Structural and functional analysis of proteins that regulate fission yeast telomerase activity

Jonathan Edward Barrow

Imperial College London

and

The Francis Crick Institute

PhD Supervisor: Steve Smerdon

A thesis submitted for the degree of

Doctor of Philosophy

Imperial College London

December 2018

Declaration

I, Jonathan Barrow, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Copyright declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

Shelterin is a protein-DNA complex found at chromosome ends, which prevents their recognition by the double strand break surveillance function of the DNA damage response and plays important roles in regulation of telomere assembly and telomerase activity. In *Schizosaccharomyces pombe*, the shelterin complex consists of Taz1, Rap1, Poz1, Tpz1, Pot1 and Ccq1. In spite of its biological importance in maintaining genomic integrity, relatively little is known about the structures of the proteins of the shelterin complex and how their interactions coordinate their regulatory functions.

This thesis describes experiments to investigate the structural and biochemical basis of shelterin assembly and regulation. Using a combination of approaches, Ccq1 was shown to behave as an elongated, modular and somewhat flexible molecule, dimerising tightly through a C-terminal coiled-coil region. Binding experiments showed that Ccq1 interacts with Tpz1 with micromolar affinity, whereas Poz1 and Tpz1 were found to bind more tightly with an apparent K_d in the low nanomolar range. This suggests a highly stable Poz1-Tpz1 interaction occurs *in vivo* in keeping with the proposed role for this sub-complex as a structural bridge within the shelterin assembly. The minimal binding motifs for Poz1 and Ccq1 were identified within the C-terminal Tpz1 region and, importantly, shown to not overlap, allowing each protein to bind to Tpz1 non-competitively to form a stable three component sub-complex. Negative stain electron microscopy of this tripartite assembly containing full-length

Poz1-Tpz1-Ccq1, co-expressed in insect cells, generated a low resolution envelope in which the three components are arranged to form an interrupted shell-like structure surrounding a central void. Most strikingly, the structure shows a markedly asymmetric architecture in spite of the 2: 2: 2 stoichiometry of Ccq1, Tpz1 and Poz1 in the complex. The molecular basis for this asymmetry remains unclear, but may have important implications for telomerase regulation.

Acknowledgements

I would like to thank Steve Smerdon for giving me the opportunity to work in his lab. I am very grateful for his support when required and also the freedom he has afforded me to direct the project.

I would also like to thank the members of my thesis committee, Xiaodong Zhang, Katrin Rittinger and Peter Thorpe. Their advice throughout has proven invaluable and has helped to concentrate this project in the correct direction on several occasions.

Thanks to all past and present lab members, your input has made the last four years a really enjoyable experience. I would like to give special thanks to Simon Pennell, who has spent a substantial amount of time proof reading this thesis and removing ‘Jonny-isms’, Sarah Webb for helping with electron microscopy image acquisition and Thomas Flower for some great ‘lab meetings’.

The core facilities available at the Francis Crick Institute have helped significantly to advance this project. Special thanks to Phil Walker, Andrew Purkiss, Steve Howell, David Frith, Laura Masino, Steve Martin, Nicola O’Reilly, Roksana Ogrodowicz, Kirsten Blake, Vangelis Christodoulou and Stephane Mouilleron. Also, thanks to Ian Taylor for help with SEC-MALLS and Dave Briggs for SAXS analysis help.

Finally, thanks to my friends and family, whose support have made this PhD possible.

Table of Contents

Abstract **4**
Acknowledgements **6**
Table of Contents **7**
Table of figures **11**
List of tables **16**
Abbreviations **18**
Chapter 1.Introduction **20**
1.1 DNA replication and the end replication problem **22**
1.1.1 DNA Replication 22
1.1.2 The end replication problem 26
1.2 Telomeres and their function **30**
1.2.1 What are telomeres? 30
1.2.2 Telomere structure 31
1.2.3 Telomerase 32
1.2.4 Telomere shape 35
1.3 Telomeres and disease prevention **40**
1.3.1 The molecular clock and the Hayflick limit 40
1.3.2 Suppression of the DNA damage response 41
1.4 Shelterin complex **44**
1.4.1 Shelterin complex components in *S. pombe* 44
1.4.2 Shelterin homology 56
1.4.3 Shelterin and telomerase regulation 57
1.4.4 Shieldin performs similar roles to shelterin, but at double strand breaks 62
1.5 Other telomere associated proteins **64**
1.5.1 Est1 and telomerase regulation 64
1.5.2 The CST complex 68
1.5.3 TZAP and telomere trimming 70
1.6 Telomere length dysregulation and disease **72**
1.6.1 Telomerase activation and cancer 72
1.6.2 Telomere maintenance disorders 73
1.6.3 T-loops and telomeropathies 76
1.7 Aims and objectives **77**
Chapter 2.Materials & Methods **78**
2.1 Molecular biology **78**
2.1.1 Vectors 78
2.1.2 Bacterial strains 79
2.1.3 Primer design 81
2.1.4 Polymerase chain reaction (PCR) 82
2.1.5 Ligation independent cloning (LIC) 83
2.1.6 Gibson assembly 85
2.1.7 Restriction enzyme based cloning and ligation 86
2.1.8 QuikChange site-directed mutagenesis 87

2.1.9 Transformation and sequencing 88
2.2 Protein production 90
2.2.1 Small scale (1 mL) protein expression in *E. coli* 90
2.2.2 Large scale (>500 mL) protein expression in *E. coli* 90
2.2.3 Expression in insect cells 91
2.2.4 Cell lysis 93
2.2.5 Histidine-tagged protein purification 94
2.2.6 Glutathione Sepharose affinity chromatography 95
2.2.7 Ion-exchange chromatography (IEX) 96
2.2.8 Size exclusion chromatography (SEC) 96
2.2.9 SDS-PAGE 97
2.2.10 Western blot 98
2.2.11 Protein concentration 98
2.2.12 Protein dialysis 99
2.3 Biophysical and biochemical techniques 100
2.3.1 Limited proteolysis 100
2.3.2 Nano differential scanning fluorimetry (nanoDSF) 100
2.3.3 Peptide array analysis 101
2.3.4 Size exclusion chromatography Multi Angle Laser Light Scattering (SEC-MALLS) 102
2.3.5 Circular dichroism 103
2.3.6 Isothermal titration calorimetry (ITC) 104
2.3.7 Biolayer interferometry (BLI) 107
2.4 Structural characterisation 110
2.4.1 Protein crystallisation 110
2.4.2 X-ray diffraction experiments 111
2.4.3 Small angle X-ray scattering (SAXS) 112
2.4.4 Glutaraldehyde fixation (GraFix) 113
2.4.5 Electron microscopy 114
2.4.6 Electron microscopy image analysis 117
Chapter 3.Protein expression and purification 119
3.1 Introduction 119
3.2 Expression protocol summary 121
3.3 Expression of Est1 123
3.3.1 *E. coli* and *S. pombe* expression 123
3.3.2 Insect cell expression 126
3.3.3 Orthologue expression 128
3.3.4 Expression summary and future uses for purified Est1 131
3.4 Expression and purification of shelterin complex components 132
3.4.1 *E. coli* expression of Ccq1 132
3.4.2 Expression summary and future uses for purified Ccq1 139
3.4.3 Small scale *E. coli* expression tests of Pot1 140
3.4.4 Tpz1 expression optimisation in *E. coli* 145
3.4.5 Expression summary and future uses for purified Tpz1 151
3.4.6 *E. coli* expression of Poz1 152
3.4.7 Expression summary and future uses for purified Poz1 159
3.5 Summary 160
Chapter 4.Structural characterisation of Ccq1 163

4.1 Introduction **163**
4.2 Biophysical characterisation of Ccq1 ..... **164**
4.2.1 Secondary structure determination 164
4.2.2 Crystallisation trials 167
4.2.3 Determination of the oligomeric state of Ccq1 171
4.2.4 Domain arrangement of Ccq1 173
4.2.5 Small angle X-ray scattering (SAXS) analysis of Ccq1 130-717 173
4.3 Conclusions **180**
Chapter 5.Biophysical and structural characterisation of shelterin sub-complexes **182**
5.1 Introduction **182**
5.2 The Tpz1-Ccq1 sub-complex **184**
5.2.1 Characterisation of a minimal Ccq1 binding domain of Tpz1 184
5.3 Characterisation of Poz1 + Tpz1 binding **191**
5.3.1 Tpz1 peptide design for Poz1 binding 191
5.3.2 Tpz1 475-508-Poz1 binding 191
5.3.3 Poz1 + Tpz1 co-expression 195
5.4 A tripartite complex of Poz1 + Tpz1 + Ccq1 ..... **199**
5.4.1 Binding affinity determination of Ccq1 for a Poz1-Tpz1 complex 199
5.4.2 Construct optimisation and tripartite complex formation 201
5.4.3 Negative stain electron microscopy analysis 202
5.4.4 Sample stabilisation by GraFix 203
5.5 Discussion **206**
5.5.1 Summary 206
5.5.2 The Poz1-Tpz1 interaction 207
5.5.3 The Ccq1-Tpz1 interaction affinity is weaker than that of Poz1-Tpz1 208
Chapter 6.Electron microscopy analysis of shelterin sub-complexes expressed in insect cells **211**
6.1 Introduction **211**
6.2 Complex expression in insect cells **212**
6.2.1 MultiBac cloning 212
6.2.2 MultiBac protein expression 214
6.3 Structural characterisation **218**
6.3.1 Complex characterisation 218
6.3.2 Complex stabilisation 219
6.3.3 Structural analysis by single particle negative stain electron microscopy 222
6.3.4 Structural asymmetry of the shelterin complex 231
6.3.5 Conformational changes are required for tripartite complex formation 232
6.4 Conclusions **235**
Chapter 7.Future perspectives **237**
7.1 Overview **237**
7.2 Further analysis of Ccq1 ..... **238**
7.2.1 Structural analysis by crystallography and SAXS 238
7.2.2 Ccq1 phosphorylation and it's interaction with Est1 240
7.3 Analysis of bipartite complexes **241**

7.4 Structural analysis of larger shelterin sub-complexes..... 243
Chapter 8.Appendix 248
Reference List 252

Table of figures

Figure 1: Formation of the eukaryotic DNA replication fork. 24

Figure 2: Leading and lagging strand synthesis during DNA replication. 25

Figure 3: The end replication problem. 27

Figure 4: The G-overhang. 32

Figure 5: T-loop formation. 38

Figure 6: Schematic showing the DNA damage response to double strand breaks in DNA. 42

Figure 7: The shelterin complex. 45

Figure 8: Interaction regions between shelterin complex components known at the outset of this project. 46

Figure 9: Domain structure of Ccq1 47

Figure 10: Domain structure of Tpz1. 48

Figure 11: Domain structure of Poz1. 50

Figure 12: Domain structure of Rap1. 51

Figure 13: Domain structure of Taz1. 53

Figure 14: Domain structure of Pot1 55

Figure 15: Shelterin abundance at telomeres is independent of telomere length. 59

Figure 16: Proposed mechanism for telomerase recruitment and stabilisation. 62

Figure 17: Domain structure of Est1. 66

Figure 18: TZAP and shelterin compete for telomere binding. 71

Figure 19: Biolayer interferometry. 109

Figure 20: Image formation by a typical transmission electron microscope. .. 115
Figure 21: Domains of Est1 and construct design. 123
Figure 22: Band ID mass spectrometry for insect cell expressed *S. pombe* Est1. 126
Figure 23: Optimisation of Est1 expression by the baculovirus expression system. 127
Figure 24: Secondary structure sequence alignment of homologous Est1 proteins from closely related yeast species using the PRALINE server [215]. 129
Figure 25: Example purification of *S. cryophilus* Est1. 130
Figure 26: Ccq1 domains and construct design. 132
Figure 27: Example purification for Ccq1 493-717. 135
Figure 28: Comparison of Ccq1 130-717 purity after 26/60 Superdex 200 SEC. 137
Figure 29: Thermal denaturation of Ccq1 130-717 as observed by nanoDSF. 138
Figure 30: pET52b Ccq1 solubility tests. 139
Figure 31: Pot1 domains and construct design. 141
Figure 32: Pot1 expression and solubility tests. 144
Figure 33: Tpz1 domains and construct design. 145
Figure 34: Tpz1 expression tests. 148
Figure 35: Tpz1 construct optimisation. 150
Figure 36: Expression test for GST tagged Tpz1 415-458. 151
Figure 37: Poz1 expression tests 153
Figure 38: Poz1 domains and construct design. 154

Figure 39: Poz1 expression tests. 156
Figure 40: Poz1 25-249 purification..... 158
Figure 41: Secondary structure prediction of an alignment of Ccq1 homologues from closely related fission yeast species using PRALINE..... 165
Figure 42: Characterisation of the secondary structure of Ccq1 130-717 and 493-717 by CD. 166
Figure 43: Initial crystallisation of Ccq1 130-717. 168
Figure 44: Crystallisation of Ccq1 130-717 using the Memgold screen (Molecular Dimensions). 169
Figure 45: Diffraction image for Ccq1 130-717..... 170
Figure 46: The oligomeric state of Ccq1 130-717 and 493-717 was characterised by SEC-MALLS. 172
Figure 47: SAXS analysis of Ccq1 130-717. 174
Figure 48: SAXS-derived envelope of Ccq1 130-717, showing the effect of enforcing P2 symmetry on the final model. 177
Figure 49: Potential modes for dimerisation of Ccq1 130-717..... 179
Figure 50: Known protein domains and regions that interact with Tpz1. 183
Figure 51: Peptide array analysis of the Ccq1-Tpz1 interaction. 185
Figure 52: Representative ITC experiment showing binding between Ccq1 130-717 and Tpz1-P24..... 187
Figure 53: Binding affinity of a biotin-Tpz1 full peptide for Ccq1 130-717. 189
Figure 54: Binding affinity of Tpz1 405-508 for Ccq1 130-717. 190
Figure 55: Schematic showing Poz1-Tpz1 interaction..... 191
Figure 56: Binding affinity of Poz1 to a Tpz1 peptide. 192

Figure 57: CD analysis of Poz1 and a Poz1-Tpz1 complex. 194
Figure 58: Size exclusion chromatography of a Poz1-Tpz1 complex. 196
Figure 59: SEC-MALLS of Poz1-Tpz1 complex. 197
Figure 60: Poz1-Tpz1 degradation during purification. 198
Figure 61: Binding affinity of a Poz1-Tpz1 complex to Ccq1. 200
Figure 62: Size exclusion chromatography of a Poz1-Tpz1-Ccq1 complex. .. 202
Figure 63: GraFix of a Poz1-Tpz1-Ccq1 heterotrimer leads to separation of different species. 205
Figure 64: Schematic showing the Poz1-Tpz1-Ccq1 interaction. 207
Figure 65: MultiBac cloning scheme 213
Figure 66: Purification of MultiBac expressed full shelterin complex. 216
Figure 67: Full shelterin complex purified sample. 217
Figure 68: SEC-MALLS of a 5-component shelterin sub-complex. 219
Figure 69: GraFix of the full shelterin complex expressed by the MultiBac insect cell system. 220
Figure 70: Schematic showing the Scipion workflow used for image analysis. 224
Figure 71: Initial image analysis of the Poz1-Tpz1-Ccq1 shelterin sub complex. 225
Figure 72: Validation of the structural analysis. 226
Figure 73: Structure of a crosslinked shelterin sub complex containing Poz1-Tpz1-Ccq1 227
Figure 74: Comparison of the structure of CPT-FL with the structure of $Poz1_{2-249}$-$Tpz1_{406-508}$-$Ccq1_{2-716}$ (CPT-ΔN) (EMDB: 8863). 229

Figure 75: A structure of similar mass to the Tpz1 extension fits into the additional volume seen in CPT-FL..... 230
Figure 76: A published EM map (EMDB: 8863) shows a bent conformation of each monomer within a Ccq1 dimer (adapted from *Scott et al.* [230])..... 233

List of tables

Table 1: Telomere sequences in different organisms. 31

Table 2: Important telomerase subunits in humans and yeast. 34

Table 3: Shelterin complex orthologues from humans and *S. pombe* 57

Table 4: Telomeropathies and the genes involved. 75

Table 5: Plasmids used for cloning and protein expression 79

Table 6: Genotypes of bacterial strains used. 80

Table 7: Reagents required for a typical PCR reaction. 82

Table 8: Thermal cycling protocol for a typical PCR using KOD Hot Start DNA Polymerase. 83

Table 9: Reagents required for T4 DNA polymerase treatment of PCR products. 84

Table 10: Typical reaction components for Gibson assembly. 86

Table 11: Reagents required for ligation with T4 DNA ligase. 87

Table 12: Summary of potential expression optimisation methods used during this project. 122

Table 13: Est1 constructs designed using secondary structure predictions. .. 125

Table 14: Ccq1 constructs designed using secondary structure predictions and tryptic digest data. 134

Table 15: Pot1 constructs designed using secondary structure predictions. .. 142

Table 16: Tpz1 constructs used in following experiments. 147

Table 17: Poz1 constructs initially designed using secondary structure predictions and then optimised by truncations. 155

Table 18: Protein fragments important for future experiments and the protocol required for purification..... 162
Table 19: SAXS-derived parameters for Ccq1 130-717. 175
Table 20: Tpz1 peptides designed using information from peptide array analysis. 186
Table 21: Uniprot IDs for proteins used during this thesis. 248
Table 22: Additives required for small-scale expression tests. 248
Table 23: T_m values measured by nanoDSF. 248

Abbreviations

Å	Angstrom
ALT	Alternative lengthening of telomeres
ATM	Ataxia telangiectasia mutated
ATP	Adenosine triphosphate
ATR	Ataxia telangiectasia and Rad3 related
BLI	Biolayer interferometry
BLM	Bloom helicase
CCD	Charge coupled device
CD	Circular dichroism
CMG	Cdc45-MCM-GINS
CST	Cdc13/Ctc1, Stn1, Ten1
DC	Dyskeratosis congenita
DNA	Deoxyribonucleic acid
DSB	Double strand break
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylene diamine tetra-acetic acid
EM	Electron microscopy
EMDB	Electron microscopy data bank
EST	Ever shorter telomeres
FSC	Fourier shell correlation
GST	Glutathione-S-transferase
HDAC	Histone deacetylase
HPLC	High pressure liquid chromatography
HRP	Horseradish peroxidase
HTH	Helix-turn-helix
IEX	Ion exchange
IMAC	Immobilised metal ion affinity chromatography
IPF	Idiopathic pulmonary fibrosis
ITC	Isothermal titration calorimetry
LB	Luria-Bertani
LIC	Ligation independent cloning
MOI	Multiplicity of infection
MRN	Mre11-Rad50-Nbs1

MWCO	Molecular weight cut off
NHEJ	Non-homologous end joining
ORC	Origin replication complex
PCR	Polymerase chain reaction
PDB	Protein data bank
RNA	Ribonucleic acid
RPA	Replication protein A
SAXS	Small angle X-ray scattering
SEC	Size exclusion chromatography
SEC-MALLS	Size exclusion chromatography multi angle laser light scattering
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
TBS	Tris buffered saline
UV	Ultraviolet

Chapter 1. Introduction

One of the defining characteristics of eukaryotic cells is the packaging of linear double-stranded genomic deoxyribonucleic acid (DNA) into chromosomes. This confers a number of advantages to the cell, not least the ability to store large quantities of genetic material in very small volumes. Chromosomes also provide opportunities for the regulation of gene expression and offer a simple mechanism for genome segregation following DNA replication during cell division [1]. Each diploid cell contains approximately two meters of DNA [2], which is wound around histones forming nucleosomes and a beads on a string structure [3]. Post translational modification of the histone tails leads to a number of different processes including loosening of the bound DNA allowing controlled expression of genes to occur [4].

However, the linear nature of chromosomes presents two major problems for the cell. The first of these is known as the ‘end replication problem’, and stems from the inability of the replication machinery to completely replicate chromosomes during the cell cycle, leading to the loss of DNA from chromosome ends [5]. The second is known as the ‘end protection problem’, which occurs because the ends of linear chromosomes resemble DNA double strand breaks (DSBs). Recognition of chromosome ends as DSBs would activate the DNA damage response and would cause the aberrant joining of chromosomes, which would have fatal consequences for the cell and organism [6]. The majority of organisms have

evolved specialised structures, known as telomeres, to deal with the end replication and end protection problems [7]. Telomeres and their associated problems will be introduced in detail later, but first it is pertinent to describe the normal mechanism of DNA replication.

1.1 DNA replication and the end replication problem

1.1.1 DNA Replication

Faithful DNA replication prior to cell division is required for doubling of the genome to occur, so each daughter cell contains a full copy of the original genetic information. During semi-conservative DNA replication, DNA polymerases synthesize a complementary DNA strand from the original template [8, 9].

Before replication can take place, the double-stranded DNA helix is unwound by helicases, leading to the formation of two pieces of single-stranded DNA and two replication forks [10]. These single-stranded DNAs can then be used to produce two copies of the original genetic material, which can be partitioned into daughter cells during mitotic cell division. This process is highly conserved and involves a number of large protein complexes.

DNA replication is initiated at specific sites known as origins, which are recognised and bound by the origin recognition complex (ORC) [11, 12]. Initially, the MCM2-7 helicase is loaded onto DNA through an interaction between a replication origin-bound ORC containing a Cdc6 subunit and Cdt1 bound to the MCM2-7 helicase hexamer [13, 14]. Loading of a second MCM2-7 helicase then occurs forming an inactive double hexameric structure [15, 16]. This complex recruits a large number of accessory proteins in a phosphorylation dependent

manner forming a preinitiation complex [17]. Rearrangement of the double hexamer occurs, resulting in the movement of each hexamer within the double hexamer onto the leading strand of two replication forks that move in opposite directions [18]. The helicase is now in its active conformation and contains Cdc45, MCM2-7 and GINS and is called the CMG helicase [19]. Each of the active CMG complexes then move along the DNA, in an ATP-dependent process, forming single-stranded DNA [20, 21], which is coated with the single-stranded DNA binding protein Replication Protein A (RPA) (see Figure 1). This single-stranded DNA is now available for priming of leading strand DNA synthesis by the Pol α-RNA primase complex. Extension of this primer on the leading strand by Pol ε can then occur [22]. The mechanism of activation and action of these complexes is the subject of intense study.

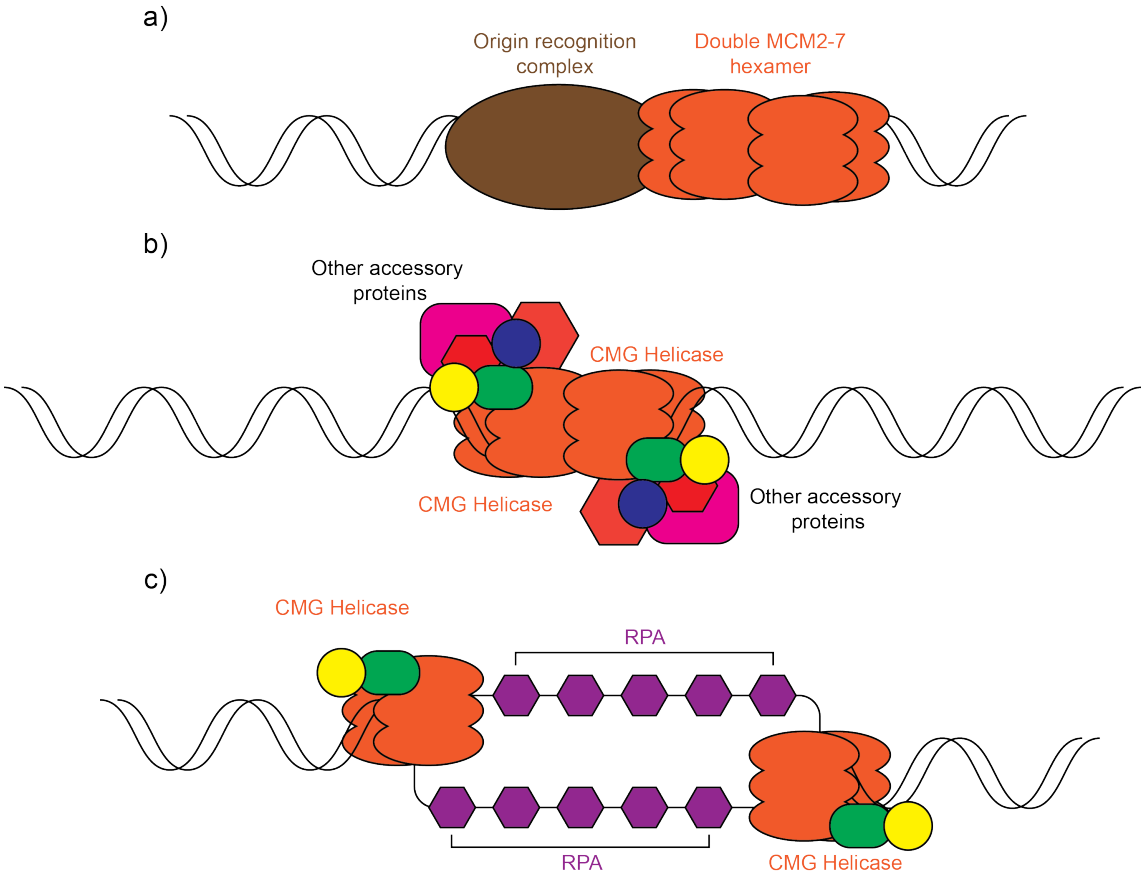


Figure 1: Formation of the eukaryotic DNA replication fork.
a) Loading of the MCM2-7 helicase onto DNA occurs at replication origins through an interaction with the origin recognition complex. A second helicase is then loaded onto the DNA forming a double hexamer structure. b) Recruitment of a large number of accessory proteins occurs through a phosphorylation dependent process. These proteins include Cdc45 and GINS, leading to formation of the CMG helicase. c) Activation of the CMG helicase causes movement of each hexamer within the double hexamer onto the leading strand of two replication forks that move in opposite directions in an ATP dependent process. RPA coats the single-stranded DNA produced, preventing DNA damage and reannealing of the complementary strands from occurring.

Leading strand synthesis occurs continuously in a 5' to 3' direction, as the polymerases and replication fork move in the same direction on the leading strand. However, synthesis on the lagging strand leads to the production of small fragments, called Okazaki fragments (see Figure 2), as polymerases move the opposite direction to the replication fork on the lagging strand. As lagging strand

synthesis must also occur in a 5' to 3' direction, it requires synthesis of short RNA primers by a Pol α-RNA primase complex and these primers are then extended by Pol δ [23] (see Figure 2). DNA synthesis by these enzymes has been shown to occur at approximately 50 nucleotides/second in cells, which is similar to the rate of replication fork movement [10]. When Pol δ reaches the end of the preceding RNA primer, displacement synthesis is initiated leading to removal of the RNA primer and its replacement with DNA [24, 25]. Once this process has been successfully completed, the two Okazaki fragments are ligated by DNA ligase [26].

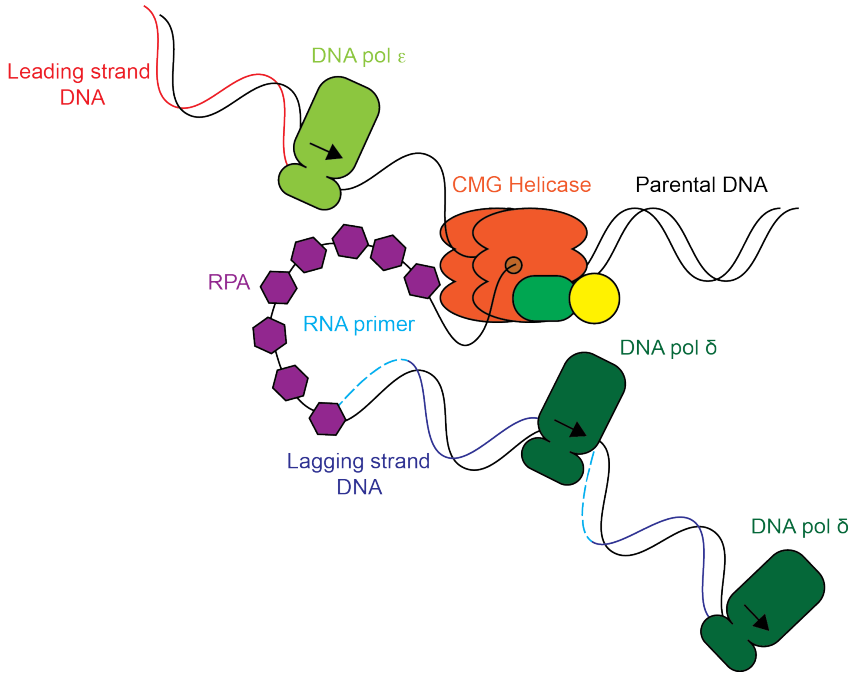


Figure 2: Leading and lagging strand synthesis during DNA replication. Single-stranded DNA is produced by the CMG helicase (orange) which unwinds the parental DNA in an ATP dependent process at the replication fork. ssDNA is then coated with RPA to prevent DNA damage and reannealing from occurring. Synthesis of DNA on the leading strand occurs continuously using Pol ε (light green) and the polymerase moves at a similar rate to the replication fork. On the lagging strand, short RNA primers (dotted blue) are added by the Pol α-RNA primase complex and these primers are then extended by Pol δ (dark green). After replacement of the RNA primer with DNA by Pol δ, ligation of two Okazaki fragments can occur using DNA ligase.

1.1.2 The end replication problem

It has long been known that DNA is lost from the ends of chromosomes during each DNA replication event within the cell cycle [5]. Initially, this process was attributed to the fact that DNA polymerase cannot fill in the final Okazaki fragment. In a model containing blunt-ended chromosomes formed on both the leading and lagging strands, leading strand synthesis continues to the end of the chromosome and removal of the final RNA primer on the lagging strand leads to loss of DNA which is at least the length of the RNA primer [27] (see Figure 3). The single-stranded DNA overhang produced from RNA primer removal cannot be converted into double-stranded DNA, due to the inability of the DNA polymerases to extend in a 3'-5' direction and this single-stranded DNA is lost producing a blunt end [27].

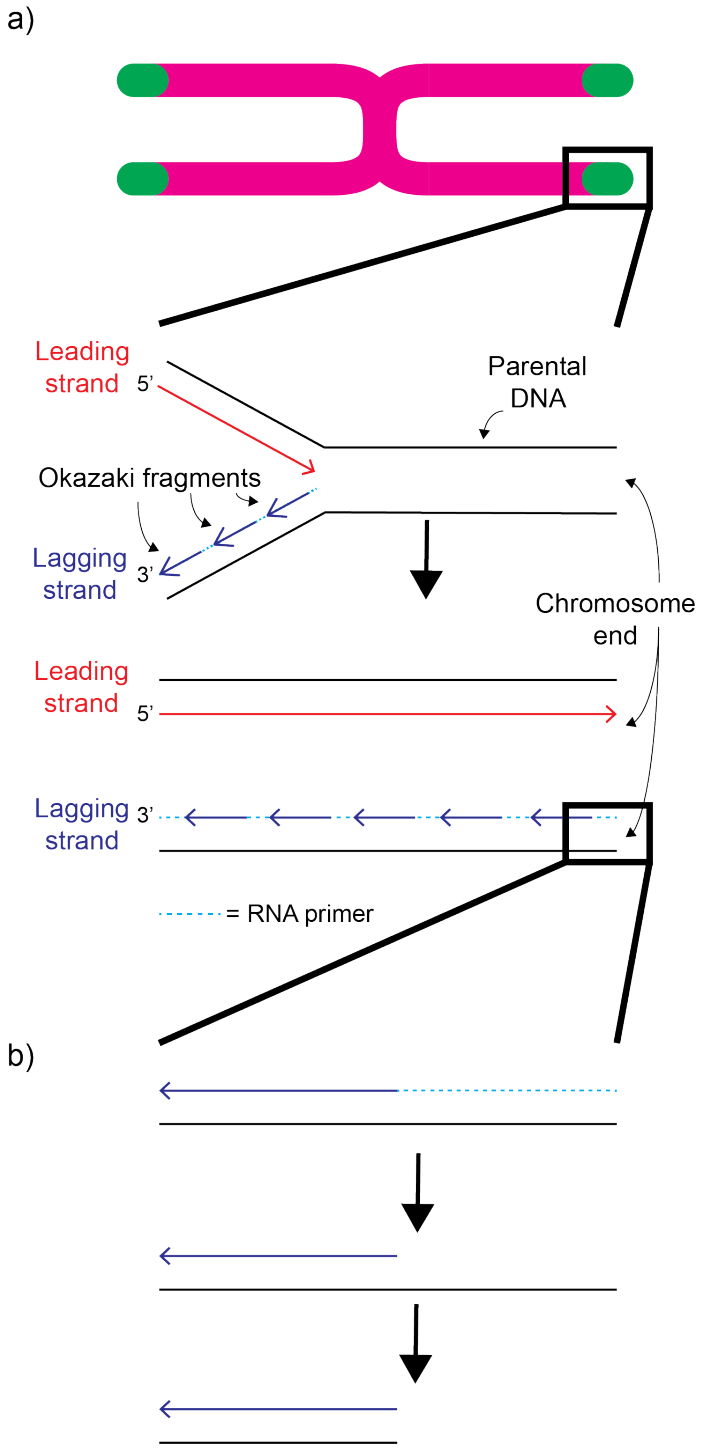


Figure 3: The end replication problem.
a) DNA replication on the leading strand occurs continuously all the way to the end of the chromosome, whereas on the lagging strand RNA primers are required. This leads to production of Okazaki fragments, as DNA synthesis can only occur in a 5'-3' direction. b) In a model containing blunt ended chromosomes, removal of the terminal RNA primer on the lagging strand leads to single-stranded DNA which cannot be filled in to re-form double-stranded DNA. Therefore, this length of DNA is lost, leading to chromosome shortening and the end replication problem.

At chromosome ends, the DNA strand containing the 3' end is guanine (G) rich and the complementary strand is cytosine (C) rich and these strands are known as the G- and C- strands respectively. Chromosome ends contain a G-rich single-stranded overhang which is required for telomeric shape and function [28] (see section 1.2.4). Production of this 3' overhang requires a more complex model for replication of DNA at chromosome ends.

Reproduction of a G-overhang on the leading and lagging strands occurs through different mechanisms. As above, leading strand extension proceeds all the way to the end of the template. To re-create the 3' overhang present before replication, resection of the C-strand occurs by the exonuclease activity of two proteins, Apollo and Exo1 [29]. At leading strand telomeres, the C-strand contains the parental template DNA. Resection produces an overhang that is longer than required, so this resected DNA is filled in, by a process resembling lagging strand synthesis [30]. The terminal RNA primer is removed, leaving an overhang [31]. This filling in process involves recruitment of the Pol α-RNA primase complex by the CST complex [32] (described in section 1.5.2).

On the lagging strand, DNA synthesis leads to production of Okazaki fragments. This means that during lagging strand replication the length of the overhang and DNA lost with each cell cycle is also set by the position of the terminal RNA primer. This final RNA primer can be placed as much as 100 bp from the end of the lagging strand, leading to the terminal 100 bp not being replicated [33]. Exo1 resection of the C-strand is also thought to occur at lagging strand telomeres.

The length of the overhang has been measured at each stage by a method which involves nuclease digestion of double-stranded DNA using a duplex-specific nuclease followed by analysis of the remaining single-stranded DNA by Southern blot-based assays [34]. Along with production of the G-overhang, this process causes successive loss of DNA during each cell cycle. In humans, this leads to loss of 50-200 bp of DNA during each cell division [33].

1.2 Telomeres and their function

1.2.1 What are telomeres?

Most organisms counter the end replication problem by having a length of non-coding DNA at the end of their chromosomes called a telomere, to prevent loss of coding DNA during replication [35] (see section 1.1.2). This non-coding DNA consists of short tandem repeats that are added by a ribonucleoprotein reverse transcriptase known as telomerase, which is found at a low abundance in cells. Expression of telomerase is supressed in somatic cells, which leads to a cell cycle dependent stepwise shortening of telomeres, known as replicative aging. When telomeres reach a specific short length cellular senescence occurs, which ultimately leads to apoptosis [36]. This mechanism provides a finite lifetime for cells (see section 1.3.1).

The exact sequence and number of telomeric repeats varies between species. Hypotrichous ciliates have telomeres <30 bp in length, human telomeres are <20 kb, while some laboratory mouse and chicken strains can have mega-telomeres that are in excess of 150 kb long [37]. It is thought that species with telomeres in excess of 20 kb in length do not supress telomerase expression in somatic cells and do not use replicative aging as a method to effectively count the number of cell divisions and prevent tumour progression [38]. An inverse correlation between telomere length and lifespan has been observed in vertebrates.

Telomerase expression was shown to have co-evolved with body size, as shorter-lived, smaller animals have long telomeres and express telomerase in somatic cells [38].

1.2.2 Telomere structure

In vertebrates the telomeric sequence on the G-strand is TTAGGG. On the other hand, the sequence is degenerate in fission yeast and contains $[G_{0-6}]$GGTTACA$[C_{0-1}]$. Other yeasts have different telomere repeat motifs, which vary in length from 5 to almost 30 bp [39] (see Table 1).

Table 1: Telomere sequences in different organisms.

Organism	**Telomeric sequence (5' → 3')**
Homo sapiens	TTAGGG
Dictyostelium	AG(1-8)
Tetrahymena	TTGGGG
Plasmodium	TTAGGG(T/C)
Bombyx mori	TTAGG
S. pombe	$[G_{0-6}]$GGTTACA$[C_{0-1}]$
S. cryophilus	GGGTTACTT
S. octosporus	GGGTTACTT
S. japonicus	GTCTTA, GGGTTTA, GTCTTAGGGTTTA
S. cerevisiae	$[TG_{2-3}][(TG)_{1-6}]$

As discussed in section 1.1.2, telomeres contain a G-rich overhang, termed the G-overhang, which is generated during DNA replication (see Figure 4). This overhang can be inserted into telomeric repeats leading to the formation of T-loops (introduced in section 1.2.4). The G-overhang also plays an important role in telomerase regulation and acts as a binding site for proteins central to chromosome end protection [40]. The regulation of the length of the G-overhang has also been shown to be a cell cycle dependent process. Telomeric repeats act as a binding site for a large number of regulatory proteins, including the shelterin complex, which have important roles in chromosome capping, shape and telomerase regulation [7].



Figure 4: The G-overhang.
Telomeres contain a G-overhang on the strand containing the 3' end. A single human telomeric repeat sequence is highlighted in bold and underlined. The G-overhang can range from 12 to several hundred nucleotides in humans [41] and the overhang length correlates with telomere length [42].

1.2.3 Telomerase

Telomeres are added to chromosome ends by a ribonucleoprotein reverse transcriptase called telomerase [43]. Blackburn and Greider first discovered this enzyme in 1984 in the ciliate *Tetrahymena thermophila*. Blackburn and Greider, along with Szostak who collaborated with Blackburn working on telomere

function, were awarded the 2009 Nobel Prize for medicine “for the discovery of how chromosomes are protected by telomeres and the enzyme telomerase [44].”

In a healthy organism telomerase activation is only observed during development, in stem cells and in other cells that are rapidly dividing. Telomerase activation in these cells is required to maintain a pluripotent population, as extension of telomeres by telomerase prevents replicative aging and re-sets the molecular clock within the cells [45].

This cell specific telomerase activation can be explained by the differential expression of telomerase in different cell types and also by the strict regulation of the enzyme by a large number of proteins. Tight regulation of telomerase prevents replacement of telomeric repeats lost during each replication cycle leading to telomere shortening. After a number of replication cycles telomeric DNA will reach a threshold short length, triggering senescence and apoptosis which prevents production of daughter cells with damaged coding DNA [46] (see section 1.3.1). Furthermore, as an organism ages, the initial telomeric length reduces as expression levels of telomerase in stem cells are not sufficient to totally prevent telomere shortening [47].

In fission yeast, telomerase comprises an enzymatic subunit (Trt1), telomeric RNA (TER1) and the accessory subunit Est1. This complex adds telomeric repeats by using TER1 RNA as a template for extension. TER1 also acts as the scaffold to bring together Est1 and Trt1 [48]. There is a high level of conservation

of the protein components of the telomerase holoenzyme within eukaryotes. In humans, the enzymatic subunit is called TERT and the telomeric RNA is called TERC. Humans also contain an Est1 homologue, also called Est1, which associates with telomeres and telomerase and is involved in telomere metabolism [49]. The mechanisms by which Est1 acts in humans remain unknown. In *S. cerevisiae*, there are three EST genes, which encode proteins that make up the budding yeast telomerase (Table 2). Deletion of two of the *S. cerevisiae* EST proteins, Est1 and Est3, leads to progressive telomere shortening, known as the ever-shorter telomere (EST) phenotype. Immunoprecipitation of Est2 also causes precipitation of the budding yeast telomeric RNA, showing this subunit to be essential for catalytic activity. Deletion of a fourth protein, CDC13, which is also known as Est4 also leads to the EST phenotype [50].

Table 2: Important telomerase subunits in humans and yeast.

	Human	***S. pombe***	***S. cerevisiae***
Telomeric RNA	TERC	TER1	TLC1
Holoenzyme subunits	TERT	Trt1, Est1	Est1, Est2, Est3

Whilst an RNA subunit is required for telomerase activity, there is a large amount of evolutionary divergence of sequence and length between different species. However, telomeric RNAs from different species do contain a number of conserved features [51]. These include a large enclosed loop, which contains the template for reverse transcription, and a downstream pseudoknot, which is

present in most eukaryotes and is required for telomerase function [51]. The pseudoknot is not found in trypanosome telomeric RNA which suggests that this feature was not present in the common ancestor between eukaryotes and trypanosomes and must have evolved later [51].

1.2.4 Telomere shape

1.2.4.1 T-loops

Electron microscopy and super resolution fluorescence imaging have shown the presence of loop structures at telomeres, called T-loops [52]. These structures play an important role in maintenance of genomic integrity at telomeres (see section 1.2.4.3). It has been shown that the N-terminal basic domain of human TRF2 is required for T-loops to form *in vivo* [53, 54]. This protein is a member of a chromosome capping complex called the shelterin complex which will be discussed in section 1.4. *In vitro* the rate of T-loop formation is very slow with just TRF2 present and so it is thought that other proteins may also be recruited to telomeres in addition to TRF2 *in vivo* to facilitate T-loop formation [55]. These proteins may include the MRN complex, which consists of MRE11, Rad50 and NBS1, and the RecQ family Werner syndrome and Bloom syndrome helicases, which could have a role in DNA unwinding prior to T-loop formation [56]. These helicases and the MRN complex have previously been shown to interact directly with TRF2, which leads to their recruitment to telomeres [57]. They also have

functions in double strand break surveillance, but during T-loop formation interactions with proteins located at chromosome ends prevent activation of a DNA damage response. Observation of T-loops in vitro requires protein-free DNA stabilised by introduction of interstrand crosslinks by psoralen and UV [58], a process that has not been successful in budding and fission yeast due to problems with purification of protein-free telomeric DNA. However, G-overhangs are observed in yeast, which points to the likely presence of T-loops [59].

1.2.4.2 T-loop evolution

It has been proposed that in a telomerase-free evolutionary ancestor, chromosome ends were stabilised and extended using a T-loop like system [60]. The first eukaryotes contained proteins involved in recombination dependent replication. Recombination-dependent replication involves invasion of DNA containing a 3' overhang into double-stranded DNA and subsequent DNA replication steps. This process occurs in prokaryotes as a method to reinitiate stalled replication forks that have encountered a lesion and also in bacteriophages for DNA replication [61]. If chromosome ends in the first eukaryotes contained a number of telomere-like repeats, then formation of a T-loop-like structure could occur. T-loops in humans are associated with proteins that have evolved from these original recombination dependent replication proteins, for example the MRN complex which is a functional homologue of T4-phage gp46/47 [60]. These proteins are involved in end resection to form the 3' overhang.

1.2.4.3 T-loop assembly

T-loop formation involves the insertion or strand invasion of the 3' overhang of the telomere into the double-stranded region of the telomere. This causes separation of the double-stranded DNA and replacement of the G-strand with the telomeric overhang to form a displacement loop (D-loop) [62]. Insertion of the 3' end occurs at a varying long distance from the end of the telomere, leading to production of a loop structure [52] (see Figure 5). T-loops conceal the 3' end of telomeres and therefore activation of ATM and the classical non-homologous end joining (c-NHEJ) pathways is avoided, preventing activation of a DNA damage response by the chromosome end. The T-loop structure also resembles an intermediate in the homologous recombination pathway known as a Holliday Junction (see Figure 5c). By comparing the structures formed in a D-loop and a Holliday Junction you can see significant similarities, and it is possible that by resembling an intermediate in a pathway, activation of a DNA damage response is avoided [60]. If TRF2 is deleted then T-loop formation is prevented and ATM and ATR activation can occur, ultimately leading to either apoptosis or senescence [63]. If a DNA damage response was not prevented, homologous recombination would lead to loss of the DNA contained in the T-loop, causing very short telomeres to be produced [64].

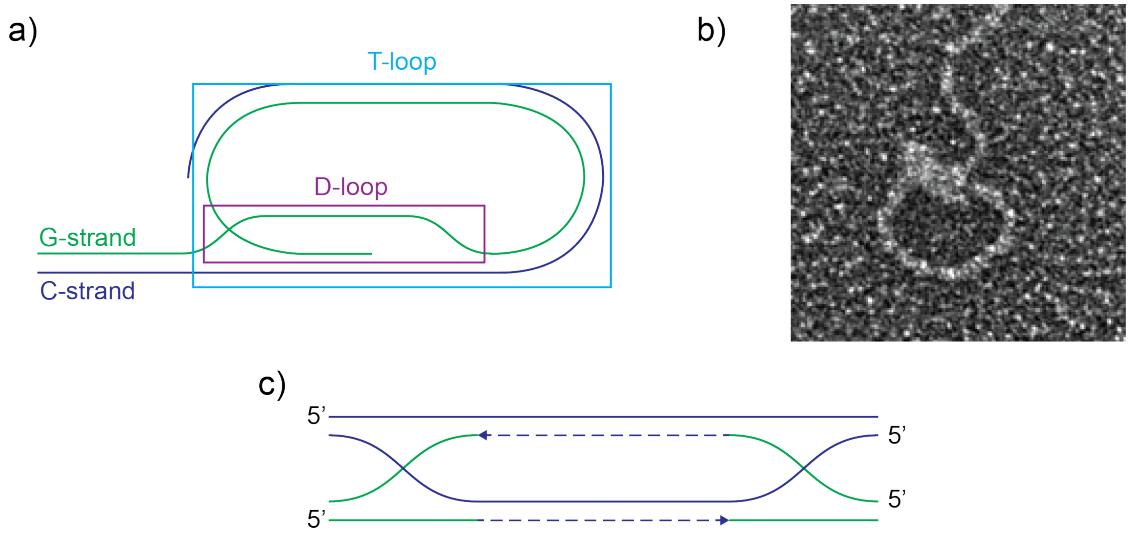


Figure 5: T-loop formation.
a) Insertion of this 3' overhang into double-stranded telomeric DNA leads to production of a D-loop as the G-overhang displaces the G-strand and base pairs with its complementary strand. This hides the chromosome end from the double strand break surveillance proteins and forms a T-loop. b) Electron micrograph showing a T-loop with TRF2 bound at its base [55]. c) Schematic showing a Holliday junction intermediate in homologous recombination. Holliday junctions and the structure formed at the base of a T-loop show significant similarities and this may be important for avoiding activation of a DNA damage response.

RTEL1 is a helicase required to promote disassembly of T-loops, which is required during S phase of the cell cycle for telomere replication to occur. RTEL1 also functions to remove G4-quadruplex structures from telomeric DNA, which can block progression of the replication fork [65]. G-quadruplex structures form because of the G-rich nature of telomeric DNA and occur due to hydrogen bonding between four guanine residues in a planar arrangement. These G-quartets can then stack on top of one another and may have functions in telomere protection and capping [66]. The RTEL1 helicase has D-loop disrupting activity, which leads to removal of T-loops [67]. RTEL1 interacts with TRF2 in a cell cycle-dependent manner, indicating that T-loop maintenance is regulated

throughout the cell cycle [68]. These actions of RTEL1 are important to maintain the integrity of telomere ends and help to prevent telomere loss and fragility [69].

1.3 Telomeres and disease prevention

1.3.1 The molecular clock and the Hayflick limit

Telomerase is expressed and active in embryonic stem cells, leading to maintenance of telomere length, allowing the cells to proliferate indefinitely. This is necessary as embryonic stem cells provide a pool of cells that can differentiate into any cell in the body. Telomerase expression is limited in somatic cells, leading to shortening of telomeres during each cell cycle and replicative aging [70]. Once the telomere reaches a critical short length, ~ 3 kb in humans, cellular senescence occurs [71]. This mechanism acts as a molecular clock and the limited capacity to divide is known as the Hayflick limit [71]. The Hayflick limit occurs after approximately 50 cell divisions in human cells [72]. This was initially discovered by the observation that cultured human cells will divide for a limited period, after which they stop growing and undergo replicative senescence or apoptosis [46]. It was observed that this molecular clock is unaffected by the cryogenic preservation of cells, as replicative senescence occurs after the same number of cell divisions independent of freezing [73].

The onset of replicative senescence after a pre-determined number of cell divisions plays an important role in disease prevention. Over time, the likelihood that DNA damage in a cell is not correctly repaired increases and this leads to a build-up of mutations. These mutations increase the risk of developing diseases

with a genetic basis, such as cancer [74]. Cancer cells do not undergo replicative senescence and one way in which they bypass the limits of the molecular clock is by reactivation of telomerase, leading to extension of telomeres [74] (see section 1.6.1).

1.3.2 Suppression of the DNA damage response

Faithful repair of DNA damage, without loss of sequence, is vital for the prolonged survival of cells. In humans, the DNA damage response is activated in response to a diverse range of DNA lesions which require a number of different pathways for repair [75, 76]. Amongst other outcomes, double strand breaks or persistent single-stranded DNA resulting from stalled replication forks can form. Chromosome ends resemble double strand breaks and so a focus will be placed on this outcome.

Repair of single-stranded DNA and double strand breaks occurs through two main pathways, the ATR pathway, which responds to persistent single-stranded DNA, and the ATM pathway, which leads to repair of double strand breaks [6] (see Figure 6). These two phosphatidyl inositol 3-kinase-related kinases are involved in recognition of DNA lesions and activation of the DNA damage checkpoint in the cell cycle. Activation of these kinases leads to accumulation of downstream effector proteins, including Chk1 and Chk2, which prevent cell cycle progression until repair of DNA damage has occurred [6]. Downstream repair of DNA double strand breaks is undertaken by homologous recombination and non-

homologous end joining. The fission yeast homologue of ATR is Rad3 and of ATM is Tel1 [77].

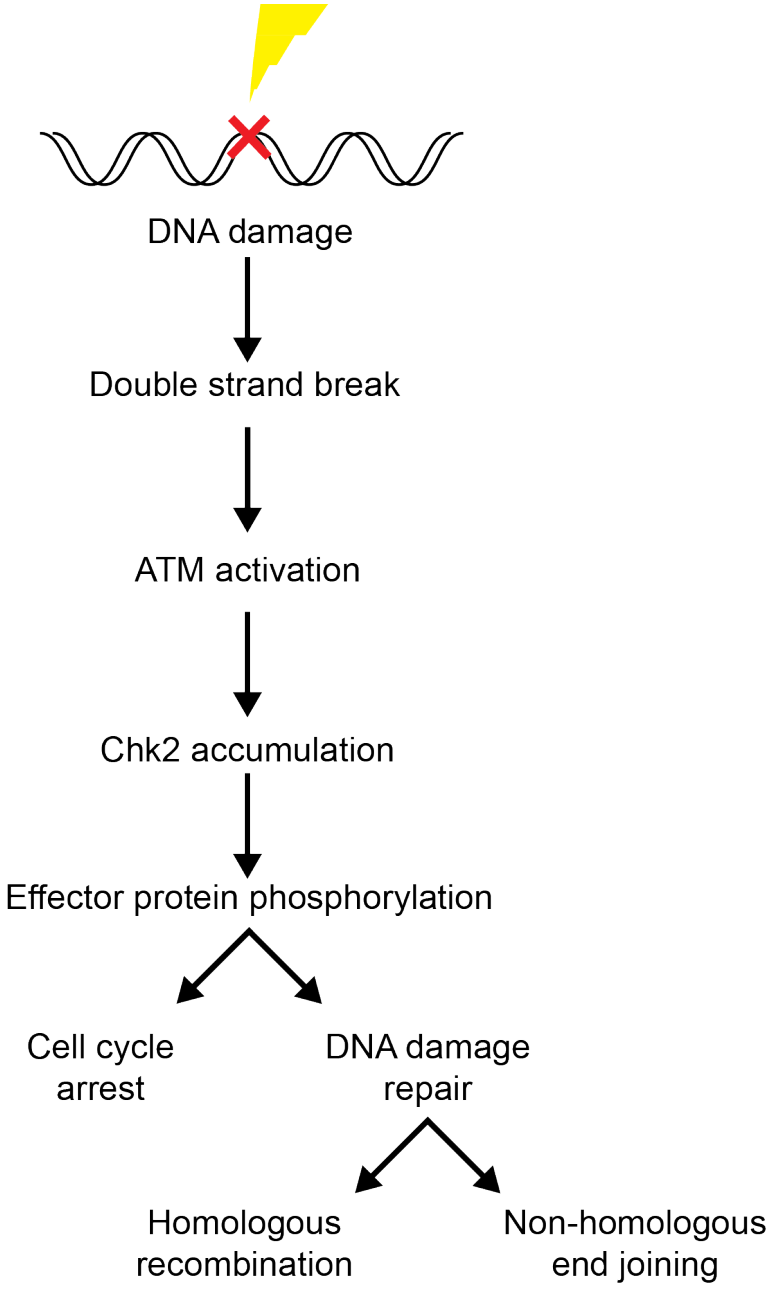


Figure 6: Schematic showing the DNA damage response to double strand breaks in DNA.
DNA damage can lead to formation of double strand breaks in the DNA. Sensor proteins, such as the Mre11-Rad50-NBS1 (MRN) complex, can detect this DNA damage and this signal can be passed on by transducers such as the ATR/ATM and Chk1/Chk2 kinases. Phosphorylation of effector proteins, such as the p53 tumour suppressor, can then occur, leading to cell cycle arrest and repair of the DNA damage by homologous recombination or non-homologous end joining.

Linear chromosome ends resemble DNA double strand breaks, but activation of the DNA damage response is prevented by the formation of T-loops and also the inhibitory effects of members of the shelterin complex (see section 1.4). In fission yeast, DNA damage response kinases Rad3 and Tel1 are also required to redundantly phosphorylate the shelterin complex component Ccq1 so it can recruit Est1 to telomeres before telomerase can become active [77] (see section 1.4.3). Even though activation of these kinases is required for phosphorylation of Ccq1 to occur, activation of the DNA damage response is prevented through a currently unknown pathway [78]. Strict regulation of activation of the ATR/ATM kinases is required, to favour activation of telomerase and prevent initiation of a DNA damage response [79].

1.4 Shelterin complex

Telomeric repeats act as a binding site for numerous regulatory proteins that have important roles in chromosome capping, shape and telomerase regulation. One such set of regulatory proteins comprises the shelterin complex [58]. The shelterin complex promotes the formation of T-loops which, along with the protein complex cap, prevents recognition of chromosome ends by double strand break repair proteins. In addition to avoiding incorrect activation of a DNA damage response, this also reduces the likelihood of homologous recombination between telomeres on different chromosomes as all share the same telomeric sequence [80].

1.4.1 Shelterin complex components in *S. pombe*

The shelterin complex contains six key regulatory proteins and forms a protein cap on the telomere. In fission yeast, the complex consists of Taz1, Rap1, Poz1, Tpz1, Pot1 and Ccq1 [81] (see Figure 7). Telomerase regulatory proteins have both positive and negative effects on telomerase activation. Members of the shelterin complex that act as negative regulators of telomerase are Taz1, Rap1 and Poz1, and Ccq1 positively regulates telomerase extension. Tpz1 and Pot1 can have both positive and negative regulatory roles.

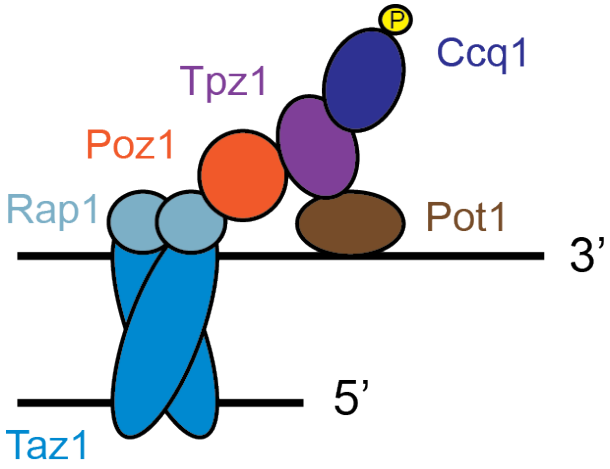


Figure 7: The shelterin complex.
In fission yeast, the shelterin complex contains 6 proteins, Taz1, Rap1, Poz1, Tpz1, Pot1 and Ccq1. Taz1 forms a homodimer and binds to double-stranded DNA and Pot1 binds to single-stranded DNA. Rap1, Poz1 and Tpz1 form a bridge between the double and single-stranded DNA binding proteins. Binding of T93 phosphorylated Ccq1 to Tpz1 can cause a switch in the telomere state from a non-extendible to an extendible state. Phosphorylated Ccq1 can interact with Est1, a telomerase accessory subunit, leading to its recruitment to telomeres.

Very little high resolution structural information for any of the shelterin components was available at the outset of this project. However, sequence analysis suggested a good deal of modularity in the architecture of all of the components which, alongside published yeast two-hybrid and biochemical studies [82, 83], provided a partial domain-wise interaction map of the complex (Figure 8).

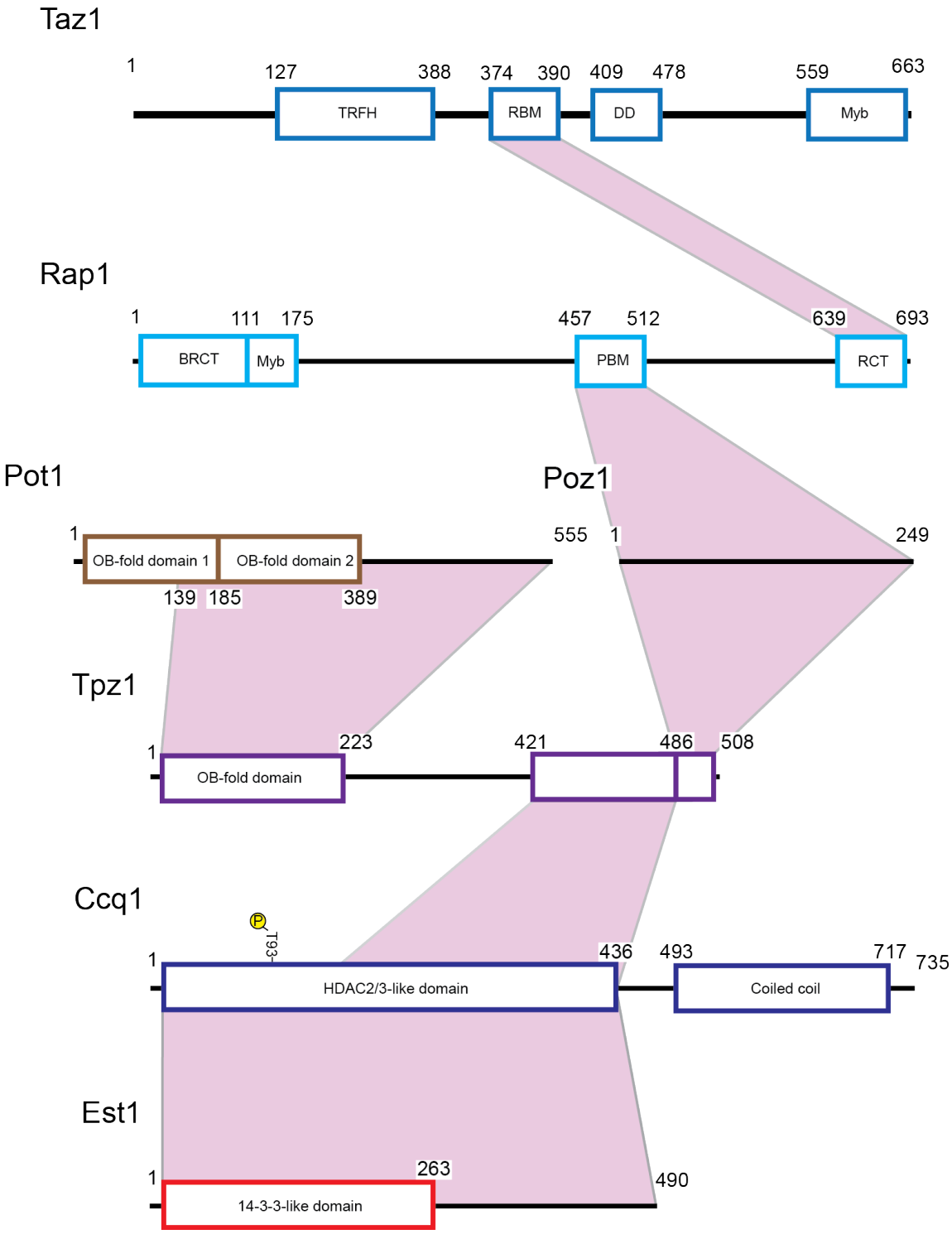


Figure 8: Interaction regions between shelterin complex components known at the outset of this project.
Single copies of each protein have been included to improve clarity. Taz1 binds to double-stranded DNA and is linked to Pot1, which binds to single-stranded DNA, through a bridge formed of Rap1, Poz1 and Tpz1. Thr93-phosphorylated Ccq1 recruits Est1 to the telomere and then binds to Tpz1.

