Structural and functional characterisation of the group II chaperonin CCT/TRiC

John Kelly

Lady Margaret Hall



Supervisors: Dr Wyatt Yue, Dr Juha Huiskonen

A thesis submitted for the degree of *Doctor of Philosophy* in Clinical Medicine at the Structural Genomics Consortium

Nuffield Department of Medicine, University of Oxford

Michaelmas term 2020

Declaration

I declare that no parts of this research or thesis have been submitted for another award, degree or diploma at any other university or learning institution. This thesis contains no other person's work, except where stated in text.

Acknowledgements

Thank you to my supervisors, Wyatt Yue and Juha Huiskonen, for their consistent source of positive feedback, patience, and encouragement to undertake new challenges and develop novel approaches that make me proud of the work reported herein. In particular to Wyatt, who took a chance on me early on in my scientific career and has been the best supervisor and mentor a student could ask for. Thank you also to the Yue group, in particular Jesse Coker and Henry Bailey, whose friendship and commiseration helped me get through when the going was tough, and celebrated the victories when they came along, and to the many members of the Structural Genomics Consortium who provided unending help and support throughout the years of my study. Thank you also to the Nuffield Department of Medicine and Pfizer, Inc for funding for the DPhil work herein.

Thank you to my parents, Carolyn and Steve, who always believed in me and encouraged me to pursue my dreams no matter how big or small. Thank you to my big brother Danny, who has been my best friend and the consummate role model my entire life.

And thank you to Abigail, my long-distance girlfriend, fiancée, and wife through this arduous journey, who has been with me through every high and every low, and without whose endless support, patience, sacrifice, and love I would not be where I am today.

Abstract

Cellular health is dependent on the proper folding of proteins and dysregulation of this pathway can lead to many diseases including neurodegeneration and cancer. This thesis describes efforts undertaken to characterise the mechanism by which one eukaryotic protein chaperonin, CCT/TRiC (TRiC), assists the folding of client substrates. TRiC folds nascent polypeptides using a built-in lid mechanism that can open and close to allow substrate access/restriction in an ATP-dependent manner. TRiC is composed of eight individual subunits, some of which can form synthetic homo-oligomers that retain biological function to a varying degree. The native TRiC complex has proven difficult to isolate recombinantly due to the presence of eight subunits in the complex. Strategies to overcome this difficulty include isolation *via* nanobody (Nb)-mediated affinity pull-down and by using CRISPR/Cas9 genome editing to insert an affinity tag in the genomic locus of a single TRiC subunit, CCT5, the latter of which can produce large quantities of pure endogenous TRiC from human cells. Structural investigations of native TRiC reveal new details into its substrate-folding mechanism by observing several intermediate states including TRiC-tubulin bound and TRiC-Actin/PhLP2A bound structures. These results suggest a mechanism whereby substrates bind TRiC in an unfolded state and achieve their near-native fold while bound in the closed state. Additionally, the interaction between the co-chaperone PhLP2A and actin shows a novel binding mechanism whereby each TRiC cavity is occupied. These results provide new insights into the TRiC-mediated substrate folding cycle and provide further therapeutic potential for human diseases caused by disruption of proteostasis.

Abbreviations

ADP – Adenosine diphosphate

AlF_x – Aluminium fluoride

AML1-ETO – Acute myeloid leukaemia 1 eight twenty-one transcription factor

AMP-PNP – Adenylyl-imidodiphosphate

AP-MS – Affinity purified mass spectrometry

ATP – Adenosine triphosphate

BDNF – Brain-derived neurotrophic factor

BeF_x – Beryllium fluoride

BLI – Bio-layer interferometry

BS3 – Bis(sulfosuccinimidyl)suberate

Cas9 – CRISPR associated protein 9

CCT – Chaperonin containing TCP-1

CDR – Complementarity-determining regions

CRISPR – Clustered regularly interspersed short palindromic repeats

CTF – Contrast transfer function

DMSO – Dimethyl sulfoxide

DSF – Differential scanning fluorimetry

DTT – Dithiothreitol

eBIC – Electron Bio-Imaging Centre

EM – Electron microscopy

EDTA – Ethylenediaminetetraacetic acid

ELISA – Enzyme-linked immunosorbent assay

F-actin – Filamentous actin

FSC – Fourier shell correlation

G-actin – Monomeric actin

GTP – Guanosine triphosphate

HAL – Homology arm left

HAR – Homology arm right

HDAC – Histone deacetylase

HDR – Homology-directed repair

H/DX-MS – Hydrogen/deuterium exchange mass spectrometry

HEK – Human embryonic kidney

HEPES – 4-(2-Hydroxyethyl)-1-piperazine ethanesulfonic acid

HIV – Human immunodeficiency virus

HMW – High molecular weight

Hsf1 – Heat shock transcription factor 1

Hsp – Heat shock protein

HγD – Human γ-D crystallin

LC-MS/MS – Liquid chromatography tandem mass spectrometry

LIC – Ligation independent cloning

LMW – Low molecular weight

Mhtt – Mutant huntingtin exon 1

mLST8 – Mammalian lethal with SEC13 protein 8

Mm-Cpn - *Methanococcus maripaludis* chaperonin

MW – Molecular weight

Nb – Nanobody

Ni-NTA – Nickle-nitrilotriacetic acid

NMR – Nuclear magnetic resonance

NPP – Nucleotide partially preloaded

NSL – Nucleotide-sensing loop

PhLP – Phosducin-like protein

PFD – Prefoldin

P-loop – phosphate-binding loop

Raptor – Regulatory associated protein of mTOR

RLS – Release loop of substrate

SAXS – Small-angle X-ray scattering

SDS-PAGE – Sodium dodecyl sulphate–polyacrylamide gel electrophoresis

SEC – Size exclusion chromatography

Sf9 – *Spodoptera frugiperda* insect cells

ssDNA – Single stranded DNA

TBD – Taxol binding domain

TCEP – Tris(2-carboxyethyl)phosphine

TEM – Tunnelling electron microscope

TEV – Tobacco Etch Virus

T_m – Thermal melt temperature

TOR – Target for rapamycin

TRiC – TCP-1 containing ring complex

V_e – Elution volume

VHL – Von Hippel Lindau tumour suppressor

XL-MS – Crosslinking coupled with mass spectrometry

Contents

Declaration 2
Acknowledgements 2
Abstract 3
Abbreviations 4
Contents 7
1. Introduction 15
1.1 Protein folding is influenced by chaperone networks involved in proteostasis. 15
1.1.1 Chaperones promote proper folding and prevent aggregation of misfolded and unfolded proteins. 15
1.1.2 Group I chaperonins 16
1.1.3 Archetypes of Group II chaperonins – thermosome, TF55, and Mm-Cpn 17
1.1.4 Eukaryotic CCT/TRiC 18
1.1.5 Over-expression of individual CCT subunits leads to homo-oligomeric complex formation 21
1.2 Substrates and cellular functions of TRiC 22
1.2.1 Early substrates identified as TRiC interacting partners 22
1.2.2 TRiC is modulated by co-chaperones including prefoldin and the phosducin-like protein family. 24
1.2.3 TRiC function is involved in multiple cellular pathways 25
1.2.4 TRiC is an important factor in a wide array of molecular diseases 27
1.2.5 TRiC is an attractive therapeutic target for disease. 28
1.3 Architecture of group II chaperonins from structural studies 29
1.3.1 Domain organization of TRiC 29
1.3.2 Organization of subunits within the TRiC complex 35
1.3.3 CCT subunit orientation within the TRiC complex 37
1.4 Role of ATP binding and hydrolysis 40
1.4.1 ATP binding and hydrolysis plays a key functional role in TRiC-mediated substrate folding 40
1.4.2 Structural studies reveal consequences of ATP binding and hydrolysis 41
1.4.3 Kinetic characterization of ATP hydrolysis in CCT subunits 45
1.4.4 ATP activity of some subunits is essential for cellular health 47
1.5 Mechanism of substrate binding 49

1.5.1 Substrate recognition at the surface apical domain region of TRiC 49
1.5.2 Kinetic and biochemical studies involving actin 51
1.5.3 Interaction of substrates inside the folding chamber revealed through structural studies 52
1.5.4 Structures of prefoldin co-chaperone bound to TRiC 56
1.6 Project aims and thesis scope 57
1.6.1 Unanswered questions from structure and biophysical studies to date 57
1.6.2 Thesis objectives 59
1.6.3 Thesis chapter layout 60
2. Methods 62
2.1 Recombinant protein production of CCT4 and CCT5 from *E. coli* 62
2.2 Recombinant protein production of CCT1-CCT8 in *Sf9* insect cells 63
2.3 Endogenous protein production 64
2.3.1 Endogenous TRiC purification from wild-type human HEK293 cells 64
2.3.2 CRISPR/Cas9-mediated genome editing to introduce affinity tag into CCT5 genetic locus 65
2.3.3 Purification of native TRiC 66
2.3.4 Nanobody mediated pull-down 66
2.4 Mass spectrometry 67
2.4.1 Intact mass spectrometry 67
2.4.2 Peptide identification by mass spectrometry 68
2.5 Biophysical and biochemical assays 70
2.5.1 Bio-layer interferometry 70
2.5.2 Thermal stability measurements by differential scanning fluorimetry 70
2.5.3 Malachite green ATPase activity assay 71
2.5.4 HγD aggregation suppression assay 71
2.6 Small angle X-ray scattering 71
2.7 Negative stain electron microscopy 72
2.8 Cryo-EM 72
2.8.1 Cryo-EM of CCT5 homo-oligomer 72
2.8.2 Cryo-EM of native TRiC 73
2.8.3 Model building of TRiC and substrates 74
3. Individual TRiC subunits 76

