dissociation measurements from a range of analyte concentrations around the K_D value and fitting all the curves of one run simultaneously (global fit) will generate kinetic rate constants with their fit errors. An estimation of the k_{on} and k_{off} values was provided in the switchBUILD software (Dynamic Biosensors GmbH) whilst constructing the method to give an idea of the range of analyte concentrations to try during assay optimisation.

The DRX^2 system (Dynamic Biosensors GmbH) is a dual-colour electro-optical instrument designed for automated switchSENSE experiments. A standard multi-purpose biochip contains four flow channels, each containing six detection electrodes in series. Each detection electrode is a gold sensor spot of ~ 0.01 mm^2, functionalised with 1:1 ratio of two different ssDNA nanolevers at approximately 50 nm separation to avoid crowding effects. The proximal end of the nanolever is immobilised onto the gold electrode surface, whilst the distal end carries the fluorophore. Dual-colour biochips contain ssDNA with either a green or a red fluorophore. One nanolever can be used to perform measurements on the target with the other one designated as the control. This configuration also makes bivalent and bispecific antibody design possible, since two targets can be present in the same model system. The ssDNA nanolevers come in two different lengths (48-mer or 96-mer ssDNA) of defined sequence. For experiments studying protein-DNA interactions, the control and target oligonucleotides can be synthesised with DNA complementary to the nanolever, and thus can be annealed to the nanolever during hybridisation.

The DNA nanolevers on the gold surface can be induced to take an up or down conformation by reversing the electrode charge. This switching process also alters the fluorescence intensity of the fluorophore. A positively charged surface attracts the DNA nanolever and quenches the fluorescence of the fluorophore due to the nonradiative energy transfer from the excited dye to the

gold metal. Conversely, a negatively charged surface repels the DNA nanolever, and at the upright position the fluorescence intensity is at the maximum as the fluorophore is furthest away from the gold surface.

6.5.4.2 SwitchSENSE data acquisition and analysis

For the switchSENSE experiments, the full and (–TRF1) shelterin^{TIN2L} complexes were purified using the two-step protocol as described in section 6.3.3. SwitchSENSE experiments were performed using the DRX^2 system in the static FPS mode, with assistance from Dr Stephen Hallett and Dr Anthony Oliver from University of Sussex. Methods specifying the experimental parameters were constructed using the switchBUILD software (Dynamic Biosensors GmbH) and executed using the switchCONTROL software (Dynamic Biosensors GmbH). The temperature of the autosampler and the biochip was maintained at 10 and 20 ºC, respectively. The telomeric and non-telomeric oligonucleotides for switchSENSE experiments were produced by annealing the reverse strand of teloDNA1 and mutDNA1 with the corresponding forward strand, which carried a 3' extension that is complementary either the red or green fluorescently labelled nanolever, respectively (Table 6.7). Immobilisation of the 52/40-mer non-telomeric and telomeric oligonucleotides was achieved by direct coupling to the complementary DNA extending from the NL-A48 and NL-B48 nanolevers on the biochip (MPC2-48-2-G1R1-S, Dynamic Biosensors GmbH), respectively, in the HE40 functionalisation buffer (Table 6.8).

Passivation was performed prior to starting an experiment. This step covered the surface of the biochip with a thio-reactive compound, which helps to protect the surface against non-specific binding of DNA and proteins. Prior to hybridising the telomeric and non-telomeric DNA, the baseline fluorescence signal (F_{down}) was low because the ssDNA was flexible, and the fluorophore was quenched by the gold biochip surface. Addition of the complementary DNA strand increased the rigidity of the DNA, which propelled the DNA away from the negatively charged surface and increased the fluorescence intensity (F_{up}). The

F_{down} and F_{up} curves were generated in the dynamic mode by applying an alternating potential to the gold surface; a wide separation between the two curves was observed indicated that the surface is clean and functional.

After the surface functionalisation step, the static FPS mode was used to interrogate DNA binding and unbinding. In this mode, the change in fluorescence intensity of the dye was used as the indicator of binding events in real time; the association of protein to telomeric DNA was detected due to quenching of the fluorescence intensity of the dye, whilst dissociation events during the buffer wash step led to an increase in the fluorescence intensity. During the association phase, SEC-purified apo full and (–TRF1) shelterin^{TIN2L} complexes (0, 62.5, 125, 250, 500, and 1000 nM) were flushed over the biochip at 100 µL min^{-1} to enable interaction with the surface-tethered oligonucleotides. During the dissociation phase (only performed after the association of 1000 nM of protein), the sample buffer was flushed over the biochip.

With time and use, the fluorescence intensity decreases and limits the lifespan of the biochip. All electrodes within the same channel were exposed to the same chemical and thermal conditions. However, only the electrodes selected for switchSENSE experiments were illuminated to minimise loss in fluorescence intensity in other electrodes due to photobleaching. Multiple steps were taken to access the quality of the channel and its sensory electrodes. Immediately after inserting the biochip, a status check was performed to provide information on the electrode quality within the selected channel. To check the chip status, the DNA nanolevers was in the double-stranded state without any ligands attached. Hybridisation was another step where the quality of the electrode was assessed. In this step, the complementary DNA strands (or DNA protein conjugates) were annealed to the immobilised DNA nanolevers on the chip surface. As the complementary DNA annealed, the nanolever was converted from a single-stranded to a double-stranded state. In this state, the nanolever showed enhanced switching, indicated by an increase in the repulsive fluorescence intensity (F_{up}) and minimal change in attractive fluorescence intensity (F_{down}).

The experimental data were analysed using the switchANALYSIS software (Dynamic Biosensors GmbH). First, the fluorescence signal from buffer-only association and dissociation runs were subtracted from the titration curves. Then, the data were normalised and fitted with a global 1:1 Langmuir binding model. The raw data points and fitted curves were exported and visualised in R.

Name	Sequence (5'–3')
mutDNA1_fwd	CAGGTACCCGGGTTATCGTTATCGTTATCGTTATCGTTATCGTTATCG TTAT
mutDNA1_rev	**ATAACGATAACGATAACGATAACGATAACCCGGGTACCTG**ATCAGT ACTTGTCAACACGAGCAGCCCGTATATTCTCCTACAGCACTA
teloDNA1_fwd	CAGGTACCCGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAG GGTTAG
teloDNA1_rev	**CTAACCCTAACCCTAACCCTAACCCTAACCCGGGTACCTG**ATCAGC GTTCGATGCTTCCGACTAATCAGCCATATCAGCTTACGACTA

Table 6.7. Telomeric and non-telomeric oligonucleotides used for the switchSENSE experiments.

The bolded regions contain the complementary sequence for annealing to the DNA nanolever. The mutDNA and teloDNA were annealed to the green and red fluorescently labelled DNA nanolever, respectively.

Buffer	Composition
Functionalisation buffer	10 mM HEPES pH 7.4, 40 mM NaCl, 0.05% Tween 20, 50 µM EDTA, 50 µM EGTA
Sample buffer	50 mM HEPES pH 8.0, 300 mM NaCl, 1 mM TCEP, 10% glycerol

Table 6.8. List of buffers used for the switchSENSE experiments.

6.6 Integrative structural biology

In this section, I introduce the concepts underlying electron microscopy and crosslinking mass spectrometry, commonly used in conjunction to probe the architecture of macromolecular complexes.

6.6.1 Electron microscopy

Electron microscopy (EM) is a rapidly advancing technique for structure determination by averaging many low-contrast images of frozen-hydrated samples. It has transformed the structural biology toolbox, facilitating structural studies of complex biological samples ranging from cell organelles using cryo-electron tomography (cryo-ET) (Hylton & Swulius, 2021), to single particles including viruses, protein complexes and helical assemblies using single-particle cryo-electron microscopy (cryo-EM) (Nogales & Scheres, 2015).

6.6.1.1 Sample preparation for negative-stain EM

Fundamentally, single-particle EM analysis relies on the computational averaging of identical particle images. One significant advantage of EM techniques compared with other established methods such as X-ray crystallography (XRC) is that sample heterogeneity (in conformation or composition) can be tolerated to a certain extent during image analysis (see section 6.6.1.6), and in some cases provides insightful mechanistic information. However, careful optimisations of sample preparation are often performed to minimise structural heterogeneity and simplify downstream image processing. Typically, a new EM project entails visualising the specimen using negative-stain EM in the initial stages. This process generates lower-resolution but high-contrast images of the sample, which helps to evaluate the structural homogeneity of the specimen.

Negative-stain is a simple and quick procedure by which samples are encased in a layer of a heavy-metal salts, typically uranium, tungsten and molybdenum (Ohi, et al., 2004). The density of the stain is close to 3-times higher than protein, which generates high-contrast images (Passmore & Russo, 2016). Conventionally, the sample is adsorbed onto a continuous carbon film deposited on a support structure, typically a 3-mm metal grid. Next, the excess sample is removed by blotting using cellulose filter paper. The grid may be washed with deionised water to remove buffer components, which may interfere with staining or generate a high background. The grid is then covered with the heavy metal stain, which is also removed by blotting and allowed to dry before being inserted into the electron microscope. Besides increasing the contrast, the layer of heavy-metal stain also provides some protection against distortions due to dehydration in the column vacuum and radiation damage during image acquisition. However, the resolution of the 3D reconstruction achieved using negative-stain EM is typically limited to ~20Å (Ohi, et al., 2004). The next step is cryo-EM, which uses a rapid freezing method to create a thin layer of biological sample in amorphous ice (Adrian, et al., 1984). This allows biological samples to be studied under native-like conditions and 3D reconstructions to be built at (near)-atomic resolution, heralding the beginning of the resolution revolution in molecular biology (Kühlbrandt, 2014).

6.6.1.2 The electron microscope

The first transmission electron microscopes were constructed in the 1930s by Ernst Ruska and Max Knoll (Ruska, 1987). This was the first major step to visualise biomolecules smaller than the wavelength of light at higher resolution. However, the earliest microscopes merely showed that electron beams could be focused to produce images of matter. Since then, there have been many advancements on the hardware front, leading to a significant drop in the theoretical resolution limit achievable by cryo-EM.

The electron microscope consists of an electron source, a set of condenser and objective lenses, a specimen stage, and an electron detector. All these components are housed in a column and operated under high vacuum to prevent the electrons from being scattered by particles in the air. Early sources of electrons were generated from passing high current through a tungsten filament. This heats up the filament, and when the electron energy is greater than the work function of tungsten, electrons are emitted from the filament (Orlova & Saibil, 2011). An improved electron source is the LaB_6 crystal, from which electrons can be produced from a smaller surface area, producing a more coherent electron beam (Orlova & Saibil, 2011). Currently, many high-performance microscopes use the field emission gun (FEG), which has an even smaller source size and produces an even more coherent and brighter beam with a narrow spread of energies (Assaiya, et al., 2021). Next, the emitted electrons are accelerated to 100 – 300 kV and converted into a parallel beam by a set of condenser lenses. The beam illuminates the sample on the specimen stage, which is sandwiched between the condenser and objective lenses. As the beam penetrates the sample, the electrons are scattered and focussed by the objective lens, which provides the primary magnification. Electrons scattering at high angles are filtered by an objective aperture in the back focal plane of the objective lenses to improve the signal-to-noise ratio of the image. The image is further magnified by immediate and projector lenses before reaching the image detector.

6.6.1.3 Electron detectors

The conventional detector in electron microscopes used to be photographic film. One of the main limitations of film is the long processing time required before the images can be digitised for analysis. This significantly limits the rate of data acquisition. The development of digital image sensors and automated data collection procedures have been a key driver in pushing the resolution limit attainable by cryo-EM. The two main types of digital detectors are charged-coupled devices (CCDs) and monolithic active pixel sensors (MAPS) synthesised with complementary metal oxide semiconductor (CMOS) technology (McMullan,

et al., 2016). CCDs are indirect electron detectors that convert the analogue photon energy into electrical charges using photosensitive elements (wells). The charges are then transferred between neighbouring pixels to an output node and turned into a digital signal (Orlova & Saibil, 2011). Scintillators are used to convert the high-energy electrons to photons, which are transferred to the CCD chip to protect the photosensitive wells from radiation damage (Orlova & Saibil, 2011). Higher accelerating voltages (>120 kV) are often necessary to image thicker specimens and insulators, including biological specimens embedded in ice, due to the reduced effect of charging in the specimen (McMullan, et al., 2016). At these high voltages, thicker scintillators are necessary to improve electron detection efficiency of CCD chips (Orlova & Saibil, 2011). However, this adds more noise to the images because the electrons are scattered in the scintillator (Orlova & Saibil, 2011). One key indicator of detector performance is its detective quantum efficiency (DQE), which is a measure of the amount of noise contributed by the detector. This is defined by the square of the ratio of output SNR, SNR_{OUT}, to that of the input, SNR_{IN}

$$DQE = (SNR_{OUT})^2 / (SNR_{IN})^2 \qquad \text{equation 16}$$

An ideal detector does not degrade the original signal in the image, and hence has an DQE of 1. However, in practice all detectors have DQE values of less than 1. For example, a CCD camera operating at 300 kV has a lower DQE than film (McMullan, et al., 2016). Current direct electron detectors use the CMOS/MAPS technology (McMullan, et al., 2016). The use of back-thinned radiation-hardened MAPS allows direct exposure to the electron beam at high accelerating voltages (Orlova & Saibil, 2011). This type of detector offers improved DQE at 300 kV over film and provides a faster readout than CCD cameras (McMullan, et al., 2014; McMullan, et al., 2016).

6.6.1.4 Image formation and contrast

Two types of contrast are generated when biological samples are imaged. The amplitude contrast results from the absorption of electrons from the incident beam. Biological samples are mostly comprised of light atoms (C, H, O, N) that rarely absorb electrons, thereby producing very little amplitude contrast. Instead, most of the contrast is generated when the electrons are scattered by atoms in the sample. The electrons in the exit wave travel at varying path lengths, producing phase contrast that can be converted into amplitude variations that can be registered by a detector. There are two types of scattering events. Inelastic scattering occurs when the incoming electrons transfer their energy to the sample. This can ionise atoms in the sample, lead to X-ray emission and produce free radicals, all of which damage the structural integrity of the sample (Orlova & Saibil, 2011). This is a significant obstacle to imaging biomolecules at atomic resolution. Typically, biological specimens are imaged with a low dose of electrons in the range of 1 to 50 electrons/$Å^2$ to minimise the extent of radiation damage. The other type of event is elastic scattering, whereby the electrons are scattered without energy loss. This is the desired form of scattering for image formation in the electron microscope. The scattering angles are proportional to the atomic number of the atom. Biological samples deflect the incoming electrons through small angles, i.e., they are weak scatterers and are considered weak phase objects (Orlova & Saibil, 2011).

One way to increase the contrast is with negative stain, which consist of atoms with high atomic number that absorb and scatter more electrons. In cryo-EM, the incident electron beam is barely changed by the biological specimen when imaging at focus; therefore, very little contrast is generated. Therefore, to increase the phase shift between incident and scattered electrons, samples are imaged under-focus. In practice, other factors such as the level of focus, lens imperfections including spherical and chromatic aberrations, and limited beam coherence affect the image generated by an electron microscope (Orlova & Saibil, 2011). All these lead to a tapering of signal transfer for high-resolution features, with the result of blurring of the finer details in the image. The effect of these

factors on image formation can be described by the contrast transfer function (CTF) (Orlova & Saibil, 2011).

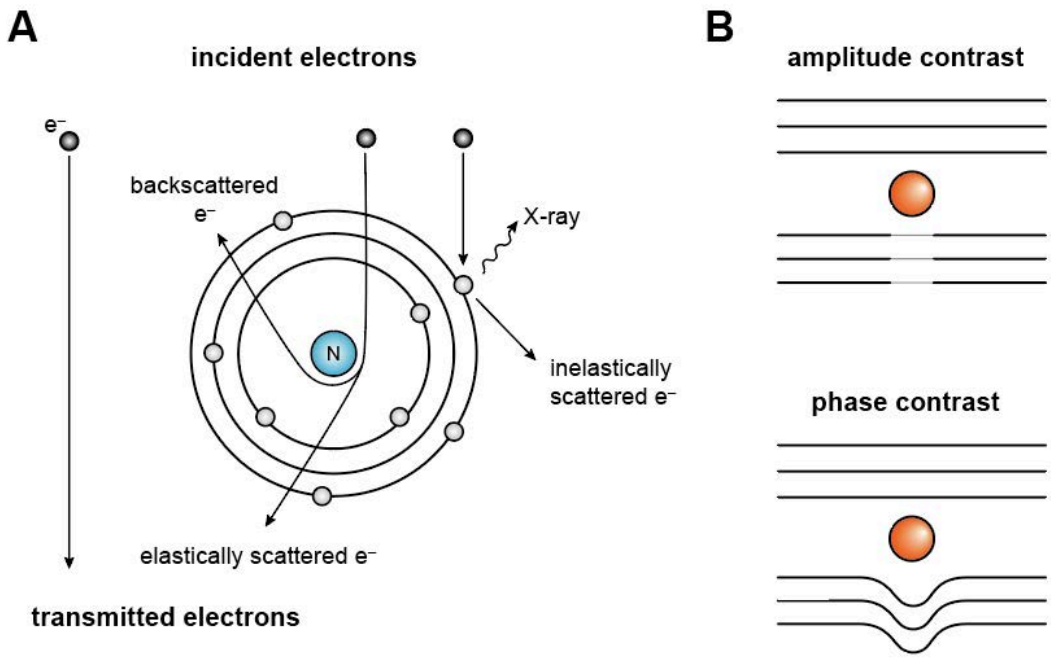

Figure 6.5. Electron scattering and image contrast.

