The molecular basis of telomere-mediated chromosome pairing and genetic exchange during prophase I of meiosis

Amy Milburn

Thesis presented for the degree of Doctor of Philosophy

Institute for Cell and Molecular Biosciences

September 2021

Abstract

Prophase I encapsulates the unique and defining events of meiosis; chromosome pairing, homologous recombination, synapsis, and subsequent segregation, to produce genetically unique haploid germ cells. These essential processes depend on a variety of protein complexes including the meiotic telomere complex (MTC), the synaptonemal complex (SC) and the meiotic recombination machinery. To carry out homology searches chromosomes must be tethered to the nuclear envelope, this is achieved by interplay between the MTC proteins (MAJIN, TERB1, and TERB2) and the shelterin complex protein, TRF1. Once attached, chromosomes undergo rapid prophase movements to find their homologous partner and begin to synapse. This involves the formation of the universally conserved SC structure along the lengths of aligned homologues. The SC provides the essential structural framework for HR and the crossover (CO) pathway. CO formation is dependent on the pro-CO machinery, including the E3 ligase proteins, HEI10 and RNF212.

The main focus of this thesis is to use biophysical and structural approaches to deepen our understanding of the roles and mechanisms of the proteins and protein complexes involved in these meiotic processes. Here, we report the structural basis of the mammalian MTC and provide a mechanistic insight into chromosome tethering at the inner nuclear membrane (INM) achieved by the MTC. We show that the MTC recruits telomere-bound TRF1, through the 2:1 TRF1:TERB1 interaction, and undergoes subsequent structural rearrangement to displace TRF1 allowing the MTC to directly bind telomeric DNA and subsequently stabilise telomere-INM connectivity. The core architecture of mammalian SC is provided through the self-assembly of the transverse filament (TF) protein, SYCP1. We provide the first structural analysis of the *D. melanogaster* SC, specifically the TF protein, C(3)G. Biophysical analysis reveals that the central α-helical domain of C(3)G form dimers in a side-by-side parallel arrangement, but has some propensity to tetramerise, which could serve as building blocks for the recruitment and assembly of the complete SC. We show that HEI10 forms an obligate tetrameric structure and RNF212:RNF212b for a highly stable 2:2 complex and propose a structural model for the human E3 ligase proteins based upon solution scattering studies. Together, these findings provide a solid foundation for elucidating the mechanisms of mammalian meiosis.

Acknowledgments

Over the last three years I have received a great deal of support and assistance. I would first like to thank my PhD supervisor, Dr Owen Davies, for giving me this opportunity and whose expertise has been invaluable during this experience. Owen has provided continuous enthusiasm for the project, encouragement, and patience, especially when things were not going to plan. Besides my supervisor, I would like to thank my progress panel, Professor Bert van den Berg and Professor Jon Higgins, for their insightful comments and encouragement.

Next, I would like to acknowledge past and current members of the Davies lab, first of all providing a welcoming working environment and their continuous support and most importantly, their humor. I would particularly like to single out Dr Orla Dunne, whom I want to thank for her patience during the first few months and answering my many questions, but more importantly being an invaluable friend both in and out of the lab. I would also like to thank fellow lab members, Dr James Dunce and Guru Manickam, for their work and support during our collaborative work. I would also like to thank Dr Helen Waller and Dr Arnaud Basle for their technical support and advice during CD and X-ray crystallography experiments.

Last but not the least, I would like to thank my family and in particular my boyfriend Alan, who has provided endless support and encouragement, as well as taking time to be involved in my work. Most importantly you have kept me sane over the last three years.

Table of Contents

Abstract……..i

Acknowledgements…….iii

Table of contents……...v

List of figures…….xiii

List of tables…….xviii

Abbreviations…….xix

Chapter 1 – Introduction…….1

1.1.1. The cell cycle, mitosis and meiosis……..2

1.1.2. Chromosome architecture……...3

1.1.3. Meiotic division…….….5

1.1.4. Meiotic Prophase I…….…….....7

1.1.5. Importance of cohesins and axial elements (AEs) in meiosis I……9

1.1.6. Rapid prophase movements (RPMs) facilitate the pairing of homologous chromosomes…....11

1.1.7. The meiotic linker of nucleoskeleton and cytoskeleton (LINC) complex……..13

1.1.8. The meiotic telomere complex (MTC)……..14

1.1.9. The synaptonemal complex (SC): A universal meiotic structure……...16

1.1.10. The SC and homologous recombination……18

1.1.11. Meiotic recombination pathway: mechanism and regulation……20

1.1.12. Defective recombination and infertility……24

1.1.13. Thesis aims and objectives……25

Chapter 2 – Methods……..29

2.1. Protein sequence analysis ……...30

2.1.1. Protein sequences analysed in this study……30

2.1.2. ***Bioinformatic analyses……...30***

2.2. Preparation of plasmids for recombinant protein expression……31

2.2.1. Plasmids used in this study……31

2.2.2. PCR amplification……...32

2.2.3. Plasmid linearisation……...33

2.2.4. Agarose gel electrophoresis……....33

2.2.5. One-step sequence and ligation independent cloning (SLIC)……..33

2.3. Transformation of competent cells……34

2.4. Recombinant protein expression and purification…… ..35

2.4.1. Over-expression of recombinant protein……35

2.4.2. Preparation of cell lysate……35

2.4.3. Nickel affinity chromatography……..35

2.4.4. Amylose affinity chromatography……..36

2.4.5. Ion exchange chromatography……36

2.4.6. Affinity-tag cleavage……..37

2.4.7. Size-exclusion chromatography……..37

2.4.8. Protein concentration and buffer exchange……37

2.4.9. Protein concentration determination……...38

2.4.10. SDS-PAGE analysis……...38

2.5. Biophysical assays……...39

2.5.1. Circular dichroism (CD) spectroscopy……39

2.5.2. Size-exclusion chromatography multi-angle light scattering (SEC-MALS)……...39

2.5.3. Size-exclusion chromatography small-angle x-ray scattering (SEC-SAXS)……..40

2.5.4. SAXS-directed protein structure modelling……41

2.5.5. Electrophoretic mobility shift assay (EMSA)……41

2.5.6. Zinc content determination assay (PAR assay)……...42

2.5.7. Limited proteolysis……43

2.5.8. Amylose pulldown……..43

2.6. X-ray crystallography……43

2.6.1. High throughput crystallisation screening……...43

2.6.2. Optimisation screening……...43

2.6.3. X-ray data collection and structure determination of $TRF1_{TRFH}$-$TERB1_{bonsai}$……..44

2.7. Yeast two hybrid (Y2H)……..45

2.7.1. Growth of strains……45

2.7.2. Transformation of yeast cells……..45

2.7.3. Yeast mating……...45

Chapter 3 - The molecular basis of the meiotic telomere tethering complex……...47

3.1. Introduction……....48

3.1.1. Telomeres are nucleoprotein caps that adorn the end of chromosomes……...48

3.1.2. Telomere repeat factor proteins, TRF1 and TRF2, initiate shelterin complex formation.49

3.1.3. Formation of the shelterin complex……51

3.1.4. Telomeres are structurally adapted for meiosis……...54

3.1.5. Telomere-led rapid prophase movements facilitate homologous chromosome pairing...54

3.1.6. LINCing the telomeres to the nuclear envelope……...57

3.1.7. The meiotic telomere complex is conserved across eukaryotes……...59

3.1.8. TERB1 functions as a molecular scaffold within the mammalian meiotic telomere complex……..61

3.1.9. TRF1-TERB1 interaction resembles TRF1-TIN2……...62

3.1.10. Meiosis specific nuclear membrane complex, MAJIN-TERB2……..65

3.1.11. The human meiotic telomere complex……69

3.1.12. Telomere cap exchange……...70

3.1.13. Chapter aims……...71

3.2. Results……72

3.2.1. Bioinformatic and structural analysis of the meiotic telomere complex components…..72

3.2.2. Purification and characterisation of full length TRF1……..78

3.2.3. The TRFH domain of TRF1 homodimerises in solution…..……81

3.2.4. TERB1$_{TRFB}$ is a soluble aggregate when expressed in isolation……...……83

3.2.5. TRF1$_{TRFH}$-TERB1$_{TRFB}$ forms a 2:1 complex in solution……..86

3.2.6. TRF1$_{TRFH}$-TERB1$_{TBM}$ complex dissociates in solution……...91

3.2.7. MBP-TERB1$_{TBM}$ is soluble when expressed in isolation of TRF1……...95

3.2.8. TERB1 $_{TRFB}$ binds TERB2 to form a 1:1 complex……...96

3.2.9. Deciphering individual TRF1 and TERB2 binding sites within the TERB1$_{TRFB}$ domain.98

3.2.10. Optimisation of the TERB2 construct……...100

3.2.11. Biophysical characterisation of TERB2$_N$-TERB1$_{T2B}$……...……..……...103

3.2.12. TERB1 T2B binding site solely binds to TERB2……..106

3.2.13. TERB1 bonsai construct stably binds TRF1 to form a 2:1 complex……..……107

3.2.14. Structural analysis of TRF1$_{TRFH}$-TERB1$_{bonsai}$……..……...…...111

3.2.15. Crystallography of TRF1$_{TRFH}$-TERB1$_{bonsai}$……114

3.2.16. Further optimisation of the TERB1$_{bonsai}$ construct…….116

3.2.17. TRF1-TERB1-TERB2 forms a stable tripartite 2:1:1 complex…….117

3.2.18. Multi-phase SEC-SAXS *ab initio* modelling of TRF1-TERB1-TERB2……...119

3.2.19. TERB1-TERB2 undergoes dimerisation when in complex with MAJIN……..121

3.2.20. Modelling the MAJIN-TERB1-TERB2 complex…….123

3.2.21. The MAJIN-TERB1-TERB2 complex binds linear dsDNA……125

3.2.22. MAJIN-TERB1-TERB2 can recruit TRF1 to form a meiotic ternary telomere complex……126

3.2.23. TERB1 T648E phosphomimetic does not disrupt the TRF1$_{TRFH}$-TERB1$_{TRFB}$ complex.130

3.2.24. TERB1$_{TBM}$ phosphomimetic blocks TRF1$_{TRFH}$ binding…….134

3.2.25. TERB1$_{TRFB}$ T648E has a 10-fold reduction in binding affinity compared to wild type……...136

3.3. Discussion…….139

3.3.1. Meiotic telomere attachment is achieved by two distinct ternary complexes……..139

3.3.2. The meiotic telomere complex undergoes rearrangement upon TRF1 attachment……144

3.3.3. The DNA-binding dynamics of the meiotic telomere complex enables telomere attachment…….144

3.3.4. Spatial displacement of TRF1 upon telomere attachment…….147

3.3.5. Elucidating the meiotic telomere attachment mechanism…….148

3.3.6. Role of the meiotic telomere complex post telomere attachment……..151

Chapter 4 - Chacterisation of the *Drosophila* transverse filament protein, C(3)G……155

4.1. Introduction……...156

4.1.1. The tripartite structure of the synaptonemal complex is conserved across eukaryotes…..156

4.1.2. The *Drosophila melanogaster* synaptonemal complex…….……157

4.1.3. Structure of transverse filament proteins is conserved across eukaryotes……159

4.1.4. Human SYCP1 transverse filament assembly……..162

4.1.5. *Drosophila melanogaster* as a model organism for studying meiosis……...165

4.1.6. C(3)G encodes a *Drosophila* transverse filament protein……..…166

4.1.7. *In vivo* analysis of three in-frame C(3)G deletions……169

4.1.8. Structure-function relationship of coiled-coils……..173

4.1.9. Small angle X-ray scattering (SAXS) can be utilised to characterise SC proteins…….175

4.1.10. Chapter aims……...……..176

4.2. Results……..….178

4.2.1. Purification and characterisation of C(3)G structural core……..…...……...178

4.2.2. C(3)G forms parallel dimeric assemblies in solution…….……..…….184

4.2.3. Dissecting the core structure of C(3)G$_{226-650}$……..186

C(3)G$_{330-650}$……...186

C(3)G$_{470-650}$……...…….189

C(3)G$_{534-650}$ and C(3)G$_{550-650}$…….192

C(3)G$_{330-533}$……..…….197

C(3)G$_{330-459}$……..……..201

4.2.4. The N-terminus of C(3)G degrades in solution……...…….…….207

4.2.5. Purification and characterisation of the unstructured C-terminus, C(3)G$_{644-744}$…….208

4.2.6. C(3)G$_{644-744}$ binds dsDNA…….……210

4.2.7. Biophysical characterision of C(3)G$_{226-650}$ in-frame deletions……..211

4.3. Discussion……..222

4.3.1. The organisation of C(3)G is similar to mammalian SYCP1……...……..222

4.3.2. C(3)G coiled-coil dimers are in a parallel arrangement…….……222

4.3.3. The globular C-terminal domain stabilises C(3)G at the lateral region of the SC……..225

4.3.4. What holds C(3)G molecules together at the midline?.....225

4.3.5. Model for the assembly of the *Drosophila* transverse filament ……..227

Chapter 5 - Characterisation of the mammalian meiotic recombination proteins: HEI10, RNF212 and RNF212b……...231

5.1. Introduction……..…….232

5.1.1. Homologous recombination during meiosis I…….……..232

5.1.2. Meiotic recombination nodules mark the sites of mature crossovers……234

5.1.3. The *S. cerevisiae* ZMM pathway: pro-CO machinery…….…….237

5.1.4. RING domain E3 ligases…….……..240

5.1.5. The Zip3- and HEI10-families of E3 ligases…….243

5.1.6. HEI10 is required for meiotic COs in mice……...……245

5.1.7. RNF212 and HEI10 define an axis associated SUMO-ubiquitin-proteasome relay……..248

5.1.8. *D. melanogaster* encodes three meiosis specific recombination proteins: Vilya, Narya and Nenya……250

5.1.9. Chapter aims…….……252

5.2. Results…….……..254

5.2.1. Bioinformatic and secondary structural analysis of the meiotic recombination proteins...254

5.2.2. Purification and characterisation of full-length HEI10……257

5.2.3. The structural core of HEI10 requires both the RING and coiled-coil domains……260

5.2.4. $HEI10_{1-196}$ requires a long-linker to facilitate cleavage by TEV-protease……..262

5.2.5. $HEI10_{1-196LL}$ forms a stable tetramer in solution……..……..262

5.2.6. $HEI10_{1-196}$ tetramer coordinates four zinc atoms……...267

5.2.7. $HEI10_{1-196LL}$ binds linear dsDNA……..269

5.2.8. Crystallography of $HEI10_{1-196LL}$……269

5.2.9. Optimisation of $HEI10_{1-196}$ expression……..……273

5.2.10. Further crystallography of $HEI10_{1-196}$……...……277

5.2.11. Optimisation of the HEI10 structural core……280

5.2.12. Biophysical analysis and characterisation of $HEI10_{1-122}$……...……282

5.2.13. Biophysical analysis and characterisation of $HEI10_{1-104}$……...286

5.2.14. Purification and characterisation of RNF212 and RNF212b RING domains..…….…….292

5.2.15. $RNF212_{1-198}$ is not stable when expressed in isolation……...……296

5.2.16. $RNF212b_{1-136}$ structural core tetramerises in solution……...……296

5.2.17. Testing possible interactions between the recombination proteins: HEI10, RNF212 and RNF212b……..298

5.2.18. RNF212:RNF212b forms a stable 2:2 complex in solution……...300

5.2.19. RNF212 and RNF212b undergo rearrangement upon complex formation……303

5.2.20. RNF212 and RNF212b interact through their coiled-coil domains……..……307

5.2.21. *S. cerevisiae* E3 ligase protein, Zip3, dimerises in solution…….……..309

5.2.22. In solution characterisation of the *Drosophila* E3 ligase proteins: Vilya, Narya and Nenya……313

5.3. Discussion……...316

5.3.1. HEI10 requires both the RING domain and putative coiled-coil region to form a stable tetramer……..…316

5.3.2. RNF212 and RNF212b form a heterodimeric 2:2 complex……...…319

5.3.3. Meiotic E3 ligase RING domains coordinate zinc ions and bind DNA…….……320

5.3.4. The recombination machinery is structurally conserved across meiotic dividing organisms ……..322

5.3.5. How is the structure of HEI10 and RNF212(b) related to their functions in meiotic recombination?.....323

Chapter 6 – General Discussion and Future Work……..……..325

References…….329

List of figures

Chapter 1

Figure 1.1.1.| The cell cycle…….2
Figure 1.1.2.| DNA architecture and telomere function……...…….....5
Figure 1.1.3.| Stages of meiotic division……..6
Figure 1.1.4.| Schematic depiction of Prophase I…….……...8
Figure 1.1.5.| Sub-phases of Prophase I……...9
Figure 1.1.6.| Meiotic chromosome organisation…….…….....10
Figure 1.1.7.| The meiotic 'bouquet'……..12
Figure 1.1.8.| Visualisation of the LINC complex coupling dynein and telomeres……...14
Figure 1.1.9.| The meiotic telomere complex is the link between shelterin and the LINC complexes.16
Figure 1.1.10.| Mammalian synaptonemal complex (SC)…….17
Figure 1.1.11.| Cytological markers of meiotic recombination in mouse spermatocyte……...20
Figure 1.1.12.| Schematic of the meiotic recombination pathway…….22
Figure 1.1.13.| E3 ligase proteins RNF212 and HEI10 are pro-CO factors……..24

Chapter 3

Figure 3.1.1.| The human shelterin complex……...49
Figure 3.1.2.| Human TRF proteins share a conserved structure……...51
Figure 3.1.3.| Interaction network of the telomeric shelterin proteins: TRF1, TRF2 and TIN2……53
Figure 3.1.4.| The chromosomal bouquet is conserved across eukaryotes……55
Figure 3.1.5.| Stages of Prophase I…….56
Figure 3.1.6.| The LINC complex proteins colocalise at chromosome ends…….58
Figure 3.1.7.| Connecting the telomeres to the cytoskeletal network……59
Figure 3.1.8.| Schematic of the *S. pombe* meiotic telomere complex……...….60
Figure 3.1.9.| Formation of the meiotic telomere attachment plate in mice spermatocytes……..61
Figure 3.1.10.| TERB1 is a meiosis-specific protein……63
Figure 3.1.11.| Crystal structure of $TRF1_{TRFH}$-$TERB1_{TBM}$…….64
Figure 3.1.12.| MAJIN-TERB2 structure……...66
Figure 3.1.13.| DNA binding potential of the MAJIN-TERB2 complex……...68
Figure 3.1.14.| Interaction network of the meiotic telomere complex……...70
Figure 3.2.1.| Sequence analysis of meiotic telomere complex proteins……...73
Figure 3.2.2.| Bioinformatic analysis of meiotic telomere complex components structures……77

Figure 3.2.3.| Purification and structural analysis of $TRF1_{FL}$……...…….79
Figure 3.2.4.| DNA binding ability of $TRF1_{FL}$……...……81
Figure 3.2.5.| Purification and biophysical analysis of $TRF1_{TRFH}$…….82
Figure 3.2.6.| SEC-SAXS analysis of $TRF1_{TRFH...}$…….84
Figure 3.2.7.| TERB1 is a soluble aggregate……..85
Figure 3.2.8.| Screening for TERB1 interactions……...87
Figure 3.2.9.| Purification and biophysical analysis of $TRF1_{TRFH}$:$TERB1_{TRFB}$……...……..89
Figure 3.2.10.| SEC-SAXS data analysis of $TRF1_{TRFH}$:$TERB1_{TRFB}$……..……91
Figure 3.2.11.| Screening for the minimum TERB1 binding site……..93
Figure 3.2.12.| SEC-MALS analysis of His-$TRF1_{TRFH}$:MBP-$TERB1_{TBM}$……..……...94
Figure 3.2.13.| MBP-$TERB1_{TBM}$ is soluble when expressed in isolation of TRF1…….,,,…….96
Figure 3.2.14.| TERB1 and TERB2 form soluble aggregates when expressed in isolation……..97
Figure 3.2.15.| Purification and biophysical analysis of $TERB1_{TRFB}$:$TERB2_{1-119LL}$……..……...99
Figure 3.2.16.| Determination of the TERB2 binding site within $TERB1_{TRFB}$……100
Figure 3.2.17.| SEC-SAXS data analysis and limited proteolysis of TERB1-:TERB2…...……102
Figure 3.2.18.| Purification and biophysical analysis of $TERB2_{N}$:$TERB1_{T2B}$……...……..104
Figure 3.2.19.| SEC-SAXS data analysis of $TERB2_{N}$:$TERB1_{T2B}$……...…….105
Figure 3.2.20.| SDS-PAGE analysis showing TERB1 has individual TRF1 and TERB2 binding sites ……..106
Figure 3.2.21.| Schematic of TERB1 bonsai constructs……..108
Figure 3.2.22.| SEC-MALS analysis of the fusion $TRF1_{TRFH}$:$TERB1_{bonsai}$ constructs……109
Figure 3.2.23.| $TRF1_{TRFH}$-$TERB1_{bonsai}$ forms a stable 2:1 complex…….111
Figure 3.2.24.| Purification and biophysical analysis of $TRF1_{TRFH}$:$TERB1_{bonsai}$…….……113
Figure 3.2.25.| Crystallography of $TRF1_{TRFH}$:$TERB1_{585\ bonsai}$……..115
Figure 3.2.26.| Purification and biophysical analysis of His-$TRF1_{TRFH}$:MBP-$TERB1_{574bonsai}$…….....117
Figure 3.2.27.| TRF1 homodimer complexes with TERB1:TERB2 to form a heterotetramer……118
Figure 3.2.28.| SEC-SAXS data analysis of $TRF1_{TRFH}$:$TERB1_{T2B}$:$TERB2_{N}$……...120
Figure 3.2.29.| Purification and biophysical analysis of $MAJIN_{core}$:$TERB1_{T2B}$:$TERB2_{FL}$……..122
Figure 3.2.30.| SEC-SAXS data analysis of $MAJIN_{core}$:$TERB1_{T2B}$:$TERB2_{FL}$……...…….124
Figure 3.2.31.| MAJIN:TERB1:TERB2 binds linear dsDNA……126
Figure 3.2.32.| Reconstitution of the full meiotic telomere recruitment complex……...127
Figure 3.2.33.| Amylose pulldowns exploring the full meiotic ternary complex……129
Figure 3.2.34.| EMSA supershift assays show DNA binding ability of MAJIN:TERB1:TERB2 upon incubation of TRF1……..130

Figure 3.2.35.| Purification and SEC-MALS analysis of TRF1$_{TRFH}$:TERB1$_{TRFB}$ (T648E) phosphomimetic……...132

Figure 3.2.36.| TERB1$_{TRFB}$ phosphomimetic does not affect ternary complex formation……...133

Figure 3.2.37.| TERB1$_{TBM}$ phosphomimetic blocks TRF1$_{TRFH}$ binding……...135

Figure 3.2.38.| Dilution series of fusion TRF1:TERB1 complexes……137

Figure 3.3.1.| Crystal structure of the TERB1:TERB2 complex……141

Figure 3.3.2.| Schematic models of the MAJIN:TERB1:TERB2 and TRF1:TERB1:TERB2 ternary complexes……143

Figure 3.3.3.| Schematic model of the full meiotic telomere complex……145

Figure 3.3.4.| Structural model of the meiotic pre-displacement complex……..146

Figure 3.3.5.| Schematic models of the pre- and post-displacement complexes……..149

Figure 3.3.6.| SpeedyA interacts with SUN1 and TRF1……..152

Chapter 4

Figure 4.1.1.| The synaptonemal complex forms a conserved tripartite structure……..….157

Figure 4.1.2.| *D. melanogaster* females have a conserved synaptonemal complex…….……158

Figure 4.1.3.| Transverse filament proteins have three distinct domains…….159

Figure 4.1.4.| Antibodies raised to different fragments of transverse filament proteins……..161

Figure 4.1.5.| SYCP1 C-terminal coiled-coil crystal structure……...….163

Figure 4.1.6.| Crystal structure of SYCP1 N-terminal coiled-coil fragment……...……164

Figure 4.1.7.| Schematic showing the central region of the synaptonemal complex…….…..165

Figure 4.1.8.| Immunolocalisation of *D. melanogaster* C(3)G……..…..167

Figure 4.1.9.| Localisation studies of C(3)G in-frame deletions……..168

Figure 4.1.10.| Schematic showing the three in-frame C(3)G deletions……..……170

Figure 4.1.11.| In frame C(3)G deletions within the coiled-coil region results in SC discontinuities.171

Figure 4.1.12.| In frame C(3)G deletions within the coiled-coil region results in reduced recombination……..…….172

Figure 4.1.13.| Coiled-coil interactions……173

Figure 4.1.14.| SAXS analysis determining the helical orientation of SYCE1 core……...…….176

Figure 4.2.1.| C(3)G amino acid sequence alignment……..179

Figure 4.2.2.| Sequence analysis of C(3)G……...……180

Figure 4.2.3.| Purification and structural analysis of C(3)G$_{226-650}$……...…….181

Figure 4.2.4.| SEC-SAXS analysis of C(3)G$_{226-650}$……...…181

Figure 4.2.5.| SEC-SAXS analysis determined the parallel orientation of C(3)G$_{226-650}$……..……185

Figure 4.2.6.| Purification and biophysical analysis of C(3)G$_{330-650}$……..……..187
Figure 4.2.7.| SEC-SAXS analysis of C(3)G$_{330-650}$……...…….…..189
Figure 4.2.8.| Purification and biophysical analysis of C(3)G$_{470-650}$……...…….191
Figure 4.2.9.| SEC-SAXS analysis of C(3)G$_{470-650}$ …..…….…….192
Figure 4.2.10.| Purification and SEC-MALS analysis of C(3)G$_{534-650}$ and C(3)G$_{550-650}$ ……..194
Figure 4.2.11.| C(3)G$_{550-650}$ crystal hits…….196
Figure 4.2.12.| Purification and biophysical analysis of C(3)G$_{330-533}$……..……198
Figure 4.2.13.| SEC-SAXS analysis of C(3)G$_{330-533}$…….200
Figure 4.2.14.| C(3)G$_{330-533}$ crystal hits……...……..201
Figure 4.2.15.| Purification and biophysical analysis of C(3)G$_{330-459}$……..……203
Figure 4.2.16.| SEC-SAXS analysis of C(3)G$_{330-459}$…….204
Figure 4.2.17.| Schematic summary of the C(3)G α-helical domain truncations…….……205
Figure 4.2.18.| Purification of MBP-C(3)G$_{1-230}$……...…….207
Figure 4.2.19.| Purification and biophysical analysis of C(3)G$_{644-744}$…….…….209
Figure 4.2.20.| SEC-SAXS analysis of C(3)G$_{644-744}$…….210
Figure 4.2.21.| EMSA demonstrate the DNA binding ability of C(3)G$_{644-744}$……..……211
Figure 4.2.22.| Purification of MBP-C(3)G deletion constructs……..……212
Figure 4.2.23.| SEC-MALS analysis of MBP-C(3)G Deletion constructs……..….213
Figure 4.2.24.| SEC-SAXS *P(r)* curves of MBP-C(3)G Deletions…….…….214
Figure 4.2.25.| SEC-MALS analysis of MBP-C(3)G Deletion constructs……..…….216
Figure 4.2.26.| Dilution series of cleaved deletion C(3)G constructs……...…218
Figure 4.2.27.| CD analysis of C(3)G Deletions…….…….219
Figure 4.2.28.| SEC-SAXS analysis of C(3)G Deletions ……221
Figure 4.3.1.| Schematic model for C(3)G structure……...…….223
Figure 4.3.2.| C(3)G spans across the central element of the SC…….226
Figure 4.3.3.| Schematic model of C(3)G transverse filament assembly……...……..229

Chapter 5

Figure 5.1.1.| Schematic of the meiotic recombination pathway……...234
Figure 5.1.2.| Recombination nodules mark the positions of chiasmata along SCs…..……..237
Figure 5.1.3.| *S. cerevisiae* ZMM interaction network…….…….239
Figure 5.1.4.| RING E3-mediated catalysis…….…….241

Figure 5.1.5.| RING domains of E3-ligases typically form homo- or hetero-dimers……243
Figure 5.1.6.| SUMO localises to synapsed chromosomes……..……..244
Figure 5.1.7.| HEI10 localisation to the synaptonemal complex and crossover sites…….……247
Figure 5.1.8.| Control of meiotic crossovers……250
Figure 5.1.9.| Protein alignment of *D. melanogaster* meiotic recombination proteins…….…….252
Figure 5.2.1.| HEI10 amino acid sequence alignment……...254
Figure 5.2.2.| Bioinformatic analysis of HEI10……256
Figure 5.2.3.| Bioinformatic analysis of meiotic E3 ligase proteins…….……257
Figure 5.2.4.| Purification and structural analysis of full length HEI10……...…….259
Figure 5.2.5.| HEI10 structural core requires both the RING and coiled-coil domain……..…….261
Figure 5.2.6.| HEI10$_{1\text{-}196}$ requires a TGS long linker for successful cleavage……....262
Figure 5.2.7.| Purification and biophysical analysis of HEI10$_{1\text{-}196\ 5xTGS}$……264
Figure 5.2.8.| SEC-SAXS analysis of HEI10$_{1\text{-}196\ 5xTGS...}$……266
Figure 5.2.9.| HEI10$_{1\text{-}196\ 5xTGS}$ zinc determination by PAR assay……...……268
Figure 5.2.10.| EMSA demonstrate the DNA binding ability of HEI10$_{1\text{-}196LL}$……269
Figure 5.2.11.| HEI10$_{1\text{-}196LL}$ diffracts anisotropically……...…….270
Figure 5.2.12.| HEI10$_{1\text{-}196LL}$ MPD H4 crystal hits……...……..272
Figure 5.2.13.| HEI10$_{1\text{-}196LL}$ MPD H7 optimisation crystal hits……273
Figure 5.2.14.| Optimisation of HEI10$_{1\text{-}196}$ expression…….……275
Figure 5.2.15.| HEI10$_{1\text{-}196}$ cloned into a pRSF-duet vector does not require a TGS long linker to facilitate cleavage……/…….……..275
Figure 5.2.16.| Purification and SEC-MALS analysis of HEI10$_{1\text{-}196}$……...……..275
Figure 5.2.17.| Changing the expression vector does not effect HEI10$_{1\text{-}196}$ zinc content…….……276
Figure 5.2.18.| HEI10$_{1\text{-}196}$ crystal hits……..….277
Figure 5.2.19.| HEI10$_{1\text{-}196\ 5xTGS}$ purified in 20mM L-arginine crystallised in MPD H4……...…..278
Figure 5.2.20.| Limited proteolysis of HEI10$_{1\text{-}196\ 5xTGS}$……..……280
Figure 5.2.21.|Purification and SEC-MALS analysis of HEI10$_{1\text{-}147...}$……...281
Figure 5.2.22.| Purification and biophysical analysis of HEI10$_{1\text{-}122}$…….……282
Figure 5.2.23.| HEI10$_{1\text{-}122}$ zinc content determination by PAR assay……...283
Figure 5.2.24.| SEC-SAXS analysis of HEI10$_{1\text{-}122...}$……284
Figure 5.2.25.| HEI10$_{1\text{-}122}$ crystals hits……..285
Figure 5.2.26.| Purification and biophysical analysis of HEI10$_{1\text{-}104}$……..286
Figure 5.2.27.| HEI10$_{1\text{-}104}$ zinc content determination by PAR assay……...……289
Figure 5.2.28.| SEC-SAXS analysis of HEI10$_{1\text{-}104}$……...…….290

Figure 5.2.29.| Biophysical analysis of RNF212$_{1-61}$ and RNF212b$_{1-56}$……..293
Figure 5.2.30.| Zinc content determination of RNF212$_{1-61}$ and RNF212b$_{1-56}$ by PAR assay……..……295
Figure 5.2.31.| Purification and SEC-MALS analysis of RNF212$_{1-198}$……..297
Figure 5.2.32.| SEC-MALS analysis of MBP-RNF212b……..…298
Figure 5.2.33.| Screening for interactions between HEI10, RNF212 and RNF212b…….299
Figure 5.2.34.| SEC-MALS analysis of MBP-RNF212$_{1-198}$-His-RNF212b$_{1-136}$……301
Figure 5.2.35.| Optimisation of RNF212 with in the fusion RNF212-RNF212b complex……302
Figure 5.2.36.| RNF212b cleavage by TEV-protease needs to be improved…….303
Figure 5.2.37.| Purification and biophysical analysis of RNF212$_{1-123}$-RNF212b$_{1-136}$……..…..304
Figure 5.2.38.| RNF212$_{1-123}$-RNF212b$_{1-136}$ zinc content determination by PAR assay….…….306
Figure 5.2.39.| SEC-SAXS analysis of RNF212$_{1-123}$-RNF212b$_{1-136.}$…….307
Figure 5.2.40.| Purification and SEC-MALS analysis of MBP-RNF212$_{1-198}$-MBP-RNF212b$_{65-136.}$….309
Figure 5.2.41.| Biophysical analysis of MBP-Zip3$_{42-185}$……312
Figure 5.2.42.| Purification and SEC-MALS analysis of *Drosophila* E3 ligase proteins……..315
Figure 5.3.1.| Schematic model of HEI10$_{1-196}$ structural core……317
Figure 5.3.2.| Alternative schematic model for the meiotic E3 ligase……...…318

List of tables

Table 2.1.1.| *D. melanogaster* sequences……..……..29
Table 2.1.2.| Properties of plasmids utilised in this study……..32
Table 3.1.1.| Components connecting chromosomes to the cytoskeleton across eukaryotic species…60
Table 3.2.1.| SEC-MALS analysis of TRF1-TERB1 fusion complexes……..110
Table 3.2.2.| Published structures of TRF1 (1H60) and TRF1-TERB1 (5WIR and 5XUP)……...114
Table 4.1.1.| Eukaryotic transverse filament proteins have conserved features……..……160
Table 4.2.1.| Summary table of C(3)G constructs…….206
Table 5.2.1.| Data collection, phasing and refinement statistics for HEI10$_{1-196}$……...……271
Table 5.2.2.| Summary table of human meiotic E3 ligase proteins……310

Abbreviations

°C	degrees Celsius
Å	Ångstrom
AE	Axial element
APS	ammonium persulphate
BiFC	bimolecular fluorescence complementation
bp	base pair
Bqt1/2	telomere bouquet protein 1/2
BSA	bovine serum albumin
C(3)G	crossover suppressor on 3 of Gowen
C. elegans	*Caenorhabditis elegans*
CD	circular dichroism
CDK	cyclin-dependant kinase
CE	central element
CO	crossover
co-IP	co-immunoprecipitation
CV	column volume
D. melanogaster	*Drosophila melanogaster*
Da	dalton
dH2O	distilled water
dHj	double Holliday junction
D-loop	displacement loop
Dmax	maximum dimension
DNA	Deoxyribonucleic acid
dRI	differential refractive index
ds	double-stranded
DSB	double-strand break
DTT	dithiothreitol
E. coli	*Escherichia coli*
EM	electron microscopy
EMSA	electrophoretic mobility shift assay
EN	early nodule
FISH	fluorescence in situ hybridisation
GST	glutathione S-transferase

IEx	ion exchange
INM	inner nuclear membrane
IPTG	isopropyl β-D-1-thiogalactopyranoside
KASH	Klarsicht/ANC-1/Syne/homology
kb	kilo-base
kDa	kilo Dalton
LB	luria-Bertani
LE	lateral element
LINC	linker of nucleoskeleton and cytoskeleton
LN	late nodule
LS	light scattering
M	molar
MAJIN	membrane-anchored junction protein
MALS	multi-angle light scattering
MBP	maltose binding protein
mg	milligram
min	minute
ml	millilitre
MLH1/3	DNA mismatch repair protein Mlh1/3
mM	millimolar
MMR	mismatch repair
MOPS	3-Morpholinopropanesulfonic acid
MPD	2-Methyl-2,4-pentanediol
MRE	mean residue ellipticity
ms	millisecond
MSA	multiple sequence alignment
MSH4/5	MutS protein homolog 4/5
MTC	meiotic telomere complex
Mw	molecular weight
MWCO	molecular weight cut-off
NCO	non-crossover
Ndj1	Non-disjunction protein 1
NE	nuclear envelope
NEB	New England Biolabs
Ni-NTA	nickel-nitrilotriacetic acid

nm	nano metre
NM	nuclear membrane
OD	optical density
ONM	outer nuclear membrane
P(r)	paired-distance distribution function
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
pH	hydrogen potential
pI	isoelectric point
PTM	post translational modification
Q	scattering vector
Rc	radius of gyration of the cross-section
REC8	Meiotic recombination protein REC8 homolog
Rg	radius of gyration
RING	really interesting new gene
RPM	rapid prophase movements
rpm	revolutions per minute
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Saccharomyces pombe*
SAC	spindle assemble checkpoint
SAXS	small-angle X-ray scattering
SC	synaptonemal complex
SDSA	synthesis dependent strand annealing
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
sfGFP	superfolder green fluorescent protein
SIX6OS1	Six6 opposite strand transcript 1
SLIC	Sequence and ligation independent cloning
SMC1/3/1β	Structural maintenance of chromosomes protein 1/3/1β
SMS	SUMO modification system
ss	single-stranded
STED	stimulated emission depletion
SUMO	small ubiquitin-like modifier
SUN1	Sad-1/UNC-84
SYCE1-3	synaptonemal complex central element protein 1-3

SYCP1-3	synaptonemal complex protein 1-3
TB	terrific broth
TEM	transmission electron microscopy
TEMED	Tetramethylethylenediamine
TERB1/2	telomere repeat-binding bouquet formation protein 1/2
TEV	tobacco etch virus
TEX12	testis-expressed protein 12
TF	transverse filament
TIN2	TERF1-interacting nuclear factor 2
T-loop	telomere loop
Tm	melting temperature
TM	transmembrane
TRF1/2	Telomeric repeat-binding factor 1/2
Tris	tris(hydroxymethyl)aminomethane
µg	microgram
µl	microlitre
µM	micromolar
UPS	ubiquitin proteosome system
UV	ultraviolet
V	volts
w/v	weight per volume concentration percentage
WT	wild type
Y2H	yeast two hybrid
ZIP1	Synaptonemal complex protein Zip1
ZIP3	Molecular zipper protein 3

Chapter 1

INTRODUCTION

1.1.1. The cell cycle, mitosis and meiosis

All living organisms are made up of one or more cells, which are required to reproduce and divide. In prokaryotes, this is achieved through binary fission, in which cell genetic material (DNA) is replicated and then divided into two parts by cytokinesis. In comparison, eukaryotes have a more complex cell cycle, consisting of four precisely timed and highly regulated stages involving: cell growth, DNA replication, distribution of the duplicated daughter cells followed by cell division. The cell cycle has two major phases depending on the cell type: interphase and mitotic phases for somatic cells or interphase and meiotic phase for gametes. Interphase can be further subdivided into three phases: G1, G2 and S phase, where the cell growth (G phases) sandwiches DNA synthesis (S phase). The replicated chromosomes are then distributed to daughter nuclei by a series of highly coordinated events, during mitosis or meiosis, preceding cell division (Figure 1.1.1.) (Alberts et al., 2003; Cooper, 2000).

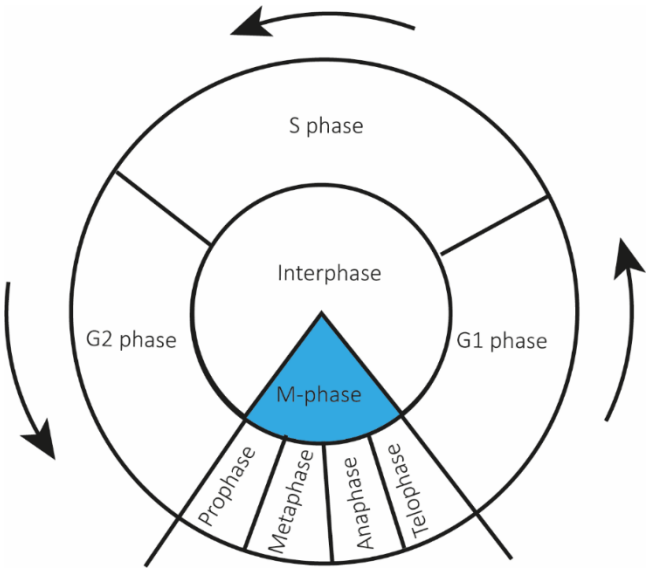


Figure 1.1.1.| The cell cycle.
Actively dividing eukaryotic cells undergo a series of stages known as the cell cycle, comprising of G1, S, G2 and M phases. During the S (synthesis) phase genetic material is duplicated, and in the M phase (either mitosis or meiosis) genetic material and the cell undergoes division. Mitosis and meiosis can be subdivided into four stages: prophase, metaphase, anaphase and telophase, named according to their chromosome appearance.

M-phase is comprised of two processes: mitosis or meiosis, in which the cells chromosomes undergo division between two or four daughter cells, respectively, and cytokinesis. The dynamic nature of M-phase is understood to involve four phases, based on the physical state of the chromosomes and spindle:

prophase, metaphase, anaphase and telophase (Figure 1.1.1.). Telophase is then followed by cytokinesis, which involves the division of the cytoplasm into daughter cells. Mitosis is a highly regulated process, required for cell renewal in plants, animals and fungi. It involves hundreds of different proteins to facilitate the equational division of somatic cells giving rise to two diploid daughter cells. Like in binary fission, the parental identity is preserved in mitosis. On the other hand, meiosis is a specialised form of eukaryotic division required for sexually-reproducing organisms used to produce gametes, such as sperm or egg cells (Baudat et al., 2013; Petronczki et al., 2003).

1.1.2. Chromosome architecture

The cell nucleus houses the majority of the cell's genetic material. The average diameter of a mammalian nucleus is approximately 6 μm, however each human cell contains around two metres of DNA (Alberts et al., 2003). Therefore, during most of the cell cycle nuclear DNA must be organised in a condensed form, known as chromatin, in order to fit into the nucleus. Chromatin structures consist of DNA-protein complexes, in which DNA is combined with histones, proteins around which DNA wraps and condenses, to facilitate compaction within the nucleus. Within cells, chromatin usually folds into characteristic formations called chromosomes (Kornberg, 1974).

During cell division, chromatin alternates between a condensed state, to facilitate chromosome segregation, and a decondensed state when DNA replicates (Sonneville et al., 2015). Upon entry of prophase I, chromosomes become more tightly packed, and their condensed form is visible under a light microscope. At the onset of meiosis, each chromosome exists in a replicated form, giving a 'X' appearance. Each identical sister chromatids produced are connected at the centromere. The short and long arms of the chromatids are called the p and q arms, respectively, and the regions at the end of the chromatids are called the telomeres (Alberts et al., 2003; Blackburn, 2000) (Figure 1.1.2.a.). Every cell of the human body contains 23 chromosomes from each biological parent, a total of 46 chromosomes. Chromosomes 1-22 are autosomes, whilst chromosome 23 corresponds to the sex chromosome (X or Y). Females have two X chromosomes (XX) and males have an X and Y chromosome (XY).

Telomeres are specific DNA-protein structures that cap both ends of each chromosome, to protect the genome from being targeted by DNA repair machinery. In most organisms, telomeres consist of double-stranded (ds) tandem arrays of a short repeat G-rich conserved DNA sequence, $(TTAGGG)_n$, and end with a short single-stranded (ss) 3' overhang (Mceachern et al., 2000). Both the dsDNA and ssDNA repeat nucleotide units recruit an array of telomere-specific DNA-binding proteins that catalyse the formation of a telomere loop (t-loop), to protect the chromosome ends (De Lange, 2004; Griffith et al., 1999). T-loop formation is facilitated by a six telomere-specific protein complex, known as shelterin. Three shelterin proteins, TRF1, TRF2 and POT1 directly bind telomeric TTAGGG repeats and recruit TIN2, TPP1 and RAP1 to form the shelterin complex, that protects chromosome ends by actively changing the architecture of telomeric DNA (De Lange, 2005; Nandakumar & Cech, 2013; Sfeir et al., 2012) (Figure 1.1.2.b.). With every cell division, telomeric ends are truncated by up to 300 bp of DNA. Together with shelterin, the enzyme, telomerase, is required to counteract telomere shortening, thus maintaining the length of telomeres (Cong et al., 2002). However, due to insufficient telomerase expression, telomeres gradually shorten therefore limiting the number of times a cell can divide. The shortening of telomeres is associated with aspects of aging and cancer (Blackburn et al., 2015; Shammas, 2011). During cell division, chromosome architecture is essential for keeping the DNA intact and evenly distributed among cells, to ensure that DNA is accurately copied and correctly distributed during cell divisions.

Studies have shown that the shelterin complex protein, TRF1, plays a role beyond telomere protection. Knockout mouse models abolishing *trf1* in germ cells, results in arrest of spermatocytes at two different stages; the pachytene-like and the meiotic division stages (Karlseder et al., 2003). The meiotic division arrest is due to chromosome end-to-end fusions however, the pachytene-like arrest is due to defective attachment of telomeres to the nuclear envelope (NE) (Karlseder et al., 2003; Shibuya & Watanabe, 2014). These results suggest that TRF1 has dual roles in meiosis and is required for meiotic progression (Wang et al., 2018).

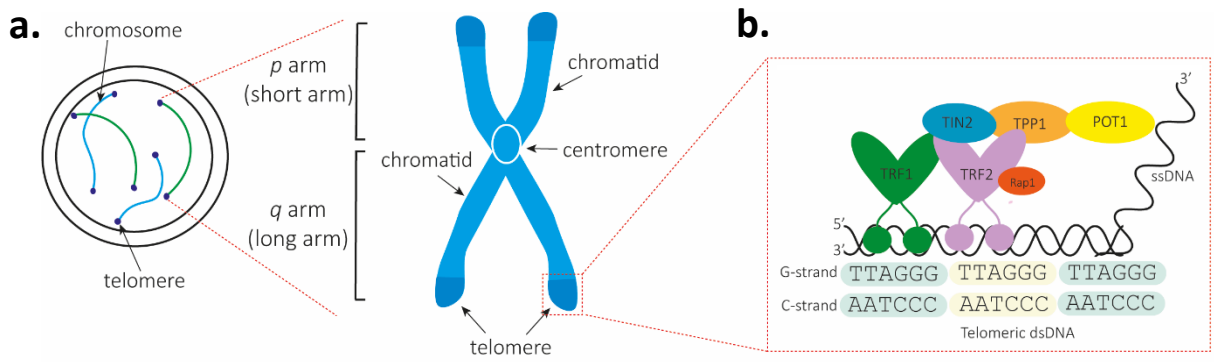


Figure 1.1.2.| DNA architecture and telomere function
Chromosomes formed of sister chromatids, produced during DNA replication, are joined together by centromeres. The short and long arms of the chromatids are called p and q arms, respectively. The telomere is a specialised DNA-protein complex that cap the ends of chromatids preventing them from being processed by the DNA damage response. Telomeric DNA consists of a GC-rich telomeric repeat which is recognised by the nucleoprotein shelterin complex, composed of six proteins: TRF1, TRF2, TIN2, TPP1, POT1 and Rap1. Shelterin is present at telomeres throughout the cell cycle and is crucial for telomere protection.

1.1.3. Meiotic division

Meiosis involves two successive rounds of cellular divisions, meiosis I and meiosis II, resulting in four genetically unique haploid sex cells or gametes. Meiosis I involves a reductional cell division, which segregates replicated homologous chromosomes to produce two haploid cells, thus reducing the chromosome number by half. Analogous to mitosis, meiosis II, is an equational division, in which the sister chromatids are separated to form four haploid daughter cells, each containing a single copy of each chromosome. Similar to the mitosis cell cycle, meiosis I and meiosis II are subdivided into four stages: prophase, metaphase, anaphase and telophase followed by cytokinesis (Alberts et al., 2003; Cooper, 2000).

During interphase, individual chromosomes are not visible but as the cell cycle enters prophase I, cohesin-mediated condensation of chromosomes takes place. In prophase I, spindle apparatus begin to form and homologous chromosomes align next to each other to form bivalents. The two chromosomes in each pair become tightly associated with each other along their lengths and a proteinaceous complex termed the synaptonemal complex (SC) assembles, which is known as synapsis (Fawcett, 1956; Moses, 1956). In addition, homologous chromosomes undergo recombination consisting of crossing over,

resulting in the exchange of genetic material and more importantly the formation of a physical link between two non-sister chromatids, known as chiasmata. This creates unique genetic material, essential for genetic diversity (Pyatnitskaya et al.,2019; von Wettstein, 1984). At the start of metaphase I, microtubules emerge from the spindle and attach to a large protein assembly known as the kinetochore, near the centromere of each chromosome. This facilitates the bivalents to align up along the equator of the cell, termed the metaphase plate (Petronczki et al., 2003). During anaphase I, microtubules disassemble and contract, in turn, pulling homologues apart so that a pair of chromosomes is at each pole. At telophase I, the cytoplasm divides in two, forming two genetically diverse cells, that each contain half the number of chromosomes as the parent cell. In some species, once meiosis I is complete, cells enter a resting phase known as interkinesis before starting the second meiotic division (Alberts et al., 2003).

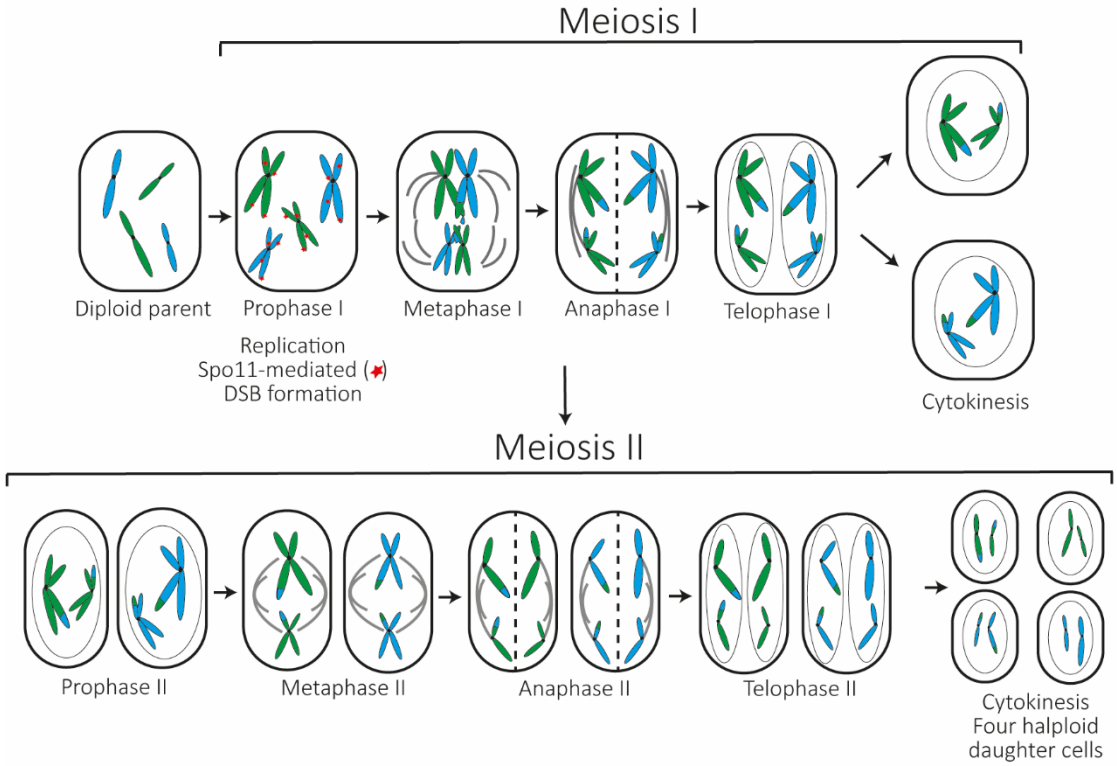


Figure 1.1.3.| Stages of meiotic cell division.
Meiosis involves DNA replication followed by two successive rounds of cell division in which the cell divides two times through meiosis I and meiosis II, to form four haploid daughter cells. In the first meiotic division, the homologues are segregated to separate daughter cells by the spindle apparatus. During prophase I, double strand breaks (DSBs) are introduced in order to initiate homologous recombination, which exchanges genetic material between homologous chromosomes and promotes accurate chromosome segregation at anaphase I. The sister chromatids then undergo a second round of division (Meiosis II) to produce a total of four genetically diverse haploid daughter cells.

During meiosis II, the two cells undergo a second round of division, prophase II, metaphase II, anaphase II and telophase II to produce four genetically unique haploid daughter cells. The stages of meiosis are schematised in Figure 1.1.3.

1.1.4. Meiotic Prophase I

Prophase I is the first and arguably the most important segment of meiosis. Molecular genetics and biochemical studies in yeasts, *C. elegans*, *Drosophila* and mouse have provided many molecular insights into the homologue interactions during prophase I. In order to achieve proper chromosome segregation, homologous chromosomes must interact and generate stable associations, chiasmata, during prophase I, that maintain bivalent configuration until metaphase I. Chiasmata are established by three stepwise processes during prophase I, summarised schematically in figure 1.1.4.:

1. Homologous pairing – the interaction of homologous chromosomes that results in the co-alignment of chromosomes along their entire length (Hunter, 2015; Lee et al., 2015)
2. Synapsis – the formation of the synaptonemal complex (SC) between each homologous pair (Page & Hawley, 2004)
3. Crossing over – the reciprocal exchange of genetic material (Baudat & De Massy, 2007; Pyatnitskaya et al., 2019)

Prophase I involves five distinct substages, based on the appearance of meiotic chromosomes derived from Greek words: leptotene ‘thin threads’, zygotene ‘paired threads’, pachytene ‘thick threads’, diplotene ‘two threads’ and diakinesis ‘across/through motion’. These stages have been observed in many organisms by electron microscopy (Figure 1.1.5.) (Zickler & Kleckner, 2015). Programmed double strand breaks (DSBs) occur during G2/leptotene, once chromosomes begin to condense (Keeney, 2008). Between leptotene and zygotene, each chromosome must search for its homologous partner and become aligned along their lengths, which is mediated by DSB formation (Figure 1.1.5.a.) (Zickler & Kleckner, 1998).

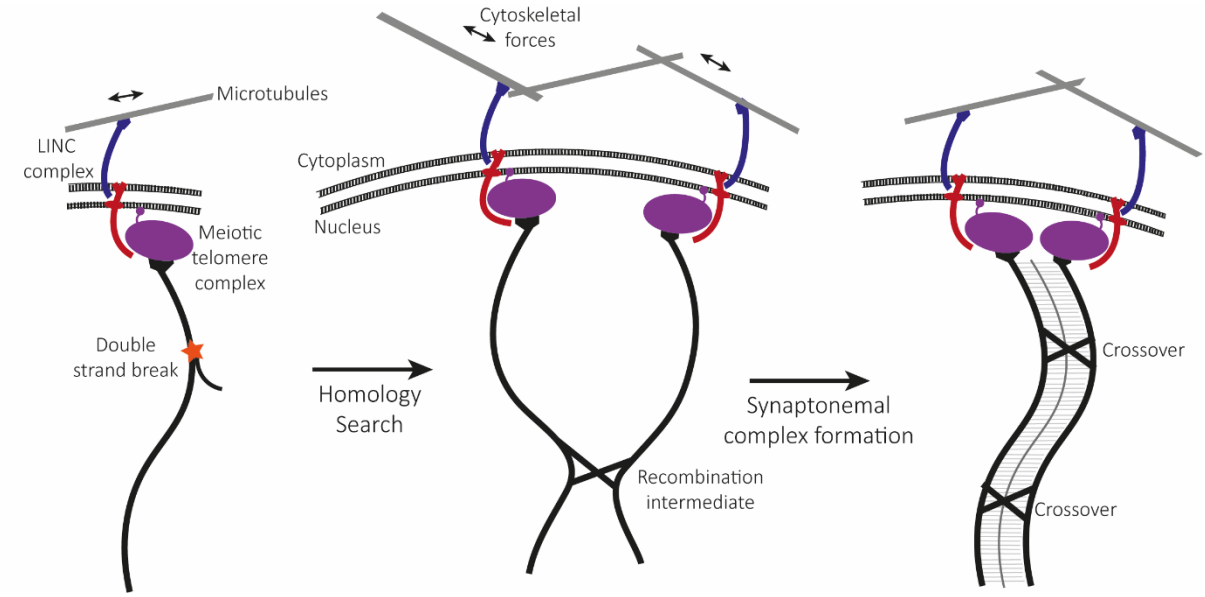


Figure 1.1.4.| Schematic depiction of Prophase I.
During prophase I, chromosomes must search for their homologous partner using a two-dimensional chromosome search. Telomeres are tethered to the inner nuclear membrane (INM) by the meiotic telomere complex and rapidly zip along the membrane using cytoskeletal forces. The transmembrane linker of nucleoskeleton (LINC) complex bridges the connection between the telomeres and the microtubules, thereby transmitting the forces. Chromosome movements facilitates pairing and brings homologous chromosomes together. The homologous pairs then undergo synapsis, a process by which the synaptonemal complex forms along the lengths of aligned chromosomes. DSBs initiate homologous recombination and the crossover pathway to establish chiasmata between homologues.

Extensive homology searching is achieved through a series of highly dynamic telomere-led chromosome choreography and the formation of the chromosomal 'bouquet' (Moiseeva et al., 2017; Zickler, 2006). Zygotene is marked by the initiation of synapsis and the assembly of the proteinaceous SC structure, along the lengths of the paired chromosomes (Figure 1.1.5.b.). By pachytene, SC formation between homologues is complete, forming a mature bivalent (Cahoon & Hawley, 2016; Zickler & Kleckner, 2015). During the zygotene and pachytene stages, a DSB repair mechanism termed homologous recombination occurs, generating crossover (CO) products towards late pachytene (Figure 1.1.5.c.) (Allers & Lichten, 2001; Baudat & De Massy, 2007; Hunter & Kleckner, 2001). Following this, the SC disassembles and homologues become more compact and separate along their lengths. Chiasmata, resulting from successful crossover events, serve to connect the paired homologues and are evident during diplotene and diakinesis (Petronczki et al, 2003; Zickler, & Kleckner, 2014) (Figure 1.1.5.d.).

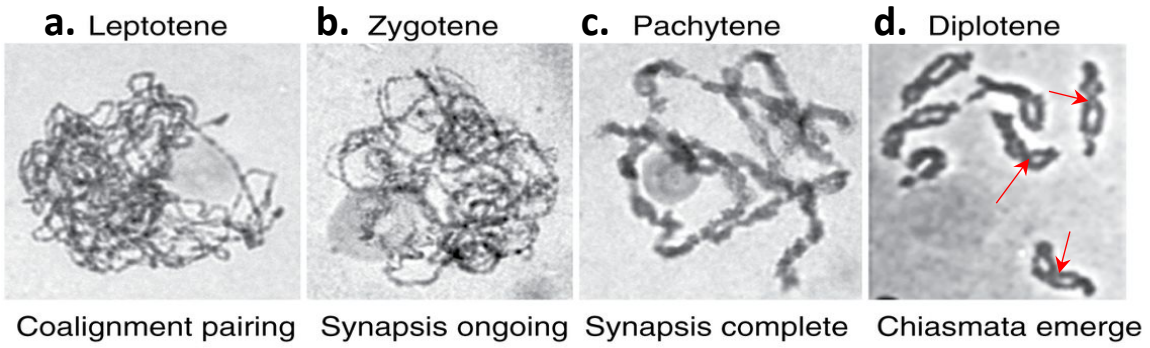


Figure 1.1.5.| Sub-phases of Prophase I.
a) Electron microscopy of surface-spread microsporocytes nuclei of rye, showing the stages of prophase I at the whole chromosome level; chromosomes are stained by hematoxylin. a) During leptotene, DSBs are induced across the genome, which initiate chromosome searches to locate their homologous partner. b-c) During zygotene homologues begin to synapse and fully synapsed chromosomes are visible in pachytene, bound by the synaptonemal complex along the lengths of aligned homologues. c-d) Once fully synapsed, DSB DNA intermediates are processed by homologous recombination, resulting in crossover products. Subsequently, the SC begins to disassemble, and homologous chromosomes are only held together by chiasmata, visible in diplotene (indicated by red arrows in d.). Figure taken from Zickler & Kleckner, 2015.

1.1.5. Importance of cohesins and axial elements (AEs) in meiosis I

In leptotene of prophase I, proteinaceous axial elements (AEs) are established along the sister chromatids of each homologous chromosome. The AE is established to organise sister chromatids into arrays of multiple DNA loops, with their bases anchored on the AEs (Figure 1.1.6.a.). Prior to synapsis, the AEs are required to regulate chromosome compaction into evenly spaced chromatin loops (Page & Hawley, 2004; Zickler & Kleckner, 1999). Subsequently the AE, also called the lateral elements (LEs), serve as the framework to assemble the SC, to create synapsis between aligned homologues (Figure 1.1.6.b.) (Moses, 1956; Westergaard & von Wettstein, 1972). The mammalian AE consists of two protein groups: cohesins and the AE proteins: synaptonemal complex proteins 2 and 3 (SYCP2/SYCP3) (Klein et al., 1999; Kouznetsova et al., 2005). SYCP3 is a major component of the AE and is required for the incorporation of SYCP2. In *Sycp3*$^{-/-}$ spermatocytes, AEs are not formed and SYCP2 fails to localise to chromosome cores, suggesting that SCYP3 is a major determinant of the AE and SYCP2 is a structural component of the AE. Moreover, deletion of either *Sycp2* or *Sycp3* leads to chromatin

elongation and chromatin loops are no longer correctly attached, thus affecting downstream events (Pelttari et al., 2001; Yuan et al., 2000).

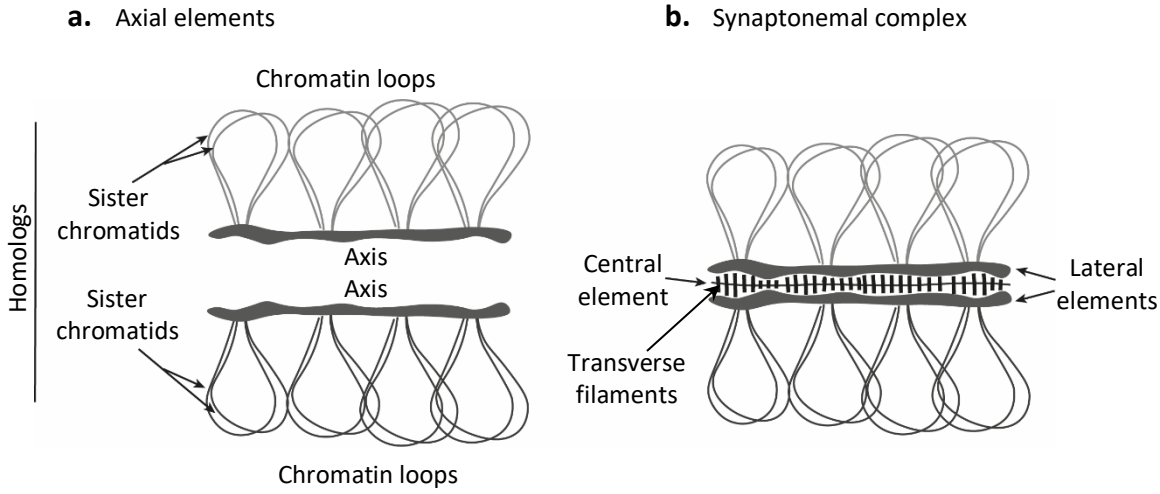


Figure 1.1.6.| Meiotic chromosome organisation.
a) Pairs of sister chromatids are organised in a series of loops (10-20 kb in *S. cerevisiae*) anchored at their bases by a proteinaceous axis called the axial elements (AEs). Axial elements are enriched with several protein components and cohesion complexes. b) At zygotene, homologous chromosomes start to synapse and the homologous AEs are brought together to form the lateral elements (LEs) of the synaptonemal complex (SC), held together by transverse filaments (TFs). The TFs and central element components make up the central region of the SC. By pachytene, the SC is fully formed.

Cohesins are ubiquitously expressed proteins, that hold sister chromatids together with a ring-like structure in both mitosis and meiosis. Cohesins form a tetrameric complex consisting of two structural maintenance of chromosome (SMC) proteins, SMC1 and SMC3, a klesin protein, RAD21 or REC8, and a STAG protein. Meiosis-specific mammalian cohesins have been identified as REC8, RAD21L, SMC1β and STAG3, and are essential structural components of the AEs (Gruber et al., 2003; Gutiérrez-Caballero et al., 2011; Parisi et al., 1999; Prieto et al., 2001). Genetic depletion studies in mice, determined that the cohesin complex is required for homologous pairing and that either klesin proteins, REC8 or RAD21L, are required for the association of SYCP3 and SYCP2 with chromosomes (Llano et al., 2012). These observations suggest that cohesins are crucial for AE formation.

In addition, meiotic cohesins function downstream of programmed DSB formation. It has been proposed that cohesins participate in homologous recombination to ensure recombination occurs between chromatids of homologues (Llano et al., 2012). During leptotene, telomeres become anchored

to the NE to facilitate homologous chromosome pairing through rapid prophase movements (RPMs) (Lee et al., 2015). The meiosis-specific cohesins, STAG3, REC8, and RAD21L localise to the telomere-ends and are proposed to be involved in stable telomere attachment (Ishiguro et al., 2011; Lee & Hirano, 2011; Prieto et al., 2001).

1.1.6. Rapid prophase movements (RPMs) facilitate the pairing of homologous chromosomes

One unique aspect of meiosis is the pairing and exchange of genetic material between homologous chromosomes. However, the mechanism in which homologous chromosomes become paired in early prophase I is still not fully understood. Homologous chromosomes must pair in order to complete the downstream events of synapsis and recombination (Zickler & Kleckner, 1998; Zickler & Kleckner, 2015). In many eukaryotic species, during early prophase I (leptotene to zygotene), chromosome arrangement and oscillation generated by telomere-led chromosome movements promote the pairing of homologous chromosomes (Bhalla & Dernburg, 2008; Chikashige et al., 1994; Lee et al., 2015; Woglar & Jantsch, 2014). Chromosome movements, coined RPMs, are achieved by the physical attachment of chromosomes to the inner nuclear membrane (INM) of the NE by their telomeric ends. Cytoskeletal forces, transmitted from the cytoskeleton, drive these chromosome movements (Lee et al., 2015). The attachment and movement of meiotic telomeres are separately regulated by two complexes, the transmembrane linker of nucleoskeleton and cytoskeleton (LINC) complex and the meiotic telomere tethering complex (MTC), respectively, schematised in figure 1.1.7.a. (Crisp et al., 2006; Shibuya & Watanabe, 2014).

The dramatic telomere-led RPMs subsequently lead to the clustering of telomeres in close proximity at the nuclear periphery, forming a floral-like formation termed a chromosomal ‘bouquet’ (Harper, 2004; Scherthan, 2001). The chromosomal bouquet is ubiquitous among eukaryotes and functions to reduce the spacing between homologues, increasing the occurrence of chromosomal interactions and promoting homologous pairing (Figure 1.1.7.b.-d.) (Berríos et al., 2014; Golubovskaya et al., 2002; Kezer et al., 1989). Bouquet formation is not a required step for homologous pairing and synapsis,

instead acts as a catalyst to speed up the process by positioning one or both ends of the chromosomes in a limited sector of the nuclear periphery (Lee et al., 2012). The rate of RPMs differs throughout the stages of prophase, during leptotene and zygotene RPMs peak, allowing for homology search and bouquet formation. Upon bouquet formation the rate of RPMs is reduced, suggesting that pairing has successfully taken place (Lee et al., 2015; Shibuya et al., 2014). Although the movement of chromosomes is dynamic, the relative positions of the chromosomes are mostly restrained suggesting that chromosomes are rotating randomly but unidirectionally, therefore minimising the search for homologous chromosomes (Lee et al., 2015).

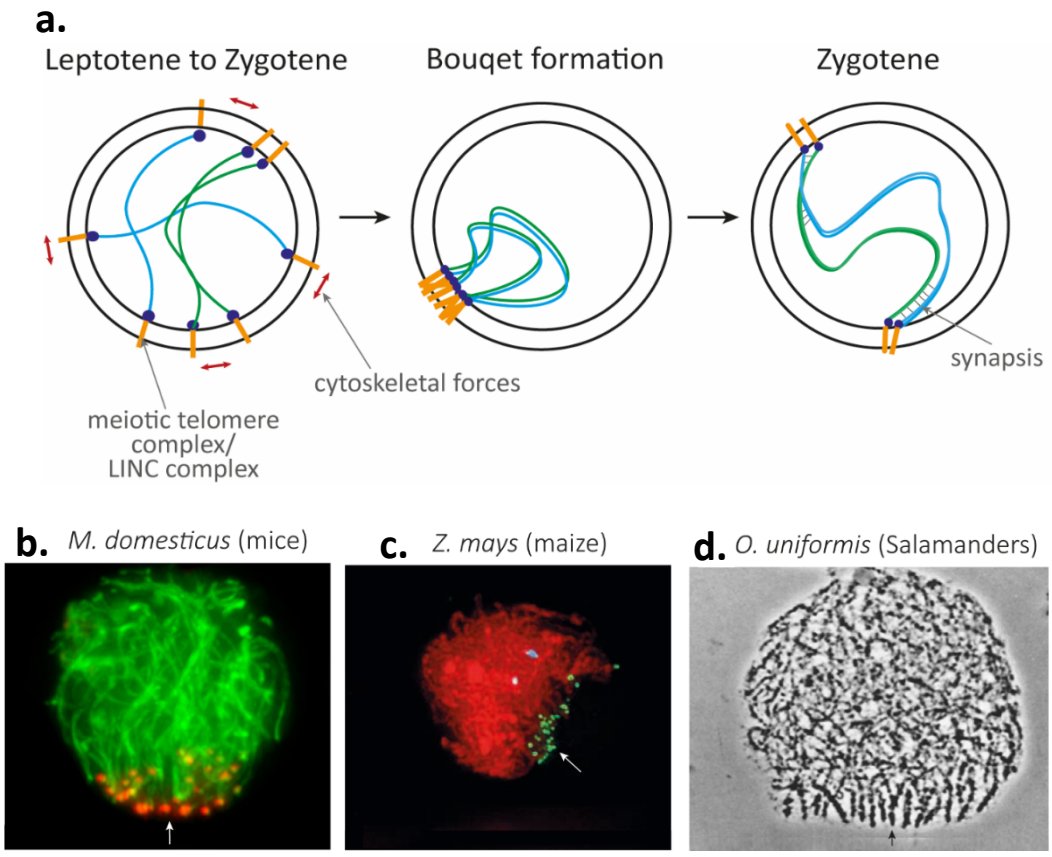


Figure 1.1.7.| The meiotic 'bouquet'.
a) Schematic representation of bouquet formation in zygotene stage of prophase I. During early meiosis chromosomes become attached to the nuclear envelope by their telomeres through interaction with the meiotic telomere complex and the LINC complex proteins. This attachment facilitates RPMs by cytoskeletal forces, allowing for homologous chromosomes to migrate and cluster together forming the chromosomal 'bouquet'. Synapsis takes place between fully aligned paired homologues. b-c) Chromosomal bouquet formation is a conserved meiotic event occurring at zygotene, in which dramatic chromosomal movements leads to the clustering of telomeric chromosome ends towards one nuclear pole. Arrows show the meiotic bouquet in b) mouse spermatocytes (Berríos et al., 2014), c) maize (Golubovskaya et al., 2002) and d) Salamanders (Kezer et al., 1989).

1.1.7. The meiotic linker of nucleoskeleton and cytoskeleton (LINC) complex

RPMs are dependent on the LINC complex proteins (Lee et al., 2015). The LINC complex is a ubiquitously conserved assembly, composed of two proteins: SUN (Sad-1/Unc-84) and KASH (Klarsichr/ANC-1/Syne/homology) (Zhang et al., 2009). In meiosis, during prophase I, the LINC complex functions to span the nuclear membrane and bridge telomeres to the cytoskeleton. Thus, it provides the molecular connections to the cytoskeletal components that originate the forces to generate RPMs of chromosomes (Hiraoka & Dernburg, 2009; Kracklauer et al., 2013). Studies in mice have shown that the dynein movement along microtubules is transmitted to telomeres through the LINC complex protein components SUN1 and the meiosis-specific KASH5, that are localised to the INM and outer nuclear membrane (ONM) of the NE, respectively (Horn et al., 2013).

Localisation studies have shown that SUN1 associates with the telomere marker, CREST, at the INM. The SUN domain of SUN1 spans the INM and the C-terminus interacts with the KASH domain of KASH5 in the perinuclear space (Figure 1.1.8.a.) (Hiraoka & Dernburg, 2009; Lee et al., 2015). The KASH5 N-terminus interacts with the cytoplasmic motors, dynein-dynactin, at the ONM, which mediates microtubule binding (Figure 1.1.8.b.) (Lee et al., 2015). Together SUN1 and KASH5 bridge the internal regions of the NE and cytoskeleton, in turn mediating the transmission of cytoskeletal forces across the NE to enable chromosome movement, schematised in figure 1.1.8.c (Lee et al., 2015; Morimoto et al., 2012). Studies have shown that deletion of either *Sun1* or *Kash5* leads to the disruption of chromosome synapsis, revealing that telomere attachment to the NE is insufficient to promote chromosome pairing and that the process of telomere attachment must be coupled to cytoplasmic forces via the LINC complex (Horn et al., 2013; Link et al., 2014).

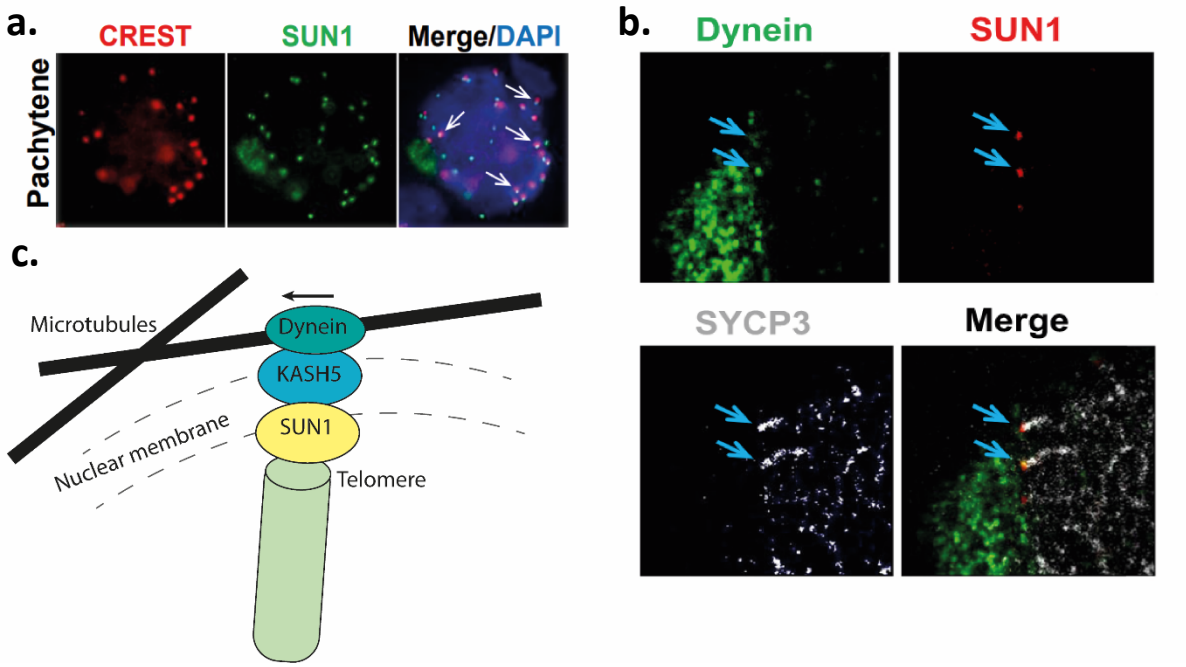


Figure 1.1.8.| Visualisation of the LINC complex coupling dynein and telomeres
a-b) Visualisation of structurally preserved spermatocytes by confocal microscopy shows the association of telomeres to cytoskeletal components (dynein) via SUN1-KASH5 nuclear envelope bridges. a) Colocalisation of telomere ends (marked by CREST; red) and SUN1 (green) at the nuclear periphery during pachytene of prophase I. Arrows indicated examples of CREST/SUN1 colocalisation. b) Colocalisation of dynein (green) and SUN1 (red) demonstrates that SUN1-KASH5-dynein connect the ends of the telomere at the INM to microtubules in the cytoplasmic space. SYCP3 used to mark chromosome cores. c) Schematic representation of the LINC complex components KASH5 and SUN1 spanning the nuclear membrane and creating a physical link between dynein on microtubules and NE-attached telomeres. Microscopy images taken from Lee et al., 2015

1.1.8. The meiotic telomere complex (MTC)

In order for chromosomes to carry out RPMs and achieve homologous pairing, telomeres must attach to the NE (Lee et al., 2012). Telomere attachment is mediated by meiosis-specific telomere adapter proteins, Bqt1/Bqt2 (*S. pombe*), Ndj1 (*S. cerevisiae*) and TERB1 (mammals) (Chikashige et al., 2006; Daniel et al., 2014; Trelles-Sticken et al., 2000). Sequence analysis revealed no similarities between these proteins, questioning whether the telomere attachment mechanisms are conserved across species. Electron microscopy (EM) studies in higher eukaryotes have revealed a conserved dense structure at the INM where chromosomes attach, termed the telomere attachment plate, suggesting the possibility of a conserved telomere attachment complex (Figure 1.1.9.a.) (Shibuya et al., 2015).

In mice, the tripartite MTC is a multi-subunit DNA-binding complex, composed of telomere repeats binding bouquet formation proteins 1/2 (TERB1, TERB2) and membrane-anchored junction protein (MAJIN) (Daniel et al., 2014; Shibuya et al., 2015; Shibuya & Watanabe, 2014). In addition to the LINC complex, the MTC functions to establish a second physical linkage for telomere attachment to the NE (Figure 1.1.9.b.). Individual knockouts of each MTC protein component led to meiotic arrest in both male and female mouse models, indicating that all three proteins are essential for fertility in mice (Shibuya & Watanabe, 2014).

Studies have mapped the binary interactions between the mammalian MTC proteins by yeast two-hybrid screens (Y2H), which revealed the connection between the MTC and the shelterin complex through a TERB1-TRF1 interaction (Daniel et al., 2014; Shibuya et al., 2014; Shibuya et al., 2015). Moreover, one Y2H study determined an interaction between TERB1 and the LINC complex protein SUN1 (Shibuya et al., 2015). This suggests that TERB1 functions as a molecular scaffold, simultaneously interacting with TERB2, SUN1 and TRF1. MAJIN and TERB1 are physically linked by TERB2, which binds MAJIN and TERB1 through its C- and N-terminus, respectively. MAJIN is a putative transmembrane (TM) protein, localised at the inner surface of the NE, that recruits TERB2 and TERB1 to the INM (Shibuya et al., 2015). In turn, localises telomeres to the INM through TERB1-TRF1. Moreover, MAJIN possesses a DNA binding domain, suggesting it may also function to stabilise telomere attachment (Dunce, Milburn et al., 2018). These observations suggest that the MTC repurposes and integrates the functions of the LINC and shelterin complexes, in order to achieve stable chromosome attachment and facilitate RPMs during prophase I, further explained in Chapter 3.

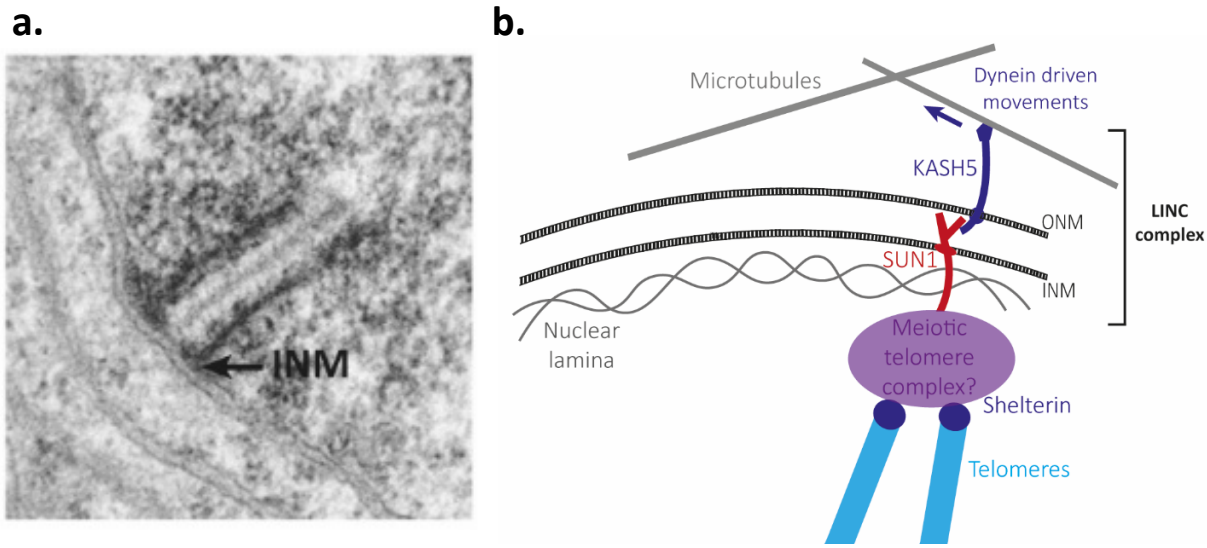


Figure 1.1.9.| The meiotic telomere complex is the link between shelterin and the LINC complexes
a) Electron micrograph image of the telomere attachment plate in mice spermatocytes at the INM. At the site of chromosome attachment to the INM there is dense thickening suggestive of the meiotic telomere complex. b) Schematic of the mammalian attachment plate in which the LINC complex (SUN1/KASH5) and the meiotic telomere complex bridge the telomeres (shelterin) to the microtubules,forces through driven movements. EM image from Shibuya et al., 2015.

1.1.9. The synaptonemal complex (SC): A universal meiotic structure

Pairing of homologous chromosomes leads to the assembly of the SC, that provides continuous synapsis between homologous chromosome axes and ensures the process of homologous recombination (Dunce et al., 2018; Page & Hawley, 2004; Westergaard & von Wettstein, 1972). The SC is a meiosis specific supramolecular protein lattice, found in almost all sexually reproducing organisms, with the exception of male fruit flies (Costa & Cooke, 2007; Fraune et al., 2012; C. Lake & Hawley, 2012; Page & Hawley, 2004). Early transmission electron microscopy (TEM) experiments demonstrate that the SC has a conserved ladder-like organisation comprised of two parallel lateral elements (LEs) which the chromatin of homologous chromosomes is attached (Fawcett, 1956; Moses, 1956; Westergaard & von Wettstein, 1972). The LEs flank the central region. The central region is composed of a midline 20-40 nm wide central element (CE) and numerous transverse filaments (TFs) that lie perpendicular to LEs. The TFs bridge the space between the CE and LEs, providing ~100 nm separation between LEs to form a zipper-like structure (Figure 1.1.10.a.) (Westergaard & von Wettstein, 1972). Thus, the SC functions

to provide continuous synapsis along the entire length of the aligned homologous chromosomes during zygotene and becomes fully assembled by pachytene (Solari & Moses, 1973).

SC protein components have been identified using immunolocalisation techniques, co-immunoprecipitation (co-IP) and genetic analyses, and the interactions between proteins have been mapped. Despite the tripartite structure being conserved across species, there is no obvious sequence similarity between SC components. In mammals, so far, eight protein components have been identified: LE proteins, SYCP2 (Yang et al., 2006) and SYCP3 (Yuan et al., 2000), the TF protein SYCP1 (Dunce et al., 2018; Meuwissen et al., 1992) and the CE proteins SYCE3 (Schramm et al., 2011), SYCE1 (Bolcun-Filas et al., 2009; Dunne & Davies, 2019), SIX6OS1 (Gómez et al., 2016), SYCE2 (Bolcun-Filas et al., 2007) and TEX12 (Davies et al., 2012; Hamer et al., 2006), schematised in Figure 1.1.10.b.

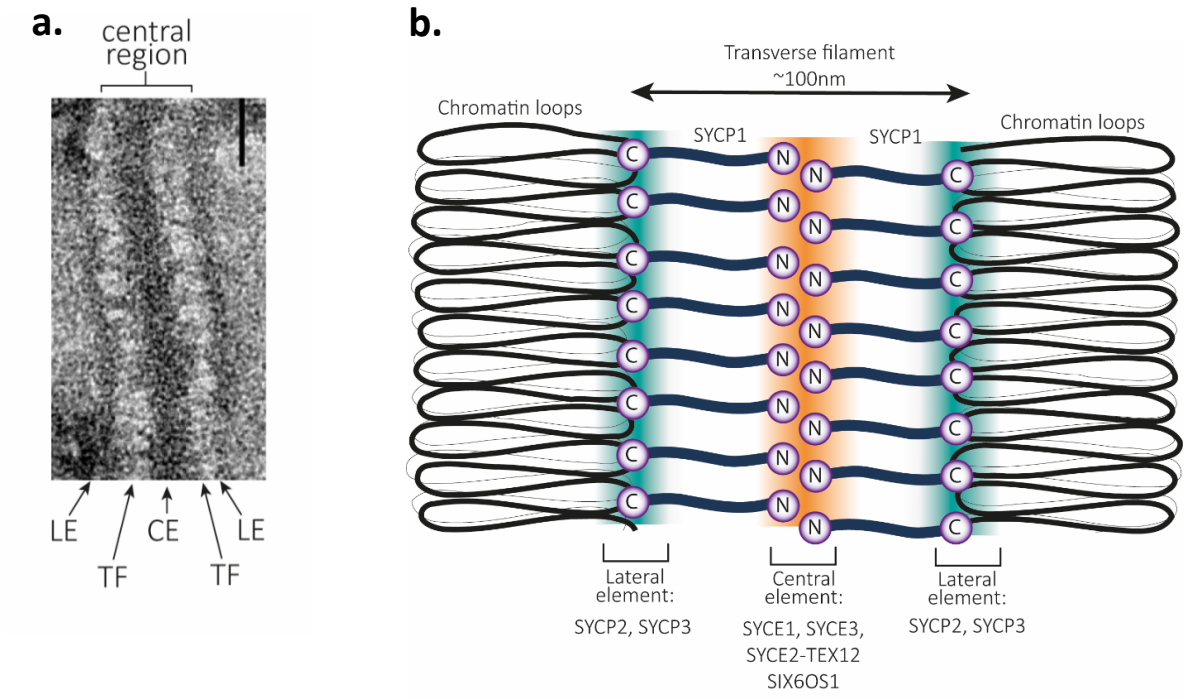


Figure 1.1.10.| Mammalian synaptonemal complex (SC).
a) Electron micrograph of mouse SC conveying the conserved tripartite structure of two chromosome-bound lateral elements that flank a midline central element and connected by a series of transverse filaments. Scale bar, 100 nm b) Schematic of the mammalian SC depicting the chromatin loops of paired homologous chromosomes with the assembly of the lateral elements, SYCP2 and SYCP3 (blue) at the axes. The transverse filament protein, SYCP1, span across the central region and bridge the lateral and central elements. The central element is comprised of SYCE1, SYCE3, SYCE2-TEX12, and SIX6OS1. Electron micrograph from Shramm et al., 2011.

Assembly and disassembly of the SC are carefully timed events within meiotic prophase I. SC assembly is initiated by the Spo11-induced DSBs and starts by the formation of AEs along each pair of homologous sister chromatids, which later become the so-called LEs, SYCP2 and SYCP3 (Keeney, 2008; Solari & Moses, 1973; Yang et al., 2006; Yuan et al., 2000). Subsequently, the paired LEs are joined together by a single TF protein, SYCP1, which undergoes self-assembly to form a supramolecular lattice, consisting of at least two layers of SYCP1 molecules that span the central region (Dunce et al., 2018; Hernández-Hernández et al., 2016; Schücker et al., 2015). The SYCP1 lattice provides the core architecture of the SC but requires the CE and LE proteins for stabilisation and chromosome structure, respectively (Lu et al., 2014; Winkel et al., 2009). CE proteins SYCE1, SYCE3 and SIX6OS1 are proposed to stabilise initial SYCP1 synapsis by providing vertical and transverse links between SYCP1 molecules (Dunce et al., 2018). In addition, CE proteins SYCE2 and TEX12 form a stable constitutive complex that spontaneously assemble into filamentous structures (Davies et al., 2012). SYCE2-TEX12 provides a longitudinal scaffold for the long-range extension of SYCP1 synapsis, ensuring continuous synapsis between homologous chromosomes (Davies et al., 2012; Dunce et al., 2020). The SC structure terminates at synapsed telomere ends through the attachment plate that are physically fused to the INM (Figure 1.1.9.a.) (Alsheimer et al., 1999). Complete SC formation is essential for successful meiotic progression and errors in SC assembly has been associated with infertility, miscarriage and aneuploidy (Hassold & Hunt, 2001).

1.1.10. The SC and homologous recombination

A universal feature of meiosis is the formation of numerous Spo11-mediated DSBs across cells, in order to initiate the DNA-repair process, homologous recombination (Gray & Cohen, 2016; Hunter, 2015; Keeney, 2008). The process of homologous recombination results in the two possible outcomes, the reciprocal or non-reciprocal exchange of genetic material via the crossover (CO) or non-crossover (NCO) pathways, respectively (Guillon et al., 2005).

Despite a high number (~200-300 in mice) of meiotic DSBs being introduced, the majority of these are destined to become NCO products, sometimes termed gene conversions, which do not result in a physical link between homologues. Only around 10% (~22 ± 3) of the DSBs undergo designation/selection to form COs (Gray & Cohen, 2016; Turner, 2007). This concept, termed CO assurance, suggests that the process of CO formation is tightly regulated, both temporally and spatially, resulting in non-randomly distributed COs, and typically one CO per chromosome arm (Broman et al., 2002; Shinohara et al., 2008). COs involve the reciprocal exchange of genetic material between homologous chromosomes, thus generating genetic variation. More importantly, COs provide the sole physical connections between homologues following the disassembly of the SC. These connections, termed chiasmata, in addition to sister chromatid cohesion, allows chromosomes to correctly align on the spindle and accurately segregate at the first meiotic division (Petronczki et al., 2003; Rasmussen & Holm, 1984).

In most organisms, synapsis and homologous recombination are interdependent, indicating that the SC plays a key role in the recombination process (Qiao et al., 2012). In *S. cerevisiae*, mice and *Arabidopsis*, SC formation often nucleates at points of DSBs, suggesting that homologous recombination promotes SC formation (Henderson & Keeney, 2004). Moreover, the SC is required for crossing over and subsequent accurate segregation of homologues (Dunne & Davies, 2019). The SC zipper-like structure physically tethers homologous chromosomes as well as facilitating the resolution of recombination intermediates (Zickler & Kleckner, 2015). Studies have shown that in mammals, deletion of CE proteins eliminates CO formation, demonstrating the SC plays a critical role in the CO pathway (Bolcun-Filas et al., 2009; Gómez-H et al., 2016; Sánchez-Sáez et al., 2020; Schramm et al., 2011; Sym & Roeder, 1994). However, in some organisms including *Drosophila* and *C. elegans*, the SC is formed in the absence of homologous recombination, suggesting that the temporal order of events in these organisms may differ (Page & Hawley, 2001; Roeder, 1997).

1.1.11. Meiotic recombination pathway: mechanism and regulation

Meiotic recombination is initiated by the introduction of numerous programmed DSBs per cell by the evolutionary conserved endonuclease, Spo11, in conjunction with additional proteins, during leptotene of prophase I (Keeney, 2008; Keeney et al., 1997; Robert et al., 2016). This can be visualized by H2AX phosphorylation (γH2AX) signals distributed throughout the nucleus of mice spermatocytes (Figure 1.1.11.a.) (Hunter, 2015). A pair of Spo11 monomers act in concert to cut both DNA strands via a topoisomerase-like reaction to generate covalent protein-DNA linkages to the 5' DNA ends either side of the break (Keeney, 2008). Once formed, DSBs are subject to rapid exonucleolytic resection of the 5' strands to produce 3' ssDNA tails of ~ 600 nucleotides. DNA end-resection is mediated by several nucleases including the MRN complex, CtIP, EXO1, BLM helicase, DNA2, and RPA (Davies et al., 2015; Nimonkar et al., 2011). The ssDNA extensions are coated with RPA to form a nucleoprotein filament capable of locating DNA homology within chromosome and mediating strand invasion, in which it invades the complementary sequence on a sister chromatid to form a DNA joint intermediate called a displacement-loop (D-loop) (Ma et al., 2017; Szostak et al., 1983). In most eukaryotes strand invasion requires two recombinases, Rad51 and Dmc1, which together mediate the search for homologous sequences (Figure 1.1.11.b) (Brooks et al., 2018; Moens et al., 1997).

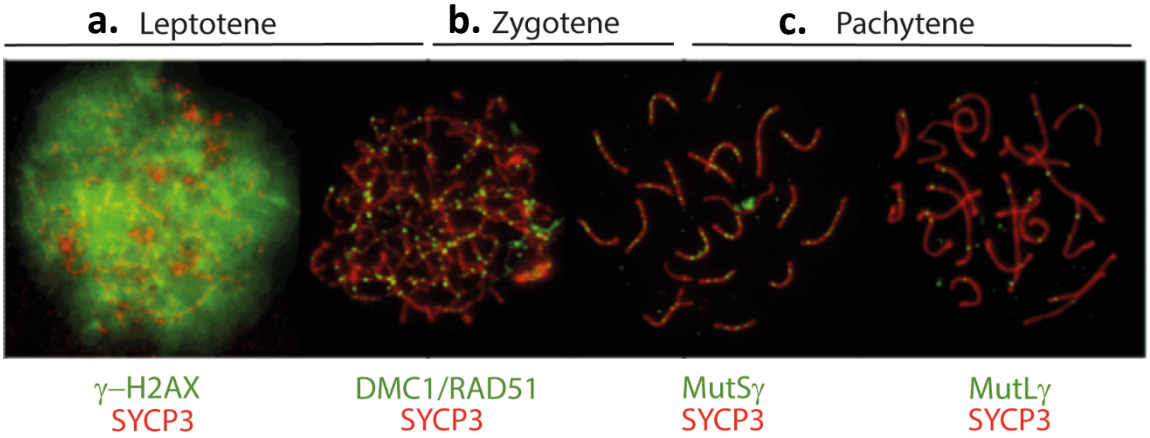


Figure 1.1.11.| Cytological markers of meiotic recombination in mouse spermatocyte. Super-resolution microscopy of mouse spermatocyte nuclei immunostained for homologue axis component SYCP3 (red) and other recombination markers throughout prophase I. a) At leptotene, DSBs are introduced throughout the nuclei, shown by γH2AX foci (green). b) Zygotene spermatocytes show a high number of DMC1/RAD51 foci (green) associated with the chromosome cores (SYCP3 – red). b-c) During zygotene-pachytene DMC1/RAD51 dissociate from CO sites and the ZMM protein including MSH4/MSH5 (MutSγ) (green) complex associate to a subset of DSB intermediates. c) At pachytene, the mutLγ (Mlh1/Mlh3) resolvase complex is recruited to dHj intermediates destined for COs. Figure adapted from Hunter, 2015.

D-loop strand invasion intermediates can be further processed in different ways to produce either CO or NCO products. The differentiation between these pathways takes place early in the process of recombination and is dependent upon a multitude of proteins (Pyatnitskaya et al., 2019). The NCO pathway involves the synthesis-dependent strand annealing (SDSA) mechanism formed by annealing of the extended single-stranded end to the ssDNA on the other end, followed by gap-filling DNA synthesis and ligation (Allers & Lichten, 2001). Alternatively, strand invasion intermediates can be channelled to form CO products. In *S. cerevisiae* and mammals there are two distinct classes of COs: interfering (Class I) and non-interfering (Class II). However, the majority of COs (90-95%) are formed by interference leading to the non-random placement of COs resulting in the even distribution of chiasmata across chromosomes (Shinohara et al., 2008). The alternative CO pathway (class II) is only observed in a distinct subset of COs and, unlike the class I pathway, exhibits no interference. This pathway is dependent upon MUS81/MMS4-based resolvase complex (De los Santos et al., 2003).

In the class I pathway the D-loop strand-invasion intermediate is stabilised by second end capture that primes DNA synthesis and ligation, generating a more stable double Holliday junction (dHj) joint molecule intermediate (Hunter & Kleckner, 2001). Resolution of dHj intermediates can either yield NCO or CO products, dependent on the orientation of each Holliday junction. CO formation leads to the exchange of genetic material between non-sister chromatids of homologous chromosomes. For most organisms, the class I CO pathway is mediated by orthologues of the *S. cerevisiae* ZMM proteins: Zip1, Zip2, Zip3, Mer3, Msh4 and Msh5, which includes the MutSγ (Msh4/Msh5) DNA mismatch repair (MMR) complex (Börner et al., 2004; Pyatnitskaya et al., 2019; Shinohara et al., 2008). The ZMM proteins, with the exception of Zip1, localise to designated CO sites and function collaboratively to promote crossover assurance and interference (Shinohara et al., 2008). *S. cerevisiae* strains carrying single mutations in these genes, have a reduction in COs and defects in the formation of the SC (Pyatnitskaya et al., 2019). However, NCOs are not affected indicating they are independent pathways. Moreover, the MutLγ (Mlh1/Mlh3) complex, which works downstream of the ZMM proteins, is required to resolve dHjs into CO products (Figure 1.1.11.c.) (Santucci-Darmanin et al., 2000; Snowden et al., 2004). The meiotic recombination pathway is summarised schematically in figure 1.1.12.

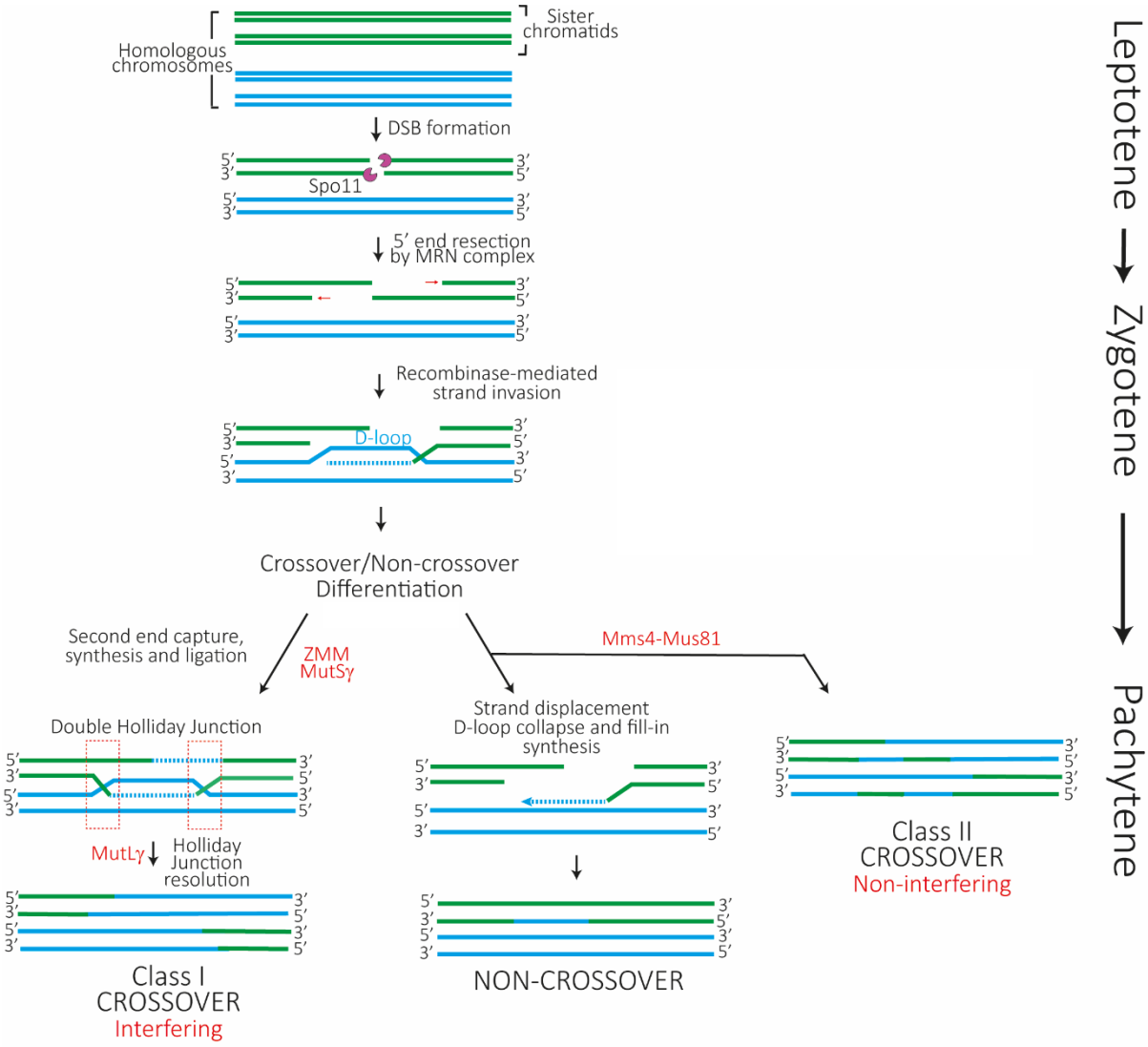


Figure 1.1.12.| Schematic of the meiotic recombination pathway.
Spo11 enzyme introduces hundreds of DSBs throughout the genome. 5' end resection and strand invasion process the DSB to form a D-loop. D-loop intermediates can be repaired into three possible outcomes: non-crossover, crossover (interfering) and crossover (non-interfering), which are formed from the middle to late pachytene. Interfering CO pathways involves the formation and resolution of a double-Holliday junction by a subset of ZMM (MutSγ) and MutLγ complex proteins. In contrast, synthesis-dependent strand annealing leads to non-crossover recombination products. Non-interfering crossovers are formed by Mms4-Mus81 endonucleases.

In yeast, the commitment towards the class I CO pathway is defined soon after DSB formation during leptotene/zygotene of prophase I, indicating that CO formation is temporally and spatially regulated (Zickler & Kleckner, 1998). How the fate of DSBs is controlled is an intricate process, which is still not fully understood. It has been proposed that the ZMM proteins play a role beyond stabilising and priming strand invasion CO precursors (Agarwal & Roeder, 2000). Moreover, studies have observed

that only a subset of DNA joint molecules, marked by the MutSγ complex, recruit the downstream MutLγ complex, required for dHj resolution, suggesting that additional processing is required (Santucci-Darmanin et al., 2002; Zakharyevich et al., 2012).

Many DNA repair processes use post-translational modifications (PTMs), including phosphorylation, ubiquitinoylation, and SUMOylation to modulate enzymatic activities and regulate protein stability, localisation and interactions (Bai et al., 2020; Psakhye & Jentsch, 2012). In addition, the ubiquitination pathway plays a central role in protein degradation in eukaryotic cell division (Kerscher et al., 2006). This may be the case during meiotic recombination, since SUMOylation/ubiquitination has been observed along chromosome axes during prophase I in yeast, *C. elegans* and mammals (Nottke et al., 2017). The yeast ZMM complex includes Zip3, an E3 ligase protein, that has shown to be involved in SUMOylation/ubiquitination pathways during recombination to control CO formation (Agarwal & Roeder, 2000). Moreover, a number of studies in mouse have shown that the Zip3-like proteins, RNF212 and HEI10, are also RING-domain E3 ligases involved in SUMOylation and ubiquitination, respectively, and are both essential for crossover recombination (Gray & Cohen, 2016; Prasada Rao et al., 2017; Qiao et al., 2014; Reynolds et al., 2013; Ward et al., 2007). In male mouse spermatocytes, RNF212 has been shown to colocalise with the majority of MutSγ foci and involved in D-loop stabilisation. As observed for MutSγ foci, the number of RNF212 foci also decreases through pachytene, thereby is inferred to catalyse SUMO conjugation and promotes selective stabilisation of MutSγ to only a subset of potential CO sites (Reynolds et al., 2013). In contrast, the ubiquitin-ligase HEI10 marks CO-sites during mid-/late-pachytene and co-localises with Mlh3 (MutLγ complex), suggesting that the ubiquitin activity of HEI10 promotes the turnover of recombination factors and promotes CO formation, schematised in Figure 1.1.13 and further explained in Chapter 5 (Qiao et al., 2014; Reynolds et al., 2013). Together, these observations suggest that meiotic recombination is tightly regulated and involves highly coordinated events, of which some are still not completely understood.

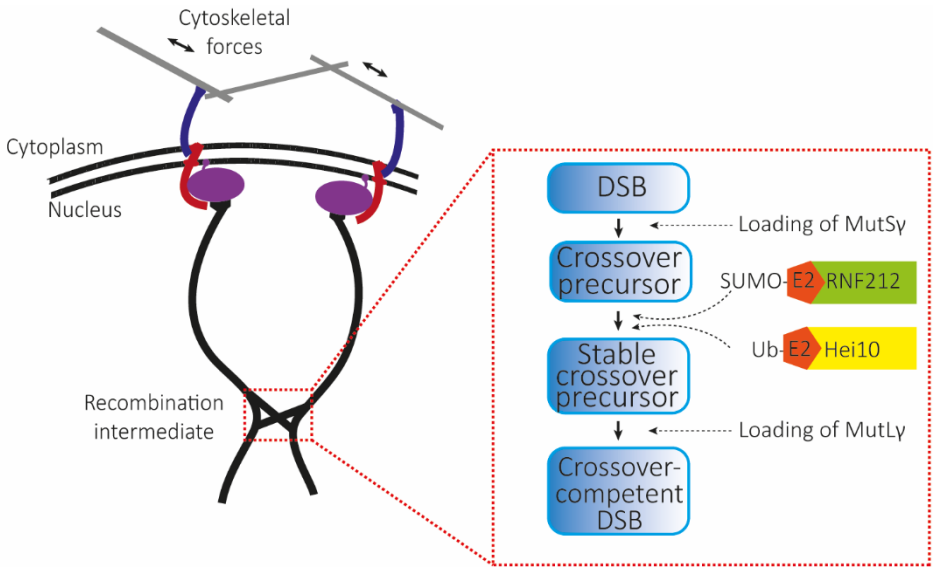


Figure 1.1.13.| E3 ligase proteins RNF212 and HEI10 are pro-CO factors. The formation of interfering COs is highly regulated process involving numerous proteins and protein complexes, including MutSγ and MutLγ complexes. Mammalian E3 ligase proteins, RNF212 and HEI10 are essential for crossover formation through the involvement of SUMOylation and ubiquitination pathways, respectively.

1.1.12. Defective recombination and infertility

Meiosis is a highly regulated sequence of events, orchestrated by a multitude of proteins and controlled by a number of checkpoints. During meiosis, genetic recombination between homologous chromosomes is essential for successful chromosome segregation at meiosis I. The spindle assembly checkpoint (SAC) is the guardian of faithful chromosome segregation (Lane & Kauppi, 2019). Meiosis II resembles mitotic divisions, where the sister chromatids separate to produce four genetically diverse haploid gametes (sex cells). Despite the regulatory checkpoint mechanisms during meiosis I and II, human reproduction does accumulate errors and ~15 % of couples are affected by infertility and 1 in 4 pregnancies end in miscarriage (Sharma et al., 2013). Meiosis I in human females can be error prone and compromised SAC function in meiosis I can lead to the formation of aneuploid (trisomy or monosomy) gametes and in most cases often results in the inability to achieve pregnancy, early spontaneous abortion, or congenital defects (Hassold & Hunt, 2001). A third of all miscarriages are caused by fatal aneuploidy. Moreover, aneuploidy is also the leading single cause in congenital birth defects (Nagaoka et al, 2012). In the female germline, two key factors, suboptimal crossing over and

advancing maternal age, have been identified as main causes of aneuploidy. Aneuploidy arises from a two-step process consisting of chromosome mis-segregation followed by inefficient SAC control. Klinefelter syndrome (trisomy - XXY) and Turner syndrome (monosomy - Xo) are two common sex chromosome aneuploidies in live-born individuals (Herbert et al., 2015; Nagaoka et al., 2012).

Crossover events between homologous chromatids are required for the exchange of genetic material but also essential to form chiasmata. Failure to form chiasmata results in homologues remaining unconnected and cannot attain a stable bipolar orientation on the meiosis I spindle. Furthermore, the failure to separate the pairs of daughter cells, chromosome nondisjunction, is a major cause of aneuploidy. The frequency of these events have shown to increase with maternal age and specifically more crossovers are detected in children born to older mothers (Bomblies et al., 2016; Herbert et al., 2015).

In addition, meiotic recombination represents a huge mutational target, and studies have shown that many variants of recombination genes lead to infertility (Hunter, 2015). In humans, it is difficult to establish the roles of specific recombination proteins due to this detrimental effect. Therefore, it is important to study these proteins *in vitro* to determine biophysical and biochemical properties, in addition to their interaction network, to deduce their mechanisms of action during homologous recombination. Moreover, by studying homologous proteins in the model organisms *D. melanogaster* and *S. cerevisiae*, and subjecting them to functional , structural, and genetic analyses, this will allow us to identify the biological role of gene products within the context of a whole organism and provide much needed insights which can also be applied to better understand human meiosis.

1.1.13. Thesis aims and objectives

Over the past half century substantial advances have been made in the field of meiosis, unravelling mechanisms, and characterising the proteins involved. Despite this, our understanding in some of the most fundamental aspects of meiosis remain unclear. In particular, many questions remain unanswered in the mechanism of homologous chromosome recognition and pairing, as well as the crossover

pathway. In the last decade, our understanding of the molecular aspects of meiosis, including the mammalian synaptonemal complex, have been widened by structural biology techniques, such as X-ray crystallography. The thesis aim was to adopt a structural biology approach using a range of biophysical techniques, including X-ray crystallography, SEC-MALS and SEC-SAXS, to further our understanding of the structure-function relationship of proteins and protein complexes involved in meiosis.

In Chapter 3, we used a plethora of structural and biochemical techniques to gain a detailed understanding of the meiotic telomere complex proteins, TERB1, TERB2, MAJIN, and the shelterin complex protein TRF1 in order to elucidate how telomere tethering to the nuclear envelope is achieved. Ultimately, we uncovered the mechanism by which telomere tethering is achieved through interplay between the shelterin complex, the meiotic telomere complex and the LINC complex proteins.

Many aspects of meiosis show strong evolutionary conservation, and the stages of meiosis have been studied in a variety of model organisms including *S. cerevisiae*, *C. elegans* and *D. melanogaster*, to provide new insights into mechanisms and regulations during prophase I. Due to the lack of a genetically-tractable cellular system for mammalian meiosis, *in vivo* analysis of the mammalian SC assembly has not been fully explored. Therefore, it is important to study meiosis in model systems to determine structural similarities and assembly mechanisms, which can then be applied to the mammalian system.

Our current understanding of the structure and biochemical properties of individual proteins involved in the assembly of the *D. melanogaster* synaptonemal complex remains limited. The aim of Chapter 4 was to to uncover the molecular structure and provide insights into the assembly mechanism of the *D. melanogaster* SC, by in-depth structural characterisation of the transverse filament protein, C(3)G. Despite little sequence conservation, the *D. melanogaster* TF protein, C(3)G, has a predicted secondary structure resembling SYCP1. Therefore, by carrying out a structural study of the *Drosophila* SC, I determined an understanding into the underlying conservation.

A recent study, using *in vivo* techniques, has shown removing parts of the predicted coiled-coil domain of C(3)G results in partial loss of SC function at different stages in early meiosis (Billmyre et al., 2019). Using a combination of biophysical techniques I analysed three in-frame deletion mutations of C(3)G to show how the in-frame deletion mutants affect the structural and biophysical properties of C(3)G.

Chapter 5 focused on the mammalian E3 ligase proteins, HEI10, RNF212 and RNF212b, which are essential for crossover designation and formation. It has been suggested that these proteins have roles in ensuring crossovers through ubiquitination and SUMOylation pathways, respectively (Prasada Rao et al., 2017). However, the structures of these proteins have yet to be characterised. I carried out a detailed characterisation of the mammalian RING-domain E3 ligase proteins using both biophysical and structural techniques. I tested for interactions between the E3 ligase proteins to provide mechanistic details into how the E3 ligase proteins may assemble and function within the recombination pathway. In addition, I analysed the meiotic E3 ligase proteins in fly and yeast model organisms, *D. melanogaster* and *S. cerevisiae*, to show that the E3 ligase structure is conserved across species.

Together, the results presented in all three chapters uncover the molecular structures of meiotic proteins and protein complexes. This has provided key insights into the crucial events during prophase I of meiosis, including the dynamic movement of chromosomes to facilitate pairing, the assembly of the synaptonemal complex and the formation of crossovers.

Chapter 2

METHODS

2.1. Protein sequence analysis

2.1.1. ***Protein sequences analysed in this study***

The meiotic telomere complex proteins TERB1, TERB2 and TRF1 sequences relate to their canonical human isoforms (accession codes Q8NA31, Q8NHR7 and P54274, respectively). MAJIN sequence utilised relates to the human MAJIN isoform X1 (254 amino acids; accession number XP_024304215.1).

Human meiotic recombination proteins, HEI10, RNF212 and RNF212b sequences utilised relate to the canonical human isoform with Uniprot accession codes Q9NPC3, Q495C1 and A8MTL3, respectively.

The *D. melanogaster* proteins C(3)G, Vilya, Narya and Nenya protein sequences were obtained from the FlyBase database (http://www.flybase.org/).

Table 2.1.1.| *D. melanogaster* sequences

Protein	Annotation symbol	FlyBase ID
C(3)G	CG17604	FBgn0000246
Vilya	CG2709	FBgn0283545
Narya	CG12200	FBgn0031018
Nenya	CG31053	FBgn0051053

2.1.2. ***Bioinformatic analyses***

For the human proteins described in this study bioinformatic analyses including secondary structure prediction, sequence conservation and predicted coiled-coil propensity were carried out. Multiple sequence alignments (MSAs) of close homologous sequences were generated by a BLAST search and retrieving protein sequences from UniProtKB (http://www.uniprot.org) and using MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/). MSAs were utilised to predict secondary structure using JPred4 and presented using Jalview and plotted graphically (Cole et al., 2008; Drozdetskiy et al., 2015). The Consurf server was used to determine protein sequence conservation based on the phylogenetic relations between homologous sequences and calculate conservation scores per residue which were

plotted graphically (Ashkenazy et al., 2016). Coiled coil propensity was predicted by the COILS server which calculates the probability that the sequence will adopt a coiled-coil conformation using a scanning window of 14, 21 and 28 residues (Lupas et al., 1991; Lupas, 1997). The score of each residue was then plotted graphically.

Bioinformatic analyses of *Drosophila melanogaster* proteins C(3)G, Vilya, Narya and Nenya were carried out using a database of *Drosophila* genes and genomes, FlyBase (https://flybase.org). An alignment of 18 *Drosophila* orthologues was manually generated for the four *D. melanogaster* proteins and used to determine the secondary structure prediction, sequence conservation and predicted coiled-coil propensity. (Ashkenazy et al., 2016; Hemmer & Blumenstiel, 2016; Miller et al., 2018).

The ExPASy Protparam tool (https://web.expasy.org/protparam/) was used to determine physical and chemical parameters for a given protein sequence including extinction coefficient, molecular mass and isoelectric point (pI).

2.2. Preparation of plasmids for recombinant protein expression

2.2.1. ***Plasmids used in this study***

In this study plasmids were used according to their in-sequence TEV-cleavable tags; pHAT4 and pMAT11 plasmids were used for single protein expression fused to a N-terminal His_6 and His_6-MBP tag, respectively (Peränen et al., 1996). The pRSF-Duet1 plasmid (Novagen) was utilised for co-expression of two proteins of interest due to having multiple insertion sites. The protein sequence, fused to affinity tag(s), were amplified from the pHAT4 or pMAT11 plasmids and inserted into the pRSF-Duet1 multiple cloning sites. The pRSF-Duet1 plasmid could be co-transformed with pMAT11 or pHAT4 plasmids facilitating the co-expression of 2 or 3 recombinant proteins, due to their compatible antibiotic resistance (Table 2.1.2.). Coding sequences were amplified by PCR and inserted into these vectors by one-step sequence and ligation independent cloning (SLIC) (Jeong et al., 2012).

Table 2.1.2.| Properties of plasmids utilised in this study. The restriction site corresponds to the restriction endonuclease enzyme used for vector linearisation. The plasmid resistance marker correlates to the antibiotic resistance they confer to transformed *E. coli*.

Plasmid	Restriction site	Resistance marker	µg/ml
pHAT4	NcoI	Ampicillin	100
pMAT11	NcoI	Ampicillin	100
pRSF-Duet1	Nco1 (MCS I) Nde1 (MCS II)	Kanamycin	50
pRARE	n/a	Chloramphenicol	34
pGBKT7	EcoR1	Kanamycin	50
pGADT7	EcoR1	Ampicillin	100

2.2.2. PCR amplification

Primers were designed and purchased from Sigma Aldrich. DNA sequences corresponding to regions of interest were amplified by a 25 µl PCR reaction catalysed by KOD DNA polymerase (Novagen) (0.2 µl) in a solution containing: 1x 'KOD DNA polymerase buffer 1 or 2' (2.5 µl), dNTPs (2.5 µl), $MgCl_2$ (1.5 µl), forward and reverse primers (0.2 µl each), template DNA (0.25 µl) and made up to 25 µl with fresh MilliQ. The recommended 3-step PCR thermocycling conditions were followed using a TPersonal Thermocycler (Biometra):

1. 95 °C for 2 minutes
2. 98 °C for 15 seconds (Denaturation)
3. 55 °C for 5 seconds (Annealing)
4. 72 °C for 20 seconds (Extension)

   } 35 cycles (10 cycles for initial rounds of overlapping PCR)

5. 74 °C for 5 minutes

Hold at 4 °C

Full PCR amplification was performed through 35 cycles of steps 2-4. For site directed mutagenesis overlapping PCR was performed, achieved through the production of overlapping fragments generated by initial PCR amplification using 10 cycles. The overlapping fragments were mixed and then used as

a template for a 35 cycle PCR amplification. PCR reactions and cycling conditions were adapted if necessary, in some cases the addition of 5 % DMSO was used to help a generate PCR product.

In the cases of TERB2, HEI10 and RNF212 the protein construct was cloned with a N-terminal 3x or 5xTGS long linker repeat sequence between the TEV cleavage site and the start of the protein sequence to aid the cleavage of the N-terminal solubility tag by TEV-protease enzyme.

2.2.3. ***Plasmid linearisation***

Plasmids need to be linearised by the according single restriction enzyme (Table 2.1.2). 50 µl reactions were set up containing 2 µg plasmid, 5 µl of the appropriate enzyme buffer, 2 µl of the restriction enzyme (NEB) and made up to 50 µl with fresh MilliQ. Reactions were incubated at 37 °C for 3 hours and run on an agarose gel (see below) for band excision and purification.

2.2.4. ***Agarose gel electrophoresis***

PCR amplified DNA products and linearised plasmids were analysed by agarose gel electrophoresis using a 0.8 % w/v gel with the addition of 1x SYBR Safe (Invitrogen). 5 µl and 45 µl DNA samples were prepared for analysis and purification, respectively, through addition of 1x gel loading dye (NEB) and loaded alongside 1 kb and 100 bp DNA ladders (NEB) for reference. Gel electrophoresis was performed at 100 V for approximately 30 minutes. Analytical DNA gels were imaged with the Gel Doc XR+ Gel Documentation System (BioRad) and purification samples were visualised by the Dark Reader *blue light* transilluminator box and purified using the GeneJet Gel Extraction Kit (Thermo Scientific) following the manufacturers protocol. DNA concentrations of PCR products and linearised vectors were determined using a nanodrop UV-Vis spectrophotometer (ThermoScientific).

2.2.5. ***One-step sequence and ligation independent cloning (SLIC)***

One-step SLIC reactions were performed, following the protocol described in Jeong et al., 2012, to insert desired DNA fragments into linearised plasmids. The volume of insert and linearised vector (100 ng) were calculated at a 3:1 molar ratio and mixed with 1 µl 1x Buffer 2.1 (NEB), 1 µl 10x BSA and made up to 10 µl with fresh MillQ. 0.3 µl of T4 DNA polymerase (NEB) was added to the reaction and

incubated at 22 °C for precisely 2 ½ minutes and transferred to ice. After 10 minutes a 50µl aliquot of DH5α *E.coli* competent cells were transformed with 2 µl of the SLIC reaction (see 2.3.).

Single colonies were selected and cultured in 5 ml LB broth supplemented with the according antibiotic (Table 2.1.1.) at 37 °C for 16 hours overnight. The recombinant plasmids were then purified using the GeneJET Plasmid Miniprep Kit (Thermo Scientific). The concentrations of the purified plasmid were measured using the Nanodrop spectrophotometer. The cloning reaction was confirmed by performing a restriction digest using the appropriate restriction enzyme for the vector and analysed by agarose gel electrophoresis to detect clones with inserts of the correct size.

Plasmids were sequenced by Source Bioscience DNA sequencing services and analysed manually using Chromas 3.1. and a BlastX NCBI search was also used to confirm the result. A list of sequenced cloned plasmids used in this study is summarised in Tables 2.1.2., 2.1.3.and 2.1.4.

2.3. Transformation of competent cells

BL21 (DE3) (Novagen) harbouring the pRARE plasmid and DH5α (Life Technologies) chemically competent *E. coli* cells were prepared by the same method using rubidium chloride competent cell preparation protocol for both protein expression and cloning, respectively. 50 µl aliquots of competent cells were flash frozen in liquid nitrogen and stored at -80 °C until required.

Chemically competent DH5α *E. coli* cells were transformed with the 2 µl of the SLIC reaction (described in 2.2.5.) to be recombined and BL21 (DE3) cells were transformed or co-transformed with 100 ng of sequenced plasmid(s) for protein over-expression. Aliquots of chemically competent cells were thawed on ice for 10 minutes before the appropriate amount of DNA was added aseptically and incubated for 20 minutes on ice. Cells were heat-shocked for 45 seconds at 42 °C and further incubated on ice for 10 minutes before the addition of 300 µl sterile SOC broth (Formedium). Transformations were incubated for 1 hour at 37 °C at 180 rpm and plated onto sterile LB agar plates supplemented with the appropriate antibiotic(s) (Table 2.1.1.). Plates were incubated overnight (~16 hours) at 37 °C.

2.4. Recombinant protein expression and purification

2.4.1. *Over-expression of recombinant protein*

BL21 (DE3) pRARE *E. coli* competent cells (Novagen) were transformed using plasmid(s) encoding the protein(s) of interest and plated on LB agar supplemented with the according antibiotic(s) (Table 2.1.2.), to allow for selection of successfully transformed clones (detailed in 2.3.). Colonies were selected aseptically and grown in autoclaved 2x YT broth (Formedium) plus the appropriate antibiotic(s) in an aerated shaker at 37°C, 180 rpm until an optical density of ~0.8 at 600 nm was reached. The culture was then induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 500 µM and grown at 25 °C for 16 hours (overnight) at 180 rpm. For the expression of the human meiotic recombination proteins, HEI10, RNF212 and RNF212b, a nutrient rich media, Terrific Broth (TB), (Formedium) was used instead of 2xYT to help promote folding of the RING domain. Once cultures had reached an OD of 0.8, zinc acetate was added to a final concentration of 100 µM and were cooled on ice for 10 minutes prior to adding IPTG. Cultures expressing TRF1 or HEI10 constructs were incubated at 15°C overnight.

2.4.2. *Preparation of cell lysate*

The 1 litre cultures were centrifuged at 4°C, 4200 rpm for 30 minutes and the supernatant was discarded. The pellet was resuspended in 30ml lysis buffer containing 20 mM Tris pH 8.0, 500 mM KCl. In the case of protein complexes including TRF1, 10 % glycerol was added to the lysis buffer. The bacterial pellets of HEI10 were resuspended in lysis buffer containing 50 mM Tris pH 8.0, 500 mM NaCl, 10 % glycerol, 5 mM DTT and 100 µM zinc acetate. Cell lysates were either frozen and stored at -20 °C for future purification or directly sonicated. Protein purification was initiated by sonication of cells followed by centrifugation at 40,000 g for 30 minutes to pellet the cell debris (insoluble material) from the protein supernatant (soluble material). All steps were carried out on ice.

2.4.3. *Nickel affinity chromatography*

Nickel and amylose affinity chromatography were performed in the cold room at 4 °C to minimise protein degradation. Ni-NTA resin (Qiagen) was packed into an econo-column (Bio-Rad Laboratories)

and equilibrated with the according lysis buffer (see 2.4.2.) using a peristaltic pump, at a flow rate of 1 ml/min. For every litre of culture, 1-1.5 ml of Ni-NTA resin was used. Cell lysate was loaded onto the pre-equilibrated column, followed by washing with 5 column volumes (CVs) of lysis buffer. The resin was then washed with 5 CVs of lysis buffer supplemented with 20 mM imidazole pH 8.0. His_6-tagged bound protein(s) were then eluted by increasing the imidazole concentration to 200 mM.

2.4.4. *Amylose affinity chromatography*

The Ni-NTA eluate of protein constructs containing an MBP-solubility tag was applied to an amylose resin column (NEB) that had been pre-equilibrated in 20mM Tris pH 8.0, 150mM KCl and 2mM DTT (equilibration buffer), following the same procedure described in 2.4.3. Once the Ni-NTA eluate had been loaded, the resin was washed with 5 CVs of equilibration buffer, followed by elution with equilibration buffer supplemented with 30mM D-maltose.

For recombinant protein pull-down assays, cell lysate was directly applied to pre-equilibrated amylose resin packed into econo-columns and incubated at 4 °C, gently rocking for 1 hour. After which, resin was allowed to settle and the lysate was dripped through the resin by gravity, followed by a wash and elution step as described above.

2.4.5. *Ion exchange chromatography*

Ion exchange (IEx) chromatography was used to further purify the solubility tagged proteins from any degradation products using either a HiTrapQ HP or HiTrap SP (GE Healthcare) 5ml anion or cation affinity column, respectively dependent upon the proteins isoelectric point (pI). IEx chromatography was performed using an ÄKTA Start purification system (GE healthcare). The system and attached column were pre-equilibrated with 20mM Tris pH 8.0, 100mM KCl and 2mM DTT (buffer A) and the amylose or Ni-NTA eluate was adjusted 100mM KCl then loaded at 2ml/min. Once loaded the column was washed with 5 CVs of buffer A and the protein was eluted across a 100 % salt gradient in 1 ml fractions through mixing buffer A with buffer B (20mM Tris pH 8.0, 1 M KCl and 2mM DTT) across a 50 ml volume.

2.4.6. *Affinity tag cleavage*

Affinity tag cleavage was achieved by using sfGFP-TEV protease, which cleaved the protein at the TEV protease recognition site between the affinity-tag and the protein sequence. The pooled IEx fractions were I ncubated with sfGFP-TEV (prepared following the procedure described in Wu et al., 2009) at a 1:30 molar ratio for 2 hours at room temperature, then at 4°C overnight gently rocking. Upon the addition of sfGFP-TEV to TRF1 and TRF1-containing complexes, some protein precipitation was observed overnight, therefore the protein-TEV solution was centrifuged at 40,000 g for 30 minutes and the supernatant was pooled. SDS-PAGE analysis of the pelleted aggregate and supernatant determined that cleaved TRF1 was present in the supernatant.

Cleaved recombinant protein(s) were separated from sfGFP-TEV protease and affinity tags during a second round of IEx chromatography usually under the same conditions as fusion protein, described in 2.4.5. In some cases, where the protein pI was high, or the protein was known to bind DNA the second IEx step utilised the HiTrapSP or HiTrap Heparin HP (GE Healthcare) exchange columns.

2.4.7. *Size-exclusion chromatography*

Preparative size-exclusion chromatography (SEC) was used as the final purification step, performed using the HiLoad S200 16/600 SEC superdex column (GE Healthcare) attached to an ÄKTA Pure purification system (GE healthcare). In most cases the system and column wer pre-equilibrated in 20mM Tris pH 8.0, 150 mM KCl and 2mM DTT (plus 10 % glycerol for HEI10 and TRF1 proteins) buffer and 1-5 ml of protein sample was injected onto the column. The protein sample was eluted in 2 ml fractions by an isocratic gradient, separating the solution of particles based on size.

2.4.8. *Protein concentration and buffer exchange*

Following recombinant protein purification, samples were concentrated by centrifugation using a Vivaspin centrifugal concentrator (Fisher Scientific) device according to the sample molecular weight (3 or 10 kDa MWCO). Samples were centrifuged at 5500 x g for 10-minute intervals at 4 °C and resuspended between each centrifugation cycle. After concentration, protein samples were centrifuged

for 30 minutes at 16000 x g at 4 °C following protein concentration determination (see 2.4.9.). Protein samples were aliquoted and stored at -80 °C following flash-freezing in liquid nitrogen.

Buffer exchange was used to transfer a protein solution into a different buffer appropriate for downstream applications including circular dichroism (CD) spectroscopy. This was achieved using NAP-5 columns (GE Healthcare), following the manufacturer's recommended protocol.

2.4.9. *Protein concentration determination*

Protein sample concentration was calculated using the Beer-Lambert law: $A = \varepsilon \times l \times c$, where A is the absorbance, ε is the extinction coefficient, l is the cell path length and c is the molar concentration. The absorbance at 280 nm was determined by UV spectroscopy using a Cary 60 UV-Vis spectrophotometer (Agilent) with micro 15 mm UV cuvettes (Sigma Aldrich) with a pathlength of 1 cm. Protein molecular weights and extinction coefficients were determined from the ExPASy Protparam tool (https://web.expasy.org/protparam/), allowing for protein concentration to be calculated.

2.4.10. *SDS-PAGE analysis*

Throughout the protein purification stages, protein samples were analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) using the Novex Bolt Bis-Tris gel system (Invitrogen) with 15% polyacrylamide gels. Protein samples were prepared by the addition of 1x LDS loading dye and 1x reducing agent (Invitrogen), then incubated at 75°C for 10 minutes before being loaded onto the gel alongside the PageRuler prestained 10-180 kDa protein ladder (ThermoScientific) for molecular weight standards. Gels were electrophoresed at 200V for approximately 30 minutes or until the dye front has reached the bottom of the gel, then stained with Coomassie Brilliant Blue (Sigma Aldrich) for visualisation.

2.5. Biophysical assays

2.5.1. Circular dichroism (CD) spectroscopy

Far-UV circular dichroism (CD) spectroscopy experiments were performed using a Jasco-810 spectropolarimeter (Institute of Cell and Molecular Biosciences, Newcastle University). Prior to analysis, protein samples were buffer exchanged (described in 2.4.8.) into 10 mM disodium phosphate pH 7.5, 150 mM sodium fluoride buffer. Protein concentrations were determined and changed accordingly to be analysed between 0.1 and 0.5 mg/ml using a 0.2 nm path length quartz cuvette (Hellma). Scans were performed at 4 °C between 260 and 185 nm at 0.2 nm intervals with a response time of 4 seconds and a bandwidth of 2 nm. Nine accumulations were measured and averaged after buffer correction and converted to mean residue ellipicity [θ] (MRE), (x1000 deg.cm^2.dmol^{-1}residue^{-1}) using the equation: $MRE\ ([\theta]) = \frac{\theta.\mathrm{MRW}}{10.P.conc}$ where θ is Theta Machine Units, measured in millidegrees, MRW is mean residue weight (calculated as protein mass (Da) / residue number), P is pathlength in cms, and conc is protein concentration measured in mg/ml (Kelly et al., 2005). The Dichroweb server (http://dichroweb.cryst.bbk.ac.uk) was used to perform deconvolution of the data to estimate the secondary structure composition including helical content using the CDSSTR algorithm.

CD thermal denaturation of samples in 20 mM Tris pH 8.0, 150 mM KCl buffer were measured by tracking the helical signal at 222 nm between 4 and 95 °C in 0.2 °C increments at a ramping rate of 1°C per minute. Data were converted to MRE ($[\theta]_{222}$) and plotted as percent folded which assumes a fully folded molecule at 5 °C and complete denaturation at 95 °C. Melting temperature (Tm) was estimated as the point at which 50% of the sample is unfolded.

2.5.2. Size-exclusion chromatography multi-angle light scattering (SEC-MALS)

Size-exclusion chromatography multi-angle light scattering (SEC-MALS) analysis was used to determine the absolute molecular weight of protein species following separation according to their shape and size. SEC was performed with a Superdex or Superose 200 10/300 GL columns (GE Healthcare) attached to an ÄKTA Pure purification system (GE Healthcare), simultaneously connected to the DAWN HELEOS II MALS light scattering detector followed by the Optilab T-rEX differential

refractometer (Wyatt Technology). The system was pre-equilibriated in 20mM Tris pH 8.0, 150 mM KCl, 2mM DTT buffer, unless stated otherwise, at flow rate of 0.5 ml/min overnight or until the differential refractive index (dRI) was stable. Meiotic telomere complex proteins (TRF1, TERB1, TERB2 and MAJIN) constructs were analysed in 20mM Tris pH 8.0, 250 mM KCl, 2mM DTT buffer. Protein samples were centrifuged at 16000 x g for 30 minutes to remove any aggregates before analysis. Bovine serum albumin (BSA) was initially analysed and used as the configuration for subsequent samples. Differential refractive index (dRI) and light scattering (LS) was collected and analysed using ASTRA® 6.1 Software (Wyatt Technology), using a nominal dn/dc value of 0.185 ml/g for Zimm plot extrapolation to determine the molecular mass of protein species (Zhu et al., 1996).

2.5.3. Size-exclusion chromatography small-angle x-ray scattering (SEC-SAXS)

Size-exclusion chromatography small-angle x-ray scattering (SEC-SAXS) analysis was used to determine low resolution structural properties of protein species following separation according to their shape and size. Data was collected on the B21 beamline at the Diamond Light Source (DLS) synchrotron facility (Oxfordshire, UK). SEC was performed with a Superdex 200 10/300 GL column (GE Healthcare) attached to an Agilent 1200 HPLC system, both pre-equilibrated in 20mM Tris pH 8.0, 150mM KCl buffer at 0.5 ml/min. Samples were transported to DLS on dry ice and before analysis were rapidly thawed and centrifuged at 16000 x g for 30 minutes to remove any aggregates. 100µl protein samples (concentrations >10mg/ml) were pipetted into a 96-well tray and data was collected using batch mode. Samples were loaded onto the SEC column and the output was directed to the SAX experimental cell. SAXS data was continuously collected in 3 second frames at 12.4 keV, with the detector distance of 4.014 m.

Processing and analysis of data were performed using ScÅtter 3.0. (www.bioisis.net/scatter). The scattering data for each frame across the protein elution peak was subtracted from buffer, followed by averaging of selected buffer-subtracted frames. Guinier analysis was performed to determine the radius of gyration (*Rg*) and the cross-sectional radius of gyration (*Rc*), where determinable. Approximate values for the maximum interatomic distance (*Dmax*) were generated by BayesApp

(http://www.bayesapp.org) and the paired-distance distribution profile (*P(r)*) generated by PRIMUS from the ATSAS suite (Konarev et al., 2003).

2.5.4. SAXS-directed protein structure modelling

Ab initio modelling was performed to produce low resolution molecular envelopes from the SAXS *P(r)* analysis. Molecular envelopes were generated using DAMMIF (ATSAS suite), multiple (10-20) independent runs were carried out and then averaged (Franke & Svergun, 2009). The DAMAVER program suite (ATSAS) used DAMSEL to compare the generated models to find the most probable model and to identify any outliers. DAMAVER then used the DAMSEL output to average the suitable models followed by DAMFILT to filter the averaged model. The final DAMFILT models were presented with surface representation (Volkov & Svergun, 2003).

Crystal structures were docked into DAMFILT molecular envelopes using SUPCOMB (Kozin & Svergun, 2001). Multi-phase *ab initio* modelling was performed using MONSA (Svergun, 1999) and CORAL (Petoukhov et al., 2012). Crystal structures and models were fitted to the experimental data using CRYSOL (Franke et al., 2017).

2.5.5. Electrophoretic mobility shift assay (EMSA)

Electrophoretic mobility shift assays (EMSAs) were performed to assess the DNA binding ability of TRF1, HEI10 and RNF212 proteins. Analysis was carried out using random linear double-stranded DNA (dsDNA) sequence of different lengths, stated in the results. For the case of TRF1, in addition to a 75 bp random dsDNA sequence, annealed (6nt linker of CGACGA) telomeric dsDNA (TTAGGG repeat) of 54bp was also used. Protein samples of increasing concentration (stated in the Results) were incubated with 25 μM (per base pair) of DNA in 20mM Tris, 250mM KCl buffer at 4°C for 60 minutes. A 0.5% TBE agarose gel containing 4 μl of SYBR safe (Life Technologies) was pre-run for 60 minutes at 20 V with current of 4 mA at 4 °C. Glycerol was added at a final concentration of 3% to the samples prior to loading and the gel was run for a further 4 hours under the same conditions.

For inhibition of DNA-binding assays, 57 bp random dsDNA sequence was pre-incubated with MAJIN-TERB1-TERB2 for 20 minutes at 4 °C prior to adding TRF1. In DNA super-shift assays, 57 bp random

dsDNA sequence was pre-incubated with TRF1 for 20 minutes at 4 °C prior to adding MAJIN-TERB1-TERB2.

EMSA experiments were utilised to determine the strength of DNA-binding of TRF1. Dr James Dunce modified the experimental parameters, detailed above, to develop a system to estimate the apparent affinity (K_D) of DNA-binding proteins to dsDNA. EMSAs were performed using a 5'-FAM (fluorescein amidites) -labelled 144 bp random sequence dsDNA at 25 nM. This concentration is below the previously determined K_D. Fluorescent signal was enhanced by SYBR™ Gold (ThermoFisher) staining. Imaging was performed using a Typhoon™ FLA 9500 (GE Healthcare) and analysed using ImageJ software to quantify the proportion of unbound dsDNA (https://imagej.nih.gov/ij/). This was converted to a percentage bound (% DNA bound) and plotted against protein concentration fitted to the Hill equation ($\%\ \text{DNA bound} = C^n/(K_D^n + C^n)$) with the apparent K_D determined.

2.5.6. Zinc content determination assay (PAR assay)

A series of standards containing 0 – 100 µM zinc acetate were prepared. Subsequently, 10 µl of each standard or sample is mixed with 80 µl PAR (4-(2-pyridylazo) resorcinol) at 50 µM dissolved in 20 mM Tris pH 8.0, 150 mM KCl (no DTT). After 5 minutes at room temperature each sample in analysed by UV spectroscopy between 200 – 600 nm. Zinc in solution is chelated by PAR resulting in an observable spectrophotometric shift from 414 to 494 nm.

10 µl protein samples at three concentrations (20, 40 and 80 µM) were digested by adding 0.3 µl proteinase K at 20 mg/ml and incubating for 1 hour at 60 ˚C. Protein samples were then mixed with 80 µl PAR, incubated and then analysed. Protein samples were then compared with the zinc standards. The absorbance of the zinc standards at 413.5 nm ($A_{413.5}$) and 491 nm (A_{491}) are plotted against the calculated zinc concentration. Trendlines were calculated by linear regression of the experimental data, with a R^2 value close to 1. Zinc standards with an absorptivity outside of the linear range are shown in grey and have been excluded from calculating the linear equation (Figure 5.2.6.c and d.). The determined linear equation allows for the calculated concentration of zinc to be accurately determined.

2.5.7. Limited proteolysis

TERB1$_{585-642}$-TERB2$_{1-107LL}$ (58.2 µM) was incubated with trypsin at a 1:1000 molar ratio (stock concentration of trypsin - 85µM). At each time point 0 (before adding trypsin) ,1,5,10,30,60,120,240 minutes and overnight 13µl of the sample was taken and 7µl of LDS loading buffer (ThermoFisher) was added, incubated at 95°C for 5 minutes and samples were analysed by SDS-PAGE.

2.5.8. Amylose pulldown

MAJIN$_{\Delta TM}$-MBP-TERB2-TERB1$_{TRFB}$ (2 µM) was pre-incubated with plasmid dsDNA (20 µM per base pair; 7987 bp) for 30 minutes at 4°C, and then with TRF1$_{TRFH}$ (4 µM) for 30 minutes at 4°C in 20 mM Tris pH 8.0, 250 mM KCl, 10 mg/ml BSA (100 µl reaction volume). Reactions were added to 40 µl of pre-equilibrated amylose resin (NEB) and incubated for 1 hour at 4°C. After centrifugation at 4,000 g, the supernatant was discarded and the resin was washed twice using buffer with BSA present, and a further four times without BSA. 50 µl 1.5x LDS loading buffer (ThermoFisher) was added to the amylose resin, incubated at 95°C for 5 minutes and samples were analysed by SDS-PAGE (2.4.10.).

2.6. X-ray crystallography

2.6.1. High throughput crystallisation screening

Protein samples were screened at concentrations between 5-50 mg/ml in their size exclusion chromatography buffer using the commercial 96-well screens: JSCG+, PACT, Morpheus, Structure (Molecular dimensions), Index, Ammonium Sulphate (Hampton Research) and MPD (NeXtal Biotech). The Mosquito nanodrop crystallisation robot (X-ray facility, Newcastle University) was utilised to set up two vapour diffusion sitting drops per condition, with 100 nl of well solution mixed with 100 nl (drop 1) or 200 nl (drop 2) of protein. Trays were promptly sealed and incubated at 20 °C or 4 °C, and were regularly monitored and imaged using a DFC400 digital microscope camera (Leica).

2.6.2. Optimisation screening

Optimisation screens of successful crystal hits were set up manually in 24-well plates by hanging drop vapour diffusion. Crystal screens were designed varying the component concentrations of the initial crystal hit as well as pH, drop size and protein concentration.

For HEI10 crystal optimisations seeding techniques were performed to improve crystal growth. Microseeding was performed using a seed stock that was created by manual shattering crystals using a seed bead tube (Hampton Research). Varying dilutions of the seed stock were streaked into pre-equilibrated drops using a cat whisker. Cross-seeding was also performed by including the seed stock in initial commercial screens to encourage crystallisation.

Crystals from initial and optimised screens were fished and cryoprotected in 20 % (v/v) PEG400 and 80 % mother liquor, then flash frozen in liquid nitrogen. Crystal conditions containing ≥ 30 % MPD did not require cryoprotecting.

2.6.3. X-ray data collection and structure determination of $TRF1_{TRFH}$-$TERB1_{bonsai}$

Crystallisation screening of $TRF1_{TRFH}$-$TERB1_{bonsai}$ was performed replicating the published crystallisation conditions of TRF1 (PDB code: 1H60) and TRF1-TERB1 (PDB codes: 5WIR and 5XUP) using hanging drop vapor diffusion. The crystal condition, 0.1 M Tris-HCl (pH 8.5), 30 % PEG 300 yielded large crystals, which were fished and cryoprotected in 20 % (v/v) PEG400 and 80 % mother liquor, then flash frozen in liquid nitrogen.

X-ray diffraction data was collected at the I24 microfocus beamline at the Diamond Light Source synchrotron facility (Oxfordshire, UK) on a Pilatus 6M detector as 3600 consecutive 0.10° frames of 0.010 s exposure, at 0.9796 Å, 100 K. The data was auto processed using the Xia2 3dii pipeline utilising XDS and XSCALE (Kabsch, 2010). The space group $P4_12_12$ was selected with a single molecule in the asymmetric unit. The scaled but unmerged data was processed with Aimless and a resolution limit of 2.10 Å was set (Adams et al., 2010; Evans, 2006). Cell content analysis was performed using the protein sequence to specify the cell contents, indicating there was only 1 molecule in the asymmetric unit and a solvent content of 34.99%. The phase problem was solved by molecular replacement using Phaser and the search model, TRF1 (PDB code - 1H6O) (Fairall et al., 2001) . This gave an initial translational function Z score (TFZ) of 7 and an overall log-likelihood gain (LLG) of 518, resulting in a unique solution. The structure was visualised using Coot, in which only showed density for the TRF1 molecule, and none for the TERB1 peptide (Emsley & Cowtan, 2004). Structural figures were made using Pymol (The PyMOL Molecular graphics system, Version 2.0 Schrodinger, LLC.)

2.7. Yeast two hybrid (Y2H)

2.7.1. Growth of strains

A yeast two hybrid (Y2H) screen was performed using the Matchmaker Gold system (Clontech), using the recommended manufacturer's instructions. Glycerol stocks of the *S. cerevisiae* strains, Y187 and Y2H Gold, were stored at -80 °C and for use were streaked onto sterile YPDA agar (Formedium) plates adjusted to pH 6.5. Plates were incubated at 30 °C for 5 days.

2.7.2. Transformation of yeast cells

pGADT7 vectors were transformed into the Y187 strain and the pGBKT7 vectors into the gold *S. cerivisiae* strain, following the Freimoser Fast Yeast Transformation protocol. Carrier salmon testes DNA (Sigma Aldrich) dissolved in phosphate buffer saline (PBS) stored at -20 °C was thawed. Once thawed each aliquot (500 µl) was heated at 95 °C for 5 minutes and then put on ice for 10 minutes, and repeated. The appropriate yeast cells were scraped and added to 50 µl of carrier DNA with 1-1.5 µg of the desired plasmid, 240 µl of filter sterilised 50% PEG 3350 and 36 µl 1M lithium acetate. The transformation mixture was thoroughly mixed and incubated at 30 °C for at least 30 minutes followed by 15 minutes at 45 °C. Cells were pelleted at 15000 G for 30 seconds and resuspended with 100 µl filter sterilised MilliQ. Y2H Gold-pGADT7 and Y187-pGBKT7 transformations were plated on SD leucine (-Leu) or SD tryptophan (-Trp) deficient media adjusted to pH 5.8, respectively, allowing for selection. Plates were incubated at 30 °C for 5 days, after which single colonies were streaked onto fresh SD plates and incubated for a further 3-5 days.

2.7.3. Yeast mating

The two transformants, Y2H Gold-pGADT7 and Y187-pGBKT7, were harvested and mated for Y2H experiments. Stocks of each transformant were made by scraping cells into 2x YPDA broth (Formedium) and 250 µl of each stock for the mating reaction were mixed and incubated for 24 hours at 30 °C, 40 rpm. The mating reaction was then pelleted (250 G for 10 minutes) and resuspended in 500 µl of 0.5x YPDA broth.