3.1 Background 76
3.2 Expression and purification of CCT4 and CCT5 in *E. coli* 77
3.3 Biophysical investigations of CCT4 and CCT5 chaperonins 80
3.3.1 CCT5 ATPase activity 80
3.3.2 CCT5 aggregation suppression of human γD crystallin 81
3.3.3 CCT5 thermal stability 83
3.4 Structural study of CCT4 and CCT5 using cryo-EM 85
3.4.1 Background of negative stain and cryo-EM 85
3.4.2 Initial screening of grids and small data collection 89
3.4.3 Large scale data collection of CCT5 homo-oligomeric complexes in open and closed conformations 90
3.5 Expression and purification of individual subunits from insect cells 95
3.6 SAXS reveals different oligomeric states 99
3.7 CCT subunits characterized by stability and activity assays. 105
3.7.1 DSF stability measured for all eight CCT subunits 105
3.7.2 ATPase activity assay measured for all subunits 106
3.8 Negative stain EM reveals homo-oligomeric CCT complexes 108
3.9 Summary 109
4. Strategies for native protein complex isolation 111
4.1 Previous attempts to purify eukaryotic TRiC 111
4.2 Endogenous purification from wild-type mammalian cells 112
4.3 Recombinant purification through Multibac vectors 116
4.3.1 MultiBac donors and acceptor vectors 117
4.3.2 Strategies for the MultiBac vector 118
4.3.3 Sub-cloning to generate constructs 120
4.4 CRISPR/Cas9 mediated genome editing of CCT3 124
4.4.1 Design of CCT3 tagging strategy 126
4.4.2 Assembly of CCT3 homology-directed repair templates 129
4.4.3 Transfection of CRISPR/Cas9 components for CCT3 genome editing 133
4.5 Selection and validation of CCT5 C-terminal affinity tag knock-in, as a viable strategy for isolating native TRiC 135
4.5.1 Validation of C-terminal affinity tagged CCT5 cell line 135

4.5.2 Small scale purification of CCT5-FLAG cell line successfully isolates native TRiC 136
4.5.3 Large scale purification from CCT5-FLAG line further confirms isolation of native TRiC 139
4.5.4 Final purification procedure for TRiC 143
4.6 Nanobody-mediated pull-down 144
4.6.1 Development of anti-CCT5 nanobody binders 145
4.6.2 Anti-CCT5 nanobodies can isolate native TRiC 148
4.6.3 Biophysical assessment of binding affinities for anti-CCT5 nanobodies 150
4.6.4 Functional analysis of TRiC through ATP activity 153
4.7 Summary 156
5. Structure and characterization of native human TRiC 157
5.1 Current understanding of TRiC-mediated substrate folding 157
5.2 Considerations for grid preparation and optimization 159
5.2.1 Sample heterogeneity 159
5.2.2 Ice thickness 160
5.2.3 Orientation bias 160
5.2.4 Conformational flexibility 162
5.2.5 Structural homology of TRiC subunits 162
5.3 Cryo-EM data collection and processing of TRiC-Nb18-ATP·AlF_x 163
5.4 Global structure of human TRiC in the closed state 166
5.4.1 Structural details of TRiC in the closed state 167
5.4.2 Inter-subunit interactions 170
5.4.3 Nanobody interaction with CCT5 173
5.5 Solution digest MSMS and crosslinking identified potential substrates bound within TRiC folding cavity 174
5.5.1 Crosslinking condition optimization 175
5.5.2 Cross-linking analysis of Nb18 and TRiC 177
5.5.3 Peptide enrichment in endogenous TRiC affinity purification 180
5.6 Substrates bound within TRiC closed state structure 182
5.6.1 3D classification of TRiC in the closed state reveals substrate densities with unique features 183
5.6.2 Identification of tubulin bound within folding cavity of TRiC in the closed state 185

5.6.3 Characterization of tubulin bound within folding cavity of TRiC in the closed state 190
5.7 Relaxed 3D classifications yield further substrate-bound reconstructions 194
5.7.1 Actin bound to the *cis* ring of TRiC in the closed state 195
5.7.2 Actin interactions with TRiC in the closed state 198
5.8 Co-chaperone PhLP2A is bound in the *trans* ring 203
5.8.1 Identification of PhLP2A as the second substrate bound within TRiC 205
5.8.2 Helical regions of PhLP2A hint at actin-regulating mechanism 209
5.8.3 Thioredoxin domain and C-terminus anchor PhLP2A in *trans* ring 212
5.9 3D reconstruction of TRiC in the open conformation 217
5.9.1 TRiC interactions with nanobody 18 and substrate in the open state 219
5.9.2 Differences between TRiC in the open and closed states 223
5.10 Functional analysis of TRiC through ATP activity **Error! Bookmark not defined.**
6. Discussion and conclusion 228
6.1 Overview of discoveries described in this thesis 228
6.2 Relevance to chaperonin protein folding and proteostasis 230
6.3 Limitations in TRiC folding pathway mechanism 232
6.4 Further work to understand eukaryotic group II chaperonin 234
References 239

List of Figures

Figure 1.1: Architecture of group I and group II chaperonins. 19
Figure 1.2: Early mechanistic understanding of eukaryotic group II chaperonin. 20
Figure 1.3: Domain architecture of actin and tubulin. 24
Figure 1.4: Domain architecture of a group II chaperonin subunit. 30
Figure 1.5: Sequence alignment of group I and group II chaperonin subunits. 33
Figure 1.6: Structural motifs in group II chaperonins. 35
Figure 1.7: Assembly of CCT subunits within TRiC complex. 37
Figure 1.8: TRiC subunit orientation and colour scheme for representation. 38
Figure 1.9: TRiC folding chamber is separated into positive and negatively charged hemispheres. 40
Figure 1.10: An ensemble of yeast TRiC structures reveals details of subunit movements throughout cycle of ATP hydrolysis. 44
Figure 1.11: Sequence conservation of nucleoside binding regions of chaperonin subunits. 46
Figure 1.12: Recent cryo-EM analysis reveals a pattern of substrate binding within human TRiC. 56
Figure 1.13: Updated mechanism of TRiC-mediated substrate folding.. 59
Figure 3.1: Purification of full-length CCT4 and CCT5. 79
Figure 3.2: CCT5 ATPase activity assay. 81
Figure 3.3: CCT5 aggregation suppression assay. 83
Figure 3.4: Thermal stability of CCT5 with ATP analogues. 85
Figure 3.5: Cryo-EM workflow. 87
Figure 3.6: Typical cryo-EM data processing workflow. 88
Figure 3.7: Preliminary grid screening and data collection of CCT4 and CCT5 homo-oligomeric complexes. 90
Figure 3.8: Representative cryo-EM micrographs from $CCT5_{16}$ grid screening. 91
Figure 3.9: $CCT5_{16}$-ATP-AlF_x data processing statistics. 92
Figure 3.10: Density map of $CCT5_{16}$-ATP-AlF_x. 94
Figure 3.11: All CCT subunits of can be solubly be over-expressed individually. 96
Figure 3.12: Large scale purifications of CCT subunits. 98
Figure 3.13: Representative SAXS analysis of CCT5 HMW and LMW samples.. 100
Figure 3.14. CCT subunit size and shape estimates using SEC-SAXS. 104
Figure 3.15: Melt temperatures (T_ms) of all eight CCT subunits. 106
Figure 3.16: CCT subunits measured for ATP hydrolysis activity. 107
Figure 3.17: Negative stain micrographs of CCT HMW species. 109
Figure 4.1: Expression levels of CCT subunits in HeLa and HEK293 cell lines. 112
Figure 4.2: Endogenous TRiC purification from WT-HEK293 cells. 114
Figure 4.3: Negative stain EM of endogenously purified TRiC from WT-HEK293 cells. 115
Figure 4.4: MultiBac cloning strategy for assembling eight-gene expression plasmid containing CCT1-CCT8. 119
Figure 4.5: Initial cloning phase of TRiC MultiBac vector assembly. 122

Figure 4.6: Recombination attempt of CCT2/CCT5 and CCT3/CCT6 vectors to form four-gene CCT2/3/5/6 pUCDM plasmid. 123
Figure 4.7: CRISPR/Cas9 mediated genome editing workflow to incorporate knock-in affinity tag at target site. 126
Figure 4.8: Target regions for tag insertions in CCT3. 128
Figure 4.9: PCR amplification of CCT3 target regions. 131
Figure 4.10: Assembly of homology-directed repair template for CCT3 R339 target site. 132
Figure 4.11: Assessment of CCT3-affinity tag insertion. 134
Figure 4.12: Identification of CCT5 C-term 3xFLAG/SpyTag knock in. 136
Figure 4.13: Small scale purification of CCT5 C-term 3xFLAG/SpyTag knock in. 138
Figure 4.14: Selection and large-scale purification of monoclonal CCT5 C-term edited cell lines. 140
Figure 4.15: Large-scale purification of TRiC from monoclonal CCT5 C-term edited cell lines. 142
Figure 4.16: Standard purification of large scale TRiC. 143
Figure 4.17: $CCT5_{16}$ purified for immunization in llamas. 146
Figure 4.18: Summary of 28 anti-$CCT5_{16}$ nanobody binders. 147
Figure 4.19: Nanobody purification and pull-down experiments. 149
Figure 4.20: Follow-up of nanobody-mediated TRiC isolation. 150
Figure 4.21: Binding kinetics of Nb-CCT5 complexes measured by BLI. 152
Figure 4.22: ATPase activity of $CCT5_{16}$ and TRiC with anti-CCT5 nanobody binders. 155
Figure 5.1: Plot of EM maps released in the public domain from 2001-2019. 159
Figure 5.2: Optimization of cryo grid conditions for TRiC. 161
Figure 5.3: Data processing of TRiC-Nb18-ATP·AlFx complex. 164
Figure 5.4: Reconstructions of TRiC-Nb18-ATP-AlFx cryo-EM data collection. 166
Figure 5.5: Atomic models of TRiC-Nb18 in the closed state. 168
Figure 5.6: Ligand binding within TRiC in the closed state. 169
Figure 5.7: Subunit assembly is mediated by β-sheet stacking between subunits. 171
Figure 5.8: N-terminal interaction network between *cis* and *trans* rings. 172
Figure 5.9: Nanobody binding to CCT5. 174
Figure 5.10: Optimization of crosslinking conditions to determine optimal ratio of crosslinker to chaperonin:nanobody complex. 177
Figure 5.11: Crosslinks identified by d0/d4 BS3 LC/MSMS. 180
Figure 5.12: Volcano plot of peptides identified in solution digest LC-MS/MS. 181
Figure 5.13: 3D classifications of closed state particles. 183
Figure 5.14: Tubulin bound within TRiC. 185
Figure 5.15: LC-MS/MS and structural analysis enable building of atomic model of tubulin. 188
Figure 5.16: Tubulin model bound to TRiC. 190
Figure 5.17: Overview of tubulin contacts with TRiC subunits. 191
Figure 5.18: Contacts between CCT subunits and tubulin within closed conformation. 194
Figure 5.19: Relaxed classifications of closed particles. 195
Figure 5.20: Actin-bound map in the closed conformation. 197
Figure 5.21: Actin-TRiC binding locations. 198