(A) Schematic of elastic and inelastic scattering of electron by the sample (adapted from (Orlova & Saibil, 2011)). **(B)** Schematic of two different types of contrast, generated by the absorption of electrons (amplitude contrast) and changing of the phase of the exit electron wave (phase contrast).

6.6.1.5 Image processing

In structure determination using EM techniques, 2D micrographs are taken of a 3D structure of interest. The recorded images are affected by the level of focus and other operating conditions of the microscope (see above). The defocus and CTF parameters of each image can be estimated by computing theoretical CTFs at different defoci and matching one to the experimental CTF of the image (Figure 6.6A). One method to correct for the CTF and restore the specimen spectrum of the images is phase correction, whereby the image phases in the negative CTF

region are flipped to the positive (Figure 6.6B) (Orlova & Saibil, 2011). Some information is lost at points where the CTF has a zero amplitude (Figure 6.6B). In practice, this is solved by taking images at a wide range of defoci, and information can be restored in each image at positions where CTF is zero during image processing (Figure 6.6C) (Orlova & Saibil, 2011).

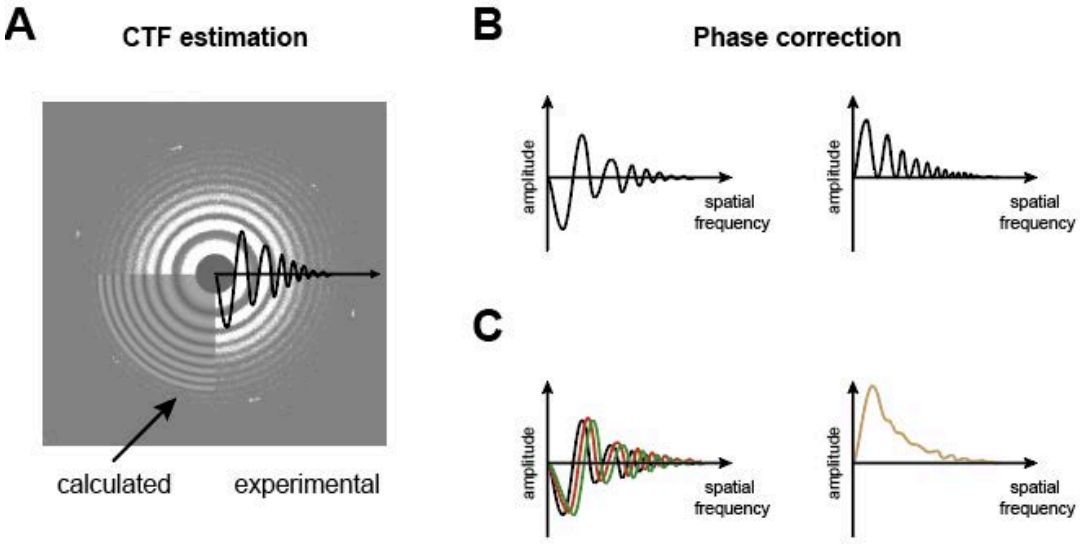

Figure 6.6. Diffraction pattern and CTF correction.

(A) Example of a diffraction pattern with the calculated CTF (bottom left) matched to the experimental CTF. **(B & C)** CTF correction using the phase correction approach using image collected at a single defocus (B) or multiple defoci (C) (adapted from (Orlova & Saibil, 2011)).

After CTF correction, the selected micrographs are used for selecting the particles of interest. Data processing suites such as RELION (Scheres, 2012b), cisTEM (Grant, et al., 2018), and cryoSPARC (Punjani, et al., 2017) all include options for both manual and automated particle selection. Various softwares, such as Gautomatch (Zhang), crYOLO (Wagner, et al., 2019), and Topaz (Bepler, et al., 2019) have also been developed to assist with this process. The particles are then extracted (boxed) as individual images. These individual particles images are noisy, due to factors such as differences in ice or stain thickness, damage to the sample during grid preparation or imaging, and noise from the electron detector. 2D classification is the stage in image processing that

generates 2D averages (classes) by comparing the particle images and superimposing those with the same features. This process reduces the random noise associated with individual particle images and generate 2D classes with higher SNR. This is often used to assess the quality of the data and as a pre-processing step leading up to 3D reconstruction.

At the start of 2D classification, the particle images are subjected to many iterations of alignment to minimise the differences between their translation and in-plane rotation. However, pairwise comparison of all the images is computationally expensive due to the high dimensionality of image data. This can be mitigated by using statistical approaches, such as multivariate statistical analysis (MSA), which essentially captures a smaller set of uncorrelated variables with the largest variations in a dataset, known as principal components (Orlova & Saibil, 2011). The coordinate system is then restructured using these major components as the axes, thereby reducing the number of variables used to describe the data. In this approach, each pixel from an image is represented as a vector of coordinates, and a set of images are considered as cloud of vectors (Figure 6.7A). Similar images will appear close together within the data cloud. Finally, the images are classified into different clusters. Two common approaches are hierarchical and K-means clustering (Figure 6.7B) (Orlova & Saibil, 2011). In hierarchical clustering, images are considered as separate classes and are grouped based on the distance, or similarity, between them in the data cloud. In K-means clustering, a set number of points (K) are picked from the data at random as seeds. Each point in the cloud is then allocated to the nearest cluster initiated by one of the K points. As this occurs, the centre of the cluster is reorganised with the goal to minimise the intra-cluster distances. Classification by K-means clustering is highly dependent on the initial assignment of the class centres and is often used for sorting data to smaller number of clusters (Orlova & Saibil, 2011). Variations of hierarchical and K-means clustering form the basis of classification in early data processing softwares including IMAGIC, which focuses on minimising the intra-class variation in a cluster and maximising the inter-class variation between the centre of different clusters (van Heel, et al., 1996).

Another classification approach implemented by increasingly popular data processing suites, including RELION (Scheres, et al., 2005) and cryoSPARC (Punjani, et al., 2020), is the maximum likelihood (ML) method. The ML algorithm is used to align and classify particle images simultaneously. In RELION, random subsets of unaligned particle images are initially selected to generate averages for multi-reference alignment (Scheres, et al., 2005). The number of references is pre-specified and can be adjusted based on the number of particles and number of different 2D views present in the data. The latter is usually unknown, and thus the optimal number of references are determined empirically. A characteristic of the ML approach is that the particle images are not assigned to any particular class. Rather, the images are being compared with all the references, and the probability of the particle being assigned to each are calculated. Class averages are computed as weighted averages over all possible assignments, which are refined over multiple iterations. An important advantage of this method is that it reduces the likelihood of premature commitment of particles to a particular class caused by noise and poor alignment during the initial stages of classification. The ML approach can be applied to the 2D and 3D classification stages (described in detail below).

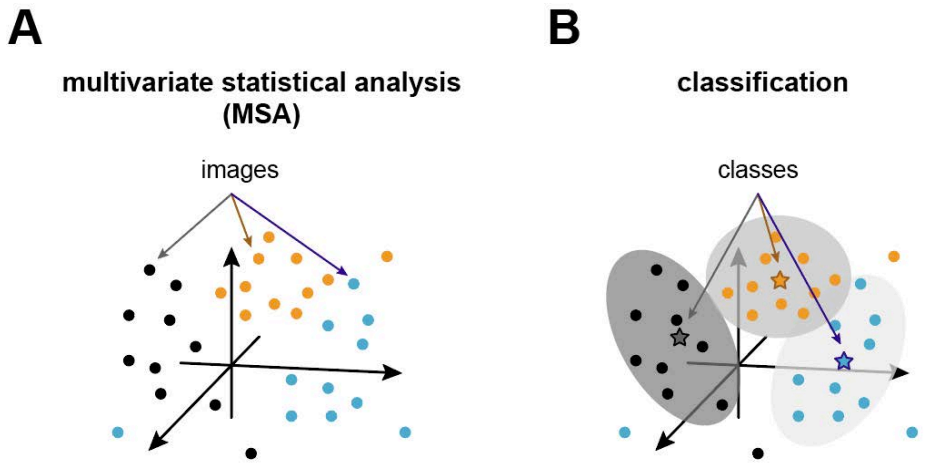

Figure 6.7. Principles of multivariate statistical analysis and classification.

(A) In MSA, each image is considered as a point in a multidimensional data cloud. **(B)** The data points are clustered into classes according to their position in the data cloud. The centre of each class (the class average) is represented by the star.

In the electron microscope, 2D images (projections) are formed from imaging a rotated and translated 3D object. These 2D projections represent many different views of the object, which can be assembled to reconstruct the original 3D density using the projection-slice theorem. This theorem states that the Fourier transform (FT) of a 2D projection image is a section through the origin of the 3D FT of the object. Therefore, if 2D projections at multiple viewing angles and their orientations are known, the corresponding 2D Fourier slices can be fitted within the 3D FT. The original 3D object can then be reconstructed in real space using the inverse FT. The biggest challenge is to determine the relative orientations of the 2D projections. One iterative approach is 'projection-matching', whereby the experimental particle images are matched to reprojections of a 3D reference structure computed at various angles (Penczek, et al., 1994). By using the best matching orientation angles, the reconstruction at each iteration should be an improvement on the initial reference. The ML approach is another iterative approach to 3D classification and reconstruction. Unlike the projection-matching approach, the ML approach does not assign individual 2D projections are not assigned to a single, best orientation. Instead, it attempts to find the 3D structure or structures that best explain the experimental images by marginalising over the class assignment and the unknown orientation (Scheres, 2012b; Punjani, et al., 2017). Instead, it integrates over the probability distribution of all potential orientations at each iteration. A regularisation term is also imposed during 3D refinement to penalise unnecessary complexity in the model and prevent overfitting (Scheres, 2012a; Scheres, 2012b).

Traditionally, most iterative refinement approaches are local optimisation methods, where starting with an accurate initial model is key in achieving a correct final 3D reconstruction. An inaccurate initialisation likely results in an incorrect structure that represents a local optimal probability within the space of all 3D structures (Punjani, et al., 2017). CryoSPARC implements a stochastic gradient descent (SGD) algorithm to reconstruct low-resolution initial structures (Punjani, et al., 2017). During 3D *ab initio* reconstruction using the SGD scheme, an arbitrary 3D density is computed and improved over many incremental steps. At each step, a random selection of particle images is used to refine the 3D structure. This is relatively inexpensive and allows the algorithm to search across

the space of all 3D structures in a stochastic manner, which renders the process insensitive to initialisation (Punjani, et al., 2017). The 3D *ab initio* reconstruction step is also a tool for exploring structural heterogeneity in the images (see more on dealing with heterogeneity below) (Punjani, et al., 2017). These low-resolution structures can then be used as the starting reference during 3D refinement to push the EM density maps to higher resolution.

6.6.1.6 Heterogeneity

In practice, the resolution of the final 3D structure determined by cryo-EM is often limited by compositional and/or structural heterogeneity in the sample, rather than due to problems in the imaging process or data processing. However, the transition of biomolecules between states of different size and conformation can provide mechanistic insight into protein function (Ripstein, et al., 2017; Ripstein, et al., 2020) and have been key in time-critical studies of human pathogens (Wrobel, et al., 2022). Several computational tools have been developed to examine both discrete and continuous heterogeneity within a sample. Discrete heterogeneity is generally explored and dealt with during extensive rounds 2D and 3D classifications (Scheres, 2016). On the other hand, continuous heterogeneity tends to be more challenging to resolve computationally. In RELION, one approach that can account for a certain extent of conformational heterogeneity is multi-body refinement. In this method, the density map containing flexible regions is divided into separate rigid ‘bodies’ for focused refinement (Nguyen, et al., 2015; Nakane, et al., 2018). In cryoSPARC, the 3D variability analysis offers a principal component analysis (PCA)-based approach to fit 3D linear subspace models to visualise continuous motion and flexibility of biomolecules (Punjani & Fleet, 2021).

6.6.1.7 Negative-stain EM sample and grid preparation

For sample preparation using GraFix (section 4.1.2), 500 µl of 6 µM affinity-purified $(–TRF1)^{TIN2L}$ complex was incubated with 7.2 µM of the 52/40-mer model telomeric DNA (teloDNA1, see Table 6.6) on ice for 60 mins. The sample was purified using Superose 6 Increase 10/300 GL, eluting at 0.3 mL/min in 200 µl fractions. Two SEC fractions from the centre of peak ***I*** was pooled and subjected to ultracentrifugation at 80,000 xg for 18 hr at 4 ºC in a 10 – 50% glycerol gradient without or with a 0 – 0.2% glutaraldehyde gradient. Next, the gradient was fractionated in 100 µL starting from the top. Fractions were analysed by SDS-PAGE to select the fractions for making negative-stain grids.

For the negative-stain EM analysis of DNA-bound $(–TRF1)^{TIN2S}$ (section 1.1.4), 500 µL of 1 µM SEC-purified $(TRF1)^{TIN2S}$ complex were incubated with 5 µM of the 40/28-mer model telomeric DNA (teloDNA2, see Table 6.6) on ice for 30 mins. The sample was crosslinked with a final concentration 0.075% of glutaraldehyde, quenched with a final concentration of 100 µM of Tris-HCl pH 8.0. The crosslinked sample was centrifuged at 15,000 xg for 10 min before purification using a Superose 6 Increase 10/300 GL column, eluting at 0.3 mL/min in 200 µL fractions. Fractions were analysed by SDS-PAGE, EMSA and mass photometry.

Quantifoil copper grids (1.2/1.3 µm holes, 400-hole mesh copper grids, Quantifoil Microtool GmbH) supporting a continuous carbon film were glow-discharged using the easiGlow Glow Discharge Cleaning System (PELCO) at 15 mA for 60 s. 2 µL of a SEC fraction were applied to the grid for 1 min. The sample was removed by blotting with cellulose blotting paper and washed with deionised water twice before staining with a drop of 2% (w/v) uranyl acetate for 30 s. The excess stain was blotted away, and the grids were left to air-dry.

6.6.1.8 Negative-stain EM data collection and analysis

For screening purposes, negative-stain grids were imaged using the Tecnai T12 microscope with a LaB_6 filament operating at an accelerating voltage of 120 kV with a 1K x 1K CCD TemCam-F114T detector (TVIPS, GmbH). For data presented in section 4.1.4, automated data collections were performed by Dr Fabienne Beuron using a FEI Tecnai TF20 transmission electron microscope operating at 200 kV accelerating voltage, via the EM-Tools software (TVIPS, GmbH). The micrographs were recorded using a TemCam-F416 4K x 4K CMOS detector (TVIPS, GmbH) at 50,000x magnification, resulting in a pixel size of 1.732 Å. Micrographs were processed using cryoSPARC v3.3.1 (Punjani, et al., 2017). The EM maps from 3D *ab initio* reconstructions were visualised using UCSF Chimera (Pettersen, et al., 2004) and ChimeraX (Pettersen, et al., 2021). Back-projections were generated from a 3D *ab initio* structure using RELION v3.1.1 (Scheres, 2012b). The negative-stain EM study of the DNA-bound full shelterin^{TIN2S} complex was performed by Dr Oviya Inian.

6.6.2 Crosslinking mass spectrometry

6.6.2.1 Background

In recent years, crosslinking mass spectrometry (XL-MS) has seen increasing integration into the pipeline for structural characterisation of proteins and other macromolecules. It provides complementary information for targets that are not amenable to conventional high-resolution structural analysis (Schmidt & Urlaub, 2017). The crosslinking reaction takes place in solution and tolerates a diverse range of buffer conditions. Moreover, it can work with heterogenous samples, from single proteins to whole cell lysates (Graziadei & Rappsilber, 2022). Rapidly expanding XL-MS methodologies are providing valuable insights into protein structure, topology, conformational changes and protein-protein and protein-DNA interactions (Graziadei & Rappsilber, 2022).