Resuspended mating reactions were plated onto corresponding medium to select for successful mating (positive mating control) and protein-protein interactions. 100µl was plated on to double drop out medium agar plates (DDO), -Trp/-Leu media , as a control for successful mating and 150µl onto quadruple drop out (QDO) agar plates, -Trp/-Leu/-His/-Ade media to select for positive interactions. X-α-gal was added to the QDO medium at a 1:500 dilution to select for mated colonies by the appearance of blue colonies. Plates were incubated at 30 °C for 5 days, then imaged against white light illumination for analysis.

Chapter 3

The molecular basis of the meiotic telomere tethering complex

3.1. INTRODUCTION

3.1.1. Telomeres are nucleoprotein caps that adorn the end of chromosomes

Eukaryotic cells have a variety of mechanisms to protect themselves from the harmful effects of DNA damage. The telomere is a specialised DNA-protein complex capping the end of eukaryotic chromosomes preventing them from being recognised by the DNA damage response machinery, and thus processed as broken DNA ends, thereby maintaining the integrity of the chromosome (Sandell & Zakian, 1993). In many eukaryotes telomeres are composed of tandem arrays of short G-rich repeats of double-stranded (ds) DNA (5-26bp), and a 3' single-stranded (ss) DNA overhang that serve as binding sites for specific proteins. Telomeric repeats are conserved across most eukaryotic species and the telomeres in mammals have a conserved $(TTAGGG)_n$ repeating unit (Cohn et al., 1998). Telomeres are maintained by the enzyme telomerase, a ribonucleoprotein complex, that catalyses the addition of telomeric repeats to the end of chromosomes (Collins et al., 1995). This provides a mechanism to maintain the length of chromosomes as well providing docking sites for telomeric proteins. Telomerase is comprised of three major subunits: a telomerase RNA component (TERC), a catalytic reverse transcriptase subunit (TERT) and a telomerase-associated multiprotein complex (shelterin) (Collins et al., 1995; Cong et al., 2002).

The human shelterin complex is a nucleoprotein composed of six proteins; telomere repeat factor 1 (TRF1), telomere repeat factor 2 (TRF2), TRF1-interacting protein 2 (TIN2), telomere-associated protein 1 (TPP1), protector of telomeres 1 (POT1) and human protein repression and activation protein 1 (Rap1) (Figure 3.1.1). Shelterin is present at telomeres throughout the cell cycle and is crucial for assuring the correct telomere length and telomere protection (De Lange, 2005; Palm & de Lange, 2008). It has been shown that deleting individual shelterin components activates the DNA damage signaling and repair pathways, leading to chromosome fusion or telomere loss, suggesting that the telomere becomes 'uncapped'. Furthermore, shelterin has the capacity to recognise telomeric DNA with at least five DNA binding domains, with one being a ssDNA binding domain (De Lange, 2005; Hu et al., 2017; Sfeir et al., 2012).

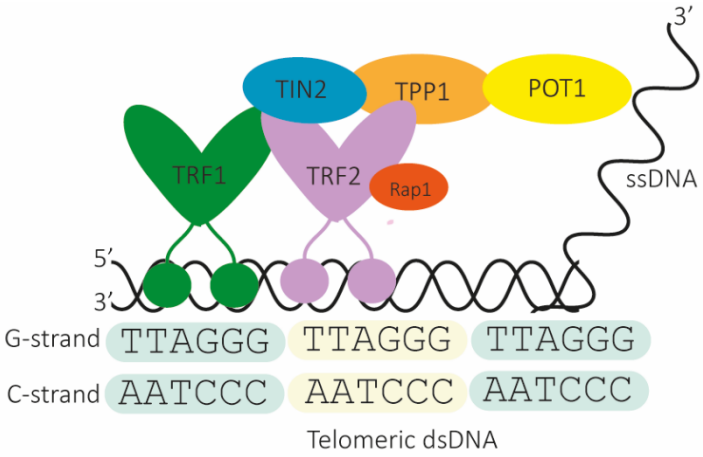


Figure 3.1.1.| The human shelterin complex
The human shelterin complex consists of six conserved proteins. Both TRF1 and TRF2 bind to telomeric dsDNA $(TTAGGG)_n$ repeats as homodimers. TIN2 is the central component of the shelterin complex, interacting with both the TRFH domains of TRF1 and TRF2 as well as interacting with the TPP1-POT1 complex. POT1 binds specifically to the 3' single-stranded telomeric extension. TRF2 also recruits Rap1 completing the shelterin complex.

3.1.2. Telomere repeat factor proteins, TRF1 and TRF2, initiate shelterin complex formation

The mammalian shelterin proteins, telomere-repeat factor proteins 1 and 2 (TRF1 and TRF2), can directly bind to the telomeric dsDNA and in turn recruit further shelterin complex proteins. Human TRF1 activity was originally detected in HeLa cell nuclear extracts by binding to dsDNA fragments containing the conserved telomeric $(TTAGGG)_n$ repeats. It has also been shown that TRF1 can not bind ssDNA sequences (Broccoli et al., 1997; Chong et al., 1995; Liu et al., 2004; van Steensel et al., 1998).

TRF1 and TRF2 have a similar overall domain architecture; a C-terminal myb-like DNA binding domain (DBD) and a central dimerisation domain, known as the TRF homology domain (TRFH) (Figure 3.1.2.a.). The C-terminus of TRF1/2 have strong conservation with DNA-binding repeats found in the family of MYB proteins. For example, c-Myb oncoprotein has two MYB repeats that fold into helix-turn-helix (HTH) motifs that are closely packed onto DNA. It has been shown that at least two myb-related factors are required for DNA binding. However, both TRF1 and TRF2 only have a single myb-like repeat. Moreover, the yeast shelterin protein yRap1, binds telomeric DNA by two MYB repeats separated by a 40 amino acid linker, very similar to c-Myb. At the C-terminus of TRF1 and

TRF2 there are three highly conserved tryptophan residues that form a single helix-turn-helix (HTH) DNA binding fold. Therefore, TRF proteins must dimerise to achieve binding to telomere repeats using two HTH motifs, similar to yRAP1 and c-Myb. Further to this, gel shift assays have shown that the single Myb-like domain of TRF1 alone is not sufficient for DNA binding, indicating that the TRFH dimerisation domain is also required so that two Myb domains can bind DNA to form a stable protein-DNA complex. Moreover, it has been suggested that telomeric binding proteins use a conserved DNA-binding motif to recognise telomeric repeats (Chong et al., 1995; Court et al., 2005; König & Rhodes, 1997; Li et al., 2000; Smogorzewska et al., 2000).

As mentioned previously, both TRF1 and TRF2 form homodimers through their TRFH domains, formed by amino acid residues 62-265 and 43-245, respectively. Both *in vitro* and *in vivo* studies have confirmed that the localisation of TRF1 and TRF2 to telomeric DNA is dependent upon their ability to homodimerise. However, TRF1 and TRF2 do not form heterodimers (Bianchi et al., 1997; Takai et al., 2010). Crystal structures of both $TRF1_{TRFH}$ and $TRF2_{TRFH}$ dimerisation domains have been solved to 2.9 and 2.2 Å, respectively (Figures 3.1.2.b.) (Fairall et al., 2001). From the structures it is possible to determine that both proteins have the same overall α-helical architecture, resembling a twisted horseshoe, despite only having 27% sequence identity. Dimerisation occurs through three α-helices from each monomer, interacting in an anti-parallel arrangement. This is thought to provide an interaction surface for other proteins. From overlaying the two structures it is possible to see small differences between the two proteins, including the length of helices and their packing (Figure 3.1.2.c.). These small differences create non-complementary dimer interfaces in both shape and charge suggesting they may interact with other proteins differently. (Chen et al., 2008; Fairall et al., 2001).

TRF1 and TRF2 differ substantially at the N-terminus, as TRF1 is acidic and TRF2 is basic, with a pI of 3.0 and 9.11, respectively. Clusters of acidic residues have been shown to mediate protein-protein interactions during transcriptional regulation, but more interestingly are found in several proteins involved in chromosome function. This further suggests that TRF1 and TRF2 may have different roles within the shelterin complex (Smogorzewska et al., 2000).

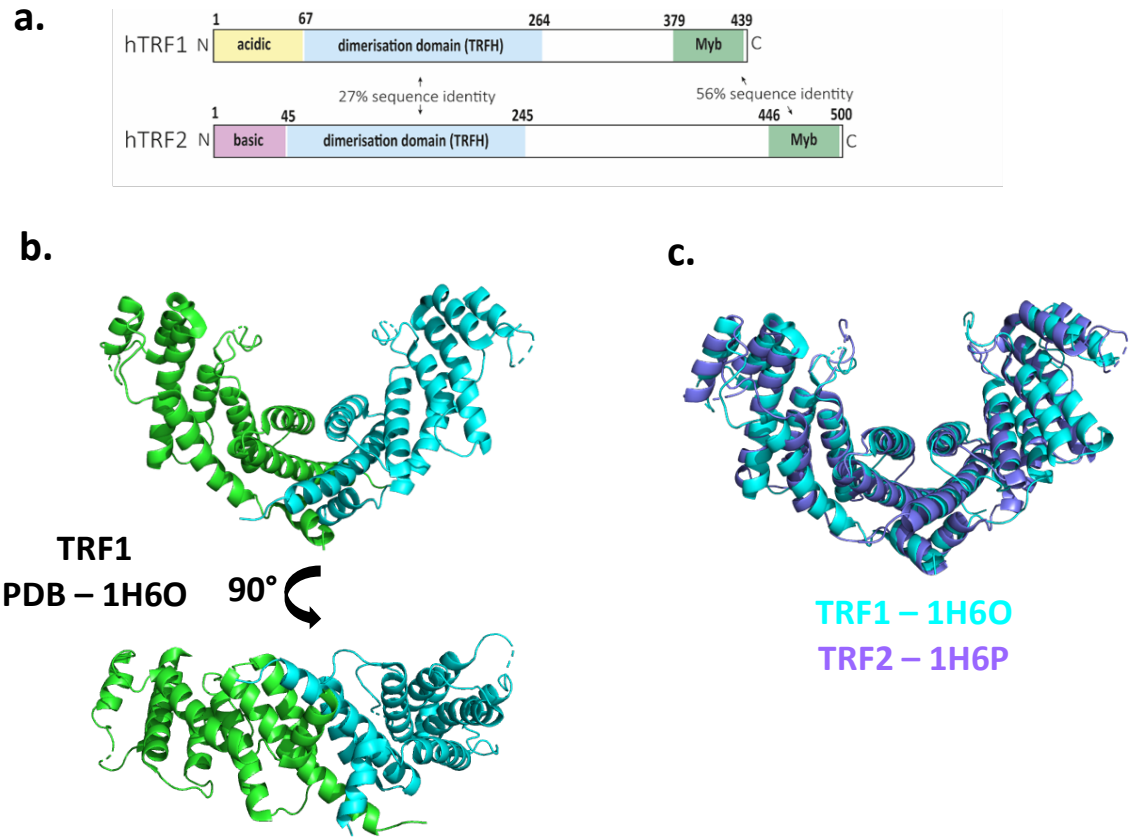


Figure 3.1.2.| Human TRF proteins share a conserved structure.
a) The human shelterin complex contains two dsDNA binding proteins, that bind specifically to telomeric DNA, TRF1 and TRF2. Both proteins share the same domain structure; a conserved C-terminal myb domain that facilitates DNA binding and a central TRFH domain, that is essential for homodimerisation. The C-terminus of TRF1 and TRF2 differ in charge, which could be important for specific interactions. b) Crystal structure of $TRF1_{TRFH}$ dimerisation domain (amino acid residues 65-265; PDB: 1H6O; Fairall et al., 2001). Each TRF1 dimer is comprised of nine α-helices that form an elongated α-helical bundle. c) Superposition of $TRF1_{65\text{-}265}$ (cyan) and $TRF2_{43\text{-}245}$ (purple, PDB: 1H6P; Fairall et al., 2001) crystal structures show that despite the TRFH domains of TRF1 and TRF2 only having a 27% sequence identity they have a very similar overall structure resembling a twisted horseshoe.

3.1.3. Formation of the shelterin complex

The central component of the shelterin complex, TIN2, interacts with both TRF1 and TRF2, as well as tethering the TPP1/POT1 heterodimer complex to TRF1 and TRF2 (Figure 3.1.3.a.) (De Lange, 2005). Pot1 is also a DNA sequence-specific binding protein and specifically binds to the ssDNA 3' overhang, further stabilising the shelterin complex onto telomeres. Human Rap1 does not bind specifically to DNA

but is recruited to the telomeric shelterin complex through an interaction with TRF2 (De Lange, 2005; Hu et al., 2017).

The TRFH domain of TRF1 and TRF2, have a conserved docking motif, FxLxP and YxLxP, respectively, known as the TRFH binding motif (TBM). This conserved motif is known to be associated with the binding of TIN2 to both TRF1 and TRF2 (Diotti & Loayza, 2011). Crystal structures of both $TRF1_{TRFH}$-$TIN2_{TBM}$ and $TRF2_{TRFH}$-$TIN2_{TBM}$ have been solved to 2Å and 2.15Å, respectively (Figure 3.1.3.b.). The structure shows that the $TIN2_{TBM}$ peptide binds to each $TRF1_{TRFH}$ monomer using the FxLxP motif, forming a 2:2 complex (Figure 3.1.3.c.). The overall crystal structures are very similar but there are some subtle differences. Isothermal titration calorimetry (ITC) deduced a binding affinity (K_d) of $TRF1_{TRFH}$-$TIN2_{TBM}$ to be 0.31μM and $TRF2_{TRFH}$-$TIN2_{TBM}$ to be 20-fold higher at 6.49 μM further confirming that the interaction of TIN2 with the TRF1 and TRF2 proteins are not identical. Further analysis of the TIN2 sequence found that the amino acid region 2-202 is closely related to the TRFH domain of TRF1 and TRF2 indicating that there is a second interaction site between TIN2 and TRF2 (Yong Chen et al., 2008; Diotti & Loayza, 2011; Hu et al., 2017).

Subcomplexes of the shelterin complex have been detected, in particular TRF2-TIN2-TPP1-POT1, which is known as the shelterin core complex ($shelterin_{core}$), this is the minimal complex of shelterin that can bind both dsDNA and ssDNA. In solution technique, size-exclusion chromatography in line with multi-angle light scattering (SEC-MALS) determined the absolute molecular weight, from which calculated the stoichiometry of $shelterin_{core}$ to be 2:1:1:1. This shows that TRF2 interacts with TIN2 with a 2:1 stoichiometry, compared to the 2:2 complex seen in the $TRF2_{TRFH}$-$TIN2_{TBM}$ crystal structure (Diotti & Loayza, 2011; Lim et al., 2017). It has been shown that both $TRF1_{TRFH}$ and $TRF2_{TRFH}$ can further recruit shelterin-accessory proteins using the conserved docking motif mentioned previously. $TRF2_{TRFH}$ interacts with Apollo during S-phase of meiosis. The crystal structure of $TRF2_{TRFH}$-Apollo reveals that the binding mechanism has similarities with $TRF1_{TRFH}$-$TIN2_{TBM}$ (Yong Chen et al., 2008). This observation suggests that there could be further interactions of TRF1 and TRF2 with non-shelterin components during meiosis (Diotti & Loayza, 2011; D. Liu et al., 2004; Nandakumar & Cech, 2013).

The binding of shelterin to telomeric DNA is a universal process in eukaryotic cells. The shelterin complex has also been well characterised in the fission yeast, *Schizosaccharomyces pombe* (*S. pombe*). Similar to the mammalian shelterin complex, the *S. pombe* shelterin is comprised of six proteins including Taz1 and Pot1 that bind telomeric ds- and ssDNA, respectively. Taz1 is a functional orthologue of mammalian TRF proteins, that binds to telomeric dsDNA through its conserved C-terminal Myb domain. Taz1 recruits Rap1, which in turn interacts with Pot1 and recruit other shelterin complex proteins. Rap1 can bind directly to multiple proteins, serving as a hub, and is involved in various other telomere functions (Chikashige & Hiraoka, 2001; Cooper et al., 1998; Deng et al., 2015).

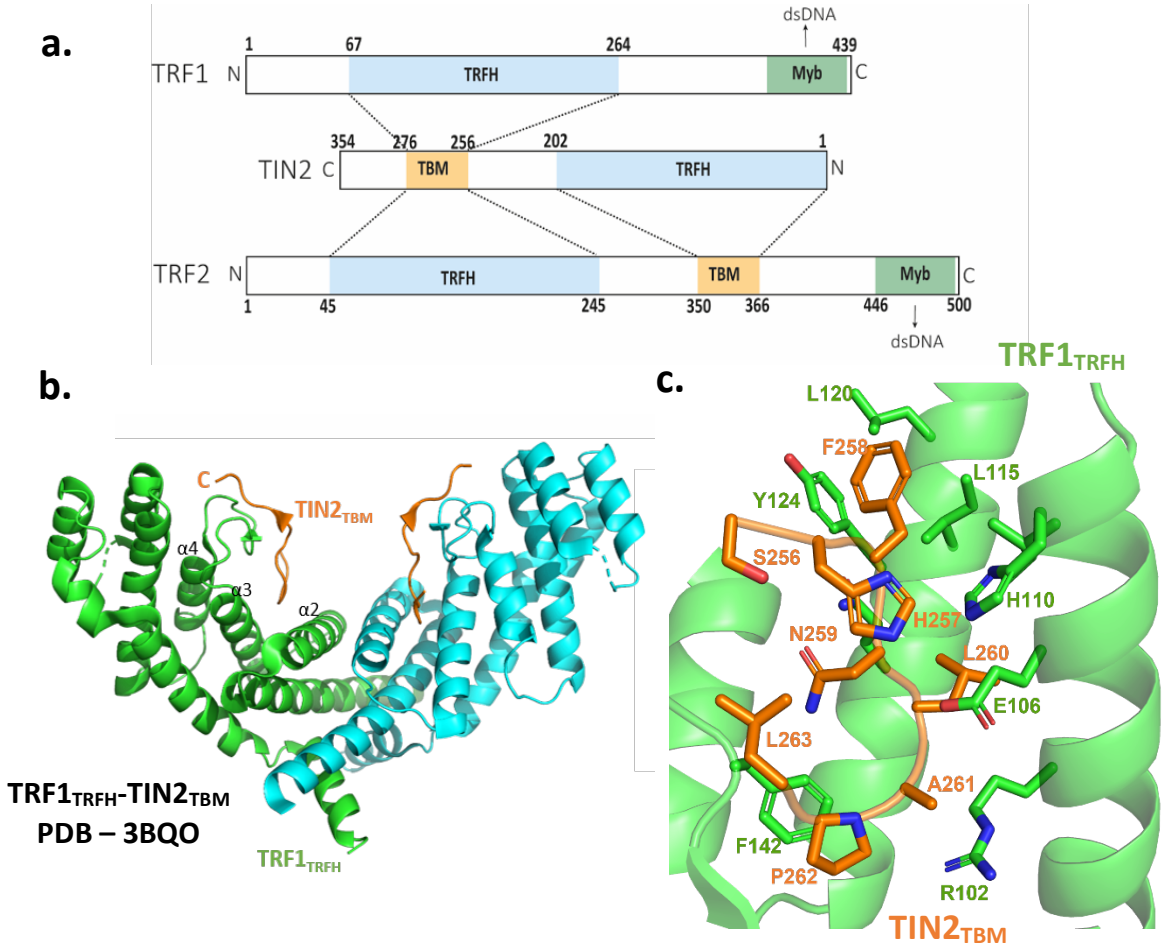


Figure 3.1.3.| Interaction network of the telomeric shelterin proteins: TRF1, TRF2 and TIN2
a) TIN2 is the central component of the shelterin complex, interacting with both TRF1 and TRF2. TIN2 has a TRF binding motif (TBM) that binds to the TRF homodimerisation (TRFH) of both TRF1 and TRF2. There is a second binding reaction between TIN2 and TRF2, in which the TIN2 TRFH domain binds to a TBM motif in TRF2. b) Crystal structure of $TRF1_{TRFH}$-$TIN2_{TBM}$ (PDB code – 3BQO; Chen et al., 2008). TIN2 C-terminal peptide (amino acid residues 256-276) co-crystalised with the TRFH domain of TRF1. The crystal structure was solved by molecular replacement using TRF1 (PDB 1H6O; Fairall et al., 2001) as the model to 2.0Å. The structure revealed that two TIN2 peptides bind to the TRF1 homodimer, forming a 2:2 complex. c) Each TIN2 peptide binds to a TRF1 monomer using the FxLxP motif (F258-P262). TIN2 F258 binds a hydrophobic pocket within TRF1.

3.1.4. Telomeres are structurally adapted for meiosis

Shelterin components can interact with non-shelterin proteins to form unique subcomplexes, allowing dynamic integration of the telomere nucleoprotein-complex with other protein complexes, including meiosis-specific proteins (Yong Chen et al., 2008; Diotti & Loayza, 2011). In prophase I of meiosis, chromosome ends are attached to the inner nuclear membrane (INM) through the association of telomeres with the transmembrane linker of nucleoskeleton and cytoskeleton (LINC) complex (Bone & Starr, 2016; C.-Y. Lee et al., 2015). This interaction is bridged by a meiosis specific complex, known as the meiotic telomere complex (MTC).

In mice germ cells, genetic ablation of TRF1 results in the meiotic telomere complex no longer being able to assemble onto telomeres, demonstrating that TRF1 is crucial for meiotic progression. Furthermore, immunostaining of TRF1-null cells shows a lack of the meiosis-specific MTC protein, TERB1, suggesting that TERB1 is recruited to telomeres through TRF1, thus initialising the recruitment of further meiotic telomere complex proteins (Zhang et al., 2017). This suggests that the shelterin complex protein, TRF1, initiates the binding of meiotic telomere complex, specifically through the interaction with TERB1, which physically connects the telomeres to the LINC complex (Daniel et al., 2014; Shibuya et al., 2014). Furthermore, *S.pombe* shelterin proteins Taz1 and Rap1 function to bridge the meiosis-specific proteins Bqt1 and Bqt2 to the spindle pole body SUN-domain protein Sad1. Disruption of Rap1 binding to Bqt1-Bqt2 complex causes defects in telomere clustering and chromosome segregation (Chikashige et al., 2007; Chikashige et al., 2006).

3.1.5. Telomere-led rapid prophase movements facilitate homologous chromosome pairing

A precondition of successful synapsis is the alignment of homologous chromosomes in the correct orientation (Scherthan et al., 1996). Chromosome are tethered to the nuclear envelope (NE) by their telomeres and move dynamically by rapid prophase movements (RPMs) along the NE creating a reduced unidirectional search for homologous chromosome pairs (Lee et al., 2015). RPMs occur during zygotene and are ubiquitous among eukaryotes and use a general mechanism involving cytoskeletal

elements that produce the forces to generate movement of the LINC complex which in turn allows for chromosome movements (Kohli, 1994; Zickler & Kleckner, 1998).

In mammalian cells, a motor driven action involving dynein in complex with dynactin generates the driving forces which are transmitted to the LINC complex and subsequent chromosome movement. Destabilisation of the dynein-dynactin complex has shown to reduce RPMs in zygotene nuclei, suggesting that microtubules are essential for chromosome movements. (Lee et al., 2015).

RPMs continue throughout leptotene-zygotene to pachytene transitions, peaking at over 100 nm s^{-1} in zygotene, then diminish in diplotene thus suggesting that RPMs function to facilitate chromosome pairing (Lee et al., 2015; Shibuya et al., 2014). Towards the leptotene-zygotene transition, RPMs lead to the clustering of telomeres into a small region of the NE or near the spindle pole body (SPB) in yeast, forming what is known as a chromosomal 'bouquet' or 'horsetail nucleus', respectively (Harper, 2004; Zickler & Kleckner, 1998). Bouquet formation confers a parallel alignment of homologues, which is highly conserved across eukaryotes suggesting that it has a significant role in prophase I of meiosis which might facilitate disentangling of interlocked chromosomes and provide a homology search mechanism (Zickler & Kleckner, 2015) (Figures 3.1.4. and 3.1.5.).

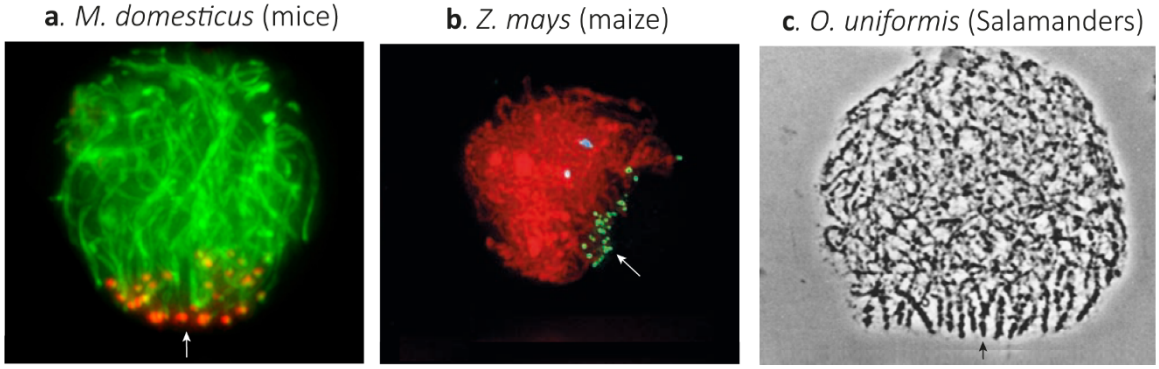


Figure 3.1.4.| The chromosomal bouquet is conserved across eukaryotes.
Images of the meiotic cells of different eukaryotic species (a) mice, (b) maize and (c) salamanders at the zygotene stage of prophase I show a conserved chromosomal bouquet structure (indicated by the arrow). (a) In mice spermatocytes the telomere ends (red) are clustered to one pole of the nucleus. (b) In maize the chromosomes (chromatin stained with DAPI – red) are clustered into one small area of the nucleus, seen by an area of telomeres (FITC – green). (c) Phase contrast micrograph at late zygotene of Salamander nuclei show clustering of telomeres to one pole, showing close association of chromosomes. Figures adapted from a) Berríos *et al*., 2013, b) Golubovskaya et al., 2002 and c) Kezer, 1989.

Bouquet formation is independent of both synapsis and recombination events. However, it plays a role in efficient chromosome pairing and subsequent synapsis. The rate and efficiency of both these processes are dramatically reduced in bouquet mutants. Despite this, attachment to the NE is a prerequisite of bouquet formation. In *S.pombe,* bouquet formation is dependent upon the presence of the both shelterin proteins, Taz1 and Rap1, as well as the LINC protein Kms1 and meiosis specific proteins, Bqt1 and Bqt2 (Chikashige & Hiraoka, 2001; Cooper et al., 1998; A Yamamoto et al., 1999).

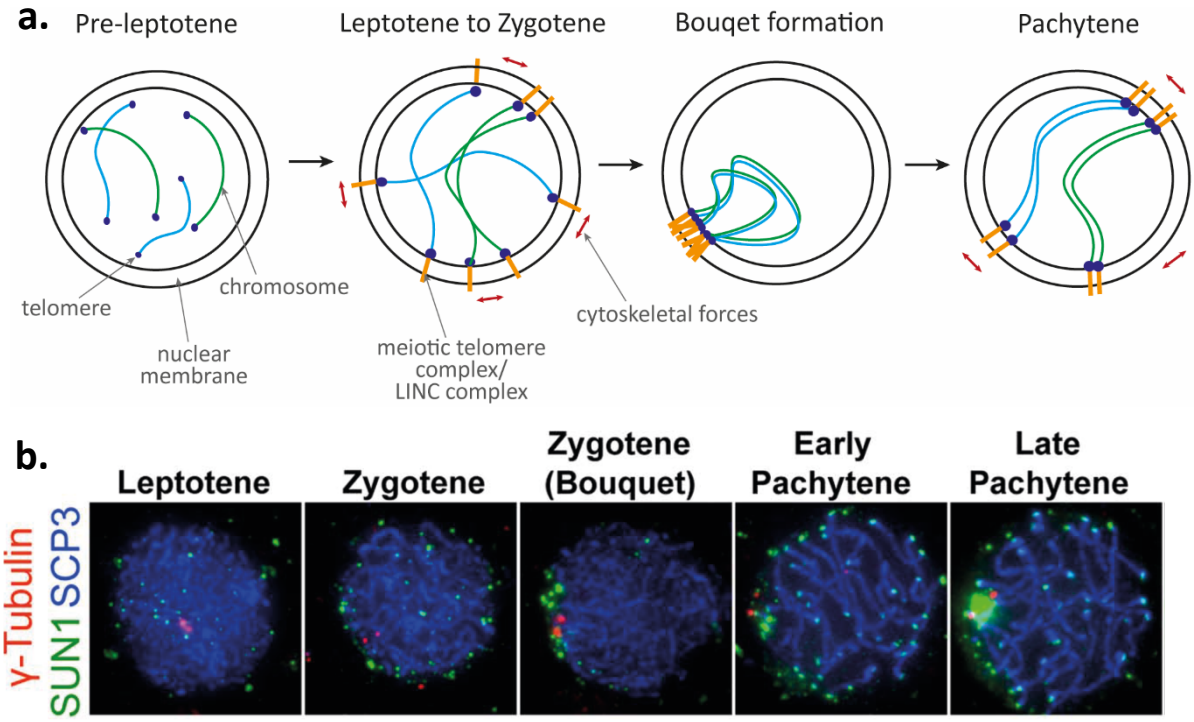


Figure 3.1.5.| Stages of Prophase I
a) By leptotene chromosomes are fully elongated and telomeres attach randomly to the nuclear envelope (NE). This is achieved through the association of three complexes: the shelterin complex, meiotic telomere complex and the LINC complex. Cytoskeletal forces are then transduced to chromosome ends facilitating the rapid movements of chromosomes. At the leptotene-zygotene transition telomeres congregate to a limited area of the NE, to form the chromosomal bouquet, most chromosomes align and allow for chromatin interactions. Between zygotene and pachytene synapsis is initiated and by pachytene homologues are completely paired along the length of the whole chromosome, this is achieved by the formation of the synaptonemal complex. b) Visualisation of the prophase stages of live mice spermatocytes which were stained for SCP3 (cyan), γ-Tubulin (red) and SUN1 (green). γ-Tubulin is used as a marker of the nuclear periphery. At zygotene the LINC complex protein SUN1 is localised at the ends of chromosomes, suggesting interaction with the ends of telomeres allowing for RPMs. This results in bouquet formation, seen by clustering of SUN1 foci. At early pachytene the telomeres of homologue pairs are scattered across the NE and the appearance of the synaptonemal complex lateral element protein, SCP3, foci.
Figure taken from Shibuya et al., 2014.

3.1.6. LINCing the telomeres to the nuclear envelope

Chromosome movements are dependent upon cytoskeletal forces generated by dynein-dynactin driven motions in the cytoskeleton (Lee et al., 2015). The LINC complex is a highly conserved protein complex that provides the molecular connection to transduce cytoskeletal forces to the end of chromosomes (Zhang et al., 2009).

The architecture of the NE is conserved in all eukaryotes; there is an inner nuclear membrane (INM) and an outer nuclear membrane (ONM), separated by the perinuclear space (PNS). In meiotic cells the NE has adapted differently compared to the NE in somatic cells, including differences to the nuclear lamina composition which aids meiotic chromosome dynamics. The nuclear lamina is a major structural feature of the NE, forming a proteinaceous network that associates with the nuclear face of the INM (Burke & Stewart, 2002; Matsumoto et al., 2015; Sato et al., 2009). The lamina of mammalian spermatocytes contains lamin B1 as well as a meiosis specific variant of lamin A; lamin C2. Unlike somatic lamins, lamin C2 forms discontinuous domains at the NE and is highly expressed at the attachment sites of telomeres . This specific structure and localisation suggests that lamin C2 may play a role in anchoring the LINC complex and the attachment of telomeres to the NE. Furthermore, it is already known that somatic lamins interact with components of the INM (Crisp et al., 2006; Haque et al., 2006; Schmitt et al., 2007).

Throughout prophase I chromosomes are tethered to the INM through their telomere ends, which is achieved by the association of telomeres with the LINC complex forming a bridge between chromosome ends and the cytoskeletal forces. The LINC complex is a mobile structure within the plane of the nuclear membrane composed of SUN and KASH domain proteins, that are ubiquitously expressed (Hiraoka & Dernburg, 2009). The mammalian meiotic KASH-domain protein, KASH5 crosses the ONM interacting with the dynein in the cytoplasm, providing a link to the microtubules (MTs) and thus a mechanism for transferring forces from the cytoplasm. The highly conserved KASH domain directly interacts with the SUN-domain of SUN1, forming a bridge across the perinuclear space (Figure 3.1.6.a.). SUN1 crosses the INM allowing for interaction with the nuclear lamina, emerin and nesprins as well as potentially interacting with meiosis-specific proteins (Haque et al., 2006; Morimoto

et al., 2012). It has been hypothesised that SUN1 may interact with the meiotic telomere associated protein, TERB1 at the INM through its N-terminal domain (Shibuya & Watanabe, 2014). Moreover, a SUN1 mutant lacking the SUN domain and the coiled-coil region can still localise to the INM, suggesting that it is the highly conserved nucleoplasmic N-terminus of SUN1 that interacts with TERB1 (Crisp et al., 2006). Fluorescence imaging studies of mice spermatocyte spreads have shown that SUN1 and KASH5 foci co-localise with the shelterin component, Rap1 at the end of SCP3 stained lateral elements (Figure 3.1.6.b.) (Horn et al., 2013). This localisation is concomitant to the proposed LINC-meiotic telomere complex interaction.

The spindle pole body (SPB) of *S.pombe* contains the SUN- and KASH-domain proteins, Sad1 and Kms1, respectively. Analogous to the mammalian LINC complex, Sad1-Kms1 provide the bridge between telomeres and the cytoplasmic actin motors to drive chromosome movements allowing for horsetail nuclear movements. Association between Sad1 and telomeres is mediated by the shelterin protein, Rap1 and Sad1 directly interacts with the meiosis-specific telomere complex Bqt1-Bqt2 (Yuji Chikashige et al., 2006; Hagan & Yanagida, 1995; Horn et al., 2013).

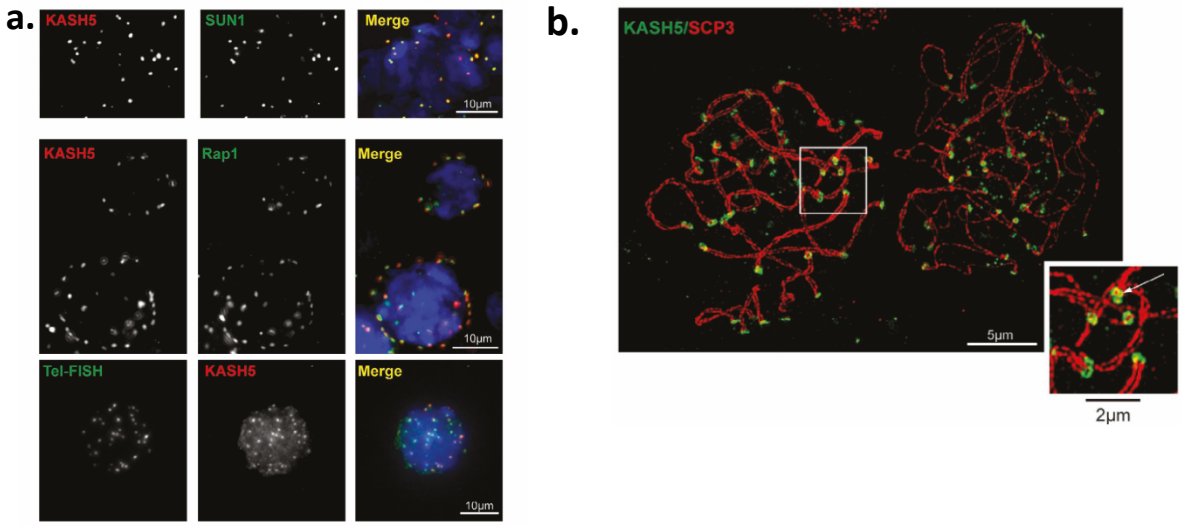


Figure 3.1.6.| The LINC complex proteins colocalise at chromosome ends.
a) Mouse spermatocyte spreads labelled with anti-KASH5 (red) with anti-SUN1 (green) (top), anti-Rap1 (middle). KASH5 colocalises with SUN1. KASH5 colocalises with shelterin protein Rap1 and telomeres, shown by Fluorescence in situ hybridisation (FISH) using a telomeric-specific oligonucleotide probe, Tel-fish (bottom). This colocalisation shows that the SUN1-KASH5 LINC complex localise to telomere ends during prophase I of meiosis. b) Structured illumination microscopy (SIM) of mice spermatocyte spreads labelled with anti-KASH (green) and anti-SCP3 (red). SIM reveals KASH5 rings at the tips of paired SCP3 axial strands. Figure taken from Horn *et al.,* 2013.

3.1.7. The meiotic telomere complex is conserved across eukaryotes

Meiotic telomere regulation is evolutionary conserved across species from humans to fission yeast. Independent of species, a conserved telomere-binding protein is required to recruit meiotic proteins to construct a meiotic telomere associated structure, which facilitates chromosome movement, summarised in table 3.1.1. (Chikashige et al., 2006; Conrad et al., 2008; Kanoh & Ishikawa, 2001; Shibuya et al., 2014). Chromosome ends must be physically associated with LINC complex during prophase I, which is facilitated by the meiotic telomere complex (MTC) (Shibuya et al., 2014). The MTC integrates the functions of both the LINC and shelterin complexes (Figure 3.1.7.).

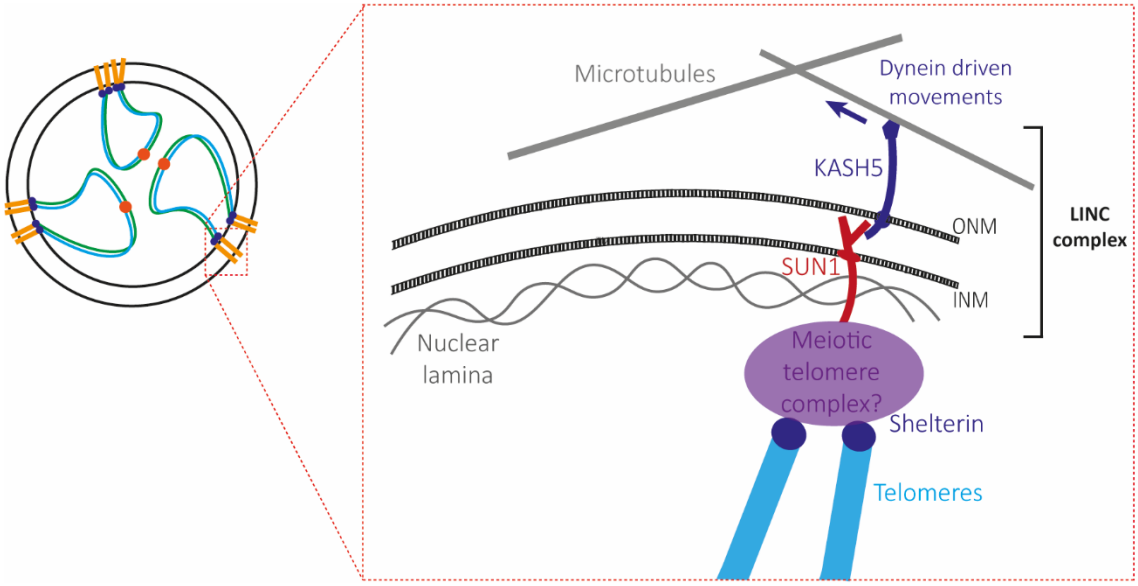


Figure 3.1.7.| Connecting the telomeres to the cytoskeletal network.
Telomere ends attach to the nuclear envelope during zygotene of prophase I to facilitate rapid prophase movements of meiotic chromosomes. Cytoskeletal forces generated by dynein-dynactin power chromosome movements, which is achieved by a cooperative protein interaction network between the LINC complex and the meiotic telomere complex.

The MTC was first characterised in the fission yeast *S. pombe* (Figure 3.1.8.) The meiosis-specific proteins Bqt1 and Bqt2 were identified to cooperate with each other in order to interact with the shelterin protein Rap1. The SUN-domain protein, Sad1 is recruited to Rap1 bound telomeres through an interaction with Bqt1-Bqt2 therefore forming a bridge between the telomere ends (Rap1) and the SPB

(Sad1). Disruption of Rap1 binding to Bqt1-Bqt2 complexes causes defects in telomere clustering and chromosome segregation (Chikashige et al., 2006; Moiseeva et al., 2017).

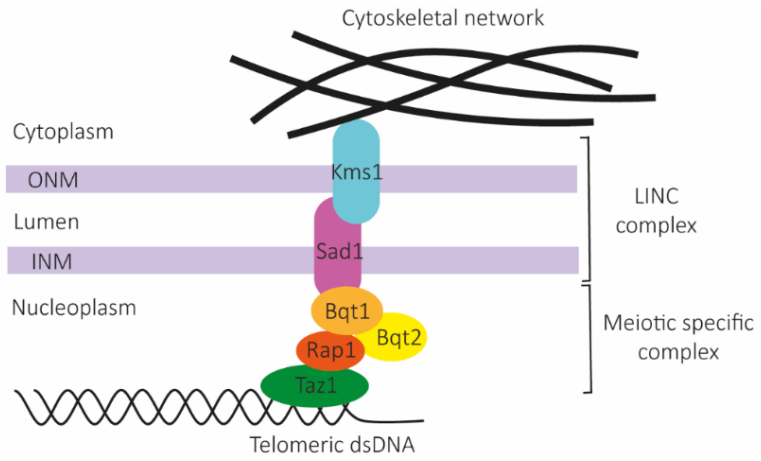


Figure 3.1.8.| Schematic of the *S. pombe* meiotic telomere complex. The yeast *S. pombe* LINC complex components Kms1 and Sad1 span across the nuclear envelope. The KASH-domain protein, Ksm1 interacts with cytoskeletal network; microtubules in *S.pombe* or actin filaments in *S. cerevisiae*, which mediate the cytoskeletal forces for chromosome movements. Sad1 interacts directly with Ksm1as well as connecting the telomeres by a mediating an interaction with the meiosis specific complex. The shelterin protein Rap1 binds to the meiosis specific Bqt1-Bqt2 complex, which also interacts with Sad1, thus connecting the telomeres to the cytoskeletal network.

Table 3.1.1.| Components connecting chromosomes to the cytoskeleton across eukaryotic species.
Despite there being very low or no sequence conservation in the meiotic proteins across species, all four species use the same general mechanism in tethering telomeres (pairing centres in *C. elegans*) to the nuclear membrane during prophase I of meiosis.

	Components connecting chromosomes to the cytoskeleton			
	S. pombe	*C. elegans*	*S. cerevisiae*	*M. musculus*
KASH-domain protein (ONM)	Kms1	ZYG-12	Cms4	Kash5
SUN-domain protein (INM)	Sad1	SUN-1	Mps3	Sun1, Sun2
Meiosis-specific connector	Bqt1, Bqt2	HIM-8, Zim proteins	Ndj1	Terb1
Chromosomal loci tethered	telomeres	pairing centres	telomeres	Telomeres
Motor	dynein, dynactin, kinesin	dynein	actin motors	dynein, dynactin

3.1.8. TERB1 functions as a molecular scaffold within the mammalian meiotic telomere complex

The gene encoding coiled-coil domain containing 79 (*ccdc79*) was found to specifically localise to several punctuated dots in mice spermatocytes, but not in mitotic germ cells, suggesting the encoded protein is meiosis specific. Furthermore GFP-tagged CCDC79 localised to chromosome ends and colocalise with the shelterin component TRF1, from leptotene to diplotene during prophase I. CCDC79 was then renamed telomere repeat binding bouquet formation protein 1 (TERB1), due its importance in bouquet formation (Daniel et al., 2014). Further to TERB1, two additional meiosis-specific proteins have been identified by co-immunoprecipitation studies (co-IP) of mouse testis extract: telomere repeat binding bouquet formation protein 2 (TERB2) and membrane anchored-junction protein (MAJIN) (Shibuya et al., 2015). Individual disruption of either of these three proteins in mice leads to meiotic arrest with failure of telomere attachments, chromosomal movements, failure of DNA double strand break repair and disordered synapsis (Figure 3.1.9.). Therefore, the MTC is crucial for meiotic progression throughout prophase I, positioning telomeres onto the INM forming a stable chromosome-INM structure known as the telomere attachment plate (Shibuya et al., 2015).

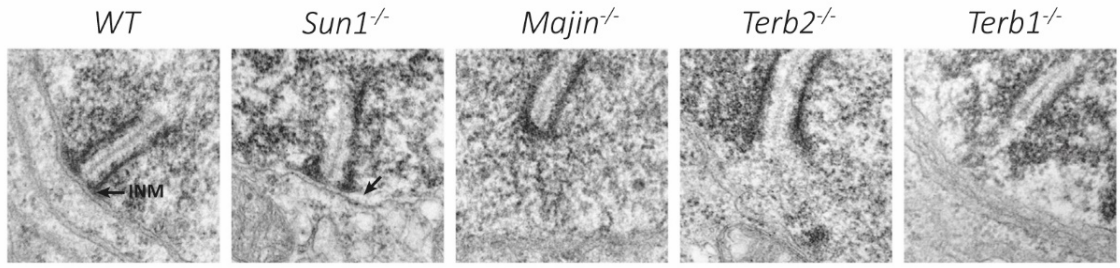


Figure 3.1.9.| Formation of the meiotic telomere attachment plate in mice spermatocytes.
a) Electron micrograph observations in male mice spermatocytes analysed the status of the telomere attachment plate in wild type (WT) spermatocytes and a series of meiotic telomere complex protein mutants. In WT and *Sun1*$^{-/-}$ spermatocytes telomeric DNA is embedded in the attachment plate, forming a clear structure of electron dense of synapsed lateral elements leading to a capping structure at the INM (shown by a black arrow). These complete structures are absent in *Majin*$^{-/-}$, *Terb2*$^{-/-}$ and *Terb1*$^{-/-}$ spermatocytes. In *Majin*$^{-/-}$ spermatocytes the capping structure is still present but is no longer located at the INM. In both the *Terb2*$^{-/-}$ and *Terb1*$^{-/-}$ spermatocytes the capping structure is disrupted however the LEs are still preserved. Figure taken from Shibuya et al., 2015.

In mice, binary interactions between the meiotic telomere proteins have been mapped through yeast two hybrid (Y2H) screens and *in vivo* interaction studies. TERB1 functions as a molecular scaffold within the MTC that localises to telomeres through its central, conserved TRF1-binding domain (TRFB), simultaneously binding the shelterin protein, TRF1 as well as TERB2 (Figure 3.1.10.a. and 3.1.10.b.) (Pendlebury et al., 2017; Shibuya et al., 2015; Shibuya & Watanabe, 2014). Coordinated assembly of the MTC is required for downstream processes, however, little is known about the underlying mechanisms by which this occurs.

Localisation studies have shown that TERB1 foci reside to telomere ends and co-localise with both TRF1 and SUN1 (Figure 3.1.10.d. and 3.1.10.e.). Furthermore, a Y2H screen determined that TERB1 N- and C-termini bind to the N-terminius of SUN1 and a meiosis-specific cohesion complex subunit, STAG3, respectively (Figure 3.1.10.c.) (Daniel et al., 2014; Shibuya & Watanabe, 2014). During prophase I, cohesins are essential for the completion of homologous chromosome pairing, recombination in addition to meiotic chromosome axis formation, further explained in 1.1.5. (Brar et al., 2009). *Terb1*$^{-/-}$ cells expressing a GFP-TERB1$^{\Delta Myb}$ showed structural defects in telomeres, suggesting that there is functional significance of cohesins within the MTC. In addition, it has been suggested that cohesins play a role protecting the structural integrity of the telomere during chromosome movements (Zhang et al., 2017).

3.1.9. TRF1-TERB1 interaction resembles TRF1-TIN2

The interaction between the homodimerisation domain of TRF1 ($TRF1_{TRFH}$) and a TERB1 peptide, amino acid residues 642-656, herein known as the TRF1 binding motif ($TERB1_{TBM}$) has been characterised both structurally and biochemically (Long et al., 2017; Pendlebury et al., 2017). From these studies it was determined that TERB1 uses a similar strategy as the shelterin complex TIN2 to bind TRF1. Sequence analysis has shown that TERB1 contains a consensus cyclin-dependent kinase (CDK) motif which overlaps with the FxLxP motif also found in TIN2, therefore suggesting that TERB1 may bind using a similar mechanism (Figure 3.1.11.d.).

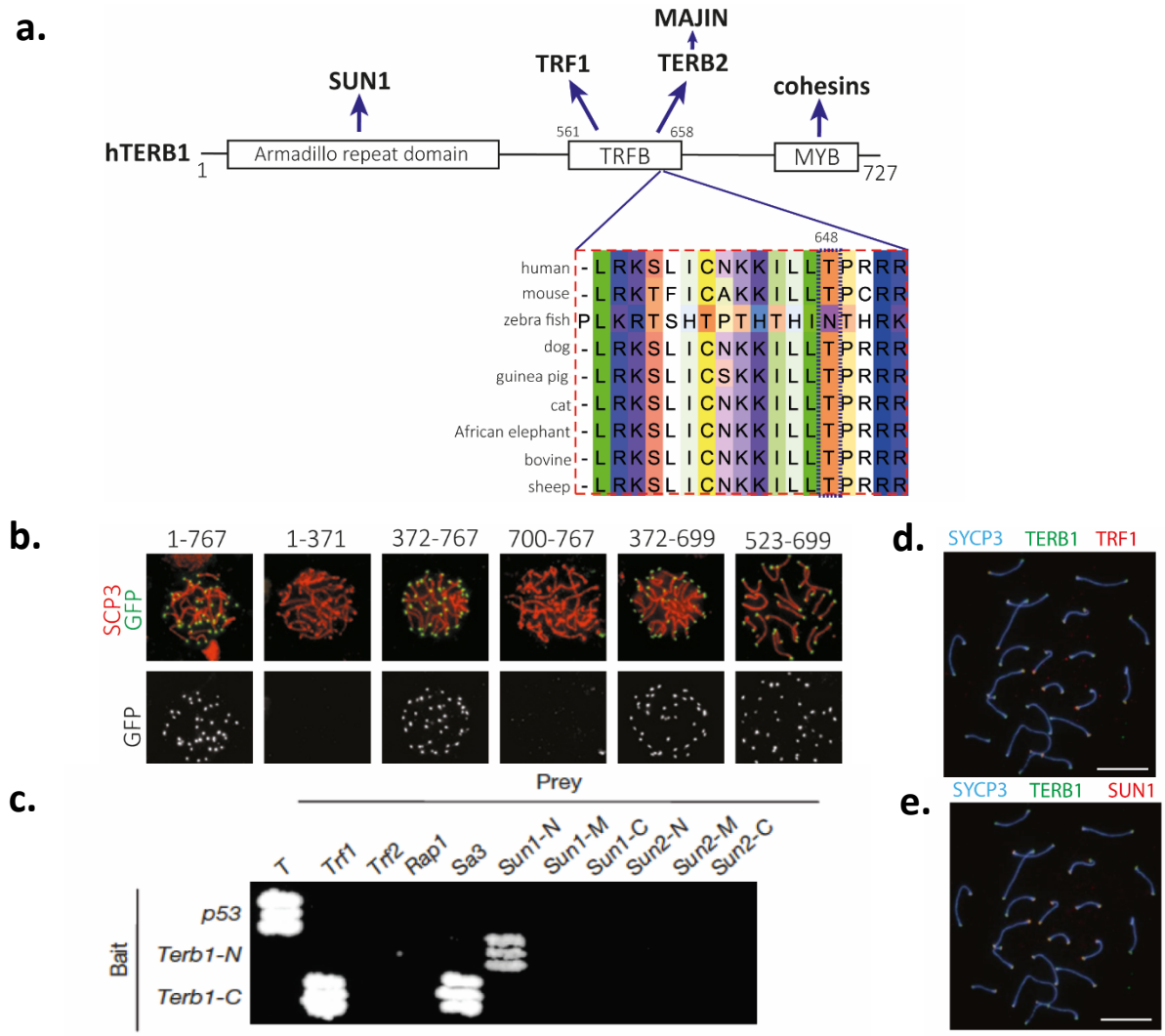


Figure 3.1.10.| TERB1 is a meiosis-specific protein

a) TERB1 has three conserved domains: N-terminal armadillo repeat domain, C-terminal myb domain and a central TRF binding (TRFB) domain that mediate different protein-protein and possibly protein-DNA interactions. The TRFB domain of TERB1 mediates the interaction with the shelterin protein TRF1. The C-terminus of the TRFB domain (amino acid residues 648-651) contains a highly conserved CDK consensus motif, (S/T)PX(K/R), this also includes the conserved threonine phosphorylation site (T648). It is also known that TRFB domain of TERB1 contains the TERB2 binding site. The N-terminal armadillo repeat domain is thought to be the SUN1 binding domain. Similar to TRF1, TERB1 also has a Myb-like domain which may bind to cohesins. b) A series of TERB1 truncations determined a minimum TERB1 construct of amino acid residues 523-699 required to localise to telomeres, suggesting this is the minimum TRF1 binding site. c) Yeast two hybrid screen determined that the N-termini armadillo repeat domain of TERB1 interacts specifically to the N-terminus of SUN1. TERB1 C-terminal Myb-like domain interacts with the cohesion protein Stag3 (Sa3). d-e) Immunofluorescence (IF) imaging of mice spermatocyte spreads at pachytene stage of prophase I. By pachytene stage of prophase the synaptonemal complex is fully formed along the lengths of synapsed chromosomes, shown bt the lateral element protein, SYCP3 coating the chromosomes. d) TRF1 foci (red) reside at telomere ends, and TERB1(green) colocalises with TRF1. e) TERB1 foci also colocalise with the LINC complex protein SUN1 (red) at chromosome ends, showing that TERB1 colocalises with both SUN1 and TRF1. Figures taken from Shibuya, 2014 (b)(c) and Daniel *et al.,* 2014 (d)(e).

Crystal structures of $TRF1_{TRFH}$- $TERB1_{TBM}$ were achieved through soaking $TERB1_{TBM}$ into TRF1 crystals. The crystal structure determined a 2:2 oligomeric stoichiometry; each TRF1 monomer binding a short TERB1 peptide (Figure 3.1.11.a.). Furthermore, the structure revealed that TERB1 binding to $TRF1_{TRFH}$ is reminiscent of $TRF1_{TRFH}$-$TIN2_{TBM}$ binding, in which TRF1 recognises the IxLxP motif on TERB1 via the peptide binding site in its TRFH domain (Figures 3.1.3.b, 3.1.11.b. and 3.1.11.c.) (Long et al., 2017; Pendlebury et al., 2017).

Preliminary studies within our group, using SEC-MALS analysis, of the TRF1-TERB1 interaction using the wider TRF binding domain (TRFB) of TERB1 revealed an oligomeric state of 2:1 for the $TRF1_{TRFH}$ -$TERB1_{TRFB}$ complex (performed by Mr. Lee Thung Sen). This suggests that a single TERB1 molecule binds to a TRF1 homodimer. Although this is in disagreement with previous structural data, the crystal structure of $TRF1_{TRFH}$-$TERB1_{TBM}$ complex indicates that larger peptide could easily be accommodate as there is 34.5 Å between the short $TERB1_{TBM}$ peptides, suggesting that the longer TERB1 construct could easily be accommodated, spanning both regions, and this is likely the true mechanism of binding. (Figure 3.1.11.e.).

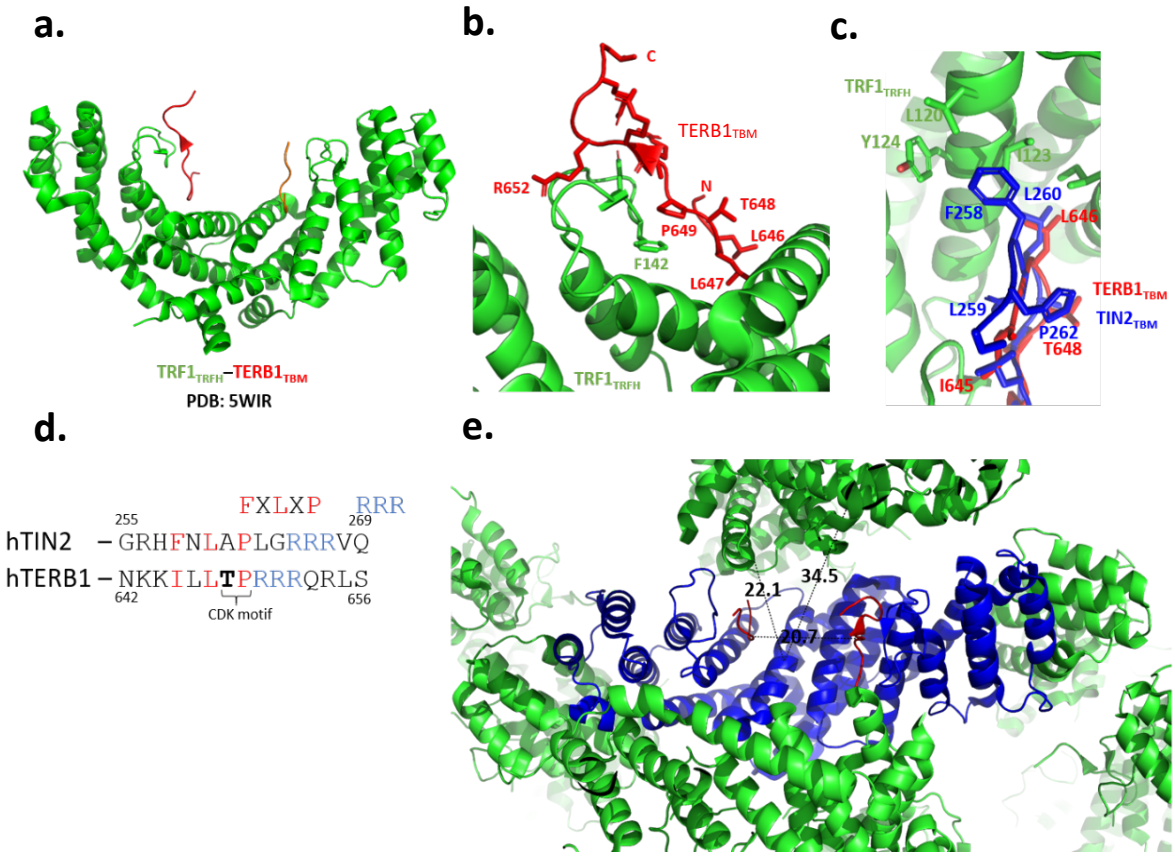


Figure 3.1.11.| Crystal structure of $TRF1_{TRFH}$-$TERB1_{TBM}$
a) The $TRF1_{TRFH}$-$TERB1_{TBM}$ structure was solved to 2.1Å through soaking $TERB1_{TBM}$ peptide (amino acid residues 642-656) into $TRF1_{TRFH}$ crystals and using molecular replacement. The crystal structure revealed a 2:2 stoichiometry reminiscent of $TRF1_{TRFH}$-$TIN2_{TBM}$. Each TERB1 peptide binding to a TRF1 monomer does not interfere with second binding site or TRF1 dimerisation. b) TERB1 peptide binds in an extended conformation across the concaved surface of each $TRF1_{TRFH}$ monomer.The interaction between TRF1 and the TERB1 peptide is stablised through hydrophobic associations. c) Superposition of the $TRF1_{TRFH}$-$TERB1_{TBM}$ (PDB: 5WIR) and $TRF1_{TRFH}$-$TIN2_{TBM}$ (PDB: 3BQO) structures show strong similarities between the two complexes. In both complexes two TBM peptides bind to the TRF1 homodimer. d) Both TERB1 and TIN2 bind to TRF1 using the I/FxLxP motif, however there are some small differences. TIN2 F258 binds the TRF1 hydrophobic pocket, but in contrast the equivalent TERB1 I645 is excluded from this pocket, and instead L646 partially occupies the hydrophobic pocket. TERB1 and TIN2 LxP residues show similar conformations. e) The crystal packing of TRF1-TERB1(blue) shows that at the dimerisation interface there is space within the packing for a larger TERB1 construct.

3.1.10. Meiosis specific nuclear membrane complex, MAJIN-TERB2

MAJIN is a putative transmembrane protein localised at the inner surface of the NE via a C-terminal transmembrane domain. Ectopic expression of MAJIN in U-2 OS cells shows clear localisation to the NE, unlike TERB2 which displays a diffused distribution when expressed alone. When both proteins are co-expressed MAJIN can recruit TERB2 to the NE, suggesting a direct interaction between the two proteins (Shibuya et al., 2015; Y. Wang et al., 2019). The N-terminal structural core of MAJIN (amino acid residues 1-112) was recombinantly expressed with the C-terminus of TERB2 (amino acid residues 168-220) and co-purified to form a equimolar complex. The presence of both proteins was essential for stability suggesting the complex is constitutive. SEC-MALS analysis determined that the absolute molecular weight corresponded to a 2:2 heterotetramer. The crystal structure of $MAJIN_{1-112}$-$TERB2_{168-220}$ ($MAJIN_{core}$-$TERB2_C$) complex has been recently solved to 2.9Å by our laboratory (work performed by Dr James Dunce and Mr Gurusaran Manickam). A further optimised complex, $MAJIN_{1-106}$-$TERB2_{168-207}$, was then solved using molecular replacement to 1.85Å. The 2:2, MAJIN-TERB2 structure entails two TERB2 chains wrapped around a core MAJIN globular dimer (Figure 3.1.12.a.). MAJIN forms a β-grasp domain where a β-sheet grasps around an α-helix of TERB2 (Figure 3.1.12.b.). The interaction is largely mediated through by the hydrophobic side chains of TERB2 inserting into pockets on the MAJIN molecular surface (Figure 3.1.12.c.). Furthermore, analysis by size-exclusion

chromatography linked to small angle x-ray scattering (SEC-SAXS) determined scattering curves and *ab initio* envelopes that closely match the crystal structure (Figure 3.1.12.d. and 3.1.12.e). Mutational analysis of MAJIN (F73C and Y75E) at the MAJIN-TERB2 dimerisation interface disrupted the 2:2 assembly, forming a 1:1 complex. This intriguing result shows that a single MAJIN-TERB2 protomer can correctly fold independent of dimerisation, suggesting that dimerisation of the MAJIN-TRB2 complex may play a role in the wider architecture of the meiotic telomere complex (Dunce, Milburn, et al., 2018).

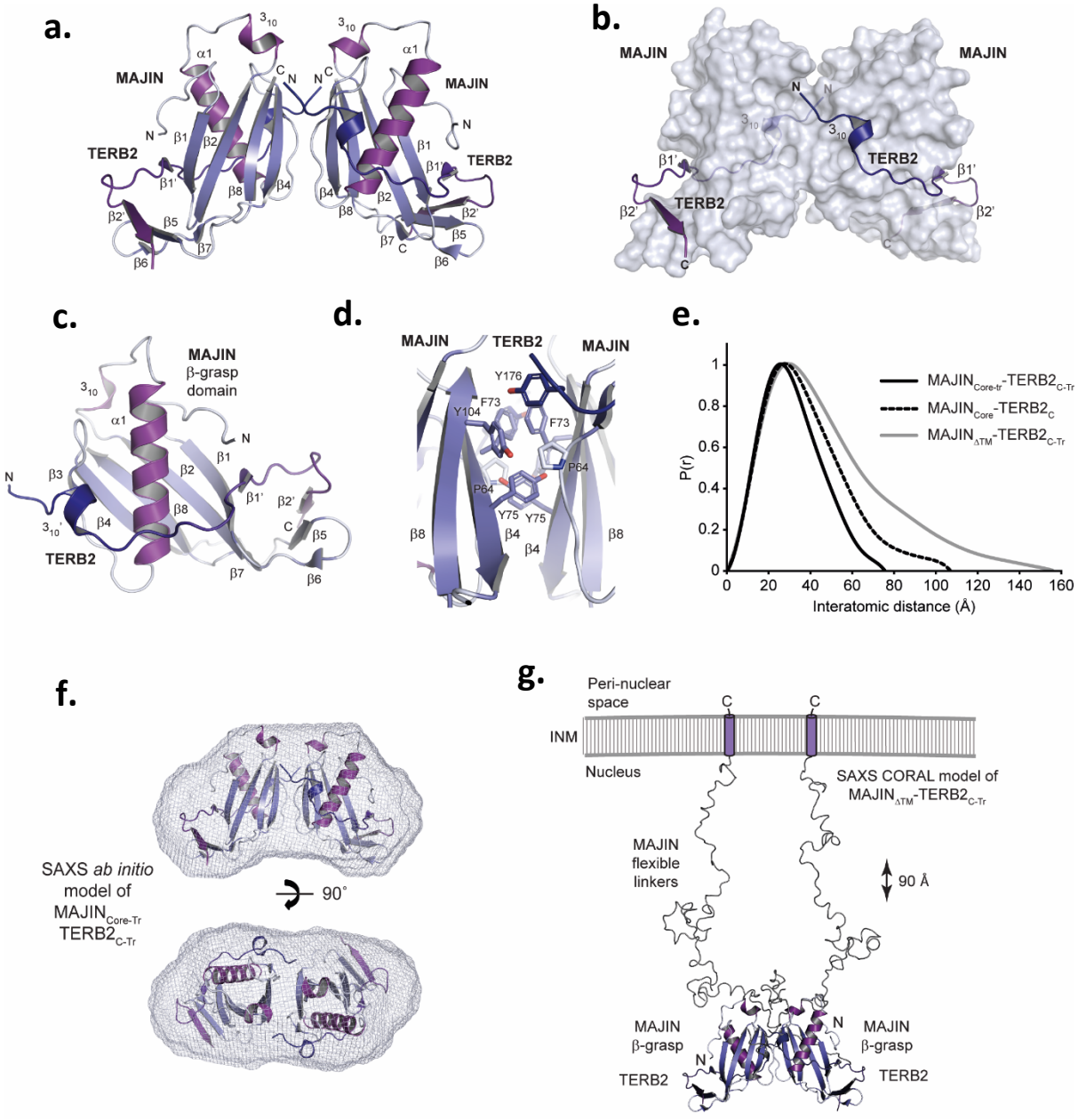


Figure 3.1.12.| MAJIN-TERB2 structure.
a) The 1.85Å crystal structure of the $MAJIN_{1-106}$-$TERB2_{168-207}$ core reveals a 2:2 hetero-tetrameric complex in which two TERB2 polypeptide chains meanders around a globular MAJIN dimer. b) Molecular surface of MAJIN with cartoon representation of TERB2 chains. c) MAJIN adopts a unique β-grasp fold, which entails a β-sheet wrapped around a central α-helix in a β(2)-α-β(3) configuration. The TERB2 chain wraps around the exposed surface of the α-helix forming a fully enclosed structure. d) The dimerisation interface of the MAJIN-TERB2 complex forms a large hydrophobic consisting of aromatic and hydrophobic interactions between P64, F73, Y75 and Y104 (MAJIN), and Y176 (TERB2). Mutations of aromatic residues, F73E and Y75E, disrupts the dimer interface. The symmetrical dimeric core complex is likely to act as the centre of the meiotic telomere complex, facilitating the formation of a wider complex. e) SEC-SAXS *P(r)* distribution of the core $MAJIN_{1-106}$-$TERB2_{168-207}$ complex (black solid) and slightly longer construct, $MAJIN_{1-112}$-$TERB2_{168-220}$ (black dashed) gave a maximum dimension of 80Å and 120Å, respectively. Increasing the length of the MAJIN construct (amino acid residues 1-233) to include the flexible region drastically increases the interatomic distance of the complex to 150 Å. f) SEC-SAXS DAMMIF *ab initio* envelope of $MAJIN_{1-106}$-$TERB2_{168-207}$ with docking of the crystal structure, showing that the model is a good match to the crystal structure. g) SAXS coral model of the wider $MAJIN_{1-233}$-$TERB2_{168-207}$ complex, combined with the solved crystal structure reveals that MAJIN flexible linkers between the structural core and the TM region may provide a distance of ~90Å between the NE and the MAJIN-TERB2 core. When stretched this separation could increase to up to 400Å. Figures made by Dr Owen Davies and taken from Dunce, Milburn *et al*., 2018.

SEC-SAXS analysis was also performed on a wider MAJIN-TERB2 complex, where the C-terminus of MAJIN was extended (amino acid residues 1-233). The scattering data for this complex in addition to the crystal structure of the core complex was used to model the larger complex. From the scattering data it was proposed that MAJIN linkers constitute flexible extensions, up to 400Å when stretched, between the MAJIN-TERB2 structural core and the TM domain embedded in the INM (Figure 3.1.12.f.) (Dunce, Milburn, et al., 2018).

The MAJIN-TERB2 structure revealed that the complex provides multiple DNA-binding surfaces suggesting a direct interaction with DNA. There is an extensive patch of highly conserved basic amino acids on the surface of each MAJIN protomer that enhance the binding of DNA. Moreover, the C-terminal end of TERB2 also includes highly conserved basic residues, which have been shown to be crucial for DNA binding (Figure 3.1.13.a.). Electromobility shift assays (EMSA) were used to determine the DNA binding affinity of MAJIN-TERB2 to an apparent K_D of 0.12μM, matching the DNA binding affinity of TRF1. From these results it was proposed that telomeric DNA is looped around MAJIN-TERB2 molecules to achieve cooperativity through binding to the interfaces of both protomers (Figures 3.1.13.b. and 3.1.13.c.). The MAJIN-TERB2 complex serves as the architectural core of the

meiotic telomere complex, establishing a physical linkage to the NE and allowing for further interactions to develop from this (Dunce, Milburn, et al., 2018).

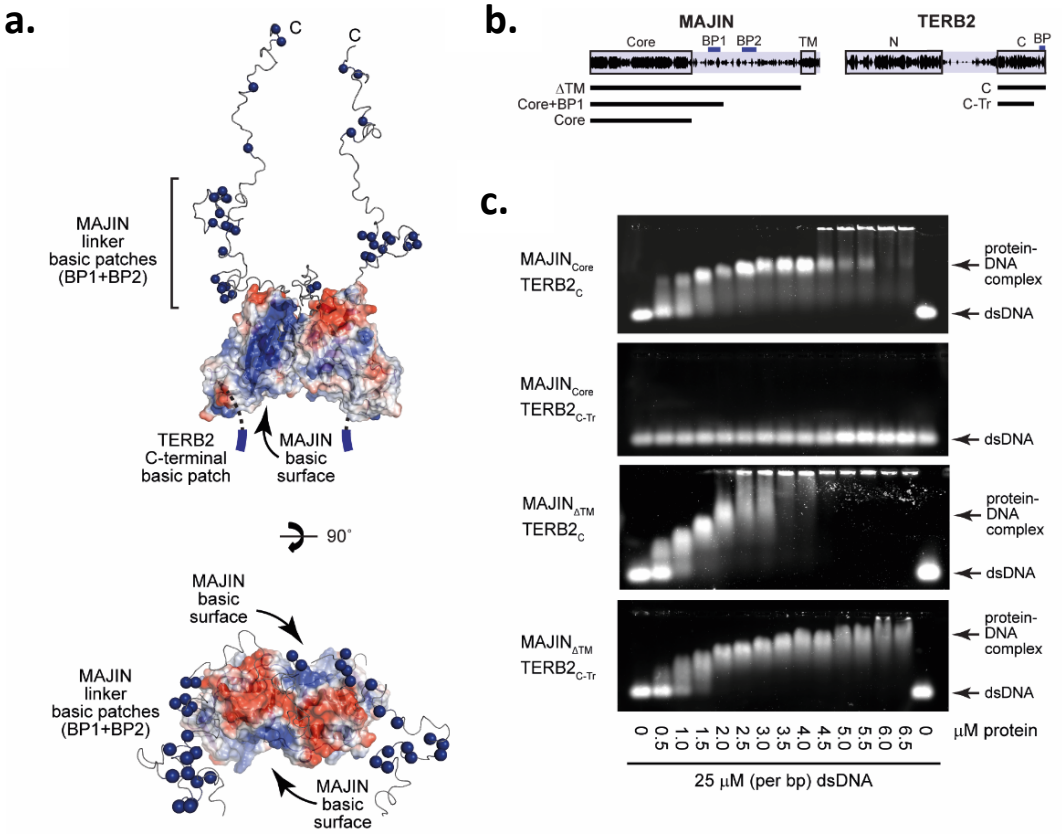


Figure 3.1.13.| DNA binding potential of the MAJIN-TERB2 complex
a) Surface electrostatic potential of the MAJIN-TERB2 core structure (red, electronegative; blue, electropositive), with the SAXS modelled MAJIN flexible linkers. Basic residues are shown by blue spheres. MAJIN contains a large basic patch on the surface of each protomer (red). In addition, the C-terminal end of TERB2 (amino acid residues 214-220) includes five highly conserved basic residues with a seven amino acid stretch suggesting that the MAJIN-TERB2 complex has DNA-binding potential. b) Schematic highlighting the protein constructs of MAJIN and TERB2 (MAJIN ΔTM (1–233), Core +BP1 (1–147) and Core (1–112); TERB2 C (168–220) and C-Tr (168–207)). The linker region of MAJIN contains patches of basic residues (BP1 and BP2). Also, TERB2 contains a basic patch (BP) at its extreme C-terminus which is deleted in the TERB2$_{C\text{-}Tr}$ construct. c) DNA binding by the MAJIN-TERB2 complex was shown using electromobility shift assays (EMSA). When the basic patches are present on both proteins (MAJIN$_{ΔTM}$-TERB2$_{C}$) DNA binding is enhanced matching the DNA binding affinity of TRF1.When the basic patches are deleted in either one of the proteins (MAJIN$_{ΔTM}$-TERB2$_{C\text{-}Tr}$ or MAJIN$_{CORE}$-TERB2$_{C}$) DNA binding is still present but reduced, indicating that both proteins bind to the DNA in a cooperative manner. Deleting the basic patches on both MAJIN and TERB2 (MAJIN$_{CORE}$-TERB2$_{C\text{-}Tr}$) abolishes DNA binding completely, as a shift is not observed when DNA is titrated into the protein complex. Figures made by Dr Owen Davies and taken from Dunce, Milburn *et al.*, 2018

3.1.11. The human meiotic telomere complex

Telomeres must be physically associated to with the nuclear envelope, which is achieved by an intricate interaction network between the shelterin complex, the meiotic telomere complex and the LINC complex. This complex connection facilitates the stable attachment of chromosome ends to the NE in prophase I and achieves successful homologous chromosome pairing and synapsis. Binary interactions between these complexes have been mapped by Y2H screens and co-immunoprecipitation (co-IP) assays (Figure 3.1.14.) (Shibuya et al., 2015).

The human MTC is comprised of three proteins, TERB1, TERB2 and MAJIN. As previously discussed the MAJIN-TERB2 constitutive complex is initially brought to the nuclear envelope by the MAJIN N-terminal transmembrane domains that reside within the INM (Dunce, Milburn, et al., 2018). In turn, TERB2 medicates association with TERB1 forming a stable tripartite meiotic telomere complex. The complex subsequently assembles onto telomeres when they come into close contact to the nuclear periphery in leptotene of early prophase I. This is achieved through the interaction between TERB1 and the shelterin complex protein TRF1 (Pendlebury et al., 2017; Shibuya et al., 2015; Y. Wang et al., 2019). Yeast two hybrid studies have proposed that the N-terminus of SUN1 interacts with the N-terminal armadillo repeat domain of TERB1, creating a direct link between the LINC complex and the meiotic telomere complex (Figure 3.1.10.c.) (Shibuya & Watanabe, 2014). The interaction network between the MTC, shelterin and the LINC complex is schematised in figure 3.1.14.

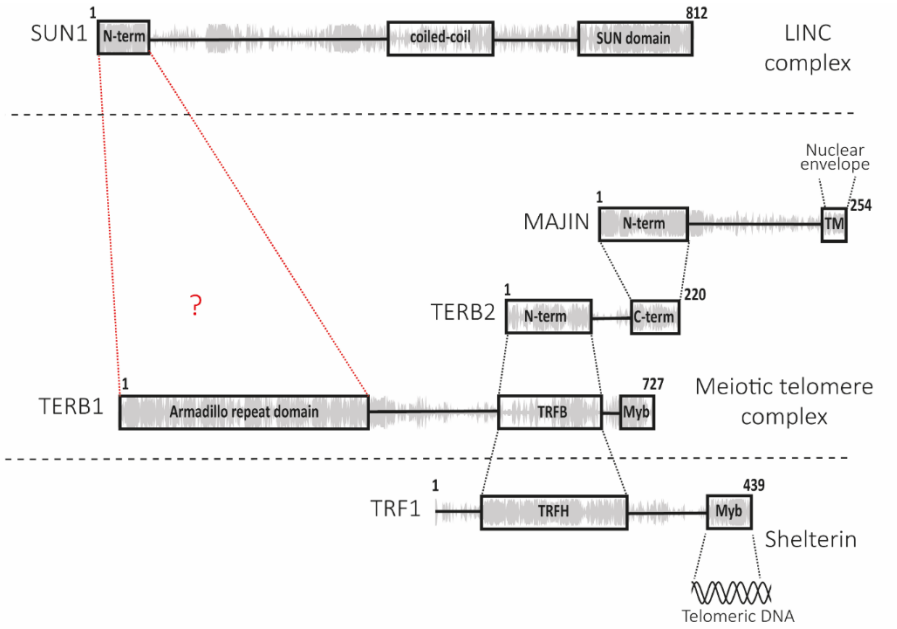


Figure 3.1.14.| Interaction network of the meiotic telomere complex.
The association of meiotic chromosomes to the nuclear envelope during prophase I of meiosis is achieved through the inter-molecular interactions between the meiotic telomere complex (MAJIN, TERB1 and TERB2), the LINC complex (SUN1 and KASH) and the shelterin complex component, TRF1. Binary interactions between these complexes have been mapped by yeast two hybrid screens and *in vivo* interaction and deletion studies. Shown in grey are the secondary structure conservation predictions of the individual proteins. Sequence conservation scores were determined per residue by Consurf using automated sequence alignments (Ashkenazy et al., 2016; Ashkenazy et al., 2010).

3.1.12. **Telomere cap exchange**

Previous studies have shown that the meiotic tripartite complex (MAJIN-TERB1-TERB2) interacts non-specifically with dsDNA, and the shelterin protein TRF1 binds telomeric dsDNA in a sequence-dependent manner (Dunce, Milburn, et al., 2018; Shibuya et al., 2015; Smogorzewska et al., 2000). This suggests that the addition of TRF1 to the meiotic telomere complex induces telomere-specific binding of the complex loading MAJIN-TERB2-TERB1 onto telomeric DNA. The mechanism involving telomere tethering to the INM has been coined telomere 'cap exchange' in which the telomere shelterin 'cap' is exchanged with the meiotic telomere complex thereby connecting the telomeres to the NE allowing chromosome movements (Shibuya et al., 2015). However, the structural mechanism in which TRF1 is displaced from the telomere is still largely unknown.

3.1.13. Chapter aims

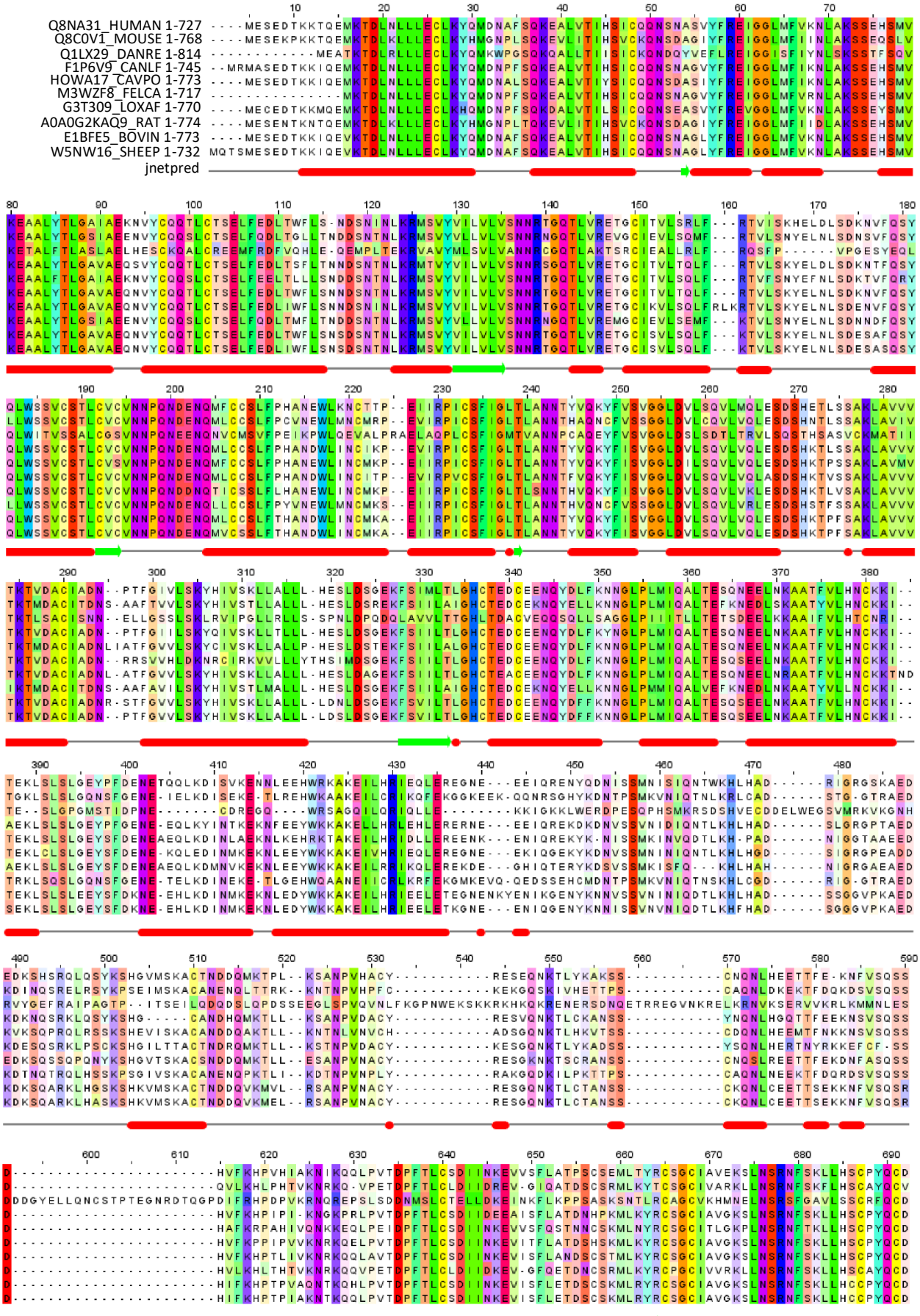
From the current literature it is currently poorly understood how the tripartite meiotic telomere complex is formed to facilitate the attachment of telomeres to the INM and to support the subsequent telomere movements. In this chapter, I aim to gain a full biophysical understanding of the molecular structure of the meiotic-telomere complex and its interplay with the LINC complex and shelterin in order to elucidate the mechanistic basis of telomere attachment and displacement during prophase I. Using a crystallographic and biophysical approach including SEC-MALS and SEC-SAXS in solution techniques I will characterise the structure of the TRF1-TERB1 complex, as well as the wider ternary TRF1-TERB1-TERB2 and MAJIN-TERB1-TERB2 complexes. Throughout this chapter, work carried out in collaboration with Dr James Dunce and Mr Gurusaran Manickam is clearly acknowledged within the text and relevant figure legends. The work presented in this chapter is part of a published paper in Nature Communications; 'Structural basis of meiotic telomere attachment to the nuclear envelope by MAJIN-TERB2-TERB1'2018, where I was joint first author.

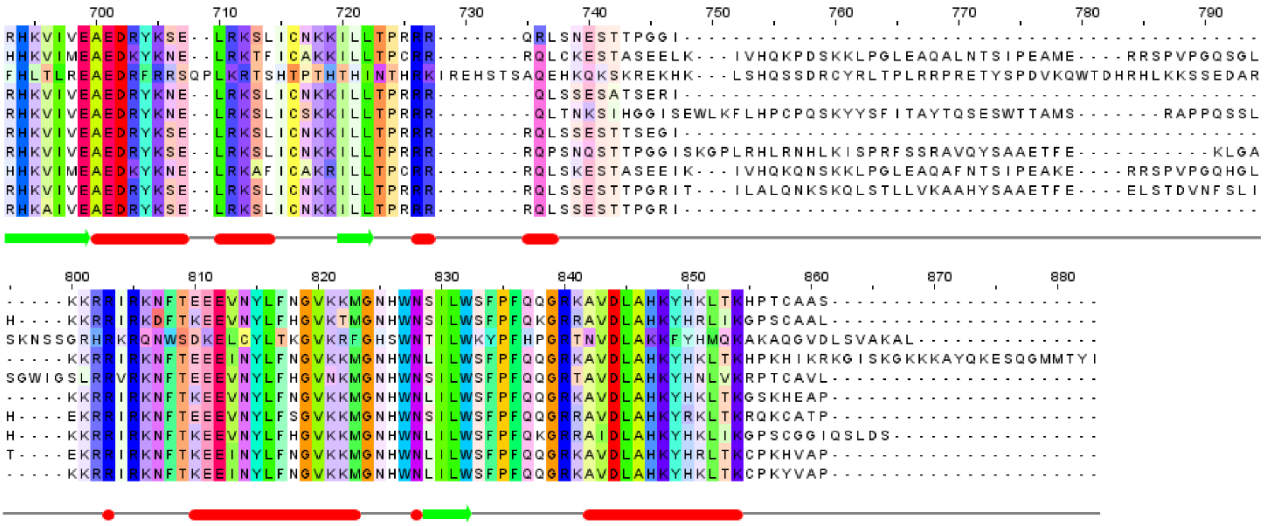
3.2. RESULTS

3.2.1. Bioinformatic and structural analysis of the meiotic telomere complex components

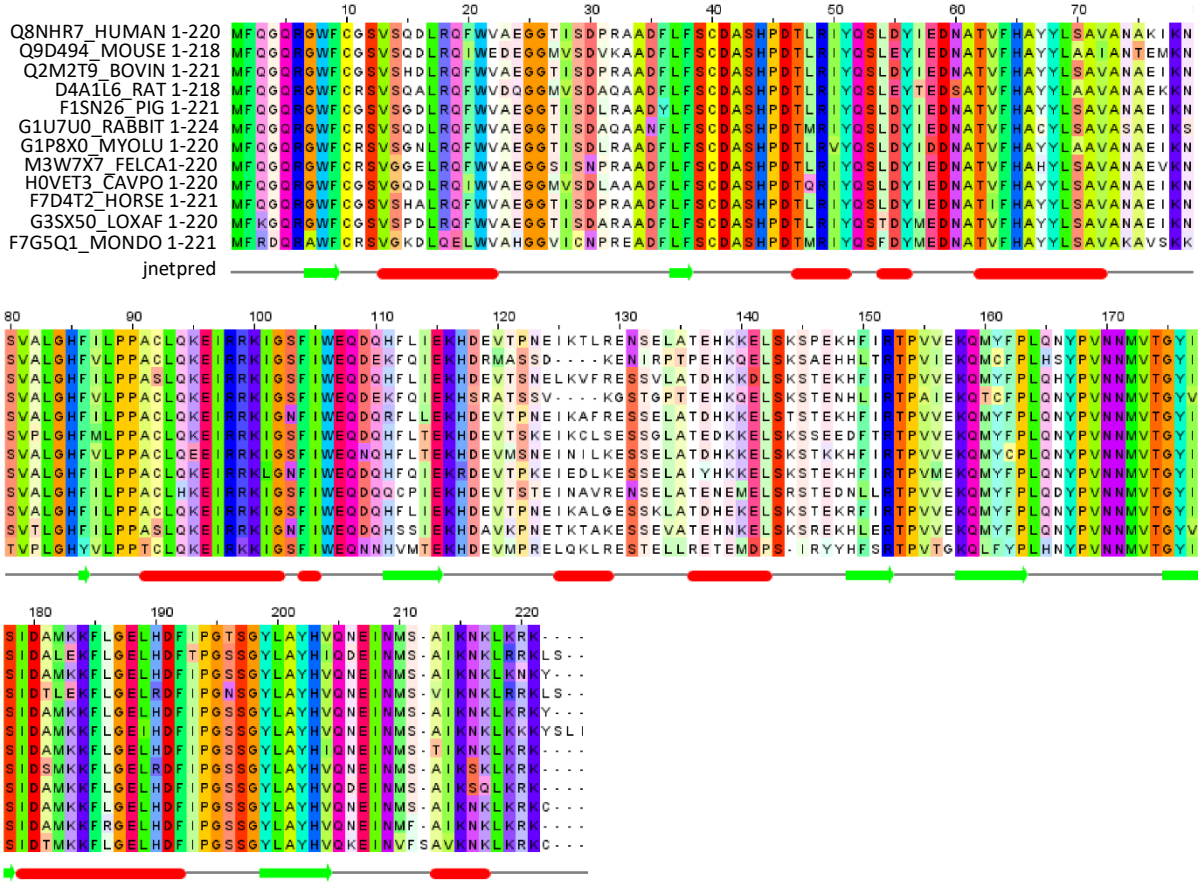
Constructs for the expression of the human meiotic telomere complex components (MAJIN, TERB1 and TERB2) and the shelterin complex component TRF1 were originally designed on their sequence conservation, secondary structure prediction, and known tertiary domain structures. Full-length multiple sequence alignments of TERB1, TERB2, MAJIN and TRF1 were performed by MUSCLE (Edgar, 2004) and JNetPRED (Drozdetskiy et al., 2015) was used for secondary structure prediction, schematised in Figures 3.2.1.a.-d. TERB1 is a 727 amino acid protein with a highly conserved N-terminal armadillo repeat domain (residues 1-424), immediately followed by a short potential coiled-coil region and a C-terminal Myb-like domain (residues 666-727). TERB1 interactions with TERB2 and TRF1 are achieved through a central TRF-binding domain (TRFB), predicted to consist of a mixed α-helical and β-sheet content (residues 561-658) (Figure 3.2.2.a.). TERB2 is a 220 amino acid protein which contains a central large unstructured region flanked by highly conserved N-terminal (residues 1-107) and C-terminal (residues 168-220) domains which interact with the TRFB domain of TERB1 and the C-terminal core domain (residues 1-112) of MAJIN, respectively (Figure 3.2.2.b.). As well as the core domain MAJIN has a C-terminal transmembrane (TM) domain, which localises in the nuclear envelope (Figure 3.2.2.c.). The shelterin complex protein TRF1 is a highly helical 439 amino acid protein, consisting of a central highly conserved homodimerisation (TRFH) domain (residues 62-268) which binds to TERB1, and a C-terminal Myb-domain that recognises telomeric repeat dsDNA (residues 375-432) (Figure 3.2.2.d.).

a. TERB1

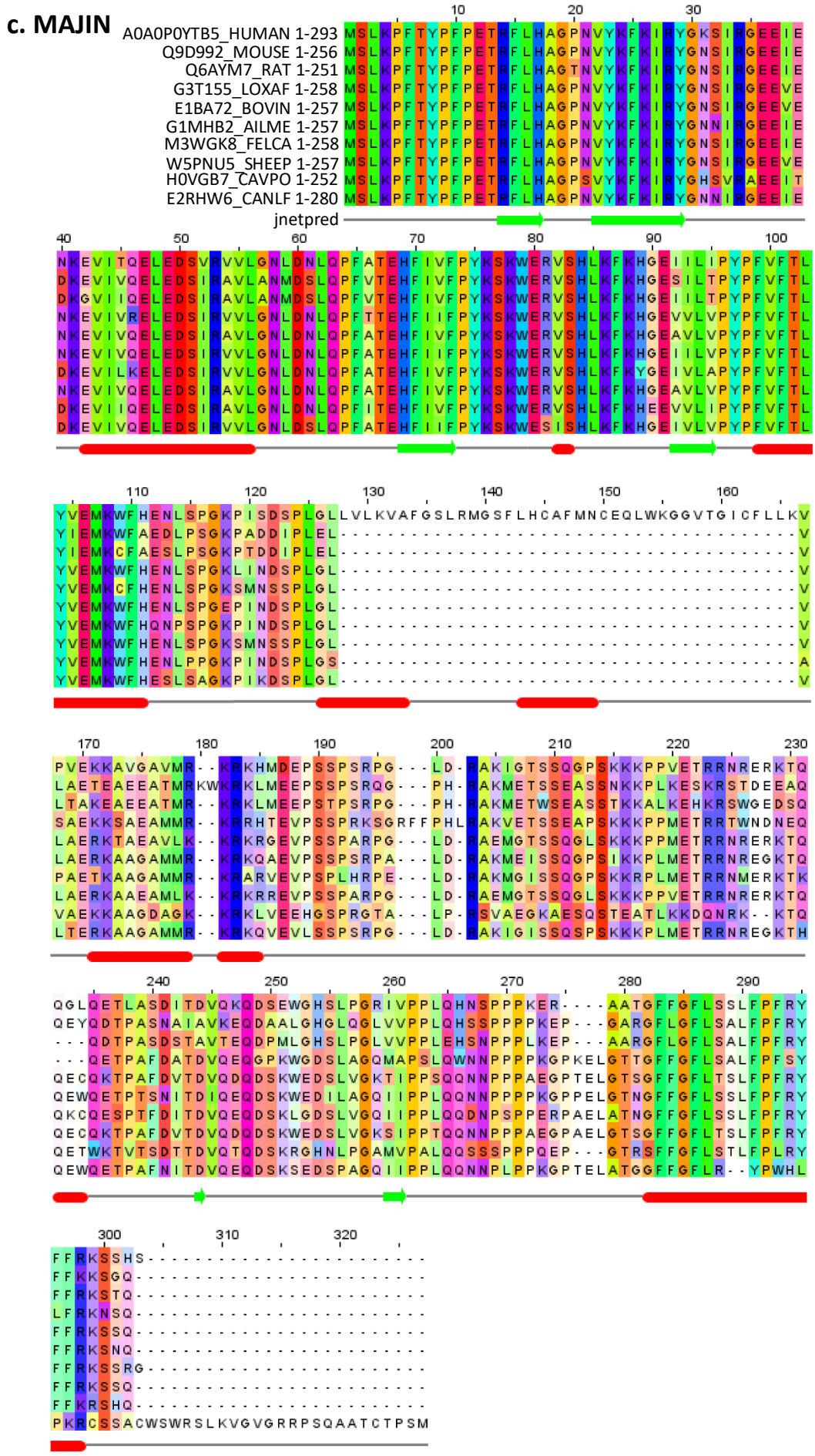

Q8NA31_HUMAN 1-727
Q8C0V1_MOUSE 1-768
Q1LX29_DANRE 1-814
F1P6V9_CANLF 1-745
HOWA17_CAVPO 1-773
M3WZF8_FELCA 1-717
G3T309_LOXAF 1-770
A0A0G2KAQ9_RAT 1-774
E1BFE5_BOVIN 1-773
W5NW16_SHEEP 1-732
jnetpred



b. TERB2


Q8NHR7_HUMAN 1-220
Q9D494_MOUSE 1-218
Q2M2T9_BOVIN 1-221
D4A1L6_RAT 1-218
F1SN26_PIG 1-221
G1U7U0_RABBIT 1-224
G1P8X0_MYOLU 1-220
M3W7X7_FELCA1-220
H0VET3_CAVPO 1-220
F7D4T2_HORSE 1-221
G3SX50_LOXAF 1-220
F7G5Q1_MONDO 1-221
jnetpred

c. MAJIN


A0A0P0YTB5_HUMAN 1-293
Q9D992_MOUSE 1-256
Q6AYM7_RAT 1-251
G3T155_LOXAF 1-258
E1BA72_BOVIN 1-257
G1MHB2_AILME 1-257
M3WGK8_FELCA 1-258
W5PNU5_SHEEP 1-257
H0VGB7_CAVPO 1-252
E2RHW6_CANLF 1-280
jnetpred

d. TRF1

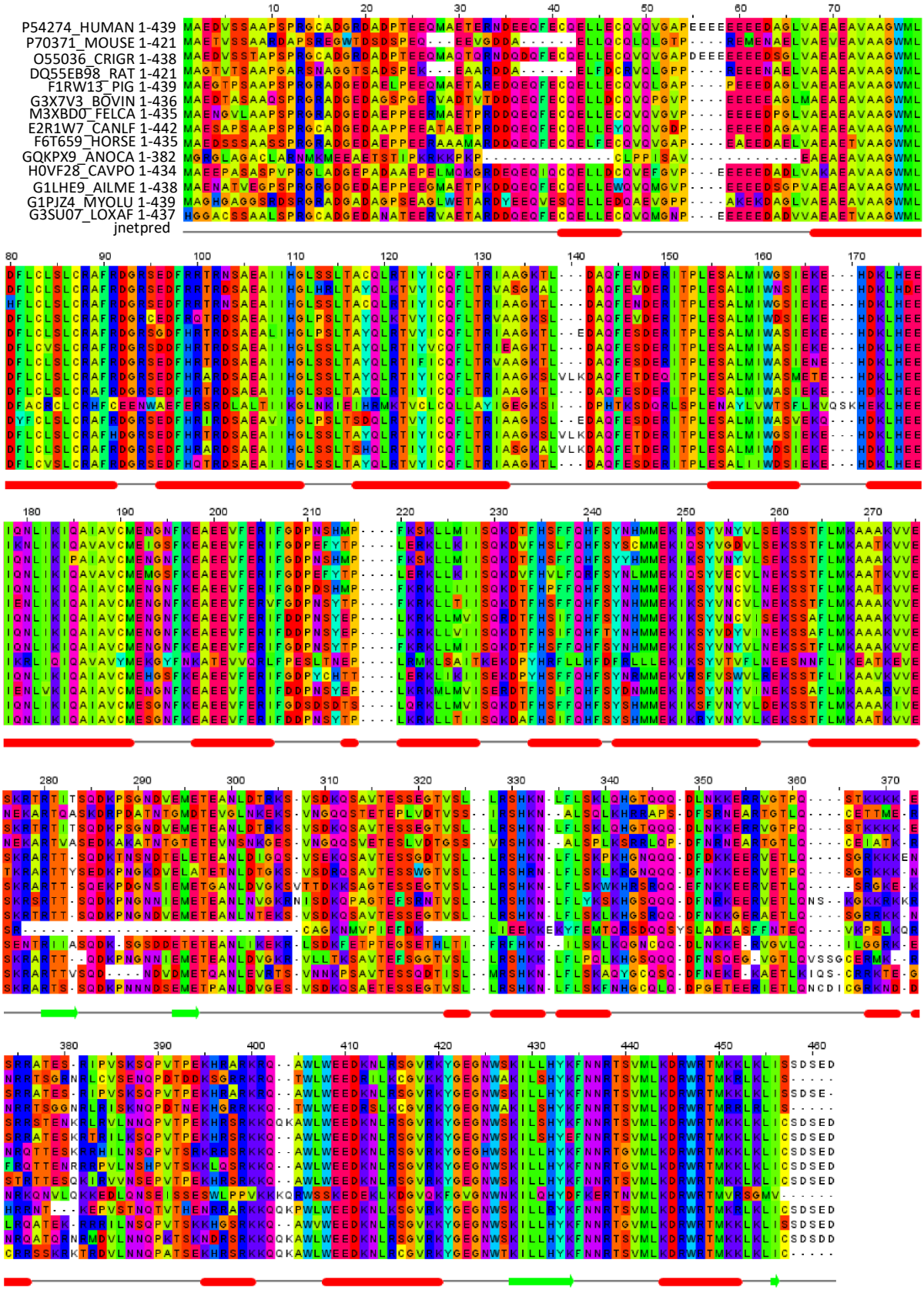

P54274_HUMAN 1-439
P70371_MOUSE 1-421
O55036_CRIGR 1-438
DQ55EB98_RAT 1-421
F1RW13_PIG 1-439
G3X7V3_BOVIN 1-436
M3XBD0_FELCA 1-435
E2R1W7_CANLF 1-442
F6T659_HORSE 1-435
GQKPX9_ANOCA 1-382
H0VF28_CAVPO 1-434
G1LHE9_AILME 1-438
G1PJZ4_MYOLU 1-439
G3SU07_LOXAF 1-437
jnetpred

Figure 3.2.1.| Seqeuence analysis of meiotic telomere complex proteins
Multiple sequence alignments generated by MUSCLE (Edgar, 2004) and secondary structure prediction performed by JNetPRED (Drozdetskiy et al., 2015) of the meiotic telomere complex proteins a) TERB1 b) TERB2 c) MAJIN and d) TRF1. Sequence analysis was visualised in Jalview and amino acids are coloured by conservation. The JNetPRED results are displayed beneath the alignment; alpha helices (red rods), beta sheets (green arrow) and unstructured regions (grey line).

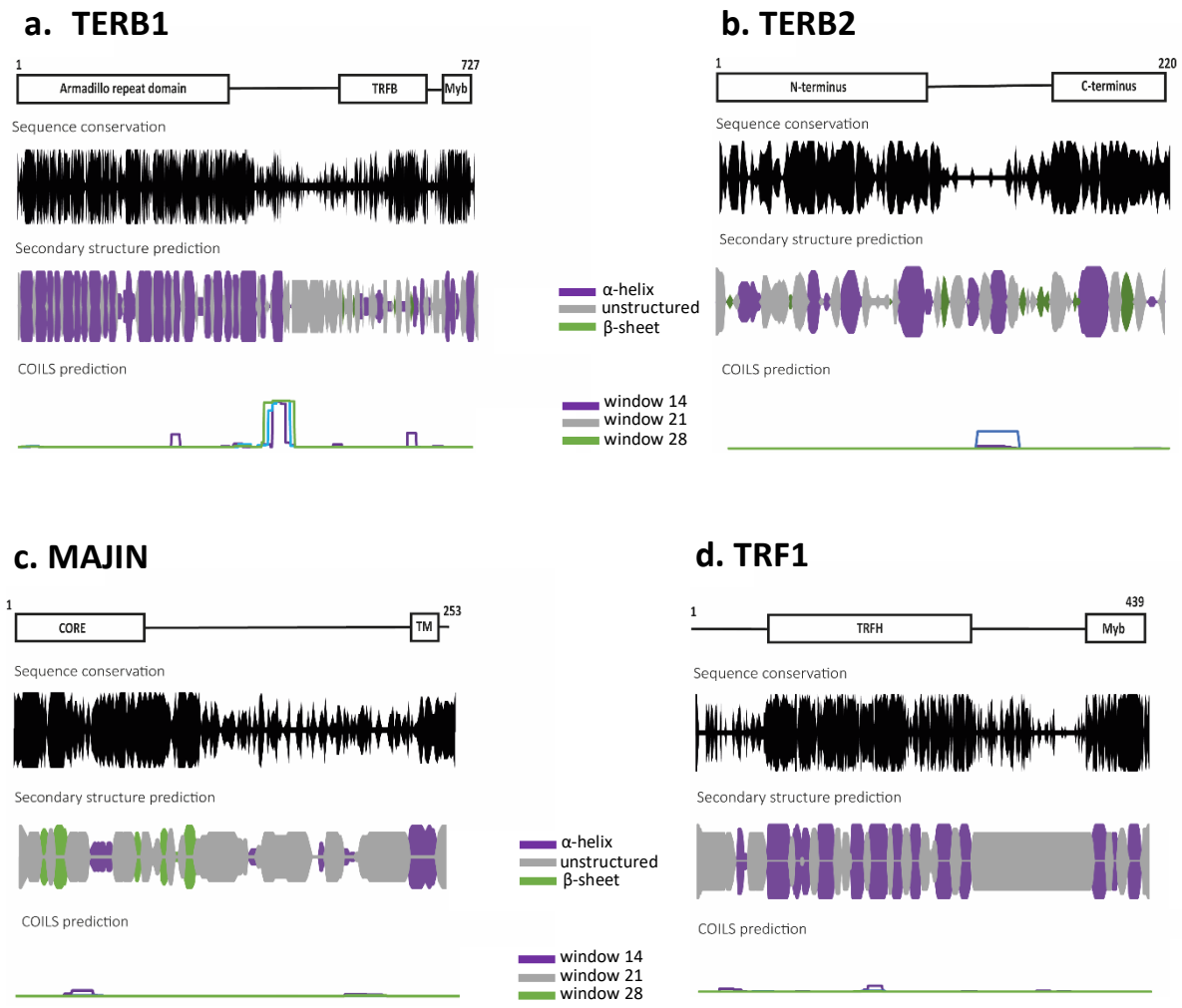


Figure 3.2.2.| Bioinformatic analysis of meiotic telomere complex components structures. Aligned sequence schematics showing the domain structures, sequence conservation, secondary structure prediction and COILS prediction of a) TERB1 b) TERB2 c) MAJIN and d) TRF1.) Aligned sequence schematics showing the domain structures, sequence conservation, secondary structure and COILS predictions. Sequence conservation scores were determined per residue by Consurf using automated sequence alignments (Ashkenazy et al., 2016). Secondary structure prediction was calculated by JPred4. α-helix in purple, β-sheet in green and unstructured in grey (Drozdetskiy et al., 2015). Coiled-coil conformation prediction was calculated by COILS which compared the sequence of interest to a database of known parallel two-stranded coiled-coils and derived a similarity score in a 14,21 or 28 amino acid window. (Lupas et al., 1991).

3.2.2. Purification and characterisation of full length TRF1

Full length TRF1, amino acid residues 1-439 ($TRF1_{FL}$), was expressed with in a N-terminal His_6-tag in *E.coli* protein expression strain BL21-DE3; cells where then lysed in a standard lysis buffer with the addition of 10% glycerol. Glycerol was included as we observed TRF1 had the tendency to aggregate in solution. All the purification steps using buffer containing 10% glycerol to minimise any aggregation and/or precipitation of TRF1. Ni-NTA affinity beads were ultilised to perform the initial protein purification step and the Ni-NTA 200mM imidazole eluate was stored on ice. We observed that $TRF1_{FL}$ precipitated when stored on ice but went back into solution at room temperature, therefore TRF1 was purified and stored the protein at room temperature. $TRF1_{FL}$ was further purified by anion exchange chromatography and the His_6-affinity tag was cleaved by enzymatic cleavage using TEV-protease. A second anion exchange step was used to separate $TRF1_{FL}$ from the His_6-tag and TEV protease, followed by size exclusion chromatography. An overview of the protein purification stages is summarised in Figure 3.2.3.a.

Size exclusion chromatography in line with multi-angle light scattering (SEC-MALS) analysis was used to determine the absolute molecular weight of the protein, from this we can confidently gain the oligomeric status. $TRF1_{FL}$ elutes as a single peak with a molecular mass of 102 kDa correlating to a dimer (dimeric theoretical molecular weight – 100 kDa) (Figure 3.2.3.b.). This finding suggests that the addition of the C-terminal Myb domain does not affect the dimerisation of TRF1.

We performed size-exclusion chromatography in line with small angle X-ray scattering (SEC-SAXS) analysis to determine low resolution structural information of $TRF1_{FL}$. SEC-SAXS analysis provides information of the shape and size (width and length) of the protein in solution. Transformation of the scattering data into real-space reveals the relative distribution of interatomic distances within a pairwise distance distribution profile known as the *p(r)* curve. The *p(r)* curve of $TRF1_{FL}$ has a profile typical of globular proteins with a smooth converged tail. The maximum interatomic distance (*Dmax*) can be identified by at which the curve intercepts the x-axis. $TRF1_{FL}$ has a *Dmax* of 130 Å (Figure 3.2.3.c.). The radius of gyration (*Rg*) provides a measure of the overall size of the molecule and can be directly obtained from the scattering data (real space *Rg*) or estimated using the Guinier approximation (Guinier

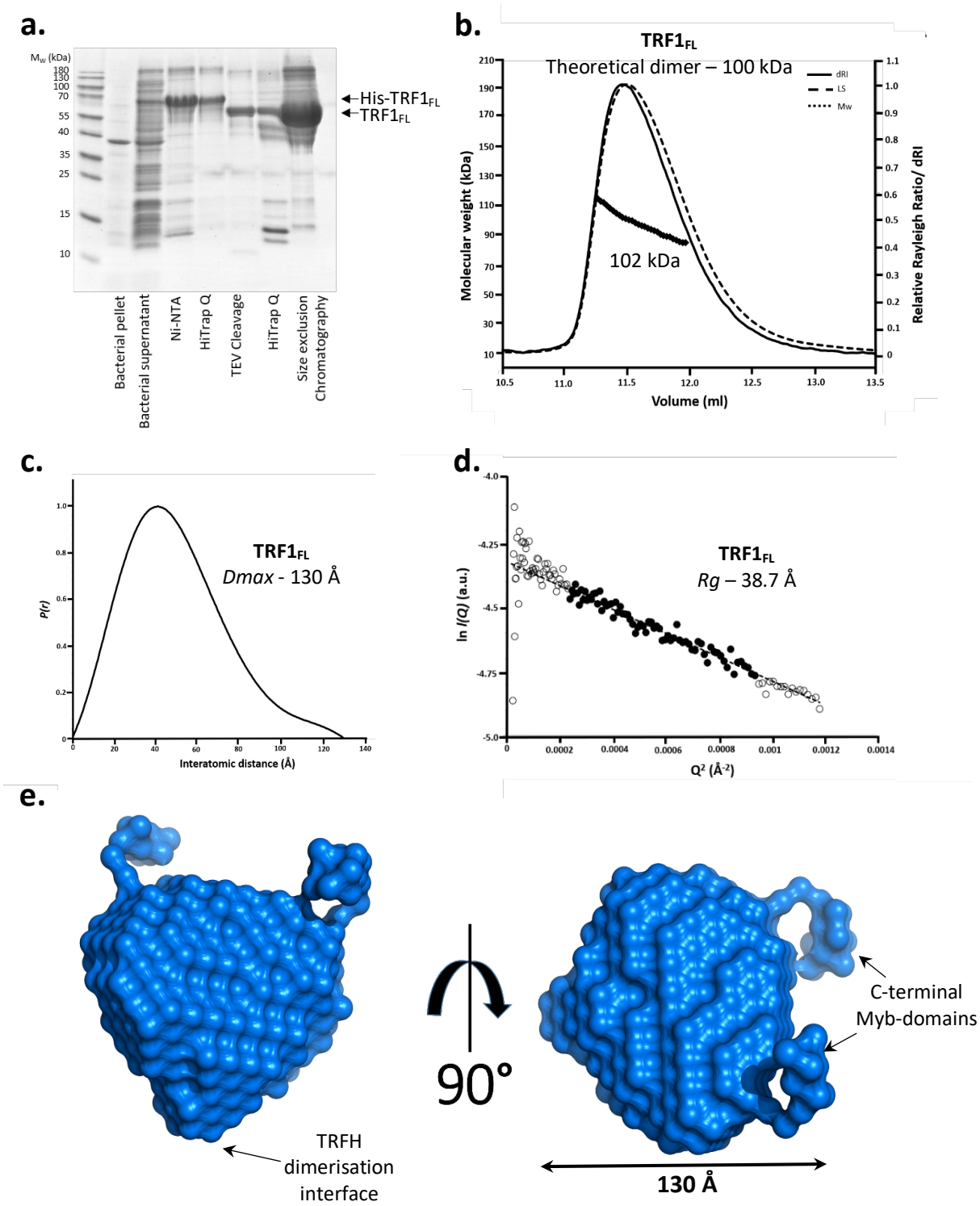


Figure 3.2.3.| Purification and structural analysis of $TRF1_{FL}$

a) SDS-PAGE showing the purification summary of $TRF1_{FL}$ through Ni-NTA and anion exchange chromatography. N-terminal His-tag was removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. b) SEC-MALS analysis showing a single peak with an experimental molecular weight of 102 kDa (theoretical dimer – 100 kDa) suggesting that full length TRF1 dimerises in solution. c-d) SAXS analysis of $TRF1_{FL}$. c) *P(r)* distribution of $TRF1_{FL}$ showing a maximum dimension of 130 Å. d) Guinier analysis determined a radius of gyration (*Rg*) value of 38.7 Å. Empty circles represent the complete dataset and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a dashed line. (*Q.Rg* values were < 1.3). e) SAXS *ab initio* DAMMIF model of $TRF1_{FL}$ presented as a molecular envelope with length of 130 Å corresponding to the determined *Dmax*.

analysis *Rg*). $TRF1_{FL}$ has a Guinier analysis determined *Rg* of 38.7 Å, which closely matches the real-space *Rg* (37.3 Å) (Figure 3.2.3.d.).

The real space SAXS data was utilised for *ab initio* modelling to create a single phase dummy atom model using DAMMIF software, explained in Methods 2.5.4. (Daniel et al.,, 2009). The low resolution molecular envelope of $TRF1_{FL}$ demonstrates an globular V-shaped molecule, with two distinct smaller domains which could possibly be the DNA binding Myb-domains (Figure 3.2.3.e.).

It is well known from the literature that TRF1 is a core shelterin complex protein, that binds directly to double stranded (ds) telomeric DNA. We analysed the ability of full length TRF1 to bind dsDNA using electrophoretic mobility shift assay (EMSA). 25 μM of random sequence dsDNA or telomeric repeat dsDNA was incubated with increasing amounts of $TRF1_{FL}$. Incubated samples were ran on an agarose gel and shifts in the bands were observed, which clearly indicated strong affinity for dsDNA (Figure 3.2.4.a. and b.). A DNA shift was observed at 6.4 and 3.2 μM for $TRF1_{FL}$ for random and telomeric DNA, respectively, indicating that $TRF1_{1\text{-}439}$ preferentially binds telomeric repeat DNA (TTAGGG)n. Previous studies have reported that the C-terminal Myb domain of TRF1 has an affinity of 0.2 μM to dsDNA (Hanaoka et al., 2005). Using a modified EMSA protocol we were able to calculate an affinity of DNA-binding proteins of TRF_{FL} to dsDNA (Work performed by Dr James Dunce, published in Dunce and Milburn et al., 2018). EMSAs were performed using 5'-FAM (fluorescein amidites)-labelled dsDNA to quantify the proportion of dsDNA that was unbound. An apparent K_D for TRF1 was determined at 0.1 ± 0.01 μM to dsDNA, which is in accordance to previously published data (Dunce, Milburn, et al., 2018; Hanaoka et al., 2005).

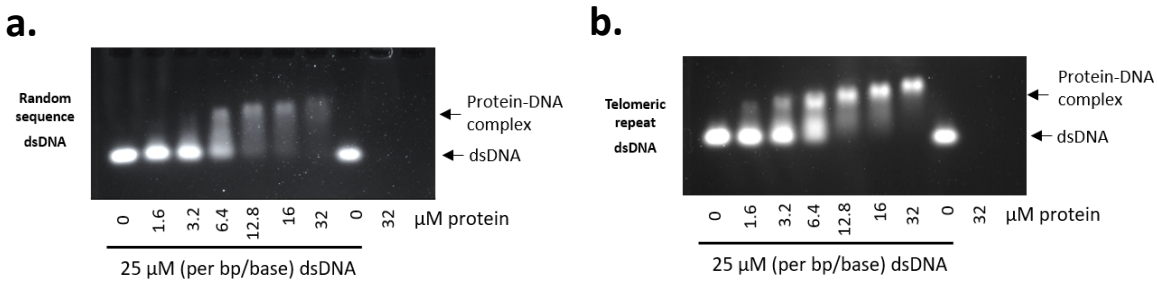


Figure 3.2.4.| DNA binding ability of $TRF1_{FL}$.
Electrophoretic mobility shift assays (EMSA) were used to demonstrate the DNA binding ability of $TRF1_{FL}$ a) for linear random sequence dsDNA and b) telomeric repeat dsDNA. TRF1 preferentially binds to telomeric repeat (TTAGGG)n DNA, as shown by forming a protein-DNA complex at a lower concentration of protein.

3.2.3. The TRFH domain of TRF1 homodimerises in solution

SEC-MALS analysis of $TRF1_{TRFH}$ was used to assess its ability to dimerise in solution. $TRF1_{TRFH}$ was purified by the same optimised procedure as full length TRF1, summarised in figure 3.2.5.a. The SEC-MALS analysis of both His_6-tagged and cleaved $TRF1_{TRFH}$ determined molecular weights of 56 and 49 kDa, respectively, corresponding to a TRF1 dimer (theoretical Mw dimer 55.5 and 48 kDa, respectively) (Figure 3.2.5.b. and 3.2.5.c.). These results determined that $TRF1_{TRFH}$ homodimerises in solution, in line with the crystal structure (PDB code – 1H6O, Yong Chen et al., 2008).

Far-UV circular dichroism (CD) spectroscopy was used to quantify the secondary structure of $TRF1_{TRFH}$. As expected, $TRF1_{TRFH}$ shows a characteristic alpha-helical spectrum with negative peaks of similar magnitude at 222 and 208 nm and deconvolution estimates an α-helical content of 64% (Figure 3.2.5.e.). The thermal stability of $TRF1_{TRFH}$ was determined using CD thermal denaturation by measuring the helical signal at 222 nm between 4 and 95 °C. Temperature is plotted against percentage unfolded and the melting temperature (T_m) is taken at 50 % unfolded. $TRF1_{TRFH}$ has a T_m of 50 °C, suggesting $TRF1_{TRFH}$ is correctly folded (Figure 3.2.5.f.).

The DNA binding ability of $TRF1_{TRFH}$ was tested by EMSA, as expected from the literature, the homodimerisation domain of TRF1 is unable to bind dsDNA, shown by a flat line of unbound (Figure 3.2.5.g). This confirms that the Myb-domain of TRF1 is required for the binding of telomeric dsDNA.

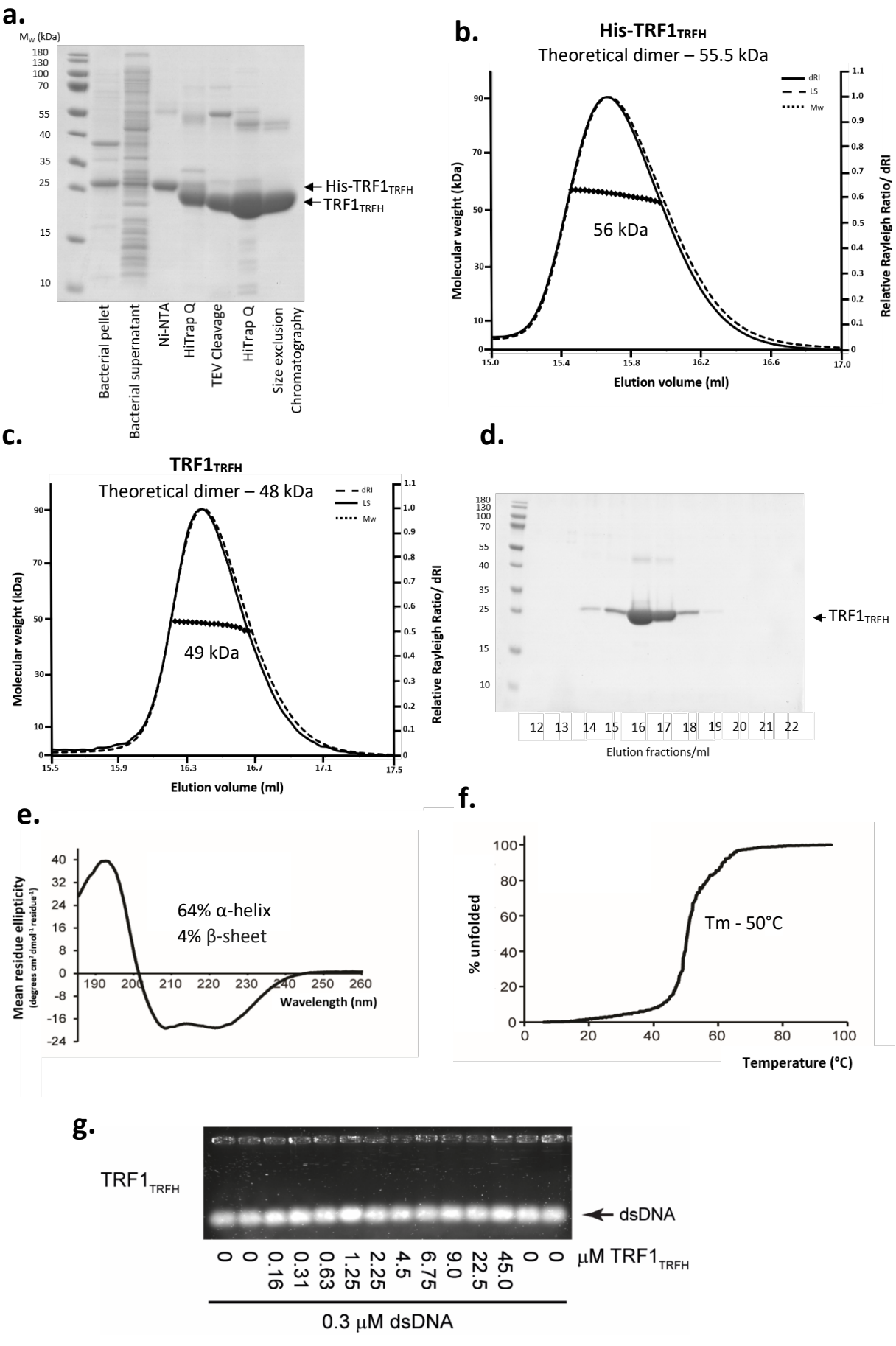

a.
Mw (kDa)
180
130
100
70
55
40
35
25
15
10
His-TRF1TRFH
TRF1TRFH
Bacterial pellet
Bacterial supernatant
Ni-NTA
HiTrap Q
TEV Cleavage
HiTrap Q
Size exclusion Chromatography
b.
His-TRF1TRFH
Theoretical dimer – 55.5 kDa
dRI
LS
Mw
56 kDa
Molecular weight (kDa)
Relative Rayleigh Ratio/ dRI
Elution volume (ml)
c.
TRF1TRFH
Theoretical dimer – 48 kDa
dRI
LS
Mw
49 kDa
Molecular weight (kDa)
Relative Rayleigh Ratio/ dRI
Elution volume (ml)
d.
TRF1TRFH
12 13 14 15 16 17 18 19 20 21 22
Elution fractions/ml
e.
Mean residue ellipticity (degrees cm2 dmol-1 residue-1)
64% α-helix
4% β-sheet
Wavelength (nm)
f.
% unfolded
Tm - 50°C
Temperature (°C)
g.
TRF1TRFH
dsDNA
0 0 0.16 0.31 0.63 1.25 2.25 4.5 6.75 9.0 22.5 45.0 0 0
μM TRF1TRFH
0.3 μM dsDNA

Figure 3.2.5.| Purification and biophysical analysis of $TRF1_{TRFH}$
a) SDS-PAGE showing the purification summary of $TRF1_{TRFH}$ through Ni-NTA and anion exchange chromatography. N-terminal His-tag was removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. b-d) SEC-MALS analysis of b) His-tagged and d) cleaved $TRF1_{TRFH}$ both showing a single peak with experimental molecular weights of 56 kDa (theoretical dimer – 55.5 kDa) and 49 kDa (theoretical dimer – 48 kDa) confirming that the homodimerisation domain of TRF1 dimerises in solution. c) SDS-PAGE analysis of the SEC-MALS output of 1ml fractions. e) Far UV CD wavelength scan between 260-185 nm of $TRF1_{TRFH}$ shows a typical α-helical trace. Deconvolution of the data estimates the secondary structure to be 64 % α-helical. f) CD thermal denaturation measured at 222 nm between 4 and 95°C, estimated a melting temperature of 50 °C. g) Electrophoretic mobility shift assays (EMSA) demonstrated that the homodimerisation domain of TRF1 does not have DNA binding ability.

SEC-SAXS analysis of $TRF1_{TRFH}$ revealed a distance distribution profile typical of a globular protein, with a *Dmax* of 97 Å. As expected, this is shorter than the full-length protein as the unstructured N-terminus and C-terminal Myb-domain have been removed (Figure 3.2.6.a.). $TRF1_{TRFH}$ has a Guinier analysis determined *Rg* of 29.9 Å, which closely matches the real-space *Rg* (30.76 Å) (Figure 3.2.6.b.). The SEC-SAXS scattering data was ultilised to create a low-resolution *ab intio* model of $TRF1_{TRFH}$. The resultant modeling demonstrates a V-shaped molecular envelope for $TRF1_{TRFH}$ that closely matches its known crystal structure (PDB: 1H6O, Fairall et al., 2001), shown by using SUPCOMB to dock the crystal structure into the *ab initio* envelope (Figure 3.2.6.c. and d.) (Kozin & Svergun, 2001).

3.2.4. $TERB1_{TRFB}$ is a soluble aggregate when expressed in isolation

We initially set out to characterise the meiotic telomere complex proteins individually. The full length TERB1 protein was cloned ultilising two IMAGE clones, containing residues 1-424 and 486-727, respectively, resulting in a 63 amino acid deletion (Δ425-487), called $TERB1_{FL\Delta}$. Through sequence analysis and secondary structure prediction two constructs were designed corresponding to the N-terminal armadillo repeat domain ($TERB1_{ARM}$, amino acid residues 1-424) and the central TRF1 binding domain ($TERB1_{TRFB}$, amino acid residues 561-658).

$TERB1_{FL\Delta}$, $TERB1_{ARM}$ and $TERB1_{TRFB}$ were expressed as His_6-MBP-fusions encoded by a pMAT11 vector in *E.coli* BL21 (DE3) cells. Both $TERB1_{FL\Delta}$ and $TERB1_{ARM}$ showed weak expression, with most

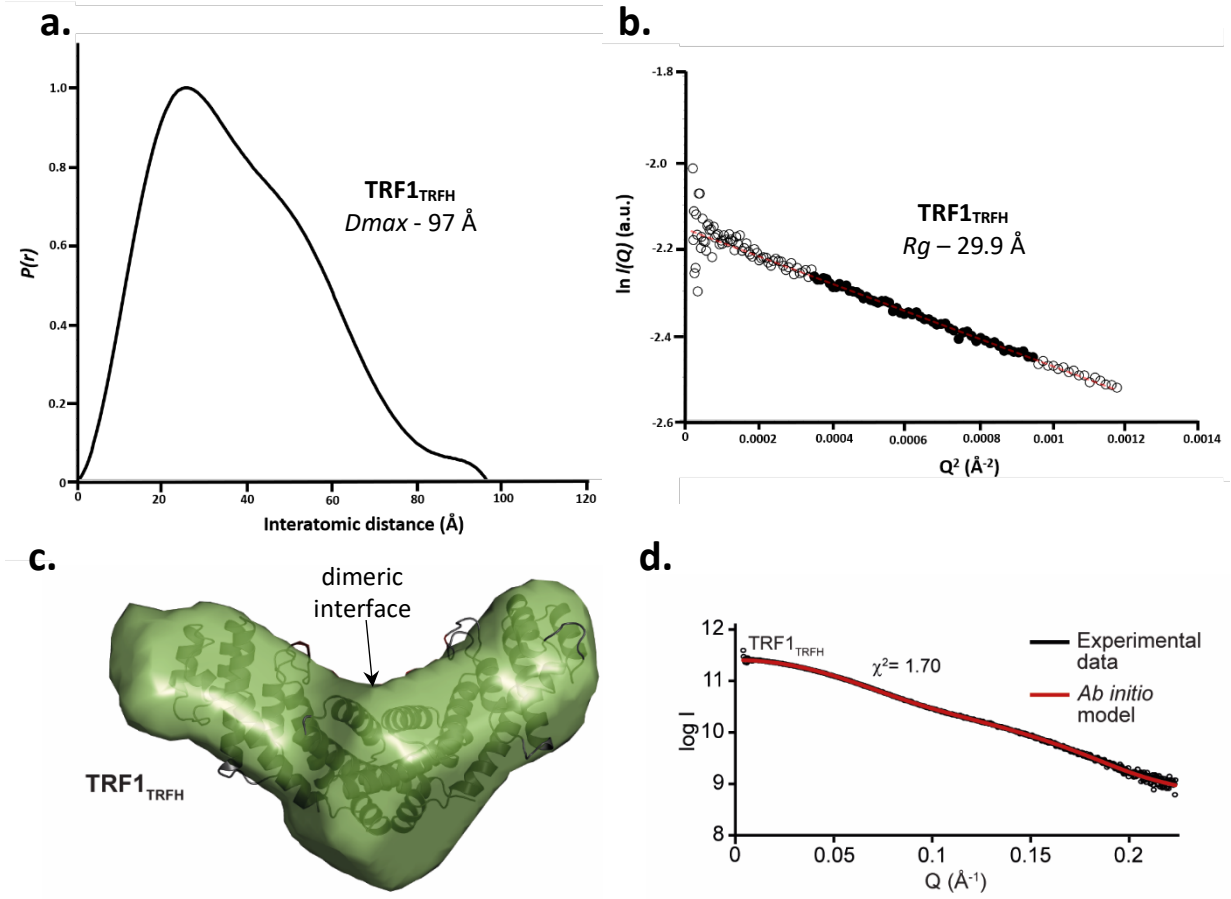


Figure 3.2.6.| SEC-SAXS analysis of TRF1$_{TRFH}$.
a-b) SAXS analysis of TRF1$_{TRFH}$. a) *P(r)* distribution of TRF1$_{62-268}$ showing a maximum dimension of 97 Å. b) Guinier analysis determined a radius of gyration (*Rg*) value of 29.9 Å. The real space *Rg* closely matches the Guinier analysis *Rg* value of 30.76 and 29.9 Å, respectively. Empty circles represent the complete dataset and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a dashed line. (*Q.Rg* values were < 1.3). c) SAXS ab initio DAMMIF model of TRF1$_{TRFH}$ presented as the molecular envelope. The crystal structure of TRF1$_{TRFH}$ (PDB code: 1H6O, Fairall et al., 2001) can be docked into the SAXS envelope. d) SAXS scattering of TRF1$_{TRFH}$ (black) overlaid with the theoretical scattering curve of the modelled structure. SAXS *ab initio* models and figures made by Dr Owen Davies, published in Dunce, Milburn et al., 2018.

of the material being insoluble, as evidenced by the presence of the protein in the lysed cell pellet. The small amount of soluble material was highly degraded upon purification by amylose affinity chromatography isolation (Figure 3.2.7.a. and b). This could suggest that the expression conditions such as temperature and lysis buffer conditions need to be optimised to increase protein solubility and reduce degradation. Post-induction temperature of the bacterial cultures was reduced from 25 to 15°C to

promote protein folding. Furthermore, glycerol and L-arginine were added to the lysis buffer. Despite these attempts, TERB1 still remained insoluble (data not shown). Expression of TERB1$_{TRFB}$ yielded soluble material. However, around 50 % of the material was degraded. The amylose eluate was further purified by anion exchange chromatography to remove unbound MBP (Figure 3.2.7.c.). MBP-TERB1$_{TRFB}$ was analysed by SEC-MALS, which determined that TERB1 forms large megadalton species suggesting aggregation (Figure 3.2.7.d.). From this result we hypothesise that TERB1 requires a binding partner to induce stability.

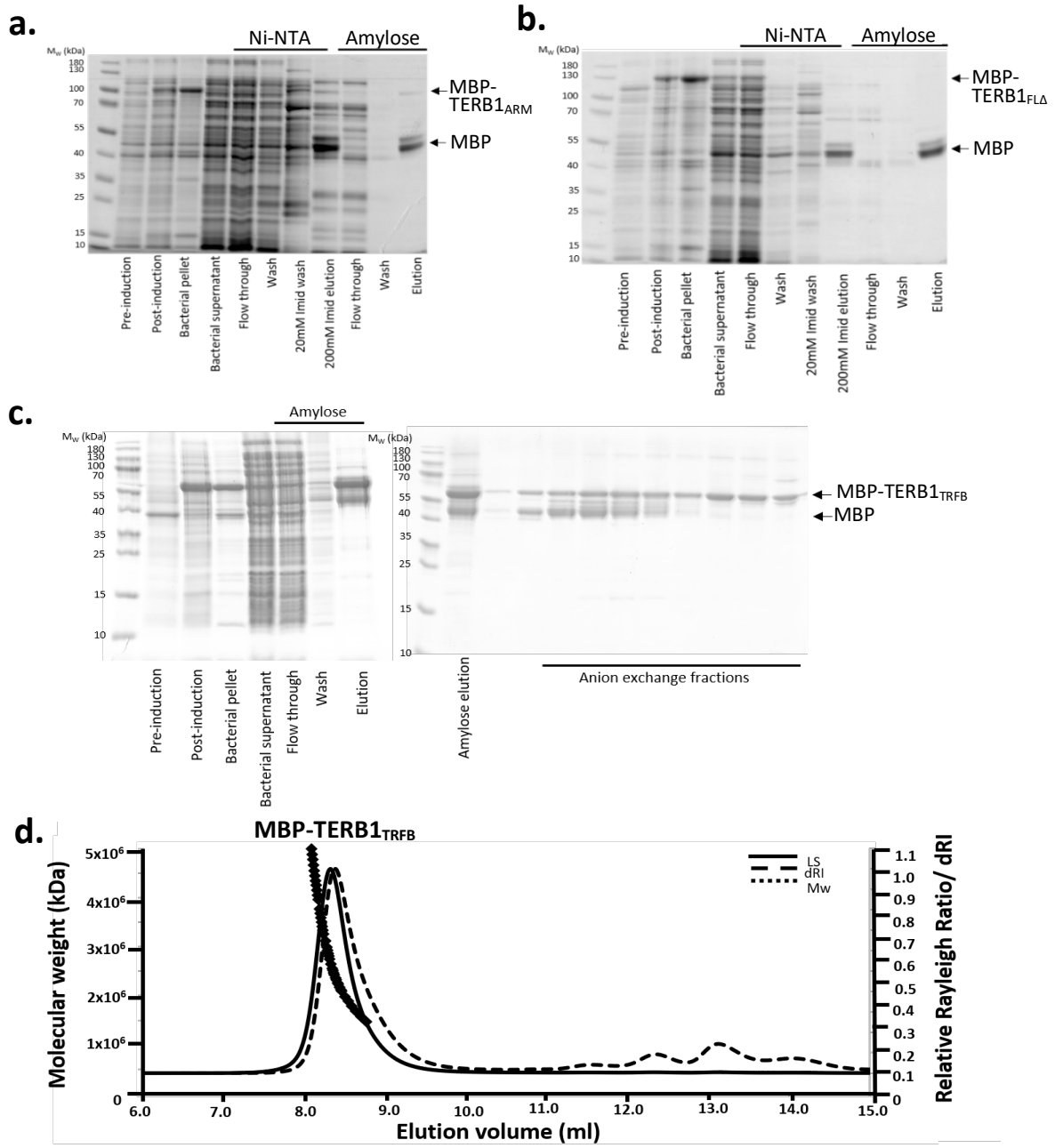


Figure 3.2.7.| TERB1 is a soluble aggregate.
a-c) SDS-PAGE analysis shows the purification steps of three TERB1 constructs with N-terminal MBP solubility tags. a) MBP-TERB1$_{ARM}$ and b) MBP-TERB1$_{\Delta FL}$ both expressed poorly, and most of the protein is insoluble as seen in the bacterial pellet. The limited amount of soluble material is highly degraded down to free-MBP suggesting the protein is highly unstable. c) MBP-TERB1$_{TRFB}$ soluble material is purified by amylose affinity and anion exchange chromatography. There is a significant amount of free-MBP present, showing that the protein is unstable and degrading. d) SEC-MALS analysis of MBP-TERB1$_{TRFB}$ in isolation is a soluble aggregate forming megadalton species.

Published studies and former Masters student, Mr Lee Thung Sen, have ultilised the Y2H system to successfully identify interactions between the meiotic proteins (Shibuya & Watanabe, 2014). We wanted to validate these findings by replicating a smaller screen focusing on the MTC protein, TERB1.

A Y2H screen was performed with the three constructs of TERB1, the LINC complex proteins, SUN1 and KASH5, and the shelterin complex protein, TRF1. The screen highlighted the previously identified TERB1-TRF1 interaction (Figure 3.2.8.). Previous Y2H experiments have shown that the N-terminal ARM-domain of TERB1 interacts with the N-terminus of SUN1. We observed an interaction between pGBKT7 TERB1$_{ARM}$ (TERB1$_{1-424}$) and pGADT7 SUN1$_{1-276}$. However, this was not present when the yeast vectors were switched. Furthermore, we observed that pGBKT7 TERB1$_{ARM}$ interacts with the negative control, suggesting that this vector autoactivates, thus any observations need to be excluded.

3.2.5. TRF1$_{TRFH}$-TERB1$_{TRFB}$ forms a 2:1 complex in solution

To investigate the interaction between TRF1 and TERB1 we co-expressed the homodimerisation domain of TRF1 with its TERB1 binding domain that was deduced by Y2H (Shibuya & Watanabe, 2014). TRF1$_{TRFH}$ and TERB1$_{TRFB}$ were fused to a His$_6$-tag and His$_6$-MBP-tag, respectively, to aid solubility of TERB1 and co-expressed in *E. coli* and grown at 15°C upon induction with IPTG to maximise the expression and solubility of TERB1. The cell pellets were lysed in standard buffer plus 10% glycerol to prevent precipitation of TRF1. The complex was co-purified through sequential Ni-NTA and amylose affinity chromatography followed by anion exchange chromatography. Affinity-tags were removed by incubating with TEV-protease overnight, followed by anion-exchange chromatography and size-exclusion chromatography, summarised in Figure 3.2.9.a. Upon the

concentration of $TRF1_{TRFH}$- $TERB1_{TRFB,}$ the solution acquired a brown-pinkish tinge, indicating a possible metal-binding ability.

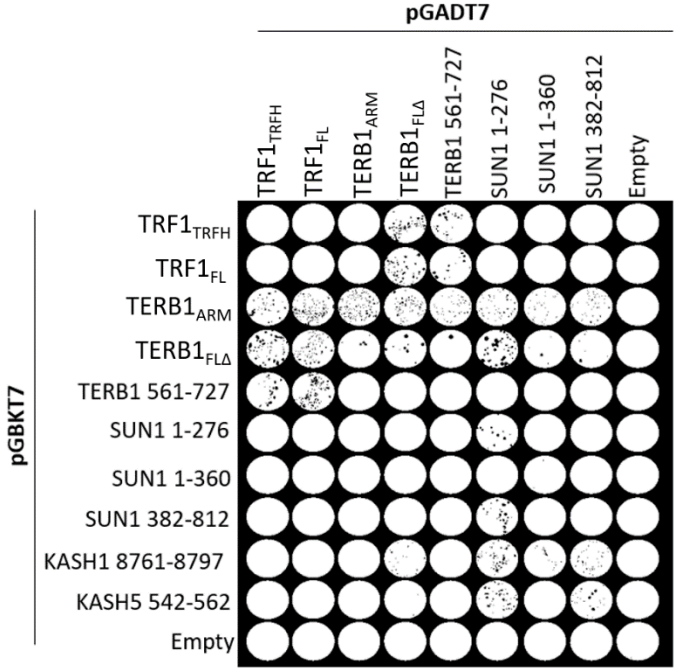


Figure 3.2.8.| Screening for TERB1 interactions.
a) Yeast two hybrid grid of the shelterin protein, TRF1, LINC complex proteins SUN1 and KASH5 and TERB1 in both the bait (pGBKT7) and target (pGADT7) yeast plasmids. Yeast strains Y2H gold (pGBKT7 - bait) and Y187 (pGADT7 - target) were mated and subsequently plated on quadruple dropout (QDO) plates for colony selection, subject to reporter gene activation upon positive interactions. Full length, N- and C-terminal constructs of TERB1; $TERB1_{FL\Delta}$, $TERB1_{ARM}$ and $TERB1_{561\text{-}727}$, respectively, were screened against N- and C-terminal constructs of SUN1, $TRF1_{FL}$ and $TRF1_{TRFH}$. The interaction between SUN1 and KASH5 has already been well characterised so these two proteins were used as a positive control. The screen confirmed an interaction between $TRF1_{TRFH}$ and $TERB1_{561\text{-}727}$, which contains the TRFB domain. From the screen it is not possible to determine an interaction between $TERB1_{ARM}$ and SUN1, as the pGBKT7 $TERB1_{ARM}$ plasmid autoactivates as there are colonies in all pGADT7 interacting plasmids.
* $TERB1_{FL\Delta}$ - contains a 63 amino acid deletion (Δ 425-487)

Fusion and cleaved material were analysed by SEC-MALS to determine absolute molecular weight of the complex and oligomeric state. Both MBP-fusion and cleaved proteins produced a single peak with a molecular mass of 109 and 56.8 kDa, respectively, corresponding to a 2:1 TRF1-TERB1 complex (theoretical 2:1 – 112 and 59 kDa) (Figure 3.2.9.b. and 3.2.9.c.). SDS-PAGE analysis of the cleaved $TRF1_{TRFH}$-$TERB1_{TRFB}$ SEC output fractions further demonstrates that the two proteins co-elute as a 2:1

complex (Figure 3.2.8.d.). Recently published studies proposed that TRF1 and TERB1 exist as an equimolar 2:2 complex, shown by a crystal structure of $TRF1_{TRFH}$ in complex with a 15 amino acid TERB1 peptide, residues 642-656 (TRF1 binding motif, $TERB1_{TBM}$). The structure determined that two $TERB1_{TBM}$ peptides bind to the $TRF1_{TRFH}$ homodimer. However, our detailed in solution studies by SEC-MALS and SEC-SAXS did not observe this 2:2 complex, instead our results show that a single $TERB1_{TRFB}$ molecule binds to a $TRF1_{TRFH}$ homodimer indicating a 2:1 stoichiometry. This suggests that $TERB1_{TBM}$ peptide is too short to show the true binding event and the longer $TERB1_{TRFB}$ construct likely blocks the second site through steric hindrance to provide a 2:1 complex.CD spectroscopy was used to study the secondary structure composition of $TRF1_{TRFH}$-$TERB1_{TRFB}$, which revealed an α-helical profile. Analysis by deconvolution estimated the alpha helical and beta sheet content to be 34 and 11%, respectively (Figure 3.2.8.e.). The high percentage α-helical content could be attributed to the $TRF1_{TRFH}$ homodimer. The CD thermal melt shows a steady progression of protein unfolding at 57°C, likely corresponding to co-operative unfolding of $TRF1_{TRFH}$-$TERB1_{TRFB}$ (Figure 3.2.8.f.). We have already shown by EMSA that $TRF1_{TRFH}$ does not bind dsDNA (Figure 3.2.4.g.), therefore we tested the DNA binding ability of $TERB1_{TRFB}$ by analysing the $TRF1_{TRFH}$-$TERB1_{TRFB}$ complex. EMSA indicated that $TERB1_{TRFB}$ does bind dsDNA, shown by the presence of a single band corresponding to unbound DNA (Figure 3.2.9.g. and h.).