Figure 5.22: Interactions between actin and CCT8. 199
Figure 5.23: Interactions between actin and CCT6. 200
Figure 5.24: Interactions between actin and CCT3. 201
Figure 5.25: Interactions between actin and CCT1. 202
Figure 5.26: Interactions between actin and CCT4 and CCT7.. 203
Figure 5.27: Relaxed class 2 closed map remaining unmodelled density. 205
Figure 5.28: Sequence alignment of PhLP family. 206
Figure 5.29: Structure-based predictions of PhLP2A as unidentified substrate bound in *trans* ring of TRiC in relaxed class 2 reconstruction. 207
Figure 5.30: PhLP2A identified as second substrate in relaxed class 2 closed map.. 208
Figure 5.31: PhLP2A helix 1 interactions with actin and TRiC. 210
Figure 5.32: PhLP2A helix 2 and helix 3 interactions with TRiC..... 212
Figure 5.33: PhLP2A thioredoxin domain interactions with *trans* ring of TRiC.. 214
Figure 5.34: PhLP2A missing C-terminal region is proximal to TRiC subunit C-termini.. 215
Figure 5.35: CCT3 C-terminus binds between groove of CCT4 and CCT1.. 216
Figure 5.36: Model of TRiC in the open conformation. 218
Figure 5.37: Ligand occupancy in TRiC subunits in the open conformation. 219
Figure 5.38: Nanobody 18 binds CCT5 in the open conformation. 220
Figure 5.39: Open conformation of TRiC contains substrate bound.. 222
Figure 5.40: N-terminal interaction network between *cis* and *trans* rings does not exist for TRiC in the open conformation. 224
Figure 5.41: Stacking of CCT subunits differs between open and closed conformations.. 226
Figure 6.1: Proposed mechanism of TRiC-mediated substrate folding. 231

List of Tables

Table 2.1. Cryo-EM data collection, refinement, and validation statistics. 75
Table 3.1: SAXS data analysis parameters for all CCT subunits. 102
Table 4.1: Steady-state kinetics of Nb-CCT5 complexes measured by BLI. 153
Table 5.1: Key residues around CCT nucleotide binding pockets. 170
Table 5.2: List of peptides identified in purified TRiC sample. 178

1. Introduction

1.1 Protein folding is influenced by chaperone networks involved in proteostasis.

The health and survival of a cell is dependent on the stability and function of proteins that carry out DNA repair, conduct cellular transport, catalyse metabolic reactions, and provide structural stability, among other roles. To maintain cellular health proteins are “chaperoned” by a subnetwork of proteins that stabilise and/or fold polypeptide chains. This network, termed protein homeostasis (proteostasis), ensures a balance between biogenesis, stability, and degradation of proteins to maintain biological processes and promote healthy ageing[1]. While many proteins are able to spontaneously achieve their native fold by virtue of their amino acid properties under the “thermodynamic hypothesis”[2], some aggregate irreversibly without proper folding conditions or require further assistance to properly fold[3]. The latter category consists of large, topologically complex polypeptides with hydrophobic regions that are susceptible to aggregation when left unassisted in the crowded cellular environment[1]. These unstable nascent polypeptides require assistance from a class of proteins called molecular chaperones[4] that function to deliver and process nascent peptides into their final structures[5].

1.1.1 Chaperones promote proper folding and prevent aggregation of misfolded and unfolded proteins.

Chaperones maintain protein stability and reduce aggregation propensity by covering hydrophobic pockets that are exposed either by synthesis, refolding after membrane transport, thermal denaturation, or changes in protein-protein interactions[6]. Several chaperone groups exist that were originally characterised by size, including

stress proteins or heat shock proteins (Hsps) that see increased expression upon cellular stress conditions[1]. Another group includes the bacterial-mitochondrial-chloroplast chaperones and was coined the “chaperonin” group to exclude nucleoplasmins and Hsp70[6]. These chaperonins were subsequently shown to be conserved in all cellular kingdoms and are further sub-grouped into group I chaperonins and group II chaperonins[7–9]. Group I chaperonins are found in bacteria as well as organelles derived by endosymbiosis including chloroplasts and mitochondria, whereas group II chaperonins exist in archaea and the eukaryotic cytosol.

1.1.2 Group I chaperonins

The bacterial chaperonin was first identified as member of the Hsp60 family as a result of experiments on the GroEL/GroES system investigating protein maturation in bacteriophages[10]. The GroEL protein was first characterised and was shown to have adenosine triphosphate (ATP)-ase activity[11]. Structural characterisation by negative stain electron microscopy (EM) revealed that the ~1MDa complex contains 14 copies of the 60 kDa GroEL subunit arranged in two double-stacked rings of seven subunits each[12]. Later, another protein, GroES, was linked through genetic and biochemical analysis in *Escherichia coli (E. coli)* to GroEL and both are required for proper bacteriophage morphogenesis[13]. Along with the Rubisco-subunit-binding protein identified in chloroplasts[6] and the mitochondrial-associated Hsp60 in eukaryotes[14], these collectively became known as the group I chaperonins. The first crystal structure of GroEL from *E. coli*[15] revealed the domain structure of chaperonins and broadly showed cylindrical, ring-like structures containing two stacked rings with a large folding cavity in the middle of the complex. Each subunit contained: a large equatorial domain

that forms the base of the subunit involved in ring assembly involving the equatorial domain of the adjacent ring, and also contained the ATP-binding site; an apical domain that forms the lid region of the cylinder; and an intermediate domain that links the equatorial and apical domains together[15]. A subsequent crystal structure of the GroEL–GroES complex illustrated how GroES works in tandem with GroEL to form a lid for protein folding that occurs within the GroEL chamber[16] (Figure 1.1A).

1.1.3 Archetypes of Group II chaperonins – thermosome, TF55, and Mm-Cpn

Understanding of chaperonins was further expanded upon the discovery of structurally similar protein-folding complexes, named thermosome in the thermophilic archaebacteria *Pyrodictium occultum*[17] (Figure 1.1B) and TF55 in *Sulfolobus shibatae*[18]. These organisms thrive at high temperatures in environments that promote destabilisation and unfolding of polypeptide chains, and thus require protein chaperone machinery to survive. Structural investigations of these complexes revealed similar double stacked rings to GroEL that contain subunits with equatorial, intermediate, and apical domains. While GroEL is composed of seven subunits per ring, these archaeal chaperonins differ in ring composition: the thermosome contains two eight-membered rings composed of two different subunits known as α and β[19], while TF55 can contain both eight- and nine-membered rings built from one to three subunits depending on variation in temperature and heat/cold stress conditions in *Sulfolobus shibatae*[20]. However, the main difference from archaeal group II and bacterial group I chaperonins lies in the domain composition of these thermophilic chaperonins, which include an apical domain with a long helical protrusion that acts as a flexible built-in lid mechanism. Due to this built-in lid these archaeal chaperonins thus do not require a partner,

compared to the GroEL/ES system found in bacteria. A third example of archaeal chaperonins with a built-in lid mechanism is present in the *Methanococcus maripaludis* chaperonin (Mm-Cpn), which forms double eight-membered rings from only one subunit, and has been widely studied as an easily isolated complex for structural and biochemical studies[21] (Figure 1.1C).

1.1.4 Eukaryotic CCT/TRiC

The eukaryotic chaperonin, belonging to group II, was first identified as a result of mutational experiments in mice involving spermatogenesis that cause a tail-less phenotype from the recessive *t* allele, hence the naming of the first identified subunit in this multi-subunit complex as tail-less complex polypeptide 1 (TCP-1 or CCT1)[22,23]. The observation that the bacterium TF55 and eukaryotic TCP-1 share 40% sequence homology suggested that TCP-1 was the eukaryotic homologue of the archaeal group II chaperonins[18]. This was confirmed upon purification and visualisation *via* negative stain EM of TCP-1 isolated from yeast that was part of a larger chaperonin complex containing eight different subunits[24]. Discovery of eight TCP-1 related subunits led to the naming of the eukaryotic chaperonin as chaperonin containing TCP-1 (CCT) or TCP-1 containing ring complex (TRiC)[22] and is herein referred to as TRiC (Figure 1.1D). The eight individual subunits CCT1–8 (sometimes referred to for bovine and human TRiC using Greek letters CCTα-β-γ-δ-ε-ζ-η-θ), share approximately 25–40% sequence identity[25] and are expected to adopt similar topology. An additional isoform of CCT6 in mouse testes led to the distinction that CCT6A is the general somatic isoform and CCT6B is present mainly in testes[26]. In this thesis CCT6 refers to CCT6A.

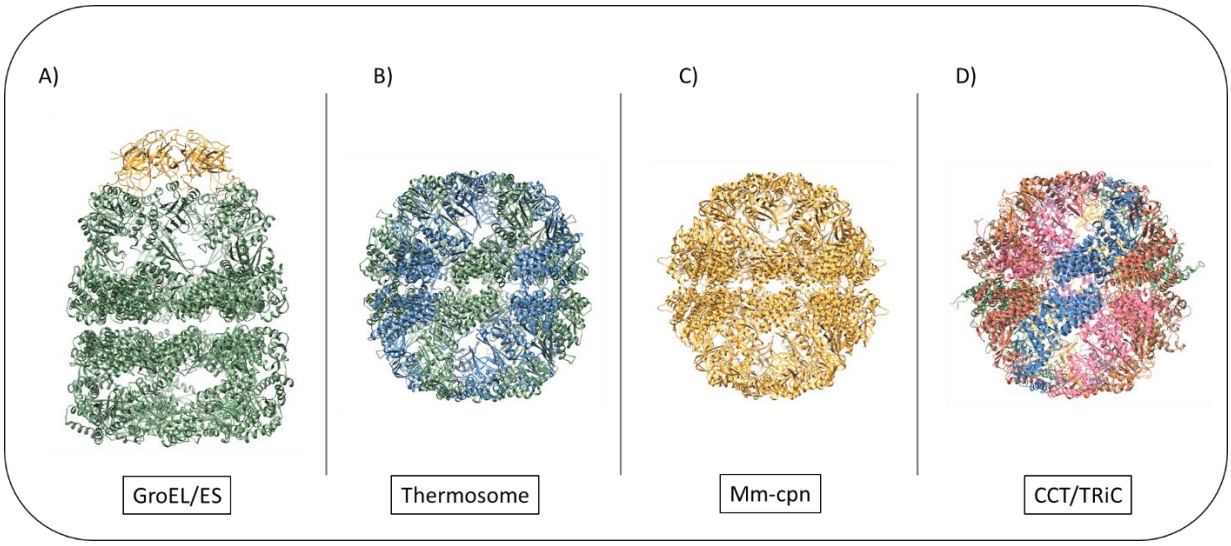


Figure 1.1: Architecture of group I and group II chaperonins. A) Crystal structure of GroEL/ES complex from *E. coli* at 3.0 Å resolution (PDB: 1AON)[16]. GroEL is in green and GroES in gold. B) Crystal structure of thermosome from *Thermoplasma acidophilum* at 2.6 Å resolution (PDB: 1A6D)[19], composed of two subunits. Alpha subunit is in green and beta subunit is in blue. C) Crystal structure of homo-oligomeric chaperonin from *Methanococcus maripaludis* (Mm-Cpn) at 3.2 Å resolution (PDB: 3KFB)[21]. D) Cryo-EM structure of human TRiC at 3.0 Å resolution (PDB: 6KS6)[27], with eight constituent subunits coloured differently. Protein models were visualised using ChimeraX throughout this thesis[28,29].