1.4.1.1 Ccq1

Ccq1 is a 735 amino acid (aa) protein that contains a predicted C-terminal coiled-coil domain and a predicted HDAC2/3-like domain towards the N-terminus [84] (see Figure 9). A central region of the protein, from residues 123-436, has previously been shown to form an interaction with Tpz1 leading to recruitment of this protein to chromosome ends [83]. The region of the protein from residue 1-436 has also been shown to interact with the accessory subunit of fission yeast telomerase, Est1. Both of these binding regions were determined by truncational analysis of the protein using yeast two hybrid analysis [85].

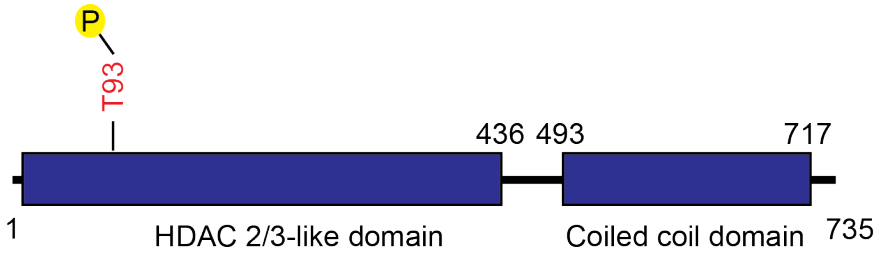


Figure 9: Domain structure of Ccq1.
Ccq1 contains an N-terminal predicted HDAC 2/3-like domain and a predicted C-terminal coiled-coil. Threonine residue 93 can be redundantly phosphorylated by Tel1 and Rad3, leading to an interaction with Est1.

Phosphorylation of Ccq1 occurs redundantly by Tel1 and Rad3 at threonine residue 93, which leads to an interaction with Est1 [85]. If Ccq1 is deleted in cells this leads to a reduction of approximately 200 bp in telomere length, due to the inability of cells to recruit Est1 [82]. Tel1/Rad3 deletion causes a failure to recruit telomerase to telomeres and total telomere loss [79, 86]. Binding of Ccq1 to Tpz1 has also been shown to cause a switch from a non-extendible to an extendible state at telomeres. This was studied using separation of function mutants of Tpz1, which disrupted specific interactions that this protein forms in the shelterin

complex [87]. Deletion of Ccq1 leads to telomere shortening *in vivo* [78]. Overall these data show that Ccq1 has a positive regulatory role for telomerase activation at telomeres. To date there is no high-resolution structural information for Ccq1 available.

1.4.1.2 Tpz1

Tpz1 is a 508 residue protein that contains a predicted OB-fold domain at its N-terminus [88] and two distinct binding sites for Ccq1 and Poz1 at its C-terminus (see Figure 10). An interaction has been shown between the N-terminal OB-fold domain and Pot1 by yeast-two hybrid analysis and co-immunoprecipitation assays [82, 83].

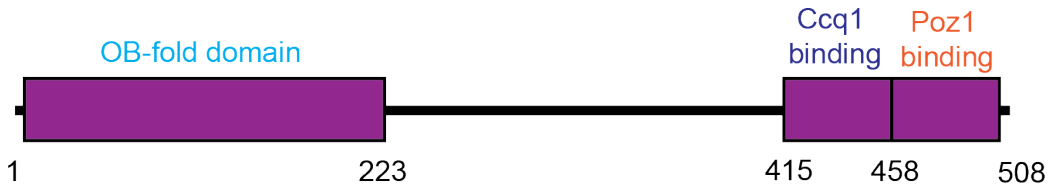


Figure 10: Domain structure of Tpz1.
Domain boundaries updated to reflect results obtained during this thesis. Tpz1 contains a predicted OB-fold domain at is N-terminus, which interacts with Pot1. Two distinct binding sites, one for Ccq1 and another for Poz1, are found towards the C-terminus of the protein.

Tpz1 plays an important role in the shelterin complex as it forms a bridge between double-stranded and single-stranded DNA binding proteins. Tpz1 can have both positive and negative regulatory roles in telomerase activation [89]. Tpz1 interacts with Ccq1 leading to the phosphorylation-dependent recruitment of Est1 to telomeres, an interaction that also leads to abrogation of the shelterin

complex-dependent repression of telomerase [83]. Binding of Ccq1 to Tpz1 has also been shown to facilitate phosphorylation of Ccq1 by Tel1/Rad3 [83]. Conversely, the interaction between Tpz1 and Poz1 promotes Poz1-telomere association, which limits phosphorylation of Ccq1 by Rad3 [82] and this prevents Est1 recruitment and telomere activation. Current published structures of Tpz1 only comprise a short 32 residue peptide that includes the Poz1 binding site (PDB: 5XXE) [90].

1.4.1.3 Poz1

Poz1 is a 249 aa protein that is predicted to be globular over its whole length. It interacts with a C-terminal region of Tpz1 and has been shown to be dimeric [83, 91]. Recently, a crystal structure of Poz1 bound to peptides from Tpz1 and Rap1 has been solved (PDB: 5XXF) [90] (see Figure 11). The structure of Poz1 resembles the structure of other TRFH domains (see section 1.4.1.5) which are mostly helical and contain a deep binding groove for their respective partners [92]. Dimerisation of Poz1 occurs through the first two α-helices of the protein found between residues 1-42. Binding of Poz1 to Tpz1 occurs in the groove between the N-terminal helices $\alpha1/\alpha2$ and $\alpha3/\alpha4$. Rap1 binds to $\alpha5$ and $\alpha9$ of Poz1 located on the opposite side of Poz1 to the Tpz1 binding site [91] (see Figure 11). Structural characterisation of the Poz1-Tpz1-Rap1 complex required production of a fusion protein containing Poz1 and Rap1, as the affinity of this interaction was not sufficient for the complex to form in solution [90].

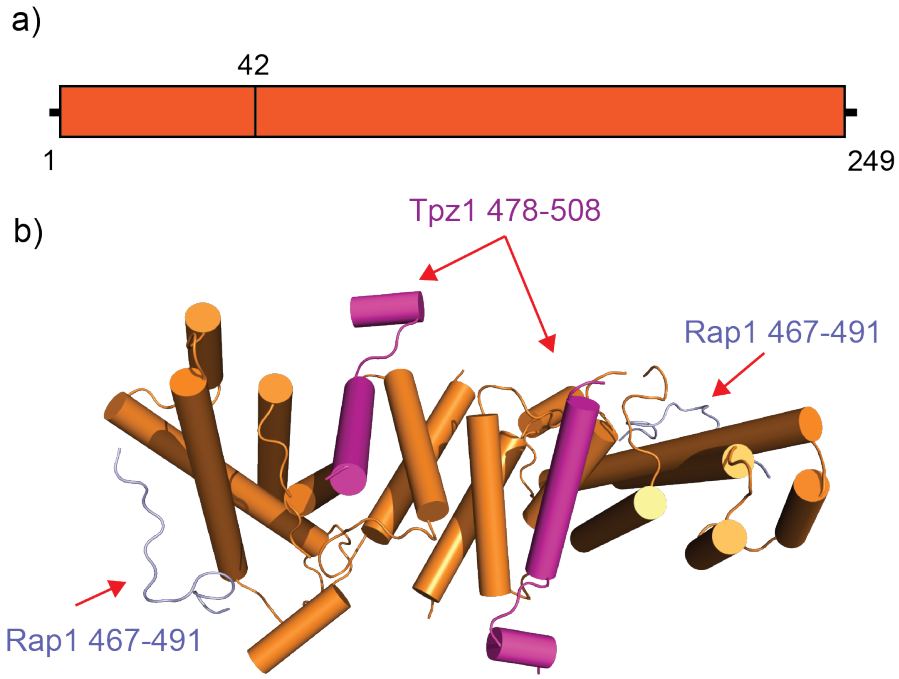


Figure 11: Domain structure of Poz1.
a) Poz1 is 249 aa in length and is predicted to be globular over its whole length. b) Crystal structure showing a homo-dimer of Poz1 (orange) with each Poz1 monomer bound to a Tpz1 peptide (purple) and a Rap1 peptide (blue) (PDB: 5XXF) [90].

Poz1 has been shown to have negative regulatory roles in telomerase activation. Deletion of Poz1 *in vivo* or disruption of the Poz1-Tpz1 interaction both lead to the production of elongated telomeres [93]. Poz1 also limits accumulation of Rad3 at telomeres which prevents Ccq1 phosphorylation [83]. These data show that Poz1 is required to maintain telomerase in an inactive state and prevent aberrant extension of telomeres.

1.4.1.4 Rap1

Rap1 is a 693 residue protein that contains binding sites for Poz1 and Taz1 towards the C-terminus (see Figure 12). The region from 457-693 has been shown to be sufficient for the telomere length control and end protection functions performed by Rap1 [94]. The BRCT-Myb domain containing region does not play a major role at telomeres [94]. Structural information for Rap1 is limited to crystallographic studies that include short peptides bound to Poz1 (see Figure 11) (PDB: 5XXF) or short Rap1-Taz1 fusion proteins (PDB: 2L3N) [95].

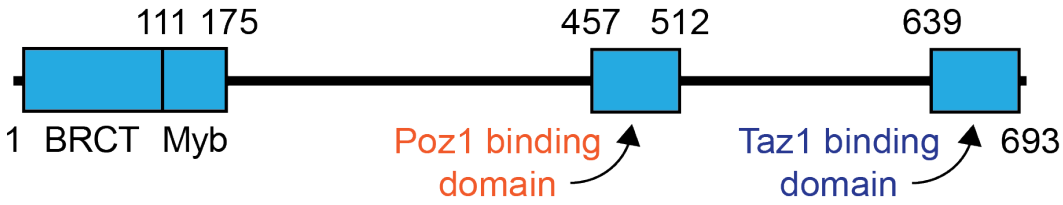


Figure 12: Domain structure of Rap1.
Rap1 contains Poz1 and Taz1 binding domains towards the C-terminus of the protein. It has been shown that only the C-terminal region (residues 457-693) are required for telomere length control and end protection, with the N-terminal region forming interactions with proteins involved in meiosis.

Rap1 is recruited to telomeres through an interaction with Taz1 and it then interacts with Poz1 through a central binding domain. Deletion of Rap1 or disruption of either the Poz1 or Taz1 binding interfaces lead to telomere elongation *in vivo*, suggesting that Rap1 also acts as a negative regulator of telomerase [96]. Disruption of Poz1 or Taz1 binding to Rap1 also leads to non-homologous end joining-dependent telomere end fusions, which resemble the fusions that occur when Rap1 or Taz1 are deleted, showing the importance of this protein for telomere end protection [93].

1.4.1.5 Taz1

Taz1 contains 663 residues and binds to double-stranded telomeric DNA through its Myb domain (see Figure 13). Unlike the TRFH domains of TRF1 and TRF2, which facilitate dimerisation of these proteins, the TRFH domain of Taz1 has a distorted configuration (PDB: 4ZMI) (see Figure 13b), which disrupts the equivalent dimerisation interface found in TRF1 and TRF2 [92, 97] and prevents it forming during Taz1 dimerisation. Taz1 dimerisation requires a coiled-coil region (see below). The TRFH domain of Taz1 has activities in telomere recruitment of Taz1 and telomere length regulation, as mutation of a patch of acidic residues within the TRFH domain of Taz1 leads to extended telomeres and phenotypes that resemble Taz1 deletion [98], however the mechanism of action of this acidic patch remains unknown.

Homodimerisation of Taz1 is mediated by a region between residues 409-478, which contains a coiled-coil (PDB: 4ZMK) [98] and this is thought to be important for the high affinity DNA binding of Taz1 (see Figure 13c). Recognition and binding to double-stranded telomeric repeats occurs through a conserved Myb module. This domain is a helix-turn-helix (HTH) domain motif, containing approximately 55 amino acids, typically found as a tandem repeat in eukaryotic transcription factors [99]. Most proteins that bind DNA with a Myb domain contain several copies of these domains. However, each Taz1 monomer contains a single Myb domain and the avidity effect is presumably mimicked by homodimerisation [98]. The structure of a Rap1 binding domain-Taz1 fusion

protein has also been solved by nuclear magnetic resonance and this shows significant structural similarity to the human Rap1-TRF2 interaction although the interaction affinity is 100-fold weaker (PDB: 2L3N) [95].

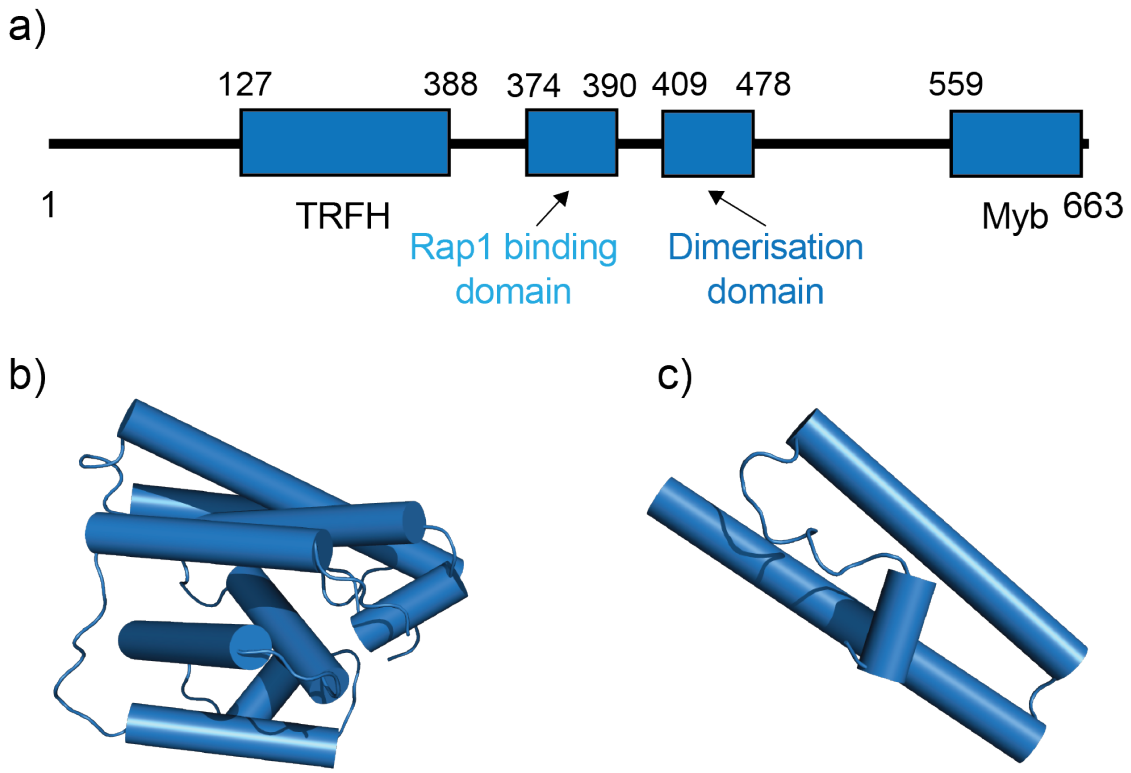


Figure 13: Domain structure of Taz1.
a) Schematic showing the domain structure of Taz1. Taz1 contains a TRFH domain towards the N-terminus, however this domain only has limited homology to other TRFH domains and it does not mediate dimerisation of the protein. Rap1 binds to a central region from residue 374-390 and homodimerisation of Taz1 is facilitated by a coiled-coil within residues 409-478. DNA binding occurs through the C-terminal Myb domain. b) Structure of the TRFH domain of Taz1 (PDB: 4ZMI). There is only limited homology between this TRFH domain and the TRFH domains of TRF1 and TRF2 which cause dimerisation of these related human proteins. This lack of homology explains why the TRFH domain of Taz1 does not facilitate dimerisation and instead has functions in telomere recruitment of Taz1 and telomere length regulation c) Structure of the dimerisation domain of Taz1 (PDB: 4ZMK) which forms an anti-parallel coiled-coil structure, with dimerisation occurring mostly through hydrophobic interactions.

Loss of Taz1 in cells leads to telomere elongation, deregulation of the G-overhang and lethal chromosome end-fusions in G1 phase of the cell cycle, showing that Taz1 acts as a negative regulator of telomerase and performs important actions in telomere end protection [96]. Phenotypes caused by deletion of Taz1 *in vivo* have some similarities to Rap1 deletion showing that some Taz1 functions, including telomere end protection, act in a Rap1-dependent pathway [96].

1.4.1.6 Pot1

The shelterin complex contains proteins that bind to both single-stranded DNA (Pot1) and double-stranded DNA (Taz1). Pot1 is a 555 aa protein that contains two OB-fold domains towards the N-terminus, which mediate binding to single-stranded DNA (see Figure 14). The structure of both of these OB fold domains in complex with DNA have previously been solved [100, 101] (PDB: 1QZG, 4HID) (see Figure 14). A C-terminal region encompassing residues 139-555 has been shown to be required for binding to Tpz1 by yeast-two hybrid analysis involving protein truncation mutants [82].

This conserved OB fold is found in other single-stranded DNA binding proteins such as Stn1-Ten1 [102]. Stn1-Ten1 is required to recruit the RNA primase-Polα complex to telomeres (see section 1.5.2). The structure of the first OB-fold domain of Pot1 shows residues important for specific recognition of telomeric repeats that rely on non-Watson-Crick base pairs formed within a specific GGTT

single-stranded DNA sequence and therefore this sequence is required for telomere-specific recognition by Pot1 [100]. Structures and biophysical analysis of the second OB-fold domain show that this can bind to a number of different single-stranded DNA sequences non-specifically with similar affinities [101]. Unlike human Pot1, the OB-folds in fission yeast Pot1 act independently and are separated by a 25 residue linker [101]. The structures of the OB-folds of Pot1 have only been solved separately.

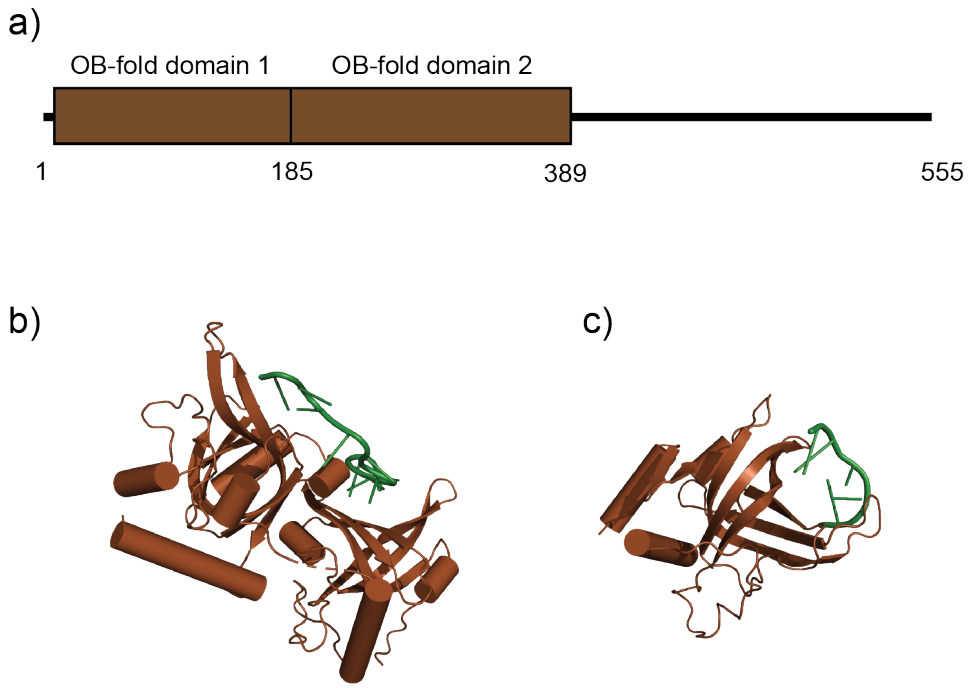


Figure 14: Domain structure of Pot1.
a) Schematic representation of the domain structure of Pot1. Pot1 contains two OB fold domains towards the N-terminus which interact with single-stranded DNA. A C-terminal region from residues 139-555 has been shown to be important for the Pot1-Tpz1 interaction. b) Structure of OB-fold domain 1 in complex with ssDNA (PDB: 1QZG). This domain confers the telomeric sequence specificity of the complex as it requires the formation of two pairs of non-Watson-Crick base pairs where guanine (G) pairs with thymine (T). A sequence containing GGTT is required for protein-DNA binding. c) Structure of OB fold domain 2 in complex with ssDNA (PDB: 4HID). This domain contains a conserved β-barrel structure found in other OB-folds, however can bind many different DNA sequences with similar affinity [101].

Pot1 has been shown to have both positive and negative regulatory roles within the shelterin complex. The linkage between Pot1 and other shelterin components is required to maintain the telomere in a non-extendible state, showing a negative role, and disruption of the Tpz1-Pot1 interaction leads to telomere elongation [87]. However, if Pot1 is deleted then telomere deprotection occurs, leading to telomere shortening and circularised chromosomes [103], showing a positive regulatory role of Pot1.

1.4.2 Shelterin homology

The fission yeast shelterin complex contains a high degree of functional, but not sequence, homology to the shelterin complex in mammals and has been more intensively studied due to the ease of genetic manipulation in the yeast system. Homologous proteins from *S. pombe* and *Homo sapiens* are shown in Table 3. The stoichiometry of human shelterin components was determined by SEC-MALLS to include 2x TRF1, 2x TRF2, 2x Rap1 and 1 of each of the other subunits [104]. Ccq1 is the only protein in the complex which lacks a clear functional homologue in mammals. However, it has been hypothesised that the human homologue of Poz1, TIN2, has an extended region, not found in Poz1, which may perform a similar role to Ccq1 in telomerase recruitment [105]. In addition, fission yeast shelterin only contains a single, homodimeric double-stranded DNA binding protein, Taz1, whereas in humans the functions of Taz1 are performed by a pair of homodimers of TRF1 and TRF2 [106-108]. TRF1 has been shown to mainly act in DNA replication at telomeres, while TRF2 is involved in T-loop formation

and prevention of ATM activation [106-108]. Mutational studies of shelterin complex proteins have identified residues critical for protein-protein interactions within the complex. However, structural information is still lacking for most proteins and their interactions within the shelterin complex.

Table 3: Shelterin complex orthologues from humans and *S. pombe*

Homo sapiens	***S. pombe***	**Homology**
TRF1/TRF2	Taz1	17 %
Rap1	Rap1	19 %
TIN2	Poz1	19 %
TPP1	Tpz1	20 %
Pot1	Pot1	22 %
-	Ccq1	-

1.4.3 Shelterin and telomerase regulation

1.4.3.1 Phosphorylation-dependent telomerase activation

In fission yeast, Ccq1 functions to recruit factors to telomere-bound telomerase and then to stabilise the complex at telomeres [85]. Phosphorylation of Ccq1 at threonine residue 93 (T93) leads to a direct interaction between Ccq1 and the 14-3-3-like domain of Est1 [85], a telomerase accessory subunit that can interact with both proteins and RNA. Yeast two-hybrid and RNA three-hybrid experiments have been used to map these interaction sites [85] (see section 1.5.1).

Ccq1 phosphorylation at T93 occurs redundantly by both Tel1 (ataxia telangiectasia mutated (ATM) homologue) and Rad3 (ataxia telangiectasia and

Rad3 related (ATR) homologue) and cells that lack both these kinases are unable to recruit telomerase to telomeres [83]. Interestingly, ATR/ATM phosphorylation has also recently been shown to play an important role in telomere association of telomerase in humans, but the substrates of these kinases involved in this process has yet to be identified [109]. TRF2, the human homologue of *S. pombe* Taz1, is important for repression of ATM signaling and non-homologous end joining [96]. Activation of these two processes is required for repair of DNA double strand breaks. Mouse Pot1a/b (*S. pombe* Pot1), along with stabilising interactions with other shelterin complex members, prevents replication protein A (RPA) binding to single-stranded DNA at telomeres [110]. This prevents ATR activation, which blocks activation of a DNA damage response [110]. A similar mode of repression of homologues of ATR and ATM might also lead to prevention of Ccq1 phosphorylation in the fission yeast system.

It has been shown that the abundance of shelterin components in mammalian cells is independent of telomere length and so there is an inverse correlation between telomere length and shelterin component density [111] (see Figure 15). If shelterin abundance also remains constant at fission yeast telomeres independent of telomere length, this would suggest that at longer telomeres, where there is a larger distance between shelterin complexes and greater accessibility of shelterin components, less Ccq1 phosphorylation could occur. The increased accessibility of shelterin components that can inhibit kinase activation might lead to less telomerase activation occurring at longer telomeres. Conversely, at shorter telomeres the shelterin complex components involved in

negative regulation of Ccq1 phosphorylation would be less accessible, due to the close proximity of other copies of the shelterin complex, and therefore may not be able to perform their inhibitory functions. This would allow accumulation of the DNA damage response kinases Rad3 and Tel1 and phosphorylation of Ccq1 to occur, leading to recruitment of Est1 to telomeres and extension by telomerase.

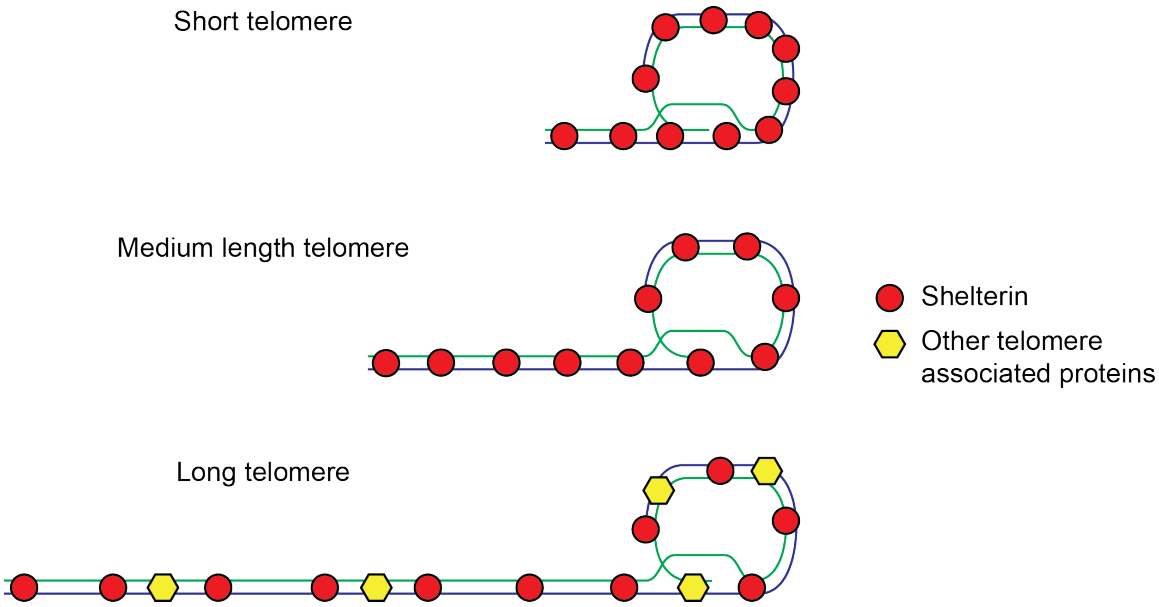


Figure 15: Shelterin abundance at telomeres is independent of telomere length. Telomeres are bound by the same number of shelterin complexes independent of their length and so at long telomeres these complexes are spaced out more along the DNA. This allows binding of other shelterin associated proteins, which are involved in trimming of long telomeres and will be discussed further in section 1.5.3.

1.4.3.2 Other regulatory actions of shelterin for telomerase recruitment

As mentioned previously, the fission yeast shelterin complex contains proteins that bind to nucleic acids. Taz1 binds to dsDNA while Pot1 binds ssDNA. These proteins are joined by a bridge, comprised of Rap1, Poz1 and Tpz1 which is important for accessibility of telomeres to telomerase and this general structure

is conserved between yeast and mammals. It has been demonstrated that the ability of Tpz1 to form part of this bridge is key for maintaining telomeres in a non-extendible state through analysis of separation of function mutants, while binding of Ccq1 can lead to switching of telomeres to an extendible state [87]. This state switch requires binding of T93-phosphorylated Ccq1 to Tpz1, but the mechanism of action of the switch remains unknown. The state switch could be due to an induced conformational change or recruitment of adaptor proteins.

Yeast two hybrid and glutathione-S-transferase (GST) affinity pull down assays have been used to determine that a C-terminal domain of Tpz1 (residues 421-486) interacts with a central region of Ccq1 (residues 123-436). Furthermore, the human homologue of Tpz1, TPP1, contains a cluster of residues also important for recruitment and stabilisation of the telomerase enzymatic subunit at telomeres, called the TEL-patch [112]. A TEL-patch-like region has been identified at the N-terminal end of Tpz1 and appears to perform similar functions through interactions with the TEN domain of Trt1 [112]. Along with these activities in telomerase recruitment, the TEN domain is thought to anchor telomerase to DNA whilst the RNA subunit is repositioned to allow reverse transcription of a new telomeric repeat to be performed [113].

The proposed mechanism for telomerase recruitment, activation and stabilisation at telomeres involves a two-step process which requires both T93-phosphorylated Ccq1 and the TEL-patch of Tpz1. Cell-cycle dependent T93-phosphorylation of Ccq1 by Tel1 and Rad3 primes telomeres for extension

and recruits the telomerase holoenzyme to telomeres by two separate interactions between Ccq1-phosphorylated T93 and Est1 [105]. Binding of the Tpz1 TEL-patch to Trt1 then further stabilises telomerase at the chromosome end (see Figure 16). Cross-linking mass spectrometry was used to confirm that both the HDAC2/3-like domain at the N-terminus of Ccq1 and phosphorylated residue T93 form important interactions with Est1 [112]. It was shown that mutations at either of these sites in isolation does not totally prevent binding between Ccq1 and Est1. Furthermore, overlap of the Ccq1-Tpz1 and Ccq1-Est1 binding sites was shown to occur, meaning that a complex between Tpz1-Ccq1-Est1 is likely to only occur transiently [112]. This adds more weight to the most recent model for telomerase recruitment that requires both a Ccq1 phosphorylation-dependent and Tpz1 TEL patch interaction with telomerase components to occur [112].

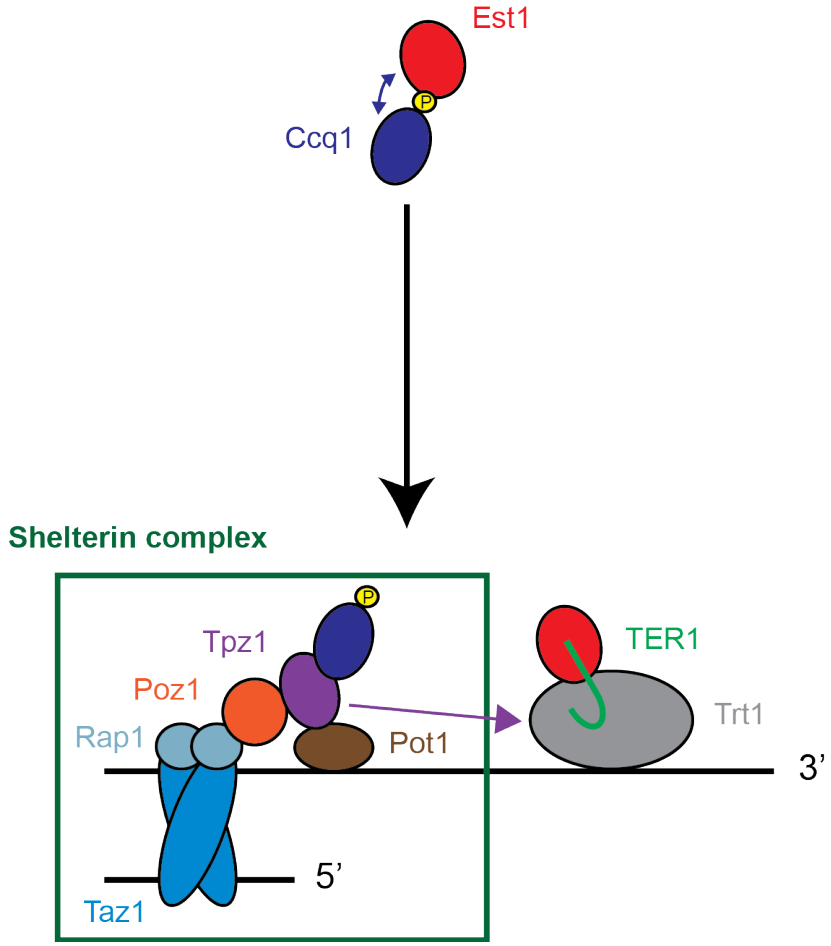


Figure 16: Proposed mechanism for telomerase recruitment and stabilisation. Ccq1 and Est1 interact, leading to recruitment of Est1 to the telomerase enzymatic subunit, Trt1. Est1 binding to TER1 stabilises telomerase at telomeres and a secondary interaction occurs between the TEL-patch of Tpz1 and the TEN domain of Trt1 (purple arrow). Binding of Ccq1 to Tpz1 leads to a switch in the telomere from a non-extendible to an extendible state.

1.4.4 Shieldin performs similar roles to shelterin, but at double strand breaks

Recently, a complex has been discovered that performs similar roles to the shelterin complex at double strand breaks in DNA rather than telomeres. This complex has been named the shieldin complex [114-117] and has been shown to prevent hyper-resection of DNA through an interaction with the CST complex (see section 1.5.2). This is similar to the process which occurs at telomeres to

form the G-overhang required for T-loop formation (see section 1.2.4). The shieldin complex further prevents over-production of single-stranded DNA by recruitment of the Pol α-primase complex leading to fill-in of resected DNA. The actions of shieldin prevents the mutagenic repair process of single-strand annealing from occurring [118] and leads to production of short single-stranded overhangs which are sufficient for homology-directed repair [119].

1.5 Other telomere associated proteins

1.5.1 Est1 and telomerase regulation

Est1 is a 490 residue protein that contains a 14-3-3-like domain towards its N-terminus (see Figure 17a). This domain mediates interactions with both T93 phosphorylated Ccq1 and TER1, the fission yeast telomeric RNA subunit [120]. This is the first 14-3-3-like domain known to interact with both phosphorylated residues and RNA.

14-3-3 domain containing proteins were the first to be identified that specifically bind to phosphorylated serine or threonine motifs [121]. This family of proteins is highly conserved and humans have seven 14-3-3 domain containing proteins [122]. Structures of all seven of these proteins have been solved and all human 14-3-3 domain containing proteins form homodimers [123, 124]. Each protein consists of a bundle of nine α-helices, of which α3, α5, α7 and α9 form a peptide binding groove required for recognition of phosphoserine or phosphothreonine by conserved residues (see Figure 17b). Helices α1, α3 and α4 are required for formation of the dimer [123]. 14-3-3 domain containing proteins do not undergo a conformational change upon ligand binding, which suggests these proteins could act through changes within the ligand upon phosphorylation dependent binding [125]. Phosphorylated residues in ligands of 14-3-3 proteins are often found in disordered regions and so changes in ligand protein structure, including

increased rigidity, are likely to occur upon binding [126]. Humans also have a number of proteins that contain 14-3-3-like domains and these include the nonsense-mediated mRNA decay proteins SMG5, SMG6 and SMG7. These proteins contain a 14-3-3-like domain adjacent to a helical hairpin domain, whereas 14-3-3 domain containing proteins usually contain a single functional domain [127]. *S. pombe* Est1 is predicted to contain a 14-3-3-like domain towards the its N-terminus and has a C-terminal region from residues 263-490 for which the confidence of secondary structure predictions is low and could potentially contain another domain. The 14-3-3-like domain is predicted to have a conserved fold and contain 3 of the 5 conserved residues required for phosphoserine/threonine binding [120] (see Figure 17c).

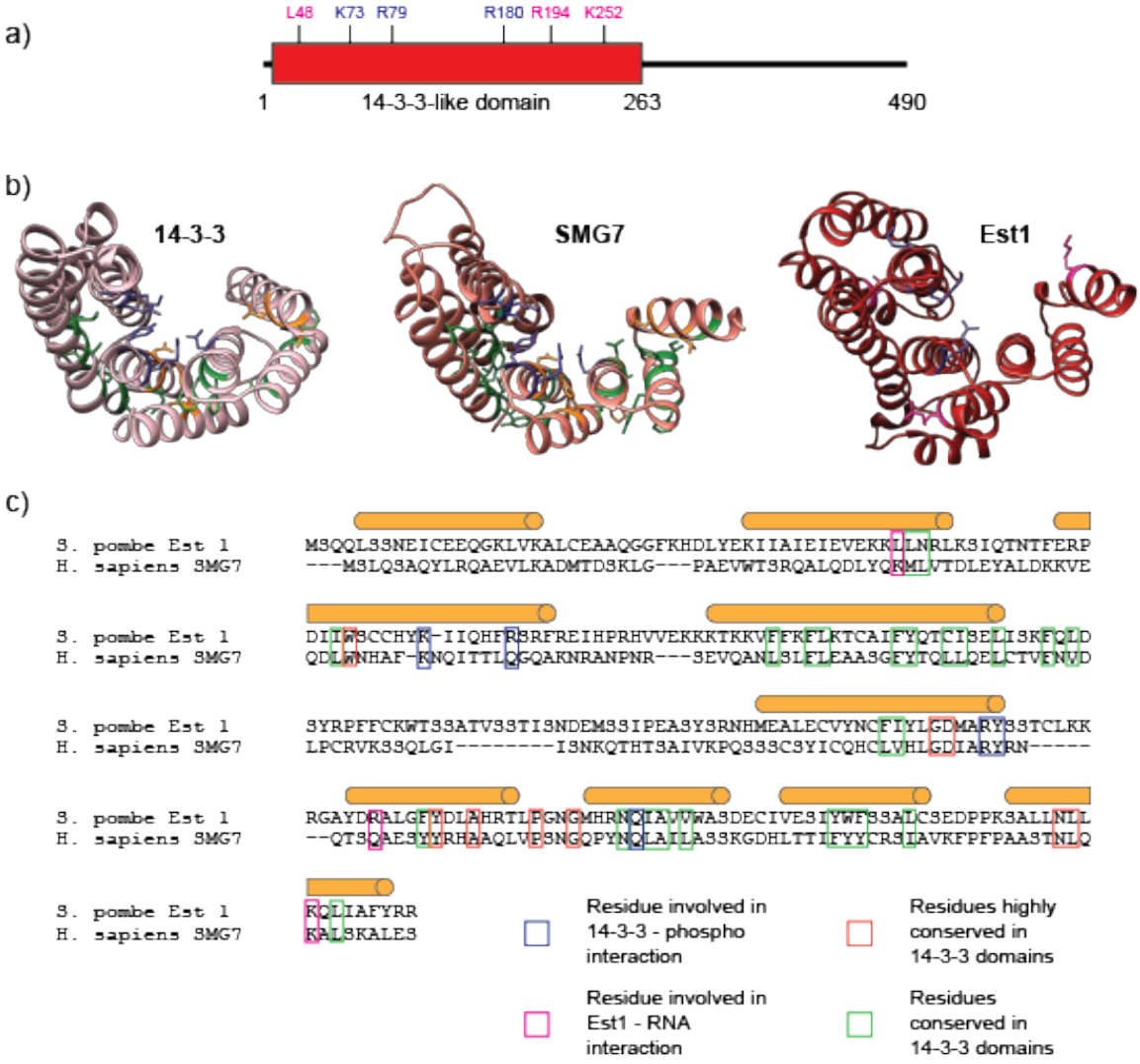


Figure 17: Domain structure of Est1.
The 14-3-3-like domain of Est1 is involved in binding of this protein to both phosphorylated Ccq1 and the TER1 telomeric RNA subunit. a) Schematic showing the known Est1 domain structure. Residues important for Ccq1 binding are shown in blue and for TER1 binding shown in pink. b) Crystal structures of a 14-3-3 domain (PDB: 1QJA) and SMG7 (PDB: 1YA0) and a Phyre2 model of Est1 showing a 14-3-3-like fold with conserved phospho-interacting residues. Colouring of residues is consistent with those shown in panel c. c) Structure-based sequence alignment for the 14-3-3-like domains of Est1 and SMG7 showing conserved residues required for the fold and phospho-binding activities.

The interaction between phosphorylated Ccq1 and Est1 has been shown by yeast two-hybrid experiments to require Est1 residues K73, R79 and R180 and mutation of these residues in fission yeast cells leads to telomere shortening [85].

This interaction, along with another between the HDAC2/3 domain of Ccq1 and Est1 have been verified by cross-linking mass spectrometry [105] (see section 1.4.3.2). The Ccq1-Est1 complex is then recruited to Trt1, the telomerase enzymatic subunit, where Est1 binds to TER1, the fission yeast telomeric RNA subunit, causing displacement of Ccq1 from Est1. Residues L48, R194 and K252 of Est1 were shown by yeast three-hybrid to be important for the TER1-Est1 interaction (see Figure 17b). Both the Ccq1 and TER1 interaction regions map to the predicted 14-3-3-like domain of Est1 and together these data suggest a partial overlap of binding sites for Ccq1 and TER1 on Est1 [120]. The identified RNA binding residues do not map to a specific region of the 14-3-3-like domain so it is likely that there are other residues that also mediate RNA binding. Est1 is required to stabilise Trt1 at telomeres and without Ccq1-dependent recruitment of Est1, Trt1 cannot become active [120].

Structural information for a human homologue of Est1, SMG7, is available (PDB: 1YA0) [127]. This protein is involved in the nonsense mediated decay pathway which functions include degradation of mRNA transcripts that contain premature stop codons [127]. Homology between SMG7 and Est1 was used to predict residues that would be important for interactions formed by Est1 [85]. There remains no structural information for fission yeast Est1.

1.5.2 The CST complex

A number of non-shelterin telomere associated proteins have been discovered. One of these is the CST complex, which in humans contains Ctc1, Stn1 and Ten1. In budding yeast Ctc1 is replaced by Cdc13, but currently a fission yeast homologue of Cdc13/Ctc1 has not been identified [128]. The CST complex is a replication protein A complex (RPA)-like complex that binds to single-stranded telomeric DNA with oligosaccharide/oligonucleotide-binding folds (OB-folds) through a similar mechanism to DNA binding by RPA [102].

The Stn1-Ten1 complex is structurally conserved between humans and yeast, but Cdc13 and Ctc1 vary in length, sequence and function [128]. Ctc1 directly inhibits human telomerase recruitment to telomeres [128]. Cdc13 interacts with a component of the budding yeast telomerase holoenzyme (Est1) leading to its telomeric recruitment and therefore Ctc1/Cdc13 have contrasting actions in humans and yeast [128]. Current data suggest that Cdc13/Ctc1 confers the telomeric sequence specificity of the complex, as the apparent binding affinity of the protein to DNA was dependent on the DNA sequence [129], with the other two proteins in the complex required to increase binding affinity and recruit telomerase and RNA primase-Polα to telomeres [130]. In humans, the telomerase enzymatic subunit (TERT) is recruited through a direct interaction between the TEN domain of TERT and the TEL patch of TPP1, independently of the CST complex [131]. The CST complex then inhibits telomerase after extension has occurred and also recruits the RNA primase-polα complex for C-

strand extension (as described in section 1.1.2). This telomerase inhibition is thought to occur via a direct interaction between CST and TPP1 [128].

Budding yeast only contains a homologue for Rap1 in the shelterin complex and so the CST complex can act as a cap for the *S. cerevisiae* telomere, preventing the accessibility of the telomere overhang to telomerase. The CST complex acts to down-regulate telomere extension, contrasting with the function of the complex in telomerase recruitment [132]. The switch between up- and down-regulation of telomere replication has been shown to be influenced by cell-cycle dependent post-translational modifications such as SUMOylation and phosphorylation of Cdc13. Binding of the budding yeast telomerase subunit Est1 to modified Cdc13 partially disrupts the Cdc13-Stn1 interaction and the CST cap structure [133, 134]. CST complex structure disruption allows telomerase to extend telomeres in budding yeast and this mechanism shows potential similarities to binding of T93 phosphorylated Ccq1 leading to a switch from a non-extendible to an extendible state [87].

Recently, it has been shown that Stn1-Ten1 in fission yeast has crucial activities in telomerase regulation and chromosome end replication. Stn1 contains a SUMO-interacting motif (SIM) and is recruited to SUMOylated Tpz1 after telomerase extension of telomeres has occurred. It is thought that recruitment of Stn1-Ten1 then inhibits telomerase activity [135]. Furthermore, it has been shown that Stn1-Ten1 promotes replication fork progression through telomeric and sub

telomeric regions and is required to restart DNA synthesis at stalled or collapsed replication forks [135].

1.5.3 TZAP and telomere trimming

A second non-shelterin telomere-associated protein, previously called ZBTB48 and renamed TZAP, has been recently been discovered in humans. This protein has been shown to bind to long telomeres that have a low density of shelterin components bound and is involved in the trimming of these long telomeres [136] (see Figure 18). This process is important as unusually long telomeric DNA leads to telomeric fragility and increase the risk of telomere fusions [137]. Trimming of telomeres leads to production of extra chromosomal telomeric DNA by removal of T-loops from the ends of chromosomes. Once telomere trimming has occurred the human shelterin component TRF2 facilitates the formation of a new T-loop, which prevents the chromosome end being exposed. TZAP competes for binding of telomeric sequence with TRF1 and TRF2 and therefore, at long telomeres more of the protein can bind [138] (see Figure 18). TZAP recognises telomeric repeats using a C-terminal C2H2 zinc finger domain and binding affinity is increased by binding of an adjacent loop [139]. TZAP is present as a homodimer and has no known enzymatic activity, so may act as an adaptor protein for binding of proteins that act in the trimming process [139].

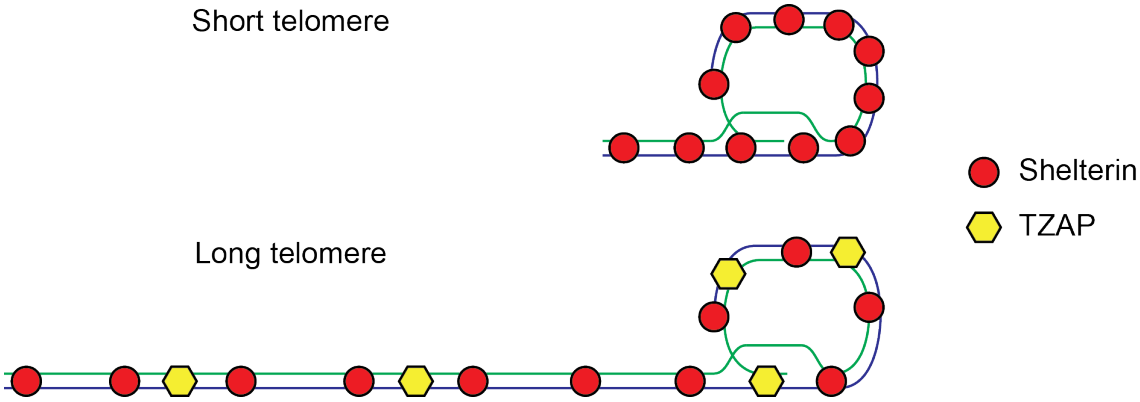


Figure 18: TZAP and shelterin compete for telomere binding.
TZAP competes with TRF1 and TRF2 for telomeric binding and so at long telomeres, where there is a lower density of shelterin bound to telomeres, more TZAP can bind. TZAP can act as an adaptor protein which leads to binding of proteins that act in telomere trimming.

1.6 Telomere length dysregulation and disease

1.6.1 Telomerase activation and cancer

Telomerase activation is normally only observed in stem cells and rapidly dividing cells [140], so DNA is lost from telomeres during each cell cycle in somatic cells. Once telomeric DNA reaches a specific short length after a number of cell divisions has occurred cells undergo senescence or apoptosis (see section 1.3.1). This will often take place through the activation of the p53 and pRB tumour suppressor pathways, leading to irreversible growth arrest. This prevents continued division of cells that have accumulated mutations over time and so acts as a mechanism to prevent diseases such as cancer [73]. If cells gain mutations in one of these tumour suppressor pathways then they can avoid senescence and continue to divide. Once many of the telomeres have reached a critically short length then cells reach a point known as crisis, during which there is increased genomic instability and large amounts of cell death. Rarely, cells may escape crisis and continue to divide, leading to cancer [141]. Escape from crisis occurs either through reactivation of telomerase or by the alternative lengthening of telomeres (ALT) pathway. Reactivation of telomerase occurs in approximately 85 – 90 % of cases where cells escape crisis [142] with ALT responsible for the remaining 10 – 15 %. The mechanism of ALT depends on homologous recombination to facilitate telomere elongation and is found in cancers with a particularly high mortality rate [143].

1.6.2 Telomere maintenance disorders

The shelterin complex is vital for maintenance of genomic integrity at telomeres, and many of the interactions within the complex have been shown to be required for survival. Deletion of shelterin components or abrogation of binding between shelterin complex proteins has been shown to be lethal [144]. In humans, cells containing a defective shelterin are subject to fast loss from the population due to strong negative selection. As a result of the requirement for a fully functioning shelterin complex, only mutations in TIN2 and TPP1 have been observed to lead to telomere disorders. Mutations in these shelterin complex proteins as well as in other parts of the telomere maintenance machinery cause a number of related rare genetic diseases, known as telomeropathies (shown in Table 4). These conditions generally cause impaired telomere maintenance and can ultimately lead to bone marrow failure, due to the dysfunction of telomerase in stem cells contained in the bone marrow [145].

The most common telomere maintenance disorder is idiopathic pulmonary fibrosis (IPF) and there are between 50000-100000 sufferers in the USA alone. Although many of the causes of IPF are unknown, one-sixth of all monogenic IPF cases are caused by mutations in TERT and TERC and the disease is characterised by parenchymal lung scarring, along with telomere shortening related defects such as bone marrow failure [146].

Dyskeratosis congenita (DC) is a rare telomere maintenance disorder, which is caused by a wide range of mutations that can be autosomal-dominant, recessive, X-linked or occur *de novo* (see Table 4). This disease gives rise to typical symptoms of mucosal leukoplakia, abnormal skin pigmentation, and nail dystrophy [147]. These symptoms have also been observed to be associated with an increased risk of cancer. The most common mutations that cause DC have been observed in the genes that encode dyskerin and TIN2. Hoyeraal–Hreidarsson syndrome (HH) is a related disorder that shares many of the same symptoms as DC. However, in many cases the genetic basis of the disease is yet to be determined. There are several other variations of DC, which have similar symptoms, for example Revesz syndrome [148]. This syndrome presents during childhood and resembles DC, but retinopathy and intracranial calcifications also occur. Mutations that cause Revesz syndrome have been mapped to the TIN2 gene and several of these have been shown to disrupt the interaction between TIN2 and TRF1 [149]. This leads to patients with the disease having very short telomeres, which may explain the early onset of the disease [149].

Among other rare telomere disorders is Werner syndrome, which is a disease characterised by premature aging of the sufferers and is caused by mutations in the RecQ family of helicases, which have actions in T-loop formation [150] (see section 1.2.4.1).

Table 4: Telomeropathies and the genes involved.
Table adapted from *Armanios et al.* [151]

Gene	Mutations cause:	References
TERT/TERC	Familial idiopathic pulmonary fibrosis	[152-162]
	Sporadic idiopathic pulmonary fibrosis	
	Aplastic anaemia	
	Autosomal dominant dyskeratosis congenita	
	Familial myelodysplastic syndrome - acute myeloid leukaemia	
DKC1 (dyskerin)	De novo dyskeratosis congenita	[163-166]
	X-linked recessive dyskeratosis congenita	
	Hoyeraal-Hreidarsson syndrome	
TINF2 (TIN2)	De novo dyskeratosis congenita	[167-169]
	Autosomal-dominant dyskeratosis congenita	
	Hoyeraal-Hreidarsson syndrome	
	Revesz syndrome	
NOP10	Autosomal-recessive dyskeratosis congenita	[170]
NHP2	Autosomal-recessive dyskeratosis congenita	[171]
TCAB1	Autosomal-recessive dyskeratosis congenita	[172]
CTC1	Coats plus syndrome	[173-175]
	Autosomal-recessive dyskeratosis congenita	
WRN	Werner syndrome	[150]
ACD (TPP1)	Aplastic anaemia	[176]

In humans, the assembly of the telomerase holoenzyme occurs in Cajal bodies after recruitment by TCAB1, which becomes part of the active telomerase enzyme. Cajal bodies are sites within the nucleus where RNA processing occurs, allowing binding between the telomeric RNA and protein subunits [177, 178]. Assembly of the enzymatic subunit (TERT) and the RNA template (TERC) with accessory proteins including dyskerin, NOP10, NHP2 and GAR1 then occurs through a currently uncharacterised process. Binding and retention of these proteins with the telomerase enzymatic subunit are required for the telomerase enzyme to become active. Mutations in these proteins lead to impaired telomerase function and shortened telomeres and have been observed in families with telomere disorders, such as DC [179].

1.6.3 T-loops and telomeropathies

In healthy cells, RTEL1 is required to remove T-loops prior to DNA replication. Mutations that prevent this process from occurring cause loss of large amounts of telomere sequence. When RTEL1 recruitment is prevented, T-loops are excised by the SLX1-SLX4 nuclease complex and the telomeric sequence contained in the loop is lost [5]. Defective recruitment of RTEL1 has been implicated in HH syndrome [69].

Along with shelterin, mutations in the CST complex, specifically CTC1, have been identified in mice that lead to telomeropathies [173-175]. These mutations have been implicated in rapid telomere loss due to the inability of the mutant complex to restart stalled replication forks and facilitate C-strand fill in, which is the process of C-strand synthesis required during G-overhang production [147]. Despite the CST complex being conserved between mice and humans, differences are observed and it is thought that telomere shortening upon loss of the CST complex in humans is due to binding of RPA to extended G-overhang regions which leads to activation of a DNA damage response [180]. In humans, mutations in the CST complex have been implicated in dyskeratosis congenita and coats plus syndrome [173-175].

1.7 Aims and objectives

At the outset of this project, there was limited structural information for many of the proteins and interactions involved in the vital processes that protect chromosome ends. Strict regulation of telomerase activity and telomere shape is crucial for maintenance of healthy cells and genomic integrity. This project aimed to provide molecular insights into the activities of proteins and complexes involved in these processes. Fission yeast was chosen as it bears significant similarities to the human system and studies in *S. pombe* allow for easy genetic manipulation during downstream *in vivo* experiments.

Initially, aims were to structurally characterise the interactions of the shelterin component Ccq1, its phosphorylation-dependent interactions with the 14-3-3-like adaptor Est1 and the Est1-TER1 protein-RNA interaction. These interactions are potentially competitive and occur upstream of telomerase activation. These goals were then expanded to include a more detailed biophysical and biochemical analysis of Ccq1, characterisation of its interactions with the Tpz1-Poz1 bridging complex and finally, structural analysis of the shelterin complex by electron microscopy.

Chapter 2. Materials & Methods

2.1 Molecular biology

2.1.1 Vectors

E. coli expression of recombinant proteins was performed using pET22b, pET24b, pET28b, pET47b, pET49b, pET52b (Novagen), pET52SUMO (Vangelis Christodoulou, Francis Crick Institute) or pGEX6p1 (GE Healthcare) plasmid vectors. Insect cell transfer vectors used were pIEx-Bac3 (Novagen), pTriEx6 (Novagen) and pFUBB (McDonald Lab, Francis Crick Institute) (see Table 5). The pET52SUMO vector was modified in house by Vangelis Christodoulou to include both N-terminal 6x His and SUMO tags. Inserts were cloned into pET47, pET49b, pET52b and pET52SUMO by ligation independent cloning (see section 2.1.5.1). Inserts were introduced into other *E. coli* expression vectors by cloning using restriction digestion of vectors and inserts followed by ligation (see section 2.1.7). In-fusion cloning was used for pIEx-Bac3 and pTriEx6 vectors (see section 2.1.5.2) and Gibson assembly for pFUBB (see section 2.1.6).

Gene expression was controlled by either the *tac* or T7 promoters. The pGEX vectors use the *tac* promoter and to allow expression of proteins encoded by these vector constructs, IPTG displaces the *lac* repressor protein (LacI) from the

lac operator sequence, which allows *E. coli* RNA polymerase to transcribe the gene of interest. pET vectors use the T7 promoter, which has a more complex mode of action. DE3 strains of *E. coli* are required to express proteins cloned into these vectors. In these cells IPTG induction causes LacI displacement from the *lac* promoter found upstream of a copy of T7 RNA polymerase which has been added into the *E. coli* genome. This allows transcription and translation of T7 RNA polymerase which can then transcribe genes under the control of a T7 promoter. In both cases transcription of the gene of interest is induced by addition of IPTG.

Table 5: Plasmids used for cloning and protein expression

Vector	Antibiotic resistance	Purification tag	System
pET22b	Ampicillin	C-terminal 6 x Histidine (His) fusion	*E. coli*
pET24b	Kanamycin	C-terminal 6 x His fusion	
pET28b	Kanamycin	N-terminal thrombin cleavable 6 x His fusion	
pET47b LIC	Kanamycin	N-terminal 3C cleavable 6 x His fusion	
pET49b LIC	Kanamycin	N-terminal 3C cleavable glutathione S-transferase (GST) and 6 x His fusion	
pET52b LIC	Ampicillin	N-terminal 3C cleavable Strep fusion	
pET52SUMO LIC	Ampicillin	N-terminal 3C cleavable SUMO and 6 x His fusion	
pGEX6p1	Ampicillin	N-terminal 3C cleavable GST fusion	
pIEx-Bac3	Ampicillin	N-terminal 10 x His fusion	Insect cells
pTriEx6	Ampicillin	C-terminal 10 x His fusion	
pFUBB	Ampicillin	None	

2.1.2 Bacterial strains

DH5α (Invitrogen) or Top10 cells (Invitrogen) were used for plasmid amplification and BL21 (DE3) (New England Biolabs) and Rosetta2 (DE3) (Novagen) cells were used for protein overexpression. BL21 Gold (DE3) (Agilent) was used for both protein overexpression and for plasmid amplification. Insect cell

transformations were performed with MultiBacDH10 cells (in house) (see Table 6).

Table 6: Genotypes of bacterial strains used.

Bacterial strain	Genotype
DH5α (Invitrogen)	F– φ80l*acZ*ΔM15 Δ(*lacZYA*-argF)U169 *recA1 endA1 hsd*R17(rK–, mK+) *phoA sup*E44 λ– thi-1 *gyr*A96 *relA1*
Top10 (Invitrogen)	F– *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ*ΔM15 Δ*lacX*74 *recA1 ara*D139 Δ(*ara*-leu)7697 *galU galK* λ– *rpsL(StrR) endA1 nupG*
BL21 (DE3) (New England Biolabs)	fhuA2 *[lon] ompT gal* (λ DE3) *[dcm]* Δ*hsdS*
Rosetta2 (DE3) (Novagen)	F- *ompT hsdSB*(rB- mB-) *gal dcm (DE3) pRARE2* (CamR)
BL21 Gold (DE3) (Agilent)	*E. coli* B F– *ompT hsdS*(r – m –) *dcm*+ *Tetr gal λ(DE3) endA Hte*
MultiBacDH10 (adapted in house)	F– *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ*ΔM15 Δ*lac*X74 *recA1 endA1 ara*D139 Δ(*ara*-leu)7697 *galU galK λ– rpsL nupG / bMON14272 / pMON7124*

Cloning strains have genetic markers that allow these *E. coli* cells to produce high quality DNA. RecA1 has a single point mutation in the *recA* gene, preventing the DNA recombination activity of this protein, stabilising DNA inserts [181]. The *endA1* marker has a mutation in the nonspecific endonuclease I gene, improving the quality of plasmid DNA purified from these cells. Strains used for protein expression have markers that allow for the induction of protein expression and prevent degradation of expressed proteins. Rosetta2 cells contain the pRARE2 plasmid that expresses tRNAs for 7 rare codons, potentially enhancing expression of eukaryotic proteins containing codons rarely used in *E. coli*. The *ompT* and *lon* mutant strains are deficient in these proteases, leading to improved

protein stability. The *lon* protease is involved in degradation of unfolded proteins [182] and *ompT* degrades foreign proteins [183]. Blue/white screening for recombinant colonies is possible when using the MultiBacDH10 cells due to the *lacZ*ΔM15 marker which involves partial deletion of the *lacZ* gene. Plasmids used for cloning contain the α-peptide which can complement that *lacZ* partial deletion, leading to production of functional β-galactosidase. This functional enzyme can hydrolyse X-gal to form a blue coloured product. A multiple cloning site is included in the centre of the α-peptide which allows cloning into this locus to occur. If successful incorporation of DNA occurs this disrupts the α-peptide, preventing functional β-galactosidase from forming [184]. Therefore, cells containing insert DNA form white colonies.

2.1.3 Primer design

DNA primers for PCR amplification were designed using the CCD crystallisation construct designer [185], which allowed translation start/stop sites to be incorporated in coding regions that do not contain predicted secondary structure. Oligonucleotide primers were synthesized by Sigma and were designed for use in the cloning protocols below. Typically, codon optimised DNA sequences were used for proteins shown in Table 21.

2.1.4 Polymerase chain reaction (PCR)

PCR reactions were typically carried out in 50 μL total volume using KOD Hot Start DNA Polymerase (Merck) in 200 μL thin-walled PCR tubes (Abgene) (Table 7).

Table 7: Reagents required for a typical PCR reaction.

Reagent	Amount
2 x KOD Hot Start Master Mix	25μL
Template	10-50 ng
Forward Primer (10 μM)	1.5 μL
Reverse Primer (10 μM)	1.5 μL
H_20	To 50 μL total volume

Thermal cycling was performed as per the manufacturer's instructions (see Table 8) using a DNA Engine Dyad thermocycler (MJ Research). The lid was set to a constant temperature of 100 °C and the annealing temperature was set at a constant temperature in a range from 55-65 °C, dependent on primer melting temperatures. After completion of PCR, the products were purified using a Qiaquick PCR purification kit (Qiagen).