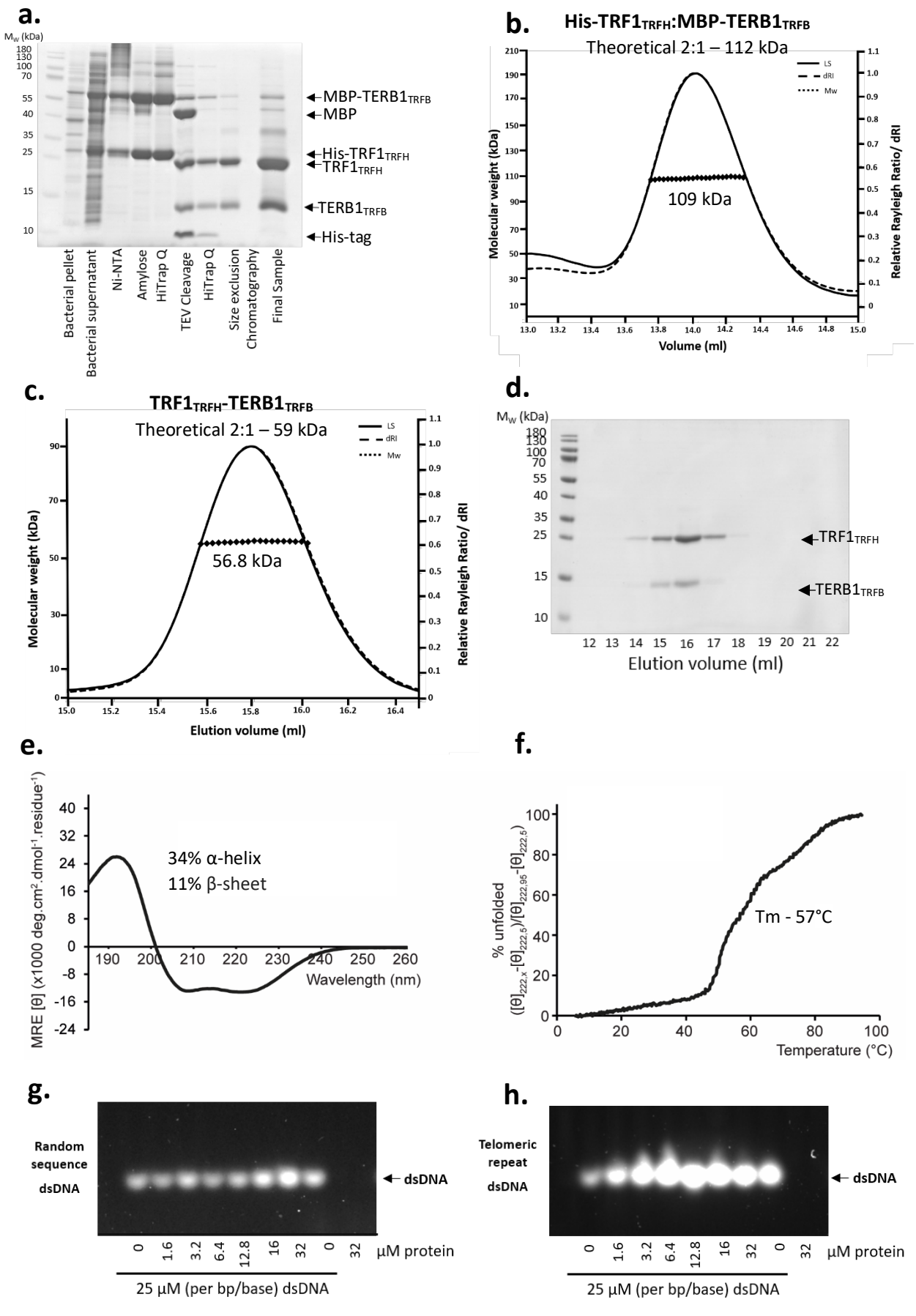

a.
M_W (kDa)
180
130
100
70
55
40
35
25
15
10
MBP-TERB1TRFB
MBP
His-TRF1TRFH
TRF1TRFH
TERB1TRFB
His-tag
Bacterial pellet
Bacterial supernatant
Ni-NTA
Amylose
HiTrap Q
TEV Cleavage
HiTrap Q
Size exclusion Chromatography
Final Sample
b.
His-TRF1TRFH:MBP-TERB1TRFB
Theoretical 2:1 – 112 kDa
LS
dRI
Mw
109 kDa
Molecular weight (kDa)
Relative Rayleigh Ratio/ dRI
Volume (ml)
c.
TRF1TRFH-TERB1TRFB
Theoretical 2:1 – 59 kDa
LS
dRI
Mw
56.8 kDa
Molecular weight (kDa)
Relative Rayleigh Ratio/ dRI
Elution volume (ml)
d.
M_W (kDa)
180
130
100
70
55
40
35
25
15
10
TRF1TRFH
TERB1TRFB
12 13 14 15 16 17 18 19 20 21 22
Elution volume (ml)
e.
MRE [θ] (x1000 deg.cm².dmol⁻¹.residue⁻¹)
34% α-helix
11% β-sheet
190 200 210 220 230 240 250 260
Wavelength (nm)
f.
% unfolded
$([\theta]_{222,x}-[\theta]_{222,5})/([\theta]_{222,95}-[\theta]_{222,5})$
Tm - 57°C
Temperature (°C)
g.
Random sequence dsDNA
dsDNA
0 1.6 3.2 6.4 12.8 16 32 0 32 µM protein
25 µM (per bp/base) dsDNA
h.
Telomeric repeat dsDNA
dsDNA
0 1.6 3.2 6.4 12.8 16 32 0 32 µM protein
25 µM (per bp/base) dsDNA

Figure 3.2.9.| Purification and biophysical analysis of $TRF1_{TRFH}$-$TERB1_{TRFB}$
a) SDS-PAGE showing the purification summary of $TRF1_{TRFH}$-$TERB1_{TRFB}$ through Ni-NTA, amylose and anion exchange chromatography. N-terminal solubility tags were removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. b-d) SEC-MALS analysis of b) fusion and d) cleaved $TRF1_{TRFH}$-$TERB1_{TRFB,}$ both elution profiles have a single peak with experimental molecular weights of 109 and 56.8 kDa, respectively. A theoretical 2:1 complex has an estimated molecular weight of 112 and 59 kDa, respectively indicating a single TERB1 molecule binds a TRF1 homodimer. c) SDS-PAGE analysis of the SEC-MALS output of 1ml fractions, shows co-elution of the TRF1-TERB1 complex. e) Far UV CD wavelength scan between 260-185 nm of $TRF1_{TRFH}$-$TERB1_{TRFB.}$ Deconvolution of the data estimates the secondary structure to consist of 34% α-helices and 11% β-sheets. f) CD thermal denaturation measured at 222nm between 4 and 95°C, estimated a melting temperature of 57°C. g-h) EMSA demonstrated that $TRF1_{TRFH}$-$TERB1_{TRFB}$ does not have DNA binding ability.

The solution structure of $TRF1_{TRFH}$-$TERB1_{TRFB}$ was determined by SEC-SAXS experiments. The SEC-SAXS scattering data allowed for the calculation of the *Dmax* and *Rg*, which were determined as 110 and 30.9 Å, respectively (Figures 3.2.10.a.-c.). These results are analogous to the SEC-SAXS of $TRF1_{TRFH}$ when analysed alone (*Dmax*: 97 Å, *Rg*: 29.9 Å), suggesting that $TERB1_{TRFB}$ may dock into the V-shaped $TRF1_{TRFH}$ interface, thus not changing the shape and size. This hypothesis could be confirmed by the X-ray crystal structure of $TRF1_{TRFH}$-$TERB1_{TRFB.}$ The normalised Kratky plot has a bell-shaped (Gaussian) peak typical of a compact globular complex with some flexibility (Figure 3.2.10.d.).

High through-put crystal screening, using commercially available screens, of $TRF1_{TRFH}$-$TERB1_{TRFB}$ were carried out but this did not result in any crystal 'hits'. This may suggest that there are flexible regions with the TERB1 construct that are potentially preventing crystal growth. We know that the TRFB domain of TERB1 contains the binding site for both TRF1 and TERB2 therefore it would be beneficial to identify the two binding sites and optimise the TERB1 construct.

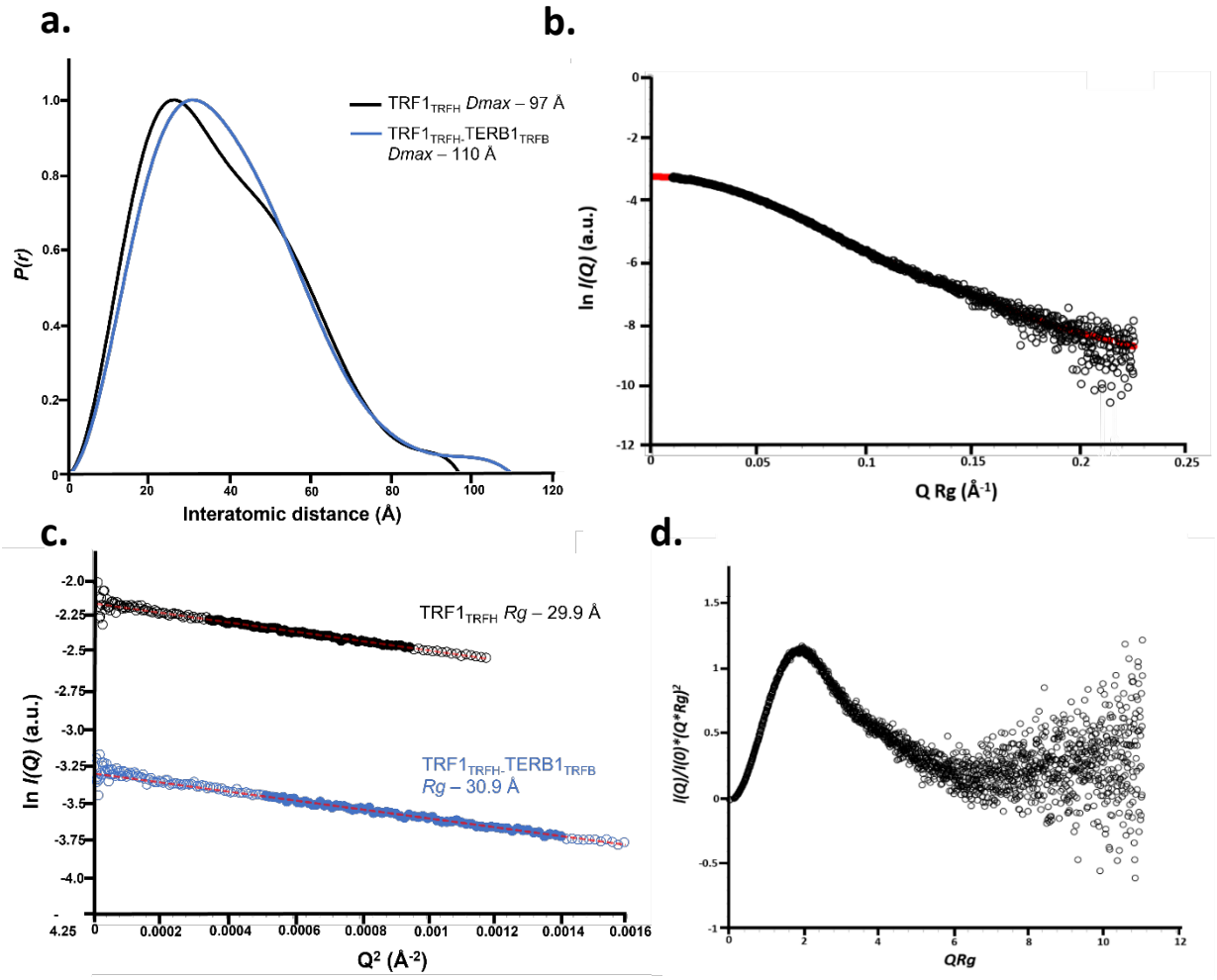


Figure 3.2.10.| SEC-SAXS data analysis of TRF1$_{TRFH}$-TERB1$_{TRFB}$
a-d) SAXS analysis of TRF1$_{TRFH}$-TERB1$_{TRFB}$. a) Overlaid *P(r)* distribution of TRF1$_{TRFH}$ (black) and TRF1$_{TRFH}$-TERB1$_{TRFB}$ showing a maximum dimension of 97 and 110 Å, respectively. b) Averaged experimental SAXS profile of TRF1$_{TRFH}$-TERB1$_{TRFB}$ with the fit used for the *P(r)* distribution shown in red. c) Overlaid guinier analysis of TRF1$_{TRFH}$ (black) and TRF1$_{TRFH}$-TERB1$_{TRFB}$ determined a *Rg* value of 29.9 and 30.9 Å, respectively. Empty circles represent the complete data set and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a dashed line. (*Q.Rg* values were < 1.3). d) Normalised Kratky plot for TRF1$_{TRFH}$-TERB1$_{TRFB}$.

3.2.6. TRF1$_{TRFH}$-TERB1$_{TBM}$ complex dissociates in solution

To probe whether the TRF1 interaction occurs within a specific region of the TERB1$_{TRFB}$ domain, several truncated versions of TERB1 were designed based on secondary structure predictions, sequence conservation and information from the known crystal structure. Based on these analyses several N-terminal truncations were made; TERB1$_{574-656}$, TERB1$_{585-656}$, TERB1$_{585-658}$, TERB1$_{600-656}$, TERB1$_{600-658}$,

TERB1$_{619-656}$ and TERB1$_{642-658}$. To assess how these truncation mutants affected the interaction between TRF1$_{TRFH}$ and TERB1, the proteins were co-expressed together as His$_6$- and MBP-tagged versions, respectively. The bacterial lysates from these co-expressions were applied to amylose resin, incubated and after washing the bound protein was eluted with D-maltose and the elutate was analysed by SDS-PAGE (Figures 3.2.11.a. and 3.2.11.b.). Initial analysis demonstrated that TERB1$_{642-658}$, previously denoted the TRF1 binding motif (TERB$_{TBM}$), was able to interact with TRF1, shown by the presence of His-TRF1$_{TRFH}$ in the amylose eluate. Co-purified His-TRF1$_{TRFH}$:MBP-TERB1 constructs were further purified by anion exchange chromatography, which purifies proteins on the basis of charge, using a salt gradient to elute the positively charged proteins from the anionic resin. Analysis of the elution fractions from the anionic exchange purification clearly shows that as MBP-TERB1 is truncated the interaction between TRF1$_{TRFH}$ and TERB1 weakens**,** as evidenced by a shift in the MBP-TERB1 elution volume to earlier fractions and the emergence of distinct peaks (Figure 3.2.11.c-f.). This suggests that the binding observed by amylose affinity chromatography for TRF1$_{TRFH}$-TERB1$_{TBM}$ is weak and a larger construct is required for stability.

The anion exchange chromatography elution fractions of His$_6$-TRF1$_{TRFH}$:MBP-TERB1$_{TBM}$ were pooled, concentrated to 50mg/ml, then analysed by SEC-MALS. The large MBP (44 kDa) solubility tag was fused to TERB1$_{TBM}$ used to study the TRF1$_{TRFH}$-TERB1$_{TBM}$ complex to provide an accurate stoichiometry, as TERB1$_{TBM}$ is very small (2.1 kDa) therefore the addition of the MBP tag will remove the ambiguity of the stoichiometry.

The observed SEC profile has two clear peaks with molecular weights of 99 and 46 kDa, respectively (Figure 3.2.12.a.). The calculated theoretical molecular masses of His$_6$-TRF1$_{TRFH}$- MBP-TERB1$_{TBM}$ as a 2:2 and 2:1 complex are 150 and 103 kDa, respectively, clearly indicating that the main peak of His$_6$-TRF1$_{TRFH}$- MBP-TERB1$_{TBM}$ forms a 2:1 complex (Figure 3.2.12.b.). The second, smaller peak has a molecular mass corresponding to a His$_6$-TRF1$_{TRFH}$ dimer (55 kDa) or an MBP-TERB1$_{TBM}$ monomer (47 kDa). SDS-PAGE analysis of the SEC-MALS elution fractions determined that the latter peak corresponds to free TERB1$_{TBM}$ (Figure 3.2.12.d.).

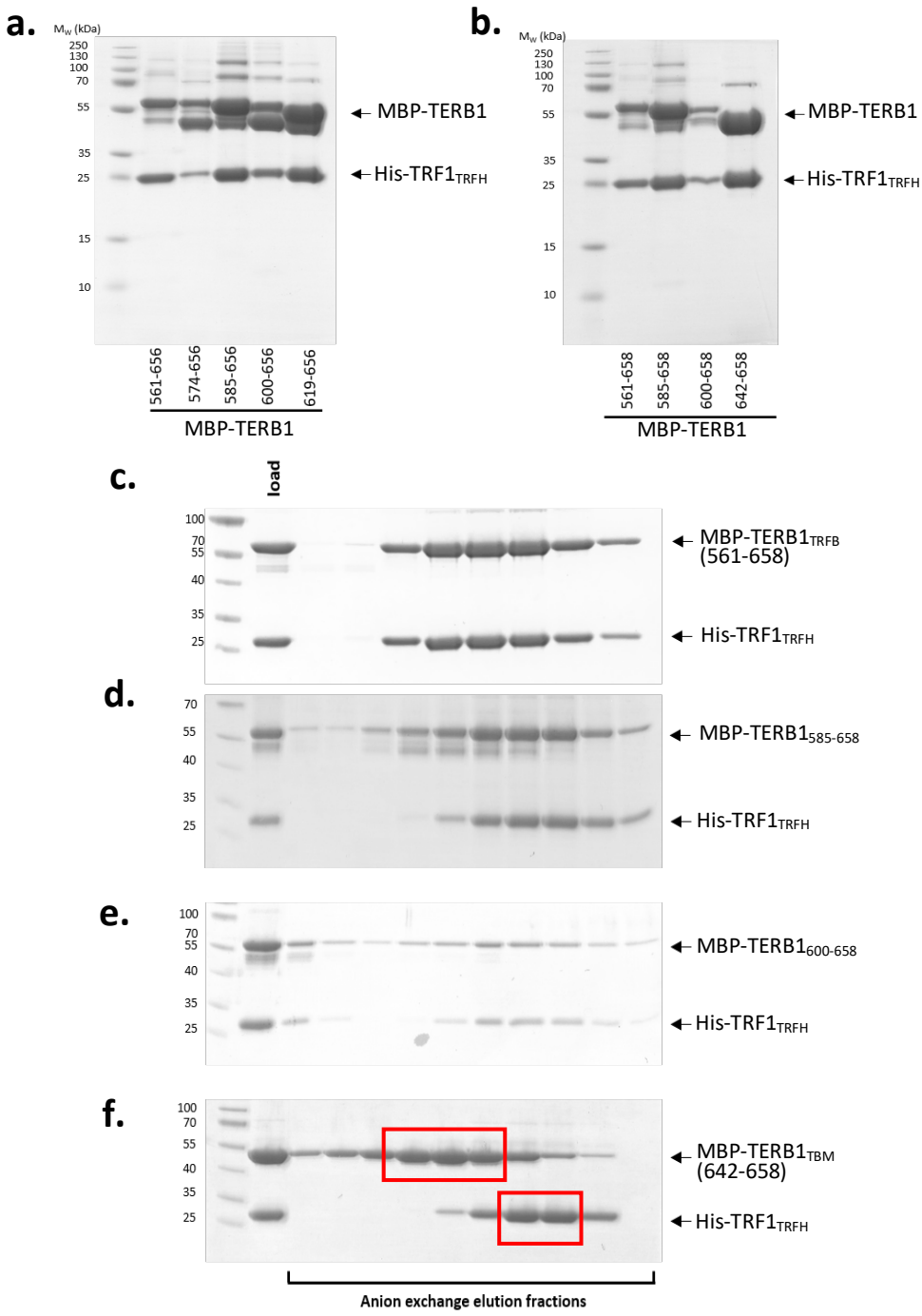


Figure 3.2.11.| Screening for the minimum TERB1 binding site.
a-b) SDS-PAGE of His-TRF1$_{TRFH}$-MBP-TERB1 amylose eluates. N-terminal truncated MBP-TERB1 constructs were co-expressed with His-TRF1 and sequentially purified by Ni-NTA and amylose resin, followed by a wash step and elution of the bound protein. This determined that His-TRF1$_{TRFH}$ can interact with all the truncated TERB1 constructs, with TERB1$_{TBM}$ being the minimal binding site. Concentrations were not normalised. c-f) SDS-PAGE showing further purification of His-TRF1$_{TRFH}$-MBP-TERB1 constructs by anion exchange chromatography. The amylose eluates from b) were loaded onto a Hi-trap-Q HP anion exchange column, eluted across a salt concentration and 1ml fractions were loaded onto a SDS-PAGE gel. c) MBP-TERB1$_{TRFB}$ coelutes with His-TRF1$_{TRFH}$. f) MBP-TERB1$_{TBM}$ no longer co-elutes with His-TRF1$_{TRFH}$, shown by His-TRF1$_{TRFH}$ eluting in later fractions.

Overlaying the SEC-MALS profiles of His_6-$TRF1_{TRFH}$-MBP-$TERB1_{TRFB}$ and His_6-$TRF1_{TRFH}$-MBP-$TERB1_{TBM}$ clearly indicates that the $TERB1_{TBM}$ peptide partially dissociates from the $TRF1_{TRFH}$ dimer further suggest that for a stable interaction the full TRFB domain of TERB1 is required. In addition, MBP is known as a stable monomeric protein, hence is expected not to interfere with the characterisation of TRF1-TERB1 binding (Momin et al., 2019).

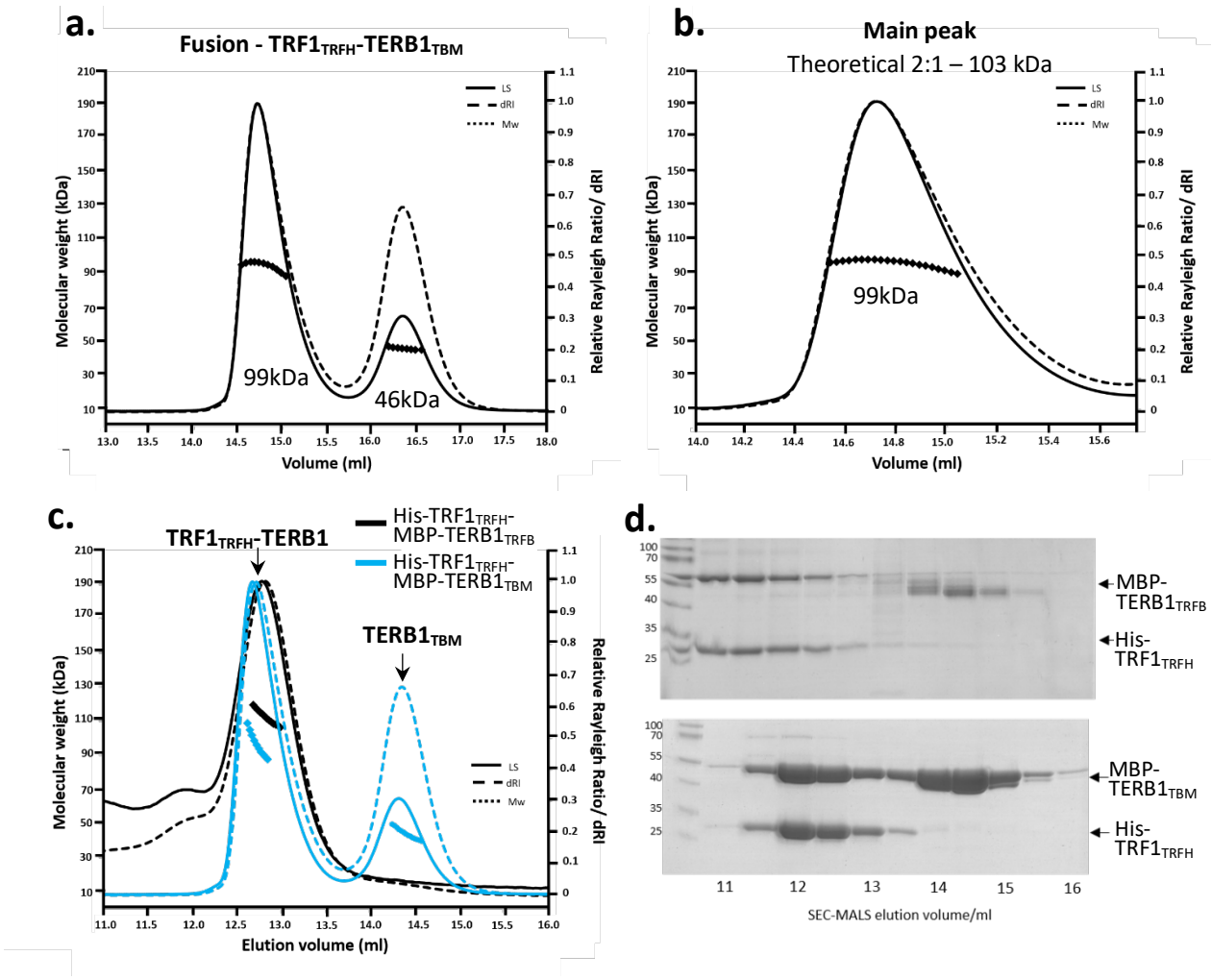


Figure 3.2.12.| SEC-MALS analysis of His-$TRF1_{TRFH}$-MBP-$TERB1_{TBM}$
a-b) SEC-MALS analysis of His-$TRF1_{TRFH}$-MBP-$TERB1_{TBM}$ complex. Anion exchange chromatography fractions (Figure 3.2.11.b.) were pooled and concentrated and analysed by SEC-MALS. a) Eluted as two distinct peaks, the first (main) peak has a Mw of 99 kDa and a second peak of 46 kDa, corresponding to a 2:1 complex (theoretical Mw 103 kDa) and free MBP-$TERB1_{TBM}$ (theoretical Mw 47.1 kDa), further confirmed by SDS-PAGE analysis (d). c) Overlaid SEC-MALS of His-$TRF1_{TRFH}$-MBP-$TERB1_{TBM}$ (blue) and of His-$TRF1_{TRFH}$-MBP-$TERB1_{TRFB}$ (black). The main peak of His-$TRF1_{TRFH}$-MBP-$TERB1_{TBM}$ perfectly overlays with the longer construct.

The 2:2 crystal structure of $TRF1_{TRFH}$-$TERB1_{TBM}$ was obtained by soaking abundant $TERB1_{TBM}$ peptide into $TRF1_{TRFH}$ crystals suggesting that this complex may not reflect the biological interaction. However, we cannot rule out the possibility that MBP has provided steric hindrance that mimics that of the wider $TERB1_{TRFB}$ sequence, hence blocking the 2:2 interaction.

3.2.7. MBP-$TERB1_{TBM}$ is soluble when expressed in isolation of TRF1

$TERB1_{TBM}$ (amino acid residues 642-658) was expressed as an MBP-fusion in isolation, which yielded soluble material. MBP-$TERB1_{TBM}$ was purified by amylose affinity chromatography and subsequent anion exchange chromatography (Figure 3.2.13.a.). Analysis by SDS-PAGE indicates that there is a small amount of protein degradation suggesting some protein instability or proteolytic cleavage. We analysed the MBP-fusion protein by SEC-MALS, which determined an absolute molecular mass of 46.7 kDa indicating a monomeric species (theoretical Mw – 47.1 kDa) (Figure 3.2.13.b.). Overlaying the SEC-MALS of His_6-$TRF1_{TRFH}$:MBP-$TERB1_{TBM}$ and MBP-$TERB1_{TBM}$ clearly indicates that of His_6-$TRF1_{TRFH}$:MBP-$TERB1_{TBM}$ dissociates over an isocratic gradient and further confirms that $TERB1_{TBM}$ is unstable when complexed to $TRF1_{TRFH}$ (Figure 3.2.12.c.). We attempted to further purify MBP-$TERB1_{TBM}$ to yield cleaved material so that we could analyse the $TRF1_{TRFH}$-$TERB1_{TBM}$ interaction through mixing experiments and analysis by SEC-MALS. However, upon TEV-cleavage and anion exchange chromatography it was not possible to separate any cleaved material from fusion material suggesting that $TERB1_{TBM}$ peptide is unstable when expressed on its own.

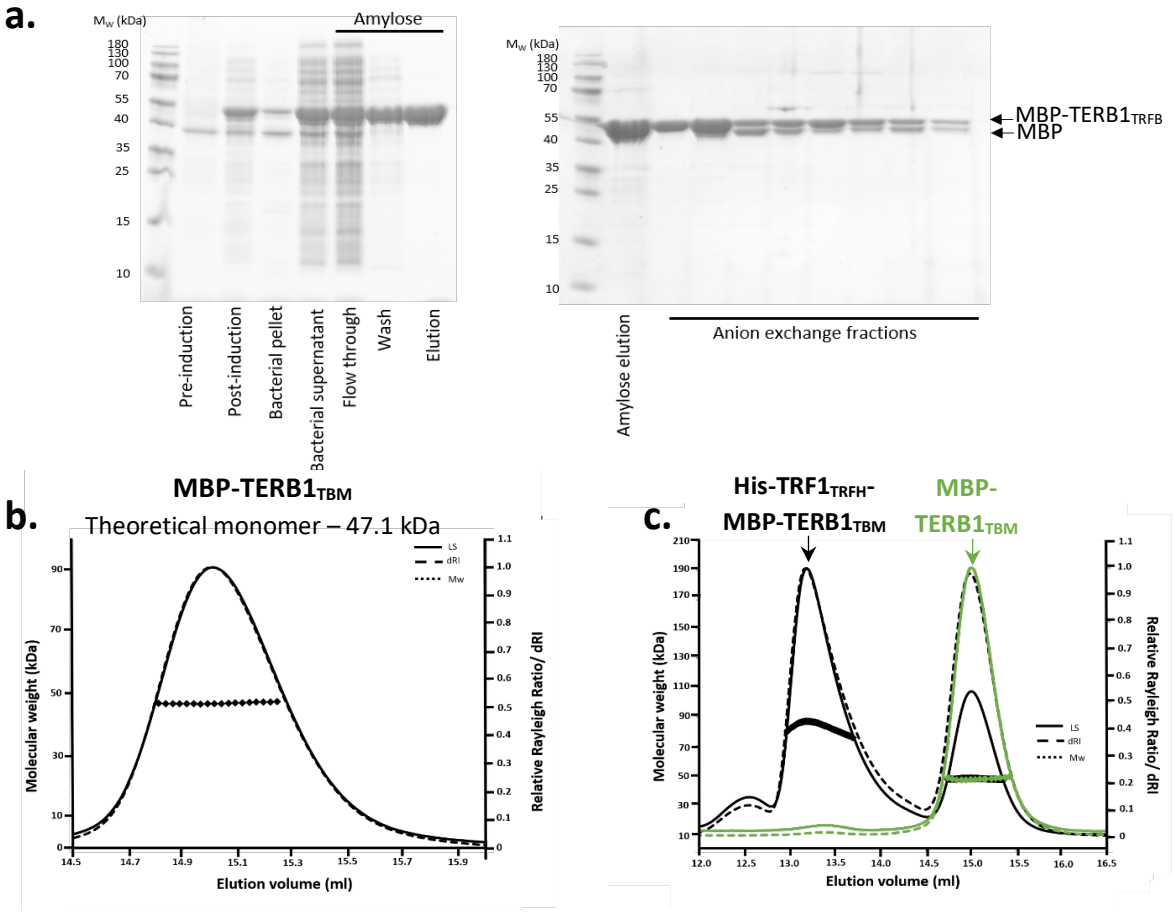


Figure 3.2.13.| MBP-TERB1$_{TBM}$ is soluble when expressed in isolation of TRF1.
a) SDS-PAGE gels showing the purification steps of TERB1$_{TBM}$ with a N-terminal MBP solubility tag. Soluble material is purified by amylose affinity and anion exchange chromatography. There is a small amount of free-MBP present indicating protein degradation suggesting protein instability. b) SEC-MALS analysis of MBP-TERB1$_{TBM}$ in isolation has an experimental Mw of 46.7 kDa, indicating a monomeric species (theoretical monomer – 47.1 kDa). c) Overlaid SEC-MALS of MBP-TERB1$_{TBM}$ and His-TRF1$_{TRFH}$-MBP-TERB1$_{TBM}$ shows clear dissociation of the TRF1-TERB1 complex when a shorter TERB1 construct is used.

3.2.8. TERB1 $_{TRFB}$ binds TERB2 to form a 1:1 complex

Y2H studies have previously shown that the TRF1 binding domain of TERB1 can bind both TRF1 and TERB2. Furthermore, when analysed by SEC-MALS it was observed that both TERB1 and TERB2 form large molecular weight aggregates when analysed individually (Figure 3.2.14.). These results suggest that TERB1 and TERB2 depend on each other for stability and that they may form a constitutive complex.

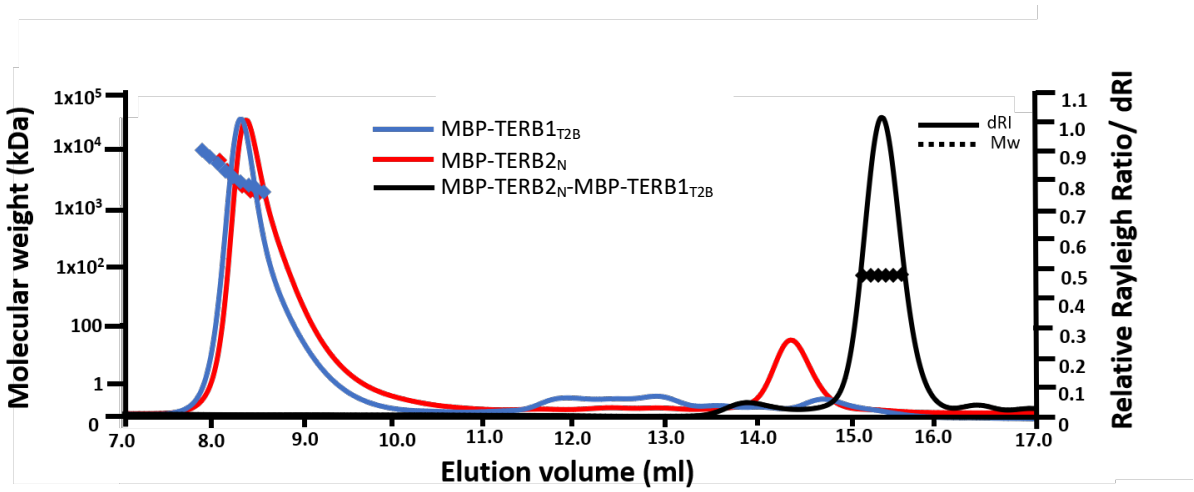


Figure 3.2.14.| TERB1 and TERB2 form soluble aggregates when expressed in isolation. SEC-MALS analysis of MBP-TERB1$_{T2B}$ (blue), MBP-TERB2$_{N}$ (red) in isolation deduced that both proteins are soluble aggregates forming megadalton species. However, when the two proteins are co-expressed and co-purified they form a soluble complex.

Previous work carried out by Mr Lee Thung Sen optimised the co-purification of TERB2 with TERB1 and mapped the interaction of TERB1$_{TRFB}$ to the N-terminus of TERB2 (amino acid residues 1-119). It was observed that upon affinity-tag cleavage TERB2 remained fused to the MBP-affinity tag independent of the amount of TEV-protease enzyme used. This suggested that the TEV-cleavage recognition site is not accessible to the TEV-protease. To address this problem a new TERB2 construct was designed with a flexible threonine-glycine-serine (TGS) repeat sequence (3xTGS) added to the beginning of the protein sequence to increase the accessibility of TEV-protease, herein known as TERB2$_{1\text{-}119LL}$. The addition of the 3xTGS linker did improve the cleavage of MBP-TERB2.

MBP-TERB2$_{1\text{-}119LL}$ and His$_6$-TERB1$_{TRFB}$ were co-expressed in *E. coli* BL21 DE3 cells and grown overnight at 15 °C upon induction with IPTG to maximise the expression and solubility of both proteins. The complex was co-purified through sequential Ni-NTA and amylose affinity chromatography. Affinity tags were removed by incubating with TEV-protease and cleaved material was purified by anion exchange chromatography which was unsuccessful in removing fusion material. A further amylose affinity purification step was utilised to remove fusion material, followed by an anion exchange step and size exclusion chromatography, summarised in figure 3.2.15.a. SEC-MALS analysis was used

to determine the absolute molecular weight of both the fusion and cleaved protein complex. Both fusion and cleaved material showed perfect co-elution of TERB1 and TERB2 with a molecular mass of 81.5 and 26.2 kDa, respectively (Figures 3.2.15.b-d.). These results indicate that TERB1-TERB2 heterodimerise, forming a 1:1 complex (theoretical molecular weights 75 and 26.2 kDa).

3.2.9. Deciphering individual TRF1 and TERB2 binding sites within the TERB1$_{TRFB}$ domain

We have shown that TERB1 forms stable bipartite complexes with both TRF1 and TERB2 suggesting that the interactions are independent but within the TRFB domain. This indicates there are separate TERB2 and TRF1 interaction sites. To probe whether the interaction of TERB2 occurs within a specific region of TERB$_{TRFB}$, N- and C-terminal truncations were designed and cloned. MBP-TERB2$_{1\text{-}119LL}$ was co-expressed with the His$_6$-TERB1 constructs and co-purified by sequential Ni-NTA and amylose affinity chromatography and the resultant amylose eluates were analysed by SDS-PAGE (Figure 3.2.16.a). Complex formation was retained TERB1$_{585\text{-}642}$-TERB21-$_{119LL}$, which contains both an N- and C-terminal truncation. A further truncated TERB1$_{585\text{-}629}$ construct resulted in a greatly reduced yield suggesting that complex stability had been compromised. Dissection of the TERB1 construct revealed that residues 585-642 are sufficient for interaction with TERB2$_{1\text{-}119LL}$, named the TERB2 binding motif (TERB1$_{T2B}$).

MBP-TERB2$_{1\text{-}119LL}$:His$_6$-TERB1$_{T2B}$ was further purified by anion exchange chromatography, which showed co-elution across the salt concentration gradient, indicating stable complex formation. Fusion proteins were subjected to TEV-cleavage and further purified by the same steps as original construct. Analysis of fusion and cleaved complexes by SEC-MALS determined experimental molecular weights of 68.6 and 18 kDa, respectively, corresponding to a 1:1 complex (theoretical 1:1 – 71.4 and 21.1 kDa) (Figure 3.2.16.b-d.). This result indicates that the TERB1 truncation does not affect the oligomeric status of TERB1-TERB2 complex. Furthermore, we determined that the TRF1 binding motif, TBM (residues 642-658), does not contribute to TERB2 binding, confirming that TERB2 and TRF1 have separate binding sites.

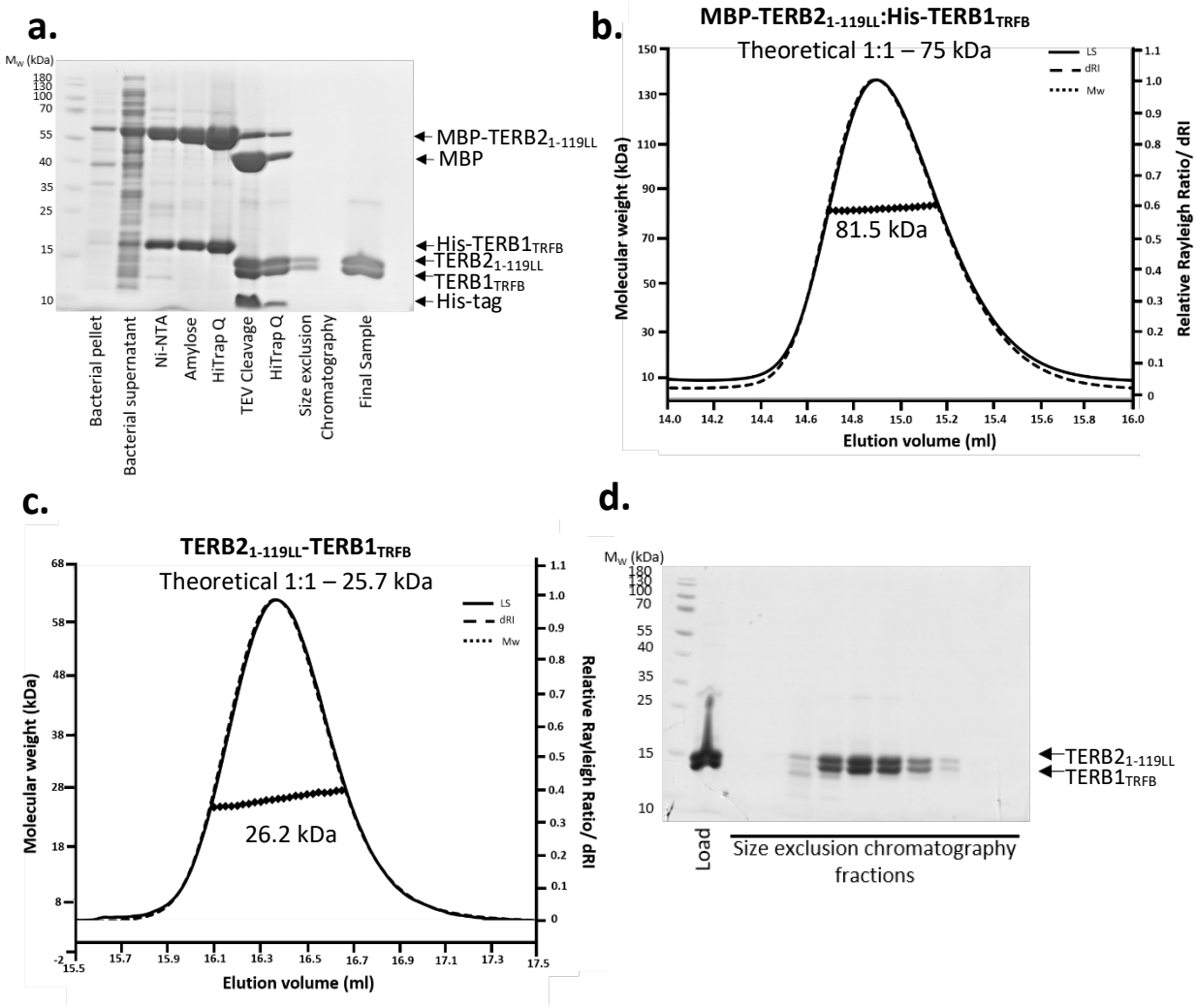


Figure 3.2.15.| Purification and biophysical analysis of TERB1$_{TRFB}$-TERB2$_{1-119LL}$.
a) SDS-PAGE showing the purification summary of TERB2$_{1-119LL}$-TERB1$_{TRFB}$ through Ni-NTA, amylose and anion exchange chromatography. N-terminal solubility tags were removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. b-d) SEC-MALS analysis of b) fusion and c) cleaved TERB2$_{1-119LL}$-TERB1$_{TRFB}$ both elution profiles have a single peak with experimental molecular weights of 81.5 and 26.2 kDa, respectively. A theoretical 1:1 complex has an estimated molecular weight of 75 and 25.7 kDa, respectively indicating a single TERB1 molecule binds a TERB2 molecule. d) SDS-PAGE analysis of the SEC-MALS shows co-elution of the TERB1-TERB2 complex.

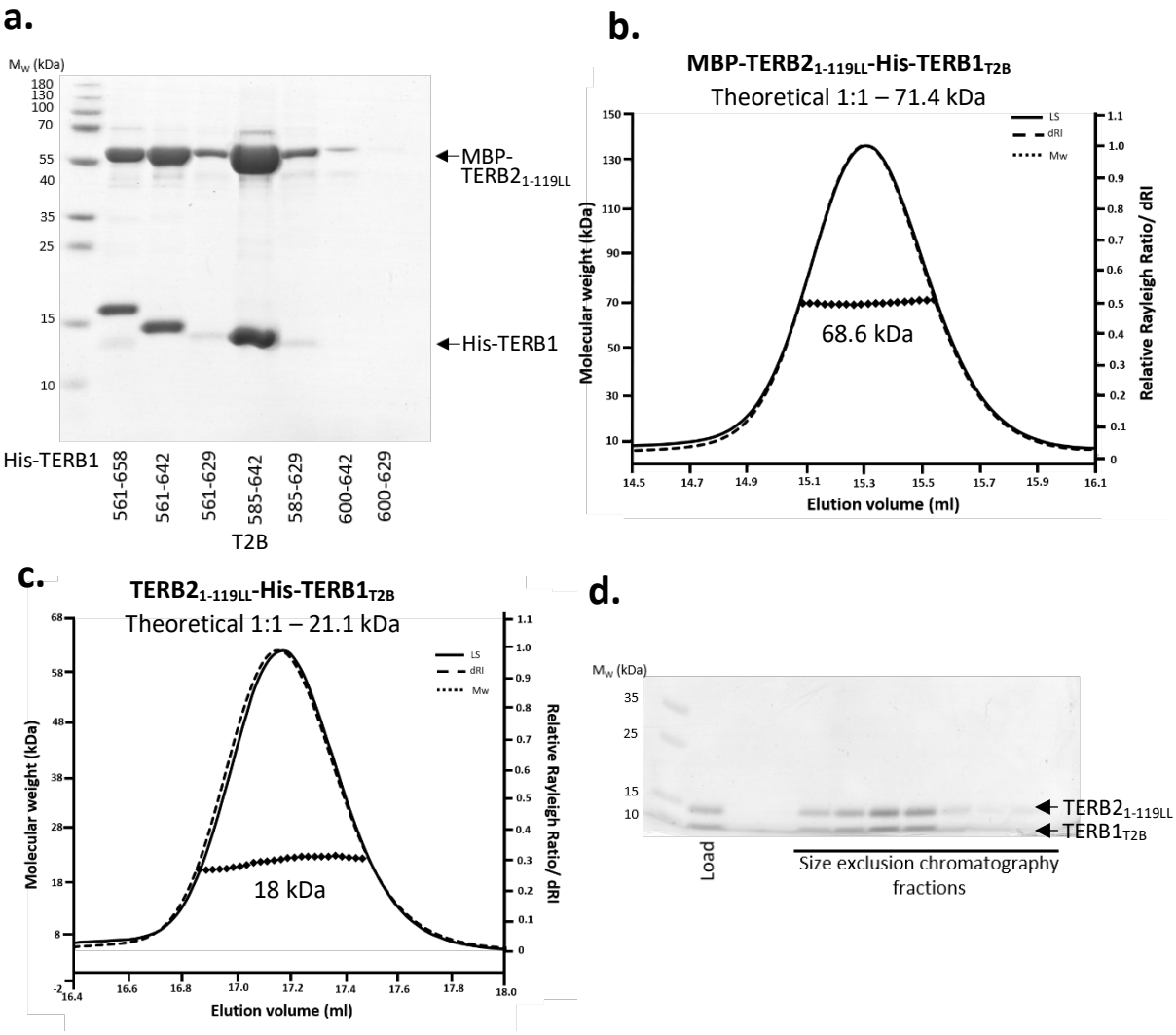


Figure 3.2.16.| Determination of the TERB2 binding site within TERB1$_{TRFB}$
a) SDS-PAGE of MBP-TERB21-$_{119LL}$-His-TERB1 amylose eluates. N- and C-terminal truncated His-TERB1 constructs were co-expressed with MBP-TERB2 and incubated with Ni-NTA resin and Ni-NTA eluate was subsequently incubated in amylose resin, followed by a wash step and elution of the bound protein. b-c) SEC-MALS analysis of b) fusion and c) cleaved TERB2$_{1-119LL}$-TERB1$_{T2B}$. Both elution profiles have a single peak with experimental molecular weights of 68.6 and 18 kDa, respectively. A theoretical 1:1 complex has an estimated molecular weight of 71.4 and 21.1 kDa, respectively. d) SDS-PAGE analysis of the SEC-MALS shows co-elution of TERB2$_{1-119LL}$-TERB1$_{T2B}$.

3.2.10. Optimisation of the TERB2 construct

We analysed the solution structure of TERB2$_{1-119LL}$-TERB1$_{TRFB}$ by SEC-SAXS (Figures 3.2.17.a-e.). SEC-SAXS analysis determined an interatomic distance distribution profile of a typical globular protein with an elongated tail, resulting in a *Dmax* of 130 Å (Figure 3.2.17.a). Further analysis of the SAXS

data including the Kratky plot ($I(q)*q^2$ vs. q) and Porod-Debye plot ($q^4I(q)$ vs. q^4) are informative to check the flexibility and globularity of proteins. These plots suggest that there is significant flexibility within $TERB2_{1\text{-}119LL}$-$TERB1_{TRFB}$ (Figures 3.2.17.d. and e.). These results suggest that the $TERB2_{1\text{-}119LL}$-$TERB1_{TRFB}$ complex is elongated and has flexibility, which we propose is due to the TERB2 C-terminus.

A limited proteolysis assay was ultilised to probe the $TERB2_{1\text{-}119LL}$ structure. As TERB2 is a soluble aggregate when expressed alone, we used the $TERB2_{1\text{-}119LL}$-$TERB1_{T2B}$ complex in which the TERB1 construct had already been optimised. $TERB2_{1\text{-}119LL}$-$TERB1_{T2B}$ was incubated with trypsin protease, to investigate the accessibility of the TERB2 construct. Samples were taken over a 16-hour time frame and an SDS-PAGE was used to analyse the result (Figure 3.2.17.f.). Analysis of the gel shows that after 180 minutes there is a single degradation product with molecular mass less than 10 kDa. To further confirm this, TERB2 secondary structure prediction and sequence conservation were analysed and a series of TERB2 truncations were designed and cloned.

His-$TERB1_{T2B}$ was co-expressed with MBP-TERB2 constructs ($TERB2_{34\text{-}119}$, $TERB2_{1\text{-}107}$ and $TERB2_{1\text{-}96}$) and co-purified by sequential Ni-NTA and amylose affinity chromatography. Resultant amylose eluates were analysed by SDS-PAGE (data not shown). From the results we determined that the N-terminal truncation, MBP-$TERB2_{34\text{-}119LL}$ and C-terminal truncation, MBP-$TERB2_{1\text{-}96LL}$ no longer interacted with TERB1. Dissection of the TERB2 construct revealed that residues 1-107 are sufficient for interaction with $TERB1_{T2B}$, herein known as $TERB2_N$.

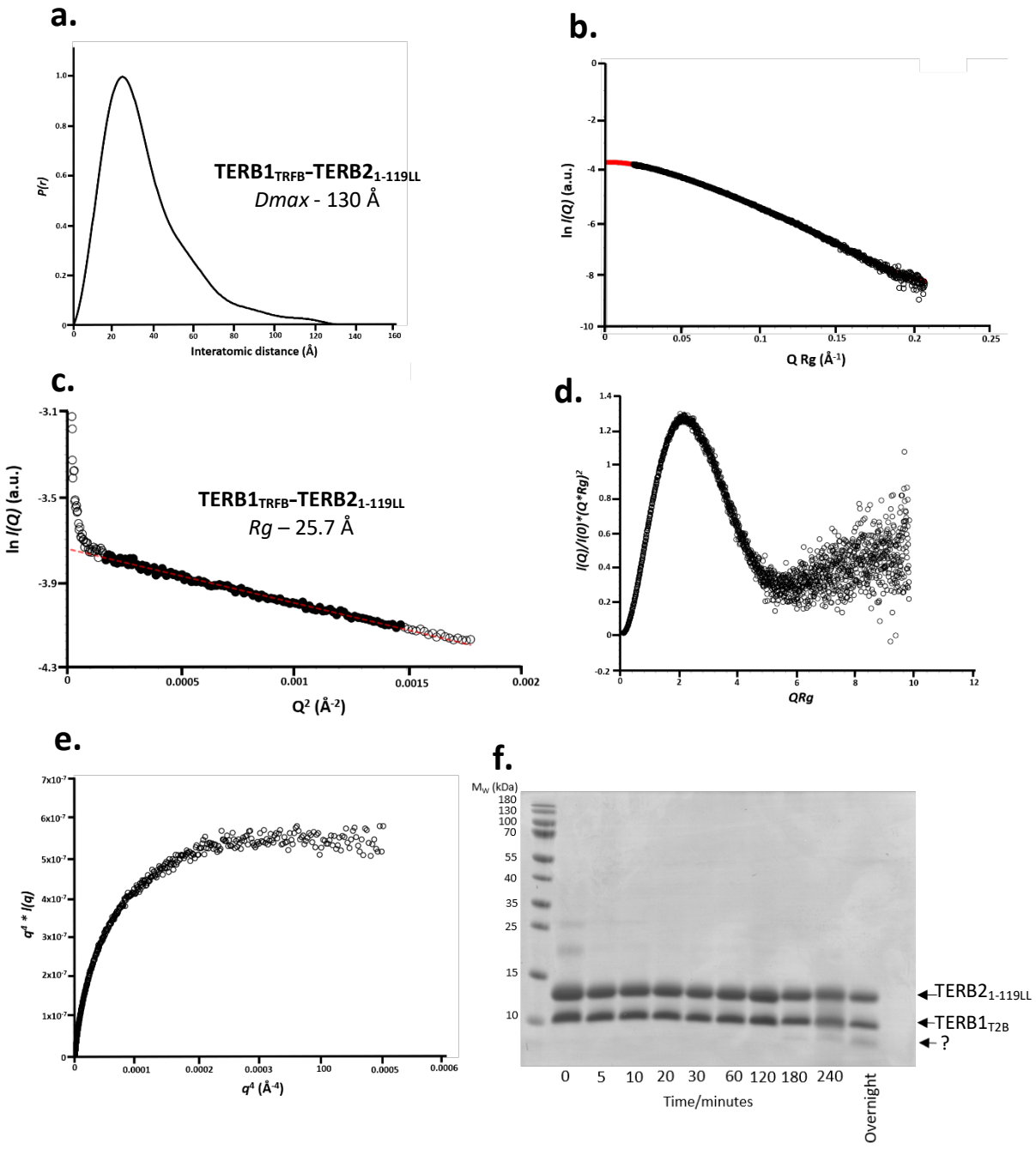


Figure 3.2.17.| SEC-SAXS analysis and limited proteolysis of TERB2-TERB1.
a-e) SAXS analysis of TERB2$_{1\text{-}119LL}$-TERB1$_{TRFB}$. a) *P(r)* distribution showing a maximum dimension of 130 Å. b) Averaged experimental SAXS profile of TERB2$_{1\text{-}119LL}$-TERB1$_{TRFB}$, with the fit used for the *P(r)* distribution shown in red. c) Guinier analysis determined a radius of gyration (*Rg*) value of 25.7 Å. Empty circles represent the complete dataset and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a dashed line. (*Q.Rg* values were < 1.3). d) Normalised Kratky plot for TERB2$_{1\text{-}119LL}$-TERB1$_{TRFB.}$ e) Porod-Debye plot showing flexibility of TERB2$_{1\text{-}119LL}$-TERB1$_{TRFB.}$ f) SDS-PAGE analysis showing limited proteolysis of TERB1-TERB2 using trypsin protease. Samples were taken at a number of time points. Indicated by an arrow is a possible digested product.

3.2.11. Biophysical characterisation of TERB2$_N$-TERB1$_{T2B}$

MBP-TERB2$_N$ was co-expressed with His$_6$-TERB1$_{T2B}$ in *E. coli* BL21 DE3 cells and co-purified using an amylose affinity column. A Ni-NTA purification was not used as in test purifications the complex eluted in the 20 mM imidazole wash step. We next performed an anion exchange chromatography step which was followed by TEV-cleavage. The cleaved material was applied to the cation exchange chromatography column (HiTrap-SP) and further purified by size exclusion chromatography. Protein purification steps summarised in figure 3.2.18.a.

Fusion and cleaved protein were analysed by SEC-MALS, which determined experimental molecular weights of 67 and 19 kDa, respectively, corresponding to a 1:1 complex (theoretical 1:1 – 70 and 20 kDa) (Figure 3.2.18.b-d). CD spectroscopy of TERB2$_N$-TERB1$_{T2B}$, revealed a mixture of alpha helical and beta sheet composition. Analysis by deconvolution estimated the alpha helical and beta sheet content to be 44% and 18%, respectively (Figure 3.2.18.e.). CD thermal melt shows a steady progression of protein unfolding at 55°C, indicating co-operative unfolding of TERB2$_N$-TERB1$_{T2B}$ (Figure 3.2.18.f.).

SEC-SAXS analysis of TERB2$_N$-TERB1$_{T2B}$ was utilised to gain low resolution structural data, and more importantly determine if this construct is a good target for crystallography. TERB2$_N$-TERB1$_{T2B}$ distance distribution profile has a globular bell-shaped curve and presents a more compact *Dmax* of 55 Å in comparison to 130 Å of TERB2$_{1\text{-}119LL}$-TERB1$_{TRFB}$ (Figure 3.2.19.a.). TERB2$_N$-TERB1$_{T2B}$ has a Guinier analysis determined *Rg* of 19 Å, which matches the real-space *Rg* (Figure 3.2.19.b.). *Ab initio* modelling of TERB2$_N$-TERB1$_{T2B}$ was performed by Dr Owen Davies to produce a low-resolution molecular envelope from the SEC-SAXS scattering data. The resultant envelope demonstrates the globular nature of the complex, with the formation of a blob-like structure indicating that TERB1 and TERB2 are closely associated (Figure 3.2.10.c). The filtered averaged *ab initio* model was generated from 30 independent DAMMIF runs. The *ab initio* modelled theoretical curve closely matches the experimental scattering data with a χ^2 value of 1.9 suggesting that the derived model is a good fit (Figure 3.2.19.d).

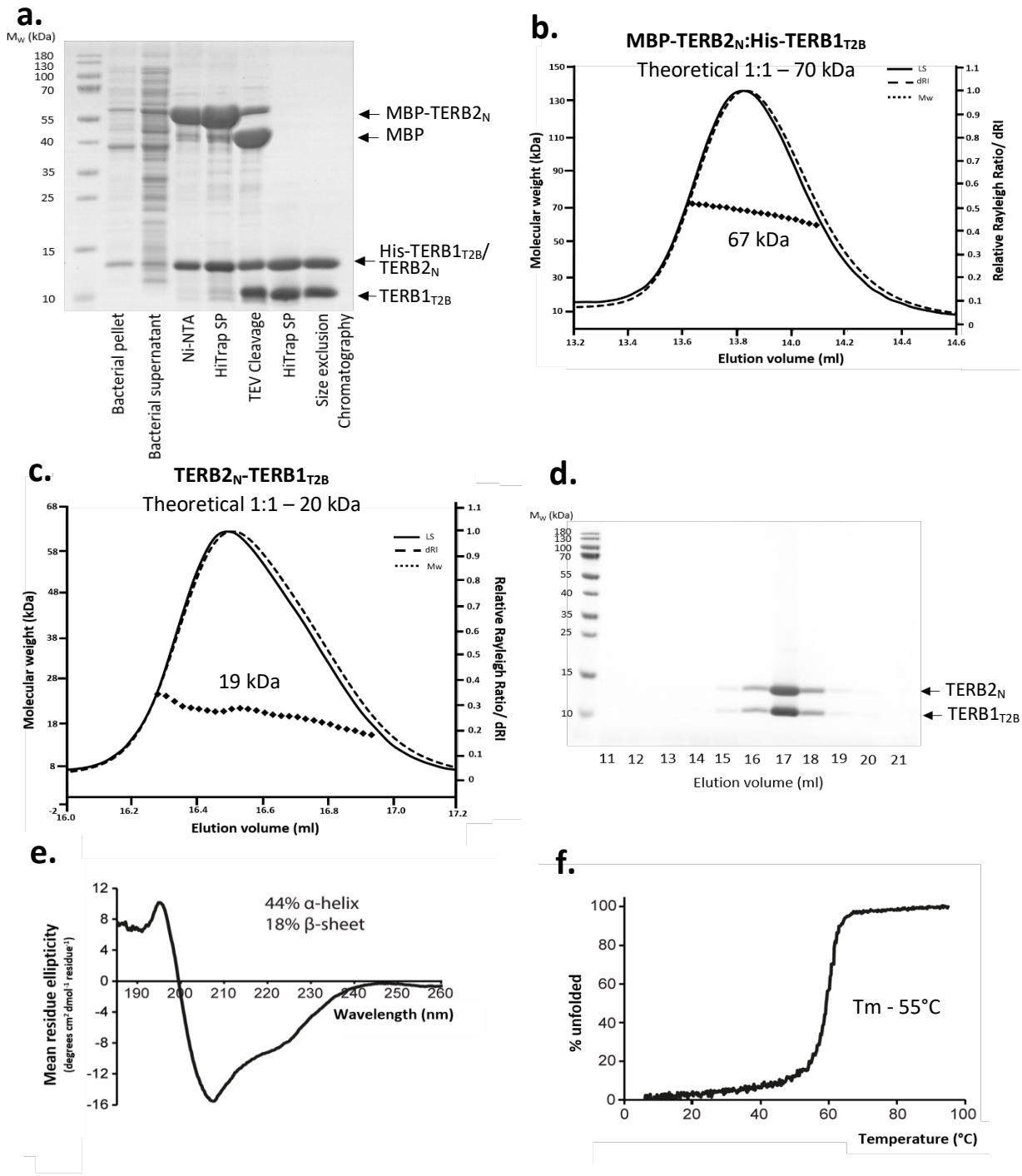


Figure 3.2.18.| Purification and biophysical analysis of TERB2$_N$-TERB1$_{T2B}$

a) SDS-PAGE showing the purification summary of TERB2$_N$-TERB1$_{T2B}$ through Ni-NTA, amylose and anion exchange chromatography. N-terminal solubility tags were removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. b-d) SEC-MALS analysis of b) fusion and d) cleaved TERB2$_N$-TERB1$_{T2B}$ both elution profiles have a single peak with experimental molecular weights of 67 and 19 kDa, respectively. A theoretical 2:1 complex has an estimated molecular weight of 70 and 20 kDa, respectively. c) SDS-PAGE analysis of the SEC-MALS output of 1ml fractions, shows co-elution of the TERB1-TERB2 complex. e) Far UV CD wavelength scan between 260-185nm of TERB2$_N$-TERB1$_{T2B}$. Deconvolution of the data estimates the secondary structure to consist of 44% α-helices and 18% β-sheets. f) CD thermal denaturation measured at 222nm between 4 and 95°C, estimated a melting temperature of 55°C.

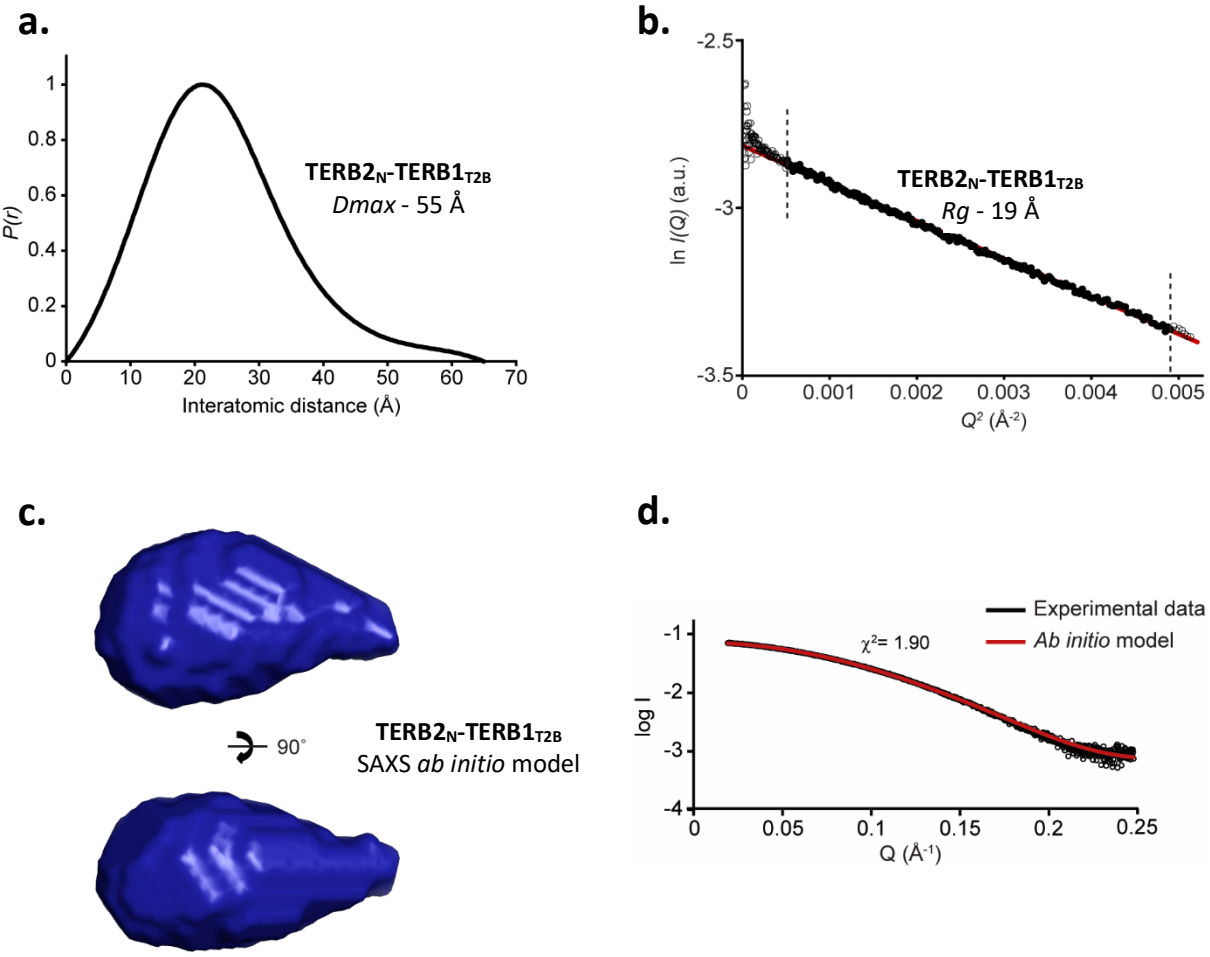


Figure 3.2.19.| SEC-SAXS data analysis of TERB2$_N$-TERB1$_{T2B}$
a-d) SAXS analysis of TERB2$_N$-TERB1$_{T2B}$ a) *P(r)* distribution of TERB2$_N$-TERB1$_{T2B}$ showing a maximum dimension of 55 Å. b) Guinier analysis determined a radius of gyration (*Rg*) value of 19 Å. The real space *Rg* and Guinier analysis *Rg* value both give a value of 19 Å. Clear circles represent the complete data set and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a dashed line. (*Q.Rg* values were < 1.3). c) SAXS *ab initio* model of TERB2$_N$-TERB1$_{T2B}$ using the modelling program, DAMMIF. TERB2-TERB1 has a globular envelope with an extended TERB2 region, also observed in the *P(r)*. d) SAXS scattering of TERB2$_N$-TERB1$_{T2B}$ (black) overlaid with the theoretical scattering curve of the modelled structure. SAXS data was analysed and figures made by Dr Owen Davies, published in Dunce, Milburn, et al., 2018.

3.2.12. TERB1 T2B binding site solely binds to TERB2

To confirm our finding that within the TERB1$_{TRFB}$ domain (amino acid residues 561-658) there is a specific TERB2 binding motif (TERB1$_{T2B}$) we co-expressed TERB1$_{T2B}$ with TRF1 to see if they would co-purify. His-TRF1$_{TRFH}$ was co-expressed with MBP-TERB1$_{T2B}$ in *E.coli* BL21 DE3 cells and co-purified by sequential Ni-NTA and amylose affinity chromatography. Analysis by SDS-PAGE revealed that the TRF1$_{TRFH}$ does not interact with TERB1$_{T2B}$, observed by the presence of TRF1$_{TRFH}$ in the amylose flow through (Figure 3.2.20.a.). Furthermore, this was tested by the co-purification of His-TRF1$_{TRFH,}$ His-TERB1$_{T2B}$ and MBP-TERB2$_{N}$. SDS-PAGE analysis determined the same result, in which His-TRF1$_{TRFH}$ is in the amylose flow through and both His-TERB1$_{T2B}$ and MBP-TERB2$_{N}$ are present in the amylose eluate, confirming an interaction (Figure 3.2.20.b.).

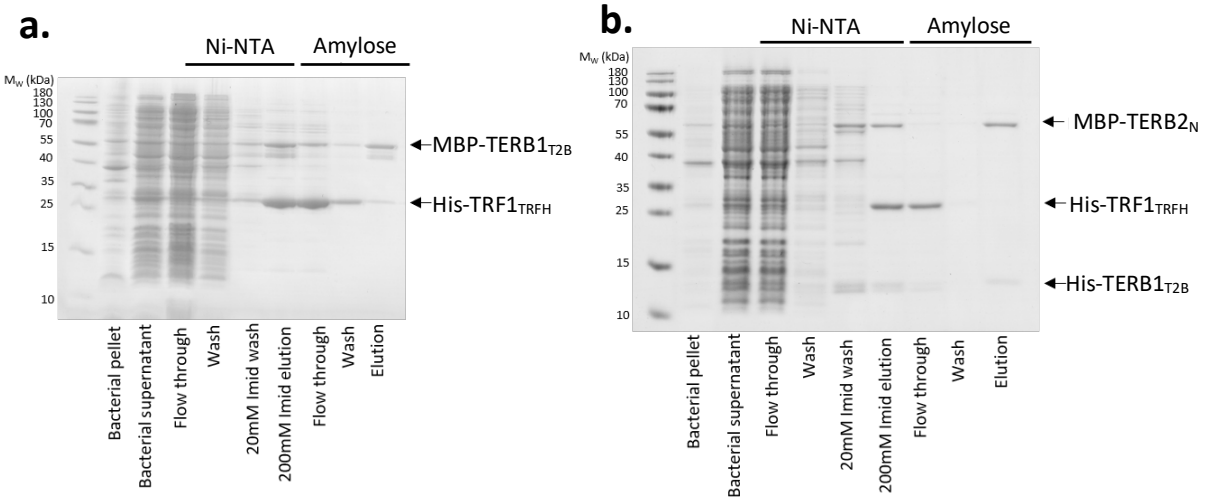


Figure 3.2.20.| SDS-PAGE analysis showing TERB1 has individual TRF1 and TERB2 binding sites.
a) SDS-PAGE showing Ni-NTA and amylose affinity chromatography steps of co-expressed His-TRF1$_{TRFH}$:MBP-TERB1$_{T2B}$. TERB1$_{T2B}$ has a N-terminal MBP and His-tag therefore both proteins can be purified by Ni-NTA resin. His-TRF1$_{TRFH}$ is present in the amylose flow through, indicating there is no binding between TRF1$_{TRFH}$ and TERB1$_{T2B}$ suggesting that this region of TERB1 is exclusively the TERB2 binding motif (T2B). b) Co-purification of His-TRF1$_{TRFH}$:His-TERB1$_{T2B}$:MBP-TERB2$_{N.}$ by amylose affinity shows that His-TRF1$_{TRFH}$ is present in the amylose wash. His-TERB1$_{T2B}$ and MBP-TERB2$_{N}$ are present in the amylose eluate indicating no binding.

3.2.13. TERB1 bonsai construct stably binds TRF1 to form a 2:1 complex

We have structurally characterised the interaction between TRF1 and TERB1 in solution. However, our main target was to determine the precise molecular basis of the TRF1-TERB1 complex. To address this crystallographic studies were initiated with the aim of obtaining a high-resolution structural information of the $TRF1_{TRFH}$-$TERB1_{TRFB}$ complex. Screening of TRF1-TERB1 was carried out at a range of concentrations (5-20 mg/ml) but no crystal hits were observed. After extensive crystallisation efforts no diffraction quality protein crystal were obtained. It is possible that a different optimised construct, with reduced flexibility for example, may be required to obtain crystals of the TRF1-TERB1 complex.

Two other research groups have successfully co-crystallised $TRF1_{TRFH}$ with a $TERB1_{TBM}$ peptide (PDB accession codes: 5XUP and 5WIR), in which two TERB1 peptides bind to the $TRF1_{TRFH}$ dimer, resulting in a 2:2 stoichiometry. Furthermore, analysis of the crystal packing of $TRF1_{TRFH}$-$TERB1_{TBM}$ (PDB accession code: 5WIR) shows that there is space in the packing for a larger TERB1 construct (Figure 3.1.11.).

We propose that residues flanking the $TERB1_{T2B}$ within the TRFB domain of TERB1 are required for the stable binding of TRF1. We hypothesise that TERB1 residues 561-585 bind to a second binding site of the $TRF1_{TRFH}$ dimer in a way which might mimic the $TERB1_{TBM}$ binding interaction, thereby forming a 2:1 complex that we have observed by SEC-MALS. To test this, four new TERB1 constructs were designed. Three TERB1 bonsai constructs were designed by linking together amino acid residues 561-585 and 642-658 with differing lengths of threonine-glycine-serine (TGS) flexible linker regions, (5xTGS, 3xTGS and 1xTGS) and finally a no linker, directly fused, construct (Figure 3.2.21.).

Bonsai constructs were achieved by overlapping PCR and cloned into pMAT11 vector. MBP-TERB1 constructs were co-expressed with His-$TRF1_{TRFH}$ and co-purified by sequential Ni-NTA and amylose affinity chromatography. SDS-PAGE analysis of the resultant amylose eluates determined that all four MBP-$TERB1_{bonsai}$ constructs interact with His-$TRF1_{TRFH}$ (Figure 3.2.22.a-d.).

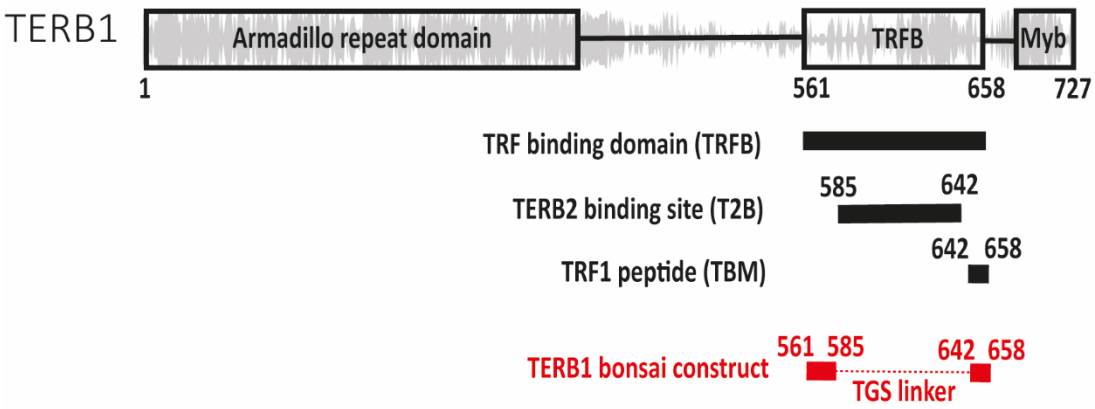


Figure 3.2.21.| Schematic of TERB1 bonsai constructs.
It is known that TERB1 interacts with both TRF1 and TERB2 through its TRFB domain ($TERB1_{561-658}$). A TERB1 bonsai construct was designed connecting the flanking regions of the T2B domain, aa. 561-585 and aa. 642-658 with a flexible TGS linker. Bonsai constructs of TERB1 was cloned with different lengths TGS linker; 5x, 3x, 1x as well as no linker present.

His-$TRF1_{TRFH}$:MBP-$TERB1_{bonsai}$ constructs were further purified by anion exchange chromatography and subjected to SEC-MALS analysis to determine the oligomeric properties (Figure 3.2.22.e.). Fusion proteins were initially studied by SEC-MALS to provide a more accurate oligomeric status as uncleaved $TERB1_{bonsai}$ is very small (5kDa) therefore the His_6-MBP tag will remove the ambiguity of the stoichiometry. The SEC-MALS profiles reveal that all four His-$TRF1_{TRFH}$-MBP-$TERB1_{bonsai}$ constructs elute as two peaks with the majority of the material forming a peak with molecular mass corresponding to a 2:1 complex, summarised in Table 3.2.1. These results suggest that the TERB1 bonsai constructs do not affect the TRF1-TERB1 stoichiometry but there is a small peak with a molecular mass of 52.2 kDa corresponding to dimeric His-$TRF1_{TRFH}$ or monomeric MBP-$TERB1_{bonsai}$ (theoretical Mw – 55.2 and 49 kDa, respectively). SEC-MALS analysis of His-$TRF1_{TRFH}$:MBP-$TERB1_{bonsai\ No\ TGS}$ determined a molecular mass of 104.3 kDa indicating 2:1 stoichiometry (theoretical 2:1 – 105.4 kDa) (Figure 3.2.22.f.). These results, summarised in Table 3.2.1., indicate that the bonsai construct does not require a TGS flexible-linker but the observation of free $TRF1_{TRFH}$ and/or MBP-$TERB1_{bonsai}$ suggests that $TRF1_{TRFH}$ in complex with $TERB1_{bonsai}$ is less stable than $TERB1_{TRFB}$. Overlaying the His-$TRF1_{TRFH}$:MBP-$TERB1_{bonsai}$ SEC-MALS elution profiles reveals that as the TGS linker is shortened the protein elutes over a broader range, which could be due to the complex falling apart and that the TERB1

can no longer bind to the two $TRF1_{TRFH}$ binding sites (Figure 3.2.22.e.). However, these observations are only speculative and further analysis of $TRF1_{TRFH}$-$TERB1_{bonsai}$ is required.

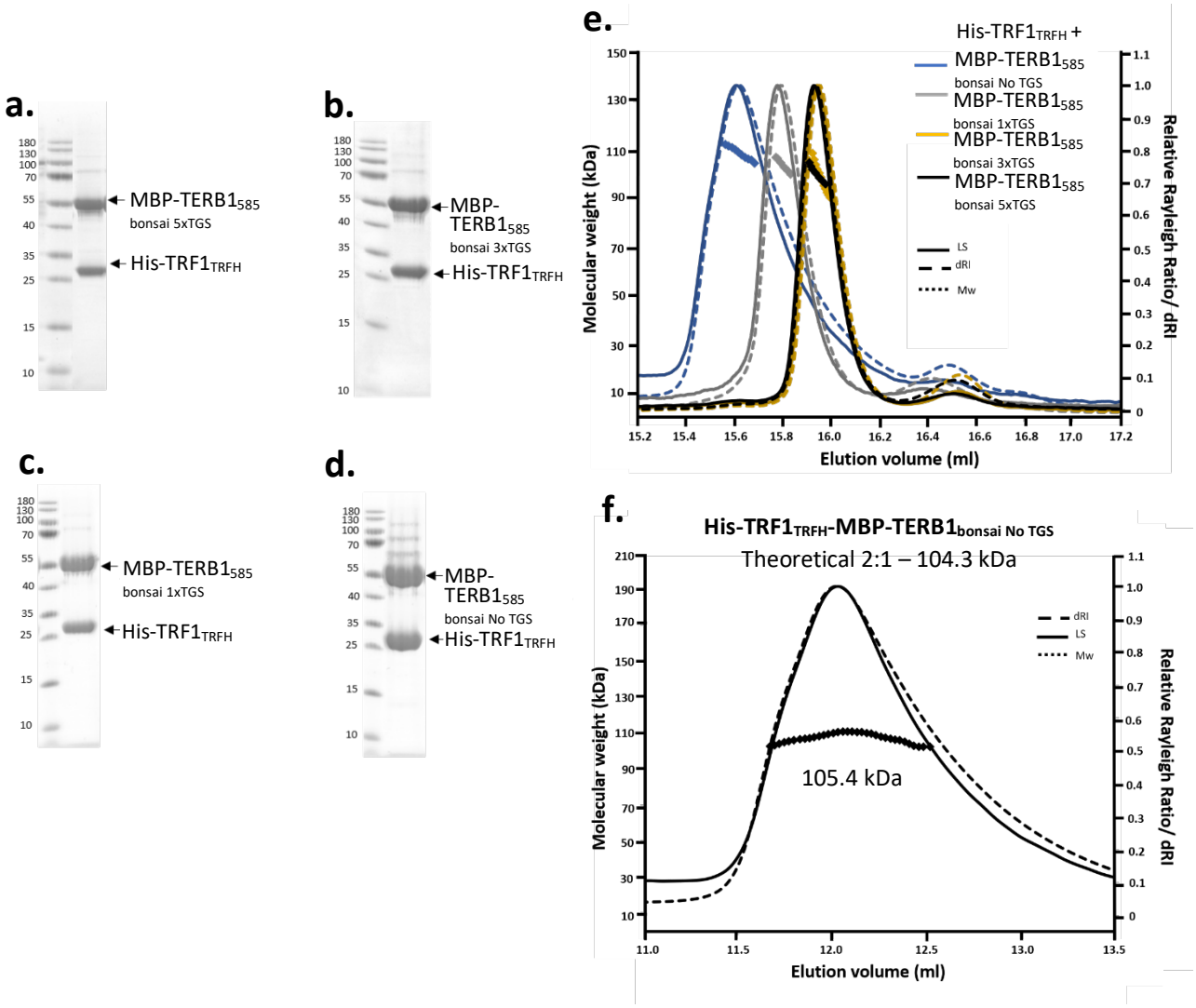


Figure 3.2.22.| SEC-MALS analysis of the fusion $TRF1_{TRFH}$-$TERB1_{bonsai}$ constructs.
a-d) SDS-PAGE analysis of amylose eluates of MBP-$TERB1_{bonsai}$:His-$TRF1_{TRFH}$ complex. Shortening the TGS linker has no effect on the co-purification of the TRF1-TERB1 complex. e-f) SEC-MALS analysis of fusion TRF1-TERB1 bonsai complexes. e) Overlaid SEC-MALS of MBP-$TERB1_{bonsai}$-His-$TRF1_{TRFH}$ complexes shortening the TGS flexible linker. All three bonsai, 5x,3x, 1x and no-TGS linker complexes have absolute molecular weights corresponding to a 2:1 stoichiometry, like the WT complex (Table 3.2.1.). This shows that removing the TERB2 binding site of TERB1 has no effect on the interaction between TRF1 and TERB1. f) SEC-MALS analysis of His-$TRF1_{TRFH}$:MBP-$TERB1_{bonsai}$, elution profile has a single peak with experimental molecular weight of 105.4 kDa. A theoretical 2:1 complex has an estimated molecular weight of 104.3 kDa indicating that the TERB1 bonsai does not affect the 2:1 stoichiometry, a single TERB1 molecule binds a TRF1 homodimer.

Table 3.2.1.| SEC-MALS analysis of TRF1-TERB1 fusion complexes

*Theoretical molecular weight assuming a 2:1 TRF1-TERB1 stoichiometry and that $TRF1_{TRFH}$ is dimeric

Complex	Experimental Mw /kDa	Theoretical Mw/kDa*	Stoichiometry
His-$TRF1_{TRFH}$:MBP-$TERB1_{TRFB}$	114.3	111.8	2:1
His-$TRF1_{TRFH}$:MBP-$TERB1_{bonsai\ 5xTGS}$	100.9	106.1	2:1
His-$TRF1_{TRFH}$:MBP-$TERB1_{bonsai\ 3xTGS}$	99.3	105.9	2:1
His-$TRF1_{TRFH}$:MBP-$TERB1_{bonsai\ 1xTGS}$	104.8	105.6	2.1
His-$TRF1_{TRFH}$:MBP-$TERB1_{bonsai\ No\ TGS}$	104.3	105.4	2:1
His-$TRF1_{TRFH}$:MBP-$TERB1_{TBM}$	95.6 kDa (peak 1) 46 kDa (peak 2)	102.6	2:1 species with free TERB1/TRF1
His-$TRF1_{TRFH}$	50.5	55.5	dimer

Analysis by SDS-PAGE of anion exchange elution profiles of His-$TRF1_{TRFH}$ -MBP-$TERB1_{TRFB}$ and His-$TRF1_{TRFH}$:MBP-$TERB1_{bonsai\ No\ TGS}$ show perfect co-elution across the salt concentration gradient, in comparison to His-$TRF1_{TRFH}$:MBP-$TERB1_{TBM}$ which dissociates (Figure 3.2.23.a.). Overlay of the SEC-MALS analysis of these three fusion complexes indicate that His-$TRF1_{TRFH}$:MBP-$TERB1_{bonsai\ No\ TGS}$ is more stable than His-$TRF1_{TRFH}$:MBP-$TERB1_{TBM}$ but still has some dissociation (Figure 3.2.23.b.).

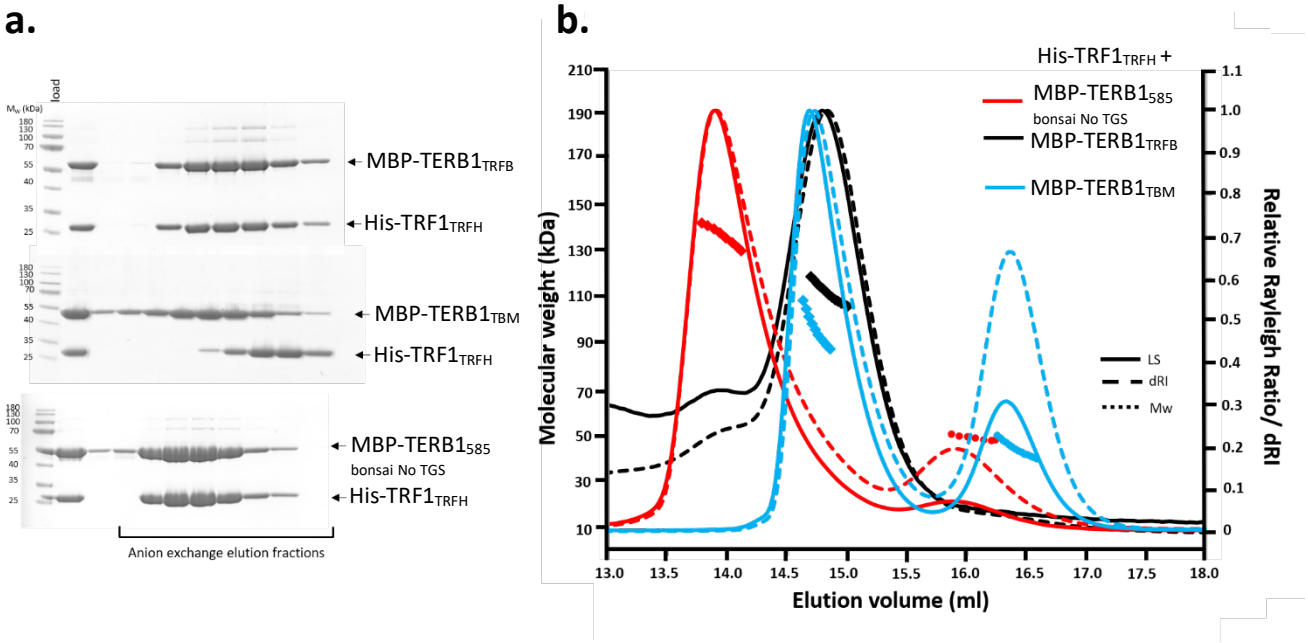


Figure 3.2.23.| TRF1$_{TRFH}$-TERB1$_{bonsai}$ forms a stable 2:1 complex.
a) SDS-PAGE analysis of His-TRF1:MBP-TERB1 anion exchange chromatography elution profiles with different TERB1 constructs. b) SEC-MALS analysis of the His-TRF1:MBP-TERB1 constructs.

3.2.14. Structural analysis of TRF1$_{TRFH}$-TERB1$_{bonsai}$

The fusion material of the shortest no linker construct of TERB1 (TERB1$_{585bonsai}$) in complex with TRF1$_{TRFH}$ was cleaved with TEV protease and subsequently purified by sequential anion exchange and size exclusion chromatography. Protein purification steps are summarised in Figure 3.2.24.a. SEC-MALS analysis of TRF1$_{TRFH}$-TERB1$_{585bonsai}$ produced a single peak with a molecular mass of 49 kDa which we predict to be a 2:1 complex (theoretical 2:1 – 53 kDa), despite a 4 kDa difference to the theoretical molecular weight (Figure 3.2.24.b.). These findings indicate that the linker insertion connecting the two TERB1 regions does not affect TERB1 binding to the TRF1 dimer. SDS-PAGE analysis of the final cleaved products of TRF1$_{TRFH}$-TERB1$_{585bonsai}$ and TRF1$_{TRFH}$-TERB1$_{TRFB}$ suggest that there could be excess TRF1$_{TRFH}$ when complexed with TERB1$_{585bonsai}$ Figure 3.2.24.c.).

We studied TRF1$_{TRFH}$-TERB1$_{585bonsai}$ by SEC-SAXS analysis to investigate the structural parameters in solution and whether the TERB1 bonsai construct had affected the structure. SEC-SAXS revealed an interatomic distance distribution profile with a typical Gaussian distribution indicative of a globular

protein, with a slight elongation suggesting there is some flexibility in the construct (Figure 3.2.24.e). The *Dmax* is 130 Å, compared to $TRF1_{TRFH}$-$TERB1_{TRFB}$ which has a *Dmax* of 110 Å. The radius of gyration (*Rg*) was determined at 31.3 Å, indicating that the overall size of is unaffected by the $TERB1_{bonsai}$ construct (Figure 3.2.24.f.). Both the *P(r)* distribution and guinier of $TRF1_{TRFH}$-$TERB1_{bonsai}$ and $TRF1_{TRFH}$-$TERB1_{TRFB}$ are comparable suggesting removing the TERB1 T2B binding site does not drastically affect the overall TRF1-TERB1 structure.

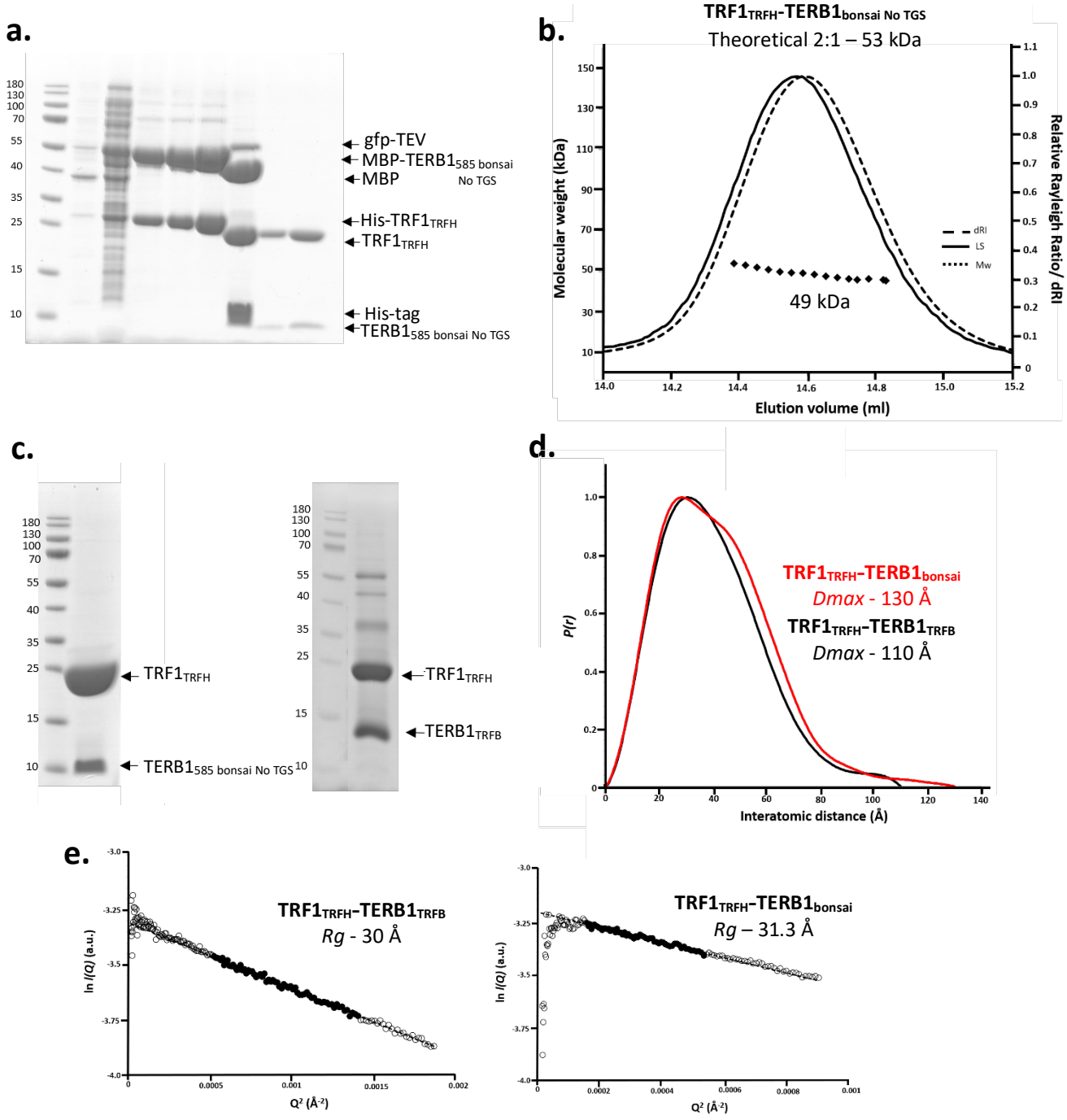


Figure 3.2.24.| Purification and biophysical analysis of TRF1$_{TRFH}$-TERB1$_{bonsai}$

a) SDS-PAGE showing the purification summary of TRF1$_{TRFH}$-TERB1$_{bonsai}$ through Ni-NTA, amylose and anion exchange chromatography. N-terminal solubility tags were removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. b) SEC-MALS analysis of TRF1$_{TRFH}$-TERB1$_{bonsai,}$ elution profile has a single peak with an experimental molecular weight of 49 kDa. A theoretical 2:1 complex has an estimated molecular weight of 53 kDa indicating 2:1 stoichiometry. c) SDS-PAGE showing TRF1$_{TRFH}$-TERB1$_{TRFB}$ (left) and TRF1$_{TRFH}$-TERB1$_{bonsai}$ (right) final protein products d) *P(r)* distribution of TRF1$_{TRFH}$-TERB1$_{bonsai}$ showing a maximum dimension of 130 Å (red) overlaid with TRF1$_{TRFH}$-TERB1$_{TRFB}$ (black) e) Guinier analysis of TRF1$_{TRFH}$-TERB1$_{bonsai}$ determined a radius of gyration (*Rg*) value of 31.3 Å. Empty circles represent the complete dataset and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a dashed line. (*Q.Rg* values were < 1.3).

3.2.15. Crystallography of $TRF1_{TRFH}$-$TERB1_{bonsai}$

High throughput commercial screening of $TRF1_{TRFH}$-$TERB1_{bonsai}$ was carried out, in addition to the published conditions of $TRF1_{TRFH}$ and $TRF1_{TRFH}$-$TERB1_{TBM}$ (summarised in Table 3.2.2.) at two protein concentrations, 7 mg/ml and 15 mg/ml (Fairall et al., 2001; Long et al., 2017; Pendlebury et al., 2017). Two of the published conditions led to the formation of crystals. However, crystals were much larger in the Tris-HCl, propanol condition. Further optimisations of the published conditions were performed to improve the crystal quality and size (Figure 3.2.25.a.). Crystals were harvested from both the initial and optimised screens and tested for X-ray diffraction (Figure 3.2.25.b.).

Table 3.2.2.| Published structures of TRF1 (1H6O) and TRF1-TERB1 (5WIR and 5XUP). Crystal conditions for both $TRF1_{TRFH}$ and $TRF1_{TRFH}$-$TERB1_{TBM}$ were replicated for $TRF1_{TRFH}$-$TERB1_{bonsai}$ (Fairall et al., 2001; Long et al., 2017; Pendlebury et al., 2017).

PDB code	Construct boundaries	Crystal condition	Resolution (Å)	Space group
1H6O	$TRF1_{62-265}$	100 mM $(CH3)_2AsO_2Na$ (pH 6.0) 5%–15% glycerol 1–5 mM $Mg(OAc)_2$ 1% PEG 8000	2.9	$P3_1$ 2 1
5WIR	$TRF1_{62-265}$ $TERB1_{642-656}$	0.1M Tris-HCL (pH 8.5) 30% PEG 300	2.1	$P6_4$
5XUP	$TRF1_{65-265}$ $TERB1_{642-656}$	10% 2-propanol 0.1M Tris-HCl (pH 8.5)	2.1	$P6_4$

The diffraction data was auto processed utilising the automated pipeline Xia2. The space group $P4_12_12$ was selected with a single molecule in the asymmetric unit. In comparison, $TRF1_{TRFH}$ and $TRF1_{TRFH}$-$TERB1_{TBM}$ crystallised in space groups $P3_12_1$ and $P6_4$, respectively. The structure was solved through molecular replacement using $TRF1_{TRFH}$ (PDB - 1H6O, Fairall et al., 2001) as a search model resulting in a 2.10 Å structure. This process is described in full in the Methods, section 2.6.3. Visualisation of the structure only shows density for the TRF1 molecule, and none for the TERB1 peptide.

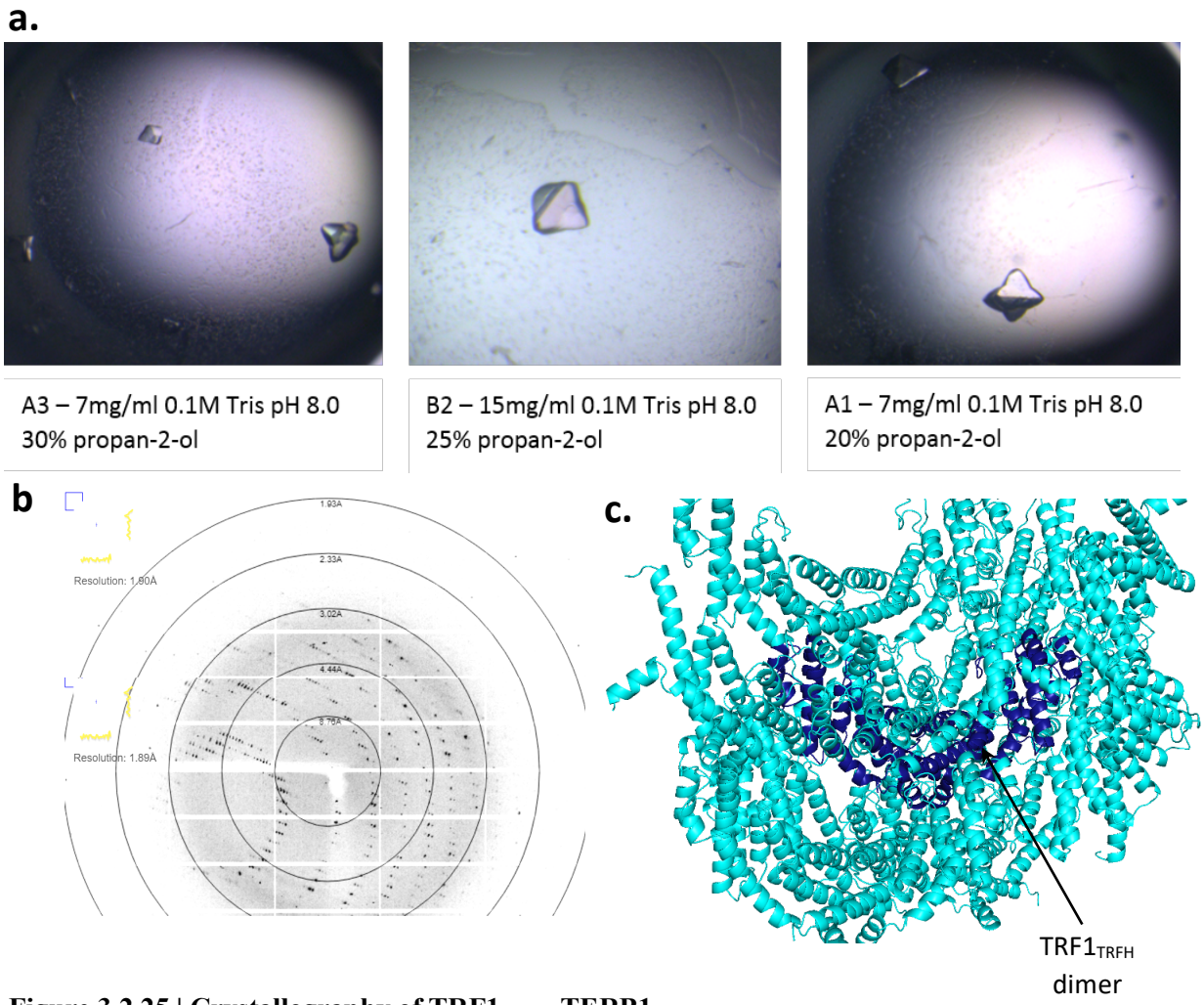


Figure 3.2.25.| Crystallography of $TRF1_{TRFH}$-$TERB1_{585\ bonsai}$
a) Crystal images of the $TRF1_{TRFH}$-$TERB1_{585bonsai}$ complex, in an optimised 10% 2-propanol 0.1M Tris-HCl (pH 8.5) condition. b) Diffraction image of a $TRF1_{TRFH}$-$TERB1_{585\ bonsai}$ crystal on the I24 beamline. The X-ray crystal structure was solved at 2.1 Å (xia2 3dii) by molecular replacement using $TRF1_{TRFH}$ (PDB: 1H6O, Fairall et al., 2001) as a model. c) The solved structure showed only density for TRF1 (blue), with a monomer in the asymmetric unit. Crystal packing also shows that the TRF1 is tightly packed, suggesting that there is no space for the TERB1 construct. This is very different to the crystal packing observed for the $TRF1_{TRFH}$-$TERB1_{TBM}$ structure (PDB: 5WIR, Pendlebury et al., 2017).

Analysis of the crystal packing shows that there is very tight stacking of the TRF1 monomer, suggesting that there is no space present for the TERB1 construct (Figure 3.2.25.c.). From this it is possible to speculate that there could be free TRF1 present in the TRF1-TERB1 complex which is crystallising quickly, and possibly inhibiting crystallisation of the TRF1-TERB1 complex. Another possibility is that the presence of the TERB1 peptide is inhibiting crystallisation or that the complex dissociates in this crystal condition. The commercial screens did not yield any new crystal conditions.

3.2.16. Further optimisation of the TERB1$_{bonsai}$ construct

In a further attempt to aid crystallisation of the TRF1-TERB1 complex, we designed and optimised the TERB1 construct. The N-terminus was further truncated to residues 561-574 and tethered to 642-658, herein known as TERB1$_{574bonsai}$. MBP-TERB1$_{574bonsai}$ was co-expressed with His-TRF1$_{TRFH}$ and co-purified by sequential Ni-NTA and amylose affinity chromatography. His-TRF1$_{TRFH}$:MBP-TERB1$_{574bonsai}$ was further purified by anion exchange chromatography and elution fractions were analysed by SDS-PAGE (Figure 3.2.26.a.). We determined that the two proteins no longer co-eluted observed by a shift in His$_6$-TRF1$_{TRFH}$ elution volume, similar to the observed TRF1$_{TRFH}$-TERB1$_{TBM}$ anion exchange, suggesting that that shortening the TERB1 bonsai construct has weakened the interaction. The anion exchange chromatography elution fractions were pooled and concentrated to 50mg/ml then analysed by SEC-MALS to determine the absolute molecular weight of the fusion complex (Figure 3.2.26.b.). His-TRF1$_{TRFH}$:MBP-TERB1$_{574bonsai}$ eluted as two overlapping peaks with an absolute molecular mass of 64.1 kDa. The expected theoretical molecular weight of a 2:1 complex 103.9 Kda (2:1 stoichiometry) suggesting that the complex is falling apart during the SEC isocratic gradient. The two separate protein components, MBP-TERB1$_{574bonsai}$ (monomeric theoretical Mw - 48.5 kDa) and the His$_6$-TRF1$_{TRFH}$ dimer (Mw 55.5 kDa) are co-eluting due to having similar molecular weights but are no longer a complex. Overlaid SEC-MALS elution profiles of His-TRF1$_{TRFH}$ and His-TRF1$_{TRFH}$:MBP-TERB1$_{574bonsai}$ also suggests that the protein complex has dissociated (Figure 3.2.26.b.). This result indicates that TERB1 amino acid residues 561-585 and 642-658 (TERB1$_{585bonsai}$) is the minimum construct required to stably bind TRF1-TRFH dimer, however a linker may be required to increase stability.

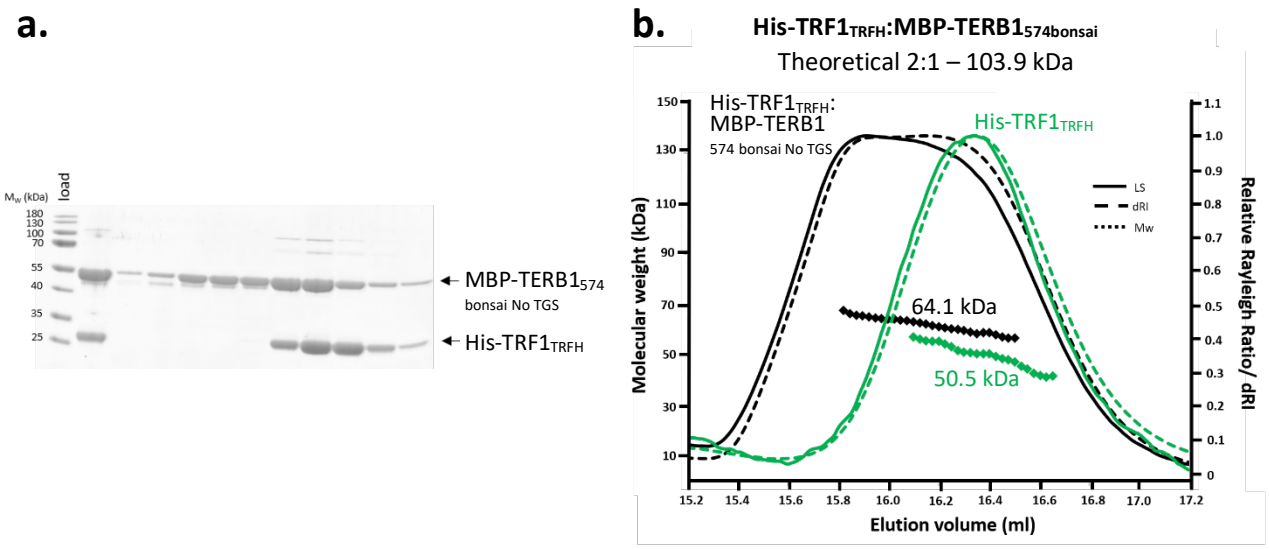


Figure 3.2.26.| Purification and biophysical analysis of His-TRF1$_{TRFH}$:MBP-TERB1$_{574bonsai}$
a) SDS-PAGE showing anion exchange elution profile of His-TRF1$_{TRFH}$:MBP-TERB1$_{574bonsai.}$ MBP-TERB1$_{574bonsai}$ no longer co-elutes with His-TRF1$_{TRFH,}$ observed by His-TRF1$_{TRFH}$ eluting in later fractions. b) SEC-MALS analysis shows that TRF1$_{TRFH}$-TERB1$_{574bonsai}$ elutes as two overlapping peaks with an experimental molecular weight of 64.1 kDa (black). MBP-TERB1$_{574bonsai}$ and His-TRF1$_{TRFH}$ dimer partly co-elute due to having similar molecular weights. Overlaid SEC-MALS of TRF1$_{TRFH}$-TERB1$_{574bonsai}$ (black) with His-TRF1$_{TRFH}$ (green) confirms this.

3.2.17. TRF1:TERB1:TERB2 forms a stable tripartite 2:1:1 complex

We have shown that the shelterin component and telomere binding protein, TRF1, directly interacts with TERB1 through the TRF1 binding domain (TRFB) that includes, and flanks, its TERB2 binding site (T2B). Furthermore, we determined by SEC-MALS analysis that TRF1-TERB1 and TERB1-TERB2 form 2:1 and 1:1 complexes in solution, respectively. To analyse the tripartite TRF1-TERB1-TERB2 complex we co-expressed His-TRF1$_{TRFH}$ with His$_6$-TERB1$_{TRFB}$ and MBP-TERB2$_N$ in *E. coli* BL21 DE3 cells. His-TRF1$_{TRFH}$:His-TERB1$_{TRFB}$:MBP-TERB2$_N$ complex was co-purified ultilising the same procedure as the TRF1 $_{TRFH}$-TERB1$_{TRFB}$ complex, with the TEV-protease cleaved complex purified by anion exchange chromatography and subsequent size exclusion chromatography, summarised in Figure 3.2.27.a. The fusion and cleaved complexes were analysed by SEC-MALS to determine the oligomeric status of the complex. The SEC profiles reveal a single protein peak, with MALS analysis determining the absolute molecular weights to be 128 and 84 kDa, respectively, corresponding to a 2:1:1 complex (theoretical 2:1:1 – 123 and 73 kDa) (Figure 3.2.27.b. and 3.2.27.c.).

SEC-MALS revealed that the TRF1:TERB1:TERB2 oligomerises in accordance with the TRF1-TERB1 and TERB1-TERB2 complex, with the ternary complex forming a 2:1:1 complex, summarised in figures 3.2.27.c. and .d.

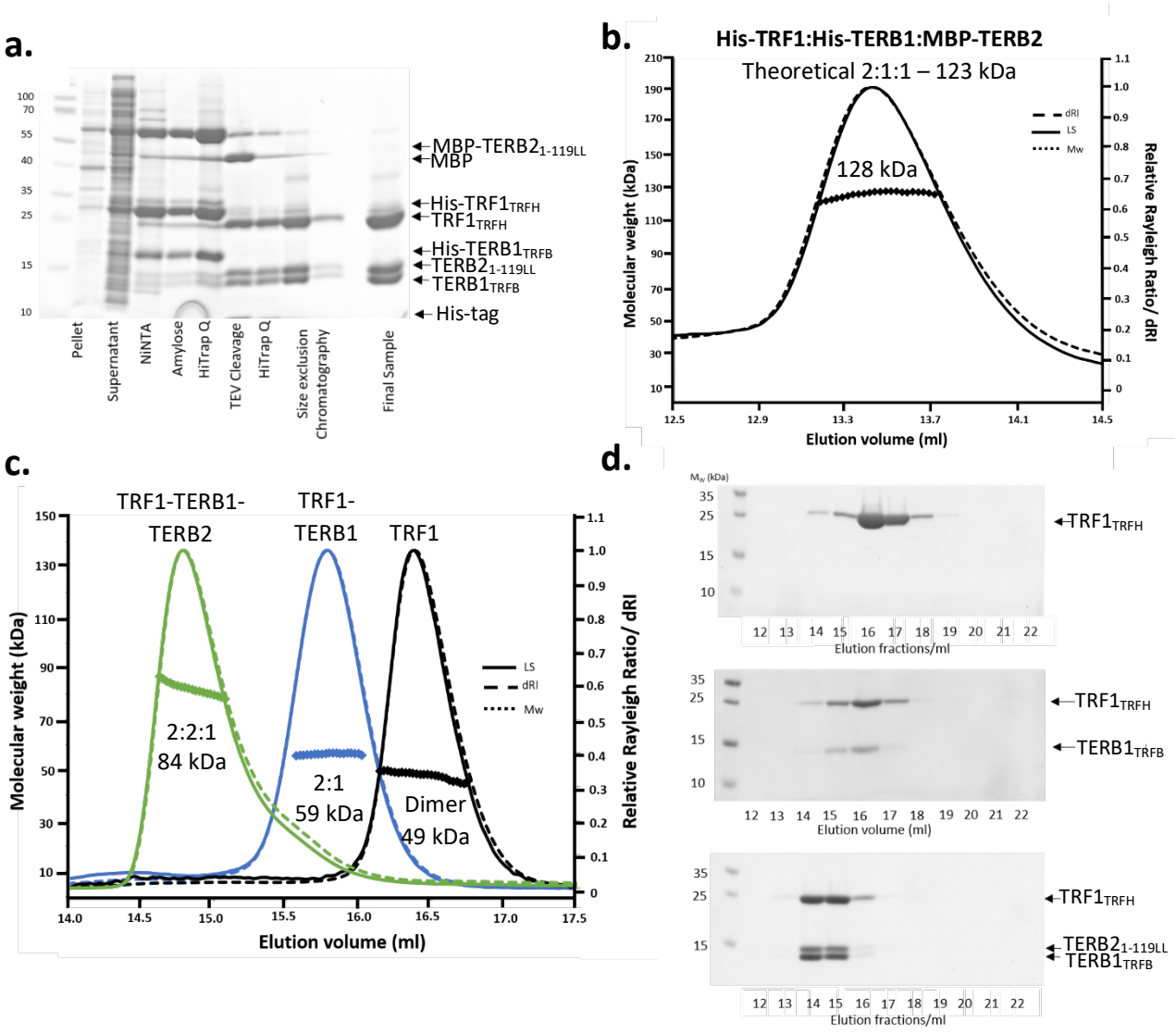


Figure 3.2.27.| TRF1 homodimer complexes with TERB1-TERB2 to form a heterotetramer. a) SDS-PAGE showing the purification summary of $TRF1_{TRFH}$-$TERB1_{TRFB}$-$TERB2_{1-119LL}$ through Ni-NTA, amylose and anion exchange chromatography. N-terminal solubility tags were removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. b) SEC-MALS analysis of His-$TRF1_{TRFH}$-His-$TERB1_{TRFB}$-MBP-$TERB2_{1-119LL}$. The elution profile has a single peak with an experimental molecular weight of 128 kDa. A theoretical 2:1:1 complex has an estimated molecular weight of 123 kDa indicating a TERB1:TERB2 1:1 complex binds a TRF1 homodimer to form a heterotetramer. c) Overlaid SEC-MALS analyses of $TRF1_{TRFH}$, $TRF1_{TRFH}$-$TERB1_{TRFB}$ and $TRF1_{TRFH}$-$TERB1_{TRFB}$-$TERB2_{1-119LL}$. $TRF1_{TRFH}$ homodimerises (black) and recruits a TERB1 monomer to form a 2:1 complex (blue). In turn, TRF1-TERB1-TERB2 forms a ternary 2:1:1 complex (green). b) SDS-PAGE analysis of the SEC-MALS output of 1ml fractions of $TRF1_{TRFH}$ (top) $TRF1_{TRFH}$-$TERB1_{TRFB}$ (middle) and $TRF1_{TRFH}$-$TERB1_{TRFB}$-$TERB2_{1-119LL}$ (bottom).

3.2.18. Multi-phase SEC-SAXS *ab initio* modelling of TRF1:TERB1:TERB2

SEC-SAXS analysis of $TRF1_{TRFH}$-$TERB1_{TRFB}$-$TERB2_N$, carried out by Dr Owen Davies, reveals a gaussian distribution profile indicative of a globular protein complex, with a *Dmax* of 114 Å, in keeping with the *Dmax* values for $TRF1_{TRFH}$ and $TRF1_{TRFH}$-$TERB1_{TRFB}$, of 100 and 105 Å, respectively (Figure 3.2.28.d., Dunce, Milburn et al., 2018). The *Rg* of $TRF1_{TRFH}$-$TERB1_{TRFB}$-$TERB2_N$ was determined to be 33 Å (Figure 3.2.28.a.). These results indicate that the addition of $TERB2_N$ does not greatly change the maximum dimensions and shape of the ternary complex, indicating that $TERB1_{TRFB}$-$TERB2_N$ closely associates with $TRF1_{TRFH}$ dimer (Figures 3.2.28.a-d.).

We determined the molecular architecture of the $TRF1_{TRFH}$-$TERB1_{TRFB}$-$TERB2_N$ ternary complex through SEC-SAXS multi-phase *ab initio* modelling. The resultant model suggests the globular $TERB1_{TRFB}$-$TERB2_N$ molecular envelopes are closely associated and dock into the $TRF1_{TRFH}$ V-shaped molecular envelope via a TERB1 monomer binding to the TRF1 dimeric cleft (Figure 3.2.28.f.). The recently reported crystal structure of $TRF1_{TRFH}$-$TERB1_{TBM}$ (PDB accession code – 5WIR) can be docked into the $TRF1_{TRFH}$ molecular envelope confirming that TERB1 binds to the TRF1 cleft but in a 2:1 stoichiometry. The model determined that the $TERB2_N$ molecular envelope is away from the $TRF1_{TRFH}$ cleft, indicating TERB1 mediates the interaction between TRF1 and TERB2. The multi-phase SAXS *ab initio* modelled theoretical curve closely matches the experimental scattering data with a χ^2 value of 1.02 suggesting that the derived model is a good fit (Figure 3.2.28.e.).

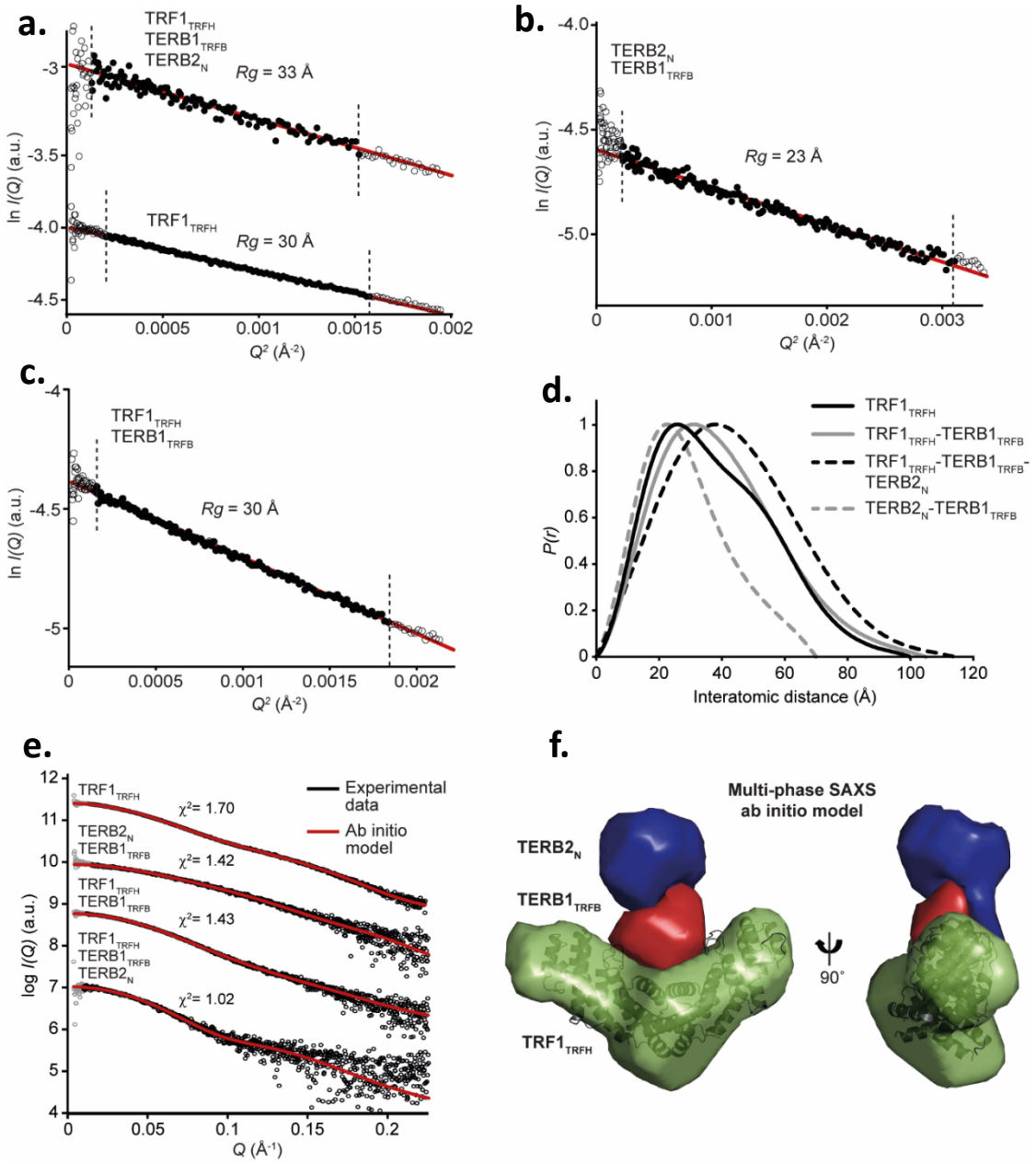
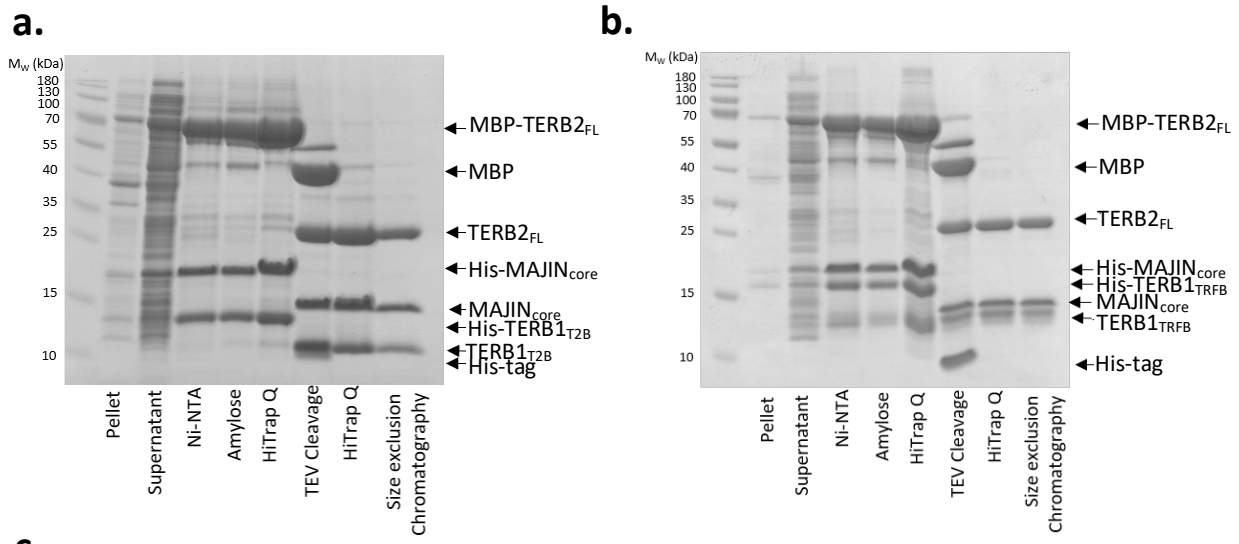
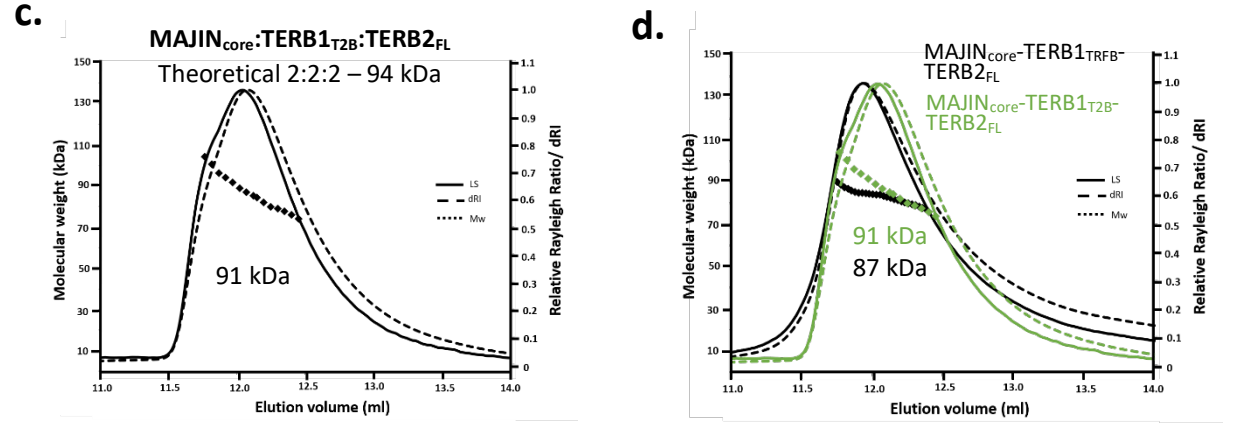


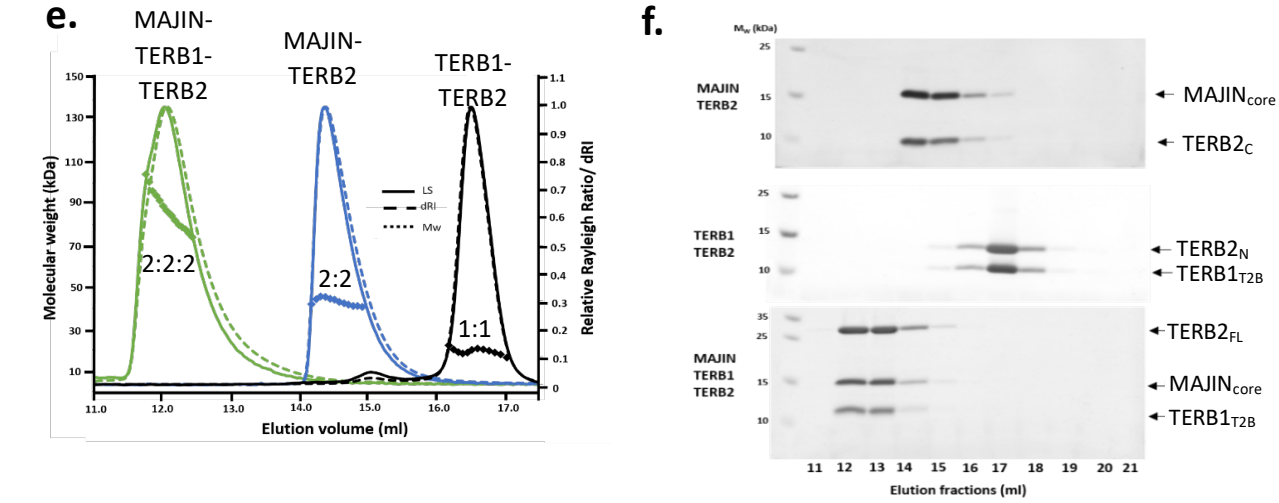
Figure 3.2.28.| SEC-SAXS data analysis of TRF1$_{TRFH}$-TERB1$_{TRFB}$-TERB2$_N$
a-c) Guinier analysis to determine the radius of gyration (*Rg*) of the components constituting the heterotetramer TRF1-TERB1-TERB2 complex. Empty circles represent the complete data set and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a dashed line. (*Q.Rg* values were < 1.3). d) Overlaid *P(r)* distribution of TRF1$_{TRFH}$ (black, solid), TRF1$_{TRFH}$-TERB1$_{TRFB}$ (grey), TERB2$_N$-TERB (grey, dashed) and TRF1$_{TRFH}$-TERB1$_{TRFB}$-TERB2$_N$ (black, dashed) showing maximum dimensions of 100, 105, 70 and 114 Å, respectively. e) SAXS scattering data of TRF1$_{TRFH}$, TERB1$_{TRFB}$-TERB2$_N$, TRF1$_{TRFH}$-TERB1$_{TRFB}$, and TRF1$_{TRFH}$-TERB1$_{TRFB}$-TERB2$_N$ (black) overlaid with the theoretical scattering curve of the modelled structure f) Multiphase SAXS *ab initio* model of TRF1$_{TRFH}$-TERB1$_{TRFB}$-TERB2$_N$ using MONSA which shows a globular model of TERB1-TERB2 (red and blue, respectively), in which TERB1 docks into the TRF1$_{TRFH}$ dimeric cleft. The TRF1$_{TRFH}$-TERB1$_{TBM}$ (PDB code: 5WIR, Pendlebury et al., 2017) is docked into the TRF1 molecular envelope (green). SAXS data analysed and figures made by Dr Owen Davies (Dunce, Milburn et al., 2018).

3.2.19. TERB1-TERB2 undergoes dimerisation when in complex with MAJIN

The MAJIN-TERB2 core structure ($MAJIN_{core}$-$TERB2_{C}$) has been extensively studied structurally and biophysically, further explained in 3.1.10. (Dunce, Milburn et al., 2018). SEC-MALS analyses of MAJIN-TERB2 and TERB1-TERB2 determined a 2:2 and 1:1 stoichiometry, respectively. We next sought to characterise the MAJIN-TERB1-TERB2 ternary meiotic complex by co-purifying the interacting regions of the bipartite complexes. His-$MAJIN_{core}$:MBP-$TERB2_{FL}$:His-$TERB1_{T2B}$ was co-expressed in *E.coli* BL21 DE3 cells and co-purified ultilising the same procedure as the MBP-$TERB2_{N}$:His-$TERB1_{T2B}$ complex, with the TEV-protease cleaved complex purified by anion exchange chromatography and subsequent size exclusion chromatography, summarised in figure 3.2.29.a. In addition, the ternary complex was purified using the full $TERB1_{TRFB}$ construct (Figure 3.2.29.b.). SEC-MALS analysis of both $MAJIN_{core}$-$TERB2_{FL}$-TERB1 ($TERB1_{TRFB}$ and $TERB1_{T2B}$) complexes revealed a single elution peak with molecular masses of 87 and 94 kDa, respectively, corresponding to a 2:2:2 stoichiometry (theoretical 2:2:2 Mw – 103 and 94 kDa) (Figure 3.2.29.c. and .d.). From the overlaid SEC-MALS profiles of $MAJIN_{core}$-$TERB2_{FL}$-$TERB1_{TRFB}$ and $MAJIN_{core}$-$TERB2_{FL}$-$TERB1_{T2B}$ we observe that the elution peaks of both complexes have asymmetry with corresponding sloping of the molecular weight fit and deviance between the light scattering and dRI in the second half of the peak. In combination, these observations suggest that there is possible dissociation between the three components. However, these elution characteristics have been observed in multiple MAJIN-TERB2 complexes including the crystallised MAJIN-TERB2 construct therefore may be an artefact (Dunce, Milburn et al., 2018). SEC-MALS analysis revealed that the TERB1-TERB2 (1:1) complex undergoes dimerisation in the wider MAJIN-TERB1-TERB2 complex to form a 2:2:2 heterohexamer (Figures 3.2.29.e. and f.). Far UV CD spectroscopy analysis determined that the $MAJIN_{core}$:$TERB2_{FL}$:$TERB1_{T2B}$ complex adopts a globular mixed alpha/beta structure (Figure 3.2.29.g.). Thermal denaturation revealed cooperative unfolding of $MAJIN_{core}$:$TERB2_{FL}$:$TERB1_{T2B}$, with an estimated Tm of 60 °C (Figure 3.2.29.h.).


a.
b.
MBP-TERB2FL
MBP
TERB2FL
His-MAJINcore
MAJINcore
His-TERB1T2B
TERB1T2B
His-tag
His-TERB1TRFB
TERB1TRFB
Pellet
Supernatant
Ni-NTA
Amylose
HiTrap Q
TEV Cleavage
HiTrap Q
Size exclusion Chromatography


c.
MAJINcore:TERB1T2B:TERB2FL
Theoretical 2:2:2 – 94 kDa
91 kDa
d.
MAJINcore-TERB1TRFB-TERB2FL
MAJINcore-TERB1T2B-TERB2FL
91 kDa
87 kDa
Molecular weight (kDa)
Relative Rayleigh Ratio/ dRI
Elution volume (ml)
LS
dRI
Mw


e.
MAJIN-TERB1-TERB2
MAJIN-TERB2
TERB1-TERB2
2:2:2
2:2
1:1
f.
MAJIN TERB2
TERB1 TERB2
MAJIN TERB1 TERB2
MAJINcore
TERB2C
TERB2N
TERB1T2B
TERB2FL
Elution fractions (ml)

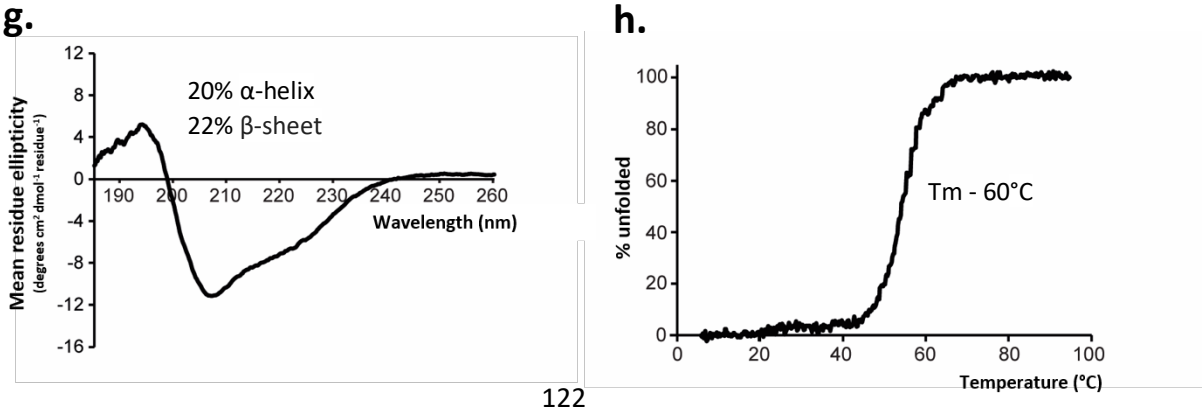

g.
20% α-helix
22% β-sheet
Mean residue ellipticity
Wavelength (nm)
h.
% unfolded
Tm - 60°C
Temperature (°C)

Figure 3.2.29.| Purification and biophysical analysis of MAJIN$_{core}$:TERB1:TERB2$_{FL}$ a-b) SDS-PAGE showing the purification summary of MAJIN$_{core}$:TERB1:TERB2$_{FL}$ a) TERB1$_{TRFB}$ and b) TERB1$_{T2B}$ through Ni-NTA, amylose and anion exchange chromatography. N-terminal solubility tags were removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. b) SEC-MALS analysis of cleaved MAJIN$_{core}$-TERB1$_{T2B}$-TERB2$_{FL,}$ the elution profile has a single peak with experimental molecular weight of 91 kDa. A theoretical 2:2:2 complex has an estimated molecular weight of 94 kDa. d) Overlaid SEC-MALS analysis of MAJIN$_{core}$:TERB1$_{TRFB}$:TERB2$_{FL}$ (black) and MAJIN$_{core}$:TERB1$_{T2B}$:TERB2$_{FL}$ (green), both having experimental molecular weights equating to a 2:2:2 complex. e) Overlaid SEC-MALS analyses of TERB1$_{T2B}$-TERB2$_{N}$, MAJIN$_{core}$-TERB2$_{C}$ and MAJIN$_{core}$:TERB1$_{T2B}$:TERB2$_{FL}$. TERB1$_{T2B}$-TERB2$_{N}$ forms a 1:1 complex (black) and MAJIN$_{core}$-TERB2$_{FL}$ forms a 2:2 complex (blue). TERB1$_{T2B}$-TERB2$_{N}$ undergoes dimerisation to form a 2:2:2 MAJIN$_{core}$:TERB1$_{T2B}$:TERB2$_{FL}$ ternary complex (green). f) SDS-PAGE analysis of the SEC-MALS output of 1ml fractions of MAJIN$_{core}$-TERB2$_{C}$ (top), TERB1$_{T2B}$-TERB2$_{N}$ (middle), and MAJIN$_{core}$:TERB1$_{T2B}$:TERB2$_{FL}$ (bottom). g) Far UV CD wavelength scan between 260-185nm of MAJIN$_{core}$:TERB1$_{T2B}$:TERB2$_{FL.}$ Deconvolution of the data estimates the secondary structure to consist of 20 % α-helices and 22 % β-sheets. h) CD thermal denaturation measured at 222 nm between 4 and 95°C, estimated a melting temperature of 60 °C.

3.2.20. Modelling the MAJIN:TERB1:TERB2 complex

The MAJIN$_{core}$-TERB2$_{C}$ and TERB1$_{T2B}$-TERB2$_{N}$ structures are connected by a poorly conserved stretch of 61 amino acid residues in the middle of the TERB2 sequence (Figure 3.2.1.b.). This observation suggests that there are three core globular domains within the 2:2:2 ternary complex separated by a TERB2 linker region. SAXS rigid body and linker modelling coupled with *ab initio* modelling to was utilised to determine a CORAL model of the relative positioning of the proteins within the multi-domain MAJIN:TERB1:TERB2 complex (Petoukhov et al., 2012). SEC-SAXS analysis and modelling was carried out by Dr Owen Davies (Dunce, Milburn et al., 2018). SEC-SAXS analysis of MAJIN$_{core}$:TERB2$_{FL}$:TERB1$_{T2B}$ revealed a *p(r)* profile typical of a globular complex with flexibility, shown by an elongated tail with a *Dmax* of 220 Å (Figure 3.2.30.c.). This result is indicative of the presence of the TERB2 linker region, suggesting that there is flexibility between the MAJIN-TERB2 and TERB1-TERB2 core regions which have an experimental *Dmax* of 120 and 65 Å, respectively when analysed as their bipartite complexes (Figures 3.2.30.c. and 3.2.30.d.). The *Rg* of MAJIN$_{core}$:TERB2$_{FL}$:TERB1$_{T2B}$ was determined to be 49 Å (Figure 3.2.30.a.). CORAL software was used to integrate multiple data sets: TERB1$_{T2B}$-TERB2$_{N}$ *ab initio* molecular envelope (Figure 3.2.19.c.),

the $MAJIN_{core}$-$TERB2_C$ crystal structure (Figure 3.1.30.a.), the SAXS scattering data of MAJIN-TERB2 and MAJIN-TERB2-TERB1 inclusive of the TERB2 flexible region.

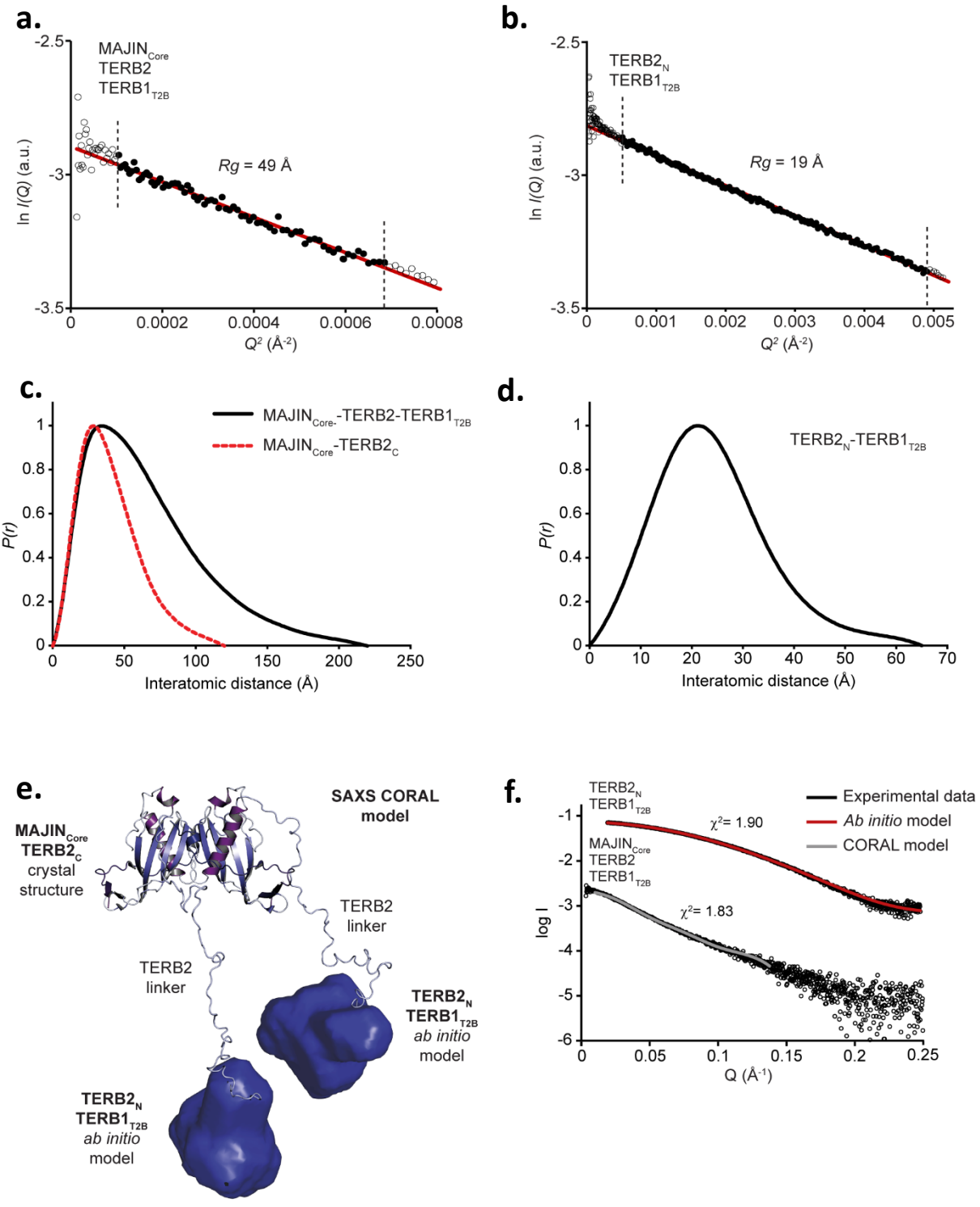


Figure 3.2.30.| SEC-SAXS data analysis of $MAJIN_{core}$:$TERB1_{T2B}$:$TERB2_{FL}$
a-b) Guinier analysis to determine the radius of gyration (*Rg*) of $MAJIN_{core}$:$TERB1_{T2B}$:$TERB2_{FL}$ and components constituting the hetero-hexameric complex. Clear circles represent the complete data set and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a dashed line. (*Q.Rg* values were < 1.3). c) Overlaid *P(r)* distribution of $MAJIN_{core}$:$TERB1_{T2B}$:$TERB2_{FL}$ (black, solid) and $MAJIN_{core}$-$TERB2_{C}$ (red, dashed) d) $TERB1_{T2B}$-$TERB2_{N}$ showing maximum dimensions of 220, 120 and 65 Å, respectively. e) SAXS CORAL model of $MAJIN_{core}$:$TERB1_{T2B}$:$TERB2_{FL}$ used rigid body modelling to combine the crystal structure of MAJIN-TERB2, TERB1-TERB2 DAMMIF *ab initio* model and modelling TERB2 flexible linkers. Determined that the MAJIN-TERB2 structural core and TERB1-TERB2 are two separate complexes linked by TERB2 flexible linkers. f) SAXS scattering data of $MAJIN_{core}$-$TERB1_{T2B}$-$TERB2_{FL}$ and $MAJIN_{core}$-$TERB2_{C}$ (black) overlaid with the theoretical scattering curve of the TERB1-TERB2 (*ab initio*, red) and $MAJIN_{core}$:$TERB1_{T2B}$:$TERB2_{1-220LL}$ (CORAL, grey) modelled structures. SAXS data analysed and figures made by Dr Owen Davies (Dunce, Milburn et al., 2018).

TERB2 linkers were modelled through fitting against the SAXS scattering curves. The scattering data are consistent with a central orientation of the MAJIN-TERB2 heterotetramer, with flexible linkers within the middle of TERB2 providing separation from two distinct TERB1-TERB2 1:1 complexes (Figure 3.2.30.e. and .f.).

3.2.21. The MAJIN:TERB1:TERB2 complex binds linear dsDNA

It was previously shown by EMSA that the $MAJIN_{1-106}$-$TERB2_{168-207}$ core complex has DNA-binding potential , with an extensive basic patch on the surface of each MAJIN protomer and a basic TERB2 C-terminal tail (Figure 3.1.13.c.) (Dunce, Milburn, et al., 2018). The addition of MAJIN linker basic patches ($MAJIN_{1-147}$-$TERB2_{168-207}$) enhances DNA binding potential with a determined apparent K_D of 0.12 μM, matching the DNA binding affinity of $TRF1_{FL}$ (0.10 μM). Furthermore, we have also shown that $TERB1_{TRFB}$ does not have DNA binding potential (Figure 3.2.8.g. and .h.).

The ability of MAJIN:TERB1:TERB2 to bind dsDNA was analysed using EMSA. Incubated samples were ran on an agarose gel, where shifts in the bands were observed which clearly indicated strong affinity for dsDNA (Figures 3.2.31.). These observations determined that MAJIN:TERB1:TERB2 ternary complex retains the DNA binding ability of $MAJIN_{core}$-$TERB2_{C}$, independent of TERB1 constructs ($TERB1_{TRFB}$ or $TERB1_{T2B}$. A modified EMSA protocol, performed by Dr James Dunce, was used to determine the binding affinity of $MAJIN_{core}$:$TERB1_{TRFB}$:$TERB2_{FL}$ to dsDNA (Dunce, Milburn

et al., 2018). An apparent K_D of 0.46 ± 0.04 μM to dsDNA was determined, indicating that the DNA-binding ability of the $MAJIN_{core}$-$TERB2_C$ structural core is retained (apparent K_D of 0.55 μM) (Dunce, Milburn et al., 2018).

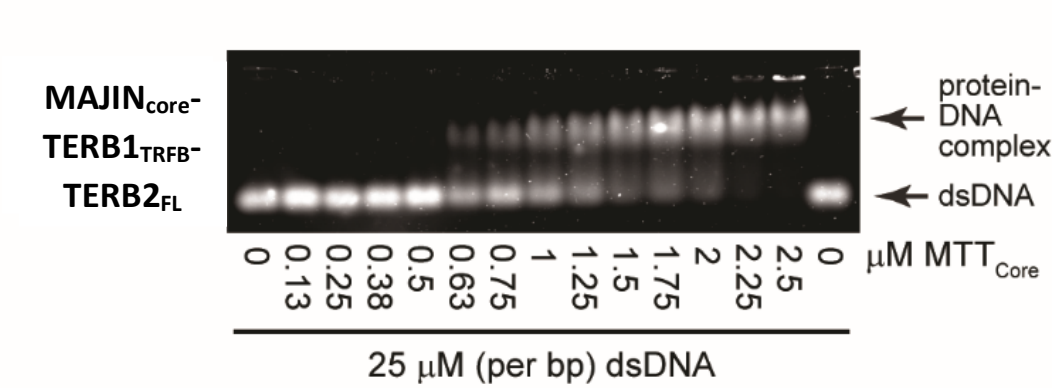


Figure 3.2.31.| MAJIN:TERB1:TERB2 binds linear dsDNA.
An EMSA was used to demonstrate the DNA binding ability of $MAJIN_{core}$:$TERB1_{TRFB}$:$TERB2_{FL}$ with dsDNA. EMSAs and quantification were carried out by Dr James Dunce.

3.2.22. MAJIN:TERB1:TERB2 can recruit TRF1 to form a meiotic ternary telomere complex

We have determined that within the meiotic telomere proteins and the shelterin complex protein, TRF1, there are two separate stable ternary complexes, TRF1:TERB1:TERB2 and MAJIN:TERB1:TERB2. From the literature, and our previous studies, we understand that telomeres are brought to the nuclear envelope by the shelterin complex protein TRF1, and its interaction with the meiotic telomere complex, MAJIN:TERB1:TERB2 which is attached to the INM. Therefore, we want to deduce how TRF1 may be recruited to the meiotic telomere complex.

$MAJIN_{core}$:$TERB1_{TRFB}$:$TERB2_{FL}$ was purified and incubated with a 2-fold molar excess of $TRF1_{TRFH}$. SEC-MALS analysis was performed to determine the absolute molecular weight of the mixed species. SEC-MALS determined an experimental molecular weight of 199 kDa, corresponding to a MAJIN:TERB2:TERB1:TRF1 complex with a stoichiometry of 2:2:2:4 (theoretical Mw – 206 kDa) indicating that we have reconstituted a full meitotic telomere complex (Figure 3.2.32.a.). The determined 2:2:2:4 stoichiometry suggests that the $MAJIN_{core}$:$TERB1_{TRFB}$:$TERB2_{FL}$ 2:2:2 complex

recruits 2 $TRF1_{TRFH}$ dimers consistent with the 2:1 $TRF1_{TRFH}$-$TERB1_{TRFB}$ stoichiometry. The SEC-MALS elution profile of the full meiotic complex shows deviance between the light scattering and dRI in the second half of the peak which suggests that there is dissociation of the reconstituted complex. SDS-PAGE analysis of the SEC-MALS elution profiles of $MAJIN_{core}$:$TERB2_{FL}$:$TERB1_{TRFB}$, $TRF1_{TRFH}$ and the reconstituted full complex, shows a shift in the elution volume of $MAJIN_{core}$:$TERB2_{FL}$:$TERB1_{TRFB}$ indicating the formation of a full meiotic complex (Figure 3.2.32.b.).