A wide range of chemical crosslinkers is available with varying characteristics, including functional group specificity, homo- or hetero-bifunctionality, variable spacer arm length, water solubility, and cleavability by mass spectrometry. Examples of popular crosslinkers are listed in Table 6.9. Chemical crosslinkers are composed of two main elements: the reactive group and a spacer arm. The crosslinking reaction generates covalent bonds between the reactive groups on the crosslinker and the target functional group in the sample. One of the most commonly targeted groups by chemical crosslinkers are the primary amines. Primary amines ($–NH_2$) are present at the N-terminus of a polypeptide chain (α-amine) and in the side chain of lysine residues (ε-amine). They are positively charged (hydrophilic) and are usually solvent-exposed under physiological conditions, which increases their accessibility for crosslinking. One example of reactive chemical groups targeting primary amines are N-hydroxysuccinimide (NHS) esters. These crosslinkers react with primary amines to form stable amide bond, which eliminates NHS in the process (MW: 115 g/mol). For *in vitro* crosslinking reactions, samples are prepared in non-reactive buffers, such as HEPES, phosphate, carbonate-bicarbonate, or borate buffers, and can be quenched using primary amine buffers such as Tris. The reaction is optimal at physiological or slightly alkaline conditions (pH 7.2 – 8.5). NHS esters are less stable at higher pH, and hydrolysis at alkaline pH competes against the primary amine reaction. The half-life for NHS-ester containing compounds in aqueous buffer is 4 to 5 hours at pH 7.0, compared with 1 hour at pH 8.0 (Cuatrecasas & Parikh, 1972; Lomant & Fairbanks, 1976). Under optimal crosslinking conditions, most crosslinkers have either reacted with primary amines or hydrolysed within 30 mins to an hour (Iacobucci, et al., 2018). Another important reason for avoiding long incubation times is the formation of protein aggregates caused by extensive crosslinking.

For folded proteins, the targeted groups are usually located at the solvent-accessible surface of proteins. Crosslinkers with reaction chemistries targeting specific groups limits the amount structural information that can be derived. However, they simplify data analysis and provide a manageable amount of information that is amenable to downstream validation techniques. The length of the spacer arm also influences spatial resolution and data complexity; a shorter

spacer arm limits the number of crosslinks generated but provides more useful information for structural modelling. These considerations highlight one of the major challenges posed by complex crosslinking reaction mixtures. MS analysis has to identify the crosslinked peptides and the exact point of connectivity. Since crosslinked peptides consist of two peptides, the search space expands exponentially and demands powerful softwares to examine all possible combination of peptide pairs that satisfy the mass of the crosslinked product (Liu, et al., 2015).

Crosslinker	Reactive moiety	Reactive towards	Spacer (Å)	Water-soluble	MS-cleavable
EDC	Carbodiimide	Amines, carboxyl	0	Yes	No
BS3	NHS ester	Amines	11.4	Yes	No
DSSO	**NHS ester**	**Amines**	**10.3**	**No**	**Yes**
DHSO	Hydrazide	Acidic residues	12.4	Yes	Yes
BM(PEG)2	Maleimide	Sulfhydryl	14.7	Yes	No

Table 6.9. Examples of homo-bifunctional chemical crosslinkers.

6.6.2.2 XL-MS sample preparation, data collection and analysis

Three XL-MS experiments are described in this thesis. An overview of the protocol is shown in Figure 6.8. The two preliminary experiments optimised the DSSO crosslinker concentration and crosslinking duration. In these two experiments, the full apo shelterin^{TIN2L} complex was purified using the two-step protocol with SEC purification as the final step (as described in section 6.3.3). In the third experiment, the full apo shelterin^{TIN2L} complex was purified using the three-step protocol (as described in section 6.3.3). From the final SEC polishing step, fractions containing the minimal shelterin^{TIN2L} complex were pooled. To reconstitute the shelterin^{TIN2L}-DNA complex, 3.5-fold molar excess of teloDNA1 was added to the SEC-purified apo shelterin^{TIN2L} on ice for 30 mins. DSSO crosslinker (Thermo Scientific, cat #A33545) was dissolved in DMSO to make a 50 mM stock. Various DSSO concentrations were added to 1 – 2 µM of SEC-purified shelterin (in 50 mM HEPES pH 8.0, 0.5 M NaCl, 1 mM TCEP, 10%

glycerol) and incubated at 20˚C for 10 to 30 mins on a shaking platform. See figure legends for the exact crosslinking conditions. The crosslinking reaction was quenched with Tris-HCl pH 8.0 (final concentration of 50 mM), and buffer exchanged into 20 mM HEPES pH 8.0, 0.5 M NaCl, 0.1 mM TCEP by dialysis.

For XL-MS analysis of the TRF2/RAP1 subcomplex, SEC-purified dStrepII-TRF2 and dStrepII-RAP1 proteins were mixed in a 2:2 molar ratio (1.6 μM) and crosslinked with 500-times molar excess of DSSO on ice for 25 mins. The samples were quenched with Tris-HCl pH 8.0 (final concentration of 50 mM) and submitted for qualitative XL-MS analysis.

Triethylammonium bicarbonate buffer (TEAB) was added to the sample to a final concentration of 100 mM. Proteins were reduced with 5 mM TCEP for 1 hr at 60 ℃, alkylated with 10 mM iodoacetamide (IAA) for 30 mins in the dark and digested overnight with trypsin at final concentration 50 ng/μL (Pierce). For the quantitative experiments, peptides were labelled with the TMT10plex reagents (Thermo Scientific) according to manufacturer's instructions. Peptides were fractionated with high-pH reversed-phase RP chromatography using an XBridge C18 column (2.1 x 150 mm, 3.5 μm, Waters) on a Dionex UltiMate 3000 HPLC system. Mobile phase A was 0.1% v/v ammonium hydroxide, and mobile phase B was acetonitrile, 0.1% v/v ammonium hydroxide. The peptides were fractionated at 0.2 mL/min with the following gradient: 5 mins at 5% B, increasing up to 12% B in 3 mins, followed by a 32-min gradient to 35% B, a 5-min gradient to 80% B, isocratic elution for 5 mins and finally re-equilibration to 5% B. Fractions were collected every 42 s, SpeedVac-dried and orthogonally pooled into up to 20 samples for MS analysis per experiment.

LC-MS analysis was performed using the Dionex UltiMate 3000 UHPLC system coupled with the Orbitrap Lumos Mass Spectrometer (Thermo Scientific). Each peptide fraction was reconstituted in 30 μL of 0.1% formic acid, and 15 μL were loaded to the Acclaim PepMap 100, 100 μm × 2 cm C18, 5 μm trapping column at a 10 μL/min flow rate of 0.1% formic acid loading buffer. Peptides were then subjected to a gradient elution on the Acclaim PepMap (75 μm × 50 cm, 2 μm, 100 Å) C18 capillary column connected to a stainless-steel emitter with

integrated liquid junction (MSWIL, #PSSELJ) on the EASY-Spray source at 45 °C. Mobile phase A was 0.1% formic acid, and mobile phase B was 80% acetonitrile, 0.1% formic acid. The gradient separation method at flow rate 300 nL/min was as follows: 95 mins gradient from 5% – 38% B, 5 mins up to 95% B, followed by 5 mins isocratic elution at 95% B, re-equilibration to 5% B in 5 mins, and finally 10 mins isocratic elution at 5% B. Precursors between 375-1,600 m/z and charge equal to or higher than +3 were selected at 120,000 resolution in the top speed mode in 5 s and were isolated for collision-induced dissociation (CID) fragmentation (collision energy, CE; 25%) with quadrupole isolation width 1.6 Th and Orbitrap detection with 30,000 resolution. Fragments with targeted mass difference of 31.9721 (DSSO crosslinker) were further subjected to CID fragmentation at the MS3 level with CE 35%, ion trap detection, max IT 50 ms, AGC 2×10^4 and MS2 isolation window 2 Th. Two precursor groups were selected with both ions in the pair. Targeted MS precursors were dynamically excluded for further isolation and activation for 30 seconds with 10 ppm mass tolerance. For the TMT experiments, an additional high-energy collisional dissociation (HCD) MS2 scan with CE 38% at 50,000 resolution was performed for every precursor ion back-to-back with the MS2-MS3 acquisition to obtain quantification of crosslinks.

Identification of crosslinked peptides was performed in Proteome Discoverer 2.4 (Thermo Scientific) with the XlinkX search engine in the MS2-MS3 mode. Precursor, Fourier transform MS (FTMS) and ion trap MS (ITMS) mass tolerances were 20 ppm, 30 ppm and 0.5 Da, respectively, with a maximum 2 trypsin missed cleavages allowed. Carbamidomethyl at C and oxidation at M were selected as static and dynamic modifications, respectively. For the quantitative experiments, TMT6plex at N-terminus and K were selected as dynamic modifications. Spectra were searched against a FASTA file containing the sequences of the proteins in the complex. Crosslinked peptides were filtered at FDR < 0.01 using the Percolator node and decoy database search. The reporter ion quantifier node included a TMT quantification method with an integration window tolerance of 15 ppm. Only peptides with average reporter signal-to-noise > 3 were used for quantification.

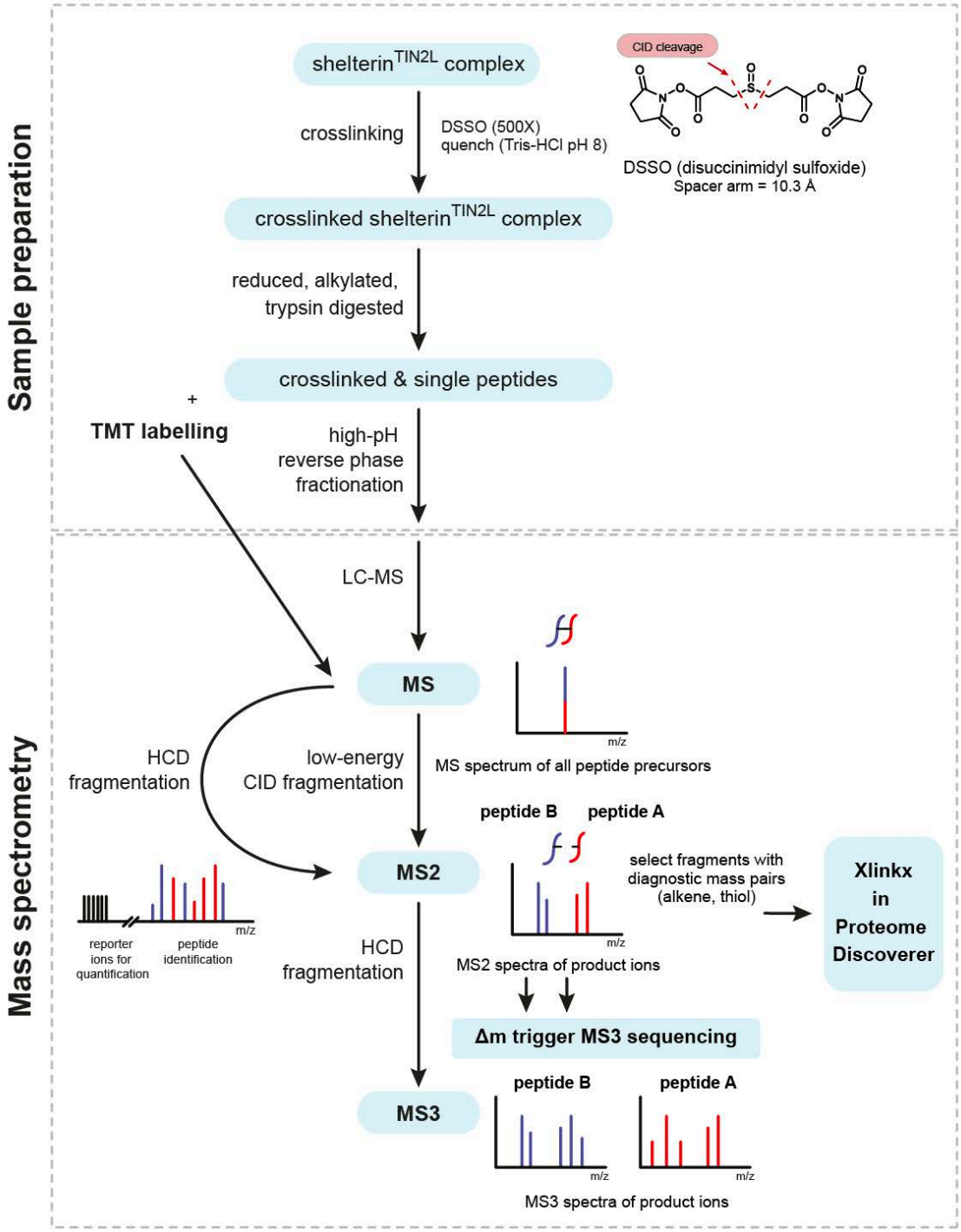

Figure 6.8. Overview of the XL-MS workflow.

7 Appendix

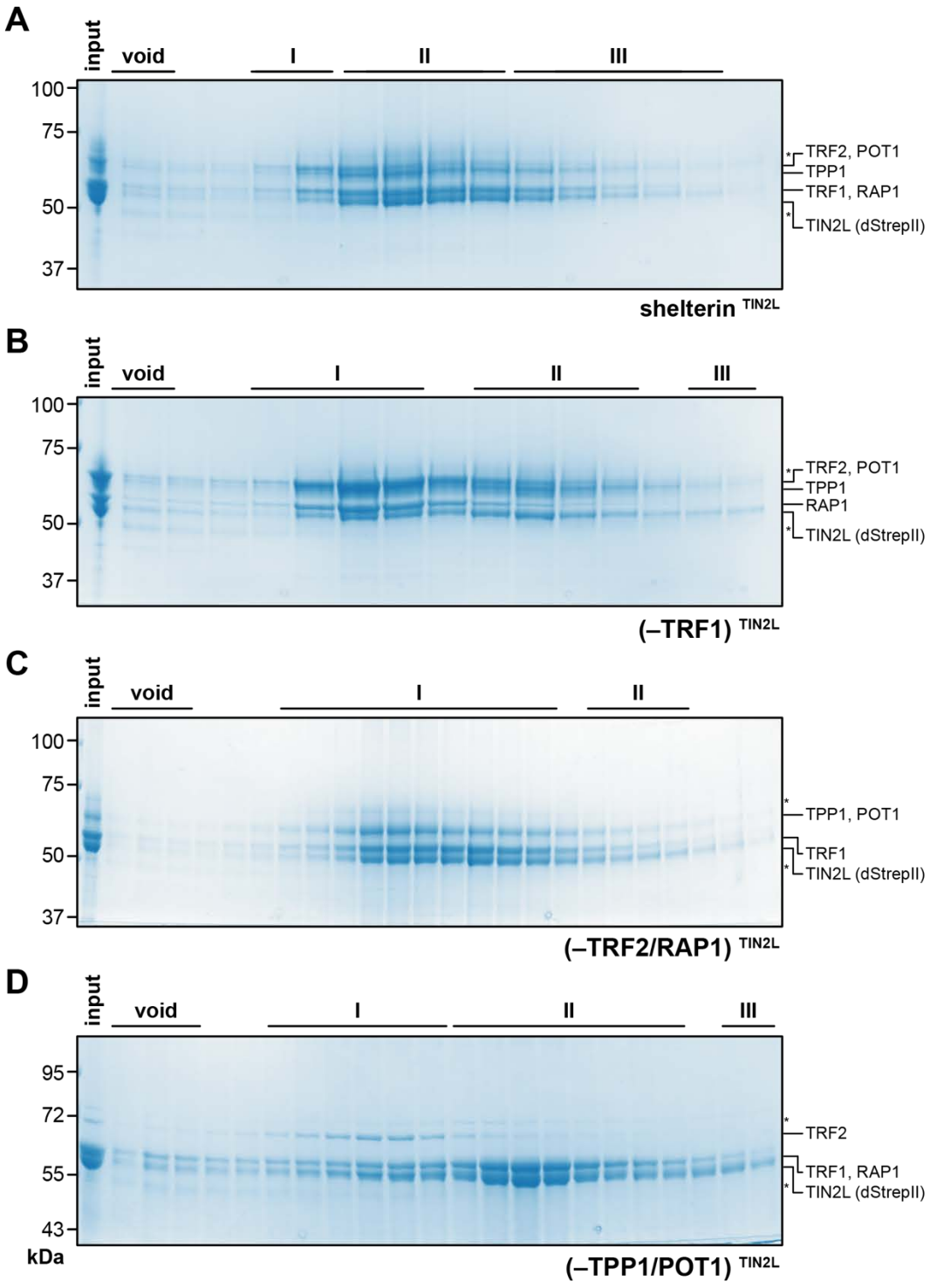

Figure 7.1. Panel of SEC-purified shelterin^{TIN2L} complexes.

SEC purification of the shelterin^{TIN2L} complex **(A)** containing all six-subunits, or lacking **(B)** TRF1, **(C)** TRF2/RAP1, **(D)** TPP1/POT1. The affinity-purified shelterin complexes were injected into a Superose 6 Increase 10/300 GL column and eluted in 200 µL fractions. Fractions across the SEC elutions were analysed by SDS-PAGE. * Contaminants (see section 2.1.2).