Analysis of DNA after PCR was performed using a 1 % agarose gel made with 1xTris-borate-ethylene diamine tetra-acetic acid (TBE) buffer containing 1x of the staining agent SYBR Safe (Thermo Fisher Scientific). Gels were developed in 1x TBE buffer at 70 V for 30 minutes, with samples alongside a 100 bp plus or 1 kb plus DNA ladder (Invitrogen).

Table 8: Thermal cycling protocol for a typical PCR using KOD Hot Start DNA Polymerase.

Step	Temperature	Time
1	98 °C	2 mins
2	98 °C	20 secs
3	55-65 °C	10 secs
4	70 °C	20 secs/kb
5	Cycle to step 2 30-35 times	

2.1.5 Ligation independent cloning (LIC)

2.1.5.1 LIC protocol (developed *in house*)

Ligation independent cloning (LIC) was used for the majority of cloning due to its suitability for high throughput studies. Large numbers of clones can be produced, without the need for the digestion or ligation steps required for restriction endonuclease-based procedures.

LIC cloning exploits the dual 3' to 5' exonuclease proofreading and 5' to 3' polymerase activity of T4 DNA polymerase. When deoxynucleotides are not present, the enzyme functions mainly as an exonuclease. If one species of deoxynucleotide is present in the reaction then the polymerase excises bases until it reaches a base that can be re-inserted by its polymerase activity. This is exploited in LIC to generate specific single-stranded DNA overhangs. If these overhangs are sufficiently long (>10 bases), then DNA fragments containing complementary overhangs can be annealed and transformed into competent *E.*

coli. The bacterial repair machinery can seal the nicks in the transformed DNA, meaning that *in vitro* ligation is not required.

Table 9: Reagents required for T4 DNA polymerase treatment of PCR products.

Reagent	Amount
Purified PCR product	0.2 pmol
T4 DNA polymerase (1 U/μL)	0.5 μL
T4 DNA polymerase buffer (NEB)	2 μL
25 mM dATP	2 μL
H_2O	To 20 μL total volume

PCR primers used for the LIC system contain 5' extensions with their first thymidine base after 10 bases. T4 DNA polymerase treatment of PCR purified products is performed with added dATP, leading to production of a single-stranded overhang of 10 bases (see Table 9). The polymerase reaction was allowed to proceed at room temperature for 30 minutes before heat inactivation of the enzyme at 75 °C. Modified LIC ready vectors were linearised by restriction digestion before being T4 DNA polymerase treated. The vector treatment was performed in the presence of dTTP, to generate complementary overhangs to the insert. 1 μL of the treated vector at 50 ng/μL was mixed with 2 μL of treated insert (0.02 pmol) and incubated at room temperature for 5 minutes. 1 μL of 25 mM EDTA was then added and the mixture was incubated for a further 5 minutes. 2 μL of this annealed product was then transformed into chemically competent *E. coli* strains such as DH5α (Invitrogen) or Top10 cells (Invitrogen) (see section 2.1.9).

2.1.5.2 In-fusion cloning

An alternative and highly efficient method of ligation-independent cloning is the In-Fusion system (Clontech). Similar to the LIC protocol (section 2.1.5.1), 15 bp overhangs are produced on the vector and insert. However, primers are designed to generate complementary overhangs during PCR amplification of both the vector and insert. PCR was undertaken using KOD HS DNA Polymerase (Merck) following the manufacturer's recommended protocol. Incorporation of the insert into the vector was then performed in the presence of In-fusion master mix and then transformed into chemically competent *E. coli* strains suitable for DNA propagation such as DH5α and TOP10. In-fusion generates a nicked product, which can then be filled in by the bacterial repair machinery.

2.1.6 Gibson assembly

Primers were designed using the NEBuilder online primer designer tool to include complementary overhangs between vectors and inserts. PCR amplification of the vector and inserts were performed with KOD HS (Merck) or PfuUltra HF (Aglient) using the manufacturer's recommended protocols. Gibson assembly of PCR purified products was then performed using the manufacturer's protocol. For 2 fragment assembly a vector: insert ratio of 1: 2 was used. When assembly of more fragments was required, equimolar ratios were used. Dependent on the DNA concentration recovered after PCR purification of fragments, 0.03-0.2 pmols

of DNA was used for 2 fragment assembly and 0.2-0.5 pmols of total DNA was used for 6 fragment assembly (see Table 10).

Table 10: Typical reaction components for Gibson assembly.

	2 fragment assembly	**6 fragment assembly**
DNA molar ratio	Vector: insert = 1: 2	Vector: insert = 1: 1
Total amount of DNA	0.03-0.2 pmols	0.2-0.5 pmols
NEBuilder HiFi DNA Assembly Master Mix	5 μL	5 μL
H_2O	To 10 μL total volume	

Samples were incubated at 50 °C for 15 minutes when two fragments were assembled and 60 minutes when up to 6 fragments were assembled. 2 μL was transformed into chemically competent *E. coli* as previously.

2.1.7 Restriction enzyme based cloning and ligation

Several of the vectors used for cloning, such as pGEX, pET22, pET24 and pET28, were not compatible with the LIC system. Incorporation of insert DNA into these plasmids was achieved by a restriction double digest followed by ligation of the digestion products. PCR primers were designed to amplify the desired DNA sequence with 5' overhangs that contain complementary restriction sites. 1 μg of purified PCR product was digested with 10 units of the required restriction endonuclease (NEB) and the target vector was treated with the same enzymes. Reactions were typically carried out for 1 hour at 37 °C using the appropriate buffer system. In some cases, treatment of the vector with Antarctic phosphatase (NEB) was required to prevent re-ligation of the vector DNA. Digestion products were separated by agarose gel electrophoresis and purified by gel extraction

using a Qiaquick gel extraction kit (Qiagen) prior to ligation. 37.5 ng of treated PCR product was mixed with 50 ng of digested vector in a total volume of 20 μL. This reaction also contained 1 μL of Quick T4 DNA Ligase (New England Biolabs) and 2 μL of the supplied ligase buffer (see Table 11). The reaction was incubated at room temperature for 10 minutes before 2 μL was transformed into DH5α (Invitrogen) or Top10 cells (Invitrogen).

Table 11: Reagents required for ligation with T4 DNA ligase.

Reagent	Amount
Digested insert DNA	37.5 ng
Digested vector DNA	50 ng
T4 DNA ligase	1 μL
10X T4 DNA ligase buffer	2 μL
H_2O	To 20 μL total volume

2.1.8 QuikChange site-directed mutagenesis

Site-directed mutagenesis was performed with the QuikChange II site-directed mutagenesis kit (Agilent) using the manufacturer's standard protocol. Primers were designed to contain the required mutation flanked by ~10 nucleotides of complementary sequence with a guanine or cytosine base at each end. PCR was performed using KOD HS polymerase (Merck), using the reaction conditions described in section 2.1.4. DpnI digestion of the plasmid template DNA was performed, leaving the unmethylated mutated PCR products intact. The PCR mixtures were incubated for 1h with 10 U of DpnI at 37 °C. 2 μL of this reaction was then transformed into chemically competent *E. coli*.

2.1.9 Transformation and sequencing

1-5 µL (5 ng) of recombinant vector was added to 25-50 µL of chemically competent *E. coli* from the desired strain (see Table 6) and incubated on ice for 15 minutes. The cells were then incubated at 42 °C for 45 seconds, which was followed by a further 5 minutes incubation on ice. 200 µL of LB was added to transformed cells and the mixture was incubated at 37 °C for 45 minutes. 50-100 µL of this transformation mixture was plated onto LB agar supplemented with the appropriate selective antibiotic and incubated at 37 °C overnight. Single colonies were picked and grown in 5 mL of antibiotic supplemented LB overnight, before bacterial pellets were harvested by centrifugation and purification of plasmid DNA using a HiSpeed MiniPrep kit (Qiagen) was performed. Purified plasmid DNA was then sent for sequence validation (GATC Biotech) using standard primers. If cultures were required for protein overexpression, a single colony was used to inoculate 50-200 mL of antibiotic supplemented LB and grown overnight as a starter culture.

MultiBacDH10 cells are electrocompetent and therefore 0.5 µL (<1 ng) of transfer vector was added to the cells and incubated on ice for 30 minutes prior to electroporation using a MicroPulser (BioRad). After electroporation, 500 µL of SOC (Novagen) was added and the cells were incubated for >4 hours at 37 °C. 50-200 µL of this transformation mixture was then spread onto LB agar plates supplemented with kanamycin, tetracycline, gentamycin, X-gal and IPTG at appropriate concentrations. Plates were incubated at 37 °C for 2-3 days. White

colonies were then picked and grown overnight in 3 mL supplemented LB cultures. Cultures were then pelleted by centrifugation followed by processing with the Qiaquick miniprep kit (Qiagen) with the following protocol variation. After addition of P1, P2 and N3 buffers, the resulting mixture was centrifuged at full speed for 20 minutes before the supernatant was transferred to a new Eppendorf tube. 800 µL of 100 % isopropanol was added and after mixing the sample was incubated on ice for 5 minutes. Pelleting of bacmid DNA was then achieved by centrifugation at 13 000 rpm for 15 minutes. The supernatant was discarded and the pellet was washed with 500 µL of 70 % ethanol. The supernatant was discarded again and the bacmid pellet was air dried. Bacmid pellets were re-dissolved directly before transfection.

2.2 Protein production

2.2.1 Small scale (1 mL) protein expression in *E. coli*

Single colonies were picked from transformation plates produced as described in section 2.1.9. A 500 µL starter culture for each construct was grown in antibiotic supplemented LB in a 96-well plate at 37 °C shaking at 400 rpm overnight. 1 mL expression cultures in LB supplemented with 505, M and $MgSO_4$ (Table 22 and *Studier* [186]) were inoculated with starter culture at a dilution factor of 1/100. After growth at 37 °C with shaking at 400 rpm for 6 hours the temperature was lowered to 18 °C and overexpression of the protein of interest was induced by the addition of isopropyl-β-D-thio-galactoside (IPTG) to a final concentration of 1 mM. Cultures were grown for a further 16 hours overnight. Cells were harvested in plate by centrifugation at 4000 rpm.

2.2.2 Large scale (>500 mL) protein expression in *E. coli*

Starter cultures were grown as described in section 2.1.9. Typically, 12 x 500 mL LB cultures were inoculated at a dilution of 1/100 and grown with shaking at 210 rpm in 2 L conical flasks at 37 °C until an OD_{600} of 0.6 - 0.8 was reached. The temperature was then lowered to 18 °C and overexpression of the protein of interest was induced by the addition of IPTG to a final concentration of 1 mM. The cultures were then grown for 16 hours overnight at 18 °C. Cells were

harvested by centrifugation at 4000 rpm for 18 minutes and pellets were either processed directly or stored at -20 °C.

2.2.3 Expression in insect cells

Bacmid DNA was produced as described in section 2.1.9. 1 mL of SF21 insect cells at 1 x 10^6 viable cells/mL, determined using a Vi-CELL XR (Beckman Coulter), was added to each well of a 6 well plate. The cells were incubated at 28 °C for an hour to allow adherence of the cells to form a monolayer. Meanwhile, the bacmid DNA was resuspended in 20 µL of water filter sterilised by centrifugation through a 0.22 µM filter. DNA concentration was then determined by UV spectroscopy using a NanoDrop 2000 spectrophotometer (Thermo Scientific).

500 ng of bacmid was added to 100 µL of SF900IISFM (Gibco) media and 10 µL of FuGENE HD (Promega) added along with a further 100 µL of media. The mixture was incubated at room temperature for >15 minutes before the addition of 800 µL of media. After cell adhesion had occurred, all media was removed from the 6 well plate, prior to addition of 1 mL of transfection mixture. This plate was incubated at 28 °C for 4-6 hours, after which 1 mL of media containing a final concentration of 5.6 µg/mL gentamycin was added. This was then incubated for 72 hours at 28 °C. The resulting virus containing supernatant, called the P1 sample, was removed and added to 50 mL of SF21 insect cells in suspension culture at 1 x 10^6 cells/mL. The sample was incubated at 28 °C with shaking at

120 rpm for 72 hours before cells were harvested by centrifugation at 1500 rpm for 20 minutes to generate a P2 amplification sample. At this point the amount of virus containing supernatant required for the next stage of amplification is dependent on the infectivity of the virus produced. This can be observed by measuring the viability of cells. The lower the viability, the less virus is required. A cell viability of between 70-85 % is expected. Between 1-50 mL of this P2 sample was added to cells at 1 x 10^6 cells/mL to a total volume of 250 mL. After a 72 hour incubation at 28 °C the cells were harvested by centrifugation to generate a P3 supernatant sample. Again, the viability of the cells determined the volume of virus for the next infection. Between 1-50 mL of this P3 sample was added to cells at 1 x 10^6 cells/mL to a total volume of 250 mL. After a further 72 hours cells were harvested to produce a P4 amplification virus stock. The viral titre was determined using the BacPAK quantitative polymerase chain reaction (qPCR) titration kit (Clontech) following the manufacturer's instructions.

The P4 virus stock was then used for protein overexpression. Between 1-5 mL of the P4 sample was added per 100 mL of cells at 1 x 10^6 cells/mL used for expression. 300-500 mL of cells were grown in each 2 L roller bottle. Several different viruses can be added simultaneously for co-infection of the cells to occur. The total volume of cells grown in each repeat was 4 L. During expression experiments cells were grown for 72 hours as previously and cells were harvested by centrifugation. The supernatant was discarded and cells were either processed immediately or frozen at -20 °C.

2.2.4 Cell lysis

Cell pellets were resuspended in 300 mL of lysis buffer 1 which typically contained 50 mM Tris pH 7-8, 300-500 mM NaCl, 0.5 mM TCEP, 10 mM $MgCl_2$ and four Complete EDTA free protease inhibitor tablets (Roche). Proteins that were purified by a first step of nickel or cobalt affinity chromatography were resuspended in the above lysis buffer supplemented with 10-25 mM imidazole pH 8. The pH of the buffer was selected to be at least 1 pH unit above or below the theoretical isoelectric point (pI) of the protein or complex being purified, as predicted by protparam (http://web.expasy.org/protparam/), while avoiding high or low pH. The isoelectric point coincides with a point at which proteins have low solubility. Cells were lysed by sonication on ice at 40 % power on a 50 % duty cycle (Branson Sonifier 450) for 8 minutes. Where large sample volumes were produced, cell lysis was performed by high pressure homogenisation in two passes at a pressure of 10000-15000 psi using an EmulsiFlex-C5 (Avestin). Lysates were clarified by centrifugation in a Beckman JA-25.50 rotor at 20000 rpm for 45 minutes at 4 °C.

Small scale (<1 mL) *E. coli* expression tests were lysed by an osmotic lysis protocol. Cells were resuspended in 0.5 mL of 20 mM Tris pH 7.7 by shaking at 10 °C for 10 minutes. 0.5 mL of 40 % sucrose, 4 mg/mL lysozyme, 4 mM EDTA was added and was shaken for a further 20 minutes at 10 °C. A further 1 mL of 20 mM Tris pH 7.7 was added and the mixture was shaken for 20 minutes at 10 °C. 2 mL of 20 mM Tris pH 7.7, 60 mM Imidazole pH 8, 600 mM NaCl, 0.04 %

Triton X-100, 20 % glycerol, 10 mM $MgCl_2$, 2 units DNase I, 2 mM TCEP, 0.2 mM PMSF was mixed with the samples and left on ice for 20-30 minutes for lysis to occur. Lysates were finally clarified by centrifugation at 4000 rpm for 1 hour at 4 °C.

For lysis of insect cells, which lack a cell wall, milder conditions using a total sonication time of 3 minutes at 40 % power and 50 % duty cycle were employed.

2.2.5 Histidine-tagged protein purification

Proteins tagged with either a 6 x or 10 x Histidine tag were subjected to nickel or cobalt affinity chromatography. 3 mL of Nickel Sepharose Fast Flow (GE Healthcare) or TALON resin (Takara Biosciences) was added to the clarified lysate and was mixed for >20 minutes at 4 °C. The mixed slurry was then poured into a gravity flow column. The flow through was reapplied to the column and the beads washed with at least 10 column volumes of imidazole containing lysis buffer to remove any non-specifically bound proteins. Typically, 25 mM imidazole pH 8 was included in the wash buffer, but this varied depending on how strongly the tagged protein interacted with the resin. Bound proteins were eluted by addition of 5 mL aliquots of 50 mM Tris pH 7 or 8, 300 or 500 mM NaCl, 0.5 mM TCEP, 250-500 mM imidazole pH 8. Fractions were collected and analysed by SDS PAGE (see section 2.2.9).

In some cases, cleavage of the His tag was performed by addition of human rhinovirus 3C protease (3C). 35 µg of a 3C-GST fusion was added per mL of affinity resin. Cleavage occurred in lysis buffer 1 (see section 2.2.4) overnight at 4 °C with rotation. 100 µL of glutathione Sepharose 4B (GE Healthcare) resin was added to the mixture to remove the protease. Elution of the cleaved protein was then performed by gravity flow. Samples were analysed by SDS PAGE. All subsequent purification steps were performed at 4 °C where possible.

2.2.6 Glutathione Sepharose affinity chromatography

All proteins that contain a glutathione-S-transferase (GST) tag were purified using glutathione Sepharose 4B resin (GE Healthcare). 3 mL of resin was added to clarified lysate and the mixture stirred slowly at 4 °C for >30 minutes. The mixed slurry was then poured into a gravity flow column and the flow through was reapplied to the column. The beads were washed with >10 column volumes of lysis buffer to remove any non-specifically bound proteins. Bound proteins were eluted by repeated addition of 5 mL aliquots of lysis buffer 1 supplemented with 10 mM reduced glutathione pH 8 and fractions were collected.

If GST tag removal was required, 3C cleavage was performed as described in section 2.2.5. In this case, 3C and free GST both bind to the glutathione sepharose resin and elution of the cleaved protein occurs. Samples were analysed by SDS PAGE.

2.2.7 Ion-exchange chromatography (IEX)

After affinity purification, some samples required further purification by ion exchange chromatography. Before ion exchange chromatography the ionic strength of the buffer was reduced by dilution in low salt buffer to allow protein binding to the column. Protein was then loaded to either Q or S columns depending on whether the theoretical pI of the protein was below or above the buffer pH respectively. Depending on the resolution required, Q, Source Q or Resource Q resin was used for purification.

Protein was loaded using a 150 mL superloop attached to an Äkta Purifier (GE Healthcare). The column was washed with >10 column volumes of loading buffer. Protein was eluted by a linear salt gradient from this initial concentration to 1 M NaCl over 10-20 column volumes. Fractions were collected and UV elution peaks were analysed by SDS PAGE. Fractions containing the protein of interest were pooled and concentrated by centrifugal concentration (see section 2.2.11).

2.2.8 Size exclusion chromatography (SEC)

Size exclusion chromatography was used as the final step of purification for all proteins to remove protein aggregates and to exchange the protein into a defined buffer. Proteins with molecular weights less than ~70 kDa were applied to a Superdex 75 column, proteins smaller than ~200 kDa were purified using a Superdex 200 column and larger complexes of proteins were purified using a Superose 6 column (all GE Healthcare). The appropriate column was attached to

an Äkta Purifier (GE Healthcare) and typically equilibrated with 50 mM Tris pH 7 or 8, 300-500 mM NaCl, 0.5 mM TCEP. Samples were loaded on to the column at a flow rate of 0.3 ml/min. When using 26/60 columns 3 mL fractions were collected and for 16/60 or 16/70 columns 1 mL fractions were collected. Peak fractions containing purified protein were observed by UV absorbance and analysed by SDS PAGE. Fractions containing the protein of interest were then pooled and concentrated to 0.5-20 mg/mL depending on sample solubility. The protein concentration was determined using a NanoDrop 2000 spectrometer (Thermo Scientific). Samples were flash frozen in liquid nitrogen and stored at -80 °C. Sample composition was confirmed by mass spectrometry.

2.2.9 SDS-PAGE

SDS-PAGE analysis of proteins was performed in 1 x MES buffer (Invitrogen) using NuPAGE 4-12 % 1.0 mm Bis-Tris gels (Thermo Fisher Scientific). Gels were run for 40 minutes at 200 volts in XCell SureLock mini-cell gel tanks (Invitrogen). 15 µL of sample was mixed with 5 µL of 5 x NuPAGE LDS sample buffer and heated to 99 °C for 5 minutes prior to loading. 15 µL of this sample was loaded into each lane. SeeBlue PlusII pre-stained protein standards (Invitrogen) were used as a molecular weight marker throughout. InstantBlue coomassie stain (Expedeon) was used to stain the gels for protein visualisation.

2.2.10 Western blot

SDS-PAGE was carried out as described in section 2.2.9. Proteins were transferred onto a nitrocellulose membrane using the iBlot system (Thermo Fisher Scientific). The membrane was blocked with TBST (20 mM Tris pH 7, 150 mM NaCl, 0.1 % Tween-20) + 5 % milk powder (w/v) for 10 minutes. A sample of positive control phosphorylated peptide was dotted onto the upper side of the membrane placed on moistened blotting paper and further incubated in blocking buffer for one hour at room temperature. A Phospho-(Ser/Thr) ATM/ATR substrate primary antibody (Cell Signalling Technology, #2851) was used at a dilution of 1:2000 to probe Ccq1 T93 phosphorylation. The membrane was then washed three times in TBST for 5 minutes each. Blots were developed using an HRP conjugated anti-rabbit secondary antibody (Cell Signalling Technology), following the manufacturer's recommended protocols. The membrane was washed three times with TBST and once with TBS. Bound secondary antibodies were visualised with an enhanced chemiluminescence (ECL) kit (Thermo Scientific) using photographic film (Kodak).

2.2.11 Protein concentration

Proteins were concentrated by centrifugation at 6000 x g using VivaSpin centrifugal concentrators (Sartorius Stedim). The molecular weight cut off (MWCO) used for each sample was not larger than half the molecular weight of the protein. Protein concentrations were determined using a NanoDrop 2000 spectrometer (Thermo Scientific) and a continuous scan from 220-350 nm was

recorded. The absorbance at 280 nm was used to derive protein concentration. Protein extinction coefficients were determined using the ProtParam server, using extinction coefficients of individual tryptophan and tyrosine residues as 5500 $M^{-1}cm^{-1}$ and 1490 $M^{-1}cm^{-1}$ respectively.

2.2.12 Protein dialysis

Proteins were dialysed using either Slide-A-Lyzer dialysis cassettes (Thermo Scientific) of various MWCO or 1 kDa MWCO Mini Dialysis caps (GE Healthcare) for small volumes (<100 µL). MWCO was selected to ensure that it was no more than half of the calculated molecular weight of the protein being dialysed. Dialysis membranes were soaked in dialysis buffer for 5 minutes prior to addition of sample. 2 L of dialysis buffer, which was constantly stirred at 4 °C overnight was used for both methods.

2.3 Biophysical and biochemical techniques

2.3.1 Limited proteolysis

Sequencing grade porcine trypsin (Promega) was used for all limited proteolysis experiments. Purified protein samples were incubated with the protease at a ratio of 500: 1 w/w and 10 µL was removed from the reaction mixture after 0, 1, 2, 5, 10, 20, 40 and 60 minutes. 5 µL of 5 x NuPAGE LDS (Invitrogen) was added to these samples immediately and placed at 100 °C for 5 minutes to stop the reaction. Analysis of this time course was performed by SDS-PAGE. For analysis by mass spectrometry, additional samples were taken at 10 and 60 minutes and reactions terminated by addition of 5 µL of 1 % trifluoroacetic acid (TFA).

2.3.2 Nano differential scanning fluorimetry (nanoDSF)

Buffer optimisation was performed using nanoDSF using a NanoTemper Prometheus NT.48 (NanoTemper). This instrument allows for label free measurement of the thermal denaturation of a protein in 48 conditions simultaneously. The ratio of fluorescence intensities at 330 and 350 nm is measured as sample temperature is increased over a linear gradient. This ratio is dependent on the immediate chemical environment around aromatic residues within proteins and so changes that occur due to protein unfolding can be detected as a change in fluorescence ratio. This allows the melting temperature

(T_m) under different conditions to be determined allowing for the condition in which the protein is most stable to be identified [187].

A commercially available buffer screen (RUBIC, Molecular Dimensions) was used for these experiments. 18 μL of each buffer condition was mixed with 2 μL of protein sample at 10 mg/mL. Samples were drawn up into capillaries by capillary action for analysis and data were visualised using the built-in PR.ThermControl software.

2.3.3 Peptide array analysis

Peptide array analysis was used to detect binding sequences between two proteins. 20 residue peptides were synthesized in house by the Peptide Synthesis facility (Francis Crick Institute) and immobilised on a cellulose membrane. Peptides were shifted by one residue along the sequence of the protein of interest.

Peptide bound membranes were activated by washing in 50 % ethanol + 10 % glacial acetic acid for 1 hour followed by washing with TBST + 0.5 % milk powder (w/v). Positive controls of purified proteins containing a His tag were added to the membrane by dot blotting two different His tag containing proteins onto the membrane. 50 μL of target protein at an initial concentration of 12 mg/mL was added to the membrane along with 5 mL of TBST + 0.5 % milk powder (w/v) and incubated with rolling at 4 °C overnight. A no-protein control was used for each

experiment. Protein binding was visualised using an anti-His antibody (Abgent) as described in section 2.2.10.

2.3.4 Size exclusion chromatography Multi Angle Laser Light Scattering (SEC-MALLS)

During SEC-MALLS proteins are separated by size using size exclusion chromatography. The eluted protein is then analysed by three different in-line detectors. The first of these shows the UV absorbance of the sample at 280 nm, which shows the elution profile of the protein. The second is a differential refractometer, which measures the difference in refractive index between the buffer and the protein sample as it elutes from the size exclusion column. This difference in refractive index is proportional to the concentration of the protein in solution, as twice as many proteins in solution will cause twice as much light scattering to occur. The final detector measures the intensity of laser light scattering at 16 angles. The light scattered at each angle is independent of the overall particle size and shape, as long as particles are much smaller than the wavelength of the incident radiation. Measuring the scattered light at multiple angles improves the signal obtained. The intensity of scattered light is related to the molecular weight of the proteins within the sample. Protein monomers in solution will scatter light out of phase with one another leading to negative interference, however if these proteins oligomerise then their scatter will become coherent and so the intensity of scatter will increase proportionately [188]. Measuring the total scattering of a protein solution and dividing through the total

protein concentration gives the average molecular weight of the protein particles at every point in the elution profile.

Purified protein samples were loaded to an equilibrated Superdex 200 10/300 Increase or Superose 6 10/300 Increase column (both GE Healthcare) at a flow rate of 0.5 mL/min on a Jasco high-performance liquid chromatography (HPLC) system. UV absorbance of the eluted proteins was measured in-line and the scattered light intensity was detected at 16 angles by a DAWN-HELEOS II multiangle laser light-scattering detector (Wyatt Technology). The refractive index was measured using an OPTILAB-rEX differential refractometer (Wyatt Technology) and used to determine the protein concentration. The apparent molecular weights of proteins in the elution profile were calculated from the scatter and concentration values using the ASTRA v6.1 software package (Wyatt Technology).

2.3.5 Circular dichroism

Plane polarised light is made up of two circularly polarised components with equal magnitudes, one is rotating clockwise (right handed) and the other anti-clockwise (left handed). Circular dichroism (CD) utilises the different absorption of this polarised light by biological macromolecules. This is due to the chirality of amino acids and can give information on the secondary structure of proteins. Experimental CD spectra are compared to reference databases containing CD

spectra for previously crystallised proteins, allowing for similarities to be drawn and prediction of the protein secondary structure.

Purified protein was diluted to 0.15 mg/mL in 50 mM Tris pH 7, 50 mM NaCl, 0.5 mM TCEP and samples were loaded into quartz cuvettes with a 1 mm path length (Hellma). Spectra were recorded at 20 °C with a Jasco-J815 spectropolarimeter, scanning from 200 to 260 nm at a rate of 100 nm/min with a 2 nm bandwidth. 10 discrete scans were averaged to give the final spectra. CD intensities are presented as the CD absorption coefficient calculated on a mean residue weight basis ($\Delta\varepsilon_{MRW}$). Secondary structure content was estimated using methods originally described by Sreerama and Woody [189]. Thermal denaturation of samples was measured over a range of 10-95 °C at a rate of 2 °C /min and a wavelength of 222 nm. This allows the temperature at which protein denaturation occurs to be established, as the CD signal changes due to a loss of protein secondary structure upon melting. Once full denaturation of the protein has occurred, the melting temperature (T_m) of the protein can be calculated from the midpoint of the CD signal change.

2.3.6 Isothermal titration calorimetry (ITC)

Quantitative studies of protein interactions were performed by ITC. This method directly measures the heat that is released or absorbed when two molecules interact. ITC is also a label free technique that does not require immobilisation of proteins and therefore measurement of binding parameters occurs with proteins

in their native states. ITC can be used to calculate the association and dissociation equilibrium constants (K_a and K_d) for an interaction as well as the change in enthalpy (ΔH), stoichiometry (N) and entropy (ΔS). Injection of ligand into the sample cell leads to a temperature peak, which can be plotted against the molar ratio of ligand: protein. This produces a binding isotherm from which the enthalpy change, binding equilibrium constant and the stoichiometry of the interaction can be derived [190]. For a simple bimolecular interaction, the stoichiometry value (N) is calculated by the molar ratio at the inflection point of the sigmoidal curve, the K_d is calculated from the gradient of the curve at the inflection point and ΔH is given by the difference between the two plateaus of the sigmoidal curve. The Gibbs free energy changes (ΔG) and entropy changes (ΔS) can then be calculated using the following equations:

$$\Delta G = \Delta H - T\Delta S$$

$$\Delta G = -RT \ln K_a$$

Where K_a is the equilibrium association constant, R is the gas constant and T is the temperature.

ITC measurements were performed using an iTC200 (MicroCal). This microcalorimeter has two cells, the sample cell which contains the sample and a reference cell which contains water. Before the experiment begins both cells are held at exactly the same temperature. During the experiment, a ligand is added

to the sample cell through sequential injections and binding events lead to changes in the temperature. These temperature changes are measured and the heat change is reversed within the calorimeter. Interactions can be endothermic or exothermic, leading to a heat increase or decrease being measured upon ligand addition.

A typical ITC experiment involves filling the sample cell with protein solution. A syringe is filled with a ligand and lowered into the sample cell. The paddle shaped syringe is spun continuously to mix the sample. Precise injections of the ligand into the sample cell are made and any heat changes are measured. The heat change is measured until the sample and reference cells return to equilibrium and another injection occurs. As the titration progresses and more ligand is injected into the sample cell, fewer sites are available for ligand binding to the sample protein. This results in a reduction in the heat change observed and once no sites are available then no further heat change is measured with subsequent injections.

Samples were dialysed into identical buffers prior to performing ITC, typically containing 50 mM Tris pH7 or 8, 300 mM NaCl, 0.5 mM TCEP. The sample concentration of the protein in the sample cell was 50 μM and the titrant in the syringe was at 500 μM. Titrations consisted of 20 x 2 μL injections into a sample cell volume of 200 μL. Measurements were performed at 12, 20 and 28 °C. A buffer blank control was performed to check for the effect of the heat of dilution on the values measured. Analysis of ITC experiments was performed using

ORIGIN software following the manufacturer's instructions and a single site binding model was assumed.

2.3.7 Biolayer interferometry (BLI)

Protein interactions were also studied by biolayer interferometry (BLI) using the Octet system (ForteBio). BLI can be used to measure the kinetics of protein-protein interactions, as well as binding affinity. It is a label free technique that requires immobilisation of protein on a biosensor tip. This technique measures the association and dissociation rate constants (k_{on} and k_{off}), which allows for calculation of the equilibrium dissociation constant (K_d) for the interaction, $K_d = k_{off}/k_{on}$. K_d can also be derived from the midpoint of the BLI response.

BLI measures the interference pattern of white light reflected by two surfaces. Light is reflected by an internal reference layer within the biosensor tip and also by the interface between the solution surrounding the biosensor and a layer of immobilised protein. Reflection of light by these two surfaces leads to constructive and destructive interference which is observed as minima and maxima across different wavelengths of light [191]. Any change in the amount of protein bound to the biosensor leads to a change in the interference pattern observed in real time (see Figure 19).

BLI was performed using a ForteBio Octet RED instrument (ForteBio) using streptavidin, GST or Ni-NTA biosensors (ForteBio) and black 96-well plates

(Greiner bio-one). Ligand samples were produced by serial dilution of an initial stock to produce a range of concentrations from 0.55-70 μM . A baseline was measured for 200 seconds in sample buffer containing 0.005 % Tween-20. Peptides or protein were loaded on to the biosensors at a concentration of 2 μg/mL. After a washing step to remove any unbound protein, the biosensors were transferred to wells containing a range of concentrations of the potential ligand and binding measured. A shift in the wavelength of the reflected light is observed upon ligand binding, which directly relates to a change in the thickness of the layer of proteins bound to the biosensor tip. Dissociation was measured by returning the biosensors to the wash buffer. Measurements were performed at 25 °C with shaking to ensure mixing and to remove effects of diffusion. The equilibrium dissociation constants (K_d) were calculated by fitting the data assuming a 1:1 interaction using software developed in house (Steve Martin, Francis Crick Institute).

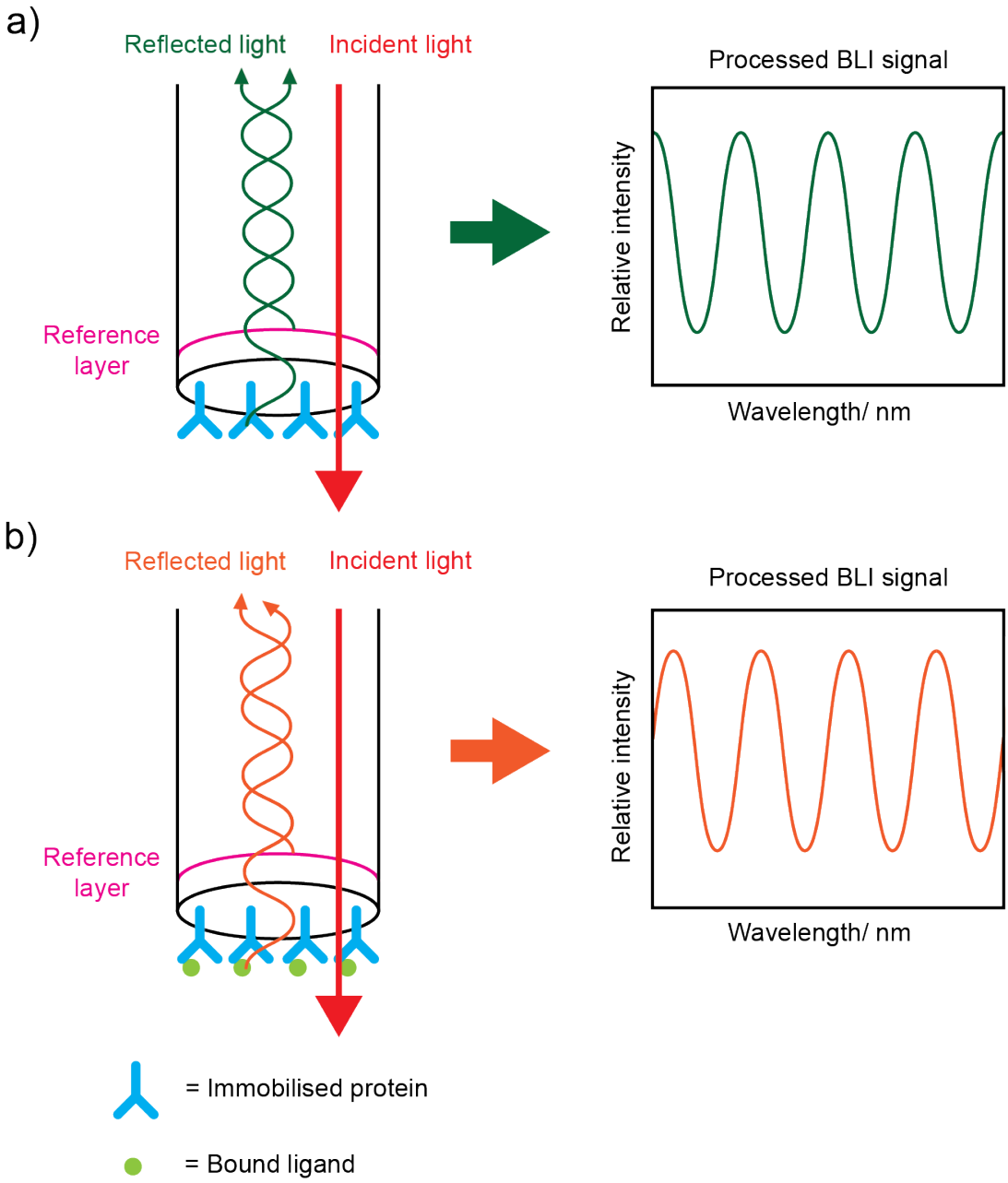


Figure 19: Biolayer interferometry.
a) Incident white light is directed down a biosensor tip that has a target protein immobilised on the end. Light is reflected by two surfaces, the internal reference layer and the interface between the solution and the layer of immobilised protein. This leads to interference between the light reflected by each layer, which can be measured in real time. b) When the biosensor is submerged in a solution containing a ligand for the immobilised protein, binding occurs between the protein and ligand, leading to an increase in the thickness of the protein layer bound to the biosensor tip. This causes a change in the distance between the layers which reflect light and so a different interference pattern is observed, leading to a wavelength shift. This allows binding events to be quantified, as the wavelength shift is a direct measurement of the amount of protein bound to the biosensor.

2.4 Structural characterisation

2.4.1 Protein crystallisation

High throughput crystallisation screening was carried out using commercially available and in-house sparse matrix and grid screens (Hampton, Qiagen, Molecular Dimensions). A 96 well multi-channel pipette was used to transfer 75 μL of each crystallisation condition into the reservoirs of a 96 well MRC 2 drop sitting drop crystallisation plate (Hampton). An OryxNano (Douglas Instruments) or Mosquito (TTP Labtech) liquid handling robot was used to dispense 200 nL drops typically containing 100 nL of protein sample and 100 nL of reservoir solution at room temperature. Plates were sealed using Crystal Clear Sealing Film (Hampton) and transferred to a RockImager plate imager (Formulatrix), which maintained plates at 18 °C and took images at set intervals. When potential crystals were observed, images were recorded under ultraviolet light to check for tryptophan fluorescence, an indicator of protein content.

Optimisation grid screens were performed for crystal hits identified from initial screening. Custom screens were produced using a Formulator (Formulatrix) and set up using the OryxNano or Mosquito, as previously. Larger scale optimisations, using a hanging drop configuration, were set up by hand using 24 well EasyXtal crystallisation plates (Qiagen). 500 μL of reservoir solution was

used and drops contained 1 μL of protein and 1 μL of well solution. These trays were incubated at 18 °C and were checked manually using a light microscope.

An attempt to improve crystal quality and numbers was performed using micro-seeding. Initial crystals were harvested in 20-50 μL of reservoir solution and transferred to a PTFE Seed Bead tube (Hampton) and crystals were crushed by vortexing. The resulting seed stock was added to crystallisation drops at several different ratios. A protein seed solution contains microcrystals, which can provide sites for initial nucleation of crystals to occur [192]. Crystallisation conditions and specific optimisation experiments are described in respective results chapters.

2.4.2 X-ray diffraction experiments

Cryoprotectants were tested prior to crystal freezing to check that ice crystals did not form. Cryoprotectants contain the original crystallisation condition supplemented with either glycerol, ethylene glycol or a low molecular weight PEG, for example PEG 400. The lowest concentration of cryoprotectant additive, which prevented formation of ice crystals, was sought so the addition of the additive was least likely to have an effect on the crystals. Crystals were transferred to a suitable cryoprotectant for one minute prior to being flash frozen in liquid nitrogen. Crystals were maintained in cryogenic conditions from this point onwards, including during data acquisition, to preserve the crystals and to reduce radiation damage. Data were collected at the Swiss Light Source PX1 Beamline

using an EIGER 16M X detector. Due to limits in the resolution and poor quality diffraction, indexing and further processing was not possible.

2.4.3 Small angle X-ray scattering (SAXS)

Small angle X-ray scattering (SAXS) is a solution-based technique for low-resolution structural characterisation of proteins. It also gives information on the oligomeric state of proteins [193]. It is based on the elastic scattering of photons by macromolecules and when plane waves pass through a sample, electrons in the atoms of the sample produce secondary waves, meaning constructive and destructive interference can then occur. As the sample is in solution, the molecules can move freely and are randomly orientated, so orientation information that is known for proteins in crystals is lost [194]. The scattered intensity of photons is collected on a 2D detector giving the radially averaged scattered intensity as a function of the momentum transfer (q).

$$q = \frac{4\pi \sin\theta}{\lambda}$$

Where λ is the wavelength of the incident beam and θ is half the angle between the scattered radiation and the incident beam.

Subtraction of the scattering caused by the buffer from scatter by the protein containing solution can then be undertaken prior to further analysis. As SAXS gives an averaged intensity, a monodisperse, fully purified sample is required

prior to experimentation to ensure that an accurate final model can be calculated [195].

SAXS for His-tagged and un-His-tagged Ccq1 130-717 was performed at the P12 beamline at the Petra III storage ring at DESY. Data was collected by Haydyn Mertens and Cy Jeffries using a Pilatus 6M detector at a distance of 3.0 m and a temperature of 20 °C. A sample concentration series from 18 mg/mL to 1 mg/mL was used for data collection in batch mode.

Primary reduction of the data was performed using ScÅtter (http://www.bioisis.net/) and data processing was carried out with ATSAS (http://www.embl-hamburg.de/biosaxs/software.html). This allowed for calculation of the radius of gyration (R_g), the maximum particle dimension (D_{max}), the excluded particle volume (Vp) and the pair distribution function (P(r)). A low resolution three-dimensional *ab initio* model was generated using DAMMIF [196] and the results of 50 independent runs were averaged using DAMAVER [197].

2.4.4 Glutaraldehyde fixation (GraFix)

GraFix was performed as described by Kastner *et al*. [198] Gradients of 5-20 %, 5-25 %, 10-25 % and 10-30 % w/v sucrose containing 50 mM HEPES pH 7.5 150 mM NaCl were prepared. Each gradient also contained a second gradient of 0-0.1 % glutaraldehyde. These different gradients were set up for different samples depending on the expected molecular weight of the complex being investigated.

A layer of buffer containing 2 % w/v sucrose was included on the top of the gradient to remove imidazole from the protein solution and exchange the sample buffer from Tris to HEPES as both Tris and imidazole can react with glutaraldehyde. Gradients were centrifuged at 40000 rpm for 18 hours or 36000 rpm for 16 hours, fractionated into 100 μL aliquots and analysed by SDS-PAGE, as described in section 2.2.9.

2.4.5 Electron microscopy

2.4.5.1 Background

In transmission electron microscopy, a high energy electron beam is used to form a direct image of a sample. This electron beam originates from a cathode, which releases electrons when heated to a sufficiently high temperature. The electron beam is focused by a condenser lens, before contacting the sample, which is fixed on a grid. Scattering of electrons occurs due to the sample and the image forming objective lens then focuses any unscattered beam back into the focal plane, as well as focusing scattered electrons out of this plane. Scattered electrons are then blocked by an aperture, preventing them from contributing to the final image. A projector lens is then required to expand the resulting beam for detection. Direct viewing of an image is possible using a fluorescent screen placed in the image plane, which converts the incident electrons into bursts of visible light (see Figure 20). The electron beam can be recorded using

photographic film, charge-coupled device (CCD) cameras or direct electron detectors [199, 200].

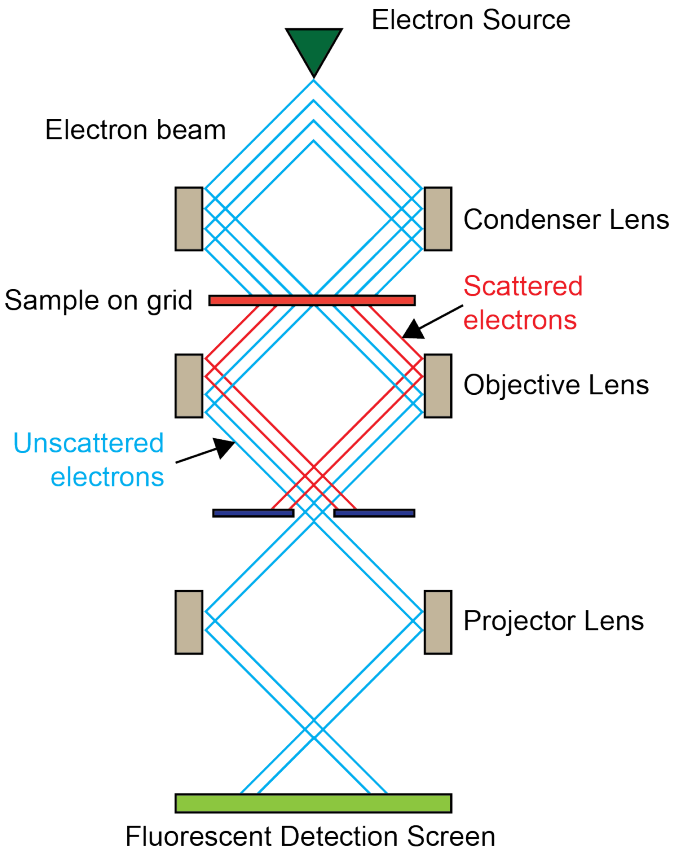


Figure 20: Image formation by a typical transmission electron microscope. The condenser lens focuses the electron beam prior to scattering by the sample in the microscope. Refocussing followed by expansion of the beam occurs before detection of an image occurs.

The recent development of direct electron detectors has led to a so called “resolution revolution” in cryo-electron microscopy due to the ability to detect incident electrons directly without the resolution limiting requirement of CCD cameras which require the incident electron beam to be turned into photons before detection can occur. Direct electron detection removes this step, preventing the spreading of the signal before detection, allowing a higher resolution to be reached [201].

As described above, scattering of electrons is caused by the sample. The amount of scattering that occurs is proportional to the electron density in the sample and in proteins this electron density is relatively low so only a small amount of scattering is observed. Consequently, micrograph contrast for protein samples is low and these samples are susceptible to radiation damage [202]. These problems are overcome by negative stain electron microscopy. As described in section 2.4.5.2 protein is incubated on a glow discharged grid before stain is added. There are several types of stain, although uranyl acetate is the most commonly used. The stain creates a shell around the protein sample which strongly scatters the electron beam due to its high electron density [203]. In micrographs the stain appears dark and the protein sample light, meaning that in negative stain electron microscopy the stain shell rather than the protein is visualised. This leads to a reduction in the possible resolution gained for the technique. Despite this, negative stain electron microscopy provides a reliable and quick technique for visualisation of proteins at low resolution [204].

2.4.5.2 Grid preparation

GraFix produced samples were diluted and concentrated several times to exchange the buffer and reduce the sucrose concentration. Continuous carbon coated 300 mesh copper grids (EMResolutions) were glow discharged using an Emitech K100X glow discharger (Quorum technologies) with a vacuum of vacuum 7 x 10^{-2} mBar and a current of 45 mA for 30 seconds. 4 μL of sample at several concentrations was incubated on the glow discharged surface of the grid

for 1 minute. Excess sample was removed from the grid by blotting and the grid washed twice with sample buffer. 1 % uranyl acetate was applied to the grid by touching the grid sample side down on a 20 μL drop of stain twice, the second with gentle agitation for 1 minute. Excess stain was removed by blotting and the grid allowed to dry.

2.4.5.3 Data acquisition

Negative stained grids were visualised using a FEI G2 Spirit TWIN transmission electron microscope (Electron microscopy Science Technology Platform, Francis Crick Institute) with a microscope voltage of 120 keV fitted with a UltraScan 894 2k CCD camera (Gatan). Datasets were collected at a magnification of 26000x, resulting in a calibrated pixel size of 4.0 Å/pixel.

2.4.6 Electron microscopy image analysis

Image analysis was performed using a Scipion pipeline. Calculation of the contrast transfer function was not performed as the resolution of negative stain data leads to production of data which is at a lower resolution than the first minimum node. Initial manual picking of particles was performed on micrographs showing good contrast between particles and the background using xmipp3 – manual-picking [205-207]. This manual picking was used to train the system for auto-picking by xmipp3 – auto-picking. A box size of 100 x 100 pixels was used. During this process particles were scored for their similarity to the reference

manually picked particles and also for their signal: noise ratio, giving an xmipp Z score [208]. The lower the score the closer the particle correlates with the manually picked particles. Particles were ranked by Z score and this allowed for removal of particles with a high score.

A 2D classification using relion – 2D classification was then performed producing 24 classes [209-211]. Default parameters were used during this classification step. Subsets of particles, averages and classes were saved for classes which showed the highest levels of contrast and the most features. A second 2D classification was run, which produced 100 classes. The resolution of this classification step was limited to 15 Å and a regularisation parameter of 1 was used to ensure that the classification was occurring strictly due to the data. Other parameters were left as the defaults. Again, subsets of the data were produced for classes which showed the highest contrast and most features. An initial 3D model was produced using eman2 – initial model [212] and 3D classification was performed using relion – 3D classification. The resolution of classification was limited to 15 Å resolution. The classes which used the largest subset of the data were then refined using relion – 3D auto-refine, before projection matching and comparisons of reprojections were made using xmipp3 – projection matching and xmipp3 – compare reprojections respectively.

Chapter 3. Protein expression and purification

3.1 Introduction

The fission yeast shelterin complex is comprised of six proteins, Taz1, Rap1, Tpz1, Poz1, Pot1 and Ccq1 [81]. These proteins have both positive and negative regulatory effects on telomerase activation. Taz1 forms a homodimer and interacts with double-stranded telomeric DNA, whereas Pot1 interacts with single-stranded telomeric sequence [87]. Tpz1 forms an important bridge between these double-stranded and single-stranded DNA binding proteins [58]. In addition, a C-terminal domain of Tpz1 interacts with Rad3/Tel1 phosphorylated Ccq1, leading to a switch from a non-extendible to an extendible state of the telomere [87]. Ccq1 is also involved in recruitment of Est1 to the telomerase enzymatic subunit, Trt1 [120]. Est1 mediated stabilisation of Trt1 at telomeres is required before the enzyme can become active [120]. A direct interaction between the TEL-like patch on Tpz1 and Trt1 also occurs, which further stabilises the telomerase complex at telomeres [112].

The models for shelterin function at telomeres have been studied by qualitative methods and therefore at the outset of this project there was very limited structural information available for these important interactions and proteins. Only structures of the DNA binding domain of Pot1 in complex with single-stranded

DNA had been reported [100, 101]. In preparation for interaction and structural studies purification of shelterin complex proteins was attempted both alone and in complex with partner proteins within the shelterin complex (Chapter 5) that had been identified by previous interaction studies [58, 105].

3.2 Expression protocol summary

Over the course of this project, expression of a large number of different proteins and constructs was tested. Details for the constructs tested are given in subsequent sections of this chapter. The DNA for these constructs were transformed into E. coli BL21 Gold (DE3) cells, which enables both DNA extraction and protein expression from the same colony. This ensured only DNA from constructs that expressed protein were sequenced. Initially, protein expression was tested in small-scale cultures (1 mL), the small-scale expression protocol is described in detail in section 2.2.1.

Protein constructs that gave the highest levels of soluble expression were then produced at a larger scale (multiple flasks containing 500 mL of *E. coli* culture). For these larger scale experiments, DNA was transformed into either BL21 (DE3) or BL21 Gold (DE3) cells. After growth of colonies on antibiotic supplemented agar plates overnight at 37 °C, single discrete colonies were picked and used to grow a starter culture consisting of antibiotic supplemented LB media. This was grown shaking overnight at 37 °C. A 1/100 dilution of this starter culture was added to 500 mL of LB and cells were grown to an optical density of 0.6-0.8. At this point, the temperature of the cultures was typically dropped to 18 °C and once the cultures had cooled, protein expression was induced by addition of IPTG to a final concentration of 1 mM. Cultures were then left to grow overnight shaking at 18 °C. For further detail on large-scale protein expression see section 2.2.2.

If these typical expression conditions did not give sufficient soluble protein expression, then the protocol was optimised. Plasmids were transformed into Rosetta2 (DE3), BL21 (DE3) pLysS, C41 or BL21 Star (DE3) cells to probe whether rare codon usage, potential expression of toxic recombinant proteins or low mRNA stability affected soluble protein expression. The temperature and time for the induction step was also optimised, with expression at either 30 °C or 37 °C for between 3-6 hours, or 16 °C or 25 °C overnight. The concentration of IPTG used to induce protein expression was also optimised, however a final concentration of 1 mM was found to give the best results in most cases. LB, Terrific Broth or autoinduction media was also tested for some proteins, although changes to the media did not have a significant impact on protein expression for any protein expressed during this project. A summary of the methods used for protein expression optimisation is shown in Table 12.

Table 12: Summary of potential expression optimisation methods used during this project.

***E. coli* cells**	**Media**	**Induction**
BL21 (DE3) BL21 Gold (DE3) BL21 Star (DE3) BL21 (DE3) pLysS Rosetta2 (DE3) C41	LB Terrific Broth Autoinduction	[IPTG] between 0.1-1 mM 25 °C O/N 18 °C O/N 16 °C O/N 37 °C 3-6 hours 30 °C 3-6 hours

3.3 Expression of Est1

3.3.1 *E. coli* and *S. pombe* expression

An initial goal was to investigate the phosphorylation dependent interaction between Est1 and Ccq1, and the Est1-TER1 interaction. Previous studies have shown that the interactions between Est1 and Ccq1 or TER1 occur in the N-terminal 14-3-3-like domain of Est1 and are partially overlapping, suggesting that these interactions may be competitive [105] (see Figure 21).

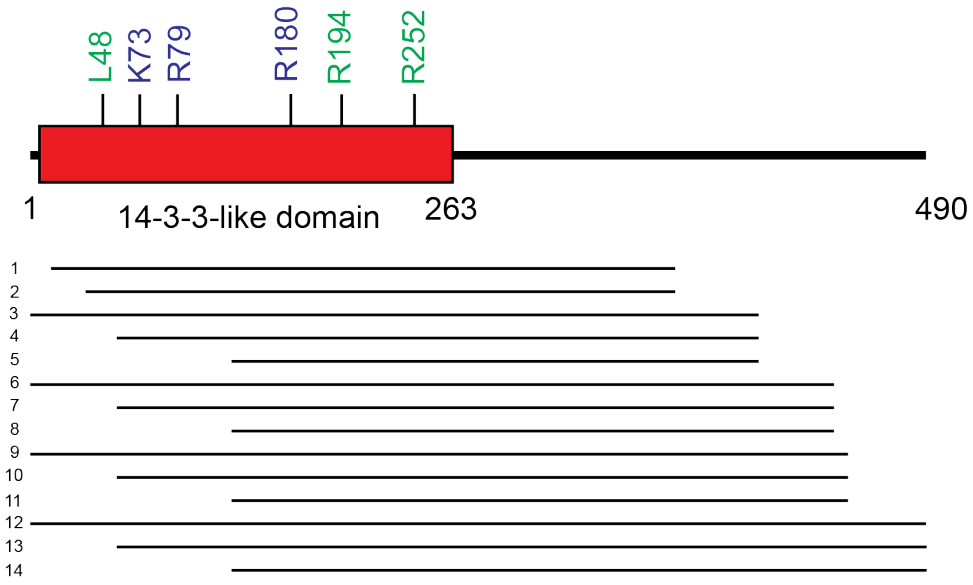


Figure 21: Domains of Est1 and construct design.
Est1 contains an N-terminal 14-3-3 domain that contains important residues for binding of Ccq1 (blue) and TER1 (green).The overlap in binding sites leads to a competitive interaction occurring. Constructs designed for protein expression are shown by black lines and relate to numbers in Table 13.

Initial *E. coli* test expression studies of full-length Est1 used a variety of vector systems (pGEX6p1, pMal, pET GST H6 and pET H6) provided by Kazunori

Tomita (UCL, UK) were unsuccessful. Further constructs were designed using the CCD structural prediction server (https://ccd.rhpc.nki.nl), which employs several secondary structure predictions to ensure constructs did not begin or end mid-way through a region of predicted secondary structure (see section 2.1.3). A large number of constructs were then screened (see Table 13) using a small-scale expression and solubility testing pipeline, developed in-house (see section 2.2.1). Fourteen constructs of varying length were chosen and PCR amplified from an *E. coli* codon optimised *Est1* gene. Inserts were cloned into a variety of expression vectors; pET47b, which provides a C-terminal 6 x His tag, pET49b (GST tag), pET22b (C-terminal 6 x His tag) and pET52SUMO (N-terminal 6 x His and SUMO tags). Constructs expressed in pET22b were designed to both include and exclude the *pelB* leader sequence (see Table 13). The *pelB* leader sequence leads to secretion of the expressed protein into the periplasmic space, which has been shown to increase the half-life of some proteins [213, 214]. The resulting plasmids were then used for small scale expression tests. Protein expression levels were monitored by SDS-PAGE of whole cell samples. Whilst no overexpression was observed in whole cell lysates, small-scale Ni-NTA affinity purification was attempted for 6 x His tagged Est1, but no soluble Est1 was recovered.

Full-length Est1 with a V5 tag was produced in *S. pombe* in collaboration with the Tomita lab at UCL. This yeast-expressed protein was purified using magnetic epoxy beads coated with mouse anti-V5 tag antibodies. Levels of expression of

Est1 were too low to be observed on a Coomassie stained gel, but expression was confirmed by Western blot.

Table 13: Est1 constructs designed using secondary structure predictions.
Ticks show if each step was successful. Where KT is written in the cloning column, constructs were provided by Kazunori Tomita (UCL, UK). # = construct number, C = cloned, E = expressed, S = soluble, P = purified. + symbols in soluble/purified column relate to the relative amounts of each construct observed or recovered with + being the smallest amount and +++ being the largest. These relative amounts only relate to other Est1 constructs.

#	Construct	Tag	C	E	S	P
12	pGEX6P Est1 1-490	N-GST	KT	-	-	-
12	pMal Est1 1-490	N-MBP	KT	-	-	-
12	pET GST H6 Est1 1-490	N-GST +His	KT	-	-	-
12	pET H6 Est1 1-490	N-His	KT	-	-	-
3	pET47b Est1 1-429	N-His	✓	-	-	-
4	pET47b Est1 56-429	N-His	✓	-	-	-
5	pET47b Est1 156-429	N-His	✓	-	-	-
6	pET47b Est1 1-451	N-His	✓	-	-	-
7	pET47b Est1 56-451	N-His	✓	-	-	-
8	pET47b Est1 156-451	N-His	✓	-	-	-
9	pET47b Est1 1-458	N-His	✓	-	-	-
10	pET47b Est1 56-458	N-His	✓	-	-	-
11	pET47b Est1 156-458	N-His	✓	-	-	-
12	pET47b Est1 1-490	N-His	✓	-	-	-
13	pET47b Est1 56-490	N-His	✓	-	-	-
14	pET47b Est1 156-490	N-His	✓	-	-	-
1	pET47b Est1 8-388	N-His	✓	-	-	-
2	pET47b Est1 32-388	N-His	✓	-	-	-
12	pET49b Est1 1-490	N-GST +His	✓	-	-	-
1	pET49b Est1 8-388	N-GST +His	✓	-	-	-
2	pET49b Est1 32-388	N-GST +His	✓	-	-	-
12	pET22b Est1 1-490 (+/- PelB)	C-His	✓	-	-	-
1	pET22b Est1 8-388 (+/- PelB)	C-His	✓	-	-	-
2	pET22b Est1 32-388 (+/- PelB)	C-His	✓	-	-	-
12	pET52SUMO Est1 1-490	N-His +SUMO	✓	-	-	-
1	pET52SUMO Est1 8-388	N-His +SUMO	✓	-	-	-
2	pET52SUMO Est1 32-388	N-His +SUMO	✓	-	-	-
12	pIEX-Bac3 Est1 1-490	N-His	✓	✓	+	+
1	pIEX-Bac3 Est1 8-388	N-His	✓	-	-	-
2	pIEX-Bac3 Est1 32-388	N-His	✓	-	-	-
12	pTriEx6 Est1 1-490	C-His	✓	✓	+	+
1	pTriEx6 Est1 8-388	C-His	✓	-	-	-
2	pTriEx6 Est1 32-388	C-His	✓	-	-	-
12	pET47b *S. cryophilus* Est1 1-490	N-His	✓	✓	++	++
12	pET47b *S. octosporus* Est1 1-490	N-His	✓	✓	++	-
12	pET47b *S. japonicus* Est1 1-490	N-His	✓	✓	++	-
12	V5- *S. pombe* Est1 1-490	N-V5 tag (PK)	✓	✓	+	+

3.3.2 Insect cell expression

As an alternative to the unsuccessful bacterial and yeast systems, insect cell expression was tested using the baculovirus system. Insect cell expression, unlike bacterial expression, can lead to addition of post-translational modifications and allows access to eukaryotic protein folding pathways, which can be beneficial for protein solubility. Full-length Est1, along with Est1 8-388 and 32-388, were cloned into pIEx-Bac3 and pTriEx6, which incorporate an N- or C-terminal 10 x His tag respectively. The virus was amplified for small-scale expression in 50 mL cell cultures. Expression was observed for both N- and C-terminally His tagged full-length Est1 (see Figure 23a) and confirmed by single Band ID mass spectrometry (see Figure 22).