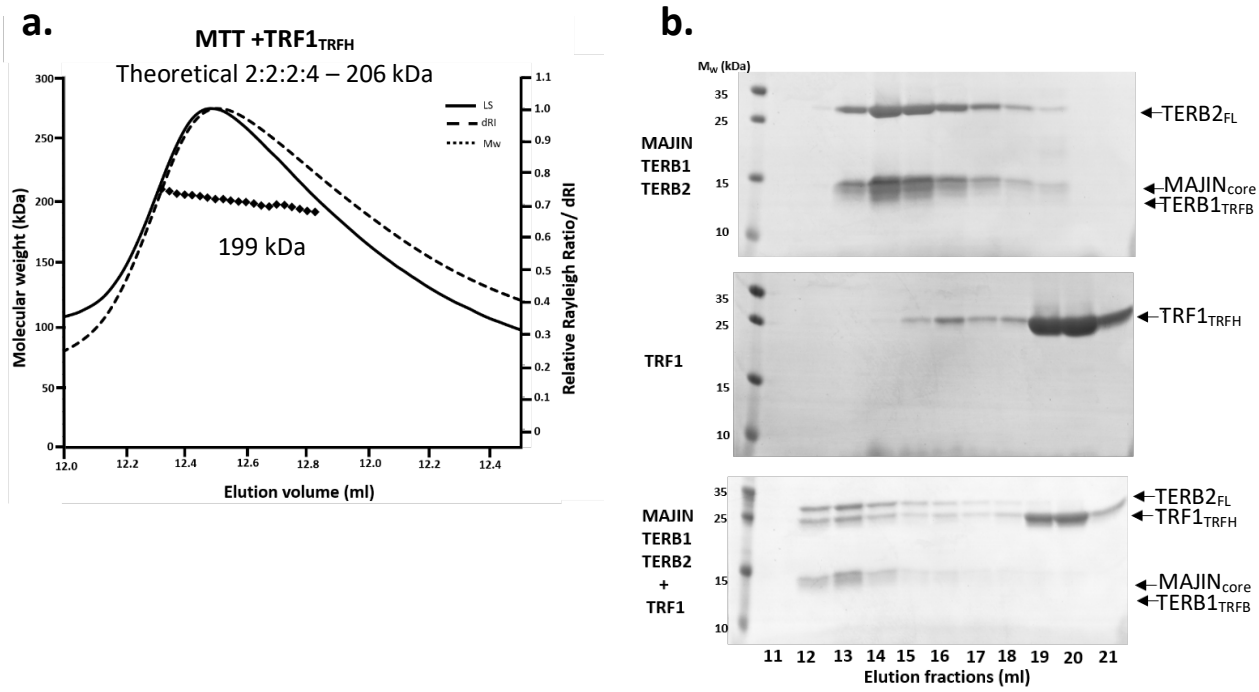


Figure 3.2.32.| Reconstitution of the full meiotic telomere recruitment complex
a) MAJIN:TERB1:TERB2 (MTT) was incubated with excess TRF1. SEC-MALS analysis of the meiotic ternary complex $MAJIN_{1-147}$:$TERB1_{TRFH}$:$TERB2_{FL}$:$TRF1_{TRFH.}$ demonstrating complex formation with molecular weight of 199 kDa constituting a 2:2:2:4 complex (theoretical molecular weight 206 kDa) b) SDS-PAGE analysis of SEC-MALS elution profiles of $MAJIN_{core}$:$TERB1_{TRFH}$:$TERB2_{FL}$ (top), $TRF1_{TRFH}$ (middle) and MAJIN:TERB1:TERB2 (MTT) was incubated with excess TRF1 (bottom).

An amylose pulldown experiment was ultilised to further explore the formation, and DNA binding ability, of the full meiotic telomere complex. Through EMSAs we have shown that MAJIN:TERB2:TERB1 complexes bind dsDNA. However, $TRF1_{TRFH}$ is unable to bind dsDNA. MBP-fusion $MAJIN_{1-233}$:$TERB2_{FL}$:$TERB1_{TRFB}$ (MBP-MTT_{1-233}) was incubated with $TRF1_{62-268}$ and applied to amylose resin that was preincubated with plasmid dsDNA. After incubation the amylose resin was washed with amylose buffer containing bovine serum albumin (BSA) to reduce non-specific protein binding to the amylose resin. Bound proteins were eluted from the resin and eluates were analysed by SDS-PAGE (Figure 3.2.33.a.). This demonstrated that MBP-MTT_{1-233} was able to pulldown $TRF1_{TRFH}$ even in the presence of dsDNA suggesting a strong interaction between MAJIN:TERB1:TERB1 and TRF1.

SEC determined the presence of a full meiotic complex, which we predict is formed by the recruitment of $TRF1_{TRFH}$ by $TERB1_{TRFB}$. However, it was postulated whether there is a secondary interaction within the full meiotic telomere complex. To test this hypothesis, we performed an amylose pulldown experiment with TERB2 constructs fused to an MBP-affinity tag incubated with His-$TRF1_{TRFH}$ in the presence or absence of $MAJIN_{core}$. SDS-PAGE analysis of the amylose eluates showed that there is a direct interaction between the C-termini of TERB2 and $TRF1_{TRFH}$ that is retained with the MAJIN-TERB2 complex (Figure 3.2.33.b.). These findings suggest that upon recruitment of TRF1 by TERB1, there is a possible conformational remodeling that allows for an interaction between TRF1 and TERB2 (This work was carried out by Dr James Dunce).

EMSAs were utilised to test the DNA binding ability of the MAJIN:TERB2:TERB1 complexes upon the addition of $TRF1_{TRFH}$. MAJIN:TERB2:TERB1 complexes were pre-incubated with random sequence dsDNA prior to the addition of $TRF1_{TRFH}$ in increasing molar amounts. Samples were then run on an agarose gel to determine the DNA-binding ability by the presence of band-shifts. As expected, we observe that $MAJIN_{core}$:$TERB2_{FL}$:$TERB1_{TRFB}$ ($MTT_{core/TRFB}$) binds dsDNA, forming a protein-DNA complex. The addition of $TRF1_{TRFH}$ starts to inhibit the DNA-binding by the MTT_{core} shown by a decrease in protein-DNA band and an increase in free-dsDNA (Figure 3.2.34.a.). This result is also

observed when using the longer MAJIN structure, which includes basic patches contained within the linker, $MAJIN_{1-147}$:$TERB2_{FL}$:$TERB1_{TRFB}$ ($MTT_{core+BP1/TRFB}$).

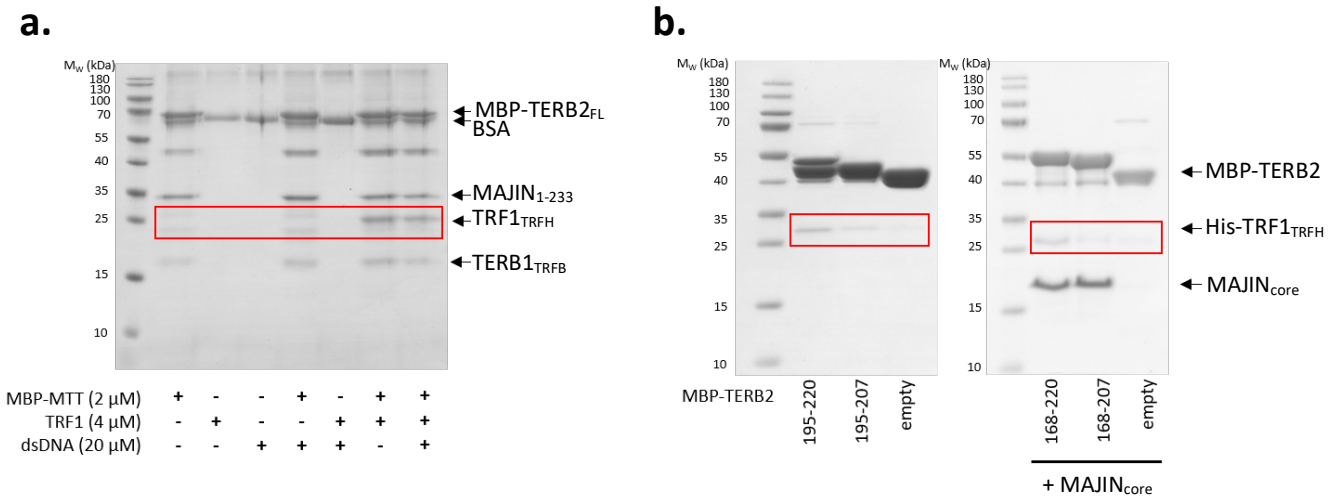


Figure 3.2.33.| Amylose pulldowns exploring the full meiotic ternary complex.
SDS-PAGE analysis of amylose pull-downs. a) Amylose pull-down of $TRF1_{62-268}$ using MBP-MTT_{1-233} with or without pre-incubation with plasmid dsDNA. 10 mg/ml BSA (molecular weight 66.4 kDa) was added to all reactions to prevent non-specific interactions. b) Recombinant co-expression amylose pull-down of His-$TRF1_{TRFH}$ with MBP-TERB2 constructs (left) with the addition of $MAJIN_{core}$ (right) as well as pLUC7. Free-MBP (empty) was used as a control to show that His-TRF1 does not interact with the MBP-tag. The amylose eluate of His-$TRF1_{TRFH}$-MBP-$TERB2_{195-220}$ shows TERB2 and a faint TRF1 band, suggesting that there is an interaction between the two proteins. This interaction is weakened as the TERB2 construct is shortened ($TERB2_{195-207}$). Presence of $TRF1_{TRFH}$ is highlighted by a red box. Recombinant amylose pulldowns were produced by Dr James Dunce.

Inclusion of the MAJIN basic patches partially compensated for the inhibition, shown by a larger amount of $TRF1_{TRFH}$ needed to inhibit DNA binding (Figure 3.2.34.b.). Furthermore, when the $MAJIN_{core}$:$TERB2_{FL}$:$TERB1_{T2B}$ ($MTT_{core/T2B}$) was used $TRF1_{TRFH}$ was unable to inhibit protein-DNA complex formation, indicating that DNA-binding by MAJIN:TERB1:TERB2 is inhibited by TRF1 in a manner dependent upon the presence of the TRF1 binding region of TERB1 ($TERB1_{TRFB/561-658}$) (Figure 3.2.33.c.). MAJIN:TERB1:TERB2 DNA-binding ability upon incubation of $TRF1_{TRFH}$ was quantified using ImageJ software to determine the % DNA bound. The % DNA bound was plotted against $TRF1_{TRFH}$:MTT molar ratio (Figure 3.2.34.d.). The slope of the $MTT_{Core/TRFB}$ plot is much steeper, indicative of a sharp decrease in $MTT_{Core/TRFB}$-DNA complex formation upon the addition of

$TRF1_{TRFH}$.We have determined that both $TRF1_{FL}$ and the ternary $MAJIN_{core}$:TERB2:$TERB1_{TRFB}$ complex are capable of binding dsDNA, with apparent Kd values of 0.10 μM and 0.46 μM, respectively. We used an EMSA to test the DNA binding ability of $TRF1_{FL}$ upon the addition of $MAJIN_{core}$:TERB2:$TERB1_{TRFB}$. $TRF1_{FL}$ was pre-incubated with random sequence linear dsDNA prior to the addition of $MAJIN_{core}$:TERB2:$TERB1_{TRFB}$ in increasing molar amounts. From the agarose gel we observe a super-shift from a $TRF1_{FL}$-DNA complex to the full meiotic telomere complex bound to DNA (Figure 3.2.34.e.). These results suggest that $MAJIN_{core}$:TERB2:$TERB1_{TRFB}$ can be recruited to DNA by full length TRF1 by direct DNA-binding through its C-terminal Myb domains.

3.2.23. TERB1 T648E phosphomimetic does not disrupt the $TRF1_{TRFH}$-$TERB1_{TRFB}$ complex

Once the telomeres have been recruited to the INM by TRF1, and the MTC (MAJIN-TERB1-TERB2) has been loaded onto the telomeres, the literature suggests that TRF1, through an unknown mechanism, is displaced. The meiotic telomere complex protein TERB1 contains a highly conserved 'TP' phosphorylation site within the TRF1 binding motif ($TERB1_{TBM}$), which can be phosphorylated by cyclin-dependent kinases (CDK) (Figure 3.1.10.a.). Previous studies have shown that TERB1 T648 residue undergoes phosphorylation *in vivo*, and this has been proposed to inhibit its interaction with TRF1 and suggested as a potential mechanism for TRF1 displacement. It has been shown *in vitro* that $TERB1_{TBM}$ containing a phosphomimetic mutation (T648E) completely disrupts the interaction with $TRF1_{TRFH}$ (Long et al., 2017; Pendlebury et al., 2017).

To explore the effect of the phosphomimetic mutation in the context of the larger TRF1 binding domain we introduced the T648E mutation into $TERB1_{TRFB}$ and cloned into the pMAT11 vector. MBP-$TERB1_{TRFB}$ T648E was co-expressed with His-$TRF1_{62-268}$ and purified by the same procedure as the wildtype complex. SDS-PAGE analysis revealed that both His-$TRF1_{TRFH}$ and MBP-$TERB1_{TRFB}$ T648E proteins are present in the amylose eluate indicating a direct interaction between the two proteins.

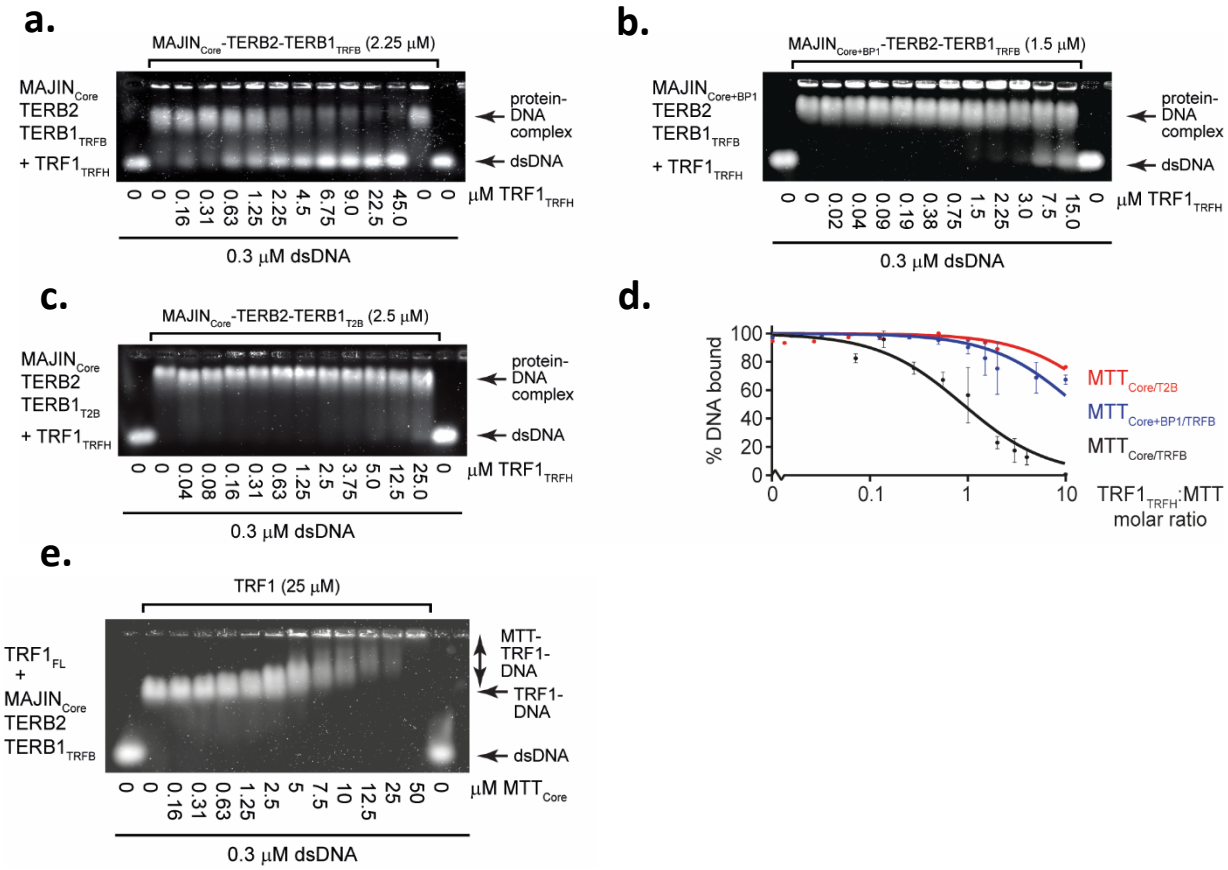


Figure 3.2.34.| EMSA supershift assays show DNA binding ability of MAJIN:TERB1:TERB2 upon incubation of TRF1.

a-d) Electrophoretic mobility shift assays (EMSA) were used to demonstrate the DNA binding ability of MAJIN:TERB1:TERB2 complexes upon the incubation of increasing amounts of TRF1$_{TRFH}$. EMSAs analysing the ability of a) MAJIN$_{Core}$:TERB2:TERB1$_{TRFB}$ (MTT$_{Core/TRFB}$), b) MAJIN$_{Core+BP1}$:TERB2:TERB1$_{TRFB}$ (MTT$_{Core/TRFB}$) and c) MAJIN$_{Core}$:TERB2:TERB1$_{T2B}$ (MTT$_{Core/TRFB}$) upon incubation with increasing amounts of TRF1$_{TRFH}$ to bind dsDNA. d) Quantification of the DNA-binding ability of MAJIN$_{Core}$:TERB2:TERB1$_{TRFB}$ (MTT$_{Core/TRFB}$, black), MAJIN$_{Core}$:TERB2:TERB1$_{T2B}$ (MTT$_{Core/T2B}$, red) and MAJIN$_{Core+BP1}$-TERB2–TERB1$_{TRFB}$ (MTT$_{Core+BP1/TRFB}$, blue) upon incubation with TRF1$_{TRFH}$ (at molar ratios indicated). e) EMSA analysing the ability of full length TRF1 (TRF1$_{FL}$) upon incubation with increasing amounts of MTT$_{Core}$ to bind dsDNA. EMSAs were replicated at least three times and performed using 25 nM (per molecule) of dsDNA. Error bars indicate standard error (n=3). Experiments were performed by Dr James Dunce and figures were made by Dr James Dunce and Dr Owen Davies.

His-TRF1$_{TRFH}$-MBP-TERB1$_{TRFB}$-T648E was further purified by anion exchange chromatography, in which both proteins perfectly co-purified across the salt concentration gradient, further suggesting stable complex formation (Figure 3.2.35.b.). Affinity tags were cleaved by TEV-protease and purified by anion exchange chromatography and subsequent size exclusion chromatography, summarised in figure 3.2.35.a.

To compare the oligomeric state of the phosphomimetic TRF1$_{TRFH}$-TERB1$_{TRFB}$ T648E to the wildtype complex SEC-MALS analysis was carried out. A single protein species was detected for both fusion and cleaved TRF1$_{TRFH}$-TERB1$_{TRFB}$ T648E complexes with a calculated molecular weight of 98 and 54 kDa, respectively (Figures 3.2.35.c. and d.). This corresponds to a 2:1 complex (theoretical 2:1 Mw – 112 and 59 kDa), as observed for the wildtype complex suggesting that the phosphomimetic mutation does not disrupt the complex formation and the 2:1 stoichiometry is maintained.

Furthermore, MAJIN$_{core}$:TERB1$_{TRFB}$:TERB2$_{FL}$ ternary complex containing the TERB1 T648E mutation was co-purified ultilising the same procedure as wildtype. SDS-PAGE analysis of the size exclusion chromatography (SEC) elution fractions determined that formation of the ternary complex is unaffected by the phosphomimetic mutation, as all three proteins co-elute (Figure 3.2.36.a).

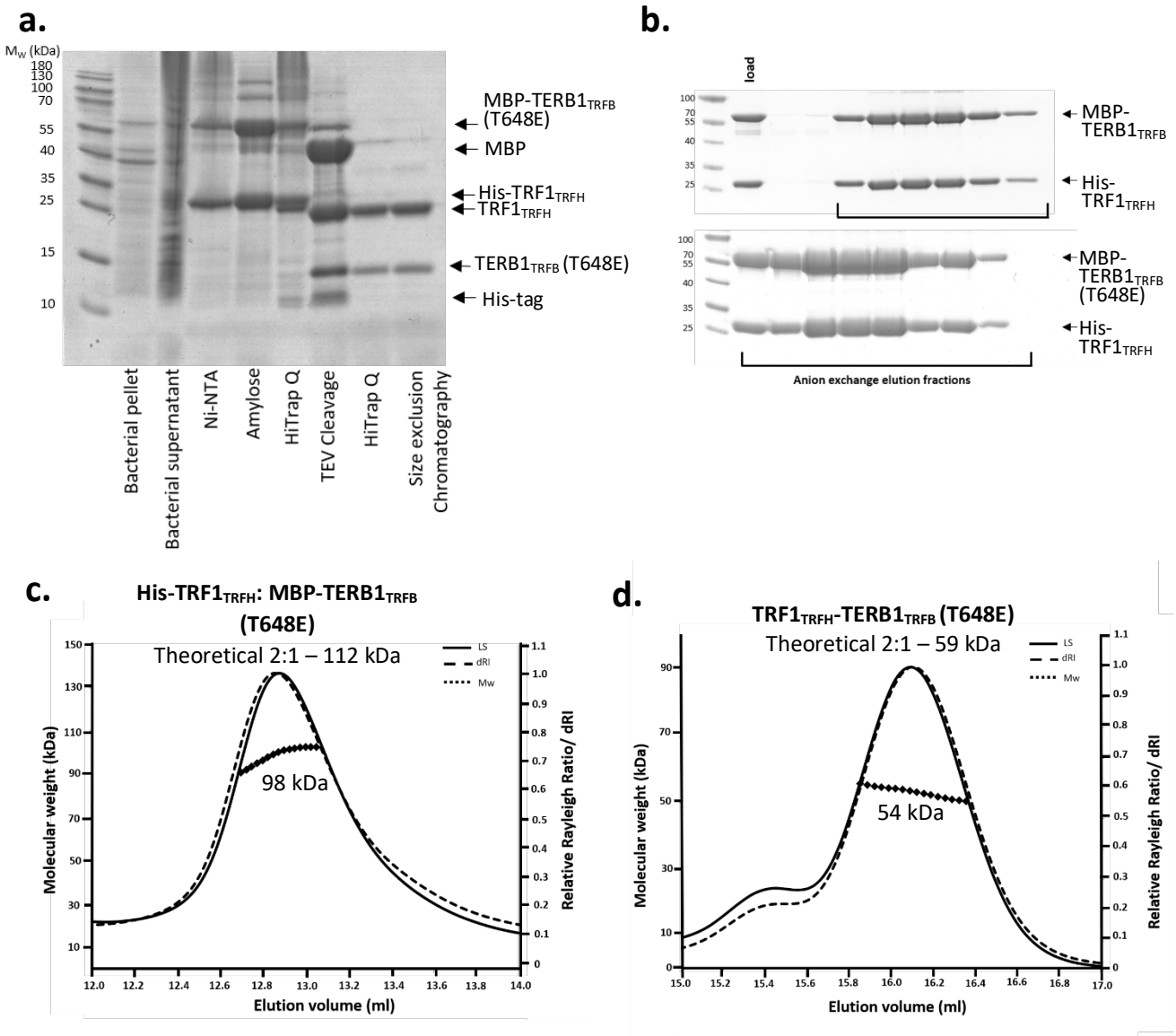


Figure 3.2.35.| Purification and SEC-MALS analysis of $TRF1_{TRFH}$-$TERB1_{TRFB}$ T648E phosphomimetic
a) SDS-PAGE showing the purification summary of $TRF1_{TRFH}$-$TERB1_{TRFB}$ T648E through Ni-NTA, amylose and anion exchange chromatography. N-terminal solubility tags were removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. b) SDS-PAGE showing anion exchange chromatography purification of fusion $TRF1_{TRFH}$-$TERB1_{TRFB}$ (top) and $TRF1_{TRFH}$-$TERB1_{TRFB}$ (T648E) (bottom). TRF1 and TERB1 in both constructs coelute, suggesting the T648E mutation has no effect on the purification steps. c-d) SEC-MALS analysis of c) fusion and d) cleaved $TRF1_{TRFH}$-$TERB1_{TRFB}$ (T648E), both elution profiles have a single peak with experimental molecular weights of 98 and 54 kDa, respectively. A theoretical 2:1 complex has an estimated molecular weight of 112 and 59 kDa, respectively.

We have previously shown that this ternary complex can recruit $TRF1_{TRFH}$ to form a stable full meiotic complex; we next tested the effect of the phosphomimetic T648E mutation to recruit $TRF1_{TRFH}$. $MAJIN_{1-147}$-$TERB1_{TRFB}$-$TERB2_{FL}$ T648E was incubated with a 2-molar excess of $TRF1_{TRFH}$ and analysed by SEC. SDS-PAGE analysis of the elution fractions indicate that full meiotic telomere complex formation is independent of the phosphomimetic mutation, further suggesting that the single phosphorylation event is insufficient to disrupt the $TRF1_{TRFH}$-$TERB1_{TRFB}$ interaction (Figure 3.2.36.b.).

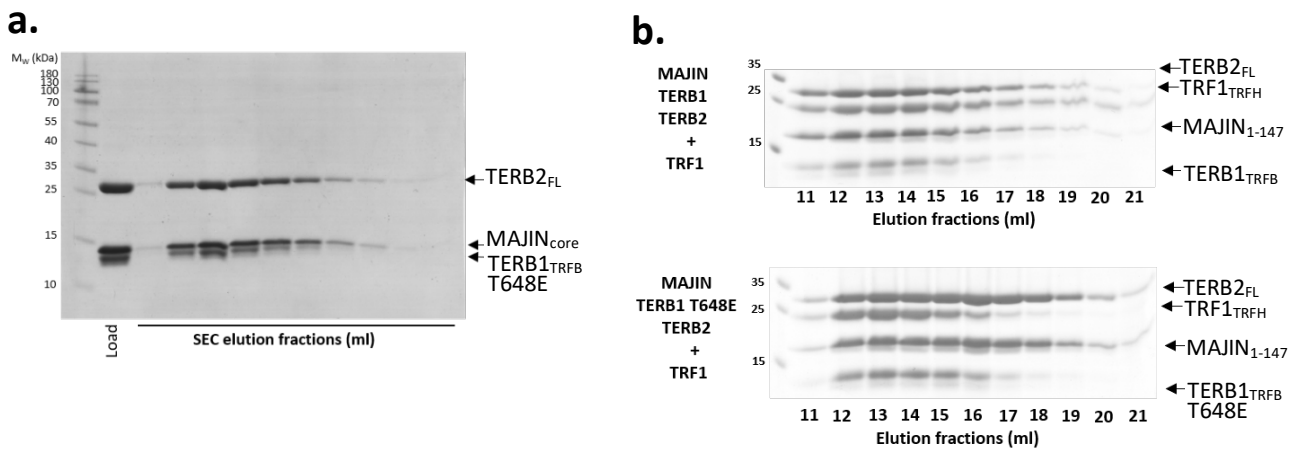


Figure 3.2.36.| $TERB1_{TRFB}$ phosphomimetic does not affect ternary complex formation.
a) SDS-PAGE analysis of the size exclusion chromatography elution fractions of $MAJIN_{core}$-$TERB1_{TRFB}$-$TERB2_{FL}$ T648E. The TERB1 T648E mutation does not affect the purification of the MAJIN-TERB1-TERB2 ternary complex. b) SDS-PAGE analysis of size exclusion chromatography elution profiles of $MAJIN_{1-147}$-$TERB1_{TRFB}$-$TERB2_{FL}$ wild type (top) and TERB1 T648E (bottom) upon incubation with a stoichiometric amount $TRF1_{TRFH}$. Ternary MAJIN-TERB1-TERB2 complexes can recruit $TRF1_{TRFH}$ thereby forming the full meiotic complex. SEC mixing experiments were performed by Mr Gurusaran Manickam.

3.2.24. TERB1$_{TBM}$ phosphomimetic blocks TRF1$_{TRFH}$ binding

A study has shown that the phosphomimetic mutation of T648E abrogated binding of TERB1$_{TBM}$ (amino acid residues 642-658) to TRF1 *in vitro* by GST (glutathione S-transferase)-pulldown assays (Pendlebury et al., 2017). We further tested this observation by performing an amylose pulldown assay. His-TRF1$_{TRFH}$ was co-expressed with MBP-fused TERB1 constructs, TERB1$_{TRFB}$ and TERB1$_{TBM}$ (wildtype and T648E). After application of clarified bacterial lysate to amylose resin and subsequent wash steps, bound protein was eluted from the resin and amylose eluates were analysed by SDS-PAGE (Figure 3.2.37.a). This analysis demonstrated that MBP-TERB1$_{TRFB}$, MBP-TERB$_{TBM}$ and MBP-TERB1$_{TRFB}$ (T648E) were able to pulldown His-TRF1$_{TRFH}$ but the MBP-TERB$_{TBM}$ (T648E) was unable to. This suggests that the phosphomimetic mutant disrupts binding when the shorter TERB1$_{TBM}$ construct was used.

We expressed TERB1$_{TBM}$ (T648E) as an MBP-fusion in isolation, which yielded soluble material. MBP-TERB1$_{TBM}$ T648E was purified by amylose affinity chromatography and subsequent anion exchange chromatography. Ultilising SEC-MALS, we tested for complex formation by mixing MBP-TERB1$_{TBM}$ T648 with 2-fold molar excess of His-TRF1$_{TRFH}$ (work performed by Dr James Dunce). The determined experimental molecular weight of 54 kDa suggested that there is no complex formation (theoretical 2:1 complex – 103 kDa) (Figure 3.2.37.b). This process was carried out with MBP-TERB1$_{TBM}$ wild type, which as expected, yielded two peaks with the latter peak overlaying with MBP-TERB1$_{TBM}$ in isolation. Overlaid His-TRF1$_{TRFH}$-MBP-TERB1$_{TBM}$ suggests that the peak corresponds to mixture of both His-TRF1$_{TRFH}$ dimer and monomeric MBP-TERB1$_{TBM}$ (theoretical molecular weights, 47 and 47.1 kDa, respectively) (Figure 3.2.37.c). We confirmed that the phosphomimetic mutation T648E disrupts TRF1 binding of TERB1$_{TBM}$, however the mutation is insufficient to disrupt TRF1-TERB1 binding when the longer TERB1$_{TRFB}$ (amino acid residues 561-658) is used.

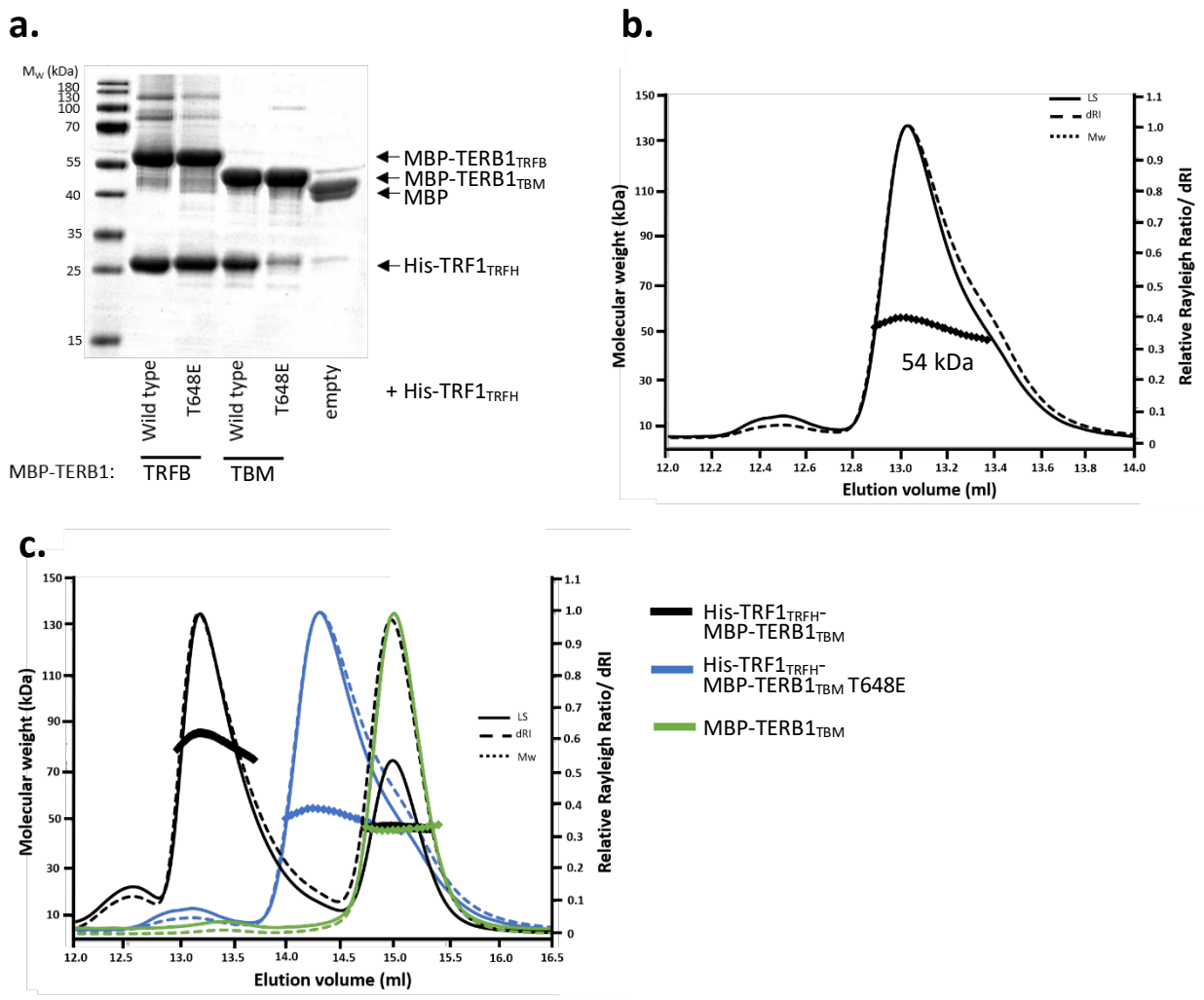


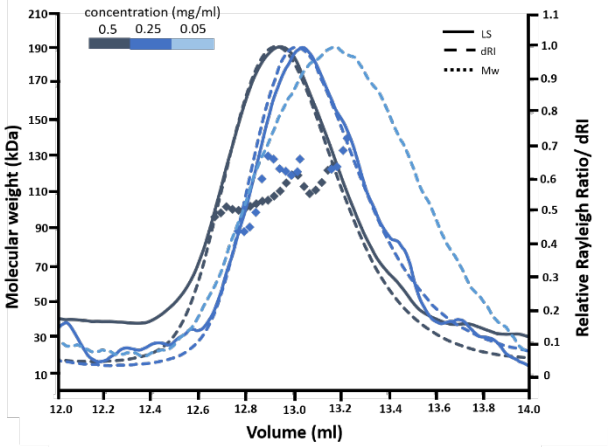
Figure 3.2.37.| TERB1$_{TBM}$ phosphomimetic blocks TRF1$_{TRFH}$ binding.
a) Amylose pulldown assay. His-TRF1$_{TRFH}$ was co-expressed with MBP-TERB1$_{TRFB}$ and MBP-TERB1$_{TBM}$ wild type and phosphomimetic T648E constructs. Lysates were incubated with amylose resin, washed and bound protein was eluted and analysed by SDS-PAGE. His-TRF1$_{TRFH}$ was found to interact with both wild type and T648 mutant TERB1$_{TRFB}$, as well as wild type TERB1$_{TBM}$. His-TRF1 was unable to interact with MBP-TERB1$_{TBM}$ T648E, suggests that the phosphomimetic mutant disrupts binding when the shorter TERB1$_{TBM}$ construct is used. b) MBP-TERB1$_{TBM}$ T648 was mixed with 2-fold molar excess of TRF1$_{TRFH}$, incubated for 1 hour then analysed by SEC-MALS. The deduced experimental molecular weight is 54 kDa, suggesting that there is no complex formation and that the peak could be TRF1$_{TRFH}$ dimer or MBP-TERB1$_{TBM}$ T648 monomer (theoretical molecular weights, 47 kDa and 47.1 kDa, respectively). c) Overlaid SEC-MALS of MBP-TERB1$_{TBM}$ wild type and T648E mutant, following incubation with 2-fold molar excess of TRF1$_{TRFH}$. As mentioned previously there is only partial 2:1 complex formation of wild type (black), shown by two peaks, the latter overlaid the MBP-TERB1$_{TRFH}$ elution profile (green). MBP-TERB1$_{TBM}$ T648 was cloned and purified by Dr James Dunce. Dr James Dunce also carried out the SEC-MALS mixing experiments and amylose pull down.

3.2.25. TERB1$_{TRFB}$ T648E has a 10-fold reduction in binding affinity to TRF1$_{TRFH}$ compared to wild type

In order to determine if the phosphomimetic mutation reduced the binding between TRF1$_{TRFH}$ and TERB1 complexes we estimated relative binding affinities by ultilising SEC-MALS. A dilution series of His-TRF1$_{TRFH}$:MBP-TERB1 complexes (TERB1$_{TRFB}$ (T648E), TERB1$_{TBM}$ and TERB1$_{TRFB}$) were applied to SEC-MALS and the dissociation was analysed. The concentration at which the protein complex is loaded onto the SEC column undergoes a ~5-fold dilution before being analysed by the MALS device. SEC-MALS analysis determined that wild type His-TRF1$_{TRFH}$:MBP-TERB1$_{TRFB}$ elutes as a single peak when loaded at 0.25 mg/ml (analysed at ~0.05 mg/ml) with a molecular mass of 102 kDa, indicating 2:1 stoichiometry is retained (theoretical 2:1 – 112 kDa) (Figure 3.2.38.a). In line with our previous observations, His-TRF1$_{TRFH}$-MBP-TERB1$_{TBM}$ dissociates by size exclusion chromatography, indicated by two peaks suggesting partial 2:1 complex formation despite being loaded at high concentration of 27.5 mg/ml (analysed at ~5 mg/ml). Further dilutions show a progressive reduction in molecular weight, and complete dissociation at 1 mg/ml (Figure 3.2.38.b.). The phosphomimetic His-TRF1$_{TRFH}$:MBP-TERB1$_{TRFB\ (}$T648E) analysed at 1 mg/ml (loaded at 5 mg/ml) shows clear complex formation, with a molecular mass of 98 kDa (theoretical 2:1 – 112 kDa). At lower protein concentrations the molecular weight progressively reduced and there is some indication of complex dissociation observed at 0.5mg/ml (Figure 3.2.38.c.). SEC-MALS revealed that TRF1$_{TRFH}$ binding to TERB1$_{TBM}$ is ~100-fold weaker than TERB1$_{TRFB}$. We observed that the phosphomimetic mutation, TERB1$_{TRFB}$ T648, does weaken binding of TRF1$_{TRFH}$. However, the affinity is at least 10-fold stronger than wild type TERB1$_{TBM}$.

a.

His-TRF1$_{TRFH}$-MBP-TERB1$_{TRFB}$

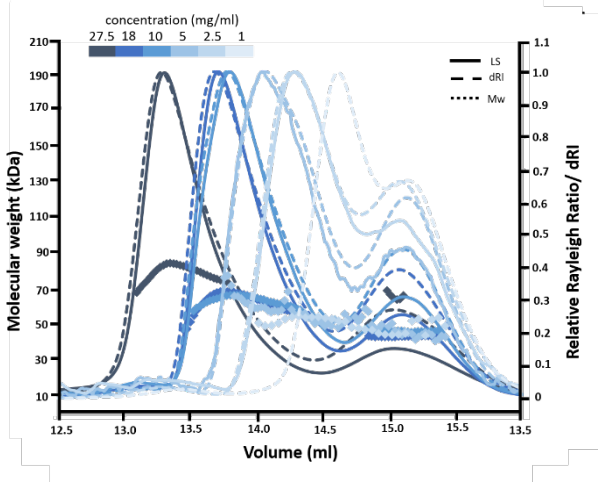


Theoretical 2:1 complex – 112 kDa

Load (mg/ml)	0.5	0.25	0.05
SEC-MALS Mw (kDa)	106	102	n.d.

b.

His-TRF1$_{TRFH}$-MBP-TERB1$_{TBM}$

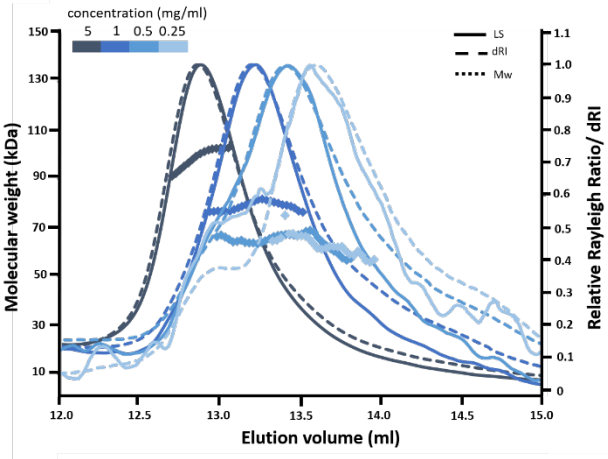


Theoretical 2:1 complex – 103 kDa

Load (mg/ml)	27.5	18	10	5	2.5	1
SEC-MALS Mw (kDa)	85	68	66.5	55	52	50

c.

His-TRF1$_{TRFH}$-MBP-TERB1$_{TRFB}$ T648E



Theoretical 2:1 complex – 112 kDa

Load (mg/ml)	5	1	0.5	0.25
SEC-MALS Mw (kDa)	98	81	67	62

Figure 3.2.38.| Dilution series of fusion TRF1-TERB1 complexes.
Dilution series of His-TRF1$_{TRFH}$ in complex with MBP-TERB1$_{TRFB}$, MBP-TERB1$_{TBM}$ and MBP-TERB1$_{TRFB}$ T648E to deduce approximate binding affinities. The protein that is loaded onto the SEC column undergoes an approximate 5-fold dilution, therefore the sample analysed by MALS is ~5-fold the concentration that was loaded. The MALS experimental molecular weight was determined for a series of dilutions. a) His-TRF1$_{TRFH}$-MBP-TERB1$_{TRFB}$ forms a clear 2:1 complex (theoretical 2:1 - 112 kDa) when loaded at both 0.5 and 0.25 mg/ml. At lower protein concentrations the molecular weight was unable to be interpreted. The complex is largely retained at very low concentrations suggesting a very stable protein interaction. b) His-TRF1$_{TRFH}$-MBP-TERB1$_{TBM}$ undergoes partial complex formation even when loaded at 27.5 mg/ml (~100-fold more than His-TRF1$_{TRFH}$-MBP-TERB1$_{TRFB}$), shown by two peaks. Further dilutions show a progressive reduction in molecular weight, and complete dissociation at 1 mg/ml. MALS elution traces suggest a 100-fold reduction in affinity relative to wildtype His-TRF1$_{TRFH}$-MBP-TERB1$_{TRFB}$. c) His-TRF1$_{TRFH}$-MBP-TERB1$_{TRFB}$ T648E loaded at 5 mg/ml a clear 2:1 complex (theoretical 2:1 - 112 kDa). At lower protein concentrations the molecular weight progressively reduced. There is some indication of complex dissociation observed at 0.5mg/ml, suggesting an approximate 10-fold reduction in affinity in respect to the wild type complex. Dilution series MALS runs were carried out by myself and Dr James Dunce.

3.3. DISCUSSION

3.3.1. Meiotic telomere attachment is achieved by two distinct ternary complexes

TERB1 is a multi-domain protein known to bind both TERB2 and the shelterin complex protein TRF1 through its central TRF1-binding (TRFB) domain. Through multiple TERB1 truncations we identified that TERB1 binds to TERB2 through a central sequence of the TRFB domain (T2B), flanked by TRF1 binding sites. $TERB1_{T2B}$ interacts with the N-termini of TERB2 to form a 1:1 complex, which can recruit a TRF1 dimer to form a ternary TRF1-TERB1-TERB2 complex with 2:1:1 stoichiometry. In solution characterisation of TRF1-TERB1-TERB2 revealed that the TERB1-TERB2 globular structure docks into the TRF1 dimerisation interface via a direct interaction between TRF1 and TERB1, schematised in figure 3.3.1.b. In line with the SAXS multi-phase *ab initio* model that TERB2 is distant from the V-shaped TRF1 cleft, we determined that there is no direct interaction between N-termini of TERB2 and TRF1 within the ternary TRF1-TERB1-TERB2 complex.

Previous studies determined the minimum TERB1 construct to bind TRF1 is a short 15 amino acid peptide ($TERB1_{TBM}$) and crystal structures of $TRF1_{TRFH}$ in complex with $TERB1_{TBM}$ revealed a 2:2 stoichiometry, in which two $TERB1_{TBM}$ peptides bind to the $TRF1_{TRFH}$ homodimer (Long et al., 2017; Pendlebury et al., 2017). However, our findings suggest that the wider TERB1 binding region ($TERB1_{TRFB}$) is required to form a stable complex and TRF1 binding by $TERB1_{TRFB}$ likely blocks the second site through steric hindrance to provide a 2:1 TRF1-TERB1 stoichiometry. Moreover, we observe that $TERB1_{TBM}$ binds to TRF1 with 100-fold reduction in affinity compared to $TERB1_{TRFB.}$ Previous studies reported K_D values for $TRF1_{TRFH}$ binding to free $TERB1_{TBM}$ and GST-$TERB1_{TBM}$ to be 5.6 μM and 75 nM, respectively (Long et al., 2017; Pendlebury et al., 2017). However, we believe that the higher apparent affinity of $TERB1_{TBM}$ to $TRF1_{TRFH}$ is likely due to the addition of the GST-tag. Native GST exists as a homodimer, which could induce two GST-tagged TERB1 peptides at both sites within a $TRF1_{TRFH}$ dimer (Maru et al., 1996; Wilce et al., 1994). We propose that $TERB1_{TRFB}$ may utilise a two-step co-operative binding mode, in which the N-terminal sequence interacts with a second binding site of the $TRF1_{TRFH}$ dimer in a manner that mimics the $TERB1_{TBM}$ binding. To gain a high

resolution molecular structure of the 2:1 TRF1-TERB1 complex we designed a TERB1 bonsai construct that lacks the TERB2 binding site but maintained stable complex formation with TRF1, and retained 2:1 stoichiometry. Crystals grew quickly and diffraction of crystals determined only TRF1 is present in the crystal structure. This outcome suggests that there could be a mixed pool of TRF1-TERB1 and free TRF1, with TRF1 alone crystallising much faster and inhibiting TRF1-TERB1 crystallisation.

Despite our efforts, we were unable to crystallise the 1:1 $TERB1_{T2B}$-$TERB2_{N}$ complex. However, recently the X-ray crystal structure of $TERB1_{589-649}$ -$TERB2_{4-110}$ was solved using single-wavelength anomalous dispersion (SAD) method at a resolution of 2.9 Å (PDB: 6J07; Wang et al., 2019). The $TERB1_{589-649}$ -$TERB2_{4-110}$ structure adopts a compact globular α-helical domain, resembling a single folding unit. The structure shows that TERB2 consists of a three β-sheets flanked by six α-helices. TERB1 is a helix-turn-helix (HtH) motif and a long tail that folds back onto the HtH motif, forming a bundle structure. TERB1 HtH motif packs against the concaved surface of TERB2, secured by a network of hydrophobic and electrostatic interactions (Figure 3.3.1.a,). Using Pymol it was possible to determine that $TERB1_{589-649}$ -$TERB2_{4-110}$ structure has a maximum dimension of 55.8 Å, in line with the in solution SAXS *Dmax* of 55 Å (Figure 3.2.19.). These results suggest that the SAXS *ab initio* model is an accurate representation of the TERB1-TERB2 structure. To verify that the SAXS molecular envelope is the same conformation as the crystal structure, the $TERB1_{589-649}$ -$TERB2_{4-110}$ crystal structure was docked into the $TERB1_{T2B}$-$TERB2_{N}$ DAMMIF envelope using SUPCOMB (Figure 3.3.1.c.) (Kozin & Svergun, 2001). The crystal structure partially fits the SAXS molecular envelope, suggesting the modelled in solution structure is a good representation of the TERB1-TERB2 structure. The key TERB1 residues responsible for the TERB1-TERB2 interaction (R605, R629, K632 and R636) are not involved in TRF1 binding, further suggesting that TRF1 and TERB2 have independent TERB1 binding sites.

Importantly, the crystal structure revealed that TERB1 coordinates a Zn^{2+} ion through two N-terminal cysteine residues, C592 and C595, and a further cysteine and histidine residue within the HtH motif, C618 and H621. Zinc coordination plays an important role in defining the overall structure of TERB1 (Figure 3.3.1.b.). From these results we can carry out mutational studies of amino acid residues

responsible for the TERB1-TERB2 interaction, as well as the TRF1-TERB1-TERB2 and MAJIN-TERB1-TERB2 tripartite complexes to gain a wider understanding for their assembly.

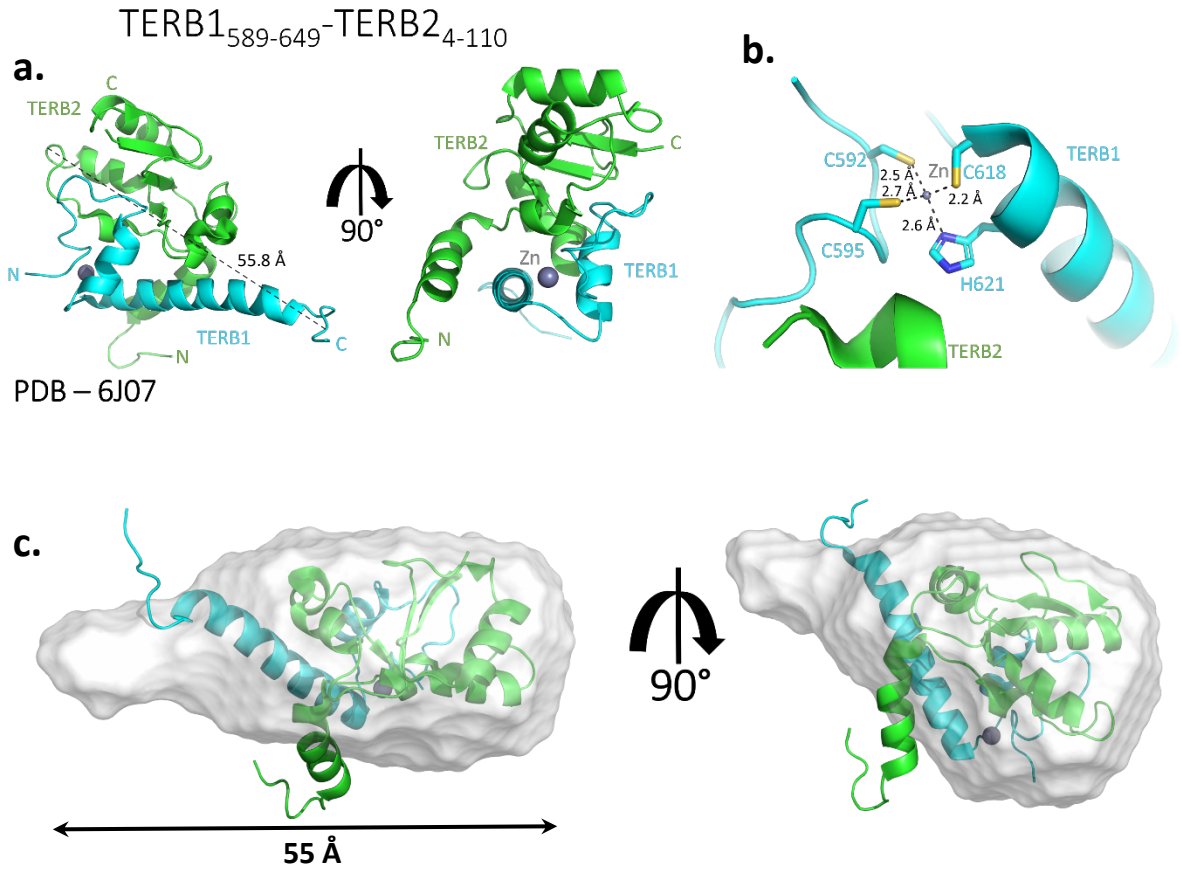


Figure 3.3.1.| Crystal structure of the TERB1-TERB2 complex
a) X-ray crystal structure of TERB1$_{589-649}$ -TERB2$_{4-110}$ complex shows TERB1 and TERB2 interaction to form a globular α-helical domain with a determined maximum dimension of 55.8 Å. The N-terminal sequence of TERB1 coordinates a single zinc ion, as depicted in b. b) TERB1 residues C592, C595, C618 and H621 coordinate a zinc ion. PDB code: 6J07. c) The TERB1$_{589-649}$ -TERB2$_{4-110}$ crystal structure was docking into the TERB1$_{585-642}$-TERB2$_{1-107LL}$ SAXS envelope, which was generated by DAMMIF *ab initio* modelling (Figure 3.2.19.) and displayed in two orientations.

The crystal structure of MAJIN-TERB2 revealed a hetero-tetrameric 2:2 complex with 2-fold symmetry. Mutational disruption of the MAJIN-TERB2 dimer interface demonstrated that MAJIN-TERB1 forms a stable 1:1 complex, suggesting that dimerisation is not a requirement for stability and likely constitutes the architectural core for the wider meiotic telomere complex (Dunce, Milburn, et al.,

2018). In the wider ternary complex in accordance to MAJIN-TERB2 (2:2), the TERB1-TERB2 (1:1) complex undergoes dimerisation to form a MAJIN-TERB1-TERB2 (2:2:2) heterohexameric complex. We utilised CORAL modelling in which determined that the MAJIN-TERB2 and TERB1-TERB2 globular domains are physically separated by TERB2 linkers, schematised in figure 3.3.2.a.

Telomeric ends of chromosomes from mouse spermatocytes were imaged using simulated illumination microscopy (SIM) as a complementary technique to our *in vitro* findings. SIM experiments were carried out by Miss Irene da Cruz and Professor Ricardo Benavente, University of Würzburg, Germany, in a collaborative project. In zygotene spreads, $MAJIN_{core}$ foci are flanked by TERB2 at the telomere ends of SYCP3 stained axial elements, indicating telomere recruitment by the meiotic telomere complex (Figure 3.3.2.c.). In late pachytene spreads we observe paired telomeres with thickened lateral element ends, in which MAJIN foci of paired telomeres overlap and TERB2 foci encircles MAJIN. In contrast, $TERB1_{T2B}$ and TERB2 foci colocalise in pachytene spreads. Staining patterns show similarities with our biophysical findings, in that the MAJIN-TERB2 and TERB1-TERB2 globular domains are spatially separated by TERB2 flexible linkers (Figure 3.3.2.d.).

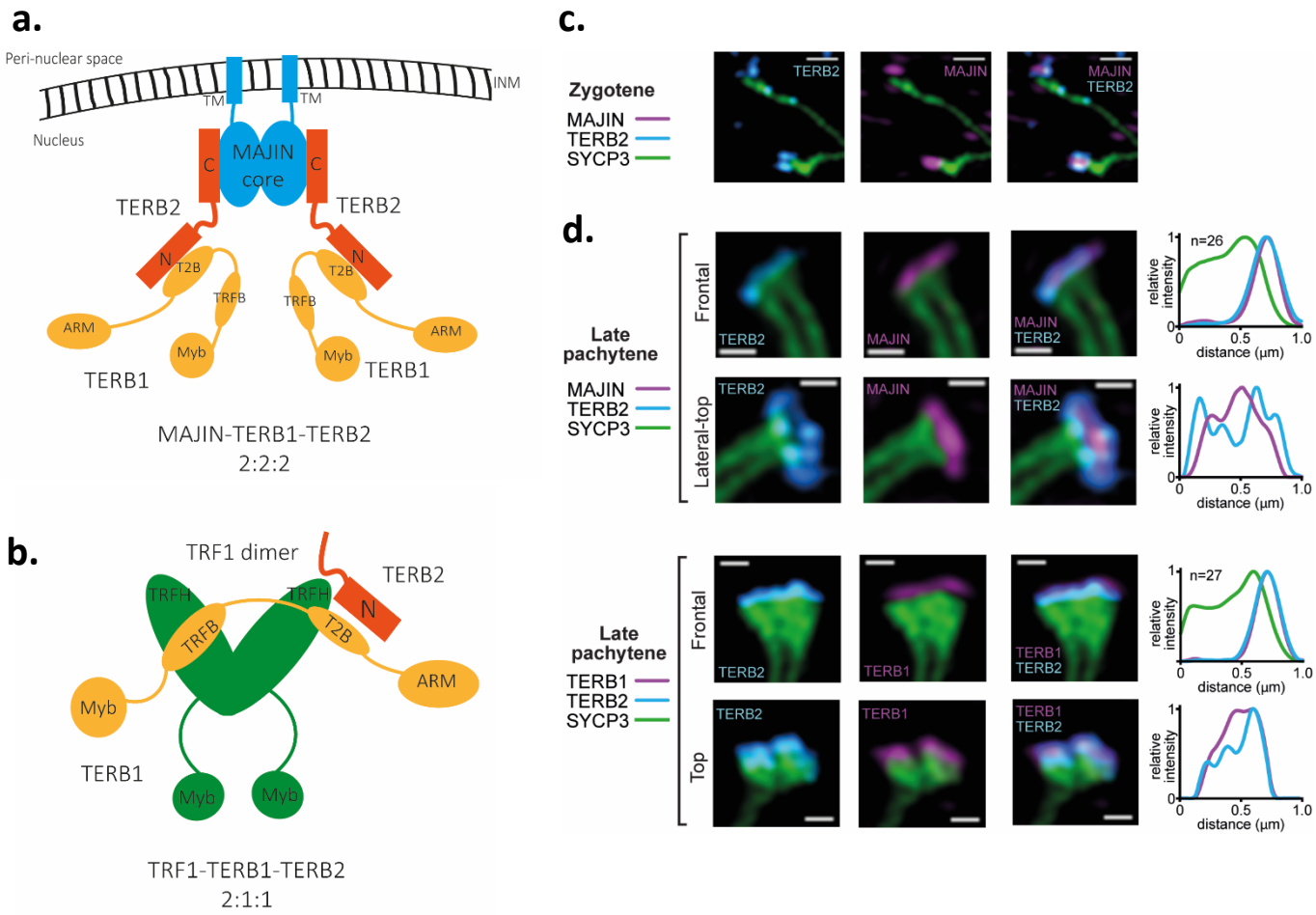


Figure 3.3.2.| Schematic models of the MAJIN-TERB1-TERB2 and TRF1-TERB1-TERB2 ternary complexes

a-b) Schematic models of the ternary complexes a) MAJIN-TERB1-TERB2 and b) TRF1-TERB1-TERB2. a) The 2:2:2 MAJIN-TERB1-TERB2 is tethered to the NE by the C-terminal transmembrane domains of MAJIN. A flexible linker region links to the core domain of MAJIN which in turn dimerises with the C-termini of TERB2. Each N-termini of TERB2 binds to a TERB1 molecule. b) The 2:2:1 TRF1-TERB1-TERB2 complex is comprised of a V-shaped $TRF1_{TRFH}$ dimer, in which the TERB1-TERB2 1:1 complex directly binds to though the $TERB1_{TRFB}$ domain. c) Structural illumination microscopy (SIM) of spread mouse zygotene spermatocyte chromosomes stained with anti-SYCP3 (green), anti-TERB2 (cyan) and anti-MAJIN (magenta). d) Spread mouse late pachytene spermatocyte chromosomes stained with anti-SYCP3 (green), anti-TERB2 (cyan) and anti-MAJIN (bottom) or anti-TERB1 (top) (magenta). Corresponding normalised intensity-distance plots are shown. Scale bars, 0.3 µm. Work was carried out by Miss Irene da Cruz as a collaborative project. Figures adapted from Dunce, Milburn et al., 2018.

3.3.2. The meiotic telomere complex undergoes rearrangement upon TRF1 attachment

We reconstituted a full meiotic telomere complex through the addition of TRF1 to the MTC (MAJIN-TERB1-TERB2), and analysis by SEC-MALS revealed a stoichiometry of 2:2:2:4. This suggests that the heterohexameric MAJIN-TERB1 TERB2 complex recruits two TRF1 dimers, consistent with our observed 2:1 TRF1-TERB1 stoichiometry (Figure 3.3.3.). Determination of the full meiotic telomere complex stoichiometry revealed similarities to the multi-protein shelterin complex (Figure 3.1.1.). Reconstitution of the core shelterin subcomplex *in vitro*, TRF2-TIN2-TPP1-POT1, determined a stoichiometry of 2:1:1:1 (Lim et al., 2017). Within the complex it is shown that TRF2 interacts with TIN2 with a 2:1 stoichiometry, not the 2:2 complex observed in the $TRF2_{TRFH}$-$TIN2_{TBM}$ crystal structure. Similar to the meiotic telomere complex, these findings suggest that within the wider complex assembly of the shelterin complex there are steric hinderances limiting binding interactions.

During zygotene, chromosomes are attached to the nuclear envelope through the myb-domains of TRF1, and possibly TERB1. Combining our findings, we propose that the full meiotic telomere complex represents a zygotene pre-displacement complex, in which the recruitment of TRF1 enables the telomeric DNA to be in close proximity to the meiotic telomere complex therefore the meiotic telomere complex indirectly interacts with telomere ends through TRF1, schematised in Figures 3.3.3. and 3.3.4.

3.3.3. The DNA-binding dynamics of the meiotic telomere complex enables telomere attachment

DNA-binding by TRF1 is achieved through its C-terminal Myb domains, and although the homodimerisation domain is required for DNA binding it does not itself bind telomeric dsDNA in isolation. Previously we have shown that the MAJIN-TERB2 core strongly binds to dsDNA and now determined that the addition of TERB1 does not affect DNA-binding of the meiotic telomere complex. Our findings suggest that direct telomeric binding by MAJIN-TERB1-TERB2 is partially inhibited by the presence of TRF1 in a TERB1 dependent manner. This suggests that upon recruitment of TRF1 to the nuclear envelope, by TERB1, there is competitive inhibition of DNA binding. Furthermore, it has

been shown that inclusion of MAJIN basic flexible regions enhances DNA-binding with an approximate affinity that matches TRF1, therefore partially compensates for DNA-binding inhibition by TERB1.

Upon TRF1 binding to the meiotic telomere complex, structural rearrangement is required to facilitate TRF1-TERB2 interaction, which also may alter the DNA-binding ability of the pre-displacement complex. EMSAs determined that the addition of $MAJIN_{Core}$–TERB2–$TERB1_{TRFB}$ to $TRF1_{FL}$:DNA results in a super-shift of MAJIN-TERB2-TERB2:DNA complexes. Our findings suggest that upon telomere attachment and TRF1 displacement, MAJIN-TERB1-TERB2 is then able to directly bind the telomeric DNA, forming the post displacement complex (Figure 3.3.4.b.). The telomeric DNA loops around the top of MAJIN-TERB2 molecules facilitating, and stabilising, telomere attachment at the nuclear envelope.

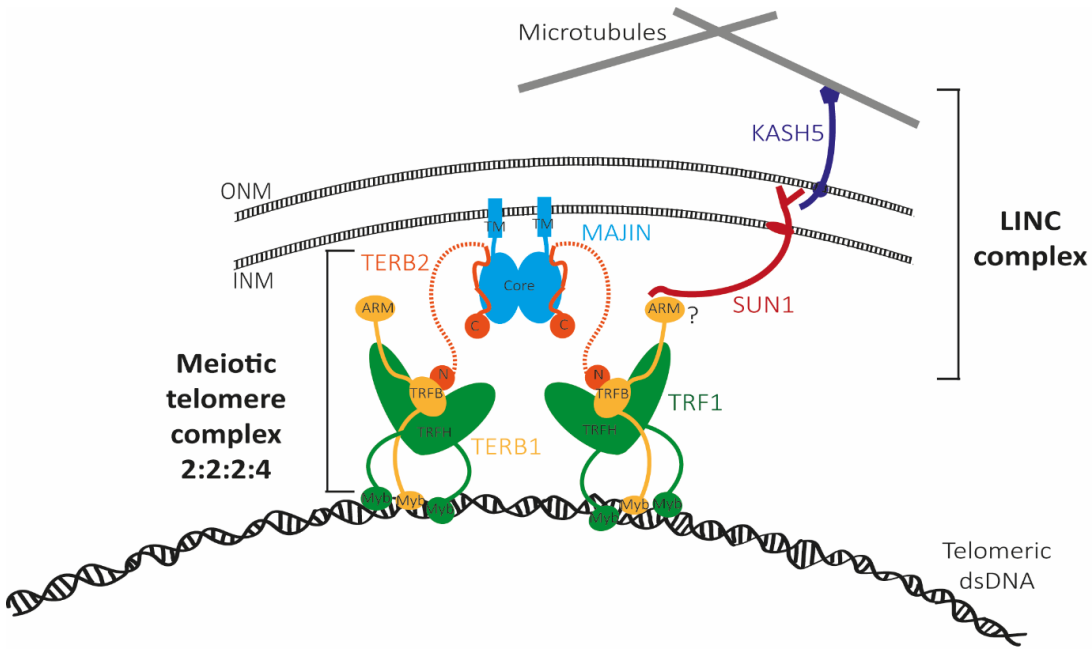


Figure 3.3.3.| Schematic model of the full meiotic telomere complex.
MAJIN-TERB2 2:2 dimeric core provides the architectural framework for meiotic telomere complex formation. Each TERB2 N-terminus can interact with of a TERB1 TRFB domain forming a 2:2:2 ternary complex. MAJIN C-terminal transmembrane domains anchor the meiotic telomere complex to the inner nuclear membrane. Each TERB1 molecule can recruit a TRF1 dimer bound to telomeres, resulting in the formation of a MAJIN-TERB2-TERB1-TRF1 2:2:2:4 full meiotic complex attached to telomeric DNA. The LINC complex, comprised of KASH5 and SUN1, spans the nuclear envelope connecting the microtubules to the meiotic telomere complex. It has been proposed that the N-terminus of SUN1 interacts with the armadillo-repeat domain of TERB1 directly linking the LINC and meiotic telomere complex. Figure adapted from Dunce, Milburn et al., 2018.

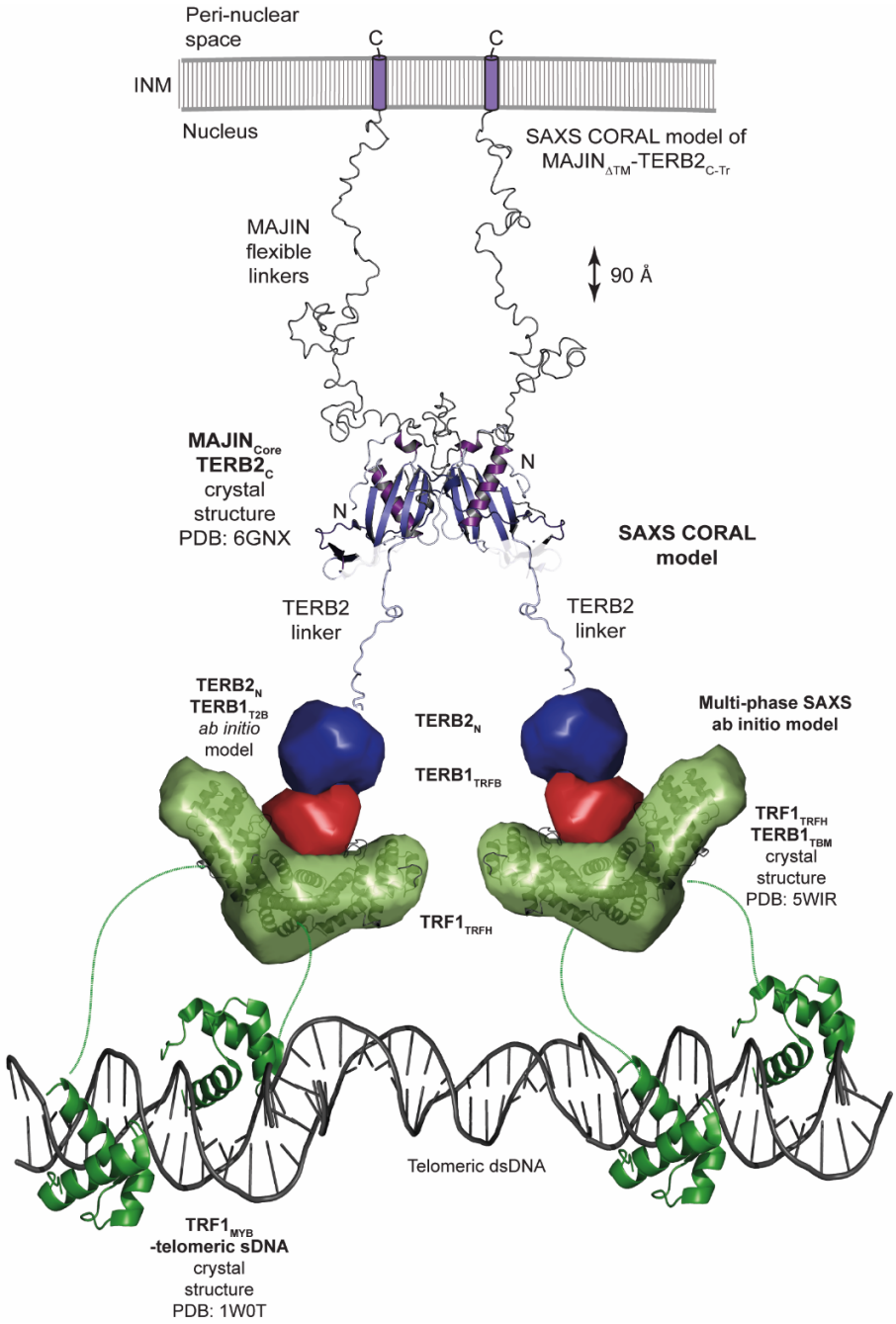


Figure 3.3.4.| Structural model of the meiotic pre-displacement complex.
In solution SAX multi-phase a*b initio* and CORAL modelling techniques and X-ray crystal structures have allowed for the determination of the full meiotic telomere complex structural model. MAJIN$_{Core}$-TERB2$_C$ crystal structure (PDB code: 6GNX) revealed a symmetrical 2:2 hetero-tetramer. SAXS CORAL model of the wider SAXS CORAL model of the wider MAJIN$_{\Delta TM}$-TERB2$_{C\text{-}Tr}$ complex, combined with the solved crystal structure determined MAJIN flexible linkers of ~90 Å connect the structural core and the TM region embedded in the NE. SAXS CORAL modelling of MAJIN$_{Core}$-TERB2-TERB1$_{T2B}$ ultilised rigid body modelling determined that the MAJIN-TERB2 structural core and TERB1-TERB2 are two separate complexes linked by TERB2 flexible linkers. Multi-phase SAXS *ab initio* of TRF1$_{TRFH}$-TERB1$_{TRFB}$-TERB2$_C$ determined a globular model of TERB1-TERB2 in which TERB1 docks into the TRF1$_{TRFH}$ dimeric cleft. The TRF1$_{TRFH}$-TERB1$_{TBM}$ (PDB code: 5WIR) is docked into the TRF1 molecular envelope. Dimerised TRF1 C-terminal Myb-domains directly binds telomeric dsDNA, thereby delivering the chromosome ends to the nuclear envelope for attachment.

3.3.4. Spatial displacement of TRF1 upon telomere attachment

Zygotene telomeres become attached to the nuclear envelope; this is achieved through the recruitment of TRF1 in complex with telomeric DNA by the meiotic telomere complex to form the full meiotic complex. This facilitates chromosome movements required for homologue pairing and recombination. In late pachytene, MAJIN-TERB1-TERB2 displaces TRF1 from the end of telomeres allowing for direct binding to telomere ends, and the maturation of the attachment plate (Shibuya et al., 2015; Zhang et al., 2017). Studies have previously coined this process as telomere cap exchange in which the shelterin complex, including TRF1, is released from telomeres and exchanged for the meiotic telomere complex (Pendlebury et al., 2017). Furthermore, another study postulated that upon TRF1 release from telomeres it remains in close proximity to enable rebinding to the telomere (Shibuya et al., 2015).

We set out to explore the off-loading mechanism of TRF1 from the meiotic telomere complex following telomere attachment. Previous studies demonstrated that phosphorylation of the conserved TERB1 T648 CDK phosphorylation site abrogates TRF1 binding to $TERB1_{TBM}$ (Pendlebury et al., 2017). By creating a phosphomimetic mutation and CDK-CyclinB1 treatment we have demonstrated that the phosphorylation of $TERB1_{T648}$ weakens the interaction between TRF1 and TERB1 but is inadequate for complete TRF1 displacement. We determined that the phosphomimetic mutation weakens binding affinity by ~10-fold compared to wild type, and the T648E mutant did not affect the ability of MAJIN-TERB2-TERB1 to recruit TRF1 *in vitro*. It is known that CDK activity increases gradually throughout prophase I of meiosis and CDK1-cyclinB localises near telomeres towards the latter stages of prophase (Moiseeva et al., 2017). Our results suggest that T648 phosphorylation must be in combination with other phosphorylation and/or binding events within TERB1 and/or surrounding proteins are required to achieve dissociation of TRF1 from the meiotic telomere complex. This is in keeping with the failure of a T648A mutation to prevent TRF1 displacement *in vivo* (Shibuya et al., 2015).

We have determined that the dynamic interaction between TERB1 and TRF1 is crucial for meiotic progression. TERB1 can recruit TRF1 in complex with telomeres to the nuclear membrane during zygotene, indicating the need for a robust interaction but can adapt to displace TRF1 in late pachytene.

SIM experiments were used to probe the TRF1 mediated attachment and displacement to the meiotic telomere complex. In zygotene spreads TRF1 and TERB1 foci are partially overlapping at the telomeric ends, indicating TRF1 attachment (Figure 3.3.5.c.). In contrast, in late pachytene spreads we observe a distinct staining pattern for TERB1 foci at the telomere ends that are surrounded by TRF1 (Figure 3.3.5.d, top). These findings are consistent with our model, in that there is subtle remodelling within the meiotic telomere complex which allows for direct binding to telomeric DNA and that TRF1 is only displaced from the extreme telomere ends of chromosomes therefore telomeric binding of the shelterin complex is retained. Furthermore, the length of telomeric DNA (10-15 kb in humans) is sufficient to account for this model and is consistent with our SIM observation that TRF1 remains in close proximity to the chromosome axis (Nandakumar & Cech, 2013). This analogy suggests that the meiotic telomere complex and shelterin complex cooperate to achieve telomere attachment and protection of the 3' end.

Furthermore, SIM reveals that in late pachytene, the localisation of telomeric DNA (TelFISH) is distal from TERB2 and the SYCP3-stained lateral elements, suggesting that telomeric DNA is looped out from the attachment plate (Figure 3.3.5.d, bottom). This localisation is in keeping with our model, which suggests that upon TRF1 displacement telomeric DNA is looped around the MAJIN-TERB2 core, as well as MAJIN linkers to enhance DNA binding and form a robust complex.

3.3.5. Elucidating the meiotic telomere attachment mechanism

Despite our detailed studies, we still have many unanswered questions regarding the complete mechanism for TRF1 displacement during the zygotene to pachytene transition. Our results indicate that the meiotic telomere complex is a highly co-ordinated yet dynamic complex, in which TERB2 and MAJIN linkers provide flexibility. We have shown within the $TRF1_{TRFH}$-$TERB1_{TRFB}$ complex, the TERB1 phosphorylation site is only partially phosphorylated by CDK1-CyclinB, suggesting that TERB1 phosphorylation in isolation is unlikely to sufficient to displace TRF1 from the meiotic telomere

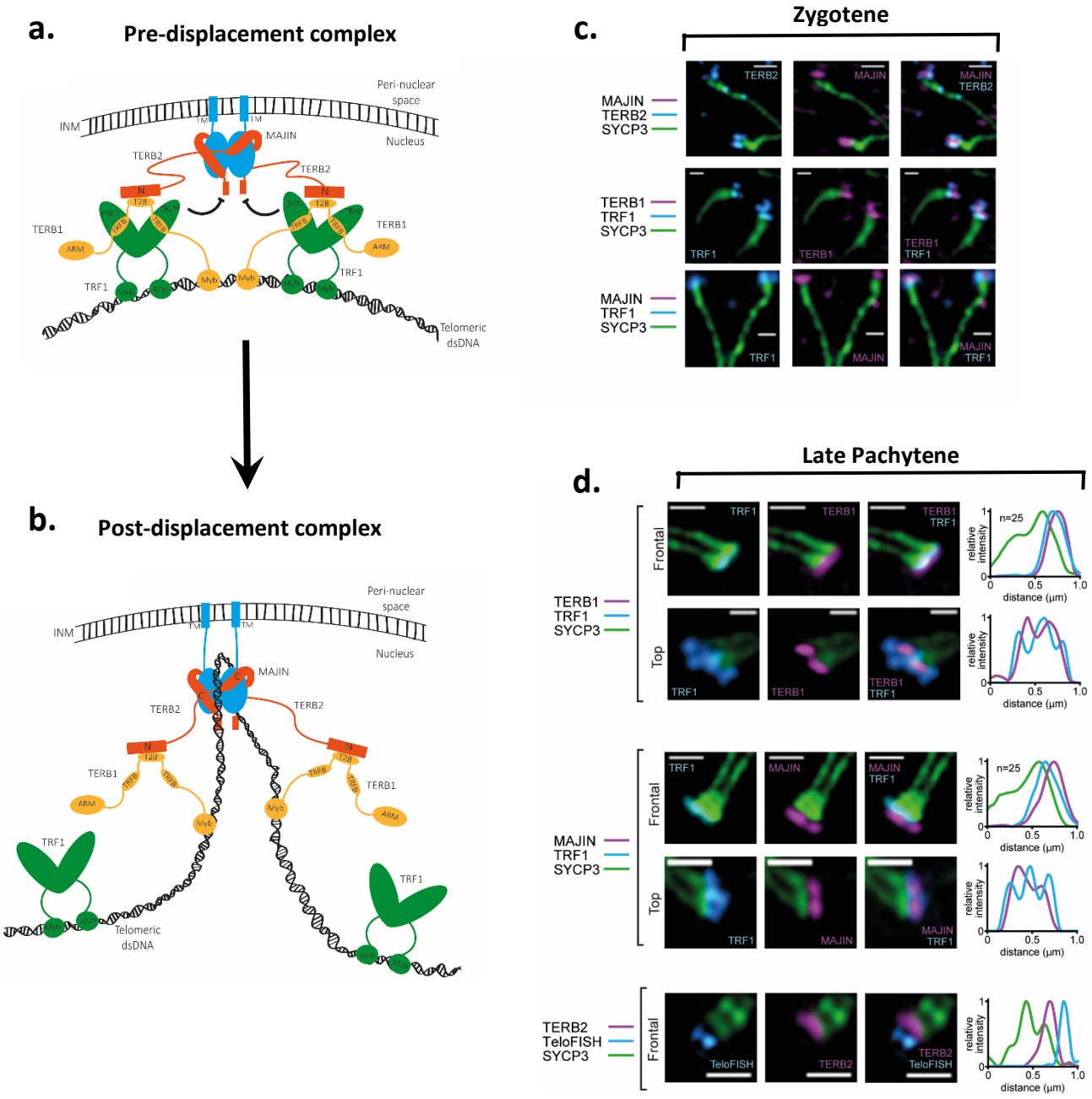


Figure 3.3.5.| Schematic models of the pre- and post-displacement complexes.
a-b) Schematic models of the pre- and post- displacement complexes. a) The full MAJIN-TERB1-TERB2-TRF1 complex represents the pre-displacement attachment complex. TRF1 bound telomeric dsDNA is recruited to the nuclear envelope by TERB1, creating an indirect interaction between the meiotic telomere complex and telomere ends. b) Upon telomere attachment, TRF1 is displaced from the meiotic telomere complex, but remains bound to the telomeric DNA. The release of TRF1 allows for meiotic telomere complex to directly the telomeric DNA. c) Structural illumination microscopy (SIM) of spread mouse zygotene spermatocyte chromosomes stained with anti-SYCP3 (green), anti-TERB2 (top) or anti-TRF1 (middle/bottom) (cyan) and anti-MAJIN (top/bottom) or anti-TERB1 (middle) (magenta). d) Spread mouse late pachytene spermatocyte chromosomes stained with anti-SYCP3 (green), anti-TRF1(top/middle) or anti-TeloFISH (bottom) (cyan) and anti-TERB1 (top) or anti-MAJIN (middle) or anti-TERB2 (bottom) (magenta). Corresponding normalised intensity-distance plots are shown. Scale bars, 0.3 μm. Work was carried out by Miss Irene da Cruz as a collaborative project. Figure adapted from Dunce, Milburn et al., 2018
.

attachment complex. In addition, we propose that the meiotic telomere complex undergoes structural rearrangements throughout prophase I, to enable TRF1 displacement.

In addition to the meiotic telomere complex and the LINC complex, other proteins have been identified to be involved in the process of telomere attachment and TRF1 displacement, including the meiosis-specific cohesin STAG3, cyclin dependent kinase 2 (CDK2) and its meiosis-specific non-canonical activator, SpeedyA (Shibuya & Watanabe, 2014; Wang et al., 2018). During prophase I of meiosis SpeedyA is localised at telomeres and in turn recruits and activates CDK2, implying that CDK2 is present at the nuclear envelope during telomere attachment (Mikolcevic et al., 2016). Studies have shown that the attachment of telomeres to the nuclear envelope is impaired in SpeedyA$^{-/-}$ knockout mice spermatocytes leading to meiotic arrest, indicating SpeedyA is essential for CDK2 recruitment to attachment sites and telomere tethering to the NE (Mikolcevic et al., 2016; Tu et al., 2017). These findings suggest that changes in telomeric protein-protein interaction affinities must be tightly regulated during prophase I, potentially by CDK2-dependent phosphorylation (Link & Jantsch, 2019).

GST-pulldown experiments revealed a direct interaction between TRF1 and SpeedyA, SpeedyA and CDK2, but no interaction was detected between CDK2 and TRF1 (Figure 3.3.6.a-c.) (L. Wang et al., 2018). These results suggest that SpeedyA functions as molecular scaffold for CDK2 recruitment to telomeres and is involved in a phosphorylation dependent mechanism for the attachment and displacement of TRF1 at the NE (McGrath et al., 2017; Tu et al., 2017; Wang et al., 2018).

A further question that remains unanswered is how the meiotic telomere complex integrates with the LINC complex. The transmembrane domain of MAJIN provides the INM connection, whilst TERB1 is proposed to interaction with SUN1 through its N-terminal armadillo repeat domain (Shibuya et al., 2014).

Despite our attempts in characterising the SUN1-TERB1 interaction, we were unable to replicate the reported Y2H interaction biochemically (Shibuya et al., 2014). More recently, immunoprecipitation assays have deduced a SUN1-TERB1, shown by a faint TERB1 band (Figure 3.3.6.d.) (Wang et al., 2020). Together, these results suggest that the SUN1-TERB1 interaction is very weak or that they

interact indirectly through a wider ternary complex or post-translation modifications are required to facilitate the interaction.

In addition to TERB1 T648, the N-terminus of SUN1 also contains a CDK phosphorylation site, implying that CDK2 may function in the inter-connection between the meiotic telomere complex and the LINC complex (Viera et al., 2015). Immunoprecipitation (co-IP) assays, Y2H screens and the crystal structure of SUN1-SpeedyA-CDK2, reveal that SpeedyA interacts with both SUN1 and TRF1 (Figure 3.3.6.f.) (Yanyan Chen et al., 2021; G. Wang et al., 2020). Despite the function of SUN1 phosphorylation is still unknown, this finding suggests that the SUN1-SpeedyA interaction provides the key linkage between the LINC, meiotic telomere complex and telomeres at the NE, thus stabilises the MTC-tethered telomeres.

Surprisingly, co-IP assays determined an interaction between SUN1 and MAJIN (Figure 3.3.6.e), which led to the proposal of a second interaction between the meiotic telomere complex and SUN1 (Wang et al., 2020). co-IP assays carried out in the presence of CDK2 inhibitors revealed that the MAJIN-SUN1 interaction is weakened. In addition, localisation studies show the colocalisation of MAJIN and SUN1 at the NE but no NE localisation of TERB1 when coexpressed with SUN1 (Wang et al., 2020). Together with our results, suggests that the possible interaction between SUN1 and TERB1 is not sufficient to localise TERB1 to the NE and SpeedyA/CDK2 may promote telomere attachment by regulating the SUN1-MAJIN interaction.

3.3.6. Role of the meiotic telomere complex post telomere attachment

We have determined that the meiotic telomere complex proteins, MAJIN, TERB1 and TERB2 play a crucial role throughout prophase I, to achieve telomere attachment in zygotene through the recruitment of TRF1. Using SIM experiments it has been shown that upon TRF1 displacement from the MTC bound telomeres TRF1 remains in close proximity to the chromosome axis, suggesting an additional role of TRF1 in meiosis (Dunce, Milburn et al., 2018). A study suggest that TRF1 and SpeedyA-CDK2 also

play a role in protecting telomeres from fusions during chromosome search and bouquet formation (Wang et al., 2018).

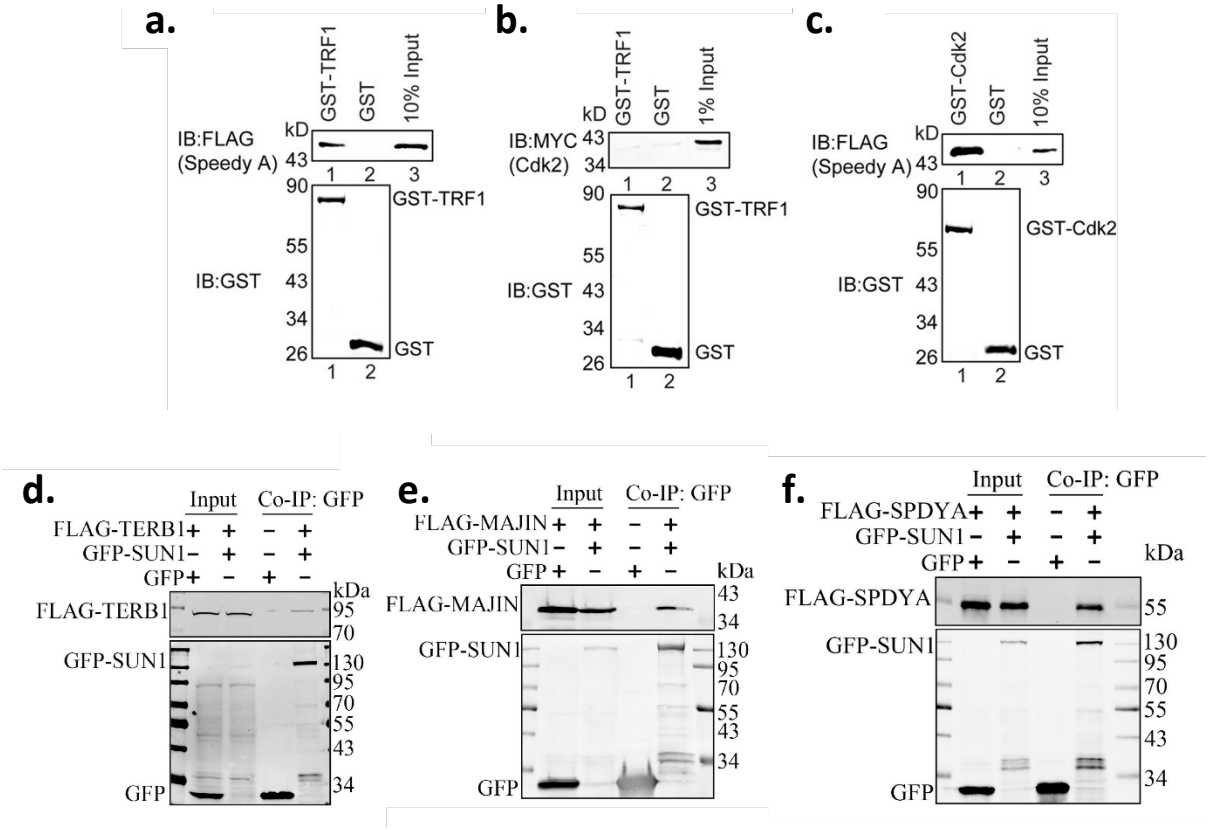


Figure 3.3.6.| SpeedyA interacts with SUN1 and TRF1
a-b) GST pull-down assays were used to analyse the interactions of TRF1 with a) SpeedyA, b) Cdk2.GST-TRF1 or GST conjugated sepharose beads were used to pulldown His-FLAG-SpeedyA or His-MYC-Cdk2. a) Direct interaction between TRF1 and SpeedyA. b) No interaction was observed between TRF1 and Cdk2. c) GST pull-down assay was used to analyse the interaction between CDK2 and SpeedyA. GST-CDK2 or GST conjugated sepharose beads were used to pulldown purified His-FLAG-SpeedyA. A direct interaction between SpeedyA and CDK2 was determined. d-f) Co-immunoprecipitation assays were performed to analyse the interaction of SUN1 with d) TERB1, e) MAJIN and f) SpeedyA (SPDYA). d) Immunoprecipitation shows that SUN1 interacts with TERB1, shown by a faint band. e) MAJIN and f) SpeedyA.
Figures from a-c) Wang et al., 2018 and d-f) Wang et al., 2020

Furthermore, we have shown that by late pachytene upon TRF1 displacement the meiotic telomere complex directly binds telomeres forming robust chromosome attachment to the nuclear membrane allowing for meiotic progression. Upon chromosome synapsis, telomere attachments mature into a disk-shaped telomere attachment plate at the inner nuclear membrane in which the synaptonemal complex fuses to. This is achieved through the interaction between the axial element proteins (AEs), SYCP2 and SYCP3, and the cohesin cores which form at the nuclear membrane during meiosis (Adelfalk et al.,

2009; Eijpe et al., 2000; Liebe., et al 2004). These observations suggest that the meiotic telomere complex may function to facilitate synaptonemal complex formation and chromosome structure through meiotic cohesins. As previously mentioned, the C-terminal Myb-like domain of TERB1 functions to recruit cohesins to telomere ends thus introducing telomere ridigity, and Y2H studies have shown a direct interaction between TERB1 and the meiosis-specific cohesin STAG3 (Shibuya & Watanabe, 2014). Furthermore, mouse immunofluorescence studies have detected co-localisation of synaptonemal complex proteins SCP2 and SCP3 and STAG3 at the attachment plate, which contains telomere repeats (Liebe et al., 2004; Prieto et al., 2001). We have proposed a model where the meiotic telomere complex assembly, with telomeric DNA, may recruit meiotic cohesins, potentially with axial element proteins, in addition to the LINC and shelterin complexes, to achieve meiotic progression by the formation of the synaptonemal complex.

Chapter 4

Chacterisation of the *Drosophila* transverse filament protein, C(3)G

4.1. INTRODUCTION

4.1.1. The tripartite structure of the synaptonemal complex is conserved across eukaryotes

At zygotene of prophase I, rapid prophase movements (RPMs) of chromosomes result in 'bouquet' formation and subsequent homologous chromosome pairing (Shibuya et al., 2014). Initially, recombination intermediates at a distance of ~ 300 nm are created between homologous chromosomes. Subsequently, the chromosomal axes become juxtaposed at ~100 nm and are bridged by a conserved proteinaceous structure known as the synaptonemal complex (SC) (von Wettstein, 1984; Walker & Hawley, 2000; D. Zickler & Kleckner, 1999). The SC structure was first visualised by electron microscopy (EM) in crayfish spermatocytes in 1956, but since then it has been cytologically observed across other eukaryotes (Moses, 1956). The molecular components of the SC have been visualised in several model organisms including mice, worms, flies, and yeast revealing that the SC has a highly conserved tripartite structure (Figure 4.1.1.) (Cahoon & Hawley, 2016; Colaiácovo et al., 2003; De Muyt et al., 2014; Hamer et al., 2006; Hemmer et al., 2014; Ortiz et al., 2016). By early pachytene, the mature SC is fully assembled and provides the essential framework required for homologous recombination and synapsis (Moses, 1956). The SC has an elaborate zipper-like structure connecting two perfectly aligned homologous chromosomes along their entire lengths. Furthermore, it is suggested that, with some exceptions, the SC has two conserved functions across eukaryotes; firstly, the SC provides a means of stabilising and maintaining associations along the lengths of homologous chromosomes together during the latter stages of prophase I. Secondly, formation of the SC, in particular the central region, promotes maturation of recombination intermediates into crossover (CO) products (de Boer & Heyting, 2006; Hayashi et al., 2010). Additionally, deletion studies of individual SC proteins cause CO disruption and chromosomal non-disjunction, indicating that the SC is critical for successful meiotic recombination.

The SC is comprised of two parallel lateral elements (LEs) that flank the chromatin and a single central element (CE). The LEs and CEs are held together by a network of transverse filaments (TFs) that lie perpendicular to the complex (D. Zickler & Kleckner, 1999). Despite the strong structural conservation,

the proteins that comprise the SC display no apparent, or very low sequence conservation, outside their respective genus. An example of this is the SCs from *Drosophila melanogaster* (*D. melanogaster*) and *Caenorhabditis elegans* (*C. elegans*) (Fraune et al., 2013; Grishaeva & Bogdanov, 2014).

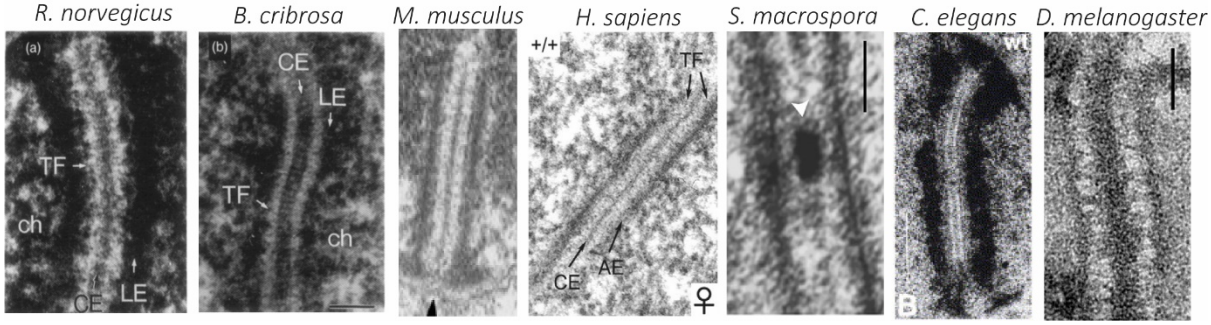
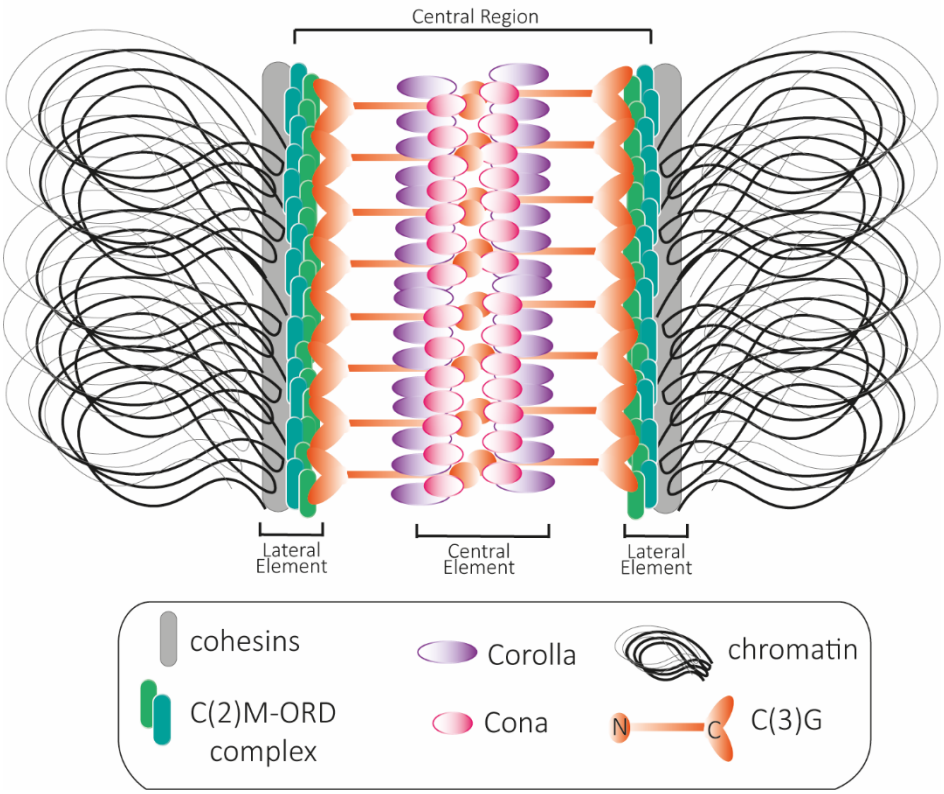


Figure 4.1.1.| The synaptonemal complex forms a conserved tripartite structure.
Electron micrographs of the synaptonemal complex across eukaryotic species shows a conserved tripartite structure. Figures taken from Cahoon & Hawley, 2016; Colaiácovo et al., 2003; De Muyt et al., 2014; Hamer et al., 2006; Hemmer et al., 2014; Ortiz et al., 2016

4.1.2. The *Drosophila melanogaster* synaptonemal complex

In contrast to mammals, where the SC is present in both males and females, the SC is found only in female *Drosophila* and is necessary for meiotic exchange (Gowen, Gowen, 1922). Several proteins of the *D. melanogaster* SC have been characterised *in vivo*, including the core components constituting the characteristic tripartite structure (Cahoon & Hawley, 2016; Hughes et al., 2018). All core proteins identified are exclusively female meiotic proteins except for the cohesion protein, ORD, which is present in both sexes (Webber et al., 2004). The LEs are comprised of cohesins and protein-associated cohesins including ORD and C(2)M. ORD localises to the arms of the chromosome during early prophase I and has been shown to be necessary for loading of the cohesion complex onto the chromosomal axis. C(2)M also resides at the chromosome axis and is required for assembling a continuous central element suggesting a possible interaction with the TF (Bickel et al., 1996; Manheim & McKim, 2003; Webber et al., 2004).

a.



b.

SC component	Human	Drosophila
Central element	SYCE1-3, TEX12, SIX6OS1	CONA, COROLLA
Transverse filament	SYCP1	C(3)G
Lateral element	SYCP2, SYCP3	C(2)M, ORD

Figure 4.1.2.| *D. melanogaster* females have a conserved synaptonemal complex
a) Despite having no sequence conservation with other eukaryotes, *D. melanogaster* forms a SC with the tripartite structure observed in mammals. The LE is composed of cohesins and the cohesion-related complexes, C(2)M and ORD, that function to link the chromatin to central element components of the SC. The CE is formed of two proteins Cona and Corolla. C(3)G lies perpendicular to the LE and spans across the central region. The N-terminal domain of C(3)G resides within the CE, suggesting that C(3)G may interact directly or indirectly with the central element proteins. The C-terminus of C(3)G is positioned within the LE. However, how the C-terminus interacts within this region remains unclear. b) Table to show the proteins constituting the synaptonemal complex of both human and *drosophila*. Both organisms have a single transverse filament protein, Sycp1 and C(3)G, respectively.

In *D. melanogaster* only a single TF protein, crossover suppressor on 3 of Gowen (C(3)G), has been identified to date. C(3)G spans the central region of the SC, linking the LEs to the middle CE (Page & Scott Hawley, 2001). CONA and Corolla have also been identified as CE proteins in *Drosophila* SC. CONA functions to promote DSB maturation into COs and colocalises with C(3)G. Deletion of either the central region proteins, *C(3)G* or *Cona*, leads to defects in homologous pairing in pachytene, as the

SC can no longer assemble (K. A. Collins et al., 2014; Page et al., 2008). Schematised in Figure 4.1.2.a. is the tripartite structure of the *Drosophila* SC, and the corresponding proteins in the mammalian SC (Figure 4.1.2.b.). In both mammalian and fly organisms, only a single protein constitutes the TF, SYCP1 and C(3)G, respectively.

4.1.3. Structure of transverse filament proteins is conserved across eukaryotes

TF proteins have been identified in several species and by electron microscopy (EM) and it has been deduced that the TF filament proteins span the central region of the SC. Furthermore, despite the low sequence conservation between species, EM has observed a highly conserved structural feature within the central region with a width of ~ 100 nm (von Wettstein, 1984). Yeast (Zip1), mammalian (SCP1/SYCP1), and fly (C(3)G) TF proteins comprise ~850 amino acids and have a conserved secondary structure formed of three domains: a central amphipathic α-helical coiled-coil domain flanked by C- and N-terminal domains of unknown structure (Figure 4.1.3.). In contrast, *C. elegans* have two transverse filament proteins, SYP-1 and SYP-2 which are much shorter but appear to retain the central coiled-coil domain, based on structural predictions.

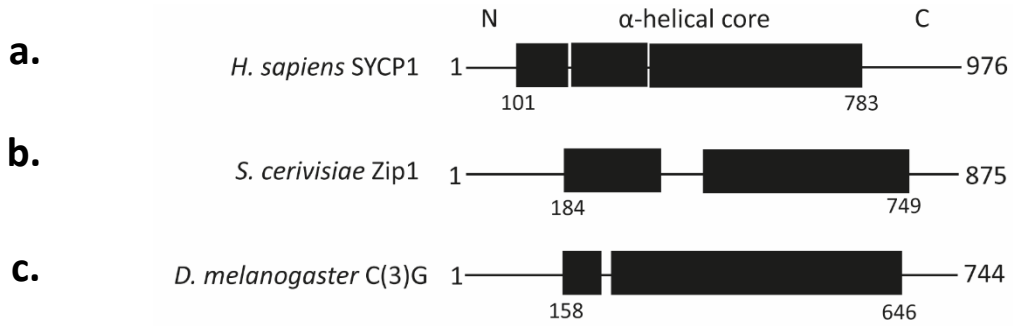


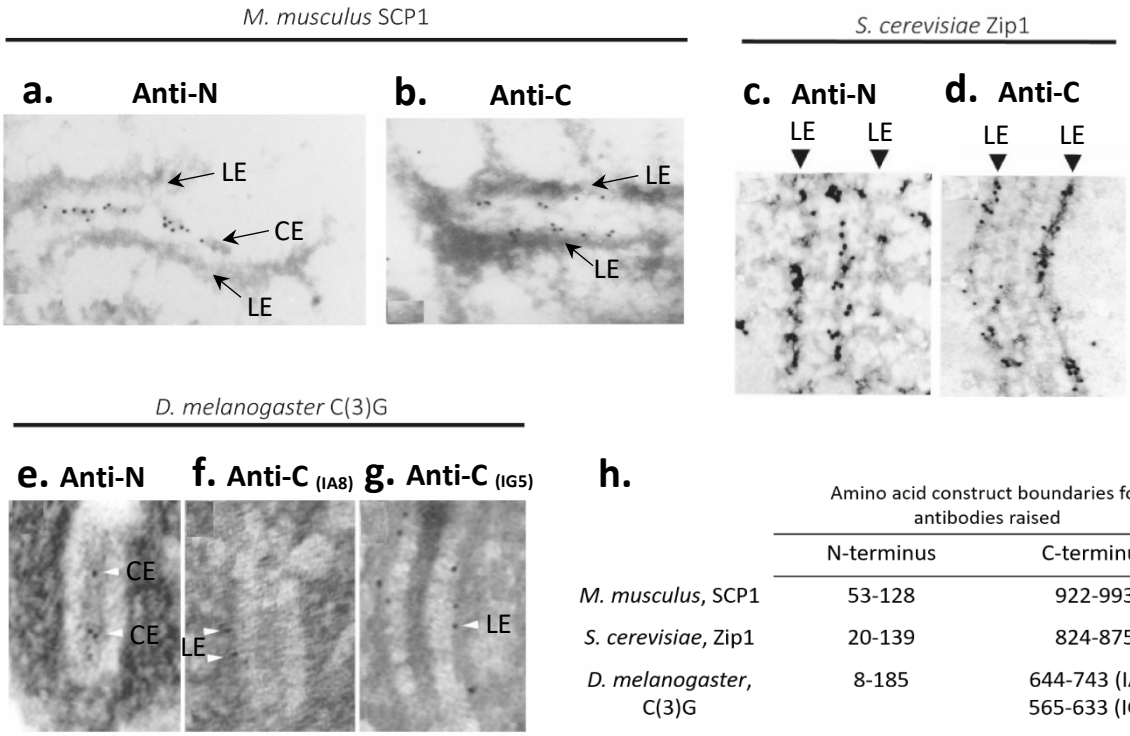
Figure 4.1.3.| Transverse filament proteins have three distinct domains.
a-c) Sequence analysis of SYCP1, Zip1 and C(3)G reveal the presence of a highly conserved α-helical core predicted to be an extended coiled-coil domain. a) The α-helical core of SYCP1 (amino acids 101-783) is highly conserved at both ends and is flanked by unstructured N- and C-terminal tails. b) The central α-helical domain of Zip1 (residues 184-749) is predicted to contain two coiled-coil regions linked by non-helical linker. c) Secondary structure predictions of C(3)G determined a central coiled-coil domain (residues 158-646) flanked by globular C- and N-terminal domains. The coiled-coil domain contains four segments of coiled-coils, with the most N-terminal (residues 158-195) is separated by a 28 amino acid linker.

In addition, the TF proteins all have a highly basic C-termini, with an isoelectric point (pI) of ~10, suggesting that the C-terminal domain may have a conserved function. Conserved features across eukaryote species are summarised in Table 4.1.1. This conserved secondary structure suggests that, despite low sequence conservation, TFs may have a similar function within the SC.

Table 4.1.1.| Eukaryotic transverse filament proteins have conserved features.
Mammalian, yeast, and fly TF proteins are a similar length comprised of a central coiled-coil domain flanked by unstructured C- and N-terminal domains. The coiled-coil region is ~650 amino acids which corresponds to a conserved central region width of ~100nm. Further to this the isoelectric point (pI) of the C- and N-terminal domains are conserved. A high pI at the C-terminus suggests possible DNA binding. In comparison, nematodes (*C. elegans*) have two TF proteins, SYP-1 and SYP-2. Despite SYP-1 being much shorter it also has a central coiled-coil domain that is proportional to the other TF proteins.

	Amino acid residues			Isoelectric point (pI)			
Species, protein	Total length	Coiled-coil domain	Central region width, nm	N-domain	Central domain	C-domain	Full length
H. sapiens, SYCP1	973	677	100	5.0	5.4	9.7	5.7
M. musculus, SYCP1	993	713	100	5.9	5.3	9.7	5.8
S. cerevisiae, Zip1	875	632	115	4.8	6.1	10.1	6.4
D. melanogaster, C(3)G	744	495	109	10.0	4.9	9.7	5.9
C. elegans, SYP-1	489	355	70-85	3.66	8.1	10.0	6.1
C. elegans, SYP-2	213	50	70-85	-	-	-	-

Immunolocalisation studies using specific C- and N-terminal antibodies and immunogold labelling have determined the positioning of SCP1, Zip1, and C(3)G, within the SC of mouse, yeast, and fly organisms, respectively (Figure 4.1.4.). Super resolution EM techniques elucidated a conserved organisation of these proteins within the SC, in which the C-termini reside within the LE and the N-termini cross the centre of the SC to form the transverse filaments (Anderson et al., 2005; Dong & Roeder, 2000; J. G. Liu et al., 1996). Furthermore, Y2H screens have shown that the N-terminus of mice SCP1 strongly interacts with itself but not with other protein domains (Liu et al., 1996). These findings suggest that the N-terminal domain is involved in establishing self-interactions within the central element to form homo-oligomers that span from the inner region of the LE to the respective far side of the CE. Mutations that either increased or decreased the length of the coiled-coil domain in Zip1 resulted in corresponding changes to the width of the SC, confirming that the TF proteins lie perpendicular to the LEs spanning the CR (Dong & Roeder, 2000; Schmekel et al., 1996).



h.

	Amino acid construct boundaries for antibodies raised	
	N-terminus	C-terminus
M. musculus, SCP1	53-128	922-993
S. cerevisiae, Zip1	20-139	824-875
D. melanogaster, C(3)G	8-185	644-743 (IA8) 565-633 (IG5)

Figure 4.1.4.| Antibodies raised to different fragments of transverse filament proteins.
In separate studies the transverse filament proteins, SCP1, Zip1 and C(3)G all have had antibodies raised against the N- and C-terminal domains to determine their organisation within the SC by immunoelectron microscopy. a-b) Immuno-gold electron microscopy analysis of spread meiotic chromosomes in mice demonstrate central element (CE) localisation of the SCP1 N terminus (a) and the SCP1 C terminus is embedded within the lateral elements (LEs) (b). c-d) Electron micrographs of the *S. cerevisiae* SC labelled with gold-conjugated antibodies specific to the N- and C-terminus. e-g) In line with both human and yeast the fly (*D. melanogaster*) transverse filament protein, C(3)G, localises in the same manner as SCP1 and Zip1. Antibodies were raised for both the N- and C-terminal domains. The C-terminal domain has two antibodies, one specific to the f) globular domain (IA8) and one for the g) C-terminal end of the coiled coil domain (IG5). Both IA8 and IG5 localise to the very edge of the central element next to the lateral element of the SC and IG5. It was possible to determine that the IA8 is slightly closer to the LE compared to IG5, as expected. h) Table to show the specific amino acid constructs used.
Figure adapted from: Liu *et al.*, 1996, Dong & Roeder, 2000 and Anderson et al., 2005.

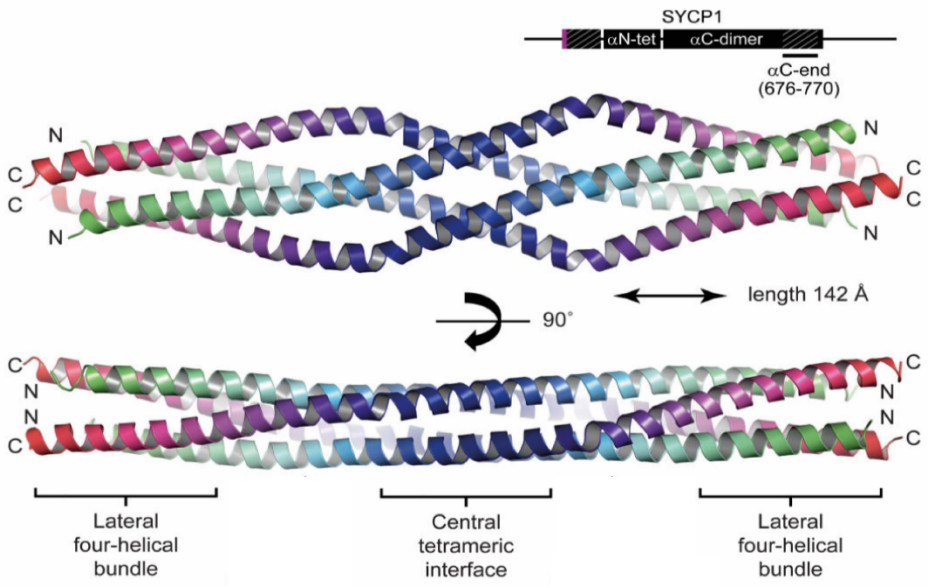
Further studies have shown that TF proteins self-interact forming bundles of thin fibres of at least two layers of TF protein molecules (Hernández-Hernández et al., 2016; Schücker et al., 2015). Ultrastructural imaging studies have observed small differences in the diameter of the TF across different eukaryote species. In the rat SC, the TFs are less frequent in comparison to those observed in insects, suggesting that the differences are in the degree of order and not the structural units themselves. These observations further suggest that the overall structure of transverse filaments is conserved (Schmekel et al, 1993; Solari & Moses, 1973).

Mutational analysis of the C-terminus in both SCP1 and Zip1 have shown that the C-terminal domain is crucial, for both formation of the SC and for the localisation of the SC to chromosomes. The C-terminal domain of SCP1 has also shown to have DNA binding characteristics (Meuwissen et al., 1992; Tung & Roeder, 1998). Together, these observations provide a possible mechanism for attachment of chromatin to the SC, in which the TF molecules form homo-oligomers between the LE and CE of the SC. It has been postulated that SYCP1 and ZIP1, could function as a molecular scaffold to orchestrate the alignment of homologous chromosome pairs during meiosis in both mice and yeast.

4.1.4. Human SYCP1 transverse filament assembly

The human TF is formed from a single 976 amino acid protein, SYCP1, which spans between the LE and CE. As previously mentioned, SYCP1 has a central α-helical core forming a predicted coiled-coil domain flanked by unstructured C- and N-terminal domains. The C- and N-termini are localised within the CE and LEs, respectively. Unlike other organisms, the mammalian TF protein has been extensively studied, including the recent crystal structures of SYCP1 coiled-coil fragments (Dunce et al., 2018; Seo et al., 2016).

The crystal structure of a C-terminal coiled-coil fragment (amino acid residues 676-770) was solved as an antiparallel tetrameric assembly (Figure 4.1.5.) (Dunce et al., 2018). Furthermore, the oligomeric state was shown to be influenced by pH; at neutral pH SYCP1 forms a dimer but at acidic pH it can assemble into a tetramer. The tetrameric crystal structure has a series of surface basic patches, suggesting a direct interaction with the DNA backbone may be possible (Dunce et al., 2018). These observations suggest that SYCP1 may be stable in different oligomeric states.



PDB code – 6F63

Figure 4.1.5.| SYCP1 C-terminal coiled-coil fragment crystal structure.
Crystal structure of SYCP1C-terminal region, $SYCP1_{676-770}$, solved to 2.15 Å, revealed an anti-parallel tetrameric assembly length of 142 Å. The central tetrameric interface consists of two stacked layers held together by a hydrogen-bond network. Each end of the tetramer is held together by a four-helical bundle formed by a hydrophobic core and anti-parallel interfaces. PDB code – 6F63. Figure taken from Dunce et al., 2018.

DNA binding studies show that SYCP1 can bind dsDNA at an acidic pH 5.5 (tetramer) but not at pH 8 (dimer) (Dunce et al., 2018). The determined structures of SYCP1 are complementary to the immunolocalisation studies, that showed the C-terminal region is positioned near the LE, suggesting that the C-terminus binds to the chromatin within the LE (J. G. Liu et al., 1996; Meuwissen et al., 1992).

The X-ray crystal structure of the N-terminal fragment of SYCP1 α-helical core ($SYCP1_{101-206}$) revealed the formation of an elongated tetramer, in which two opposing dimers interact through a head-to-head assembly, demonstrating a 'dimer of dimers' assembly (Figure 4.1.6.). Biophysical characterisation of the N-terminal region of $SYCP1_{101-206}$ by SEC-MALS determined that residues 101-111 assemble into large molecular weight species, suggesting that this region is essential for the self-assembly of the α-helical core (Dunce et al., 2018).

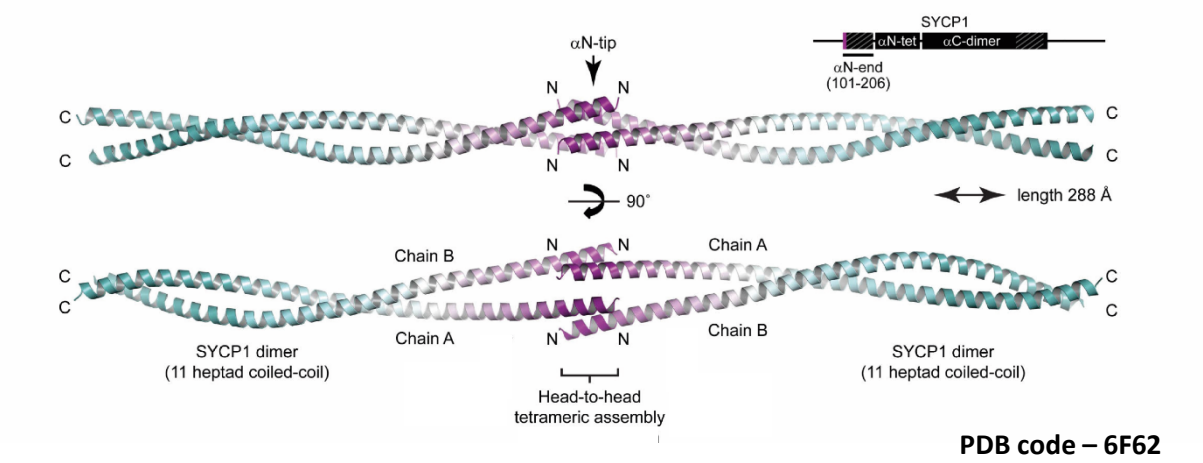


Figure 4.1.6.| Crystal structure of SYCP1 N-terminal coiled-coil fragment.
a) Crystal structure of $SYCP1_{101-206}$, solved to 2.07 Å, revealed a head-to-head tetramer formed from two opposing parallel dimers, a 'dimer of dimers'. Each parallel α-helical dimer consists of 11 heptad repeats, spanning a total length of 288 Å. PDB code – 6F62. Figure taken from Dunce et al,. 2018.

Combining these structural findings of SYCP1 with previous studies, a model was proposed for the formation of a SC precursor lattice, in which the self-assembly of SYCP1 plays a crucial role (Dunce et al., 2018). Initially the C-terminus dimeric assembly of the coiled-core core is recruited to the LEs via DNA binding of chromatin and stabilised by interacting with the LE protein SYCP2 (Winkel et al., 2009). This is followed by the self-assembly of SYCP1 to form a back-to-back tetrameric assembly at the CE, where, at the midline, SYCP1 parallel dimers form a head-to-head 'dimer of dimer' assembly. The self-assembly of SYCP1 is suggested to act as a preliminary SC lattice which provides a landing pad for further central element components to assemble a mature SC (Dunce et al., 2018). The overall structure of the SYCP1 preliminary SC lattice is schematised in Figure 4.1.7.

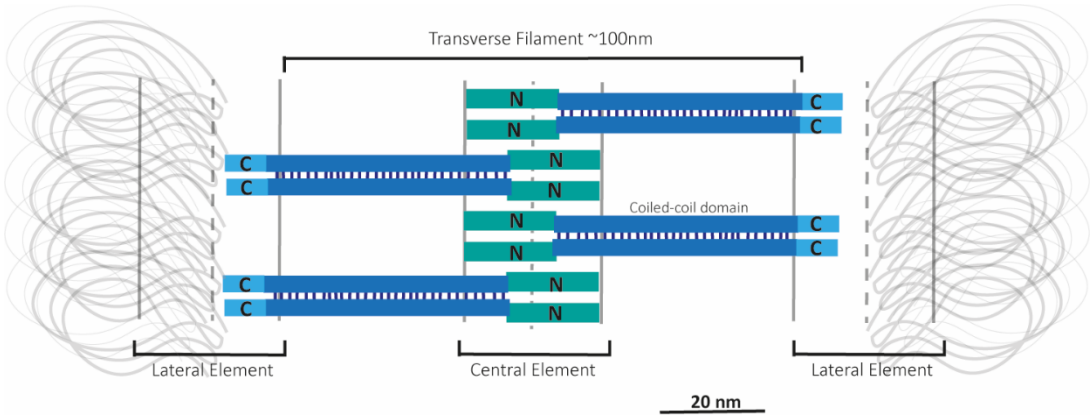


Figure 4.1.7.| Schematic showing the central region of the synaptonemal complex.
The structure of eukaryote transverse filament proteins are conserved suggesting a conserved function. Through numerous studies including immunolocalisation, super resolution microscopy, and X-ray crystal structures it has been possible to conclude an assembly mechanism. The C-termini resides within the boundary of the central region and lateral element. The crystal structure of the C-terminal end determined an anti-parallel dimeric assembly. It is known that the dimeric structure can bind to DNA suggesting that the TF filament proteins bind to the chromatin within the LE. Furthermore, the TF protein can self-assemble to form larger tetrameric assemblies or a 'dimer of dimers'. The central coiled domain spans across the central region forming layers. The N-termini reside within the central element and possibly interacts with other central element proteins. With the anti-parallel assembly, the distance between the C-termini of a dimer can be measured to determine the width of the transverse filament region to be ~100nm.

4.1.5. *Drosophila melanogaster* as a model organism for studying meiosis

The genetic systems in eukaryotic organisms, including both humans and fruit flies, share many similarities, making *Drosophila* an extremely useful organism to study common genetic processes, including meiosis. The fruit fly is the currently the main invertebrate model to study developmental genetics. Our current understanding of the human SC is growing, through in depth *in vitro* biophysical analysis of individual protein and protein complexes, *in vivo* imaging techniques and complementary mouse models (Davies et al., 2012; Dunce et al., 2018; Dunne & Davies, 2019; Lu et al., 2014; Syrjänen et al., 2014). Although there are many advantages of mouse models, they can be time consuming and very expensive. Around 75% of known human diseases have recognisable genes in the fruit fly genome, and around 50% of fly protein sequences have mammalian homologues.

The fruit fly makes a good model organism as it has a short lifecycle, taking only 10 days from fertilisation for an adult fly to develop. Furthermore, a female can produce up to 1500 eggs in its lifespan, which means a large quantity of flies can be produced over a short period of time. Finally, fruit flies are facile to maintain in a laboratory setting (Ables, 2015; Hughes et al., 2018).

Despite *D. melanogaster* only having four chromosomes, they still display sexual dimorphism. The complete genome has been sequenced and annotated, and tools exist to genetically manipulate the genome. Most of our current knowledge involved in female *Drosophila* meiosis has come from the analysis of mutations that disrupt various aspects of this process, including the SC proteins. In combination with our knowledge about the mammalian SC, we can create complementary *Drosophila* genetic manipulations using a wealth of genetic tools including the CRISPR/Cas9 technique to further our *in vivo* understanding. Genetic screens in *D. melanogaster* continue to be used to identify new genes that play a role in meiotic processes (Ables, 2015).

4.1.6. C(3)G encodes a *Drosophila* transverse filament protein

C(3)G is a 744 amino acid meiosis specific protein, essential for meiotic exchange in female *Drosophila* (Gowen and Gowen, 1922). C(3)G is predicted to have a central coiled-coil region (amino acid residues 158-646) flanked by N- and C-terminal globular domains, schematised in Figure 4.1.3.a. Secondary structure predictions revealed that the amphipathic α-helical core is comprised of four stretches of coiled-coil structure, with the most N-terminal of these (residues 158-195) separated by a 28 amino acid linker. The remaining three coiled coil segments could function as one long stretch that forms the TF structure (Jeffress et al., 2007; Lupas et al., 1991; Page & Scott Hawley, 2001). Based on the conserved secondary structure, it was proposed that C(3)G serves a structural role in *Drosophila* SC, like that seen by Zip1 and SYCP1 (SCP1) in yeast and mammals, respectively.

Studies, using anti-C(3)G antibodies, stain meiotic chromosomes in a thread-like pattern along the lengths of the chromosomes with clear association of C(3)G in the centre of the paired chromosomes, suggesting that they reside within the central region of the SC (Cahoon & Hawley, 2016; Page & Hawley, 2004). These observations are similar to the ones seen in both transverse filament proteins Zip1 and Syn1 in yeast and mice, respectively (Dong & Roeder, 2000; Liu et al., 1996; Schmekel et al., 1996).

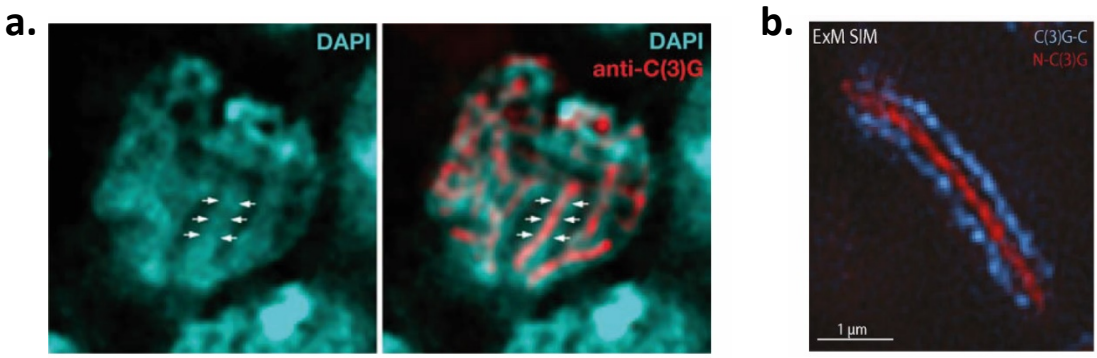


Figure 4.1.8.| Immunolocalisation of *D. melanogaster* C(3)G.
a) Deconvoluted optical section of a *D. melanogaster* pro-oocyte nucleus stained with DAPI (cyan) and anti-C(3)G (red) immunofluorescence. The transverse filament protein C(3)G localises along the length of synapsed homologous chromosomes in meiotic prophase in thread-like structures associated at the centre of the DAPI-stained meiotic chromosomes. b) Expansion microscopy (ExM) linked with structured illumination microscopy (SIM) determined super resolution microscopy of the transverse filament protein C(3)G. The C-terminal antibody (blue) labels the LE-CR boundary and the N-terminal antibody (blue) resides within the CR of the SC. It can be clearly seen that the C-terminal antibody foci forms two elongated tracks along the lengths of the SC, with the N-terminal foci forming a central track. The distance between the C-termini in opposite tracks is ~120nm. Figures taken from a) Hawley, 2004. and b) Cahoon and Hawley., 2016.

Additional studies have determined that null mutants of *C(3)G* do not diminish DSB formation. However, meiotic COs are abolished (Jang, et al., 2003; Page & Hawley, 2001). This result suggests that in Drosophila maturation of DSBs is dependent upon TF formation, indicating C(3)G has a crucial role in SC formation. Furthermore, an in-frame deletion within the coiled coil region, $C(3)G^{x204}$ (deletion of amino acid residues 340-552), still maintained recombination. However, high levels of non-synapsed or partially synapsed chromosomes were observed (Page & Hawley, 2001). A separate study tested the ability of four transgenic constructs of C(3)G, that express an in-frame deletion, to promote

synapsis and recombination in meiosis (Figure 4.1.9.a.). Using immunofluorescence microscopy it was possible to determine that the N-terminal globular domain of C(3)G and the N-terminal part of the coiled-coil segment are critical for meiotic recombination and SC formation, shown by aberrated localisation of C(3)G in pro-oocyte nuclei (Figure 4.1.9.b.-e.). Furthermore, deletion of the C-terminal domain also disrupted the thread-like localisation pattern of C(3)G, suggesting that the C-terminus is crucial for SC formation (Figure 4.1.9.f.) (Jeffress et al., 2007).

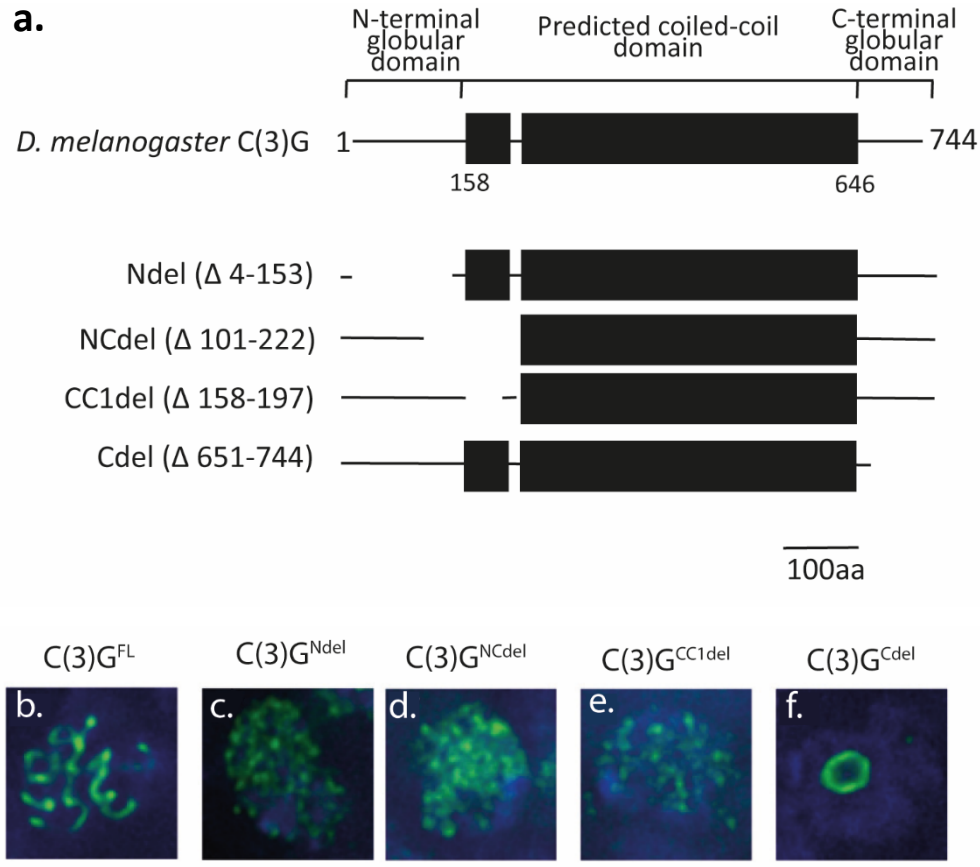


Figure 4.1.9.| Localisation studies of C(3)G in-frame deletions.
a) Schematic showing the domain structure of C(3)G and the four transgenic in-frame deletions. Ndel (Δ 4-153) removes most of the globular N-terminal domain. NCdel (Δ 101-222) and CC1del (Δ 158-197) have removed the first coiled-coil segment as well as part of the N-terminal domain in NCdel. The final deletion, Cdel, removes the C-terminal domain (Δ 651-744). b-f) Immunofluorescence images of *Drosophila* pro-oocyte nuclei expressing full length protein or the in-frame deletions of C(3)G stained with anti-C(3)G antibody (green). Nuclear DNA stained with DAPI (blue). b) In $C(3)G^{FL}$ a clear localisation pattern can be determined, in which C(3)G displays a thread-like appearance. c-e) All C(3)G N-terminal deletion constructs revealed an altered spotty non-uniform pattern of C(3)G localisation within the nucleus. f) Immunofluorescence of $C(3)G^{Cdel}$ shows a globular ring-like pattern of staining within the nucleus, possibly a globular aggregate. Figures adapted from Jeffress et al., 2007.

Together, these results suggest that the three domains of the C(3)G transverse filament protein plays a crucial role in the meiosis, in particular synaptonemal complex formation. Despite these extensive studies of C(3)G *in vivo*, the structure of C(3)G is yet to be characterised biophysically *in vitro*.

4.1.7. *In vivo* analysis of three in-frame C(3)G deletions

An *in vivo* study of C(3)G, analysed the SC maintenance throughout prophase I of meiosis using immunostaining and immuno-fluorescence techniques. Three in-frame deletions were created within the coiled-coil domain of C(3)G using the CRISPR/Cas9 technique, to investigate the role of C(3)G in the formation and function of the SC (Billmyre et al., 2019). The three in-frame deletions were designed around the predicted coiled-coil structure using the COILS server (Lupas et al., 1991; Lupas, 1997). Deletion 1 (Δ 340-552) removed around half of the predicted central coiled-coil domain whilst, deletion 2 (Δ 346-361) and deletion 3 (Δ 465-471) removed small regions, of 15 and 7 amino acid residues, respectively. These were areas that the COILS score decreased suggesting loss of coiled-coil structure (Figure 4.1.10.) (Lupas et al., 1991). It was determined that all three deletions caused a partial loss of SC function at different stages of early meiosis (Billmyre et al., 2019).

The localisation and length of the SC was observed by staining for Corolla and measured using stimulated emission depletion (STED) in the three deletions and was compared with wild type protein. In homozygous deletion 1 females, SC formation in early pachytene was comparable to wild type, but by mid-pachytene some discontinuities of the SC were seen, including a decrease in length. These results suggest that the formation of the SC is not disrupted by this large deletion but instead, the width of the SC between the C-termini C(3)G decreased by around half matching the degree of the coiled-coil region that had been removed. Analysis of C(3)G deletion 2 revealed a similar phenotype to deletion 1, despite only 15 amino acids being deleted compared to the 213 amino acid residues of deletion 1. Surprisingly, the smallest 7 amino acid deletion, deletion 3 was unable to form a fully assembled SC (Figure 4.1.11.a.-c.). From these results it was concluded that C(3)G may have a specific role between early to mid-pachytene, and deleting part of the coiled-coil region may have affected the ability of

C(3)G to recruit other proteins or self-interact to form a stable synaptonemal complex (Billmyre et al., 2019).

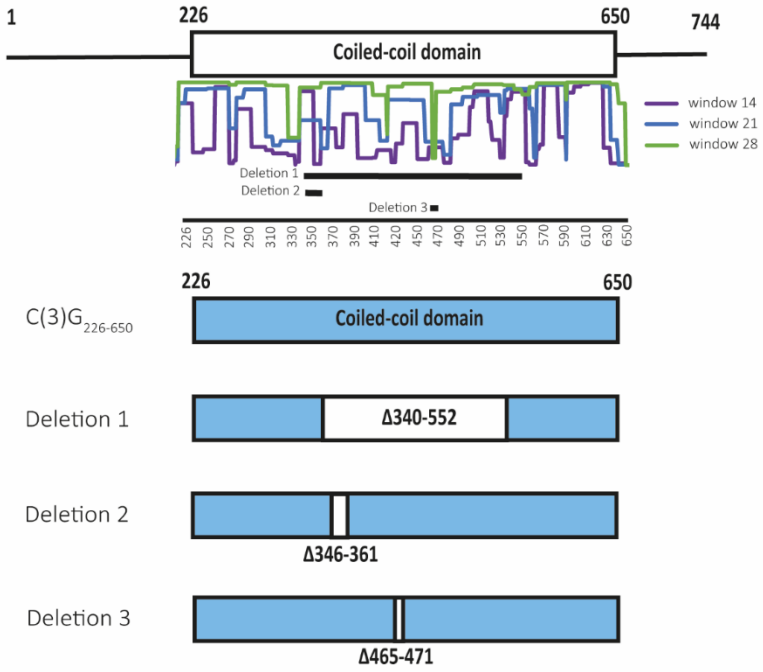


Figure 4.1.10.| Schematic showing the three in-frame C(3)G deletions
a) C(3)G is 744 amino acids long and contains a large coiled-coil domain (amino acid residues 226-650) flanked by globular C- and N-termini. C(3)G has been predicted to form parallel dimers that span the width of the synaptonemal complex. The secondary structure coiled-coil prediction server COILS show that central region is formed of four stretches of coiled-coils which may function separately or all together. The three in-frame deletions were designed using the COILS prediction, both the regions of deletions 2 and 3 have a dip in the COILS score, indicating a potential loss of coiled-coil structure. Deletion 1 is a 213 amino acid deletion of the central region of the predicted coiled-coil domain. Deletions 2 and 3 are much smaller deleting only 15 and 7 amino acid residues, respectively. It has been predicted that deletion 3 removes a single heptad.

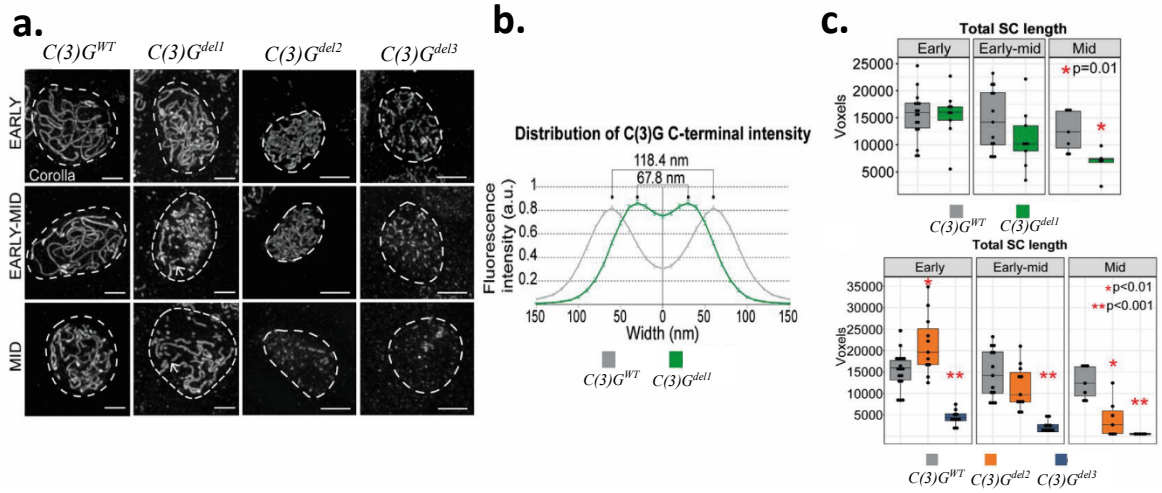


Figure 4.1.11.| In frame C(3)G deletions within the coiled-coil region results in SC discontinuities.

a) Localisation of the SC central element protein, corolla in wild type *(C(3)G^{WT})* and the three in-frame deletions, *C(3)G^{del1}*, *C(3)G^{del2}* and *C(3)G^{del3}* throughout pachytene in prophase I. Central element protein, corolla, was used as a marker of SC formation and the dashed line indicates the location of the nucleus. In WT nuclei by pachytene the SC is fully assembled and central element proteins including corolla are associated with the paired chromosomes shown by long thread-like structures. The SC remains assembled throughout pachytene to assist homologous recombination and crossing over. Towards mid to late pachytene the SC then begins to disassemble. In homozygous C(3)G^{del1} females early pachytene looks normal as the SC has assembled, however during early-mid pachytene there is a significant decrease in SC length and there are obvious discontinuities in the SC present (white arrows). C(3)G^{del2} is also able to form a SC which has a similar length to WT in early pachytene but also decreases in length throughout pachytene. In contrast C(3)G^{del3} is unable to form a fully assembled SC. (Scale bars, 2µm). b) STED data was used to determine the average distance between the C-terminal domains of a C(3)G homodimer presuming that they form anti-parallel dimers. The distance between C-termini of C(3)G^{WT} (grey) is 118.4 ± 0.6 nm (SEM) consistent with previous EM data, however in C(3)G^{del1} (green) this is reduced by nearly half to 67.8 ± 0.1 nm (SEM). The average distribution was generated by averaging 46 line profiles from 8 C(3)G^{WT} nuclei and 35 line profiles from 12 C(3)G^{del1} nuclei. c) Quantification of the total track length of SC in wild type C(3)G and deletions throughout pachytene. In early pachytene the SC length in C(3)G^{del1} (green) is comparable to WT but throughout pachytene the SC length decreases with a significant decrease in mid pachytene, this is also the case for C(3)G^{del2} (orange) SC. In C(3)G^{del3} (blue) the SC is never full assembled. *P = 0.01 and **P < 0.001 by t test. C(3)G^{WT}: n=17 (early), n=13 (early to mid), and n=7 (mid); C(3)G^{del1}: n=9 (early), n=9 (early to mid), and n=5 (mid). C(3)G^{del2}: n=11 (early), n=11 (early to mid), and n=7 (mid); C(3)G^{del3}: n=10 (early), n=10 (early to mid), and n=5 (mid). Figure from Billmyre *et al.,* 2019

Other studies have shown that in the SC is crucial for successful homologous recombination in *Drosophila* (Jeffress et al., 2007; Page & Hawley, 2001). Therefore, Billmyre et al. examined if the coiled-coil deletions of C(3)G have an effect on recombination in chromosomes. It was determined that both deletion 1 and deletion 3 exhibit defects in recombination with the X-chromosome being most effected. In comparison, deletion 2 showed nearly normal levels of recombination (Figure 4.1.12.a.-c.).

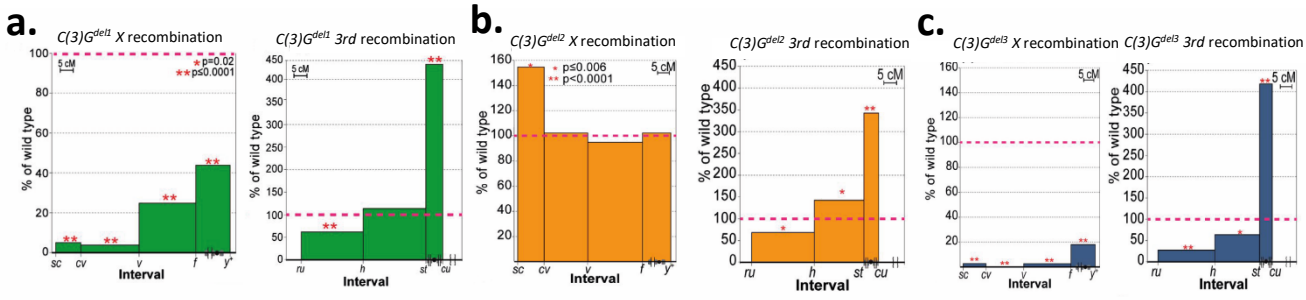


Figure 4.1.12.| In frame C(3)G deletions within the coiled-coil region results in reduced recombination.

Drosophila females have four chromosomes and an X-chromosome however chromosome 4 does not undergo recombination. Recombination on the X and 3rd chromosome of wild type and deletion females are plotted as a percentage of wild type versus chromosome location. a) In C(3)G^{del1} females recombination on the X-chromosome decreased along the entire length from 63 to 11.8cm. In comparison, recombination on the 3rd chromosome did not decrease, but there was a large increase in the proximal euchromatin region. This was also the case for d) C(3)G^{del3} mutants, however the decrease was more severe on the X-chromosome as recombination was reduced to only 4.5% of wild type. b) In C(3)G^{del2} females recombination is maintained on both the X- and 3rd chromosome. Figure taken from Billmyre *et al.,* 2019

Unlike most organisms the SC in *Drosophila* is fully formed prior to DSBs (Lindsley & Sandler, 1977). Disruption of the SC results in significant downstream effects; abolished crossovers and significantly reduced homologous pairing, suggesting that the SC is crucial for successful pairing and crossing over of homologous chromosomes (K. A. Collins et al., 2014; Hughes et al., 2018). Furthermore, deletion the SC central region proteins also results in a high frequency of unpaired homologues, indicating a crucial role for these proteins (Sherizen, et al., 2005; Takeo et al., 2011). Billmyre *et al*. used fluorescence in situ hybridisation (FISH) to examine homologue pairing in SC formed with C(3)G^{WT}, and the deletion mutants: C(3)G^{Del1}, C(3)G^{Del2} and C(3)G^{Del3}. It was determined that C(3)G^{Del2} has a very slight pairing defect, however C(3)G^{Del1}and C(3)G^{Del3} have a progressive loss of homologous chromosome pairing. In Drosophila C(3)G^{Del3}, where the SC does not assemble, it was determined that homologous pairing is reduced to 37% (Billmyre et al., 2019). These results are complementary with the observed reduced recombination in C(3)G^{Del1}and C(3)G^{Del3} mutants and suggests that a full-length SC is required for maintaining homologue pairing and successful recombination. Furthermore, these results suggest that the C(3)G may be involved in the assembly of the full-length SC, this could be similar to the experimentally determined self-assembly mechanism of SYCP1 (Dunce et al., 2018).

4.1.8. Structure-function relationship of coiled-coils

Coiled-coils are α-helical structures formed by the association of two or more amphipathic α-helices through hydrophobic interactions, thus inducing supercoiling. The amphipathic character of α-helices is achieved by the regular spacing of hydrophobic residues. This pattern is known as a heptad, where seven consecutive residues $(abcdefg)_n$ will form two turns of an α-helix and the hydrophobic residues, at positions *a* and *d*, will form a hydrophobic ridge on one side of the molecule (Crick, 1952). Additionally, residues at positions *e* and *g*, are typically charged residues that stabilise the association of α-helices through ionic interactions (Figure 4.1.13.a.). This conserved pattern allows for helices to pack together in a 'knobs-into-holes' arrangement via their hydrophobic faces which drives coiled-coil formation. A regular α-helix (undistorted α-helix) has 3.6 amino acids per turn (periodicity), meaning that the side chains of amino acids three or four residues apart are brought together and stabilised by hydrogen bonds. However, in coiled-coil structures, the periodicity is reduced to 3.5, with respect to the supercoil axis (Figure 4.1.13.b.) (Crick, 1952; McLachlan & Stewart, 1975; Truebestein & Leonard, 2016). COILS is a secondary structure prediction server used to predict the presence of coiled-coil

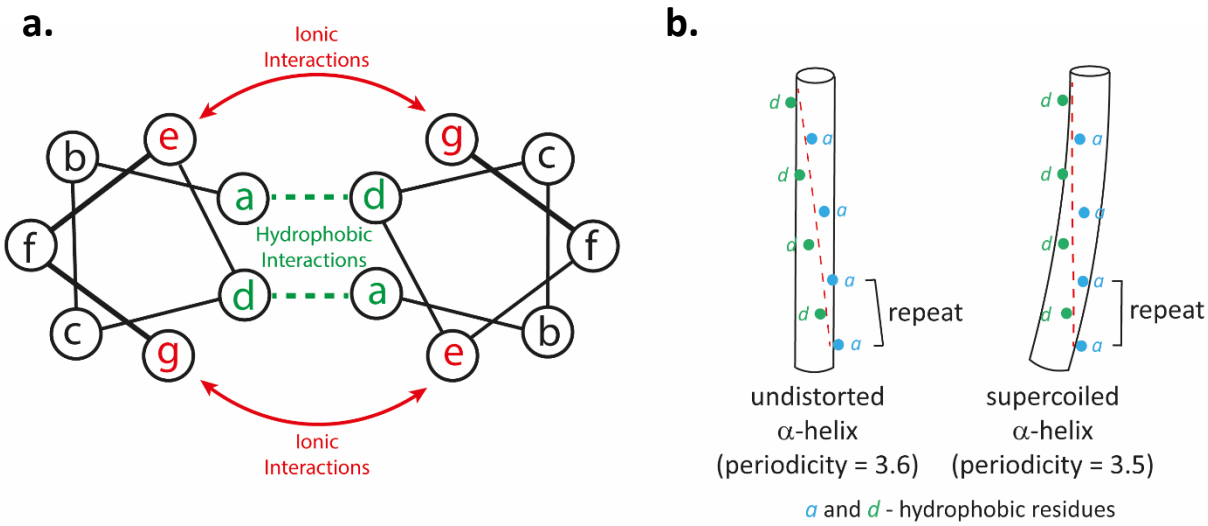


Figure 4.1.13.| Coiled-coil interactions.
a) Helical wheel plot showing the heptad repeat of a coiled-coil dimer. A canonical coiled-coil is characterised by a heptad $(abcdefg)_n$, in which hydrophobic residues are conserved at positions *a* and *d* facilitating coiled-coil formation. Amino acid residues *e* and *g* are charged residues that stabilise dimerisation through electrostatic interactions. b) Undistorted α-helices arranged side by side cannot back together due to the non-integral nature of the helix. A regular α-helix has a periodicity of 3.6. In contrast, two or more adjacent α-helices possessing the heptad repeat structure can pack together, known as supercoiling, reducing the periodicity to 3.5. Figure adapted from Truebestein & Leonard, 2016.

domains by analysing the amino acid sequence for heptad repeating units (Lupas, 1997).