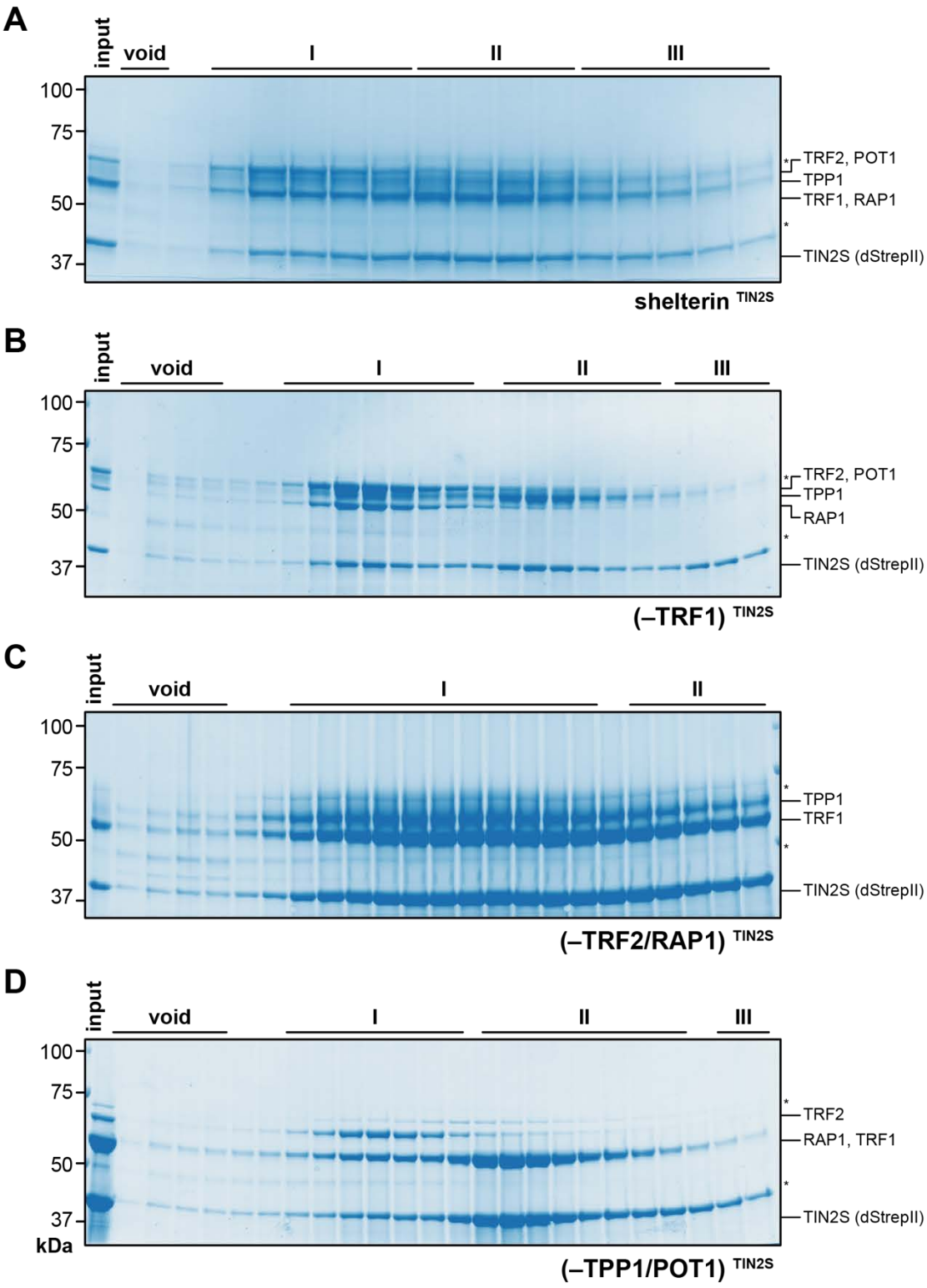

Figure 7.2. Panel of SEC-purified shelterin^{TIN2S} complexes.

SEC purification of the shelterin^{TIN2S} complex **(A)** containing all six-subunits, or lacking **(B)** TRF1, **(C)** TRF2/RAP1, **(D)** TPP1/POT1. The affinity-purified shelterin complexes were injected into a Superose 6 Increase 10/300 GL column and eluted in 200 µL fractions. Fractions across the SEC elutions were analysed by SDS-PAGE. * Contaminants (see section 2.1.2).

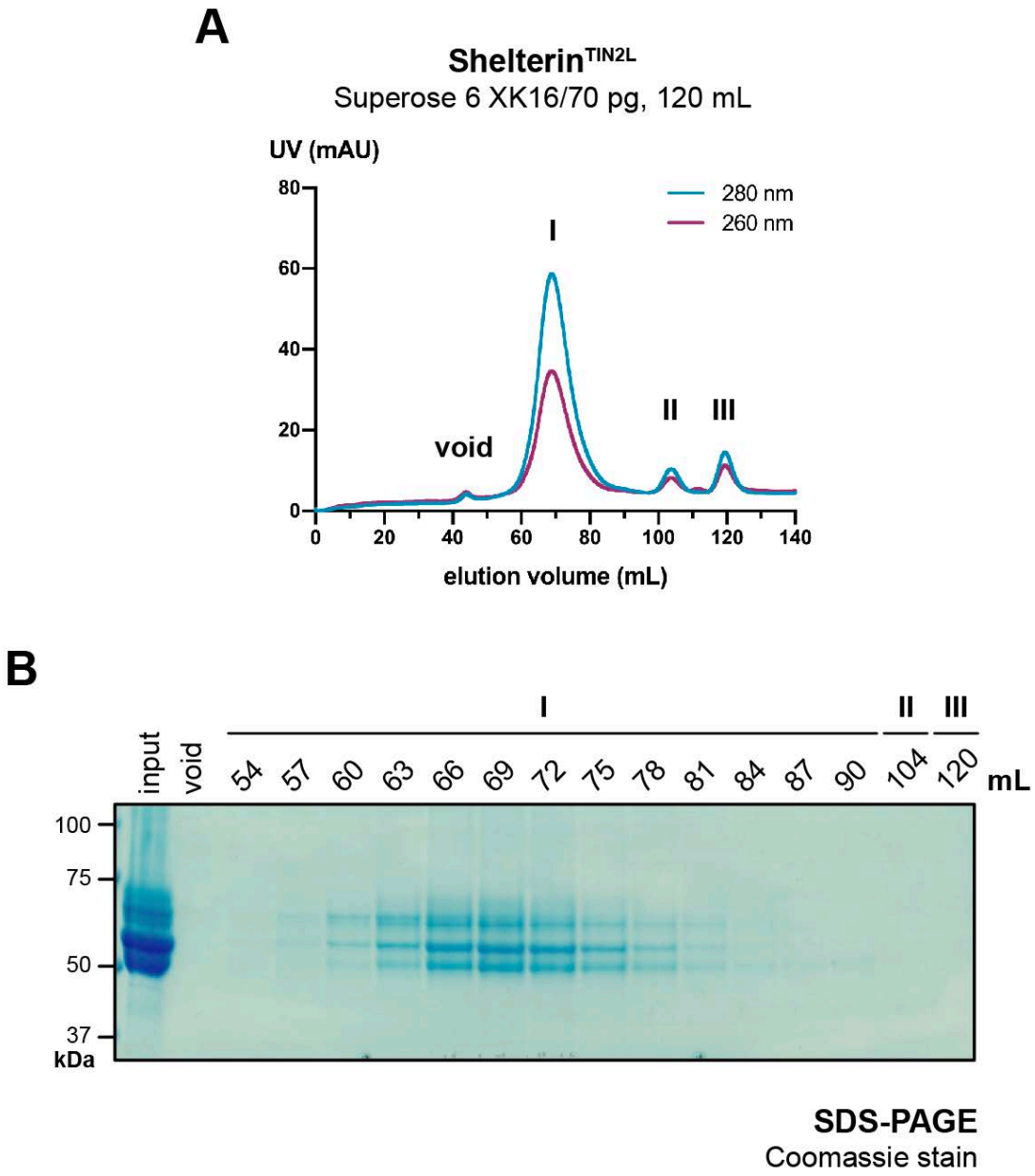

Figure 7.3. Elution profile of the full shelterin complex (Superose 6 XK 16/70 pg).

(A) Shelterin affinity elution fractions were pooled, and 6.42 mg of protein was incubated with TEV protease overnight (32:1 ratio) at 4˚C, centrifuged at 15k x g to remove any precipitate, and injected into a Superose 6 XK 16/70 pg 120 mL column and eluted at 0.4 mL/min. **(B)** SDS-PAGE analysis of SEC elution fractions using 4-15% Tris-Glycine precast gel.

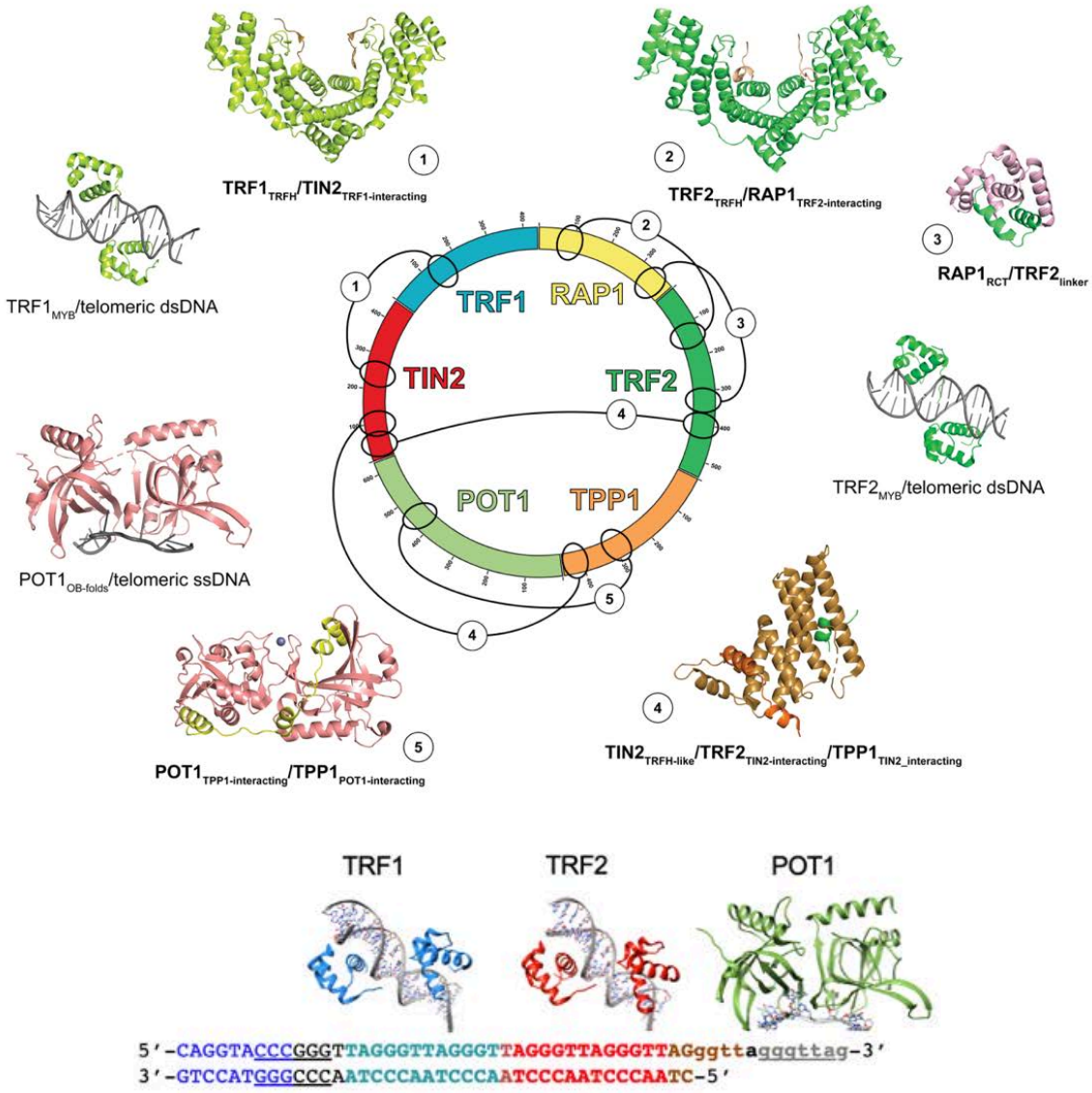

Figure 7.4. Existing crystal structures capturing protein-protein and protein-DNA interactions within the full shelterin complex.

Crystal structures illustrating the known protein-protein interactions involved in assembly of the full shelterin^{TIN2L} complex, and the protein-DNA interactions that recruit it to telomeres. PDB IDs (clockwise from interaction pair 1): 3BQO, 4RQI, 3K6G, 1W0U, 5XYF, 5UN7, 1XJV, and 1W0T.

The telomeric DNA sequence used for reconstitutions of DNA-bound shelterin complexes. TeloDNA1 has 4.5 telomeric dsDNA repeats for binding of TRF1 and TRF2 Myb domains, and the minimal ssDNA region optimised for POT1 binding.

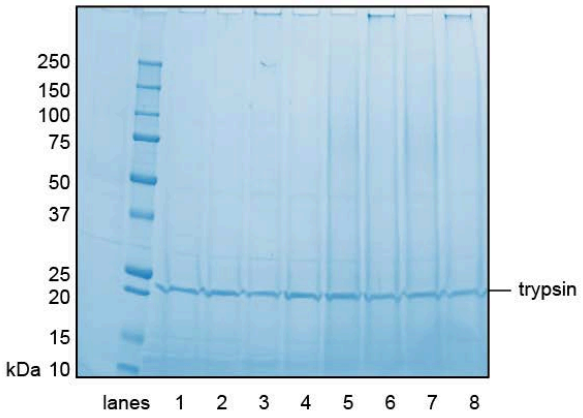

Figure 7.5. Trypsin digestion of apo shelterin complex for XL-MS analysis.

SDS-PAGE analysis of the apo shelterin^{TIN2L} complex after overnight trypsin digestion. SEC-purified shelterin^{TIN2L} complex was incubated with DMSO (control, lanes 1-4) or 500-fold molar excess of DSSO at room temperature for 10 mins (lanes 5-6) or 30 mins (lanes 7-8).

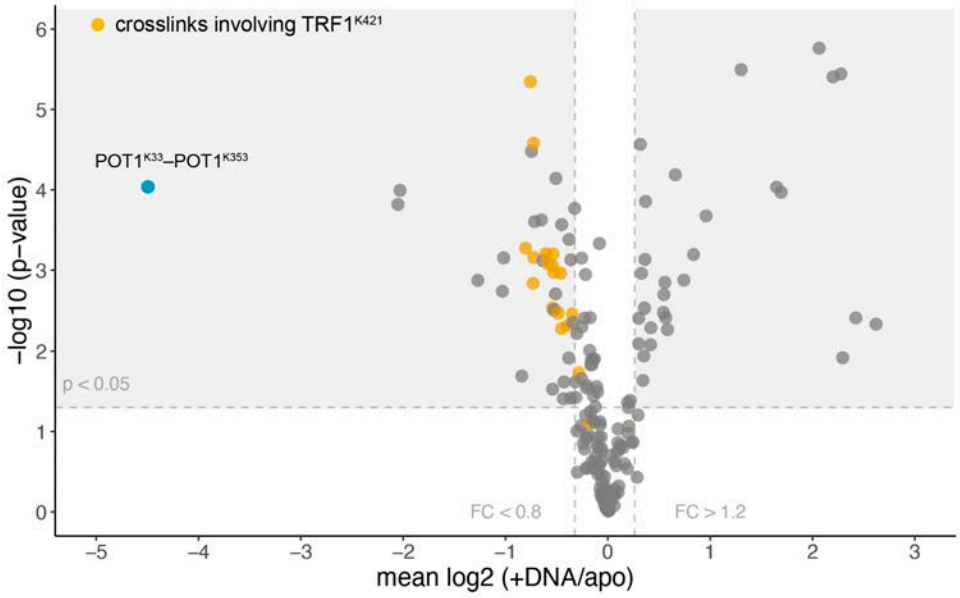

Figure 7.6. Comparing the abundance of crosslinks between the apo and DNA-bound shelterin complex.

Changes in abundance of crosslinks in the DNA-binding domains of TRF1 and POT1. Volcano plot shows the mean log2(+DNA/apo), which compares the abundance of crosslinks detected in the DNA-bound vs. apo shelterin^{TIN2L} complex (vertical dashed lines indicating thresholds of < 0.8 or > 1.2-fold-change), and the -log10(p-value) (above horizontal dashed line: $p < 0.05$).

Excised band	Accession	Description	Score	Coverage (%)	# PSMs	# Unique Peptide	MW (kDa)
50 kDa	A0A2H1WYZ0	Tubulin alpha chain GN=SFRICE_029378	178.36	62	162	5	49.9
	A0A2H1WFF5	Tubulin alpha chain GN=SFRICE_041750	161.79	55	147	2	49.8
	A0A2H1WFD3	Tubulin beta chain GN=SFRICE_019068	112.73	54	87	1	49.5
	A0A2H1W1I1	Tubulin beta chain (Fragment) GN=SFRICE_032934	110.72	72	83	2	34.4
	A0A2H1WFA2	Heat shock 70 kDa protein cognate 4 (Fragment) GN=SFRICE003522.2	72.88	38	35	16	71.5
70 kDa	A0A2H1WFA2	Heat shock 70 kDa protein cognate 4 (Fragment) GN=SFRICE003522.2	378.49	69	230	32	71.5
	A0A5Q0TYX6	Heat shock 70 kDa protein cognate 5	156.95	45	77	26	74.8
	A0A7S6RMB6	Heat shock cognate 70 protein GN=HSC70	143.28	46	76	20	73.1
	A0A7G9U6I8	Heat shock protein GN=Hsp70C	89.42	35	47	6	75.6
	A0A5Q0TYW7	Heat shock protein 68-like transcript variant 1	79.91	36	46	5	69.8

Table 7.1. Identification of contaminants in purification of shelterin complexes by mass spectrometry.