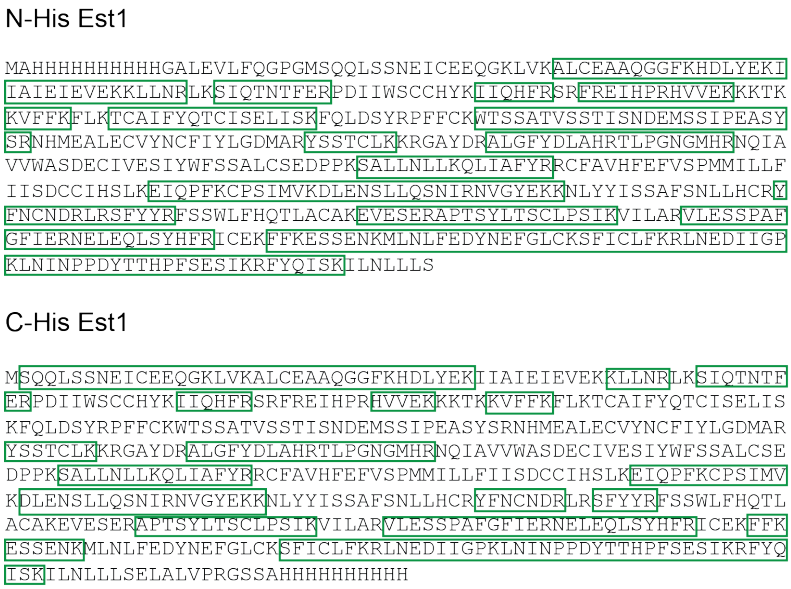


Figure 22: Band ID mass spectrometry for insect cell expressed *S. pombe* Est1. Green boxes show peptides identified by mass spectrometry. Sequence coverage of N-His Est1 and C-His Est1 was 59 % and 47 % respectively. Since, in both cases peptides close to both the N- and C-termini were present, it is likely that the semi-purified proteins were full-length. 15 and 18 unique peptides were identified for the N- and C-terminally tagged proteins respectively at a false discovery rate of 1 %, giving a negligible cumulative false discovery rate and a high likelihood the identifications are correct.

Further viral amplification produced sufficient virus for large-scale expression. Prior to expression, viral titres were normalised between different amplification reactions using qPCR and the titre used for large scale expression was optimised. For each large-scale expression experiment, between 2-5 L of Est1 expressing cells were grown, leading to production of a small amount of soluble protein. However, further increases in the volume of cells grown did not quantitatively increase Est1 yield (see Figure 23b).

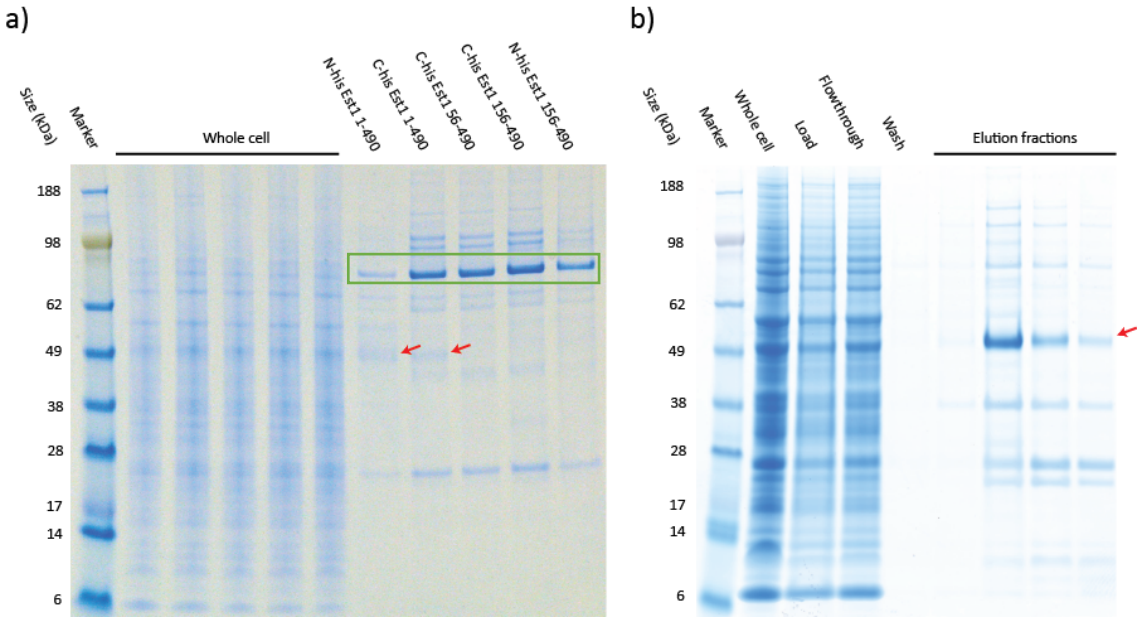


Figure 23: Optimisation of Est1 expression by the baculovirus expression system. Red arrows mark bands corresponding to Est1. Green box shows a major insect cell contaminant, assumed to be HSP90. a) Small scale expression tests produced a soluble protein band corresponding to full-length Est1 when tagged with either an N- or C-terminal 10 x His tag. Validation that this band corresponds to Est1 was performed by mass spectrometry. b) Batch affinity purification of Est1 using TALON resin. Optimisation of expression conditions led to an increase in protein recovery. However, further increases in expression scale did not increase the amount of soluble Est1 produced.

3.3.3 Orthologue expression

Expression of orthologues of *S. pombe* Est1 from closely related yeast species was tested in *E. coli*. The orthologues tested were from *Schizosaccharomyces cryophilus*, *Schizosaccharomyces octosporus* and *Schizosaccharomyces japonicus*. These proteins share a 36 %, 34 % and 23 % identity with *S. pombe* Est1 respectively. Although there is significant sequence divergence, secondary structure predictions showed conserved elements (see Figure 24). Interestingly, *S. japonicus* Ccq1 is also the most divergent in length and sequence, and does not contain a conserved threonine residue equivalent to T93 in *S. pombe* Ccq1 (see Figure 41). This suggests that Est1 recruitment must occur through a different mechanism in *S. japonicus*.

S. cryophilus Est1 exhibited the highest levels of soluble Est1 over-expression of the three orthologues. Protein was purified by a three-step protocol, involving nickel ion affinity chromatography, followed by 1 mL Q HP anion exchange chromatography and size exclusion chromatography (see Figure 25). This protein could be purified to near homogeneity, but the final yield was too low for biochemical and structural studies as < 20 μg of protein was purified from 6 L of cells. Again, mass spectrometry was used to confirm the identity of the purified protein.

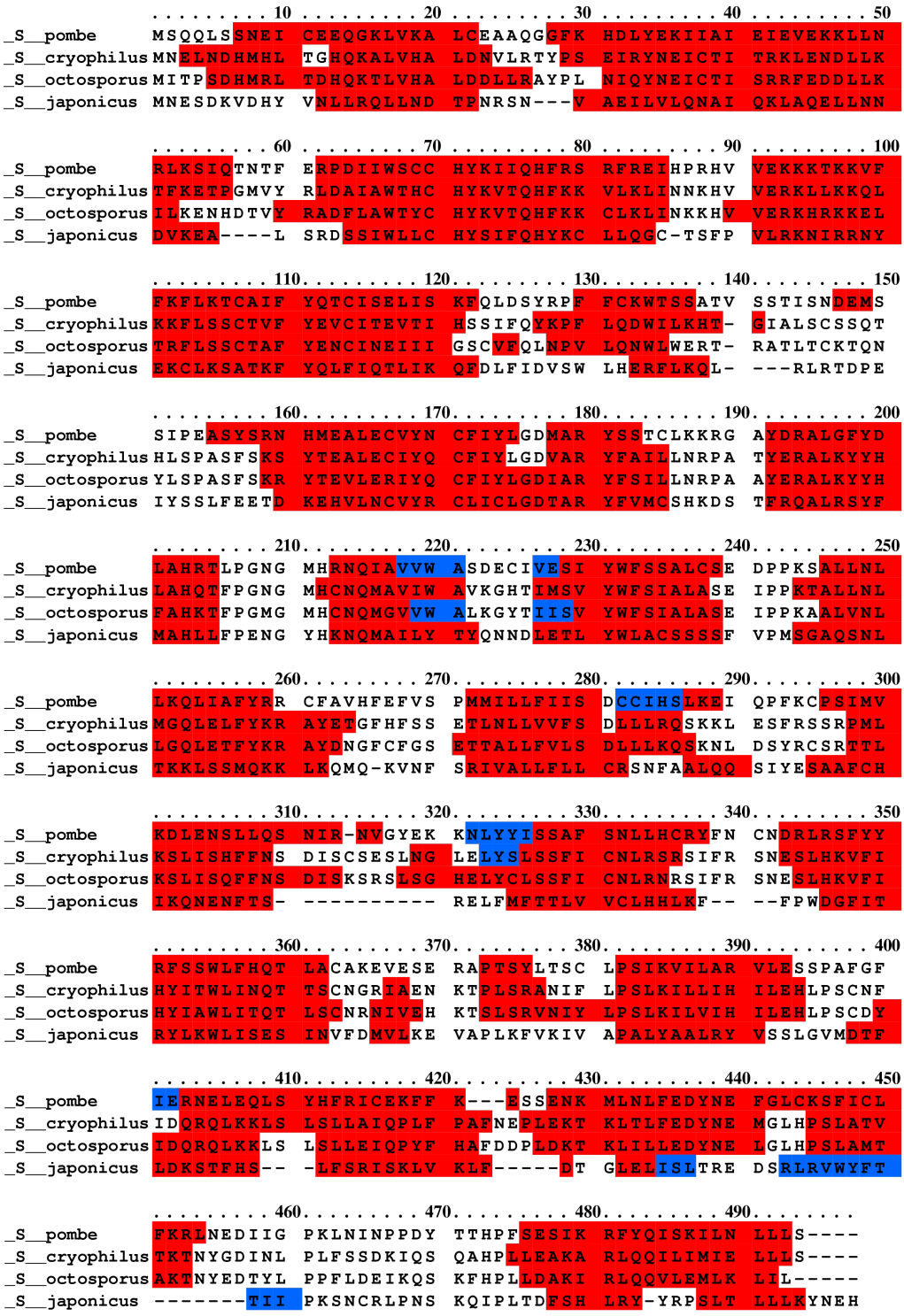


Figure 24: Secondary structure sequence alignment of homologous Est1 proteins from closely related yeast species using the PRALINE server [215].
Residues in red are predicted to be helical, while residues in blue are predicted to be in a β-sheet. Despite the sequence divergence between these four proteins, they are all predicted to be predominantly helical, as would be expected for a 14-3-3-like domain containing proteins.

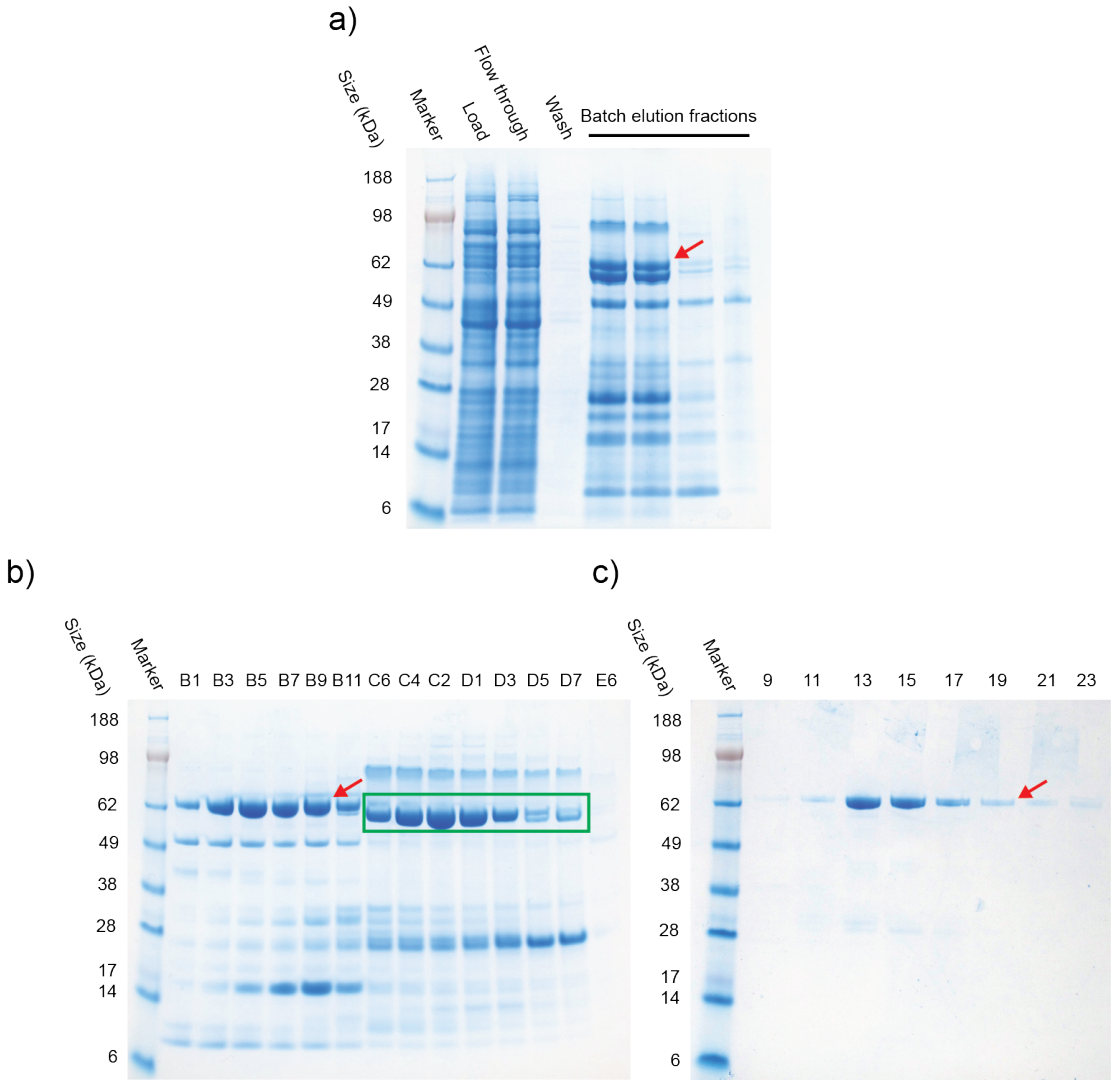


Figure 25: Example purification of *S. cryophilus* Est1.
S. cryophilus Est1 is marked by a red arrow just above the 62 kDa marker in all gels. A major contaminant is marked by a green box, this is presumed to be a truncated version of Est1. a) Cell lysate was bound to 3 mL of Nickel-NTA resin and then washed with buffer containing 25 mM imidazole. 5 mL of elution buffer containing 500 mM imidazole was added and incubated with the beads at 4 °C for 15 minutes to elute the protein. A second elution was performed by addition of a further 5 mL elution buffer. Each elution fraction was run separately on the gel. b) Est1 was bound to a 1 mL Q HP anion exchange column before elution using a linear gradient from 50 mM to 1 M NaCl in the buffer. An ÄKTA Purifier system was used for this step with fractionation into a 96-well block. Full length Est1 eluted from the column towards the beginning of the gradient at a lower NaCl concentration. c) Est1 containing fractions were pooled and concentrated before further purification using a 16/60 Superdex 75 size exclusion column run on an ÄKTAPrimeplus system, resulting in purified Est1 protein. The final yield was ~ 20 μg of protein from 6 L of cultured cells.

3.3.4 Expression summary and future uses for purified Est1

27 constructs of *S. pombe* Est1 were designed and successfully cloned for expression in *E. coli*, however none of these gave any soluble expression of the protein of interest as detected on a Coomassie stained SDS-PAGE gel. Six constructs were designed and subsequently cloned for expression in insect cells and two of these gave soluble expression, albeit in very low amounts. Expression of three homologues of Est1 from closely related yeast species was tested for full length constructs, with all homologues testing giving soluble expression. Sufficient Est1 for future experiments was purified from *E. coli* expressing *S. cryophilus* Est1.

Est1 expression remained challenging for the majority of this project and so no further experiments were possible with *E. coli* or insect cell expressed *S. pombe* Est1. Successful expression and purification of homologues of *S. pombe* Est1 was only possible at the very end of this project and so no further experiments were undertaken with this protein. However, detail for potential future experiments for Est1 homologues are included in Chapter 7.

3.4 Expression and purification of shelterin complex components

3.4.1 *E. coli* expression of Ccq1

Full-length *S. pombe* Ccq1 was initially expressed as a fusion with either a 3C cleavable N-terminal 6 x His tag or a 3C cleavable N-terminal GST tag followed by a 6 x His tag. Only 6 x His tagged constructs produced soluble protein, albeit in very small quantities.

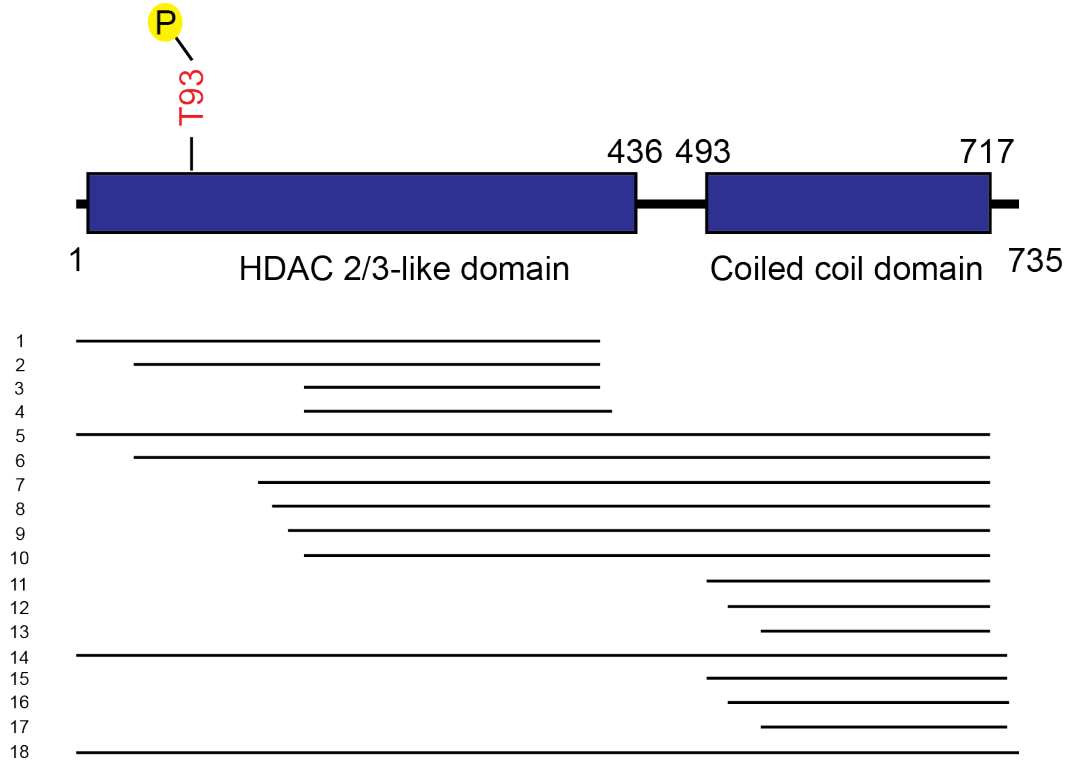


Figure 26: Ccq1 domains and construct design.
Ccq1 contains a predicted N-terminal HDAC2/3-like domain and a C-terminal coiled-coil. Binding of Ccq1 to Tpz1 and Est1 occurs within the HDAC2/3-like domain. Phosphorylation at T93 (red) is required for the interaction with Est1. Fragments used for protein expression screens are indicated by black lines.

Limited tryptic proteolysis of full-length expressed protein gave two main fragments which were identified by mass spectrometry, a minimal protease resistant domain (residues 493-717) and a larger fragment (residues 130-717), identified as a stable intermediate in the digestion reaction (see Figure 27a). The minimal protease resistant domain was predicted to consist of a C-terminal coiled-coil domain by secondary structure predictions, using PSIpred and Phyre2 [216, 217] and coiled-coil prediction programs such as LOGICOIL and COILS [218, 219]. These data were used to design further constructs encompassing other predicted domains or important regions of Ccq1, such as threonine residue 93, which is required for the phosphorylation-dependent interaction between Ccq1 and Est1 (see Figure 26). These constructs were cloned into pET47b and protein expression was tested (see Table 14).

Table 14: Ccq1 constructs designed using secondary structure predictions and tryptic digest data.
Ticks show if each step was successful. Constructs provided by Kazunori Tomita (UCL) are indicated by KT in the cloning column. # = construct number, C = cloned, E = expressed, S = soluble, P = purified. + symbols in soluble/purified column relate to the relative amounts of each construct observed or recovered with + being the smallest amount and +++ being the largest.

#	Construct	Tag	C	E	S	P
18	pET GST H6 Ccq1 1-735	N-GST +His	KT	✓	+	+
4	pET47b Ccq1 130-418	N-His	✓	-	-	-
1	pET47b Ccq1 1-416	N-His	✓	✓	++	++
2	pET47b Ccq1 40-416	N-His	✓	✓	++	++
3	pET47b Ccq1 130-416	N-His	✓	-	-	-
5	pET47b Ccq1 1-717	N-His	✓	-	-	-
6	pET47b Ccq1 40-717	N-His	✓	-	-	-
10	pET47b Ccq1 130-717	N-His	✓	✓	+++	+++
11	pET47b Ccq1 493-717	N-His	✓	✓	+++	+++
12	pET47b Ccq1 499-717	N-His	✓	-	-	-
13	pET47b Ccq1 511-717	N-His	✓	-	-	-
15	pET47b Ccq1 493-730	N-His	✓	✓	+++	+++
16	pET47b Ccq1 499-730	N-His	✓	-	-	-
17	pET47b Ccq1 511-730	N-His	✓	✓	++	++
18	pET47b Ccq1 1-735	N-His	✓	✓	++	++
18	pIEX-Bac3 Ccq1 1-735	N-His	✓	✓	++	++
18	pET49b Ccq1 1-735	N-GST +His	✓	✓	+	
18	pET47b Ccq1 1-735 T93E	N-His	✓	✓	++	++
7	pET52b Ccq1 120-717	N-Strep	✓	✓	++	++
8	pET52b Ccq1 123-717	N-Strep	✓	✓	+	-
9	pET52b Ccq1 126-717	N-Strep	✓	✓	+	
10	pET52b Ccq1 130-717	N-Strep	✓	✓	++	++
14	pET47b Ccq1 1-730	N-His	✓	-	-	-

Ccq1 residues 130-717, 493-717 and 493-730 gave the highest levels of soluble expression and so these three constructs were pursued further.

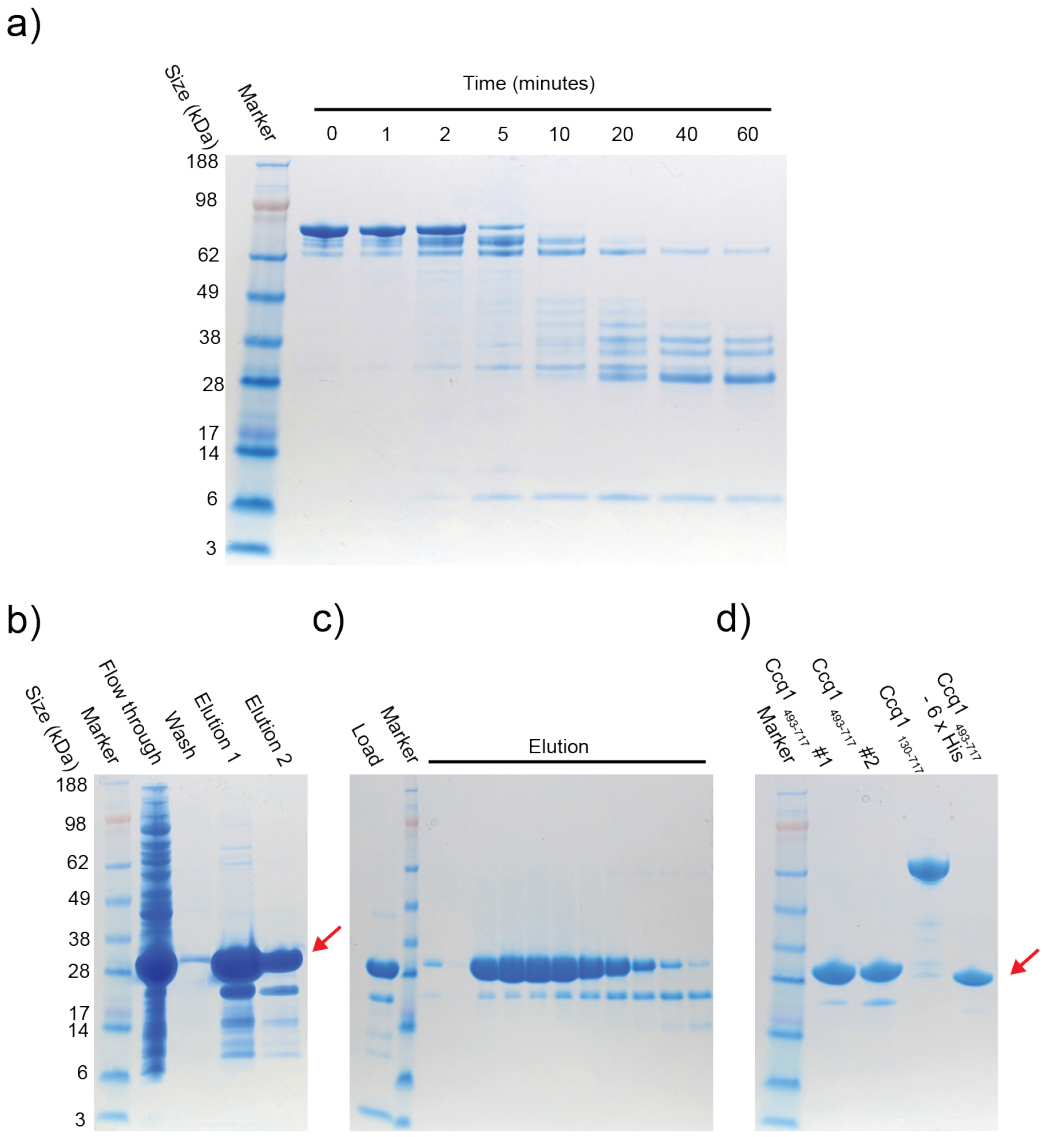


Figure 27: Example purification for Ccq1 493-717.
a) Limited tryptic proteolysis of purified full-length Ccq1. Proteolysis products were identified by mass spectrometry and this was used to inform construct design. b) Nickel immobilized metal ion affinity chromatography (IMAC) purification. The bound sample was washed with a low concentration imidazole containing buffer, before elution in an imidazole gradient from 25 mM to 500 mM. Ccq1 493-717 is marked by red arrows. c) Separation of the nickel eluate by 26/60 Superdex 200 size exclusion chromatography (SEC). The band at ~28 kDa corresponds to Ccq1 493-717. d) Concentrated protein samples used for initial crystallisation optimisation experiments.

Ccq1 residues 130-717, 493-717 and 493-730 were initially purified by a two-step process consisting of nickel immobilised metal ion affinity chromatography

(IMAC), followed by size exclusion chromatography (SEC) (see Figure 27). The purified protein contained few contaminants, as shown by SDS-PAGE analysis (see Figure 27d), which are potentially truncated versions of each construct due to degradation. Despite purification to near homogeneity, crystallisation screens failed to generate any crystals.

Addition of an anion exchange chromatography (IEX) step between nickel IMAC and SEC led to further improvements in the homogeneity of sample produced for Ccq1 130-717 (see Figure 28). For biochemical experiments that required the production of untagged protein, the His tag was removed by on column cleavage with 3C protease (see section 2.2.5). 3C protease was removed from the cleaved sample by addition of a small amount of glutathione Sepharose beads as it contained a GST tag.

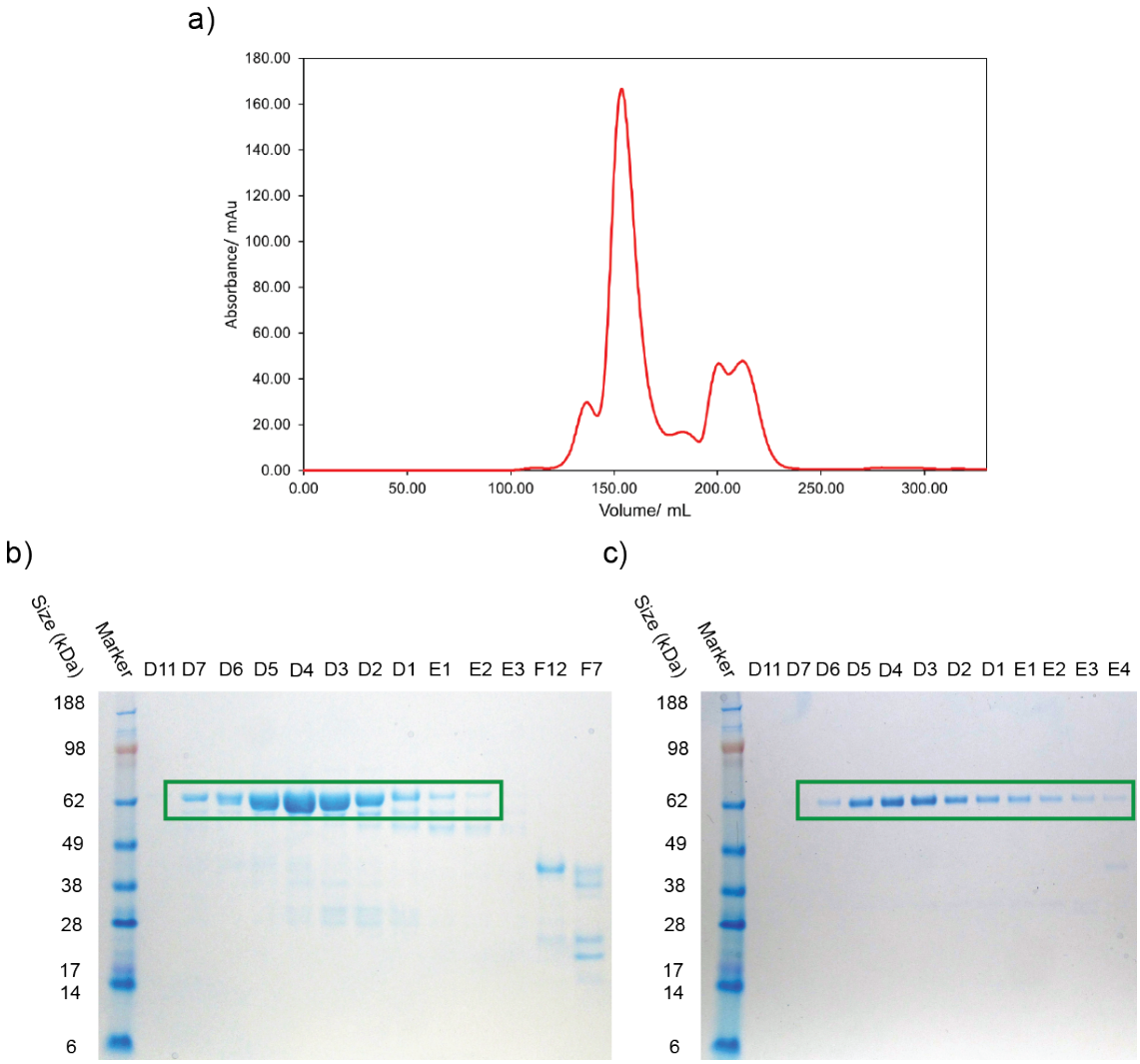


Figure 28: Comparison of Ccq1 130-717 purity after 26/60 Superdex 200 SEC. The green boxes show bands corresponding to Ccq1 130-717. a) Typical size exclusion chromatography trace for purification of Ccq1 130-717. The major peak runs close to the void volume for the column (˜110 mL), with the centre of this peak (˜155 mL) corresponding to fraction D4 on the gels below. Purification was repeatable, with the major species eluting at the same point in each repetition. b) SEC elution fractions when Ccq1 130-717 was purified by a two-step protocol (Ni IMAC followed by 26/60 Superdex 200 SEC). There are some contaminating protein bands that are smaller than the purified construct, which are likely to be due to sample degradation. c) Ccq1 130-717 SEC elution fractions after purification by a three-step protocol (Ni IMAC followed by anion exchange chromatography and 26/60 Superdex 200 SEC).

Ccq1 stability was tested in 96 different conditions by nano differential scanning fluorimetry (nanoDSF), using a NanoTemper Prometheus. This assay determines

the protein thermal denaturation curves across a range of buffer conditions. The highest T_m was observed for protein buffered with 50 mM HEPES pH 7.5, 1 M NaCl or 50 mM Tris pH 8.0, 1 M NaCl (see Figure 29 and Table 23). This is very similar to the buffer that had been used for initial purification experiments (50 mM Tris pH 8.0, 300 mM NaCl, 0.5 mM TCEP) and therefore no further optimisation of the buffer was possible.

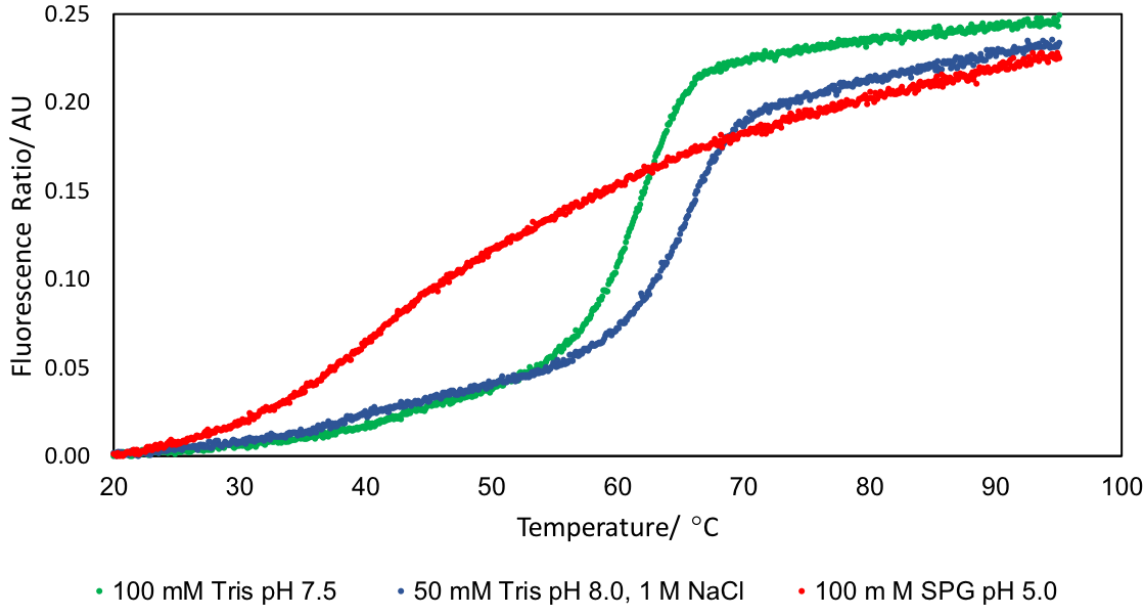


Figure 29: Thermal denaturation of Ccq1 130-717 as observed by nanoDSF. Ratio of fluorescence plotted against temperature for Ccq1 130-717 in three different buffer conditions. Increased pH and salt concentration lead to a shift in the plotted curve relating to a change in fluorescence ratio at higher temperatures, showing an increase in protein stability in 50 mM Tris pH 8, 1 M NaCl (blue curve) relative to the condition containing 100 mM SPG pH 5.0 (red curve). A low salt condition containing 100 mM Tris pH 7.5, has an intermediate stability (green curve), showing that both pH and ionic strength of the buffer impact Ccq1 130-717 stability.

To attempt to improve crystalisability, Ccq1 constructs with small N-terminal extensions were designed. Sequences corresponding to Ccq1 residues 120-717, 123-717, 126-717 and 130-717 were cloned into pET52b, which provides a 3C-cleavable Strep tag, and expressed as described previously. Of these, Ccq1 120-717 gave the highest levels of soluble expression and was selected for large

scale expression and purification (see Figure 30) by a similar three-step process to those used in previous Ccq1 purification protocols, except that nickel IMAC was replaced with Strep-tag affinity chromatography.

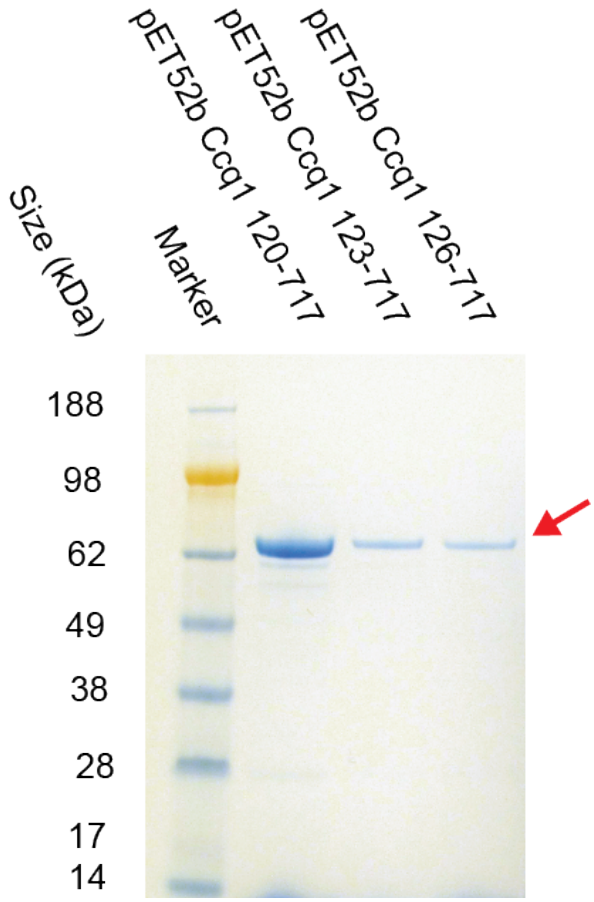


Figure 30: pET52b Ccq1 solubility tests.
Ccq1 is marked by a red arrow. Of the constructs expressed purified in parallel, Ccq1 120-717 gave the highest yield.

3.4.2 Expression summary and future uses for purified Ccq1

23 constructs were designed and successfully cloned for Ccq1. Soluble expression was observed for 15 of these cloned proteins and 12 of these soluble

proteins were purified. Purification of five of these proteins was optimised and then these were subsequently used for a number of downstream experiments.

The five lengths were used to structurally characterise Ccq1 and analyse it's interaction with Tpz1. Purified Ccq1 was used for SEC-MALLS, CD, and both high and low resolution structural characterisation by X-ray crystallography and SAXS (see Chapter 4). The binding affinity of Ccq1 to Tpz1 was measured by BLI and ITC (see section 5.2). Recombinant *S. pombe* Ccq1 was also purified from *E. coli* and insect cells in complex with other members of the shelterin complex for further structural analysis by electron microscopy (see section 5.4 and Chapter 6).

3.4.3 Small scale *E. coli* expression tests of Pot1

Pot1 contains several previously characterised domains (see Figure 31). Two N-terminal domains, which have previously been crystallised, are involved in recognition of single-stranded telomeric DNA but do not form interactions with other members of the shelterin complex [100, 101]. Instead, Tpz1 interacts with a C-terminal domain of Pot1 and this interaction has not been structurally characterised.

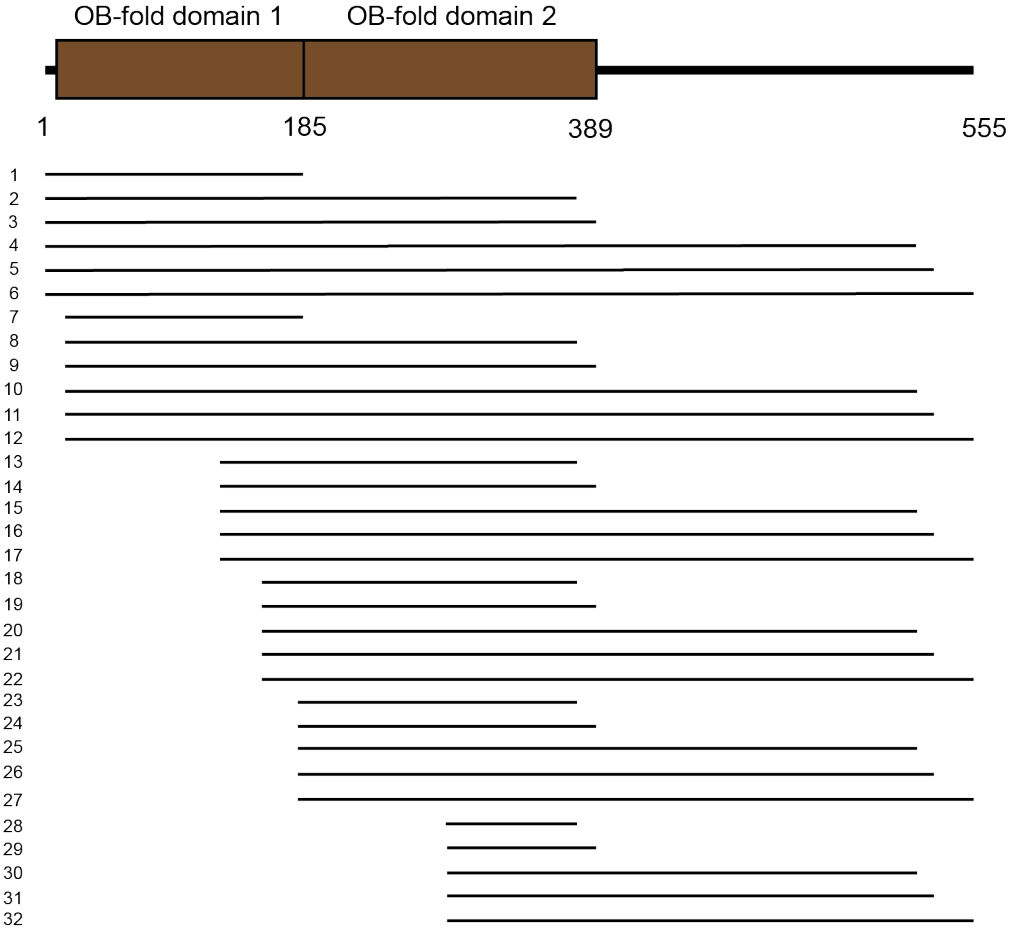


Figure 31: Pot1 domains and construct design.
The interaction between Pot1 and single-stranded DNA occurs through the OB-fold domains at the N-terminus of Pot1. Crystal structures are available for the interaction between the OB-fold domains and ssDNA (PDB: 1QZH, 4HJ5). An interaction with Tpz1 has been shown to require residues 139-555 of Pot1. However, a smaller more specific region may be identified. Constructs designed for protein expression are shown as black lines.

Twenty sequence-verified Pot1 constructs, with either an N-terminal 6 x His or GST tag, were expressed in *E. coli* at a small scale (see Table 15). High levels of protein expression were observed for His tagged constructs (see Figure 32) with negligible expression seen for constructs containing a GST tag, except for the construct containing residues 1-185. Solubility tests for His tagged constructs showed little or no protein recovery, except Pot1 residues 1-185, corresponding to the previously crystallised region (PDB: 1QZG). 64 constructs were designed

for Pot1 and 41 were successfully cloned. Expression was observed for 18 of these clones, although soluble protein was only recovered for 7 of these proteins. Soluble protein corresponds to the OB-fold region of Pot1, which interacts with DNA. As regions of Pot1 that interact with other members of the shelterin complex could not be purified, no further experiments were undertaken with Pot1.

Table 15: Pot1 constructs designed using secondary structure predictions. Successful cloning and expression are indicated by a tick, whereas unsuccessful cloning and expression is indicated by a dash. # = construct number, C = cloned, E = expressed, S = soluble, P = purified. + symbols in soluble/purified column relate to the relative amounts of each construct observed or recovered with + being the smallest amount and +++ being the largest. These relative amounts only relate to other Pot1 constructs.

#	Construct	Tag	C	E	S	P
1	pET47b Pot1 1-185	N-His	✓	✓	+++	-
2	pET47b Pot1 1-380	N-His	✓	✓	+	-
3	pET47b Pot1 1-389	N-His	✓	✓	+	-
4	pET47b Pot1 1-497	N-His	-	-	-	-
5	pET47b Pot1 1-501	N-His	-	-	-	-
6	pET47b Pot1 1-555	N-His	-	-	-	-
7	pET47b Pot1 5-185	N-His	-	-	-	-
8	pET47b Pot1 5-380	N-His	✓	✓	+	-
9	pET47b Pot1 5-389	N-His	-	-	-	-
10	pET47b Pot1 5-497	N-His	-	-	-	-
11	pET47b Pot1 5-501	N-His	✓	✓	-	-
12	pET47b Pot1 5-555	N-His	✓	✓	-	-
13	pET47b Pot1 131-380	N-His	✓	✓	-	-
14	pET47b Pot1 131-389	N-His	✓	✓	+	-
15	pET47b Pot1 131-497	N-His	✓	✓	-	-
16	pET47b Pot1 131-501	N-His	✓	✓	-	-
17	pET47b Pot1 131-555	N-His	✓	✓	-	-
18	pET47b Pot1 156-380	N-His	✓	✓	+	-
19	pET47b Pot1 156-389	N-His	✓	✓	-	-
20	pET47b Pot1 156-497	N-His	✓	-	-	-
21	pET47b Pot1 156-501	N-His	✓	-	-	-
22	pET47b Pot1 156-555	N-His	✓	✓	-	-
23	pET47b Pot1 184-380	N-His	-	-	-	-
24	pET47b Pot1 184-389	N-His	✓	✓	+	-
25	pET47b Pot1 184-497	N-His	✓	-	-	-
26	pET47b Pot1 184-501	N-His	✓	✓	-	-
27	pET47b Pot1 184-555	N-His	-	-	-	-
28	pET47b Pot1 289-380	N-His	-	-	-	-
29	pET47b Pot1 289-389	N-His	-	-	-	-
30	pET47b Pot1 289-497	N-His	✓	✓	-	-
31	pET47b Pot1 289-501	N-His	-	-	-	-

#	Construct	Tag	C	E	S	P
32	pET47b Pot1 289-555	N-His	-	-	-	-
1	pET49b Pot1 1-185	N-GST+ His	✓	✓	-	-
2	pET49b Pot1 1-380	N-GST+ His	✓	-	-	-
3	pET49b Pot1 1-389	N-GST+ His	✓	-	-	-
4	pET49b Pot1 1-497	N-GST+ His	✓	-	-	-
5	pET49b Pot1 1-501	N-GST+ His	-	-	-	-
6	pET49b Pot1 1-555	N-GST+ His	-	-	-	-
7	pET49b Pot1 5-185	N-GST+ His	-	-	-	-
8	pET49b Pot1 5-380	N-GST+ His	✓	-	-	-
9	pET49b Pot1 5-389	N-GST+ His	-	-	-	-
10	pET49b Pot1 5-497	N-GST+ His	✓	-	-	-
11	pET49b Pot1 5-501	N-GST+ His	✓	-	-	-
12	pET49b Pot1 5-555	N-GST+ His	✓	-	-	-
13	pET49b Pot1 131-380	N-GST+ His	✓	-	-	-
14	pET49b Pot1 131-389	N-GST+ His	✓	-	-	-
15	pET49b Pot1 131-497	N-GST+ His	✓	-	-	-
16	pET49b Pot1 131-501	N-GST+ His	-	-	-	-
17	pET49b Pot1 131-555	N-GST+ His	✓	-	-	-
18	pET49b Pot1 156-380	N-GST+ His	✓	-	-	-
19	pET49b Pot1 156-389	N-GST+ His	✓	-	-	-
20	pET49b Pot1 156-497	N-GST+ His	-	-	-	-
21	pET49b Pot1 156-501	N-GST+ His	✓	-	-	-
22	pET49b Pot1 156-555	N-GST+ His	✓	-	-	-
23	pET49b Pot1 184-380	N-GST+ His	-	-	-	-
24	pET49b Pot1 184-389	N-GST+ His	-	-	-	-
25	pET49b Pot1 184-497	N-GST+ His	✓	-	-	-
26	pET49b Pot1 184-501	N-GST+ His	✓	-	-	-
27	pET49b Pot1 184-555	N-GST+ His	✓	-	-	-
28	pET49b Pot1 289-380	N-GST+ His	-	-	-	-
29	pET49b Pot1 289-389	N-GST+ His	✓	-	-	-
30	pET49b Pot1 289-497	N-GST+ His	-	-	-	-
31	pET49b Pot1 289-501	N-GST+ His	-	-	-	-
32	pET49b Pot1 289-555	N-GST+ His	✓	-	-	-

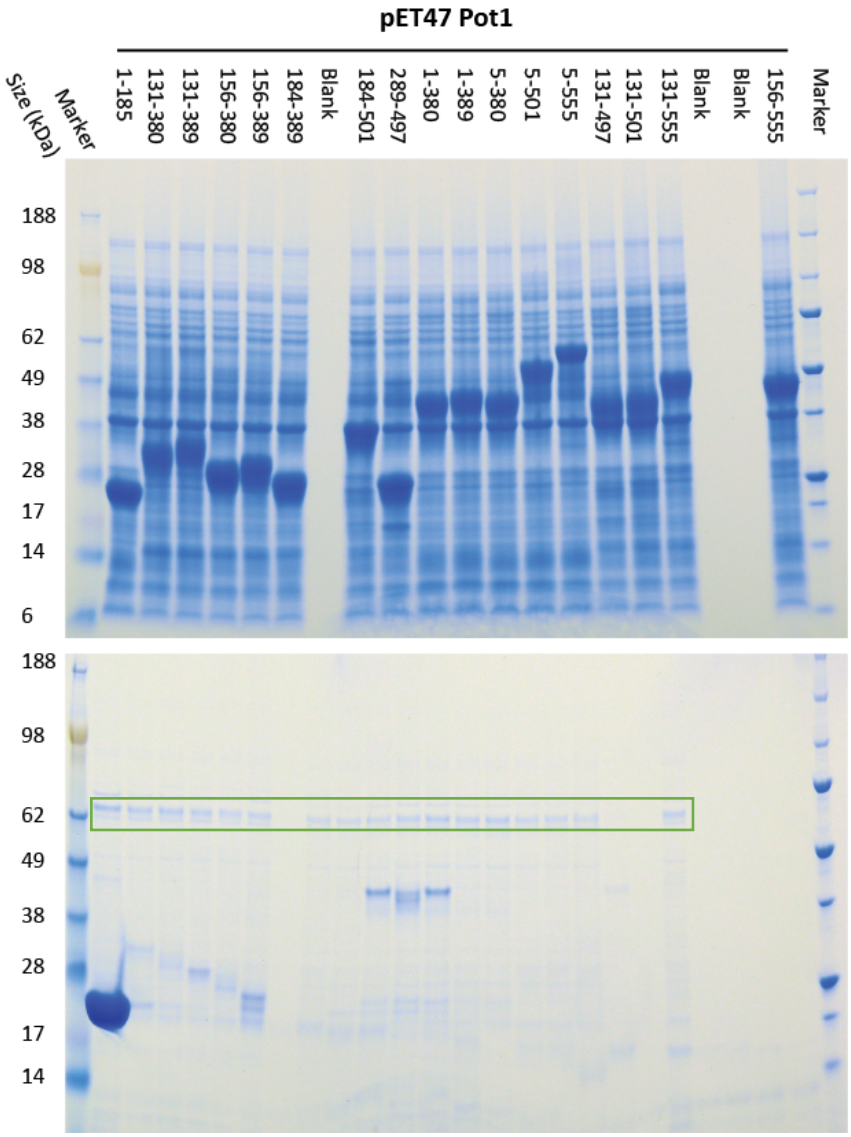


Figure 32: Pot1 expression and solubility tests.
Small scale Pot1 expression tests showed high levels of Pot1 expression, but very little soluble protein was recovered. An exception was the construct containing residues 1-185, which contains the ssDNA binding domain of Pot1 and a crystal structure of this domain has previously been solved (PDB: 1QZG). A consistent contaminant is observed at ~ 70 kDa which is presumed to be HSP70 (green box).

3.4.4 Tpz1 expression optimisation in *E. coli*

Tpz1 contains an N-terminal region that interacts with Pot1. The C-terminus of Tpz1 also contains two distinct binding sites for Ccq1 and Poz1, which bind separately, yet simultaneously, to these two sites and play important roles in telomere regulation and shelterin function [83] (see Figure 33).

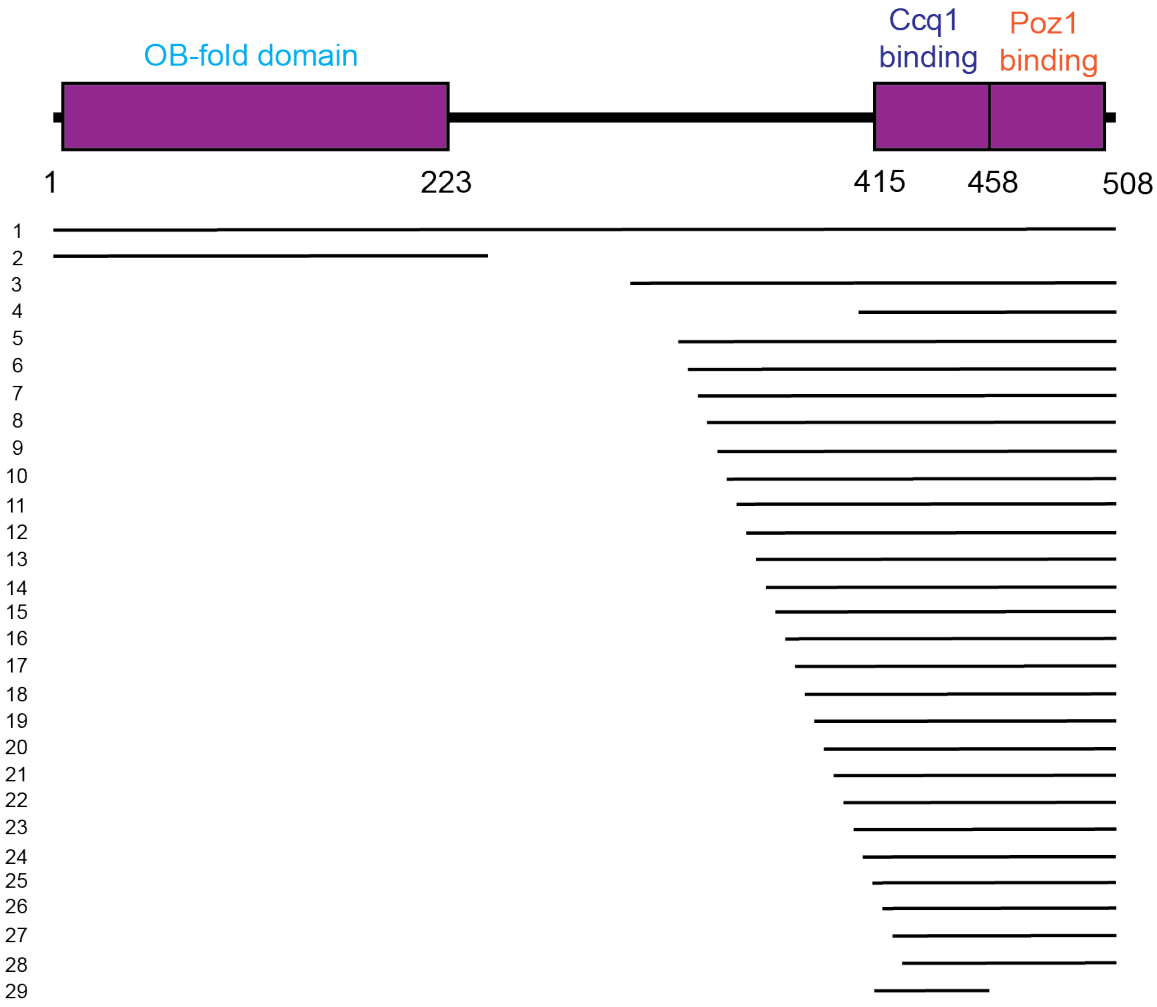


Figure 33: Tpz1 domains and construct design.
The N-terminal OB-fold domain interacts with Pot1, while the C-terminal region forms two separate interactions with Ccq1 and Poz1. Residues Y439, L445 and L449 are important for Ccq1 binding and W498 and I501 are involved in the Poz1 interaction. Constructs designed for protein expression are shown by black lines.

Four Tpz1 constructs were designed to include the Pot1 and Ccq1/Poz1 binding domains. These constructs were cloned into both pET47b and pET49b, which encode 3C cleavable 6 x His or GST tags respectively. For initial expression trials, plasmids were transformed into BL21 (Gold) and Rosetta2 cells and protein expression induced at both 16 and 25 °C. His tagged constructs exhibited the best protein over-expression with minimal differences observed between each media and induction temperature investigated. Once again, negligible soluble protein was recovered after an initial affinity purification step (see Figure 34). Since His tagged Tpz1 residues 405-508, which contains the binding sites for Poz1 and Ccq1, gave the highest yield of soluble protein this construct was chosen for further optimisation.

Table 16: Tpz1 constructs used in following experiments.
Successful cloning and expression are indicated by a tick, whereas unsuccessful cloning and expression is indicated by a dash. # = construct number, C = cloned, E = expressed, S = soluble, P = purified.+ symbols in soluble/purified column relate to the relative amounts of each construct observed or recovered with + being the smallest amount and +++ being the largest. These relative amounts only relate to other Tpz1 constructs.

#	Construct	Tag	C	E	S	P
1	pET47b Tpz1 1-508	N-His	✓	✓	-	-
2	pET47b Tpz1 1-230	N-His	✓	✓	-	-
3	pET47b Tpz1 331-508	N-His	✓	✓	-	-
4	pET47b Tpz1 405-508	N-His	✓	✓	+	-
1	pET49b Tpz1 1-508	N-GST+ His	✓	-	-	-
2	pET49b Tpz1 1-230	N-GST+ His	✓	-	-	-
3	pET49b Tpz1 331-508	N-GST+ His	✓	✓	-	-
4	pET49b Tpz1 405-508	N-GST+ His	✓	-	-	-
4	pGEX6p1 Tpz1 405-508	N-GST	✓	✓	+	+++
5	pET47b Tpz1 355-508	N-His	✓	✓	+	-
6	pET47b Tpz1 358-508	N-His	✓	✓	+	-
7	pET47b Tpz1 361-508	N-His	✓	✓	++	-
8	pET47b Tpz1 364-508	N-His	✓	✓	+++	-
9	pET47b Tpz1 367-508	N-His	✓	✓	+++	++
10	pET47b Tpz1 370-508	N-His	✓	✓	+++	++
11	pET47b Tpz1 373-508	N-His	✓	✓	+++	-
12	pET47b Tpz1 376-508	N-His	✓	✓	+++	-
13	pET47b Tpz1 379-508	N-His	✓	✓	+++	-
14	pET47b Tpz1 382-508	N-His	✓	✓	+++	-
15	pET47b Tpz1 385-508	N-His	✓	✓	+++	-
16	pET47b Tpz1 388-508	N-His	✓	✓	++	-
17	pET47b Tpz1 391-508	N-His	✓	✓	++	-
18	pET47b Tpz1 394-508	N-His	✓	✓	+++	-
19	pET47b Tpz1 397-508	N-His	✓	✓	+	-
20	pET47b Tpz1 400-508	N-His	✓	✓	+	-
21	pET47b Tpz1 403-508	N-His	✓	✓	+	-
22	pET47b Tpz1 406-508	N-His	✓	✓	+	-
23	pET47b Tpz1 409-508	N-His	✓	✓	+	-
24	pET47b Tpz1 412-508	N-His	✓	✓	-	-
25	pET47b Tpz1 415-508	N-His	✓	✓	-	-
26	pET47b Tpz1 418-508	N-His	✓	✓	-	-
27	pET47b Tpz1 421-508	N-His	✓	✓	-	-
28	pET47b Tpz1 424-508	N-His	✓	✓	-	-
29	pGEX6p1 Tpz1 415-458	N-GST	✓	✓	+	+

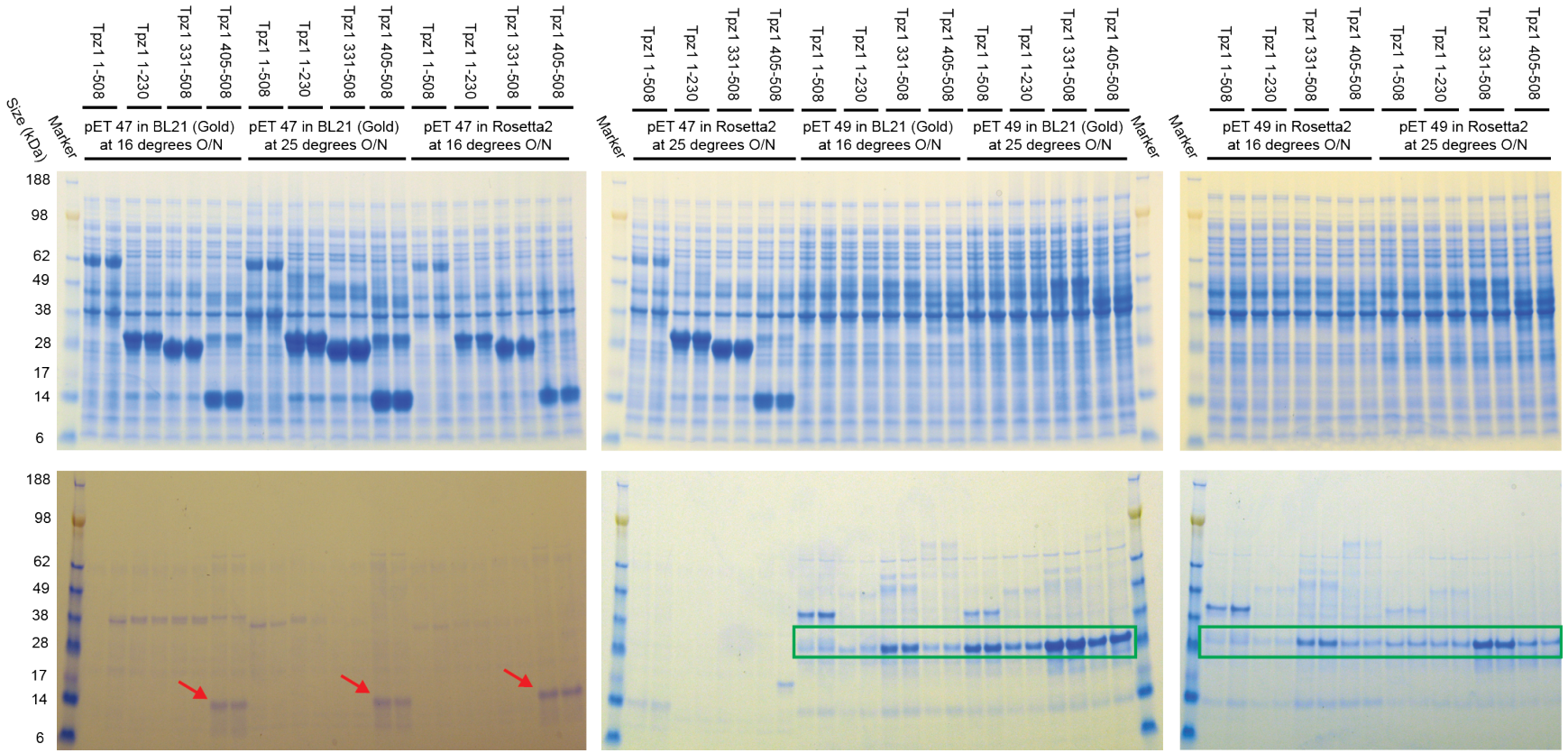


Figure 34: Tpz1 expression tests.