Coiled-coil domains have a wide variety of biological functions including molecular spacers that separate functional domains, or as a scaffold for large macromolecular complexes. The physical properties of coiled-coil domains, including their length and flexibility, have important structural and functional properties (Truebestein & Leonard, 2016). The *T. maritma* outer membrane protein, Omp-α, contains a predicted coiled-coil domain that spans the periplasmic space. EM studies determined that Omp-α is a rod-like structure with a length of 50 nm, corresponding to the distance between inner and outer membranes, proposing its function as a molecular spacer (Engel et al., 1992). In addition, coiled-coil domains have shown to have flexibility, thereby facilitating conformational changes. Motor proteins, including Kinesin-1, possess coiled-coil domains that undergo dynamic movements that drive downstream processes (Marx et al., 2009).

The SC is the most universally conserved structure of meiosis, containing multiple proteins with predicted coiled-coil domains. Previous studies of mammalian SC protein structures of SYCP1, SYCE3, SYCE2-TEX12, SYCP3, and SYCE1 have shown that these proteins have long stretches of amphipathic α-helical regions, predicted to be coiled-coil domains. In all cases, the proteins have a typical elongated α-helical rod-like structure, forming homo-dimeric or homo-tetrameric configurations with precise lengths. These studies suggest that the coiled-coil domains may facilitate homo-oligomerisation and function as molecular spacers, as well as interaction domains for other SC proteins (Davies et al., 2012; Dunce et al., 2018; Dunne & Davies, 2019; Lu et al., 2014; Syrjänen et al., 2014). Thus, the SC proteins likely exploit the coiled-coil characteristics to define the SC geometry.

The TF proteins across multiple organisms share no sequence homology but have a highly conserved secondary structure including a predicted central coiled-coil domain. Mammalian SYCP1 consists of a central α-helical region with a clear heptad repeat pattern. SYCP1 has sequence similarity to the coiled-coil region of the motor proteins myosin and tropomyosin (McLachlan & Stewart, 1975; Meuwissen et al., 1992). Studies have shown that the TF protein functions as a physical spacer that spans across two synapsed homologous chromosomes and also governs the width of the SC (Meuwissen et al., 1992; Öllinger et al., 2005).

4.1.9. Small angle X-ray scattering (SAXS) can be utilised to characterise SC proteins.

Size exclusion chromatography in line with small angle X-ray scattering (SEC-SAXS) is a robust structural technique used to determine a low-resolution structure of proteins in solution. SEC-SAXS utilises the X-ray scattering profile of species separated by SEC to determine size and shape parameters (width and length) and flexibility of a protein. Transformation of the X-ray scattering data yields the real-space paired distribution, *P(r)*, profile; a histogram of the interatomic vectors within the molecule, which is used to estimate the maximum dimension (*Dmax*) of the protein. X-ray scattering is proportional to the interatomic distances in the sample, therefore large globular molecules, such as maltose binding protein (MBP), will have a predominant scattering profile compared to elongated proteins.

Oligomeric coiled-coil proteins can adopt parallel or anti-parallel configurations. By utilising SEC-SAXS analysis it is possible to predict the orientation of the α-helices. The protein of interest is analysed as N- and C- terminal MBP-fusions, thereby exploiting the strong scattering of the MBP-affinity tag compared to the coiled-coils. From the *P(r)* distribution profiles it is possible to determine the relative position of MBP-molecules, thereby predict the N- and C-terminal orientation within the coiled-coil fusion protein. Both N- and C-terminal MBP fusion constructs have only short interatomic distances in parallel coiled-coils and long interatomic distances in anti-parallel coiled-coils. An MBP-tag at both termini is used as a positive control and have both inter-long and short distances for both parallel and anti-parallel orientated coiled coils (Figure 4.1.14.a.).

This technique has previously been used to determine the orientation of helices of mammalian SC proteins, including SYCE1 and SYCP1 (Dunce et al., 2018; Dunne & Davies, 2019). The central element SC protein, SYCE1, has an N-terminal α-helical structural core (amino acid residues 25-179) predicted to be a coiled-coil domain. The real space *P(r)* distributions of MBP-SYCE1 and SYCE1-MBP demonstrate the presence of a long inter-MBP peak at 125-175 Å and lack a short inter-MBP peak. In contrast, MBP-SYCE1-MBP has both a short and long inter-MBP peak at ~70 Å and 125-175 Å, respectively, demonstrating that SYCE1 structural core is in anti-parallel orientation (Figure 4.1.14.b.).

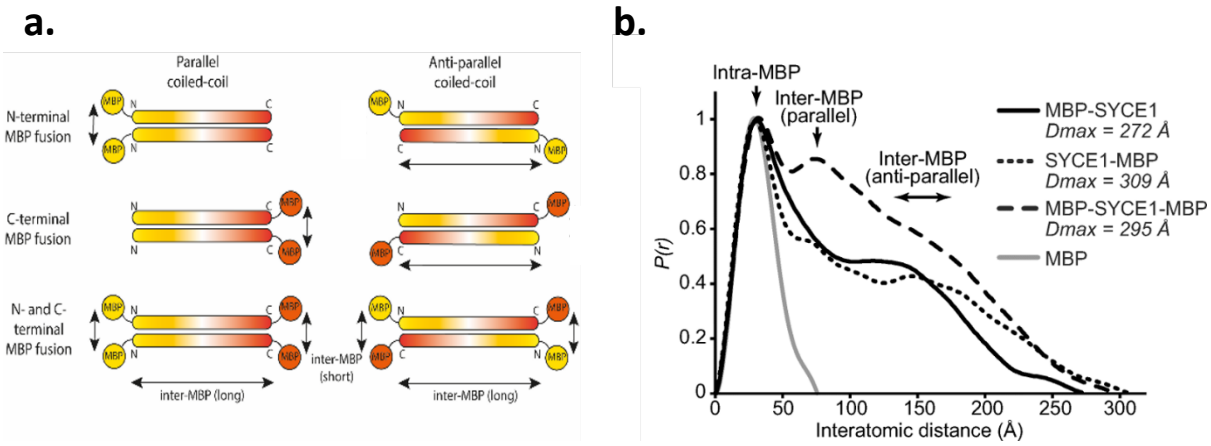


Figure 4.1.14.| SAXS analysis determining the helical orientation of SYCE1 core.
a) A range of MBP-tagged SYCE1$_{25-179}$ fusions were generated. Utilising the globular nature of the MBP-affinity tag it is possible to determine the distance between MBP-tags within the SYCE1 dimer. In both the single MBP-fusions for a parallel orientated dimer only short inter-MBP distances should be observed. In contrast, for anti-parallel coiled-coil dimers single MBP-fusions demonstrate both long and short inter-MBP distances. N- and C-terminal double MBP-fusion construct is used as a positive control. In both parallel and anti-parallel configurations double MBP-fusion possess both long and short inter-MBP distances. b) SEC-SAXS *P(r)* distributions of MBP-tagged SYCE1 fusions, demonstrating anti-parallel orientation of the coiled-coil domain. The *P(r)* distribution profiles of MBP-SYCE1 (black), SYCE1-MBP (black narrow dashes) and MBP-SYCE1-MBP (black wide dashes) all possess a peak at ~70 Å and 125-175 Å demonstrating that SYCE1 structural core is in anti-parallel orientation. Maximum dimensions (*Dmax*) are indicated and the positions of intra-MBP and both parallel and anti-parallel inter-MBP peaks are highlighted. Figure b) taken from Dunne and Davies, 2019.

4.1.10. Chapter aims

The literature has shown that the proteins constituting the SC are quite diverse across organisms, despite having a general conserved tripartite structure. The overall structure and organisation of *D. melanogaster* SC has been determined by super resolution microscopy. However, individual proteins have yet to be characterised *in vitro* (Cahoon et al., 2017). This project was initiated as a collaboration with Prof. Scott Hawley (Stowers Institute for Medical Research, Kansas City) to gain a biophysical and structural understanding of the *Drosophila* transverse filament protein (C(3)G) and to gather complementary biophysical data to the in-frame deletions determined by Dr. Kathleen Billmyre, summarised in 4.1.7. (Billmyre et al., 2019). In addition, we set out to biophysically characterise the predicted coiled-coil and the globular C- and N-terminal domains. *In vivo* immunogold localisation and expansion super resolution microscopy studies have shown that C(3)G shares a similar localisation

pattern within the SC as its *S. cerevisiae* and mammalian counterparts. It is assumed that C(3)G has shares a conserved structure to Zip1 or SYCP1 but currently there is no literature providing an insight into the structure of C(3)G *in vitro*. Therefore, I set out to explore the domain structure of C(3)G in detail and to ultimately deduce a complete molecular understanding of the *Drosophila* transverse filament protein.

4.2. RESULTS

4.2.1. Purification and characterisation of C(3)G structural core

C(3)G is 744 amino acids long and predicted to have an α-helical structure from residues 112-650 (Figures 4.2.1. and 4.2.2.). This region was cloned and expressed as a N-terminal MBP-fusion. The MBP-C(3)$G_{112-650}$ construct expressed well but was highly insoluble, suggesting that the construct needed to be optimised. An N-terminal region of 114 amino acids was removed (C(3)$G_{226-650}$ and was tested for its protein expression and solubility. In comparison to the MBP-C(3)$G_{112-650}$, MBP-C(3)$G_{226-650}$ yielded highly soluble material, herein termed the helical core domain.

The MBP-C(3)$G_{226-650}$ protein was purified through sequential amylose affinity and anion exchange chromatography. The MBP tag was cleaved by incubating with TEV-protease and then an anion exchange chromatography step was performed to separate cleaved C(3)$G_{226-650}$ from the MBP-tag and TEV protease, followed by size exclusion chromatography (SEC) as a final purification step. The purification process is summarised in Figure 4.2.3.a.

The in-solution technique, SEC-MALS, was used to determine the absolute molecular mass of the His_6-MBP-C(3)$G_{226-650}$ and the MBP free species to determine the oligomeric state. Both the MBP fusion and MBP free C(3)$G_{226-650}$ elute as a single peak, with molecular masses of 176 and 97.3 kDa, respectively, corresponding to a dimer in both instances (dimeric theoretical molecular weights – 189 and 99.6 kDa, respectively) (Figure 4.2.3.b and c).

Far-UV CD spectroscopy was used to quantify the secondary structure composition of C(3)$G_{226-650}$. Analysis revealed a characteristic α-helical spectrum with negative peak of similar magnitude at 208 and 222 nm and a positive peak at 193 nm (Figure 4.2.3.d.). Deconvolution by DichroWeb estimated an α-helical content of 77%, meaning 327 out of the 425 amino acid sequence comprise the α-helical region. This result suggest that the α-helical core domain of C(3)G consists of discontinuous α-helical domains instead of a rigid coiled-coil domain.

CD spectroscopy was also utilised to determine the thermal stability of C(3)G$_{226-650}$ by tracking the helical signal at 222 nm between 5 and 95°C. A two-step co-operative unfolding event of C(3)G$_{226-650}$ revealed a melting temperature (T_m) of 37.4°C (Figure 4.2.2.e.).

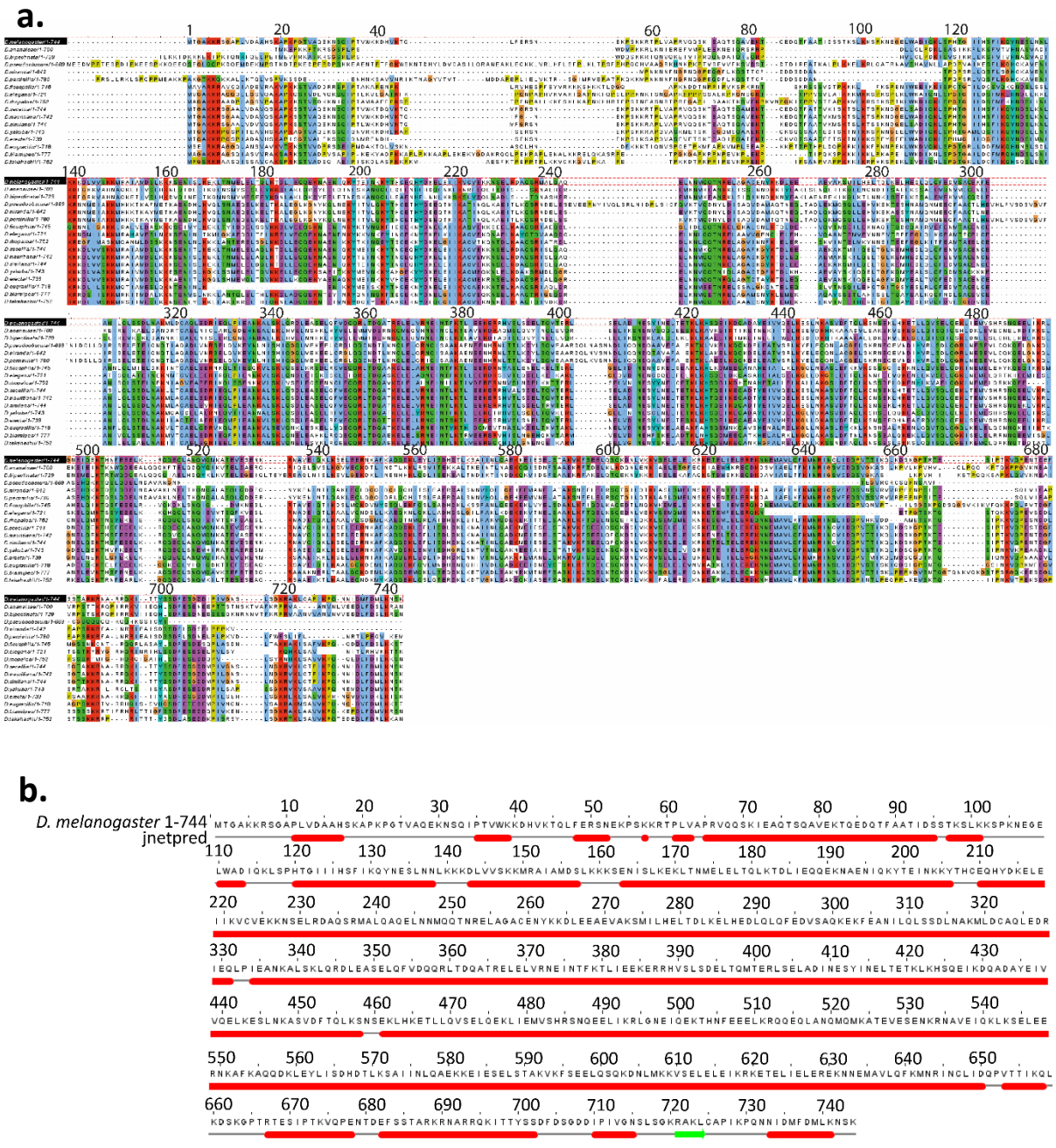


Figure 4.2.1.| C(3)G amino acid sequence alignment
a) C(3)G multiple sequence alignment generated by MUSCLE of 18 *Drosophila* orthologues. Sequence analysis was visualised in Jalview and amino acids are coloured by conservation with labelled amino acid positions corresponding to *D. melanogaster* C(3)G sequence. b) Secondary structure prediction of *D. melanogaster* C(3)G performed by JNetPRED (Drozdetskiy et al., 2015). Alpha helices (red rods), beta sheets (green arrow) and unstructured regions (grey line).

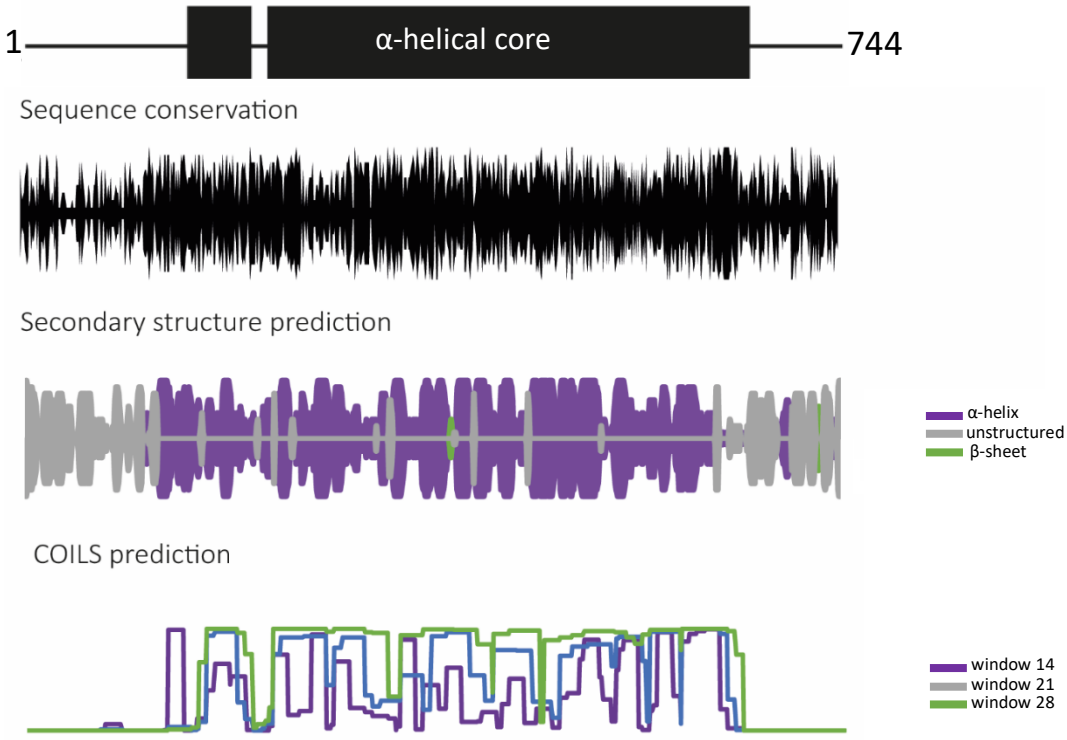


Figure 4.2.2.| Sequence analysis of C(3)G
Aligned sequence schematics showing the domain structures, sequence conservation, secondary structure and COILS predictions of C(3)G. Sequence conservation scores were determined per residue by Consurf using the same manually created sequence alignment of C(3)G. Secondary structure prediction was calculated by JPred4 using an alignment of the same 18 sequences (Apendix 1). α-helix in purple, β-sheet in green and unstructured in grey (Drozdetskiy et al., 2015). Coiled-coil conformation prediction was calculated by COILS which compared the C(3)G sequence to a database of known parallel two-stranded coiled-coils and derived a similarity score in a 14,21 or 28 amino acid window. (Lupas et al. , 1991)

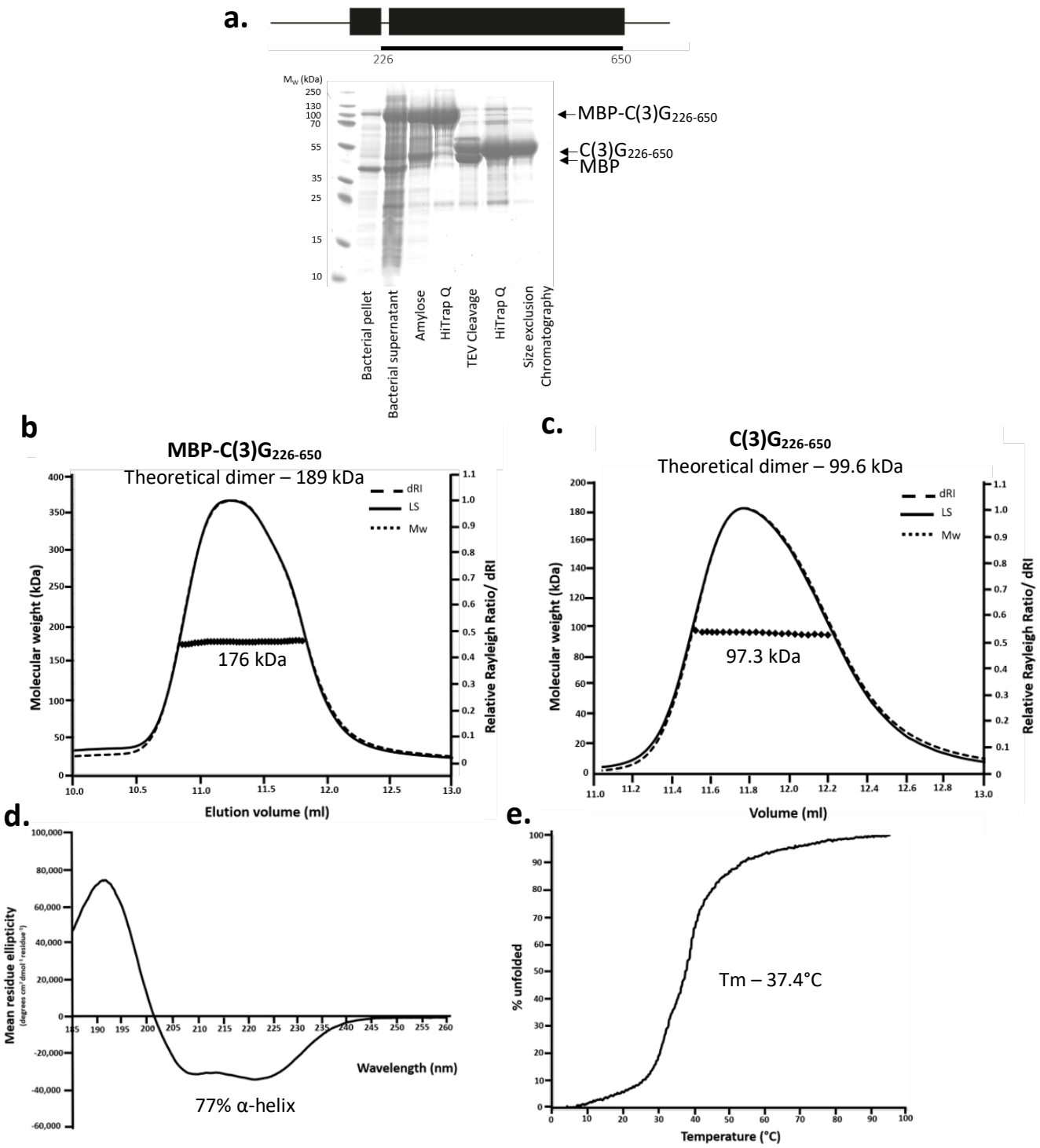


Figure 4.2.3.| Purification and structural analysis of C(3)G$_{226-650}$

a) SDS-PAGE showing the purification summary of C(3)G$_{226-650}$ through sequential amylose affinity and anion exchange chromatography. N-terminal MBP-tag was removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. b-c) SEC-MALS analysis of MBP- C(3)G$_{226-650}$ and cleaved C(3)G$_{226-650}$ elute as a single peak with experimental molecular weights of 176 and 97.3 kDa, respectively, (theoretical dimer – 189 and 99.6 kDa) indicating that C(3)G$_{226-650}$ core domain dimerises in solution. d) Far UV CD wavelength scan between 260-185nm of C(3)G$_{226-650}$ shows a typical α-helical trace. Deconvolution of the data estimates the secondary structure to be 77% α-helical. f) CD thermal denaturation measured at 222nm between 4 and 95°C, estimated a melting temperature of 37.4°C.

SEC-SAXS analysis is a powerful tool to characterise the size and shape of elongated coil-coils. The averaged SAXS profile of C(3)$G_{226-650}$ shows a steep gradient within the low q region (small angle and low resolution region known as the Guinier region), indicating an elongated structure (Figure 4.2.4.a). Fourier transform of the scattering curve allows for calculation of the pairwise distribution profile (*P(r)* curve). The *P(r)* distribution shows a positive skew with a shape corresponding to a profile of an elongated rod structure. The maximum dimension (*Dmax*) of C(3)$G_{226-650}$ was determined to be 680 Å (Figure 4.2.4.b.). The experimental determined length of a coiled-coil construct (*Dmax*) can be compared with the theoretical sequence length estimating 1.5 Å per residue (Lupas & Gruber, 2005). The determined *Dmax* is within the range of the theoretical length of an extended α-helical coiled-coil of this length (637.5 Å). The radius of gyration (*Rg*) was determined as 151 Å, also consistent for an elongated structure (Figure 4.2.4.c.). Determination of the cross-sectional radius of gyration (*Rc*) reveals the thickness of elongated molecules, which can be used to calculate the number of helices within the coiled-coil. A typical dimeric coiled-coil has a measured *Rc* of ~8 Å and a tetrameric coiled-coil of ~10 Å. The *Rc* of the C(3)$G_{226-650}$ dimer was observed to be 9 Å (Figure 4.2.4.d.), consistent with a dimeric coiled-coil (Solari & Moses, 1973; Truebestein & Leonard, 2016).

The real space SAXS data was utilised for *ab initio* modelling to create a single phase dummy atom model using DAMMIN software (Atsas suite). The low resolution molecular envelope of C(3)$G_{226-650}$ suggests an elongated rod-like molecule typical of a coiled-coil domain, with a *Dmax* of 680 Å (Figure 4.2.4.e.).

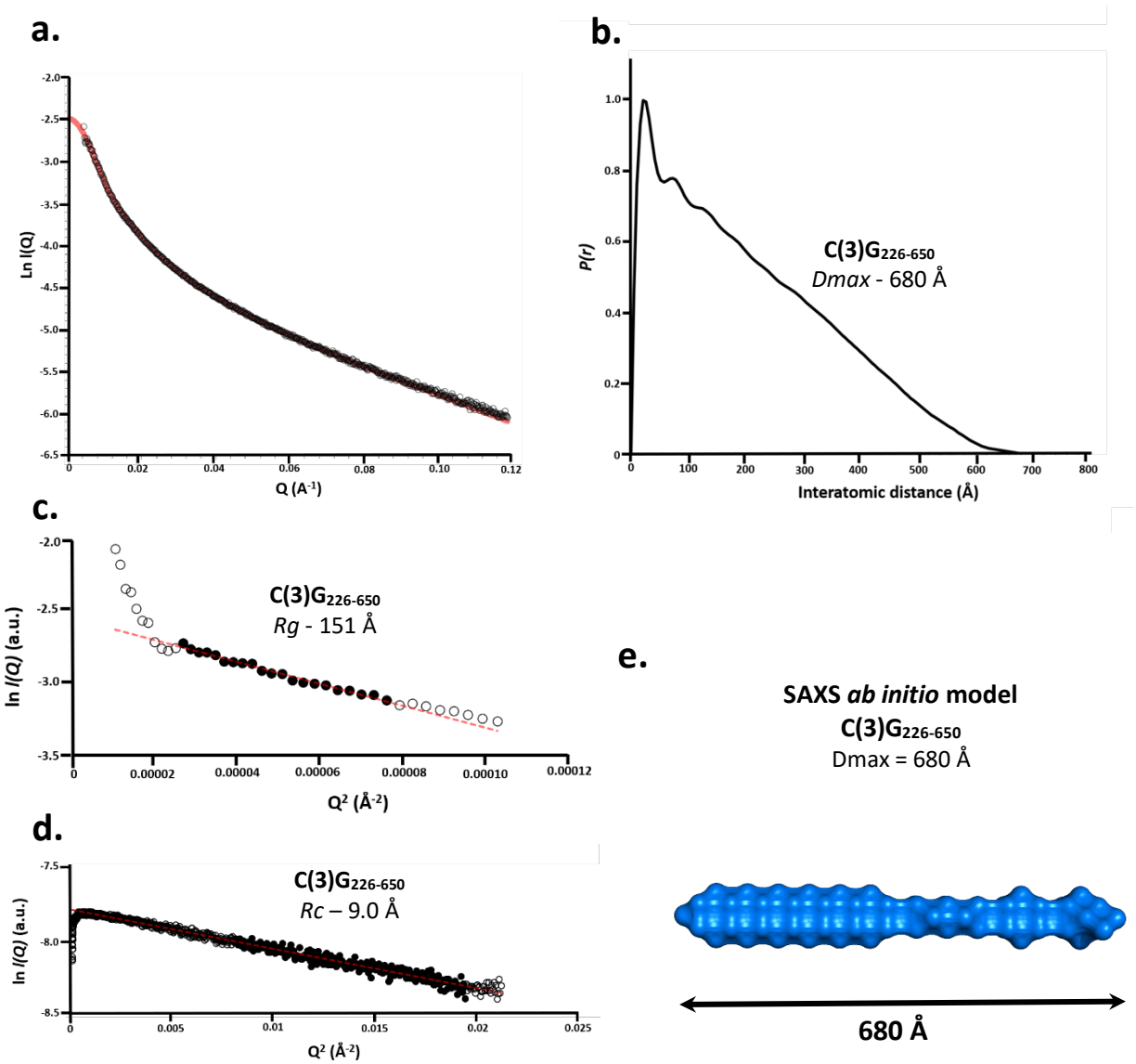


Figure 4.2.4.| SEC-SAXS analysis of C(3)G$_{226-650}$

a) Averaged SAXS profile of C(3)G$_{226-650}$, with the fit used for *P(r)* distribution in red. b) *P(r)* distribution of C(3)G$_{226-650}$ showing a maximum dimension (*Dmax*) of 680 Å. c) Guinier analysis determined a radius of gyration (*Rg*) value of 151 Å. The real space *Rg* closely matches the Guinier analysis *Rg* value of 175 Å and 151 Å, respectively. Clear circles represent the complete data set and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3). d) Guinier analysis determined the radius of gyration of the cross-section (*Rc*) of C(3)G$_{330-459}$ to be 9.0 Å. Clear circles represent the complete data and solid circles represent the region used for the fit. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3). e) SAXS *ab initio* DAMMIF model of C(3)G$_{226-650}$ presented as a molecular envelope with length of 650 Å corresponding to the determined *Dmax*.

4.2.2. C(3)G forms parallel dimeric assemblies in solution

The assembly and orientation of predicted coiled-coil α-helical domains have fundamental roles in their protein function (Park, 2020). In order to determine the orientation of helices within the coiled-coil domains of human SC proteins, including SYCP1 and SYCE1, the Davies' lab produced a method utilising small angle x-ray scattering (SAXS), as described in 4.1.9 (Dunce et al., 2018; Dunne & Davies, 2019). This entails analysing the distance distribution profile (*P(r)*) of a series of MBP-fused protein constructs, determining helical positioning by measuring the distance between the dominant globular MBP-affinity tag peaks. In parallel configurations, individual N- and C-terminal fusions demonstrate short inter-MBP distances, whilst an anti-parallel configuration demonstrate long inter-MBP distances. The double C- and N-terminal MBP-fusion is used as a positive control and demonstrates both long and short inter-MBP distances in both configurations (Figure 4.2.5.a.). Through interpretation of these results, it is possible to predict the relative orientation of helices.

C(3)G$_{226\text{-}650}$ was cloned with the addition of a non-cleavable C-terminal MBP-solubility tag allowing for the purification of C-terminal and double MBP fusions of C(3)G$_{226\text{-}650}$ and an N-terminal MBP-C(3)G$_{226\text{-}650}$ (Figure 5.2.5.b.). SEC-MALS analysis was performed to confirm that the addition of tags did not disrupt dimer formation. N-terminal MBP-tagged C(3)G$_{226\text{-}650}$ (MBP-C(3)G), C-terminal MBP-tagged C(3)G$_{226\text{-}650}$ (C(3)G-MBP) and double MBP tagged C(3)G$_{226\text{-}650}$ (MBP-C(3)G-MBP) all elute as a single peak, with determined molecular weights of 263, 176 and 169 kDa, respectively, corresponding to dimer formation in solution (theoretical dimers – 232, 189 and 186 kDa) (Figure 4.2.5.c.).

For SEC-SAXS analysis the measured intensities, *I(q)*, is plotted as log10 of *I(q)* as intensities decay quickly, furthermore the larger the object the faster the *I(q)* decay, decreasing the quality of data within the low q range. The low-q region of the scattering curve is characteristic of the overall dimension of the analysed particles, including the paired distribution, *p(R)* profile. This suggests analysis of MBP-fusion protein constructs of C(3)G may be too large for high quality SAXS analysis.

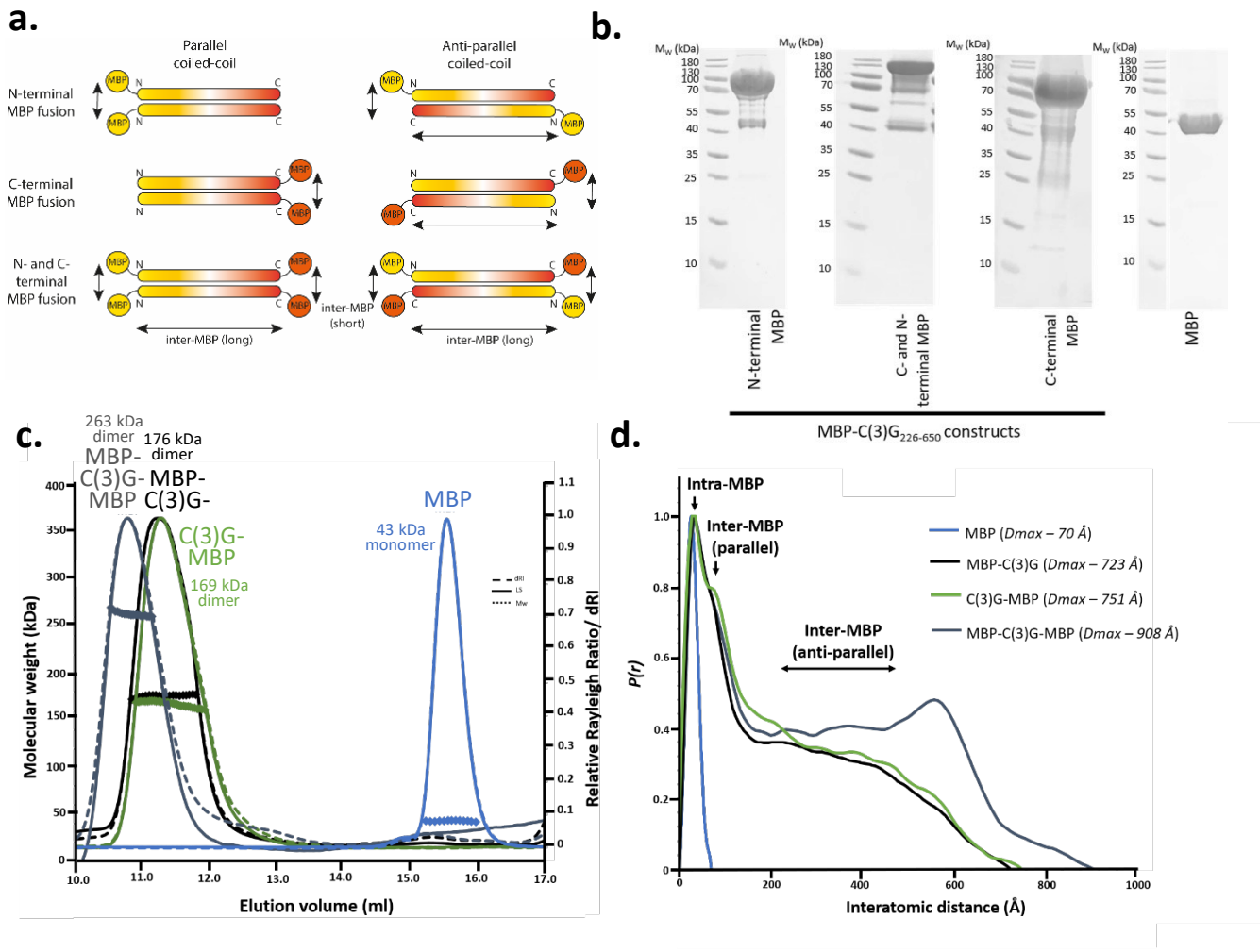


Figure 4.2.5.| SEC-SAXS analysis determined the parallel orientation of C(3)G$_{226-650}$
a) Schematic showing the inter-MBP distances for C-terminal, N-terminal, and double MBP-fusions for parallel and anti-parallel orientations. b) SDS-PAGE analysis showing the final purified MBP-tagged C(3)G$_{226-650}$ constructs and MBP. c) SEC-MALS analysis of C(3)G$_{226-650}$ fusions, indicating that the additional MBP-tags does not disrupt dimer formation. MBP-C(3)G-MBP (grey), MBP-C(3)G (black) and C(3)G-MBP (green) showed molecular weights of 263, 176 and 169 kDa, respectively (theoretical dimers – 232, 189 and 186 kDa) and the MBP monomer (43 kDa) is shown in blue. c) SEC-SAXS *P(r)* distributions of C(3)G$_{226-650}$ MBP-fusions demonstrating that C(3)G coiled-coil domain forms an parallel dimeric configuration. MBP-C(3)G (black) and C(3)G-MBP (green) have maximum dimensions of 723 and 751 Å, respectively and both possess an inter-MBP peak at ~ 80 Å. MBP-C(3)G-MBP (grey) has a maximum dimension of 908 Å and has both a short and long inter-MBP peak at ~80 and ~600 Å. The MBP control has a maximum dimension of 70 Å (blue).

The SEC-SAXS real space analysis of MBP-C(3)G, C(3)G-MBP and MBP-C(3)G-MBP estimated the *Dmax* to be 723, 751 and 908 Å, respectively. These are comparable with the SAX data collected for C(3)G$_{226-650}$ with extra ~70 Å due to the addition of the MBP-tag (Figure 4.2.5.d.). All three MBP-fused C(3)G constructs possess an inter-MBP peak at ~ 80 Å and the double MBP-tagged construct also has

a second inter-MBP peak at ~ 600 Å. Therefore, the single MBP-C(3)G construct possesses only a short (80 Å) inter-MBP peak and the double MBP-C(3)G has both short and long inter-MBP peaks. These observations suggest that the dimeric α-helical chains of $C(3)G_{226-650}$ are in a parallel configuration.

4.2.3. Dissecting the core structure of $C(3)G_{226-650}$

We have shown that the α-helical core region of C(3)G forms a stable dimeric coiled-coil structure in solution. Moreover, we sought to determine the minimal protein sequence required to form a stable dimer, with the possibility to use the optimised structure as a crystal target. Both N-terminal, C-terminal and N- and C-terminal truncation constructs were designed using α-helical secondary structure predictions as well as the limited sequence conservation (Figures 4.2.1. and 4.2.2.).

$C(3)G_{330-650}$

A further N-terminally truncated construct, removal of another 104 amino acids, $C(3)G_{330-650}$, was expressed as a MBP-fusion and purified in the same manner as described above for $C(3)G_{226-650}$. An overview of the purification steps is given in Figure 4.2.6.a. To determine that this truncation of C(3)G did not disrupt the dimer in solution, we analysed both MBP-fusion and cleaved $C(3)G_{330-650}$ by SEC-MALS. MBP-$C3G_{330-650}$ eluted as a single peak with a molecular mass of 153 kDa, corresponding to a dimer (theoretical dimer – 165 kDa) (Figure 4.2.6.b.). Cleaved $C(3)G_{330-650}$ was applied to SEC-MALS at 13mg/ml and eluted as multiple peaks over a broad elution range, however a molecular mass of 72.4 kDa is constant across the peak suggesting that the sample is a monodisperse dimer (theoretical dimer – 76 kDa). Moreover, loading at lower concentrations, $C(3)G_{330-650}$ eluted as a single symmetrical bell-shaped curve (Figure 4.2.6.c.). Irregularity of the SEC profile at the higher concentration is likely explained by overloading the column due to the elongated rod-like shape.

Far UV CD spectroscopy revealed that $C(3)G_{330-650}$ is almost entirely α-helical, with deconvolution estimating an α-helical content of 88% (282 helical residues out of 320) (Figure 5.2.6.d.) The CD

thermal melt profile shows a steady progression of protein unfolding, with a T_m of 38.6 °C, consistent with the longer $C(3)G_{226\text{-}650}$ construct (Figure 5.2.6.e.)

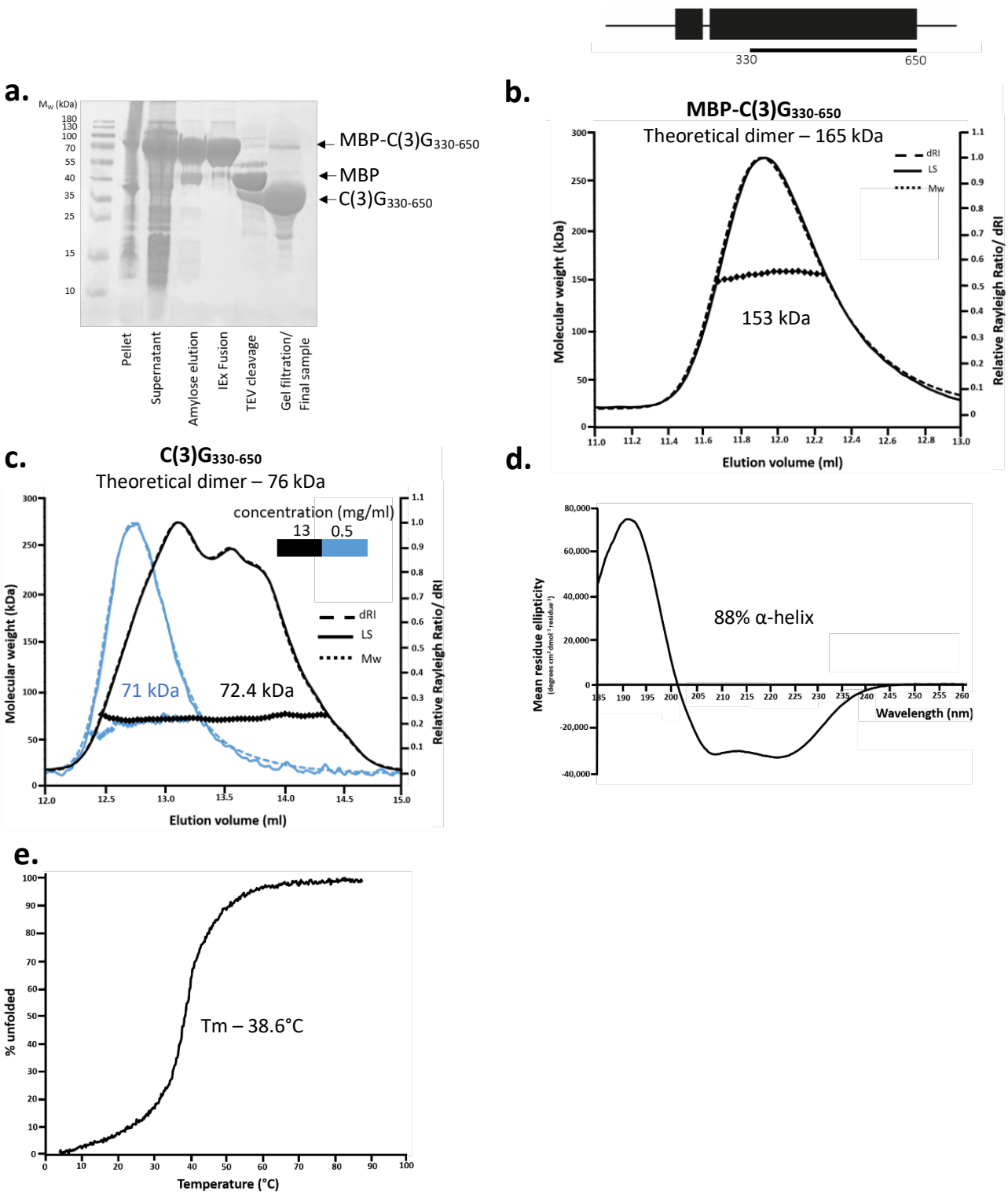


Figure 4.2.6.| Purification and biophysical analysis of C(3)G$_{330-650}$
a) SDS-PAGE showing the purification summary of C(3)G$_{330-650}$ through sequential amylose affinity and anion exchange chromatography. The N-terminal MBP-tag was removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. b) SEC-MALS analysis of MBP-C(3)G$_{330-650}$ eluted as a peak with a molecular weight of 153 kDa corresponding to a dimer (theoretical dimer - 165 kDa). c) SEC-MALS analysis of C(3)G$_{330-650}$ loaded at 13 mg/ml (black) eluted as a broad peak, with a constant molecular weight of 72.4 kDa, corresponding to a dimer (theoretical dimer - 76 kDa). Loading at a lower concentration, 0.5 mg/ml (blue), produced a much more symmetrical bell-shaped peak with a molecular weight of 71 kDa. d) Far UV CD wavelength scan between 260-185nm of C(3)G$_{330-650}$ shows a typical α-helical trace. Deconvolution of the data estimates the secondary structure to be 88% α-helical e) CD thermal denaturation measured at 222 nm between 4 and 95°C, estimated a melting temperature of 38.6 °C.

SEC-SAXS analysis of C(3)G$_{330-650}$ revealed characteristics of a typical elongated structure. The radius of gyration (*Rg*) and radius of gyration of the cross section (*Rc*) were determined as 129 Å and 8.2 Å, respectively, fitting with a dimeric coiled coil (Figures 5.2.7.a. and b.). Inspection of the distance distribution profile of C(3)G$_{330-650}$ shows elongated characteristics including a skewed distribution and a tailing profile (Figure 5.2.7.c.). Real space analysis estimates the maximum dimension (*Dmax*) to be 560 Å.

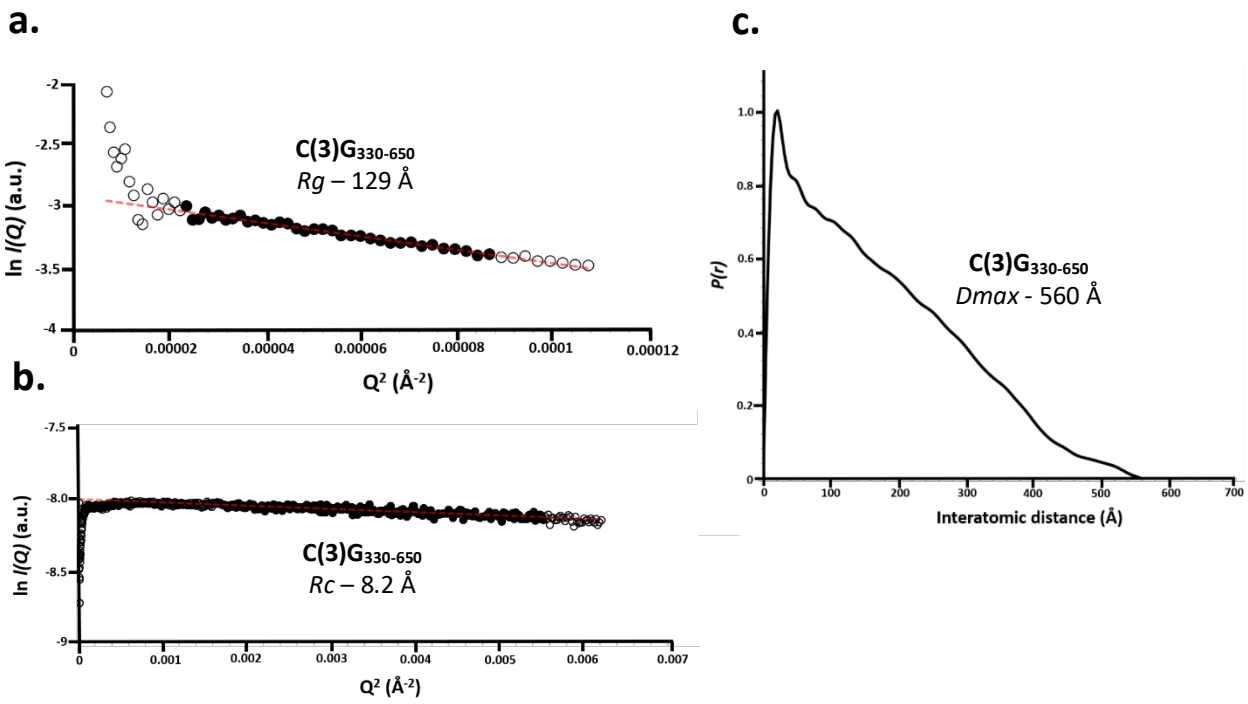


Figure 4.2.7.| SEC-SAXS analysis of C(3)G$_{330-650}$
a) Guinier analysis determined a radius of gyration (*Rg*) value of 129 Å. The real space *Rg* closely matches the Guinier analysis *Rg* value of 146 Å and 129 Å, respectively. Clear circles represent the complete data set and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3). b) *P(r)* distribution of C(3)G$_{330-650}$ showing a maximum dimension of 650 Å. c) Guinier analysis determined the radius of gyration of the cross-section (*Rc*) of C(3)G$_{330-650}$ to be 8.2 Å. Clear circles represent the complete data and solid circles represent the region used for the fit. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3).

C(3)G$_{470-650}$

We truncated a further 140 amino acid residues from the N-terminus, generating the truncated protein C(3)G$_{470-650}$, and analysed this new construct as described above. MBP-C(3)G$_{470-650}$ was purified using the same steps, summarised in Figure 4.2.8.a. SEC-MALS analysis of MBP-tagged and cleaved C(3)G$_{470-650}$ revealed absolute molecular masses of 117 and 37 kDa, respectively, indicating dimeric formation is retained (theoretical dimer - 132.2 and 43.2 kDa, respectively) (Figures 4.2.8.b. and c.). Quantification of C(3)G$_{470-650}$ secondary structure by Far-UV CD spectroscopy determined a characteristic α-helical spectrum with negative peaks at 208 and 222 nm and a positive peak at 193 nm (Figure 4.2.8.d.). Deconvolution estimated the α-helical content to be 83 %, equating to 149 helical

residues out of 180, consistent with the longer C(3)G$_{226-650}$ construct. Analysis of the thermal denaturation spectrum at 222 nm revealed a cooperative one-step unfolding event and a Tm of 34.8 °C (Figure 4.2.8.e.). Overall, this suggests that truncating the original C(3)G coiled-coil domain by 244 amino acids at the N-terminus does not affect dimer formation or protein stability as the results are consistent with C(3)G$_{226-650}$, despite removing a large part of the α-helical region.

SEC-SAXS analysis of C(3)G$_{470-650}$, as expected, revealed characteristics of an elongated rod-like molecule with an *Rg* and *Rc* of 63.65 and 9.1 Å, respectively (Figures 4.2.9.a. and b.). The *P(r)* curve has a profile typical of an elongated structure, with a *Dmax* of 290 Å (Figure 4.2.9.c.). In combination these results match the theoretical parameters of a coiled-coil of this length. We utilised the SAXS data to carry out low resolution *ab initio* modelling of C(3)G$_{470-650}$ using DAMMIF. The modelled envelope demonstrates an elongated rod-like molecule typical of a coiled-coil domain, with a *Dmax* of 290 Å (Figure 4.2.9.d.).

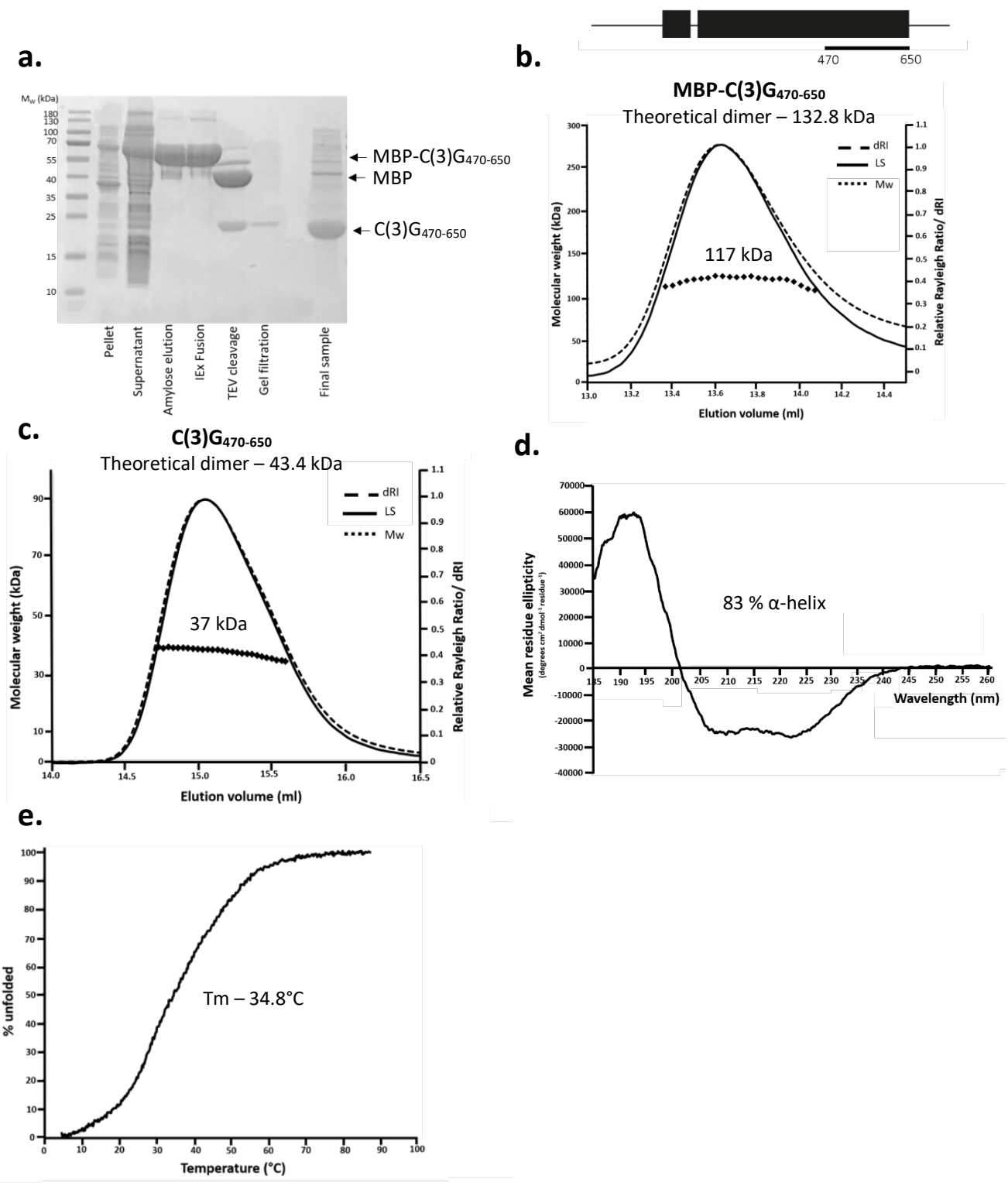


Figure 4.2.8.| Purification and biophysical analysis of C(3)G$_{470-650}$

a) SDS-PAGE showing the purification summary of C(3)G$_{470-650}$ through sequential amylose affinity and anion exchange chromatography. N-terminal MBP-tag was removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. b-c) SEC-MALS analysis of b) MBP-C(3)G$_{470-650}$ and c) cleaved C(3)G$_{470-650}$ determined a single peak with a molecular weights of 117 and 37 kDa, respectively corresponding to a dimer (theoretical dimer – 132. and 43.4 kDa, respectively). d) Far UV CD wavelength scan between 260-185nm of C(3)G$_{330-650}$ shows a typical α-helical trace. Deconvolution of the data estimates the secondary structure to be 83 % α-helical e) CD thermal denaturation measured at 222nm between 4 and 95°C, estimated a melting temperature of 34.8 °C.

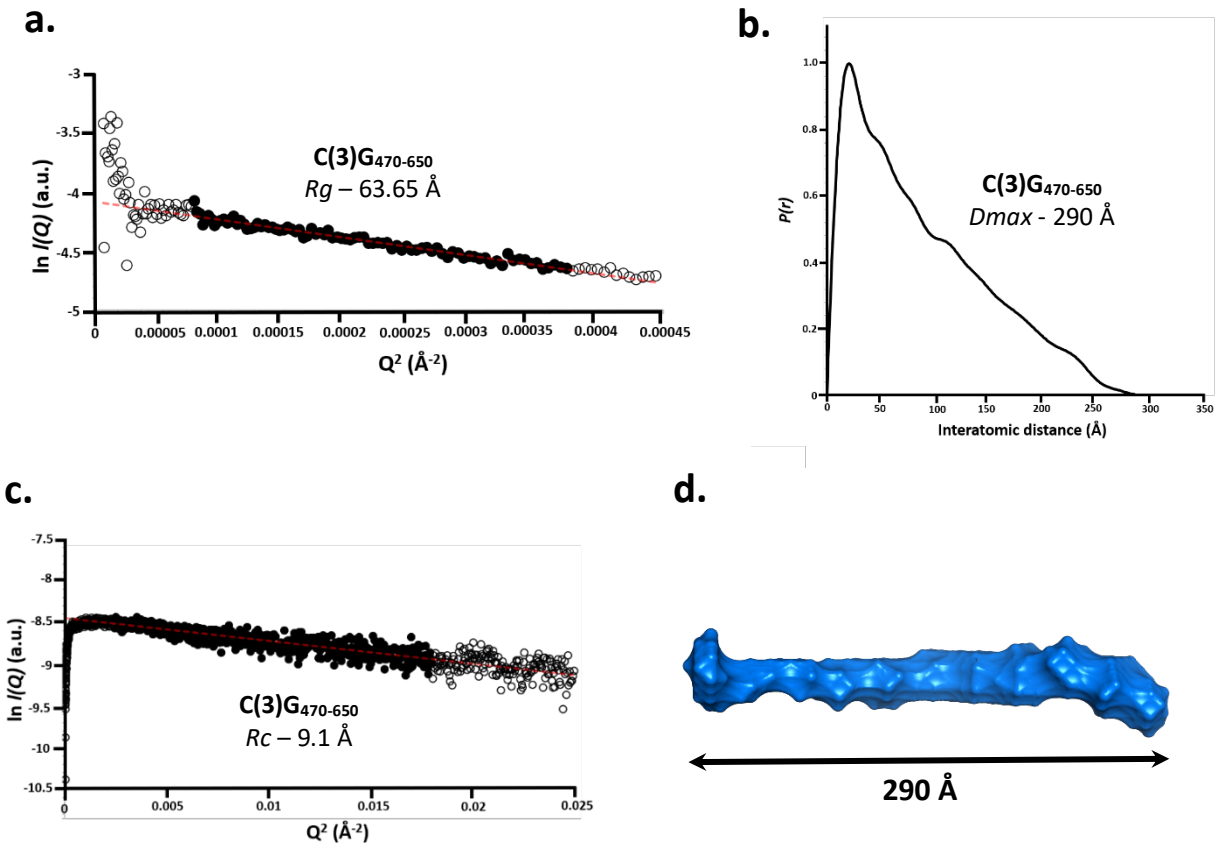


Figure 4.2.9.| SEC-SAXS analysis of C(3)G$_{470-650}$
a) Guinier analysis determined a radius of gyration (*Rg*) value of 63.65 Å. The real space *Rg* closely matches the Guinier analysis *Rg* value of 75.2 Å and 63.65 Å, respectively. Clear circles represent the complete data set and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3). b) *P(r)* distribution of C(3)G$_{470-650}$ showing a maximum dimension of 290 Å. c) Guinier analysis determined the radius of gyration of the cross-section (*Rc*) of C(3)G$_{470-650}$ to be 9.1 Å. Clear circles represent the complete data and solid circles represent the region used for the fit. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3). d) SAXS *ab initio* model of C(3)G$_{470-650}$ using the modelling program, DAMMIF determined a molecular envelope with a *Dmax* of 290 Å , as observed by the paired distribution profile.

C(3)G$_{534-650}$ and C(3)G$_{550-650}$

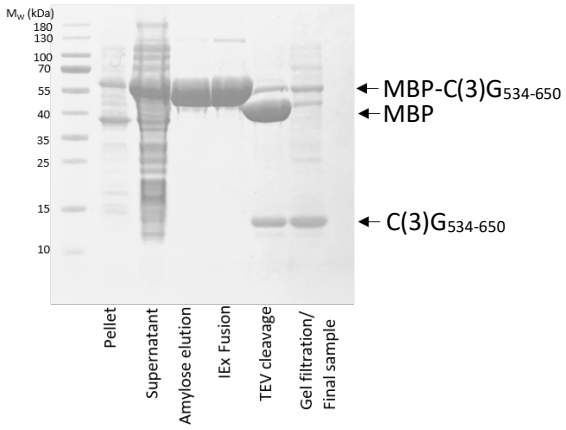
Two further C(3)G N-terminal truncations, C(3)G$_{534-650}$ and C(3)G$_{550-650}$, were designed based on secondary structure prediction. The shorter construct, C(3)G$_{550-650}$, was designed to be an optimised X-ray crystallography target; a shorter more compact structure may crystallise more easily.

Both constructs were expressed as MBP-fusions and purified using the same steps as the longer constructs, summarised in figures 4.2.10.a and b. We utilised SEC-MALS analysis to determine the absolute molecular mass of both C(3)G$_{534-650}$ and C(3)G$_{550-650}$ as MBP-fusions and cleaved MBP free

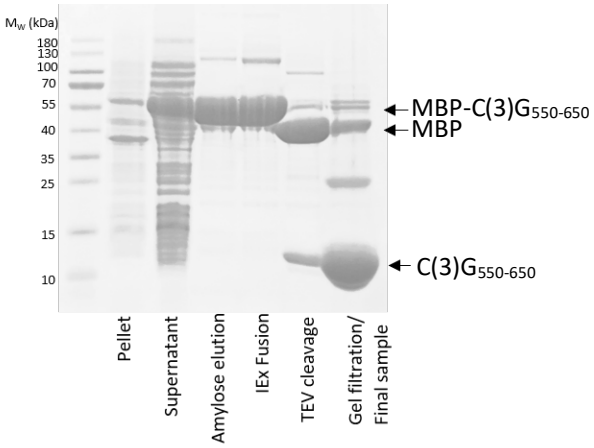
proteins. MBP-C(3)$G_{534-650}$ eluted as a broad asymmetrical peak with the second half of the peak sloping, showing deviation between the dRI and light scattering (LS) (Figure 4.2.10.c.). The molecular weight decreases across the peak, starting at 97.9 kDa and decreasing to 59.5 kDa (shown by orange arrows), indicating a mixture of dimer and monomer (theoretical monomeric molecular weight 58.7 kDa). This suggests that C(3)$G_{534-650}$ could be unstable as a dimer and subsequently dissociating. The SEC-MALS profile of the cleaved material showed a similar peak profile, with a broad single peak which indicated a molecular weight across the peak of 21.3 kDa (Figure 4.2.10.e.). The resultant molecular weight, 21.3 kDa is midway between a monomer and dimer (theoretical monomer – 14.1 kDa), so we cannot confidently determine the oligomeric state due to the small differences between the theoretical molecular weights.

SEC-MALS analysis of the slightly shorter construct, C(3)$G_{550-650}$, was carried out using a Superdex 200 increase column, instead of the Superose column which was used for the other C(3)G constructs. The Superdex column enables better resolution, therefore the opportunity to separate out mixed species. MBP-C(3)$G_{550-650}$ eluted as three distinct peaks with molecular weights of 232.5, 127.5 and 62.2 kDa corresponding to a mixture of tetrameric, dimeric and monomeric species, respectively (theoretical monomer – 56.8 kDa) (Figure 4.2.10.d). SEC-MALS analysis of cleaved C(3)$G_{550-650}$ revealed two peaks with molecular weights of 41.7 and 12.1 kDa (Figure 4.2.10.f.). The latter, and the predominant, peak corresponds to a monomeric species (theoretical monomer – 12.2 kDa) and the small 41.1 kDa peak could be C(3)$G_{550-650}$ forming a tetramer, again suggesting a mixture of oligomeric states.

a.



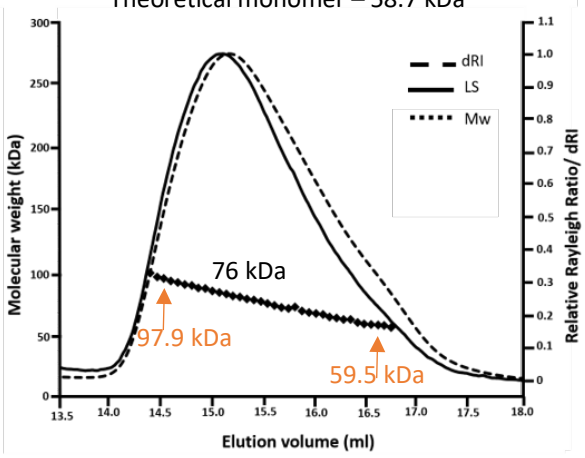
b.



c.

MBP-C(3)G534-650

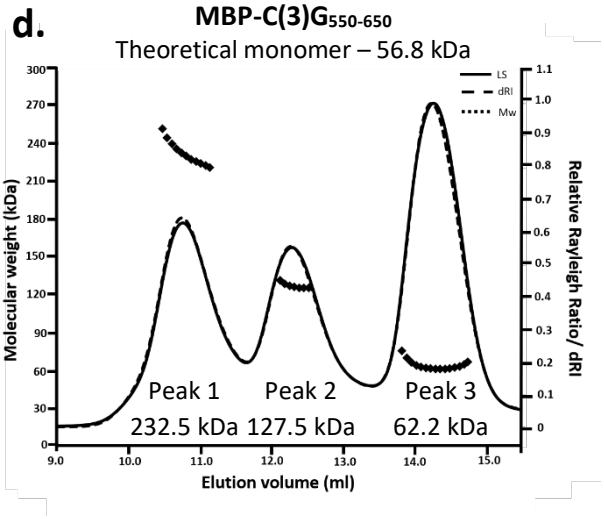
Theoretical monomer – 58.7 kDa



d.

MBP-C(3)G550-650

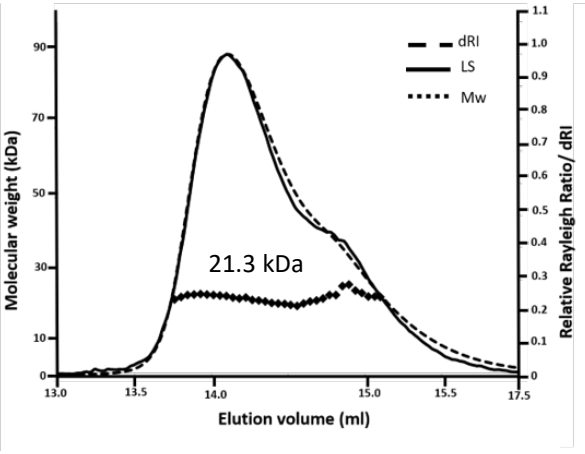
Theoretical monomer – 56.8 kDa



e.

C(3)G534-650

Theoretical monomer – 14.1 kDa



f.

C(3)G550-650

Theoretical monomer – 12.2 kDa

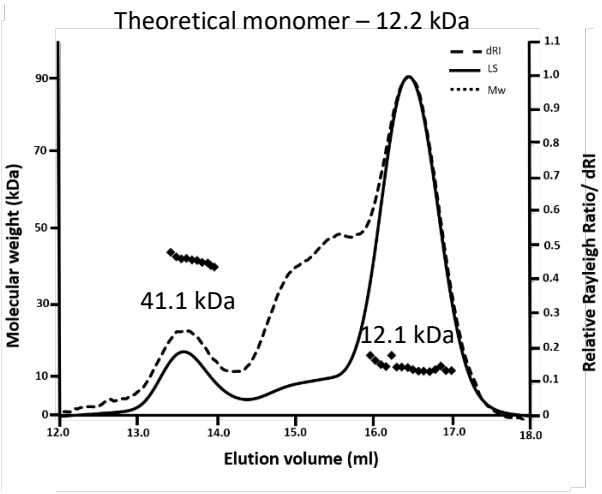
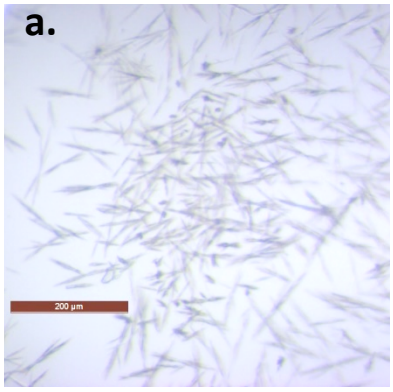
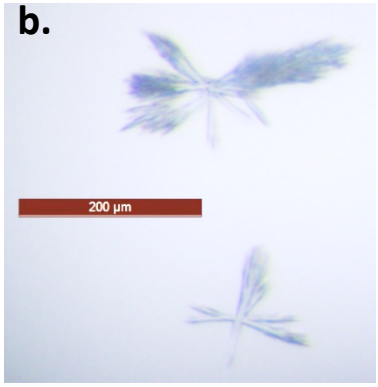


Figure 4.2.10.| Purification and SEC-MALS analysis of C(3)G$_{534-650}$ and C(3)G$_{550-650}$
a-b) SDS-PAGE showing the purification summary of a) C(3)G$_{534-650}$ and b) C(3)G$_{550-650}$ through sequential amylose affinity and anion exchange chromatography. N-terminal MBP-tag was removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. c-f) SEC-MALS analysis of C(3)G constructs c) MBP-C(3)G$_{534-650}$ and e) C(3)G$_{534-650}$ demonstrated a shouldered single peak molecular weights of 76 and 21.3 kDa, respectively (theoretical monomer – 58.7 and 28.2 kDa). d) MBP-C(3)G$_{550-650}$ eluted across three peaks with molecular weights of 232.5, 127.5 and 62.5 kDa (theoretical monomer – 56.8 kDa). f) cleaved C(3)G$_{550-650}$ eluted as two peaks with the main peak having a molecular weight of 12.1 kDa corresponding to a monomer (theoretical monomer – 12.2 kDa)

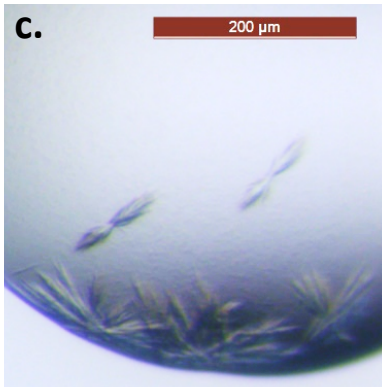
In order to gain a high resolution structure of C(3)G coiled coil domain we initiated crystallographic studies with the shortest C(3)G construct, C(3)G$_{550-650}$, with the hope to solve an X-ray crystal structure. Four initial commercial screens, JCSG+, Index, PACT and MPD were set up at the highest concentration we could yield (11.5 mg/ml). Three out of the four screens gave rise to crystalline growth, with MPD yielding multiple hits (Figure 4.2.11.). Despite this, all of the conditions yielded the same needle-like 'sea urchin' spherulites which grew overnight. To gain a different crystalline form multiple hanging drop and optimisations were set up in an attempt to improve crystal growth. Screening involved varying protein concentration, drop size, buffer component concentrations and the pH. In addition, screens were set up and stored at 4°C with the hope to slow down nucleation events. Despite multiple crystallisation efforts, no other crystal form was obtained. It is possible that a different protein construct is required for successful crystallisation, or a greater chemical space needs to be explored.



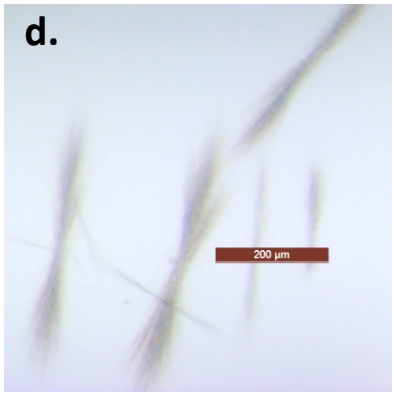
MPD G5: 0.2M Ammonium acetate. 0.1M Sodium citrate pH 5.6 15% MPD



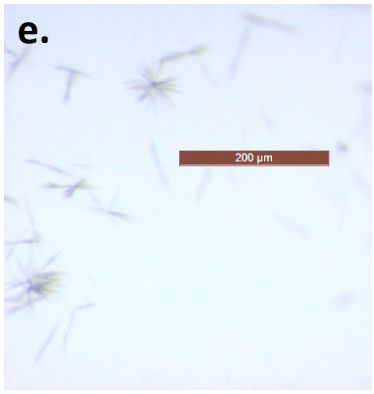
MPD G2: 0.05M Magnesium chloride. 0.1M Tris pH 8.5 12% MPD



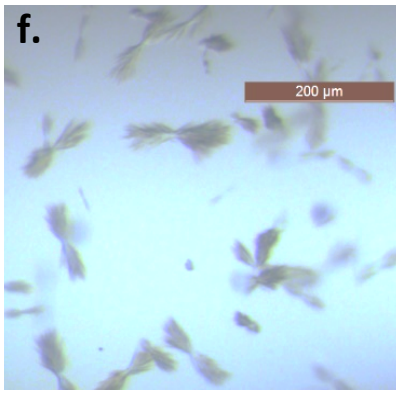
MPD G11: 0.02M calcium chloride 0.1M sodium acetate pH 4.6 30% MPD



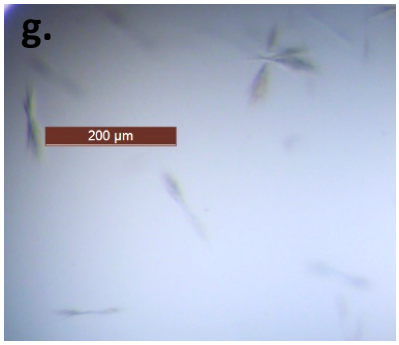
MPD G7: 0.2M Sodium citrate. 0.1M HEPES pH 7.5 15% MPD



MPD F5: 0.1M Tris pH 8.0. 40% MPD



Index C12: 15% v/v TacsimateTM pH 7.0, 0.1 M HEPES pH 7.0, 2% w/v Polyethylene glycol 3,350



JCSG+ C9: 0.1 M Na/K phosphate pH 6.2 25% v/v 1,2 propandiol, 10% v/v Glycerol

Figure 4.2.11.| C(3)G$_{550-650}$ crystal hits.
Commercial screens of C(3)G$_{550-650}$ were set up at 11.5 mg/ml. Crystals hits were observed in a-e) MPD, f) Index and g) JCSG+. All crystal hits were in a needle-like spherulites of varying size.

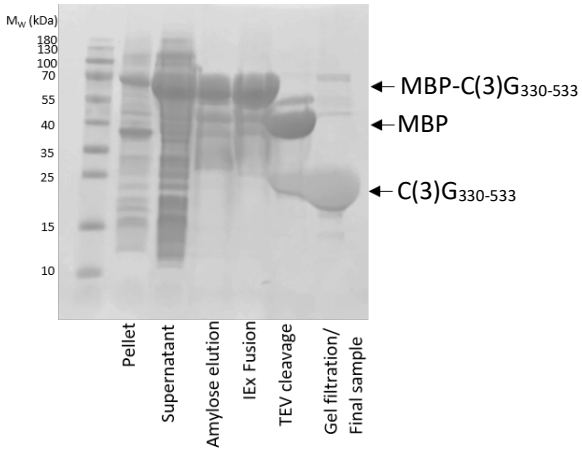
C(3)G$_{330-533}$

A further 117 amino acids were truncated from the C-terminus of C(3)G$_{330-650}$ to produce a shorter protein construct for X-ray crystallography. Purification of C(3)G$_{330-533}$ was achieved through the same steps as the original construct, summarised in Figure 4.2.12.a. This gave greatly improved protein yields compared to the other constructs analysed, suggesting a possible increase in stability. SEC-MALS analysis was performed on both MBP-fusion and cleaved C(3)G$_{330-533}$ to determine the absolute molecular weight of the protein. The SEC profile demonstrates that MBP-C(3)G$_{330-533}$ elutes as two peaks with molecular weights of 251.5 and 130.5 kDa, corresponding to a tetramer and a dimer, respectively (theoretical dimer – 138 kDa) (Figure 4.2.12.b.). Cleaved C(3)G$_{330-533}$ also eluted as two peaks, with the predominant peak corresponding to a dimer (peak 2) and a small amount of tetramer (peak 1), with calculated molecular weights of 45.2 and 87.4 kDa, respectively (theoretical dimer 48.6 kDa) (Figure 4.2.12.c). In order to determine the stability of the tetramer a dilution series of C(3)G$_{330-533}$ was applied to SEC-MALS. The concentration at which the protein complex was loaded onto the SEC column undergoes a ~5-fold dilution before being analysed by the MALS device. SEC-MALS analysis determined that even when C(3)G$_{330-533}$ was loaded at 0.5 mg/ml (analysed at 0.1 mg/ml) two peaks were observed, suggesting the tetramer is stable despite only being present in low amounts (Figure 4.2.12.d.).

CD far-UV spectroscopy of C(3)G$_{330-533}$ revealed a characteristic spectrum of a highly α-helical protein with negative peaks at 208 and 222 nm and a positive peak at 193 nm (Figure 4.2.12.e.). Deconvolution suggested an α-helical content of 85%, corresponding to equating to 173 α-helical residues out of 203. Thermal denaturation revealed single step unfolding of C(3)G$_{330-533}$ with an estimated Tm, at 50% unfolded, of 33.2°C, consistent with the longer construct (Figure 4.2.12.f.).

Solution studies by SEC-SAXS revealed that C(3)G$_{330-533}$ eluted as two peaks, consistent with the SEC-MALS data with the first peak being uninterpretable. Analysis of the latter, predominant, peak revealed an *Rg* and *Rc* of 74 and 10.4 Å, respectively (Figure 4.2.13.a. and c.). Since the *Rc* reveals the thickness of elongated molecules, we can infer this to calculate the number of helices within the coiled-coil.

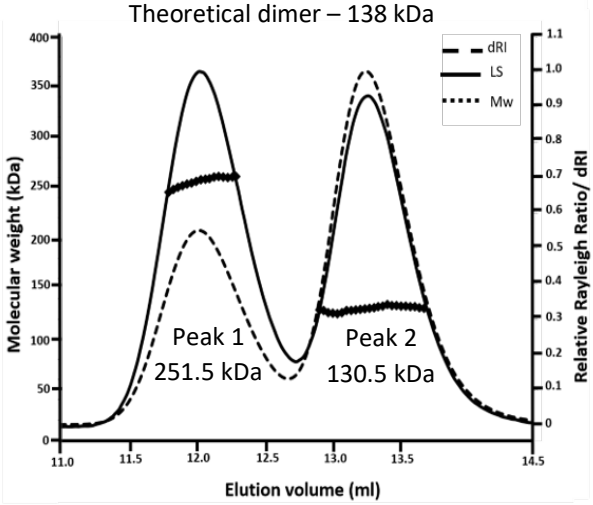
a.



b.

MBP-C(3)$G_{330-533}$

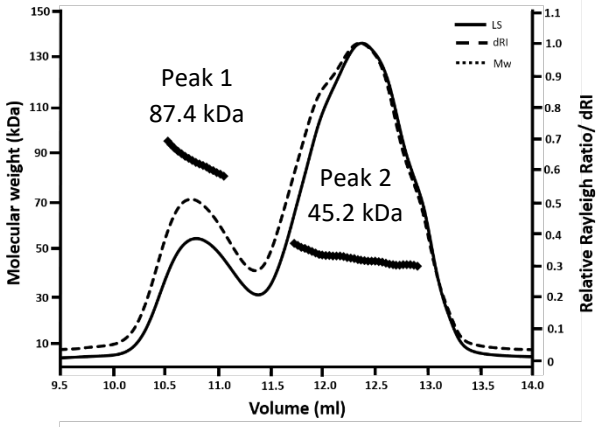
Theoretical dimer – 138 kDa



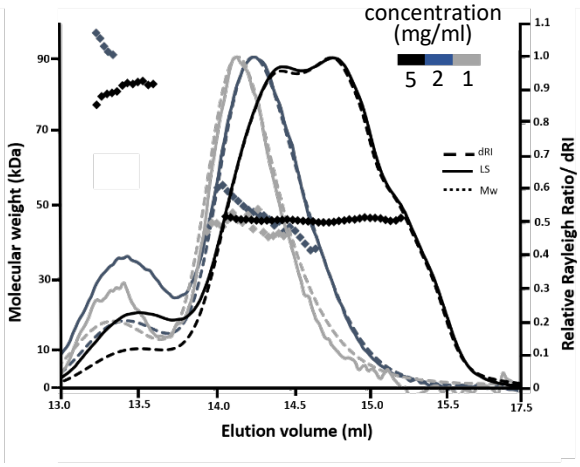
c.

C(3)$G_{330-533}$

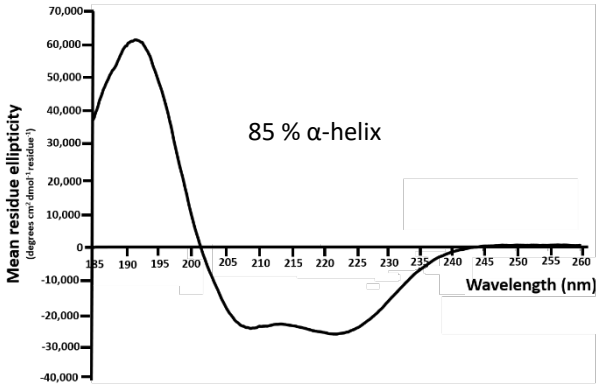
Theoretical dimer – 48.6 kDa



d.



e.



f.

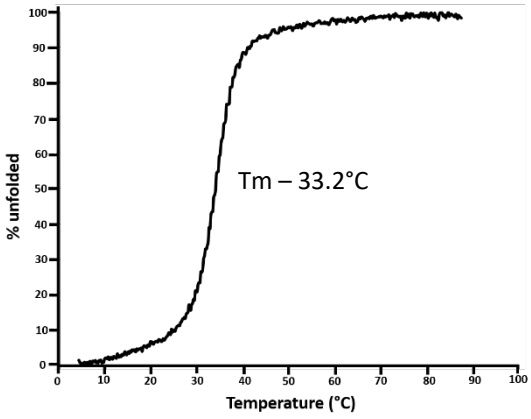


Figure 4.2.12.| Purification and biophysical analysis of C(3)G$_{330-533}$
a) SDS-PAGE showing the purification summary of C(3)G$_{330-533}$ through sequential amylose affinity and anion exchange chromatography. N-terminal MBP-tag was removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. b) SEC-MALS analysis of MBP-C(3)G$_{330-533}$ eluted as two clear peaks with molecular weights of 251.5 and 130.5 kDa corresponding to a tetramer and a dimer, respectively (theoretical dimer - 138 kDa). c) SEC-MALS analysis of cleaved C(3)G$_{330-533}$ also eluted as two peak, with molecular weights of 87.4 and 45.2 kDa, corresponding to a tetramer and dimer (theoretical dimer – 48.6 kDa). d) SEC-MALS dilution series of C(3)G$_{330-533}$ revealed that even when loaded at 0.5 mg/ml C(3)G$_{330-533}$ elutes as two peaks, with the predominant peak being the dimer. The protein that is loaded onto the SEC column undergoes an approximate 5-fold dilution, therefore the sample analysed by MALS is ~5-fold the concentration that was loaded. e) Far UV CD wavelength scan between 260-185nm of C(3)G$_{330-533}$ shows a typical α-helical trace. Deconvolution of the data estimates the secondary structure to be 85 % α-helical e) CD thermal denaturation measured at 222 nm between 4 and 95°C, estimated a melting temperature (T_m) of 33.2 °C.

The observed *Rc* of C(3)G$_{330-533}$ is consistent with a tetrameric coiled-coil (measured *Rc* of 10 Å) (Solari & Moses, 1973; Truebestein & Leonard, 2016). Furthermore, the Porod-Debye plot was used to derive the volume of the scattering particle and estimate the molecular mass. Inspection of the Kratky and Porod plots of C(3)G$_{330-533}$ reveals a particle with a defined Porod-Debye plateau and an associated volume of 173430 Å^3, consistent with a 102 kDa protein (Figure 4.2.13.d.). This result indicates that the SAXS analysis of C(3)G$_{330-533}$ is of the tetramer. C(3)G$_{330-533}$ has a *P(r)* profile typical of an elongated protein shown by an elongated tail with a *Dmax* of 342 Å (Figure 4.2.13.b.).

The SEC-SAXS scattering data was ultilised to create a low-resolution *ab intio* model of C(3)G$_{330-533}$. *Ab initio* modelling produces a molecular envelope of densely packed beads with a defined volume (*Rg*) and dimension (*Dmax*). The resultant modelling demonstrates a rod-shaped molecular envelope for C(3)G$_{330-533}$ typical of an extended α-helical coiled-coil protein (Figure 4.2.13.e.).

High throughput screening of C(3)G$_{330-533}$ was carried out with the commercial crystallistaion screens JCSG+, MPD, Index, and Proplex at a concentration of 112 mg/ml. The JCSG+ commercial condition led to the formation of crystals in several drops. However, all crystals had the same needle-like morphology, similar to that observed with C(3)G$_{550-650}$ crystals images (Figure 4.2.14.a.-e). Hanging drop optimisations of these conditions were set up, including lowering the protein concentration, changing the drop size and lowering the precipitant. Optimisation of the JCSG+ condition also yielded needle-like ‘sea urchin’ crystals (Figure 4.2.14.f.)..

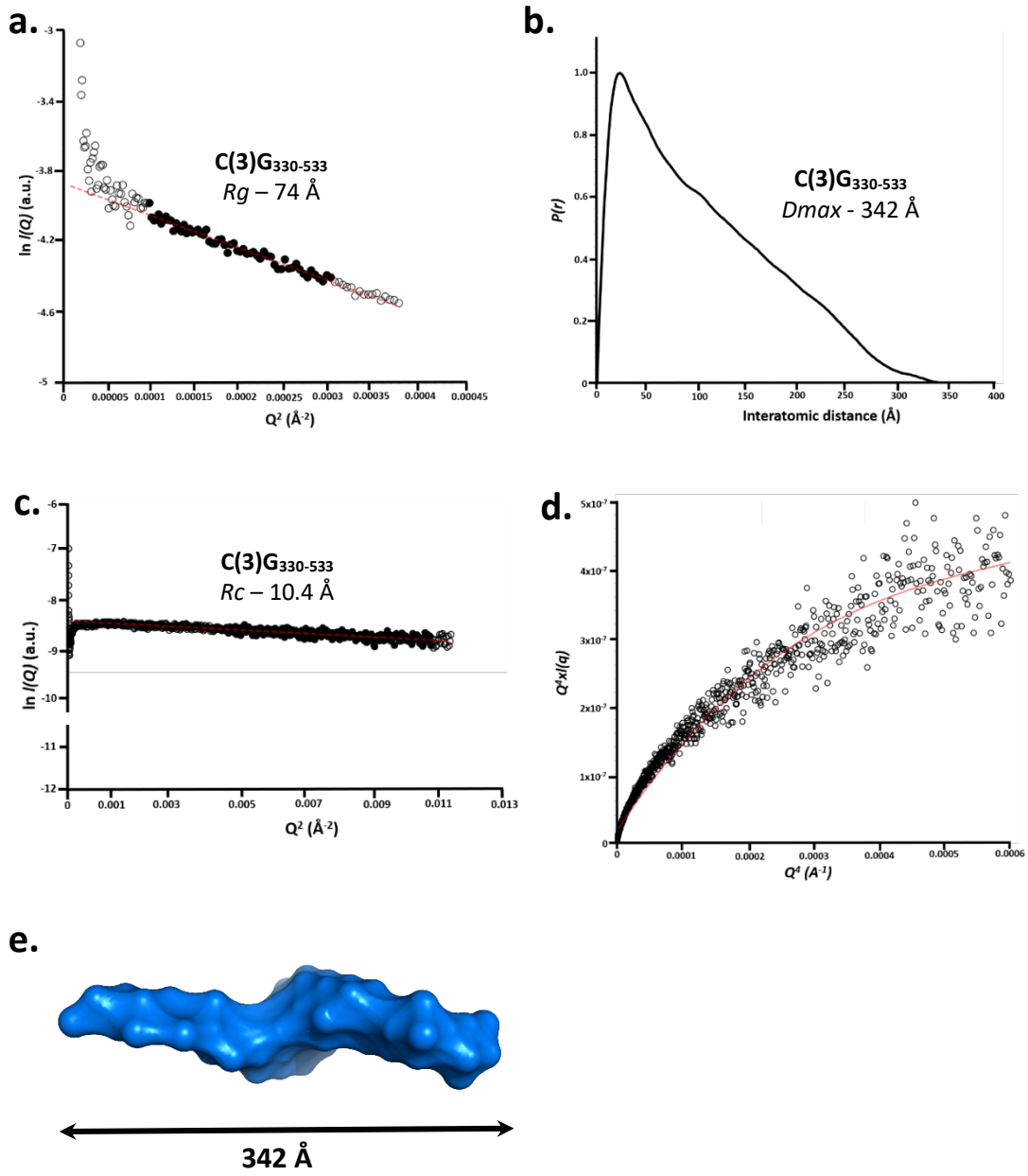
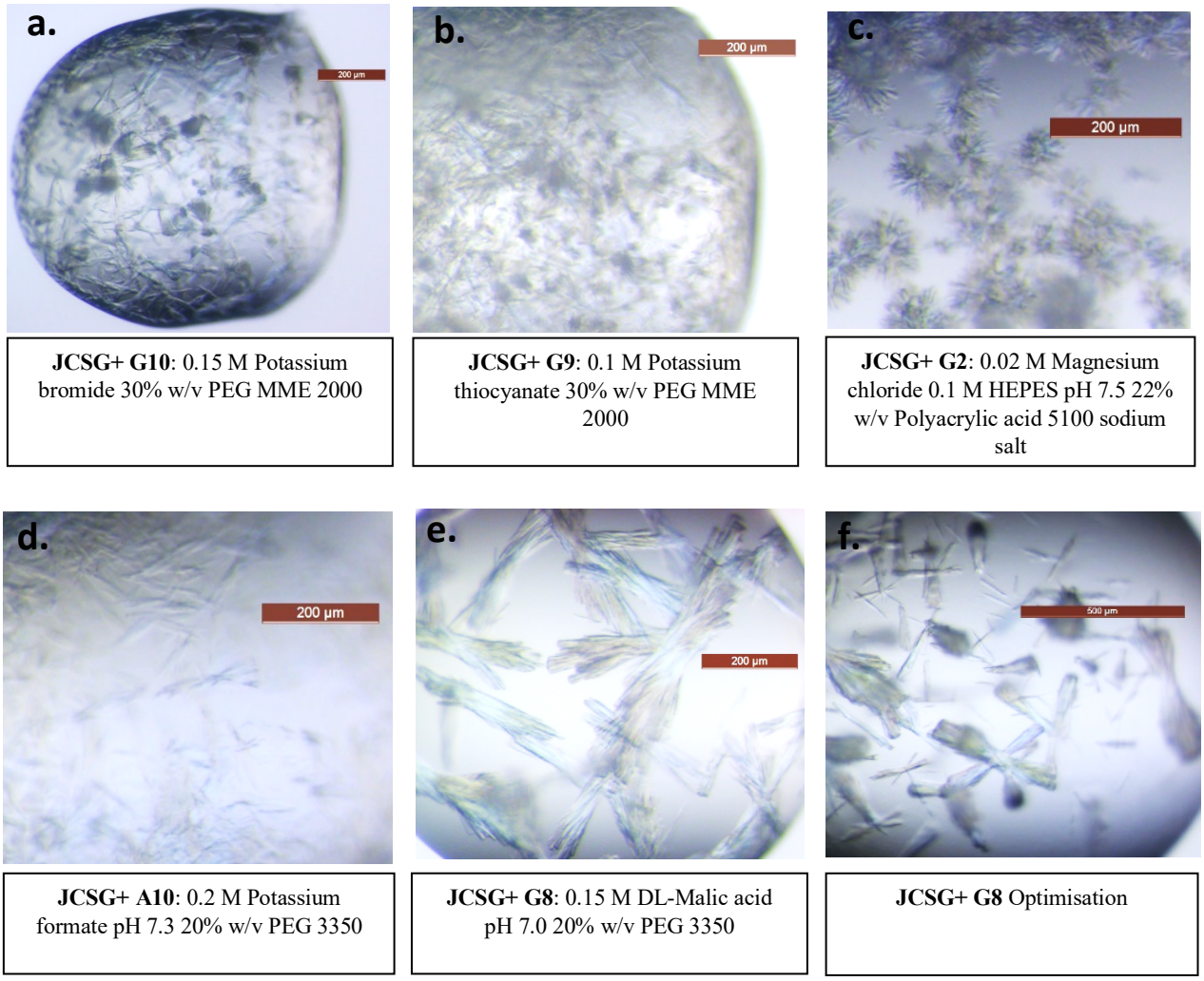


Figure 4.2.13.| SEC-SAXS analysis of C(3)G$_{330-533}$
a) Guinier analysis determined a radius of gyration (*Rg*) value of 74 Å. The real space *Rg* closely matches the Guinier analysis *Rg* value of 90 Å and 74 Å, respectively. Clear circles represent the complete data set and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3). b) *P(r)* distribution of C(3)G$_{330-533}$ showing a maximum dimension of 342 Å. c) Guinier analysis determined the radius of gyration the *Rc*) of C(3)G$_{330-533}$ to be 10.4 Å. Clear circles represent the complete data and solid circles represent the region used for the fit. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3). d) Porod-Debye plot of C(3)G$_{330-533}$ determined an associated volume of 173430 Å^3. e) DAMMIF *ab initio* model of C(3)G$_{330-533}$ determined an elongated molecular envelope with a *Dmax* of 342 Å, as observed by the paired distribution profile.

Lowering the precipitant in the optimisation gave empty drops. We used the crystals in JCSG+ G8 to create a seed stock solution, which was then used to streak seed through these empty drops to initiate nucleation events and start crystal growth. However, this did not yield any crystals.



JCSG+ G10: 0.15 M Potassium bromide 30% w/v PEG MME 2000

JCSG+ G9: 0.1 M Potassium thiocyanate 30% w/v PEG MME 2000

JCSG+ G2: 0.02 M Magnesium chloride 0.1 M HEPES pH 7.5 22% w/v Polyacrylic acid 5100 sodium salt

JCSG+ A10: 0.2 M Potassium formate pH 7.3 20% w/v PEG 3350

JCSG+ G8: 0.15 M DL-Malic acid pH 7.0 20% w/v PEG 3350

JCSG+ G8 Optimisation

Figure 4.2.14.| C(3)G$_{330-533}$ crystal hits.
High throughput screens of C(3)G$_{330-533}$ were set up at 112 mg/ml in JCSG+, MPD, Index and Proplex. a-e) Crystal hits were only observed in JCSG+. f) Optimisation hanging drop screen of JCSG+ also produced large needle-like clusters of crystals.

C(3)G$_{330-459}$

C(3)G$_{330-459}$ is a C-terminal truncation of 74 amino acids from the previously studied C(3)G$_{330-533}$ construct. C(G)G$_{330-459}$ was expressed with an N-terminal MBP-tag and purified following the same method, summarised in figure 4.2.15.a.. Both MBP-C(G)G$_{330-459}$ and C(G)G$_{330-459}$ were analysed by

SEC-MALS to determine the oligomeric state. MBP-C(3)$G_{330-459}$ and C(G)$G_{330-459}$ eluted as a single peak with molecular masses of 55.7 and 22.6 kDa, respectively. MALS of MBP-C(3)$G_{330-459}$ determined a molecular weight corresponding to a monomer (theoretical monomer – 60.1 kDa). However, MALS data from the cleaved protein did not allow confident determination between a monomer and dimer (theoretical monomer and dimer -15.4 and 30.8 kDa, respectively) (Figure 4.2.15.b. and c.). This result suggests that there could be a mixture of both monomer and dimer species, and that the further C-terminal truncation possibly weakens dimerisation, indicating that residues 459-533 are important for stable dimer formation.

Far-UV CD spectroscopy determined that C(G)$G_{330-459}$ has an α-helical structure with deconvolution suggesting 82 % helicity (Figure 4.2.15.d.). Thermal denaturation revealed a single stage unfolding event of C(G)$G_{330-459}$, with an estimated Tm of 25.6°C (Figure 4.2.15.e.). This observed T_m is much lower than both C(3)$G_{226-650}$ and C(3)$G_{330-533}$, which are stable dimeric coiled coils, having T_m of 38.6 and 33.2°C, respectively, which suggests that C(3)$G_{330-459}$ is less stable than the two longer constructs.

SEC-SAXS analysis of C(3)$G_{330-459}$ was used to gain in solution low resolution structural information. The *Rg* and *Rc* were determined to be 53.5 and 9.5 Å, respectively (Figure 4.2.16.a. and c.).These results are consistent with an elongated structure, with the *Rc* indicating a dimeric coiled-coil suggesting that the SAXS data is dominated by the larger dimeric species. The real space distribution curve shows an extended tail indicative of an elongated structure, with an *Dmax* of 216 Å (Figure 4.2.16.b.).

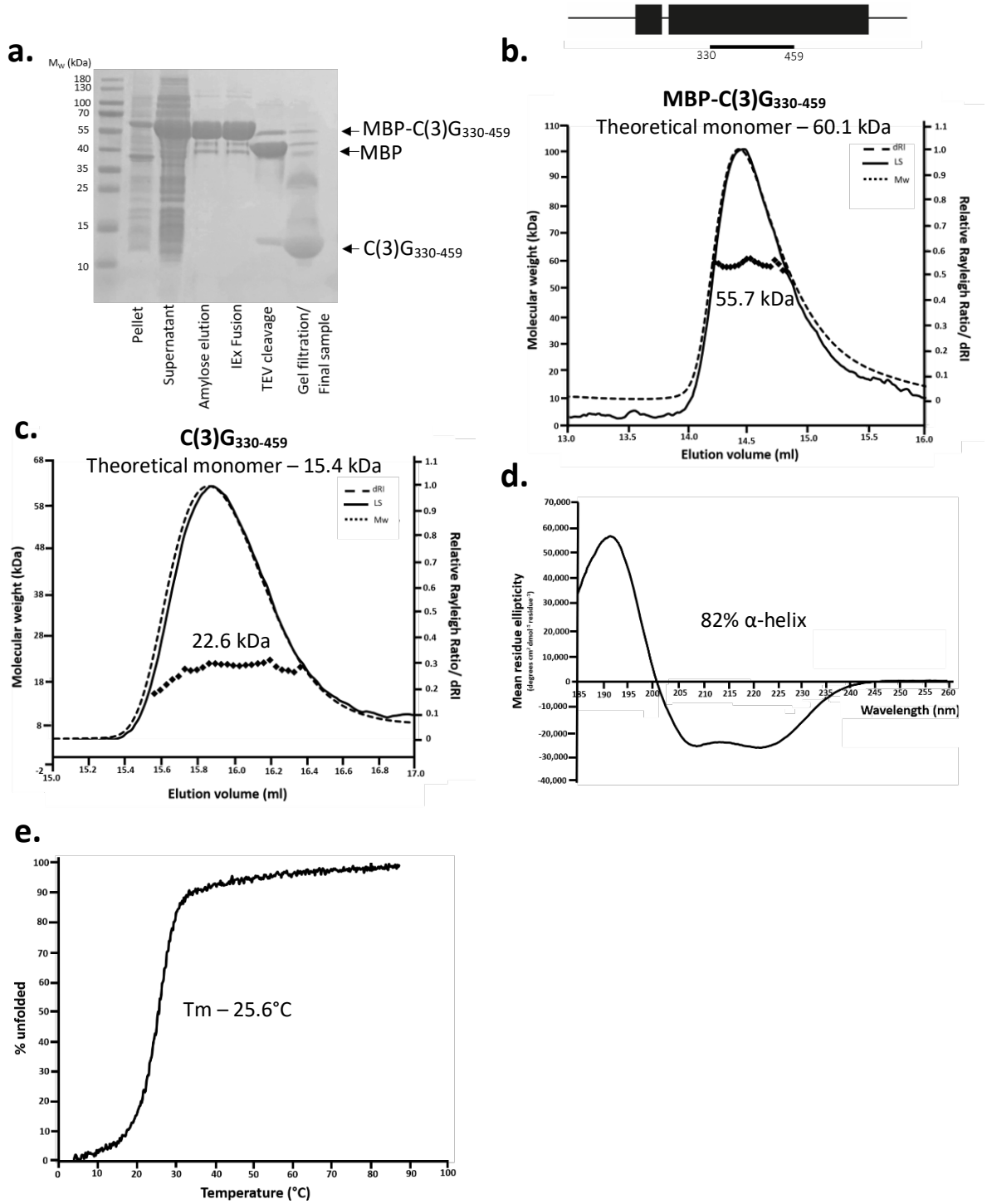


Figure 4.2.15.| Purification and biophysical analysis of C(3)G$_{330-459}$

a) SDS-PAGE showing the purification summary of C(3)G$_{330-459}$ through sequential amylose affinity and anion exchange chromatography. N-terminal MBP-tag was removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. b-c) SEC-MALS analysis of b) MBP-C(3)G$_{330-459}$ and c) cleaved C(3)G$_{330-459}$ determined a single peak with a molecular weights of 55.7 and 22.6 kDa, respectively corresponding to a monomer (theoretical monomer – 60.1 and 15.4 kDa, respectively.) d) Far UV CD wavelength scan between 260-185nm of C(3)G$_{330-459}$ shows a typical α-helical trace. Deconvolution of the data estimates the secondary structure to be 82 % α-helical e) CD thermal denaturation measured at 222 nm between 4°C and 95°C, estimated a melting temperature of 25.6 °C.

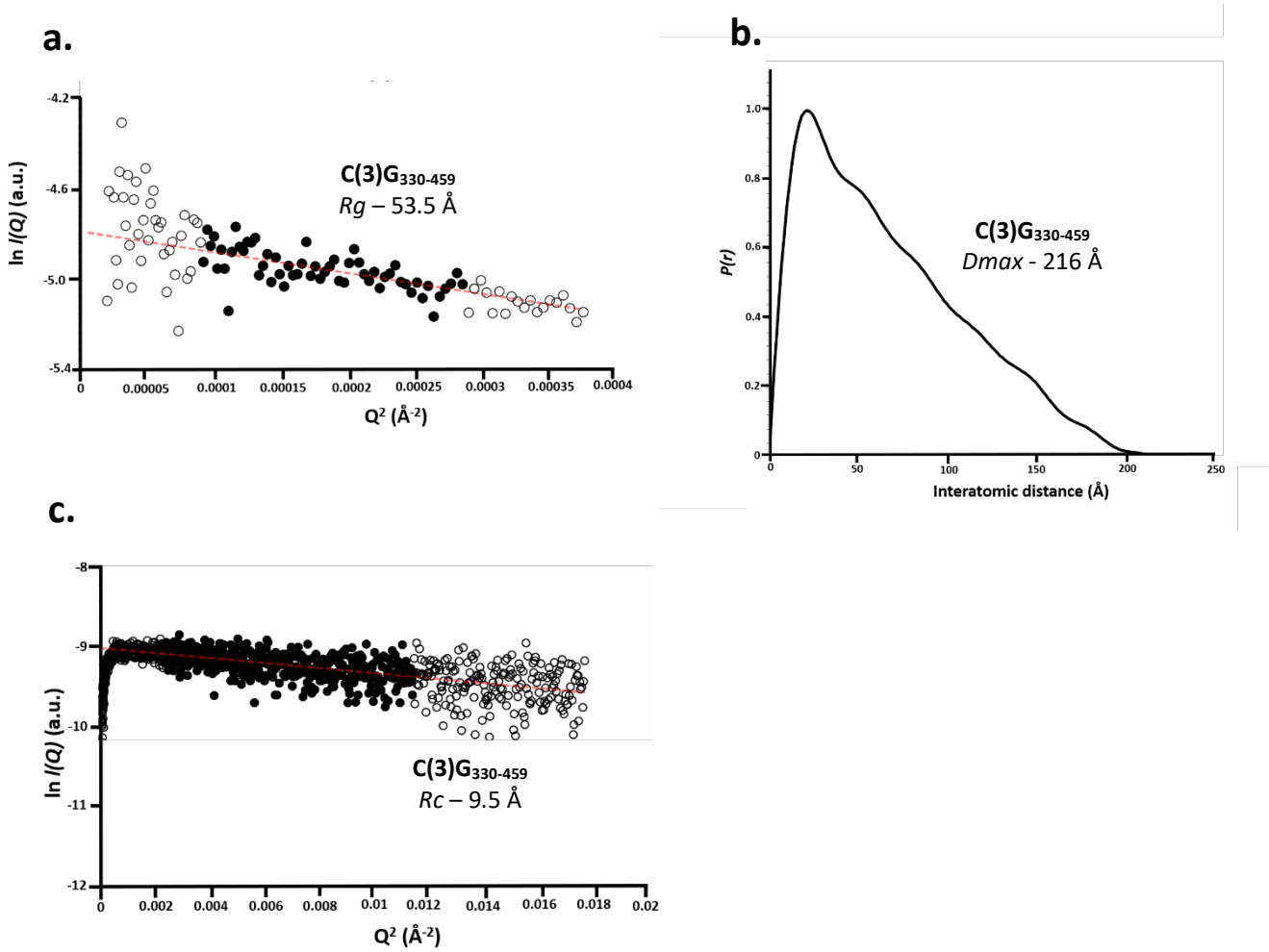


Figure 4.2.16.| SEC-SAXS analysis of C(3)G$_{330-459}$
a) Guinier analysis determined a radius of gyration (*Rg*) value of 53.5 Å. The real space *Rg* closely matches the Guinier analysis *Rg* value of 56.9 and 53.5 Å, respectively. Clear circles represent the complete data set and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3). b) *P(r)* distribution of C(3)G$_{330-459}$ showing a maximum dimension of 216 Å. c) Guinier analysis determined the *Rc* to be 9.5 Å. Clear circles represent the complete data and solid circles represent the region used for the fit. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3).

We have determined C(3)G$_{226-650}$ to be the longest soluble construct of the α-helical domain, forming a stable dimer in solution. Purification of C(3)G fragments, and subsequent biophysical analysis, demonstrates the α-helical nature of C(3)G. SEC-MALS analysis determined that the C(3)G constructs were predominantly dimeric, with some propensity for tetramerisation, driven by the central amino acid residues 330-533. SEC-SAXS analysis of C(3)G$_{330-533}$ determined parameters consistent with a tetrameric coiled-coil. We propose that the C(3)G dimers bundle together to form a tetramer.

Analysis of two C(3)G N-terminal truncations, C(3)G$_{330-650}$ and C(3)G$_{470-650}$, revealed absolute molecular weights corresponding to dimeric species, suggesting that the N-terminus of the α-helical domain is not essential for dimerisation. However, analysis of shorter constructs, C(3)G$_{330-459}$ and C(3)G$_{534-650}$, revealed molecular weights between a monomer and dimer, suggesting that these two constructs form unstable dimers that dissociate in solution.

The CD thermal melt of C(3)G$_{330-459}$ was determined to be much lower than the longer dimeric constructs, suggesting reduced stability. SEC-MALS analysis of the C-terminal 100 amino acids, C(3)G$_{550-650}$, determined a monomer, suggesting that the central dimeric/tetrameric α-helical domain may be flanked by a monomeric C-terminus. The oligomeric status, determined by SEC-MALS, of the truncated C(3)G constructs and biophysical findings are summarised in Figure 4.2.17. and Table 4.2.1.

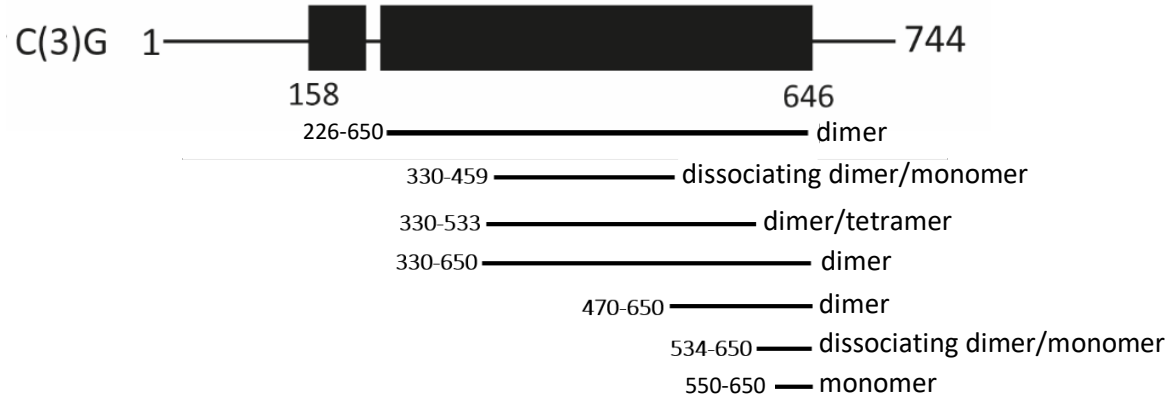


Figure 4.2.17.| Schematic summary of the C(3)G α-helical domain truncations.
A series of N-terminal, C- and N-terminal, and C-terminal truncations of C(3)G$_{226-650}$ were analysed by SEC-MALS to determine their oligomeric state.

Table 4.2.1.| Summary table of C(3)G constructs

* - $C(3)G_{330-533}$ – SEC-SAXS analysis of the main peak

** - C(3)G644-744 far-UV spectroscopy deconvolution revealed: 5% α-helix, 35% β-sheet, 24% turn and 36% unstructured.

C(3)G Biophysical Data Summary												
				CD			SEC-MALS			SEC-SAXS		
Construct	Number of residues	Mw (kDa)		Helicity		Tm (°C)	Mw (kDa)		Oligomer	*Rg* (Å)	*Rc* (Å)	*Dmax* (Å)
		cleaved	fusion	%	a.a.		Fusion	Cleaved				
$C(3)G_{226-650}$	427	49.8	94.7	77	329	37.4	176	97.3	dimer	151	9.0	680
$C(3)G_{330-650}$	323	38.0	82.7	88	282	41.2	153	72.4	dimer	129	8.2	560
$C(3)G_{330-533}$	206	24.3	68.9	85	173	33.2	Peak 1 – 251.5 Peak 2 – 130.5	Peak 1 – 87.4 Peak 2 – 45.2	tetramer/ dimer	74*	10.4*	342*
$C(3)G_{330-459}$	132	15.4	60.1	82	107	25.6	55.7	22.6	monomer/dimer	53.5	9.5	216
$C(3)G_{470-650}$	183	21.7	66.4	83	149	36.8	117	37	dimer	63.5	9.1	290
$C(3)G_{534-650}$	119	14.1	58.7	-	-	41.8	71	21.3	monomer/dimer	-	-	-
$C(3)G_{550-650}$	103	12.2	56.8	-	-	44.2	Peak 1 – 127.5 Peak 2 – 62.2	12.1	monomer	-	-	-
Deletion 1 $C(3)G_{226-650\ \Delta 340-552}$	215	24.8	69.5	86	185	22.4	68.1	25.1	monomer	60.27	14.1	278
Deletion 2 $C(3)G_{226-650\ \Delta 346-361}$	412	48.1	92.8	83.9	346	42.0	175.8	91.1	dimer	151	9	674
Deletion 3 $C(3)G_{226-650\ \Delta 465-471}$	421	49.2	93.9	86	359	38.0	174.3	95.6	dimer	142.2	8.6	650
$C(3)G_{644-744}$	103	11.6	56.3	5**	-	-	55.5	11.6	monomer	32	-	130

4.2.4. The N-terminus of C(3)G degrades in solution

Super-resolution EM studies have shown that the N-terminus of the mammalian transverse filament protein (SYCP1) resides at the SC midline (J. G. Liu et al., 1996). It has also been shown that SYCP1 molecules self-associate through their N-terminus (Dunce et al., 2018; Liu et al., 1996). Immunolocalisation studies of the N-terminus of C(3)G revealed similar functions with that of SYCP1, with the N-terminus residing within the midline of the SC (Anderson et al., 2005; Cahoon et al., 2017). Secondary structure prediction of C(3)G determined that the N-terminal 100 amino acids are highly unstructured with only amino acid residues 100-230 predicted to display α-helical character. We, however, set out to explore the structural properties of the C(3)G N-terminus and its role in the SC.

$C(3)G_{1\text{-}230}$ was expressed with an N-terminal His_6-MBP solubility tag encoded by the pMAT11 vector in *E. coli*. MBP-$C(3)G_{1\text{-}230}$ (71.1 kDa) showed weak expression. The small amount of soluble material was also highly degraded upon purification by amylose affinity chromatography (Figure 4.2.18.a.). The presence of insoluble material observed in the pellet (Figure 4.2.18.a.), combined with high levels of degradation, could suggest that the expression conditions such as temperature and lysis buffer need to be optimised. Upon induction, the growth temperature of the bacterial cultures was reduced from 25 to 15 °C to promote protein folding.

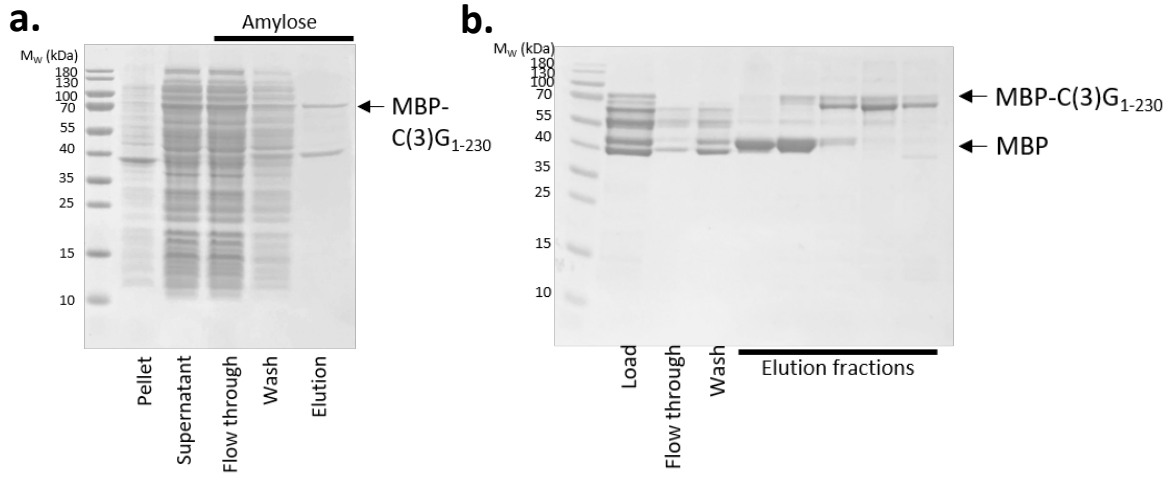


Figure 4.2.18.| Purification of MBP-$C(3)G_{1\text{-}230}$
a-b) The small amount of MBP-$C(3)G_{1\text{-}230}$ soluble material was purified by amylose affinity (a) and subsequent anion exchange chromatography. There is a significant amount of free-MBP present, this indicates that the N-terminus of C(3)G undergoes degradation suggesting the protein is unstable.

Despite this, the expression of C(3)G_{1-230} did not improve. The amylose eluate was further purified by anion exchange chromatography to remove free-MBP and degradation products (Figure 4.2.18.b). Further degradation was observed suggesting that the N-termini of C(3)G is not stable when expressed in isolation.

4.2.5. Purification and characterisation of the unstructured C-terminus, C(3)$G_{644-744}$

Recent studies have shown that the mammalian TF protein, SYCP1, is recruited to the chromosomal axes through its C-terminal DNA binding domain. This SYCP1-DNA interaction subsequently allows for the proper alignment of SYCP1, which span across the central region of the SC, and form head-to-head interactions at the N-termini (Dunce et al., 2018). To investigate whether the C-terminus of C(3)G behaves in a similar manner to SYCP1, we expressed the C-terminus for purification and analysis.

The C-terminus of C(3)G (amino acid residues 644-744) was expressed with an N-terminal MBP solubility tag in *E. coli* BL21 DE3 cells. C(3)$G_{644-744}$ was purified by sequential amylose affinity and anion exchange chromatography. The MBP affinity tag was cleaved by TEV-protease and a cation exchange step was performed to separate C(3)$G_{644-744}$ from the MBP-tag and TEV protease, followed by SEC. An overview of the protein purification steps is summarised in figure 4.2.19.a. The finding that C(3)$G_{644-744}$ was able to bind to a cation HiTrap Heparin HP column indicates its DNA binding ability as heparin mimics the polyanionic structure of the nucleic acid.

SEC-MALS analysis was used to determine the absolute molecular weight of C(3)$G_{644-744}$ and estimate the oligomeric state. Both MBP-C(3)$G_{644-744}$ and C(3)$G_{644-744}$ elute as a single peak with molecular mass of 55.5 and 11.6 kDa, respectively, correlating to a monomer (theoretical monomer – 56.3 and 11.6 kDa, respectively) (Figure 4.2.19.b. and c.). This finding suggests that the C-terminus of C(3)G does not form a dimer like the central coiled-coil domain.

Far-UV CD spectroscopy was used to gain an insight into the secondary structure of C(3)$G_{644-744}$. The CD scan between 260 to 185 nm of C(3)$G_{644-744}$ revealed a negative peak at ~200 nm, suggesting a

possible β-sheet structure. Deconvolution estimates a secondary structure content of 5% α-helix, 35% β-sheet, 24% turn and 36% unstructured (Figure 4.2.19.d.). Due to a very low amount of helical structure, it was not possible to determine the thermal stability of C(3)G$_{644-744}$, as this measured at the helical signal (222nm).

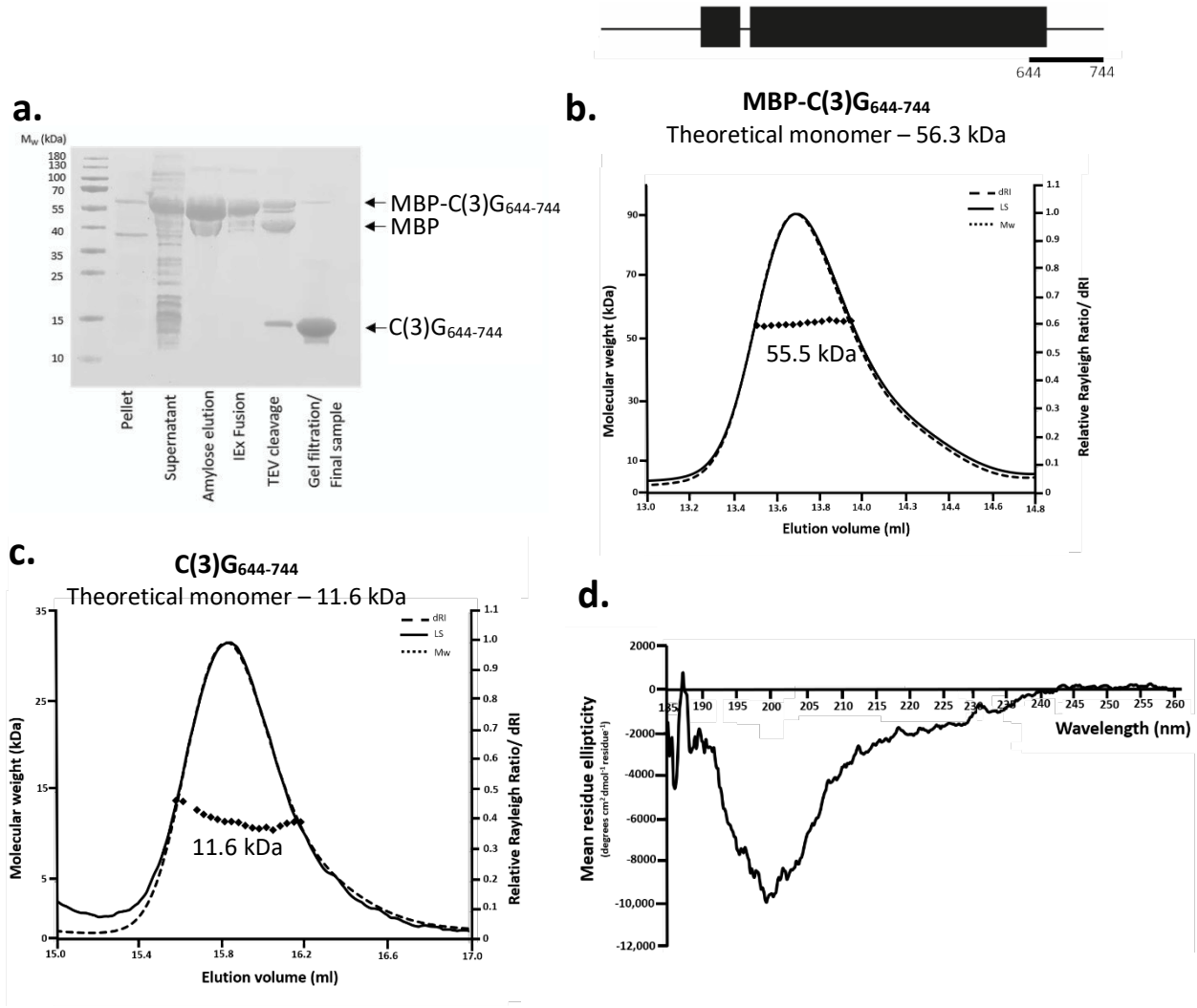


Figure 4.2.19.| Purification and biophysical analysis of C(3)G$_{644-744}$
a) SDS-PAGE showing the purification summary of C(3)G$_{644-744}$ through sequential amylose affinity and anion exchange chromatography. N-terminal MBP-tag was removed by incubation with TEV protease followed by cation exchange and subsequent size exclusion chromatography. b-c) SEC-MALS analysis of b) MBP-C(3)G$_{644-744}$ and c) cleaved C(3)G$_{644-744}$ determined a bell-shaped curve corresponding to molecular weights of 55.5 and 11.6 kDa, respectively corresponding to monomeric species (theoretical Mw – 56.3 and 11.6 kDa. d) Far UV CD wavelength scan between 260-185nm of C(3)G$_{644-744}$ revealed a negative band at ~200nm. Deconvolution estimates a secondary structure content of 35% β-sheet, 24% turn, 5% α-helix and 36% disordered.

We performed SEC-SAXS analysis to determine low resolution structural information of C(3)G$_{644-744}$. The *Rg* was determined to be 32 Å, which closely matches the real-space *Rg* (34.2 Å) (Figure 4.2.20.a.). The real space distribution curve shows a profile typical of a more compact globular protein, with a *Dmax* of 130 Å (Figure 4.2.20.b.).

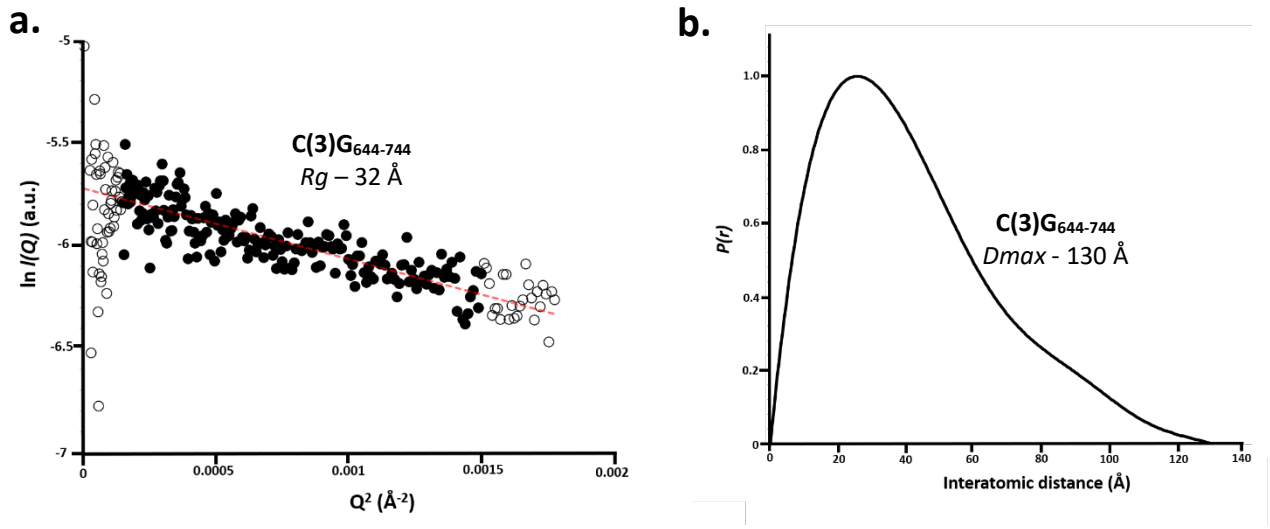


Figure 4.2.20.| SEC-SAXS analysis of C(3)G$_{644-744}$
a) Guinier analysis determined a radius of gyration (*Rg*) value of 32 Å. The real space *Rg* closely matches the Guinier analysis *Rg* value of 34.2 and 32 Å, respectively. Clear circles represent the complete data set and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3). b) *P(r)* distribution of C(3)G$_{644-744}$ showing a maximum dimension of 130 Å.

4.2.6. C(3)G$_{644-744}$ binds dsDNA

The unstructured C-terminus of C(3)G is highly basic (theoretical pI of 9.59); indicating a possibility to bind DNA. We analysed the ability of cleaved C(3)G$_{644-744}$ to bind dsDNA using an electrophoretic mobility shift assay (EMSA). Random sequence dsDNA, at 25 µM, of 3 different lengths (60, 303 and 690 bp) were incubated with increasing amounts of C(3)G$_{644-744}$ (0-32µM). Incubated samples were then run on an agarose gel, where shifts in the bands were observed, which clearly indicated strong affinity for dsDNA (Figure 4.2.21.). A DNA shift was observed at 6.4 µM of C(3)G$_{644-744}$ for 60 bp dsDNA and 3.2 µM for 303 and 690 bp, indicating that C(3)G$_{644-744}$ preferentially binds to the longer length dsDNA.

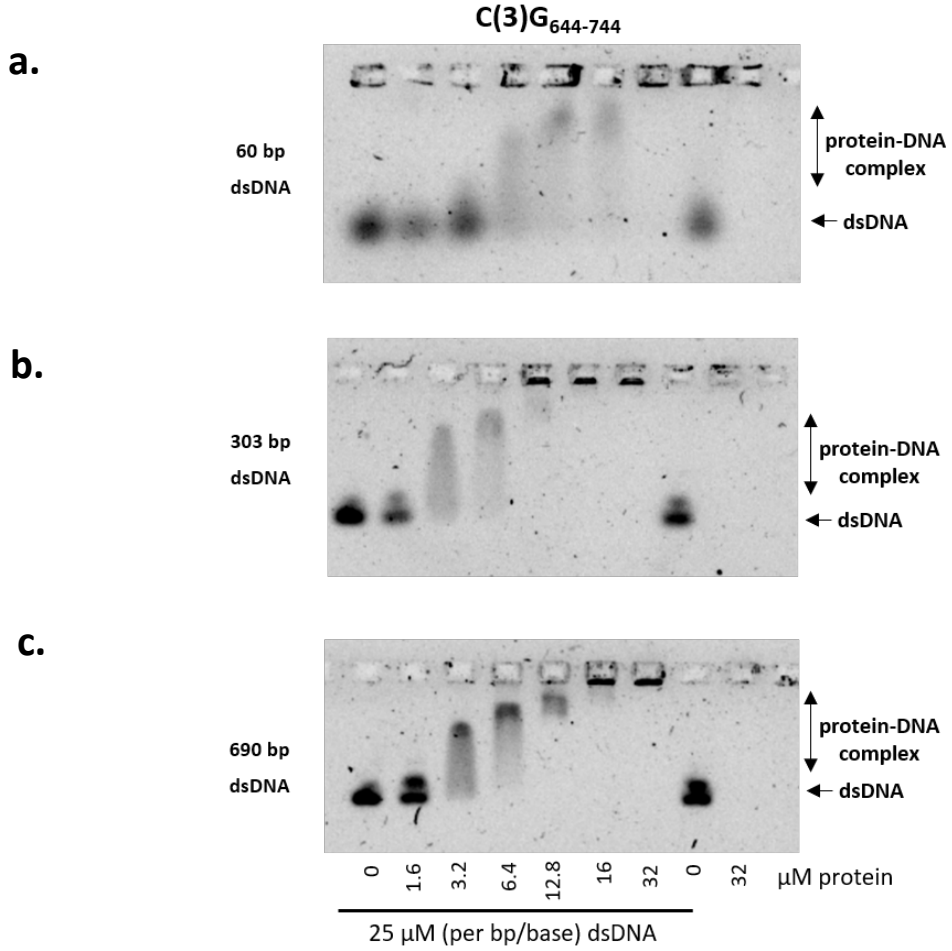


Figure 4.2.21.| EMSA demonstrate the DNA binding ability of C(3)G$_{644-744}$
a-c) Electrophoretic mobility shift assays (EMSA) demonstrate that the C-terminus of C(3)G has the ability to bind random sequence dsDNA of different lengths. a) 60 bp dsDNA, b) 303 bp dsDNA and c) 690 bp dsDNA were used to test the DNA binding ability of C(3)G$_{644-744}$ to bind dsDNA, shown by forming a protein-DNA complex between 3.2-6.4 µM protein.

4.2.7. Biophysical characterision of C(3)G$_{226-650}$ in-frame deletions

To investigate a possible structural function of the C(3)G α-helical coiled coil domain, C(3)G was characterised *in vivo* using deletion studies and high resolution EM techniques by Dr. Katherine Billmyre (Billmyre et al., 2019). Three in-frame deletion mutations; deletion 1 (Δ 340-552), deletion 2 (Δ 346-361) and deletion 3 (Δ 465-471), herein known as C(3)G^{Del1}, C(3)G^{Del2}, and C(3)G^{Del3}, within the highly conserved α-helical domain were designed using the coiled-coil prediction server COILS, detailed in 4.1.7. C(3)G^{Del1} is a 213 amino acid deletion of the central region of the predicted coiled-

coil domain. C(3)G^{Del2} and C(3)G^{Del3} have much smaller deletion regions; 15 and 7 amino acid residues, respectively (Figure 4.2.22.a.). The deleted regions in C(3)G^{del2} and C(3)G^{Del3} both display a dip in the COILS score, suggesting they may be part of a possible linker region between the coiled-coil structures (Lupas et al., 1991). In summary, it was determined that all three deletions caused a partial loss of SC function at different stages of early meiosis. However, surprisingly the smallest 7 amino acid deletion, C(3)G^{Del2} (Δ465-471), was unable to form a fully assembled SC (Billmyre et al., 2019). In order to further dissect the role of the C(3)G α-helical core we set out to characterise the same three in-frame deletions biophysically. All three deletions were introduced into the core coiled-coil construct, C(3)G$_{226-650}$, by overlapping PCR and cloned into the pMAT11 vector and expressed in *E. coli* BL21 DE3 cells. MBP-C(3)G deletions were purified through sequential amylose affinity and anion exchange chromatography, final products are shown in figure 4.2.22.b.

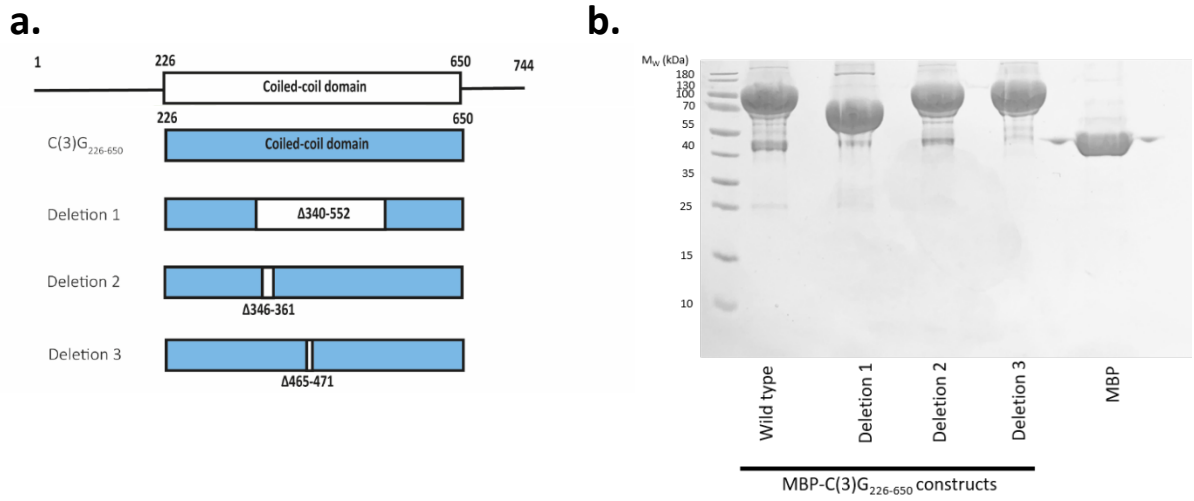
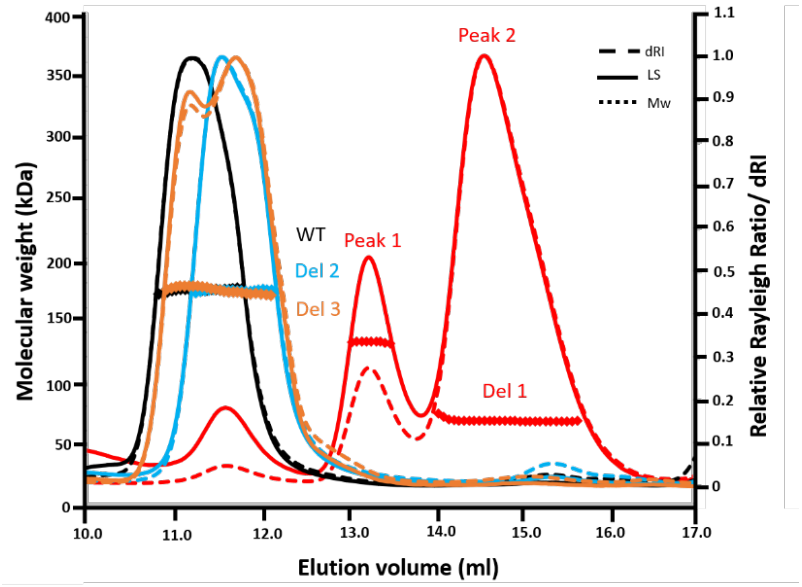


Figure 4.2.22.| Purification of MBP-C(3)G deletion constructs
a) Schematic showing the three in-frame deletions within the α-helical predicted coiled-coil domain. C(3)G^{Del1} removed around half of the α-helical domain (amino acid residues Δ340-552). C(3)G^{Del2} and C(3)G^{Del3} are much smaller deletions, removing 15 *(Δ346-361)* and 7 *(Δ465-471)* amino acid residues respectively, within the same region as C(3)G^{Del1}. b) SDS-PAGE showing the pooled anion exchange fusion fractions of MBP-C(3)G$_{226-650}$ (WT), MBP-C(3)G$_{Del1}$, MBP-C(3)G$_{Del2}$ and MBP-C(3)G$_{Del3}$ that was purified by sequential amylose affinity and anion exchange chromatography.

MBP-C(3)G deletion fusions were analysed by SEC-MALS to determine the absolute molecular weights and the oligomeric states. MBP-C(3)$G_{226-650}$ (MBP-C(3)G^{WT}) was also analysed as a control (Figure 4.2.23.). MBP-C(3)G^{Del1} (red trace) eluted as two peaks with molecular weights of 132.6 kDa (peak 1) and 68.1 kDa (peak 2), corresponding to a dimer and a monomer, respectively (theoretical monomer – 69.5 kDa). The latter monomeric peak is the dominant species, suggesting that the 213 amino acid deletion has disrupted dimer formation, however protein folding is maintained. Both MBP-C(3)G^{Del2} (cyan trace) and MBP-C(3)G^{Del3} (orange trace) elute as a single peak, overlaying with MBP-C(3)G^{WT} (black trace). The molecular weights were constant across the peaks and determined to be 175.8 and 174.3 kDa, respectively, indicating that deletions 2 and 3 do not disrupt dimer formation (MBP-C(3)G^{Del2} and MBP-C(3)G^{Del3} theoretical dimer – 185.6 and 187.8 kDa, respectively).



Construct	Theoretical Mw (kDa)	Experimental Mw (kDa)	Oligomeric state
Wild type	94.5	176	Dimer
Deletion 1	69.5	Peak 1 – 132.6 Peak 2 – 68.1	Dimer Monomer
Deletion 2	92.8	175.8	Dimer
Deletion 3	93.9	174.3	Dimer

Figure 4.2.23.| SEC-MALS analysis of MBP-C(3)G Deletion constructs.
SEC-MALS analysis of MBP-C(3)G^{WT} (black), MBP-C(3)G^{Del2} (cyan) and MBP-C(3)G^{Del3} (orange) eluted as single peaks with molecular weights of 176, 175.8 and 174.3 kDa, respectively, corresponding to dimeric species. MBP-C(3)G^{Del1} (red) elutes as two peaks with molecular weights of 132.6 (Peak 1) and 68.1 kDa (Peak 2), corresponding to a dimer and a monomer, respectively.

SEC-SAXS analysis of the MBP-C(3)G deletion fusions was used to determine the maximum dimensions (*Dmax*) of the constructs and more importantly derive the orientation of helices within the dimer. We have previously shown by SAXS distance distribution profiles of MBP-fusion constructs that C(3)G^{WT} forms a parallel dimer (Figure 4.2.5.d.), therefore we wanted to determine if the deletion mutants disrupt or change this parallel orientation. The distance distribution profiles of the MBP-deletion constructs were typical of a α-helical coiled-coil structure with elongated tails, consistent with MBP-C(3)G^{WT}.

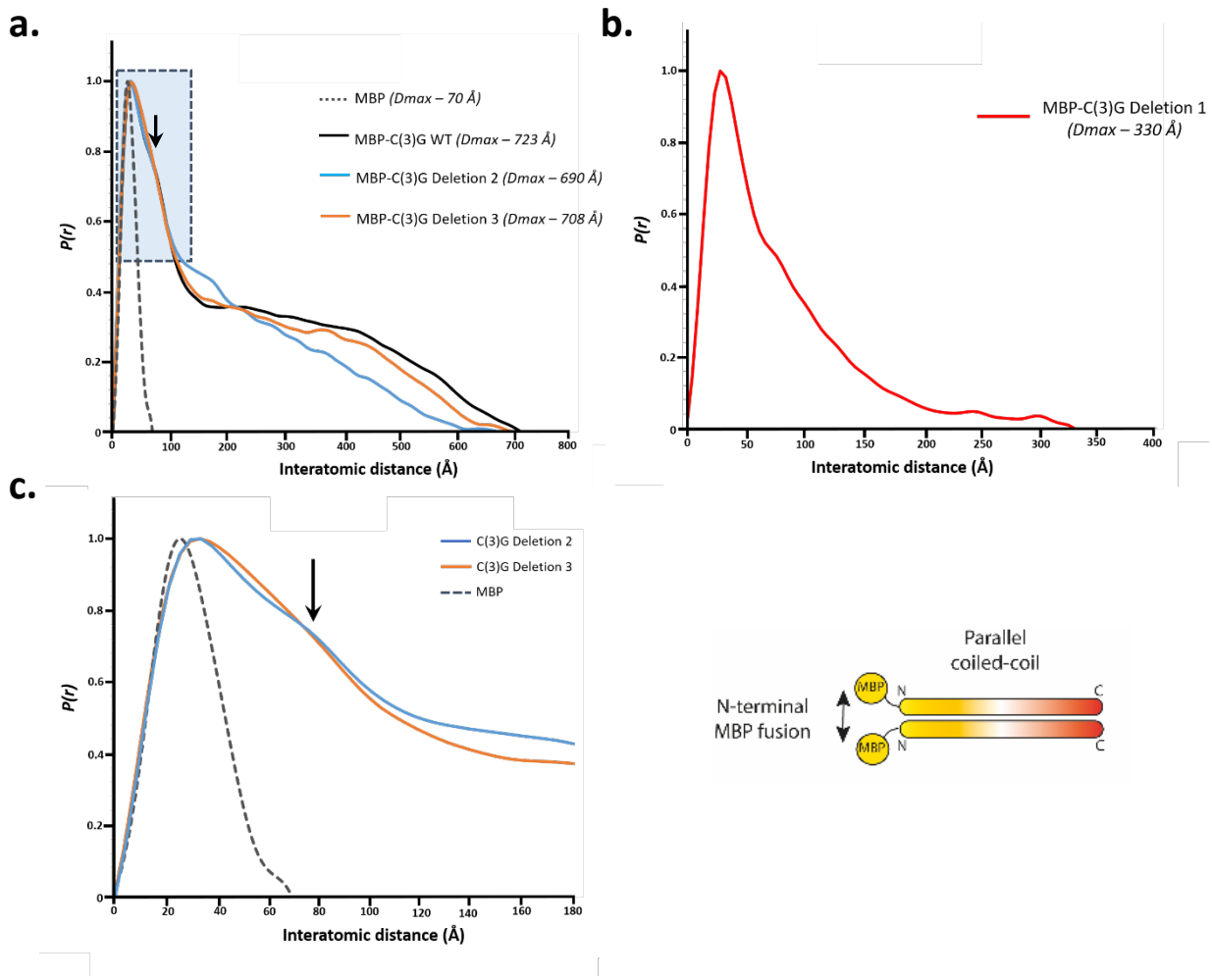
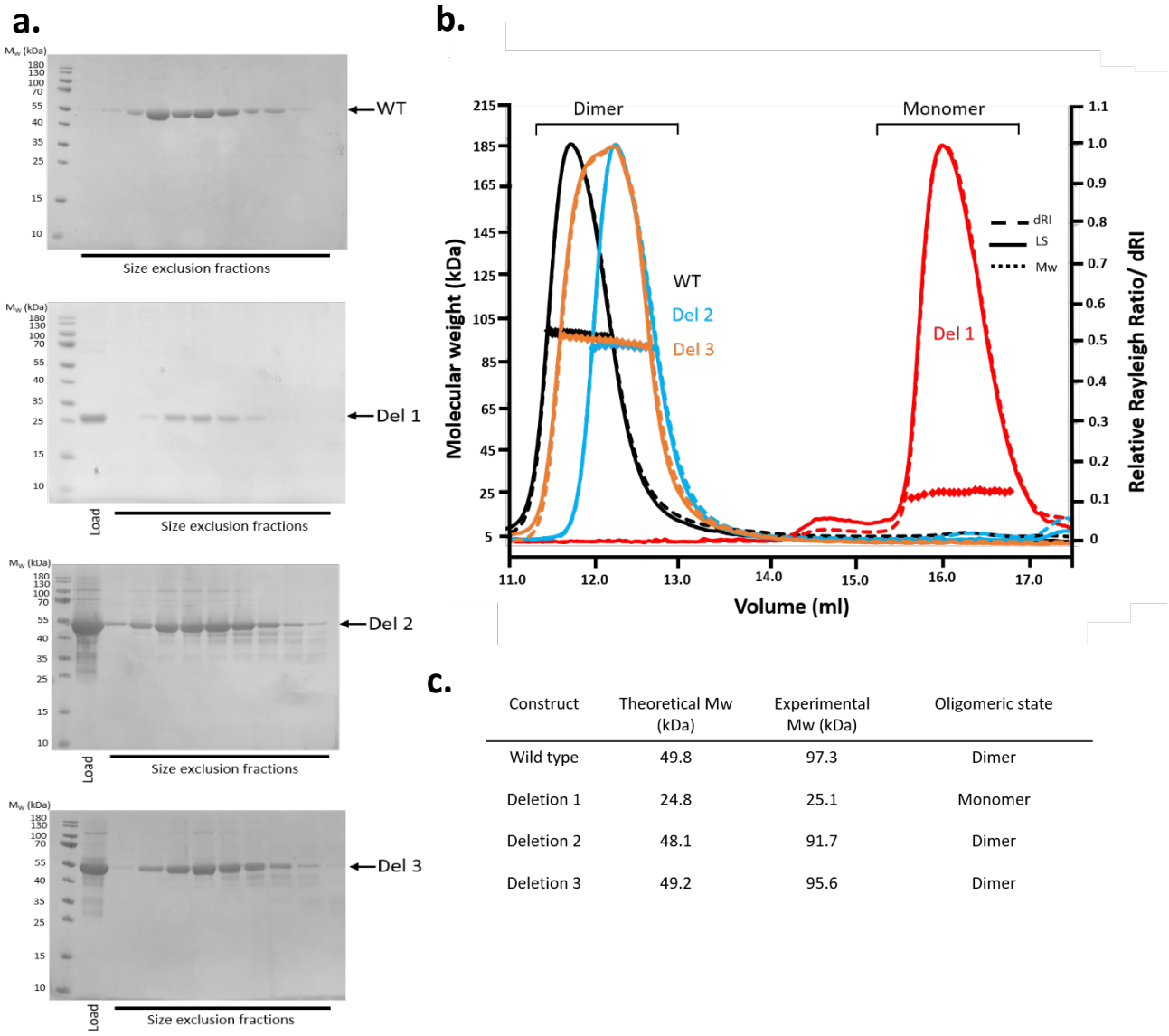


Figure 4.2.24.| SEC-SAXS *P(r)* curves of MBP-C(3)G Deletions
a) Overlaid *P(r)* distributions of MBP-C(3)G^{Del2} (aqua) and MBP-C(3)G^{Del3} (orange) with MBP (dashed-grey) and MBP-C(3)G^{WT} (black). MBP-C(3)G^{Del2} and MBP-C(3)G^{Del3} have maximum dimensions of 690 and 708 Å, respectively, similar to that of MBP-C(3)G^{WT} (723 Å). Indicated by an arrow is an inter-MBP distance peak at ~80 Å. The MBP control has a maximum dimension of 70 Å (blue); b) *P(r)* distribution of MBP-C(3)G^{Del1} revealed a *Dmax* of 330 Å; c) The same *P(r)* distribution profile of MBP-C(3)G^{Del2} and MBP-C(3)G^{Del3} as (a), magnified between 0-180 Å, reveals the inter-MBP peak of MBP-C(3)G^{Del2} (aqua) and MBP-C(3)G^{Del3} (orange) at ~80 Å (highlighted by an arrow).