Affinity-purified shelterin^{TIN2L} was analysed by SDS-PAGE and Coomassie staining. The bands containing the contaminants near the 50 kDa and 75 kDa protein markers (see Figure 2.1C) were excised and subjected to trypsin digestion and peptide identification by mass spectrometry. The contaminants from Sf9 cells near the 50 kDa and 75 kDa markers were Tubulin α/β chains and HSP70 proteins, respectively.

Accession	Description	Score	Coverage (%)	# Unique Peptides	# PSMs	MW (kDa)
Q15554	Telomeric repeat-binding factor 2 [TERF2_HUMAN]	672.52	64.76	36	557	59.6
Q9BSI4	TERF1-interacting nuclear factor 2 [TINF2_HUMAN]	1436.6	77.61	30	1142	50.0
Q9NUX5	Protection of telomeres protein 1 [POTE1_HUMAN]	409.72	70.35	30	311	71.4
Q96AP0	Adrenocortical dysplasia protein homolog [ACD_HUMAN]	560.18	72.98	28	378	57.7
Q9NYB0	Telomeric repeat-binding factor 2-interacting protein 1 [TE2IP_HUMAN]	540.2	62.41	17	286	44.2

Table 7.2. Identification of shelterin subunits in recombinant shelterin (–TRF1)^{TIN2L} by mass spectrometry.

Affinity-purified apo shelterin (–TRF1)^{TIN2L} was subjected to trypsin digestion and peptide identification by mass spectrometry. No peptides were identified for TRF1.

8 References

Abreu, E. et al., 2010. TIN2-tethered TPP1 recruits human telomerase to telomeres in vivo. *Molecular and Cellular Biology,* 30(12), pp. 2971-2982.

Adrian, M., Dubochet, J., Lepault, J. & McDowall, A., 1984. Cryo- electron microscopy of viruses.. *Nature,* Volume 308, p. 32–36.

Ahel, I. et al., 2008. Poly(ADP-ribose)-binding zinc finger motifs in DNA repair/checkpoint proteins.. *Nature,* 451(7174), p. 81–85.

Alter, B. P., Giri, N., Savage, S. A. & Rosenberg, P. S., 2009. Cancer in dyskeratosis congenita. *Blood,* p. 6549--6557.

Aravind, L. et al., 2014. The Natural History of ADP-Ribosyltransferases and the ADP-Ribosylation System. In: *Endogenous ADP-Ribosylation.* s.l.:Springer International Publishing, p. 3--32.

Arques, O. et al., 2015. Tankyrase Inhibition Blocks Wnt/β-Catenin Pathway and Reverts Resistance to PI3K and AKT Inhibitors in the Treatment of Colorectal Cancer. *Clinical Cancer Research,* 22(3), p. 644--656.

Assaiya, A., Burada, A., Dhingra, S. & Kumar, J., 2021. An overview of the recent advances in cryo-electron microscopy for life sciences. *Emerg Top Life Sci.,* 5(1), p. 151-168.

Awasthi, P., Foiani, M. & Kumar, A., 2015. ATM and ATR signaling at a glance. *J Cell Sci,* p. 4255-4262.

Azad, G. K. & Tomar, R. S., 2016. The multifunctional transcription factor Rap1 a regulator of yeast physiology. *Frontiers in Bioscience,* 21(5), p. 918-930.

Bae, N. & Baumann, P., 2007. A RAP1/TRF2 complex inhibits nonhomologous end-joining at human telomeric DNA ends. *Molecular Cell,* 26(3), p. 323-334.

Bandaria, J. et al., 2016. Shelterin protects chromosome ends by compacting telomeric chromatin. *Cell,* 164(4), p. 735-746.

Barkauskaite, E., Jankevicius, G. & Ahel, I., 2015. Structures and Mechanisms of Enzymes Employed in the Synthesis and Degradation of PARP-Dependent Protein ADP-Ribosylation. *Molecular Cell,* 58(6), p. 935--946.

Baumann, P. & Cech, T. R., 2001. Pot1, the Putative Telomere End-Binding Protein in Fission Yeast and Humans. *Science,* p. 1171--1175.

Baykal, C., Kavak, A., Gulcan, P. & Buyukbabani, N., 2003. Dyskeratosis congenita associated with three malignancies. *Journal of the European Academy of Dermatology and Venereology,* p. 216--218.

Beattie, T. L., Zhou, W., Robinson, M. O. & Harrington, L., 2001. Functional Multimerization of the Human Telomerase Reverse Transcriptase. *Molecular and Cellular Biology,* 21(18), p. 6151--6160.

Benarroch-Popivker, D. et al., 2016. TRF2-Mediated Control of Telomere DNA Topology as a Mechanism for Chromosome-End Protection. *Molecular Cell,* 61(2), p. 274-286.

Bepler, T. et al., 2019. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods,* Volume 16, p. 1153–1160.

Bianchi, A. et al., 1997. TRF1 is a dimer and bends telomeric DNA. *The EMBO journal,* p. 1785--1794.

Bianchi, A. et al., 1999. TRF1 binds a bipartite telomeric site with extreme spatial flexibility.. *The EMBO Journal,* 18(20), p. 5735-5744.

Bieniossek, C., Richmond, T. & Berger, I., 2008. MultiBac: multigene baculovirus-based eukaryotic protein complex production. *Current Protocols in Protein Science,* 51(1), p. 5.20.1-5.20.26.

Bilaud, T. et al., 1997. Telomeric localization of TRF2, a novel human telobox protein. *Nature Genetics,* 17(2), p. 236--239.

Bilaud, T. et al., 1996. The Telobox, a Myb-Related Telomeric DNA Binding Motif Found in Proteins from Yeast, Plants and Human. *Nucleic Acids Research,* 24(7), p. 1294--1303.

Blackburn, E. H., 1991. Structure and function of telomeres. *Nature,* p. 569--573.

Blackburn, E. H. & Gall, J. G., 1978. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *Journal of Molecular Biology,* p. 33--53.

Blackford, A. & Jackson, S., 2017. ATM, ATR, and DNA-PK: The Trinity at the heart of the DNA damage response. *Molecular Cell,* 66(6), p. 801-817.

Boskovic, J. et al., 2016. Molecular Architecture of Full-length TRF1 Favors Its Interaction with DNA. *Journal of Biological Chemistry,* p. 21829--21835.

Broccoli, D., Smogorzewska, A., Chong, L. & Lange, T. d., 1997. Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nature Genetics,* 17(2), p. 231--235.

Buchman, A., Lue, N. & Kornberg, R., 1988. Connections between transcriptional activators, silencers, and telomeres as revealed by functional analysis of a yeast DNA-binding protein. *Molecular and Cellular Biology,* 8(12), p. 5086-5099.

Canela, A., Vera, E., Klatt, P. & Blasco, M. A., 2007. High-throughput telomere length quantification by FISH and its application to human population studies.. *Proceedings of the National Academy of Sciences,* 104(13), p. 5300–5305.

Canudas, S. et al., 2007. Protein requirements for sister telomere association in human cells. *The EMBO Journal,* 26(23), p. 4867--4878.

Cao, K. et al., 2011. Progerin and telomere dysfunction collaborate to trigger cellular senescence in normal human fibroblasts. *Journal of Clinical Investigation,* 121(7), p. 2833--2844.

Carrel, A., 1912. On the permanent life of tissues outside of the organism.. *The Journal of Experimental Medicine,* 15(5), p. 516--528.

Carrel, A. & Ebeling, A. H., 1921. Age and multiplication of fibroblasts. *The Journal of Experimental Medicine,* 34(6), p. 599--623.

Celli, G. & de Lange, T., 2005. DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion. *Nature Cell Biology,* 7(7), p. 712-718.

Chai, W., Du, Q., Shay, J. W. & Wright, W. E., 2006. Human Telomeres Have Different Overhang Sizes at Leading versus Lagging Strands. *Molecular Cell,* 21(3), p. 427--435.

Chang, P., Coughlin, M. & & Mitchison, T. J., 2005. Tankyrase-1 polymerization of poly(ADP-ribose) is required for spindle structure and function.. *Nature Cell Biology,* 7(11), p. 1133--1139.

Chang, W., Dynek, J. N. & Smith, S., 2003. TRF1 is degraded by ubiquitin-mediated proteolysis after release from telomeres. *Genes & Development,* 17(11), p. 1328-1333.

Chang, W., Dynek, J. N. & Smith, S., 2005. NuMA is a major acceptor of poly(ADP-ribosyl)ation by tankyrase 1 in mitosis. *Biochemical Journal,* 391(2), p. 177--184.

Chen, C. et al., 2017. Structural insights into POT1-TPP1 interaction and POT1 C-terminal mutations in human cancer. *Nature Communications,* 8(1), p. 1-15.

Chen, D. et al., 2011. Identification of Macrodomain Proteins as Novel O-Acetyl-ADP-ribose Deacetylases. *Journal of Biological Chemistry,* 286(15), p. 13261--13271.

Chen, J. L. & Greider, C. W., 2004. Telomerase RNA structure and function: implications for dyskeratosis congenita.. *Trends in Biochemical Sciences,* 29(4), p. 183–192.

Chen, Y., 2019. The structural biology of the shelterin complex. *Biological Chemistry,* 400(4), p. 457--466.

Chen, Y. et al., 2011. A conserved motif within RAP1 has diversified roles in telomere protection and regulation in different organisms. *Nature Structural & Molecular Biology,* 18(2), p. 213-221.

Chen, Y. et al., 2008. A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins. *Science,* 319(5866), p. 1092-1096.

Chiang, Y. J. et al., 2008. Tankyrase 1 and tankyrase 2 are essential but redundant for mouse embryonic development. *PLoS ONE,* p. e2639.

Chiang, Y. J. et al., 2004. Telomere-associated protein TIN2 is essential for early embryonic development through a telomerase-independent pathway. *Molecular and Cellular Biology,* 24(15), p. 6631-6634.

Chi, N.-W. & Lodish, H. F., 2000. Tankyrase Is a Golgi-associated Mitogen-activated Protein Kinase Substrate That Interacts with IRAP in GLUT4 Vesicles. *Journal of Biological Chemistry,* 275(49), p. 38437--38444.

Chong, L. et al., 1995. A Human Telomeric Protein. *Science,* 270(5242), p. 1663--1667.

Cho-Park, P. & Steller, H., 2013. Proteasome Regulation by ADP-Ribosylation. *Cell,* 153(3), p. 614--627.

Cleal, K., Norris, K. & Baird, D., 2018. Telomere length dynamics and the evolution of cancer genome architecture. *International journal of molecular sciences,* p. 482.

Cohen, M. S. & Chang, P., 2018. Insights into the biogenesis, function, and regulation of ADP-ribosylation. *Nature Chemical Biology,* 14(3), p. 236--243.

Cole, H. N., 1930. Dyskeratosis Congenita with Pigmentation, Dystrophia Unguium, and Leukoplakia Oris. *Archives of Dermatology,* p. 71.

Collier, R., 2001. Understanding the mode of action of diphtheria toxin: a perspective on progress during the 20th century. *Toxicon,* 39(11), p. 1793--1803.

Collins, K. & Gandhi, L., 1998. The reverse transcriptase component of the Tetrahymena telomerase ribonucleoprotein complex. *Proceedings of the National Academy of Sciences,* 95(15), p. 8485--8490.

Conrad, M. N., Wright, J. H., Wolf, A. J. & Zakian, V. A., 1990. RAP1 protein interacts with yeast telomeres in vivo: Overproduction alters telomere structure and decreases chromosome stability. *Cell,* 63(4), p. 739-750.

Cook, B. D. et al., 2002. Role for the Related Poly(ADP-Ribose) Polymerases Tankyrase 1 and 2 at Human Telomeres. *Molecular and Cellular Biology,* 22(1), p. 332--342.

Coren, J. S. & Vogt, V. M., 1992. Purification of a telomere-binding protein from Physarum polycephalum.. *Biochimica et Biophysica Acta - Gene Structure and Expression,* 1171(2), p. 162–166.

Court, R. et al., 2005. How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: A view from high-resolution Crystal Structures. *EMBO reports,* 6(1), p. 39-45.

Court, R., Chapman, L., Fairall, L. & Rhodes, D., 2005. How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures.. *EMBO Reports,* 6(1), p. 39–45.

Crabbe, L., 2004. Defective Telomere Lagging Strand Synthesis in Cells Lacking WRN Helicase Activity. *Science,* 306(5703), p. 1951--1953.

Cristofari, G. et al., 2007. Human Telomerase RNA Accumulation in Cajal Bodies Facilitates Telomerase Recruitment to Telomeres and Telomere Elongation. *Molecular Cell,* 27(6), p. 882–889.

Crooks, G. E., 2004. WebLogo: A Sequence Logo Generator. *Genome Research,* p. 1188--1190.

Cuatrecasas, P. & Parikh, I., 1972. Adsorbents for affinity chromatography. Use of N-hydroxysuccinimide esters of agarose. *Biochemistry,* 11(12), p. 2291–2299.

Dai, X. et al., 2010. Molecular steps of G-overhang generation at human telomeres and its function in chromosome end protection. *The EMBO Journal,* 29(16), p. 2788--2801.

DaRosa, P. A., Klevit, R. E. & Xu, W., 2018. Structural basis for tankyrase-RNF146 interaction reveals noncanonical tankyrase-binding motifs. *Protein Science,* 27(6), p. 1057--1067.

de Lange, T., 2004. T-loops and the origin of telomeres. *Nature Reviews Molecular Cell Biology,* 5(4), p. 323--329.

de Lange, T., 2005. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes & Development,* p. 2100--2110.

de Lange, T., 2018. Shelterin-Mediated Telomere Protection. *Annual Review of Genetics,* 52(1), p. 223--247.

Denchi, E. & de Lange, T., 2007. Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. *Nature,* 448(7157), p. 1068-1071.

Denchi, E. L. & de Lange, T., 2007. Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. *Nature,* p. 1068--1071.

Deng, Y., Guo, X., Ferguson, D. & Chang, S., 2009. Multiple roles for mre11 at Uncapped Telomeres. *Nature,* 460(7257), p. 914-918.

Dokal, I., 2000. Dyskeratosis congenita in all its forms. *British Journal of Haematology,* p. 768--779.

Doksani, Y., 2019. The Response to DNA Damage at Telomeric Repeats and Its Consequences for Telomere Function. *Genes,* p. 318.

Doksani, Y. & de Lange, T., 2016. Telomere-internal double-strand breaks are repaired by homologous recombination and PARP1/Lig3-dependent end-joining. *Cell Reports,* 17(6), p. 1646-1656.

Doksani, Y., Wu, J. Y., de Lange, T. & Zhuang, X., 2013. Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent T-loop formation. *Cell,* 155(2), p. 345-356.

Donigian, J. R. & de Lange, T., 2007. The Role of the Poly(ADP-ribose) Polymerase Tankyrase1 in Telomere Length Control by the TRF1 Component of the Shelterin Complex.. *Journal of Biological Chemistry,* 282(31), p. 22662–22667.

Dregalla, R. C. et al., 2010. Regulatory roles of tankyrase 1 at telomeres and in DNA repair: suppression of T-SCE and stabilization of DNA-PKcs. *Aging,* 2(10), p. 691--708.

Du, X. et al., 2004. Telomere Shortening Exposes Functions for the Mouse Werner and Bloom Syndrome Genes. *Molecular and Cellular Biology,* 24(19), p. 8437--8446.

Dynek, J. N. & Smith, S., 2004. Resolution of sister telomere association is required for progression through mitosis. *Science,* p. 97--100.

Edelhoch, H., 1967. Spectroscopic determination of tryptophan and tyrosine in proteins.. *Biochemistry,* Volume 6, p. 1948-1954.

Edwards, D. N., Orren, D. K. & Machwe, A., 2014. Strand exchange of telomeric DNA catalyzed by the Werner syndrome protein (WRN) is specifically stimulated by TRF2. *Nucleic Acids Research,* 42(12), p. 7748--7761.

Egan, E. & Collins, K., 2012. An enhanced H/ACA RNP assembly mechanism for human telomerase RNA. *Molecular and Cellular Biology,* 32(13), p. 2428-2439.

Eisemann, T. et al., 2016. Tankyrase-1 Ankyrin Repeats Form an Adaptable Binding Platform for Targets of ADP-Ribose Modification. *Structure,* 24(10), p. 1679--1692.

Erdel, F. et al., 2017. elomere Recognition and Assembly Mechanism of Mammalian Shelterin. *Cell Reports,* 18(1), p. 41--53.