Bands of expected size for Tpz1 constructs are marked with a red arrow. A contaminant that is ~26kDa is marked with a green box and presumed to be GST. Expression of protein from two distinct colonies for pET47 (N-terminal 6 x his tag) or pET49 (N-terminal GST tag) Tpz1 constructs was tested. Whole cell samples show overexpression of Tpz1 (top panel), however little or no soluble protein was recovered after a nickel IMAC or glutathione sepharose affinity step (bottom panel). pET47 Tpz1 405-508 gives the largest amount of soluble protein for all tested conditions.

To this end, 24 constructs spanning the C-terminal region of Tpz1 were generated by PCR and cloned into pET47b. Constructs were designed with sequential 3 residue extensions, in the range 355-508 to 424-508. A peak in soluble expression levels was observed for constructs starting between 361 to 394 (see Figure 35). Constructs starting between 364-379 appear to be the most soluble and two constructs, at the peak solubility, encompassing residues 367-508 and 370-508 were chosen for expression and purification at a larger scale. Further improvements in solubility of Tpz1 were seen when the protein was expressed in a pGEX6P1 vector with an N-terminal 3C cleavable GST tag. This vector contains a shorter linker between the GST tag and the protein of interest than pET49. Despite increased solubility, purification of Tpz1 from contaminating degradation products remained a challenge. Co-expression experiments designed to stabilise Tpz1 are discussed in section 5.3.3.

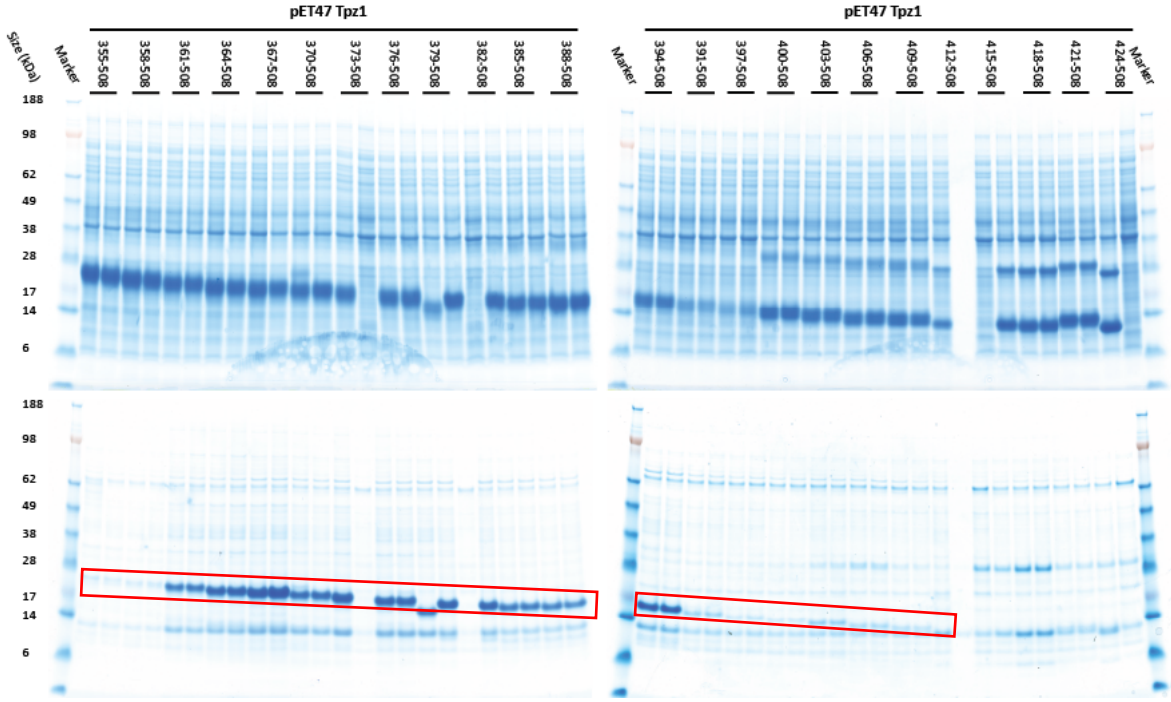


Figure 35: Tpz1 construct optimisation.
Twenty-four constructs were tested for expression and solubility. pET47 Tpz1 405-508 was used as a starting point and this construct was sequentially lengthened or shortened by three residues in an attempt to observe an improvement in soluble protein recovery. Constructs beginning between 361-394 and extending to the native C-terminus gave an increase in soluble expression. Protein bands within the red box correspond to soluble Tpz1 and were used to were used to give relative protein solubility amounts shown in Table 16.

Previous reports have shown that the region from Tpz1 residues 415-458 interacts with Ccq1 [89]. This peptide was successfully cloned and expressed with an N-terminal GST fusion in pGEX6p1, for use in Ccq1-Tpz1 interaction experiments. The GST-peptide fusion was purified by glutathione Sepharose affinity chromatography and the resulting protein analysed by SDS-PAGE. Two species were identified, GST-Tpz1 415-458 and a second protein assumed to be free GST (Figure 36).

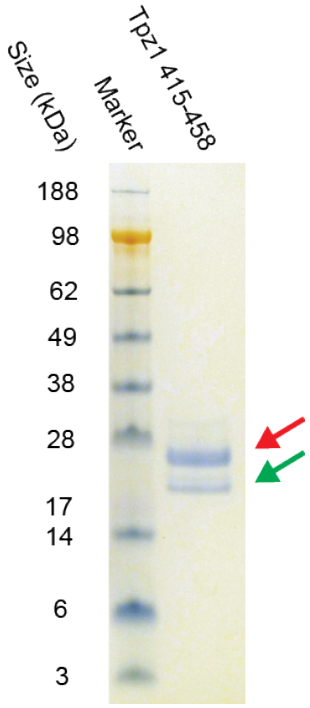


Figure 36: Expression test for GST tagged Tpz1 415-458.
Tpz1 415-458 was expressed with an N-terminal GST tag in pGEX6p1. The protein shown was purified by batch binding to 500 μL of glutathione Sepharose, with protein eluted by addition of 10 mM reduced glutathione pH 8.0 to the elution buffer. GST-Tpz1 415-458 is marked with a red arrow, the smaller contaminant (marked with a green arrow) is assumed to be free GST.

3.4.5 Expression summary and future uses for purified Tpz1

34 constructs were designed and successfully cloned for Tpz1, with expression observed for 31 of these proteins. Soluble protein was recovered after initial purification for 22 proteins and four protein fragments were fully purified for use in future structural analysis

Tpz1 purified protein was used to characterise interaction with its partners. The binding affinity of Tpz1 for both Poz1 and Ccq1 was tested using BLI and ITC and

the stoichiometry of the interaction between Ccq1 and Tpz1 was also probed using SEC-MALLS. Crystallisation experiments were attempted for the bipartite complex containing Ccq1-Tpz1 at several different ratios. Co-expression and co-purification of Tpz1 with other members of the shelterin complex was also attempted to see if this improved solubility of the proteins, in comparison to when expressed separately (see Chapter 5). Structural characterisation of these complexes by electron microscopy was attempted with and without stabilisation by cross-linking (see Chapter 6).

3.4.6 *E. coli* expression of Poz1

Poz1 was predicted to contain a majority of helical secondary structure using PSIpred [216] and the regions/domains that mediate interactions with other shelterin proteins, such as Tpz1, were largely unstudied (Figure 38). The expression and solubility of 7 different Poz1 constructs with either N-terminal His or GST tags was tested. His tagged Poz1 1-205, 70-205, 1-249 and 70-249 and GST tagged Poz1 1-205, 42-205, 70-205, 1-249 and 42-249 showed overexpression in whole cell samples. Higher overall overexpression levels compared to background protein expression were observed for His tagged proteins, although soluble protein was only recovered for His tagged full-length protein (see Figure 37).

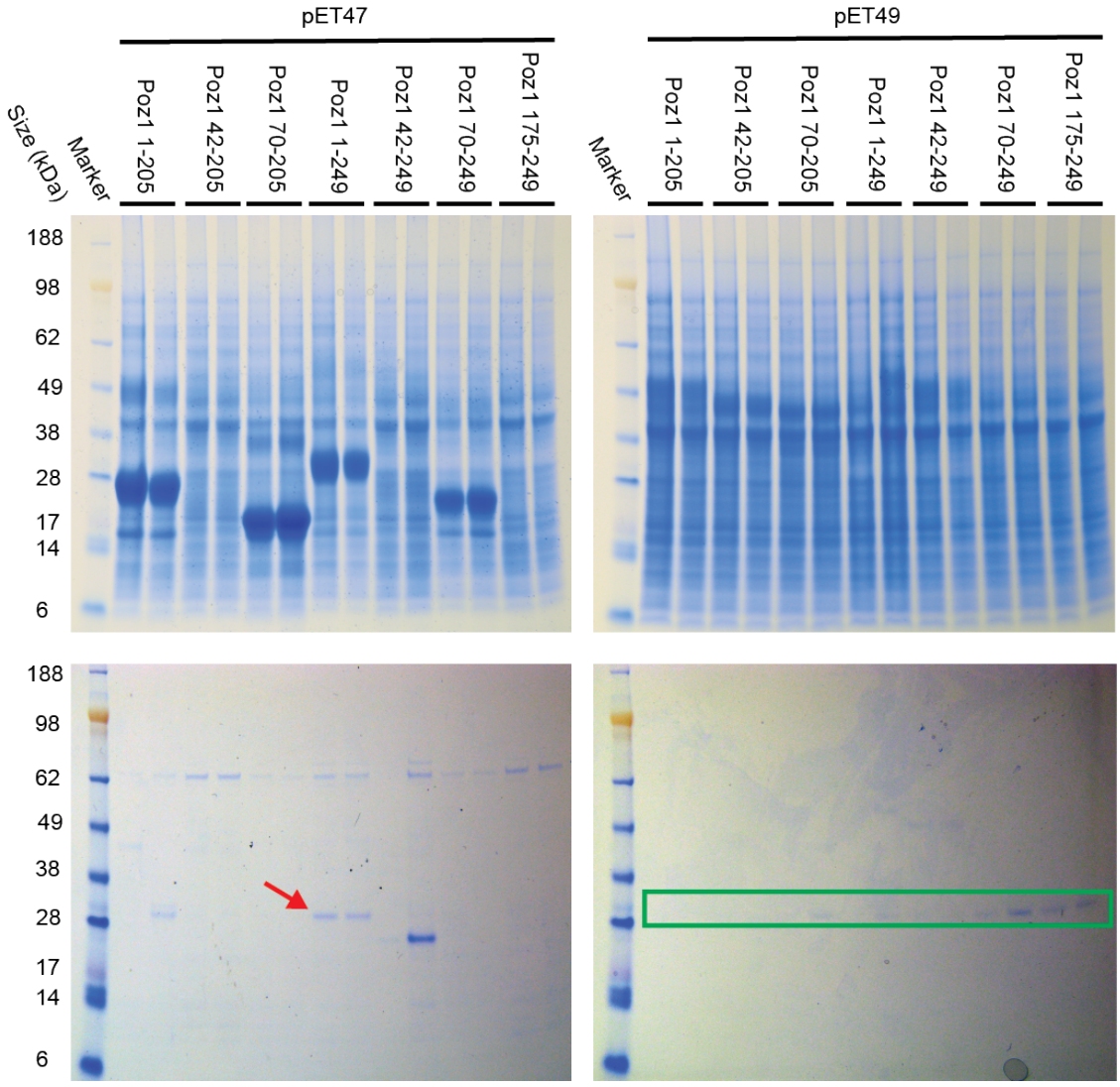


Figure 37: Poz1 expression tests
Expression tests were carried out at 1 mL scale at 37 °C with expression induced by addition of 1 mM IPTG once cells reached a density of OD_{600} 0.6-0.8. Cells were grown at 18 °C for 16-18 hours after induction, before harvesting. Soluble protein bands that are the expected size for Poz1 constructs are marked with a red arrow. Expression of each construct was checked for two distinct colonies of pET47 (N-terminal 6 x his tag) and pET49 (N-terminal GST tag) Poz1 constructs. Whole cell samples show intense bands relating to overexpression of His tagged constructs and less intense overexpression bands for constructs with a GST tag (top panel). Soluble protein is only recovered after a nickel IMAC or glutathione sepharose affinity step for pET47b Poz1 1-249 (bottom panel). There is a prominent protein contaminant in one lane corresponding to purified pET47b Poz1 42-249. This band is not at the expected size for Poz1 42-249 and occurs when no overexpression is observed in the related whole cell sample. The green box surrounds a contaminant in the pET49 related lanes which is assumed to be GST.

Yields of the full-length protein were, nonetheless, very low and so further constructs were produced in an attempt to improve solubility (see Table 17). A lack of consensus between a number of secondary structure prediction algorithms suggested that the N-terminal region was unstructured and its deletion might lead to an improvement in Poz1 solubility. Furthermore, a minimal Tpz1 binding domain in Poz1 had not been identified and therefore sequential 5 residue truncations of the N-terminus of Poz1 were produced and these experiments showed increased solubility and purity for both Poz1 25-249 and Poz1 30-249 after a single affinity step (see Figure 39).



Figure 38: Poz1 domains and construct design.
Poz1 was predicted to be to contain a majority of helical secondary structure. A minimum binding domain for Tpz1 had not been identified. Constructs designed for protein expression are depicted by black lines.

Table 17: Poz1 constructs initially designed using secondary structure predictions and then optimised by truncations.
Successful cloning and expression are indicated by a tick, whereas unsuccessful cloning and expression is indicated by a dash. # = construct number, C = cloned, E = expressed, S = soluble, P = purified. + symbols in soluble/purified column relate to the relative amounts of each construct observed or recovered with + being the smallest amount and +++ being the largest. These relative amounts only relate to other Poz1 constructs.

#	Construct	Tag	C	E	S	P
1	pET47b Poz1 1-205	N-His	✓	✓	-	-
2	pET47b Poz1 42-205	N-His	✓	-	-	-
3	pET47b Poz1 70-205	N-His	✓	✓	-	-
4	pET47b Poz1 1-249	N-His	✓	✓	+	+
5	pET47b Poz1 42-249	N-His	✓	-	-	-
6	pET47b Poz1 70-249	N-His	✓	✓	-	-
7	pET47b Poz1 175-249	N-His	✓	-	-	-
1	pET49b Poz1 1-205	N-GST+ His	✓	✓	-	-
2	pET49b Poz1 42-205	N-GST+ His	✓	✓	-	-
3	pET49b Poz1 70-205	N-GST+ His	✓	✓	-	-
4	pET49b Poz1 1-249	N-GST+ His	✓	✓	-	-
5	pET49b Poz1 42-249	N-GST+ His	✓	✓	-	-
6	pET49b Poz1 70-249	N-GST+ His	✓	✓	-	-
7	pET49b Poz1 175-249	N-GST+ His	✓	✓	-	-
8	pET47b Poz1 5-249	N-His	✓	✓	+	-
9	pET47b Poz1 10-249	N-His	✓	✓	+	-
10	pET47b Poz1 15-249	N-His	✓	✓	+	-
11	pET47b Poz1 20-249	N-His	✓	✓	++	-
12	pET47b Poz1 25-249	N-His	✓	✓	++	++
13	pET47b Poz1 30-249	N-His	✓	✓	++	++
14	pET47b Poz1 35-249	N-His	✓	✓	++	-
15	pET47b Poz1 40-249	N-His	✓	✓	++	-
16	pET47b Poz1 45-249	N-His	✓	✓	+++	-
17	pET47b Poz1 50-249	N-His	✓	✓	+++	+++
4	pET24b Poz1 1-249	C-His	✓	✓	-	-
4	pET28b Poz1 1-249	N-His	✓	✓	-	-

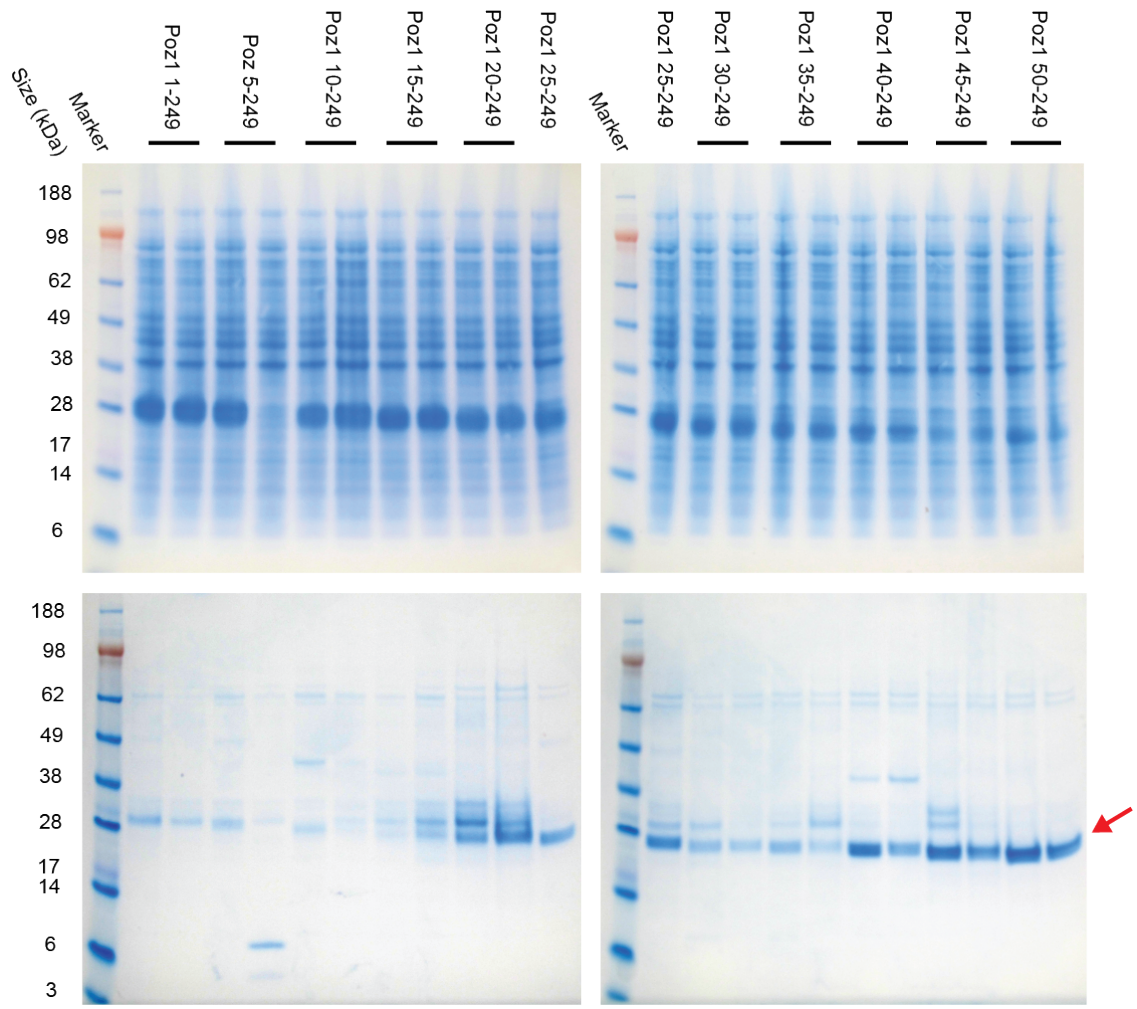


Figure 39: Poz1 expression tests.
Truncations were made to the N-terminus of Poz1 in an attempt to improve protein solubility. The shortest truncation, giving the Poz1 construct closest to full-length protein, was sought. Whole cell samples showed Poz1 overexpression had occurred (top panel). Poz1 25-249 and 30-249 gave improved solubility and the protein was relatively pure after a single affinity purification step (bottom panel). The most intense band in line with the red arrow corresponds to the target protein. Sample identity was confirmed by mass spectrometry.

Poz1 25-249 and 30-249 were expressed and purified to near homogeneity at a larger scale using IMAC, ion-exchange and size exclusion chromatography (see Figure 40). Although this protein could be concentrated to 0.5 mg/mL, at higher concentrations some precipitation was observed.

The pH of the buffer used in this protocol was critical to enable maximum protein recovery during the affinity and ion exchange steps. The theoretical pI of the protein was calculated to be 7.13 [220]. At pH 8, anion exchange chromatography was required, which led to loss of significant amounts of protein as Poz1 only weakly associates with the anion exchange resin. It was beneficial to lower the pH to 6.5 after affinity chromatography and use cation exchange chromatography to maximise protein yield. If a buffer pH of 6.5 was used for the whole purification scheme, binding of protein to the nickel IMAC was sub-optimal. Acidic pH interferes with binding of the His tag to the nickel column, reducing the yield of Poz1 through this step. Although protein precipitation was observed during the buffer exchange dilution step, this led to the minimum loss of protein during purification.

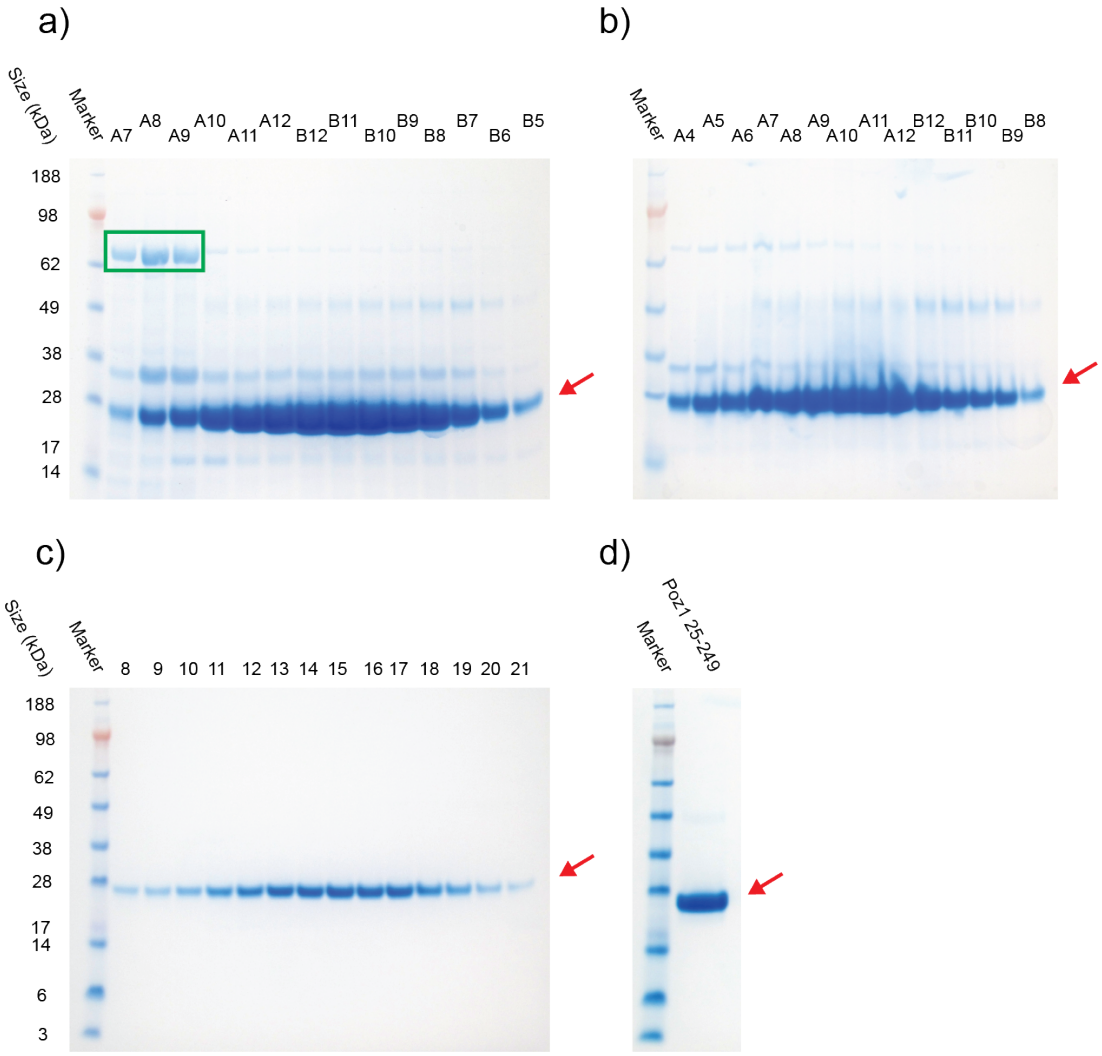


Figure 40: Poz1 25-249 purification
The red arrows indicate bands corresponding to the expected size for Poz1 25-249 through each stage of the purification. Band identity was confirmed by mass spectrometry after full purification a) The first step of Poz1 25-249 purification was performed by nickel IMAC. This involved a 5 mL HisTrap column, with protein eluted by a linear imidazole gradient from 25 mM to 500 mM. The band relating to Poz1 is observed just below the 28 kDa marker band. The green box shows a major contaminant, expected to be HSP70. b) Buffer exchange by dilution to lower the pH from 8 to 6.5 was performed. The protein was then loaded on to a 60 mL SourceS cation exchange column and protein eluted by a linear salt gradient from 250 mM to 1 M NaCl. c) Separation of the ion exchange eluate by 16/60 Superdex 75 size exclusion chromatography. d) Sample after concentration, showing protein purified to near homogeneity.

3.4.7 Expression summary and future uses for purified Poz1

26 constructs were designed and successfully cloned for Poz1, with 23 of these giving protein expression. 11 soluble proteins were recovered after initial tests and four proteins were purified for use in downstream structural analysis.

The affinity of the interaction between Poz1 and Tpz1 was characterised by BLI and any effects on the secondary structure of these two proteins due to occurrence of an interaction was probed by CD. Due to a high affinity interaction between Poz1 and Tpz1, co-expression of these proteins was tested to probe if this improved stability and solubility of the proteins during purification (see Chapter 5). Initial structural characterisation of the interaction by X-ray crystallography and electron microscopy was attempted, however this was more successful for larger complexes of shelterin components co-expressed in insect cells using the MultiBac system (see Chapter 6).

3.5 Summary

The first step towards gaining an understanding of telomerase regulation by Est1 and the shelterin complex was to express and purify the necessary components. A number of problems were encountered during recombinant expression and several different strategies were used to optimise the process. Synthetic genes, which had been optimised for codon usage in a specific expression system, were used in place of the native sequence, as different species have differing codon usage. This prevents blocks to translation occurring due to lack of the required tRNAs and/or interference with ribosome access to initiation sites by secondary structures formed in the mRNA. Different promoter systems were used for protein expression, such as the T7 and *tac* promoters. Genes under control of a T7 promoter require expression of T7 polymerase for induction to occur and so this prevents leaky expression, which can be beneficial for cell growth if the recombinant protein is toxic. The affinity tags on each fusion protein were also optimised. Different tags, such as GST, can improve the solubility of recombinant proteins. Also, the positioning of an affinity tag at the N- or C-terminus of a protein can improve expression levels and stability in situations such as where one of the termini are buried in the final folded structure. Fine screening of constructs and truncational analysis, often informed by proteolysis studies, was used to remove terminal unstructured regions that can often lead to aggregation. *E. coli*, yeast and insect cells were used for protein expression. Bacteria were used initially due to the ease and speed of this system and expression in insect cells was then

tested to see if access to eukaryotic protein folding machinery and post-translational modifications led to improvements in protein solubility. Finally, it is often the case that expression of protein orthologues from closely related species can be successful where there initial target is not, an approach inspired by John Kendrew's use of a panel of cetacean myoglobins during his Nobel Prize recognised work on myoglobin structure determination by X-ray crystallography [221].

As discussed, expression of Est1 in *E. coli* or *S. pombe* was unsuccessful, small amounts of Est1 were produced using baculovirus expression. Increasing the scale of expression experiments did not improve yield. To overcome problems related to *S. pombe* Est1 expression, orthologues of Est1 from related fission yeast species were trialed. Protein expression levels were more promising, with small amounts of *S. cryophilus* Est1 expressed and purified.

A large number of constructs of Ccq1, Poz1 and Tpz1 were prepared for biophysical and structural analysis. Four Ccq1 constructs could be successfully expressed and purified, resulting in a high yield. As a result, Ccq1 proteins could be used for structural characterisation and biophysical analysis of Ccq1 alone, as well as the Ccq1-Tpz1 interaction. The yield of Poz1 and Tpz1 proteins was much lower, restricting characterisation of the proteins individually. These proteins were used in biophysical experiments that require smaller amounts of protein, such as peptide arrays and biolayer interferometry. Protein fragments and their purification protocols are summarised in Table 18.

Table 18: Protein fragments important for future experiments and the protocol required for purification.
Several other Tpz1 fragments used in future experiments were produced, but these required co-expression with binding partners for improved solubility. Ni^{2+}: nickel affinity chromatography, Q/S IEX: ion exchange chromatography using Q or S resin, S75/S200: size exclusion chromatography using a Superdex S75 or S200 column (GE Healthcare).

Protein	Tag	Purification
Ccq1 130-717	N-His, 3C site	Ni^{2+}, Q IEX, S200
Ccq1 493-717	N-His, 3C site	Ni^{2+}, Q IEX, S200
Ccq1 1-735	N-His, 3C site	Ni^{2+}, Q IEX, S200
Ccq1 120-717	N-Strep, 3C site	Strep-Tactin, Q IEX, S200
Poz1 1-249	N-His, 3C site	Ni^{2+}, S IEX, S75
Poz1 25-249	N-His, 3C site	Ni^{2+}, S IEX, S75
Poz1 30-249	N-His, 3C site	Ni^{2+}, S IEX, S75
Tpz1 415-458	N-GST, 3C site	GSH

Chapter 4. Structural characterisation of Ccq1

4.1 Introduction

In fission yeast, phosphorylated Ccq1 recruits Est1 to the telomerase enzymatic subunit, Trt1 [120]. This is required for stabilisation of the telomerase holoenzyme at chromosome ends prior to its activation [85]. Subsequent binding of phosphorylated Ccq1 to Tpz1 has important regulatory functions and leads to the removal of repression of telomerase activity by the shelterin complex [87]. A human homologue for Ccq1 has not been discovered. It is possible that a C-terminal extension of TIN2 performs the actions of Ccq1 at human telomeres, due to identified interaction partners although it has no significant homology with Ccq1 [105]. A large number of disease causing mutations map to the TIN2 gene [145] and characterisation of Ccq1 could lead to confirmation that TIN2 can perform the vital regulatory interactions of Ccq1 at human telomeres.

Structural and biophysical analysis of Ccq1 included initial characterisation of several Ccq1 constructs by circular dichroism (CD) and size exclusion chromatography multi angle laser light scattering (SEC-MALLS). This was followed by extensive crystallisation screening and optimisation. Small angle X-ray scattering was used to determine a low resolution structure for Ccq1.

4.2 Biophysical characterisation of Ccq1

4.2.1 Secondary structure determination

Purification of full-length Ccq1 led to the recovery of sufficient amounts of protein to allow domain mapping by limited proteolysis and mass spectrometry, with a view to producing new constructs optimised for structural and biochemical studies (Figure 27a). Proteolysis produced two main species, Ccq1 130-717 as an intermediate produced during digestion and 493-717, which contained the minimum protease resistant domain. Secondary structure predictions show the presence of a mostly helical conserved region towards the C-terminus of the protein extending from approximately residues 490-720, with the rest of the protein predicted to contain a more mixed secondary structure (see Figure 41). The region from residues 140-180 was predicted, using the PRALINE prediction server, to contain another section of conserved α-helical secondary structure, with the region between 180 and 490 containing a mixture of α-helix, β-sheet and unstructured residues. The region N-terminal to residue 140 is much less conserved and less structured, consistent with its rapid degradation by trypsin.

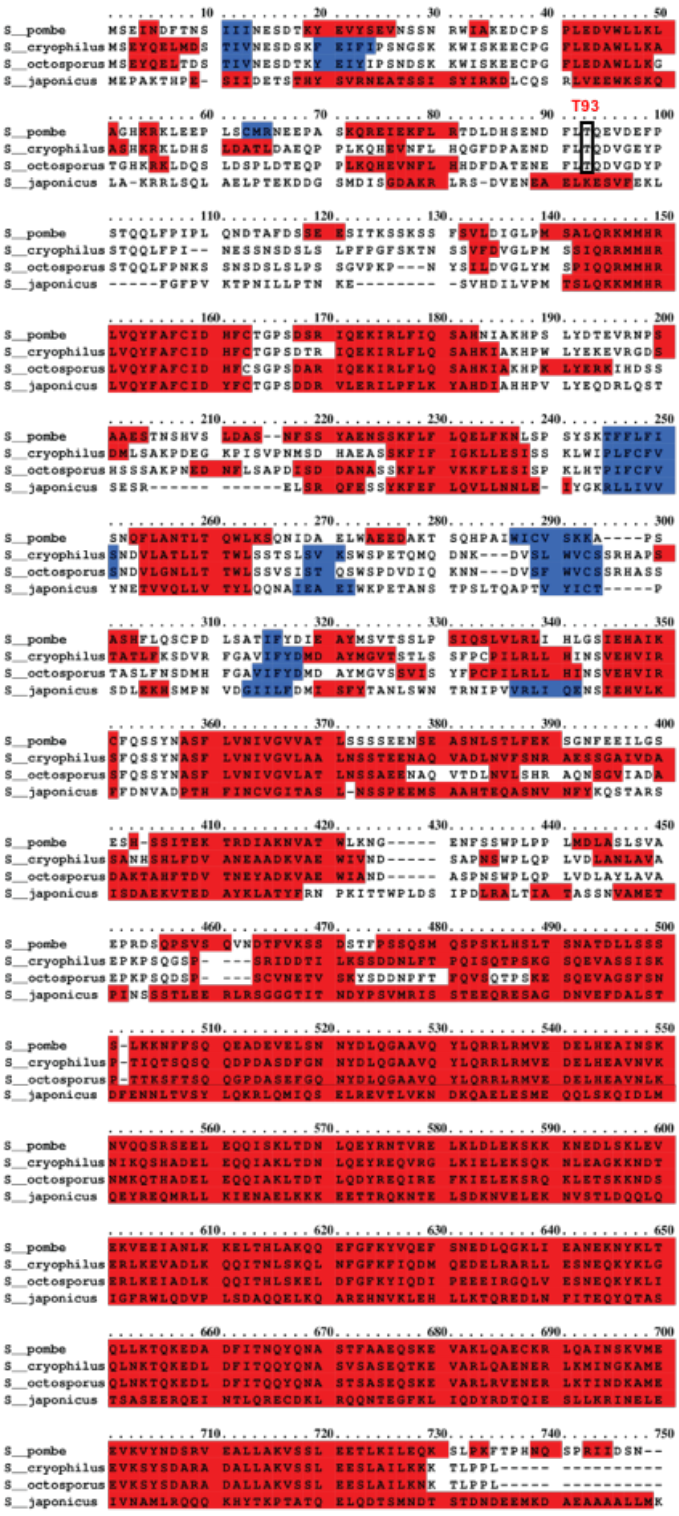


Figure 41: Secondary structure prediction of an alignment of Ccq1 homologues from closely related fission yeast species using PRALINE.

Red = α-helix, Blue = β-sheet, Black box highlights T93. A conserved α-helical region is observed at the C-terminus and the rest of the protein contains some conserved secondary structure features, but has a higher proportion of unstructured and β-sheet regions.

Secondary structural content of the Ccq1 130-717 and 493-717 proteins was determined using far UV circular dichroism spectroscopy (CD). Comparison of these spectra showed that Ccq1 493-717 was predicted to contain 83 % α-helices, with the extended protein from residues 130-717 predicted to contain a comparative higher proportion of β-sheets and random coil (see Figure 42) largely in agreement with the predicted secondary structural content (Figure 41).

	% α helical	% β sheet	% turn	% random
Ccq1 $_{493-717}$	83.35	0.1	5.6	11.1
Ccq1 $_{130-717}$	53.33	6.67	14.03	27.1

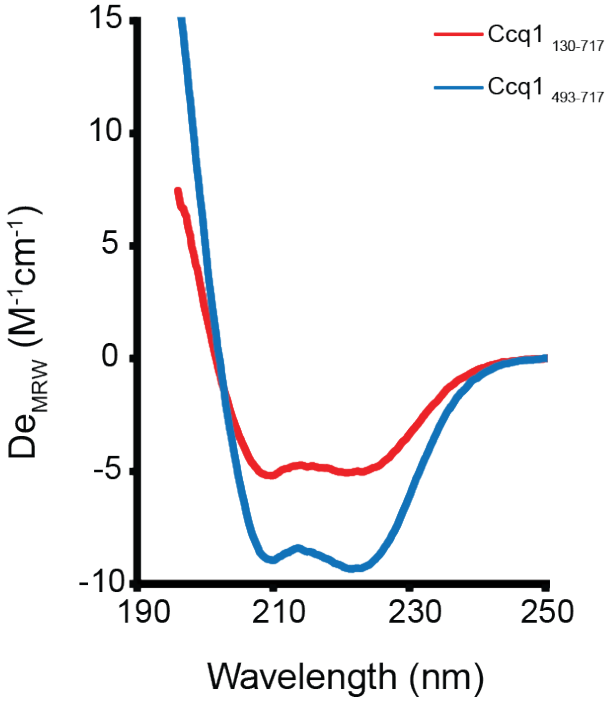


Figure 42: Characterisation of the secondary structure of Ccq1 130-717 and 493-717 by CD.
CD data showing secondary structure of the two constructs. The trace shows raw data, with values in the table relating to the alignment of this data to reference spectra.

4.2.2 Crystallisation trials

Both Ccq1 130-717 and 493-717 were used for crystallisation trials. Initial trials using protein purified by a two-step purification scheme, consisting of affinity chromatography followed by size exclusion chromatography, were unsuccessful. However, addition of an ion exchange step and resulting improvements in sample purity, yielded preparations of Ccq1 130-717 that produced small crystals in both 0.1 M sodium citrate pH 6.5, 1.4 M ammonium sulphate and 0.1 M sodium citrate pH 6.5, 16 % w/v PEG 3350 at a protein concentration of 8 mg/mL. These initial crystals were used to produce a seed solution for further optimisation experiments (see section 2.4.1) leading to larger crystals with a similar needle shaped morphology. PEG or ammonium sulphate concentration and buffer pH were varied in parallel and improvements in crystal formation were observed for PEG containing optimisation screens. Crystals only formed in conditions with a pH close to neutral and reducing the PEG concentration reduced nucleation and gave rise to larger crystals. These fluoresced under ultraviolet light illumination confirming that they were proteinaceous (see Figure 43), but failed to show any diffraction when tested using the in-house X-ray source.

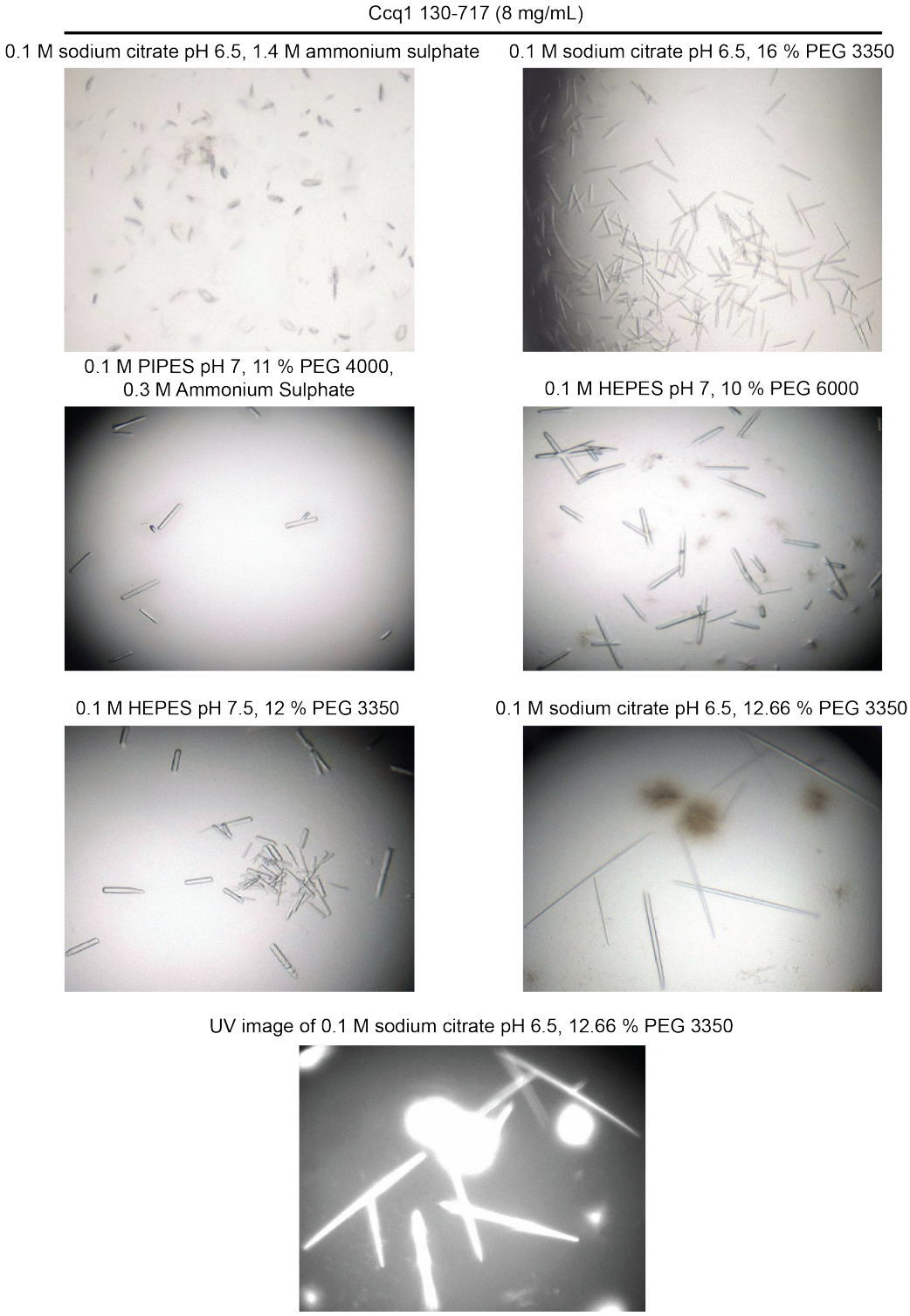


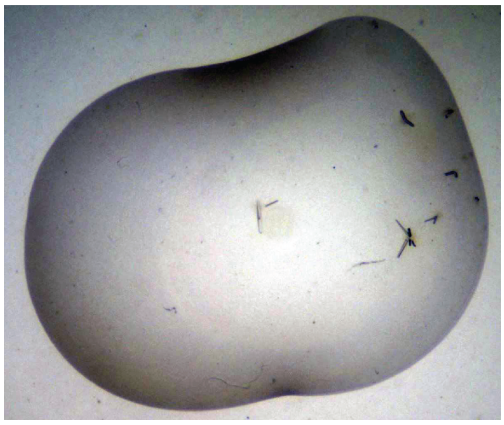
Figure 43: Initial crystallisation of Ccq1 130-717.
Crystallisation condition screening with a sample with improved purity gave small crystals in both 0.1 M sodium citrate pH 6.5, 1.4 M ammonium sulphate and 0.1 M sodium citrate pH 6.5, 16 % PEG 3350. Further condition screening and optimisation led to a reduction in crystal number and increase in crystal size. Protein was used at a concentration of 8 mg/mL.

To improve crystal quality, a number of different constructs of Ccq1 (120-717, 123-717 and 126-717) were tested with either a His or Strep affinity tag, but none of these crystallised across a broad range of standard screens. Crystallisation experiments were then repeated with both His tagged and untagged Ccq1 130-717, at concentrations of both 4 and 8 mg/mL and temperatures of 4 °C and 18 °C. This led to the appearance of crystals in 6.6 % PEG 6K, 0.225 M MES pH 6.6, 0.225 M Bis-Tris pH 6.6 for the 4 mg/mL 18 °C incubated His tag-containing protein sample. Again, these crystals showed UV (see Figure 44) and this condition was optimised by varying pH from pH 5.5-7 and PEG 6K concentration from 5-20 %, but attempts to reproduce these crystals were initially unsuccessful. Optimisation was carried out by hand with 1 μL drops and simultaneously with 200 nL drops set up using a Mosquito robot.

Ccq1 130-717 (4 mg/mL)

6.6 % PEG 6K, 0.225 M MES pH 6.6, 0.225 M Bis-Tris pH 6.6

Visible light image



UV image

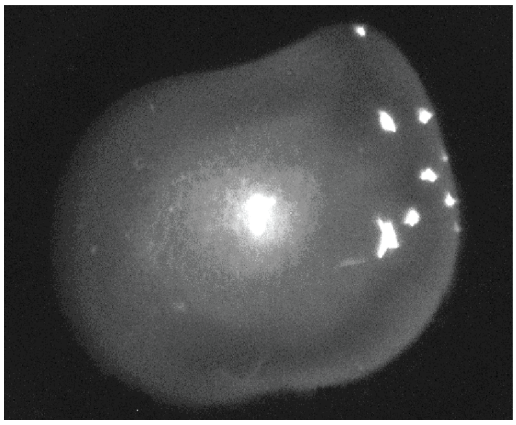


Figure 44: Crystallisation of Ccq1 130-717 using the Memgold screen (Molecular Dimensions).
The visible light image shows formation of a number of small needle shaped crystals in this crystallisation condition for His tagged Ccq1 130-717. These crystals showed tryptophan fluorescence when imaged with ultraviolet light.

However, addition of protein seeds, derived from the original crystallisation trials, led to formation of crystals in conditions containing 0.2 M $CaCl_2$, 0.1 M MES pH 6.5, 26 % PEG MME 350 and 0.1 M $CaCl_2$, 0.1 M Tris pH 6.5, 13 % PEG MME 2K. Optimisation of the three components of each condition led to an improvement in the size and number of crystals formed. When tested using an in-house X-ray source no diffraction was observed, but diffraction was observed for the original crystals (Figure 44) at the Swiss Light Source. Diffraction was limited to 8 Å resolution (see Figure 45) and this low-quality data prevented indexing of the pattern. The diffraction spots observed are close together and due to the inverse relationship between the distance between diffraction spots and the size of the unit cell, this suggests that the unit cell is large in at least one dimension.

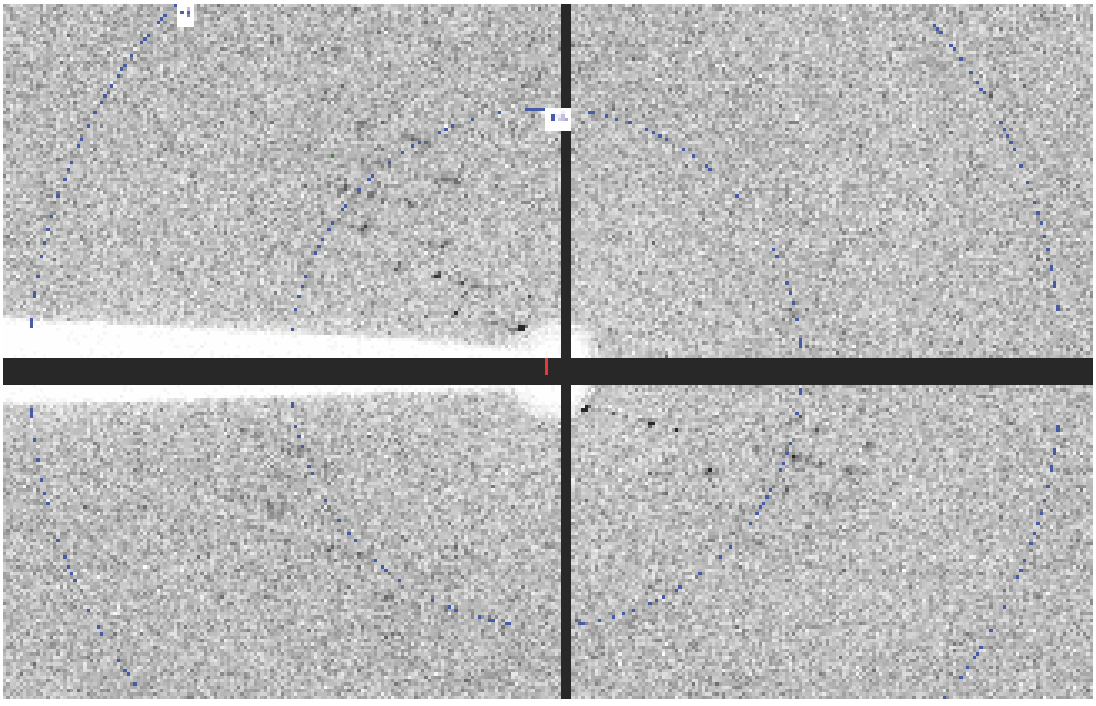


Figure 45: Diffraction image for Ccq1 130-717.
Diffraction was observed to approximately 8 Å resolution, however indexing of these data was not possible. The inner resolution circle is at 12 Å resolution and the outer circle at 6 Å resolution.

4.2.3 Determination of the oligomeric state of Ccq1

SEC-MALLS was used to determine the oligomeric state of Ccq1 in solution. This technique measures changes in the refractive index and light scattering properties of a protein solution, after separation by size exclusion chromatography [188] (see section 2.3.4). Three different loading concentrations (1, 3 and 6 mg/mL equivalent to 38, 114 and 228 μM for Ccq1 493-717 and 15, 45 and 90 μM for Ccq1 130-717) were tested to determine whether any concentration dependent oligomerisation occurs. Ccq1 493-717 has a calculated mass of 26.2 kDa, but SEC-MALLS reported a mass of ~50 kDa. Ccq1 130-717 has a calculated mass of 66.4 kDa, with SEC-MALLS showing a mass of ~130 kDa. Thus, both proteins are dimeric in solution. Furthermore, no significant difference in apparent mass was observed between the three concentrations for either sample, indicating that no dissociation occurs in this concentration range (see Figure 46). Some further oligomerisation of Ccq1 130-717 was observed towards the beginning of the elution profile. The concentration of Ccq1 493-717 in the peak of the SEC-MALLS resulting from loading at 1 mg/mL was ~1 μM and for the 130-717 construct was ~0.2 μM. As no dissociation of the dimer was observed, this suggests that the affinity of dimerisation is likely in the nanomolar range or tighter.

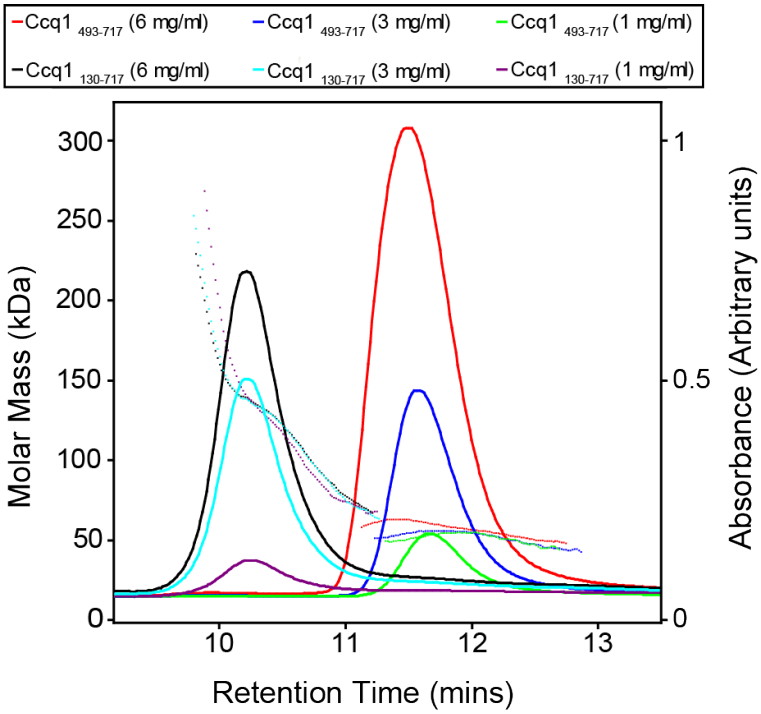


Figure 46: The oligomeric state of Ccq1 130-717 and 493-717 was characterised by SEC-MALLS.
SEC-MALLS trace showing the solution composition of both Ccq1 130-717 and 493-717 at three protein concentrations. Absorbance is shown by a continuous line with molar mass depicted by points. Ccq1 493-717 is 26.2 kDa in size with SEC-MALLS giving a mass of approximately 50 kDa, with Ccq1 130-717 being 66.4 kDa and SEC-MALLS showing a mass of ~130 kDa.

When analysed by size exclusion chromatography, Ccq1 130-717 eluted from a Superdex 200 column very close to the void volume. Comparing this elution volume to the volume at which protein standards of a known mass elute from the same column, Ccq1 elutes at a volume that does not match either its theoretical or determined mass. Protein standards are typically globular, near spherical proteins. Elongated proteins with a similar mass appear larger than spherical proteins, as they have an increased radius of gyration. The most plausible explanation for this increase in apparent size is that dimeric Ccq1 has an extended conformation.

4.2.4 Domain arrangement of Ccq1

Both the CD analysis of Ccq1 130-717 and the secondary structure prediction with PRALINE show a mixed α/β secondary structure. Consistent with previous observations [83, 84, 142], alignments generated using HHpred [222] predicted the presence of two domains, the first of which aligned most closely with an HDAC domain found in budding yeast, which has a mixed α/β architecture. This region also aligned with similar confidence to a number of other structurally related proteins involved in chromatin remodeling. The second predicted domain was a C-terminal coiled-coil region consistent with the high helical content observed by CD analysis. The top ten matches in HHpred consisted of a mixture of parallel and antiparallel coiled-coils and so no confident predictions could be made as to the orientation of the domain in Ccq1.

4.2.5 Small angle X-ray scattering (SAXS) analysis of Ccq1 130-717

Due to the inability to produce crystals that diffracted to a sufficient resolution within the required timeframe, a low-resolution structure of Ccq1 130-717 was determined by SAXS analysis. Data were collected in batch mode at the P12 beamline at the Petra III storage ring at DESY at concentrations of 18.03, 9.42, 4.57, 2.31, 2.22, and 1.21 mg/mL. Initial analysis was performed on each data set separately using ScÅtter software [223], before the median data set was calculated (Figure 47a). Low resolution data points for high concentration samples and high-resolution data points for low concentration samples were

removed from the median data set, improving the accuracy of the final generated model. This prevented any effect on the calculated model occurring due to any aggregation at high concentrations in the low-resolution data region and also removed the increased noise at higher resolution found in low concentration samples.

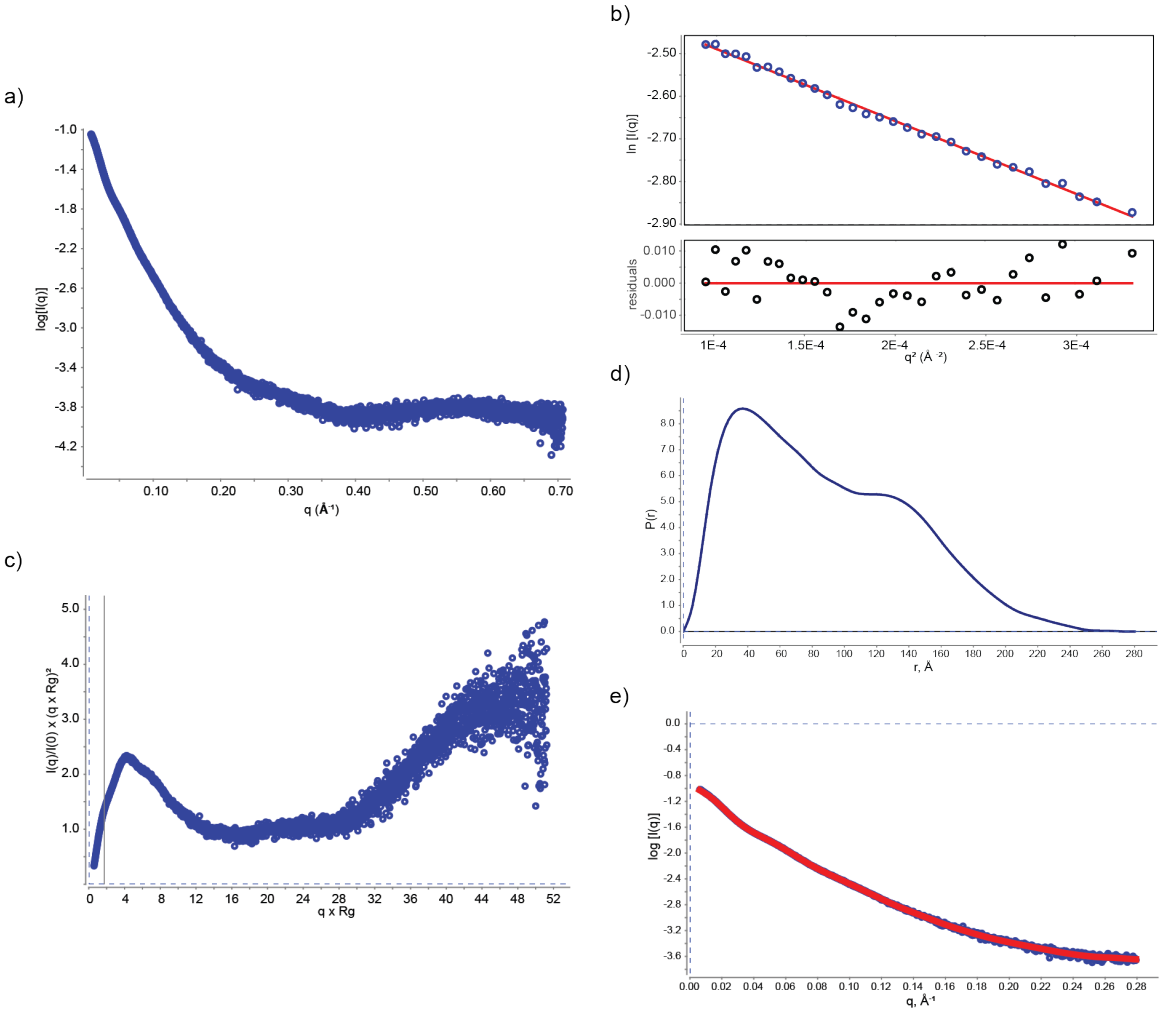


Figure 47: SAXS analysis of Ccq1 130-717.
ScÅtter [223] was used for all initial data analysis.
a) Median scattering data obtained for a concentration series of Ccq1 130-717. b) The low-*q* linear regions of the scattering curve used for Guinier analysis to determine the I(0) and R_g values. c) Kratky analysis of the data set showing the presence of an extended species, with possible regions of flexibility. d) Normalised interatomic pairwise distribution function, P(r), showing a maximum particle size, D_{max}, of 280 Å. Each peak within the profile relates to a region of structure within the protein and the distance between each region. The asymmetrical shape shows presence of an extended species. e) The P(r) curve led to calculation of a scattering curve (red) which showed a similar shape to the measured scattering data (blue).

Parameters derived from the analysis of the scattering curves are shown in Table 19. Data were collected to a *q* of 0.7 $Å^{-1}$, however for analysis this data range was cut to 0.006 – 0.28 $Å^{-1}$.

Table 19: SAXS-derived parameters for Ccq1 130-717.
q is the momentum transfer function, which is measured during the scattering experiment which equals $4\pi\sin(\theta)/\lambda$, where θ is the half the angle of scatter and λ is the wavelength of the incident beam.

Structural parameters	**Values**
q range ($Å^{-1}$)	0.006 – 0.28
I(0) (cm^{-1}/absorbance)	0.099
R_g (Å)	72.28
D_{max} (Å)	280
Porod Volume estimate ($Å^3$)	302011
SAXS-derived molecular mass (ScÅtter) (kDa)	132
SAXS-derived molecular mass (SAXSMoW) (kDa)	144.3
Calculated molecular mass (from sequence) (kDa)	66.35

Data analysis using ScÅtter [223] gave a value for the radius of gyration (R_g) of 72.28 Å, a maximum intensity (I(0)) of 0.099 cm^{-1}/absorbance and a maximum particle dimension (D_{max}) of 280 Å. R_g and I(0) were calculated by Guinier analysis and the D_{max} was calculated from the P(r) distribution (see Figure 47b and d). These values show that Ccq1 130-717 has an extended conformation, which is likely to be proportionally narrow in the perpendicular dimension. The distance distribution function (P(r)) showed an asymmetrical shape, which also suggests that the particle is elongated. Furthermore, the Kratky plot does not return to zero,

suggesting an elongated particle shape that may have some flexibility. The molecular weight of the particle was determined using both ScÅtter [223] and SAXSMoW [224], giving values of 132 kDa and 144.3 kDa respectively. The molecular mass of a monomer of Ccq1 130-717 is 66.35 kDa, so the SAXS-derived envelope contains a homodimer of Ccq1 and is consistent with the MALLS analysis.