The eukaryotic chaperonin TRiC also showed biochemical evidence of ATP activity and refolding of actin, tubulin, and luciferase *in vitro*, illustrating that this chaperonin binds to nascent polypeptides in their unfolded state[24,30,31] and is required for tubulin biogenesis[32]. Further biochemical analysis of the yeast TRiC complex with actin revealed two separate states based on binding of ATP that became known as the open and closed conformations[33,34]. Bovine TRiC was structurally shown in cryo-electron microscopy (cryo-EM) experiments to also contain a helical lid protrusion that adopts open and closed conformations based on ATP binding and hydrolysis[35]. Later, additional substrates beyond actin and tubulin were identified for TRiC, including cell-cycle regulators, mitochondria-targeted proteins, and ribosomal biogenesis factors[36–38] (for full list see Appendix 1). These early results suggested a mechanism whereby nascent polypeptides are synthesized by the ribosome (Figure 1.2: step 1) and engage

the eukaryotic group II chaperonin (Figure 1.2: step 2), which subsequently binds and hydrolyses ATP (Figure 1.2: step 3) to induce closure of the subunits, thus trapping nascent polypeptides within the chamber created by the ring-like formation of the complex to promote native substrate folding and release into the cytosol (Figure 1.2: step 4). However, these details required further evidence to confirm the substrate folding mechanism and to fill in the intermediary gaps including how substrate is delivered and details of the substrate folding intermediates while engaged with the chaperonin.

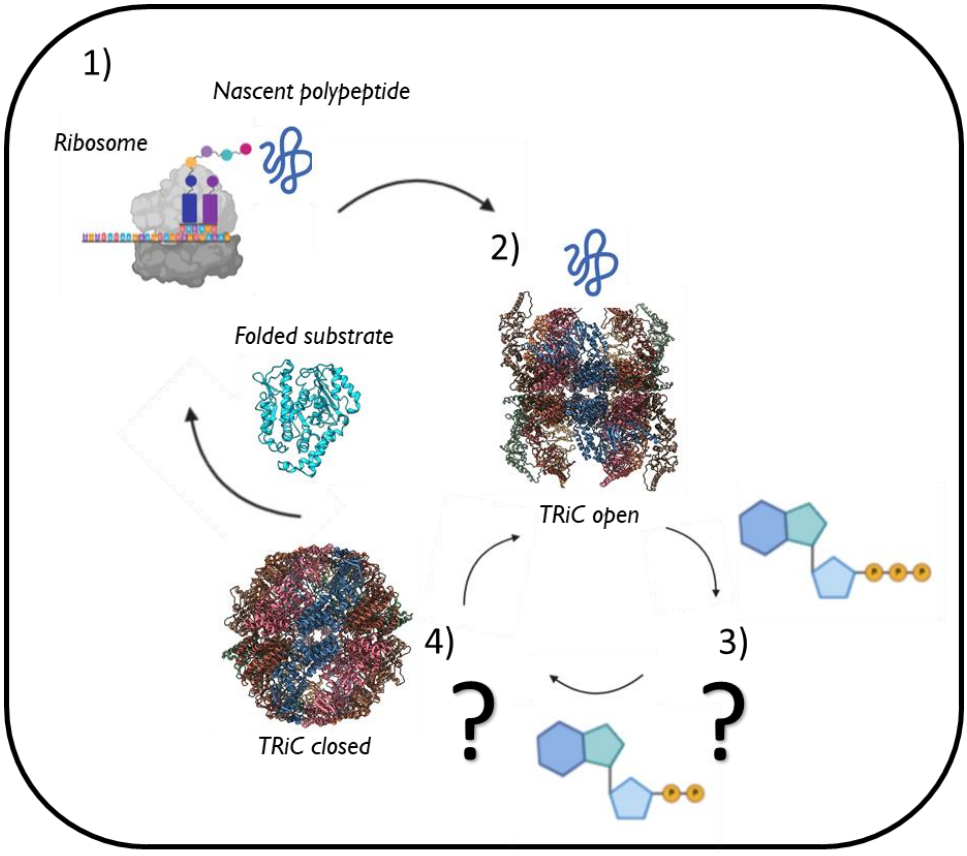


Figure 1.2: Early mechanistic understanding of eukaryotic group II chaperonin. 1) Nascent polypeptides are synthesized by the ribosome and 2) engage chaperonin in the open state. 3) ATP then binds and is hydrolysed to close the complex. 4) Substrate achieves its native fold and is released to the cytosol.

1.1.5 Over-expression of individual CCT subunits leads to homo-oligomeric complex formation

Recombinant expression of individual CCT subunits provided an avenue to reconstitute the human TRiC complex *in vitro* for characterisation, while avoiding the arduous task of isolating all eight subunits of the TRiC complex together by co-expression. Recombinant over-expression of individual subunits in *E. coli* revealed a surprising result that some subunits, notably CCT4 and CCT5, can form 16-subunit homo-oligomeric complexes *in vitro* that are similar in size and architecture to TRiC and retain ATP hydrolysis activity[39]. Studies using Mm-Cpn provided biochemical assays and structural information[40–42] that were developed to characterise group II chaperonins and could be extended to study these $CCT4_{16}$ and $CCT5_{16}$ homo-oligomers. One such assay investigated the ability of chaperonin to reduce the aggregation of certain client substrates, a phenomenon originally observed for polyglutamine tracts[43,44]. The substrate human γ-D crystallin (HγD) is known to associate with group II chaperonins including Mm-Cpn and was used as an example in this assay[40]. Mm-Cpn is able to reduce aggregation kinetics of HγD *in vitro*, an observation that was extended to include $CCT4_{16}$ and $CCT5_{16}$, therefore providing evidence that these homo-oligomers are functionally active[39].

The recombinant expression system established for individual CCT subunits also allowed for controlled, pair-wise co-expression to investigate the mechanism of TRiC complex assembly from eight subunits[45]. Using a bicistronic expression vector, each CCT subunit was sequentially co-expressed with CCT5 one at a time. All subunits except CCT6 experienced a shift in molecular weight in sucrose gradient

ultracentrifugation, indicative of complex formation. The same experiment with CCT4 only showed complex formation with subunits CCT5 and CCT8 to form chaperonin rings. This analysis suggests CCT5 is the most important subunit for assembly of TRiC, however more work is needed to determine the hierarchical process by which CCT subunits come together to form an eight-membered ring[45].

1.2 Substrates and cellular functions of TRiC

1.2.1 Early substrates identified as TRiC interacting partners

High-throughput mass spectrometry and proteomics advances have identified a host of potential TRiC substrates and modulators in yeast and mammalian cells. The proteins identified were so numerous that TRiC is expected to be involved in interacting with ~10% of the proteome across a wide spectrum of protein families[36,46–48]. These potential client/substrate proteins span many cellular functions including cytoskeletal assembly, cell cycle regulation, protein degradation, DNA repair, mitosis/meiosis, metabolism, RNA processing, transcription/translation, cellular trafficking, and viral assembly[49] (Appendix 1). Importantly, no overall conservation in structure or function of client substrates exists, other than a propensity for beta-sheet formation and hydrophobic residues in their sequences[36,50,51]. More investigation is therefore needed to understand the degree to which TRiC operates in the cytosol to act as a modulator of proteostasis.

The most abundant and characterised TRiC substrates include actin and tubulin[44,49,52–56]. Actin and tubulin are structural proteins that form the cellular cytoskeleton through assembly of microfilaments and microtubules, respectively. Filamentous actin (F-actin) is the most abundant eukaryotic protein and is comprised of

actin monomers (G-actin) that contain a small and large lobe that are discontinuous in sequence. These lobes are further divided into two subdomains (1 and 2; 3 and 4) separated by a cleft that is the site of nucleotide-Mg^{2+}/Ca^{2+} binding[57] (Figure 1.3A). Six actin isoforms exist in mammals including: α-actin, of which there is skeletal, cardiac, and smooth actin that are each expressed in their respective muscle tissues; smooth γ-actin, which is expressed in smooth muscle tissues; and β- and γ-cyto actins that are ubiquitously expressed[58]. Tubulin microtubules are formed by polymers of α- and β-tubulin that bind guanosine triphosphate (GTP) to promote assembly of microtubules (Figure 1.3B: top). A tubulin monomer comprises an N-terminal nucleotide-binding Rossmann fold domain (aa 1-205), an intermediate Taxol-binding domain (TBD; aa 206-381), and a C-terminal motor protein binding domain[59] (Figure 1.3B: bottom).

Alongside early associations of TRiC with actin and tubulin, TRiC was also identified as the molecular target of the antihistamine drug clemastine in *Plasmodium falciparum*, which was shown to bind to CCT subunits and destabilise the mitotic spindle through a disruption of TRiC-mediated tubulin folding[60]. Similarly, the sphingolipid 1-deoxydihydroceramide was shown to cause anoxic-related death by impairing TRiC-mediated actin folding[61]. These pathway disrupting agents provide further evidence of the importance of TRiC for actin and tubulin biogenesis.