Fairall, L. et al., 2001. Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2. *Molecular Cell,* 8(2), p. 351–361.

Fang, G., Gray, J. T. & Cech, T. R., 1993. Oxytricha telomere-binding protein: separable DNA-binding and dimerization domains of the alpha-subunit.. *Genes & Development,* 7(5), p. 870--882.

Forney, J., Henderson, E. R. & Blackburn, E. H., 1987. Identification of the telomeric sequence of the acellular slime molds Didymium iridis and Physarum potycephalum. *Nucleic Acids Research,* 15(22), p. 9143--9152.

Frescas, D. & de Lange, T., 2014. TRF2-tethered TIN2 can mediate telomere protection by TPP1/POT1. *Molecular and Cellular Biology,* 34(7), p. 1349-1362.

Fumagalli, M. et al., 2012. Telomeric DNA damage is irreparable and causes persistent DNA-damage-response activation. *Nature cell biology,* p. 355-365.

Gaiser, O. J. et al., 2004. Solution structure, backbone dynamics, and association behavior of the C-terminal BRCT domain from the breast cancer-associated protein BRCA1. *Biochemistry,* 43(51), p. 15983-15995.

Garten, A. et al., 2009. Nampt: linking NAD biology, metabolism and cancer. *Trends in Endocrinology & Metabolism,* 20(3), p. 130--138.

Gasteiger, E. et al., 2005. Protein Identification and Analysis Tools on the ExPASy Server. In: *The Proteomics Protocols Handbook.* -: Humana Press, p. 571--607.

Gaullier, G. et al., 2016. A higher-order entity formed by the flexible assembly of RAP1 with TRF2.. *Nucleic Acids Research,* 44(4), p. 1962–1976.

Ghanim, G. et al., 2021. Structure of human telomerase holoenzyme with bound telomeric DNA. *Nature,* 593(7859), p. 449-453.

Gillis, A. J., Schuller, A. P. & Skordalakes, E., 2008. Structure of the Tribolium castaneum telomerase catalytic subunit TERT. *Nature,* 455(7213), p. 633--637.

Giraldo, R. & Rhodes, D., 1994. The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *The EMBO journal,* p. 2411--2420.

Giraldo, R., Suzuki, M., Chapman, L. & Rhodes, D., 1994. Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: a circular dichroism study. *Proceedings of the National Academy of Sciences,* p. 7658--7662.

Goodsell, D. S., Autin, L. & Olson, A. J., 2019. Illustrate: Software for Biomolecular Illustration. *Structure,* 27(11), p. 1716--1720.e1.

Gottschling, D. E. & Cech, T. R., 1984. Chromatin structure of the molecular ends of Oxytricha macronuclear DNA: phased nucleosomes and a telomeric complex. *Cell,* p. 501--510.

Gottschling, D. E. & Zakian, V. A., 1986. Telomere proteins: Specific recognition and protection of the natural termini of Oxytricha macronuclear DNA. *Cell,* 47(2), p. 195--205.

Grant, T., Rohou, A. & Grigorieff, N., 2018. cisTEM, user-friendly software for single-particle image processing. *eLife,* Volume 7, p. e35383.

Graziadei, A. & Rappsilber, J., 2022. Leveraging crosslinking mass spectrometry in structural and Cell Biology. *Structure,* 30(1), p. 37-54.

Greider, C. W. & Blackburn, E. H., 1985. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell,* p. 405--413.

Greider, C. W. & Blackburn, E. H., 1987. The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell,* p. 887--898.

Griffith, J., Bianchi, A. & de Lange, T., 1998. TRF1 promotes parallel pairing of telomeric tracts in vitro. *Journal of molecular biology,* 278(1), p. 79-88.

Griffith, J. et al., 1999. Mammalian telomeres end in a large duplex loop. *Cell,* 97(4), p. 503-514.

Grill, S., Tesmer, V. M. & Nandakumar, J., 2018. The N terminus of the OB domain of telomere protein TPP1 is critical for telomerase action. *Cell Reports,* 22(5), p. 1132-1140.

Grimm, M., Zimniak, T., Kahraman, A. & Herzog, F., 2015. xVis: a web server for the schematic visualization and interpretation of crosslink-derived spatial restraints.. *Nucleic Acids Research,* 43(W1), p. W362–W369.

Guettler, S. et al., 2011. Structural Basis and Sequence Rules for Substrate Recognition by Tankyrase Explain the Basis for Cherubism Disease. *Cell,* 147(6), p. 1340--1354.

Gutierrez, C. et al., 2016. Developing an Acidic Residue Reactive and Sulfoxide-Containing MS-Cleavable Homobifunctional Cross-Linker for Probing Protein-Protein Interactions. *Analytical Chemistry,* 88(16), p. 8315–8322.

Haider, S. et al., 2018. Holliday Junctions Formed from Human Telomeric DNA. *Journal of the American Chemical Society,* p. 15366-15374.

Haikarainen, T. et al., 2013. Structural Basis and Selectivity of Tankyrase Inhibition by a Wnt Signaling Inhibitor WIKI4. *PLoS ONE,* 8(6), p. e65404.

Hanahan, D. & Weinberg, R. A., 2011. Hallmarks of cancer: the next generation. *Cell,* p. 646-674.

Hanaoka, S., Nagadoi, A. & Nishimura, Y., 2005. Comparison between TRF2 and TRF1 of their telomeric DNA-bound structures and DNA-binding activities.. *Protein Sci.,* Volume 14, p. 119–130.

Hanaoka, S. et al., 2001. NMR structure of the hrap1 myb motif reveals a canonical three-helix bundle lacking the positive surface charge typical of myb DNA-binding domains. *Journal of Molecular Biology,* 312(1), p. 167–175.

Harkisheimer, M., Mason, M., Shuvaeva, E. & Skordalakes, E., 2013. A Motif in the Vertebrate Telomerase N-Terminal Linker of TERT Contributes to RNA Binding and Telomerase Activity and Processivity.. *Structure,* 21(10), p. 1870--1878.

Harley, C. B., 1991. Telomere loss: mitotic clock or genetic time bomb?. *Mutation Research/DNAging,* Volume 256, p. 271--282.

Harley, C., Futcher, A. & Greider, C., 1990. Telomeres shorten during ageing of human fibroblasts. *Nature,* 345(6274), p. 458-460.

Haupt, C. et al., 2017. Combining Chemical Cross-linking and Mass Spectrometry of Intact Protein Complexes to Study the Architecture of Multi-subunit Protein Assemblies. *Journal of Visualized Experiments,* Issue 129, p. e56747.

Hayflick, L., 1965. The limited in vitro lifetime of human diploid cell strains. *Experimental Cell Research,* p. 614--636.

Hayflick, L. & Moorhead, P. S., 1961. The serial cultivation of human diploid cell strains. *Experimental cell research,* p. 585--621.

Henderson, E. R. & Larson, D. D., 1991. Telomeres — what's new at the end?. *Current Opinion in Genetics & Development,* p. 538 - 543.

Hockemeyer, D. & Collins, K., 2015. Control of telomerase action at human telomeres.. *Nature Structural & Molecular Biology,* 22(11), p. 848–852.

Hockemeyer, D., Daniels, J.-P., Takai, H. & deLange, T., 2006. Recent Expansion of the Telomeric Complex in Rodents: Two Distinct POT1 Proteins Protect Mouse Telomeres. *Cell,* 126(1), p. 63--77.

Hockemeyer, D. et al., 2005. POT1 protects telomeres from a transient DNA damage response and determines how human chromosomes end.. *The EMBO Journal,* 24(14), p. 2667–2678.

Hoffman, H., Rice, C. & Skordalakes, E., 2017. Structural Analysis Reveals the Deleterious Effects of Telomerase Mutations in Bone Marrow Failure Syndromes. *Journal of Biological Chemistry,* 292(11), p. 4593--4601.

Horvath, M. P. et al., 1998. Crystal Structure of the Oxytricha nova Telomere End Binding Protein Complexed with Single Strand DNA. *Cell,* 95(7), p. 963--974.

Hottiger, M. O. et al., 2010. Toward a unified nomenclature for mammalian ADP-ribosyltransferases.. *Trends in Biochemical Sciences,* 35(4), p. 208–219.

Houghtaling, B. R., Cuttonaro, L., Chang, W. & Smith, S., 2004. A dynamic molecular link between the telomere length regulator TRF1 and the chromosome end protector TRF2. *Current Biology,* 14(18), p. 1621-1631.

Huang, J. et al., 2014. Structural basis for protein-RNA recognition in telomerase. *Nature Structural & Molecular Biology,* 21(6), p. 507--512.

Huang, S.-M.A. et al., 2009. Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling.. *Nature,* 461(7264), p. 614–620.

Hu, C. et al., 2017. Structural and functional analyses of the mammalian TIN2-TPP1-TRF2 telomeric complex. *Cell Research,* p. 1485--1502.

Hud, N. & Plavec, J., 2006. The role of cations in determining quadruplex structure and stability. In: S. Neidle & S. Balasubramanian, eds. *Quadruplex Nucleic Acids..* Cambridge: Royal Society of Chemistry, p. 100–130.

Hustedt, N. & Durocher, D., 2016. The control of DNA repair by the cell cycle. *Nature Cell Biology,* 19(1), p. 1-9.

Hylton, R. & Swulius, M., 2021. Challenges and triumphs in Cryo-Electron Tomography. *iScience,* 24(9), p. 102959.

Iacobucci, C. et al., 2018. A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein–protein interactions.. *Nature Protocols,* 13(12), p. 2864–2889.

James, R. G. et al., 2012. WIKI4, a Novel Inhibitor of Tankyrase and Wnt/β-Catenin Signaling. *PLoS ONE,* 7(12), p. e50457.

Jankevicius, G. et al., 2013. A family of macrodomain proteins reverses cellular mono-ADP-ribosylation. *Nature Structural & Molecular Biology,* 20(4), p. 508--514.

Janovič, T. et al., 2019. Human telomere repeat binding factor TRF1 replaces TRF2 bound to Shelterin core hub TIN2 when TPP1 is absent. *Journal of Molecular Biology,* 431(17), p. 3289-3301.

Jeong, S. A. et al., 2015. Akt-mediated phosphorylation increases the binding affinity of hTERT for importin α to promote nuclear translocation. *Journal of Cell Science,* Volume 128, p. 2287-2301.

Jiang, J. et al., 2018. Structure of Telomerase with Telomeric DNA.. *Cell,* 173(5), p. 1179–1190.e13.

König, P., Giraldo, R., Chapman, L. & Rhodes, D., 1996. The Crystal Structure of the DNA-Binding Domain of Yeast RAP1 in Complex with Telomeric DNA. *Cell,* p. 125--136.

Kühlbrandt, W., 2014. The Resolution Revolution. *Science,* 343(6178), p. 1443-1444.

Kabir, S., Hockemeyer, D. & de Lange, T., 2014. TALEN gene knockouts reveal no requirement for the conserved human shelterin protein Rap1 in telomere protection and length regulation.. *Cell Reports,* 9(4), p. 1273-1273.

Kaiser, W. & Rant, U., 2010. Conformations of End-Tethered DNA Molecules on Gold Surfaces: Influences of Applied Electric Potential, Electrolyte Screening, and Temperature. *Journal of the American Chemical Society,* 132(23), p. 7935--7945.

Kaminker, P. G. et al., 2001. TANK2, a New TRF1-associated Poly(ADP-ribose) Polymerase, Causes Rapid Induction of Cell Death upon Overexpression.. *Journal of Biological Chemistry,* 276(38), p. 35891–35899.

Kaminker, P., Kim, S.-H., Desprez, P.-Y. & Campisi, J., 2009. A novel form of the telomere-associated protein TIN2 localizes to the nuclear matrix. *Cell Cycle,* 8(6), p. 931-939.

Kao, A. et al., 2010. Development of a Novel Cross-linking Strategy for Fast and Accurate Identification of Cross-linked Peptides of Protein Complexes.. *Molecular & Cellular Proteomics,* 10(1), p. M110.002212.

Kao, A. et al., 2011. Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Molecular & Cellular Proteomics,* 10(1), p. M110.002212.

Kao, A. et al., 2011. Development of a novel cross-linking strategy for fast and accurate identification of cross-linked peptides of protein complexes. *Molecular & Cellular Proteomics,* 1-(1), p. M110.002212.

Karlberg, T. et al., 2010. Structural Basis for the Interaction between Tankyrase-2 and a Potent Wnt-Signaling Inhibitor. *Journal of Medicinal Chemistry,* 53(14), p. 5352-5355.

Karlseder, J. et al., 1999. P53- and ATM-dependent apoptosis induced by telomeres lacking TRF2. *Science,* 283(5406), p. 1321-1325.

Karlseder, J. et al., 2004. The telomeric protein TRF2 binds the ATM kinase and can inhibit the ATM-dependent DNA damage response. *PLoS Biology,* 2(8), p. e240.

Karlseder, J. et al., 2003. Targeted deletion reveals an essential function for the telomere length regulator Trf1. *Molecular and Cellular Biology,* 23(18), p. 6533-6541.

Karlseder, J., Smogorzewska, A. & Lange, T. d., 2002. Senescence Induced by Altered Telomere State, Not Telomere Loss. *Science,* 295(5564), p. 2446--2449.

Kastner, B. et al., 2008. GraFix: sample preparation for single-particle electron cryomicroscopy. *Nature Methods,* 5(1), p. 53-55.

Kibe, T., Zimmermann, M. & de Lange, T., 2016. TPP1 blocks an ATR-mediated resection mechanism at telomeres. *Molecular Cell,* p. 236--246.

Kim, N. et al., 1994. Specific association of human telomerase activity with immortal cells and cancer. *Science,* 266(5193), p. 2011-2015.

Kim, S.-h.et al., 2004. TIN2 mediates functions of TRF2 at human telomeres. *Journal of Biological Chemistry,* 279(42), p. 43799-43804.

Kim, S.-h., Kaminker, P. & Campisi, J., 1999. TIN2, a new regulator of telomere length in human cells. *Nature Genetics,* 23(4), p. 405-412.

Kim, W. et al., 2016. Regulation of the human telomerase gene TERT by telomere position effect—over long distances (TPE-old): Implications for aging and cancer. *PLOS Biology,* 14(12), p. pbio.2000016.

Klobutcher, L. A., Swanton, M. T., Pierluigi, D. & Prescott, D. M., 1981. 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,* p. 3015-3019.

Knezevic, J. et al., 2012. Quantitation of Affinity, Avidity, and Binding Kinetics of Protein Analytes with a Dynamically Switchable Biosurface. *Journal of the American Chemical Society,* 132(37), p. 15225--15228.

Knijnenburg, T. A. et al., 2018. Genomic and molecular landscape of DNA damage repair deficiency across The Cancer Genome Atlas. *Cell Reports,* 23(1), p. 239--254.

Koenig, P., Giraldo, R., Chapman, L. & Rhodes, D., 1997. *DNA-binding domain of RAP1 in complex with telomeric DNA site.* [Online]
Available at: https://doi.org/10.2210%2Fpdb1ign%2Fpdb
[Accessed 22 05 2020].

Lai, C. K., Mitchell, J. R. & Collins, K., 2001. RNA Binding Domain of Telomerase Reverse Transcriptase. *Molecular and Cellular Biology,* 21(4), p. 990-1000.

Lam, Y. C. et al., 2010. SNMIB/Apollo protects leading-strand telomeres against NHEJ-mediated repair. *The EMBO Journal,* 29(13), p. 2230--2241.

Langer, A. et al., 2013. Protein analysis by time-resolved measurements with an electro-switchable DNA chip.. *Nature Communications,* 4(1), p. 1--8.

Langer, A. et al., 2015. Polymerase/DNA interactions and enzymatic activity: multi-parameter analysis with electro-switchable biosurfaces. *Scientific Reports,* 5(1), p. 12066.

Latrick, C. M. & Cech, T. R., 2010. POT1–TPP1 enhances telomerase processivity by slowing primer dissociation and aiding translocation. *The EMBO Journal,* 29(5), p. 924-933.

Lau, T. et al., 2013. A Novel Tankyrase Small-Molecule Inhibitor Suppresses APC Mutation-Driven Colorectal Tumor Growth. *Cancer Research,* 73(10), p. 3132--3144}.

Lehtiö, L. et al., 2008. Zinc Binding Catalytic Domain of Human Tankyrase 1. *Journal of Molecular Biology,* 379(1), p. 136-145.

Lei, M., Podell, E. R. & Cech, T. R., 2004. Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nature Structural & Molecular Biology,* Volume 11, p. 1223-1229.

Lei, M., Zaug, A., Podell, E. & Cech, T., 2005. Switching human telomerase on and off with hPOT1 protein in vitro. *Journal of Biological Chemistry,* 280(21), p. 20449-20456.

Leung, C. & Glover, J., 2011. BRCT domains. *Cell Cycle,* 10(15), p. 2461-2470.

Li, B. & de Lange, T., 2003. Rap1 affects the length and heterogeneity of human telomeres. *Molecular biology of the cell,* p. 5060--5068.