DAMMIF was used for *ab initio* calculation of models showing a low resolution molecular shape for Ccq1 [196] and the program was executed with and without enforcing P2 symmetry (see Figure 48). This program more heavily weights low resolution data during structural calculations, so even though data to q = 0.28 Å^{-1} were included for analysis, the higher resolution data does not have a large effect on the calculated model. Averaging 50 individual runs, using DAMAVER [197], produced the final models shown in Figure 48. The model produced without imposing symmetry more closely resembles the data collected, as shown by lower χ^2 values (mean $\chi^2 = 0.67 \pm 0.19$). The model with P2 symmetry enforced has an average χ^2 value of 1.68 ± 0.60. However, fitting of the symmetrical and un-symmetrical maps, using chimera [225], gave a correlation of fit of 0.91 (Figure 48b), showing that they are highly similar and so it is possible that features of both maps are correct.

Both SAXS-generated maps show an extended dimeric structure for Ccq1. When coupled with the SEC-MALLS, CD and secondary structure predictions this points towards the likely presence of a coiled-coil domain towards the C-terminus of the

protein, which is responsible for dimerisation. The presence of a C-terminal coiled-coil was also predicted using PCOILS [226], the COILS server [218] and LOGICOIL [219].

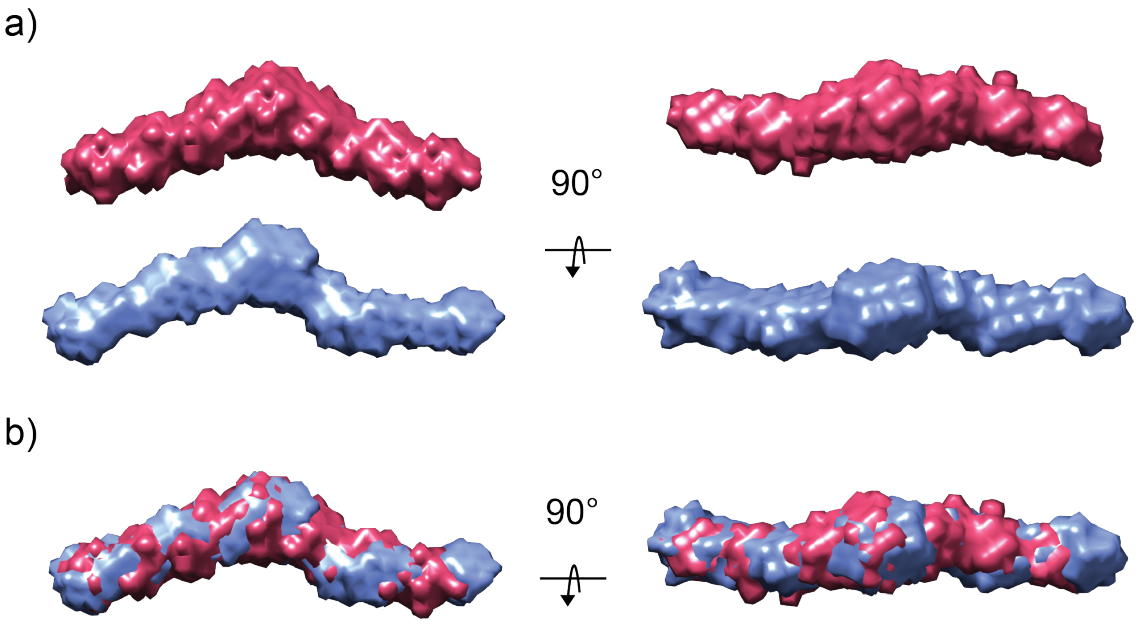


Figure 48: SAXS-derived envelope of Ccq1 130-717, showing the effect of enforcing P2 symmetry on the final model.
Blue model = no symmetry enforced. Red model = P2 symmetry enforced.
a) Both envelopes derived from the SAXS data show an extended conformation, with a slight bend at the centre. Ccq1 is dimeric and this appears to be mediated by a putative C-terminal coiled coil. b) An overlay of the two structures shows a high level of similarity and the fit has a correlation of 0.91.

The most likely conformation of Ccq1 is an antiparallel coiled-coil arrangement (Figure 49 – Model A). This would lead to the globular HDAC-like domains being found at the opposite ends of the structure, with dimerisation mediated by the coiled-coil domain. Coiled-coil dimerisation would correlate with the low nanomolar affinity of the interaction observed in section 4.2.3. The *ab initio* SAXS model has sufficient volume to fit a single globular domain at most points in the structure and a similar kinked conformation to the model shown in Figure 48 has been observed previously for an antiparallel coiled-coil [227]. The Kratky plot

produced during SAXS analysis of Ccq1 (see Figure 47c) shows an elevated baseline at high q values, which qualitatively suggests flexibility in the protein, but does not allow for quantitative analysis of the degree of flexibility.

Other models are possible but do not agree with several constraints. Firstly, parallel dimerisation of the coiled-coil region of Ccq1, which would lead to the N-terminal globular domains of each monomer being in close proximity to one another, would further require dimerisation or close association of the N-terminal domains (Figure 49 – Model B). Although dimeric association of HDAC-family domains has been observed [228], this would result in a structure too large to fit in any part of the envelope. Finally, although consistent with the shape of the SAXS envelope, Model C requires that dimerisation be mediated mostly by the N-terminal globular region rather than the coiled-coil and can therefore be reasonably disregarded.

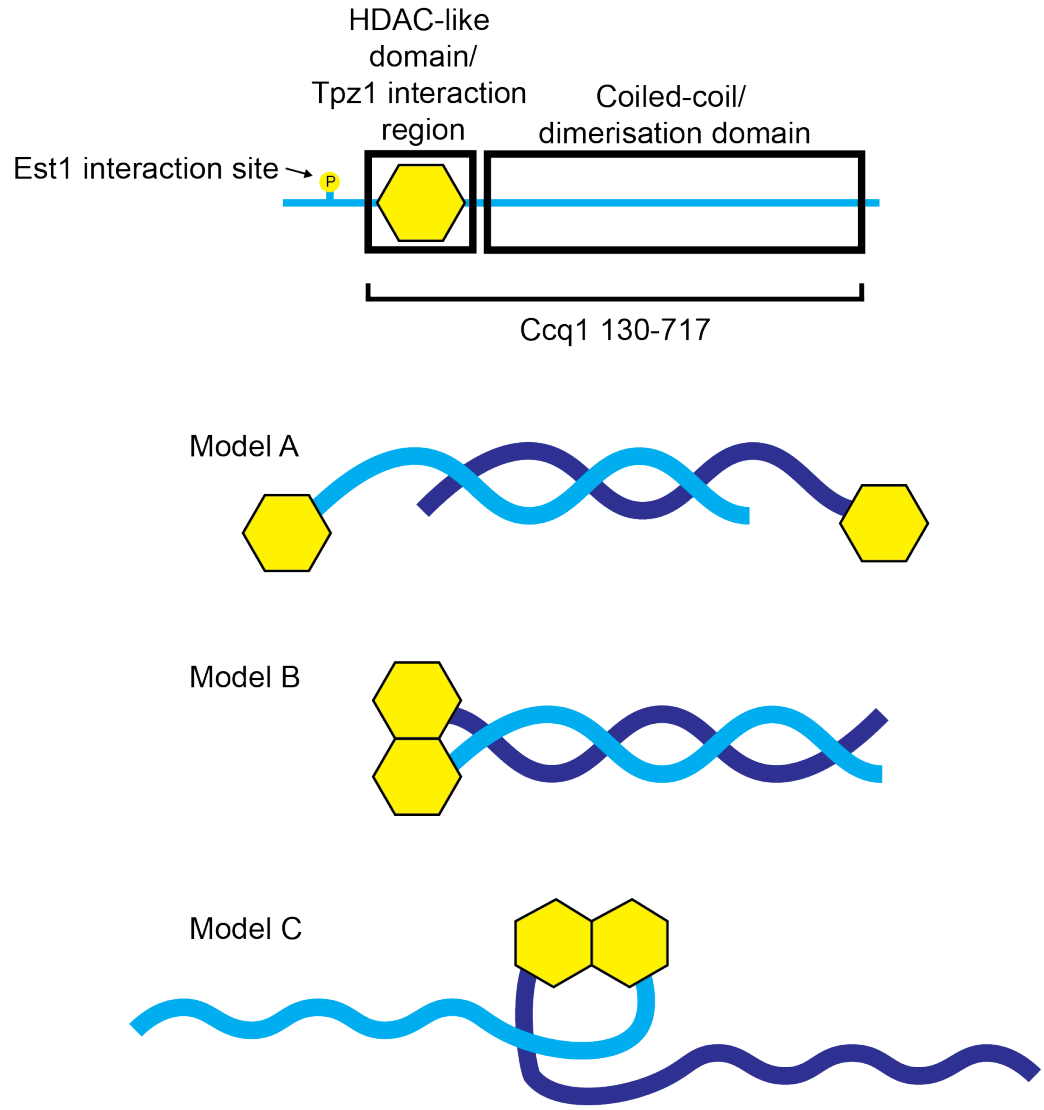


Figure 49: Potential modes for dimerisation of Ccq1 130-717.
Globular HDAC-like domains are shown by yellow hexagons with the coiled-coil region shown by intertwined blue lines. Est1 interacts with threonine residue 93 which is not in the construct characterised by SAXS. Model A shows anti-parallel dimerisation of the coiled-coil, leading to the globular domains being found at opposite poles of the structure. In model B, parallel dimerisation of the coiled-coil occurs, leading to dimerisation of the globular domains. Model C shows dimerisation of the globular domains and anti-parallel dimerisation of a small region of the coiled-coil leading to the globular domains being found at the centre of the dimer at the symmetry axis.

4.3 Conclusions

Redundant phosphorylation of Ccq1 by Tel/Rad3 leads to recruitment of Est1 to the telomerase enzymatic subunit [85]. This required for stabilisation of telomerase at telomeres. Binding of Ccq1 to the shelterin complex causes a switch from a non-extendible to an extendible state [87]. At present, a human homologue for Ccq1 remains unknown and structural characterisation of *S. pombe* Ccq1 would show the mechanism by which this protein acts, which could lead to identification of potential human proteins that carry out a similar process.

Ccq1 was initially characterised by limited tryptic proteolysis and CD. Proteolysis produced two protein fragments, Ccq1 130-717 and 493-717, which were used for future experiments. CD showed the C-terminal region of Ccq1 was mostly helical and the region from residues 130-493 was less structured. Consistent with the likely presence of a coiled-coil region at the C-terminus of Ccq1, tight dimerisation of both Ccq1 species was observed by SEC-MALLS. Crystallisation trials gave rise to a number of crystals under several conditions for Ccq1 130-717, which were optimised. Diffraction to low resolution was obtained for a single crystal form but was insufficient to allow indexing. A wide range of further trials were undertaken for Ccq1 493-717 and no crystals formed. Hydrodynamic analyses using SEC-MALLS and SAXS show that the protein has an extended conformation with dimerisation likely occurring at the centre of the envelope most probably through an anti-parallel coiled-coil assembly.

The coiled-coil dimerisation mode remains unknown. Anti-parallel dimerisation, together with the flexibility implied from the SAXS analysis, would suggest a model in which the globular HDAC-like domains are positioned at either end of a tube-like structure and are connected by a linker or ‘hinge’ region to a central dimeric coiled-coil. Together these data and the possible models for Ccq1 organisation have implications for its overall function and structural placement in the shelterin complex. Methods to verify these potential models are described in Chapter 7.

Chapter 5. Biophysical and structural characterisation of shelterin sub-complexes

5.1 Introduction

Tpz1 forms a bridge within the shelterin complex between single and double-stranded DNA binding components [83]. To perform this role, Tpz1 contains domains that interact with Pot1, Poz1 and Ccq1. A complete link in the shelterin complex between double and single-stranded binding components has been shown to be required for maintenance of telomeres in a repressed non-extendible state and the Tpz1-Ccq1 interaction releases this repression [87]. Disruption of binding of Tpz1 to Pot1 or Poz1 leads to unregulated extension of telomeres *in vivo* [93, 103].

The aim of this part of the project was to structurally and biophysically characterise these interactions. Prior studies by yeast-two hybrid and pull-down analysis have shown qualitatively that Pot1 interacts with an N-terminal OB-fold domain within Tpz1, and Poz1 and Ccq1 bind to two distinct regions in the C-terminus of Tpz1 (Figure 50) [82, 83, 87]. Due to the limited soluble expression of Tpz1 and Poz1, different approaches, such as biolayer interferometry and co-expression of sub-complexes, were required to characterise and quantify these interactions.

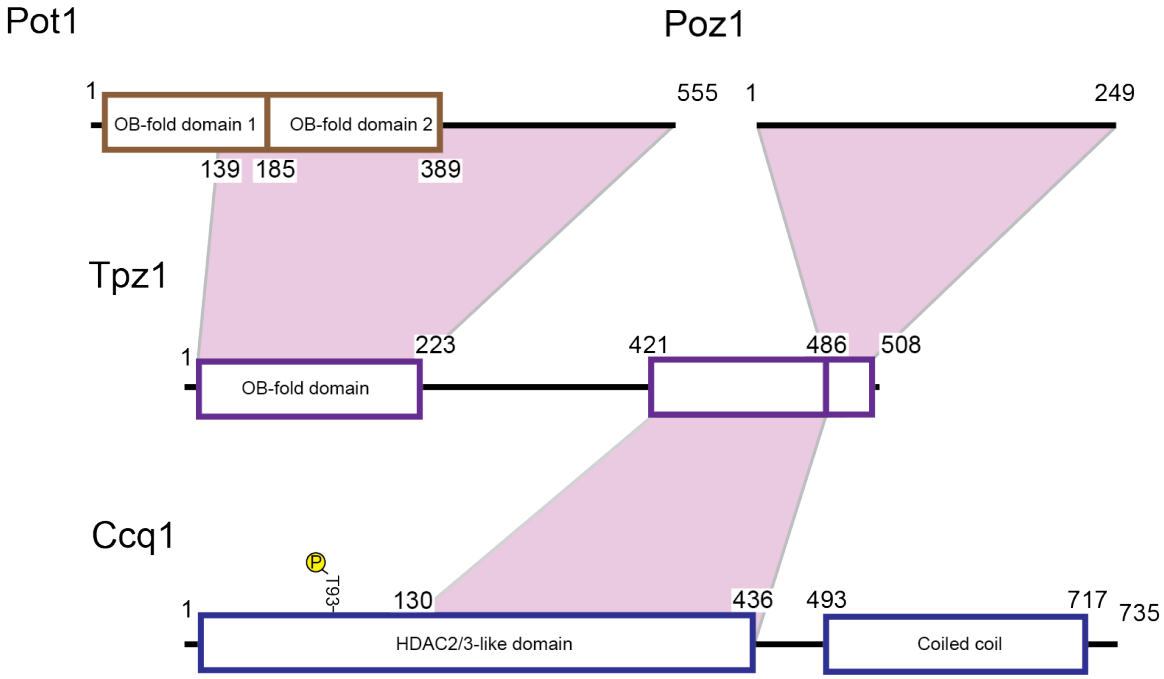


Figure 50: Known protein domains and regions that interact with Tpz1.
Pot1 interacts with an N-terminal domain of Tpz1, while Poz1 and Ccq1 bind to two separate regions at the C-terminus of Tpz1.

5.2 The Tpz1-Ccq1 sub-complex

5.2.1 Characterisation of a minimal Ccq1 binding domain of Tpz1

Binding of Ccq1 to the C-terminal region of Tpz1 has previously been observed by yeast two-hybrid analysis of the interaction [82]. In order to identify a minimal Tpz1 peptide required for Ccq1, binding was investigated by peptide array analysis. A cellulose filter array containing 96 20 residue long peptides that covered the C-terminus of Tpz1, each with a 1 residue shift, was produced by the peptide chemistry facility in house. 50 μL Ccq1 130-717 was added to the milk blocking step of the protocol at an initial concentration of 12 mg/mL and binding of the protein to the array was visualised with an anti-His antibody, as described in section 2.3.3.

Initially a large amount of non-specific binding of antibody to the peptides immobilised on the membrane was observed, so several different anti-His antibodies were tested. Additionally, 10 mM DTT was included in the binding step to improve specificity of binding of Ccq1 to the peptide array. DTT ensures that the sample remains reduced and prevents interactions occurring between cysteine-containing peptides within the array and was removed before the antibody was added, to avoid denaturation of the antibodies. A negative control with no added Ccq1 prior to antibody binding and two His tagged proteins, Ccq1 and yeast Xrs2 (Sarah Webb, Francis Crick Institute), as positive controls for

anti-His antibody activity which were dotted onto the peptide array membrane, before following the same downstream blotting protocol. Specific Ccq1 binding to the array between Tpz1 residues 437-460 was observed and repeats showed a reproducible binding pattern which led to design of binding peptides of Tpz1 (see Figure 51). The no protein negative control showed no specific antibody binding to peptides within the array.

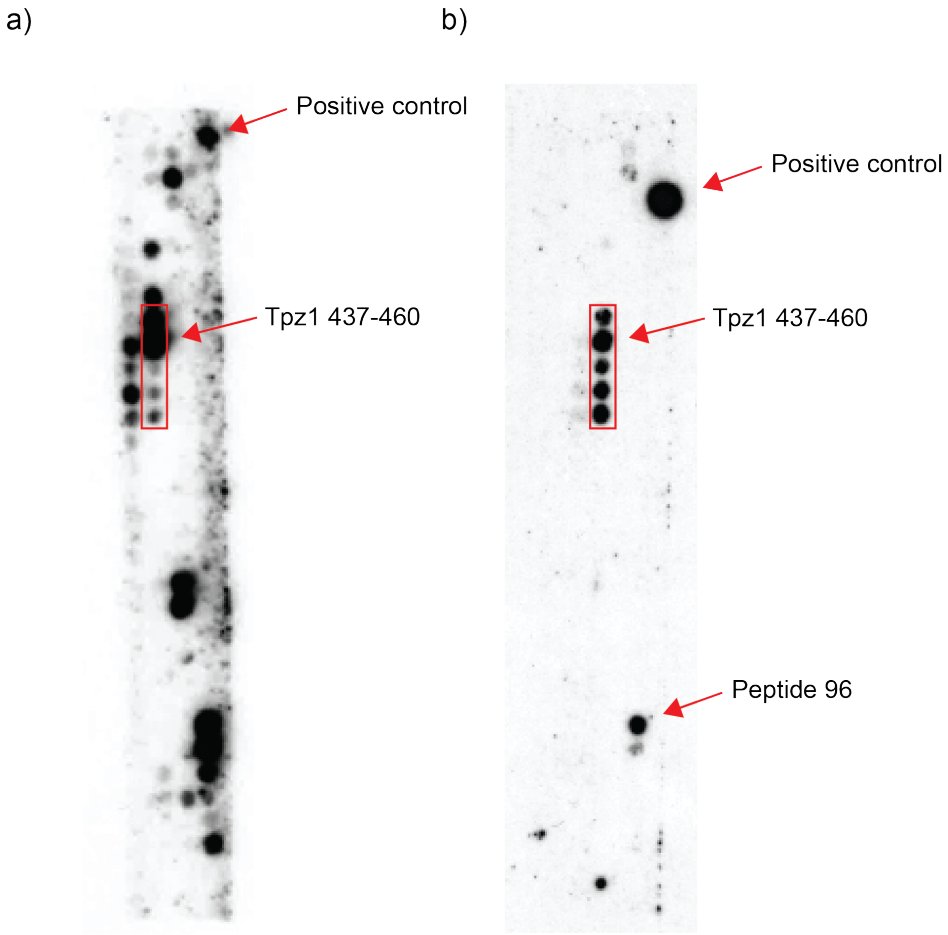


Figure 51: Peptide array analysis of the Ccq1-Tpz1 interaction.
Ccq1 130-717 binding to peptide arrays was visualised using an anti-His primary antibody and blots were developed using an ECL kit. A positive control of a different His-tagged protein was spotted on to the peptide array membrane prior to the experiment. a) Peptide array showing binding of His tagged Ccq1 130-717 to immobilised 20 residue Tpz1 C-terminal peptides. Binding of Ccq1 to a region of Tpz1 from residue 437 to 460 was observed, however there was also non-specific binding. b) Peptide array showing binding of His tagged Ccq1 130-717 to immobilised Tpz1 peptides after optimisation in the presence of 10 mM DTT. Adding DTT to the binding step reduced non-specific binding. Binding of Ccq1 to peptide # 96 still occurred.

On the basis of these results, peptides were designed for quantitative interaction studies. A minimum binding peptide was designed, containing only the 16 residues that were common to every peptide that showed an interaction with Ccq1 in the peptide array experiment (Tpz1-P16). A longer 24 residue peptide, containing all residues within any of the peptides able to bind to Ccq1, was synthesised (Tpz1-P24) along with peptide # 96, which appeared to bind as a singleton, and peptide # 31 that showed no binding as a negative control for (see Table 20).

Table 20: Tpz1 peptides designed using information from peptide array analysis.

Name	Sequence
Tpz1-P16 (residues 441-456)	FMKGLETTLQELYVEH
Tpz1-P24 (residues 437-460)	PDYDFMKGLETTLQELYVEHQSKK
Negative control (# 31) (residues 420-439)	SDICNEQIELEYKRKPIPDY
Peptide # 96 (residues 485-504)	GLPHGSFFELLRDWKKIEEF

The binding affinity between Ccq1 130-717 and these peptides was probed initially by isothermal titration calorimetry (ITC). Prior to all titrations the protein and peptide were dialysed into a buffer containing 50 mM Tris pH 7, 150 mM NaCl, 0.5 mM TCEP. Ccq1 protein was used as the analyte, with the Tpz1 peptide being the titrant in each case.

Interestingly, Tpz1-P16, which contained the core region common to all interacting peptides in the peptide array experiment, showed no detectable binding to Ccq1. Peptide # 96 did not bind to Ccq1, showing that the binding suggested by the filter array experiments were likely due to non-specific interaction of either the anti-His antibody or Ccq1. No binding was also observed

between peptide # 31 and Ccq1. In contrast, Ccq1 binding was seen for Tpz1-P24 with an affinity of approximately 4.3 μM (Figure 52). The interaction is enthalpically driven as shown by the negative ΔH value of -15.4 kcalmol^{-1}. However, the entropic component is unfavourable as -TΔS is positive (8.04 kcalmol^{-1}), which might be indicative of a conformational change or increased rigidity of the proteins upon ligand binding.

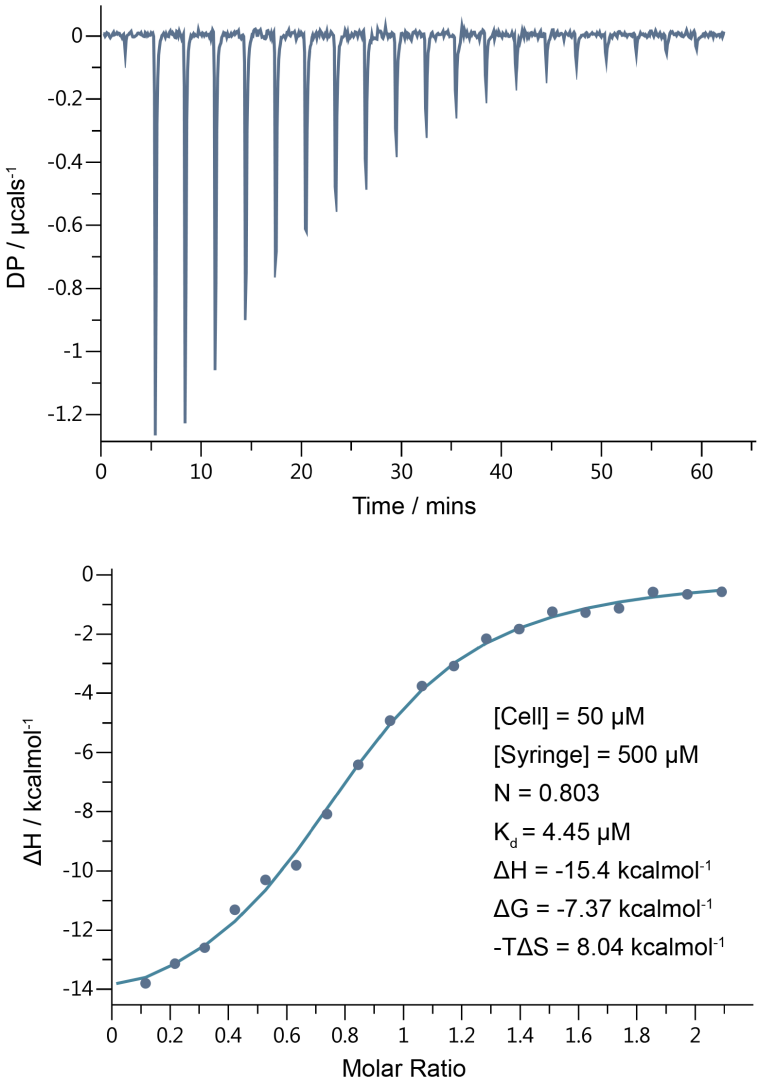


Figure 52: Representative ITC experiment showing binding between Ccq1 130-717 and Tpz1-P24.
An exothermic binding interaction occurs leading to negative peaks (top panel). Integration of these peaks allows plotting of a binding isotherm (bottom panel). A K_d of 4.45 μM was calculated for this experimental repeat.

Biolayer interferometry (BLI) was used to confirm the binding affinity calculated by ITC and to allow direct comparisons between this and the other interactions studied, for which sufficient protein for ITC could not be purified. Each biotinylated Tpz1 peptide was immobilised on streptavidin biosensors and binding to a concentration series of Ccq1 130-717 was tested. Once again, no binding was observed for the Tpz1-P16 peptide. Tpz1-P24 bound with low micromolar affinity with a K_d ranging from 2.8 μM to 6.5 μM affinity over several repeats, in keeping with the K_d value measured by ITC (see Figure 53). Again, no binding to the control peptide was observed.

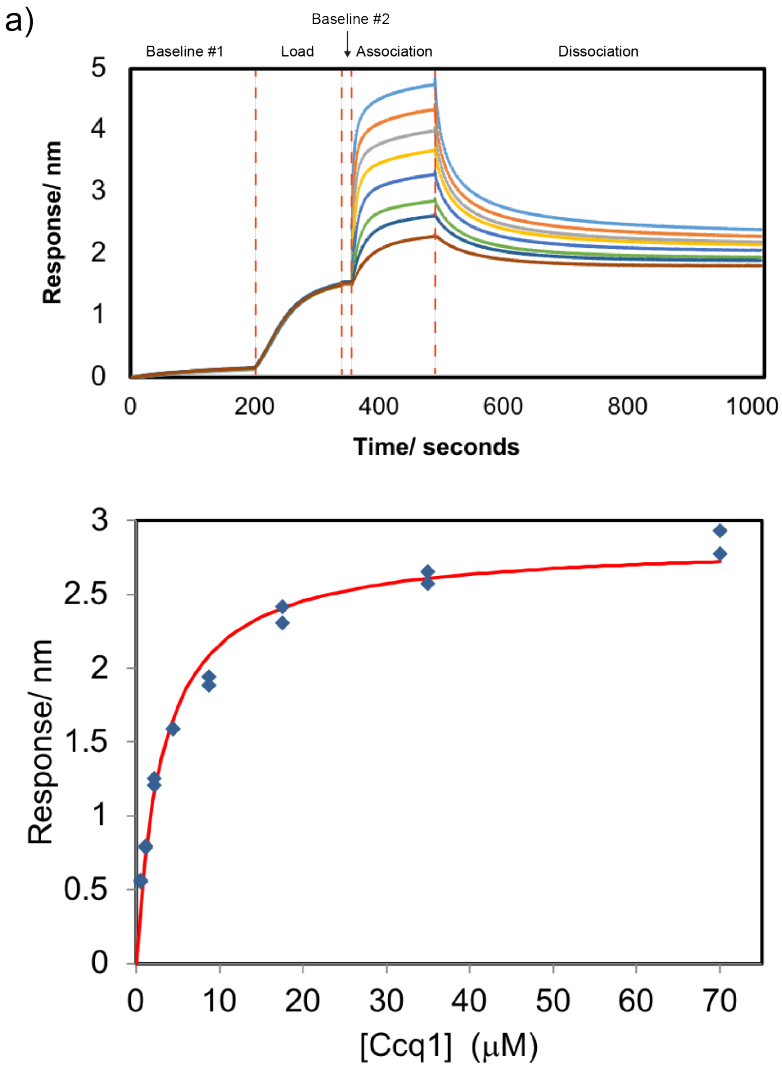


Figure 53: Binding affinity of a biotin-Tpz1 full peptide for Ccq1 130-717.
a) Sensorgrams showing the raw data collected for Tpz1 binding to a serial dilution series of an initial stock of Ccq1 at 70 μM. b) The Kd for the interaction determined by BLI was 3.2 ± 0.5 μM. This is calculated from the midpoint of the response curve. Biotinylated Tpz1 full peptide was immobilised on biosensors and binding to a serial dilution series of Ccq1 130-717 was measured.

The binding affinity between GST-Tpz1 405-508 and Ccq1 130-717 was measured to check if the measured binding affinity between the peptide and Ccq1 resembles the binding affinity of Ccq1 for longer sections of Tpz1. To circumvent degradation during purification (see section 3.4.4), GST-Tpz1 405-508 was crudely purified from *E. coli* using glutathione Sepharose affinity chromatography

and was then immediately immobilised on GST biosensors. The binding affinity of this construct for Ccq1 was then measured by BLI to be approximately 2 μM, which is consistent with the Tpz1-P24-Ccq1 interactions and suggests that this smaller peptide constitutes the full interacting region between Ccq1 and Tpz1 in the shelterin complex.

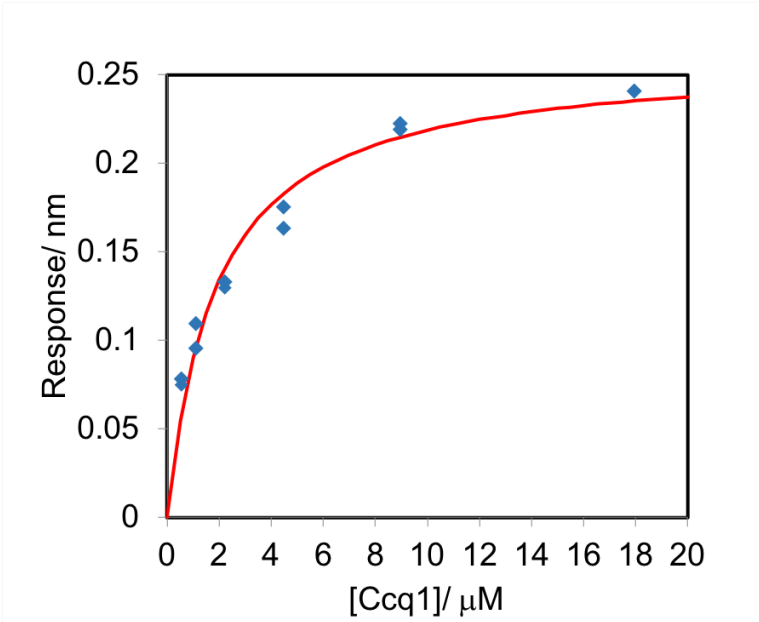


Figure 54: Binding affinity of Tpz1 405-508 for Ccq1 130-717.

The K_d for the interaction determined by this BLI experiment was 1.9 ± 0.5 μM. This is consistent with the affinity determined for the Tpz1 full peptide-Ccq1 130-717 interaction, showing that the peptide contains the residues important for wild type binding to occur.

5.3 Characterisation of Poz1 + Tpz1 binding

5.3.1 Tpz1 peptide design for Poz1 binding

The smallest peptide apparently able to support Tpz1 binding to Poz1, Tpz1 486-508, was identified by yeast-two hybrid and GST pull down experiments [82, 83, 87] (see Figure 55). Further sequence analysis and secondary structure prediction suggested the presence of an additional alpha helical region N-terminal to residue 486, starting at N475 and therefore a peptide spanning residues 475-508 (the C-terminal residue of Tpz1) was synthesised with and without a biotin tag for further binding analysis.

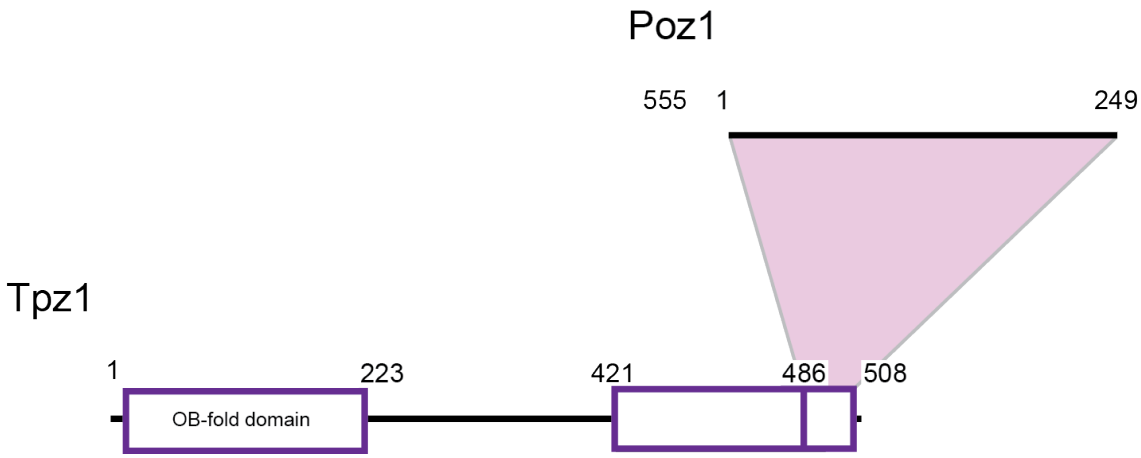


Figure 55: Schematic showing Poz1-Tpz1 interaction.
Previous studies have shown that full-length Poz1 interacts with a region of Tpz1 from residues 486-508.

5.3.2 Tpz1 475-508-Poz1 binding

Biolayer interferometry experiments were used to measure the affinity of the interaction between biotinylated Tpz1 475-508 and Poz1 25-249. Biotinylated

Tpz1 was attached to streptavidin sensors and binding to a serial dilution series of Poz1 was observed. This gave a dissociation constant (K_d) 27 +/- 7 nM, showing that a high affinity interaction occurs between Poz1 and Tpz1 (see Figure 56). Poz1 25-249 appeared unstable at room temperature and the signal measured by BLI reduced in proportion with time. This was potentially due to aggregation of the protein occurring at room temperature.

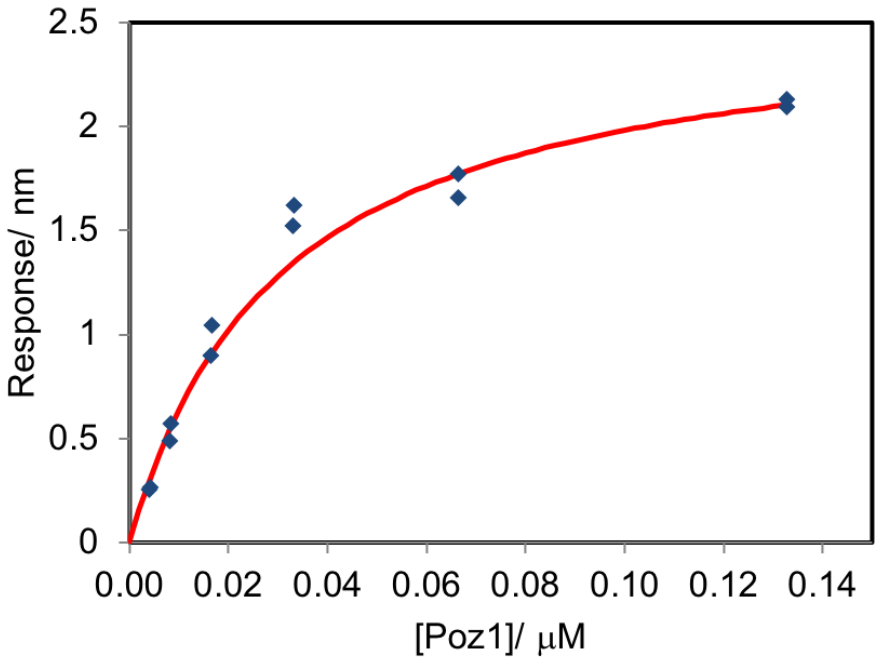


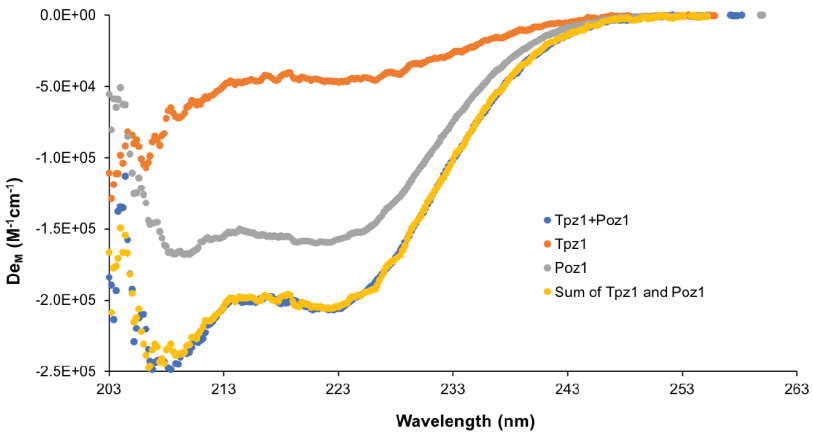
Figure 56: Binding affinity of Poz1 to a Tpz1 peptide.
Variation of response as a function of Poz1 concentration. K_d is determined by total response/2 and this analysis gave a K_d of 27 +/- 7 nM.

As formation of a Poz1-Tpz1 complex occurs, any effects on the secondary structure of Poz1 or Tpz1 upon addition of Tpz1 peptide was measured by CD spectroscopy (see Figure 57a). CD of Poz1 25-249 with and without the Tpz1 475-508 peptide showed that addition of the individual spectra for Poz1 and Tpz1 overlays with the spectra produced for the complex. This shows a very minimal change in the measured secondary structure of the protein was caused by

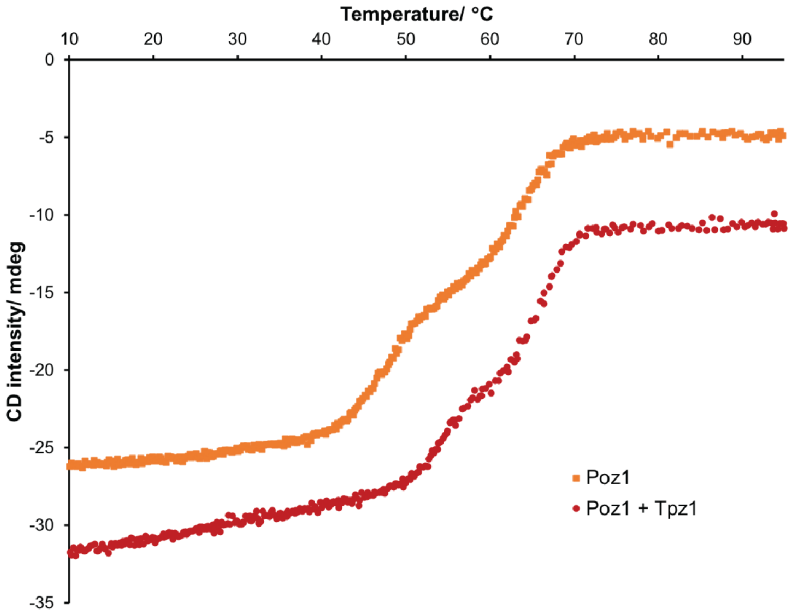
addition of the peptide and that Poz1 remained predominantly α-helical, in keeping with secondary structure predictions (see section 3.4.6).

To assess the effects of Tpz1 peptide binding on Poz1 stability, a thermal denaturation experiment was performed using CD. The melting curves for Poz1 and a complex of Tpz1 + Poz1 are both biphasic, meaning that two denaturation events occur in series (see Figure 57b). A shift in both T_m values was observed, with the first value increasing by 6.4 °C and the second by 3.2 °C. This shows that Tpz1 binding stabilises both transitions such that the T_m for both transitions has increased but the unfolding events remain uncoupled, even in the presence of the Tpz1 peptide. Each transition may relate to denaturation of a distinct domain and as Tpz1 binding stabilises both transitions, Tpz1 may bind between these two domains without causing them to denature simultaneously during CD.

a)



b)



c)

	T_m # 1	ΔT_m #1	T_m # 2	ΔT_m #2
Poz1	48.8	-	63.2	-
Poz1 + Tpz1	52.2	6.4	66.4	3.2

Figure 57: CD analysis of Poz1 and a Poz1-Tpz1 complex.
T_m values shown have units of °C. a) CD data showing that addition of spectra measured for Poz1 and a Tpz1 peptide measured separately overlaps with a spectrum measured for a complex of Poz1+Tpz1. b) Thermal denaturation curves show an increase in T_m caused by addition of the Tpz1 peptide to Poz1, meaning that the Tpz1 peptide stabilises Poz1. A biphasic denaturation curve is observed, showing that two separate denaturation events occur. The Poz1 + Tpz1 complex is formed by a 10 fold excess of peptide to protein. A change in the starting and finishing points of the CD intensity curves is due to the additive effect of the secondary structure of the Tpz1 peptide with Poz1 25-249. c) Table showing T_m values and the difference in T_m between +/- peptide samples. Addition of Tpz1 peptide to Poz1 leads to an increase in both T_m #1 and T_m #2, although has a larger effect on T_m #1.

5.3.3 Poz1 + Tpz1 co-expression

The high affinity interaction between Poz1 and the Tpz1 peptide and the improved solution properties of the complex suggests that this interaction may be constitutive *in vivo*. To test this hypothesis, co-expression of Poz1-Tpz1 was attempted in *E. coli* using standard conditions (described in section 2.2.2). Poz1 was expressed using a kanamycin resistant vector with a cleavable N-terminal 6x His tag and Tpz1 was expressed using an ampicillin resistant vector with a cleavable N-terminal GST tag. This led to formation of a complex that could be purified by GST affinity chromatography and size exclusion chromatography (see Figure 58). Tpz1 and Poz1 elute as a complex in the same fraction after size exclusion chromatography, in agreement with the high affinity interaction observed in section 5.3.2. Sample protein composition was verified by mass spectrometry.

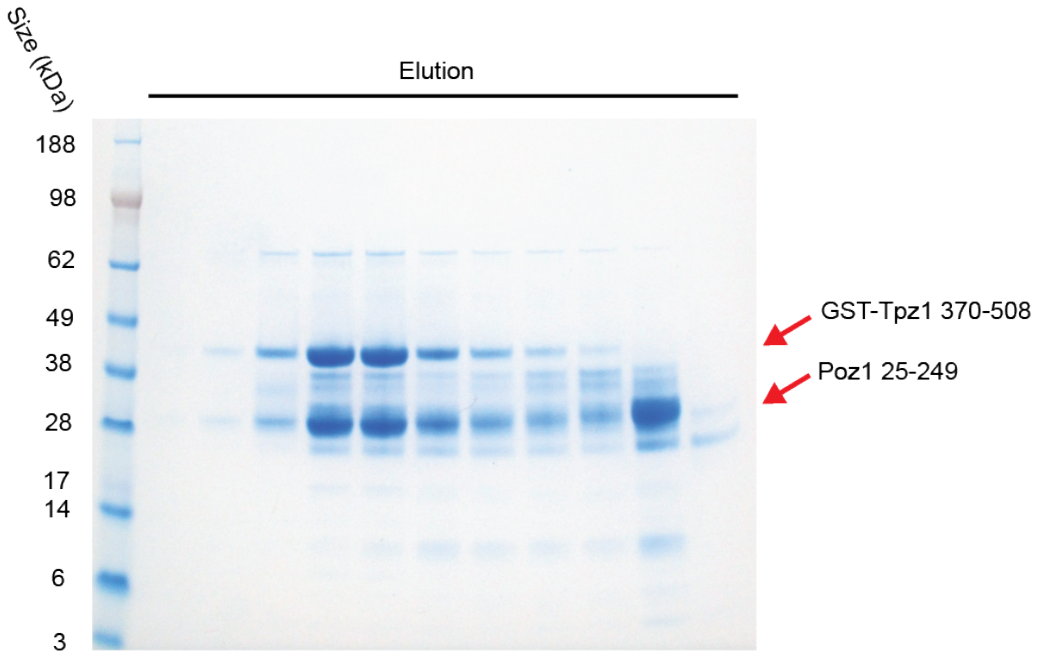


Figure 58: Size exclusion chromatography of a Poz1-Tpz1 complex.
Batch purification of GST-Tpz1 370-508 co-expressed with his-Poz1 25-249 used 3 mL glutathione Sepharose. Protein was eluted by addition of buffer containing 10 mM reduced glutathione. This sample was concentrated to 3 mL before being loaded to a 26/60 Superdex 200 size exclusion chromatography column. 3 mL elution fractions were collected and these are shown above.

Further purification of this complex, by addition of an ion exchange step to the protocol, was not possible as a reduction in buffer salt concentration below 300 mM NaCl led to protein precipitation. The protein complex was prone to aggregation and precipitation when concentrated in excess of 2 mg/mL. Further stabilisation of this complex may have been possible by production of a fusion protein containing both Poz1 and Tpz1 joined by a flexible linker sequence or by changing the constructs used during expression. Inclusion of a detergent in the purification buffers may have also increased solubility, although this would likely interfere with downstream experiments.

Affinity tags were removed by addition of 3C protease, primarily because the GST affinity tag typically causes dimerisation of proteins to occur [229]. SEC-MALLS of a Poz1-Tpz1 complex was attempted (see Figure 59). Difficulties with buffer mismatch due to the limited solubility of the GST-free complex, along with degradation of Tpz1 370-508, prevented an accurate evaluation of complex molecular weight. However, the mass reported for the center of the main elution peak was approximately 75 kDa, suggesting that two Poz1 monomers (~27 kDa) bind to one Tpz1 (~16 kDa) monomer, leading to production of a heterotrimeric species. This is consistent with the relative Coomassie staining of each protein after SDS-PAGE analysis, as both proteins contain a similar number of basic residues, but Poz1 stains more intensely.

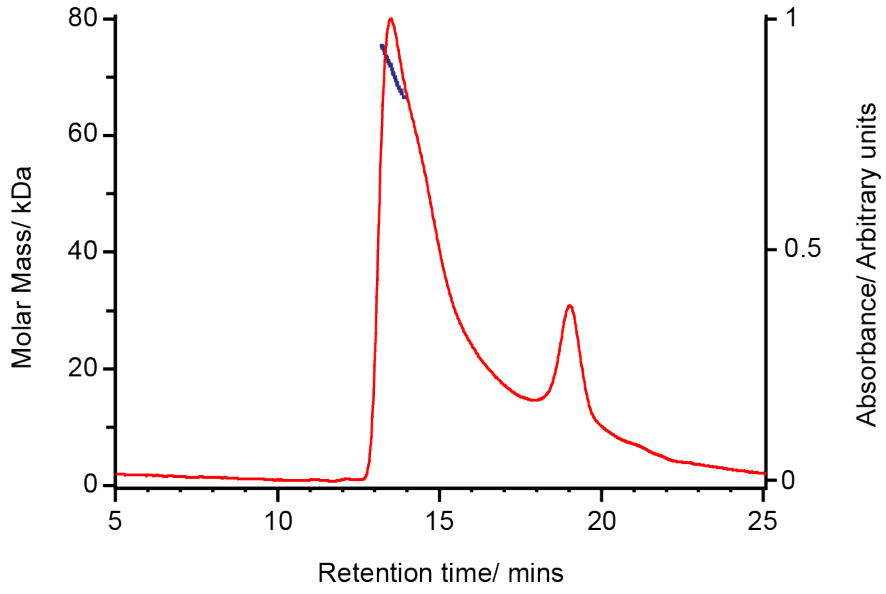


Figure 59: SEC-MALLS of Poz1-Tpz1 complex.
SEC-MALLS was run with an initial protein concentration of 2 mg/mL. A range of molar masses from 65-78 kDa were calculated for species within the sample. This SEC-MALLS gives inconclusive results due to a possible mismatch in the buffers used, which gives a large peak for refractive index after sample elution. This prevents optimal baseline subtraction and gives rise to a variation in observed protein mass across the peak. Repeats were not possible due to lack of sample availability, resulting from sample stability.

The solubility issues described above hampered further experimentation. As described, the best expressing GST-Tpz1 370-508 protein was found to rapidly undergo degradation during purification even when co-expressed with Poz1. Degradation occurred particularly readily after affinity tag cleavage (see Figure 60), which prevented addition of more steps to the purification scheme. Degradation of Tpz1 may be caused by the lack of expression of some binding partners, such as Ccq1, leading to increased availability of Tpz1 for degradation by bacterial proteases. It is possible that this degradation could have been reduced by including protease inhibitors in the buffers used during purification steps, not just the lysis buffer.

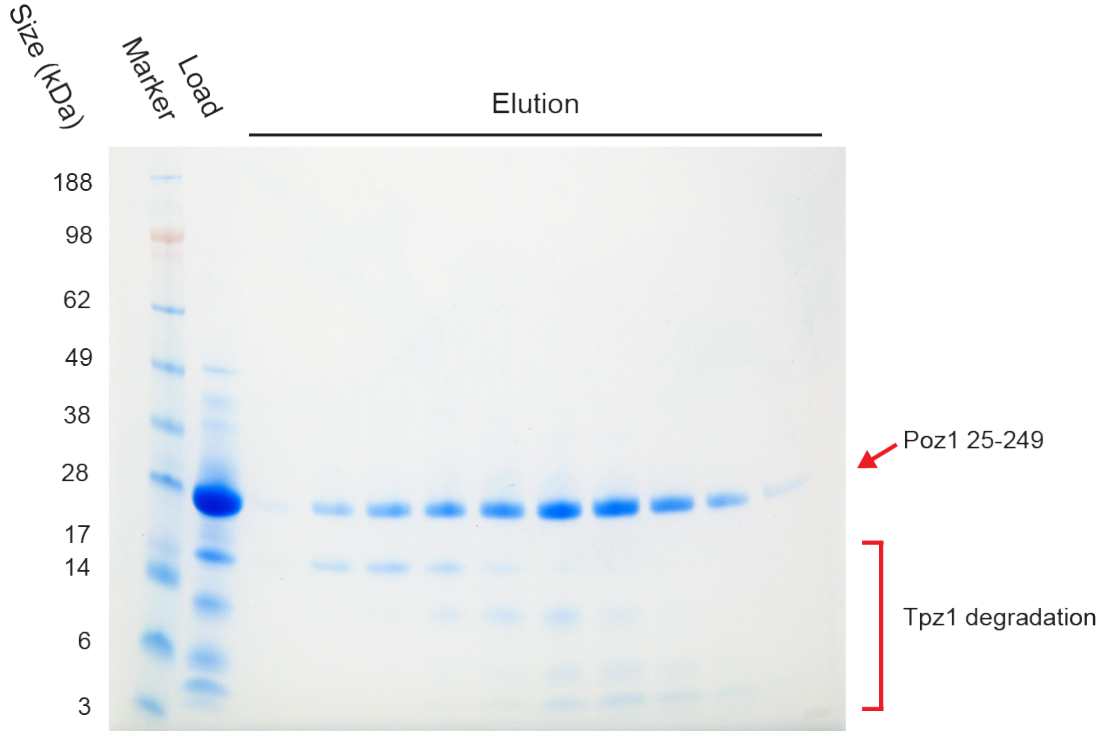


Figure 60: Poz1-Tpz1 degradation during purification.
Size exclusion chromatography elution fractions showing elution of Poz1 with several different sized degradation products of Tpz1 after cleavage of affinity tags using 3C protease. Presence of Poz1 and Tpz1 in the observed bands was verified by mass spectrometry.

5.4 A tripartite complex of Poz1 + Tpz1 + Ccq1

5.4.1 Binding affinity determination of Ccq1 for a Poz1-Tpz1 complex

It was hypothesized that by increasing the number of shelterin components present, degradation and stability issues might be resolved. Firstly, the affinity of the interaction between Ccq1 and the complex containing Poz1-Tpz1 was probed by biolayer interferometry to investigate whether the presence of Poz1 had any effect on the affinity of the interaction between Tpz1 and Ccq1. Ccq1 130-717 was purified separately and the His tag removed by 3C cleavage. Uncleaved Ccq1 was removed from the sample by addition of nickel sepharose resin to the purified cleaved protein and subsequent removal of the resin by gravity flow column purification. The Poz1-Tpz1 complex was bound either through the His-tag of Poz1 to Ni-NTA biosensors or to GST biosensors through the affinity tag fused to Tpz1 (Figure 61). The experiment was repeated with a sample that did not contain a GST tag attached to Tpz1 370-508 and this showed that dimerisation of GST was not affecting the measured binding affinity.

When Poz1-Tpz1 was immobilised on Ni-NTA sensors without removal of the GST tag, an affinity of 6 μM was measured and after GST tag cleavage an affinity of 3.55 μM was measured. Immobilisation on GST sensors gave an affinity of 5.6 μM, showing that immobilisation of the Poz1-Tpz1 complex by each component did not cause a significant change to the affinity measured. These

values were consistent with the affinities determined for the Tpz1-P24-Ccq1 and GST-Tpz1$_{370-508}$-Ccq1 interactions (section 5.2.1), showing that the presence of Poz1 bound to Tpz1 has no direct or allosteric effect on Ccq1 binding and that an interaction between Poz1 and Ccq1 is unlikely to occur, despite their close proximity to one another when bound to the C-terminus of Tpz1. This was confirmed by BLI, which showed no detectable interaction between Poz1 and Ccq1 in the absence of Tpz1. Experiments were performed directly after initial purification, to reduce the time for potential protein degradation to occur.

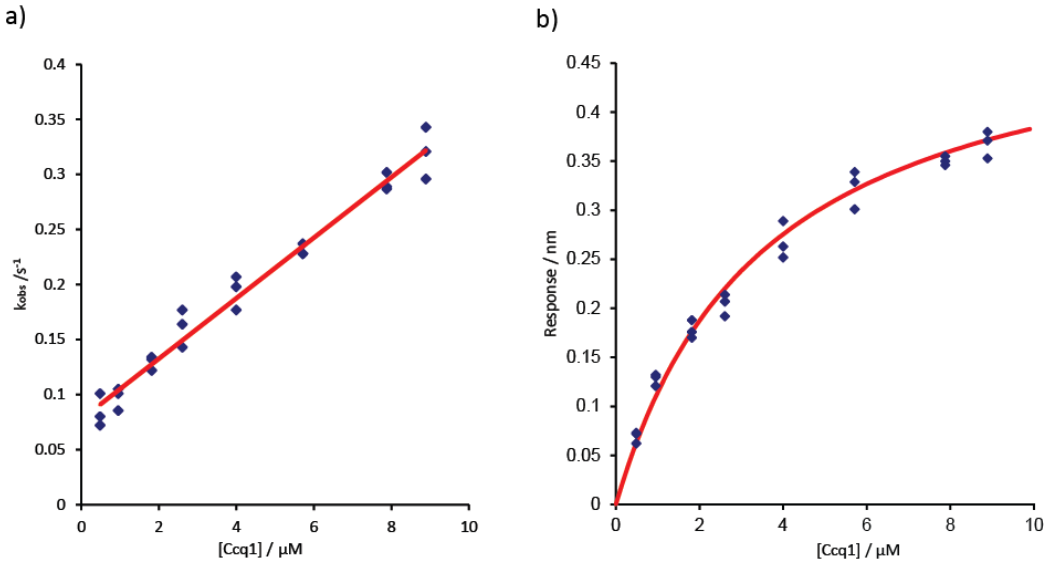


Figure 61: Binding affinity of a Poz1-Tpz1 complex to Ccq1.
The analysis shown is from an experiment where the GST tag had been removed and the complex was immobilised by an interaction between His-Poz1 and Ni-NTA biosensors. a) Observed rate (k_{obs}) *versus* Ccq1 concentration. This graph gives the association and dissociation rate constants (k_{on} and k_{off}) as the slope and intercept respectively. The equilibrium dissociation constant is determined as k_{off}/k_{on}. K_d = 2.84 +/- 0.26 µM. b) Variation of response as a function of Ccq1 concentration. K_d is determined by total response/2 and this analysis gave K_d = 3.55 +/- 0.45 µM.

5.4.2 Construct optimisation and tripartite complex formation

Several methods were employed to attempt to improve protein solubility and yields. Firstly, crudely purified proteins were mixed and further purified. Secondly, His-Poz1 1-249 + GST-Tpz1 370-508 were coexpressed and purified to generate a complex, which was then mixed with purified Ccq1 130-717. Finally, Ccq1 expressing cells were co-lysed with Poz1-Tpz1 co-expressing cells. Initially, degradation issues were still observed for all protocols and incomplete cleavage of the affinity tags during purification occurred.

Optimisation of protein constructs led to improvements in the purification scheme. Poz1 was re-cloned into pET28b (N-terminal thrombin-cleavable 6 x His tag). Tpz1 tag cleavage required addition of 3C protease, so cloning of Poz1 into pET28b allowed for stepwise removal of affinity tags from Poz1 and Tpz1 by different proteases during the purification protocol. His tagged Ccq1 was used to pull down an untagged complex of Poz1-Tpz1 purified by stepwise tag removal and the resulting 3-component complex was then further purified by size exclusion chromatography (see Figure 62). This produced a fully cleaved sample which could be used for preliminary negative stain electron microscopy studies (section 5.4.3). The fractions marked by Box 1 show purification of the full complex and Box 2 shows the presence of a second peak in protein concentration due to dissociation of Poz1-Tpz1 from Ccq1, as this peak contains an excess of the Poz1-Tpz1 complex. This dissociation agrees with BLI data presented in sections 5.3.2 and 5.4.1 that shows that the Poz1-Tpz1 interaction affinity is in

the low nanomolar range, while the interaction of this bipartite complex with Ccq1 is in the low micromolar range.

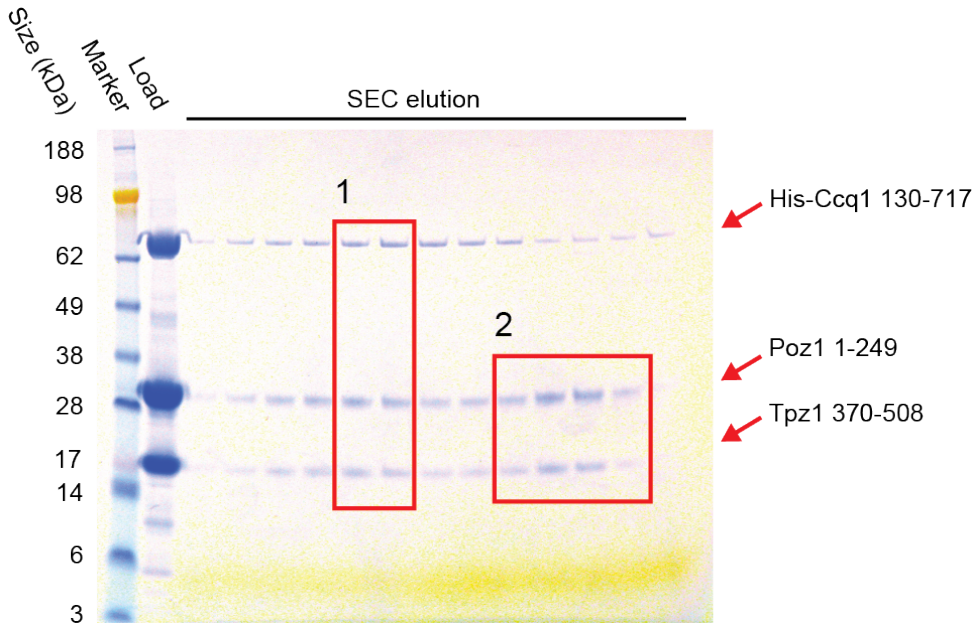


Figure 62: Size exclusion chromatography of a Poz1-Tpz1-Ccq1 complex. Poz1-Tpz1 was purified and affinity tags were removed in a stepwise process by addition of different proteases. Ni-NTA bound His Ccq1 130-718 was used to pull down the resulting untagged Poz1-Tpz1 complex and the tripartite complex was further purified by size exclusion chromatography. The full tripartite complex was purified in region 1 and region 2 shows dissociation of Poz1-Tpz1 from Ccq1 leads to a second peak in gel filtration that contains an excess of Poz1-Tpz1.

5.4.3 Negative stain electron microscopy analysis

A complex containing $Poz1_{1-249}$-$Tpz1_{370-508}$-$Ccq1_{130-717}$ was purified previously from *E. coli* cells by several methods (see Figure 62). SEC-MALLS and SAXS of Ccq1 alone showed that it formed a homodimeric complex and SEC-MALLS of a Poz1-Tpz1 complex had shown that it was likely to contain more than one copy of each protein. This suggested that the tripartite complex containing Poz1-Tpz1-Ccq1 would be sufficiently large to begin structural characterisation by negative

stain electron microscopy. Due to low protein recovery after purification and time constraints, only analysis by negative stain electron microscopy was possible.

The complex was applied to continuous carbon coated copper grids and stained with 1 % uranyl acetate (section 2.4.5.2) to allow initial electron microscopy analysis. The resulting micrographs suggested substantial complex dissociation during the grid preparation and staining process evidenced by their “speckled” appearance and the absence of any large particles. The protein concentration required for EM could not be measured by UV spectrophotometry.

5.4.4 Sample stabilisation by GraFix

To counter the observed dissociation, samples were subjected to crosslinking using the GraFix method of Kastner *et al* [198] (see section 2.4.4). This technique uses gentle glutaraldehyde fixation coupled to ultracentrifugation through a sucrose gradient. This can lead to stabilisation of proteins within a complex and allows separation of different complexes from one another.