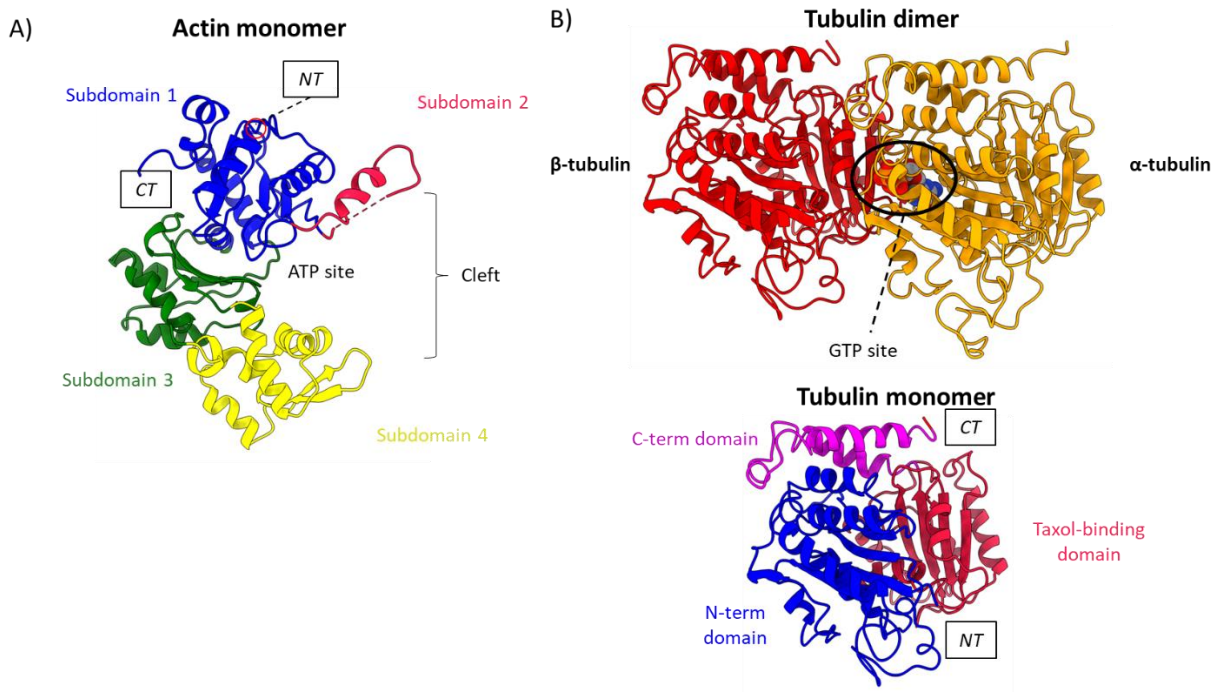


Figure 1.3: Domain architecture of actin and tubulin. A) Structure of Actin monomer from *Oryctolagus cuniculus* by X-ray crystallography at 2.7 Å resolution (PDB: 2Q1N)[62] displaying subdomains 1 (blue, 2 (red), 3 (green), and 4 (yellow) with ATP site and cleft indicated. B) Top: Structure of dimer of α- (orange) and β-tubulin (red) from *Sus scrofa* by electron crystallography at 3.7 Å resolution (PDB: 1TUB)[59]. GTP molecule (silver) is rendered as spheres. Bottom: tubulin monomer displaying N-terminal domain (blue), Taxol-binding domain (red), and C-terminal domain (magenta).

1.2.2 TRiC is modulated by co-chaperones including prefoldin and the phosducin-like protein family.

To maintain proteostasis in the cell, TRiC can recruit several chaperone partners (co-chaperones) that assist or modulate its protein folding activity. Two known co-chaperones of TRiC include prefoldin (PFD) and the phosducin-like proteins (PhLPs). Prefoldin is an ATP-independent co-chaperone around 100 kDa in size that is composed of two α and four β subunits[63] (for structure see Figure 1.12D). Prefoldin acts by binding nascent polypeptides to shield them from aggregation between ribosomal synthesis and delivery to the TRiC apical domains[64]. This co-chaperone was originally

shown to bind actin and tubulin, and, similar to TRiC, has been more recently shown to interact with many other proteins including those involved in cancer, neurodegeneration, and other cellular pathways[65–70]. While PFD is not essential for eukaryotic cell viability, disruption of PFD binding to TRiC causes widespread protein aggregation *in vivo* in yeast, establishing its role as an essential chaperone for proteostasis[71].

The phosducin-like protein family contains several members that play a role in G protein signalling including phosducin-like protein (*PDCL*, or PhLP1), phosducin-like protein 2 (*PDCL2*, or PhLP2B), phosducin-like protein 3 (*PDCL3*, or PhLP2A), and thioredoxin domain-containing protein 9 (TXND9)[72]. Both PhLP1 and PhLP2A were identified as ATPase regulators of TRiC and act as inhibitors of substrate folding when highly expressed by reducing TRiC ATPase activity[73–75]. These co-chaperones of TRiC are necessary for correct cytoskeletal assembly likely by acting antagonistically with PFD to establish a balanced modulation of cytoskeletal biogenesis[73,75].

1.2.3 TRiC function is involved in multiple cellular pathways

The cellular roles of TRiC are continuously updated as more evidence is discovered of processes that are mediated by its protein folding activity, and of client proteins identified to require TRiC. Studies in *Saccharomyces cerevisiae*[76] and *Drosophila melanogaster* demonstrated that TRiC is required for cell growth and actin cytoskeleton organisation through its interaction with the target for rapamycin (TOR) pathway, specifically TOR, Rheb, and S6K, as exhibited by a reduction in growth upon knock-down of CCT4 expression[77]. Reduction in TRiC also led to a decrease in phosphorylation of S6K and S6, and insulin/TOR signalling was shown as necessary to induce expression levels of TRiC[77]. The mechanistic target for rapamycin (mTOR) forms

a complex (mTORC1) with mammalian lethal with SEC13 protein 8 (mLST8) and the regulatory associated protein of mTOR (Raptor)[78], both of which were suggested to interact with TRiC to achieve their native folds before forming the mTORC1 complex with mTOR[76,79–81]. The mTOR protein kinase is a major gene regulator and thus this association with TRiC is a potential target for drug discovery[82].

A role in cytoskeletal regulation has also been reported in which CCT1 has been shown to associate with the human protease HtrA3 *in vivo*[83]. Additionally, depletion of any CCT subunit leads to a headless phenotype in the *Drosophila melanogaster* pupal stage, indicative of an important role in organ growth. CCT subunits have also been studied regarding their role in axonal transport in neurons. In mouse models CCT1 and CCT3 expression have been shown to be linked to anterograde and retrograde transport of brain-derived neurotrophic factor (BDNF)[84]. Expression of CCT5 is also linked to BDNF in wild type neurons and, interestingly, the tau protein, known for its involvement in Alzheimer's disease, is required for this effect and led to increased CDK5 activity, tau phosphorylation, and effects on microtubule bundling[85]. TRiC is also implicated in regulation of heat shock transcription factor 1 (HSF1), and the small-molecule activator of HSF1, HSF1A, inhibited TRiC activity in an ATP independent manner[86].

The TRiC folding pathway is also important for assembly of protein complexes. For example, TRiC is involved in post-translational modifications though its role as an obligate folder of histone deacetylase (HDAC) assembly through a direct interaction with HDAC1, and the N-terminus of HDAC1 was shown to be important for substrate release from the TRiC folding chamber[87]. Recent studies have also shown that TRiC and Hsp70

are recruited to specific topologically complex domains of nascent polypeptides, with Hsp70 acting first to bind select domain sites followed by TRiC engagement with the near-natively folded polypeptide second, in a manner that correlates with slowdown of translation as a further mechanism to regulate protein folding[88]. G proteins form complexes consisting of Gα, Gβ, and Gγ subunits that first require formation of a Gβγ dimer. However, each protein is unstable on its own and requires TRiC and PhLP1 to form a stable dimer[74,89]. Additionally, the acute myeloid leukaemia 1 eight twenty-one (AML1-ETO), a fused transcription factor caused by the chromosomal translocation t (8;21)(q22;q22), has been implicated in acute myeloid leukemia[90] and was shown to achieve its native fold through interactions with Hsp70 and TRiC[91]. The tumour suppressor p53 is also associated with TRiC as deletion of TRiC leads to an increase in misfolded p53, suggesting that TRiC functions in p53 related cancers[92]. Taken together, TRiC is instrumental in a wide range of important cellular functions, many of which have yet to be fully characterised (Appendix 1).

1.2.4 TRiC is an important factor in a wide array of molecular diseases

Many cancers have been associated with TRiC function or malfunction; an unsurprising feature due to the large number of substrates implicated in the substrate folding pathway (Appendix 1). Overexpression of CCT subunits was observed in triple-negative breast cancer cell lines[93–96], colorectal cancer cell lines[69,94,97], glioblastoma[98], non-small[99] and small cell lung cancer[100], multiple myeloma[101], liver cancer[102–104], and gastric cancer cell lines[94,105]. Consequently, a therapeutic peptide called CT20p has been investigated as a potential therapy for cancer cell lines implicated in TRiC, due to its selective cytotoxicity through inhibition of TRiC in small cell lung cancer[93,100]. Aside

from common disease, hereditary sensory neuropathy (OMIM 256840) is the only known genetic disease directly caused by autosomal recessive mutations on genes encoding CCT subunits. These mutations are located in the equatorial domains of TRiC subunits, and are present on the exterior of the given subunit including a C450Y substitution in CCT4 found in Sprague–Dawley rats[106] and a H147R substitution in CCT5 found in humans[107]. Modulation of TRiC through individual CCT subunits therefore has potential therapeutic benefit for human diseases.

1.2.5 TRiC is an attractive therapeutic target for disease.

A major disease area that has been linked to TRiC, with great therapeutic need, includes diseases caused by protein aggregation. TRiC can suppress aggregation of toxic proteins involved in neurodegenerative diseases including mutant huntingtin. Studies have shown depletion of TRiC in yeast and human cells leads to accumulation of aggregated mutant huntingtin exon 1 (mhtt), particularly through interactions with CCT1 and the N-terminal region of mhtt[108,109]. *In vivo* experiments confirmed that exogenous delivery of a CCT1 fragment was sufficient to reduce huntingtin aggregation in rat cells expressing mhtt[110]. Structural studies using tomography also showed that $CCT5_{16}$ and TRiC reduce aggregation of mhtt not by actively sequestering or refolding mhtt as expected, but instead by acting as a cap at the end of the fibrils to block further aggregation[111,112]. Additionally, TRiC was shown to inhibit aggregation of α-synuclein associated with Parkinson's disease[113] and is associated with Alzheimer's disease in genome-wide association studies[114]. Investigations in *Drosophila* and mouse tissues shed light on the mechanism by which TRiC reduces aggregation of toxic substrates. Instead of acting directly on substrates, the loss of function of specific subunits

(CCT2/5/7) caused unfolding and depolarisation of actin assembly that led to downstream effects on autophagic flux and caused subsequent accumulation of mutant huntingtin, polyA stretches, mutant ATXN3, and p62[115]. TRiC is also implicated in prion protein binding, which binds tightly to TRiC and prevents its ability to fold native substrates[116]. TRiC can therefore act as a potential therapeutic agent to mediate neurodegenerative diseases linked to protein aggregation.