Li, B., Oestreich, S. & Lange, T. d., 2000. Identification of Human Rap1. *Cell,* 101(5), p. 471--483.

Li, B. et al., 2016. Crystal structure of a tankyrase 1/telomere repeat factor 1 complex. *Acta Crystallographica Section F Structural Biology Communications,* 72(4), p. 320--327.

Licht, J. D. & Collins, K., 1999. elomerase RNA function in recombinant Tetrahymena telomerase. *Genes & Development,* 13(9), p. 1116--1125.

Lieber, M. R., 2010. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual Review of Biochemistry,* p. 181-211.

Li, J. et al., 2005. Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. *Nucleic acids research,* p. 4649--4659.

Lim, C. J. & Cech, T. R., 2021. Shaping human telomeres: from shelterin and CST complexes to telomeric chromatin organization. *Nature Reviews Molecular Cell Biology,* 22(4), p. 283-298.

Lim, C. J., Zaug, A. J., Kim, H. J. & Cech, T. R., 2017. Reconstitution of human shelterin complexes reveals unexpected stoichiometry and dual pathways to enhance telomerase processivity. *Nature Communications,* 8(1075), p. 1-14.

Lingner, J. et al., 1997. Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science,* p. 561--567.

Lipps, H. J. & Rhodes, D., 2009. G-quadruplex structures: in vivo evidence and function. *Trends in cell biology,* p. 414--422.

Liu, B. et al., 2022. Structure of active human telomerase with telomere shelterin protein TPP1. *Nature,* 604(7906), p. 578-583.

Liu, D., O'Connor, M. S., Qin, J. & Songyang, Z., 2004b. Telosome, a Mammalian Telomere-associated Complex Formed by Multiple Telomeric Proteins. *Journal of Biological Chemistry,* 279(49), p. 51338-51342.

Liu, D. et al., 2004a. PTOP interacts with POT1 and regulates its localization to telomeres. *Nature Cell Biology,* 6(7), p. 673-680.

Liu, F., Rijkers, D. T. S., Post, H. & Heck, A. J. R., 2015. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nature Methods,* 12(12), p. 1179–1184.

Lizio, M. et al., 2018. Update of the FANTOM web resource: expansion to provide additional transcriptome atlases. *Nucleic Acids Research,* p. D752-D758.

Loayza, D. & de Lange, T., 2003. POT1 as a terminal transducer of TRF1 telomere length control. *Nature,* Volume 423, p. 1013-1018.

Loayza, D. et al., 2004. DNA binding features of human POT1. *Journal of Biological Chemistry,* 279(13), p. 13241-13248.

Lomant, A. & Fairbanks, G., 1976. Chemical probes of extended biological structures: Synthesis and properties of the cleavable protein cross-linking reagent [35S]dithiobis(succinimidyl propionate).. *Journal of Molecular Biology,* 104(1), p. 243–261.

Lonsdale, J. et al., 2013. The genotype-tissue expression (GTEx) project. *Nature genetics,* p. 580.

Lopes, J. et al., 2011. G-quadruplex-induced instability during leading-strand replication. *The EMBO Journal,* 30(19), p. 4033--4046.

Lue, N. F., 2005. A Physical and Functional Constituent of Telomerase Anchor Site. *Journal of Biological Chemistry,* 280(28), p. 26586--26591.

López-Otín, C. et al., 2013. The hallmarks of aging. *Cell,* 153(6), p. 1194-1217.

Makarov, V. L., Hirose, Y. & Langmore, J. P., 1997. Long G Tails at Both Ends of Human Chromosomes Suggest a C Strand Degradation Mechanism for Telomere Shortening. *Cell,* 88(5), p. 657--666.

Makovets, S., Herskowitz, I. & Blackburn, E. H., 2004. Anatomy and Dynamics of DNA Replication Fork Movement in Yeast Telomeric Regions.. *Molecular and Cellular Biology,* 24(9), p. 4019–4031.

Mariotti, L. et al., 2016. Tankyrase requires SAM domain-dependent polymerization to support Wnt-β-catenin signaling. *Molecular Cell,* 63(3), p. 498-513.

Masutomi, K. et al., 2000. Telomerase Activity Reconstituted in Vitro with Purified Human Telomerase Reverse Transcriptase and Human Telomerase RNA Componen. *Journal of Biological Chemistry,* 275(29), p. 22568--22573.

Matot, B. et al., 2011. The orientation of the C-terminal domain of the Saccharomyces cerevisiae Rap1 protein is determined by its binding to DNA. *Nucleic Acids Research,* p. 3197--3207.

Mattern, K. et al., 2004. Dynamics of protein binding to telomeres in living cells: Implications for telomere structure and function. *Molecular and Cellular Biology,* 24(12), p. 5587-5594.

McClintock, B., 1941. The stability of broken ends of chromosomes in Zea mays. *Genetics,* p. 234.

McCord, R. P. et al., 2012. Correlated alterations in genome organization, histone methylation, and DNA-lamin A/C interactions in Hutchinson-Gilford progeria syndrome. *Genome Research,* 23(2), p. 260--269.

McMullan, G., Faruqi, A., Clare, D. & Henderson, R., 2014. Comparison of optimal performance at 300 keV of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy,* Volume 147, p. 156–163.

McMullan, G., Faruqi, A. & Henderson, R., 2016. Direct Electron Detectors. *Methods in Enzymology,* Volume 579, p. 1-17.

Meselson, M. & Stahl, F., 1958. The replication of DNA in Escherichia coli.. *Proceedings of the National Academy of Sciences,* 44(7), p. 671-682.

Metcalfe, J. A. et al., 1996. Accelerated telomere shortening in ataxia telangiectasia. *Nature Genetics,* 13(3), p. 350--353.

Mie, G., 1908. Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen.. *Ann. Physik,,* Volume 25,, p. 377– 445.

Mitchell, M. et al., 2010. Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA. *Nature Structural & Molecular Biology,* 17(4), p. 513--518.

Miyoshi, T., Kanoh, J., Saito, M. & Ishikawa, F., 2008. Fission Yeast Pot1-Tpp1 Protects Telomeres and Regulates Telomere Length. *Science,* p. 1341--1344.

Mohaghegh, P. et al., 2001. The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic acids research,* 29(13), p. 2843--2849.

Morin, I. et al., 2008. Checkpoint-dependent phosphorylation of Exo1 modulates the {DNA} damage response. *The EMBO Journal,* 27(18), p. 2400--2410.

Morrone, S. et al., 2012. Crystal structure of a Tankyrase-Axin complex and its implications for Axin turnover and Tankyrase substrate recruitment. *Proceedings of the National Academy of Sciences,* 109(5), p. 1500--1505.

Moser, B. & Nakamura, T., 2009. Protection and replication of telomeres in fission yeast. *Biochemistry and Cell Biology,* 87(5), p. 747--758.

Muller, H., 1938. The remaking of chromosomes. *Collecting Net,* p. 181-198.

Murzin, A. G., 1993. OB(oligonucleotide/oligosaccharide binding)-fold: Common structural and functional solution for non-homologous sequences.. *The EMBO Journal,* 12(3), p. 861-867.

Nagy, Z. et al., 2016. Tankyrases Promote Homologous Recombination and Check Point Activation in Response to DSBs. *PLoS Genetics,* 12(2), p. e1005791.

Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S., 2018. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife,* Volume 7, p. e36861.

Nandakumar, J. et al., 2012. The TEL patch of telomere protein TPP1 mediates telomerase recruitment and processivity. *Nature,* Volume 492, p. 285-289.

Nandakumar, J., Podell, E. & Cech, T., 2010. How telomeric protein POT1 avoids RNA to achieve specificity for single-stranded DNA. *Proceedings of the National Academy of Sciences,* 107(2), p. 651-656.

Nelson, N. et al., 2018. The C-terminal extension unique to the long isoform of the shelterin component TIN2 enhances its interaction with TRF2 in a phosphorylation- and dyskeratosis congenita cluster-dependent fashion. *Molecular and Cellular Biology,* 38(12), p. MCB.00025-18.

Nguyen, T. et al., 2015. The architecture of the spliceosomal U4/U6.U5 tri-snrnp. *Nature,* 523(7558), p. 47-52.

Nguyen, T. H. D. et al., 2018. Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature,* 557(7704), p. 190--195.

Nguyen, T. et al., 2018. Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature,* 557(7704), p. 190-195.

Nogales, E. & Scheres, S., 2015. Cryo-EM: A unique tool for the visualization of macromolecular complexity. *Molecular Cell,* 58(4), p. 677-689.

O'Connor, M. et al., 2004. The human Rap1 protein complex and modulation of telomere length. *Journal of Biological Chemistry,* 279(27), p. 28585-28591.

O'Connor, M. S. et al., 2006. A critical role for TPP1 and TIN2 interaction in high-order telomeric complex assembly. *Proceedings of the National Academy of Sciences,* 103(32), p. 11874-11879.

Ohi, M., Li, Y., Cheng, Y. & Walz, T., 2004. Negative staining and image classification — powerful tools in modern electron microscopy. *Biological Procedures Online,* 6(1), p. 23-34.

Okamoto, K. et al., 2013. A two-step mechanism for TRF2-mediated chromosome-end protection. *Nature,* 494(7438), p. 502-505.

Okazaki, R. et al., 1968. Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains.. *Proceedings of the National Academy of Sciences,* 59(2), p. 598-605.

Olovnikov, A. M., 1973. 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,* p. 181--190.

Opresko, P. L. et al., 2004. The Werner Syndrome Helicase and Exonuclease Cooperate to Resolve Telomeric D Loops in a Manner Regulated by TRF1 and TRF2. *Molecular Cell,* 14(6), p. 763--774.

Opresko, P. & Shay, J., 2017. Telomere-associated aging disorders. *Ageing Research Reviews,* Volume 33, p. 52-66.

Orlova, E. V. & Saibil, H. R., 2011. Structural Analysis of Macromolecular Assemblies by Electron Microscopy. *Chem. Rev.,* Volume 111, p. 7710–7748.

Ortega Arroyo, J., Cole, D. & Kukura, P., 2016. Interferometric scattering microscopy and its combination with single-molecule fluorescence imaging. *Nature Protocols,* 11(4), p. 617-633.

Ortega-Arroyo, J. & Kukura, P., 2012. Interferometric scattering microscopy (iSCAT): new frontiers in ultrafast and ultrasensitive optical microscopy. *Physical Chemistry Chemical Physics,* 14(45), p. 15625-15636.

Overbeek, M. v. & de Lange, T., 2006. Apollo, an Artemis-Related Nuclease, Interacts with TRF2 and Protects Human Telomeres in S Phase. *Current Biology,* 16(13), p. 1295--1302.

Ozaki, Y. et al., 2012. Poly-ADP Ribosylation of Miki by tankyrase-1 Promotes Centrosome Maturation. *Molecular Cell,* 47(5), p. 694--706.

Pace, C. et al., 1995. How to measure and predict the molar absorption coefficient of a protein.. *Protein Sci. ,* Volume 11, p. 2411-2423.

Paeschke, K. et al., 2005. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nature Structural & Molecular Biology,* p. 847--854.

Palm, W. & de Lange, T., 2008. How Shelterin Protects Mammalian Telomeres. *Annual Review of Genetics,* 42(1), p. 301-334.

Palm, W., Hockemeyer, D., Kibe, T. & Lange, T. d., 2009. Functional Dissection of Human and Mouse POT1 Proteins. *Molecular and Cellular Biology,* 29(2), p. 471--482.

Passmore, L. & Russo, C., 2016. Specimen preparation for high-resolution cryo-EM. *Methods in Enzymology,* Volume 579, p. 51-86.

Penczek, P., Grassucci, R. & Frank, J., 1994. The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy,* 53(3), p. 251-270.

Perina, D. et al., 2014. Distribution of protein poly(ADP-ribosyl)ation systems across all domains of life. *DNA Repair,* Volume 23, p. 4–16.

Peterson, F. C. et al., 2011. Orphan macrodomain protein (human C6orf130) is an O-Acyl-ADP-ribose deacylase. *Journal of Biological Chemistry,* 286(41), p. 35955--35965.

Pettersen, E. F. et al., 2004. UCSF Chimera--a visualization system for exploratory research and analysis.. *Journal of Computational Chemistry,* 25(13), p. 1605-1612.

Pettersen, E. F. et al., 2021. UCSF ChimeraX: Structure visualization for researchers, educators, and developers.. *Protein Science. ,* 30(1), p. 70-82.

Pike, A. et al., 2019. TIN2 functions with TPP1/POT1 to stimulate telomerase processivity. *Molecular and Cellular Biology,* 39(21), p. 00593-18.

Pleschke, J. M., Kleczkowska, H. E., Strohm, M. & Althaus, F. R., 2000. Poly(ADP-ribose) Binds to Specific Domains in DNA Damage Checkpoint Proteins.. *Journal of Biological Chemistry,* 275(52), p. 40974--40980.

Podlevsky, J. D. et al., 2008. The Telomerase Database. *Nucleic Acids Research,* 36(Database), p. D339--D343.

Pollock, K., Mariotti, L. & Guettler, S., 2017. Regulation of WNT/β-catenin signalling by tankyrase-dependent poly(adp-ribosyl)ation and scaffolding. *British Journal of Pharmacology,* 174(24), p. 4611-4636.

Price, C. M., 1990. Telomere structure in Euplotes crassus: characterization of DNA-protein interactions and isolation of a telomere-binding protein.. *Molecular and Cellular Biology,* 10(7), p. 3421--3431.

Punjani, A. & Fleet, D., 2021. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *Journal of Structural Biology,* 213(2), p. 107702.

Punjani, A., Rubinstein, J., Fleet, D. & Brubaker, M., 2017. cryosPArc: algorithms for rapid unsupervised cryo-em structure determination. *Nature Methods,* 14(3), p. 290-297.

Punjani, A., Zhang, H. & Fleet, D., 2020. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nature Methods,* Volume 17, p. 1214–1221.

Rant, U. et al., 2004. Structural Properties of Oligonucleotide Monolayers on Gold Surfaces Probed by Fluorescence Investigations. *Langmuir,* 20(23), p. 10086--10092.

Rant, U. et al., 2004. Structural Properties of Oligonucleotide Monolayers on Gold Surfaces Probed by Fluorescence Investigations. 20(23), p. 10086-10092.

Rhodes, D. & Lipps, H. J., 2015. G-quadruplexes and their regulatory roles in biology. *Nucleic acids research,* p. 8627--8637.

Rice, C. et al., 2017. Structural and functional analysis of the human POT1-TPP1 telomeric complex. *Nature Communications,* 8(1), p. 14928.

Ripstein, Z. et al., 2017. Structure of a AAA+ unfoldase in the process of unfolding substrate. *eLife,* Volume 6, p. e25754.

Ripstein, Z. et al., 2020. A processive rotary mechanism couples substrate unfolding and proteolysis in the ClpXP degradation machinery. *eLife,* Volume 9, p. e52158.

Robart, A. R. & Collins, K., 2010. Investigation of Human Telomerase Holoenzyme Assembly, Activity, and Processivity Using Disease-linked Subunit Variants. *Journal of Biological Chemistry,* 285(7), p. 4375-4386.

Robart, A. R. & Collins, K., 2011. Human Telomerase Domain Interactions Capture DNA for TEN Domain-Dependent Processive Elongation. *Molecular Cell,* 42(3), p. 308--318.

Ruis, P. & Boulton, S. J., 2021. The end protection problem—an unexpected twist in the tail. *Genes & Development,* Volume 35, p. 1-21.

Ruska, E., 1987. The development of the electron microscope and of electron microscopy (Nobel lecture). *Angewandte Chemie International,* 26(7), p. 595-706.

Saint-Leger A, Koelblen M, Civitelli L, Bah A, Djerbi N, et al. 2014. The basic N-terminal domain of TRF2 limits recombination endonuclease action at human telomeres. *Cell Cycle,* 13:2469–74.

Samassekou, O., Gadji, M., Drouin, R. & Yan, J., 2010. Sizing the ends: normal length of human telomeres. *Annals of Anatomy-Anatomischer Anzeiger,* p. 284--291.

Sarek, G. et al., 2019. Cdk phosphorylation of TRF2 controls T-loop dynamics during the cell cycle. *Nature,* 575(7783), p. 523-527.

Sarek, G. et al., 2015. TRF2 Recruits RTEL1 to Telomeres in S Phase to Promote T-Loop Unwinding.. *Molecular Cell,* 57(4), p. 622–635.

Sarkar, J. et al., 2015. SLX4 contributes to telomere preservation and regulated processing of telomeric joint molecule intermediates.. *Nucleic Acids Research,* 43(12), p. 5912–5923.

Sarthy, J., Bae, N., Scrafford, J. & Baumann, P., 2009. Human RAP1 inhibits non-homologous end joining at telomeres. *The EMBO Journal,* 28(21), p. 3390-3399.

Sauerwald, A. et al., 2013. Structure of active dimeric human telomerase. *Nature Structural & Molecular Biology,* 20(4), p. 454--460.

Savage, S. A. et al., 2007. TINF2, a Component of the Shelterin Telomere Protection Complex, Is Mutated in Dyskeratosis Congenita. *Blood,* p. 835--835.