Sucrose gradient optimisation produced a single species with the minimal number of crosslinks required for complex stabilisation. Initially three sucrose gradients, 5-20%, 5-25% and 10-25%, were tested, including a 0-0.1 % gradient of the crosslinking agent glutaraldehyde. Protein purification for samples subjected to GraFix required a switch to HEPES buffer, as Tris buffer also interacts with glutaraldehyde due to its amine groups, which inhibit protein cross-

linking. 100 μL of protein at 0.25 mg/mL was applied to each gradient and samples were centrifuged at 40000 rpm for 18 hours. The gradients were then fractionated into 100 μL aliquots and analysed by SDS-PAGE with silver staining to visualise the complexes produced.

A 5-20% gradient gave the best separation between different species and therefore was used for further sample preparation. During sample preparation a higher initial concentration of protein (~1 mg/mL) was used, so coomassie stained SDS-PAGE was used for post GraFix analysis of the resultant fractions. Fractions corresponding to the largest non-aggregated complex were pooled and concentrated (see Figure 63) and used for preliminary single particle negative stain electron microscopy experiments. The largest species present at the bottom of the gradient was not used as it likely contains aggregates and heterogeneously crosslinked species. Unfortunately, micrographs still showed a heterogeneous mixture of differently sized particles, presumably representing differently crosslinked species or sub-complexes. In addition, the contrast between particles and background staining was also not consistent across the EM grid. Due to these problems, only a small number images were taken.

Repetition failed to improve EM results and so the decision was taken to switch to insect cell expression in an attempt to generate larger, more stable complexes with greater numbers of shelterin components. These experiments are discussed in section 6.3.

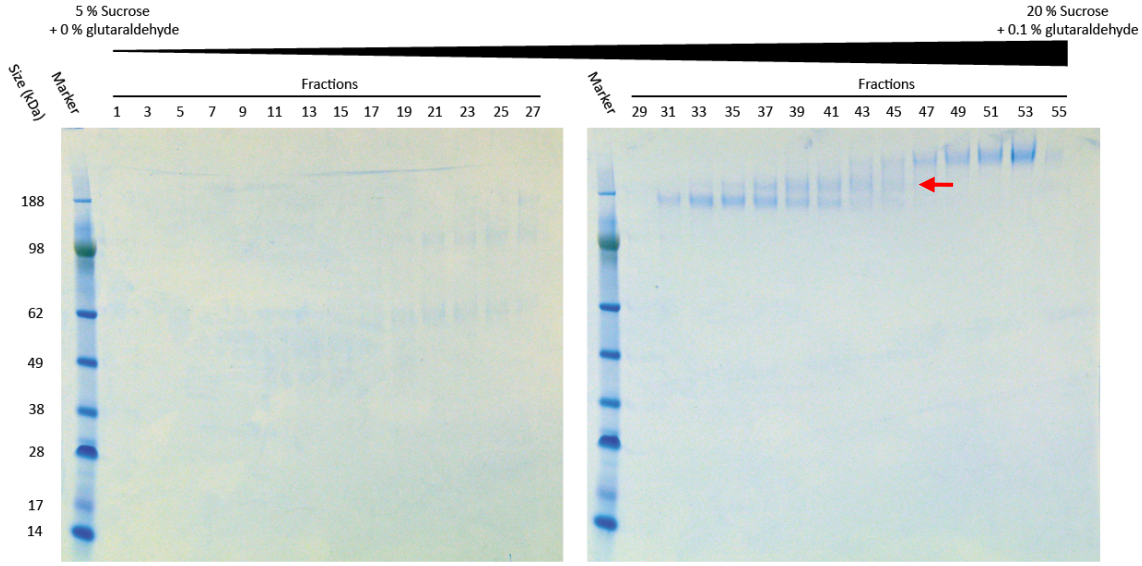


Figure 63: GraFix of a Poz1-Tpz1-Ccq1 heterotrimer leads to separation of different species.
GraFix was performed using a sucrose gradient from 5-20% centrifuged at 40000 rpm for 18 hours as this gave the largest separation of different species after fractionation. The red arrow shows the band chosen for EM analysis. The largest species on the gel was found at the bottom of the gradient so could contain a mixture of different proteins and so was not used. In this case, fractions 39-45 were pooled and concentrated for preliminary single particle negative stain electron microscopy experiments. Sub-optimally crosslinked species are seen on the first gel between fractions 17-27, due to the lower glutaraldehyde concentration in this region of the gradient.

5.5 Discussion

5.5.1 Summary

During this project, Ccq1 and Poz1 were shown to bind independently to a C-terminal region of Tpz1 (see Figure 64). Binding analysis of a Tpz1 peptide and a Poz1-Tpz1 complex to Ccq1 showed similar affinities in the low micromolar range. The binding affinity of Poz1 for Tpz1 was also measured and was found to be in the low nanomolar range. Co-expression of Poz1 with Tpz1 improved protein recovery during purification but degradation of Tpz1 occurred rapidly. Co-lysis of Ccq1 with Poz1-Tpz1 co-expressing cells led to further improvements in protein solubility. Initial analysis by negative stain electron microscopy of *E. coli* expressed $Poz1_{1-249}$-$Tpz1_{370-508}$-$Ccq1_{130-717}$ showed that complex stabilisation was required prior to visualisation, but glutaraldehyde fixation failed to significantly improve particle homogeneity and image contrast.

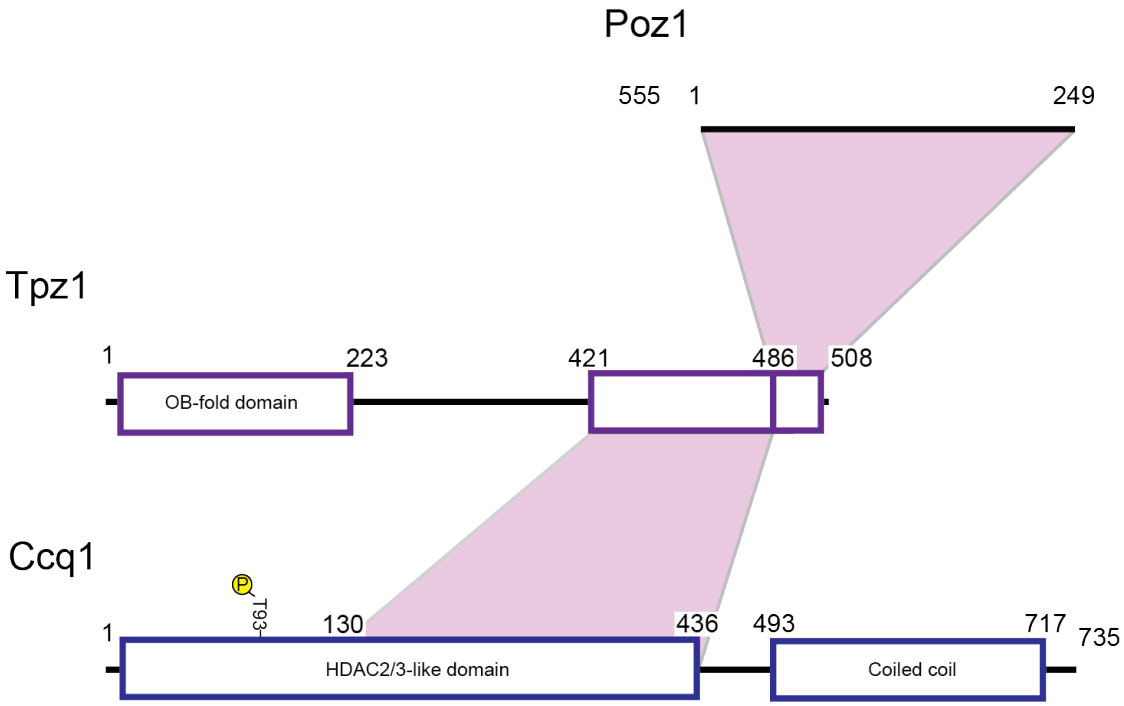


Figure 64: Schematic showing the Poz1-Tpz1-Ccq1 interaction.
Both Poz1 and Ccq1 have been shown to interact independently with a region at the C-terminus of Tpz1. Full-length Poz1 has been shown to interact with Tpz1 residues 486-508 and a central region of Ccq1 from residues 130-436 interacts with Tpz1 421-486.

5.5.2 The Poz1-Tpz1 interaction

When expressed and purified separately, neither Tpz1 or Poz1 were very stable and only small amounts of protein could be recovered for biophysical analysis. A CD thermal denaturation experiment showed an increase in the stability of Poz1 when a Tpz1 peptide was added. Binding of the Tpz1 peptide stabilises both Poz1 unfolding transitions, consistent with Tpz1 binding at the interface between two folded domains seen in the crystal structure of the complex [91]. The binding affinity between Poz1 and Tpz1 was measured by biolayer interferometry and showed a tight affinity in the low nanomolar range. These data suggest that Poz1 and Tpz1 may form a constitutive complex in cells.

Poz1 and Tpz1 are at the centre of the shelterin complex and form a bridge between the single-stranded and double-stranded DNA binding proteins, Pot1 and Taz1 respectively. A full link between Pot1 and Taz1 has been shown to be important for repression of telomerase activity and maintenance of the telomere in a non-extendible state [87]. Mutations that disrupt the interaction between Poz1 and Tpz1 lead to uncontrolled telomere extension in cells, showing this interaction is required for normal telomere length regulation and that it plays a negative regulatory role [230]. It has also been shown that it is the link made by Poz1 and Tpz1 between single-stranded and double-stranded DNA and not the proteins themselves that are important, as replacement of these proteins with short linkers still leads to maintenance of telomeres at wild type lengths in cells. However, other end protection functions, such as protection against end fusions, in cells containing these 'minishelterins' are compromised [93]. The presence of a fully functioning link in the shelterin complex is important for its function and formation of a constitutive Poz1-Tpz1 interaction, as suggested by the data presented here, would lead to correct regulation of telomere length. In somatic cells, telomerase is not active and formation of a constitutive complex between Poz1 and Tpz1 may have an important role in this repression.

5.5.3 The Ccq1-Tpz1 interaction affinity is weaker than that of Poz1-Tpz1

Binding of Ccq1 to Tpz1 occurs with a lower affinity, in the low micromolar range, compared to the high affinity Poz1-Tpz1 interaction. The presence of Poz1, also bound to a C-terminal region of Tpz1 close to the Ccq1 binding site, has no effect

on Ccq1-Tpz1 binding affinity. This suggests that although both Ccq1 and Poz1 interact with the C-terminus of Tpz1, no direct binding between Ccq1 and Poz1 occurs.

Binding of phosphorylated Ccq1 to Tpz1 leads to a switch in the state of the telomere from a non-extendible to an extendible state. The interaction of Ccq1 with Tpz1 is required to remove the repression of telomerase facilitated by the linkages formed between single-stranded and double-stranded DNA by the other shelterin complex components [87]. This then allows extension of telomeres by telomerase to commence. This process is highly regulated and can only occur for a specific length of time, as otherwise overly extended telomeres would be produced. The affinity of the interaction between a peptide and an extended length of Tpz1 with Ccq1 was probed by ITC and BLI, with both techniques giving an affinity in the low micromolar range. As the measured affinity for the short peptide and a longer section of Tpz1 for Ccq1 were consistent, this shows that the identified peptide was likely to contain all of the residues important interaction within the shelterin complex. The reduced affinity of this interaction might indicate a weaker, possibly transient, interaction between Ccq1 and Tpz1 that could be required to prevent prolonged activation of the extendible state at telomeres and resulting production of long telomeres that are unstable and prone to fusion, leading to genomic instability [231].

Due to the critical requirement for prevention of over-extension of telomeres, there are likely to be several mechanisms acting to prevent production of long

telomeres and a weak, potentially transient, interaction between Ccq1 and Tpz1 could act as a backstop, preventing over-activation of telomeres into an extendible state and over-extension if other regulatory mechanisms fail.

Chapter 6. Electron microscopy analysis of shelterin sub-complexes expressed in insect cells

6.1 Introduction

The co-expression and co-lysis experiments described in Chapter 5 led to improvements in yield and stability of shelterin complex components when expressed in *E. coli*. This may be indicative of constitutive interactions between at least some of the shelterin proteins and would explain many of the difficulties observed when expressing each protein in isolation. *E. coli* co-expression and co-lysis experiments of truncated Ccq1-Tpz1-Poz1 fragments produced a complex large enough to be studied in preliminary electron microscopy experiments, but significant sample heterogeneity was observed.

Given the improvements in protein stability observed when the number of shelterin components present was increased, the decision was made to pursue co-expression of full-length components in insect cells using the MultiBac system with a view to electron microscopy (EM) studies of the entire fission yeast shelterin complex.

6.2 Complex expression in insect cells

6.2.1 MultiBac cloning

DNA sequences encoding the shelterin proteins were cloned into a modified pFUBB transfer vector (McDonald Lab, Francis Crick Institute) in pairs based on sub-complex interactions described in the preceding chapters. Rap1/ Taz1, Poz1/ Tpz1 and Ccq1/ Pot1 were cloned into three separate vectors (see Figure 65). The preparation of three vectors allowed testing of different combinations of shelterin components to identify larger stable sub-complexes in addition to the full assembly. Along with the genes of interest, two separate terminator sequences and a dual promoter were also cloned into the vectors in between two Tn7 sites, which were required for site specific transposition of the target DNA [232]. Specific overhangs were added to the vector and 5 insert fragments for each reaction to ensure that incorporation by the Gibson assembly could only proceed if all inserts were present in the correct orientation and order. Each gene within a pair of proteins of interest were arranged on opposite DNA strands, as translation from the dual promoter occurs in both directions.

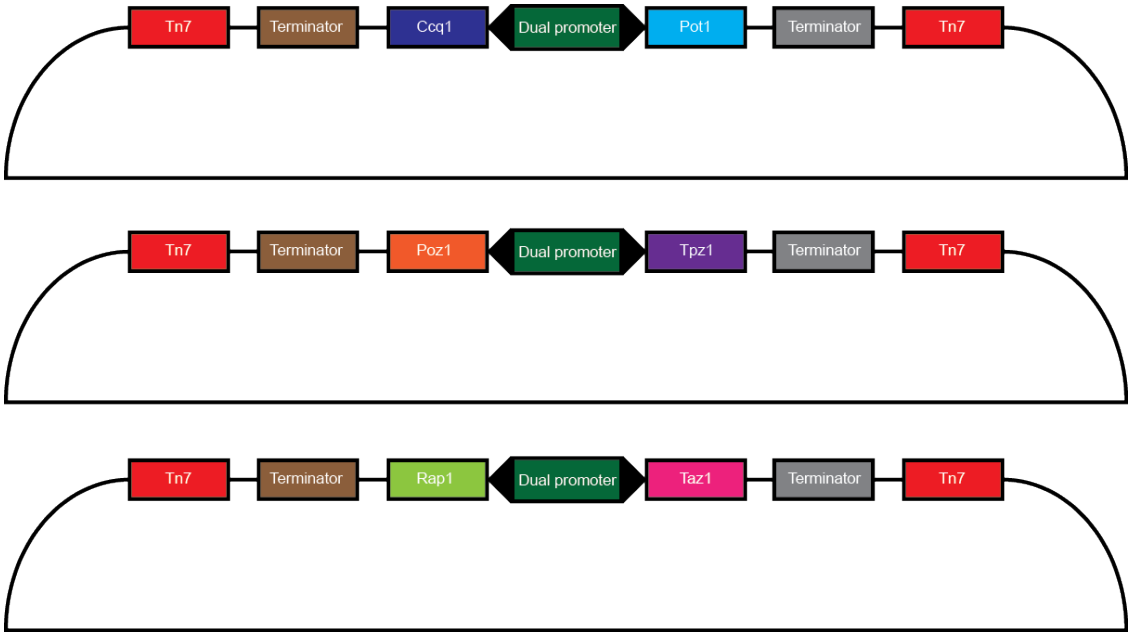


Figure 65: MultiBac cloning scheme.
The pFUBB vector backbone, two separate terminator sequences, dual promoter and genes of interest were amplified by PCR with specific overhangs, which would only allow for incorporation by Gibson assembly in the order and orientation shown. Each fragment was a different length and so different concentrations of each PCR product were required for successful plasmid assembly to occur. Several repeats of the reaction were required to optimise the ratio of each PCR fragment necessary.

The pET47b clones for *E. coli* expression of both full-length Ccq1 and Poz1 were used as the template for PCR amplification and primers were designed to include the N-terminal 3C cleavable 6 x His tag with each protein in the fragment amplified. The other proteins were cloned from codon-optimised synthetic DNA sequences (GeneArt) without affinity tags. Successful cloning required the use of the high fidelity polymerase PfuUltraHF (Agilent) due to the large size of the plasmids produced.

6.2.2 MultiBac protein expression

The three recombinant pFUBB vectors were transformed into separate aliquots of electrocompetent MultiBacDH10 cells, to generate bacmids by recombination using the Tn7 sites within the pFUBB vector. Tn7 transposition leads to disruption of the *lacZ* gene, allowing blue-white selection to identify successful incorporation into the bacmid (see section 2.1.2 for details). DNA was then isolated by a modified mini-prep protocol to prevent shearing and nicking of the large recombinant bacmids. Each bacmid was transfected into a separate culture of SF21 insect cells and this was followed by three subsequent viral amplification steps (see section 2.2.3).

The third amplification step produced 250 mL of cell culture expressing the pair of proteins found within each bacmid, known as the P4 amplification. These cells were pelleted by centrifugation, leaving the virus-containing supernatant, which was used for future infections. The three cell pellets containing either Ccq1-Pot1, Poz1-Tpz1 or Rap1-Taz1 were co-lysed in 50 mM Tris pH 8, 500 mM NaCl, 0.5 mM TCEP, 25 mM imidazole pH 8, before initial purification using a 1 mL HisTrap HP column. Both Poz1 and Ccq1 constructs contained a His tag and so were expected to bind to nickel affinity resin. Shelterin complex proteins that bind to these proteins should also co-purify. Proteins were eluted from the nickel affinity column using an increasing linear gradient of imidazole from 25 mM to 500 mM. This gave a very heterogenous product with no obvious bands relating to overexpressed proteins when analysed by SDS-PAGE. As an alternative

strategy, each pair of proteins were expressed on a 1 L scale using the P4 viral stocks for initial insect cell infections. The cell pellets were then mixed together, co-lysed and purified by batch binding to 3 mL of Ni-sepharose resin, washing and elution with lysis buffer supplemented with 500 mM imidazole pH 8. Protein was eluted in 5 mL fractions until all bound protein had been removed from the resin, as monitored by SDS-PAGE. The elution fractions were pooled and concentrated to 100 μL, before being loaded to a 10/300 Superose6 column for final purification and buffer exchange. Although the gel filtration profile showed a number of overlapping species, fractions at the peak UV intensity nonetheless contained shelterin complex components Ccq1, Poz1, Tpz1, Rap1 and Taz1, as confirmed by Band ID mass spectrometry. Although the coverage of each protein was relatively low (20-30 %), peptides were observed close to the N- and C-termini of all 5 proteins, showing that full-length protein for each component was likely to be present. Expression of Pot1 was not observed (see Figure 66).

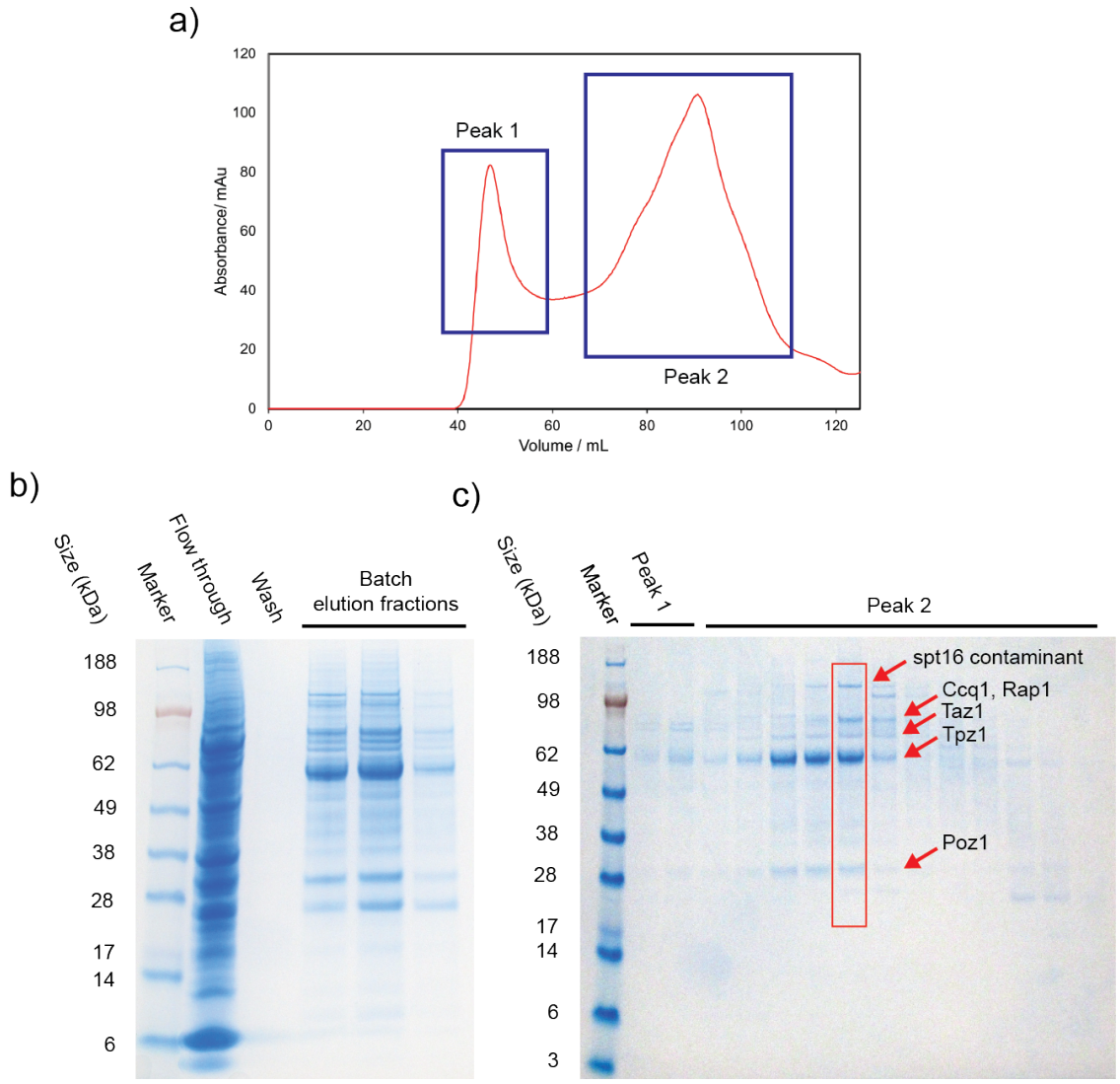


Figure 66: Purification of MultiBac expressed full shelterin complex.
Ccq1 (83kDa), Rap1 (80 kDa), Taz1 (75 kDa), Tpz1 (58 kDa) and Poz1 (30 kDa). a) Representative size exclusion chromatogram for the full shelterin complex, showing a void peak (peak 1) followed by a peak containing multiple different species (Peak 2). The maximum point of Peak 2 corresponds to the 5 component shelterin complex shown on gels below. b) Nickel affinity chromatography of the full shelterin complex. Ccq1 and Poz1 both contain 6 x His tags and therefore should bind to the column and co-elute with other interacting proteins. c) Size exclusion chromatography of the nickel eluate. 10/300 Superose6 size exclusion chromatography led to separation of a putative complex containing shelterin components from contaminants found in the nickel affinity elution. The red box shows the fraction containing the highest concentration of proteins giving bands of sizes expected for shelterin complex components.

As several contaminating protein species were also present, improvements to sample purity were sought by addition of a second nickel affinity chromatography step, after the initial rounds of nickel and size exclusion chromatography. This second affinity purification used a 1 mL HisTrap HP column with elution by a linear gradient of imidazole from 25 mM to 500 mM. A further Superose6 size exclusion chromatography step was then performed. Repeating this protocol with increased culture volumes improved protein recovery and overall purity (see Figure 67). This sample was then used for biophysical analysis and single particle negative stain electron microscopy.

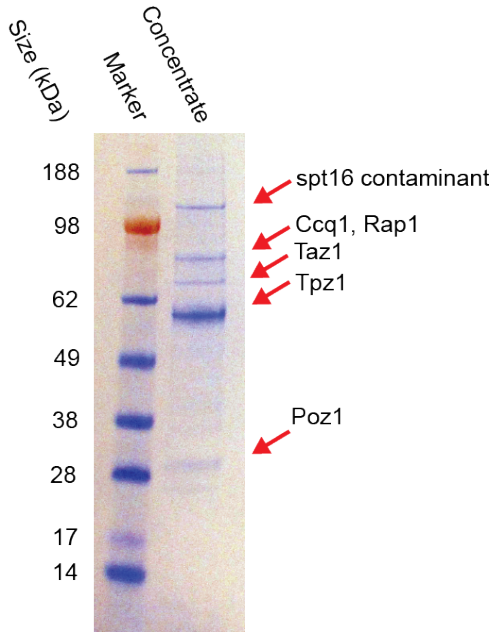


Figure 67: Full shelterin complex purified sample.
Band ID mass spectrometry identified several protein bands contain shelterin complex components. Bands from the gel were excised and protein content was confirmed by mass spectrometry. Several contaminating proteins were also present and the high molecular weight contaminant that is larger than 100 kDa was identified by mass spectrometry as spt16, a component of the FACT complex. Ccq1, Rap1, Taz1, Tpz1 and Poz1 were detected. Pot1 expression and purification was not detected.

6.3 Structural characterisation

6.3.1 Complex characterisation

SEC-MALLS analysis was used in an attempt to define the total mass and stoichiometry of the 5-component complex. The UV absorbance curve resembled the curve observed during the final size exclusion step of the purification protocol and displayed several overlapping species (see Figure 68). The overlap between species affected the mass determined for each species as the calculated mass was an average of the mass of the different species present, making any estimate of complex stoichiometry inaccurate. The multiple species present may represent different complexes that contain different oligomeric states or complexes missing one or more of the proteins.

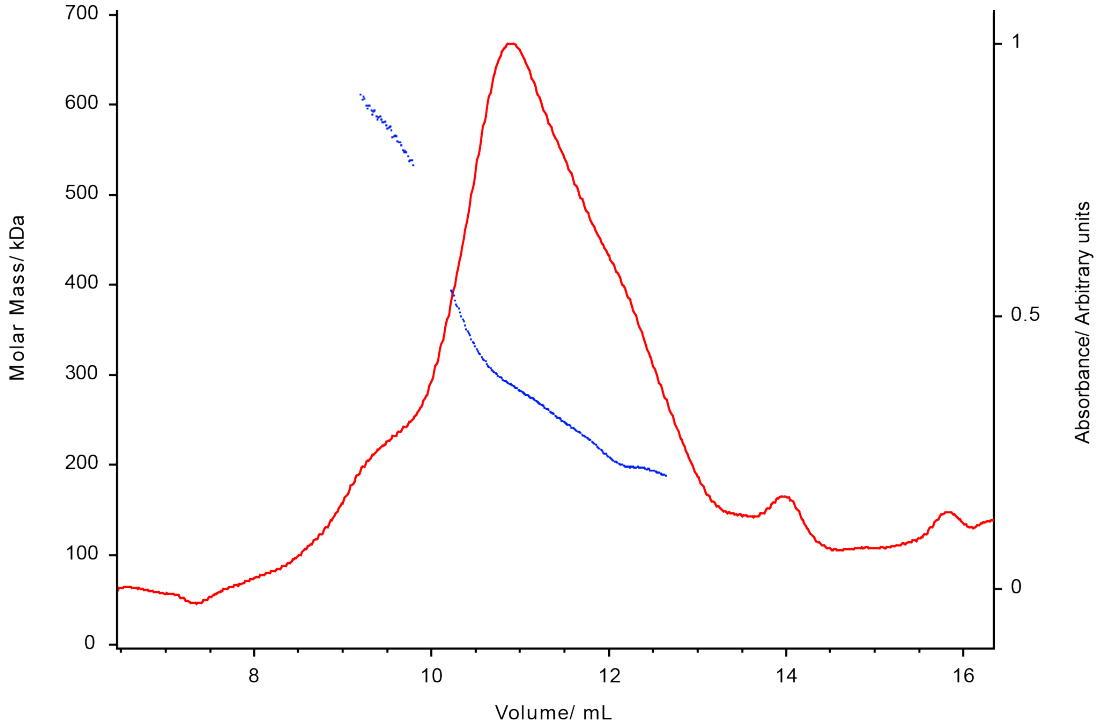


Figure 68: SEC-MALLS of a 5-component shelterin sub-complex.
SEC-MALLS was run with an initial protein concentration of 1 mg/mL. Three different major overlapping species were present in the sample leading to calculation of several potential complex masses. Species with a mass of ~ 600 kDa, ~ 300 kDa and ~200 kDa were calculated. Due to the overlapping purification of these different species the molar masses calculated are sloping. These three species may relate to different complexes of shelterin components containing different oligomeric states of the proteins or different numbers of components.

6.3.2 Complex stabilisation

The 5-component complex was used to prepare stained grids for single particle negative stain EM. Micrographs from these initial experiments again showed heterogeneity in particle size with the majority of particles smaller than expected for a 5-component shelterin complex (particle diameter in excess of ~ 120 Å). This was presumably due to dissociation of the complex components from one

another during the staining process, as was observed for the three-component Poz1$_{1-249}$-Tpz1$_{370-508}$-Ccq1$_{130-717}$ complex described in section 5.4.4.

GraFix was once again employed in an attempt to stabilise the complex. Again, several different glutaraldehyde-containing sucrose gradients and centrifugation durations were tested to optimise complex crosslinking and species separation.

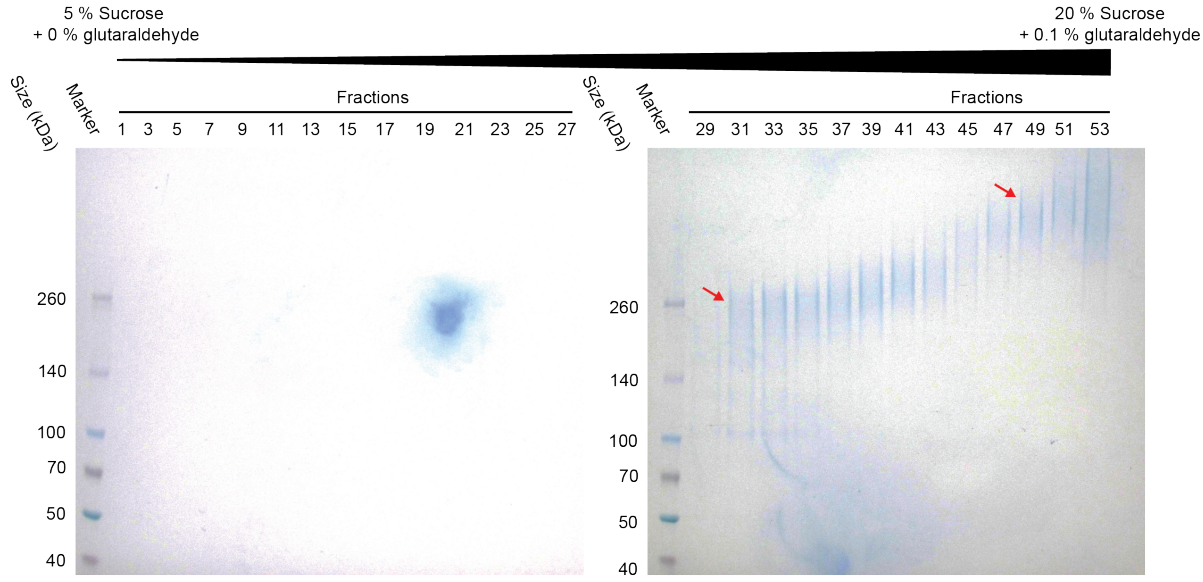


Figure 69: GraFix of the full shelterin complex expressed by the MultiBac insect cell system.
GraFix was performed using a sucrose gradient from 10-30% centrifuged at 36000 rpm for 16 hours as this gave the largest separation of different species after fractionation. The gradient was fractionated into 100 μL aliquots and every second fraction was visualised by coomassie stained SDS-PAGE analysis. Red arrows show the two species that were used for future visualisation by single particle negative stain electron microscopy. Fractions 31-35 and 49-53 were pooled for these experiments as sample A and sample B respectively.

The final GraFix procedure incorporated a sucrose gradient from 10-30 % and centrifugation at 36000 rpm for 16 hours. These conditions gave rise to two main species, both of which were larger in size than the largest protein marker band

(larger than 260 kDa) (see Figure 69) and consistent with the size ranges observed by SEC-MALLS. Proteins present in each crosslinked species were identified by Band ID mass spectrometry. Interestingly, only Poz1-Tpz1-Ccq1 were identified, meaning that either Rap1-Taz1 had been lost during the GraFix step or that the crosslinking chemistry may have affected the trypsin digestion step prior to mass spectrometry.

To confirm Rap1-Taz1 was not present in the final sample, shelterin complex components were expressed by co-infection with just two viruses, containing Poz1-Tpz1 and Ccq1-Pot1. Purification and GraFix treatment of these samples produced a similar pattern after SDS-PAGE analysis to when all three viruses were co-infected, supporting the hypothesis that Rap1-Taz1 was not present in either crosslinked species. Subsequent purification of the three-component complex only used these two viruses. Using this approach, a complex containing Poz1-Tpz1-Ccq1 was purified by a 4-step protocol as described previously, using the N-terminal His tags present on Ccq1 and Poz1. The purified complexes were then stabilised by GraFix, which provided some additional purification. In each repeat a similar pattern of bands was visualised by SDS-PAGE analysis after GraFix stabilisation.

6.3.3 Structural analysis by single particle negative stain electron microscopy

Samples from non-overlapping regions of the two species produced by GraFix (Figure 69) were pooled and concentrated before staining and visualisation by electron microscopy. The smaller species (sample A) produced by GraFix was very difficult to analyse after electron microscopy image acquisition due to a lack of contrast between this relatively small complex (~260 kDa) and the background noise. Only 32 micrographs were taken for this complex and image processing yielded a relatively unfeatured volume. No further image acquisition or processing of this sample was attempted.

The protein concentration and staining protocol used for producing grids of the larger species (sample B) was optimised and then approximately 250 images of this sample were taken as described in section 2.4.5. Electron microscopy grid preparation and data collection were performed with the assistance and guidance of Sarah Webb, Francis Crick Institute. This sample showed high contrast between particles and background and the data were further analysed using the Scipion suite (see Figure 70), as described in section 2.4.6. In brief, 211 micrographs were imported for analysis and ~1000 particles were manually picked from 10 micrographs. These manually selected particles were used as references to allow auto-picking by xmipp3 for the remaining micrographs, leading to the extraction of 30351 particles. Ranking and discarding of particles based on Z score (a measure of particle image quality and similarity to reference

particles) led to removal of 1180 particles with a score larger than 3 from the dataset, leaving 29171 particles for initial 2D classification into 24 classes. 8 of these were discarded, due to these classes containing either small numbers of particles or the production of an unfeatured class average. This left 16 classes containing 18551 particles. The remaining particles were used in a second 2D classification using RELION to produce 100 classes. 46 of these were discarded, leaving 54 classes containing 14445 particles (representative particles and 2D class averages are shown in Figure 71). 10 initial models were generated *ab initio* using EMAN2 and the most featured output was used as an initial model for 3D classification using RELION. This produced 6 3D classes and the class containing the most particles (3530 particles) was then refined with relion auto-refine to produce the final model shown. This produced two maps each derived from half of the data, which were then combined and the resolution of this refined volume at a Fourier Shell Correlation (FSC) = 0.5 was approximately 31 Å (Figure 72b). Angular distribution plots show orientations where data has been collected for the complex. Where there are gaps in the sphere of red dots, this shows that data was not collected for particles in that orientation. The size of the dots relates to the relative number of particles collected for each view. These plots show that the protein complex was found in preferred orientations on grids, as for 2 planes of views of the complex there was minimal data collected. The views where data are missing correspond to the larger flattened sides of the complex (see Figure 72c and Figure 73), potentially adversely affecting the accuracy of the map in these orientations.

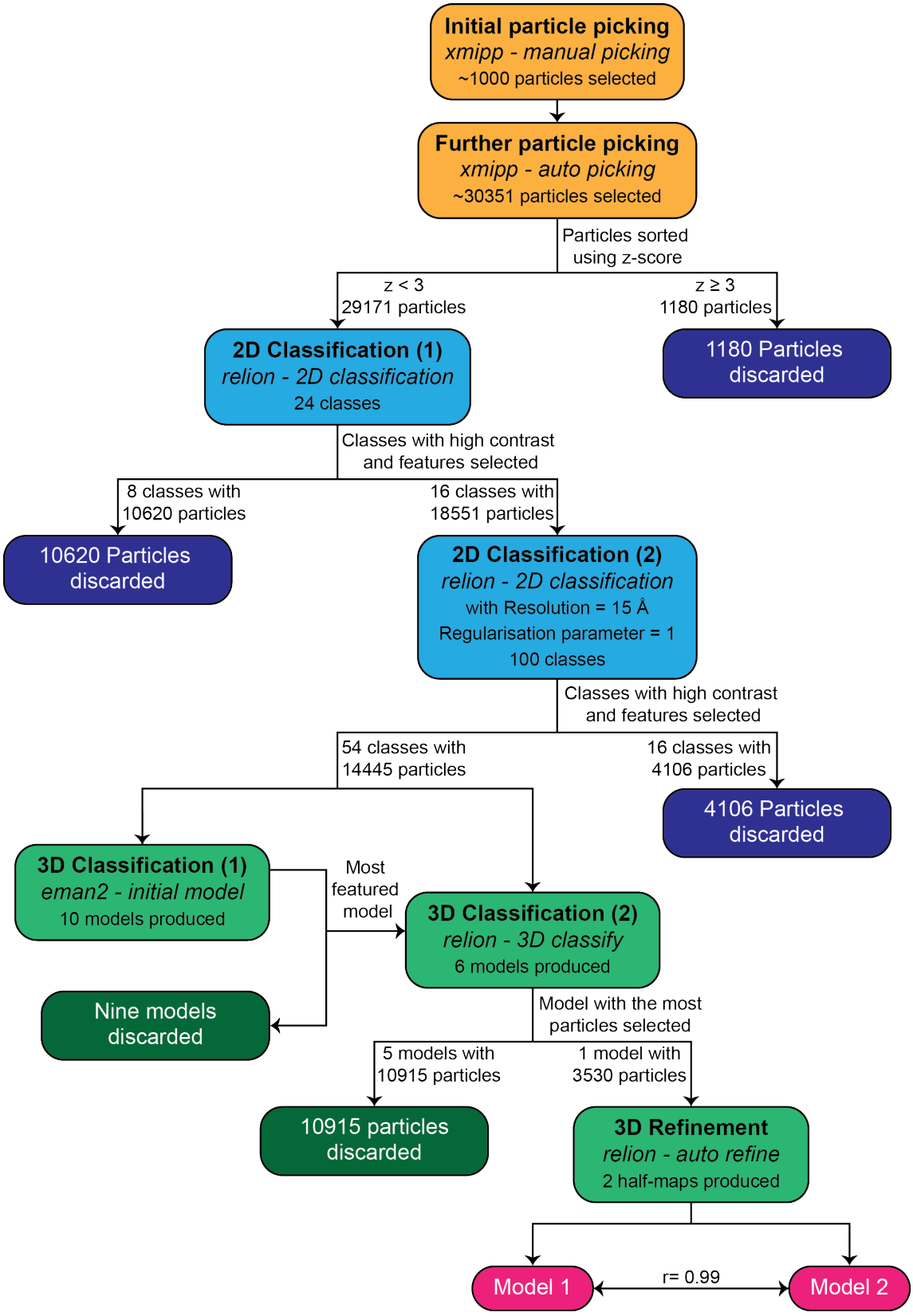
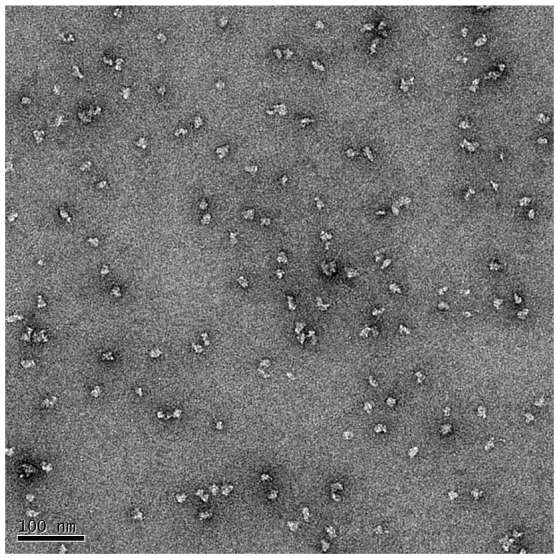


Figure 70: Schematic showing the Scipion workflow used for image analysis. Programs used for each step are shown in *italics*.

a)



b)

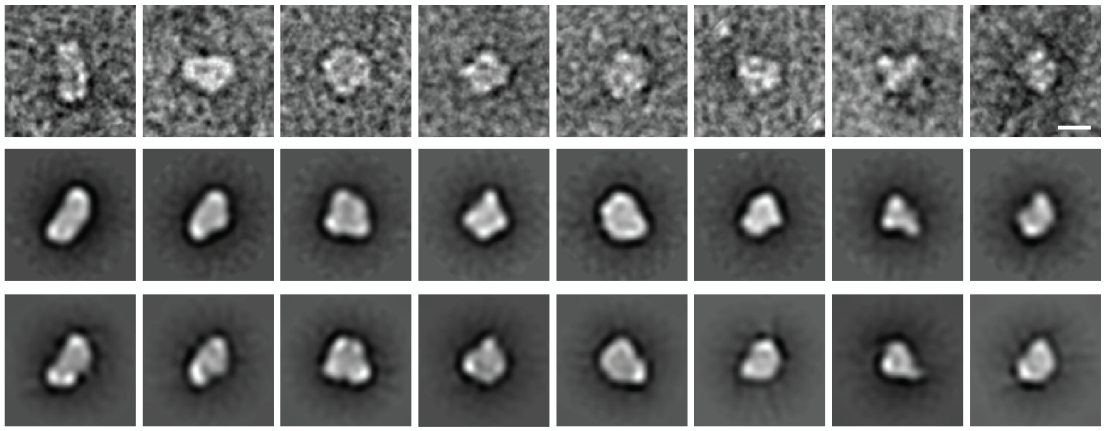


Figure 71: Initial image analysis of the Poz1-Tpz1-Ccq1 shelterin sub complex.
a) Representative field of a raw negatively stained micrograph obtained for a purified crosslinked complex of Poz1-Tpz1-Ccq1. b) Gallery of individual particles from crosslinked full-length Poz1-Tpz1-Ccq1 (top panel) contributing to a corresponding selection of 2D averages (middle panel). These class averages were used to calculate an initial 3D structure and reprojections of this structure (bottom panel). The number of particles contributing to each class average from left to right is 350, 352, 254, 338, 392, 291, 294 and 237. The scale bar shows 10 nm.

Half maps were calculated during the auto-refine process, leading to the production of two maps calculated from separate subsets of the dataset. These half maps appear identical (see Figure 72a) with a correlation of fit of 0.99.

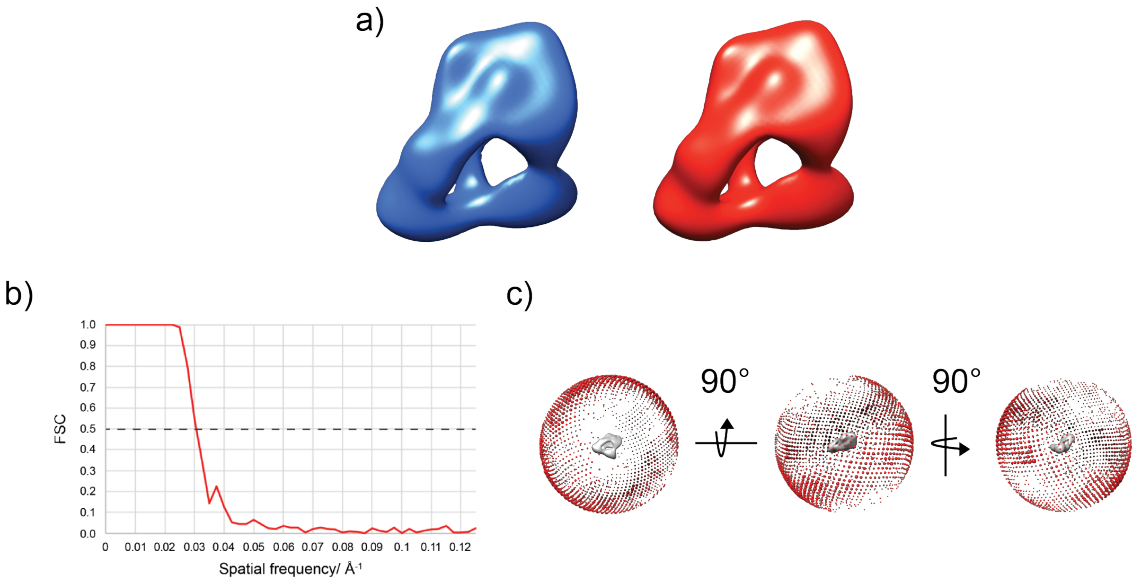


Figure 72: Validation of the structural analysis.
a) Relion auto-refine splits the data into two equally sized subsets and then refines each model separately. Both models appear identical and chimera gives a correlation of fit of 0.99, showing only minimal differences between the maps. This agreement shows that the calculated map fits the data used for classification. b) Fourier-Shell Correlation (FSC) for the crosslinked Poz1-Tpz1-Ccq1 complex between two maps calculated from separate subsets of particle images. At FSC = 0.5, the resolution of this map is approximately 31 Å. c) Orthogonal angular distribution plots, showing that the Poz1-Tpz1-Ccq1 has a preferred orientation when imaged by negative stain electron microscopy. Some views of the complex have only a small number of particles related to them.

The calculated volume, hereafter named CPT-FL, has an asymmetrical shape with a large hole on one side and three further holes on the opposite side. These are linked by a central void around which the volume wraps (see Figure 73). The protein complex has approximate dimensions of 140 x 72 x 190 Å. This asymmetrical shape is somewhat unexpected given the proposed 2: 2: 2 stoichiometry of subunits within the complex. Ccq1 was shown to be dimeric by SAXS and SEC-MALLS and a recent crystal structure has shown Poz1-Tpz1 is found in a 2: 2 complex [90]. A 2: 2: 2 arrangement of CPT-FL is consistent with the proposed stoichiometry of an EM study published during the course of this project (see below) [230].

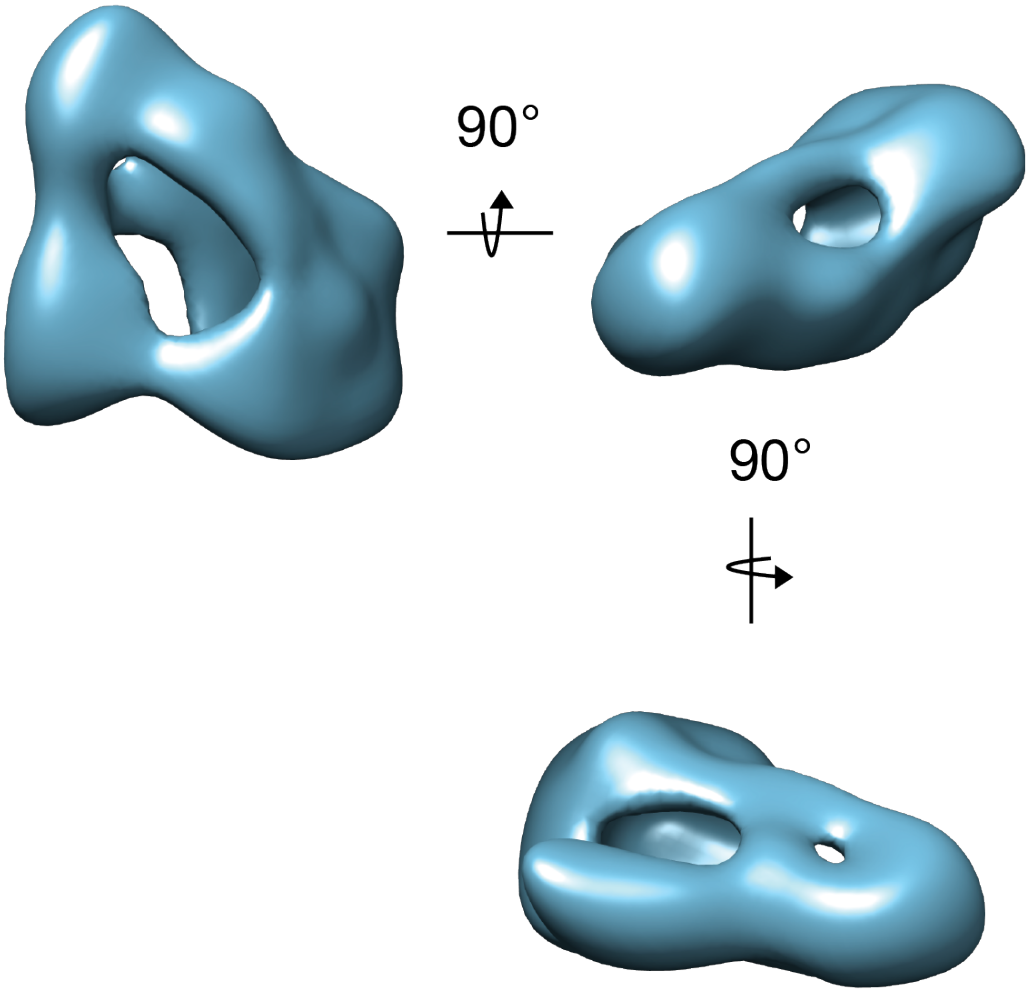


Figure 73: Structure of a crosslinked shelterin sub complex containing Poz1-Tpz1-Ccq1.
Orthogonal surface representations of the 3D map of crosslinked full-length Poz1-Tpz1-Ccq1 (CPT-FL).

CPT-FL shows some similarities to a recently reported EM analysis of a smaller $Poz1_{2\text{-}249}$-$Tpz1_{406\text{-}508}$-$Ccq1_{2\text{-}716}$ complex [230], hereafter called CPT-ΔN, but also contains an additional 820 residues of Tpz1 representing two copies of the N-terminal region. The CPT-ΔN structure was generated by Scott *et al.* by negative stain EM, after co-expression of the proteins in *E. coli*, followed by purification of the proteins by nickel affinity, ion exchange and size exclusion chromatography and stabilisation of the complex by addition of DSSO (for full details see [230]).

The absolute hand of the CPT-FL has not been determined and the best match between CPT-ΔN and CPT-FL was obtained using a z-flipped volume. Manual alignment of these two structures show similarities in some features, but also clear differences. These could be due to errors, but also may be due to rearrangements in the structure caused by the increased size of the Tpz1 protein used. Several regions of each structure have been labelled for clarity of interpretation (shown in Figure 74). A large hole is found on the front of both structures in the top panel of Figure 74 through which holes in the opposite site of the structure, labelled a and b, are visible. These are present in similar places in both structures. A third hole is also present that is not visible in the views shown. Region 1 relates to where Poz1 and Tpz1 are thought to be localised in CPT-ΔN and this region appears to be similarly placed in CPT-FL (blue). Region 2 contains an arm in the structure which appears to be the most consistent feature between both volumes. In CPT-ΔN this was proposed to contain the N-terminus of one Ccq1 monomer and in CPT-FL this region has shifted compared to other regions in the structure. Region 3 shows a protrusion that appears to be present in both structures, which was predicted to contain Ccq1 in CPT-ΔN. Region 4 corresponds to where the proposed interface between the N-terminus of Ccq1 and Tpz1-Poz1 occurs in CPT-ΔN. This comprises a substantial volume in CPT-ΔN, whereas in CPT-FL this density has reduced in size and shifted, suggesting that both N-termini of Ccq1 are likely to occupy different positions in CPT-ΔN and CPT-FL.

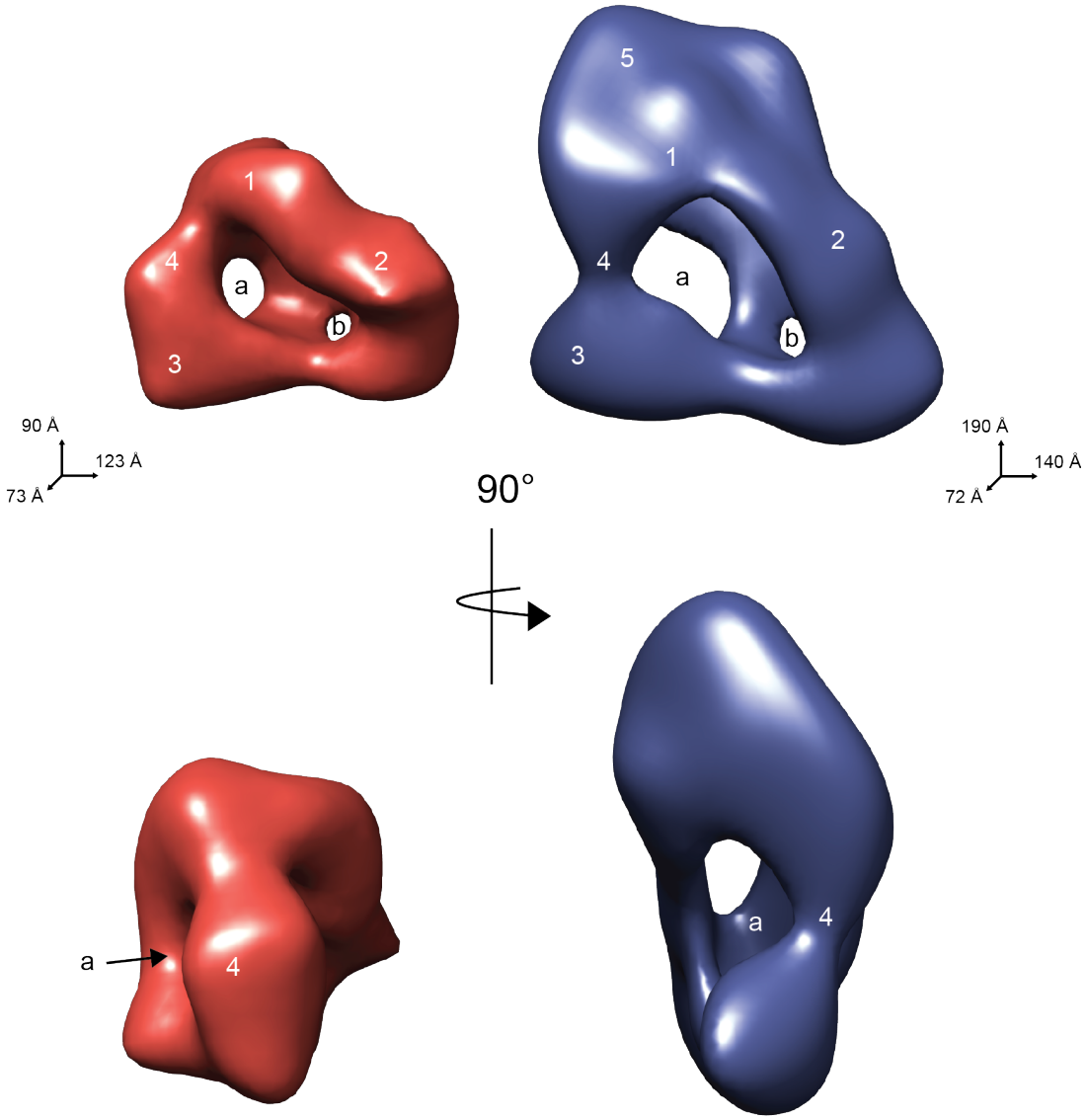


Figure 74: Comparison of the structure of CPT-FL with the structure of Poz1$_{2-249}$-Tpz1$_{406-508}$-Ccq1$_{2-716}$ (CPT-ΔN) (EMDB: 8863).
CPT-ΔN is shown in red and CPT-FL is shown in blue. Optimal alignment of the two volumes required z-flipping of the CPT-FL structure. Letters and numbers mark regions of similarity and differences between the two structures. Volume visualisation was performed using chimera.

CPT-ΔN (red) was shown to contain Poz1-Tpz1$_{406-508}$ in region 1 by tagging the N-terminus of each protein sequentially with MBP and visualising its position. In

CPT-FL, an extended protrusion (labelled region 5) close to region 1 is present, and is sufficiently large to accommodate the additional 95 kDa of Tpz1 (see Figure 75).

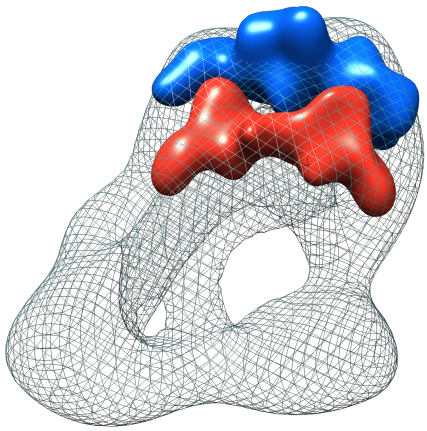


Figure 75: A structure of similar mass to the Tpz1 extension fits into the additional volume seen in CPT-FL.
The volume shown in red contains the crystal structure of Poz1-Tpz1 (PDB 5XXE) and the volume show in blue is an edited version of SMG7 (PDB 1YA0) that has a molecular weight of approximately 95 kDa. SMG7 was used for analysis of the volume size as it has a similar molecular weight to the N-terminal extension of Tpz1 and also has a similar shape to the increased volume in this area of the EM envelope. SMG7 fits into the region proposed to contain the 95 kDa N-terminal extension of Tpz1 present in CPT-FL compared to CPT-ΔN.

As described in section 1.4.1.3, a crystal structure of $Poz1_{1\text{-}249}$-$Tpz1_{478\text{-}508}$-$Rap1_{467\text{-}491}$ was recently solved (see Figure 11) [90]. In order to attempt docking of this structure into both EM volumes, the Rap1 peptide was deleted prior to conversion of this structure to a volume with a similar resolution to the electron microscopy models using the xmipp3- convert a PDB and xmipp3- filter volumes programs within Scipion. Automatic fitting of the resulting volume to CPT-ΔN and CPT-FL was unsuccessful, but the structures could be aligned manually based on common features. The published CPT-ΔN complex shows Tpz1 is present on the inward facing side of the central void close

to region 1 (see Figure 74). For the N-terminal extension of Tpz1 to be found in region 5, either a long linker between the N- and C-terminal domains of Tpz1 is necessary or conformational changes in the CPT-FL complex, causing changes to the position of Tpz1, must occur. Stable association of the N-terminal domains of Tpz1 with the core of the CPT-FL complex seems to occur, as a defined volume is observed by electron microscopy.

6.3.4 Structural asymmetry of the shelterin complex

The previously reported EM structure of $Poz1_{2-249}$-$Tpz1_{406-508}$-$Ccq1_{2-716}$ [230] (see Figure 74) was shown to contain 2 copies of each protein by size exclusion chromatography and SDS-PAGE of a crosslinked protein complex [230]. Interestingly, despite this 2: 2: 2 stoichiometry, neither the published structure, nor the envelope determined during this project show 2-fold symmetry and the molecular basis of the asymmetric architecture remains unknown. Interestingly, the stoichiometry of the human shelterin complex has been shown to be 2 RAP1: 2 TRF2: 1 TIN2: 1 TPP1: 1 POT1 [104] suggesting that this too might be structurally asymmetric. From a functional perspective, the asymmetric distribution of subunits and domains implied by the EM studies described here and elsewhere [230] suggest differences in the accessibility of shelterin complex components. This in turn has implications for their ability to perform regulatory functions, such as the prevention of Ccq1 phosphorylation that occurs through the blocking of Tel/Rad3 accumulation at telomeres by Poz1 [83].

6.3.5 Conformational changes are required for tripartite complex formation

Binding of Poz1 to Tpz1 has previously been shown to allosterically increase binding of Poz1 to Rap1. Similarly, TPP1-TIN2 binding in humans promotes association with TRF2 [91]. This conserved cooperativity may be important for shelterin function and it is possible that binding of Ccq1 to Tpz1 causes further conformational changes to occur that may regulate binding to other shelterin complex components. Allosteric effects may be signification for models for removal of telomerase repression upon binding of Ccq1 to Poz1-Tpz1. Firstly, binding of Ccq1 could disrupt the Poz1-Tpz1 bridge and lead to removal of telomerase repression. However, no evidence of a competitive interaction was observed (section 5.4.1) and Ccq1 and Tpz1 bind to two distinct sites within the C-terminus of Tpz1, consistent with previous studies [83]. A second model involves reorganisation of the Poz1-Tpz1 complex upon binding of Ccq1. A recent crystal structure shows that Tpz1 points out on opposite sides of the Poz1 dimer [90] (see Figure 11). However, the EM analysis of CPT-ΔN, which contains bound Ccq1, positions both Tpz1 peptides on the inner surface of the volume (see Figure 74) suggesting that a conformational change in the relative orientation of the Tpz1 peptides would likely be necessary to reconcile the two structures. This model seems more likely and would resolve the ambiguity generated by the competing models of Poz1-Tpz1 organisation.

Both CPT-ΔN and CPT-FL show a relatively compact but hollow configuration with the structure calculated here having approximate dimensions of 140 x 72 x 190 Å. The D_{max} measured for the Ccq1 dimer by SAXS was 280 Å, which does not fit into the EM volume in any orientation. However, the Kratky analysis indicated that the extended structure is not rigid and that significant flexibility is evident. The CPT-ΔN analysis positions the C-terminal coiled-coil of the Ccq1 dimer in a region of density at the centre of the map common to both structures where three components coalesce, with the two Ccq1 N-terminal arms extending out in different directions [230]. It was proposed that there was a bend in the structure of each arm that occurs before the N-terminal domain of each Ccq1 monomer binds to Tpz1 (see Figure 76). This fits with the model described in section 4.2.5, where coiled-coil dimerisation occurs in an antiparallel configuration and a flexible region between the coiled-coil and HDAC domains is required for complex formation.