TRiC has also been investigated as an anti-viral target due to its role in viral replication and has been implicated in the assembly of viral proteins. This includes the hepatitis C virus genome between CCT5 and the viral NS5B protein[117], the Epstein-Barr virus through an interaction between CCT5 and the viral nuclear protein EBNA-3[118], CCT2 and the Zika virus protein NS1[119], CCT3 and nucleo- and phosphoproteins of rabies virus[120], CCT3 and p4 of Mason-Pfizer monkey virus[121], and CCT3 with HIV-1 protein p6[121–123]. TRiC is also implicated in viral capsid assembly in reovirus[124,125] and hepatitis B[126] and could be a target for therapy to treat a host of viral pathogens. As mentioned before (page 23) TRiC has been identified as a target of the antihistamine clemastine in *Plasmodium falciparum*, suggesting a possible treatment for malaria[60]. Additionally, subunits CCT4 and CCT5 are implicated in the uptake of toxic bacterial glycosylating proteins from *Clostridium difficile* and could be a target for toxic bacterial effectors[127]. The role of TRiC in infectious disease is has potential therapeutic benefit, however further investigation into these folding mechanisms is required.

1.3 Architecture of group II chaperonins from structural studies

1.3.1 Domain organisation of TRiC

The general architecture of group II chaperonins was first identified through negative stain and TRiC from yeast, mouse, and human tissues[12,15,24,128]. Structures were improved in subsequent X-ray crystallographic and cryo-EM experiments with the thermosome, TF55, Mm-Cpn, and TRiC[16,19–21,129–133]. These studies revealed conserved structural homology among both group I and group II chaperonin subunits. Each subunit contains an ATP-binding equatorial domain, an intermediate linker domain, and a substrate binding apical domain, with the N- and C-termini of the polypeptide chain beginning and ending in the equatorial domain (Figure 1.4) with varying sequence identity between the eight subunits (Figure 1.5).

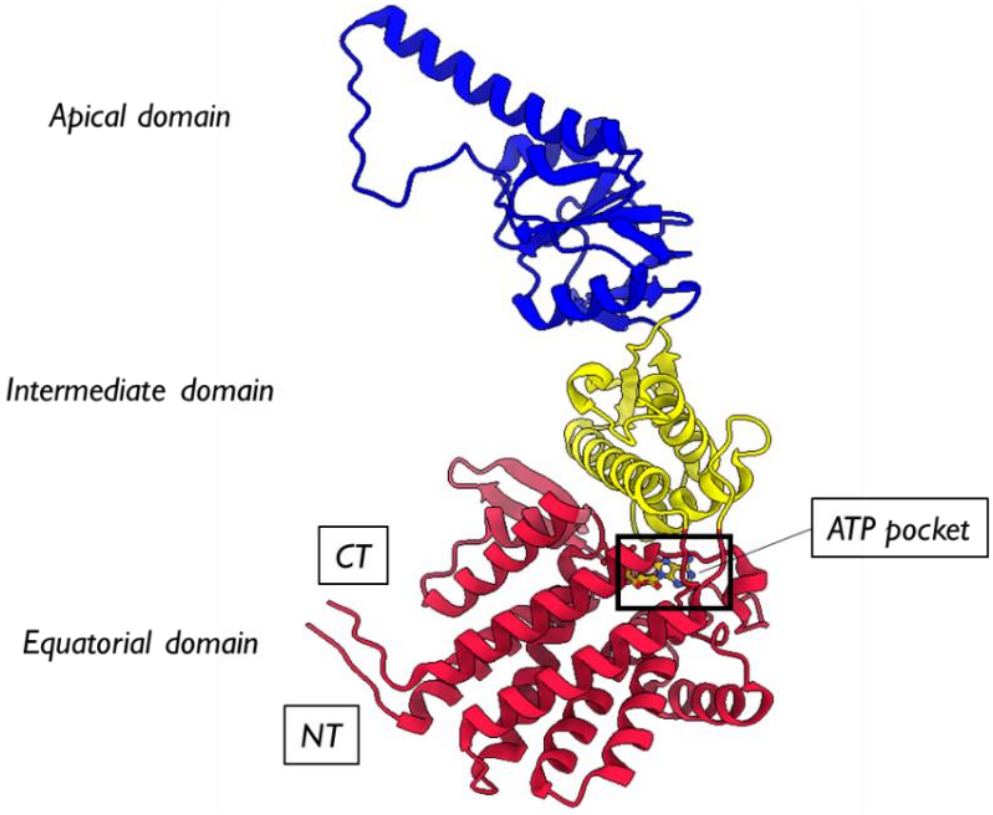
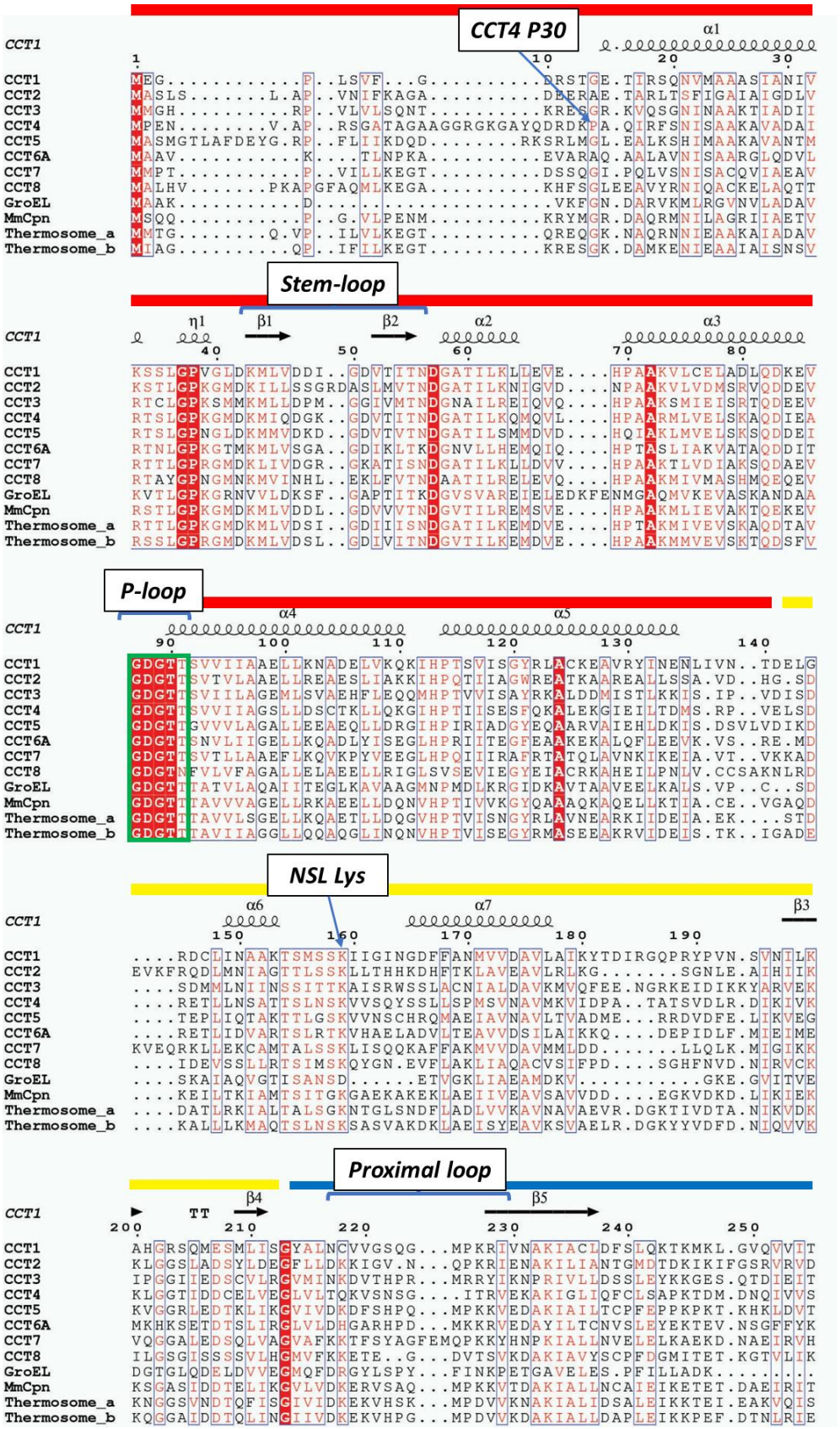
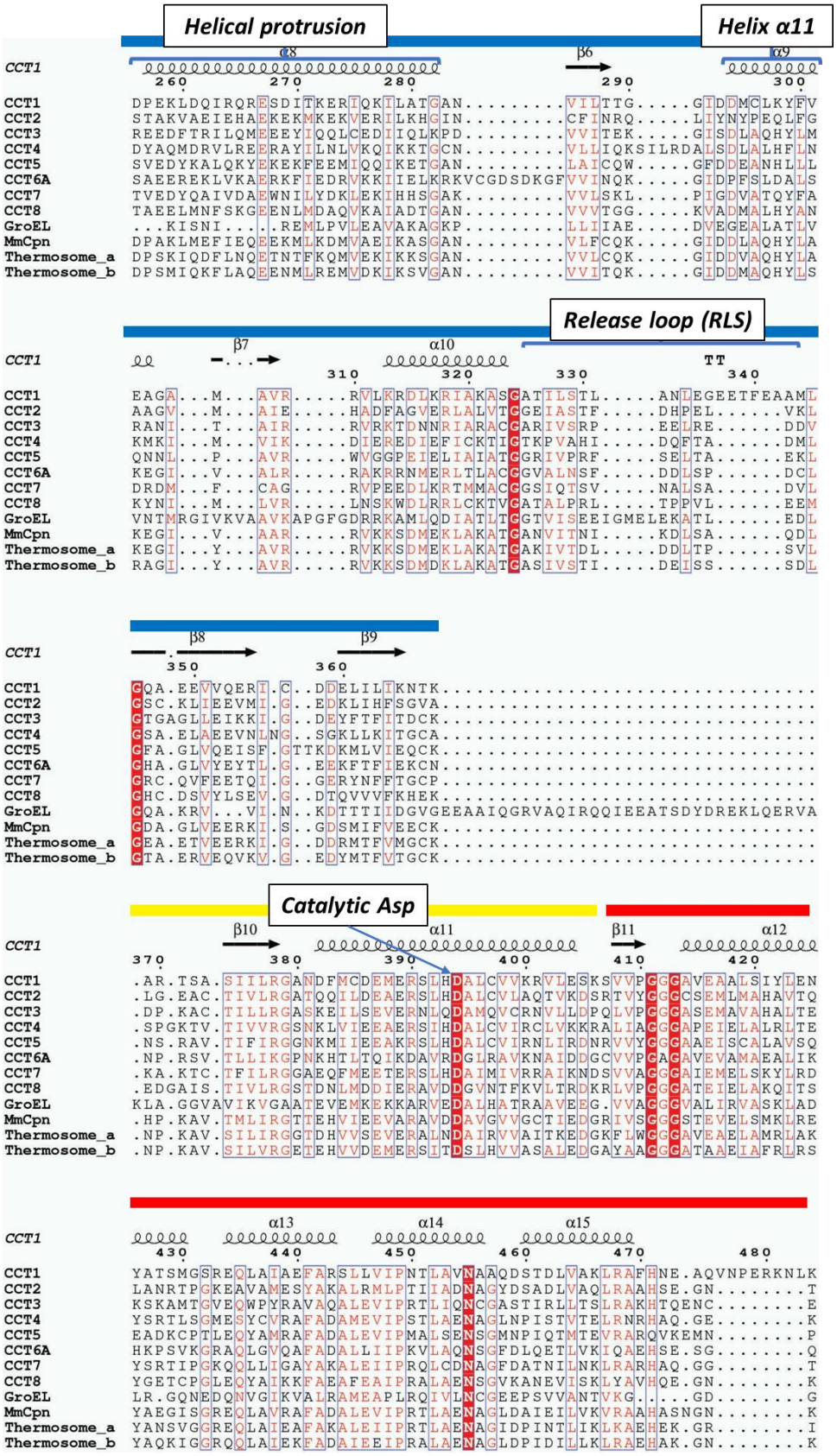