Savage, S. & Bertuch, A., 2010. The genetics and clinical manifestations of telomere biology disorders. *Genetics in Medicine,* 12(12), p. 753-764.

Sbodio, J. & Chi, N.-W., 2002. Identification of a tankyrase-binding motif shared by IRAP, TAB182, and human TRF1 but not mouse TRF1. *Journal of Biological Chemistry,* 277(35), p. 31887-31892.

Scheres, S., 2012a. A Bayesian view on cryo-EM structure determination.. *J. Mol. Biol.,* Volume 415, p. 406–418.

Scheres, S., 2012b. RELION: implementation of a Bayesian approach to cryo-EM structure determination.. *J. Struct. Biol.,* Volume 180, p. 519–530.

Scheres, S., 2016. Processing of structurally heterogeneous cryo-EM data in RELION. *Methods in Enzymology,* Volume 579, p. 125-157.

Scheres, S. et al., 2005. Maximum-likelihood multi-reference refinement for electron microscopy images.. *J. Mol. Biol. ,* Volume 348, p. 139–149.

Schmidt, C. & Urlaub, H., 2017. Combining cryo-electron microscopy (cryo-EM) and cross-linking mass spectrometry (CX-MS) for structural elucidation of large protein assemblies. *Current Opinion in Structural Biology,* Volume 46, p. 157-168.

Schmutz, I. et al., 2017. TRF2 binds branched DNA to safeguard telomere integrity. *Nature Structural & Molecular Biology,* 24(9), p. 734–742.

Scully, R., Panday, A., Elango, R. & Willis , N., 2019. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nature Reviews Molecular Cell Biology,* 20(11), p. 698-714.

Seimiya, H., Muramatsu, Y., Smith, S. & Tsuruo, T., 2004. Functional Subdomain in the Ankyrin Domain of Tankyrase 1 Required for Poly(ADP-Ribosyl)ation of TRF1 and Telomere Elongation.. *Molecular and Cellular Biology,* 24(5), p. 1944–1955.

Sekne, Z., Ghanim, G., van Roon, A.-M. & Nguyen, T., 2022. Structural basis of human telomerase recruitment by TPP1-POT1. *Science,* 375(6585), p. 1173-1176.

Sen, D. & Gilbert, W., 1990. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature,* 344(6265), p. 410-414.

Sfeir, A., 2012. Telomeres at a glance. *Journal of Cell Science,* 125(18), p. 4173-4178.

Sfeir, A. & de Lange, T., 2012. Removal of shelterin reveals the telo- mere end-protection problem.. *Science,* 336(6081), p. 593-597.

Sfeir, A. J., Chai, W., Shay, J. W. & Wright, W. E., 2005. Telomere-End Processing: the Terminal Nucleotides of Human Chromosomes. *Molecular Cell,* 18(1), p. 131--138.

Sfeir, A. et al., 2010. Loss of Rap1 induces telomere recombination in the absence of NHEJ or a DNA damage signal.. *Science,* 327(5973), p. 1657-1661.

Sfeir, A. et al., 2009. Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell,* 138(1), p. 90-103.

Shampay, J., Szostak, J. W. & Blackburn, E. H., 1984. DNA sequences of telomeres maintained in yeast. *Nature,* p. 154--157.

Shay, J. & Bacchetti, S., 1997. A survey of telomerase activity in human cancer. *European journal of cancer,* p. 787--791.

Shultz, M. D. et al., 2013. Identification of NVP-TNKS656: The Use of Structure–Efficiency Relationships To Generate a Highly Potent, Selective, and Orally Active Tankyrase Inhibitor.. *Journal of Medicinal Chemistry,* 56(15), p. 6495–6511.

Sirohi, D. et al., 2016. The 3.8 Å resolution cryo-EM structure of Zika virus. *Science,* 352(6284), p. 467-470.

Slade, D. et al., 2011. The structure and catalytic mechanism of a poly(ADP-ribose) glycohydrolase. *Nature,* 477(7366), p. 616--620.

Smith, E. et al., 2022. Insights into POT1 structural dynamics revealed by Cryo-EM. *PLOS ONE,* 17(2), p. e0264073.

Smith, S. & de Lange, T., 2000. Tankyrase promotes telomere elongation in human cells. *Current Biology,* 10(20), p. 1299-1302.

Smith, S., Giriat, I., Schmitt, A. & de Lange, T., 1998. Tankyrase, a poly (ADP-ribose) polymerase at human telomeres. *Science,* 282(5393), p. 1484-1487.

Smogorzewska, A. et al., 2000. Control of Human Telomere Length by TRF1 and TRF2. *Molecular and Cellular Biology,* 20(5), p. 1659--1668.

Soltermann, F., 2020. *Biophysical Characterization of Protein-Protein Interactions with Mass Photometry,* s.l.: University of Oxford.

Soltermann, F. et al., 2020. Quantifying protein-protein interactions by molecular counting with mass photometry. *Angewandte Chemie,* 132(27), p. 10866-10871.

Song, K. et al., 2000. Interaction of human Ku70 with TRF2. *FEBS Letters,* 481(1), p. 81-85.

Sonn-Segev, A. et al., 2020. Quantifying the heterogeneity of macromolecular machines by mass photometry. *Nature Communications,* 11(1772), p. 1-10.

Spielmann, H. P., Dwyer, T. J., John E. Hearst & Wemmer, D. E., 1995. Solution structures of psoralen monoadducted and cross-linked DNA oligomers by NMR spectroscopy and restrained molecular dynamics. *Biochemistry,* p. 12937–12953.

Stansel, R., de Lange, T. & Griffith, J., 2001. T-loop assembly in vitro involves binding of TRF2 near the 3′ telomeric overhang.. *The EMBO journal,* 20(19), p. 5532-5540.

Szostak, J. W. & Blackburn, E. H., 1982. Cloning yeast telomeres on linear plasmid vectors. *Cell,* p. 245--255.

Tagari, M. et al., 2002. New electron microscopy database and deposition system. *Trends in Biochemical Sciences,* 27(11), p. 589.

Takai, H., Smogorzewska, A. & de Lange, T., 2003. DNA damage foci at dysfunctional telomeres. *Current Biology,* 13(17), p. 1549-1556.

Takai, K. K. et al., 2010. In vivo stoichiometry of shelterin components. *Journal of Biological Chemistry,* 285(2), p. 1457-1467.

Takai, K. K. et al., 2011. Telomere protection by TPP1/POT1 requires tethering to TIN2. *Molecular Cell,* 44(4), p. 647-659.

The Cancer Genome Atlas Research Network, 2011. Integrated genomic analyses of ovarian carcinoma.. *Nature,* 474(7353), p. 609--615.

Timashev, L. A. & de Lange, T., 2020. Characterization of t-loop formation by TRF2. *Nucleus,* 11(1), p. 164-177.

Tomar, R. S., 2016. The multifunctional transcription factor Rap1 a regulator of yeast physiology. *Frontiers in Bioscience,* p. 918--930.

Tomaska, L. et al., 2004. Taz1 Binding to a fission yeast model telomere formation of telomeric loops and higher order structures. *Journal of Biological Chemistry,* p. 50764--50772.

Tomita, K. & Cooper, J. P., 2008. Fission yeast Ccq1 is telomerase recruiter and local checkpoint controller. *Genes & Development,* 22(24), p. 3461-3474.

Touzot, F. et al., 2010. Function of Apollo (SNM1B) at telomere highlighted by a splice variant identified in a patient with Hoyeraal-Hreidarsson syndrome. *Proceedings of the National Academy of Sciences,* 107(22), p. 10097--10102.

Tribolium Genome Sequencing Consortium, 2008. The genome of the model beetle and pest Tribolium castaneum. *Nature,* 452(24), p. 949--955.

Uhlen, M. et al., 2015. Tissue-based map of the human proteome. *American Association for the Advancement of Science,* Volume 347, p. 1260419-1260419.

Ulaner, G. et al., 2001. Tissue-specific alternate splicing of human telomerase reverse transcriptase (hTRET) influences telomere lengths during human development. *International Journal of Cancer,* 91(5), p. 644-649.

van Heel, M. et al., 1996. A new generation of the IMAGIC image processing system. *Journal of Structural Biology,* 116(1), p. 17-24.

Van Ly, D. et al., 2018. Telomere loop dynamics in chromosome end protection. *Molecular Cell,* 71(4), p. 510--525.e6.

van Steensel, B. & Belmont, A. S., 2017. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell,* 169(5), p. 780--791.

van Steensel, B. & de Lange, T., 1997. Control of telomere length by the human telomeric protein TRF1. *Nature,* 385(6618), p. 740-743.

Van Steensel, B. & de Lange, T., 1997. Control of telomere length by the human telomeric protein TRF1. *Nature,* p. 740--743.

Van Steensel, B., Smogorzewska, A. & de Lange, T., 1998. TRF2 protects human telomeres from end-to-end fusions. *Cell,* p. 401--413.

Vannier, J.-B.et al., 2012. RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. *Cell,* 149(4), p. 795-806.

Venteicher, A. et al., 2009. A human telomerase holoenzyme protein required for cajal body localization and telomere synthesis. *Science,* 323(5914), p. 644-648.

Vera, E. et al., 2012. The rate of increase of short telomeres predicts longevity in mammals. *Cell Reports,* 2(4), p. 732–737.

Verschueren, H., 1985. Interference reflection microscopy in cell biology: methodology and applications. *Journal of Cell Science,* 75(1), p. 279-301.

Voronkov, A. et al., 2013. Structural Basis and SAR for G007-LK, a Lead Stage 1,2,4-Triazole Based Specific Tankyrase 1/2 Inhibitor.. *Journal of Medicinal Chemistry,* 56(7), p. 3012–3023.

Vyas, S. et al., 2013. A systematic analysis of the PARP protein family identifies new functions critical for cell physiology.. *Nature Communications,* 4(2240).

Waaler, J. et al., 2011. Novel Synthetic Antagonists of Canonical Wnt Signaling Inhibit Colorectal Cancer Cell Growth. *Cancer Research,* 71(1), p. 197--205.

Waaler, J. et al., 2012. A Novel Tankyrase Inhibitor Decreases Canonical Wnt Signaling in Colon Carcinoma Cells and Reduces Tumor Growth in Conditional APC Mutant Mice. *Cancer Research,* 72(11), p. 2822--2832.

Wagner, T. et al., 2019. Sphire-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Communications Biology,* Volume 2, p. 218.

Wahlberg, E. et al., 2012. Family-wide chemical profiling and structural analysis of PARP and tankyrase inhibitors.. *Nature Biotechnology,* 30(3), p. 283–288.

Walne, A. J. et al., 2008. TINF2 mutations result in very short telomeres: analysis of a large cohort of patients with dyskeratosis congenita and related bone marrow failure syndromes. *Blood,* p. 3594--3600.

Wang, F. et al., 2007. The POT1-TPP1 telomere complex is a telomerase processivity factor. *Nature,* Volume 445, p. 506-510.

Wang RC, Smogorzewska A, de Lange T. 2004. Homologous recombination generates T-loop-sized deletions at human telomeres. *Cell*, 119:355–68

Wang, Y. & Patel, D. J., 1993. Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure,* 1(4), p. 263--282.

Wang, Z. et al., 2012. Recognition of the iso-ADP-ribose moiety in poly(ADP-ribose) by WWE domains suggests a general mechanism for poly(ADP-ribosyl)ation-dependent ubiquitination.. *Genes & Development,* 26(3), p. 235–240.

Wan, M., Qin, J., Songyang, Z. & Liu, D., 2009. OB Fold-containing Protein 1 (OBFC1), a Human Homolog of Yeast Stn1, Associates with TPP1 and Is Implicated in Telomere Length Regulation. *Journal of Biological Chemistry,* 284(39), p. 26725--26731.

Watson, J. & Crick, F., 1953. Molecular structure of Nucleic Acids: A structure for deoxyribose nucleic acid. *Nature,* 171(4356), p. 737-738.

Watson, J. D., 1972. Origin of Concatemeric T7 DNA. *Nature New Biology,* p. 197--201.

Weinrich, S. L. et al., 1997. Reconstitution of human telomerase with the template RNA component hTR and the catalytic protein subunit hTRT. *Nature Genetics,* 17(4), p. 498--502.

Weismann, A., 1891. *Essays upon heredity and kindred biological problems..* 2 ed. s.l.:Oxford, Clarendon Press.

Williamson, J. R., Raghuraman, M. & Cech, T. R., 1989. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell,* p. 871--880.

Wold, M., 1997. Replication protein A: A heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annual Review of Biochemistry,* 66(1), p. 61-92.

Wong, K.-K.et al., 2003. Telomere dysfunction and Atm deficiency compromises organ homeostasis and accelerates ageing. *Nature,* 421(6923), p. 643--648.

Wood, L. D. et al., 2001. Characterization of ataxia telangiectasia fibroblasts with extended life-span through telomerase expression. *Oncogene,* 21(3), p. 278--288.

Wright, W. et al., 1996. Telomerase activity in human germline and embryonic tissues and cells. *Developmental Genetics,* 18(18), p. 173-179.

Wrobel, A. G. et al., 2022. Evolution of the SARS-CoV-2 spike protein in the human host. *Nature Communications,* Volume 13, p. 1178.

Wu, L. et al., 2006. Pot1 Deficiency Initiates DNA Damage Checkpoint Activation and Aberrant Homologous Recombination at Telomeres. *Cell,* 126(1), p. 49--62.

Wu, P., Overbeek, M. v., Rooney, S. & Lange, T. d., 2010. Apollo Contributes to G Overhang Maintenance and Protects Leading-End Telomeres. *Molecular Cell,* 39(4), p. 606--617.

Xin, H. et al., 2007. TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase. *Nature,* p. 559--562.

Yang, E. et al., 2016. Wnt pathway activation by ADP-ribosylation. *Nature Communications,* p. 1-14.

Yeager, T. R. et al., 1999. Telomerase-negative immortalized human cells contain a novel type of promyelocytic leukemia (PML) body. *Cancer Research,* 59(17), p. 4175--4179.

Ye, J. Z.-S. & de Lange, T., 2004c. TIN2 is a tankyrase 1 PARP modulator in the TRF1 telomere length control complex. *Nature Genetics,* 36(6), p. 618-623.

Ye, J. Z.-S. et al., 2004a. TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres. *Journal of Biological Chemistry,* 279(45), p. 47264-47271.

Ye, J. Z.-S. et al., 2004b. POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex. *Genes & Development,* 14(18), p. 1649-1654.

Young, G. et al., 2018. Quantitative mass imaging of single biological macromolecules.. *Science,* 360(6387), p. 423-427.

Young, G. & Kukura, P., 2019. Interferometric scattering microscopy. *Annual Review of Physical Chemistry,* 70(1), p. 301-322.

Young, G. & Kukura, P., 2019. Interferometric Scattering Microscopy. *Annual Review of Physical Chemistry,* 70(1), p. 301-322.

Yu, C. & Huang, L., 2017. Cross-linking mass spectrometry: An emerging technology for interactomics and structural biology. *Analytical Chemistry,* 90(1), p. 144-165.

Zaug, A., Podell, E. & Cech, T., 2005. Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro.. *Proceedings of the National Academy of Sciences,* 102(31), p. 10864-10869.

Zhang, F., Chen, Y., Li, M. & Yu, X., 2014. The oligonucleotide/oligosaccharide-binding fold motif is a poly(ADP-ribose)-binding domain that mediates DNA damage response.. *Proceedings of the National Academy of Sciences,* 111(20), p. 7278--7283.

Zhang, W. et al., 2011. Solution structure of Rap1 BRCT domain from Saccharomyces cerevisiae reveals a novel fold. *Biochemical and Biophysical Research Communications,* 404(4), p. 1055-1059.

Zhang, Y. et al., 2011. RNF146 is a poly(ADP-ribose)-directed E3 ligase that regulates axin degradation and Wnt signalling.. *Nature Cell Biology,* 13(5), p. 623–629.

Zhong, F. L. et al., 2012. TPP1 OB-Fold Domain Controls Telomere Maintenance by Recruiting Telomerase to Chromosome Ends.. *Cell,* 150(3), p. 481-494.

Zhong, Z., Shiue, L., Kaplan, S. & de Lange, T., 1992. A mammalian factor that binds telomeric TTAGGG repeats in vitro.. *Molecular and Cellular Biology,* 12(11), p. 4834-4843.

Zhu, X.-D.et al., 2000. Cell-cycle-regulated association of RAD50/mre11/nbs1 with TRF2 and human telomeres. *Nature Genetics,* 25(3), p. 347-352.

Zimmermann, M., Kibe, T., Kabir, S. & de Lange, T., 2014. TRF1 negotiates TTAGGG repeat-associated replication problems by recruiting the BLM helicase and the TPP1/POT1 repressor of ATR. *Genes & Development,* 28(22), p. 2477-2491.

Zinder, J. C. et al., 2022. Shelterin is a Dimeric Complex with Extensive Structural Heterogeneity. *bioRxiv preprint.*