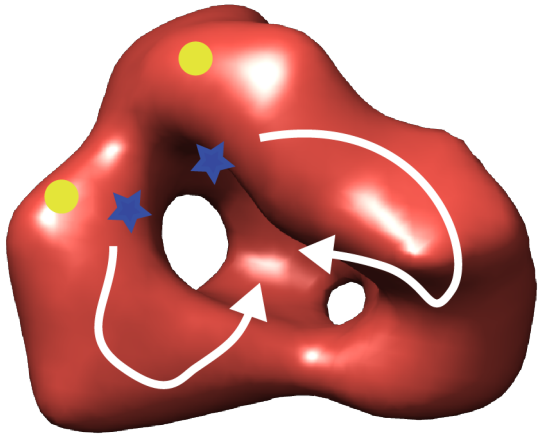


Figure 76: A published EM map (EMDB: 8863) shows a bent conformation of each monomer within a Ccq1 dimer (adapted from *Scott et al.* [230]).
The bent conformation of Ccq1 is coordinated by dimerisation of the C-terminal coiled-coil regions, as shown by the arrows in the model above.

The white arcs shown in the above structure show the published proposed conformation of Ccq1, blue stars show the position of Tpz1 monomers and yellow circles show Poz1 monomer positions within a structure of a complex of $Poz1_{2-249}$-$Tpz1_{406-508}$-$Ccq1_{2-716}$ [230]. The arrowheads show the proposed site of C-terminal coiled-coil dimerisation. If this model is correct Ccq1 must have a bent conformation when an N-terminal region is bound to Tpz1 in the shelterin complex.

6.4 Conclusions

Protein expression was optimised using the MultiBac system to produce full-length proteins in insect cells. Initial EM experiments showed protein dissociation during the staining and grid preparation process. The complex was stabilised using GraFix, allowing collection of data and subsequent image analysis, resulting in a 31 Å resolution map (FSC = 0.5). This volume resembles a recently published structure for $Poz1_{2-249}$-$Tpz1_{406-508}$-$Ccq1_{2-716}$ (CPT-ΔN), but significant differences are evident, possibly as a result of conformational rearrangements resulting from the additional N-terminal region of Tpz1 in the CPT-FL complex. Both CPT-ΔN and CPT-FL show an unexpected asymmetry, which could have functional importance.

A recent crystal structure of $Poz1_{1-249}$-$Tpz1_{478-508}$-$Rap1_{467-491}$ could also be manually fitted into both maps with the orientation and placement of the Tpz1 peptide guided by the MBP-tag experiments described above. Placed in this way, the Tpz1 region points into the void of the volumes. Therefore, in order to accommodate the Tpz1 N-terminal domains, either flexible linkers must be present between the N- and C-termini of Tpz1 or binding of Ccq1 to Tpz1 causes a change in the conformation of the Poz1-Tpz1 complex. Binding of Ccq1 to Tpz1 has previously been shown to remove repression of telomere extension by telomerase [87], however the structural basis of this change remained unknown. Taken together with the data showing a high affinity interaction between Tpz1-

Poz1 and a lower affinity interaction between Tpz1-Ccq1, it seems possible that removal of repression upon Ccq1 binding occurs due to a conformational change in the shelterin bridge rather than by disruption of the interactions within this bridge.

Chapter 7. Future perspectives

7.1 Overview

Telomerase, telomeres and telomere-associated proteins perform vital functions in the maintenance of genomic integrity by solving the end protection and end replication problems associated with the organisation of DNA into chromosomes. The shelterin complex caps chromosome ends and is involved in telomere regulation [7]. The incorrect regulation of telomeres and their extension can lead to cancer and a large number of rare telomere disorders, known as telomeropathies [147].

At the outset of this project very little was known about the structure of the shelterin proteins and their complexes. Knowledge of protein interactions was limited to mutational and yeast-two hybrid analyses, which had identified regions critical for maintenance of wild-type function [48, 83, 87, 120]. This project aimed to extend these initial observations towards a more detailed and comprehensive characterisation of inter-subunit interactions within the fission yeast shelterin complex, and to investigate the structural basis of the mutually exclusive interactions of phosphorylated Ccq1 and TER1 RNA with Est1.

7.2 Further analysis of Ccq1

7.2.1 Structural analysis by crystallography and SAXS

Experiments presented in this thesis have shown that Ccq1 is dimeric in solution and has an extended conformation. The SAXS analysis discussed in Chapter 4 showed a tube-shaped structure with a kink at the centre (see Figure 48). This, together with the CD, limited proteolysis and further bioinformatic analysis (section 4.2), suggests three possible conformations of the Ccq1 dimer: antiparallel coiled-coil dimerisation, parallel coiled-coil dimerisation or dimerisation of the globular domains along with a small region of coiled-coil dimerisation immediately C-terminal to the globular domains (see Figure 49). As discussed in section 4.2.5, the most likely orientation of Ccq1 is an antiparallel dimeric coiled-coil. Two possible experiments to validate this model involve adding or subtracting from the SAXS volume by using different lengths of protein or adding tags to the expressed protein. Changing the length of the protein would result in disappearance or addition of regions of the calculated SAXS volume, and adding a tag would result in addition to the volume. These changes could then be attributed to the changes in the expressed protein, allowing for localisation of these regions to the calculated SAXS volume and the protein orientation to be determined.

There is no high-resolution structural data available for Ccq1 and the published domain information is limited to bioinformatic studies [85]. Ccq1 was predicted to contain a mainly helical coiled-coil region towards the C-terminus and an HDAC2/3-like domain towards the N-terminus (see Figure 9). Crystallisation of Ccq1 130-717 was possible but these crystals diffracted to approximately 8 Å resolution and require optimisation (see Figure 45).

Changes to the purified protein used for crystallisation may also lead to improvements in crystal quality. For example, the purification strategy could be altered by changing purification tags or using different purification buffers. Different protein tags could change the purity of the sample, as different contaminants may be present, while changing the buffer system or the ionic strength changes the stability of the protein in solution, which may lead to an increased likelihood of nucleation. If there were no improvements in diffraction for Ccq1 130-717 observed, further construct optimisation for Ccq1 or co-crystallisation of Ccq1 with one or more of its binding partners might be a productive strategy. High resolution structural information for Ccq1 could be used to direct rational mutagenesis of the protein and the effects tested biochemically *in vitro*. Identification of disruptive mutations would then allow for further testing of the phenotypic effect of these mutations on yeast cell growth, viability and telomere homeostasis.

7.2.2 Ccq1 phosphorylation and it's interaction with Est1

Previously, it has been shown that Ccq1 containing a T to E phosphor-mimic mutation at threonine residue 93 does not recapitulate the binding to Est1 characteristic of T93 phosphorylation and does not lead to wild type telomere maintenance [85]. To fully test the interaction between Ccq1 and Est1, production of phosphorylated Ccq1 would be required. In preliminary experiments, Ccq1 was not phosphorylated when expressed in insect cells under normal conditions. It is possible that subjecting insect cells to DNA damage would lead to activation of ATR/ATM which could cause phosphorylation of Ccq1 to occur. This phosphorylation could then be probed by Western blotting as described. Alternatively, native chemical ligation of a phosphorylated peptide onto *E. coli* expressed, N-terminally truncated Ccq1 might be used to produce phosphorylated protein which could be then used to probe the Ccq1-Est1 interaction and ultimately the competitive Ccq1-Est1-TER1 interaction. Latterly during the project, it was shown that *S. cryophilus* Est1 could be purified in small amounts that are nonetheless sufficient for initial binding analysis to the highly homologous *S. pombe* or *S. cryophilus* Ccq1.

7.3 Analysis of bipartite complexes

Other proteins that form part of the shelterin complex were expressed and purified and the minimum regions required for binding between a number of these proteins were determined (see Chapter 5). BLI and ITC were used to measure the binding affinity between these partners, however initial crystallisation experiments were unsuccessful. Elucidating how these protein interact is an essential step in understanding how the shelterin complex associates and functions within the cell. Building on experiments in this thesis, further crystallisation experiments could be undertaken. In order to do these experiments more protein would be required. Protein yields could be increased by co-expressing binding partners and changing the purification scheme. Expressing the minimum binding domains of shelterin components with their binding partners may increase the solubility of the expressed protein. This may be particularly pertinent in the case of Tpz1 and Poz1, where the interaction is of strong nanomolar affinity. As was already described for Ccq1 (see section 7.2.1), the stability of the sample may be altered by changing the buffers and purification scheme. In the case of Poz1, removing the nickel affinity step during purification would prevent the need to change the pH of the buffer during the purification and would reduce the precipitation seen throughout the process.

If the protein yield and stability could be improved then further crystallisation and analysis of the stoichiometry of shelterin sub-complexes by techniques such as

SEC-MALLS could be undertaken. SAXS could also be used to analyse the low-resolution structure of monodisperse shelterin sub-complexes. In the case of the Poz1-Tpz1 complex, further analysis could utilise AUC in combination with SEC-MALS to resolve the ambiguity in subunit stoichiometry and determine if the complex forms an A2B heterotrimer or a 2: 2 heterotetramer.

7.4 Structural analysis of larger shelterin sub-complexes

During this project, a 31 Å resolution envelope of full length Ccq1-Poz1-Tpz1 was obtained. The resolution of this structure is relatively low but has the potential to be improved. Most importantly, the quality of the sample used for electron microscopy needs to be improved by optimizing expression, purification and sample preparation protocols. At present the complex is purified by nickel affinity chromatography, then size exclusion chromatography, followed by another round of nickel affinity and size exclusion chromatography. Fusion of other affinity purification tags such as GST or MBP to shelterin complex components may improve their solubility, as well as simplifying the purification protocol, reducing the number of steps required. This would allow more complex to be produced and hopefully result in a more stable and soluble complex. In parallel, stability of the complex may also be improved by optimising crosslinking by using a different cross-linking agent, such as disuccinimidyl sulfoxide (DSSO) or bissulfosuccinimidyl suberate (BS3), and varying concentrations and reaction times.

As discussed in Chapter 6, during this project Scott *et al.* published a low-resolution EM structure containing truncated Ccq1-Poz1-Tpz1 [230], which was named CPT-ΔN in this thesis. Scott *et al.* expressed the proteins in Rosetta-BL21 (DE3) *E. coli* using a modified pET28-6His-SUMO vector. Expression was induced by addition of 0.4 mM IPTG for 5 hours at 30°C. The complex was

purified though a standard protocol consisting of nickel affinity and anion exchange chromatography, followed by chemical crosslinking using 10 mM DSSO and then size exclusion chromatography. With this information in hand, further *E. coli* expression of larger complexes of shelterin components could be attempted. The convenience and speed of expression in *E. coli* compared to insect cells makes this system more attractive for high throughput construct and expression optimisation experiments. Another possible method to improve shelterin complex purification would be to purify the complex directly from *S. pombe*. Purification from yeast would require a technique called tandem affinity purification [233]. The TAP tagged protein, and any proteins it is interacting with, could then be immunoprecipitated, allowing for potential purification of the whole shelterin complex in its native state.

If sufficient quantities of shelterin sub-complexes could be reproducibly purified, these complexes could be characterised by SEC-MALLS and analytical ultracentrifugation. These techniques would also allow for determination of subunit composition stoichiometry that might inform the EM studies. To improve the quality of EM data, the staining of EM grids could be optimised by changing the stain used, the incubation time with the stain, the washing protocol or the charge on the grid by using different glow-discharge parameters. These changes may influence the thickness of stain on the grid, which may lead to improved contrast between particles and the background staining, making the particles more visible. As the particle contrast improves, particles become easier to classify and align with similar views during data analysis. The angular distribution

plots for the shelterin envelope indicate lack of data in some orientations, suggesting that the complex may have a preferred orientation on the grid. The amount of data acquired in these orientations could be partly improved by tilting the specimen during image acquisition [234].

Ultimately, the best way to improve the resolution of the model is to image the sample using cryo-electron microscopy (cryo-EM). When the sample is frozen the protein is being directly imaged as opposed to the shell created by the stain, which leads to more accurate details being recorded. Cryo-EM may also overcome issues associated with preferred orientation, as in cryo-EM the sample is frozen within holes of the grid. If high-resolution cryo-EM was successful and resulted in a molecular resolution structure being determined, the location of each of the component proteins could be ascertained through data analysis. To localise proteins within the current low-resolution volume, shelterin components could be tagged with a large globular protein, such as MBP, or an antibody could be bound to an epitope of the protein or a tag such as a His tag. Any extra regions of density in the calculated envelope could be attributed to the tag or antibody, suggesting the location and orientation for that particular protein. It may also be possible to express and purify larger complexes containing more shelterin components. The determined envelopes could then be compared and the extra proteins located in the regions of density not present in the smaller particle. To further aid in the localisation of shelterin components, the interaction interfaces within the complex could be probed by cross-linking mass spectrometry. This analysis would map areas of the complex that are in close proximity, which,

together with electron microscopy experiments and knowledge of pairwise-interactions within the complex, could be used to build a model for how the complex fits together.

In order to fully understand the function of shelterin, it would be useful to determine the structure of shelterin in complex with DNA. DNA substrates containing telomeric sequence and a telomere-like G-overhang could be added to the complex prior to electron microscopy. This information would provide insights into how shelterin interacts with DNA and potential reveal the mechanisms by which shelterin represses or activates telomere extension.

If data produced by cryo-electron microscopy was not sufficient to provide atomic-level resolution, then high resolution structural characterisation of the constituent proteins using X-ray crystallography could be pursued, followed by docking of the structures into a low-resolution EM map to provide a pseudo-atomic model. In the case of shelterin, it is possible that this approach may not be successful, as data in this thesis suggest that shelterin complex proteins may undergo conformational changes upon binding to one another (discussed further in Section 6.3.5). A low-resolution negative stain EM envelope of shelterin could be validated by SAXS of the equivalent un-crosslinked complex. The comparison of these models would show if crosslinking of the proteins leads to production of any artefacts, such as oligomerisation of proteins, which are not present when proteins are in their native state.

Finally, given the biological importance of shelterin function, it is perhaps surprising that the component subunits of the human and yeast complexes show such a high degree of divergence. High resolution structural information for the fission yeast and human shelterin complexes would show, for example, if current mechanisms of recruitment of telomerase accessory subunits required for telomerase stabilisation [105] are conserved between humans and fission yeast. Furthermore, high resolution structural data for the shelterin complex could be used to compare and contrast with mechanisms of action of the recently discovered shieldin complex, which has a similar protective action to shelterin at double strand breaks [114, 119].

Chapter 8. Appendix

Table 21: Uniprot IDs for proteins used during this thesis.

Protein	Uniprot ID
Est1 (*S. pombe*)	O74804
Est1 (*S. cryophilus*)	S9W5V7
Est1 (*S. octosporus*)	S9R276
Est1 (*S. japonicus*)	B6JX13
Ccq1	Q10432
Tpz1	O14246
Poz1	O13852
Pot1	O13988
Rap1	Q96TL7
Taz1	P79005

Table 22: Additives required for small-scale expression tests.
Additives were prepared as described by *Studier* [186].

Additive	Constituent	Concentration
50 x M	Na_2HPO_4	25 mM
	KH_2PO_4	25 mM
	NH_4Cl	50 mM
	Na_2SO_4	5 mM
100 x 505	Glycerol	0.5 %
	Glucose	0.05 %

Table 23: T_m values measured by nanoDSF.
T_m values are calculated as the inflection point of the curve of the ratio of fluorescence at 330 and 350 nm against temperature. Where no value is given measurement was not successful. The higher the T_m, the more stable Ccq1 130-717 is in the related condition.

Condition	T_m
100 mM Citric Acid pH 4.0	-
100 mM Citric Acid 250 mM NaCl pH 4.0	-
100 mM NaAcetate pH 4.5	-
100 mM NaAcetate 250 mM NaCl pH 4.5	-
100 mM SPG pH 5.0	41.0°C
100 mM Citric Acid pH 5.0	42.0°C
100 mM MES pH 6.0	43.5°C

Condition	**T_m**
50 mM MES / Bis-Tris pH 6.0	45.1°C
100 mM Citric Acid 250 mM NaCl pH 5.0	45.5°C
100 mM SPG pH 4.0	45.5°C
100 mM MES 250 mM NaCl pH 6.0	49.0°C
100 mM SPG pH 4.5	49.8°C
100 mM SPG pH 5.5	52.0°C
50 mM Bis-Tris / PIPES pH 6.5	53.5°C
50 mM MES / Imidazole pH 6.5	53.7°C
500 mM Imidazole 100 mM NaCl pH 7.5	57.1°C
100 mM MES pH 6.5	57.6°C
100 mM KPhosphate (monobasic) pH 6.0	58.7°C
100 mM KPhosphate 250 mM NaCl pH 6.0	58.7°C
350 mM Imidazole 100 mM NaCl pH 7.5	58.9°C
50 mM MOPS / Bis-Tris propane pH 7.0	59.4°C
250 mM Imidazole 100 mM NaCl pH 7.5	59.9°C
100 mM Bis-Tris pH 6.5	59.9°C
100 mM Imidazole pH 7.5	60.4°C
100 mM AmAcetate pH 7.3	60.4°C
100 mM MOPS pH 7.0	60.6°C
100 mM KPhosphate 250 mM NaCl pH 7.0	60.6°C
100 mM SPG pH 6.0	60.9°C
100 mM SPG pH 8.5	61.1°C
125 mM Imidazole 100 mM NaCl pH 7.5	61.2°C
50 mM MOPS / HEPES-Na pH 7.5	61.2°C
100 mM KPhosphate (monobasic) pH 7.0	61.4°C
100 mM MES 250 mM NaCl pH 6.5	61.6°C
100 mM Tris-HCl pH 7.5	61.8°C
100 mM SPG pH 10.0	61.8°C
Water 250 mM NaCl	61.9°C
100 mM HEPES pH 7.0	62.0°C
100 mM CHES 250 mM NaCl pH 9.0	62.3°C
Water	62.3°C
100 mM Imidazole 250 mM NaCl pH 7.5	62.3°C
20 mM HEPES pH 7.5	62.4°C
20 mM Na_2Phosphate (dibasic) pH 7.5	62.4°C
50 mM HEPES 50 mM NaCl pH 7.5	62.5°C
50 mM Imidazole 100 mM NaCl pH 7.5	62.5°C
100 mM HEPES pH 8.0	62.5°C

Condition	T_m
50 mM HEPES pH 7.5	62.6°C
100 mM CHES pH 9.0	62.7°C
100 mM Bis-Tris 250 mM NaCl pH 6.5	62.8°C
100 mM Tris-HCl pH 8.5	62.9°C
100 mM Tris-HCl pH 8.0	62.9°C
50 mM Tris-HCl pH 8.0	62.9°C
20 mM Tris-HCl pH 8.0	63.0°C
125 mM HEPES pH 7.5	63.0°C
100 mM SPG pH 6.5	63.1°C
50 mM Na_2Phosphate (dibasic) pH 7.5	63.1°C
100 mM Tricine pH 8.0	63.2°C
50 mM Tris-HCl 50 mM NaCl pH 8.0	63.2°C
100 mM Citric Acid 250 mM NaCl pH 6.0	63.2°C
100 mM Bicine pH 8.5	63.3°C
100 mM MOPS 250 mM NaCl pH 7.0	63.3°C
125 mM Tris-HCl pH 8.0	63.3°C
100 mM AmAcetate 250 mM NaCl pH 7.3	63.4°C
50 mM NaPhosphate / citric acid pH 7.5	63.4°C
50 mM HEPES 125 mM NaCl pH 7.5	63.5°C
0.1 M Bicine / Trizma base pH 8.5	63.5°C
100 mM Tris-HCl 250 mM NaCl pH 8.5	63.5°C
100 mM SPG pH 9.0	63.5°C
100 mM HEPES 250 mM NaCl pH 7.0	63.6°C
100 mM Bicine pH 8.0	63.6°C
100 mM Na_2Phosphate (dibasic) pH 7.5	63.7°C
250 mM Tris-HCl pH 8.0	63.7°C
100 mM SPG pH 8.0	63.8°C
100 mM SPG pH 7.0	63.8°C
125 mM Na_2Phosphate (dibasic) pH 7.5	63.8°C
50 mM Tris-HCl 125 mM NaCl pH 8.0	63.9°C
100 mM SPG pH 7.5	63.9°C
100 mM Citric Acid pH 6.0	63.9°C
100 mM Na_2Phosphate (dibasic) pH 7.0	64.0°C
100 mM Tris-HCl 250 mM NaCl pH 7.5	64.0°C
250 mM Na_2Phosphate (dibasic) pH 7.5	64.1°C
50 mM HEPES 250 mM NaCl pH 7.5	64.1°C
100 mM HEPES 250 mM NaCl pH 8.0	64.2°C
100 mM Bicine 250 mM NaCl pH 8.0	64.2°C

Condition	**T_m**
250 mM HEPES pH 7.5	64.3°C
100 mM Tricine 250 mM NaCl pH 8.0	64.3°C
50 mM Tris-HCl 250 mM NaCl pH 8.0	64.4°C
100 mM Bicine 250 mM NaCl pH 8.5	64.4°C
100 mM Tris-HCl 250 mM NaCl pH 8.0	64.4°C
100 mM Na_2Phosphate 250 mM NaCl pH 7.5	64.5°C
100 mM Na_2Phosphate 250 mM NaCl pH 7.0	64.8°C
50 mM HEPES 500 mM NaCl pH 7.5	65.0°C
50 mM Tris-HCl 750 mM NaCl pH 8.0	65.3°C
50 mM HEPES 750 mM NaCl pH 7.5	65.3°C
50 mM Tris-HCl 1000 mM NaCl pH 8.0	65.4°C
50 mM Tris-HCl 500 mM NaCl pH 8.0	65.4°C
50 mM HEPES 1000 mM NaCl pH 7.5	65.6°C

Reference List

1. Alberts, B.J., A.; Lewis, J.; Raff, M.; Roberts, K.; Walter P.;, *Molecular Biology of the Cell*. 5th ed. 2008: Garland Science.
2. Bloom, K. and A. Joglekar, *Towards building a chromosome segregation machine.* Nature, 2010. **463**(7280): p. 446-56.
3. Olins, A.L. and D.E. Olins, *Spheroid chromatin units (v bodies).* Science, 1974. **183**(4122): p. 330-2.
4. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.
5. Maestroni, L., S. Matmati, and S. Coulon, *Solving the Telomere Replication Problem.* Genes (Basel), 2017. **8**(2).
6. Zhou, B.B. and S.J. Elledge, *The DNA damage response: putting checkpoints in perspective.* Nature, 2000. **408**(6811): p. 433-9.
7. de Lange, T., *How telomeres solve the end-protection problem.* Science, 2009. **326**(5955): p. 948-52.
8. Leman, A.R. and E. Noguchi, *The replication fork: understanding the eukaryotic replication machinery and the challenges to genome duplication.* Genes (Basel), 2013. **4**(1): p. 1-32.
9. Aze, A. and D. Maiorano, *Recent advances in understanding DNA replication: cell type-specific adaptation of the DNA replication program.* F1000Res, 2018. **7**.
10. Burgers, P.M.J. and T.A. Kunkel, *Eukaryotic DNA Replication Fork.* Annu Rev Biochem, 2017. **86**: p. 417-438.
11. Bell, S.P. and B. Stillman, *ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex.* Nature, 1992. **357**(6374): p. 128-34.
12. Fragkos, M., et al., *DNA replication origin activation in space and time.* Nat Rev Mol Cell Biol, 2015. **16**(6): p. 360-74.
13. Donovan, S., et al., *Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast.* Proc Natl Acad Sci U S A, 1997. **94**(11): p. 5611-6.
14. Sun, J., et al., *Cryo-EM structure of a helicase loading intermediate containing ORC-Cdc6-Cdt1-MCM2-7 bound to DNA.* Nat Struct Mol Biol, 2013. **20**(8): p. 944-51.
15. Evrin, C., et al., *A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication.* Proc Natl Acad Sci U S A, 2009. **106**(48): p. 20240-5.
16. Remus, D., et al., *Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing.* Cell, 2009. **139**(4): p. 719-30.
17. Yeeles, J.T., et al., *Regulated eukaryotic DNA replication origin firing with purified proteins.* Nature, 2015. **519**(7544): p. 431-5.

18. Noguchi, Y., et al., *Cryo-EM structure of Mcm2-7 double hexamer on DNA suggests a lagging-strand DNA extrusion model.* Proc Natl Acad Sci U S A, 2017. **114**(45): p. E9529-E9538.
19. Costa, A., et al., *The structural basis for MCM2-7 helicase activation by GINS and Cdc45.* Nat Struct Mol Biol, 2011. **18**(4): p. 471-7.
20. Costa, A., et al., *DNA binding polarity, dimerization, and ATPase ring remodeling in the CMG helicase of the eukaryotic replisome.* Elife, 2014. **3**: p. e03273.
21. Georgescu, R., et al., *Structure of eukaryotic CMG helicase at a replication fork and implications to replisome architecture and origin initiation.* Proc Natl Acad Sci U S A, 2017. **114**(5): p. E697-E706.
22. Garg, P. and P.M. Burgers, *DNA polymerases that propagate the eukaryotic DNA replication fork.* Crit Rev Biochem Mol Biol, 2005. **40**(2): p. 115-28.
23. Johansson, E. and N. Dixon, *Replicative DNA polymerases.* Cold Spring Harb Perspect Biol, 2013. **5**(6).
24. Stith, C.M., et al., *Flexibility of eukaryotic Okazaki fragment maturation through regulated strand displacement synthesis.* J Biol Chem, 2008. **283**(49): p. 34129-40.
25. Liu, B., et al., *Direct Visualization of RNA-DNA Primer Removal from Okazaki Fragments Provides Support for Flap Cleavage and Exonucleolytic Pathways in Eukaryotic Cells.* J Biol Chem, 2017. **292**(12): p. 4777-4788.
26. Zheng, L. and B. Shen, *Okazaki fragment maturation: nucleases take centre stage.* J Mol Cell Biol, 2011. **3**(1): p. 23-30.
27. Sfeir, A.J., J.W. Shay, and W.E. Wright, *Fine-tuning the chromosome ends: the last base of human telomeres.* Cell Cycle, 2005. **4**(11): p. 1467-70.
28. Paeschke, K., K.R. McDonald, and V.A. Zakian, *Telomeres: structures in need of unwinding.* FEBS Lett, 2010. **584**(17): p. 3760-72.
29. Wu, P., H. Takai, and T. de Lange, *Telomeric 3' overhangs derive from resection by Exo1 and Apollo and fill-in by POT1b-associated CST.* Cell, 2012. **150**(1): p. 39-52.
30. Longhese, M.P., et al., *Mechanisms and regulation of DNA end resection.* EMBO J, 2010. **29**(17): p. 2864-74.
31. Soudet, J., P. Jolivet, and M.T. Teixeira, *Elucidation of the DNA end-replication problem in Saccharomyces cerevisiae.* Mol Cell, 2014. **53**(6): p. 954-64.
32. Dai, X., et al., *Molecular steps of G-overhang generation at human telomeres and its function in chromosome end protection.* EMBO J, 2010. **29**(16): p. 2788-801.
33. Chow, T.T., et al., *Early and late steps in telomere overhang processing in normal human cells: the position of the final RNA primer drives telomere shortening.* Genes Dev, 2012. **26**(11): p. 1167-78.
34. Zhao, Y., J.W. Shay, and W.E. Wright, *Telomere G-overhang length measurement method 1: the DSN method.* Methods Mol Biol, 2011. **735**: p. 47-54.

35. Cech, T.R., *Beginning to understand the end of the chromosome.* Cell, 2004. **116**(2): p. 273-9.
36. Saretzki, G. and T. Von Zglinicki, *Replicative aging, telomeres, and oxidative stress.* Ann N Y Acad Sci, 2002. **959**: p. 24-9.
37. Zijlmans, J.M., et al., *Telomeres in the mouse have large inter-chromosomal variations in the number of T2AG3 repeats.* Proc Natl Acad Sci U S A, 1997. **94**(14): p. 7423-8.
38. Gomes, N.M., et al., *Comparative biology of mammalian telomeres: hypotheses on ancestral states and the roles of telomeres in longevity determination.* Aging Cell, 2011. **10**(5): p. 761-8.
39. Sepsiova, R., et al., *Evolution of Telomeres in Schizosaccharomyces pombe and Its Possible Relationship to the Diversification of Telomere Binding Proteins.* PLoS One, 2016. **11**(4): p. e0154225.
40. Giraud-Panis, M.J., et al., *Structural identity of telomeric complexes.* FEBS Lett, 2010. **584**(17): p. 3785-99.
41. Chai, W., et al., *Human telomeres have different overhang sizes at leading versus lagging strands.* Mol Cell, 2006. **21**(3): p. 427-35.
42. Lee, M.E., et al., *Variation of the 3' telomeric overhang lengths in human cells.* Cancer Lett, 2008. **264**(1): p. 107-18.
43. Blackburn, E.H., *Structure and function of telomeres.* Nature, 1991. **350**(6319): p. 569-73.
44. The Nobel Prize. *The Nobel Prize in Physiology or Medicine 2009 - Summary*. 2009 [cited 2018 October]; Available from: https://www.nobelprize.org/prizes/medicine/2009/press-release/.
45. Hiyama, E. and K. Hiyama, *Telomere and telomerase in stem cells.* Br J Cancer, 2007. **96**(7): p. 1020-4.
46. Hayflick, L., *The Limited in Vitro Lifetime of Human Diploid Cell Strains.* Exp Cell Res, 1965. **37**: p. 614-36.
47. Collins, K. and J.R. Mitchell, *Telomerase in the human organism.* Oncogene, 2002. **21**(4): p. 564-79.
48. Leonardi, J., et al., *TER1, the RNA subunit of fission yeast telomerase.* Nat Struct Mol Biol, 2008. **15**(1): p. 26-33.
49. Reichenbach, P., et al., *A human homolog of yeast Est1 associates with telomerase and uncaps chromosome ends when overexpressed.* Curr Biol, 2003. **13**(7): p. 568-74.
50. Chen, H., et al., *Structural Insights into Yeast Telomerase Recruitment to Telomeres.* Cell, 2018. **172**(1-2): p. 331-343 e13.
51. Egan, E.D. and K. Collins, *Biogenesis of telomerase ribonucleoproteins.* RNA, 2012. **18**(10): p. 1747-59.
52. Griffith, J.D., et al., *Mammalian telomeres end in a large duplex loop.* Cell, 1999. **97**(4): p. 503-14.
53. Saint-Leger, A., et al., *The basic N-terminal domain of TRF2 limits recombination endonuclease action at human telomeres.* Cell Cycle, 2014. **13**(15): p. 2469-74.
54. Necasova, I., et al., *Basic domain of telomere guardian TRF2 reduces D-loop unwinding whereas Rap1 restores it.* Nucleic Acids Res, 2017. **45**(21): p. 12170-12180.

55. Stansel, R.M., T. de Lange, and J.D. Griffith, *T-loop assembly in vitro involves binding of TRF2 near the 3' telomeric overhang.* EMBO J, 2001. **20**(19): p. 5532-40.
56. de Lange, T. and J.H. Petrini, *A new connection at human telomeres: association of the Mre11 complex with TRF2.* Cold Spring Harb Symp Quant Biol, 2000. **65**: p. 265-73.
57. Opresko, P.L., et al., *POT1 stimulates RecQ helicases WRN and BLM to unwind telomeric DNA substrates.* J Biol Chem, 2005. **280**(37): p. 32069-80.
58. de Lange, T., *Shelterin: the protein complex that shapes and safeguards human telomeres.* Genes Dev, 2005. **19**(18): p. 2100-10.
59. Wellinger, R.J., A.J. Wolf, and V.A. Zakian, *Saccharomyces telomeres acquire single-strand TG1-3 tails late in S phase.* Cell, 1993. **72**(1): p. 51-60.
60. de Lange, T., *T-loops and the origin of telomeres.* Nat Rev Mol Cell Biol, 2004. **5**(4): p. 323-9.
61. Mosig, G., *Recombination and recombination-dependent DNA replication in bacteriophage T4.* Annu Rev Genet, 1998. **32**: p. 379-413.
62. Amiard, S., et al., *A topological mechanism for TRF2-enhanced strand invasion.* Nat Struct Mol Biol, 2007. **14**(2): p. 147-54.
63. Doksani, Y., et al., *Super-resolution fluorescence imaging of telomeres reveals TRF2-dependent T-loop formation.* Cell, 2013. **155**(2): p. 345-356.
64. Wang, R.C., A. Smogorzewska, and T. de Lange, *Homologous recombination generates T-loop-sized deletions at human telomeres.* Cell, 2004. **119**(3): p. 355-68.
65. Valton, A.L. and M.N. Prioleau, *G-Quadruplexes in DNA Replication: A Problem or a Necessity?* Trends Genet, 2016. **32**(11): p. 697-706.
66. Rhodes, D. and H.J. Lipps, *G-quadruplexes and their regulatory roles in biology.* Nucleic Acids Res, 2015. **43**(18): p. 8627-37.
67. Barber, L.J., et al., *RTEL1 maintains genomic stability by suppressing homologous recombination.* Cell, 2008. **135**(2): p. 261-71.
68. Sarek, G., et al., *TRF2 recruits RTEL1 to telomeres in S phase to promote t-loop unwinding.* Mol Cell, 2015. **57**(4): p. 622-635.
69. Vannier, J.B., G. Sarek, and S.J. Boulton, *RTEL1: functions of a disease-associated helicase.* Trends Cell Biol, 2014. **24**(7): p. 416-25.
70. Cong, Y.S., W.E. Wright, and J.W. Shay, *Human telomerase and its regulation.* Microbiol Mol Biol Rev, 2002. **66**(3): p. 407-25, table of contents.
71. Shay, J.W. and W.E. Wright, *Hayflick, his limit, and cellular ageing.* Nat Rev Mol Cell Biol, 2000. **1**(1): p. 72-6.
72. Rubin, H., *The disparity between human cell senescence in vitro and lifelong replication in vivo.* Nat Biotechnol, 2002. **20**(7): p. 675-81.
73. Shay, J.W. and W.E. Wright, *Role of telomeres and telomerase in cancer.* Semin Cancer Biol, 2011. **21**(6): p. 349-53.
74. Kim, N.W., et al., *Specific association of human telomerase activity with immortal cells and cancer.* Science, 1994. **266**(5193): p. 2011-5.

75. Sirbu, B.M. and D. Cortez, *DNA damage response: three levels of DNA repair regulation.* Cold Spring Harb Perspect Biol, 2013. **5**(8): p. a012724.
76. Jackson, S.P. and J. Bartek, *The DNA-damage response in human biology and disease.* Nature, 2009. **461**(7267): p. 1071-8.
77. Yamazaki, H., Y. Tarumoto, and F. Ishikawa, *Tel1(ATM) and Rad3(ATR) phosphorylate the telomere protein Ccq1 to recruit telomerase and elongate telomeres in fission yeast.* Genes Dev, 2012. **26**(3): p. 241-6.
78. Tomita, K. and J.P. Cooper, *Fission yeast Ccq1 is telomerase recruiter and local checkpoint controller.* Genes Dev, 2008. **22**(24): p. 3461-74.
79. Moser, B.A., et al., *Fission yeast Tel1(ATM) and Rad3(ATR) promote telomere protection and telomerase recruitment.* PLoS Genet, 2009. **5**(8): p. e1000622.
80. Hockemeyer, D. and K. Collins, *Control of telomerase action at human telomeres.* Nat Struct Mol Biol, 2015. **22**(11): p. 848-52.
81. Moser, B.A. and T.M. Nakamura, *Protection and replication of telomeres in fission yeast.* Biochem Cell Biol, 2009. **87**(5): p. 747-58.
82. Miyoshi, T., et al., *Fission yeast Pot1-Tpp1 protects telomeres and regulates telomere length.* Science, 2008. **320**(5881): p. 1341-4.
83. Harland, J.L., et al., *Tpz1-Ccq1 and Tpz1-Poz1 interactions within fission yeast shelterin modulate Ccq1 Thr93 phosphorylation and telomerase recruitment.* PLoS Genet, 2014. **10**(10): p. e1004708.
84. Moser, B.A., O.N. Raguimova, and T.M. Nakamura, *Ccq1-Tpz1TPP1 interaction facilitates telomerase and SHREC association with telomeres in fission yeast.* Mol Biol Cell, 2015. **26**(21): p. 3857-66.
85. Moser, B.A., et al., *Tel1ATM and Rad3ATR kinases promote Ccq1-Est1 interaction to maintain telomeres in fission yeast.* Nat Struct Mol Biol, 2011. **18**(12): p. 1408-13.
86. Naito, T., A. Matsuura, and F. Ishikawa, *Circular chromosome formation in a fission yeast mutant defective in two ATM homologues.* Nat Genet, 1998. **20**(2): p. 203-6.
87. Jun, H.I., et al., *Tpz1 controls a telomerase-nonextendible telomeric state and coordinates switching to an extendible state via Ccq1.* Genes Dev, 2013. **27**(17): p. 1917-31.
88. Armstrong, C.A., et al., *Telomerase activation after recruitment in fission yeast.* Curr Biol, 2014. **24**(17): p. 2006-11.
89. Liu, J., et al., *Dissecting Fission Yeast Shelterin Interactions via MICro-MS Links Disruption of Shelterin Bridge to Tumorigenesis.* Cell Rep, 2015. **12**(12): p. 2169-80.
90. Xue, J., et al., *Structure of the fission yeast S. pombe telomeric Tpz1-Poz1-Rap1 complex.* Cell Res, 2017. **27**(12): p. 1503-1520.
91. Kim, J.K., et al., *Structural Basis for Shelterin Bridge Assembly.* Mol Cell, 2017. **68**(4): p. 698-714 e5.
92. Fairall, L., et al., *Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2.* Mol Cell, 2001. **8**(2): p. 351-61.
93. Pan, L., et al., *Minishelterins separate telomere length regulation and end protection in fission yeast.* Genes Dev, 2015. **29**(11): p. 1164-74.

94. Fujita, I., M. Tanaka, and J. Kanoh, *Identification of the functional domains of the telomere protein Rap1 in Schizosaccharomyces pombe.* PLoS One, 2012. **7**(11): p. e49151.
95. Chen, Y., et al., *A conserved motif within RAP1 has diversified roles in telomere protection and regulation in different organisms.* Nat Struct Mol Biol, 2011. **18**(2): p. 213-21.
96. Miller, K.M., M.G. Ferreira, and J.P. Cooper, *Taz1, Rap1 and Rif1 act both interdependently and independently to maintain telomeres.* EMBO J, 2005. **24**(17): p. 3128-35.
97. Court, R., et al., *How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures.* EMBO Rep, 2005. **6**(1): p. 39-45.
98. Deng, W., et al., *Fission yeast telomere-binding protein Taz1 is a functional but not a structural counterpart of human TRF1 and TRF2.* Cell Res, 2015. **25**(7): p. 881-4.
99. Loayza, D., et al., *DNA binding features of human POT1: a nonamer 5'-TAGGGTTAG-3' minimal binding site, sequence specificity, and internal binding to multimeric sites.* J Biol Chem, 2004. **279**(13): p. 13241-8.
100. Lei, M., et al., *DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA.* Nature, 2003. **426**(6963): p. 198-203.
101. Dickey, T.H., M.A. McKercher, and D.S. Wuttke, *Nonspecific recognition is achieved in Pot1pC through the use of multiple binding modes.* Structure, 2013. **21**(1): p. 121-132.
102. Miyake, Y., et al., *RPA-like mammalian Ctc1-Stn1-Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway.* Mol Cell, 2009. **36**(2): p. 193-206.
103. Wu, L., et al., *Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres.* Cell, 2006. **126**(1): p. 49-62.
104. Lim, C.J., et al., *Reconstitution of human shelterin complexes reveals unexpected stoichiometry and dual pathways to enhance telomerase processivity.* Nat Commun, 2017. **8**(1): p. 1075.
105. Armstrong, C.A., et al., *Fission yeast Ccq1 is a modulator of telomerase activity.* Nucleic Acids Res, 2018. **46**(2): p. 704-716.
106. Sfeir, A., et al., *Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication.* Cell, 2009. **138**(1): p. 90-103.
107. Smogorzewska, A., et al., *Control of human telomere length by TRF1 and TRF2.* Mol Cell Biol, 2000. **20**(5): p. 1659-68.
108. Sfeir, A. and T. de Lange, *Removal of shelterin reveals the telomere end-protection problem.* Science, 2012. **336**(6081): p. 593-7.
109. Tong, A.S., et al., *ATM and ATR Signaling Regulate the Recruitment of Human Telomerase to Telomeres.* Cell Rep, 2015. **13**(8): p. 1633-46.
110. Palm, W., et al., *Functional dissection of human and mouse POT1 proteins.* Mol Cell Biol, 2009. **29**(2): p. 471-82.
111. Takai, K.K., et al., *In vivo stoichiometry of shelterin components.* J Biol Chem, 2010. **285**(2): p. 1457-67.

112. Hu, X., et al., *Multi-step coordination of telomerase recruitment in fission yeast through two coupled telomere-telomerase interfaces.* Elife, 2016. **5**.
113. Lue, N.F., *A physical and functional constituent of telomerase anchor site.* J Biol Chem, 2005. **280**(28): p. 26586-91.
114. Gupta, R., et al., *DNA Repair Network Analysis Reveals Shieldin as a Key Regulator of NHEJ and PARP Inhibitor Sensitivity.* Cell, 2018. **173**(4): p. 972-988 e23.
115. Noordermeer, S.M., et al., *The shieldin complex mediates 53BP1-dependent DNA repair.* Nature, 2018. **560**(7716): p. 117-121.
116. Ghezraoui, H., et al., *53BP1 cooperation with the REV7-shieldin complex underpins DNA structure-specific NHEJ.* Nature, 2018. **560**(7716): p. 122-127.
117. Dev, H., et al., *Shieldin complex promotes DNA end-joining and counters homologous recombination in BRCA1-null cells.* Nat Cell Biol, 2018. **20**(8): p. 954-965.
118. Bhargava, R., D.O. Onyango, and J.M. Stark, *Regulation of Single-Strand Annealing and its Role in Genome Maintenance.* Trends Genet, 2016. **32**(9): p. 566-575.
119. Mirman, Z., et al., *53BP1-RIF1-shieldin counteracts DSB resection through CST- and Polalpha-dependent fill-in.* Nature, 2018. **560**(7716): p. 112-116.
120. Webb, C.J. and V.A. Zakian, *Schizosaccharomyces pombe Ccq1 and TER1 bind the 14-3-3-like domain of Est1, which promotes and stabilizes telomerase-telomere association.* Genes Dev, 2012. **26**(1): p. 82-91.
121. Muslin, A.J., et al., *Interaction of 14-3-3 with signaling proteins is mediated by the recognition of phosphoserine.* Cell, 1996. **84**(6): p. 889-97.
122. Wang, W. and D.C. Shakes, *Molecular evolution of the 14-3-3 protein family.* J Mol Evol, 1996. **43**(4): p. 384-98.
123. Yang, X., et al., *Structural basis for protein-protein interactions in the 14-3-3 protein family.* Proc Natl Acad Sci U S A, 2006. **103**(46): p. 17237-42.
124. Xiao, B., et al., *Structure of a 14-3-3 protein and implications for coordination of multiple signalling pathways.* Nature, 1995. **376**(6536): p. 188-91.
125. Yaffe, M.B., *How do 14-3-3 proteins work?-- Gatekeeper phosphorylation and the molecular anvil hypothesis.* FEBS Lett, 2002. **513**(1): p. 53-7.
126. Johnson, C., et al., *Bioinformatic and experimental survey of 14-3-3-binding sites.* Biochem J, 2010. **427**(1): p. 69-78.
127. Fukuhara, N., et al., *SMG7 is a 14-3-3-like adaptor in the nonsense-mediated mRNA decay pathway.* Mol Cell, 2005. **17**(4): p. 537-47.
128. Rice, C. and E. Skordalakes, *Structure and function of the telomeric CST complex.* Comput Struct Biotechnol J, 2016. **14**: p. 161-7.
129. Eldridge, A.M., W.A. Halsey, and D.S. Wuttke, *Identification of the determinants for the specific recognition of single-strand telomeric DNA by Cdc13.* Biochemistry, 2006. **45**(3): p. 871-9.
130. Feng, X., et al., *CTC1-STN1 terminates telomerase while STN1-TEN1 enables C-strand synthesis during telomere replication in colon cancer cells.* Nat Commun, 2018. **9**(1): p. 2827.

131. Nandakumar, J., et al., *The TEL patch of telomere protein TPP1 mediates telomerase recruitment and processivity.* Nature, 2012. **492**(7428): p. 285-9.
132. Puglisi, A., et al., *Distinct roles for yeast Stn1 in telomere capping and telomerase inhibition.* EMBO J, 2008. **27**(17): p. 2328-39.
133. Hang, L.E., et al., *SUMOylation regulates telomere length homeostasis by targeting Cdc13.* Nat Struct Mol Biol, 2011. **18**(8): p. 920-6.
134. Li, S., et al., *Cdk1-dependent phosphorylation of Cdc13 coordinates telomere elongation during cell-cycle progression.* Cell, 2009. **136**(1): p. 50-61.
135. Matmati, S., et al., *The fission yeast Stn1-Ten1 complex limits telomerase activity via its SUMO-interacting motif and promotes telomeres replication.* Sci Adv, 2018. **4**(5): p. eaar2740.
136. Li, J.S., et al., *TZAP: A telomere-associated protein involved in telomere length control.* Science, 2017. **355**(6325): p. 638-641.
137. Pickett, H.A., et al., *Control of telomere length by a trimming mechanism that involves generation of t-circles.* EMBO J, 2009. **28**(7): p. 799-809.
138. Li, J.S.Z. and E.L. Denchi, *How stem cells keep telomeres in check.* Differentiation, 2018. **100**: p. 21-25.
139. Zhao, Y., et al., *The 11th C2H2 zinc finger and an adjacent C-terminal arm are responsible for TZAP recognition of telomeric DNA.* Cell Res, 2018. **28**(1): p. 130-134.
140. Flores, I., M.L. Cayuela, and M.A. Blasco, *Effects of telomerase and telomere length on epidermal stem cell behavior.* Science, 2005. **309**(5738): p. 1253-6.
141. Hornsby, P.J., *Telomerase and the aging process.* Exp Gerontol, 2007. **42**(7): p. 575-81.
142. Armstrong, C.A. and K. Tomita, *Fundamental mechanisms of telomerase action in yeasts and mammals: understanding telomeres and telomerase in cancer cells.* Open Biol, 2017. **7**(3).
143. Cesare, A.J. and R.R. Reddel, *Alternative lengthening of telomeres: models, mechanisms and implications.* Nat Rev Genet, 2010. **11**(5): p. 319-30.
144. Kim, H., et al., *Systematic analysis of human telomeric dysfunction using inducible telosome/shelterin CRISPR/Cas9 knockout cells.* Cell Discov, 2017. **3**: p. 17034.
145. Townsley, D.M., B. Dumitriu, and N.S. Young, *Bone marrow failure and the telomeropathies.* Blood, 2014. **124**(18): p. 2775-83.
146. Armanios, M., *Telomerase and idiopathic pulmonary fibrosis.* Mutat Res, 2012. **730**(1-2): p. 52-8.
147. Holohan, B., W.E. Wright, and J.W. Shay, *Cell biology of disease: Telomeropathies: an emerging spectrum disorder.* J Cell Biol, 2014. **205**(3): p. 289-99.
148. Ballew, B.J. and S.A. Savage, *Updates on the biology and management of dyskeratosis congenita and related telomere biology disorders.* Expert Rev Hematol, 2013. **6**(3): p. 327-37.

149. Sasa, G.S., et al., *Three novel truncating TINF2 mutations causing severe dyskeratosis congenita in early childhood.* Clin Genet, 2012. **81**(5): p. 470-8.
150. Agrelo, R., et al., *A novel Werner Syndrome mutation: pharmacological treatment by read-through of nonsense mutations and epigenetic therapies.* Epigenetics, 2015. **10**(4): p. 329-41.
151. Armanios, M. and E.H. Blackburn, *The telomere syndromes.* Nat Rev Genet, 2012. **13**(10): p. 693-704.
152. Hao, L.Y., et al., *Short telomeres, even in the presence of telomerase, limit tissue renewal capacity.* Cell, 2005. **123**(6): p. 1121-31.
153. Vulliamy, T., et al., *The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita.* Nature, 2001. **413**(6854): p. 432-5.
154. Armanios, M., et al., *Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita.* Proc Natl Acad Sci U S A, 2005. **102**(44): p. 15960-4.
155. Yamaguchi, H., et al., *Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia.* N Engl J Med, 2005. **352**(14): p. 1413-24.
156. Alder, J.K., et al., *Short telomeres are a risk factor for idiopathic pulmonary fibrosis.* Proc Natl Acad Sci U S A, 2008. **105**(35): p. 13051-6.
157. Armanios, M.Y., et al., *Telomerase mutations in families with idiopathic pulmonary fibrosis.* N Engl J Med, 2007. **356**(13): p. 1317-26.
158. Tsakiri, K.D., et al., *Adult-onset pulmonary fibrosis caused by mutations in telomerase.* Proc Natl Acad Sci U S A, 2007. **104**(18): p. 7552-7.
159. Yamaguchi, H., et al., *Mutations of the human telomerase RNA gene (TERC) in aplastic anemia and myelodysplastic syndrome.* Blood, 2003. **102**(3): p. 916-8.
160. Kirwan, M., et al., *Defining the pathogenic role of telomerase mutations in myelodysplastic syndrome and acute myeloid leukemia.* Hum Mutat, 2009. **30**(11): p. 1567-73.
161. Vulliamy, T., et al., *Association between aplastic anaemia and mutations in telomerase RNA.* Lancet, 2002. **359**(9324): p. 2168-70.
162. Fogarty, P.F., et al., *Late presentation of dyskeratosis congenita as apparently acquired aplastic anaemia due to mutations in telomerase RNA.* Lancet, 2003. **362**(9396): p. 1628-30.
163. Heiss, N.S., et al., *X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions.* Nat Genet, 1998. **19**(1): p. 32-8.
164. Mitchell, J.R., E. Wood, and K. Collins, *A telomerase component is defective in the human disease dyskeratosis congenita.* Nature, 1999. **402**(6761): p. 551-5.
165. Knight, S.W., et al., *Unexplained aplastic anaemia, immunodeficiency, and cerebellar hypoplasia (Hoyeraal-Hreidarsson syndrome) due to mutations in the dyskeratosis congenita gene, DKC1.* Br J Haematol, 1999. **107**(2): p. 335-9.

166. Yaghmai, R., et al., *Overlap of dyskeratosis congenita with the Hoyeraal-Hreidarsson syndrome.* J Pediatr, 2000. **136**(3): p. 390-3.
167. Savage, S.A., et al., *TINF2, a component of the shelterin telomere protection complex, is mutated in dyskeratosis congenita.* Am J Hum Genet, 2008. **82**(2): p. 501-9.
168. Walne, A.J., et al., *TINF2 mutations result in very short telomeres: analysis of a large cohort of patients with dyskeratosis congenita and related bone marrow failure syndromes.* Blood, 2008. **112**(9): p. 3594-600.
169. Chiang, Y.J., et al., *Telomere-associated protein TIN2 is essential for early embryonic development through a telomerase-independent pathway.* Mol Cell Biol, 2004. **24**(15): p. 6631-4.
170. Walne, A.J., et al., *Genetic heterogeneity in autosomal recessive dyskeratosis congenita with one subtype due to mutations in the telomerase-associated protein NOP10.* Hum Mol Genet, 2007. **16**(13): p. 1619-29.
171. Vulliamy, T., et al., *Mutations in the telomerase component NHP2 cause the premature ageing syndrome dyskeratosis congenita.* Proc Natl Acad Sci U S A, 2008. **105**(23): p. 8073-8.
172. Zhong, F., et al., *Disruption of telomerase trafficking by TCAB1 mutation causes dyskeratosis congenita.* Genes Dev, 2011. **25**(1): p. 11-6.
173. Gu, P., et al., *CTC1 deletion results in defective telomere replication, leading to catastrophic telomere loss and stem cell exhaustion.* EMBO J, 2012. **31**(10): p. 2309-21.
174. Keller, R.B., et al., *CTC1 Mutations in a patient with dyskeratosis congenita.* Pediatr Blood Cancer, 2012. **59**(2): p. 311-4.
175. Anderson, B.H., et al., *Mutations in CTC1, encoding conserved telomere maintenance component 1, cause Coats plus.* Nat Genet, 2012. **44**(3): p. 338-42.
176. Guo, Y., et al., *Inherited bone marrow failure associated with germline mutation of ACD, the gene encoding telomere protein TPP1.* Blood, 2014. **124**(18): p. 2767-74.
177. Morris, G.E., *The Cajal body.* Biochim Biophys Acta, 2008. **1783**(11): p. 2108-15.
178. Cristofari, G., et al., *Human telomerase RNA accumulation in Cajal bodies facilitates telomerase recruitment to telomeres and telomere elongation.* Mol Cell, 2007. **27**(6): p. 882-9.
179. Venteicher, A.S. and S.E. Artandi, *TCAB1: driving telomerase to Cajal bodies.* Cell Cycle, 2009. **8**(9): p. 1329-31.
180. Feng, X., et al., *CTC1-mediated C-strand fill-in is an essential step in telomere length maintenance.* Nucleic Acids Res, 2017. **45**(8): p. 4281-4293.
181. Bryant, F.R., *Construction of a recombinase-deficient mutant recA protein that retains single-stranded DNA-dependent ATPase activity.* J Biol Chem, 1988. **263**(18): p. 8716-23.
182. Ratelade, J., et al., *Production of recombinant proteins in the lon-deficient BL21(DE3) strain of Escherichia coli in the absence of the DnaK chaperone.* Appl Environ Microbiol, 2009. **75**(11): p. 3803-7.

183. McCarter, J.D., et al., *Substrate specificity of the Escherichia coli outer membrane protease OmpT.* J Bacteriol, 2004. **186**(17): p. 5919-25.
184. Moosmann, P. and S. Rusconi, *Alpha complementation of LacZ in mammalian cells.* Nucleic Acids Res, 1996. **24**(6): p. 1171-2.
185. Mooij, W.T., E. Mitsiki, and A. Perrakis, *ProteinCCD: enabling the design of protein truncation constructs for expression and crystallization experiments.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W402-5.
186. Studier, F.W., *Protein production by auto-induction in high density shaking cultures.* Protein Expr Purif, 2005. **41**(1): p. 207-34.
187. Martin, L.S., S.; Breitsprecher, D. *Analyzing Thermal Unfolding of Proteins: The Prometheus NT.48*. 2014.
188. Wen, J., T. Arakawa, and J.S. Philo, *Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions.* Anal Biochem, 1996. **240**(2): p. 155-66.
189. Sreerama, N. and R.W. Woody, *Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set.* Anal Biochem, 2000. **287**(2): p. 252-60.
190. Wiseman, T., et al., *Rapid measurement of binding constants and heats of binding using a new titration calorimeter.* Anal Biochem, 1989. **179**(1): p. 131-7.
191. Do, T., et al., *A rapid method for determining dynamic binding capacity of resins for the purification of proteins.* Protein Expr Purif, 2008. **60**(2): p. 147-50.
192. Bergfors, T., *Seeds to crystals.* J Struct Biol, 2003. **142**(1): p. 66-76.
193. Mertens, H.D. and D.I. Svergun, *Structural characterization of proteins and complexes using small-angle X-ray solution scattering.* J Struct Biol, 2010. **172**(1): p. 128-41.
194. Jacques, D.A. and J. Trewhella, *Small-angle scattering for structural biology--expanding the frontier while avoiding the pitfalls.* Protein Sci, 2010. **19**(4): p. 642-57.
195. Blanchet, C.E. and D.I. Svergun, *Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution.* Annu Rev Phys Chem, 2013. **64**: p. 37-54.
196. Franke, D. and D.I. Svergun, *DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering.* J Appl Crystallogr, 2009. **42**(Pt 2): p. 342-346.
197. Volkov, V.V.S., D. I., *Uniqueness of ab initio shape determination in small-angle scattering.* J Appl Crystallogr, 2003. **36**: p. 860-864.
198. Kastner, B., et al., *GraFix: sample preparation for single-particle electron cryomicroscopy.* Nat Methods, 2008. **5**(1): p. 53-5.
199. Faruqi, A.R. and G. McMullan, *Electronic detectors for electron microscopy.* Q Rev Biophys, 2011. **44**(3): p. 357-90.
200. Ruskin, R.S., Z. Yu, and N. Grigorieff, *Quantitative characterization of electron detectors for transmission electron microscopy.* J Struct Biol, 2013. **184**(3): p. 385-93.

201. Kuhlbrandt, W., *Biochemistry. The resolution revolution.* Science, 2014. **343**(6178): p. 1443-4.
202. Karuppasamy, M., et al., *Radiation damage in single-particle cryo-electron microscopy: effects of dose and dose rate.* J Synchrotron Radiat, 2011. **18**(Pt 3): p. 398-412.
203. De Carlo, S. and J.R. Harris, *Negative staining and cryo-negative staining of macromolecules and viruses for TEM.* Micron, 2011. **42**(2): p. 117-31.
204. Booth, D.S., A. Avila-Sakar, and Y. Cheng, *Visualizing proteins and macromolecular complexes by negative stain EM: from grid preparation to image acquisition.* J Vis Exp, 2011(58).
205. Sorzano, C.O., et al., *Automatic particle selection from electron micrographs using machine learning techniques.* J Struct Biol, 2009. **167**(3): p. 252-60.
206. de la Rosa-Trevin, J.M., et al., *Xmipp 3.0: an improved software suite for image processing in electron microscopy.* J Struct Biol, 2013. **184**(2): p. 321-8.
207. Sorzano, C.O., et al., *Semiautomatic, high-throughput, high-resolution protocol for three-dimensional reconstruction of single particles in electron microscopy.* Methods Mol Biol, 2013. **950**: p. 171-93.
208. de la Rosa-Trevin, J.M., et al., *Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy.* J Struct Biol, 2016. **195**(1): p. 93-9.
209. Scheres, S.H., *A Bayesian view on cryo-EM structure determination.* J Mol Biol, 2012. **415**(2): p. 406-18.
210. Scheres, S.H., *RELION: implementation of a Bayesian approach to cryo-EM structure determination.* J Struct Biol, 2012. **180**(3): p. 519-30.
211. Scheres, S.H. and S. Chen, *Prevention of overfitting in cryo-EM structure determination.* Nat Methods, 2012. **9**(9): p. 853-4.
212. Tang, G., et al., *EMAN2: an extensible image processing suite for electron microscopy.* J Struct Biol, 2007. **157**(1): p. 38-46.
213. Yoon, S.H., S.K. Kim, and J.F. Kim, *Secretory production of recombinant proteins in Escherichia coli.* Recent Pat Biotechnol, 2010. **4**(1): p. 23-9.
214. Sockolosky, J.T. and F.C. Szoka, *Periplasmic production via the pET expression system of soluble, bioactive human growth hormone.* Protein Expr Purif, 2013. **87**(2): p. 129-35.
215. Simossis, V.A. and J. Heringa, *PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W289-94.
216. Jones, D.T., *Protein secondary structure prediction based on position-specific scoring matrices.* J Mol Biol, 1999. **292**(2): p. 195-202.
217. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis.* Nat Protoc, 2015. **10**(6): p. 845-58.
218. Lupas, A., M. Van Dyke, and J. Stock, *Predicting coiled coils from protein sequences.* Science, 1991. **252**(5009): p. 1162-4.
219. Vincent, T.L., P.J. Green, and D.N. Woolfson, *LOGICOIL--multi-state prediction of coiled-coil oligomeric state.* Bioinformatics, 2013. **29**(1): p. 69-76.

220. Wilkins, M.R., et al., *Protein identification and analysis tools in the ExPASy server.* Methods Mol Biol, 1999. **112**: p. 531-52.
221. de Chadarevian, S., *John Kendrew and myoglobin: Protein structure determination in the 1950s.* Protein Sci, 2018. **27**(6): p. 1136-1143.
222. Soding, J., A. Biegert, and A.N. Lupas, *The HHpred interactive server for protein homology detection and structure prediction.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W244-8.
223. Rambo, R.P. and J.A. Tainer, *Accurate assessment of mass, models and resolution by small-angle scattering.* Nature, 2013. **496**(7446): p. 477-81.
224. Fischer, H.d.O.N., M.; Napolitano, H. B.; Polikarpov, I.; Craievich, A. F., *Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale.* J. Appl. Cryst., 2010. **43**: p. 101-109.
225. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.
226. Zimmermann, L., et al., *A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core.* J Mol Biol, 2018. **430**(15): p. 2237-2243.
227. Tarricone, C., et al., *The structural basis of Arfaptin-mediated cross-talk between Rac and Arf signalling pathways.* Nature, 2001. **411**(6834): p. 215-9.
228. Khan, D.H., et al., *Protein kinase CK2 regulates the dimerization of histone deacetylase 1 (HDAC1) and HDAC2 during mitosis.* J Biol Chem, 2013. **288**(23): p. 16518-28.
229. Fabrini, R., et al., *Monomer-dimer equilibrium in glutathione transferases: a critical re-examination.* Biochemistry, 2009. **48**(43): p. 10473-82.
230. Scott, H., et al., *Spatial Organization and Molecular Interactions of the Schizosaccharomyces pombe Ccq1-Tpz1-Poz1 Shelterin Complex.* J Mol Biol, 2017. **429**(19): p. 2863-2872.
231. Garcia-Exposito, L. and R.J. O'Sullivan, *TZAP-ing telomeres down to size.* EMBO Rep, 2017. **18**(6): p. 861-863.
232. Craig, N.L., *Tn7: a target site-specific transposon.* Mol Microbiol, 1991. **5**(11): p. 2569-73.
233. Puig, O., et al., *The tandem affinity purification (TAP) method: a general procedure of protein complex purification.* Methods, 2001. **24**(3): p. 218-29.
234. Naydenova, K. and C.J. Russo, *Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy.* Nat Commun, 2017. **8**(1): p. 629.