The two small deletions, MBP-C(3)G^{Del2} and MBP-C(3)G^{Del3} yielded *Dmax* values similar to that of MBP-C(3)G$_{WT}$ (black trace – 723 Å) at 690 Å (cyan trace) and 708 Å (orange trace), respectively (Figure 4.2.24.a). As expected, the MBP-C(3)G^{Del1} monomer has a much smaller *Dmax*, determined at 330 Å (Figure 2.2.24.b). Both MBP-C(3)G^{Del2} and MBP-C(3)G^{Del3} *P(r)* curves have a short inter-MBP distance peak at ~ 80 Å (shown by an arrow, Figure 2.2.24.c), correlating to the distance between N-terminal MBP-molecules within the C(3)G dimer. This observed distance indicates that the helices of C(3)G^{Del2} and C(3)G^{Del3} are in a parallel arrangement, suggesting that the two in-frame deletions do not disrupt the orientation of α-helices. However, further studies using the C-terminal and double MBP constructs of the C(3)G deletions also need to be analysed by SEC-SAXS to increase confidence in this result.

The MBP-tag were removed by incubating with TEV-protease followed by anion-exchange chromatography and SEC. SDS-PAGE analysis of the SEC profile of the C(3)G deletions and wild-type are shown in Figure 4.2.25.a. SEC-MALS analysis was used to determine the absolute molecular weights of the cleaved C(3)G constructs, to confirm the findings of the MBP-fusion constructs. As expected, C(3)G^{Del2} (cyan trace) and C(3)G^{Del3} (orange trace) form a dimer in solution with calculated molecular weights of 91.7 and 95.6 kDa, respectively (theoretical dimers – 96.2 and 98.4 kDa) (Figure 4.2.25.b). C(3)G^{Del1} elutes as a single peak with a molecular mass of 24.8 kDa (theoretical monomer – 25.1 kDa), indicating that deleting a large region of the coiled-coil domain disrupts dimer formation (red trace). The experimental and theoretical molecular masses of the C(3)G deletions are summarised in Figure 4.2.25.c.

In order to determine if the three C(3)G coiled-coil deletions reduced the stability of the C(3)G dimer, we ultilised SEC-MALS to analyse the protein constructs at different concentrations. A dilution series of C(3)G^{WT} and C(3)G deletion mutants was subjected to analysis by SEC-MALS and the stability was analysed by molecular mass determination. The concentration at which the protein complex is loaded onto the SEC column undergoes a ~5-fold dilution before being analysed by the MALS device.



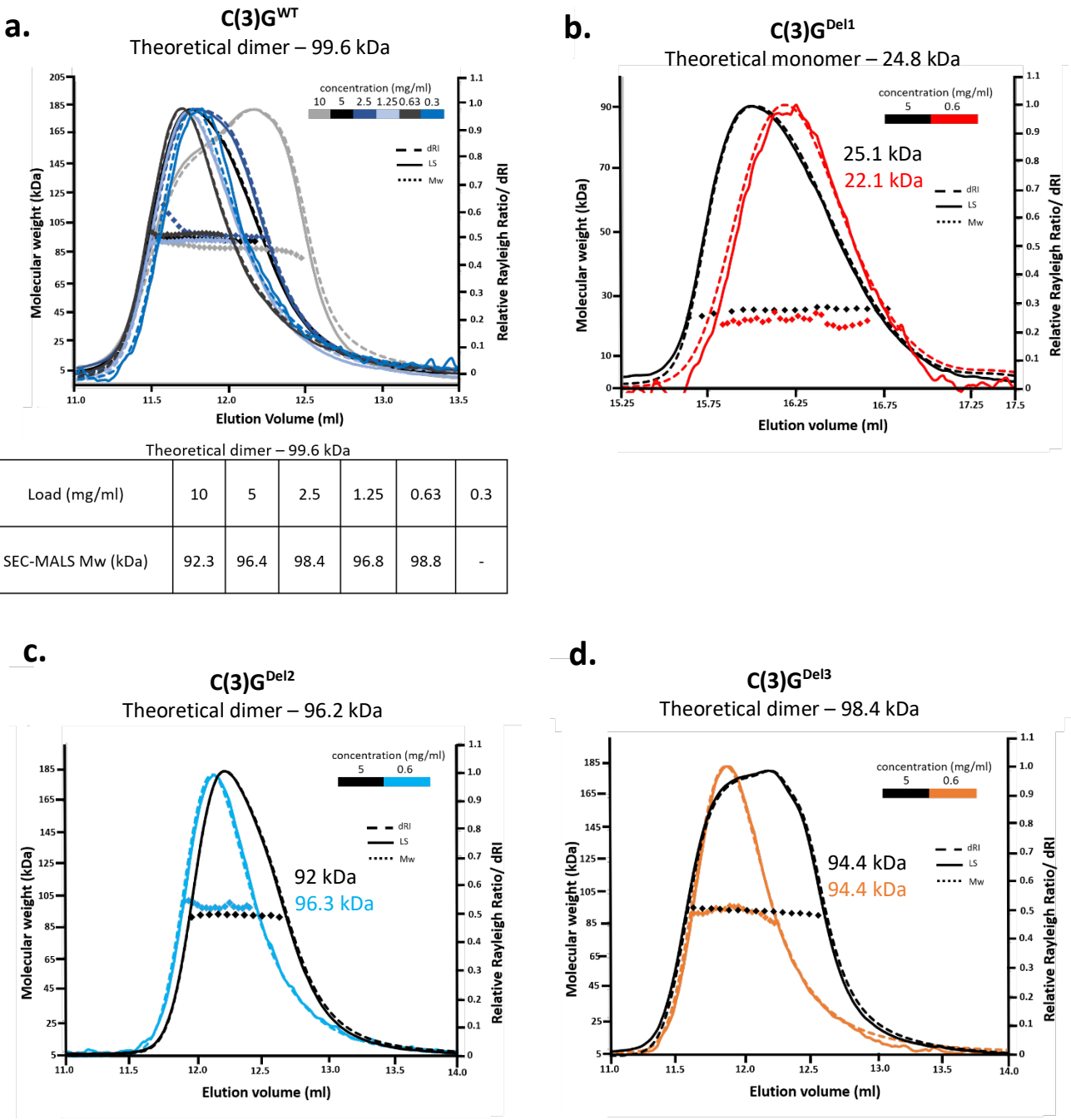
c.

Construct	Theoretical Mw (kDa)	Experimental Mw (kDa)	Oligomeric state
Wild type	49.8	97.3	Dimer
Deletion 1	24.8	25.1	Monomer
Deletion 2	48.1	91.7	Dimer
Deletion 3	49.2	95.6	Dimer

Figure 4.2.25.| SEC-MALS analysis of C(3)G deletion constructs
a) SDS-PAGE analysis of SEC elution profile of $C(3)G_{226\text{-}650}$ (WT), $C(3)G_{Del1}$, $C(3)G^{Del2}$, and $C(3)G^{Del3}$. b) SEC-MALS analysis of $C(3)G^{WT}$ (black), $C(3)G^{Del2}$ (cyan) and $C(3)G^{Del3}$ (orange) show single peaks with molecular weights of 97.3 kDa, 91.7 kDa and 95.6 kDa, respectively, corresponding to dimeric species. $C(3)G^{Del1}$ (red) also elutes as a single peak with molecular weight of 25.1 kDa corresponding to a monomer. c) Table showing the theoretical and experimental molecular weights of the C(3)G constructs and their oligomeric state.

SEC-MALS analysis determined that wild type $C(3)G_{226\text{-}650}$ elutes as a single peak when loaded at 0.625 mg/ml (analysed at 0.125 mg/ml) with a molecular mass of 98.8 kDa, indicating that the dimer retained (theoretical dimer – 99.6 kDa) (Figure 4.2.26.a). Further dilution to 0.3 mg/ml showed a progressive reduction in molecular weight, however this could be due to the limitations of the MALS. $C(3)G^{Del1}$

loaded at 5 mg/ml and 0.6 mg/ml shows a clear monomer, with a molecular mass of 25.1 and 22.1 kDa (theoretical monomer – 24.8 kDa) (Figure 4.2.26.b.). C(3)G^{Del2} and C(3)G^{Del3} loaded at 0.6 mg/ml determined stable dimers with molecular weights of 96.3 and 94.4 kDa, respectively (theoretical dimers – 96.2 and 98.4 kDa), suggesting that the 15 and 7 amino acid deletions to the coiled coil region of C(3)G do not affect dimerisation and stability (Figure 4.2.26.c. and d.).



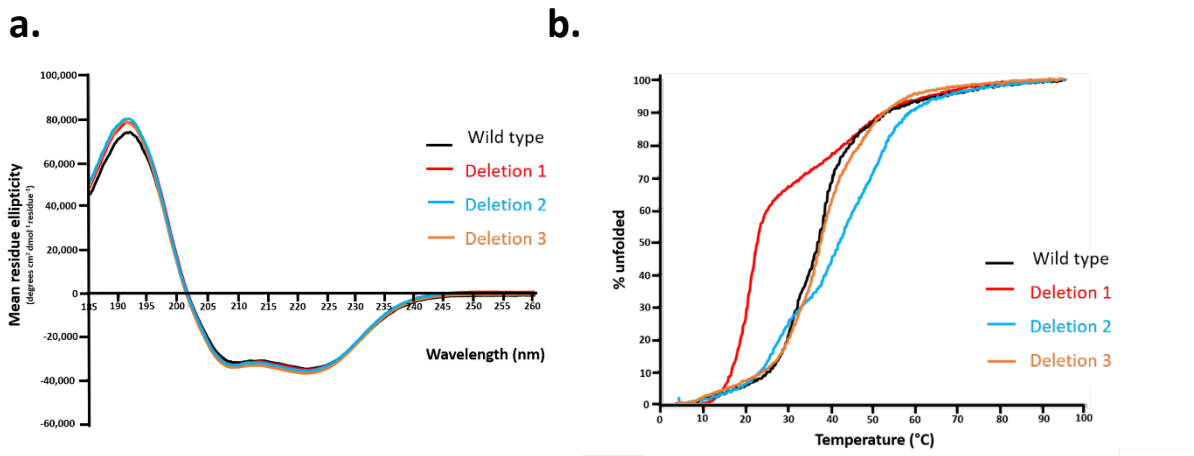
Theoretical dimer – 99.6 kDa

Load (mg/ml)	10	5	2.5	1.25	0.63	0.3
SEC-MALS Mw (kDa)	92.3	96.4	98.4	96.8	98.8	-

Figure 4.2.26.| Dilution series of cleaved deletion C(3)G constructs

Dilution series of C(3)G^{WT}, C(3)G^{Del1}, C(3)G^{Del2}, and C(3)G^{Del3} to deduce the whether the deletions within the coiled-coil domain reduce protein stability. The protein that is loaded onto the SEC column undergoes an approximate 5-fold dilution, therefore the sample analysed by MALS is ~5-fold the concentration that was loaded. The MALS experimental molecular weight was determined for a series of dilutions. a) C(3)G^{WT} forms a clear dimer (theoretical dimer – 99.6 kDa) when loaded at both 1.25 and 0.63 mg/ml. At 0.3 mg/ml the molecular weight was unable to be interpreted. The dimer is largely retained at very low concentrations suggesting high stability. b) C(3)G^{Del1} forms a monomer when loaded at 5 and 0.5 mg/ml, suggesting that the large deletion causes loss of dimerisation but a very stable monomer is formed. c-d) C(3)G^{Del2} (c) and C(3)G^{Del3} (d) maintain dimer formation (theoretical dimer - 96.2 and 98.4 kDa, respectively) at 0.5 mg/ml suggesting that the two deletion constructs are very stable.

Far-UV CD spectroscopy of C(3)G deletions, as expected, revealed characteristic spectrum of α-helical proteins with minima 208 nm and 222 nm, closely matching the spectrum of $C(3)G^{WT}$ (black) (Figure 4.2.27.a.). Deconvolution of C(3)G deletions 1-3 suggested an α-helical content of 86%, 83.9% and 86%, respectively. Thermal denaturation revealed single step unfolding events for $C(3)G^{Del2}$ (cyan) and $C(3)G^{Del3}$ (orange) with an estimated T_m of 42 and 38°C, respectively, at 50% unfolded (Figure 4.2.27.b.). $C(3)G^{Del1}$ has a much lower T_m of 22.4 °C, which is likely due to deletion 1 disrupting the dimer formation (red).



c.

Construct	% α-helix	Melting temp (°C) (50% unfolded)
Wild type	77	34
Deletion 1	86	22.4
Deletion 2	83.9	42
Deletion 3	86	38

Figure 4.2.27.| CD analysis of C(3)G Deletions.
a) Overlaid far UV CD wavelength scans between 260-185nm of $C(3)G_{WT}$ and C(3)G deletions. All four constructs show a typical α-helical trace, with negative peaks at 208 and 222 nm and a positive peak at 193 nm. Deconvolution of the data estimates the secondary structure to range between 77-86% α-helical. b) Overlaid CD thermal denaturation of $C(3)G^{WT}$ and C(3)G deletions measured at 222 nm between 4 and 95°C, estimated melting temperatures of 34, 42, and 38°C for $C(3)G^{WT}$ (black), $C(3)G^{Del2}$ (cyan) and $C(3)G^{Del3}$ (orange), respectively. The $C(3)G^{Del1}$ monomer (red) has a much lower melting temperature of 22.4 °C. c) Table of summarised melting temperatures and percentage helicity of C(3)G constructs.

SEC-SAXS analysis was performed to determine if the three deletion constructs alter the size and shape parameters of the coiled-coil region of C(3)G. The SEC-SAXS data allowed for the calculation of the *Rg* for C(3)G^{Del1}, C(3)G^{Del2}, and C(3)G^{Del3} to be 60.27, 151, and 144.2 Å, respectively, indicating an elongated rod-shape structure (Figure 4.2.28.a.-c.). The *Rc* was measured and determined to be 14.1, 9.0, and 8.6 Å, respectively (Figure 4.2.28.d-f.). The *Rc* of C(3)G^{WT} was measured at 9.0 Å (Figure 4.2.2.d.). This suggests that the *Rc* of both C(3)G^{Del2} and C(3)G^{Del3} are likely to be a dimeric coiled coil. The determined *Rc* of C(3)G^{Del1} at 14.1 Å, suggests that this deletion causes aberrant folding of C(3)G, as the *Rc* measures the widest part of the species. Furthermore, far-UV CD spectroscopy of C(3)G^{Del1} revealed a T_m of 22.4 °C, indicating that the stability of this construct is greatly reduced. SEC-SAXS is performed at room temperature (20-22 °C), suggesting that C(3)G^{Del1} may be partially unfolded before being analysed, and the SAXS data is dominated by the larger unfolded species.

The *p(r)* profiles of C(3)G^{Del1}, C(3)G^{Del2}, and C(3)G^{Del3} reveal a typical elongated distribution with a sloping tail to an x-axis intercept (*Dmax*) values of 278, 674, and 650 Å, respectively (Figure 4.2.28.g-i.). Overall, SEC-SAXS data suggests that C(3)G^{Del2} and C(3)G^{Del3} in solution structures match that of the wild type construct, possessing lengths (*Dmax*) and widths (*Rc*) typical of elongated α-helical structures.

Through biophysical analysis of the three C(3)G deletion mutations, originally described *in vivo*, we have determined that both C(3)G^{Del2} (Δ 346-361) and C(3)G^{Del3} (Δ 465-471) are structurally comparable to the wild type construct, C(3)G$_{226-650}$. This is contradictive of the *in vivo* characterisation, suggesting that the structure of C(3)G is not causing the severe phenotypes observed in homozygous *C(3)G^{Del2}* and *C(3)G^{Del3}* females (Billmyre et al., 2019). Biophysical findings of the C(3)G deletions are summarised in Table 4.2.1.

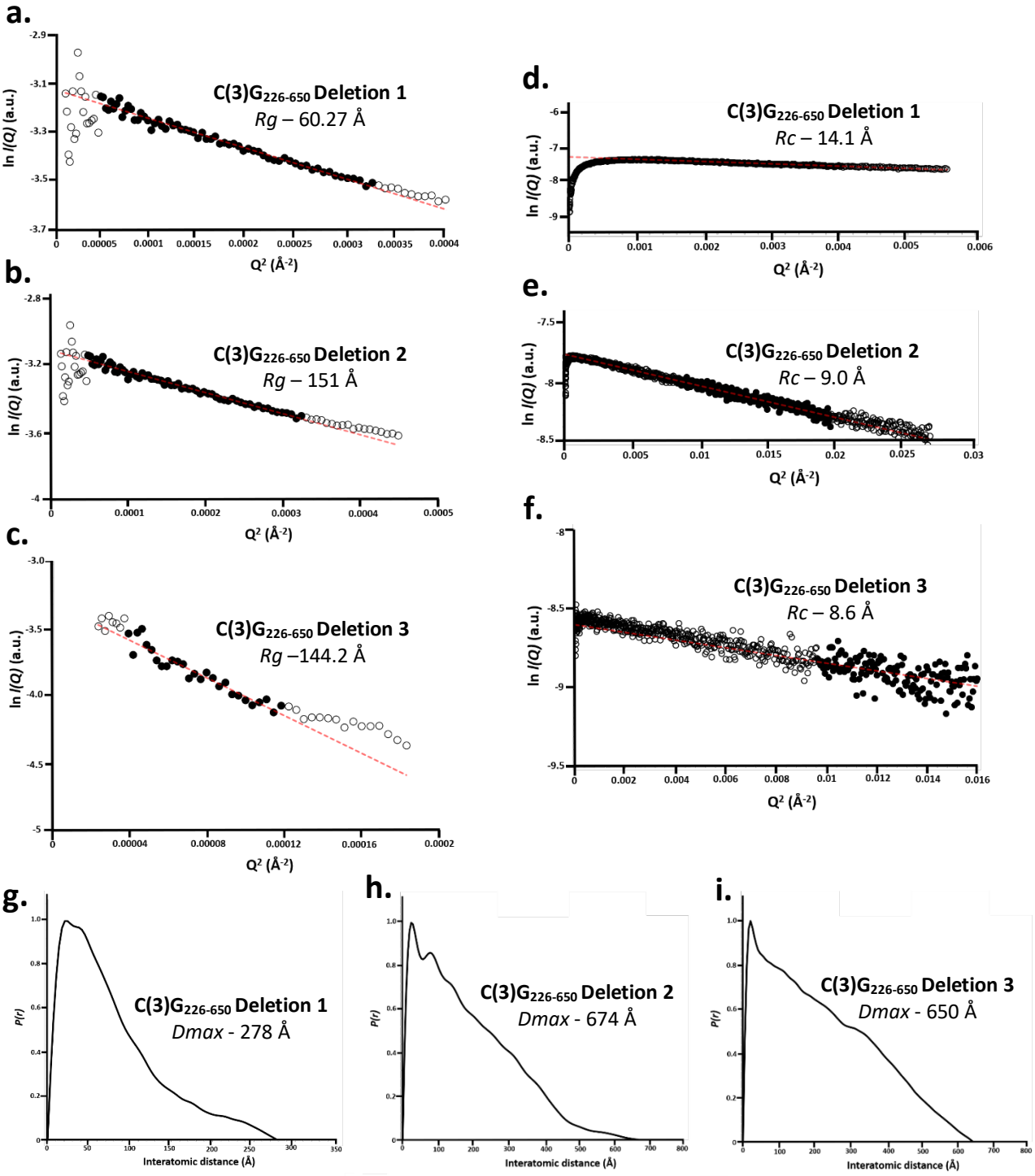


Figure 4.2.28.| SEC-SAXS analysis of C(3)G Deletions

a-c) Guinier analysis of C(3)G^{Del1} (a), C(3)G^{Del2} (b) and C(3)G^{Del3} (c) determined radius of gyration (*Rg*) values of 60.27 , 151, and 144.2 Å, respectively. The real space *Rg* values are 68.7, 158, and 188 and 53.5 Å, respectively. Clear circles represent the complete dataset and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3). d-f) Guinier analysis determined the radius of gyration of the cross-section (*Rc*) of C(3)G deletions. The *Rc* of d) C(3)G^{Del1}, C(3)G^{Del2} and C(3)G^{Del3} was determined to be 14.1, 9.0 and 8.6 Å, respectively. Empty circles represent the complete data and solid circles represent the region used for the fit. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3). g-i) *P(r)* distribution of g) C(3)G^{Del1}, h) C(3)G^{Del2} and i) C(3)G^{Del3} showing maximum dimensions of 278 , 674 and 650 Å, respectively.

4.3. DISCUSSION

4.3.1. The organisation of C(3)G is similar to mammalian SYCP1

C(3)G is a core component of the *Drosophila* SC, constituting the TF. A handful of studies have analysed C(3)G *in vivo*, determining its localisation and showing that it is required for successful SC formation. However, no *in vitro* characteristation of its biochemical or biophysical properties has been performed meaning its mechanism of action remains opaque (Manheim & McKim, 2003; Page et al., 2007; Page & Hawley, 2001). Amino acid sequence analysis and secondary structure predictions have deduced that C(3)G has a central amphipathic α-helical domain, predicted to form discontinuous coiled-coils. Both the N- and C-termini are predicted to be unstructured. This domain architecture is similar to that of its mammalian and yeast counterparts, SYCP1 and ZIP1, respectively. Therefore, it has been generally assumed that C(3)G would have a similar structure, despite having very poor sequence conservation. Through substantial biophysical analysis using SEC-MALS, SEC-SAXS and CD solution studies we have provided the first structural insight into C(3)G. In summary, we have determined that the central α-helical domain forms a stable coiled-coil dimer oriented in a parallel conformation, with parameters similar to that of human SYCP1 (Figure 4.3.2.). Moreover, analysis of the C-terminal globular domain determined a stable monomer that has the ability to complex with dsDNA, also suggesting that C(3)G has a similar structure-function relationship to SYCP1.

4.3.2. C(3)G coiled-coil dimers are in a parallel arrangement

We have shown that the longest soluble construct of the α-helical core, $C(3)G_{226-650}$, forms a dimer in solution. SEC-MALS analysis determined that when analysed at a very low concentration (0.25 mg/ml) dimerisation is not disrupted, suggesting $C(3)G_{226-650}$ is a stable construct. SEC-SAXS analysis in combination with *ab initio* modelling of the α-helical core determined an elongated rod-shaped envelope, with a maximum dimension of 680 Å and cross-sectional radius of 9.0 Å. These dimensions are consistent with individual TF of hamsters and crickets, determined by imaging techniques to be 1.6 and 1.8 nm, respectively (Solari & Moses, 1973). This observed similarity between different species

suggests that such filaments form a conserved structure within SC. We propose that the C(3)G$_{226-650}$ dimers constitute the transverse filaments, visualised by EM, that span the width of the central region of the SC.

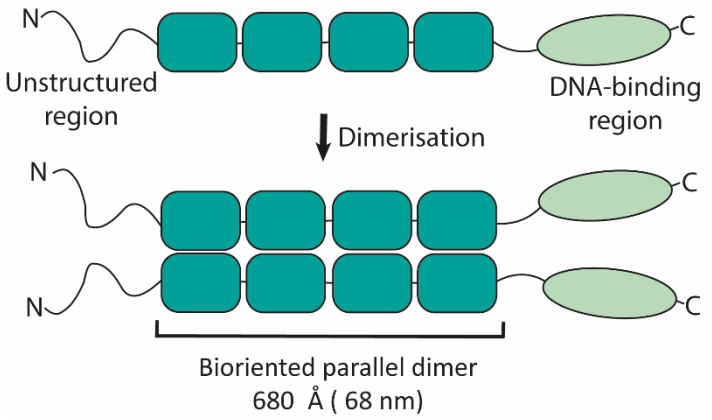


Figure 4.3.1.| Schematic model for C(3)G structure.
Schematic of the full length C(3)G, proposing that the central α-helical domain, residues 226-650, forms dimeric coiled-coil interactions in parallel orientation. The N-terminus of C(3)G (amino acids 1-111) is largely unstructured, with residues 112-226 is contains some helical prediction. The C-terminal domain, residues 644-744 was determined to be a mixture of α-helices and β-sheets, which forms a monomer in solution. The C-terminus of C(3)G has DNA binding ability.

Through analysis of a series of truncations and deletions within the α-helical domain, we have shown that it is possible to disrupt the homodimer. SEC-MALS analysis of C(3)G$_{330-459}$ and C(3)G$_{534-650}$ determined molecular weights in between monomer and dimer, suggesting these two constructs do not form stable dimers. Moreover, deleting over half of the α-helical domain, C(3)G^{Del1} (Δ 340-552), disrupts homodimerisation in solution, suggesting that the α-helical domain is crucial for dimer formation. From the Far-UV CD and denaturation studies, it was observed that all three constructs are highly α-helical and C(3)G^{Del1} and C(3)G$_{330-459}$ have low estimated melting temperatures of 22.4 and 25.6 °C, further suggesting that these fragments are monomeric at room temperature.

Previous studies have used immunoelectron microscopy experiments to show the organisation of C(3)G within the tripartite SC, in which the C- and N-termini reside in the LEs and CEs of the SC, respectively. Despite this, the organisation of C(3)G coiled-coil dimers have not been shown *in vitro*. Utilising SEC-SAXS, we determined that the C(3)G structural core is formed of parallel coiled-coils, in keeping with

the orientation of C(3)G molecules within the SC observed by SIM and super resolution EM studies (Anderson et al., 2005; Cahoon et al., 2017). Furthermore, these results are similar to the determined structure of mammalian SYCP1 (Dunce et al., 2018; Seo et al., 2016).

Dissection of the α-helical domain determined that the central region, amino acid residues 330-533, can tetramerise in solution. The tetrameric state of C(3)G could exist in two forms; two dimers could be arranged in an antiparallel orientation spanning the full width of the SC or alternatively two dimers associate in a parallel side-by-side arrangement that spans half the width of the SC. SEC-SAXS analysis of $C(3)G_{330-533}$ tetramer revealed parameters suggesting the latter. This observation suggests that the tetrameric form may be required as an intermediate in SC assembly, in order to achieve lattice formation.

From the CD denaturation studies it was observed that the dimeric C(3)G α-helical domain constructs had estimated melting temperatures between 33.2 and 42 °C. Endothermic insects, including fruit flies, have a maximum muscle temperature between 40-45 °C, suggesting that the melting temperature of C(3)G is close to the *Drosophila* body temperature (Heinrich, 1974). Furthermore, CD denaturation studies of Zip1 (residues 1-348; N-terminal and part of the coiled-coil domain) and SYCP1 (residues 1-783; N-terminus and α-helical core) determined melting temperatures of 34 and 41 °C, respectively (Dr Lucy Salmon's thesis, Dunce et al., 2018). This conserved feature of low melting temperature suggests a possible function of TF proteins. In addition, the determined maximum dimension of C(3)G α-helical core of 680 Å, is slightly longer than its theoretical length as an extended α-helical coiled-coil (637.5 Å) suggesting that this domain is not a continuous coiled-coil (Lupas & Gruber, 2005). Together, these observations suggest that the α-helical core of C(3)G is an interchanging coiled-coil, providing flexibility and upon stress can unwind into an extended protein providing an ideal preliminary framework for SC formation.

4.3.3. The globular C-terminal domain stabilises C(3)G at the lateral region of the SC

Studies of human SYCP1 have shown that the C-terminus directly interacts with DNA and suggest that SYCP1 is recruited to the chromosomal axes through SYCP1-DNA interactions. This subsequently allows for the alignment, and self-interaction, of the N-terminus within the midline of the SC (Dunce et al., 2018). The C-termini of SYCP1 and C(3)G are highly basic, both have a theoretical pI of 9.7, suggesting that C(3)G C-terminus may also bind DNA and undergo a similar mechanism of assembly within the SC. We have shown that the C-terminus of C(3)G forms a stable monomer in solution and can bind directly to dsDNA, with preference to longer dsDNA constructs. From this result, we can conclude that the C(3)G central dimeric coiled-coil splays apart at the C-termini and binds chromosomal dsDNA as a monomer. We propose that C(3)G may use a similar recruitment mechanism to the chromosomal axes as SYCP1 (Dunce et al., 2018).

The globular nature of the C-terminal tail may mediate, currently unidentified, interactions with LE cohesin proteins, such as C(2)M and ORD, which are required in early steps of SC assembly (Anderson et al., 2005; Webber et al., 2004). These possible interactions could act in concert with DNA binding to achieve the recruitment of C(3)G to the chromosome axis and subsequent stabilisation. To test this interaction studies between the C(3)G C-terminus and *Drosophila* LE proteins need to be explored in the presence and absence of dsDNA, allowing for the determination of a DNA-bound ternary complex. It has been hypothesised that human SYCP1-DNA complexes form integrated assemblies with the axial element proteins SYCP2-SYCP3 (Feng et al., 2017; Winkel et al., 2009).

4.3.4. What holds C(3)G molecules together at the midline?

The N-terminus of mouse SCP1 protein has been shown to self-interact in Y2H screens (J. G. Liu et al., 1996). In addition, a recent study of human SYCP1 protein has shown that the N-terminus forms a tetramer in solution with two dimers interacting through their N-termini to form a head-to-head assembly of a ‘dimers of dimers’. This interaction is thought to facilitate the coiled-coils of SYCP1 dimers that span between the LEs and the midline (Dunce et al., 2018). Despite our efforts, we were

unable to produce soluble material of the N-terminal domain of C(3)G to analyse biophysically. To improve the expression of C(3)G N-termini, construct boundaries could be optimised as it could be possible that the N-terminal region of the α-helical domain may be required for solubility. Another possibility could be to use a eukaryotic expression system such as insect cells, which can introduce post-translational modifications that may be required for protein folding and/or stability. However, it is also a possibility that the N-terminus of C(3)G is not stable when expressed on its own, suggesting that it may require a constitutive binding partner.

This idea could suggest that the C(3)G dimers are attached to the LE through their C-terminal domain and held together at the central region midline by another protein which interacts with the N-termini of C(3)G or the coiled-coil domain to serve as a bridge between them. Structured illumination microscopy (SIM) has shown that Corolla localises to the midline of the central region of the SC (Figure 4.3.2.) (Cahoon et al., 2017; K. A. Collins et al., 2014). Furthermore, a Y2H screen observed a direct interaction between the central element proteins Corolla and Cona, possibly localising C(3)G N-termini with Corolla and Cona at the central element of the SC (K. A. Collins et al., 2014). From these observations, we hypothesise that C(3)G, Corolla and Cona may form a tripartite structure with the central region of the SC.

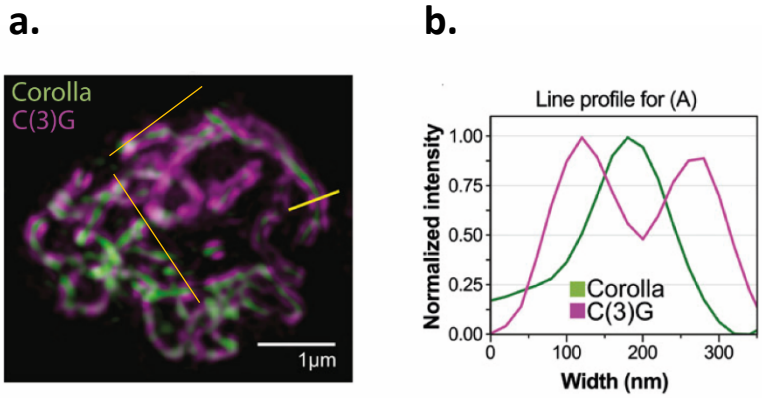


Figure 4.3.2.| C(3)G spans across the central element of the SC.
a) DeltaVision OMX microscopy of C(3)G (magenta) and Corolla (green) of pachytene pro-oocyte. C(3)G forms two parallel tracks with Corolla sandwiched inbetween. b) Representative line profile plot showing the normalised intensity for Corolla and C(3)G. C(3)G (magenta) has a double peak wih a single peak for Corolla (green).

4.3.5. Model for the assembly of the *Drosophila* transverse filament

In order to understand the complete architecture of C(3)G analysis of the full-length protein would have been desirable. In the absence of the complete protein structure, we can propose a model through combining the characterisation of the C-terminal region, the central α-helical domains, and the studies of the human TF protein, SYCP1 (Dunce et al., 2018). We have deduced that C(3)G shares common features with previously characterised mammalian SC proteins in that it adopts an α-helical homo-oligomeric coiled-coil structure. Structural analysis of the three in-frame deletions of C(3)G have deduced that the α-helical domain determines the diameter of the SC, as well as providing stability. Stimulated emission depletion (STED) analysis of homozygous $C(3)G^{Del1}$ females show a reduction in distance between C(3)G C-termini from 118.4 nm in the wild type to 67.8 nm in the $C(3)G^{Del1}$ mutant (Billmyre et al., 2019). This result matches the theoretical reduction of distance between the C-termini of C(3)G, thus inferring that C(3)G dimers are in an parallel arrangement. Furthermore, this data is complementary to the observed parameters of $C(3)G^{Del1}$ when analysed by SEC-SAXS.

Analysis of the C(3)G α-helical core in isolation to the N- and C-terminal domains, suggest that these globular end domains are not required for homo-oligomerisation. This finding supports the idea that C(3)G dimers are in a side-by-side parallel arrangement, and the C- and N-terminal domains reside in the lateral and central regions of the SC, respectively (Figure 4.3.3.a.). We propose that possible tetramerisation occurs through stacking of dimers to form a four helical bundle which spans across half the width of the central region and serves as a building block for the recruitment and assembly of the complete SC.

We propose that C(3)G is recruited to the chromosomal axes through its globular C-terminal domain by direct interaction with the chromosomal loops, and possibly stabilised through additional interactions with LE proteins ORD and C(2)M. Together these interactions anchor C(3)G which spans across the central element. Despite not being able to analyse the unstructured N-terminal domain, we predict that it may interact with the central region of the proteins Cona and Corolla. EM studies have proposed that the CE protein, Cona, is required for zipping together the N-termini of C(3)G filaments at the midline of the SC (Page et al., 2008). This suggests that C(3)G N-termini may self-assemble in an interlocking

pattern with the addition of Cona. Furthermore, Corolla is predicted to contain coiled-coil domains which, may interact with the coiled-coil domains of C(3)G and/or Cona to stabilise the C(3)G transverse filament network (Figure 4.3.3.b.) (K. A. Collins et al., 2014; Newman, Wolf, & Kim, 2000). This proposal is supported by the in-frame deletion mutations, $C(3)G^{Del2}$ (Δ 346-361) and $C(3)G^{Del3}$ (Δ 465-471) which appeared to be structurally unaffected and comparable to wild type. However, were unable to assemble a full-length SC *in vivo* suggesting that this region may be the interaction site for Corolla or Cona (Billmyre et al., 2019). Therefore, we propose a key difference between the TFs of *Drosophila* and human, is that *Drosophila* may require additional proteins, such as the CE proteins Cona and Corolla, to build a stable SC unlike human SYCP1 which forms a preliminary lattice in the absence of any other SC proteins indicating that the assembly of the SC may be different in lower eukaryotic organisms.

In *Drosophila*, we propose that once the central region network of proteins have been assembled, C(3)G extended coiled-coil domain has the ability to undergo conformational changes which may enable communication between synapsed chromosomes during homologous recombination and crossover events. In *drosophila* females the SC is necessary for meiotic exchange, shown by the elimination of meiotic exchange and gene conversion in $C(3)G^{-/-}$ mutant females (Gowen and Gowen, 1922).

Furthermore, studies in mice have shown that the SCs tripartite structure, including the TF protein SYCP1, is required, either directly or indirectly, for both homologous recombination and crossover formation. However, the complex relationship between the SC and recombination initiation and crossing over is still poorly understood (de Boer & Heyting, 2006; Hayashi et al., 2010).

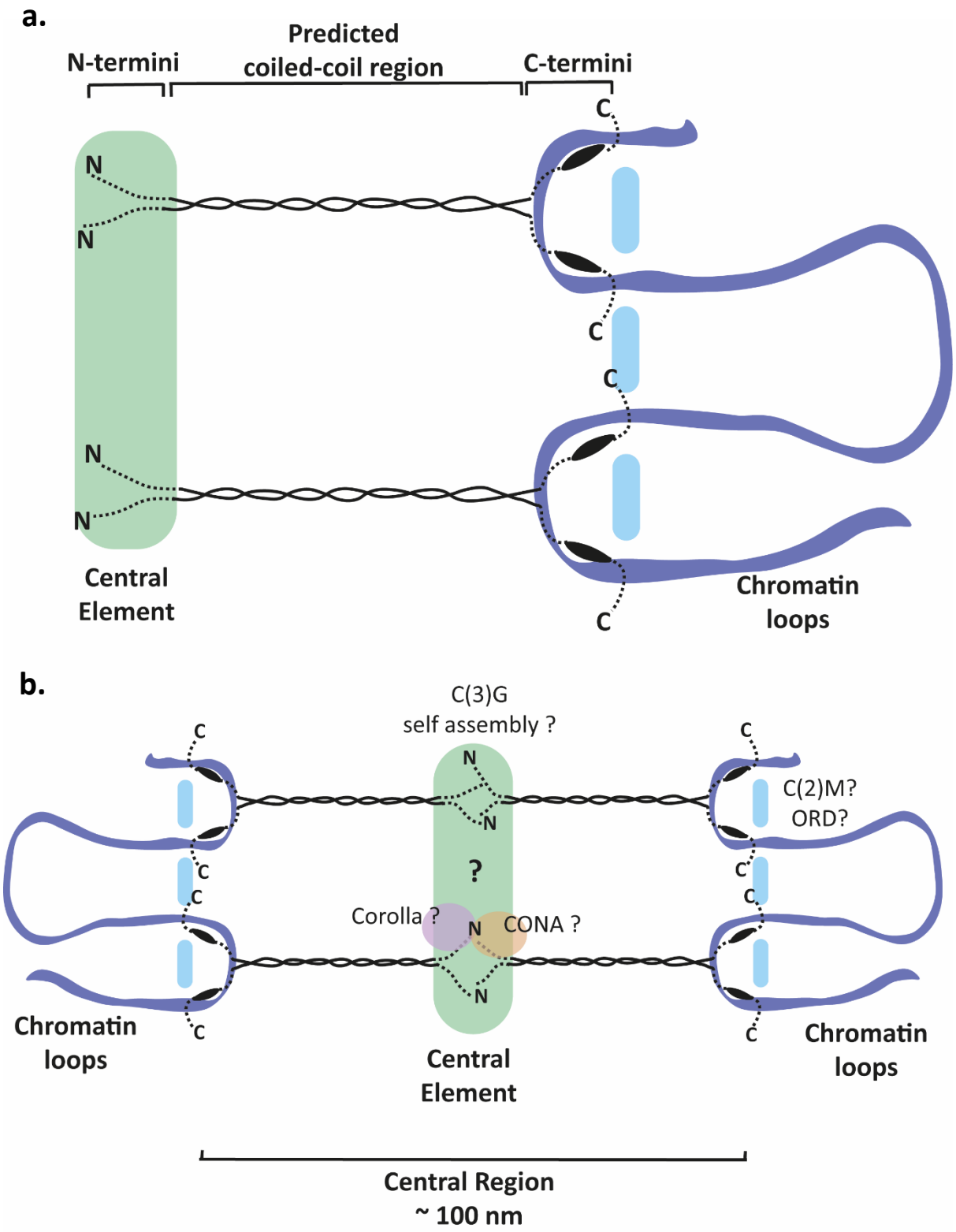


Figure 4.3.3.| Schematic model of C(3)G transverse filament assembly.
a) Model of C(3)G parallel dimer spanning between the central element (N-terminus) and the chromatin loops within the lateral element (C-terminus). The central α-helical domain forms a stable coiled-coil parallel dimer that stretches across the central region. We suggest that this is stabilised through the C-termini directly binding to the chromatin loops. b) Proposed model for the transverse filament structure of C(3)G, forming a preliminary lattice for SC formation.

Chapter 5

Characterisation of the mammalian meiotic recombination proteins: HEI10, RNF212 and RNF212b.

5.1. INTRODUCTION

5.1.1. Homologous recombination during meiosis I

Fundamental to meiosis is the process of homologous recombination between paternal and maternal chromosomes, leading to the reciprocal exchange of genetic material by crossover (CO) events (Page & Hawley, 2003). CO formation promotes the genetic heterogeneity of gametes but also provides physical connections among homologous chromosomes, known as chiasmata. Chiasmata ensure the proper alignment of homologous chromosomes once the SC disassembles, and subsequent reductional segregation of homologous pairs (Petronczki et al., 2003). Each step within recombination is tightly regulated by a subset of specific protein-protein and protein-DNA interactions.

In all organisms, the initiation of meiotic homologous recombination is achieved by the formation of hundreds of double-strand breaks (DSBs) across the genome. DSB formation is catalysed by the topoisomerase VI-like enzyme, Spo11, which generates a transient, covalent protein-DNA intermediate. Following DSB formation, Spo11 dissociates from the DNA strands, and the 5' end of each strand undergoes end resection by the MRN (Rad32-Rad50-Nbs1) nuclease complex to produce 3' single-stranded overhangs (Keeney, 2008; Keeney et al., 1997). This is proceeded by recombinase-mediated strand invasion to the homologous strand and joint-molecule formation, known as a displacement-loop (D-loop), between homologous chromosomes. After strand exchange, current models propose that homologous recombination intermediates are processed by one of two distinct pathways; the CO or non-crossover pathway (NCO) (Pyatnitskaya et al., 2019). The complete mechanism of crossover designation still remains unclear, however both CO and NCO pathways are highly differentiated with respect to molecular intermediates as well as genetic requirements.

Differential stabilisation of strand invasion intermediates underlies the choice between differentiation into COs or NCOs. D-loop intermediates, resulting from DNA strand exchange of a single break end, can be unwound by helicases, Sgs1 (yeast) or BLM (human), which result in D-loop collapse and re-joining of the broken chromosome by synthesis-dependent strand annealing (SDSA) to generate NCO products (Allers & Lichten, 2001; Hunter & Kleckner, 2001). Alternatively, DNA synthesis extends the

D-loop region, thereby providing a single stranded site for annealing of the DSB second end, known as second end capture. Ligation of DNA ends leads to the formation of a double Holliday junction (dHj) and subsequent resolution results in the exchange of genetic material, thus CO formation (Lao et al., 2008; Schwacha & Kleckner, 1995).

In both yeast and mammals, there are two classes of mechanisms, class I and II, that can transform DNA strand invasion intermediates into COs. Class II 'non-interfering' COs are achieved by Mus81-Mms4 endonuclease proteins which act independently of dHj resolution to promote a distinct subset of COs (De los Santos et al., 2003; Guillon et al., 2005). Alternatively, COs are created by the predominant class I 'interfering' pathway, with between 75-100% of total COs being resolved by this mechanism. CO interference is the non-random placement of COs along the chromosomes resulting in the chiasmata being evenly distributed across chromosomes, usually forming a single CO per chromosome arm (Berchowitz & Copenhaver, 2010; Broman et al., 2002).

The factors that control the fate of the highly dynamic D-loop intermediate to promote class I CO formation are facilitated by the well conserved pro-CO factors, known as the ZMM (Zip1-4, Msh4-5, Mer3 and Spo16) proteins in yeast (Shinohara et al., 2008). Strains carrying a mutation for any one of these genes have a reduction in CO products, whereas the number of NCOs are not affected, indicating the ZMM proteins are CO specific (Guillon et al., 2005). Moreover, distinct protein-protein interactions and biochemical activities have been identified among the ZMM proteins, thus suggests that they work at different steps of CO maturation; D-loop stabilisation, second-end capture and dHj stabilisation (Pyatnitskaya et al., 2019).

In budding yeast, most of the D-loops that are stabilised by ZMMs are processed as COs. However, in several other species it has been observed that ZMM foci outnumber the processed COs suggesting that ZMM bound D-loops are not exclusively processed as COs. This suggests that additional steps and proteins, including the MutLγ complex, and post-translational modifications (PTMs), are also required for CO assurance (Kulkarni et al., 2020; Qiao et al., 2014; Snowden et al., 2004; Viera et al., 2009). The meiotic recombination pathway is summarised schematically in Figure 5.1.1.

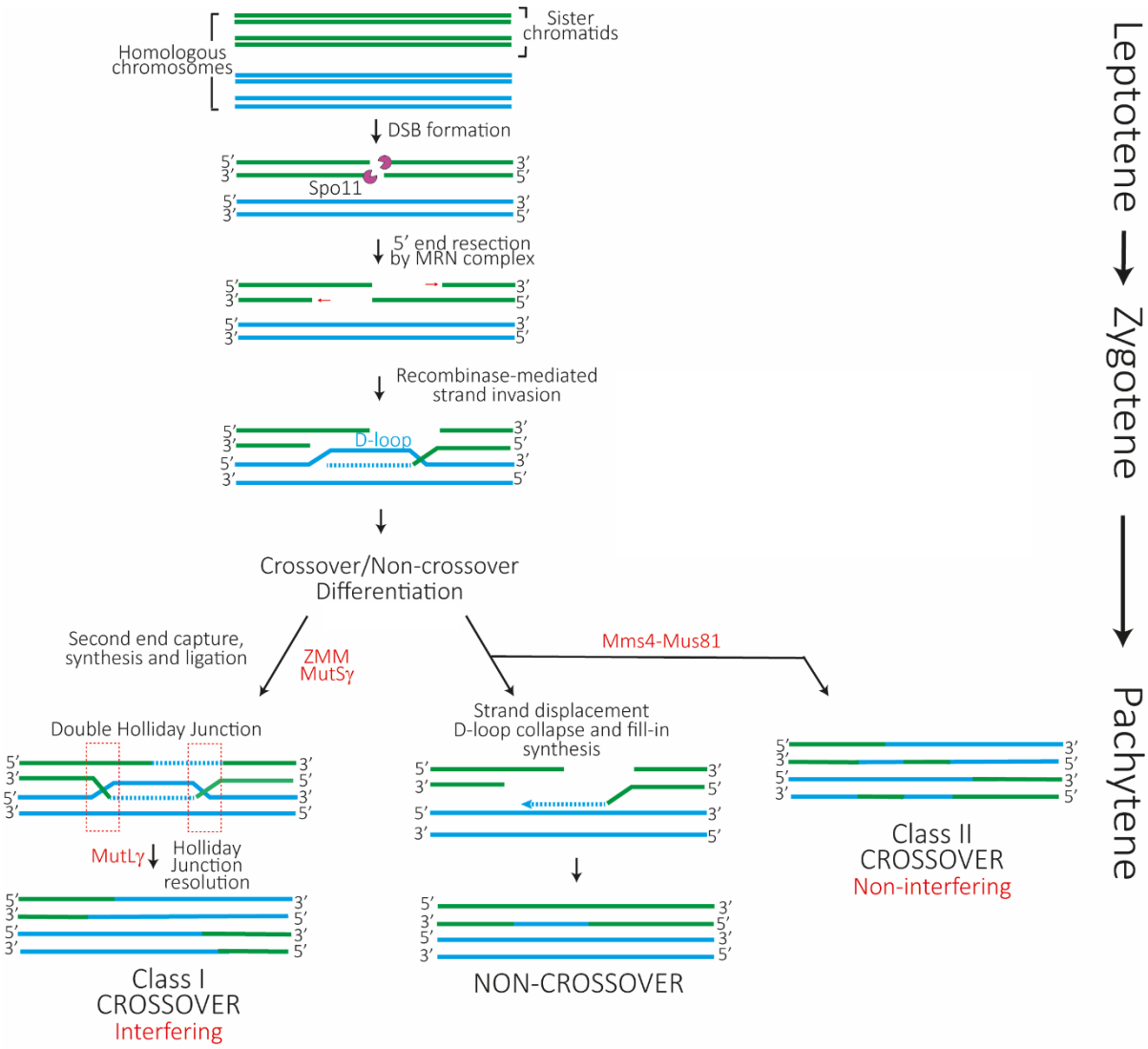


Figure 5.1.1.| Schematic of the meiotic recombination pathway.
The spo11 enzyme introduces hundreds of DSBs throughout the genome. 5' end resection and strand invasion process the DSB to form a D-loop. D-loop intermediates can be repaired into three possible outcomes: non-crossover (NCO), interfering crossover (CO) and non-interfering CO, which are formed from the middle to late pachytene. The interfering CO pathway involves the formation and resolution of a double-Holliday junction by a subset of ZMM proteins, including the complex MutSγ, and later the MutLγ complex. In contrast, synthesis-dependent strand annealing leads to NCO recombination products. Non-interfering COs are formed by Mms4-Mus81 endonucleases.

5.1.2. Meiotic recombination nodules mark the sites of mature crossovers

In budding yeast, mice and humans, despite 200-300 programmed DSBs being introduced per nucleus at leptotene, by the end of pachytene only a subset of these (~75) proceed to form CO products (Figure 5.1.2.) (Lam & Keeney, 2015). Furthermore, the COs produced exhibit wide spacing along the

chromosomes, and higher eukaryotes typically only have a single CO per chromosome arm (Codina-Pascual et al., 2006). These observations suggest that CO formation from DSBs is not a random process, but instead, a highly regulated and controlled process known as CO assurance and interference (Broman et al., 2002; Shinohara et al., 2008). The mechanisms determining which early recombination events are 'selected' to form COs is still not fully understood. Thus far, it has been determined that CO formation is mediated by preferential binding of a subset of protein complexes to maturing CO sites to form structures known as nodules (D. Zickler & Kleckner, 1999). Electron micrographs have shown that these dense structures associate with the chromosome structural axes along the synaptonemal complex (SC) from zygotene to late pachytene of prophase I (Figure 5.1.2.b. and c.) (Albini & Jones, 1987; Del Cacho et al., 2005).

Upon DSB formation by Spo11 and 5' end resection, ssDNA is coated with two eukaryotic members of the RecA recombinase family: the ubiquitously expressed RAD51 DNA recombinase (RAD51) and meiosis-specific recombinase dosage suppressor of Mck1 homolog (DMC1). The RAD51-DMC1 complex binds to 3' ssDNA tails to form helical nucleoprotein filaments, which perform searches for homologous dsDNA and subsequent strand invasion of ssDNA into D-loops (Crickard et al., 2018; Moens et al., 1997). Electron micrographs of zygotene spreads using immunogold localisation and antibodies show the colocalisation of DMC1 and RAD51 to labelled early nodules (ENs), with phosphorylated histone variant γH2AX marking the sites of ssDNA (Anderson et al., 1997). Furthermore, deletion of either *Rad51* or *Dmc1* causes meiotic arrest, associated with abnormal synapsis, indicating that both recombinases play a key role in downstream events. Upon successful strand invasion, the RAD51/DMC1 complex dissociates from ENs at the chromosomal axes (Figure 5.1.2.a. and. b.).

Around half of the ENs successively acquire replication protein A (RPA), BLM helicase and the ZMM MutSγ complex (MSH4/MSH5) and simultaneously lose the RAD51/DMC1 complex from chromosomal axes to form transition nodules (TNs) (Figure 5.1.2.a.) (Neyton et al., 2004). Both RPA and BLM may function to prevent the formation of unnecessary reciprocal recombination events, perhaps by dissolving dHjs into NCO products (Moens et al., 2007). MSH4 and MSH5 proteins share

homology with the DNA mismatch recognition factor MutS, of the bacterial mismatch repair (MMR) complex, and form the MutSγ heterodimer, which is part of the early ZMM machinery (Santucci-Darmanin et al., 2000). The MutSγ complex is conserved amongst species, and in mice it has been shown to be essential for chromosomal synapsis and CO formation. The MutSγ complex binds to joint-molecule (JM) structures, localising to most meiotic recombination sites, where it stabilises nascent JMs to facilitate chromosome synapsis and dHj formation (Kulkarni et al., 2020). Imaging studies have shown that the initial number of MutSγ foci is greater than the foci corresponding to sites of maturing COs, indicating that additional processing by further recombination proteins are required for CO formation (Kneitz et al., 2000).

CO formation relies on the downstream endonuclease MutLγ complex consisting of the MLH1-MLH3 heterodimer (Guillon et al., 2005). The MutLγ complex only accumulates at 10-15 % of MSH4 foci sites, and biases dHj resolution to specifically form COs, thereby forming mature recombination nodules (RNs) (Figure 5.1.2.) (Santucci-Darmanin et al., 2002; Santucci-Darmanin et al., 2000). Co-immunostaining analyses of mouse chromosome spreads for MSH4 (MutSγ) and MLH1 (MutLγ) show that MutLγ foci appear at mid-pachytene and only colocalise to a subset of MutSγ foci. The observed colocalisation was transient and by late pachytene MSH4 foci disappeared but MLH1 foci persisted (Kulkarni et al., 2020). These observations suggest that the MutSγ complex recruits and activates the MutLγ endonuclease required for CO formation (Figure 5.1.2.a.).

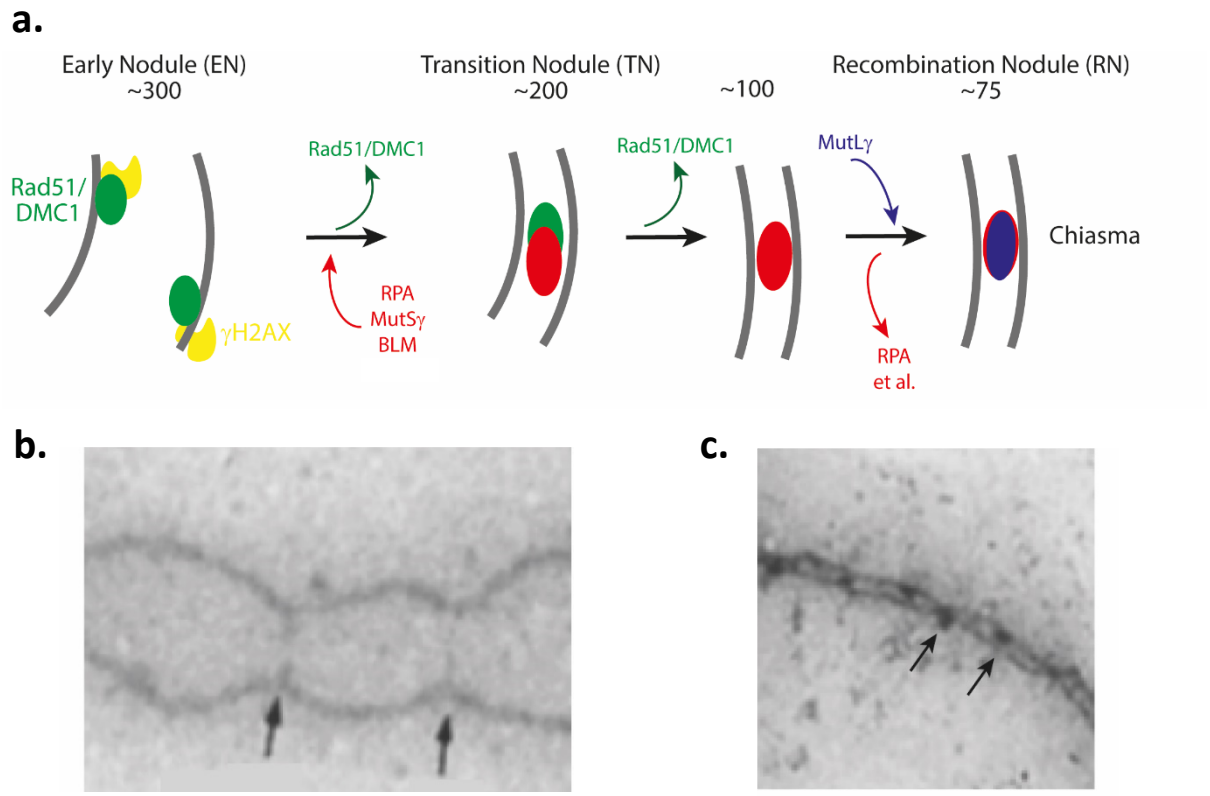


Figure 5.1.2.| Recombination nodules mark the positions of chiasmata along SCs.
a) Schematic showing the development of late recombination nodules (RN). During zygotene, between 200-400 early nodules (ENs) consisting of the recombinase proteins Rad51 and DMC1 (green) mark the sites of strand invasion DSB intermediates, shown by association with ssDNA marker γH2AX (yellow). ENs are transformed into ~200 SC-associated transition nodules (TN) by the recruitment of ZMM proteins (red) whilst losing RAD51-DMC1 recombinases. In mice spermatocytes, ~25 of the TNs recruit the MutLγ complex and in turn lose the ZMM proteins forming recombination nodules (RN). RNs mark the sites of chiasmata and COs (blue). The remaining TNs are resolved as non-crossovers. b-c) Electron micrographs showing the formation of RNs in *S. cerevisiae*. b) Close axis association via interhomologue bridges (highlighted by arrows) initiated by the recruitment of ZMM proteins. c) Fully aligned synapsed chromosomes with recombination sites shown by electron dense RNs (highlight by arrows) between the synapsed homologous chromosomes. RNs are spherical structures with a diameter of ~100 nm. Electron micrograph images taken from Albini & Jones, 1987; Del Cacho, et al., 2005.

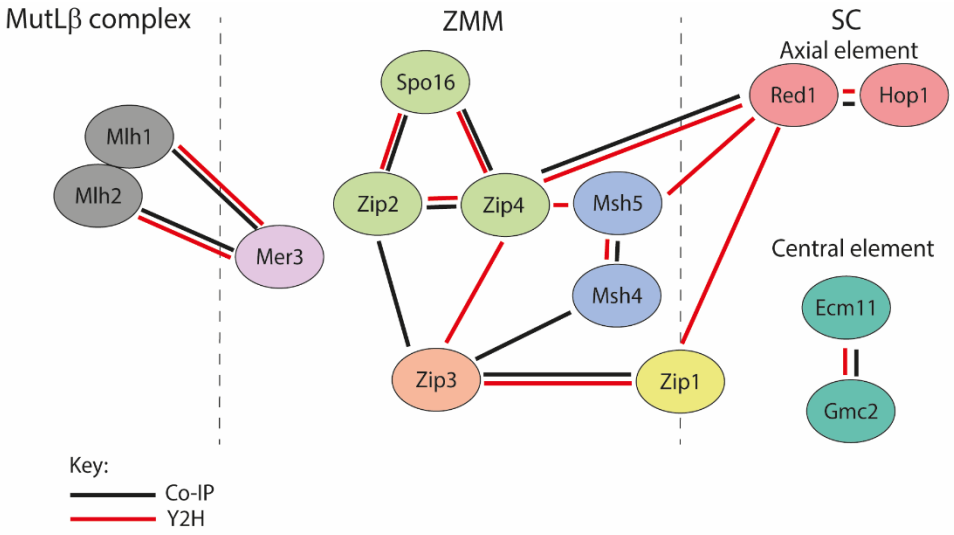
5.1.3. The *S. cerevisiae* ZMM pathway: pro-CO machinery

The commitment towards the class I CO pathway is defined at the time, or soon after, DSB formation during leptotene/zygotene, and is mediated by the ZMM 'pro-CO machinery' proteins (Börner et al., 2004). Thus far, in budding yeast (*S. cerevisiae*) the ZMM group comprises eight proteins; Zip1, Zip2, Zip3, Zip4, Msh4, Msh5, Mer3 and Spo11, which have functions including DNA helicases,

recombinases, scaffold proteins, an SC component and an E3 ligase. Collectively the ZMM proteins function to direct the fate of DSBs towards the CO class I pathway (Guillon et al., 2005; Pyatnitskaya et al., 2019; Shinohara et al., 2008). The action by ZMM components begins soon after DNA strand invasion and it has been shown that the ZMM proteins do not work independently, but instead form subcomplexes to recognise and stabilise specific DNA structures allowing for subsequent CO formation, possibly through transient protein-protein interactions (Guillon, et al., 2005). Moreover, it is proposed that the ZMM proteins also play a role in establishing SC polymerisation by bringing the chromosomal axes (axial elements) of homologous chromosomes together (Zhang et al., 2014). Through multiple studies, using both co-immunoprecipitation (co-IP), Y2H screens and colocalisation imaging techniques the interaction network between the ZMM proteins and the SC have been uncovered in *S. cerevisiae*, summarised in figure 5.1.3. (Pyatnitskaya et al., 2019).

In *S. cerevisiae*, the SC transverse filament protein, Zip1, is one of the earliest pro-CO factors to act during meiotic recombination. Zip1 is proposed to recruit other ZMM proteins to DSB sites, as well as establishing stable homologue juxtaposition through bridging the opposing chromosomal axes (de Boer & Heyting, 2006; Hayashi et al., 2010). Zip2, Zip4 and Spo11 proteins form the 'ZZS' complex, and work collectively to promote the formation of COs. Both Zip2 and Spo11 share homology to the XPF/ERCC1 family of structure-selective endonucleases, and directly bind branched DNA molecules including D-loops and Holliday junctions (De Muyt et al., 2018; Shinohara et al., 2008). Zip4 is proposed to act as a chaperone to mediate interactions with both the SC axial element (AE) protein, Red1, as well as other ZMM proteins including Zip3 and MSH5 (De Muyt et al., 2018). In addition, the meiosis-specific helicase (Mer3) also binds to D-loops, and promotes single-end invasion (SEI) resulting in D-loop stabilisation (Nakagawa et al., 2001). This stable intermediate is then processed into a protected dHj, which are bound and stabilised by the MutSγ DNA clamp complex. The ZSS complex proteins, Mer3 helicase and the MutSγ complex all work cooperatively to promote stable SEI formation. Furthermore, it has been proposed that together these proteins function to bring the chromosome axes in close association (~100 nm), which in turn initiates SC polymerisation (Figure 5.1.2.b. and c.) (Storlazzi et al., 2010).

a.



b.

ZMM protein pathway

S. cerevisiae	Molecular function	*S. macrospara*	*M. Musculus*	*A. thaliana*	*C. elegans*	*D. Melanogaster*
Zip1	SC coiled-coil central element protein	Sme4	SYCP1	ZYP1a ZYP1b	SYP1 SYP2	C(3)G
Mer3	DNA helicase, recruits the MutLβ complex	Mer3	HFM1	AtMER3	-	-
Spo16	Part of the Zip2-Zip4-Spo11 (ZZS) complex	-	Spo16	PTD	-	-
Zip2	Part of the ZSS complex	Zip2	SHOC1/mZIP2	SHOC1	-	-
Zip4	Scaffold protein that has multiple interactions with ZMM and axis proteins	Zip4	Tex11	Zip4	-	-
Zip3	RING finger E3 ligase protein, involved in SUMOylation and/or ubiqutinylation	Hei10	Hei10 RNF212 RNF212b	Hei10	ZHP-3	Vilya Narya Nenya
Msh4/Msh5	MutSγ complex, binds to D-loops and Holliday junction recombination intermediates	Msh4	Msh4/Msh5	AtMsh4/ AtMsh5	-	

Figure 5.1.3.| *S. cerevisiae* ZMM interaction network.
a) Physical interactions observed by co-immunoprecipitation (black lines) or yeast two hybrid (red lines) experiments between ZMM complex of proteins with the MutLβ complex and synaptonemal complex proteins. b) Table summarising ZMM proteins in *S. cerevisiae* and their homologous proteins in the model organisms: *S. macrospara*, *M. musculus*, *A. thaliana*, *C. elegans* and *D. melanogaster*. Figure adapted from Pyatnitskaya et al., 2019.

In addition, a subset of the MutSγ complex foci can recruit and activate the MutLγ heterodimer (MLH1/MLH3) endonuclease complex, which in turn resolves dHjs to exclusively form COs (Kulkarni et al., 2020; Snowden et al., 2004). However, the mechanism for this selection is not completely known. The ZMM protein, Zip3, is an E3 ubiquitin ligase RING-finger protein (Agarwal & Roeder, 2000). Zip3 has been shown to interact with the MutSγ complex (MSH4), Zip1, and the ZSS proteins (Zip2 and Zip4), suggesting Zip3 plays an important role in the recruitment of many ZMM proteins (Figure

5.1.3.a.) (Agarwal & Roeder, 2000; De Muyt et al., 2018). In addition, it has been proposed that Zip3 may function to control CO maturation via SUMOylation and/or /ubiquitination pathways (De Muyt et al., 2014).

5.1.4. RING domain E3 ligases

Many of the meiotic proteins have been characterised and assigned functions within meiosis. Nevertheless, the complex regulation of homologous recombination is still not well understood. Similar to many other cellular processes, the progression of distinct phases of meiosis I, including homologous recombination, is thought to be orchestrated by the synthesis and degradation of cyclins and PTMs, including phosphorylation, ubiquitination and SUMOylation (Dye & Schulman, 2007; Josefsberg & Dekel, 2002).

E3 ubiquitin ligases are a large family of enzymes that catalyse the covalent attachment of a small protein modifier, ubiquitin, to many substrates in eukaryotic cellular processes. The attachment of ubiquitin is achieved by sequential actions of a ubiquitin-activating enzyme (E1), a ubiquitin-conjugating enzyme (E2) and an E3 ligase, summarised in Figure 5.1.4.a. (Buetow & Huang, 2016). Successive ubiquitination rounds can produce substrates with polyubiquitin chains which act as markers for downstream processes including induction of protein-protein interactions, altered localisation and signaling or degradation by the 26S proteosome (Dye & Schulman, 2007; Nath & Shadan, 2009). There are three families of E3 ligases found in eukaryotes, defined by the presence of a HECT, RING or U-box domain, with both RING-finger and U-box E3 ligases facilitating the transfer of ubiquitin by a precise spatial orientation of the E2 and the substrate (Buetow & Huang, 2016; Vander Kooi et al., 2006). RING-domain containing proteins represent the largest family of E3 ubiquitin ligases and are involved in multiple stages of meiosis (Freemont, 2000).

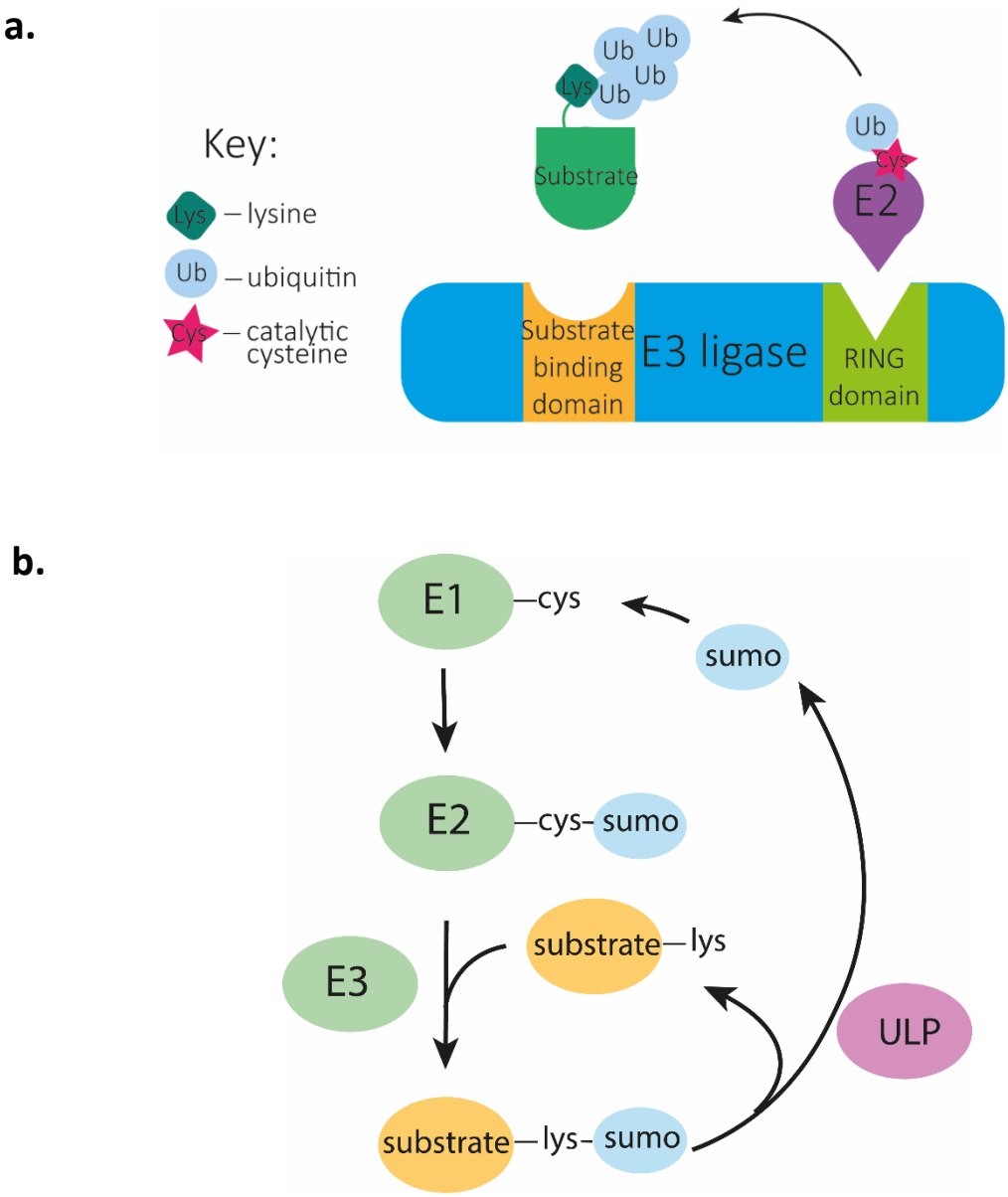


Figure 5.1.4.| RING E3-mediated catalysis.
a) The ubiquitin pathway relies on the sequential actions of a ubiquitin-activating enzyme (E1), a ubiquitin-conjugating enzyme (E2) and an E3 ligase. The E3 ligase harbours a RING domain (E2-ubiquitin binding domain) and a substrate binding domain to facilitate the direct transfer of ubiquitin from E2 to the substrate. This is achieved by the formation of an isopeptide bond between the lysine side chain and carboxyl terminus on the substrate and ubiquitin, respectively. Successive rounds of E3 catalysed reactions can form polyubiquitin labelled substrates recognised by various effectors for downstream signals. In many E3 ligases the substrate targeting mechanism is unknown. b) Overview of the SUMO pathway. SUMO is activated by a cysteine-specific protease (ULP in yeast), which permits binding to the reactive cysteine of the E1 activating enzyme. SUMO is then transferred to the E2 conjugating enzyme and subsequent E3 ligase. SUMO is then transferred to the target substrate by a lysine residue. Protein sumoylation is a dynamic and reversible process. Sumoylated proteins can be desumoylated by the same proteases that convert the inactive SUMO precursor to its reactive form.

The SUMO (Small Ubiquitin-like Modifier) protein pathway is the most studied ubiquitin-like pathway, known to regulate a wide range of cellular events including DNA repair and chromosome segregation (Hooker & Roeder, 2006; Nottke et al., 2017). SUMO proteins share structural similarities to ubiquitin, and like the ubiquitin conjugating pathway, involves the transfer of SUMO by the sequential action of E1, E2 and E3 enzymes to conjugate SUMO molecules onto the lysine residue of the target substrate. SUMO proteins are translated in their inactive form so need to be processed by a cysteine-specific protease, ULP or SENPs in yeast and mammals, respectively. ULPs remove the C-terminal peptide to expose a diglycine motif to generate the active form required for covalent attachment by the E1 activating enzyme. SUMO conjugation is a reversible process achieved by the SUMO proteases which initially activate SUMO (Gareau & Lima, 2010). The SUMO pathway is summarised schematically in Figure 5.1.4.b.

The RING-domain of E3 ligases contain a canonical sequence of eight Zn^{2+} coordinating residues, generally cysteine and histidine, with conserved spacing, Cys-X_2-Cys-$X_{(9\text{-}39)}$-Cys-$X_{(1\text{-}3)}$-His-$X_{(2\text{-}3)}$-Cys-X_2-Cys-$X_{(4\text{-}48)}$-Cys-X_2-Cys (where X is any amino acid). However, it has been observed that variations to the canonical sequence do exist (Deshaies & Joazeiro, 2009) (Figure 5.1.5.a. and b.). Structural studies have determined that the conserved cysteine and histidine residues are buried within the domain core, where they maintain the overall structure. This is achieved by the coordination of two zinc ions in a cross-braced arrangement, which creates a binding platform for the E2 enzyme (Metzger et al., 2014). In addition, a notable feature of RING-domain E3 ligases is their tendency to form homodimers or heterodimers, either through their RING domain or flanking α-helical domains (Metzger et al., 2014). Crystal structures of BRCA1-BARD1 and BIRC7 (PDB codes: 4AUQ and 1JM7, respectively) determined that dimerisation occurs through the RING domains and the flanking C- or N-terminal regions of the RING domain, respectively (Figure 5.1.5.) (Brzovic et al., 2001; Dou et al., 2014). Functional studies of RING-domain E3 ligases with mutated zinc binding residues have shown to disrupt the RING domain structure and subsequently inactivate or reduce ubiquitination activity. These observations suggest that zinc coordination is crucial for folding and ligase activity (Buetow & Huang, 2016; Deshaies & Joazeiro, 2009).

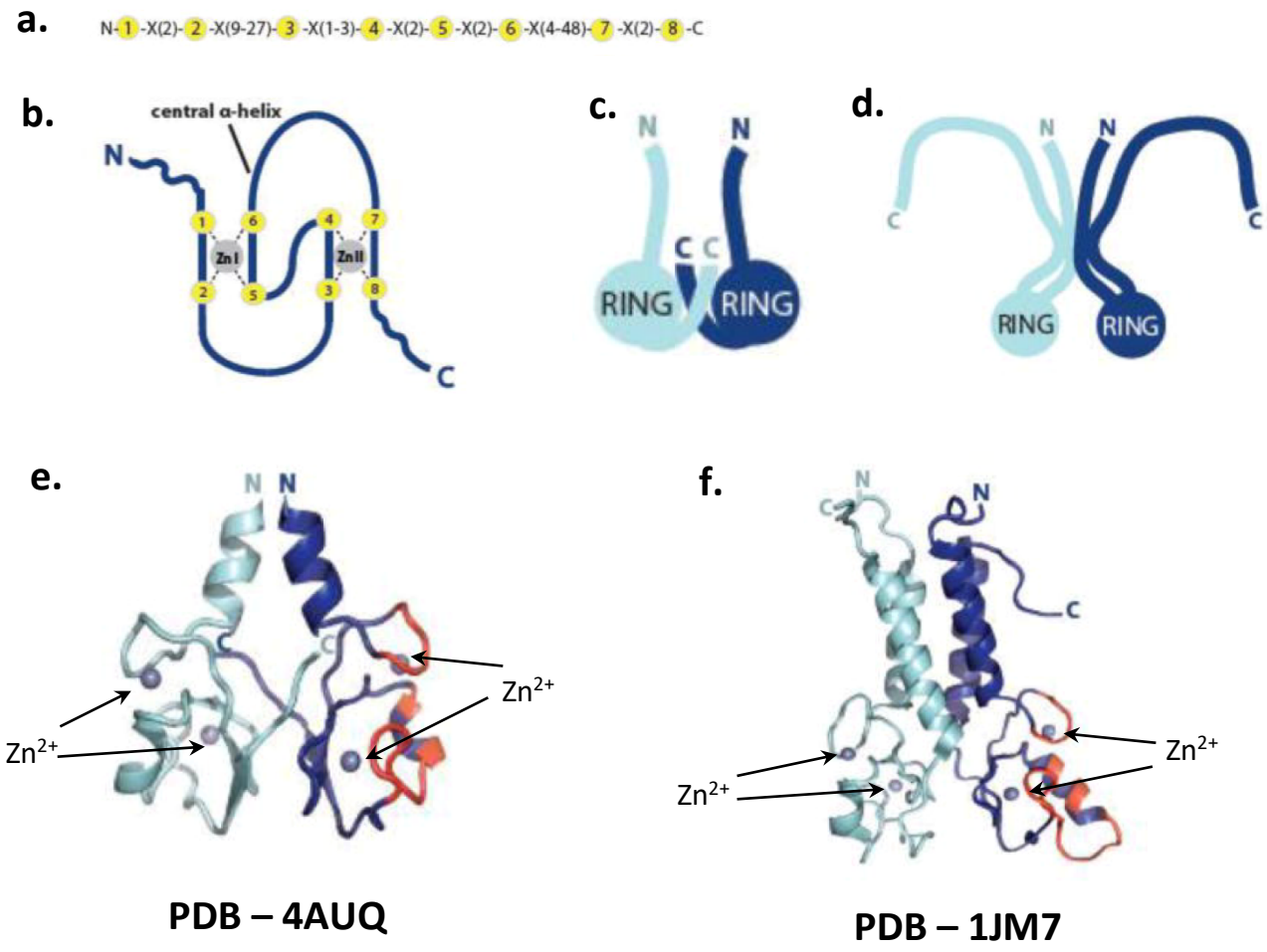


Figure 5.1.5.| RING domains of E3-ligases typically form homo- or hetero-dimers.
a) Canonical RING domains have a conserved sequence containing eight histidine or cysteine Zn^{2+} coordinating residues separated by conserved spacing. b) Zn^{2+} coordinating residues form a cross-braced arrangement with a central α-helix, creating a binding platform for E2s. c-d) Schematic showing the two ways E3 ligases dimerise. Dimerisation can occur through the c) RING domain or d) N- or C-terminal sequences flanking the RING domain. e-f) X-ray crystal structures of e) BRCA1-BARD1 heterodimer (PDB code – 4AUQ) and f) BIRC7 homodimer (PDB code – 1JM7), showing dimerisation occurs through the RING domain and the flanking domains, respectively. Both dimeric structures show that each RING domain monomer coordinates two Zn^{2+} atoms.

5.1.5. The Zip3- and Hei10-families of E3 ligases

In mice, the SUMO-modification and ubiquitin-proteosome systems (SMS and UPS, respectively) are key regulation systems of cellular proteostasis and are implicated in major events in meiosis (Eifler & Vertegaal, 2015; Hooker & Roeder, 2006). Moreover, SMS and UPS have essential roles during

prophase I of meiosis including axis morphogenesis, homologous synapsis and recombination in yeast, fungi, worm and mouse. However, these roles remain poorly characterised (Prasada Rao et al., 2017).

The meiotic recombination process has now been studied in a number of model organisms, each providing a mechanistic insight, whilst determining unique features across different species. In budding yeast, the ZMM protein, Zip3, encodes a RING finger motif E3 ligase, that is recruited to early recombination nodules (RNs) and acts upstream of all other ZMM proteins. Moreover, Zip3 has been proposed to couple the processes of SC formation with recombination during prophase I (Shinohara et al., 2008). Zip3 E3 ligase SUMO activity and phosphorylation is required for meiotic progression, as well as SC polymerisation by the accumulation of Zip1. Meiotic chromosome surface spreads reveal that SUMO localises in a Zip1-dependent manner, specifically to synapsed regions of meiotic chromosomes (Figure 5.1.6.) (Hooker & Roeder, 2006). In addition, chromatin immunoprecipitation (ChIP) sequencing analysis demonstrates that further Zip3 recruitment is dependent on Zip1, leading to a second wave of Zip3 confined to dHjs, which interacts with the MutSγ complex (Gray & Cohen, 2016). These observations suggest that Zip3 may be involved in biasing the class I CO pathway.

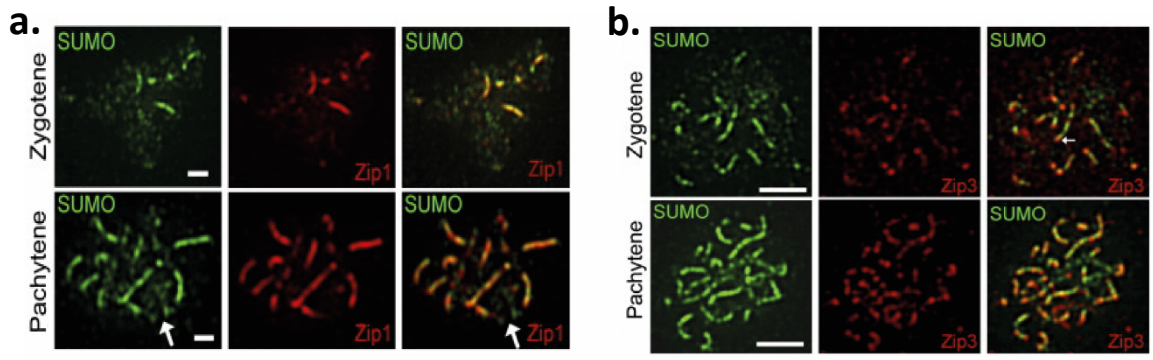


Figure 5.1.6.| SUMO localises to synapsed chromosomes
Surface-spread nuclei from wildtype meiotic cells were stained with antibodies to SUMO (green) and GFP for cells expressing a) Zip1-GFP (red) and b) Zip3-GFP (red) during zygotene and pachytene of meiosis I nuclei. Regions of colocalisation are shown in the merged images. a) During zygotene, chromosomes begin to synapse and SUMO (green) localises to foci in short stretches. By pachytene chromosomes are fully synapsed and SUMO coats the lengths of the chromosomes. Zip1 shares a similar staining pattern to SUMO, with clear colocalisation at zygotene and pachytene. b) In both zygotene and pachytene stages there is partial colocalisation of Zip3 (red) and SUMO (green), suggesting that SUMO is involved in synapsis initiation. The scale bar represents a) 1 µm and b) 2.5 µm. Image adapted from Hooker & Roeder, 2006.

Zip3 orthologues have been discovered in other organisms including lower eukaryotes and vertebrates; ZHP-3 (*C. elegans*), Vilya (*D. melanogaster*) and HEI10 (mouse, human and *Arabidopsis*). In all of these organisms, E3 ligases are required for wildtype crossover levels. However despite structural similarities, there are some subtle differences in their functions (Agarwal & Roeder, 2000; Bhalla et al., 2008; Chelysheva et al., 2012; Lake et al., 2015; Shinohara et al., 2008; Strong & Schimenti, 2010).

The E3 ligase family can be divided into two major subgroups; the Hei10-like and the Zip3-like family, dependent on their UPS and SUMO activity, respectively (Chelysheva et al., 2012; De Muyt et al., 2014). Plants, yeast, fungi and worms encode a single E3 ligase protein, either a Zip3 or a HEI10 orthologue, that act solely as a SUMO or ubiquitin E3 ligase (Bhalla et al., 2008; Chelysheva et al., 2012; Cheng et al., 2006). In comparison, vertebrates encode a protein from each family, which have specific SUMOylation or ubiquitination activity (Qiao et al., 2014; Reynolds et al., 2013). Moreover, multiple Zip3-related proteins have been identified in mammals (RNF212 and RNF212b) and *Drosophila* (Vilya, Narya and Nenya), suggesting that the function and complexity of E3 ligase proteins may differ between species (Johnston et al, 2020; Lake et al., 2019).

In mice, the E3 ligase proteins, RNF212 and HEI10, have been proposed to function in biasing the CO pathway by stabilising specific recombination factors by a RNF212-mediated SUMOylation and Hei10-mediatated ubiquitinoylation relay (Rao et al., 2017; Qiao et al., 2014). Thus, protecting the MutSγ complex from proteolysis and recruiting the MutLγ complex to resolve dHj intermediates into COs and subsequently forming chiasmata. Furthermore, in *Arabidopsis*, HEI10 has been shown to interact with MSH5 (MutSγ complex), suggesting that a similar mechanism also occurs in plants (Chelysheva et al., 2012).

5.1.6. HEI10 is required for meiotic COs in mice

The human E3 ubiquitin ligase protein, human enhancer of invasion-10 (HEI10) was first identified from the HeLa cell library based on its ability to promote yeast agar invasion and filamentation (Toby et al., 2003). A Y2H screen of human HEI10 identified mitotic cyclin B1 and an E2 ubiquitin-

conjugating enzyme (UbcH7) as HEI10-interacting proteins. Mutational studies within the RING finger motif and Cdc2/cyclin binding and phosphorylation sites caused a delay in G2/M transition in mitotic cells, suggesting HEI10 alters mitotic progression through regulation of cyclin B levels by a ubiquitin system (Toby et al., 2003). Furthermore, forward genetic mutagenesis screens and positional cloning in mice, revealed that a single point mutation, mei4, creates a mutant allele of *Hei10*. The mei4 mutation creates a 24 amino acid deletion within the highly conserved N-terminal region of HEI10 (amino acid residues 76-100), resulting in sterile females and recessive males (Strong & Schimenti, 2010; Ward et al., 2007). This indicates that in addition to its mitotic role, HEI10 is also essential for successful meiosis.

Immunocytological analyses of HEI10 *mei4/mei4* homozygous mice meiocytes revealed no abnormalities in NCO formation or chromosome synapsis. However, homologues failed to maintain interhomolog associations. This suggests the absence of COs and subsequent chiasmata (Qiao et al., 2014; Strong & Schimenti, 2010). Moreover, MutLγ foci were absent but the RNF212 foci appeared higher than WT spermatocytes. This suggests that loss of MLH3 (MutLγ component) disrupts the concentration of HEI10 at recombination sites, thus suggesting that HEI10 acts both up and downstream of the MutLγ complex in CO differentiation and maturation. In addition, HEI10 foci were observed to be present from late zygotene with numbers peaking at late pachytene (Figure 5.1.7.a. and b.) (Qiao et al., 2014; Strong & Schimenti, 2010). HEI10 is also required for the turnover of RNF212 after synapsis, that accumulates selectively at future CO sites (Qiao et al., 2014). At mid-pachytene, super-resolution structured illumination microscopy (SIM) studies show that HEI10 colocalises with the early recombination markers, MSH4, RNF212, CDK2 as well as the recombinase MutLγ complex protein MLH1, which mediates the final step of CO formation (Figure 5.1.7.c.-f.) (De Muyt et al., 2014; Qiao et al., 2014).

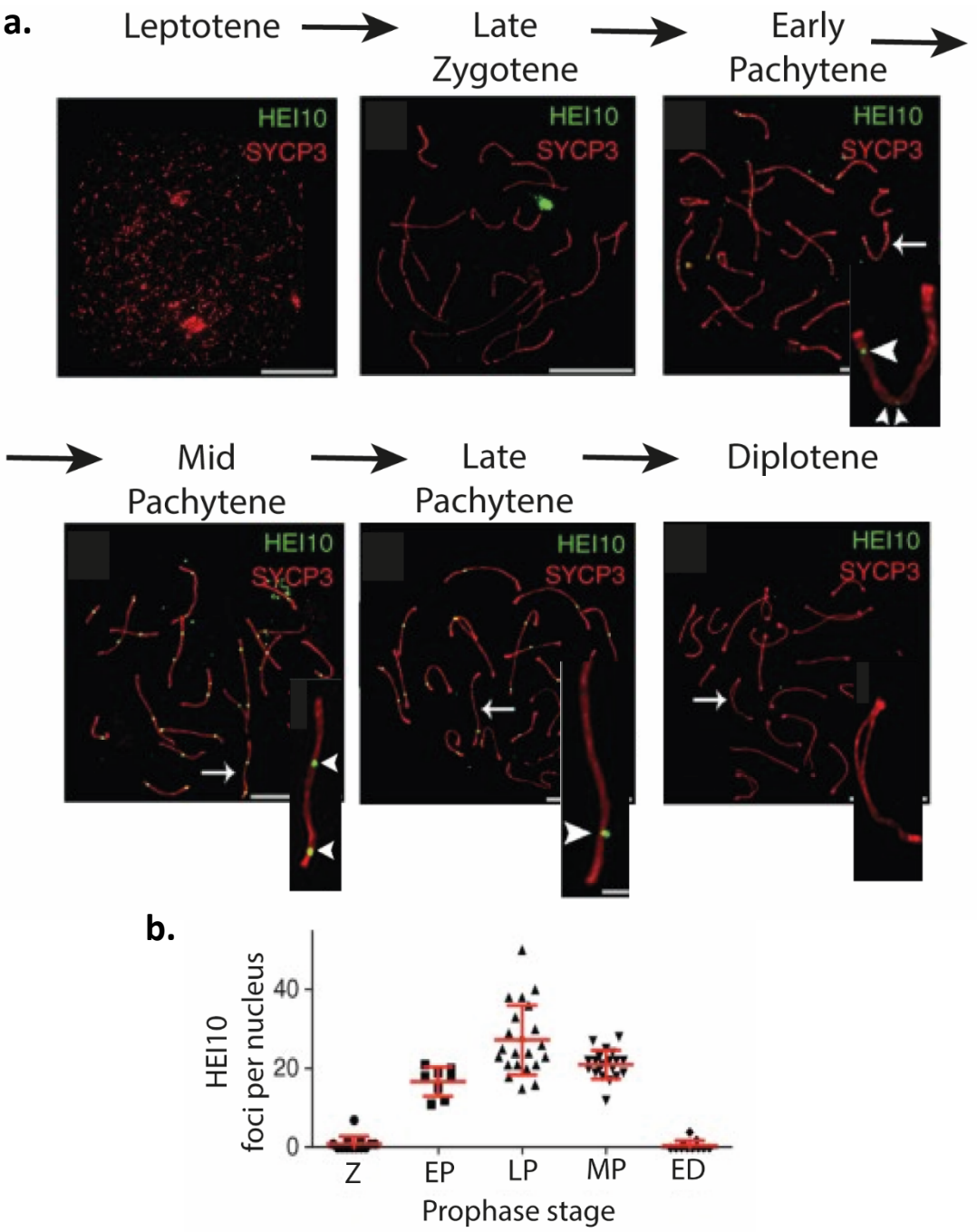

a.
Leptotene
Late Zygotene
Early Pachytene
Mid Pachytene
Late Pachytene
Diplotene
HEI10
SYCP3
b.
HEI10 foci per nucleus
40
20
0
Z
EP
LP
MP
ED
Prophase stage
Mid-pachytene
c.
HEI10
RNF212
SYCP3
d.
HEI10
MSH4
SYCE1
e.
HEI10
CDK2
SYCE1
f.
HEI10
MLH1
SYCE1

Figure 5.1.7.| Hei10 localisation to the synaptonemal complex and crossover sites.
a) SIM images of wild-type spermatocyte nuclei immunostained for Hei10 (green) and SYCP3 (green) throughout successive stages of prophase in meiosis I. Magnified views of chromosomes indicated by arrows highlight the positioning of Hei10 on chromosomes (arrowheads). b) Numbers of Hei10 foci per nucleus at prophase stages; zygotene (Z), early pachytene (EP), mid-pachytene (ED), late pachytene (LD) and early diplotene (ED). Horizontal bars represent mean ± s.d. c-f) Mid-pachytene spermatocyte immunostained for Hei10, RNF212, SYCP3, SYCE1, CDK2 and MLH1. Magnified images of chromosomes highlighted by arrows are shown below. Figure adapted from Qiao et al., 2014.

Y2H screens performed with HEI10 from *Arabidopsis* identified an interaction with HEIP1 (HEI10 interaction protein) and was validated by a bimolecular fluorescence complementation (BiFC) assay (Li et al., 2018). Furthermore, HEIP1 colocalises with HEI10 along meiotic chromosomes and at CO sites from late pachytene to diplotene, and interacts with Zip4 and MSH5 (MutSγ complex) (Li et al., 2018). These observations suggest that HEI10 may not work alone during ubiquitylation processes. It has been proposed that in *Arabidopsi*s HEIP1 together with HEI10 function to promote CO formation (Li et al., 2018). This could also be the case with mammalian HEI10, suggesting that HEI10 does not work independently in the interfering CO pathway.

5.1.7. RNF212 and Hei10 define an axis associated SUMO-ubiquitin-proteasome relay

The MutSγ complex binds and stabilises DNA strand invasion intermediates, but more importantly plays a specific role in CO differentiation through the recruitment of the recombinase MutLγ complex to a subset of recombination sites (Guillon et al., 2005; Santucci-Darmanin et al., 2002; Santucci-Darmanin et al., 2000). In mice, it has been proposed that this is achieved through a feedback loop involving the E3 ligase proteins HEI10 and RNF212 through ubiquitination and SUMOylation, respectively. HEI10 and RNF212 are thought to stabilise the ZMM proteins, including MutSγ, as well as promote the recruitment of the MutLγ complex (Holloway et al., 2014; Rao et al., 2017; Qiao et al., 2014). However, how this is mechanistically achieved is still not fully understood.

It has been proposed that cyclin-CDK activity, specifically CDK2, may also be required for the conversion of recombination intermediates to form CO products (Palmer et al., 2020). Numerous studies

have shown the localisation of CDK2 to meiotic telomeres as well as RNs, suggesting multiple roles of CDK2 in meiosis (Ashley et al., 2001; Viera et al., 2009). The RING domain sequence of HEI10 contains a consensus CDK phosphorylation 'RXL' site, suggesting a possible docking site for the binding of cyclin-CDK proteins. Furthermore, cyclin-family proteins, COSA-1 (crossover site associated 1) and CNTD1 (cyclin N-terminal domain-containing 1) have been identified as pro-CO factors and required to coordinate meiotic crossover formation in *C. elegans* and mammals, respectively (Gray et al., 2020). Upon deletion of RN-associated proteins, RNF212, HEI10 or CNTD1, the localisation of CDK2 to RNs is severely impaired. In contrast, CDK2 localisation to telomeres is unaffected in the three deletion mutants. Furthermore, deletion of HEI10 or $CDK2^{T160A}$ (T-loop mutagenesis of CDK2), results in undisrupted synapsis, but RNF212 foci remain abundant. This suggests that CDK2 may be required for the localisation of HEI10 to designated CO sites and promote the selection of RNF212-marked sites by ubiquitin-marked degradation to form late recombination nodules (Palmer et al., 2020).

It is hypothesised that the SUMOylation of early meiotic recombination proteins by RNF212 mediates their stability. In contrast, ubiquitination by HEI10 is thought to mediate their degradation, to limit the colocalisation of RNF212 to a subset of MutSγ-associated recombination sites. Colocalisation studies have shown that RNF212 colocalises to a subset of MutSγ foci. Furthermore, as prophase I progresses from early to mid-pachytene RNF212 foci decrease, suggesting that RNF212 marks late recombination nodules. This suggests that there is a SUMO-ubiquitin relay, mediated by HEI10, RNF212 and a cyclin-like protein which together regulate the number of COs, schematised in Figure 5.1.8.

Moreover, the mammalian RNF212 paralogue, RNF212b, has been shown to be strongly expressed in mice testis (Johnston et al., 2020; Kadri et al., 2016). This suggests that RN212b also has a role in meiotic recombination. However, this function is currently unknown but may provide additional complexity to the SUMO/ubiquitin switch.

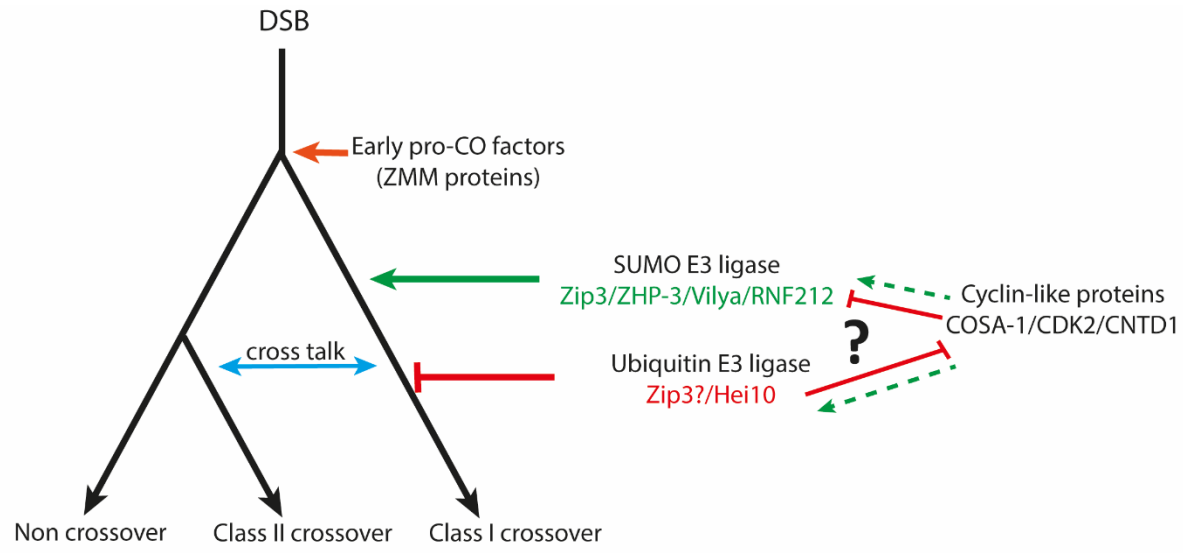


Figure 5.1.8.| Control of meiotic crossovers.
The meiotic recombination pathway is initiated upon DSBs and early pro-CO factors, including the ZMM proteins. The class I crossover pathway is regulated by SUMO and ubiquitin E3 ligase proteins and cyclin-like proteins. The ubiquitin and SUMO activity of E3 ligases act antagonistically to ensure the control the maturation of DSBs to crossovers. There is observed cross talk between class I and II crossovers.

5.1.8. *D. melanogaster* encodes three meiosis specific recombination proteins: Vilya, Narya and Nenya.

Vilya encodes a Zip3-like RING containing E3 ligase protein, shown to be essential for meiotic recombination in *Drosophila*. Despite this, no homologues for Vilya have been identified in other organisms, including mice and humans (Lake et al., 2015). Secondary structure alignment of Vilya to meiotic E3 ligases shows a strong relationship with other Zip3-like orthologues, possessing a putative SUMO-interacting motif (SIM) and three RXL cyclin binding domains (Lake et al., 2015). Evolutionary analysis revealed that Vilya is closely related to Zip3 and RNF212, suggesting involvement in SUMO pathways within meiotic recombination (Figure 5.1.9.a.) (Chelysheva et al., 2012; Lake et al., 2015; Reynolds et al., 2013).

In *Drosophila*, DSB formation occurs after the full-length SC is constructed and in the absence of SC formation DSB formation is significantly reduced (McKim et al., 1998). In comparison to the previously described E3 ligases, Vilya plays a role beyond CO formation. Vilya has been shown to be required in DSB formation by interacting with a DSB protein partner of Spo11, Mei-p22. Also, *Drosophila* appear to lack homologues of the mismatch repair complex MutSγ and the endonuclease MutLγ complex and instead use an endonuclease complex consisting of MEI-9, Ercc1, Mus312 and Hdm (Lake et al., 2015). This further suggests that *Drosophila* have a unique way of coupling DSB formation to CO events. Nevertheless, immuno-EM localised Vilya to recombination nodules and the SC central region. It has been proposed that recombination sites that recruit sufficient Vilya are able to be resolved into COs, suggesting a similar role to Zip3 and ZHP-3 in yeast and nematodes, respectively (Bhalla et al., 2008; Börner et al., 2004).

In addition, it has been recently discovered that *Drosophila* encodes two further E3 ligase proteins, Narya and Nenya. Sequence alignment of Vilya, Narya and Nenya determined sequence conservation of a C3HC4 RING finger domain and a predicted coiled-coil domain (Figure 5.1.9.b.) (Lake et al., 2019). These observations indicate structural similarities to Vilya and other E3 ligases. All three *Drosophila* proteins cluster within the Zip3-RNF212 subgroup of E3 ligases, suggesting a shared function.

Localisation studies have shown that Narya colocalises with Vilya to DSBs throughout pachytene, suggesting that Narya and possibly Nenya are components of the recombination nodule. Separation of function assays show Narya or Nenya are required for DSB formation and CO maturation, indicating that *narya* and *nenya* encode functionally redundant proteins and only the absence of both proteins results in an increase in meiotic nondisjunction (Lake et al., 2019). These observations seen in *Drosophila* may also be the case for the mammalian paralogous gene products RNF212 and RNF212b.

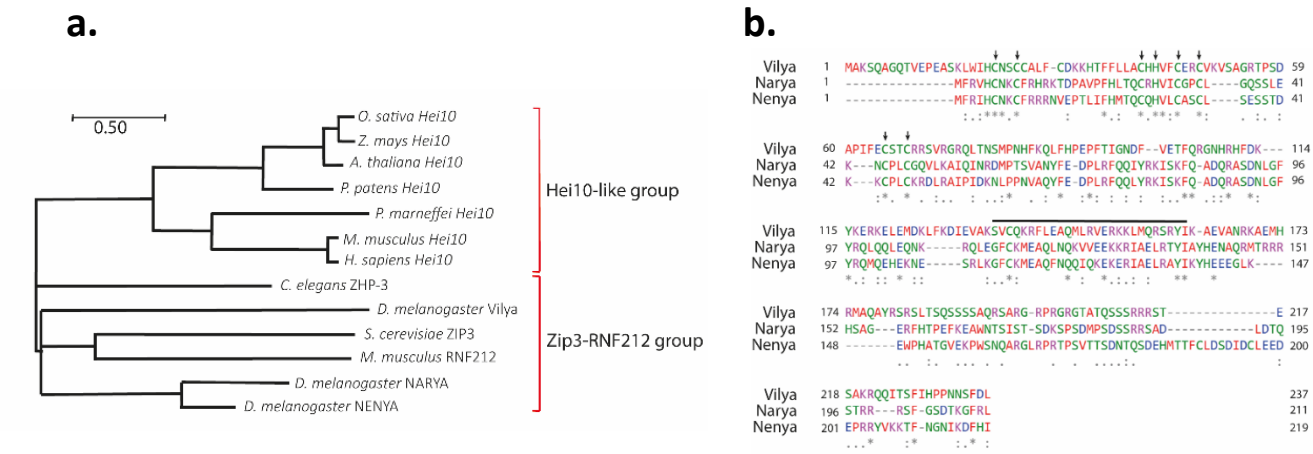


Figure 5.1.9.| Protein alignment of *D. melanogaster* meiotic recombination proteins.
a) Maximum-likelihood tree of the sequences from members of the Zip3-RNF212- and Hei10-like subgroups of E3 ligase proteins. Results show that all three Drosophila E3 ligase proteins cluster with the Zip3-RNF212 subgroup. Scale bar indicates the number of nucleotide changes per site. b) Sequence alignment of D. melanogaster meiotic recombination proteins: Vilya, Narya and Nenya. Proteins were aligned using MUSCLE and ClustalX. Arrows highlight the conserved residues in the C_3HC_4 RING finger domain. The residues predicted to form a coiled-coil domain are below the black line. Figure reproduced from Lake et al., 2019**.**

5.1.9. Chapter aims

This project was initiated by an Erasmus student, Miss Carmen Espejo Serrano, under the supervision of Dr James Dunce, to biophysically characterise the mammalian meiotic recombination proteins HEI10, RNF212 and RNF212b *in vitro*, complementary to the *in vivo* analysis carried out by The Hunter Lab (Professor Neil Hunter, University of California, Davis). Carmen made headway in the purification and analysis of the RNF212 and RNF212b RING domains in isolation and initial analysis of HEI10. Dr James Dunce then pursued the crystallography of HEI10. I set out to carry out a full biophysical characterisation of the mammalian meiotic recombination proteins in isolation to gain a wider understanding of the molecular structures, and thus similarities and differences between the SUMO- and ubiquitin- E3 ligases. Furthermore, I aimed to deduce an interaction network between the meiotic recombination proteins which could provide further insight into the structure-function role of the meiotic E3 ligase proteins in crossover formation during meiosis prophase I in humans.

In addition, I carried out the biophysical characterisation of the yeast and *Drosophila* E3 ligase proteins, to determine if the structure of the RNF212-like homologues are conserved across species. Any data presented in this chapter which has been collected by Miss Carmen Espejo Serrano or Dr James Dunce is clearly annotated. However, I have analysed all the data presented.

5.2. RESULTS

5.2.1. Bioinformatic and secondary structural analysis of the meiotic recombination proteins

The human meiotic recombination protein, HEI10 (human enhancer of invasion clone 10), encodes a 277 amino acid open reading frame (Figure 5.2.1.) (Toby et al., 2003). Primary and secondary structure prediction determined that HEI10 has a conserved α-helical central region, amino acids 108-196, with some coiled-coil prediction. Upstream of this region is a N-terminal 'really interesting new gene' (RING) domain, characteristic of E3 ligases (Figures 5.2.1. and 5.2.2.a.).

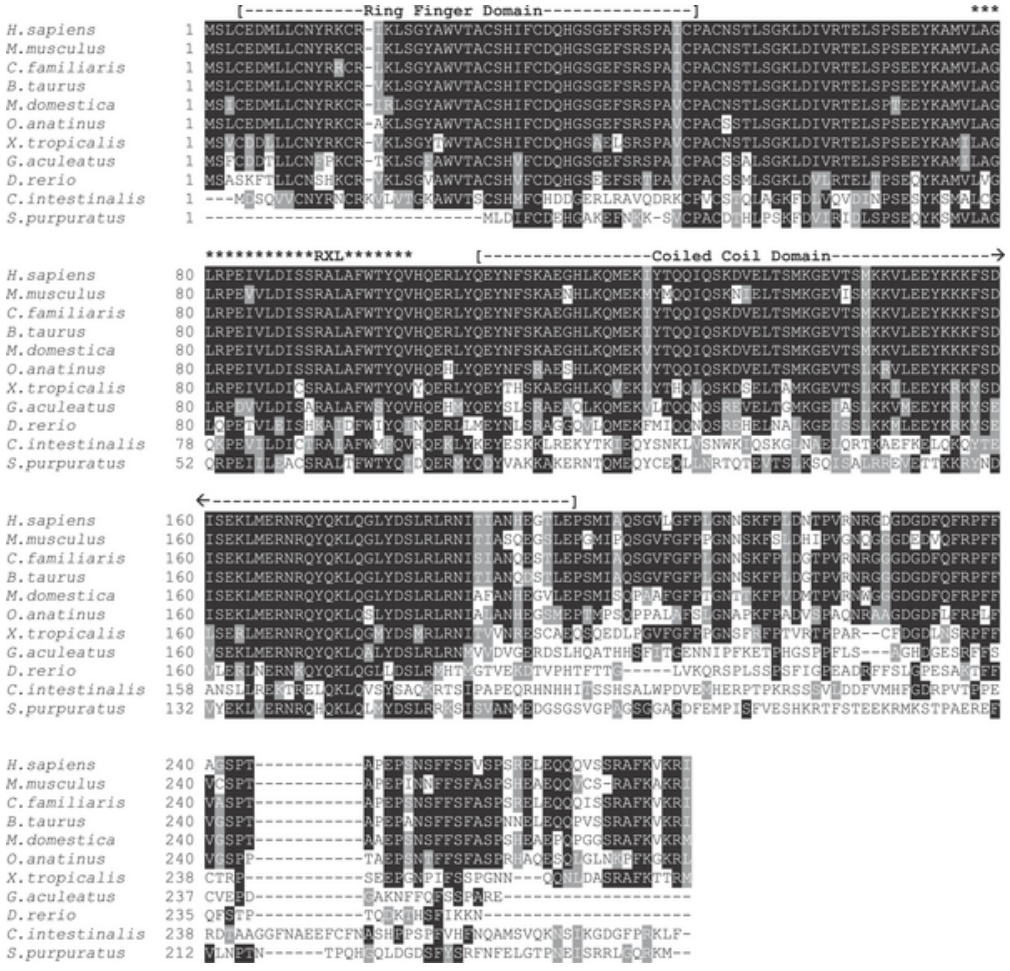


Figure 5.2.1.| HEI10 amino acid sequence alignment
ClustalW protein sequence alignment of Hei10 orthologues from Deuterostome lineages revealsing a N-terminal C3HCHC2 sequence of highly conserved histidine and cysteine residues, similar to previously characterised zinc finger C3HC4 RING finger domains. Hei10 possesses a potential RXL cyclin interaction motif and a conserved α-helical domain with coiled-coil prediction. Highlighted by asterisks is the 24 amino acid (residues 76-100) deletion encoded by the *Hei10^{mei4}* allele. ClustalW alignment taken from Ward et al., 2007.

Sequence analysis and secondary structure prediction of the HEI10 RING domain, amino acid residues 4-50, shows a structure prediction closely related to a typical C3HC4 RING domain (Barlow et al., 1994; Deshaies & Joazeiro, 2009). Sequence alignment of the HEI10 RING domain with previously characterised E3 ligase RING fingers revealed some divergence from the established RING consensus. Instead, HEI10 has a C3HCHC2 sequence and possesses slightly different spacing between the conserved cysteine residues (Figure 5.2.2.b.). These observations suggest that the function and structure of HEI10 may differ slightly to the previously characterised RING E3 ligases. The C-terminus of HEI10 (amino acid residues 198-end) is predicted to be unstructured and has reduced sequence identity when compared to the N-terminus.

In addition to HEI10, mammals encode two further E3 ligase proteins, RNF212 and its paralogue RNF212b, with lengths of 297 and 300 amino acids, respectively (Kong et al., 2008; Reynolds et al., 2013). Both proteins share the same domain structure as HEI10, consisting of a N-terminal RING domain (residues 1-61 and 1-56, respectively) followed by a α-helical domain and unstructured C-termini (Figure 5.2.3.a. and b.). Closely related protein sequences have been identified in numerous vertebrate species, although not in lower eukaryotes. Despite little or no sequence conservation, HEI10-like and/or RNF212-like homologues have been identified in plants, fungi, *C. elegans* and *Drosophila* (Bhalla et al., 2008; Cheng et al., 2006; Lake et al., 2019, 2015). *D. melanogaster* Vilya is a member of the Zip3-like family of E3 ligases and is involved in the SUMOylation pathway. Vilya is 297 amino acids in length and is highly conserved across the *Drosophila* genus. Secondary structure prediction determined a similar domain structure to the mammalian E3 ligases, suggesting that Vilya may have preserved structure and function (C. M. Lake et al., 2015) (Figure 5.2.3.c.).

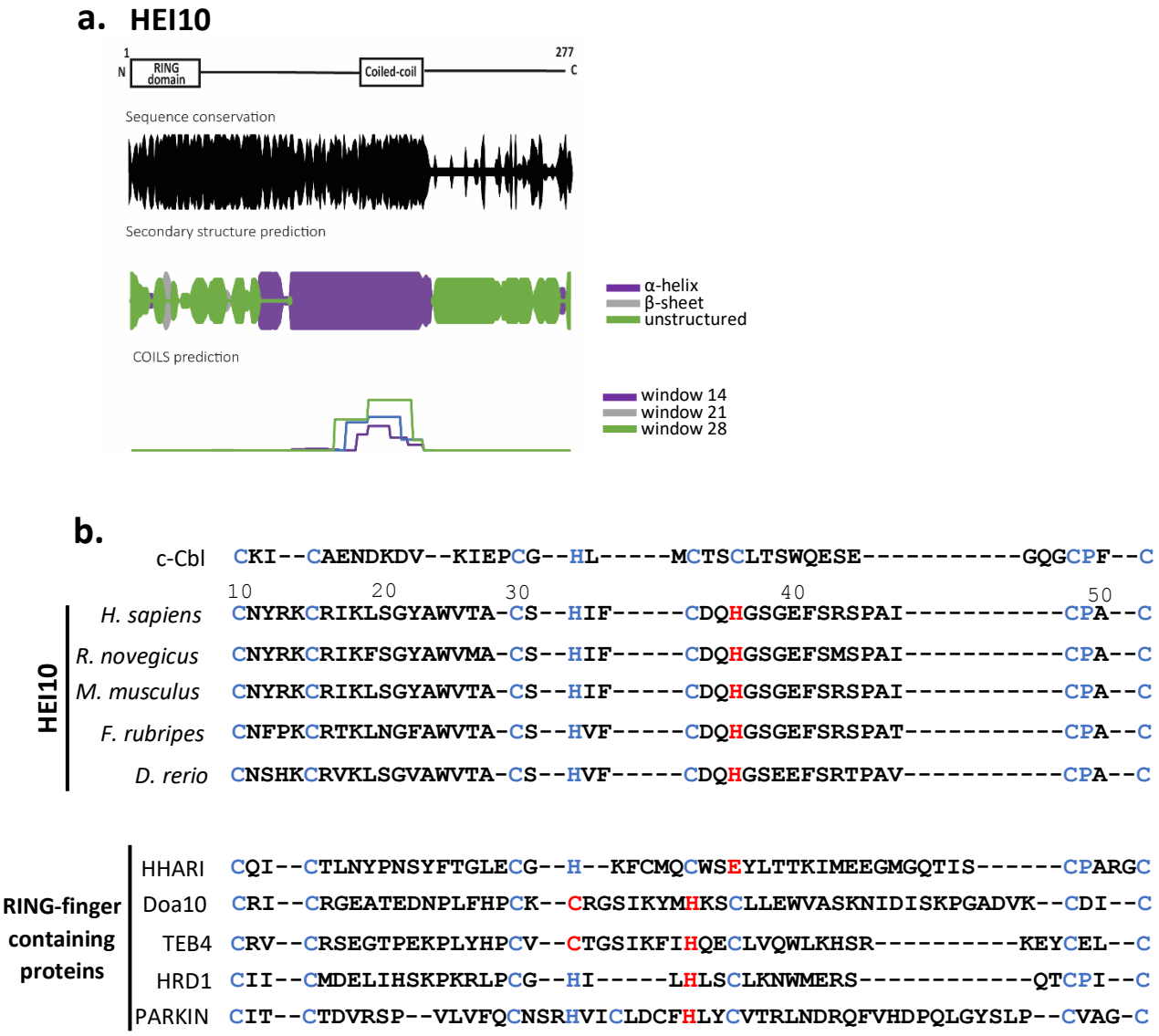


Figure 5.2.2.| Bioinformatic analysis of HEI10.
a) Aligned sequence schematics showing the domain structure, sequence conservation, secondary structure and COILS prediction of the mammalian E3 ligases protein, HEI10. HEI10 has a conserved domain structure consisting of a N-terminal RING domain, an α-helical domain with coiled-coil prediction and an unstructured C-terminus. Secondary structure prediction was calculated by JPred4. α-helix in purple, β-sheet in green and unstructured in grey (Drozdetskiy et al., 2015). Coiled-coil conformation prediction was calculated by COILS which compared the sequence of interest to a database of known parallel two-stranded coiled-coils and derived a similarity score in a 14,21 or 28 amino acid window. (Lupas et al., 1991). b) Sequence alignment of HEI10 N-terminal RING domain of five different vertebrate species versus other RING fingers, including the RING finger protein, c-Cbl, which contains the consensus C3HC4 sequence. Core cysteines indicated in blue and deviations from the consensus sequence indicated in red. Other RING fingers include the proteins HHAR1, Doa10, TEB4, HRD1, and PARKIN. Alignment of HEI10 sequences shows a conserved C3HCHC2 sequence. Figure adapted from Toby et al., 2003.

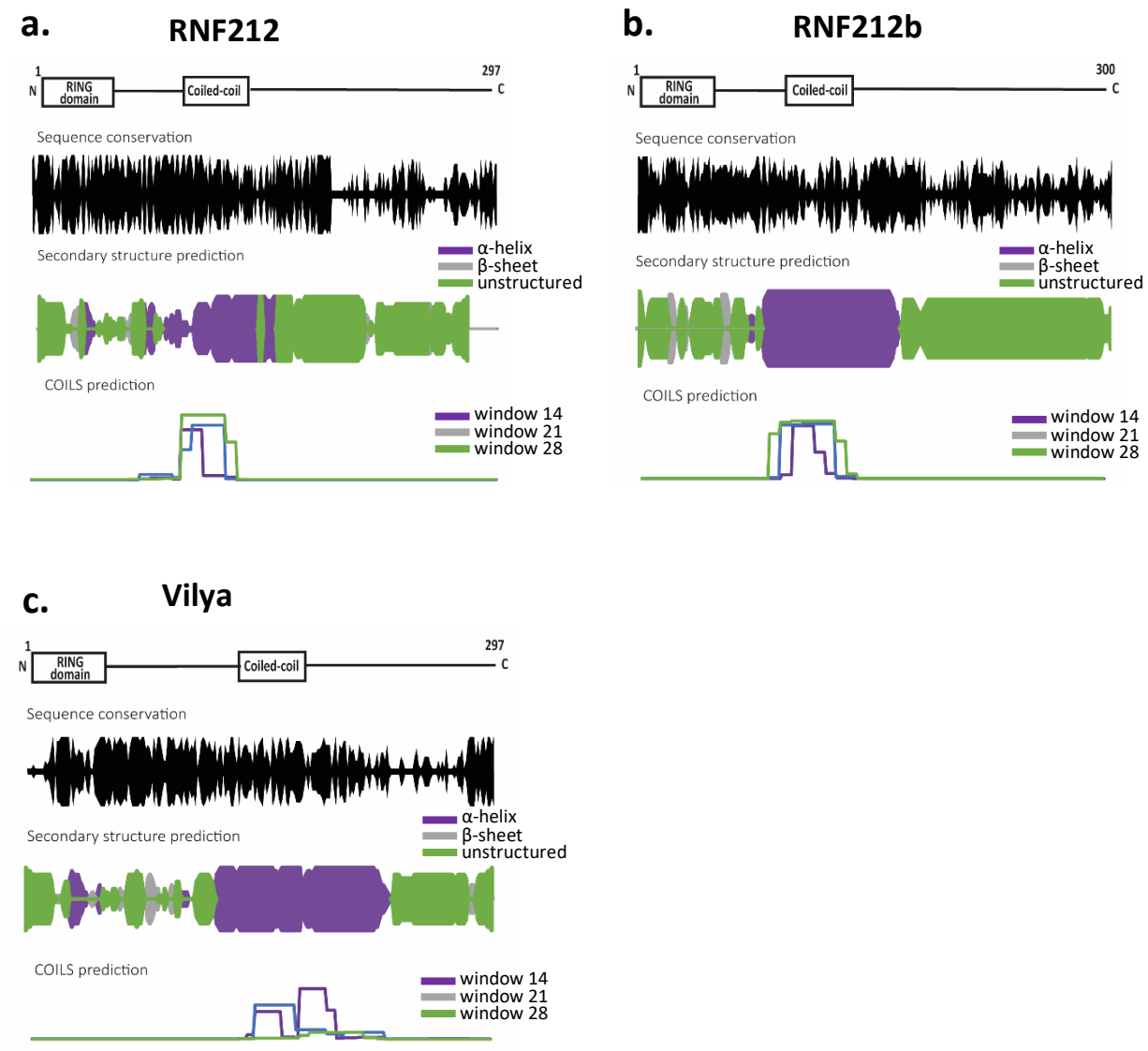


Figure 5.2.3.| Bioinformatic analysis of meiotic E3 ligase proteins.
Aligned sequence schematics showing the domain structures, sequence conservation, secondary structure and COILS prediction of the mammalian E3 ligases proteins a) RNF212 b) RNF212b and *D. melanogaster* c) Vilya. Secondary structure prediction was calculated by JPred4. α-helix in purple, β-sheet in green and unstructured in grey (Drozdetskiy et al., 2015). Coiled-coil conformation prediction was calculated by COILS which compared the sequence of interest to a database of known parallel two-stranded coiled-coils and derived a similarity score in a 14,21 or 28 amino acid window. (Lupas et al., 1991)

5.2.2. Purification and characterisation of full-length HEI10

We first set out to biophysically characterise the full-length protein, amino acid residues 1-277, which includes the conserved N-terminal RING domain and central predicted coiled-coil domain, both characteristic of E3 ligases. $HEI10_{1-277}$ was expressed with in a N-terminal His_6-MBP-tag in *E.coli* BL21

(DE3) cells and grown in terrific broth (TB) growth media. Upon induction, ZnOAc was added to a final concentration of 100 μM, to ensure sufficient zinc was present during protein expression. The growth temperature of the bacterial culture was reduced to 15°C to promote protein folding. Cells were lysed in standard lysis buffer with the addition of 10% glycerol as it was observed that HEI10 aggregated in solution. The addition of glycerol increased the stability of the protein. All purification steps were performed with buffer containing 10% glycerol to minimise protein precipitation. High level expression was obtained for $HEI10_{1\text{-}277}$, however most of the expressed protein was found to be insoluble, even when the expression temperature was lowered to 15 °C, suggesting difficulties in folding (Figure 5.2.4.a). The soluble material was purified by amylose affinity chromatography and the amylose eluate contained a large amount of degraded material. The amylose eluate was further purified by anion exchange chromatography to remove the free-MBP and degradation products (Figure 5.2.4.b.).

SEC-MALS analysis of MBP-$HEI10_{1\text{-}277}$ determined a single peak. However, the molecular weight decreases across the peak suggesting that the full-length protein is unstable in solution. The molecular weight at the middle of the peak is 312.3 kDa (theoretical tetramer – 305.6 kDa), indicating that HEI10 forms a tetramer in solution (Figure 3.2.4.c.). Despite this determination, we were unable to cleave and further analyse full length protein, further indicating protein instability. This could be due to the presence of the C-terminus (amino acid residues 197-277) which has poor sequence conservation and has very little predicted secondary structure.

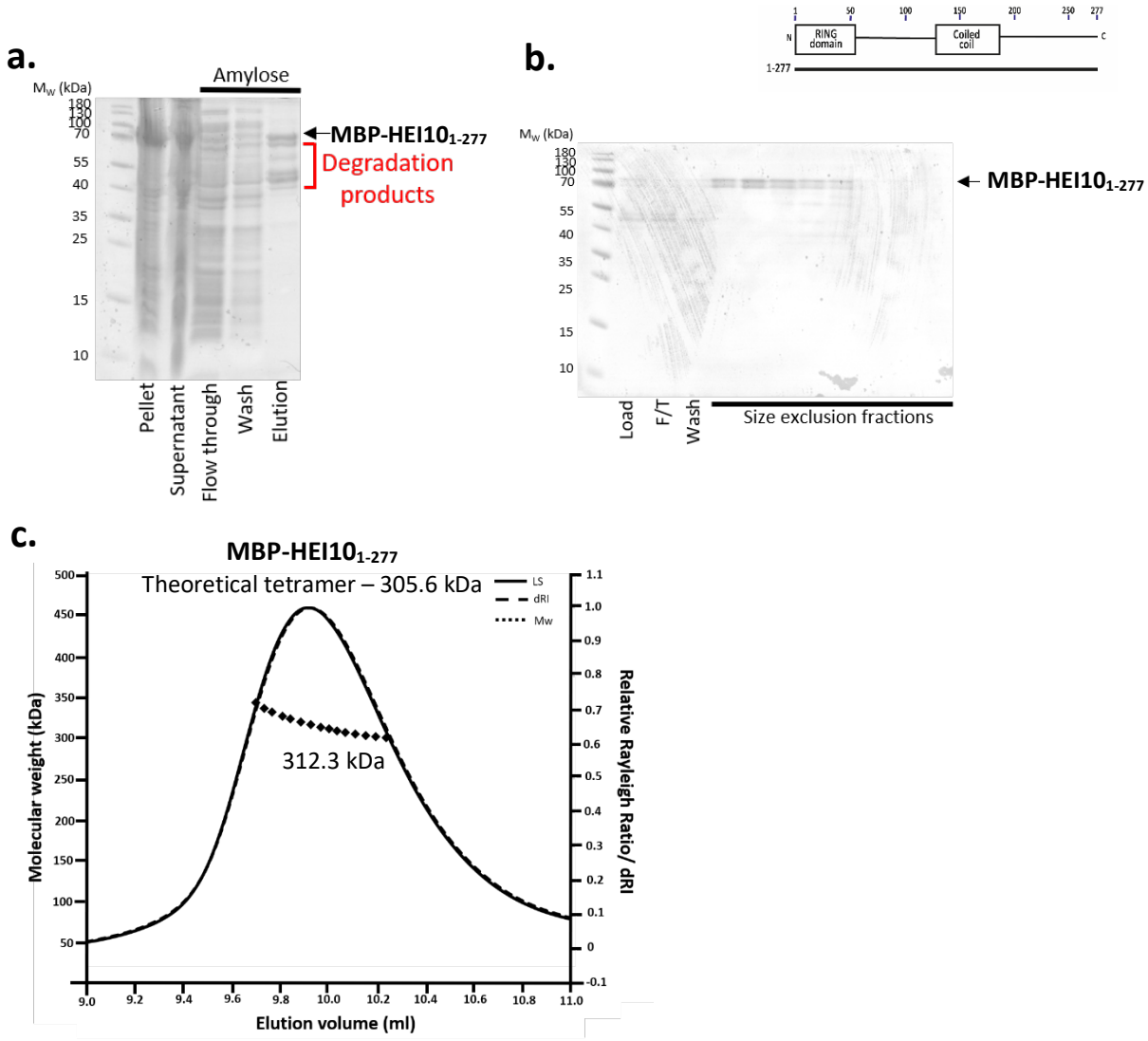


Figure 5.2.4.| Purification and structural analysis of full-length HEI10

a-b) SDS-PAGE analysis showing the purification steps of HEI10$_{1\text{-}277}$ expressed as an N-terminal MBP fusion. a) MBP-Hei10$_{1\text{-}277}$ expressed well but most of the protein was insoluble as shown by an intense band in the bacterial pellet at ~75 kDa. The limited amount of soluble material was purified by amylose affinity chromatography. MBP-HEI10$_{1\text{-}277}$ highly degrades down to free-MBP, shown by multiple bands in the amylose eluate, suggesting that the full-length protein is highly unstable. b) MBP-HEI10$_{1\text{-}277}$ was further purified by anion exchange chromatography to remove degradation products. c) SEC-MALS analysis of MBP-HEI10$_{1\text{-}277}$ revealed a single symmetrical peak with an experimental molecular weight of 312.3 kDa. A theoretical tetramer has an estimated molecular weight of 305.6 kDa indicating that HEI10 forms a tetramer in solution.

5.2.3. The structural core of HEI10 requires both the RING and coiled-coil domains.

We set out to analyse the N-terminal RING domain and the central α-helical domain in isolation. Using the secondary structure prediction and sequence conservation we designed two constructs, HEI10$_{1-70}$ and HEI10$_{69-196}$, corresponding to the RING and coiled-coil domain, respectively. Both constructs were expressed as MBP-fusions in *E.coli* BL21 (DE3) cells, referred to MBP-HEI10 constructs. High level of expression was obtained for both constructs. However, like the full-length construct around 50% of the expressed protein was in the bacterial pellet. The soluble material was purified by sequential amylose affinity and anion exchange affinity chromatography (Figure 5.2.5.a. and b.). SEC-MALS analysis of MBP-HEI10$_{69-196}$ determined large megadalton species, which suggests the protein construct is aggregated (Figure 5.2.5.d.). Analysis of MBP-HEI10$_{1-70}$ determined a molecular mass of 317 kDa, indicating oligomerisation (theoretical monomer – 54 kDa). From this result, we hypothesise that HEI10 requires both the RING domain and central α-helical domain for solubility.

HEI10 structural core (amino acid residues 1-196) was cloned and expressed in *E. coli* and grown following the same protocol as the full-length protein. MBP-HEI10$_{1-196}$ expressed well but ~50% of the material is insoluble (Figure 5.2.5.c.). This seems to be a consistent observation for HEI10 constructs analysed thus far. The soluble material of MBP-HEI10$_{1-196}$ was purified through sequential amylose affinity and anion exchange chromatography.

SEC-MALS was used to determine absolute molecular weight of the complex and oligomeric state. MBP-HEI10$_{1-196}$ eluted as a single peak with a molecular mass of 258.7 kDa, corresponding to a tetramer (theoretical tetramer – 275.2 kDa) (Figure 5.2.5.d.). These observations suggest that HEI10 requires both both the RING domain and central α-helical domain for stability.

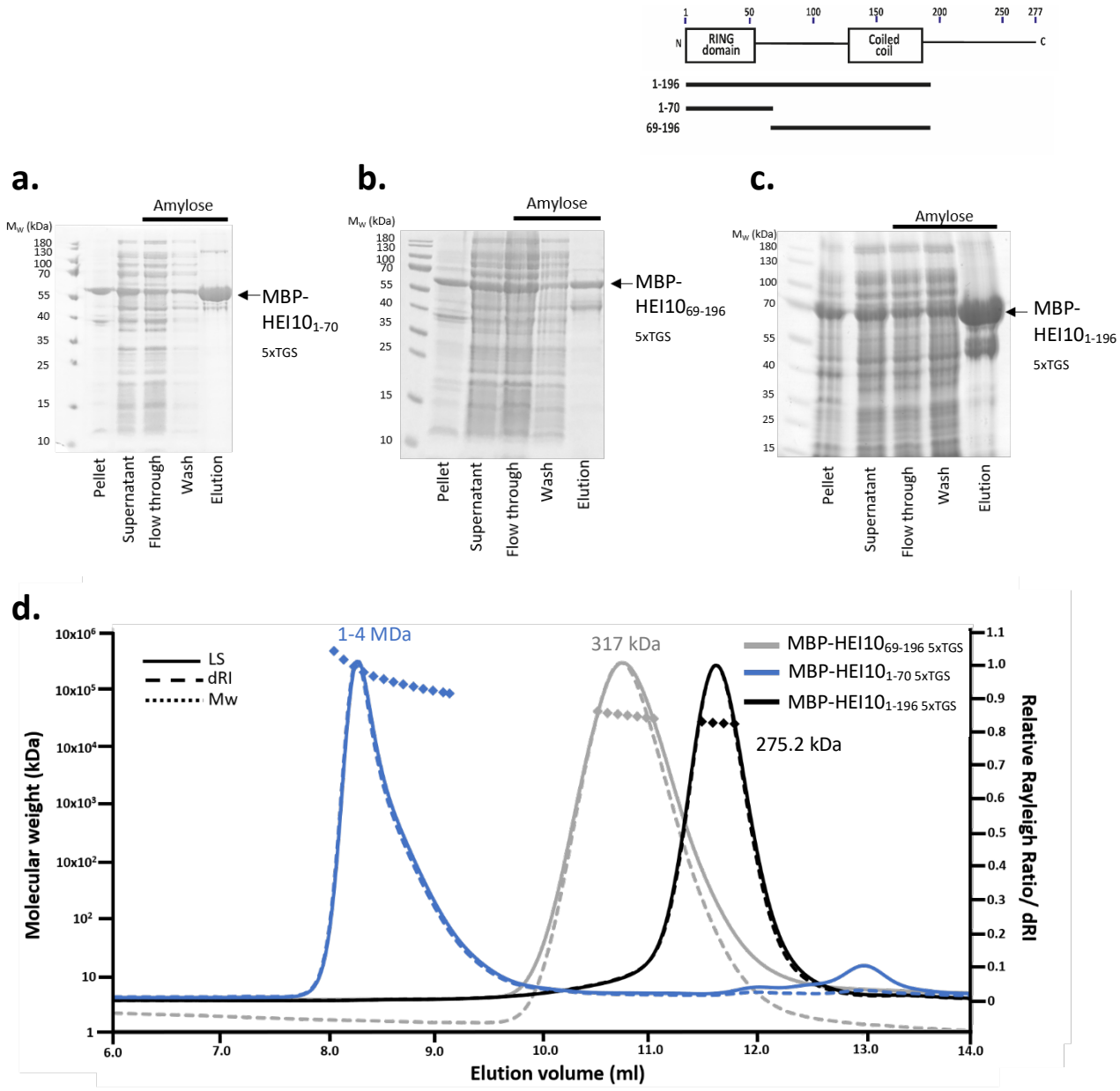


Figure 5.2.5.| HEI10 structural core requires both the RING and coiled-coil domain.
a-c) SDS-PAGE analysis of MBP-tagged Hei10 constructs purified by amylose affinity chromatography. All three constructs a) $HEI10_{1-70}$ (RING domain), b) $HEI10_{69-196}$ (α-helical, predicted coiled-coil domain) and c) $HEI10_{1-196}$ have a predominant band (~50% of material) in the bacterial pellet suggesting a large amount of insoluble material. Both $HEI10_{69-196}$ and $HEI10_{1-196}$ have a significant amount of free-MBP present, suggesting degradation. d) SEC-MALS analysis of MBP-$HEI10_{69-196}$ (blue) revealed large megadalton species suggesting it is a soluble aggregate. MBP-$HEI10_{1-70}$ (grey) has a molecular weight of 317 kDa, corresponding to a 6-mer (theoretical monomer – 54 kDa). MBP-$HEI10_{1-196}$ (black) elutes as a single peak with a molecular weight of 258.7 kDa corresponding to a tetramer (theoretical tetramer – 275.2 kDa).

5.2.4. $HEI10_{1-196}$ requires a long-linker to facilitate cleavage by TEV-protease

In order to analyse cleaved material, $HEI10_{1-196}$ was incubated with TEV-protease to cleave the N-terminal MBP-solubility tag. $HEI10_{1-196}$ remained fused to the MBP-affinity tag, independent of the TEV-protease enzyme concentration. This suggested that the TEV-cleavage recognition site is not accessible by TEV-protease. To address this problem, a flexible threonine-glycine-serine (TGS) repeat sequence was added to the beginning of the protein sequence to increase the accessibility of TEV-protease to the TEV-recognition site. Initially a 3xTGS linker was added but a large proportion of fusion protein remained uncleaved (Figure 5.2.6.a.). A longer 5xTGS linker was used, $HEI10_{1-196LL}$, which significantly improved the cleavage of MBP-$HEI10_{1-196}$, with only minor fusion protein remaining (Figure 5.2.6.b.).

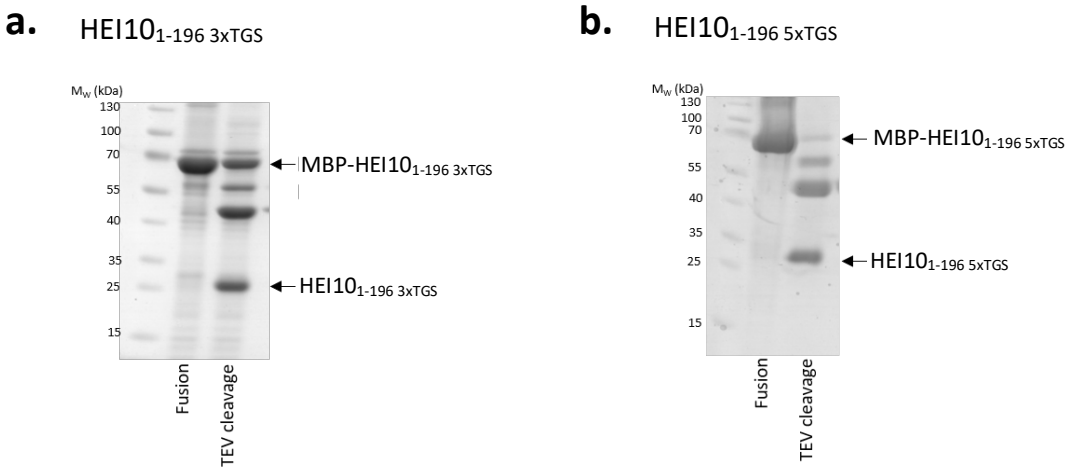


Figure 5.2.6.| $HEI10_{1-196}$ requires a TGS long linker to allow for successful cleavage.
a-b) SDS page analysis of a) MBP-$HEI10_{1-196\ 3xTGS}$ and b) MBP-$HEI10_{1-196\ 5xTGS}$. a) A predominant band at ~70 kDa suggests that the 3xTGS linker does not improve cleavage. b) MBP-$HEI10_{1-196\ 5xTGS}$ cleavage was more successful, shown by a faint band at ~70 kDa indicating most of the protein has been cleaved.

5.2.5. $HEI10_{1-196LL}$ forms a stable tetramer in solution

The N-terminal MBP-tag of $HEI10_{1-196LL}$was cleaved by TEV-protease and further purified through sequential cation exchange and size exclusion chromatography. Purification steps of $HEI10_{1-196LL}$ are summarised in Figure 5.2.7.a. SEC-MALS analysis was used to determine the absolute molecular mass

of HEI10$_{1\text{-}196LL}$ and subsequently determine the oligomeric state. HEI10$_{1\text{-}196LL}$ eluted as a single peak, with a molecular mass of 95.7 kDa, corresponding to a tetramer (theoretical molecular weight – 96.7 kDa) (Figure 5.2.7.b and c).

Far-UV CD spectroscopy was used to quantify the secondary structure composition of HEI10$_{1\text{-}196LL}$. Analysis revealed a characteristic α-helical spectrum with negative peaks of similar magnitude at 208 and 222 nm and a positive peak at 193 nm (Figure 5.2.7.c.). Deconvolution by DichroWeb estimates an α-helical content of 41%, corresponding to 80 helical residues out of 196, which corresponds to the predicted coiled-coil domain. CD spectroscopy was also utilised to determine the thermal stability of HEI10$_{1\text{-}196LL}$ by tracking the helical signal at 222 nm between 4 and 95°C. Co-operative unfolding revealed a melting temperature (T_m) of 52 °C (Figure 5.2.7.d.). Together, these observations suggest that HEI10$_{1\text{-}196LL}$ is folded.

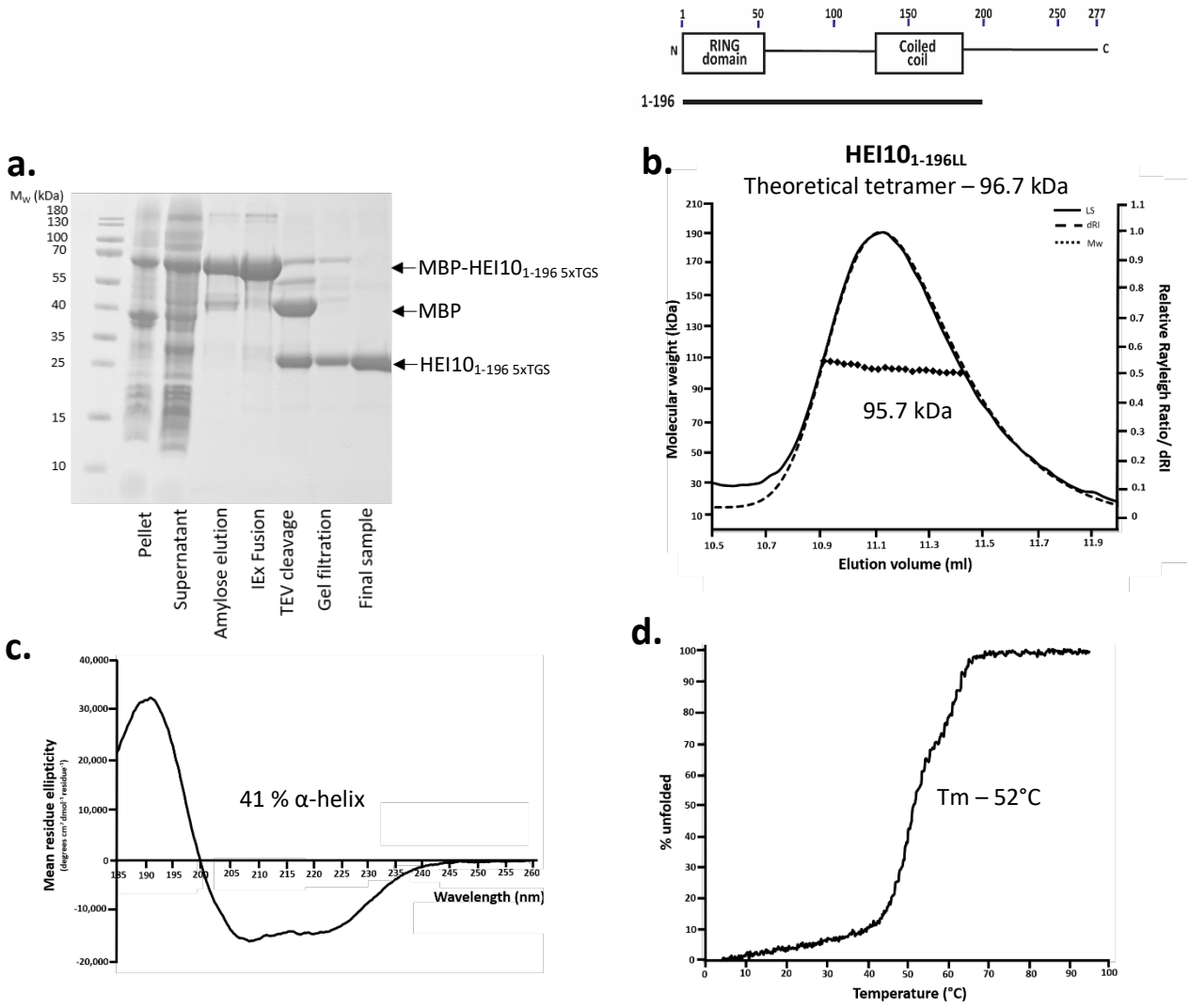


Figure 5.2.7.| Purification and biophysical analysis of HEI10$_{1\text{-}196\ 5xTGS}$
a) SDS-PAGE showing the purification summary of HEI10$_{1\text{-}196LL}$ through sequential amylose affinity and anion exchange chromatography. N-terminal MBP-tag was removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. b) SEC-MALS analysis of HEI10$_{1\text{-}196\ 5xTGS}$ revealed a single peak with an experimental molecular weight of 95.7 kDa (theoretical tetramer – 96.7 kDa) indicating that HEI10$_{1\text{-}196}$ structural core tetramerises in solution. c) Far UV CD wavelength scan between 260-185nm of HEI10$_{1\text{-}196LL}$ shows a typical α-helical trace. Deconvolution of the data estimates the secondary structure to be 41 % α-helical. d) CD thermal denaturation measured at 222nm between 4 and 95°C, indicated a melting temperature of 52 °C.

SEC-SAXS analysis of HEI10$_{1\text{-}196LL}$ was utilised to gain low resolution structural data. HEI10$_{1\text{-}196LL}$ has a Guinier analysis determined *Rg* of 75.2 Å, which matches the real-space *Rg* of 74.5 Å, for an elongated structure (Figure 5.2.8.a.). Determination of the cross-sectional radius of gyration (*Rc*) reveals the thickness of elongated molecules. The *Rc* of HEI10$_{1\text{-}196LL}$ was observed to be 11.2 Å (Figure 5.2.8.b.).

As scattering is dominated by larger species, the *Rc* value could be dominated by the RING domain, therefore not be a true value for the whole construct.

Fourier transform of the scattering curve allows for calculation of the pairwise distribution profile (*P(r)* curve). The *P(r)* distribution shows a positive skew with an elongated tail, characteristic of proteins possessing coiled-coil domains. The maximum dimension (*Dmax*) of $HEI10_{1\text{-}196LL}$ was determined to be 290 Å (Figure 5.2.8.c.). The experimental length can be compared with the theoretical length of the sequence as an individual α-helical chain (~ 1.5 Å per residue). The theoretical length of the HEI10 α-helical domain (residues 69-196) is 189 Å, therefore corresponds to the majority of the protein length.

Ab initio modelling of $HEI10_{1\text{-}196LL}$ was performed to produce a low-resolution molecular envelope from the SEC-SAXS scattering data. The resultant DAMMIF envelope suggests that the HEI10 core tetramer consists of globular N-terminal RING domains connected by the elongated helical coiled-coil domain, with tetramerisation occurring through the α-helical domain, forming a dumbbell shape (Figure 5.2.8.d).

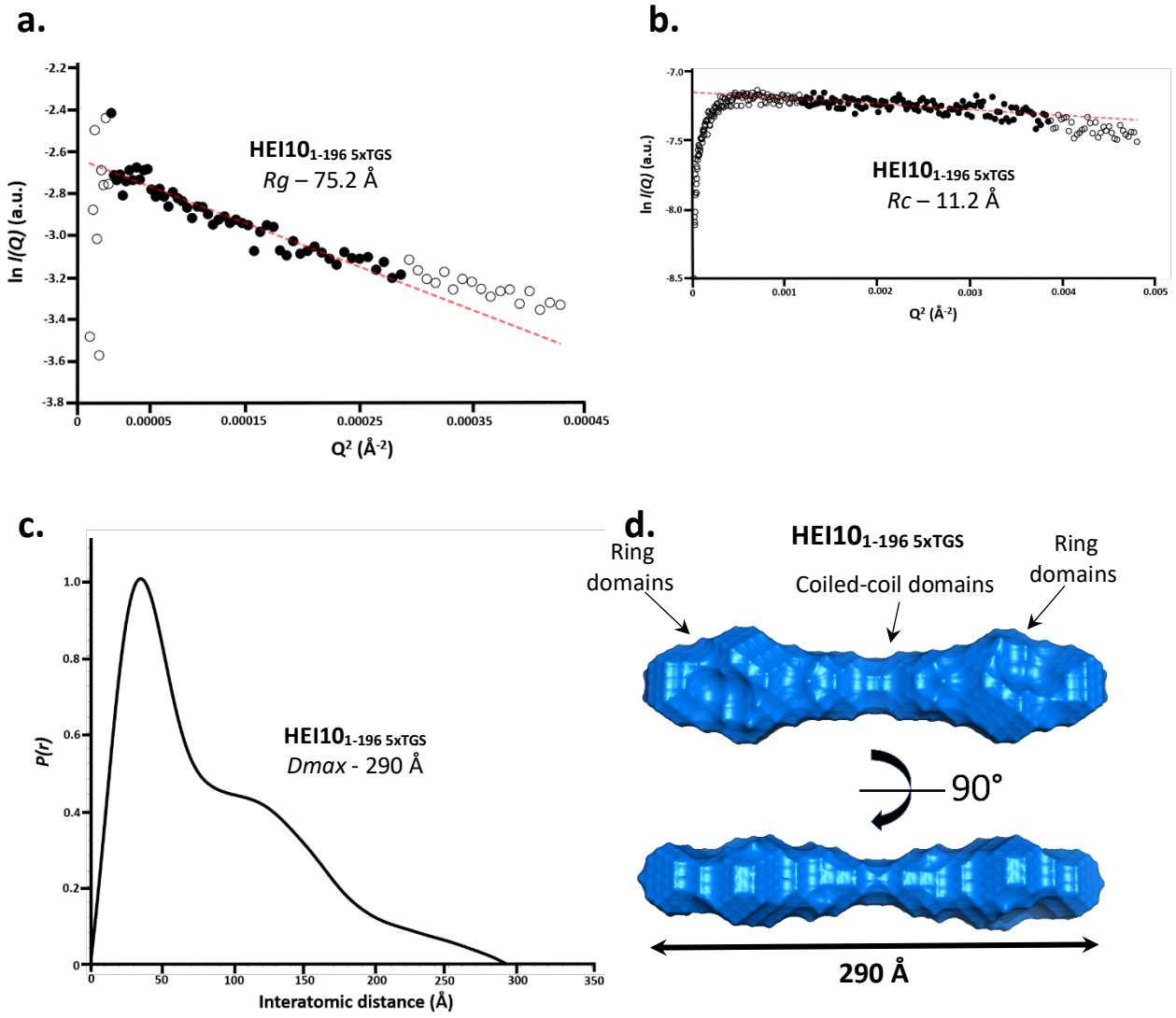
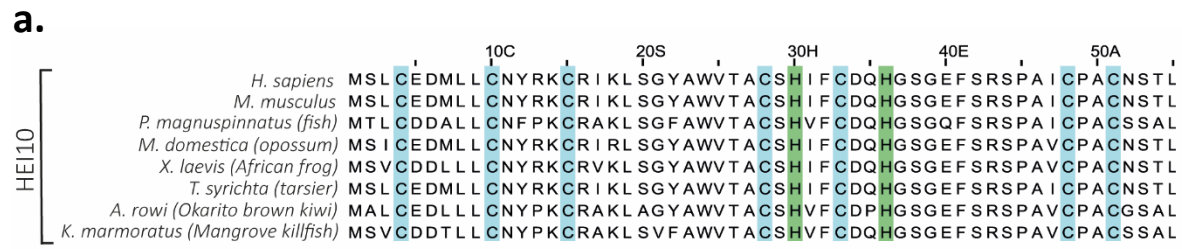


Figure 5.2.8.| SEC-SAXS analysis of HEI10$_{1\text{-}196\ 5xTGS}$.
a) Guinier analysis determined a radius of gyration (*Rg*) value of 75.2 Å. The real space *Rg* closely matches the Guinier analysis *Rg* value of 74.5 and 75.2 Å, respectively. Clear circles represent the complete data set and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3). b) Guinier analysis determined the radius of gyration of the cross-section (*Rc*) of HEI10$_{1\text{-}196\ 5xTGS}$ to be 11.2 Å. Clear circles represent the complete data and solid circles represent the region used for the fit. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3). c) *P(r)* distribution of HEI10$_{1\text{-}196\ 5xTGS}$ showing a maximum dimension (*Dmax*) of 290 Å. d) SAXS *ab initio* DAMMIF model of HEI10$_{1\text{-}196\ 5xTGS}$ presented as a dumbbell-shaped molecular envelope with length of 290 Å.