Figure 1.4: Domain architecture of a group II chaperonin subunit. CCT5 subunit from yeast TRiC at 3.0 Å resolution (PDB: 6KS6)[27] containing ATP-binding equatorial domain (red), intermediate domain (yellow), and substrate-binding apical domain (blue). NT and CT indicate location of N- and C-termini within the equatorial domain. Black box denotes ATP binding pocket.


CCT4 P30
Stem-loop
P-loop
NSL Lys
Proximal loop
CCT1
CCT2
CCT3
CCT4
CCT5
CCT6A
CCT7
CCT8
GroEL
MmCpn
Thermosome_a
Thermosome_b


Helical protrusion
Helix α11
Release loop (RLS)
Catalytic Asp
CCT1
CCT2
CCT3
CCT4
CCT5
CCT6A
CCT7
CCT8
GroEL
MmCpn
Thermosome_a
Thermosome_b

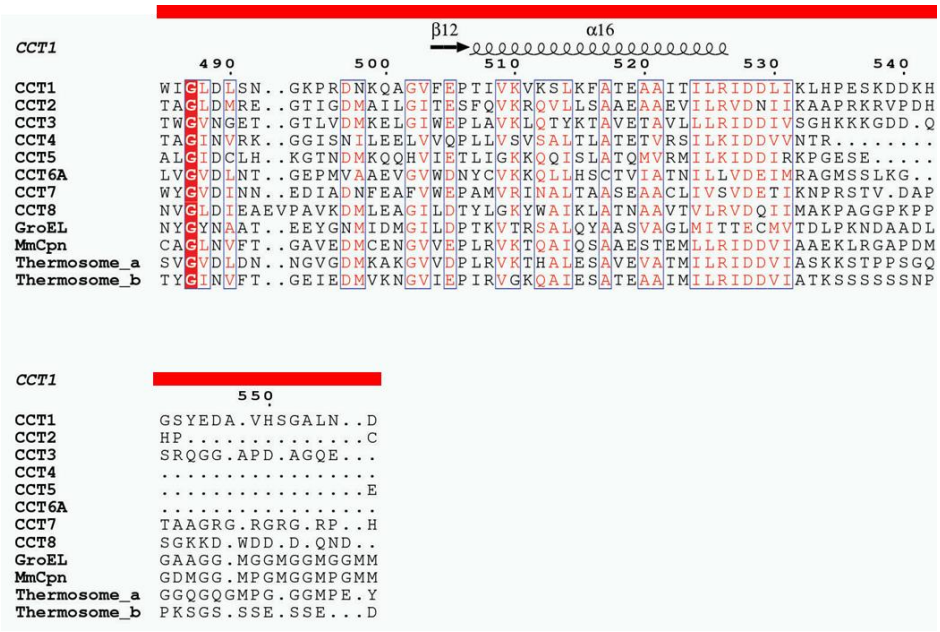


Figure 1.5: Sequence alignment of group I and group II chaperonin subunits. Sequence alignment of TRiC subunits from human, GroEL from *E. coli*, Mm-Cpn from *Methanococcus maripaludis,* and thermosome subunits from *Thermoplasma acidophilum*. Homology is indicated by coloured text and boxes in the sequence. Secondary structure is labelled on top of alignment using CCT1 as reference (PDB: 6NRA chain A). Domains are indicated by bars above alignment including equatorial (red), intermediate (yellow), and apical (blue). Conserved regions of chaperonins are indicated above alignment with blue arrows for single residues and blue brackets for motifs. Conserved P-loop containing GDGTT is further indicated by green box.

The equatorial domains from the eight constituent TRiC subunits are highly conserved in amino acid sequence, and include a loop containing the residues GDGTT near the ATP binding pocket called the phosphate-binding loop (P-loop; Figure 1.5), while sequence divergence is high among the apical domains of subunits[134]. The ATP binding pocket also contains a conserved catalytic aspartic acid that lies in an intermediate domain helix positioned directly above the bound ligand (Figure 1.5).

The highest resolution structure for any TRiC complex to date is from yeast TRiC, determined in the closed conformation at 3.0 Å resolution (PDB: 6KS6)[27]. This level of detail revealed homologous structural motifs that have been observed across the chaperonin spectrum. These include: a helical protrusion that forms the built-in lid,

the helix α11 and proximal loop motifs that are involved in substrate recognition, and the release loop of substrate (RLS) that is involved in substrate release from the apical domains; a nucleotide-sensing loop (NSL) that contains a lysine that interacts with ATP to sense occupancy, and a catalytic aspartic acid that helps coordinate hydrolysis of ATP in the intermediate domains; the P-loop motif in the equatorial domain for ATP binding, and a stem-loop that contains two β-strands that is important for subunit stacking as well as substrate recognition in the equatorial domains[27,135] (Figures 1.5 and 1.6). Mutations in the sensor and apical loops as well as the catalytic Asp were not able to rescue cells subjected to CCT2 RNA interference, suggesting that these residues are essential for TRiC activity and cell survival[136]. An allosteric network formed by the terminal regions of the CCT subunits has also been described (see next section) that likely plays a role in complex formation and ring communication, however the exact details of this interaction network remain unclear[27].

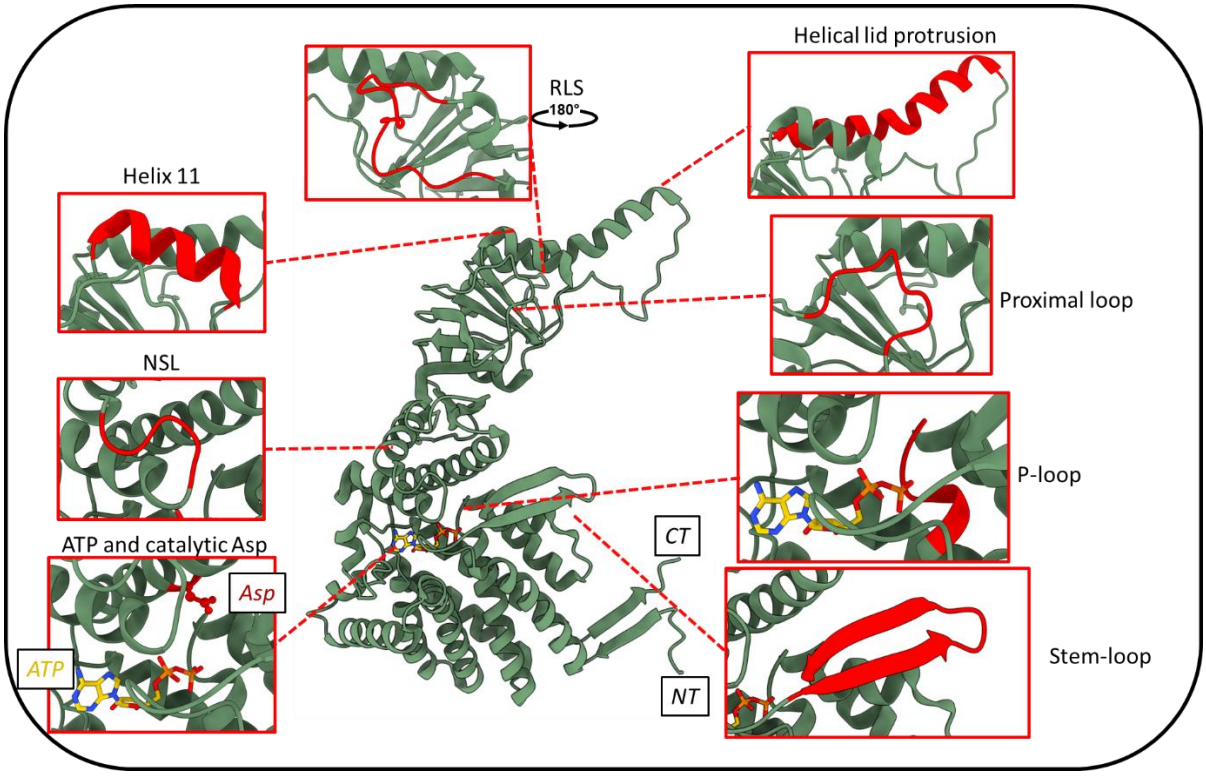


Figure 1.6: Structural motifs in group II chaperonins. CCT structural motifs represented by CCT1 (green) from yeast TRiC at 3.0 Å resolution (PDB: 6KS6)[27] including: β-hairpin stem-loop (aa 40-60), P-loop (aa ~87-107), nucleotide-sensing loop (aa ~159-180), proximal loop (aa ~210-230), helical lid protrusion (aa ~250-290), Helix 11 (aa ~290-305), release loop of substrate (RLS, aa ~320-350, and catalytic Asp (red ball-and-stick, aa ~391-407). Insets show zoomed image of motif coloured in red. ATP (gold) rendered as sticks.