5.2.6. $HEI10_{1-196}$ tetramer coordinates four zinc atoms

Sequence alignment of HEI10 N-terminal RING domain reveals the presence of seven cysteine and two histidine residues within the first 60 amino acids (Figure 5.2.9.a.). Although this is different to the canonical RING domain C3HC4 sequence, it still suggests zinc coordination of HEI10. This prompted us to assay for metal binding of the HEI10 core structure. We tested for the presence of zinc by PAR assay. This assay quantifies metal ions in metalloproteins by exploiting the differences in the absorption spectra of the chromophore chelator, 4-(2-pyridylazo) resorcinol (PAR), resulting from its complexation to Zn^{2+} and/or Cu^{2+} ions. Proteolytic digestion of HEI10 released the zinc into solution, which was chelated by PAR resulting in an observable spectrophotometric shift from 414 nm (PAR alone) to 494 nm (PAR complexed with either divalent metal ions). The absorbance spectra of zinc standards (0-100 μM) are used to estimate the number of zinc ions present (Figure 5.2.9.b.).

Absorbance spectra of digested $HEI10_{1-196LL}$ at two concentrations, 20 and 40 μM, have a trace closely matching that of the 20 and 40 μM zinc standards, respectively. The absorbance of the zinc standards at 413.5 nm ($A_{413.5}$) and 491 nm (A_{491}) were plotted against the calculated zinc concentration (black - x). Trendlines were calculated by linear regression of the experimental data, with a R^2 value close to 1. Zinc standards with an absorptivity outside of the linear range are shown in grey and have been excluded from calculating the linear equation (Figure 5.2.9.c. and d.). The determined linear equation allows for the calculated concentration of zinc to be accurately determined for HEI10. Using this approach, the experimental spectrum of $HEI10_{1-196LL}$ at 20 and 40 μM have a calculated concentration of 19.98 and 33.66 μM respectively. This result infers that a single zinc atom is coordinated by a HEI10 molecule, suggesting that 4 zinc atoms are coordinated by the HEI10 tetramer. Furthermore, this result confirms that the structural core of HEI10 is folded.



e.

Concentration HEI10 1-196$_{5xTGS}$ (µM)	Calculated concentration Zn (µM) at $A_{413.5}$	Calculated concentration Zn (µM) at A_{491}	Inferred Zn atoms per molecule
20	18.71	21.26	1
40	32.65	34.67	1

Figure 5.2.9.| HEI10$_{1\text{-}196\ 5xTGS}$ zinc determination by PAR assay.
a) Sequence alignment of the N-terminus of HEI10 reveals a C3HCHC2 sequence of highly conserved histidine and cysteine residues. Highlighted in blue and green are the conserved histidine and cysteine residues, respectively. b) Absorbance spectrum between 300-600 nm of PAR with the addition of 10-100 µM ZnOAc (zinc standards - grey). The spectrum of PAR alone (0 µM) has a single absorption peak at 414 nm. The addition of micromolar concentrations of Zn^{2+} causes a decrease in the PAR absorbance at 414 nm concomitantly with an increase in the absorbance at higher wavelengths. At 100 µM ZnOAc there is a single absorption peak at 494 nm. The absorbance spectrum of HEI10$_{1\text{-}196\ 5xTGS}$ was determined at two concentrations: 20 µM (light blue) and 40 µM (dark blue). c-d) Experimental spectrum at absorbance 413.5 (c) and 491 (d) were fitted as a linear standard, with R^2 values of 0.9948 and 0.9951, respectively. Zinc standards with absorptivity outside of the linear range are shown in grey and have been excluded from calculating the linear equation. HEI10$_{1\text{-}196\ 5xTGS}$ samples are plotted (blue dots). e) Concentration of Zn can be calculated in the protein samples from the linear plot and inferred a single Zn atom per HEI10 molecule, suggesting 4 Zn atoms per HEI10 tetramer. (n=1).

5.2.7. HEI10$_{1\text{-}196LL}$ binds linear dsDNA

The HEI10 structural core is highly basic (theoretical pI of 8.1) and binds to the cation HiTrap Heparin HP column. Heparin mimics the polyanionic structure of the nucleic acid, suggesting DNA binding ability. Furthermore, cysteine-rich amino acid sequence motifs, including RING domains, are common in a number of diverse proteins thought to interact with DNA (Freemont, Hanson, & Trowsdale, 1991).

We analysed the ability of HEI10$_{1\text{-}196LL}$ to bind dsDNA using electrophoretic mobility shift assay (EMSA). 25 μM (per bp) of linear dsDNA was incubated with increasing amounts of HEI10$_{1\text{-}196LL}$ (0-250 μM). Incubated samples were then run on an agarose gel, where shifts in the bands were observed which clearly indicated an affinity for dsDNA (Figure 5.2.10.). A DNA shift was observed at 25 μM of HEI10 and between 75-250 μM of HEI10 there is an observed smearing between free DNA and protein-DNA complex. This observation suggests unstable protein-DNA complexes form and that HEI10 only has a weak affinity to DNA.

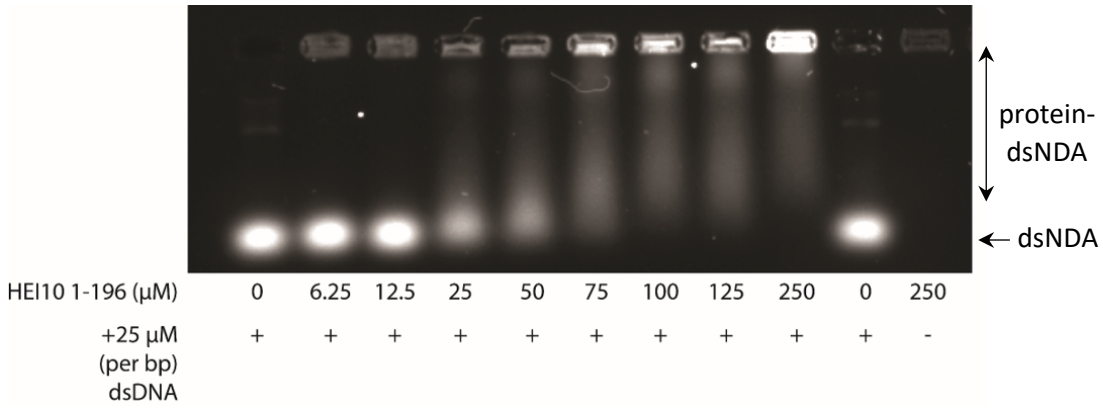


Figure 5.2.10.| EMSA demonstrate the DNA binding ability of HEI10$_{1\text{-}196LL}$.
Electrophoretic mobility shift assays (EMSA) demonstrate that the structural core of HEI10 has the ability to interact with 25 μM (per base pair) linear double-stranded DNA (dsDNA).

5.2.8. Crystallography of HEI10$_{1\text{-}196LL}$

Crystallographic studies of HEI10 were initiated by an Erasmus student, Carmen Espejo Serrano, who made a good start to the project. High throughput screening of HEI10$_{1\text{-}196LL}$ was carried out in all 7 commercial screens JCSG+, MPD, Index, Proplex, Ammonium sulphate, PACT and Morpheus. Crystal

hits were yielded in multiple screens. However, all crystals were very small, even after detailed optimisation. Microseed matrix seeding (MMS) method was used, where a seed stock dilution of 1:10,000 was systematically added to the commercial crystal screens in a ratio of 1:1:0.3 nl (protein: drop: seed stock). By using this method it was possible to grow larger single crystals of better quality, likely due to the crystals growing at lower levels of saturation.

Crystals of HEI10$_{1\text{-}196LL}$ from MPD H4 (0.1 M HEPES sodium salt pH 7.5 30% (w/v) MPD; 5% (w/v) PEG 4000) commercial screen were harvested and sent to Diamond Light Source (DLS) for X-ray diffraction (Figure 5.2.11.a.). The crystals diffracted to 2.37 Å but diffraction was highly anisotropic (Figure 5.2.11.b.). Despite multiple efforts it was not possible to gain any useable data to solve the structure. The X-ray diffraction data is detailed in Table 5.2.1. All data processing was performed by Dr Owen Davies.

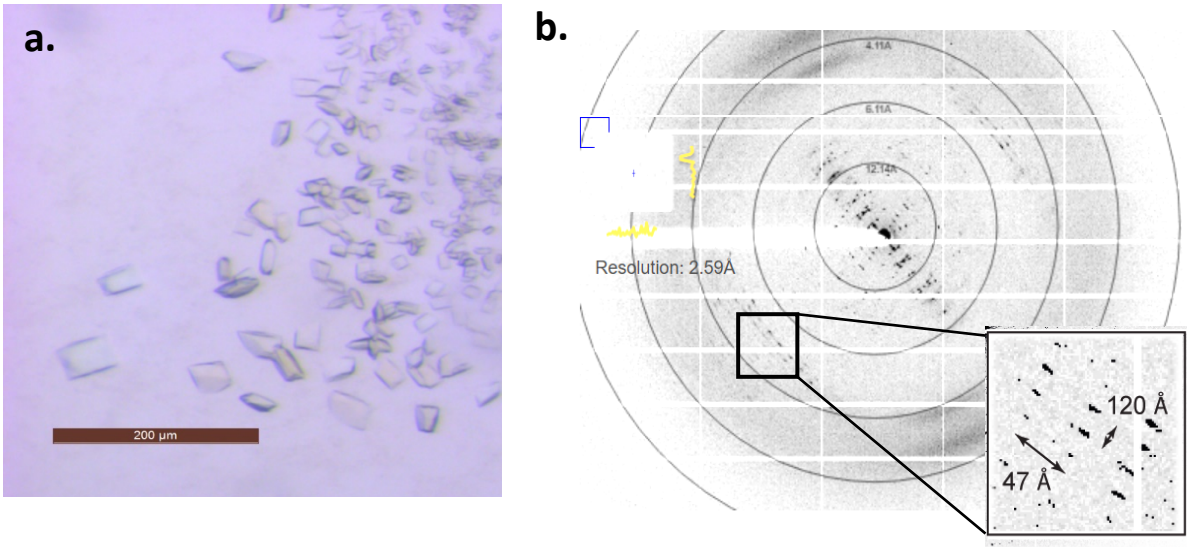


Figure 5.2.11.| HEI10$_{1\text{-}196LL}$ diffracts anisotropically.
a) Crystal images of HEI10$_{1\text{-}196\ 5xTGS}$, crystallised in MPD H4 (0.1 M HEPES sodium salt pH 7.5 30% (w/v) MPD; 5% (w/v) PEG 4000) using a 1:1000 seed stock solution. b) Diffraction image of a diffracting HEI10$_{1\text{-}196LL}$ crystal on the I24 beamline, the crystal diffracted to anisotropically to 2.37 Ä.

Table 5.2.1.| Data collection, phasing and refinement statistics for HEI10$_{1-196}$

	HEI10 1-196
Data collection	
Space group	$P2_1$
Cell dimensions	
a, *b*, *c* (Å)	53.23, 73.94, 124.10
α, β, γ (°)	90, 99.28, 90
Resolution (Å)	63.30 – 2.43 (2.75 – 2.43)*
Ellipsoidal resolution (Å)	3.319 (0.639 a* - 0.770 c*)
(direction)	5.213 (b*)
	2.285 (0.164 a* + 0.986 c*)
R_{meas}	0.135 (1.330)
R_{pim}	0.052 (0.496)
$I / \sigma(I)$	9.3 (1.7)
$CC_{1/2}$	0.997 (0.622)
Completeness (spherical) (%)	32.3 (5.2)
Completeness (ellipsoidal) (%)	87.0 (72.4)
Redundancy	6.6 (7.1)

*Values in parentheses are for highest-resolution shell.

HEI10$_{1-196LL}$ MPD H4 optimisations from multiple protein purifications were set up, varying the buffer and precipitant conditions. Single crystals of ~100 µm length and 50 µM diameter were harvested and sent to DLS for observation however crystals did not diffract or diffracted poorly (Figure 5.2.12.). The MPD commercial screen also yielded HEI10$_{1-196LL}$ crystals in H7 (35% MPD) (Figure 5.2.13.a.) Optimisations of these crystals were carried out, varying the percentage of MPD as well as changing the seed stock dilution (Figure 5.2.13.b.-g.). Both crystals from the initial screen and optimisations were tested for X-ray diffraction at DLS. Crystals diffracted poorly or not at all, indicating that we needed to alter the construct or purification process of HEI10 to yield better diffracting crystals.

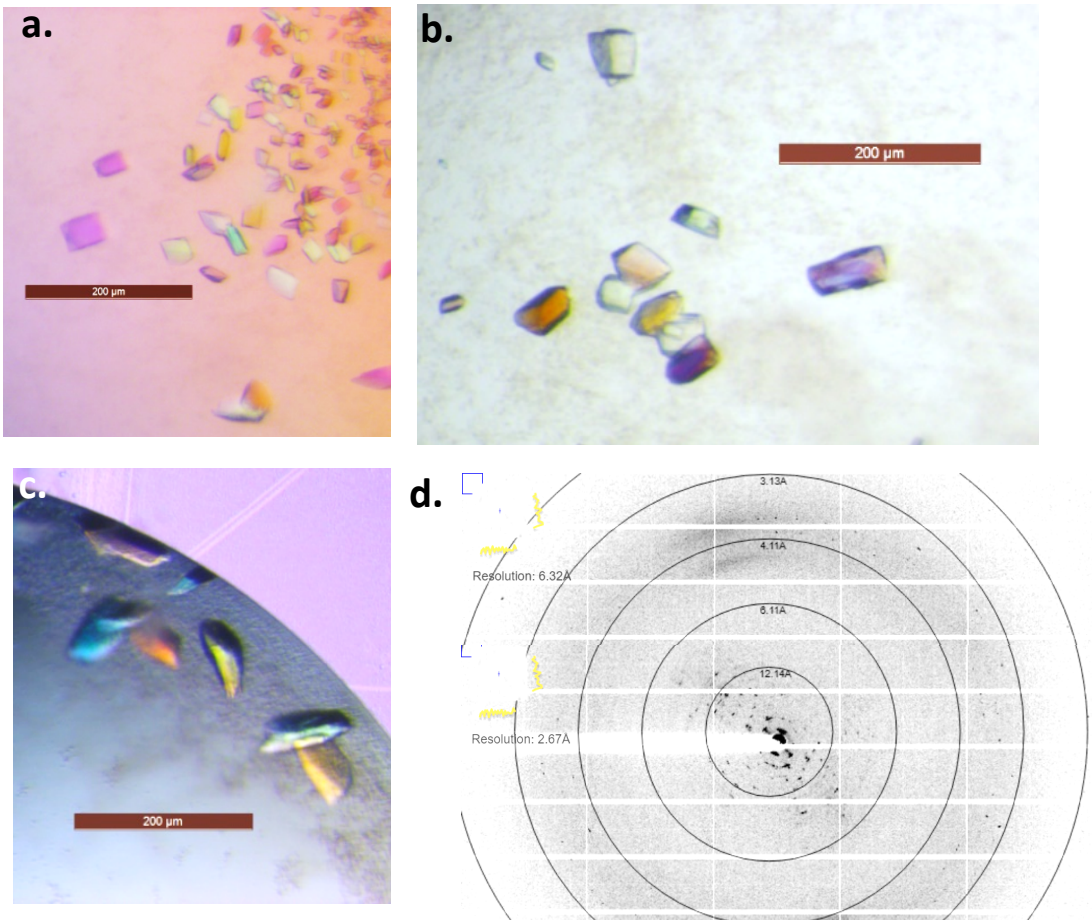


Figure 5.2.12.| HEI10$_{1\text{-}196LL}$ MPD H4 crystal hits.
a) MPD H4 crystals of HEI10$_{1\text{-}196LL}$ set up at 13 mg/ml (0.1M Hepes sodium salt pH 7.5, 30% MPD, 5% PEG 4000) using a 1 in 10,000 seed stock. b-c) Replication of MPD H4 crystal hit. d) X-ray diffraction image of HEI10$_{1\text{-}196}$ shows very poor diffraction.

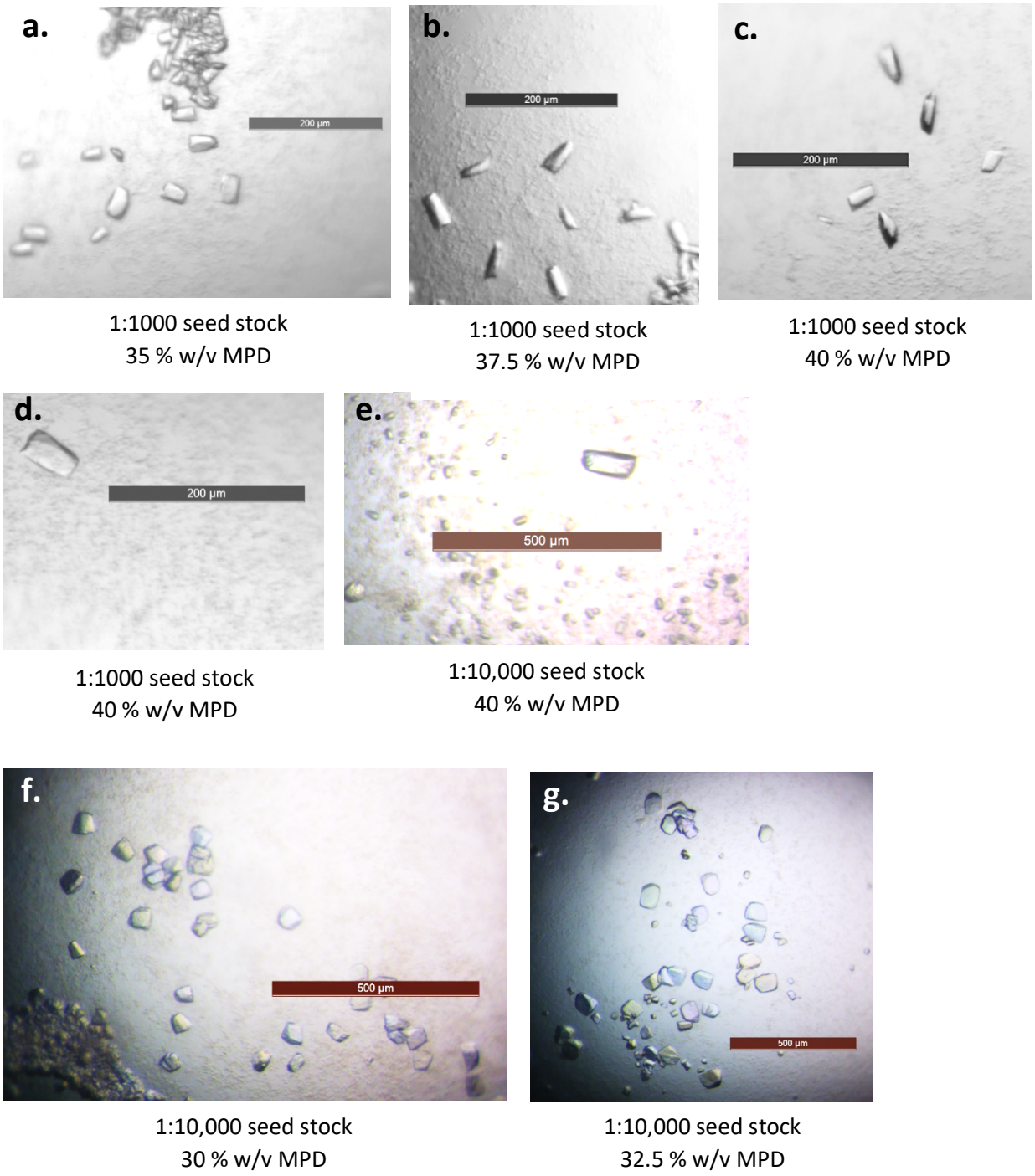


Figure 5.2.13.| HEI10$_{1\text{-}196LL}$ MPD H7 optimisation crystal hits.
a-g) Optimisation of HEI10$_{1\text{-}196\ 5xTGS}$ (pMAT11 vector) MPD H7 crystals (30% MPD), varying the MPD w/v percentage and the seed stock dilution.

5.2.9. Optimisation of HEI10$_{1\text{-}196}$ expression

Data of the HEI10 structural core thus far has been shown using HEI10$_{1\text{-}196}$ with a long TGS-linker, cloned into the pMAT11 plasmid which contains a N-terminal His$_6$-MBP-solubility tag. From the

expression tests and initial purification steps, we determined that around half of the expressed material is insoluble, as observed in the pellet (Figure 5.2.14.a-c., bands at ~70 kDa are highlighted). In addition, changing lysis conditions did not significantly improve solubility, suggesting that HEI10 is difficult to fold. This could be due to the intricate folding and zinc coordination of the RING domain. We tested the expression of HEI10 cloned into a low copy number plasmid, pRSF duet, with an N-terminal MBP tag. MBP-HEI10$_{1\text{-}196LL}$ and MBP-HEI10$_{1\text{-}196}$ were expressed in *E.coli* BL21 (DE3) cells and grown using the same method as previously shown. Cells were lysed and SDS-PAGE analysis revealed that high expression is maintained, and the amount of insoluble material is reduced, more so in MBP-HEI10$_{1\text{-}196}$ (Figures 5.2.14.d.-e.). This analysis suggests that HEI10$_{1\text{-}196}$ cloned into a low copy number plasmid improves protein solubility.

To further test this hypothesis, we purified MBP-HEI10$_{1\text{-}196LL}$ and MBP-HEI10$_{1\text{-}196}$ by sequential amylose affinity and anion exchange affinity chromatography. Both fusion constructs were incubated with TEV-protease to cleave the N-terminal His$_6$-MBP-solubility tag. SDS-PAGE analysis revealed that both constructs of HEI10 successfully cleaved, indicating that when HEI10$_{1\text{-}196}$ is cloned into the pRSF vector, a long TGS linker is not required to facilitate cleavage of the solubility tag (Figure 5.2.15.). We can hypothesise that poor cleavage was likely due to a portion of aggregated protein rather than steric hindrance per se.

The N-terminal MBP-tag of HEI10$_{1\text{-}196LL}$ and HEI10$_{1\text{-}196}$ were cleaved by TEV-protease and further purified through sequential cation exchange and size exclusion chromatography. Purification steps of both constructs are summarised in Figure 5.2.16.a. and c. SEC-MALS analysis was used to confirm if the oligomeric state of the HEI10 structural core is maintained. Both HEI10$_{1\text{-}196LL}$ and HEI10$_{1\text{-}196}$ eluted as a single peak, with a molecular mass of 104 and 109.2 kDa, respectively. These results indicate that the tetrameric state is maintained (tetrameric theoretical molecular weight – 96.7 and 90.7 kDa, respectively) (Figure 5.2.16.b. and d.).

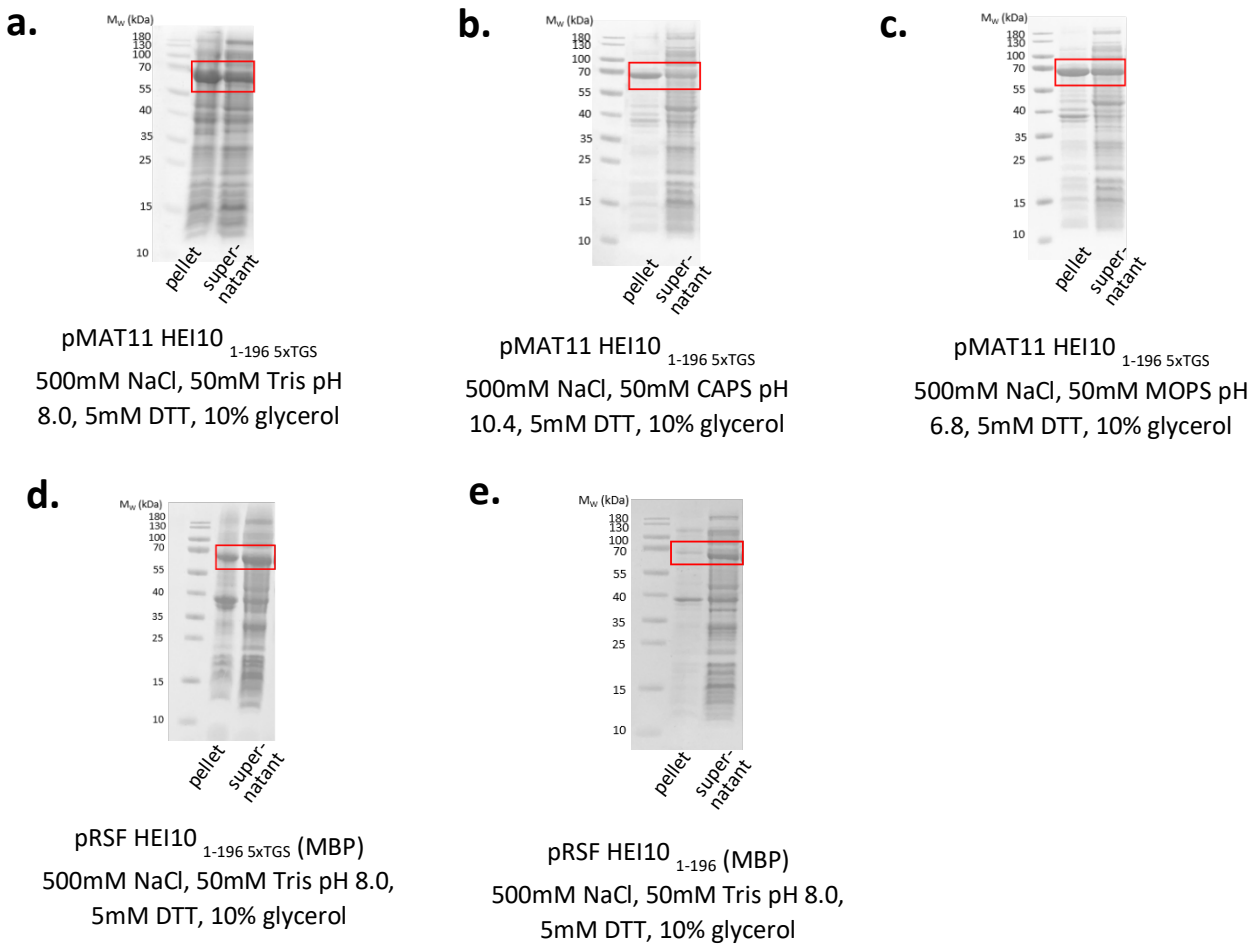


Figure 5.2.14.| Optimisation of HEI10$_{1\text{-}196}$ expression.
a-c) SDS page analysis of HEI10$_{1\text{-}196\ 5xTGS}$ lysis with various pH conditions; a) pH 8.0 (Tris) b) pH 6.8 (MOPS) c) pH 10.4 (CAPS). d-e) HEI10$_{1\text{-}196}$ with (d) and without (e) a 5xTGS linker was cloned into the pRSF duet vector with an N-terminal MBP-tag and expressed using the same conditions. Highlighted in red are bands at ~70 kDa in the pellet and supernatant of the bacterial lysate of HEI10$_{1\text{-}196}$.

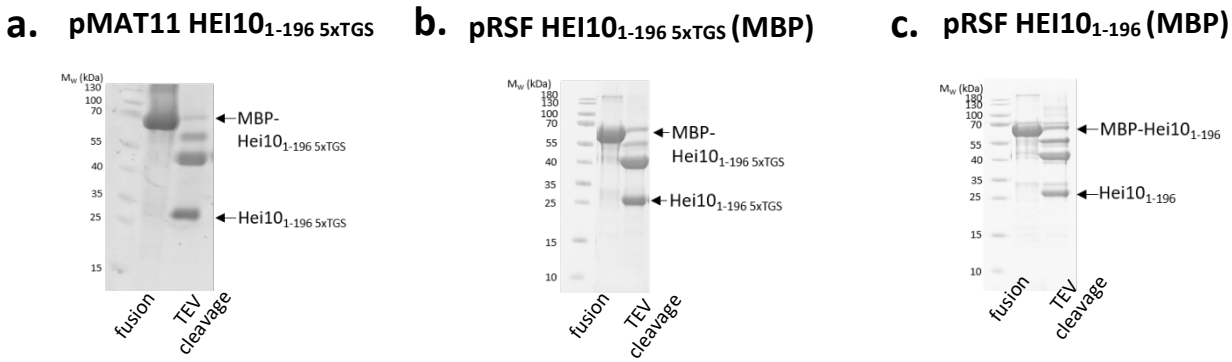


Figure 5.2.15.| HEI10$_{1\text{-}196}$ cloned into a pRSF-duet vector does not require a TGS long linker to facilitate cleavage.
a-c) SDS page analysis of MBP-HEI10$_{1\text{-}196}$ TEV-protease cleavage. a) HEI10$_{1\text{-}196}$ cloned in pMAT11 vector requires a 5xTGS flexible long linker to achieve TEV-cleavage. b-c) HEI10$_{1\text{-}196}$ cloned in pRSF duet vector with (b) and without (c) a TGS linker.

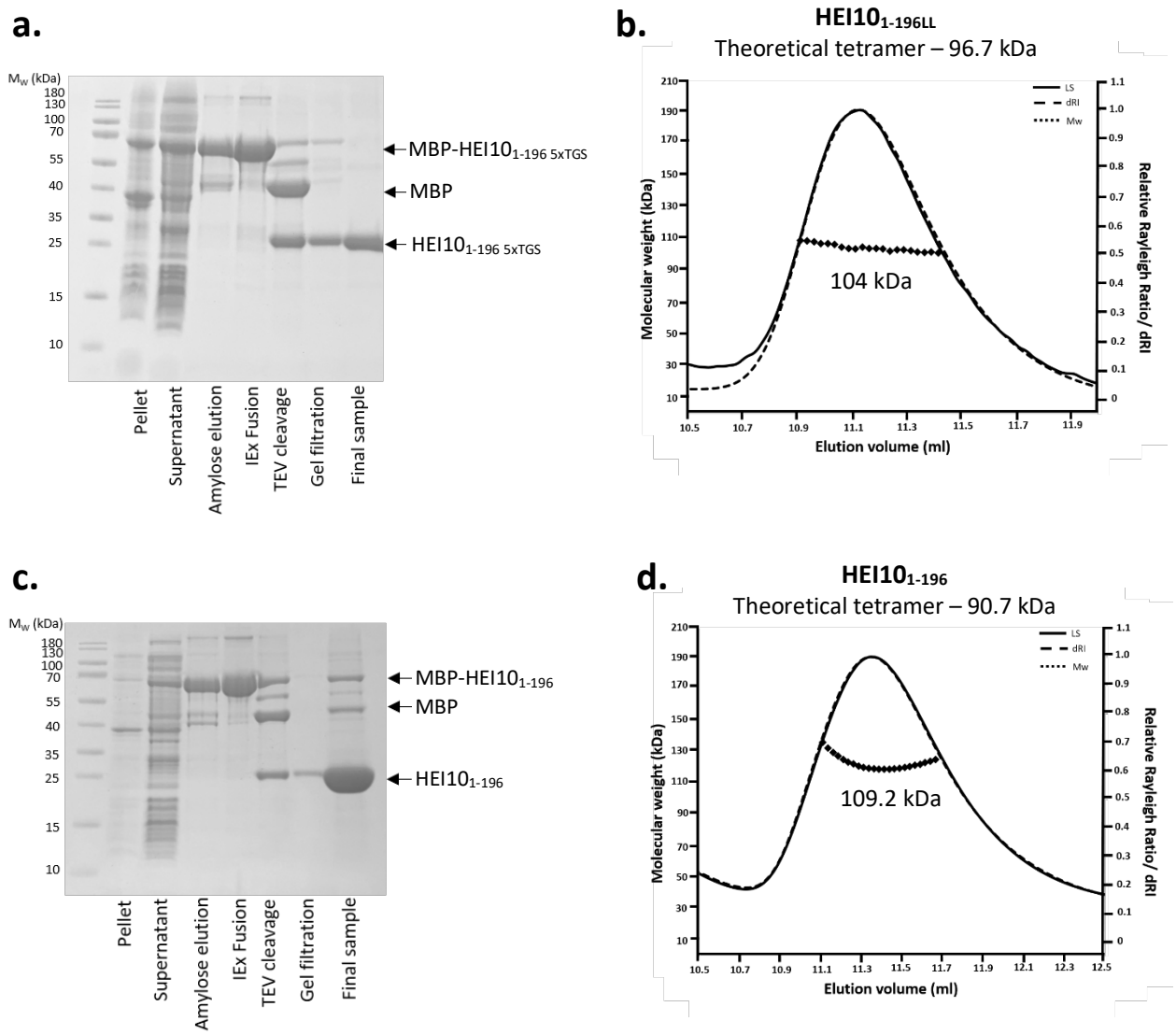
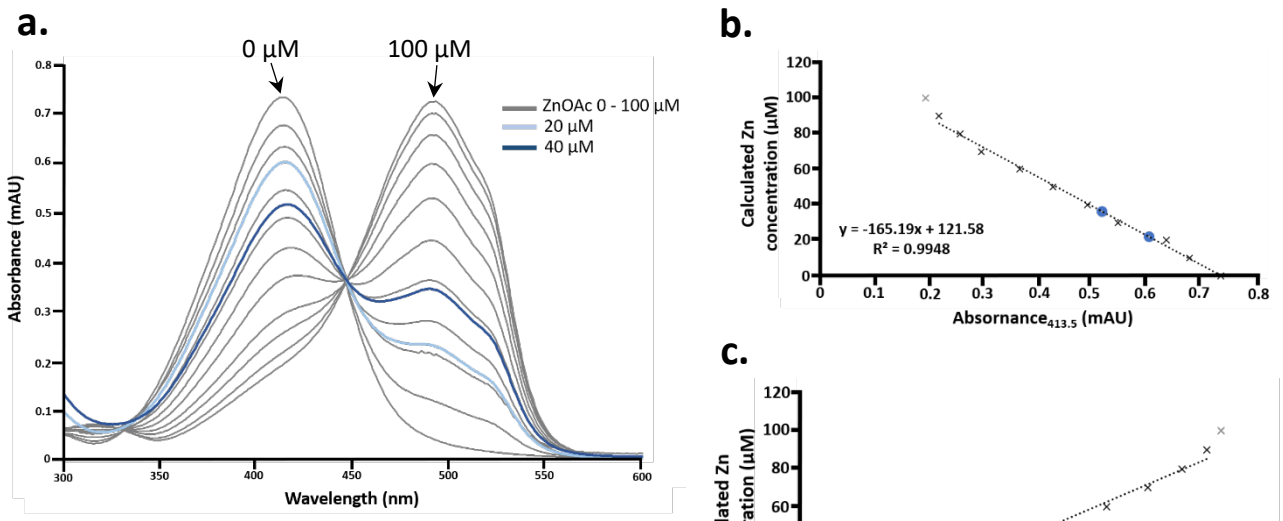


Figure 5.2.16.| Purification and SEC-MALS analysis of HEI10$_{1-196}$
a and c) SDS-PAGE showing the purification summary of HEI10$_{1-196\ 5xTGS}$ (a) and HEI10$_{1-196}$ (c) through sequential amylose affinity and anion exchange chromatography. N-terminal MBP-tag was removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. b) and d) SEC-MALS analysis of HEI10$_{1-196\ 5xTGS}$ (b) and HEI10$_{1-196}$ (d) revealed a single peak with an experimental molecular weight of 104 and 109.2 kDa, respectively (theoretical tetramer – 96.7 and 90.7 kDa) indicating that the HEI10$_{1-196}$ tetramer is not disrupted when using pRSF-duet.

To further determine that HEI10$_{1-196}$ was correctly folded, a PAR assay was utilised to determine the zinc atom content. Co-ordinated zinc atoms were released from the protein through proteolytic digestion and chelated by PAR, changing its spectrophotometric properties. The absorbance spectra of digested HEI10$_{1-196}$ was determined at two concentrations (20 and 40 μM) have a trace closely matching that of the 20 and 40 μM zinc standards, respectively (Figure 5.2.17.a.). Using the linear plots, the zinc content

of HEI10$_{1-196}$ can be determined at 22.75 and 37.42 µM, respectively (Figure 5.2.17.b.-d.). This result matches the inferred zinc content of HEI10$_{1-196LL}$ of 1 zinc atom per HEI10 molecule.



d.

Concentration HEI10 1-196 (µM)	Calculated concentration Zn (µM) at $A_{413.5}$	Calculated concentration Zn (µM) at A_{491}	Inferred Zn atoms per molecule
20	22.05	23.46	1
40	36.40	38.45	1

Figure 5.2.17.| Changing the expression vector does not effect HEI10$_{1-196}$ zinc content
a) Absorbance spectrum between 300-600 nm of zinc standards (0-100 µM - grey). The absorbance spectrum of HEI10$_{1-196}$ was determined at two concentrations: 20 µM (light blue) and 40 µM (dark blue). b-c) Experimental spectrum at absorbance 413.5 (a) and 491 (b) were fitted as a linear standard, with R^2 values of 0.9948 and 0.9951, respectively. Zinc standards with absorptivity outside of the linear range are shown in grey and have been excluded from calculating the linear equation. HEI10$_{1-196}$ samples are plotted (blue dots). d) Concentration of Zn can be calculated in the protein samples from the linear plot and inferred a single Zn atom per HEI10 molecule, suggesting 4 Zn atoms per HEI10 tetramer. (n=1)

5.2.10. Further crystallography of HEI10$_{1-196}$

Commercial screens of the HEI10 structural core without the TGS long linker were set up at 12.3 mg/ml, without a seed stock solution. Removal of the long linker (15 amino acid residues) is a significant truncation so may improve crystal growth. Crystal hits were obtained in three new conditions (Morpheus G2, C2 and C6), as well as the previous determined MPD H4 (Figure 5.2.18.a.-e.). The

crystal hits in morpheus were very small and did not improve upon optimisation. HEI10$_{1-196}$ MPD H4 optimisations were set up at 11.5 mg/ml, varying the buffer and precipitant conditions (Figure 5.2.18.a.-e.). The biggest crystals were harvested and sent to DLS for testing but did not diffract (Figure 5.2.18.f.). This was the same case as the previously tested crystals.

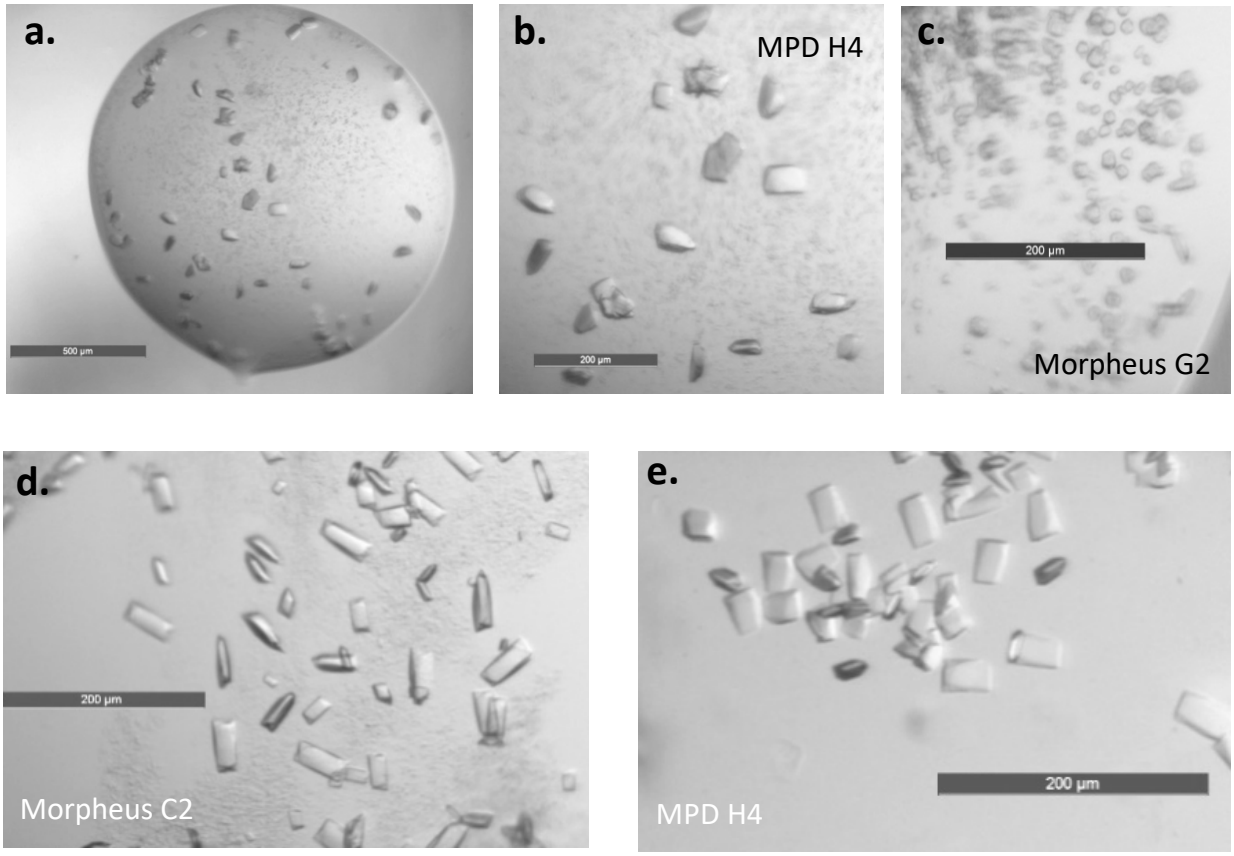


Figure 5.2.18.| HEI10$_{1-196}$ Crystal hits.
Commercial screens of HEI10$_{1-196}$ were set up at 12.3 mg/ml. Crystals hits were determined in a-b) MPD H4 and c-d) Morpheus G2 and C2. e) MPD H4 (0.1M HEPES sodium salt pH 7.5, 30% MPD, 5% PEG 4000) optimisations set up at 11.5 mg/ml.

So far all protein preparations have been using buffers containing 10 % glycerol to enhance the stability of HEI10 and prevent protein aggregation. However, the presence of glycerol can be disadvantageous for protein crystallisation. Lattices in protein crystals are stabilised by protein-protein contacts but also by interactions established by the hydration shells surrounding the protein molecules. The addition of glycerol can alter the hydrogen bond network that strengthen crystal contacts, which may negatively affect nucleation events and the rate of crystal growth (Vera et al., 2011).

We have shown that HEI10 does require glycerol, therefore we tested the stability of $HEI10_{1-196}$ in different buffer conditions including the cosolvent L-arginine (pH 8.0). L-arginine is used as an aggregation suppressor, by coating the exposed hydrophobic surfaces of the protein (Arakawa et al., 2007). $Hei10_{1-196}$ (purified in buffer containing 10% glycerol) was buffer exchanged into standard buffer with the addition of 20mM L-arginine (pH 8.0) and purified by SEC (grey trace). The SEC chromatogram overlays with $HEI10_{1-196}$ purified in glycerol (control - black trace) suggesting that protein stability is maintained (Figure 5.2.19.a.). Furthermore, we buffer exchanged into a high salt concentration (500 mM) as protein stability is suggested to increase with salt concentration (Mao et al., 2007). The SEC profile of HEI10 purified in 500 mM NaCl also overlays with the control, suggesting stability is maintained (Figure 5.2.19.a., blue trace).

MPD H4 (0.1M HEPES sodium salt pH 7.5, 30% MPD, 5% PEG 4000) crystal optimisation screens were set up for $HEI10_{1-196}$ purified in 20mM L-arginine (pH 8.0) at 11.5 mg/ml and 500mM NaCl at 7 mg/ml. HEI10 purified in L-arginine yielded plate crystals with length and diameter of ~100nm (Figure 5.2.19.b.). Crystals were harvested and tested for diffraction but did not diffract.

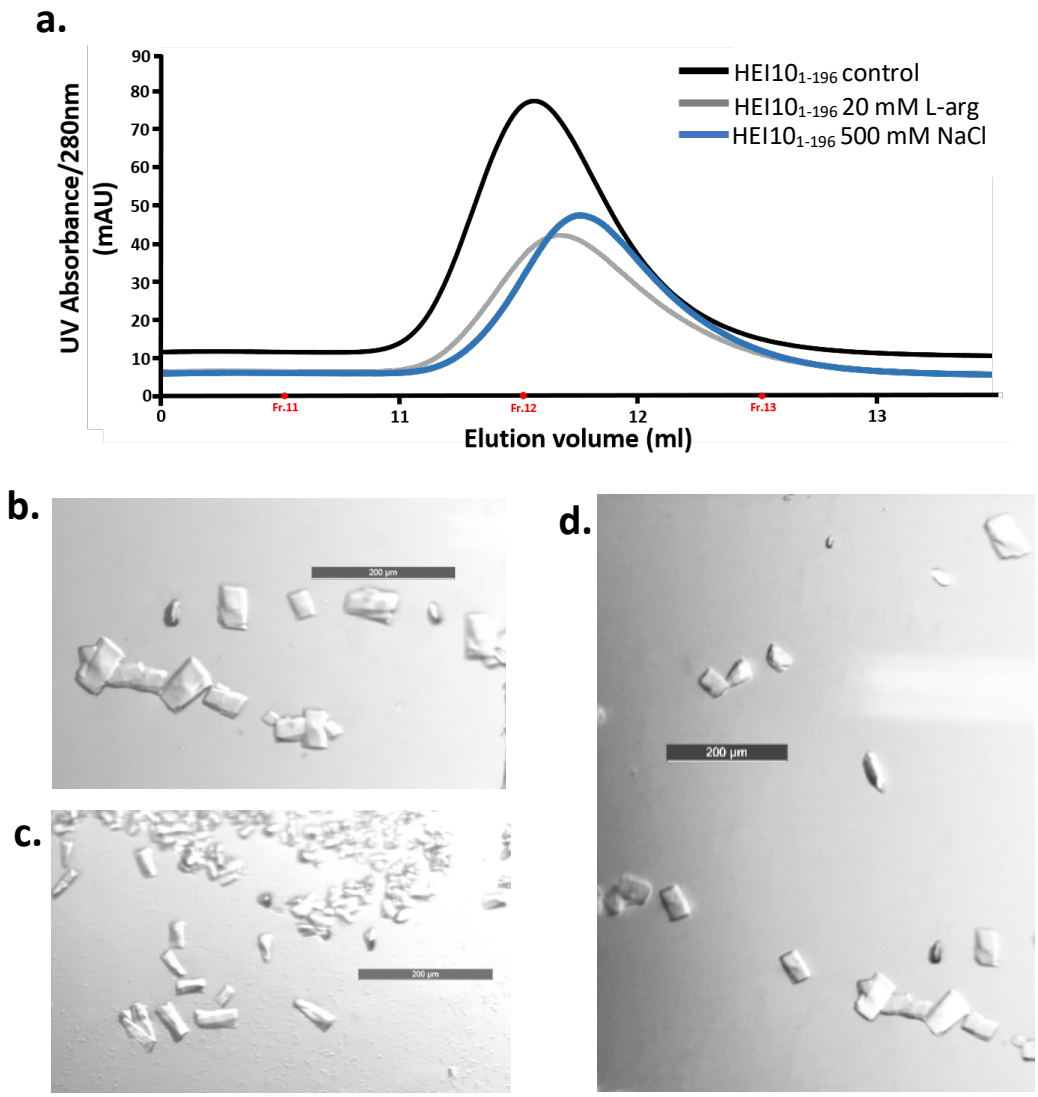
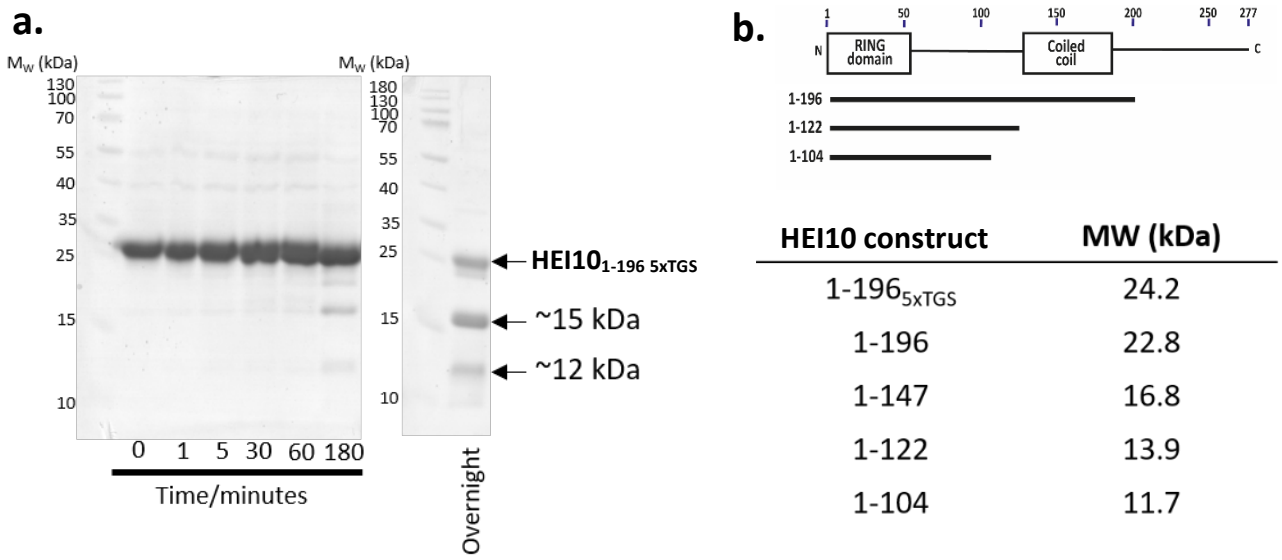


Figure 5.2.19.| HEI10$_{1\text{-}196\ 5xTGS}$ purified in 20mM L-arginine crystallised in MPD H4.
a) Overlay size-exclusion chromatograms (SEC) of HEI10$_{1\text{-}196}$ purified in standard buffer (150 mM NaCl, 50 mM Tris, 5 mM DTT) with the addition of 20mM L-arginine (pH 8.0) (grey), and 500mM NaCl (blue) and 10% glycerol (black – control). b-d) Crystal optimisation screens were set up for HEI10$_{1\text{-}196}$ purified in 20mM L-arginine (pH 8.0) at 11.5 mg/ml. Crystal growth in MPD H4 (0.1M HEPES sodium salt pH 7.5, 30% MPD, 5% PEG 4000).

5.2.11. Optimisation of the HEI10 structural core

A limited proteolysis assay was ultilised to probe the HEI10 structural core. Limited proteolysis can support information regarding protein structure and conformational changes. This method can be utilised to establish the compact domain of a protein for crystallisation studies. HEI10$_{1\text{-}196}$ was incubated with trypsin enzyme at a 1:1000 molar ratio. During digestion, samples were taken over a 16-hour time

frame and analysed by SDS-PAGE to identify cleavage products (Figure 5.2.20.a.). We observed there are multiple degradation products with molecular mass of ~15 and 12 Kda, indicating that HEI10$_{1-196}$ is exposed to the trypsin protease. To further confirm this HEI10 secondary structure prediction and sequence conservation were analysed and a series of truncations were designed and cloned corresponding to these observed masses (Figure 5.2.20.b.).



HEI10 construct	MW (kDa)
1-196$_{5xTGS}$	24.2
1-196	22.8
1-147	16.8
1-122	13.9
1-104	11.7

Figure 5.2.20.| Limited proteolysis of HEI10$_{1-196\ 5xTGS}$
a) SDS-PAGE analysis showing the limited proteolysis assay of HEI10$_{1-196\ 5xTGS}$ using trypsin protease. Trypsin was added and samples were taken at a number of time points. Overnight analysis identified two cleavage products with molecular weights of ~15 and 12 kDa. b) Using α-helical predictions and sequence conservation three new constructs were designed: 1-147, 1-122 and 1-104, corresponding to the limited proteolysis products.

MBP-HEI10 constructs (1-147, 1-122, 1-104) were cloned into the pRSF vector and expressed in *E.coli* BL21 (DE3) cells. Bacterial cell lysate of MBP-HEI10$_{1-147}$ was purified by amylose affinity and anion exchange chromatography (Figure 5.2.21.a.). The MBP-tag was cleaved by TEV protease and further purified by cation exchange chromatography and SEC, final purified product is shown in figure 5.2.21.b. It is evident from the SDS-PAGE that there is a clear degradation product of HEI10$_{1-147}$, suggesting that this construct is not stable. Furthermore, SEC-MALS analysis revealed two peaks with molecular weight of 68.4 and 39.8 kDa, respectively (Figure 5.2.21.c.). This result suggests that HEI10$_{1-}$

$_{147}$ forms a tetramer (theoretical tetramer – 67 kDa), in addition to the degradation product or a possibility of dimeric species (theoretical dimer – 37.5 kDa).

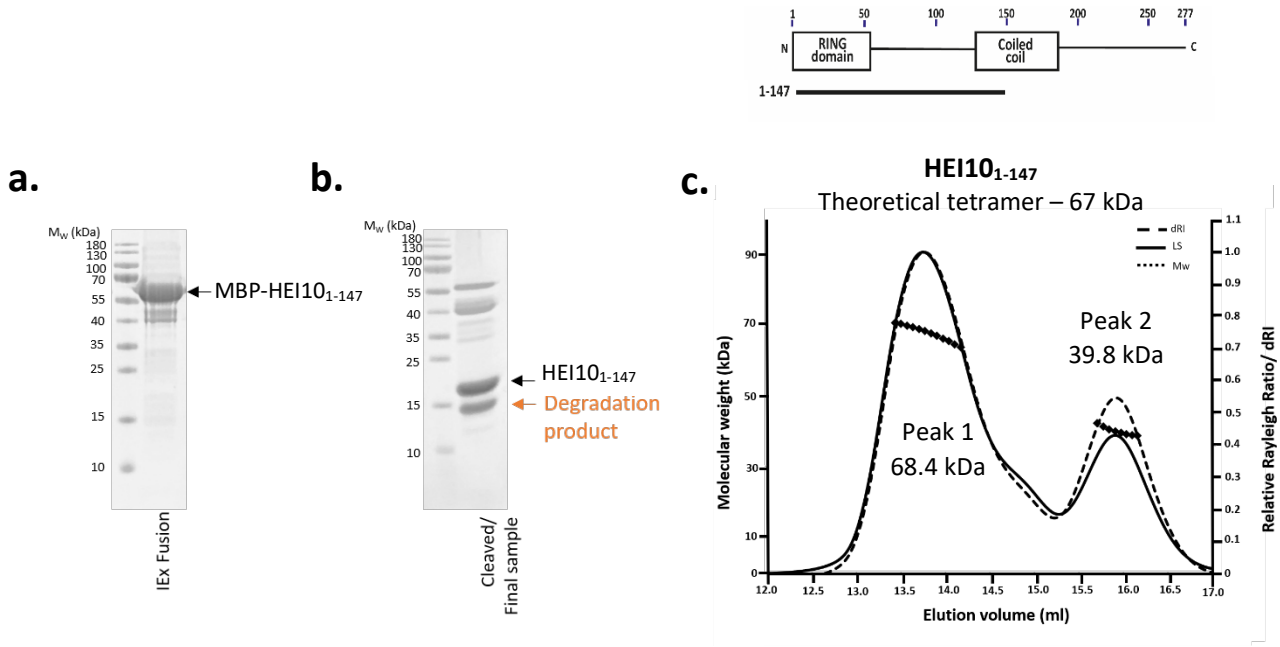


Figure 5.2.21.|Purification and SEC-MALS analysis of HEI10$_{1-147}$.
a-b) SDS-PAGE analysis of purified MBP-HEI10$_{1-147}$ (a) and cleaved HEI10$_{1-147}$ (b). Cleaved HEI10$_{1-147}$ has a clear band at ~ 14 kDa, indicating HEI10$_{1-147}$ is not a stable construct. c) SEC-MALS analysis of HEI10$_{1-147}$ revealed two peaks with molecular weights of 68.4 and 39.8 kDa.

5.2.12. Biophysical analysis and characterisation of HEI10$_{1-122}$

A C-terminal truncation of 74 amino acids, HEI10$_{1-122}$, was expressed with an MBP-tag encoded by the pRSF vector and purified in the same manner as HEI10$_{1-196}$. Purification steps are summarised in Figure 5.2.22.a. SEC-MALS analysis of HEI10$_{1-122}$ was used to determine the oligomeric state. HEI10$_{1-122}$ eluted as a single peak with a molecular mass of 53.6 kDa, corresponding to a tetramer (theoretical tetramer – 55.6 kDa), indicating tetramerisation is maintained (Figure 5.2.22.b.).

Far-UV CD spectroscopy was used to quantify the secondary structure composition of HEI10$_{1-122}$. Analysis revealed negative peaks of similar magnitude at 208 and 222 nm and a positive peak at 193 nm (Figure 5.2.22.c.). Deconvolution by DichroWeb estimates an α-helical content of 40%, corresponding to 49 helical residues out of 122, as expected with truncated α-helical domain. CD spectroscopy was also utilised to determine the thermal stability of HEI10$_{1-122}$ by tracking the helical

signal at 222 nm between 4 and 95°C. Co-operative unfolding revealed a melting temperature (T_m) of 67.4 °C (Figure 5.2.22.d.). Together these observations suggest that $HEI10_{1-122}$ is much more stable in comparison to $HEI10_{1-196}$ construct.

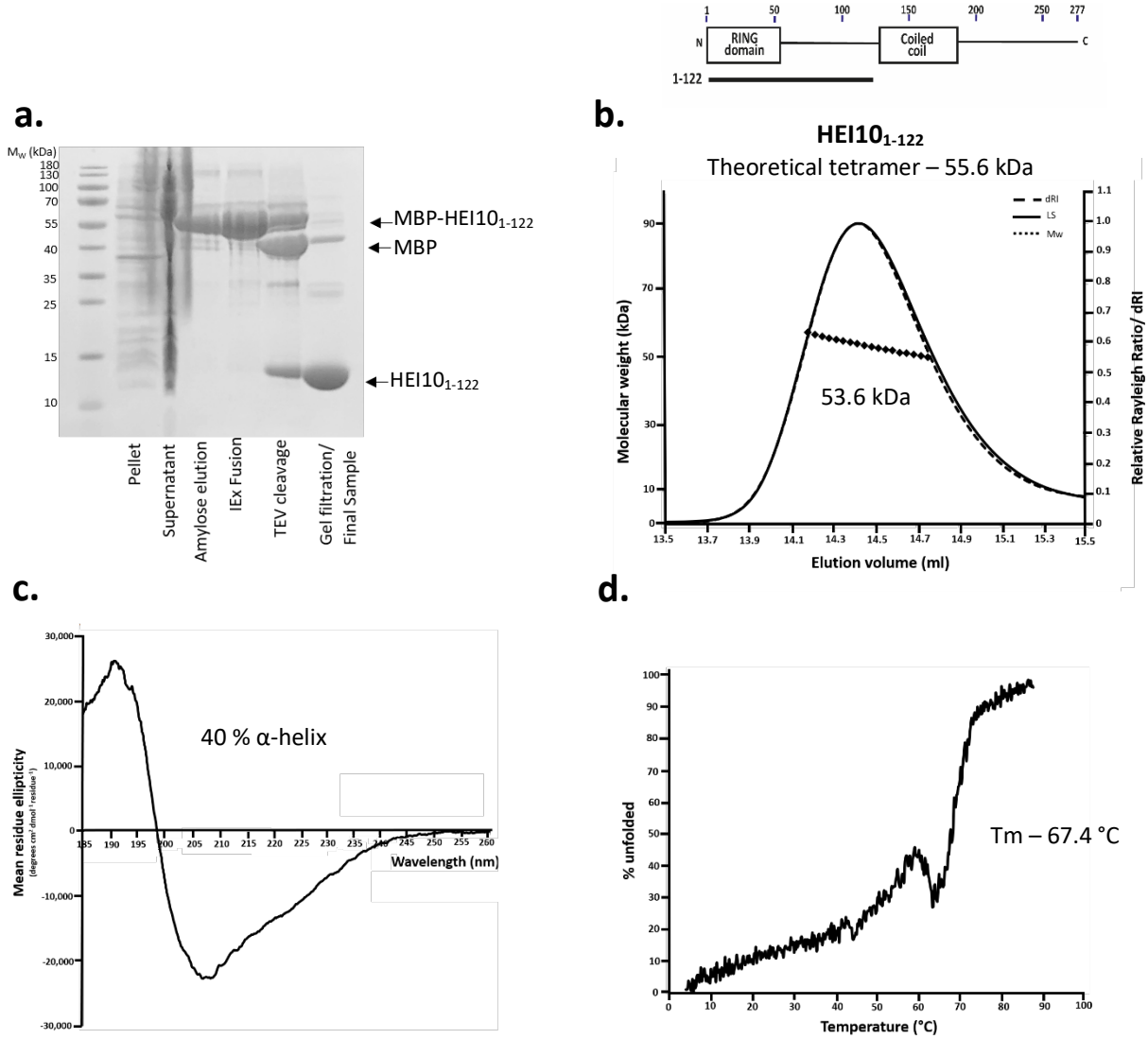
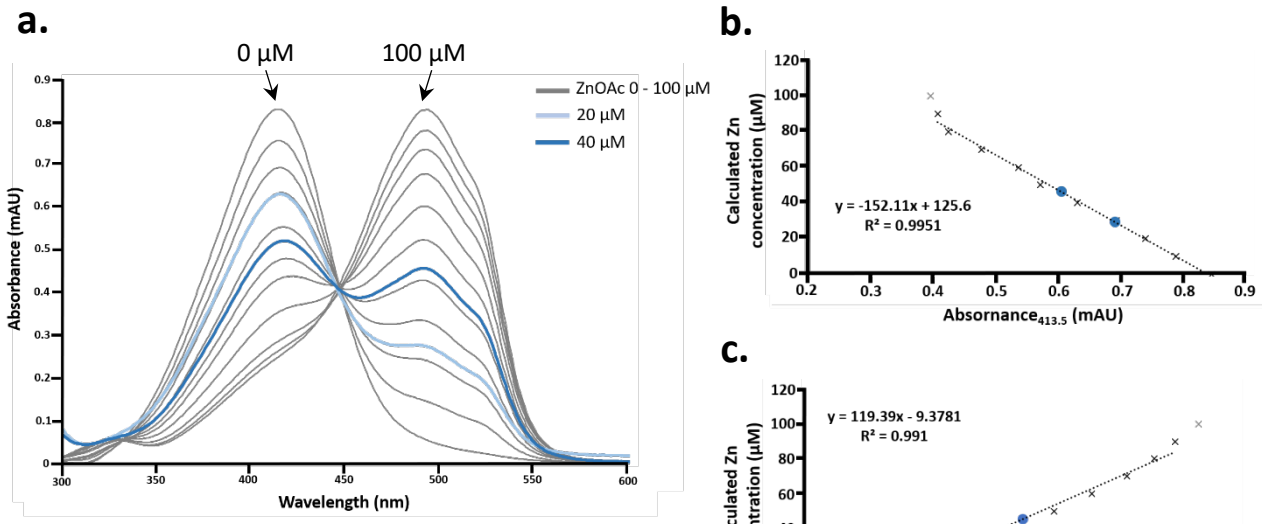


Figure 5.2.22.| Purification and biophysical analysis of $HEI10_{1-122}$ a) SDS-PAGE showing the purification summary of $HEI10_{1-122}$ through sequential amylose affinity and anion exchange chromatography. N-terminal MBP-tag was removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. b) SEC-MALS analysis of $HEI10_{1-122}$ revealed a single peak with an experimental molecular weight of 53.6 kDa (theoretical tetramer – 55.6 kDa). c) Far UV CD wavelength scan between 260-185 nm of $HEI10_{1-122}$ shows an α-helical trace. Deconvolution of the data estimates the secondary structure to be 40 % α-helical. d) CD thermal denaturation measured at 222 nm between 4 and 95 °C, estimated a melting temperature of 67.4 °C.

PAR assay of HEI10$_{1-122}$ was used to determine the zinc atom content. The absorbance spectra of zinc standards (0-100 µM) and HEI10$_{1-122}$ at two concentrations (20 and 40 µM) were determined. The protein traces closely match zinc standards 20 and 40 µM, respectively (Figure 5.2.23.a.). Using the linear plots, the zinc content of HEI10$_{1-122}$ can be determined at 26.53 and 45.89 µM, respectively (Figure 5.2.23.b.-d.). This result matches the inferred zinc content of HEI10$_{1-196}$ of 1 zinc atom per HEI10 molecule.



d.

Concentration HEI10 1-122 (µM)	Calculated concentration Zn (µM) at $A_{413.5}$	Calculated concentration Zn (µM) at A_{491}	Inferred Zn atoms per molecule
20	29.25	23.81	1
40	46.28	45.50	1

Figure 5.2.23.| HEI10$_{1-122}$ zinc content determination by PAR assay.
a) Absorbance spectrum between 300-600 nm of zinc standards (0-100 µM - grey. The absorbance spectrum of HEI10$_{1-122}$ was determined at two concentrations: 20 µM (light blue) and 40 µM (dark blue). b-c) Experimental spectrum at absorbance (a) 413.5 and (b) 491 were fitted as a linear standard, with R^2 values of 0.9951 and 0.991, respectively. Zinc standards with absorptivity outside of the linear range are shown in grey and have been excluded from calculating the linear equation. HEI10$_{1-122}$ samples are plotted (blue dots). d) Concentration of Zn can be calculated in the protein samples from the linear plot and inferred a single Zn atom per HEI10 molecule, suggesting 4 Zn atoms per HEI10 tetramer. (n=1).

SEC-SAXS analysis of the truncated HEI10$_{1-122}$ construct was utilised to gain low resolution structural data. HEI10$_{1-122}$ has a Guinier analysis determined *Rg* of 29.5 Å, which closely matches the real-space *Rg* of 30.87 Å (Figure 5.2.24.a.). The maximum dimension (*Dmax*) of HEI10$_{1-122}$ was determined to be

154 Å (Figure 5.2.24.b.). HEI10$_{1-122}$ distance distribution profile presents a more compact *Dmax* of 154 Å in comparison to 290 Å of HEI10$_{1-196}$, in line with the C-terminal truncation of the α-helical domain. The theoretical length of an α-helical structure of 122 amino acids would be 183 Å, based on 1.5 Å per residue, suggesting the HEI10$_{1-122}$ is predominantly globular.

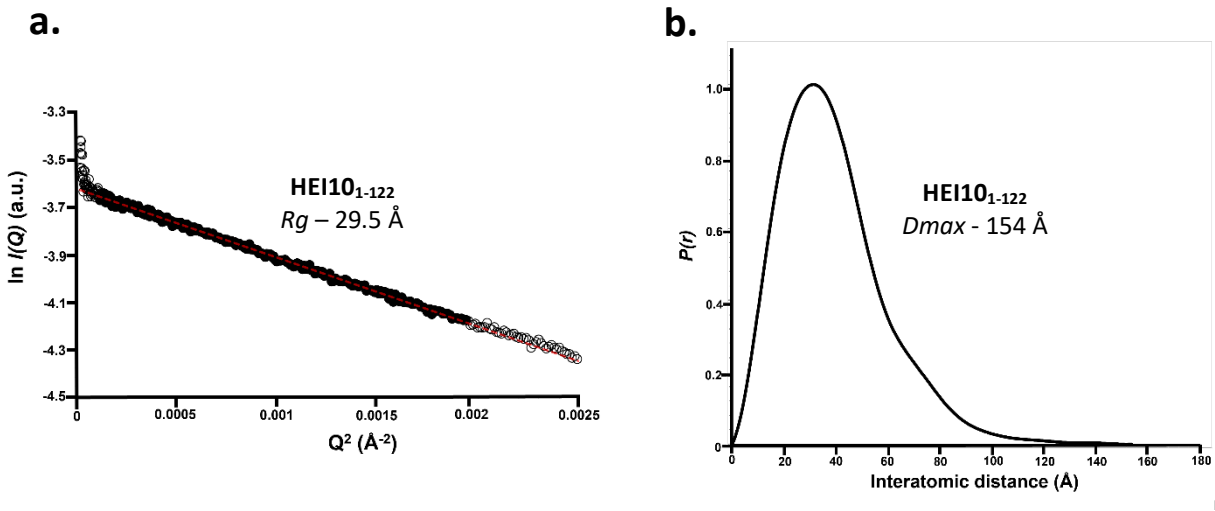


Figure 5.2.24.| SEC-SAXS analysis of HEI10$_{1-122}$.
a) Guinier analysis determined a radius of gyration (*Rg*) value of 29.5 Å. The real space *Rg* closely matches the Guinier analysis *Rg* value of 30.87 Å and 29.5 Å, respectively. Open circles represent the complete data set and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3). b) *P(r)* distribution of HEI10$_{1-122}$ showing a maximum dimension (*Dmax*) of 154 Å.

Commercial screens of the HEI10$_{1-122}$ were set up at 7 mg/ml. Crystal hits were determined in the conditions; Structure A11, MPD E2 and MPD G3, however crystal quality was very poor (Figure 5.2.25.). Optimisations of the conditions were set up, but crystals could not be reproduced.

The crystal conditions of Structure A11 and MPD G3 are very similar but with a different precipitant, 2-propanol and MPD, respectively. Optimisations of HEI10$_{1-122}$ were set up using different precipitants including different PEGs and alcohols but no crystals grew. The crystals in MPD E3 were used to create a seed stock (1:1000 and 1:10,000) and streak seeding was performed in the empty optimisation drops, with the hope to initiate nucleation events. Despite these attempts, no crystals were obtained.

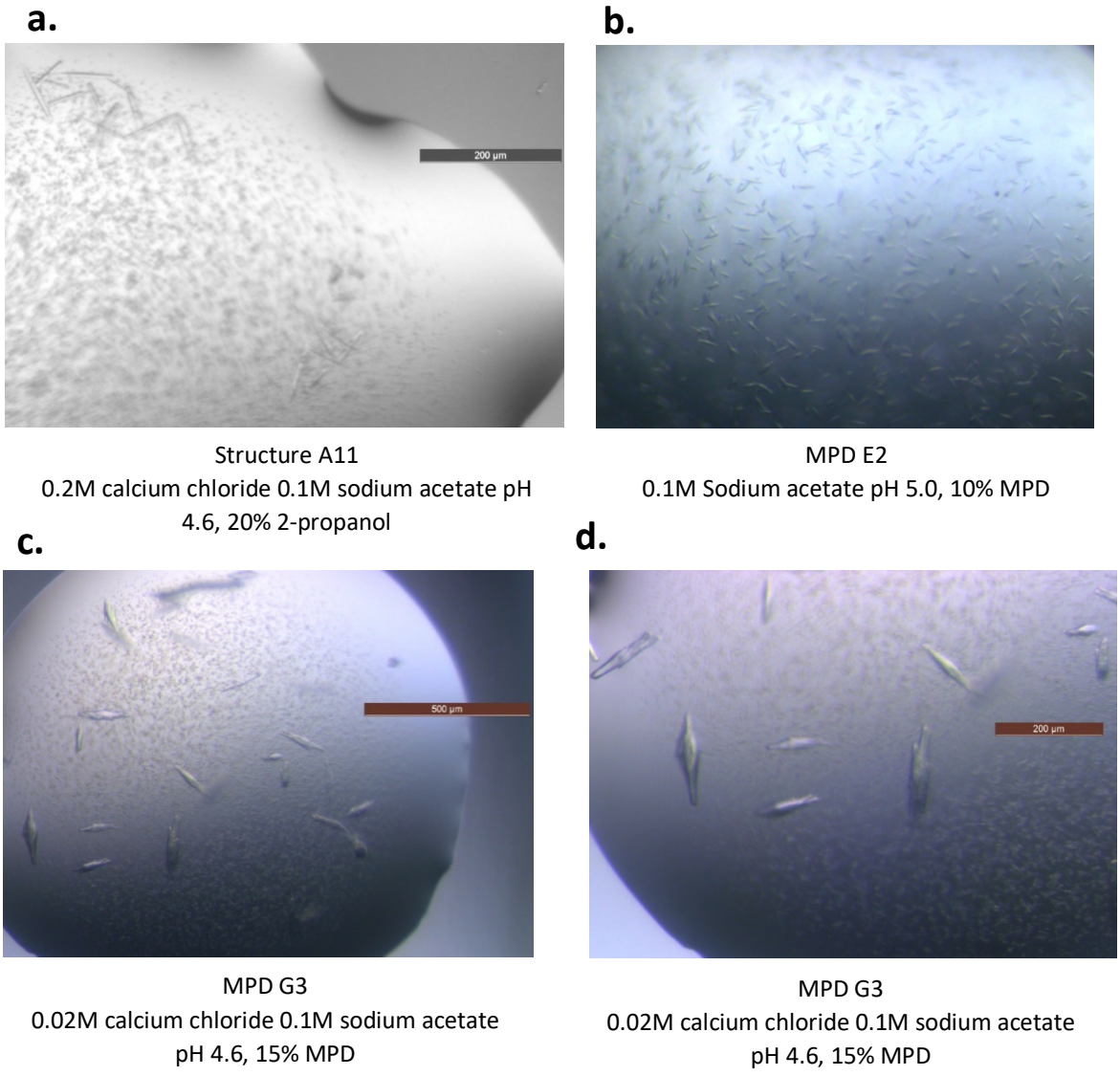


Figure 5.2.25.| HEI10$_{1-122}$ crystals hits.
Commercial screens of HEI10$_{1-122}$ were set up at 7 mg/ml. Crystals hits were obtained in a) Structure A11, b) MPD E2 and c-d) MPD G3.

5.2.13. Biophysical analysis and characterisation of HEI10$_{1-104}$

A further 18 amino acids were truncated from the C-terminus of the HEI10 structural core. HEI10$_{1-104}$ was expressed with an MBP-tag encoded by the pRSF vector and purified in the same manner as HEI10$_{1-196}$. An overview of the purification steps is given in Figure 5.2.26.a. SEC-MALS analysis of HEI10$_{1-104}$ revealed two overlapping peaks with molecular mass of 46.7 and 32.8 kDa, respectively. The first peak corresponds to a tetramer (theoretical tetramer – 46.8 kDa) but the latter peak suggests that the tetramer dissociates, possibly into dimers (theoretical dimer – 23.4 kDa). This result suggests that

HEI10$_{1\text{-}104}$ construct may have truncated too much from the α-helical coiled-coil domain (Figure 5.2.26.b.).

Far-UV CD spectroscopy of HEI10$_{1\text{-}104}$ maintained an α-helical spectrum (Figure 5.2.26.c.). Deconvolution by DichroWeb estimates an α-helical content of 32 %, corresponding to 33 helical residues out of 104. This result of a lower α-helical content is consistent with truncating the α-helical predicted coiled-coil region. CD spectroscopy was also utilised to determine the thermal stability of HEI10$_{1\text{-}104}$ by tracking the helical signal at 222 nm between 5 and 95°C. Co-operative unfolding revealed a melting temperature (T_m) of 62.2 °C (Figure 5.2.26.d.). These observations suggest that HEI10$_{1\text{-}104}$ has comparable properties to the longer HEI10$_{1\text{-}196}$ and HEI10$_{1\text{-}122}$ construct.

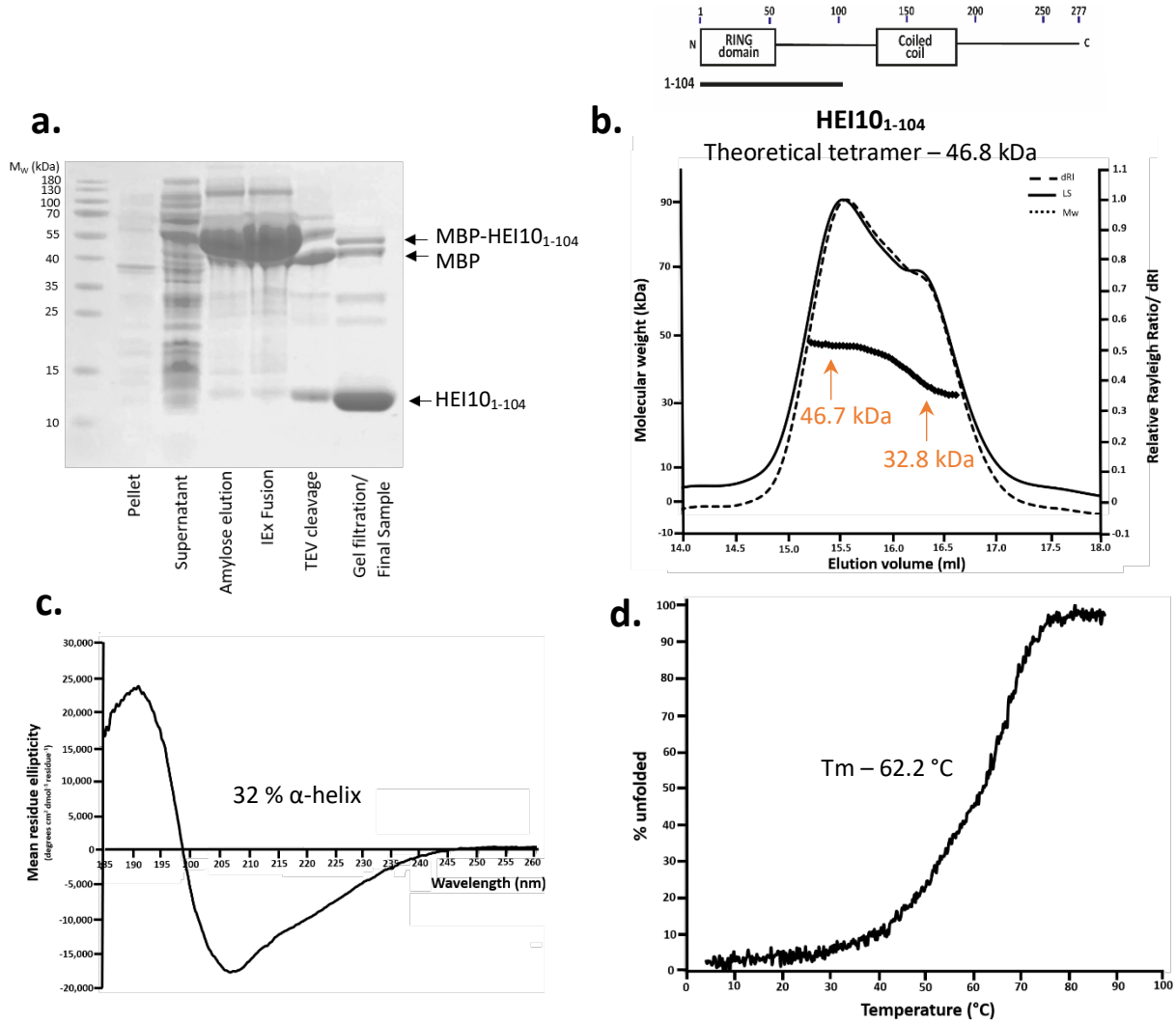
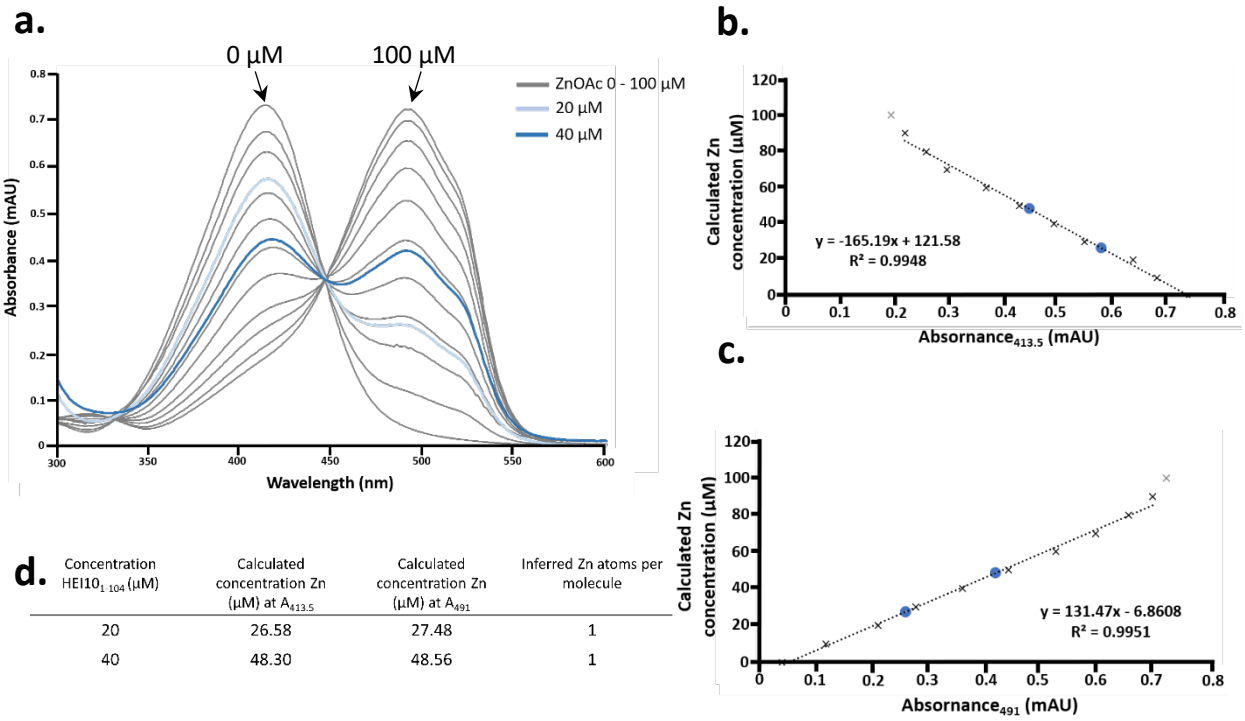


Figure 5.2.26.| Purification and biophysical analysis of HEI10$_{1-104}$
a) SDS-PAGE showing the purification summary of HEI10$_{1-104}$ through sequential amylose affinity and anion exchange chromatography. N-terminal MBP-tag was removed by incubation with TEV protease followed by anion exchange and subsequent size exclusion chromatography. b) SEC-MALS analysis of HEI10$_{1-104}$ revealed two peaks with molecular weights of 46.7 kDa (theoretical tetramer – 46.8 kDa) and 32.8 kDa. c) Far UV CD wavelength scan between 260-185nm of HEI10$_{1-104}$ shows an α-helical trace. Deconvolution of the data estimates the secondary structure to be 32 % α-helical. d) CD thermal denaturation measured at 222nm between 4 and 95°C, estimated a melting temperature of 62.6 °C.

A PAR assay of HEI10$_{1-104}$ was used to determine the zinc atom content. The absorbance spectra of zinc standards (0-100 µM) and HEI10$_{1-104}$ at two concentrations (20 and 40 µM) were determined. The protein traces match zinc standards 20 and 40 µM, respectively (Figure 5.2.27.a.). Using the linear plots, the zinc content of HEI10$_{1-104}$ can be determined at 27.03 and 44.30 µM, respectively (Figure 5.2.27.b.-

d.). This result suggests an inferred zinc content of HEI10$_{1\text{-}196}$ and HEI10$_{1\text{-}122}$ of 1 zinc atom per HEI10 molecule.



d.

Concentration HEI10$_{1\text{-}104}$ (µM)	Calculated concentration Zn (µM) at $A_{413.5}$	Calculated concentration Zn (µM) at A_{491}	Inferred Zn atoms per molecule
20	26.58	27.48	1
40	48.30	48.56	1

Figure 5.2.27.| HEI10$_{1\text{-}104}$ zinc content determination by PAR assay.
a) Absorbance spectrum between 300-600 nm of zinc standards (0-100 µM - grey. The absorbance spectrum of HEI10$_{1\text{-}104}$ was determined at two concentrations: 20 µM (light blue) and 40 µM (dark blue). b-c) Experimental spectrum at absorbance 413.5 (a) and 491 (b) were fitted as a linear standard, with R^2 values of 0.9948 and 0.9951, respectively. Zinc standards with absorptivity outside of the linear range are shown in grey and have been excluded from calculating the linear equation. HEI10$_{1\text{-}104}$ samples are plotted (blue dots). d) Concentration of Zn can be calculated in the protein samples from the linear plot and inferred a single Zn atom per HEI10 molecule, suggesting four Zn atoms per HEI10 tetramer. (n=1).

SEC-SAXS analysis of HEI10$_{1\text{-}104}$ also revealed two overlapping peaks, consistent with the SEC-MALS data. The SAXS data corresponding to the first, smaller peak (pre-peak) contains noise but the experimental data was still interpretable and revealed a *Rg* and *Dmax* of 29.07 and 98 Å, respectively (Figure 5.2.28.b and c.). The dominant HEI10$_{1\text{-}104}$ peak (main peak) has a Guinier analysis *Rg* of 23.29 Å, which matches the real-space *Rg* of 23.5 Å. The *Dmax* of the main peak was determined to be 88 Å. The maximum dimension of the pre-peak and main peak of HEI10$_{1\text{-}104}$ are of similar values suggesting that despite a disrupted tetramer, the length of the protein does not change. Moreover, the scattering

data was utilised to derive the volume of the molecular mass of $HEI10_{1-104}$ (Figure 5.2.28.c.). Inspection of the Kratky and Porod plots of the two $HEI10_{1-104}$ peak reveals a particle with a defined Porod-Debye plateau and an associated volume of 118280 (pre-peak) and 68324 $Å^3$ (main peak), consistent with a 69.5 and 40 kDa protein, respectively. This result suggests that the main peak of $HEI10_{1-104}$ is no longer a tetramer, and possibly could be a dimer, indicating that the coiled-coil region of HEI10 is crucial for tetramerisation and is weakened in the $HEI10_{1-104}$ construct. However, mass determination by SAXS is very inaccurate, so are interpretation of the data is only suggestive.

The SEC-SAXS scattering data was ultilised to create a low-resolution *ab initio* model of $HEI10_{1-104}$. The resultant modeling demonstrates a more compact molecular envelope for $HEI10_{1-104}$ in comparison to longer $HEI10_{1-196}$ construct (Figure 5.2.28.d.).

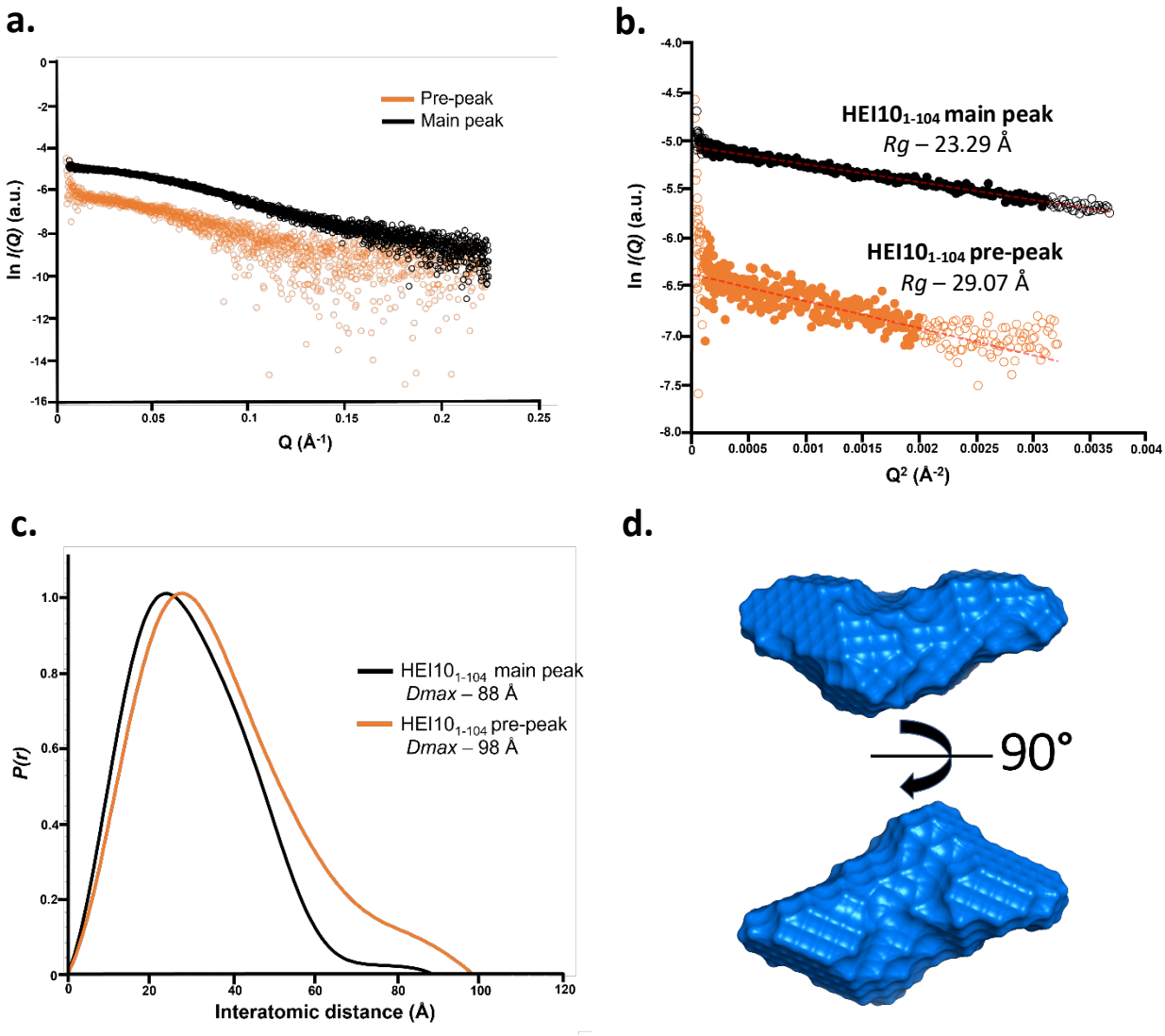


Figure 5.2.28.| SEC-SAXS analysis of HEI10$_{1\text{-}104}$.
a) SAXS scattering profiles of HEI10$_{1\text{-}104}$ pre-peak (orange) and main peak (black). b) Guinier analysis determined a radius of gyration (*Rg*) value of 29.07 (orange) and 23.29 Å (black), respectively. The real space *Rg* was determined as 28.9 and 23.5 Å, respectively, which closely match the Guinier analysis *Rg* value. Empty circles represent the complete dataset and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3). c) *P(r)* distribution of HEI10$_{1\text{-}104}$ pre-peak (orange) and main peak (black) showing a maximum dimension (*Dmax*) of 98 and 88 Å. d) SAXS *ab initio* DAMMIF model of HEI10$_{1\text{-}104}$ (main peak) presented as a molecular envelope.

5.2.14. Purification and characterisation of RNF212 and RNF212b RING domains

In addition to HEI10, a second E3 ligase family protein, RNF212, has been shown to be essential in crossing over in mice (Qiao et al., 2014; Reynolds et al., 2013). Moreover, a RNF212 paralogue, RNF212b, has also been identified in mammals (Johnston et al., 2020). Both RNF212 and RNF212b share predicted structural homology to the E3 ligase family of proteins; N-terminal zinc-binding RING domain and a central coiled-coil domain, schematised in Figure 5.2.3.

Both localisation and functional studies of RNF212 have been carried out and have shown that RNF212 foci mark crossover sites during mid-pachytene of prophase I, and co-localises with HEI10 (Qiao et al., 2014). Despite these studies, RNF212 has not previously been characterised *in vitro* and no direct interactions have been determined between RNF212 and other meiotic proteins. A biophysical study of RNF212 and RNF212b, in isolation and possible complexes may permit the structural characterisation and provide an understanding for function of RNF212 and RNF212b within meiotic recombination. Initially we set out to investigate the RING and coiled-coil domains of RNF212 and RNF212b in isolation. Preliminary data for RNF212 and RNF212 was collected by Carmen Espejo Serrano, but I processed the data and made the figures.

Construct boundaries for both proteins were designed by sequence conservation and secondary structure predictions. The N-terminal RING of RNF212 and RNF212b (amino acid residues 1-61 and 1-56, respectively) were expressed with in a N-terminal MBP-solubility tag in *E.coli* encoded by the pMAT11 vector and grown following the same protocol as HEI10. MBP-RNF212$_{1-61}$ and MBP-RNF212b$_{1-56}$ were purified through sequential amylose affinity and anion exchange chromatography. Solubility tags were cleaved by TEV protease and further purified by anion exchange and SEC. Final purified material (Figure 5.2.29.a. and b.) was analysed by SEC-MALS to determine the absolute molecular weight. Both RNF212$_{1-61}$ and RNF212b$_{1-56}$ eluted as single peaks with determined molecular weights of 8.6 and 11.4 kDa, corresponding to a monomer and dimer respectively (theoretical monomer – 8.5 and 6.4 kDa, respectively) (Figure 5.2.29.c.).

.

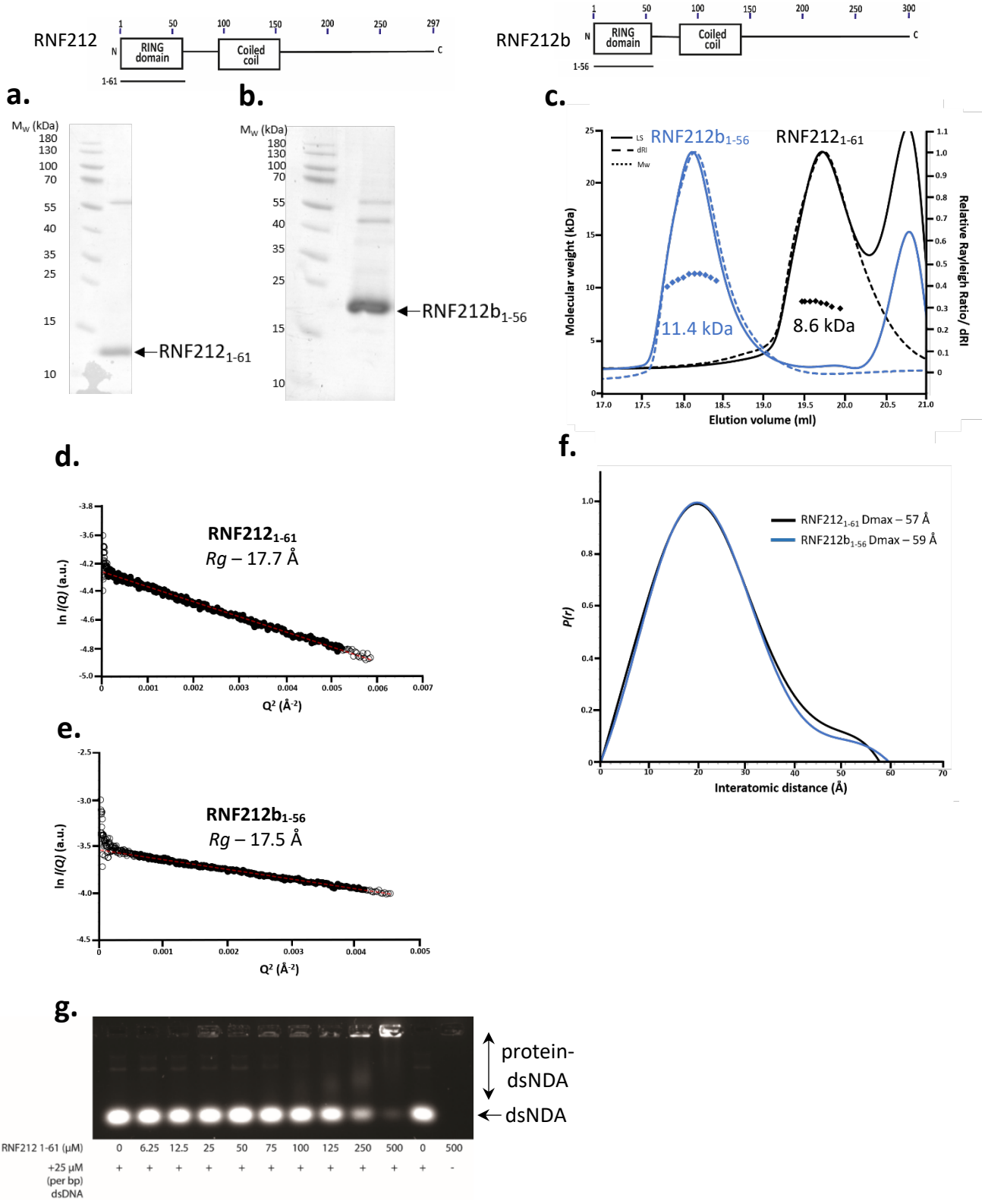
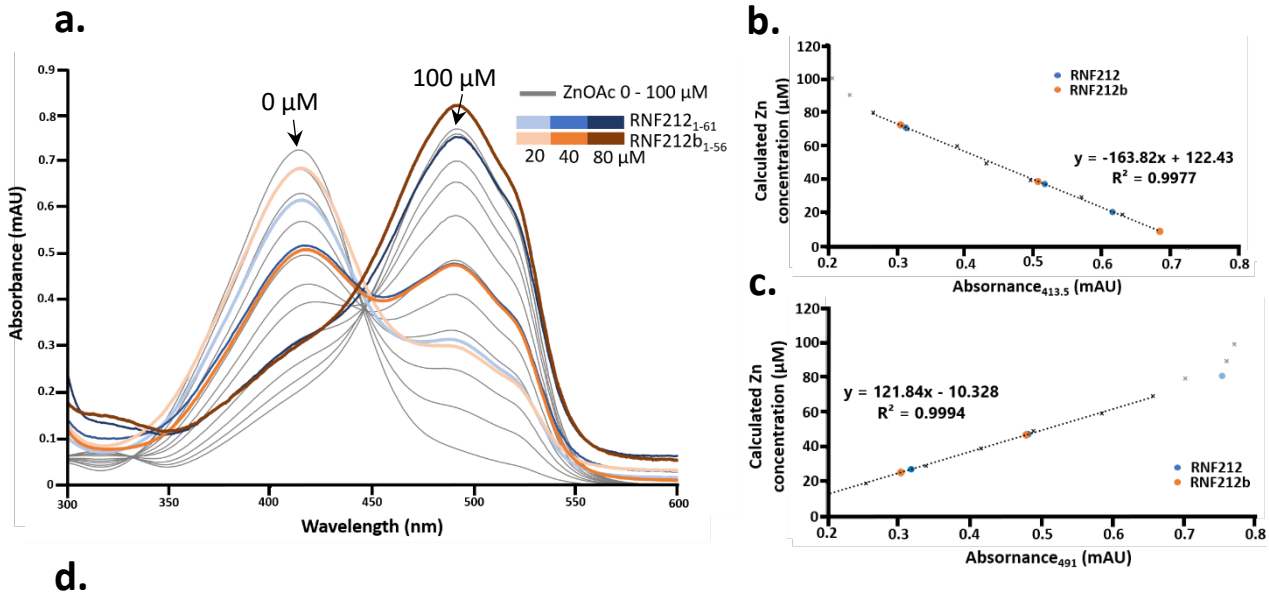

RNF212
RING domain
Coiled coil
1-61
RNF212b
1-56
a.
b.
c.
M_W (kDa)
RNF212$_{1-61}$
RNF212b$_{1-56}$
LS
dRI
Mw
11.4 kDa
8.6 kDa
Molecular weight (kDa)
Relative Rayleigh Ratio/ dRI
Elution volume (ml)
d.
RNF212$_{1-61}$
Rg – 17.7 Å
ln I(Q) (a.u.)
Q² (Å⁻²)
e.
RNF212b$_{1-56}$
Rg – 17.5 Å
f.
RNF212$_{1-61}$ Dmax – 57 Å
RNF212b$_{1-56}$ Dmax – 59 Å
P(r)
Interatomic distance (Å)
g.
protein-dsNDA
dsNDA
RNF212 1-61 (μM)
0 6.25 12.5 25 50 75 100 125 250 500 0 500
+25 μM (per bp) dsDNA
+ + + + + + + + + + + -

Figure 5.2.29.| Biophysical analysis of RNF212$_{1-61}$ and RNF212b$_{1-56}$
a-b) SDS-PAGE showing the final purified of a) RNF212$_{1-61}$ and b) RNF212b$_{1-56}$. c) Overlaid SEC-MALS analysis of RNF212$_{1-61}$ (black) and RNF212b$_{1-56}$ (blue) determined a main peak with molecular weight of 8.6 and 11.4 kDa, respectively revealing that the RING domain of RNF212 and RNF212b form a monomer and dimer when analysed in isolation, respectively (theoretical monomer – 8.5 and 6.4 kDa). d-f) SEC-SAXS analysis of RNF212$_{1-61}$ and RNF212b$_{1-56}$. Guinier analysis determined a radius of gyration (*Rg*) value of 17.7 and 17.8 Å, respectively. The real space *Rg* determined at 17.8 Å for both constructs closely matches the Guinier analysis *Rg*. Empty circles represent the complete data set and the solid circles represent the Guinier region, data used for determination of the *Rg*. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3). e) Overlaid *P(r)* distribution of RNF212$_{1-61}$ (black) and RNF212$_{1-56}$ (blue) determine a *Dmax* of 57 and 59 Å, respectively. e) EMSA demonstrate that the RING domain of RNF212$_{1-61}$ can interact with 25 μM (per base pair) linear double-stranded DNA (dsDNA).

SEC-SAXS analysis was performed to determine low resolution structural information of RNF212$_{1-61}$ and RNF212b$_{1-56}$. The radius of gyration (*Rg*) provides a measure of the overall size of the molecule and can be directly obtained from the scattering data or estimated using the Guinier approximation. RNF212$_{1-61}$ and RNF212b$_{1-56}$ have a Guinier analysis determined *Rg* of 17.7 and 17.8 Å, respectively, which closely matches the real-space *Rg* (17.8 and 17.6 Å, respectively) (Figure 5.2.29.d. and e.). The *p(r)* curve of both RNF212$_{1-61}$ and RNF212b$_{1-56}$ reveal a profile typical of globular proteins with a smooth converged tail. The maximum interatomic distance (*Dmax*) of RNF212$_{1-61}$ and RNF212b$_{1-56}$ were determined to be 57 and 59 Å, respectively (Figure 5.2.26.c.). Despite the SEC-MALS revealing that RNF212$_{1-61}$ and RNF212b$_{1-56}$ form a monomer and dimer, respectively, in solution, the SAXS data is very similar, suggesting that the RING domains may stack on top of each other.

We have previously shown through EMSA that the structural core of HEI10 binds linear dsDNA (Figure 5.2.10.). Using the same technique, we analysed the ability of the RNF212 RING domain in isolation (theoretical pI of 8.66) to bind dsDNA. 25 μM (per bp) of linear dsDNA was incubated with increasing amounts of RNF212$_{1-61}$ (0-500 μM). Incubated samples were then run on an agarose gel, where shifts in the bands were observed which clearly indicated an affinity for dsDNA (Figure 5.2.29.e.). A DNA shift was observed at 125 μM of RNF212. This observation suggests that the meiotic recombination proteins can bind to dsDNA through their RING domain.

A PAR assay was used to determine the zinc atom content of digested RNF212 and RNF212b RING domains. The absorbance spectra of zinc standards (0-100 µM) and $RNF212_{1-61}$/$RNF212b_{1-56}$ at three concentrations (20, 40 and 80 µM) were determined. The protein traces closely match zinc standards 20, 40 and 80 µM, respectively (Figure 5.2.30.a.). Using the linear plots, the zinc content of $RNF212_{1-61}$ was determined to be 24.77, 42.97 and 76.35 µM, respectively and $RNF212b_{1-56}$ to be 26.29, 47.8 and 89.77 µM (Figure 5.2.30.b.-d.). This result matches the inferred zinc content of HEI10 of 1 zinc atom per molecule.



d.

Concentration $RNF212_{1-61}$ (µM)	Calculated concentration Zn (µM) at $A_{413.5}$	Calculated concentration Zn (µM) at A_{491}	Inferred Zn atoms per molecule
20	21.49	28.05	1
40	37.79	48.15	1
80	71.22	81.47	1

Concentration $RNF212b_{1-56}$ (µM)	Calculated concentration Zn (µM) at $A_{413.5}$	Calculated concentration Zn (µM) at A_{491}	Inferred Zn atoms per molecule
20	10.14	26.29	1
40	39.36	47.80	1
80	72.49	89.77	1

Figure 5.2.30.| Zinc content determination of $RNF212_{1-61}$ and $RNF212b_{1-56}$ by PAR assay.
a) Absorbance spectrum between 300-600 nm of zinc standards (0-100 µM - grey). The absorbance spectrum of $RNF212_{1-61}$ and $RNF212b_{1-56}$ were determined at three concentrations: 20, 40 and 80 µM. b-c) Experimental spectrum at absorbance 413.5 (a) and 491 (b) were fitted as a linear standard, with R^2 values of 0.9977 and 0.9994, respectively. Zinc standards with absorptivity outside of the linear range are shown in grey and have been excluded from calculating the linear equation. RNF212 and RNF212b samples are plotted (blue and orange circles, respectively). d) Concentration of Zn can be calculated in the protein samples from the linear plot and inferred a single Zn atom per RNF212 or RNF212b molecule. (n=1).

5.2.15. RNF212$_{1-198}$ is not stable when expressed in isolation

The RING and coiled-coil domain of RNF212, amino acid residues 1-198, was cloned as a MBP-fusion in *E.coli* BL21 (DE3) cells. Expression of MBP-RNF212$_{1-198}$ yielded soluble material but around 50% of the purified material was degraded (Figure 5.2.31.a.). The amylose eluate was further purified by anion exchange chromatography to remove any free-MBP (Figure 5.2.31.b.). Fusion material was analysed by SEC-MALS, which determined that RNF212 elutes as two overlapping peaks. The first peak forms large megadalton species and the second peak has a determined molecular weight of 281 kDa, suggesting a mixture of soluble aggregate and tetramer (theoretical tetramer – 268.8 kDa) (Figure 5.2.31.c.). From this result we hypothesise that RNF212 requires a binding partner to induce stability and folding, likely to interact through the α-helical predicted coiled-coil domain.

5.2.16. RNF212b$_{1-136}$ structural core tetramerises in solution

We set out to characterise the coiled-coil domain of RNF212b in isolation and in combination with the N-terminal RING domain, to investigate the mechanistic assembly of RNF212b. We have previously shown by SEC-MALS that the RING domain of RNF212b forms a dimer in solution. Construct boundaries were determined using the secondary structure α-helical prediction for the coiled-coil region. RNF212b$_{65-136}$ (coiled-coil) and RNF212b$_{1-136}$ (structural core) were cloned with an N-terminal His$_6$-MBP-tag encoded by the pMAT11 vector in *E .coli*. MBP-RNF212b$_{65-136}$ and MBP-RNF212b$_{1-136}$ were purified by sequential amylose affinity and anion exchange chromatography, final fusion material is shown in Figure 5.2.32.a. and b. Fusion material was analysed by SEC-MALS. RNF212b$_{65-136}$ and RNF212b$_{1-136}$ fusion constructs eluted as single peaks, with molecular weights of 124 and 246 kDa, respectively (Figure 5.2.32.c.). These results correspond to the coil-coil domain forming a dimer in isolation (theoretical dimer - 106.6 kDa), and together with the RING domain forms a tetramer (theoretical tetramer – 248 kDa). This suggests that the structural core of RNF212b requires both the RING and coiled-coil domain for tetramer formation. Moreover, these results are complementary to the observed tetramer formation of HEI10$_{1-196}$.

We set out to analyse cleaved material of RNF212b$_{65-136}$ and RNF212b$_{1-136}$, however once subjected to TEV-cleavage both constructs were highly degraded. This observation suggests that the MBP-tag was providing stability and without it the constructs are highly unstable. Furthermore, this may suggest that RNF212b requires a binding partner to facilitate stability.

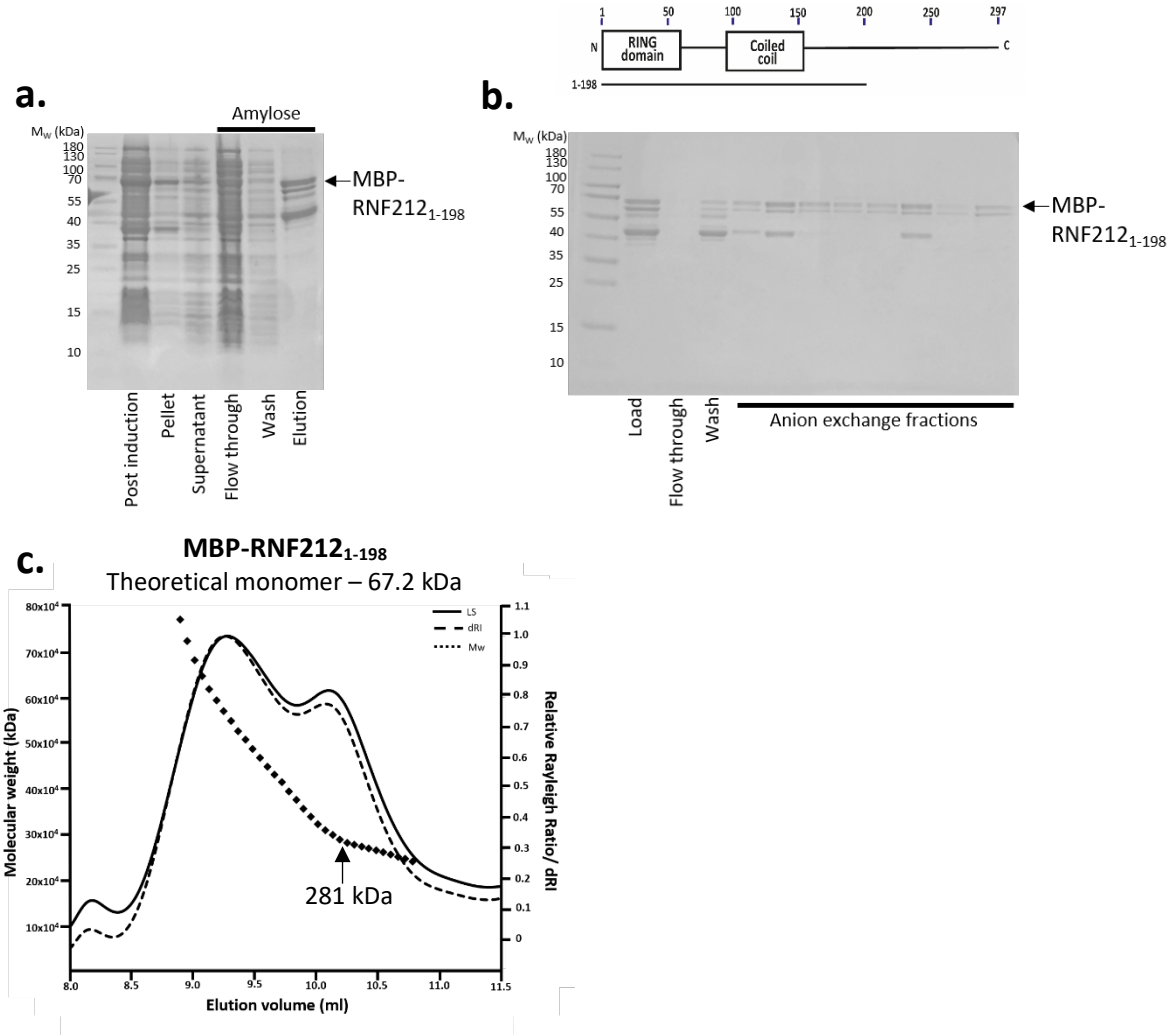


Figure 5.2.31.| Purification and SEC-MALS analysis of RNF212$_{1-198}$
a-b) SDS-PAGE analysis showing the purification steps of RNF212 structural core (RNF212$_{1-198}$) expressed as an N-terminal MBP fusion. a) The limited amount of soluble material of MBP-RNF212$_{1-198}$ was purified by amylose affinity chromatography. MBP- RNF212$_{1-198}$ highly degrades down to free-MBP, shown by a predominant band at 44 kDa in the amylose eluate. b) MBP-RNF212$_{1-198}$ was further purified by anion exchange chromatography to remove degradation products. c) SEC-MALS analysis revealed that MBP-RNF212$_{1-198}$ elutes across a broad range of two overlapping peaks with the first peak in the void volume and the second peak has a molecular weight of 281 kDa, suggesting that the folded material forms a tetramer in solution (theoretical tetramer – 268.8 kDa).

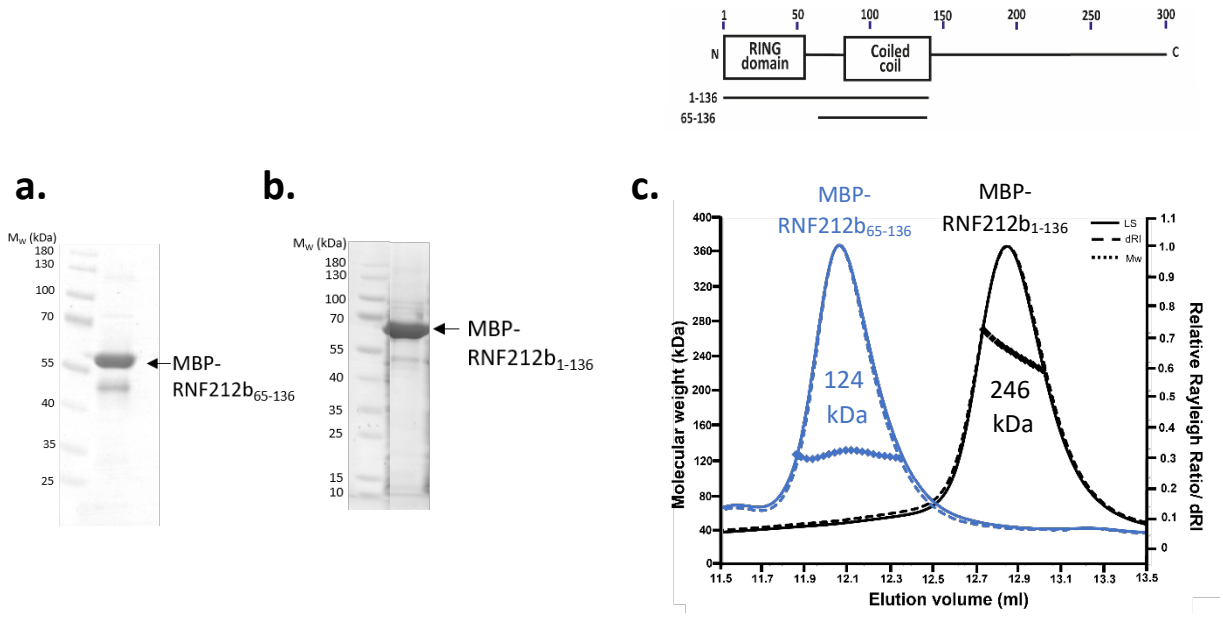


Figure 5.2.32.| SEC-MALS analysis of MBP-RNF212b
a-b) SDS-PAGE showing the final product of a) MBP-RNF212b$_{65-136}$ and b) MBP-RNF212b$_{1-136}$, purified through sequential amylose affinity and anion exchange chromatography. c) Overlaid SEC-MALS analysis of MBP-RNF212b$_{65-136}$ (blue) and MBP-RNF212b$_{1-136}$ (black) revealed a single peak with molecular weight of 124 kDa and 246 kDa, inferring that the coiled-coil domain of RNF212b forms a dimer in isolation (theoretical dimer – 106.6 kDa) and a tetramer when combined with the RING domain (theoretical tetramer – 248 kDa).

5.2.17. Testing possible interactions between the recombination proteins: HEI10, RNF212 and RNF212b

Another consideration for the poor expression and insolubility of HEI10 could be that a binding partner is required for stabilisation. Studies have shown *in vivo* that HEI10 and RNF212 colocalise during mid-pachytene at the SC (Qiao et al., 2014). We tested the possibility of an interaction between the recombination proteins structural cores by amylose affinity chromatography.

Amylose pulldowns were performed with HEI10$_{1-196}$ fused to a MBP-affinity tag and co-expressed with His$_6$-tag fused RNF212$_{1-198}$ or RNF212b$_{1-136}$ in *E.coli* BL21 (DE3) cells. The bacterial lysate was applied to amylose resin, incubated, and after washing, the bound protein(s) were eluted. Both HEI10 and RNF212 expressed, as bands of their respective molecular weights are seen in both the bacterial pellet and supernatant. His$_6$-RNF212$_{1-198}$ is present in both the flow through and wash steps but the eluate. The amylose eluate contains only MBP-HEI10$_{1-196,}$ indicating that there is no direct interaction between

HEI10 and RNF212 (Figure 5.2.33.a.). SDS-PAGE showing the amylose profile of MBP-HEI10$_{1-196}$ co-expressed with His$_6$-RNF212b$_{1-136}$ shows that His$_6$-RNF212b is insoluble and only MBP-HEI10 is in the amylose eluate, also suggesting no interaction (Figure 5.2.33.b.).

RNF212$_{1-198}$ and RNF212b$_{1-136}$ were fused to a MBP- and His$_6$-affinity tag, respectively, and co-expressed in *E. coli* BL21 (DE3) cells. RNF212 and RNF212b were co-purified by amylose affinity chromatography. SDS-PAGE analysis revealed that both proteins are present in the amylose eluate indicating that the fusion proteins interact *in vitro* (Figure 5.2.33.c). The fusion protein complex was further purified by anion exchange chromatography, where the two proteins co-eluted in a single peak across a salt gradient (Figure 5.2.33.d.).

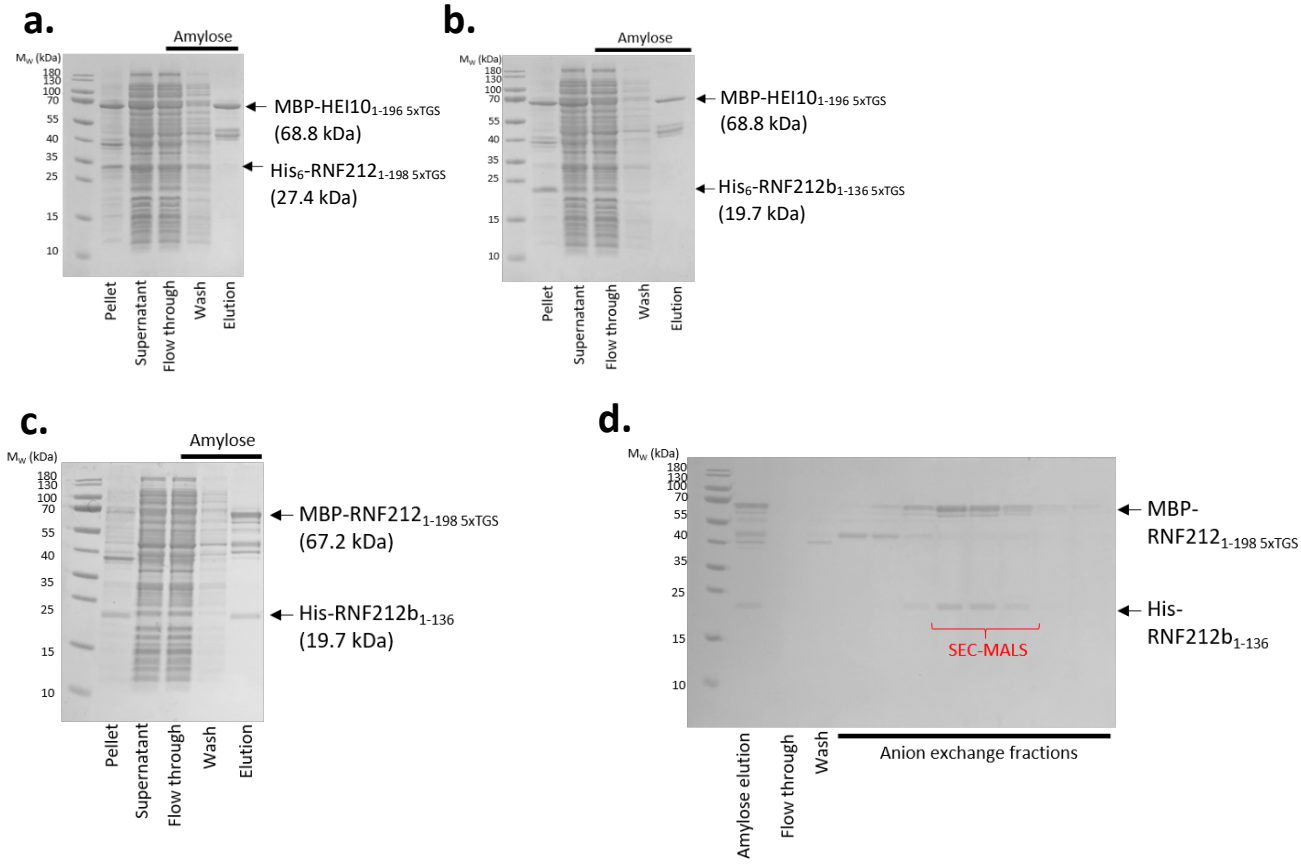


Figure 5.2.33.| Screening for interactions between HEI10, RNF212 and RNF212b
a-b) SDS-PAGE analysis of MBP-HEI10$_{1-196}$ co-expressed with a) His-RNF212$_{1-198}$ and b) His-RNF212b$_{1-136}$. Determined that MBP-HEI10$_{1-196}$ does not interact with either His-RNF212$_{1-198}$ or His-RNF12b$_{1-136}$ as only MBP-HEI10$_{1-196}$ is seen in the amylose eluate. c) SDS-PAGE analysis of the co-expression of MBP-RNF212$_{1-198}$ and His-RNF212b$_{1-136}$. Both proteins are present in the amylose eluate, indicating an interaction between RNF212 and RNF212b. d) SDS-PAGE showing co-elution of MBP-RNF212$_{1-198}$ and His-RNF212b$_{1-136}$ by anion exchange chromatography.

5.2.18. RNF212:RNF212b forms a stable 2:2 complex in solution

SEC-MALS analysis was used to determine the absolute molecular weight of the RNF212:RNF212b fusion complex. In comparison to the broad peak of MBP-RNF212 analysed in isolation, the fusion complex proteins eluted much later as a single symmetrical peak (Figure 5.2.34.a. and b.). SDS-PAGE analysis of the MALS elution profile reveals that MBP-RNF212$_{1-198}$ and His-RNF212b$_{1-136}$ perfectly co-elute, indicating that the interaction between the two proteins is robust (Figure 5.2.34.c.). An experimental molecular mass of 170.4 kDa corresponds to a 2:2 complex (theoretical 2:2 complex – 173.8 kDa).

Through sequence analysis and secondary structure prediction we designed truncated constructs of RNF212, to closely match the RNF212b$_{1-136}$ construct. MBP-RNF212 constructs (amino acid residues: 1-162, 1-154 and 1-123) were co-expressed with His$_6$-RNF212b$_{1-136}$ and co-purified through sequential amylose affinity and anion exchange chromatography. Also, His$_6$-RNF212$_{1-129}$ was co-expressed and co-purified with MBP-RNF212b$_{1-136}$. SDS-PAGE analysis of the anion exchange elution profile determined that co-elution is maintained between the truncated MBP-RNF212$_{1-123}$ and His$_6$-RNF212b$_{1-36}$ (Figure 5.2.35.a-c.). This was also the case for MBP-RNF212b$_{1-136}$:His$_6$-RNF212$_{1-129}$ (Figure 5.2.35.d.).

SEC-MALS was utilised to determine the oligomeric state of MBP-RNF212b$_{1-136}$:His$_6$-RNF212$_{1-129}$ to examine whether the truncated RNF212 constructs alter the oligomeric state. All protein complexes eluted as single peaks, with molecular weights corresponding to a 2:2 complex, suggesting that amino acid residues 124-196 of RNF212 are not required for interaction with RNF212b (Figure 5.2.35.e. and f.).

In order to analyse cleaved material TEV-protease enzyme was added to the RNF212:RNF212b complex. Despite testing different cleavage conditions, the N-terminal tag of RNF212b$_{1-136}$ remained fused (Figure 5.2.36.a. and b.). In other cases, we have added a long flexible TGS-linker to aid cleavage, however this may entail downstream consequences in crystallisation. Instead, RNF212b$_{1-136}$ was recloned without a N-terminal tag and co-purified with MBP-RNF212 following the same protocol, further indicating the strong interaction between the two proteins. MBP-RNF212:RNF212b$_{1-136}$ was

subjected to TEV-cleavage, SDS-PAGE analysis revealed that RNF212 successfully cleaved (Figure 5.2.36.c.)

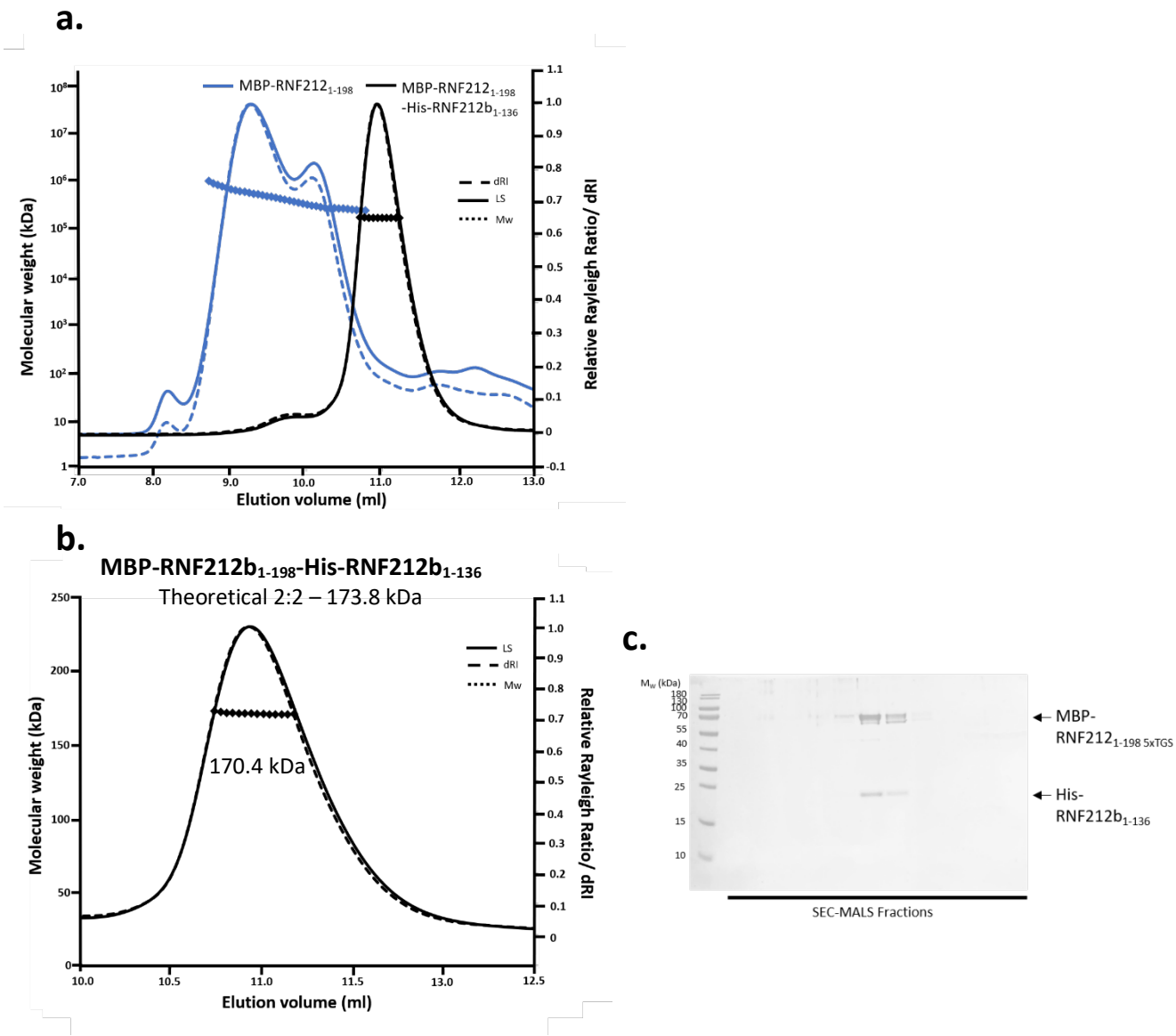
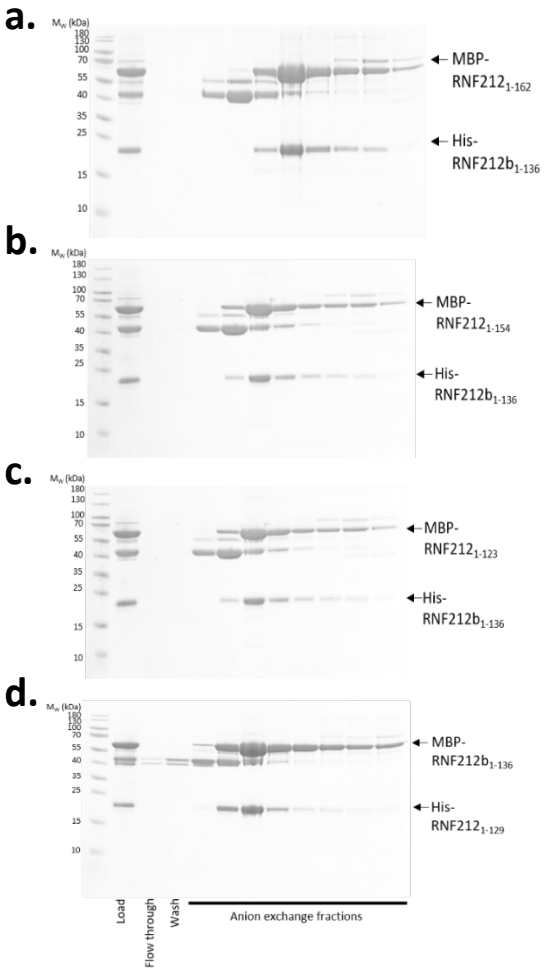
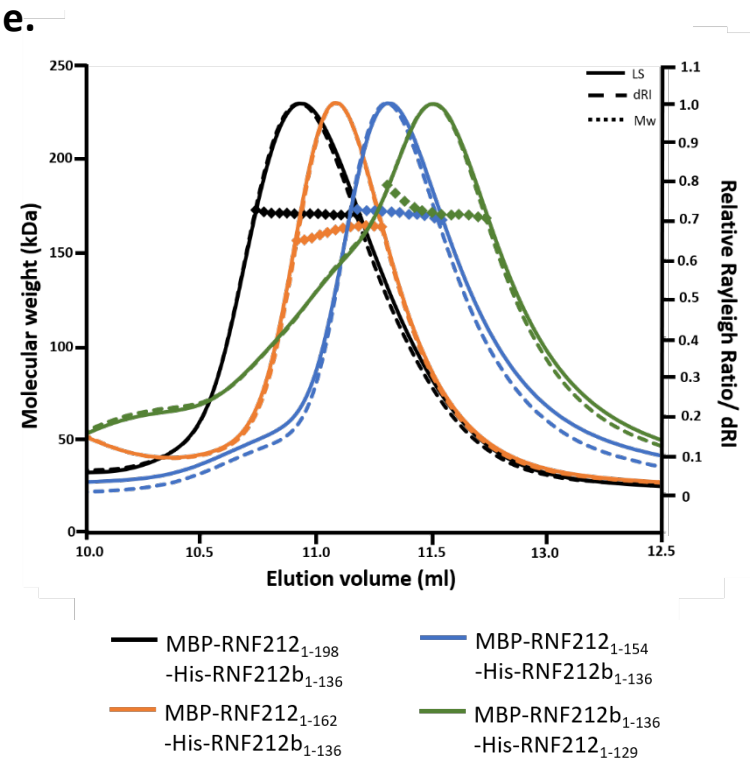


Figure 5.2.34.| SEC-MALS analysis of MBP-RNF212$_{1-198}$:His-RNF212b$_{1-136}$
a) Overlaid SEC-MALS analysis of MBP-RNF212$_{1-198}$ (blue) in isolation and MBP-RNF212$_{1-198}$ complexed with His-RNF212b$_{1-136}$ (black). MBP-RNF212$_{1-198}$ elutes across a broad range of two overlapping peaks with the first peak in the void volume (548 kDa). The second peak has a molecular weight of 281 kDa, suggesting that the folded material forms a tetramer in solution (theoretical tetramer – 268.8 kDa). MBP-RNF212$_{1-198}$:His-RNF212b$_{1-136}$ fusion complex elutes later as a single peak. b) SEC-MALS analysis of MBP-RNF212$_{1-198}$:His-RNF212b$_{1-136}$ determined molecular weight of 170.4 kDa correlating to a 2:2 complex (theoretical 2:2 – 173.8 kDa). c) SDS-PAGE of the MALS profile shows that MBP-RNF212$_{1-198}$ coelutes with His-RNF212b$_{1-136}$.





f.

RNF212-RNF212b construct	Theoretical 2:2 (kDa)	Experimental Mw (kDa)
MBP-RNF212$_{1-198}$ – His-RNF212b$_{1-136}$	173.8	170.4
MBP-RNF212$_{1-162}$ – His-RNF212b$_{1-136}$	166.4	161.5
MBP-RNF212$_{1-154}$ – His-RNF212b$_{1-136}$	164.5	170.7
MBP-RNF212b$_{1-136}$ – His-RNF212$_{1-129}$	159.4	165

Figure 5.2.35.| Optimisation of RNF212 with in the fusion RNF212-RNF212b complex
a-c) SDS-PAGE analysis of anion exchange chromatography elution profiles of His$_6$-RNF212b$_{1-136}$ co-expressed with MBP-RNF212 constructs. d) SDS-PAGE analysis of the anion exchange chromatography elution profile of MBP-RNF212B$_{1-136}$:His$_6$-RNF212$_{1-129}$. The C-terminal truncation of RNF212, RNF212$_{1-129}$, remains complexed with RNF212b$_{1-136}$. e) Overlaid SEC-MALS analysis His-RNF212b$_{1-136}$ complexed with MBP-RNF212 constructs and MBP-RNF212b$_{1-136}$-His-RNF212$_{1-129}$. f) SEC-MALS experimental molecular weights of RNF212:RNF212b complexes.

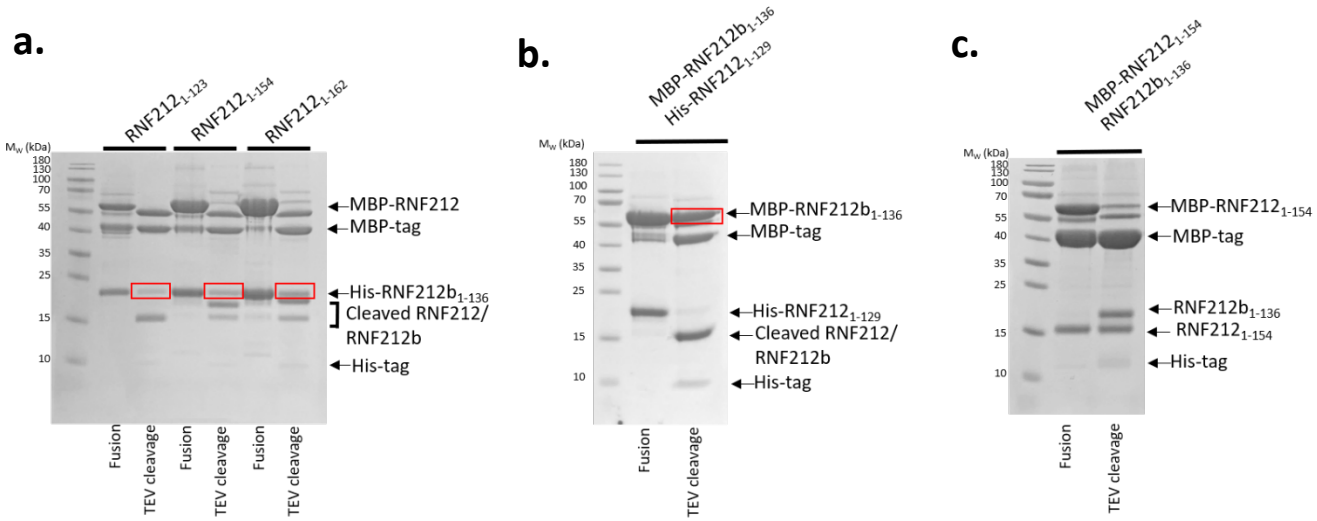


Figure 5.2.36.| RNF212b cleavage by TEV-protease needs to be improved.
SDS page analysis of RNF212-RNF212b TEV cleavage. a) His_6-RNF212b$_{1-136}$:MBP-RNF212 complexes subjected to TEV-protease cleavage. All three protein complexes show successful cleavage of MBP-RNF212 but around 50% of His_6-RNF212b$_{1-136}$ remains uncleaved. b) TEV-cleavage of MBP-RNF212b$_{1-136}$:His_6-RNF212$_{1-129}$ also determined that RNF212 is cleaved. Uncleaved RNF212b material is highlighted by red boxes. c) TEV-cleavage of MBP-RNF212$_{1-154}$ co-purified with untagged RNF212b$_{1-136}$.

5.2.19. RNF212 and RNF212b undergo rearrangement upon complex formation

We further analysed the shortest construct of RNF212, RNF212$_{1-123}$, co-purified with RNF212b$_{1-136}$. The affinity tag from MBP-RNF212$_{1-123}$ was removed by TEV protease and further purified by anion exchange chromatography and SEC, summarised in Figure 5.2.36.a. SEC-MALS analysis of RNF212$_{1-123}$:RNF212b$_{1-136}$ revealed a single peak with a molecular weight of 61.2 kDa, demonstrating that in solution RNF212 and RNF212b undergo structural rearrangement from individual tetramers to a 2:2 complex (theoretical 2:2 complex – 60.6 kDa) (Figure 5.2.37.b.). The SEC-MALS elution peak shows some deviance of the light scattering and dRI in the second half of the peak and slight sloping of the MW fit. These observations suggest a possible dissociation between the two proteins. Furthermore, due to both RNF212 and RNF212b being of similar MW, SDS-PAGE analysis of the SEC-MALS elution fractions does not help to determine co-elution (Figure 5.2.37.c.). SDS-PAGE analysis of the longer RNF212$_{1-154}$:RNF212b$_{1-136}$ SEC-MALS elution profiles reveals perfect co-elution (Figure 5.2.37.d.).

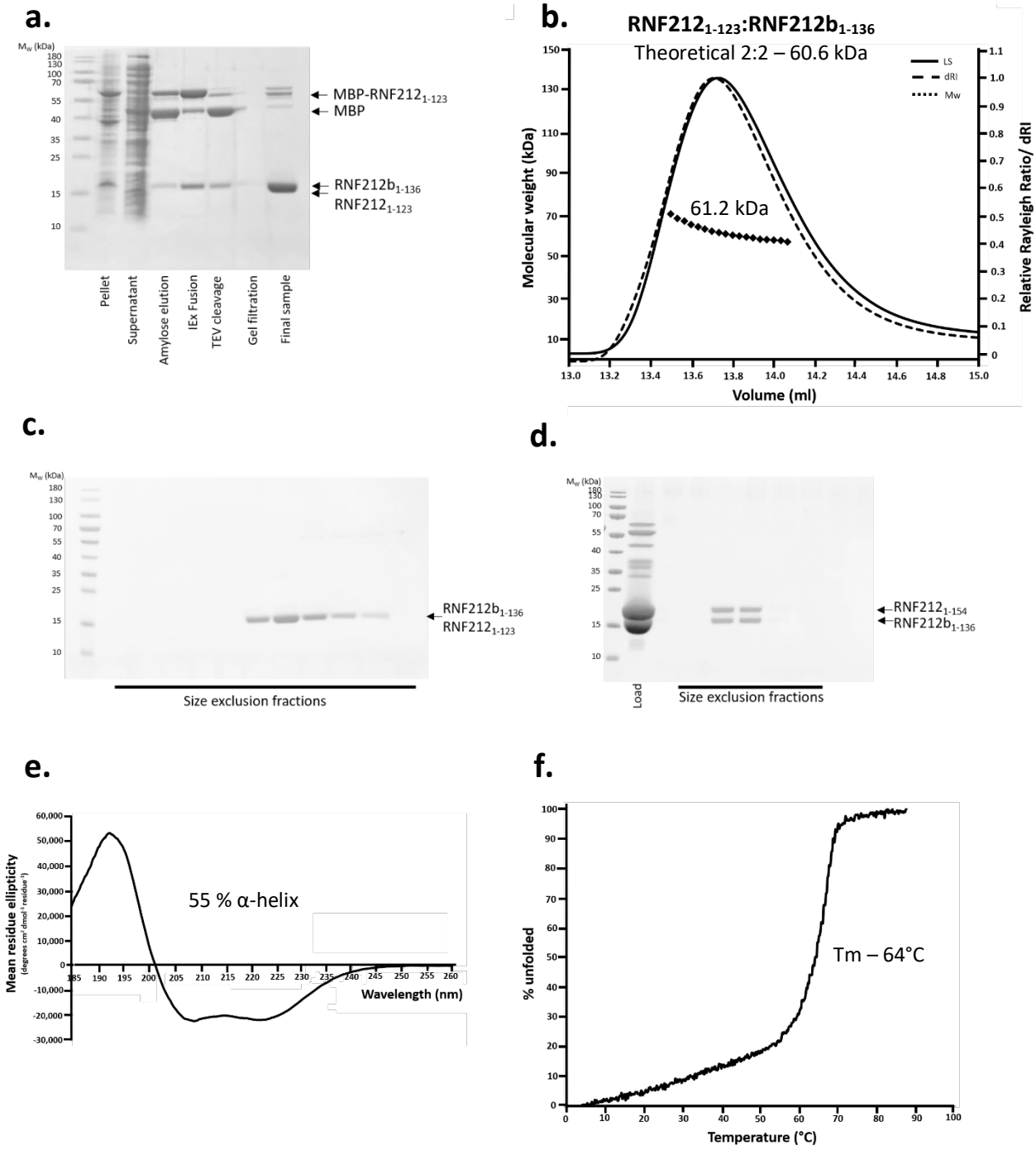
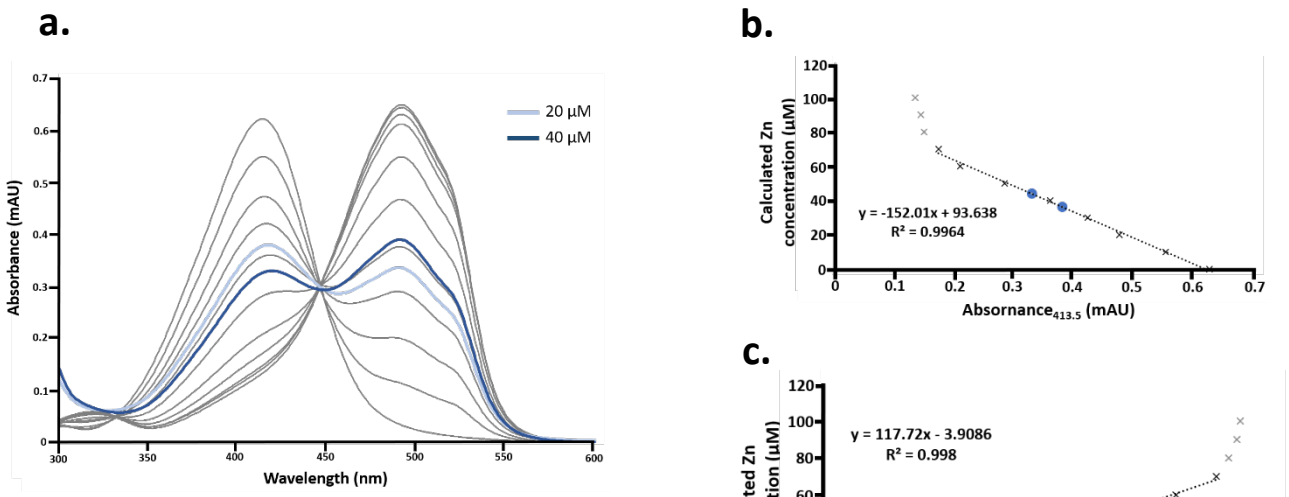


Figure 5.2.37.| Purification and biophysical analysis of RNF212$_{1-123}$:RNF212b$_{1-136}$
a) SDS-PAGE showing the purification summary of RNF212$_{1-123}$:RNF212b$_{1-136}$ through sequential amylose affinity and anion exchange chromatography. N-terminal MBP-tag of RNF212$_{1-123}$ was removed by incubation with TEV protease followed by anion exchange and subsequent SEC. b) SEC-MALS analysis of RNF212$_{1-123}$:RNF212b$_{1-136}$ revealed a single peak with an experimental molecular weight of 61.2 kDa (theoretical tetramer – 60.6 kDa) indicating the formation of a 2:2 complex in solution. c-d) SDS-PAGE showing the SEC elution profile of c) RNF212$_{1-154}$-RNF212b$_{1-136}$ and d) RNF212$_{1-123}$:RNF212b$_{1-136}$, indicates that the two proteins perfectly co-elute. e) Far UV CD wavelength scan between 260-185nm of RNF212$_{1-123}$:RNF212b$_{1-136}$ shows a typical α-helical trace. Deconvolution of the data estimates the secondary structure to be 55 % α-helical. f) CD thermal denaturation measured at 222nm between 4 and 95°C, estimated a melting temperature of 64 °C.

Far-UV CD spectroscopy was used to quantify the secondary structure of $RNF212_{1-123}$:$RNF212b_{1-136}$. A characteristic α-helical spectrum, with negative peaks at 208 and 222 nm, allowed for deconvolution by DichroWeb to determine an α-helical content of 55% (Figure 5.2.37.e.). Thermal denaturation of $RNF212_{1-123}$:$RNF212b_{1-136}$ determined a T_m of 64 °C, which suggests a co-operative unfolding event (Figure 5.2.37.f.).

PAR assay was carried out to determine the zinc atom content and to confirm that RNF212:RNF212b is folded correctly. The absorbance spectra of $RNF212_{1-123}$:$RNF212b_{1-136}$ was determined at two concentrations (20 and 40 μM) and have traces closely matching that of the 20 and 40 μM zinc standards, respectively (Figure 5.2.38.a.). Using the linear plots, the zinc content of $RNF212_{1-123}$:$RNF212b_{1-136}$ can be determined at 35.9 and 42.6 μM, respectively (Figure 5.2.38.b.-d.). From these results we cannot confidently determine the inferred zinc content of $RNF212_{1-123}$:$RNF212b_{1-136}$. It is possible that there is 1 zinc atom per RNF212:RNF212b 1:1 complex, suggesting that in the solution 2:2 complex there are 2 zinc atoms, in comparison to the four zinc atoms in the HEI10 tetramer.



Concentration $RNF212_{1-123}$-$RNF212b_{1-136}$ (µM)	Calculated concentration Zn (µM) at $A_{413.5}$	Calculated concentration Zn (µM) at A_{491}	Inferred Zn atoms per molecule
20	36.14	35.70	~1
40	43.18	42.05	1

Figure 5.2.38.| $RNF212_{1-123}$:$RNF212b_{1-136}$ zinc content determination by PAR assay.
a) Absorbance spectrum between 300-600 nm of zinc standards (0-100 µM - grey). The absorbance spectrum of $RNF212_{1-123}$:$RNF212b_{1-136}$ was determined at two concentrations: 20 µM (light blue) and 40 µM (dark blue). b-c) Experimental spectra at absorbance 413.5 and 491 nm were fitted as a linear standard, with R^2 values of 0.9964 and 0.998, respectively. Zinc standards with absorptivity outside of the linear range are shown in grey and have been excluded from calculating the linear equation. $RNF212_{1-123}$:$RNF212b_{1-136}$ samples are plotted (blue circles). d) Concentration of Zn can be calculated in the protein samples from the linear plot and inferred a single Zn atom per $RNF212_{1-123}$-$RNF212b_{1-136}$ complex, suggesting 2 Zn atoms within the 2:2 complex. (n=1).

SEC-SAXS analysis of $RNF212_{1-123}$:$RNF212b_{1-136}$ was utilised to gain low resolution structural data. The distance distribution profile has a bell-shaped curve with an elongated tail, resulting in a *Dmax* of 127 Å (Figure 5.2.39.b.). $RNF212_{1-123}$:$RNF212b_{1-136}$ has a Guinier analysis determined *Rg* of 38.07 Å, which matches the real-space *Rg* of 39.24 Å (Figure 5.2.39.a.). *Ab initio* modelling was performed to produce a low-resolution molecular envelope from the SEC-SAXS scattering data. The resultant envelope demonstrates the globular nature of the complex, with the formation of a cross-shaped structure indicating that RNF212 and RNF212b RING domains are closely associated, flanked by the coiled-coil domains (Figure 5.2.39.c).

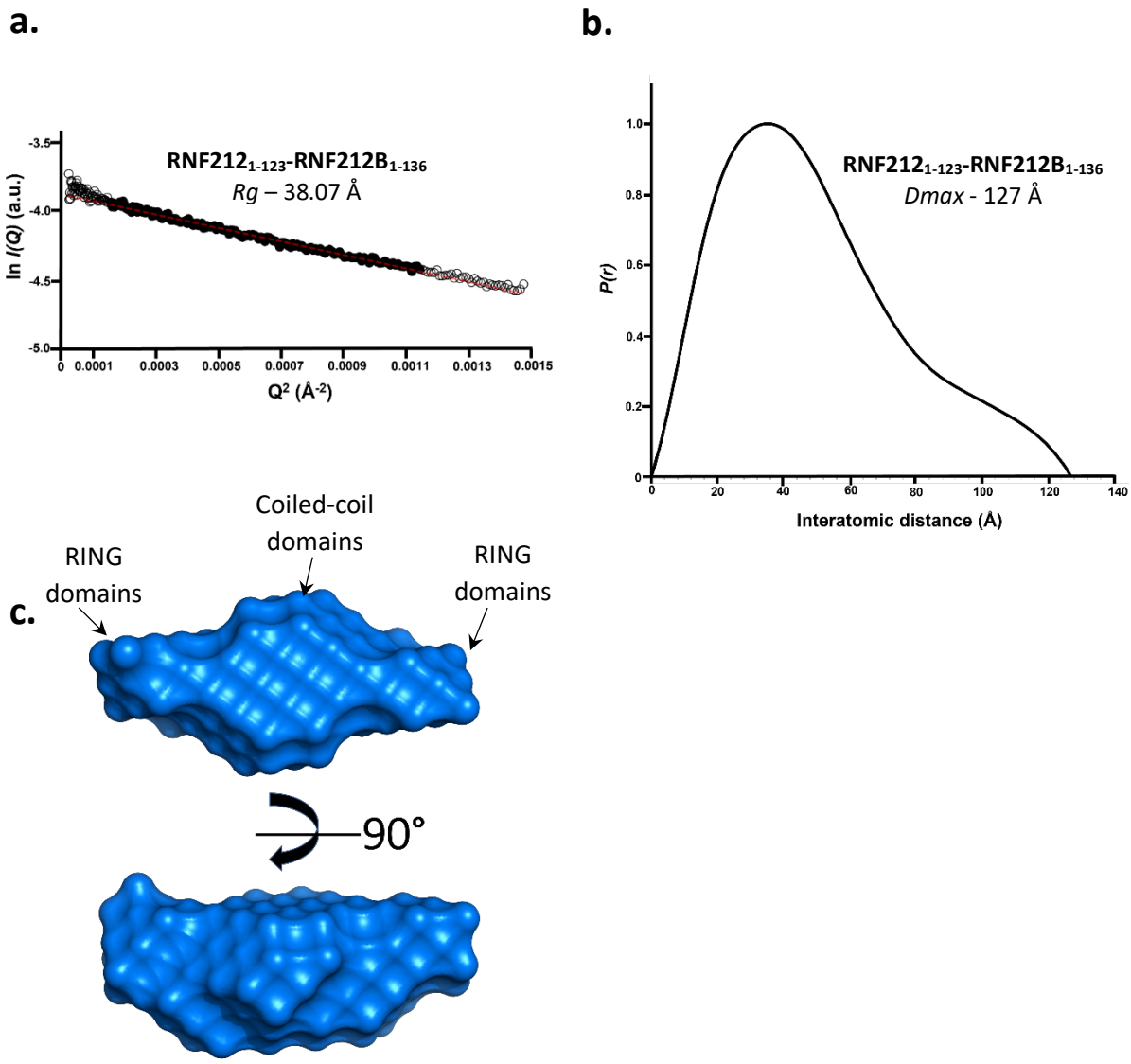


Figure 5.2.39.| SEC-SAXS analysis of RNF212$_{1-123}$:RNF212b$_{1-136}$.
a) Guinier analysis determined a radius of gyration (*Rg*) value of 38.07 Å. The real space *Rg* closely matches the Guinier analysis *Rg* value of 39.24 Å and 38.07 Å, respectively. Empty circles represent the complete data set and the solid circles represent the Guinier region. The linear fit is shown by a red dashed line. (*Q.Rg* values were < 1.3). b) *P(r)* distribution of RNF212$_{1-123}$:RNF212b$_{1-136}$ showing a maximum dimension of 127 Å. c) SAXS *ab initio* DAMMIF model of RNF212$_{1-123}$:RNF212b$_{1-136}$ presented as a molecular envelope.

5.2.20. RNF212 and RNF212b interact through their coiled-coil domains

We have determined that the structural cores of RNF212 and RNF212, residues 1-123 and 1-136, respectively, form a stable 2:2 complex in solution but do not know the specific interacting regions. The coiled-coil domain of RNF212b (amino acid residues 65-136) was co-expressed with the original

structural core construct of RNF212 (amino acid residues 1-198), both as MBP-fusions. The fusion proteins were purified by amylose affinity and anion exchange chromatography. SDS-PAGE analysis of the anion exchange elution profile determined that the fusion proteins co-elute across the salt elution gradient indicating an interaction between $RNF212_{1-198}$ and $RNF212b_{65-136}$ (Figure 5.2.40.a.).

SEC-MALS analysis of MBP-$RNF212_{1-198}$:MBP-$RNF212b_{65-136}$ was used to test if the proteins remain in complex across an isocratic gradient and to determine the absolute molecular weight of the identified protein species. MBP-$RNF212_{1-198}$:MBP-$RNF212b_{65-136}$ eluted as single peak with a determined molecular weight of 225.6 kDa (Figure 5.2.40.b.). Previous analysis of the coiled-coil domain (MBP-$RNF212b_{65-136}$) and the structural core (MBP-$RNF212_{1}$-198) by SEC-MALS determined a dimer and tetramer, respectively (Figure 5.2.32.).

The theoretical molecular weight of a 2:2 complex of MBP-$RNF212_{1-198}$:MBP-$RNF212b_{65-136}$ would be 241 kDa, therefore suggesting that RNF212:RNF212b form a 2:2 complex through their coiled-coil domains. TEV-protease was added to the fusion complex to allow for the cleaved complex to be analysed. However, upon removal of the MBP-solubility tags, both proteins were highly degraded and destabilised. This observation suggests that the coiled-coil domain of RNF212b is sufficient to interact with RNF212 core, but it is a weak interaction and requires the RING domain to form a stable complex. This observation suggests that the RNF212-RNF212b structure may be analogous to the HEI10 tetramer. The experimental data for the human recombination E3 ligase proteins is summarised in Table 5.2.2.

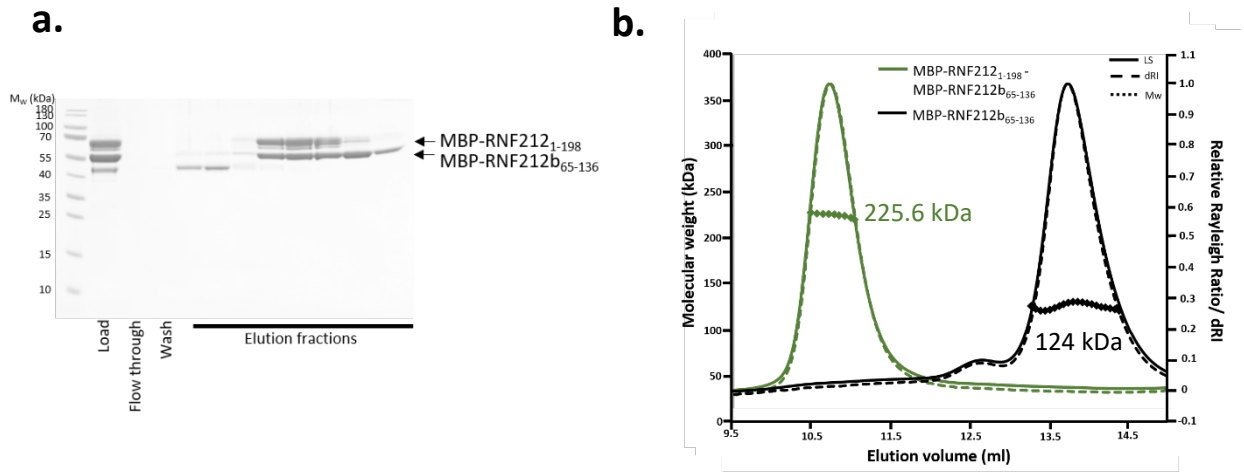


Figure 5.2.40.| Purification and SEC-MALS analysis of MBP-RNF212$_{1-198}$:MBP-RNF212b$_{65-136}$.

a) SDS-PAGE showing the purification of MBP-RNF212$_{1-198}$:MBP-RNF212b$_{65-136}$ by anion exchange chromatography. Determined that the two proteins coelute across the salt concentration gradient suggesting an interaction. b) Overlay SEC-MALS analysis of MBP-RNF212$_{1-198}$-MBP-RNF212b$_{65-136}$ complex (green) and MBP-RNF212b$_{65-136}$ in isolation (black).

5.2.21. *S. cerevisiae* E3 ligase protein, Zip3, dimerises in solution

To investigate whether the budding yeast (*S. cerevisiae*) and fly (*D. melanogaster*) E3 ligase proteins, Zip3 and Vilya/Narya/Nenya, respectively, share similar structural properties to the mammalian proteins, we expressed correlating fragments and constructs in *E. coli* for analysis.

Zip3 encodes a 482 amino acid protein that has a N-terminal RING-finger domain which contains a predicted SUMOylation target consensus sequence (amino acid residues 54-91) (Perry et al., 2005). Furthermore, secondary structure prediction revealed an extended α-helical region (amino acid residues 101-185) with coiled-coil prediction. A Zip3 construct was designed using these predictions to include both the RING and α-helical domain, amino acid residues 42-186.

Zip3$_{42-186}$ was expressed as a N-terminal MBP-fusion to promote solubility. Fusion protein was purified by sequential amylose affinity and anion exchange chromatography, final fusion protein shown in Figure 5.2.41.a.

Table 5.2.2.| Summary table of biophysical properties of human meiotic E3 ligase proteins

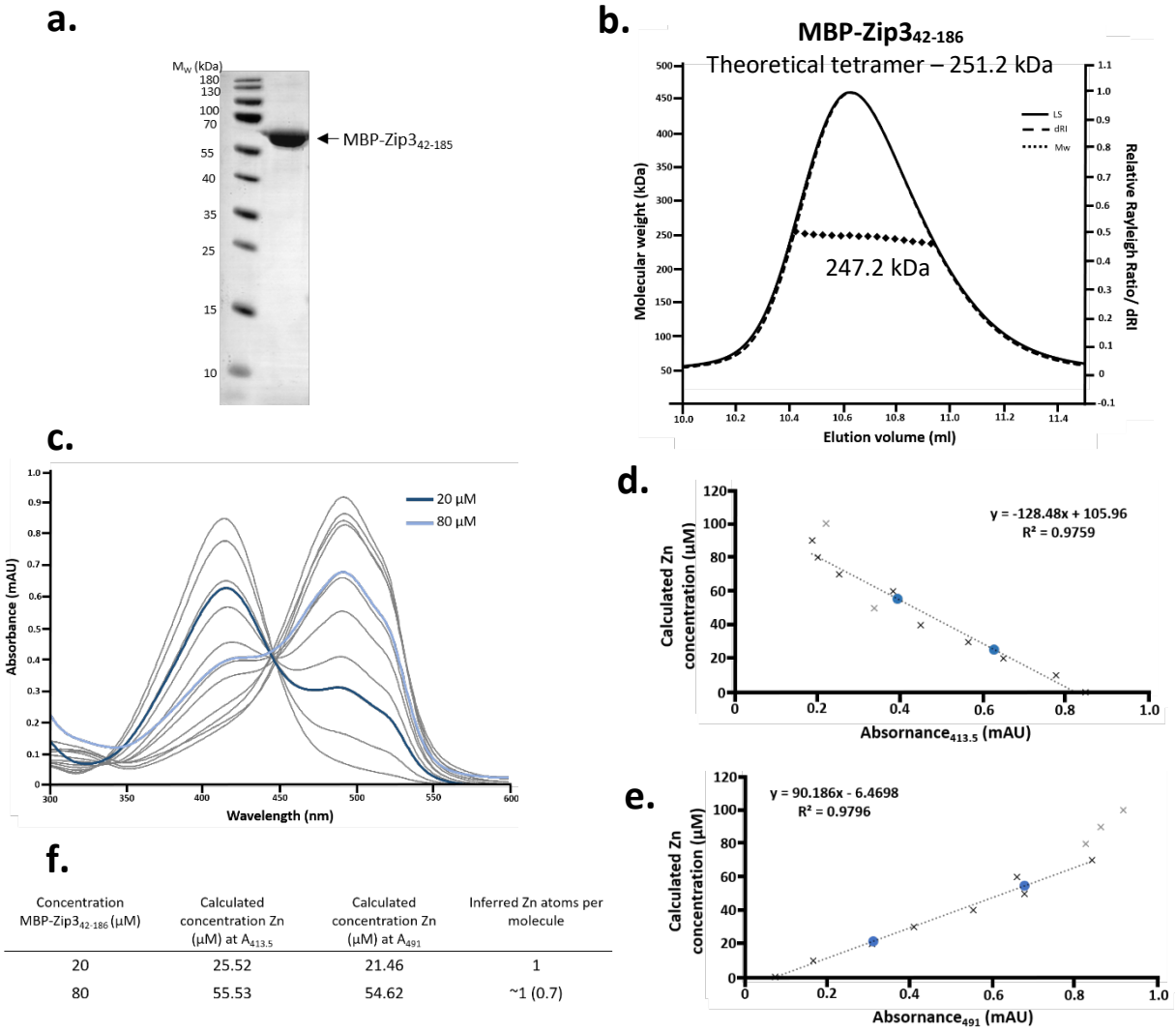
***** - HEI10$_{1-104}$ – SEC-SAXS analysis of the main peak

Human Recombination Proteins Data Summary												
				CD			SEC-MALS			SEC-SAXS		
Construct	Number of residues	Mw (kDa)		Helicity		Tm (°C)	Mw (kDa)		Oligomer	*Rg* (Å)	*Rc* (Å)	*Dmax* (Å)
		cleaved	fusion	%	a.a.		Fusion	Cleaved				
HEI10$_{1-196LL\ (5xTGS)}$ (pMAT11)	211	24.2	68.8	41	80	52	258.7	95.7	tetramer	75.2	11.2	290
HEI10$_{1-196\ (No\ linker)}$ (pRSF)	196	22.7	65.7	-	-	-	-	109.2	tetramer	-	-	-
HEI10$_{1-122}$	122	13.9	56.9	40	49	67.4	-	53.6	tetramer	29.5	-	154
HEI10$_{1-104}$	104	11.7	54.7	32	33	62.2	-	Peak 1 – 46.7 Peak 2 – 32.8	tetramer dimer?	23.29*	-	88*
RNF212$_{1-61}$	61	8.5	51.5	-	-	-	-	11.4	dimer	17.7	-	57
RNF212b$_{1-56}$	56	6.4	49.4	-	-	-	-	8.6	monomer	17.5	-	59
RNF212$_{1-198}$	198	22.6	67.2	-	-	-	Peak 1 – mDa Peak 2 - 281	-	aggregate tetramer	-	-	-
RNF212b$_{1-136}$	136	17.4	62	-	-	-	246	-	tetramer	-	-	-
RNF212b$_{56-136}$	80	8.7	53.3	-	-	-	124	-	dimer	-	-	-
RNF212$_{1-123}$-RNF212b$_{1-136}$	259	30.3 (1:1)	73.3 (1:1)	55	142	64	-	61.2	2:2	38.07		127
RNF212$_{1-198}$-RNF212b$_{56-136}$	278	34.5 (1:1)	120.5 (1:1)	-	-	-	225.6	-	2:2	-	-	-

SEC-MALS analysis of His_6-MBP- $Zip3_{42-185}$ revealed a single peak with a calculated molecular weight of 247.2 kDa, corresponding to a tetramer (theoretical tetramer - 251.2 kDa) (Figure 5.2.41.b.). This result indicates that the structure of budding yeast E3 ligase may share similarities to the HEI10 tetramer observed for mammalian E3 ligases.

We utilised the PAR assay to determine the zinc content of the fusion $Zip3_{42-185}$ construct and potentially give an insight into the folding of Zip3. The absorbance spectra of MBP-$Zip3_{42-185}$ determined at two concentrations (20 and 80 μM). Protein trace at 80 μM indicate the presence of zinc, shown by a peak at 514 nm, matching the 60 μM zinc standard the closest (Figure 5.2.40.c.). Using the linear plots, the zinc content of MBP-$Zip3_{42-185}$ was determined at 23.5 and 55.1 μM, respectively (Figure 5.2.40.d.-f.). From these results we can estimate the inferred zinc content of Zip3 to be 1 zinc atom per protein molecule, suggesting the Zip3 tetramer coordinates 4 zinc atoms.

In order to analyse cleaved material, MBP-$Zip3_{42-185}$ was subjected to affinity-tag cleavage by TEV protease. Despite cleavage optimisations and the addition of a TGS-linker tag between the TEV-recognition site and the start of the protein, little cleavage was observed. This result suggests that the material is possibly aggregated upon TEV cleavage and requires the MBP-affinity tag for solubility. Furthermore, it may suggest that Zip3 requires a binding partner to form a constitutive complex.



f.

Concentration MBP-Zip3 42-186 (µM)	Calculated concentration Zn (µM) at $A_{413.5}$	Calculated concentration Zn (µM) at A_{491}	Inferred Zn atoms per molecule
20	25.52	21.46	1
80	55.53	54.62	~1 (0.7)

Figure 5.2.41.| Biophysical analysis of MBP-Zip3$_{42-185}$

a) SDS-PAGE showing the final product of MBP-Zip3$_{42-185}$, purified through sequential amylose affinity and anion exchange chromatography. b) SEC-MALS analysis of MBP-Zip3$_{42-185}$ revealed a single peak with molecular weight of 247.2 kDa, indicating that the structural core of Zip3 forms a tetramer in solution (theoretical tetramer – 251.2 kDa). c-f) MBP-Zip3$_{42-185}$ zinc content determination by PAR assay. c) Absorbance spectrum between 300-600 nm of zinc standards (0-100 µM - grey). The absorbance spectrum of MBP-Zip3$_{42-185}$ was determined at two concentrations: 20 µM (dark blue) and 80 µM (light blue). d-e) Experimental spectrum at absorbance 413.5 (d) and 491 (e) were fitted as a linear standard, with R^2 values of 0.9759 and 0.9796, respectively. Zinc standards with absorptivity outside of the linear range are shown in grey and have been excluded from calculating the linear equation. MBP-Zip3$_{42-185}$ samples are plotted (blue circles). f) Concentration of Zn can be calculated in the protein samples from the linear plot and inferred a single Zn atom per MBP-Zip3$_{42-185}$ molecule, suggesting 4 Zn atoms are co-ordinated by the Zip3 tetramer. (n=1).

5.2.22. In solution characterisation of the *Drosophila* E3 ligase proteins: Vilya, Narya and Nenya

D. melanogaster possess three E3 ligase family proteins; Vilya and more recently identified Narya and Nenya. Vilya is a conserved component of the *Drosophila* recombination nodule and has been shown to be required for DSB formation and involved in CO determination (Lake et al., 2019, 2015). Vilya is a Zip3-like RING containing protein, with a predicted coiled-coil domain and a serine-rich C-terminal tail. Localisation studies have shown that Vilya shares a similar localisation pattern to some Zip-family members including Zhp-3 (worm), RNF212 (mouse) and HEI10 (rice and *A. thaliana*) (Lake et al., 2015). In addition to Vilya, two further E3 ligase Zip3-like proteins have been identified in *Drosophila*, Narya and Nenya. The protein paralogues, Narya and Nenya, have structural similarities to Vilya and have homology to the Zip3-like family (Lake et al., 2019).

Secondary structure predictions and *in vivo* interactions suggest that Vilya, Narya and Nenya have functional similarities to the mammalian Zip3- and HEI10-like family of E3 ligases. We set out to biophysically characterise these proteins *in vitro* to compare their structure with HEI10 and RNF212/RNF212b.

Constructs of Vilya, Narya and Nenya including the N-terminal RING domain and predicted coiled-coil domain were designed using sequence conservation and secondary structure predictions. Vilya$_{1\text{-}181}$, Narya$_{1\text{-}162}$ and Nenya$_{1\text{-}148}$ were cloned with a N-terminal MBP-solubility tag and expressed in *E. coli* BL21 (DE3) cells. Both MBP-Vilya$_{1\text{-}181}$ and MBP-Narya$_{1\text{-}162}$ poorly expressed, the small amount of soluble material was subjected to amylose affinity chromatography. Both constructs show high amounts of degradation, as multiple bands were observed in the amylose eluate gel sample (Figure 5.2.41.a. and c.). This finding suggests that both Narya and Vilya are unstable when expressed in isolation. In comparison, MBP-Nenya$_{1\text{-}148}$ highly expressed and was purified by sequential amylose affinity and anion exchange chromatography (Figure 5.2.42.d.). Fusion material was analysed by SEC-MALS to deduce the absolute molecular weight and determine the oligomeric state. MBP-Nenya$_{1\text{-}148}$ eluted as a single peak with a calculated molecular weight of 117.5 kDa, corresponding to a dimer (theoretical dimer – 125.2 kDa) (Figure 5.2.42.f.). This result differs to our findings of the HEI10 and Zip3 structural

cores, suggesting a possibility of a different mechanism of function, and like RNF212 may require a binding partner. Furthermore, upon TEV-cleavage of Nenya$_{1\text{-}148}$, we observed poor cleavage and aggregation, this suggests that the MBP-affinity tag was providing stability and Nenya may require a binding partner.

Furthermore, the N-terminal RING-domain of Vilya, residues 13-79, was expressed and purified in isolation. MBP-Vilya$_{13\text{-}79}$ was purified by amylose affinity chromatography followed by anion exchange chromatography to remove any free MBP (Figure 5.2.42.b.). SEC-MALS analysis of MBP-Vilya$_{13\text{-}79}$ determined a single peak with a calculated molecular weight of 53 kDa, corresponding to a monomer (theoretical monomer – 52.6 kDa). (Figure 5.2.42.e.). This result suggests that the RING domain of Vilya may share structural similarities to the RING domain of RNF212, as both are monomeric in solution. However, further analyses would be needed to confirm this.

Analysis of both Zip3-like and HEI10-like E3 ligases has determined that despite sharing a similar secondary structure domain structure, recombination proteins of different model organisms have shown differences in their structural properties suggesting that there might be subtle differences in how the E3 ligase proteins function across species.

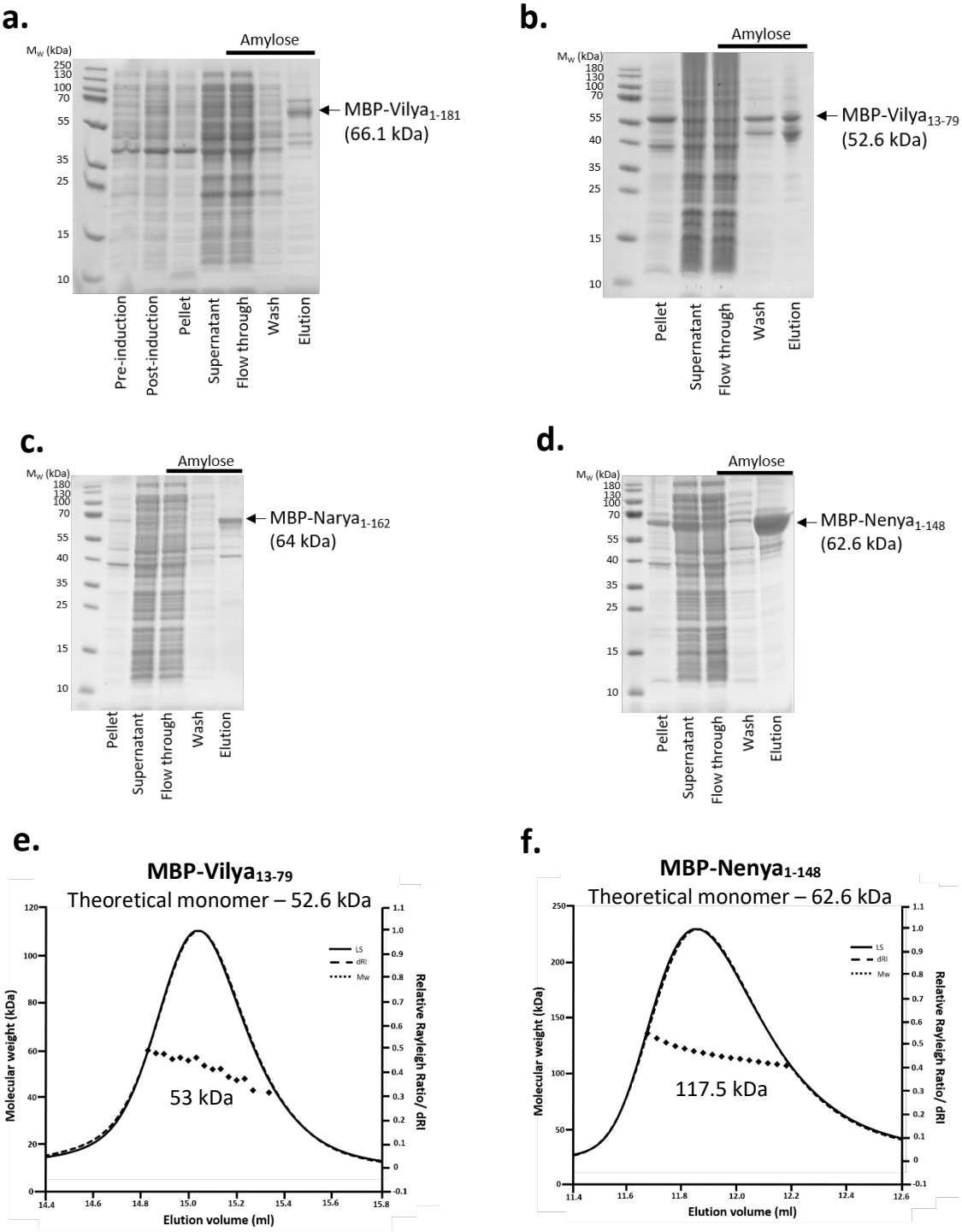


Figure 5.2.42.| Purification and SEC-MALS analysis of *Drosophila* E3 ligase proteins.
a-d) SDS-PAGE analysis showing the amylose purification step of *Drosophila* E3 ligase proteins. a) The limited amount of soluble material of MBP-Vilya$_{1-181}$ was purified by amylose affinity chromatography. MBP-Vilya$_{1-181}$ highly degrades shown by multiple bands in the amylose eluate. b) Amylose affinity chromatography of the ring domain of Vilya, MBP-Vilya$_{13-79}$. c) MBP-Narya$_{1-162}$ expressed poorly with a very small amount of material in the amylose eluate. d) Amylose affinity purification ofMBP-Nenya$_{1-148}$. e) SEC-MALS analysis of MBP-Vilya$_{13-79}$ determined a single peak with molecular weight of 53 kDa, suggesting that RING domain is a monomer in solution when expressed in isolation (theoretical monomer – 52.6 kDa). f) SEC-MALS analysis of MBP-Nenya$_{1-148}$ determined a single peak with molecular weight of 117.5 kDa, suggesting that the structural core of Nenya forms a dimer in solution (theoretical dimer – 125.2 kDa).

5.3. DISCUSSION

The roles played by HEI10 and RNF212 in meiotic prophase I have been extensively studied *in vivo* using an immunocytological approach (Holloway et al, 2014; Qiao et al., 2014; Reynolds et al., 2013; Toby et al., 2003). However, there is currently no literature on the structures of these proteins. All three human E3 ligase proteins, HEI10, RNF212 and RNF212b share a conserved tripartite secondary structure; a N-terminal RING domain, an α-helical predicted coiled-coil domain and a unstructured C-terminal tail domain. Similarly, the meiotic E3 ligase proteins of *Drosophila* and *S. cerevisiae*: Vilya, Narya, Nenya and Zip3 also have the same conserved secondary structure. This chapter aimed to establish the first structural insight into the human, *Drosophila* and yeast meiotic E3 ligase proteins through solution biophysical methods, including SEC-MALS and SEC-SAXS.

5.3.1. HEI10 requires both the RING domain and putative coiled-coil region to form a stable tetramer

Despite there being over 600 human RING-type E3 ligases genes, only a few dozen structures have been solved (Deshaies & Joazeiro, 2009). We postulate that this may be due to the difficulties in obtaining sufficient protein for structural and functional studies. From the start of the project, it quickly became apparent that HEI10 is poorly soluble, unstable and has a tendency to aggregate.

Expression and purification of full length HEI10 shows a substantial amount of insoluble material and high levels of degradation, possibly due to the relatively unstructured C-terminus. Biophysical analysis of $HEI10_{1-297}$ fused to an MBP-solubility tag revealed that HEI10 tetramerises in solution, but no analysis of the cleaved protein was obtained. Analysis of the N-terminal RING domain (amino acid residues 1-70) and the predicted coiled-coil region (amino acid residues 69-196) in isolation revealed megadalton species, suggesting that both the RING and coiled-coil domains constitute its structural core and are both required for protein folding.

Initial analysis of the HEI10 structural core was carried out using a high copy number plasmid (pMAT11) and required a long linker to aid cleavage. However, switching to a low copy number

plasmid (pRSF-duet) showed an improvement in both solubility and cleavage. This could be due to minimised transcription-replication conflicts in the low copy number plasmid which aids the production of non-truncated, thus correctly folded soluble material (Merrikh et al., 2012).

We have determined that the HEI10 structural core, amino acid residues 1-196, exists as a stable tetramer. Despite failed attempts of $HEI10_{1-196}$ crystallisation to gain a high resolution structure, SEC-SAXS analysis has provided an insight into its structure. SEC-SAXS determined scattering data indicative of an elongated molecule, 290 Å long with a 11.2 Å cross-sectional radius. The *ab-initio* low resolution molecular shape shows an elongated envelope, with two oval 'lobes' connected by a filamentous bridge, analogous to a dumbbell-like shape. We propose a model for the organisation of the HEI10 tetrameric structural core which consists of two dimers interacting through their coiled-coil domain to form a four helical bundle structure, flanked by the globular dimeric RING domains, schematised in Figure 5.3.1. This structure allows for the RING domains to interact with E2-enzymes and to carry out ubiquitination on target proteins (De Muyt et al., 2014).

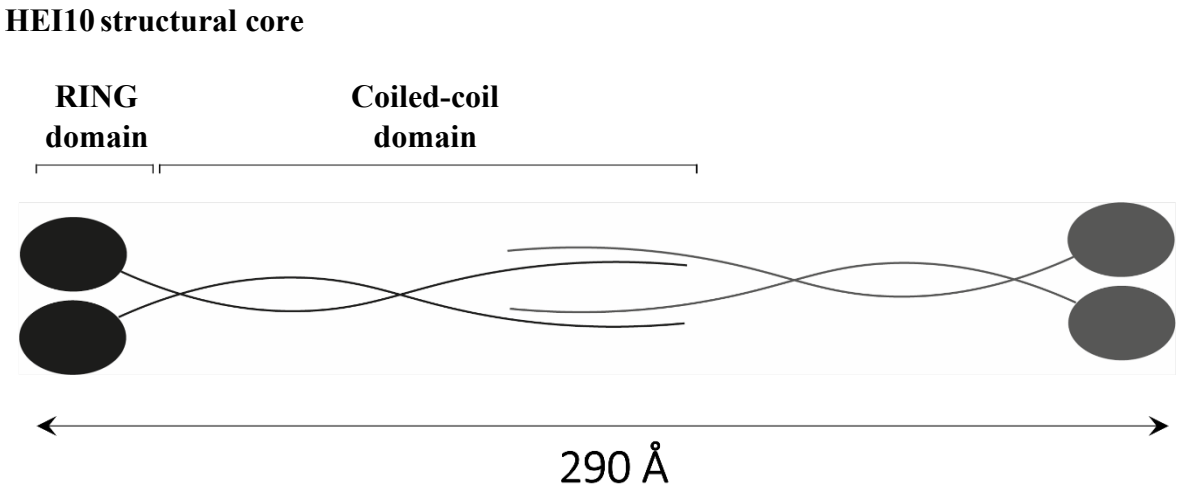


Figure 5.3.1.| Schematic model of $HEI10_{1-196}$ structural core.
HEI10 core, residues 1-196, forms a stable tetramer in solution. The coiled-coil domain of each dimer interact to form a dimer of dimers or tetramer. The coiled-coil domains are flanked by the N-terminal RING domains, which we suspect interact to initiate dimerisation.

In order to increase our chances of obtaining HEI10 crystallisation and diffraction, we analysed two shorter constructs. The two constructs, $HEI10_{1-122}$ and $HEI10_{1-104,}$ truncated the α-helical domain and

linking region, to form a compact 'folding domain', and reduce tendencies towards aggregation. Molecular weight analysis revealed that the two truncated constructs were predominantly tetrameric however, SEC-MALS result of $HEI10_{1-104}$ shows a wide peak with a non-separated shoulder, suggesting reduced stability compared to the longer HEI10 structural core. The calculated molecular mass of the main peak and the shoulder correspond to a tetramer and a possible dimer, respectively. This suggests that the α-helical domain is responsible for tetramerisation of HEI10.

SAXS analysis of both $HEI10_{1-104}$ peaks have a primarily globular structure, corresponding to the RING domains. Surprisingly, the maximum distance (*dmax*) of both the first (pre-peak) and second (main) peak were 98 and 88 Å, respectively. This observation suggests that the maximum length of $Hei10_{1-104}$ remains unchanged despite changing between a dimer and a tetramer. The molecular mass of the protein constructs constituting the two peaks were estimated by calculating the volume of correlation (*Vc*), which determined that the main peak could be in fact be a dimer, suggesting this is the predominant species. Therefore, we propose an alternative model for HEI10, in which the RING domains form a head-to-head dimer and tetramerisation is achieved through the coiled-coil domains, schematised in Figure 5.3.2.). However, in reality the molecular organisation of the HEI10 may be a combination of both models, through cooperative interaction of the α-helices and dimerisation of the RING domains. Importantly, both models suggest that the RING and coiled-coil domain are required for stable tetrameristion.

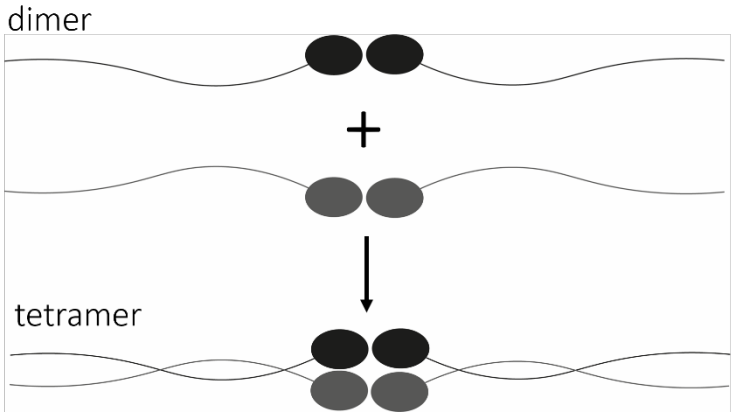


Figure 5.3.2.| Alternative schematic model for the meiotic E3 ligase.
SEC-SAXS analysis of $HEI10_{1-104}$ revealed two peaks, predicted to be a dimer and a tetramer but with similar lengths. These results suggest an alternative model of the HEI10 tetramer, in which the RING domains initiate dimerisation, and tetramerisation occurs through the α-helical domain.

Although the SAXS curve can provide a concentration-independent estimate of mass and low resolution structural properties, the ultimate aim is to still gain a crystal structure for HEI10 to provide a definite structure and confirm how the domain structure of HEI10 is utilised to form the tetramer.

5.3.2. RNF212 and RNF212b form a heterodimeric 2:2 complex

Structural comparisons between the ubiquitin and SUMO E3 ligases, HEI10 and RNF212, respectively, may be important to identify similarities as well as differences in their mechanism of function and activity.

We have studied both RNF212 and its paralogue RNF212b, in isolation. In comparison to HEI10, the RING domains of RNF212 and RNF212b were soluble and formed a dimer and monomer, respectively. Moreover, both RING domains co-ordinate a single zinc atom per protein molecule. We can infer from the oligomeric differences, that as RNF212 is able to oligomerise in isolation, it may be functional on its own, in comparison to RNF212b which may need to interact with another protein to function. Paralogues are homologous proteins that are related by a gene duplication event and may be functionally redundant (Koonin, 2005). Therefore, RNF212b may be free to acquire sequence changes that alter its structure, which introduces subtle differences to the RNF212 and HEI10 strucuture. Functional studies in mice indicate that in addition to HEI10, RNF212 is also essential for crossover formation, suggesting a possibility of redundancy for RNF212b (Reynolds et al., 2013).

Analysis of the structural cores (RING and α-helical domain) of RNF212 and RNF212b revealed that both proteins form a tetramer in solution. However, when the solubility tag was cleaved both proteins were structurally unstable. These observations also suggest disorder is due to requiring a binding partner to aid stability and folding. Despite the literature demonstrating *in vivo* the co-localisation of HEI10 and RNF212 during mid-pachytene, we were unable to detect a direct interaction between the structural cores of HEI10 and RNF212 or RNF212b by Y2H screens or amylose affinity chromatography (Prasada Rao et al., 2017; Qiao et al., 2014). This suggests that the HEI10 and RNF212 may not directly

interact, but an interaction may take place in a larger multi-protein complex, including other ZMM recombination pathway proteins and the SC proteins.

A direct interaction was determined between the structural cores of RNF212 and RNF212b. Through biophysical characterisation, RNF212 and RNF212b were found to form a 2:2 complex in solution. Upon complex formation both RNF212 and RNF212b must undergo remodeling from individual tetramers to facilitate an interaction. SEC-SAXS analysis of the 2:2 $RNF212_{1-123}$:$RNF212b_{1-136}$ complex determined structural parameters similar to that of the HEI10 construct of similar length ($HEI10_{1-122}$).

Removal of the RNF212b RING domain (amino acid residues 1-56) did not disrupt the 2:2 complex formation with the RNF212 core. This implies that the RING domains of RNF212 and RNF212b are not required for complex formation and instead is achieved through the predicted coiled-coil domain. From these results, we propose that the RNF212:RNF212b complex has a structure similar to the HEI10 tetramer. Furthermore, the protein structure dictates biochemical function, suggesting that both RNF212 and RNF212b are required for the differentiation of crossover and non-crossover sites (Qiao et al., 2014; Reynolds et al., 2013).

Numerous studies of the mammalian synaptonemal complex proteins have shown the formation of tetrameric coiled-coil building blocks that self-assemble. Our study of the meiotic recombination proteins have shown that they also have a highly conserved α-helical domain with coiled-coil prediction and are crucial for the formation of stable tetramers (Dunce et al., 2018; Syrjänen et al., 2014; West et al., 2019). Moreover, coiled-coils are known to be widely used to facilitate oligomerisation and provide flexibility which may be important for both structure and function at crossover sites (Park, 2020).

5.3.3. Meiotic E3 ligase RING domains coordinate zinc ions and bind DNA

A defining property of zinc-finger domains, including RING domains, is zinc coordination by the cysteine and/or histidine residues, which in turn provides a structural framework to stabilise their tertiary structure (Deshaies & Joazeiro, 2009). Moreover, it has been shown that the coordination of the

two zinc ions by the RING domain is crucial for E3 ligase activity, and eliminating the zinc coordination severely disrupts the ubiquitination activity (Chasapis et al., 2010).

We utilised the PAR assay to experimentally determine the presence of zinc and estimate the number of zinc atoms present per protein molecule. The results of the PAR assays of all three HEI10 constructs demonstrate that zinc is released. We were able to infer that a single zinc atom is coordinated by a HEI10 molecule, suggesting that four zinc atoms are coordinated by the HEI10 tetramer.

Thus far, most zinc-finger proteins have shown to coordinate either a single or two zinc ions, depending on the number of cysteine and histidine residues and their position within the sequence (Deshaies & Joazeiro, 2009). As previously mentioned, the RING domain sequence of HEI10 is slightly different to the canonical sequence, suggesting divergence. Therefore, our finding of a HEI10 monomer coordinating a single zinc ion is not unexpected but suggests the need for oligomerisation.

In addition, analysis of the RNF212 and RNF212b RING domains in isolation and the RNF212-RNF212b complex by PAR also revealed zinc binding, with a ratio of one zinc per protein molecule for the RING domains and 2 zinc atoms in the RNF212-RNF212b complex. These results suggest that the meiotic E3 ligase protein HEI10 and the RNF212-RNF212b complex may have differences in their structure and zinc coordination, possibly related to their functions in the ubiquitin and SUMO pathways, respectively.

We have shown by EMSA that the HEI10 structural core and RNF212b RING domain bind dsDNA. These results suggest that the E3 ligase RING domains of HEI10, RNF212b and presumably RNF212 mediate DNA-binding. Both RNF212 and HEI10 accumulate at crossover sites and colocalise with MutSγ and MutLγ, respectively. It has been proposed that HEI10 and RNF212 work antagonistically at recombination sites to promote crossover formation. HEI10 is thought to promote the dissociation of RNF212 to allow the progression of recombination (Qiao et al., 2014). This could be achieved through their binding potential to the recombination intermediates. To test this hypothesis, EMSAs could be performed to test the DNA binding ability of HEI10 upon the addition of RNF212-RNF212b. Additionally, we could perform EMSAs with the Holliday junction DNA intermediate as the substrate.

5.3.4. The recombination machinery is structurally conserved across meiotic dividing organisms

In addition to the human meiotic E3 ligases, we have provided a structural insight into both the *Drosophila* and *S. cerevisiae* E3 ligases. Studies have shown that the *D. melanogaster* E3 ligase protein, Vilya, appears to be homologous to the Zip3-like (SUMO) protein family and is involved in crossover fate as well as DSB formation (C. M. Lake et al., 2015b). Since it is generally recognised that proteins with similar functions tend to have a similar structure, we set out to explore this. Purification of Vilya N-terminal RING domain, and subsequent molecular weight analysis determined that $Vilya_{13\text{-}79}$ is a stable monomer. Analysis of the Vilya core structure resulted in poor expression, thus showing strong similarities with RNF212b. More recently, two paralogous proteins, Narya and Nenya, have been identified in *Drosophila* and shown to have strong sequence conservation with Vilya. Expression of Narya yielded insoluble material, suggesting that it requires a constitutive binding partner. In *Drosophila* females, both Narya and Vilya are required for DSB formation and repair and shown to localise to the SC central region and to DSBs (C. M. Lake et al., 2019, 2015b). In comparison, analysis of Nenya structural core demonstrated dimerisation, indicating structural similarities with the Zip3-like family of E3 ligase proteins. As Narya and Nenya are related by a gene duplication event, there is a possibility of redundancy, therefore less evolutionary pressure leading to sequence changes and thus structural differences between the three E3 ligase proteins in *Drosophila* (Koonin, 2005).

Purification of *S. cerevisiae* Zip3 structural core and SEC-MALS analysis revealed a mass corresponding to a dimeric species. Moreover, it was not possible to cleave the MBP-tag, suggesting possible protein aggregation. These results suggest, similar to RNF212, that Zip3 may need an interacting partner for stability and function. *In vitro* studies have shown that Zip3 co-localises with Zip2, and deletion of Zip3 results in the failure of SC formation, suggesting possible interacting partners (Agarwal & Roeder, 2000). Yeast two hybrid screens and co-immunoprecipitation assays of Zip3 with SC components and ZMM proteins determined interactions with Zip2, Zip4, Msh4 and the SC transverse filament protein Zip1 (Pyatnitskaya et al., 2019). These possible interactions need to be tested biochemically to determine a Zip3-interacting partner.

Despite not gaining a comprehensive study of the E3 ligases in yeast and fly, we have provided an initial insight into the conservation of structure and possible interactions across species. As expected, the Zip3-like family of SUMO E3 ligases have a conserved core dimeric structure. In comparison, HEI10, a ubiquitin E3 ligase is tetrameric and stable when expressed and purified in isolation. These findings suggest that the SUMO-modification and ubiquitin-proteosome systems may have subtle differences in their structures to aid their specific function in the two pathways.

5.3.5. How is the structure of HEI10 and RNF212(b) related to their functions in meiotic recombination?

Using a combination of biophysical and biochemical techniques we have shown that the mammalian E3 ligase proteins, HEI10, RNF212 and RNF212b, form tetrameric structures in solution that are able to co-ordinate zinc atoms through their N-terminal RING domains and bind dsDNA. The conserved arrangement of four RING domains on a coiled-coil architecture suggest that their structure may play a crucial role in the recruitment of E2 conjugating enzymes, substrate recognition or the direct catalysis of ubiquitin/SUMO transfer from the E2 enzyme to the protein substrate (Metzger et al., 2014).

Coiled-coil proteins exhibit a large diversity of biological functions. The physical properties of heptad repeats, especially their length and flexibility, can provide molecular spacing that physically separate functional domains or scaffold large macromolecular complexes (Truebestein & Leonard, 2016). In addition, coiled-coil domains have been identified in enzymes, where they function as molecular spacers, positioning the catalytic domains at fixed distances (Rose & Meier, 2004).

We propose that the coiled-coil motif of the meiotic E3 ligases is responsible for the oligomerisation to form an active tetramer, separating the RING domains by a set length. The activity of E3 ligases require tight regulation to prevent undesired activity (Garcia-Barcena et al., 2020). It is possible that the oligomerision and/or complex formation of HEI10 and RNF212-RNF212b may modulate the activity of the RING E3 ligases. Moreover, the tetrameric coiled-coil could function as a molecular scaffold that organises the biochemical activities of the RING domain and confines enzymatic activity spatially with

respect to its substrate. Thus, their structures allows for the RING domains to interact with E2-enzymes and to carry out the highly controlled ubiquitination or SUMOylation on target proteins (De Muyt et al., 2014).

6.0. General discussion and future work

The study of mammalian meiosis *in vivo* is very challenging, due to the lack of a genetically-tractable system. Thus, our approach of using biophysical and structural techniques, to characterise mammalian meiotic proteins *in vitro*, facilitates the elucidation of molecular details within meiosis. This thesis investigated the molecular architecture and function of three essential meiotic multiprotein complexes; the meiotic telomere complex (MTC), the synaptonemal complex (SC) and the meiotic recombination machinery, which together ensure the intricate chromosome choreography required for chromosome synapsis and segregation. Misregulation of these processes can result in infertility and various genetic disorders including aneuploidy and azoospermia. Through SEC-MALS and SEC-SAXS techniques we determined the stoichiometry, topology, size, shape, and dimensions of individual proteins and protein complexes in solution. This has provided insights into the structure-function relationship of the SC and meiotic telomere complex.

The MTC integrates functions of the LINC and shelterin complexes to achieve telomere attachment. Our work provides a molecular model of the connection between telomeres and the nuclear envelope. We have shown that the MTC proteins, MAJIN-TERB1-TERB2, can recruit TRF1, leading to the formation of the meiotic telomere recruitment complex. Thus, bringing telomeric DNA into close proximity, which in turn, enables its loading onto the MTC and subsequent TRF1 displacement.

Previous crystallographic studies have shown that TRF1-TERB1 forms a 2:2 complex, in which two $TERB1_{TBM}$ peptides bind to the $TRF1_{TRFH}$ dimer. Moreover, the 2:2 complex was shown to be readily disrupted by CDK2 phosphorylation of TERB1, leading to the proposal of a TRF1 displacement mechanism (Long et al., 2017; Pendlebury et al., 2017). However, we have shown that a longer $TERB1_{TRFB}$ construct binds to the $TRF1_{TRFH}$ dimer, forming a 2:1 complex, which is undisrupted by the TERB1 phosphomimetic mutation and CyclinB-CDK1 treatments. Therefore, we propose that CDK2 phosphorylation must work in combination with other signalling and phosphorylation events of TERB1 and/or surrounding proteins to achieve TRF1 displacement.

In the presented work, we have analysed the meiotic telomere recruitment complex in isolation. Recent studies have shown that CDK2 and its regulator SpeedyA play important roles in telomere attachment to the nuclear envelope (Mikolcevic et al., 2016; Tu et al., 2017; Viera et al., 2015). The identification of SpeedyA-TRF1 and SpeedyA-SUN1 interactions suggest that SpeedyA might function as the missing link between the LINC complex and the telomeres, and play a role in stabilising the MTC-telomere NE attachment and TRF1 displacement (Chen et al., 2021; Wang et al., 2018). An interesting route to increase our understanding of the TRF1 displacement mechanism could be to form a large multi-protein complex of the MTC, including the LINC complex protein, SUN1, CDK2 and SpeedyA, to achieve a holistic understanding of the telomere attachment mechanism.

In the last decade, through structural biology, our molecular understanding of the protein components and complexes constituting the mammalian SC has dramatically increased (Dunce et al., 2018; Dunne & Davies, 2019; Lu et al., 2014; Sánchez-Sáez et al., 2020; Syrjänen et al., 2014; West et al., 2019). Structural analysis of mammalian SC proteins revealed that they predominantly consist of coiled-coil domains which have the propensity to self-assemble. We hypothesise that these conserved structures are strongly coupled to their functions during meiosis.

Central to the tripartite SC structure are the transverse filaments (TFs), which resemble train tracks, connecting the two parallel lateral elements (LEs). The human TF protein, SYCP1, has been studied in great detail. SYCP1 is a tetramer that undergoes self-assembly to form a supramolecular lattice between synapsed chromosomes (Dunce et al., 2018). Here, we have provided the first structural insight into the *D. melanogaster* TF protein, C(3)G, and have shown that the structure of TFs is in fact conserved across species. Biophysical and biochemical analysis of C(3)G demonstrates that the characteristic central α-helical domain constitutes discontinuous coiled-coils that form a dimeric parallel configuration. During prophase I, in addition to the MTC, the TFs are subjected to substantial forces required for the rapid telomere-led prophase movements (Lee et al., 2015). Therefore, the discontinuous coiled-coils of C(3)G, which can form semi-flexible rods, provide elasticity and flexibility to withstand the mechanical strain of the RPMs. Moreover, the central region of the α-helical domain has the

propensity to form a tetrameric structure, suggesting a possible self-assembly mechanism, which could be similar to that of SYCP1 (Dunce et al., 2018).

It is proposed that like SYCP1, C(3)G contains DNA binding sites within its unstructured C-terminal tail, however there was no experimental evidence showing this. Here, we have shown that the globular, monomeric C-terminus of C(3)G does in fact undergo DNA binding therefore propose that C(3)G uses a similar recruitment mechanism to the chromosomal axes as SYCP1 (Dunce et al., 2018).

In addition, we have biophysically analysed three in-frame deletion mutations in the predicted coiled-coil domain of C(3)G. A recent study, using *in vivo* techniques, has shown that all three of these deletions cause a partial loss of SC function at different stages in early meiosis (Billmyre et al., 2019). We demonstrated that the three in-frame deletions did not affect protein folding and stability, however $C(3)G^{del1}$ did disrupt dimerisation, further indicating that the central α-helical domain is crucial for dimer formation. In summary, our biophysical findings are contradictive of the *in vivo* analysis, suggesting that the deletions within the C(3)G coiled-coil domain do not affect the structure but are affecting something else, such as an interaction site, to lead to the loss of SC function observed in homozygous deletion females (Billmyre et al., 2019). To further investigate this, we need to explore the interaction network of the *Drosophila* SC and characterise the multi-protein complexes to widen our structural understanding of the SC in different model organisms.

Although our understanding of the structure and mechanisms of functional roles within the human SC has significantly increased, the complex relationship between the SC and recombination initiation and crossing over is still poorly understood (de Boer & Heyting, 2006; Hayashi et al., 2010). Moreover, the mechanisms responsible for crossover assurance and interference remain unknown due to the complexity of studying these processes. Here, we focused on gaining a structural understanding of the human E3 ubiquitin and SUMOylation ligase proteins, HEI10, RNF212 and RNF212b. We have shown that the HEI10 structural core exists as a stable tetramer. SEC-SAXS analysis and examination of the solution structures of numerous HEI10 constructs has provided an insight into the structure of HEI10 and we have proposed two possible models for their organisation. In addition, we determined a novel protein interaction between the E3 SUMOylation ligase proteins RNF212 and RNF212b, suggesting

that both proteins play roles in crossing over, and are not functionally redundant as initially proposed (Johnston et al., 2020; Kadri et al., 2016). Through SEC-MALS and SEC-SAXS analysis, we find that RNF212-RNF212b forms a 2:2 complex, possibly through their coiled-coil domains, and has a solution structure analogous to the HEI10 tetramer. These observations suggest that both HEI10 and RNF212-RNF212b have corresponding functions to achieve crossover formation by ubiquitination or SUMOylation, respectively. However, the accuracy of our proposed model will be significantly increased with the crystal structures of HEI10 and RNF212:RNF212b. An interesting avenue would be to carry out a Y2H study of HEI10/RNF212(b) to screen for target proteins, thus test their ability to ubiquitinate/SUMOylate their binding partners, as well as determining the mechanistic basis of the interplay between the recombination machinery and SC to achieve crossover designation.

The work presented here provides a solid foundation to build upon in future studies. We have investigated individual proteins and multiprotein complexes which are crucial for meiosis. Hopefully in the future, our findings can be combined with *in vivo* techniques leading to a complete structural and functional understanding of the meiotic prophase mechanisms.

References

Ables, E. T. (2015). Drosophila Oocytes as a model for understanding meiosis: An educational primer to accompany ???Corolla is a novel protein that contributes to the architecture of the synaptonemal complex of drosophila??? *Genetics*, *199*(1), 17–23. https://doi.org/10.1534/genetics.114.167940

Adams, P. D., Afonine, P. V, Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., … Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 2), 213–221. https://doi.org/10.1107/S0907444909052925

Adelfalk, C., Janschek, J., Revenkova, E., Blei, C., Liebe, B., Göb, E., … Jessberger, R. (2009). Cohesin SMC1beta protects telomeres in meiocytes. *The Journal of Cell Biology*, *187*(2), 185–199. https://doi.org/10.1083/jcb.200808016

Agarwal, S., & Roeder, G. S. (2000). Zip3 provides a link between recombination enzymes and synaptonemal complex proteins. *Cell*, *102*(2), 245–255. https://doi.org/10.1016/S0092-8674(00)00029-5

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2003). *Molecular biology of the cell. 4th edn.*

Albini, S. M., & Jones, G. H. (1987). Synaptonemal complex spreading in Allium cepa and A. fistulosum - I. The initiation and sequence of pairing. *Chromosoma*, *95*(5), 324–338. https://doi.org/10.1007/BF00293179

Allers, T, & Lichten, M. (2001). Differential timing and control of noncrossover and crossover recombination during meiosis. *Cell*, *106*(1), 47–57. https://doi.org/10.1016/s0092-8674(01)00416-0

Allers, Thorsten, & Lichten, M. (2001). Differential timing and control of noncrossover and crossover recombination during meiosis. *Cell*, *106*(1), 47–57. https://doi.org/10.1016/S0092-8674(01)00416-0

Alsheimer, M., Von Glasenapp, E., Hock, R., & Benavente, R. (1999). Architecture of the nuclear periphery of rat pachytene spermatocytes: Distribution of nuclear envelope proteins in relation to synaptonemal complex attachment sites. *Molecular Biology of the Cell*, *10*(4), 1235–1245. https://doi.org/10.1091/mbc.10.4.1235

Anderson, L. K., Offenberg, H. H., Verkuijlen, W. M. H. C., & Heyting, C. (1997). RecA-like proteins are components of early meiotic nodules in lily. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(13), 6868–6873. https://doi.org/10.1073/pnas.94.13.6868

Anderson, L. K., Royer, S. M., Page, S. L., McKim, K. S., Lai, A., Lilly, M. A., & Hawley, R. S. (2005). Juxtaposition of C(2)M and the transverse filament protein C(3)G within the central region of Drosophila synaptonemal complex. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(12), 4482–4487. https://doi.org/10.1073/pnas.0500172102

Arakawa, T., Ejima, D., Tsumoto, K., Obeyama, N., Tanaka, Y., Kita, Y., & Timasheff, S. N. (2007). Suppression of protein interactions by arginine: A proposed mechanism of the arginine effects. *Biophysical Chemistry*, *127*(1–2), 1–8. https://doi.org/10.1016/j.bpc.2006.12.007

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research*, *44*(W1), W344-50. https://doi.org/10.1093/nar/gkw408

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., & Ben-Tal, N. (2010). ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Research*, *38*(SUPPL. 2), 529–533. https://doi.org/10.1093/nar/gkq399

Ashley, T., Walpita, D., & De Rooij, D. G. (2001). Localization of two mammalian cyclin dependent kinases during mammalian meiosis. *Journal of Cell Science*, *114*(4), 685–693.

Bai, M., Ti, D., Mei, Q., Liu, J., Yan, X., Chen, D., … Han, W. (2020). The Role of Posttranslational Modifications in DNA Repair. *BioMed Research International*, *2020*, 7493902. https://doi.org/10.1155/2020/7493902

Barlow, P. N., Luisi, B., Milner, A., Elliott, M., & Everett, R. (1994). Structure of the C3HC4 Domain by 1H-nuclear Magnetic Resonance Spectroscopy.: A New Structural Class of Zinc-finger. *Journal of Molecular Biology*, *237*(2), 201–211. https://doi.org/https://doi.org/10.1006/jmbi.1994.1222

Baudat, F., & De Massy, B. (2007). Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. *Chromosome Research*, *15*(5), 565–577. https://doi.org/10.1007/s10577-007-1140-3

Baudat, F., Imai, Y., & De Massy, B. (2013). Meiotic recombination in mammals: Localization and regulation. *Nature Reviews Genetics*, *14*(11), 794–806. https://doi.org/10.1038/nrg3573

Berchowitz, L., & Copenhaver, G. (2010). Genetic Interference: Dont Stand So Close to Me. *Current Genomics*, *11*(2), 91–102. https://doi.org/10.2174/138920210790886835

Berríos, S., Manieu, C., López-Fenner, J., Ayarza, E., Page, J., González, M., … Fernández-Donoso, R. (2014). Robertsonian chromosomes and the nuclear architecture of mouse meiotic prophase spermatocytes. *Biological Research*, *47*(1), 1–13. https://doi.org/10.1186/0717-6287-47-16

Bhalla, N., & Dernburg, A. . (2008). Prelude to a division. *Annu Rev Cell Dev Biol*, *24*, 397–424. https://doi.org/10.1146/annurev.cellbio.23.090506.123245.Prelude

Bhalla, Needhi, Wynne, D. J., Jantsch, V., & Dernburg, A. F. (2008). ZHP-3 acts at crossovers to couple meiotic recombination with synaptonemal complex disassembly and bivalent formation in C. elegans. *PLoS Genetics*, *4*(10). https://doi.org/10.1371/journal.pgen.1000235

Bianchi, A., Smith, S., Chong, L., Elias, P., & de Lange, T. (1997). TRF1 is a dimer and bends telomeric DNA. *The EMBO Journal*, *16*(7), 1785–1794. https://doi.org/10.1093/emboj/16.7.1785

Bickel, S. E., Wyman, D. W., Miyazaki, W. Y., Moore, D. P., & Orr-Weaver, T. L. (1996). Identification of ORD, a drosophila protein essential for sister chromatid cohesion. *EMBO Journal*, *15*(6), 1451–1459. https://doi.org/10.1002/j.1460-2075.1996.tb00487.x

Billmyre, K. K., Cahoon, C. K., Matthew Heenan, G., Wesley, E. R., Yu, Z., Unruh, J. R., … Scott Hawley, R. (2019). X chromosome and autosomal recombination are differentially sensitive to disruptions in SC maintenance. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(43), 21641–21650. https://doi.org/10.1073/pnas.1910840116

Blackburn, E. H. (2000). Telomere states and cell fates. *Nature*, *408*(6808), 53–56. https://doi.org/10.1038/35040500

Blackburn, Elizabeth H., Epel, E. S., & Lin, J. (2015). Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. *Science*, *350*(6265), 1193–1198. https://doi.org/10.1126/science.aab3389

Bolcun-Filas, E., Costa, Y., Speed, R., Taggart, M., Benavente, R., De Rooij, D. G., & Cooke, H. J. (2007). SYCE2 is required for synaptonemal complex assembly, double strand break repair, and homologous recombination. *The Journal of Cell Biology*, *176*(6), 741–747.

https://doi.org/10.1083/jcb.200610027

Bolcun-Filas, E., Hall, E., Speed, R., Taggart, M., Grey, C., de Massy, B., … Cooke, H. J. (2009). Mutation of the mouse Syce1 gene disrupts synapsis and suggests a link between synaptonemal complex structural components and DNA repair. *PLoS Genetics*, *5*(2), e1000393. https://doi.org/10.1371/journal.pgen.1000393

Bomblies, K., Jones, G., Franklin, C., Zickler, D., & Kleckner, N. (2016). The challenge of evolving stable polyploidy: could an increase in "crossover interference distance" play a central role? *Chromosoma*, *125*(2), 287–300. https://doi.org/10.1007/s00412-015-0571-4

Bone, C. R., & Starr, D. A. (2016). Nuclear migration events throughout development. *Journal of Cell Science*, *129*(10), 1951–1961. https://doi.org/10.1242/jcs.179788

Börner, G. V., Kleckner, N., & Hunter, N. (2004). Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. *Cell*, *117*(1), 29–45. https://doi.org/10.1016/S0092-8674(04)00292-2

Brar, G. A., Hochwagen, A., Ee, L. S., & Amon, A. (2009). The multiple roles of cohesin in meiotic chromosome morphogenesis and pairing. *Molecular Biology of the Cell*, *20*(3), 1030–1047. https://doi.org/10.1091/mbc.e08-06-0637

Broccoli, D., Smogorzewska, A., Chong, L., & de Lange, T. (1997). Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nature Genetics*, *17*(2), 231–235. https://doi.org/10.1038/ng1097-231

Broman, K. W., Rowe, L. B., Churchill, G. A., & Paigen, K. (2002). Crossover interference in the mouse. *Genetics*, *160*(3), 1123–1131. https://doi.org/10.1093/genetics/160.3.1123

Brooks Crickard, J., Kaniecki, K., Kwon, Y., Sung, P., & Greene, E. C. (2018). Spontaneous self-segregation of Rad51 and Dmc1 DNA recombinases within mixed recombinase filaments. *Journal of Biological Chemistry*, *293*(11), 4191–4200. https://doi.org/10.1074/jbc.RA117.001143

Brzovic, P. S., Rajagopal, P., Hoyt, D. W., King, M., & Klevit, R. E. (2001). *letters Structure of a BRCA1 – BARD1 heterodimeric RING – RING complex © 2001 Nature Publishing Group http://structbio.nature.com*. 833–837.

Buetow, L., & Huang, D. T. (2016). Structural insights into the catalysis and regulation of E3 ubiquitin ligases. *Nature Reviews Molecular Cell Biology*, *17*(10), 626–642. https://doi.org/10.1038/nrm.2016.91

Burke, B., & Stewart, C. L. (2002). Life at the edge: the nuclear envelope and human disease. *Nature Reviews. Molecular Cell Biology*, *3*(8), 575–585. https://doi.org/10.1038/nrm879

Cahoon, C. K., & Hawley, R. S. (2016). Regulating the construction and demolition of the synaptonemal complex. *Nature Structural and Molecular Biology*, *23*(5), 369–377. https://doi.org/10.1038/nsmb.3208

Cahoon, C. K., Yu, Z., Wang, Y., Guo, F., Unruh, J. R., Slaughter, B. D., & Hawley, R. S. (2017). Superresolution expansion microscopy reveals the three-dimensional organization of the *Drosophila* synaptonemal complex. *Proceedings of the National Academy of Sciences*, 201705623. https://doi.org/10.1073/pnas.1705623114

Chasapis, C. T., Loutsidou, A. K., Orkoula, M. G., & Spyroulias, G. A. (2010). Zinc binding properties of engineered ring finger domain of arkadia e3 ubiquitin ligase. *Bioinorganic Chemistry and Applications*, *2010*. https://doi.org/10.1155/2010/323152

Chelysheva, L., Vezon, D., Chambon, A., Gendrot, G., Pereira, L., Lemhemdi, A., … Grelon, M. (2012). The Arabidopsis HEI10 is a new ZMM protein related to Zip3. *PLoS Genetics*, *8*(7).

https://doi.org/10.1371/journal.pgen.1002799

Chen, Yanyan, Wang, Y., Chen, J., Zuo, W., Fan, Y., Huang, S., … Lei, M. (2021). The SUN1-SPDYA interaction plays an essential role in meiosis prophase I. *Nature Communications*, *12*(1), 3176. https://doi.org/10.1038/s41467-021-23550-w

Chen, Yong, Yang, Y., Van Overbeek, M., Donigian, J. R., Baciu, P., De Lange, T., & Lei, M. (2008). A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins. *Science*, *319*(5866), 1092–1096. https://doi.org/10.1126/science.1151804

Cheng, C. H., Lo, Y. H., Liang, S. S., Ti, S. C., Lin, F. M., Yeh, C. H., … Wang, T. F. (2006). SUMO modifications control assembly of synaptonemal complex and polycomplex in meiosis of Saccharomyces cerevisiae. *Genes and Development*, *20*(15), 2067–2081. https://doi.org/10.1101/gad.1430406

Chikashige, Y, Ding, D. Q., Funabiki, H., Haraguchi, T., Mashiko, S., Yanagida, M., & Hiraoka, Y. (1994). Telomere-led premeiotic chromosome movement in fission yeast. *Science*, *264*(5156), 270 LP – 273. https://doi.org/10.1126/science.8146661

Chikashige, Yuji, Haraguchi, T., & Hiraoka, Y. (2007). Another way to move chromosomes. *Chromosoma*, *116*(6), 497–505. https://doi.org/10.1007/s00412-007-0114-8

Chikashige, Yuji, & Hiraoka, Y. (2001). Telomere binding of the Rap1 protein is required for meiosis in fission yeast. *Current Biology*, *11*(20), 1618–1623. https://doi.org/10.1016/S0960-9822(01)00457-2

Chikashige, Yuji, Tsutsumi, C., Yamane, M., Okamasa, K., Haraguchi, T., & Hiraoka, Y. (2006). Meiotic Proteins Bqt1 and Bqt2 Tether Telomeres to Form the Bouquet Arrangement of Chromosomes. *Cell*, *125*(1), 59–69. https://doi.org/10.1016/j.cell.2006.01.048

Chong, L., van Steensel, B., Broccoli, D., Erdjument-Bromage, H., Hanish, J., Tempst, P., & de Lange, T. (1995). A human telomeric protein. *Science (New York, N.Y.)*, *270*(5242), 1663–1667. https://doi.org/10.1126/science.270.5242.1663

Codina-Pascual, M., Campillo, M., Kraus, J., Speicher, M. R., Egozcue, J., Navarro, J., & Benet, J. (2006). Crossover frequency and synaptonemal complex length: Their variability and effects on human male meiosis. *Molecular Human Reproduction*, *12*(2), 123–133. https://doi.org/10.1093/molehr/gal007

Cohn, M., McEachern, M. J., & Blackburn, E. H. (1998). Telomeric sequence diversity within the genus Saccharomyces. *Current Genetics*, *33*(2), 83–91. https://doi.org/10.1007/s002940050312

Colaiácovo, M. P., MacQueen, A. J., Martinez-Perez, E., McDonald, K., Adamo, A., La Volpe, A., & Villeneuve, A. M. (2003). Synaptonemal complex assembly in C. elegans is dispensable for loading strand-exchange proteins but critical for proper completion of recombination. *Developmental Cell*, *5*(3), 463–474. https://doi.org/10.1016/S1534-5807(03)00232-6

Cole, C., Barber, J. D., & Barton, G. J. (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Research*, *36*(Web Server issue), W197-201. https://doi.org/10.1093/nar/gkn238

Collins, K. A., Unruh, J. R., Slaughter, B. D., Yu, Z., Lake, C. M., Nielsen, R. J., … Scott Hawley, R. (2014). Corolla is a novel protein that contributes to the architecture of the synaptonemal complex of Drosophila. *Genetics*, *198*(1), 219–228. https://doi.org/10.1534/genetics.114.165290

Collins, K., Kobayashi, R., & Greider, C. W. (1995). Purification of Tetrahymena telomerase and cloning of genes encoding the two protein components of the enzyme. *Cell*, *81*(5), 677–686. https://doi.org/10.1016/0092-8674(95)90529-4

Cong, Y.-S., Wright, W. E., & Shay, J. W. (2002). Human telomerase and its regulation. *Microbiology and Molecular Biology Reviews : MMBR*, *66*(3), 407–425, table of contents.

https://doi.org/10.1128/mmbr.66.3.407-425.2002

Conrad, M. N., Lee, C. Y., Chao, G., Shinohara, M., Kosaka, H., Shinohara, A., … Dresser, M. E. (2008). Rapid Telomere Movement in Meiotic Prophase Is Promoted By NDJ1, MPS3, and CSM4 and Is Modulated by Recombination. *Cell*, *133*(7), 1175–1187. https://doi.org/10.1016/j.cell.2008.04.047

Cooper, G. M. (2000). *The Cell, 2nd edition*.

Cooper, J. P., Watanabe, Y., & Nurse, P. (1998). Fission yeast Taz1 protein is required for meiotic telomere clustering and recombination. *Nature*, *392*(6678), 828–831. https://doi.org/10.1038/33947

Costa, Y., & Cooke, H. J. (2007). Dissecting the mammalian synaptonemal complex using targeted mutations. *Chromosome Research*, *15*(5), 579–589. https://doi.org/10.1007/s10577-007-1142-1

Court, R., Chapman, L., Fairall, L., & Rhodes, D. (2005). How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: A view from high-resolution crystal structures. *EMBO Reports*, *6*(1), 39–45. https://doi.org/10.1038/sj.embor.7400314

CRICK, F. H. (1952). Is alpha-keratin a coiled coil? *Nature*, *170*(882), 3.

Crisp, M., Liu, Q., Roux, K., Rattner, J. B., Shanahan, C., Burke, B., … Hodzic, D. (2006). Coupling of the nucleus and cytoplasm: role of the LINC complex. *The Journal of Cell Biology*, *172*(1), 41–53. https://doi.org/10.1083/jcb.200509124

Daniel, K., Tränkner, D., Wojtasz, L., Shibuya, H., Watanabe, Y., Alsheimer, M., & Tóth, A. (2014). Mouse CCDC79 (TERB1) is a meiosis-specific telomere associated protein. *BMC Cell Biology*, *15*(1), 1–12. https://doi.org/10.1186/1471-2121-15-17

Davies, O. R., Forment, J. V., Sun, M., Belotserkovskaya, R., Coates, J., Galanty, Y., … Pellegrini, L. (2015). CtIP tetramer assembly is required for DNA-end resection and repair. *Nature Structural and Molecular Biology*, *22*(2), 150–157. https://doi.org/10.1038/nsmb.2937

Davies, O. R., Maman, J. D., & Pellegrini, L. (2012). Structural analysis of the human SYCE2-TEX12 complex provides molecular insights into synaptonemal complex assembly. *Open Biology*, *2*(JULY). https://doi.org/10.1098/rsob.120099

de Boer, E., & Heyting, C. (2006). The diverse roles of transverse filaments of synaptonemal complexes in meiosis. *Chromosoma*, *115*(3), 220–234. https://doi.org/10.1007/s00412-006-0057-5

De Lange, T. (2004). T-loops and the origin of telomeres. *Nature Reviews Molecular Cell Biology*, *5*(4), 323–329. https://doi.org/10.1038/nrm1359

De Lange, T. (2005). Shelterin: The protein complex that shapes and safeguards human telomeres. *Genes and Development*, *19*(18), 2100–2110. https://doi.org/10.1101/gad.1346005

De los Santos, T., Hunter, N., Lee, C., Larkin, B., Loidl, J., & Hollingsworth, N. M. (2003). The MUS81/MMS4 endonuclease acts independently of double-holliday junction resolution to promote a distinct subset of crossovers during meiosis in budding yeast. *Genetics*, *164*(1), 81–94. https://doi.org/10.1093/genetics/164.1.81

De Muyt, A., Pyatnitskaya, A., Andréani, J., Ranjha, L., Ramus, C., Laureau, R., … Borde, V. (2018). A meiotic XPF–ERCC1-like complex recognizes joint molecule recombination intermediates to promote crossover formation. *Genes and Development*, *32*(3–4), 283–296. https://doi.org/10.1101/gad.308510.117

De Muyt, A., Zhang, L., Piolot, T., Kleckner, N., Espagne, E., & Zickler, D. (2014). E3 ligase Hei10: A multifaceted structure-based signaling molecule with roles within and beyond meiosis. *Genes and Development*, *28*(10), 1111–1123. https://doi.org/10.1101/gad.240408.114

Del Cacho, E., Pages, M., Gallego, M., Monteagudo, L., & Sánchez-Acedo, C. (2005). Synaptonemal complex karyotype of Eimeria tenella. *International Journal for Parasitology*, *35*(13), 1445–1451. https://doi.org/10.1016/j.ijpara.2005.06.009

Deng, W., Wu, J., Wang, F., Kanoh, J., Dehe, P.-M., Inoue, H., … Lei, M. (2015, July). Fission yeast telomere-binding protein Taz1 is a functional but not a structural counterpart of human TRF1 and TRF2. *Cell Research*, Vol. 25, pp. 881–884. https://doi.org/10.1038/cr.2015.76

Deshaies, R. J., & Joazeiro, C. A. P. (2009). RING domain E3 ubiquitin ligases. *Annual Review of Biochemistry*, *78*, 399–434. https://doi.org/10.1146/annurev.biochem.78.101807.093809

Diotti, R., & Loayza, D. (2011). Shelterin complex and associated factors at human telomeres. *Nucleus*, *2*(2), 119–135. https://doi.org/10.4161/nucl.2.2.15135

Dong, H., & Roeder, G. S. (2000). Organization of the yeast Zip1 protein within the central region of the synaptonemal complex. *Journal of Cell Biology*, *148*(3), 417–426. https://doi.org/10.1083/jcb.148.3.417

Dou, H., Buetow, L., Sibbet, G. J., Cameron, K., & Huang, D. T. (2014). *Europe PMC Funders Group Structure of BIRC7 – E2 ubiquitin conjugate reveals the mechanism of ubiquitin transfer by a RING dimer*. *19*(9). https://doi.org/10.1038/nsmb.2379.Structure

Drozdetskiy, A., Cole, C., Procter, J., & Barton, G. J. (2015). JPred4: a protein secondary structure prediction server. *Nucleic Acids Research*, *43*(W1), W389-94. https://doi.org/10.1093/nar/gkv332

Dunce, J., Dunne, O., Ratcliff, M., Millán, C., Madgwick, S., & Usón, I. et al. (2018). Structural basis of meiotic chromosome synapsis through SYCP1 self-assembly. *Nat Struct. Mol. Biol*, *25*(7), 557–569. https://doi.org/10.1038/s41594-018-0078-9.Structural

Dunce, J. M., Dunne, O. M., Ratcliff, M., Millán, C., Madgwick, S., Usón, I., & Davies, O. R. (2018). Structural basis of meiotic chromosome synapsis through SYCP1 self-assembly. *Nature Structural and Molecular Biology*, *25*(July), 1–13. https://doi.org/10.1038/s41594-018-0078-9

Dunce, J. M., Milburn, A. E., Gurusaran, M., da Cruz, I., Sen, L. T., Benavente, R., & Davies, O. R. (2018). Structural basis of meiotic telomere attachment to the nuclear envelope by MAJIN-TERB2-TERB1. *Nature Communications*, *9*(1), 1–18. https://doi.org/10.1038/s41467-018-07794-7

Dunce, J. M., Salmon, L. J., & Davies, O. R. (2020). Structural basis of meiotic chromosome synaptic elongation through hierarchical fibrous assembly of SYCE2-TEX12. *BioRxiv*, 2020.12.30.424799. https://doi.org/10.1101/2020.12.30.424799

Dunne, O. M., & Davies, O. R. (2019). Molecular structure of human synaptonemal complex protein SYCE1. *Chromosoma*, *128*(3), 223–236. https://doi.org/10.1007/s00412-018-00688-z

Dye, B. T., & Schulman, B. A. (2007). Structural mechanisms underlying posttranslational modification by ubiquitin-like proteins. *Annual Review of Biophysics and Biomolecular Structure*, *36*, 131–150. https://doi.org/10.1146/annurev.biophys.36.040306.132820

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Eifler, K., & Vertegaal, A. C. O. (2015). SUMOylation-mediated regulation of cell cycle progression and cancer SUMO: a ubiquitin-like modifier that regulates nuclear processes. *Trends in Biochemical Science*, *40*(12), 779–793. https://doi.org/10.1016/j.tibs.2015.09.006.SUMOylation-mediated

Eijpe, M., Heyting, C., Gross, B., & Jessberger, R. (2000). Association of mammalian SMC1 and SMC3 proteins with meiotic chromosomes and synaptonemal complexes. *Journal of Cell*

Science, *113 (Pt 4*, 673–682.

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica. Section D, Biological Crystallography*, *60*(Pt 12 Pt 1), 2126–2132. https://doi.org/10.1107/S0907444904019158

Engel, A. M., Cejka, Z., Lupas, A., Lottspeich, F., & Baumeister, W. (1992). Isolation and cloning of Ompα, a coiled-coil protein spanning the periplasmic space of the ancestral eubacterium Thermotoga maritima. *EMBO Journal*, *11*(12), 4369–4378. https://doi.org/10.1002/j.1460-2075.1992.tb05537.x

Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallographica. Section D, Biological Crystallography*, *62*(Pt 1), 72–82. https://doi.org/10.1107/S0907444905036693

Fairall, L., Chapman, L., Moss, H., De Lange, T., & Rhodes, D. (2001a). Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2. *Molecular Cell*, *8*(2), 351–361. https://doi.org/10.1016/S1097-2765(01)00321-5

Fairall, L., Chapman, L., Moss, H., De Lange, T., & Rhodes, D. (2001b). Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2. *Molecular Cell*, *8*(2), 351–361. https://doi.org/10.1016/S1097-2765(01)00321-5

FAWCETT, D. W. (1956). The fine structure of chromosomes in the meiotic prophase of vertebrate spermatocytes. *The Journal of Biophysical and Biochemical Cytology*, *2*(4), 403–406. https://doi.org/10.1083/jcb.2.4.403

Feng, J., Fu, S., Cao, X., Wu, H., Lu, J., Zeng, M., … Shen, Y. (2017). Synaptonemal complex protein 2 (SYCP2) mediates the association of the centromere with the synaptonemal complex. *Protein and Cell*, *8*(7), 538–543. https://doi.org/10.1007/s13238-016-0354-6

Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., … Svergun, D. I. (2017). ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *Journal of Applied Crystallography*, *50*, 1212–1225. https://doi.org/10.1107/S1600576717007786

Franke, Daniel, & Svergun, D. I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography*, *42*(2), 342–346. https://doi.org/10.1107/S0021889809000338

Fraune, J., Alsheimer, M., Volff, J. N., Busch, K., Fraune, S., Bosch, T. C. G., & Benavente, R. (2012). Hydra meiosis reveals unexpected conservation of structural synaptonemal complex proteins across metazoans. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(41), 16588–16593. https://doi.org/10.1073/pnas.1206875109

Fraune, J., Brochier-Armanet, C., Alsheimer, M., & Benavente, R. (2013). Phylogenies of central element proteins reveal the dynamic evolutionary history of the mammalian synaptonemal complex: Ancient and recent components. *Genetics*, *195*(3), 781–793. https://doi.org/10.1534/genetics.113.156679

Freemont, P. S. (2000). Ubiquitination: RING for destruction? *Current Biology*, *10*(2), 84–87. https://doi.org/10.1016/s0960-9822(00)00287-6

Freemont, P. S., Hanson, I. M., & Trowsdale, J. (1991). A novel gysteine-rich sequence motif. *Cell*, *64*(3), 483–484. https://doi.org/10.1016/0092-8674(91)90229-R

Garcia-Barcena, C., Osinalde, N., Ramirez, J., & Mayor, U. (2020). How to Inactivate Human Ubiquitin E3 Ligases by Mutation. *Frontiers in Cell and Developmental Biology*, *8*(February), 1–14. https://doi.org/10.3389/fcell.2020.00039

Gareau, J. R. & Lima, C. D. 2010. (2010). The SUMO pathway: emerging mechanisms that shape

specificity, conjugation and recognition. *Nat Rev Mol Cell Biol*, *11*, 861–871. https://doi.org/10.1111/j.1470-8744.1991.tb00167.x

Golubovskaya, I. N., Harper, L. C., Pawlowski, W. P., Schichnes, D., & Cande, W. Z. (2002). The pam1 gene is required for meiotic bouquet formation and efficient homologous synapsis in maize (Zea mays L.). *Genetics*, *162*(4), 1979–1993. https://doi.org/10.1093/genetics/162.4.1979

Gómez-H, L., Felipe-Medina, N., Sánchez-Martín, M., Davies, O. R., Ramos, I., García-Tuñón, I., … Pendas, A. M. (2016). C14ORF39/SIX6OS1 is a constituent of the synaptonemal complex and is essential for mouse fertility. *Nature Communications*, *7*, 13298. https://doi.org/10.1038/ncomms13298

Gowen, Marie S . Gowen, J. W. . G. (1922). Complete Linkage in Drosophila melanogaster. *The America Naturalist*, *56*(644), 286–288.

Gray, S., & Cohen, P. E. (2016). Control of Meiotic Crossovers: From Double-Strand Break Formation to Designation. *Annu Rev Genet*, *23*(50), 175–210. https://doi.org/10.1146/annurev-genet-120215-035111.Control

Gray, S., Santiago, E. R., Chappie, J. S., & Cohen, P. E. (2020). Cyclin N-Terminal Domain-Containing-1 Coordinates Meiotic Crossover Formation with Cell-Cycle Progression in a Cyclin-Independent Manner. *Cell Reports*, *32*(1), 107858. https://doi.org/10.1016/j.celrep.2020.107858

Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H., & De Lange, T. (1999). Mammalian telomeres end in a large duplex loop. *Cell*, *97*(4), 503–514. https://doi.org/10.1016/S0092-8674(00)80760-6

Grishaeva, T. M., & Bogdanov, Y. F. (2014). Conservation and Variability of Synaptonemal Complex Proteins in Phylogenesis of Eukaryotes. *International Journal of Evolutionary Biology*, *2014*, 1–16. https://doi.org/10.1155/2014/856230

Gruber, S., Haering, C. H., & Nasmyth, K. (2003). Chromosomal cohesin forms a ring. *Cell*, *112*(6), 765–777. https://doi.org/10.1016/S0092-8674(03)00162-4

Guillon, H., Baudat, F., Grey, C., Liskay, R. M., & De Massy, B. (2005). Crossover and noncrossover pathways in mouse meiosis. *Molecular Cell*, *20*(4), 563–573. https://doi.org/10.1016/j.molcel.2005.09.021

Gutiérrez-Caballero, C., Herrán, Y., Sánchez-Martín, M., Suja, J. Á., Barbero, J. L., Llano, E., & Pendás, A. M. (2011). Identification and molecular characterization of the mammalian α-kleisin RAD21L. *Cell Cycle*, *10*(9), 1477–1487. https://doi.org/10.4161/cc.10.9.15515

Hagan, I., & Yanagida, M. (1995). The product of the spindle formation gene sad1+ associates with the fission yeast spindle pole body and is essential for viability. *The Journal of Cell Biology*, *129*(4), 1033–1047. https://doi.org/10.1083/jcb.129.4.1033

Hamer, G., Gell, K., Kouznetsova, A., Novak, I., Benavente, R., & Höög, C. (2006). Characterization of a novel meiosis-specific protein within the central element of the synaptonemal complex. *Journal of Cell Science*, *119*(19), 4025–4032. https://doi.org/10.1242/jcs.03182

Hanaoka, S., Nagadoi, A., & Nishimura, Y. (2005). Comparison between TRF2 and TRF1 of their telomeric DNA-bound structures and DNA-binding activities. *Protein Science : A Publication of the Protein Society*, *14*(1), 119–130. https://doi.org/10.1110/ps.04983705

Haque, F., Lloyd, D. J., Smallwood, D. T., Dent, C. L., Shanahan, C. M., Fry, A. M., … Shackleton, S. (2006). SUN1 interacts with nuclear lamin A and cytoplasmic nesprins to provide a physical connection between the nuclear lamina and the cytoskeleton. *Molecular and Cellular Biology*, *26*(10), 3738–3751. https://doi.org/10.1128/MCB.26.10.3738-3751.2006

Harper, L. (2004). A bouquet of chromosomes. *Journal of Cell Science*, *117*(18), 4025–4032. https://doi.org/10.1242/jcs.01363

Hassold, T., & Hunt, P. (2001). To err (meiotically) is human: the genesis of human aneuploidy. *Nature Reviews. Genetics*, *2*(4), 280–291. https://doi.org/10.1038/35066065

Hayashi, M., Mlynarczyk-Evans, S., & Villeneuve, A. M. (2010). The synaptonemal complex shapes the crossover landscape through cooperative assembly, crossover promotion and crossover inhibition during Caenorhabditis elegans meiosis. *Genetics*, *186*(1), 45–58. https://doi.org/10.1534/genetics.110.115501

Heinrich, B. (1974). Thermoregulation in endothermic insects. *Science*, *185*(4153), 747–756. https://doi.org/10.1126/science.185.4153.747

Hemmer, L. W., & Blumenstiel, J. P. (2016). Holding it together: Rapid evolution and positive selection in the synaptonemal complex of Drosophila. *BMC Evolutionary Biology*, *16*(1), 1–17. https://doi.org/10.1186/s12862-016-0670-8

Hemmer, L. W., Blumenstiel, J. P., Alsheimer, M., Von Glasenapp, E., Hock, R., Benavente, R., … Roeder, G. S. (2014). The Dissection of Meiotic Chromosome Movement in Mice Using an In Vivo Electroporation Technique. *Annual Review of Genetics*, *10*(1), 1611–1622. https://doi.org/10.1073/pnas.1414814112

Henderson, K. A., & Keeney, S. (2004). Tying synaptonemal complex initiation to the formation and programmed repair of DNA double-strand breaks. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(13), 4519–4524. https://doi.org/10.1073/pnas.0400843101

Herbert, M., Kalleas, D., Cooney, D., Lamb, M., & Lister, L. (2015). Meiosis and maternal aging: insights from aneuploid oocytes and trisomy births. *Cold Spring Harbor Perspectives in Biology*, *7*(4), a017970. https://doi.org/10.1101/cshperspect.a017970

Hernández-Hernández, A., Masich, S., Fukuda, T., Kouznetsova, A., Sandin, S., Daneholt, B., & Höög, C. (2016). The central element of the synaptonemal complex in mice is organized as a bilayered junction structure. *Journal of Cell Science*, *129*(11), 2239–2249. https://doi.org/10.1242/jcs.182477

Hiraoka, Y., & Dernburg, A. F. (2009). The SUN Rises on Meiotic Chromosome Dynamics. *Developmental Cell*, *17*(5), 598–605. https://doi.org/10.1016/j.devcel.2009.10.014

Holloway, J. K., Sun, X., Yokoo, R., Villeneuve, A. M., & Cohen, P. E. (2014). Mammalian CNTD1 is critical for meiotic crossover maturation and deselection of excess precrossover sites. *Journal of Cell Biology*, *205*(5), 633–641. https://doi.org/10.1083/jcb.201401122

Hooker, G. W., & Roeder, G. S. (2006). A Role for SUMO in Meiotic Chromosome Synapsis. *Current Biology*, *16*(12), 1238–1243. https://doi.org/10.1016/j.cub.2006.04.045

Horn, H. F., Kim, D. I., Wright, G. D., Wong, E. S. M., Stewart, C. L., Burke, B., & Roux, K. J. (2013). A mammalian KASH domain protein coupling meiotic chromosomes to the cytoskeleton. *Journal of Cell Biology*, *202*(7), 1023–1039. https://doi.org/10.1083/jcb.201304004

Hu, C., Rai, R., Huang, C., Broton, C., Long, J., Xu, Y., … Chen, Y. (2017). Structural and functional analyses of the mammalian TIN2-TPP1-TRF2 telomeric complex. *Cell Research*, *27*(12), 1485–1502. https://doi.org/10.1038/cr.2017.144

Hughes, S. E., Miller, D. E., Miller, A. L., & Hawley, R. S. (2018). Female meiosis: Synapsis, recombination, and segregation in Drosophila melanogaster. In *Genetics* (Vol. 208). https://doi.org/10.1534/genetics.117.300081

Hunter, N. (2015a). Meiotic recombination: The essence of heredity. *Cold Spring Harbor Perspectives in Biology*, *7*(12), 1–36. https://doi.org/10.1101/cshperspect.a016618

Hunter, N. (2015b). Meiotic recombination: The essence of heredity. *Cold Spring Harbor Perspectives in Biology*, *7*(12). https://doi.org/10.1101/cshperspect.a016618

Hunter, N., & Kleckner, N. (2001). The single-end invasion: An asymmetric intermediate at the double-strand break to double-holliday junction transition of meiotic recombination. *Cell*, *106*(1), 59–70. https://doi.org/10.1016/S0092-8674(01)00430-5

Ishiguro, K. I., Kim, J., Fujiyama-Nakamura, S., Kato, S., & Watanabe, Y. (2011). A new meiosis-specific cohesin complex implicated in the cohesin code for homologous pairing. *EMBO Reports*, *12*(3), 267–275. https://doi.org/10.1038/embor.2011.2

Jang, J. K., Sherizen, D. E., Bhagat, R., Manheim, E. A., & McKim, K. S. (2003). Relationship of DNA double-strand breaks to synapsis in Drosophila. *Journal of Cell Science*, *116*(15), 3069–3077. https://doi.org/10.1242/jcs.00614

Jeffress, J. K., Page, S. L., Royer, S. M., Belden, E. D., Blumenstiel, J. P., Anderson, L. K., & Hawley, R. S. (2007). The formation of the central element of the synaptonemal complex may occur by multiple mechanisms: The roles of the N- and C-terminal domains of the Drosophila C(3)G protein in mediating synapsis and recombination. *Genetics*, *177*(4), 2445–2456. https://doi.org/10.1534/genetics.107.078717

Jeong, J. Y., Yim, H. S., Ryu, J. Y., Lee, H. S., Lee, J. H., Seen, D. S., & Kang, S. G. (2012). One-step sequence-and ligation-independent cloning as a rapid and versatile cloning method for functional genomics Studies. *Applied and Environmental Microbiology*, *78*(15), 5440–5443. https://doi.org/10.1128/AEM.00844-12

Johnston, S. E., Stoffel, M. A., & Pemberton, J. M. (2020). Variants at RNF212 and RNF212B are associated with recombination rate variation in Soay sheep (Ovis aries). *BioRxiv*, 1–23. https://doi.org/10.1101/2020.07.26.217802

Josefsberg, L. B. Y., & Dekel, N. (2002). Translational and post-translational modifications in meiosis of the mammalian oocyte. *Molecular and Cellular Endocrinology*, *187*(1–2), 161–171. https://doi.org/10.1016/S0303-7207(01)00688-8

Kabsch, W. (2010). XDS. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 2), 125–132. https://doi.org/10.1107/S0907444909047337

Kadri, N. K., Harland, C., Faux, P., Cambisano, N., Karim, L., Coppieters, W., … Druet, T. (2016). Coding and noncoding variants in HFM1, MLH3, MSH4, MSH5, RNF212, and RNF212B affect recombination rate in cattle. *Genome Research*, *26*(10), 1323–1332. https://doi.org/10.1101/gr.204214.116

Kanoh, J., & Ishikawa, F. (2001). spRap1 and spRif1, recruited to telomeres by Taz1, are essential for telomere function in fission yeast. *Current Biology : CB*, *11*(20), 1624–1630. https://doi.org/10.1016/s0960-9822(01)00503-6

Karlseder, J., Kachatrian, L., Takai, H., Mercer, K., Hingorani, S., Jacks, T., & de Lange, T. (2003). Targeted Deletion Reveals an Essential Function for the Telomere Length Regulator Trf1. *Molecular and Cellular Biology*, *23*(18), 6533–6541. https://doi.org/10.1128/mcb.23.18.6533-6541.2003

Keeney, S. (2008). Spo11 and the formation of DNA double-strand breaks in meiosis. *Genome Dynamics and Stability*, *2*, 81–123. https://doi.org/10.1007/7050_2007_026

Keeney, S., Giroux, C. N., & Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell*, *88*(3), 375–384. https://doi.org/10.1016/S0092-8674(00)81876-0

Kelly, S. M., Jess, T. J., & Price, N. C. (2005). How to study proteins by circular dichroism. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1751*(2), 119–139. https://doi.org/10.1016/j.bbapap.2005.06.005

Kerscher, O., Felberbaum, R., & Hochstrasser, M. (2006). Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annual Review of Cell and Developmental Biology*, *22*, 159–180. https://doi.org/10.1146/annurev.cellbio.22.010605.093503

Kezer, J., Sessions, S. K., & León, P. (1989). The meiotic structure and behavior of the strongly heteromorphic X/Y sex chromosomes of neotropical plethodontid salamanders of the genus Oedipina. *Chromosoma*, *98*(6), 433–442. https://doi.org/10.1007/BF00292789

Klein, F., Mahr, P., Galova, M., Buonomo, S. B. C., Michaelis, C., Nairz, K., & Nasmyth, K. (1999). A central role for cohesins in sister chromatid cohesion, formation of axial elements, and recombination during yeast meiosis. *Cell*, *98*(1), 91–103. https://doi.org/10.1016/S0092-8674(00)80609-1

Kneitz, B., Cohen, P. E., Avdievich, E., Zhu, L., Kane, M. F., Hou, H. J., … Edelmann, W. (2000). MutS homolog 4 localization to meiotic chromosomes is required for chromosome pairing during meiosis in male and female mice. *Genes & Development*, *14*(9), 1085–1097.

Kohli, J. (1994). Meiosis. Telomeres lead chromosome movement. *Current Biology : CB*, *4*(8), 724–727. https://doi.org/10.1016/s0960-9822(00)00160-3

Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., & Svergun, D. I. (2003). PRIMUS: A Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography*, *36*(5), 1277–1282. https://doi.org/10.1107/S0021889803012779

Kong, A., Thorleifsson, G., Stefansson, H., Masson, G., Helgason, A., Gudbjartsson, D. F., … Stefansson, K. (2008). Sequence variants in the RNF212 gene associate with genome-wide recombination rate. *Science (New York, N.Y.)*, *319*(5868), 1398–1401. https://doi.org/10.1126/science.1152422

König, P., & Rhodes, D. (1997). Recognition of telomeric DNA. *Trends in Biochemical Sciences*, *22*(2), 43–47. https://doi.org/10.1016/s0968-0004(97)01008-6

Koonin, E. V. (2005). Orthologs, paralogs, and evolutionary genomics. *Annual Review of Genetics*, *39*, 309–338. https://doi.org/10.1146/annurev.genet.39.073003.114725

Kornberg, R. D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science*, *184*, 868–871.

Kouznetsova, A., Novak, I., Jessberger, R., & Höög, C. (2005). SYCP2 and SYCP3 are required for cohesin core integrity at diplotene but not for centromere cohesion at the first meiotic division. *Journal of Cell Science*, *118*(10), 2271–2278. https://doi.org/10.1242/jcs.02362

Kozin, M. B., & Svergun, D. I. (2001). Automated matching of high- and low-resolution structural models. *Journal of Applied Crystallography*, *34*(1), 33–41. https://doi.org/10.1107/S0021889800014126

Kracklauer, M., Link, J., & Alsheimer, M. (2013). LINCing the nuclear envelope to gametogenesis. *Curr Top Dev Biol*, *102*, 127–157.

Kulkarni, D. S., Owens, S. N., Honda, M., Ito, M., Yang, Y., Corrigan, M. W., … Hunter, N. (2020). PCNA activates the MutLγ endonuclease to promote meiotic crossing over. *Nature*, *586*(7830), 623–627. https://doi.org/10.1038/s41586-020-2645-6

Lake, C., & Hawley, R. (2012). The molecular control of meiotic chromosomal behavior: events in early meiotic prophase in Drosophila oocytes. *Annu Rev Physiol*, *74*, 425–451.

Lake, C. M., Nielsen, R. J., Bonner, A. M., Eche, S., White-Brown, S., McKim, K. S., & Hawley, R.

S. (2019). Narya, a RING finger domain-containing protein, is required for meiotic DNA double-strand break formation and crossover maturation in Drosophila melanogaster. *PLoS Genetics*, *15*(1), 1–27. https://doi.org/10.1371/journal.pgen.1007886

Lake, C. M., Nielsen, R. J., Guo, F., Unruh, J. R., Slaughter, B. D., & Hawley, R. S. (2015a). Vilya, a component of the recombination nodule, is required for meiotic double-strand break formation in Drosophila. *ELife*, *4*(OCTOBER2015), 1–26. https://doi.org/10.7554/eLife.08287

Lake, C. M., Nielsen, R. J., Guo, F., Unruh, J. R., Slaughter, B. D., & Hawley, R. S. (2015b). Vilya, a component of the recombination nodule, is required for meiotic double-strand break formation in Drosophila. *ELife*, *4*(OCTOBER2015), 1–26. https://doi.org/10.7554/eLife.08287

Lam, I., & Keeney, S. (2015). Mechanism and regulation of meiotic recombination initiation. *Cold Spring Harbor Perspectives in Biology*, *7*(1). https://doi.org/10.1101/cshperspect.a016634

Lane, S., & Kauppi, L. (2019). Meiotic spindle assembly checkpoint and aneuploidy in males versus females. *Cellular and Molecular Life Sciences : CMLS*, *76*(6), 1135–1150. https://doi.org/10.1007/s00018-018-2986-6

Lao JP, Oh SD, Shinohara M, Shinohara A, H. N. (2008). Rad52 promotes post-invasion steps of meiotic double-strand_break repair. *Mol Cell*, *29*(4), 517–524.

Lee, C.-Y., Horn, H. F., Stewart, C. L., Burke, B., Bolcun-Filas, E., Schimenti, J. C., … Pezza, R. J. (2015). Mechanism and regulation of rapid telomere prophase movements in mouse meiotic chromosomes. *Cell Reports*, *11*(4), 551–563. https://doi.org/10.1016/j.celrep.2015.03.045

Lee, C. Y., Conrad, M. N., & Dresser, M. E. (2012). Meiotic chromosome pairing is promoted by telomere-led chromosome movements independent of bouquet formation. *PLoS Genetics*, *8*(5). https://doi.org/10.1371/journal.pgen.1002730

Lee, C. Y., Horn, H. F., Stewart, C. L., Burke, B., Bolcun-Filas, E., Schimenti, J. C., … Pezza, R. J. (2015). Mechanism and Regulation of Rapid Telomere Prophase Movements in Mouse Meiotic Chromosomes. *Cell Reports*, *11*(4), 551–563. https://doi.org/10.1016/j.celrep.2015.03.045

Lee, J., & Hirano, T. (2011). RAD21L, a novel cohesin subunit implicated in linking homologous chromosomes in mammalian meiosis. *Journal of Cell Biology*, *192*(2), 263–276. https://doi.org/10.1083/jcb.201008005

Li, B., Oestreich, S., & de Lange, T. (2000). Identification of human Rap1: implications for telomere evolution. *Cell*, *101*(5), 471–483. https://doi.org/10.1016/s0092-8674(00)80858-2

Li, Y., Qin, B., Shen, Y., Zhang, F., Liu, C., You, H., … Cheng, Z. (2018). HEIP1 regulates crossover formation during meiosis in rice. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(42), 10810–10815. https://doi.org/10.1073/pnas.1807871115

Liebe, B., Alsheimer, M., Höög, C., Benavente, R., & Scherthan, H. (2004a). Telomere attachment, meiotic chromosome condensation, pairing, and bouquet stage duration are modified in spermatocytes lacking axial elements. *Molecular Biology of the Cell*, *15*(2), 827–837. https://doi.org/10.1091/mbc.e03-07-0524

Liebe, B., Alsheimer, M., Höög, C., Benavente, R., & Scherthan, H. (2004b). Telomere attachment, meiotic chromosome condensation, pairing, and bouquet stage duration are modified in spermatocytes lacking axial elements. *Molecular Biology of the Cell*, *15*(2), 827–837. https://doi.org/10.1091/mbc.E03-07-0524

Lim, C. J., Zaug, A. J., Kim, H. J., & Cech, T. R. (2017). Reconstitution of human shelterin complexes reveals unexpected stoichiometry and dual pathways to enhance telomerase processivity. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-01313-w

Lindsley, D. L., & Sandler, L. (1977). The genetic analysis of meiosis in female Drosophila

melanogaster. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *277*(955), 295–312. https://doi.org/10.1098/rstb.1977.0019

Link, J., & Jantsch, V. (2019). Meiotic chromosomes in motion: a perspective from Mus musculus and Caenorhabditis elegans. *Chromosoma*, *128*(3), 317–330. https://doi.org/10.1007/s00412-019-00698-5

Link, J., Leubner, M., Schmitt, J., Göb, E., Benavente, R., Jeang, K. T., … Alsheimer, M. (2014). Analysis of Meiosis in SUN1 Deficient Mice Reveals a Distinct Role of SUN2 in Mammalian Meiotic LINC Complex Formation and Function. *PLoS Genetics*, *10*(2), 14–21. https://doi.org/10.1371/journal.pgen.1004099

Liu, D., O'Connor, M. S., Qin, J., & Songyang, Z. (2004). Telosome, a mammalian telomere-associated complex formed by multiple telomeric proteins. *Journal of Biological Chemistry*, *279*(49), 51338–51342. https://doi.org/10.1074/jbc.M409293200

Liu, J. G., Yuan, L., Brundell, E., Björkroth, B., Daneholt, B., & Höög, C. (1996). Localization of the N-terminus of SCP1 to the central element of the synaptonemal complex and evidence for direct interactions between the N- termini of SCP1 molecules organized head-to-head. *Experimental Cell Research*, *226*(1), 11–19. https://doi.org/10.1006/excr.1996.0197

Llano, E., Herrán, Y., García-Tuñón, I., Gutiérrez-Caballero, C., de Álava, E., Barbero, J. L., … Pendás, A. M. (2012). Meiotic cohesin complexes are essential for the formation of the axial element in mice. *Journal of Cell Biology*, *197*(7), 877–885. https://doi.org/10.1083/jcb.201201100

Long, J., Huang, C., Chen, Y., Zhang, Y., Shi, S., Wu, L., … Lei, M. (2017). Telomeric TERB1-TRF1 interaction is crucial for male meiosis. *Nature Structural and Molecular Biology*, *24*(12), 1073–1080. https://doi.org/10.1038/nsmb.3496

Lu, J., Gu, Y., Feng, J., Zhou, W., Yang, X., & Shen, Y. (2014). Structural insight into the central element assembly of the synaptonemal complex. *Scientific Reports*, *4*, 1–6. https://doi.org/10.1038/srep07059

Lupas, A. N., & Gruber, M. (2005). The structure of alpha-helical coiled coils. *Advances in Protein Chemistry*, *70*, 37–78. https://doi.org/10.1016/S0065-3233(05)70003-6

Lupas, A, Van Dyke, M., & Stock, J. (1991). Predicting coiled coils from protein sequences. *Science (New York, N.Y.)*, *252*(5009), 1162–1164. https://doi.org/10.1126/science.252.5009.1162

Lupas, Andrei. (1997). Predicting coiled-coil regions in proteins. *Current Opinion in Structural Biology*. https://doi.org/10.1016/S0959-440X(97)80056-5

Ma, C. J., Gibb, B., Kwon, Y., Sung, P., & Greene, E. C. (2017). Protein dynamics of human RPA and RAD51 on ssDNA during assembly and disassembly of the RAD51 filament. *Nucleic Acids Research*, *45*(2), 749–761. https://doi.org/10.1093/nar/gkw1125

Manheim, E. A., & McKim, K. S. (2003). The synaptonemal complex component C(2)M regulates meiotic crossing over in Drosophila. *Current Biology*, *13*(4), 276–285. https://doi.org/10.1016/S0960-9822(03)00050-2

Mao, Y. J., Sheng, X. R., & Pan, X. M. (2007). The effects of NaCl concentration and pH on the stability of hyperthermophilic protein Ssh10b. *BMC Biochemistry*, *8*, 1–8. https://doi.org/10.1186/1471-2091-8-28

Maru, Y., Afar, D. E., Witte, O. N., & Shibuya, M. (1996). The dimerization property of glutathione S-transferase partially reactivates Bcr-Abl lacking the oligomerization domain. *Journal of Biological Chemistry*, *271*(26), 15353–15357. https://doi.org/10.1074/jbc.271.26.15353

Marx, A., Hoenger, A., & Mandelkow, E. (2009). Structures of kinesin motor proteins. *Cell Motility*

and the Cytoskeleton, *66*(11), 958–966. https://doi.org/10.1002/cm.20392

Matsumoto, A., Hieda, M., Yokoyama, Y., Nishioka, Y., Yoshidome, K., Tsujimoto, M., & Matsuura, N. (2015). Global loss of a nuclear lamina component, lamin A/C, and LINC complex components SUN1, SUN2, and nesprin-2 in breast cancer. *Cancer Medicine*, *4*(10), 1547–1557. https://doi.org/10.1002/cam4.495

Mceachern, M. J., Krauskopf, A., & Blackburn, E. H. (2000). *Telomeres and their control*.

McGrath, D. A., Fifield, B.-A., Marceau, A. H., Tripathi, S., Porter, L. A., & Rubin, S. M. (2017). Structural basis of divergent cyclin-dependent kinase activation by Spy1/RINGO proteins. *The EMBO Journal*, *36*(15), 2251–2262. https://doi.org/10.15252/embj.201796905

McKim, K. S., Green-Marroquin, B. L., Sekelsky, J. J., Chin, G., Steinberg, C., Khodosh, R., & Scott Hawley, R. (1998). Meiotic synapsis in the absence of recombination. *Science*, *279*(5352), 876–878. https://doi.org/10.1126/science.279.5352.876

McLachlan, A. D., & Stewart, M. (1975). Tropomyosin coiled-coil interactions: Evidence for an unstaggered structure. *Journal of Molecular Biology*, *98*(2), 293–304. https://doi.org/10.1016/S0022-2836(75)80119-7

Merrikh, H., Zhang, Y., Grossman, A. D., & Wang, J. D. (2012). Replication-transcription conflicts in bacteria. *Nature Reviews Microbiology*, *10*(7), 449–458. https://doi.org/10.1038/nrmicro2800

Metzger MB, Pruneda JN, Klevit RE, W. A. (2014). RING-type E3 ligases: master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination. *Biochim Biophys Acta.*, *1843*(1), 47–60. https://doi.org/10.1016/j.bbamcr.2013.05.026.RING-type

Meuwissen, R. L., Offenberg, H. H., Dietrich, A. J., Riesewijk, A., van Iersel, M., & Heyting, C. (1992). A coiled-coil related protein specific for synapsed regions of meiotic prophase chromosomes. *The EMBO Journal*, *11*(13), 5091–5100. https://doi.org/10.1002/j.1460-2075.1992.tb05616.x

Micsonai, A., Wien, F., Bulyáki, É., Kun, J., Moussong, É., Lee, Y.-H., … Kardos, J. (2018). BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Research*, *46*(W1), W315–W322. https://doi.org/10.1093/nar/gky497

Micsonai, A., Wien, F., Kernya, L., Lee, Y.-H., Goto, Y., Réfrégiers, M., & Kardos, J. (2015). Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(24), E3095-103. https://doi.org/10.1073/pnas.1500851112

Mikolcevic, P., Isoda, M., Shibuya, H., del Barco Barrantes, I., Igea, A., Suja, J. A., … Nebreda, A. R. (2016). Essential role of the Cdk2 activator RingoA in meiotic telomere tethering to the nuclear envelope. *Nature Communications*, *7*(1), 11084. https://doi.org/10.1038/ncomms11084

Miller, D. E., Staber, C., Zeitlinger, J., & Hawley, R. S. (2018). Highly Contiguous Genome Assemblies of 15 Drosophila Species Generated Using Nanopore Sequencing. *G3 (Bethesda, Md.)*, *8*(10), 3131–3141. https://doi.org/10.1534/g3.118.200160

Moens, P. B., Chen, D. J., Shen, Z., Kolas, N., Tarsounas, M., Heng, H. H. Q., & Spyropoulos, B. (1997). Rad51 immunocytology in rat and mouse spermatocytes and oocytes. *Chromosoma*, *106*(4), 207–215. https://doi.org/10.1007/s004120050241

Moens, P. B., Marcon, E., Shore, J. S., Kochakpour, N., & Spyropoulos, B. (2007). Initiation and resolution of interhomolog connections: Crossover and non-crossover sites along mouse synaptonemal complexes. *Journal of Cell Science*, *120*(6), 1017–1027. https://doi.org/10.1242/jcs.03394

Moiseeva, V., Amelina, H., Collopy, L. C., Armstrong, C. A., Pearson, S. R., & Tomita, K. (2017). The telomere bouquet facilitates meiotic prophase progression and exit in fission yeast. *Cell Discovery*, *3*, 17041. https://doi.org/10.1038/celldisc.2017.41

Momin, A. A., Hameed, U. F. S., & Arold, S. T. (2019). Passenger sequences can promote interlaced dimers in a common variant of the maltose-binding protein. *Scientific Reports*, *9*(1), 20396. https://doi.org/10.1038/s41598-019-56718-y

Morimoto, A., Shibuya, H., Zhu, X., Kim, J., Ishiguro, K. I., Han, M., & Watanabe, Y. (2012). A conserved KASH domain protein associates with telomeres, SUN1, and dynactin during mammalian meiosis. *Journal of Cell Biology*, *198*(2), 165–172. https://doi.org/10.1083/jcb.201204085

MOSES, M. J. (1956). Chromosomal structures in crayfish spermatocytes. *The Journal of Biophysical and Biochemical Cytology*, *2*(2), 215–218. https://doi.org/10.1083/jcb.2.2.215

Nagaoka, S. I., Hassold, T. J., & Hunt, P. A. (2012). Human aneuploidy: mechanisms and new insights into an age-old problem. *Nature Reviews. Genetics*, *13*(7), 493–504. https://doi.org/10.1038/nrg3245

Nakagawa T, Flores-Rozas H, K. R. (2001). The MER3 Helicase Involved in Meiotic Crossing Over Is Stimulated by Single-stranded DNA-binding Proteins and Unwinds DNA in the 3′ to 5′ Direction. *J Biol Chem*, *276*(34), 31487–31493. https://doi.org/10.1074/jbc.M104003200.The

Nandakumar, J., & Cech, T. R. (2013). Finding the end: recruitment of telomerase to telomeres. *Nature Reviews. Molecular Cell Biology*, *14*(2), 69–82. https://doi.org/10.1038/nrm3505

Nath, D., & Shadan, S. (2009). The ubiquitin system. *Nature*, *458*(7237), 421. https://doi.org/10.1038/458421a

Newman, J. R. S., Wolf, E., & Kim, P. S. (2000). A computationally directed screen identifying interacting coiled coils from Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(24), 13203–13208. https://doi.org/10.1073/pnas.97.24.13203

Neyton, S., Lespinasse, F., Moens, P. B., Paul, R., Gaudray, P., Paquis-Flucklinger, V., & Santucci-Darmanin, S. (2004). Association between MSH4 (MutS homologue 4) and the DNA strand-exchange RAD51 and DMC1 proteins during mammalian meiosis. *Molecular Human Reproduction*, *10*(12), 917–924. https://doi.org/10.1093/molehr/gah123

Nimonkar, A. V, Genschel, J., Kinoshita, E., Polaczek, P., Campbell, J. L., Wyman, C., … Kowalczykowski, S. C. (2011). BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. *Genes & Development*, *25*(4), 350–362. https://doi.org/10.1101/gad.2003811

NOTTKE, A. C., KIM, H. M. & COLAIACOVO, M. P. (2017). The Role of SUMO in Mitosis. *Advances in Experimental Medicine and Biology*, *963*(March 2018), 171–184. https://doi.org/10.1007/978-3-319-50044-7

Nottke, A. C., Kim, H.-M., & Colaiácovo, M. P. (2017). Wrestling with Chromosomes: The Roles of SUMO During Meiosis. *Advances in Experimental Medicine and Biology*, *963*, 185–196. https://doi.org/10.1007/978-3-319-50044-7_11

Öllinger, R., Alsheimer, M., & Benavente, R. (2005). Mammalian protein SCP1 forms synaptonemal complex-like structures in the absence of meiotic chromosomes. *Molecular Biology of the Cell*, *16*(1), 212–217. https://doi.org/10.1091/mbc.E04-09-0771

Ortiz, R., Kouznetsova, A., Echeverría-Martínez, O. M., Vázquez-Nin, G. H., & Hernández-Hernández, A. (2016). The width of the lateral element of the synaptonemal complex is determined by a multilayered organization of its components. *Experimental Cell Research*,

344(1), 22–29. https://doi.org/10.1016/j.yexcr.2016.03.025

Page, S. L., & Hawley, R. S. (2003). Chromosome choreography: The meiotic ballet. *Science*, *301*(5634), 785–789. https://doi.org/10.1126/science.1086605

Page, S. L., & Hawley, R. S. (2004). the Genetics and Molecular Biology of the Synaptonemal Complex. *Annual Review of Cell and Developmental Biology*, *20*(1), 525–558. https://doi.org/10.1146/annurev.cellbio.19.111301.155141

Page, S. L., Khetani, R. S., Lake, C. M., Nielsen, R. J., Jeffress, J. K., Warren, W. D., … Hawley, R. S. (2008). corona is required for higher-order assembly of transverse filaments into full-length synaptonemal complex in Drosophila oocytes. *PLoS Genetics*, *4*(9), 1–12. https://doi.org/10.1371/journal.pgen.1000194

Page, S. L., Nielsen, R. J., Teeter, K., Lake, C. M., Ong, S., Wright, K. R., … Scott Hawley, R. (2007). A germline clone screen for meiotic mutants in Drosophila melanogaster. *Fly*, *1*(3), 172–181. https://doi.org/10.4161/fly.4720

Page, S. L., & Scott Hawley, R. (2001). c(3)G encodes a Drosophila synaptonemal complex protein. *Genes and Development*, *15*(23), 3130–3143. https://doi.org/10.1101/gad.935001

Palm, W., & de Lange, T. (2008). How Shelterin Protects Mammalian Telomeres. *Annual Review of Genetics*, *42*(1), 301–334. https://doi.org/10.1146/annurev.genet.41.110306.130350

Palmer, N., Talib, S. Z. A., Singh, P., Goh, C. M. F., Liu, K., Schimenti, J. C., & Kaldis, P. (2020). A novel function for CDK2 activity at meiotic crossover sites. In *PLoS Biology* (Vol. 18). https://doi.org/10.1371/journal.pbio.3000903

Parisi, S., McKay, M. J., Molnar, M., Thompson, M. A., van der Spek, P. J., van Drunen-Schoenmaker, E., … Kohli, J. (1999). Rec8p, a Meiotic Recombination and Sister Chromatid Cohesion Phosphoprotein of the Rad21p Family Conserved from Fission Yeast to Humans. *Molecular and Cellular Biology*, *19*(5), 3515–3528. https://doi.org/10.1128/mcb.19.5.3515

Park, W. M. (2020). Coiled-coils: The molecular zippers that self-assemble protein nanostructures. *International Journal of Molecular Sciences*, *21*(10). https://doi.org/10.3390/ijms21103584

Pelttari, J., Hoja, M.-R., Yuan, L., Liu, J.-G., Brundell, E., Moens, P., … Höög, C. (2001). A Meiotic Chromosomal Core Consisting of Cohesin Complex Proteins Recruits DNA Recombination Proteins and Promotes Synapsis in the Absence of an Axial Element in Mammalian Meiotic Cells. *Molecular and Cellular Biology*, *21*(16), 5667–5677. https://doi.org/10.1128/mcb.21.16.5667-5677.2001

Pendlebury, D. F., Fujiwara, Y., Tesmer, V. M., Smith, E. M., Shibuya, H., Watanabe, Y., … Lei, M. (2017). Dissecting the telomere-inner nuclear membrane interface formed in meiosis. *Nature Structural and Molecular Biology*, *24*(12), 1073–1080. https://doi.org/10.1038/nsmb.3496

Peränen, J., Rikkonen, M., Hyvönen, M., & Kääriäinen, L. (1996). T7 Vectors with a Modified T7/ac Promoter for Expression of Proteins in Escherichia coli. *Analytical Biochemistry*, *236*(2), 371–373. https://doi.org/10.1006/abio.1996.0187

Perry, J., Kleckner, N., & Börner, G. V. (2005). Bioinformatic analyses implicate the collaborating meiotic crossover/chiasma proteins Zip2, Zip3, and Spo22/Zip4 in ubiquitin labeling. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(49), 17594–17599. https://doi.org/10.1073/pnas.0508581102

Petoukhov, M. V, Franke, D., Shkumatov, A. V, Tria, G., Kikhney, A. G., Gajda, M., … Svergun, D. I. (2012). New developments in the ATSAS program package for small-angle scattering data analysis. *Journal of Applied Crystallography*, *45*(Pt 2), 342–350. https://doi.org/10.1107/S0021889812007662

Petronczki M, Siomos MF, N. K. (2003). Un ménage à quatre: the molecular biology of chromosome segregation in meiosis. *Cell*, *112*(4), 423–440.

Prasada Rao, H. B. D., Qiao, H., Bhatt, S. K., Bailey, L. R. J., Tran, H. D., Bourne, S. L., … Hunter, N. (2017). A SUMO-ubiquitin relay recruits proteasomes to chromosome axes to regulate meiotic recombination. *Science*, *355*(6323), 403–407. https://doi.org/10.1126/science.aaf6407

Prieto, I., Suja, J. A., Pezzi, N., Kremer, L., Martínez-A, C., Rufas, J. S., & Barbero, J. L. (2001). Mammalian STAG3 is a cohesin specific to sister chromatid arms in meiosis I. *Nature Cell Biology*, *3*(8), 761–766. https://doi.org/10.1038/35087082

Psakhye, I., & Jentsch, S. (2012). Protein group modification and synergy in the SUMO pathway as exemplified in DNA repair. *Cell*, *151*(4), 807–820. https://doi.org/10.1016/j.cell.2012.10.021

Pyatnitskaya, A., Borde, V., & De Muyt, A. (2019). Crossing and zipping: molecular duties of the ZMM proteins in meiosis. *Chromosoma*, *128*(3), 181–198. https://doi.org/10.1007/s00412-019-00714-8

Qiao, H., Chen, J. K., Reynolds, A., Höög, C., Paddy, M., & Hunter, N. (2012). Interplay between synaptonemal complex, homologous recombination, and centromeres during mammalian meiosis. *PLoS Genetics*, *8*(6), e1002790. https://doi.org/10.1371/journal.pgen.1002790

Qiao, H., Prasada Rao, H. B. D., Yang, Y., Fong, J. H., Cloutier, J. M., Deacon, D. C., … Hunter, N. (2014a). Antagonistic roles of ubiquitin ligase HEI10 and SUMO ligase RNF212 regulate meiotic recombination. *Nature Genetics*, *46*(2), 194–199. https://doi.org/10.1038/ng.2858

Qiao, H., Prasada Rao, H. B. D., Yang, Y., Fong, J. H., Cloutier, J. M., Deacon, D. C., … Hunter, N. (2014b). Antagonistic roles of ubiquitin ligase HEI10 and SUMO ligase RNF212 regulate meiotic recombination. *Nature Genetics*, *46*(2), 194–199. https://doi.org/10.1038/ng.2858

Rasmussen, S. W., & Holm, P. B. (1984). The synaptonemal complex, recombination nodules and chiasmata in human spermatocytes. *Symposia of the Society for Experimental Biology*, *38*, 271–292.

Reynolds, A., Qiao, H., Yang, Y., Chen, J. K., Jackson, N., Biswas, K., … Hunter, N. (2013). RNF212 is a dosage-sensitive regulator of crossing-over during mammalian meiosis. *Nature Genetics*, *45*(3), 269–278. https://doi.org/10.1038/ng.2541

Robert, T., Nore, A., Brun, C., Maffre, C., Crimi, B., Bourbon, H.-M., & de Massy, B. (2016). The TopoVIB-Like protein family is required for meiotic DNA double-strand break formation. *Science (New York, N.Y.)*, *351*(6276), 943–949. https://doi.org/10.1126/science.aad5309

Roeder, G. S. (1997). Meiotic chromosomes: it takes two to tango. *Genes & Development*, *11*(20), 2600–2621. https://doi.org/10.1101/gad.11.20.2600

Rose, A., & Meier, I. (2004). Scaffolds, levers, rods and springs: diverse cellular functions of long coiled-coil proteins. *Cellular and Molecular Life Sciences : CMLS*, *61*(16), 1996–2009. https://doi.org/10.1007/s00018-004-4039-6

Sánchez-Sáez, F., Gómez-H, L., Dunne, O. M., Gallego-Páramo, C., Felipe-Medina, N., Sánchez-Martín, M., … Davies, O. R. (2020). Meiotic chromosome synapsis depends on multivalent SYCE1-SIX6OS1 interactions that are disrupted in cases of human infertility. *BioRxiv*, (September), 1–15. https://doi.org/10.1101/2020.02.04.934372

Sandell, L. L., & Zakian, V. A. (1993). Loss of a yeast telomere: arrest, recovery, and chromosome loss. *Cell*, *75*(4), 729–739. https://doi.org/10.1016/0092-8674(93)90493-a

Santucci-Darmanin, S., Neyton, S., Lespinasse, F., Saunières, A., Gaudray, P., & Paquis-Flucklinger, V. (2002). The DNA mismatch-repair MLH3 protein interacts with MSH4 in meiotic cells, supporting a role for this MutL homolog in mammalian meiotic recombination. *Human*

Molecular Genetics, *11*(15), 1697–1706. https://doi.org/10.1093/hmg/11.15.1697

Santucci-Darmanin, S., Walpita, D., Lespinasse, F., Desnuelle, C., Ashley, T., & Paquis-Flucklinger, V. (2000). MSH4 acts in conjunction with MLH1 during mammalian meiosis. *The FASEB Journal*, *14*(11), 1539–1547. https://doi.org/10.1096/fj.99-0851com

Sato, A., Isaac, B., Phillips, C. M., Rillo, R., Carlton, P. M., Wynne, D. J., … Dernburg, A. F. (2009). Cytoskeletal forces span the nuclear envelope to coordinate meiotic chromosome pairing and synapsis. *Cell*, *139*(5), 907–919. https://doi.org/10.1016/j.cell.2009.10.039

Scherthan, H. (2001). A bouquet makes ends meet. *Nature Reviews Molecular Cell Biology*, *2*(8), 621–627. https://doi.org/10.1038/35085086

Scherthan Weich, S., Schwegler, H., Heyting, C., Harle, M., H., & Cremer, T. (1996). Centromere and telomeremovements during early meiotic prophase of mouse and man areassociated with the onset of pairing. *J.Cell.Biol*, *134*(5), 1109–1125. https://doi.org/10.1083/jcb.134.5.1109

Schmekel, K., Meuwissen, R. L. J., Dietrich, A. J. J., Vink, A. C. G., Van Marle, J., Van Veen, H., & Heyting, C. (1996). Organization of SCP1 protein molecules within synaptonemal complexes of the rat. *Experimental Cell Research*, *226*(1), 20–30. https://doi.org/10.1006/excr.1996.0198

Schmekel, Karin, Wahrman, J., Skoglund, U., & Daneholt, B. (1993). The central region of the synaptonemal complex in Blaps cribrosa studied by electron microscope tomography. *Chromosoma*, *102*(10), 669–681. https://doi.org/10.1007/BF00650893

Schmitt, J., Benavente, R., Hodzic, D., Hoog, C., Stewart, C. L., & Alsheimer, M. (2007). Transmembrane protein Sun2 is involved in tethering mammalian meiotic telomeres to the nuclear envelope. *Proceedings of the National Academy of Sciences*, *104*(18), 7426–7431. https://doi.org/10.1073/pnas.0609198104

Schramm, S., Fraune, J., Naumann, R., Hernandez-Hernandez, A., Höög, C., Cooke, H. J., … Benavente, R. (2011). A novel mouse synaptonemal complex protein is essential for loading of central element proteins, recombination, and fertility. *PLoS Genetics*, *7*(5), e1002088. https://doi.org/10.1371/journal.pgen.1002088

Schücker, K., Holm, T., Franke, C., Sauer, M., & Benavente, R. (2015a). Elucidation of synaptonemal complex organization by super-resolution imaging with isotropic resolution. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(7), 2029–2033. https://doi.org/10.1073/pnas.1414814112

Schücker, K., Holm, T., Franke, C., Sauer, M., & Benavente, R. (2015b). Elucidation of synaptonemal complex organization by super-resolution imaging with isotropic resolution. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(7), 2029–2033. https://doi.org/10.1073/pnas.1414814112

Schwacha, A., & Kleckner, N. (1995). Identification of double holliday junctions as intermediates in meiotic recombination. *Cell*, *83*(5), 783–791. https://doi.org/10.1016/0092-8674(95)90191-4

Seo, E. K., Choi, Y., Jeong, J. H., Kim, Y. G., & Park, H. H. (2016). Crystal structure of C-terminal coiled-coil domain of SYCP1 reveals non-canonical anti-parallel dimeric structure of transverse filament at the synaptonemal complex. *PLoS ONE*, *11*(8), 1–13. https://doi.org/10.1371/journal.pone.0161379

Sfeir, Agnel J, de Lange, T. (2012). Removal of Shelterin Revelas the Telomere End-Replication Problem. *Science*, *336*(6081), 593–597. https://doi.org/10.1126/science.1218498.Removal

Shammas, M. A. (2011). Telomeres, lifestyle, cancer, and aging. *Current Opinion in Clinical Nutrition and Metabolic Care*, *14*(1), 28–34. https://doi.org/10.1097/MCO.0b013e32834121b1

Sharma, R., Biedenharn, K. R., Fedor, J. M., & Agarwal, A. (2013). Lifestyle factors and reproductive

health: taking control of your fertility. *Reproductive Biology and Endocrinology : RB&E*, *11*, 66. https://doi.org/10.1186/1477-7827-11-66

Sherizen, D., Jang, J. K., Bhagat, R., Kato, N., & Mckim, K. S. (2005). *Meiotic Recombination in Drosophila Females Depends on Chromosome Continuity Between Genetically Defined Boundaries*. *781*(February), 767–781. https://doi.org/10.1534/genetics.104.035824

Shibuya, H., Hernández-Hernández, A., Morimoto, A., Negishi, L., Höög, C., & Watanabe, Y. (2015a). MAJIN Links Telomeric DNA to the Nuclear Membrane by Exchanging Telomere Cap. *Cell*, *163*(5), 1252–1266. https://doi.org/10.1016/j.cell.2015.10.030

Shibuya, H., Hernández-Hernández, A., Morimoto, A., Negishi, L., Höög, C., & Watanabe, Y. (2015b). MAJIN Links Telomeric DNA to the Nuclear Membrane by Exchanging Telomere Cap. *Cell*, *163*(5), 1252–1266. https://doi.org/10.1016/j.cell.2015.10.030

Shibuya, H., Ishiguro, K. I., & Watanabe, Y. (2014). The TRF1-binding protein TERB1 promotes chromosome movement and telomere rigidity in meiosis. *Nature Cell Biology*, *16*(2), 145–156. https://doi.org/10.1038/ncb2896

Shibuya, H., Morimoto, A., & Watanabe, Y. (2014). The dissection of meiotic chromosome movement in mice using an in vivo electroporation technique. *PLoS Genetics*, *10*(12), e1004821. https://doi.org/10.1371/journal.pgen.1004821

Shibuya, H., & Watanabe, Y. (2014). The meiosis-specific modification of mammalian telomeres. *Cell Cycle (Georgetown, Tex.)*, *13*(13), 2024–2028. https://doi.org/10.4161/cc.29350

Shinohara, M., Oh, S. D., Hunter, N., & Shinohara, A. (2008). Crossover assurance and crossover interference are distinctly regulated by the ZMM proteins during yeast meiosis. *Nature Genetics*, *40*(3), 299–309. https://doi.org/10.1038/ng.83

Smogorzewska, A., van Steensel, B., Bianchi, A., Oelmann, S., Schaefer, M. R., Schnapp, G., & de Lange, T. (2000). Control of human telomere length by TRF1 and TRF2. *Molecular and Cellular Biology*, *20*(5), 1659–1668. https://doi.org/10.1128/mcb.20.5.1659-1668.2000

Snowden, T., Acharya, S., Butz, C., Berardini, M., & Fishel, R. (2004). hMSH4-hMSH5 recognizes holliday junctions and forms a meiosis-specific sliding clamp that embraces homologous chromosomes. *Molecular Cell*, *15*(3), 437–451. https://doi.org/10.1016/j.molcel.2004.06.040

Solari, A. J., & Moses, M. J. (1973). The structure of the central region in the synaptonemal complexes of hamster and cricket spermatocytes. *Journal of Cell Biology*, *56*(1), 145–152. https://doi.org/10.1083/jcb.56.1.145

Sonneville, R., Craig, G., Labib, K., Gartner, A., & Blow, J. J. (2015). Both Chromosome Decondensation and Condensation Are Dependent on DNA Replication in C.elegans Embryos. *Cell Reports*, *12*(3), 405–417. https://doi.org/10.1016/j.celrep.2015.06.046

Storlazzi A, Gargano S, Ruprich-Robert G, Falque M, David M, Kleckner N, Z. D. (2010). Recombination proteins mediate meiotic spatial chromosome organization and pairing. *Cell*, *141*(1), 94–106. https://doi.org/10.1016/j.cell.2010.02.041.Recombination

Strong, E. R., & Schimenti, J. C. (2010). Evidence implicating CCNB1IP1, a RING domain-containing protein required for meiotic crossing over in mice, as an E3 SUMO ligase. *Genes*, *1*(3), 440–451. https://doi.org/10.3390/genes1030440

Svergun, D. (1999). Restoring low resolution structure of biological macromolecules from solution. *Biophysical Journal*, *76*(June), 2879–2886. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0006349599774436

Sym, M., & Roeder, G. S. (1994). Crossover interference is abolished in the absence of a synaptonemal complex protein. *Cell*, *79*(2), 283–292. https://doi.org/10.1016/0092-

8674(94)90197-x

Syrjänen, J. L., Pellegrini, L., & Davies, O. R. (2014). A molecular model for the role of SYCP3 in meiotic chromosome organisation. *ELife*, *3*, 1–18. https://doi.org/10.7554/elife.02963

Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., & Stahl, F. W. (1983). The double-strand-break repair model for recombination. *Cell*, *33*(1), 25–35. https://doi.org/10.1016/0092-8674(83)90331-8

Takai, K. K., Hooper, S., Blackwood, S., Gandhi, R., & de Lange, T. (2010). In vivo stoichiometry of shelterin components. *Journal of Biological Chemistry*, *285*(2), 1457–1467. https://doi.org/10.1074/jbc.M109.038026

Takeo, S., Lake, C. M., Morais-De-Sá, E., Sunkel, C. E., & Hawley, R. S. (2011). Synaptonemal complex-dependent centromeric clustering and the initiation of synapsis in Drosophila oocytes. *Current Biology*, *21*(21), 1845–1851. https://doi.org/10.1016/j.cub.2011.09.044

Toby, G. G., Gherraby, W., Coleman, T. R., & Golemis, E. A. (2003). A Novel RING Finger Protein, Human Enhancer of Invasion 10, Alters Mitotic Progression through Regulation of Cyclin B Levels. *Molecular and Cellular Biology*, *23*(6), 2109–2122. https://doi.org/10.1128/mcb.23.6.2109-2122.2003

Trelles-Sticken, E., Dresser, M. E., & Scherthan, H. (2000). Meiotic telomere protein Ndj1p is required for meiosis-specific telomere distribution, bouquet formation and efficient homologue pairing. *Journal of Cell Biology*, *151*(1), 95–106. https://doi.org/10.1083/jcb.151.1.95

Truebestein, L., & Leonard, T. A. (2016). Coiled-coils: The long and short of it. *BioEssays*, *38*(9), 903–916. https://doi.org/10.1002/bies.201600062

Tu, Z., Bayazit, M. B., Liu, H., Zhang, J., Busayavalasa, K., Risal, S., … Liu, K. (2017). Speedy A-Cdk2 binding mediates initial telomere-nuclear envelope attachment during meiotic prophase I independent of Cdk2 activation. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(3), 592–597. https://doi.org/10.1073/pnas.1618465114

Tung, K. S., & Roeder, G. S. (1998). Meiotic chromosome morphology and behavior in zip1 mutants of Saccharomyces cerevisiae. *Genetics*, *149*(2), 817–832. https://doi.org/10.1093/genetics/149.2.817

Turner, J. M. A. (2007). Meiosis 2007--where have we got to and where are we going? *Chromosome Research : An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, *15*(5), 517–521. https://doi.org/10.1007/s10577-007-1152-z

van Steensel, B., Smogorzewska, A., & de Lange, T. (1998). TRF2 protects human telomeres from end-to-end fusions. *Cell*, *92*(3), 401–413. https://doi.org/10.1016/s0092-8674(00)80932-0

Vander Kooi CW, Ohi MD, Rosenberg JA, Oldham ML, Newcomer ME, Gould KL, C. W. (2006). The Prp19 U-box Crystal Structure Suggests a Common Dimeric Architecture for a Class of Oligomeric E3 Ubiquitin Ligases. *Biochemistry*, *45*(10), 121–130. https://doi.org/10.1021/bi051787e.The

Vera, L., Czarny, B., Georgiadis, D., Dive, V., & Stura, E. A. (2011). Practical Use of Glycerol in Protein Crystallization. *Crystal Growth & Design*, *11*(7), 2755–2762. https://doi.org/10.1021/cg101364m

Viera, A., Alsheimer, M., Gómez, R., Berenguer, I., Ortega, S., Symonds, C. E., … Suja, J. A. (2015). CDK2 regulates nuclear envelope protein dynamics and telomere attachment in mouse meiotic prophase. *Journal of Cell Science*, *128*(1), 88–99. https://doi.org/10.1242/jcs.154922

Viera, A., Rufas, J. S., Martínez, I., Barbero, J. L., Ortega, S., & Suja, J. A. (2009). CDK2 is required for proper homologous pairing, recombination and sex-body formation during male mouse

meiosis. *Journal of Cell Science*, *122*(12), 2149–2159. https://doi.org/10.1242/jcs.046706

Volkov, V. V., & Svergun, D. I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography*, *36*(3 I), 860–864. https://doi.org/10.1107/S0021889803000268

von Wettstein, D. (1984). The synaptonemal complex and genetic segregation. *Symposia of the Society for Experimental Biology*, *38*, 195–231. https://doi.org/10.1146/annurev.genet.18.1.331

Walker, M. Y., & Hawley, R. S. (2000). *Hanging on to your homolog: the roles of pairing, synapsis and recombination in the maintenance of homolog adhesion*. 3–9.

Wang, G., Wu, X., Zhou, L., Gao, S., Yun, D., Liang, A., & Sun, F. (2020). Tethering of Telomeres to the Nuclear Envelope Is Mediated by SUN1-MAJIN and Possibly Promoted by SPDYA-CDK2 During Meiosis. *Frontiers in Cell and Developmental Biology*, *8*(September), 1–11. https://doi.org/10.3389/fcell.2020.00845

Wang, L., Tu, Z., Liu, C., Liu, H., Kaldis, P., Chen, Z., & Li, W. (2018). Dual roles of TRF1 in tethering telomeres to the nuclear envelope and protecting them from fusion during meiosis. *Cell Death and Differentiation*, *25*(6), 1174–1188. https://doi.org/10.1038/s41418-017-0037-8

Wang, Y., Chen, Y., Chen, J., Wang, L., Nie, L., Long, J., … Lei, M. (2019). The meiotic TERB1-TERB2-MAJIN complex tethers telomeres to the nuclear envelope. *Nature Communications*, *10*(1), 564. https://doi.org/10.1038/s41467-019-08437-1

Ward, J. O., Reinholdt, L. G., Motley, W. W., Niswander, L. M., Deacon, D. C., Griffin, L. B., … Schimenti, J. C. (2007). Mutation in mouse Hei10, an E3 ubiquitin ligase, disrupts meiotic crossing over. *PLoS Genetics*, *3*(8), 1550–1563. https://doi.org/10.1371/journal.pgen.0030139

Webber, H. A., Howard, L., & Bickel, S. E. (2004). The cohesion protein ORD is required for homologue bias during meiotic recombination. *Journal of Cell Biology*, *164*(6), 819–829. https://doi.org/10.1083/jcb.200310077

West, A. M., Rosenberg, S. C., Ur, S. N., Lehmer, M. K., Ye, Q., Hagemann, G., … Corbett, K. D. (2019). A conserved filamentous assembly underlies the structure of the meiotic chromosome axis. *ELife*, *8*. https://doi.org/10.7554/eLife.40372

Westergaard, M., & von Wettstein, D. (1972). The synaptonemal complex. *Annual Review of Genetics*, *6*, 71–110.

Wilce, M. C. J., Feil, S. C., Board, P. G., & Parker, M. W. (1994). Crystallization and preliminary X-ray diffraction studies of a glutathione S-Transferase from the Australian sheep blowfly, Lucilia cuprina. *Journal of Molecular Biology*, *236*(5), 1407–1409. https://doi.org/10.1016/0022-2836(94)90067-1

Winkel, K., Alsheimer, M., Öllinger, R., & Benavente, R. (2009). Protein SYCP2 provides a link between transverse filaments and lateral elements of mammalian synaptonemal complexes. *Chromosoma*, *118*(2), 259–267. https://doi.org/10.1007/s00412-008-0194-0

Woglar, A., & Jantsch, V. (2014). Chromosome movement in meiosis I prophase of Caenorhabditis elegans. *Chromosoma*, *123*(1–2), 15–24. https://doi.org/10.1007/s00412-013-0436-7

Wu, X., Wu, D., Lu, Z., Chen, W., Hu, X., & Ding, Y. (2009). A novel method for high-level production of TEV protease by superfolder GFP tag. *Journal of Biomedicine and Biotechnology*, *2009*. https://doi.org/10.1155/2009/591923

Yamamoto West, R.R., McIntosh, J.R., Hiraoka, Y., a. (1999). A cytoplasmic dynein heavy chain is required for oscillatory nuclear movement of meiotic prophase and efficient meiotic recombination in fission yeast. *J. Cell Biol.*, *145*(6), 1233–1249. https://doi.org/10.1083/jcb.145.6.1233

Yang, F., De La Fuente, R., Leu, N. A., Baumann, C., McLaughlin, K. J., & Wang, P. J. (2006). Mouse SYCP2 is required for synaptonemal complex assembly and chromosomal synapsis during male meiosis. *The Journal of Cell Biology*, *173*(4), 497–507. https://doi.org/10.1083/jcb.200603063

Yuan, L., Liu, J. G., Zhao, J., Brundell, E., Daneholt, B., & Höög, C. (2000). The murine SCP3 gene is required for synaptonemal complex assembly, chromosome synapsis, and male fertility. *Molecular Cell*, *5*(1), 73–83. https://doi.org/10.1016/S1097-2765(00)80404-9

Zakharyevich, K., Tang, S., Ma, Y., & Hunter, N. (2012). Delineation of joint molecule resolution pathways in meiosis identifies a crossover-specific resolvase. *Cell*, *149*(2), 334–347. https://doi.org/10.1016/j.cell.2012.03.023

Zhang, J., Tu, Z., Watanabe, Y., & Shibuya, H. (2017). Distinct TERB1 Domains Regulate Different Protein Interactions in Meiotic Telomere Movement. *Cell Reports*, *21*(7), 1715–1726. https://doi.org/10.1016/j.celrep.2017.10.061

Zhang, L., Espagne, E., De Muyt, A., Zickler, D., & Kleckner, N. E. (2014). Interference-mediated synaptonemal complex formation with embedded crossover designation. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(47), E5059–E5068. https://doi.org/10.1073/pnas.1416411111

Zhang, X., Lei, K., Yuan, X., Wu, X., Zhuang, Y., Xu, T., … Han, M. (2009). SUN1/2 and Syne/Nesprin-1/2 complexes connect centrosome to the nucleus during neurogenesis and neuronal migration in mice. *Neuron*, *64*(2), 173–187. https://doi.org/10.1016/j.neuron.2009.08.018

Zhu, H., Ownby, D. W., Riggs, C. K., Nolasco, N. J., Stoops, J. K., & Riggs, A. F. (1996). Assembly of the gigantic hemoglobin of the earthworm Lumbricus terrestris. Roles of subunit equilibria, non-globin linker chains, and valence of the heme iron. *Journal of Biological Chemistry*, *271*(47), 30007–30021. https://doi.org/10.1074/jbc.271.47.30007

Zickler, D., & Kleckner, N. (1999). Meiotic Chromosomes: Integrating Structure and Function. *Annual Review of Genetics*, *33*(1), 603–754. https://doi.org/10.1146/annurev.genet.33.1.603

Zickler, D, & Kleckner, N. (1998). The leptotene-zygotene transition of meiosis. *Annual Review of Genetics*, *32*, 619–697. https://doi.org/10.1146/annurev.genet.32.1.619

Zickler, Denise. (2006). From early homologue recognition to synaptonemal complex formation. *Chromosoma*, *115*(3), 158–174. https://doi.org/10.1007/s00412-006-0048-6

Zickler, Denise, & Kleckner, N. (2015). Recombination, Pairing, and Synapsis of Homologs during Meiosis. *Cold Spring Harbor Perspectives in Biology*, *7*(6). https://doi.org/10.1101/cshperspect.a016626