1.3.2 Organisation of subunits within the TRiC complex

The subunits of TRiC are arranged such that one ring contains eight subunits (*cis*/intra-ring) that stack on top of the equatorial domain of a given subunit from the opposite ring (*trans*/inter-ring). The intra-ring assembly is mediated by a stack of four β-strands that are formed by the N- and C-termini of one subunit and the β-hairpin stem-loop of an adjacent subunit (Figure 1.6). This β-sheet formation applies to all subunits except CCT4 in which the N-terminus instead extends away from the folding cavity towards the outside of the complex. This change in positioning can be explained by a

proline present instead of a glycine at the N-terminal junction to helix 1[132] (Figure 1.5), however the purpose for this difference is not yet understood. The assembly of each ring forms a central folding cavity protected from the crowded cellular environment that provides a location for substrates to achieve their native folds (Figure 1.7B). The stacking of two rings creates a septum between the two subunits that divides the TRiC complex such that two folding chambers (*cis* top ring and *trans* bottom ring) exist per molecule (Figure 1.7C). Although the theoretical substrate size limit based on the dimensions of the TRiC folding cavity is similar to that of GroEL at around 70 kDa[16,137] and spans a diameter of approximately 60 Å[138], several substrates including the heavy meromyosin sub-fragment of myosin and hSnu114 that are larger than this cut-off have been reported, whereby domain-specific folding occurs[139,140].

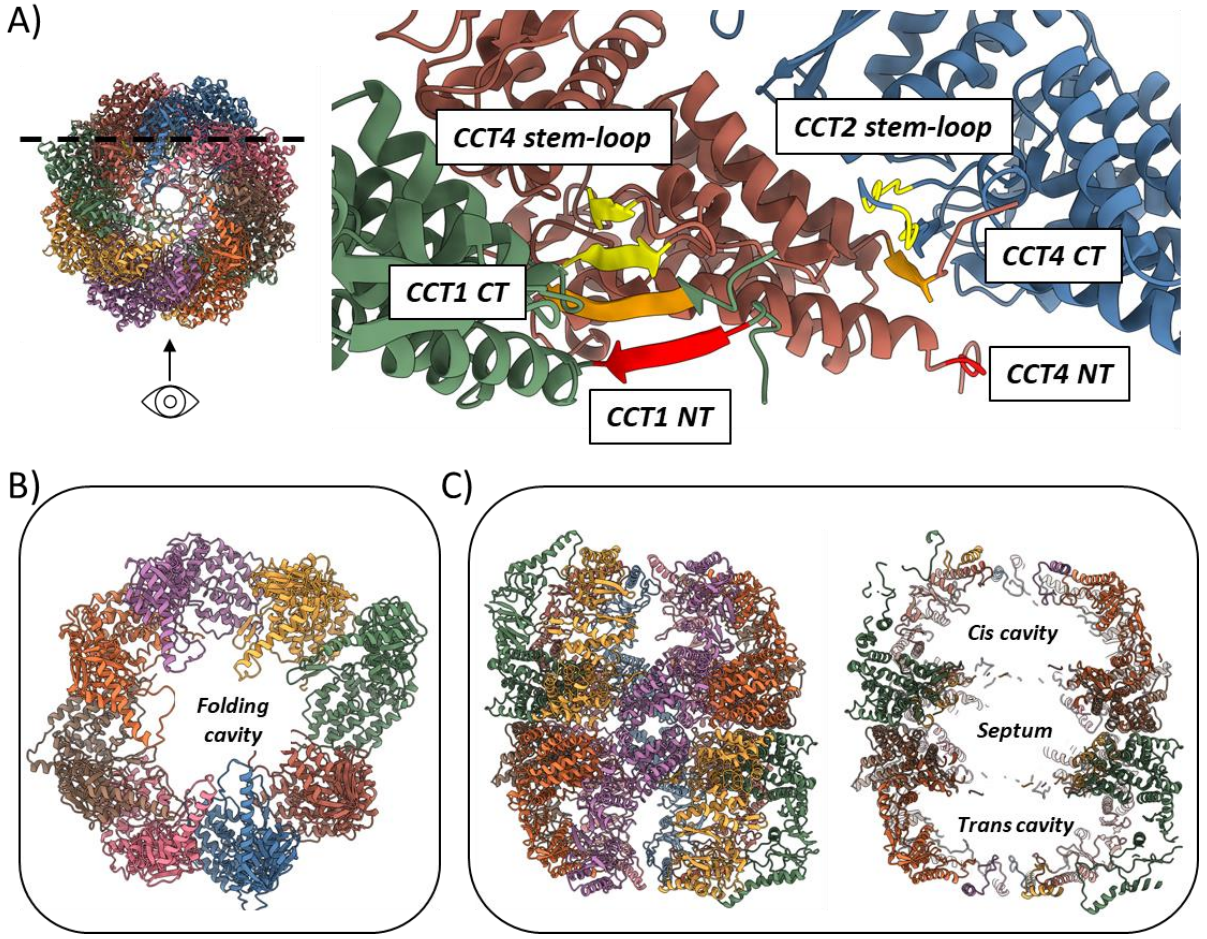


Figure 1.7: Assembly of CCT subunits within TRiC complex. A) Stacking of CCT1 (green), CCT4 (dark red), and CCT2 (blue) subunits within one ring from yeast TRiC at 3.0 Å resolution (PDB: 6KS6)[27]. N-terminal (red), C-terminal (orange), and stem-loop (yellow) beta strands form the inter-subunit beta-sheet stacking mechanism that promotes subunit assembly in the closed state. Stacking between CCT1 and CCT4 shows typical four β-sheet pattern for all subunits except between CCT4 and CCT2 where the N-terminal β-strand of CCT4 is not involved in stacking. B) Top-down view of one TRiC ring of eight CCT subunits in the open state (PDB: 6NRA)[71] showing the protein folding cavity created in the middle of the ring. C) Side view of both rings of TRiC in the open state (PDB: 6NRA)[71] including full view (left) and slice (right) revealing septum formed by the double stacking of two rings as well as *cis* and *trans* folding chambers.

1.3.3 CCT subunit orientation within the TRiC complex

Due to the structural homology between the eight CCT subunits, the definitive subunit orientation within the CCT ring assembly, and if one existed at all, was a source of debate due to ambiguity in early structural data[130,141]. Two studies investigating

subunit orientation, based on chemical cross-linking coupled with mass spectrometry (XL-MS), concluded the clockwise orientation of one ring to be CCT1-CCT4-CCT2-CCT5-CCT7-CCT8-CCT6-CCT3 (Figure 1.8). Subunits CCT2 and CCT6 create C2 symmetry by facing opposite one another while also stacking on its identical pair in the opposite ring[132,142] (Figure 1.8). This proposed subunit orientation was later experimentally confirmed by cryo-EM studies of yeast TRiC in which individual subunits contained bulky green fluorescent protein tags that could be used to orient particles and identify unique features of individual subunits[138,143,144]. This orientation consensus allowed for additional study of the protein folding chamber by informing future structural investigations of the expected ring order and identifying features of individual subunits.

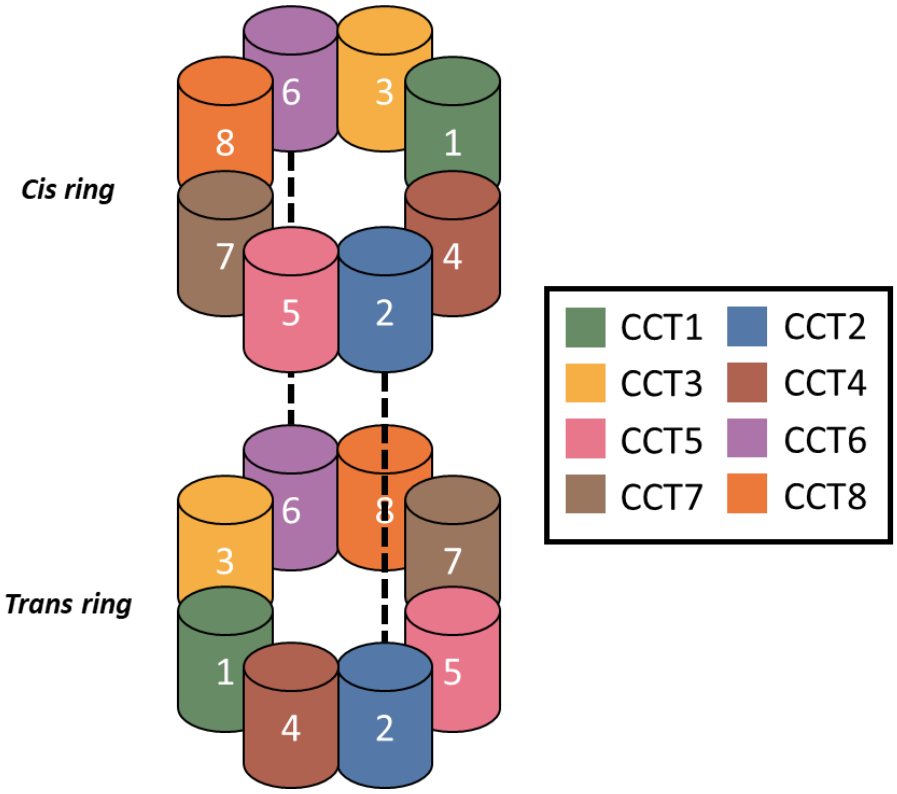


Figure 1.8: TRiC subunit orientation and colour scheme for representation. The consensus TRiC subunit orientation for each ring is clockwise: CCT1-CCT4-CCT2-CCT5-CCT7-CCT8-CCT6-CCT3. The two rings stack under this orientation, such that CCT2 and CCT6 nest on each other in the *cis* and *trans* rings (dotted lines). The colour scheme for subunits shown here is used throughout figures in the thesis.

Analysis of the residues facing this internal chamber between different subunits revealed a partitioning of TRiC into two hemispheres based on electrostatic properties. Within the intermediate domains the subunits CCT4-CCT2-CCT5 contain positively charged residues and form one side of the folding chamber (CCT2 hemisphere), while CCT8-CCT6-CCT3 contain negatively charged resides and form the other side (CCT6 hemisphere)[132] (Figure 1.9). These two hemispheres with asymmetric charges are suspected to play a role in guiding substrate folding between entering the chamber and subsequent lid closure after ATP hydrolysis. The negatively charged CCT6 hemisphere is particularly likely to play a role in substrate folding as the cavity exposed region of GroEL is also negatively charged[145]. Thus, a pattern could be envisioned in which substrates are partitioned to the CCT6 hemisphere by the positively charged CCT2 hemisphere. To further characterise the differences between the intra-ring subunits additional experiments were performed to investigate whether CCT subunits also had asymmetric ATP binding properties.

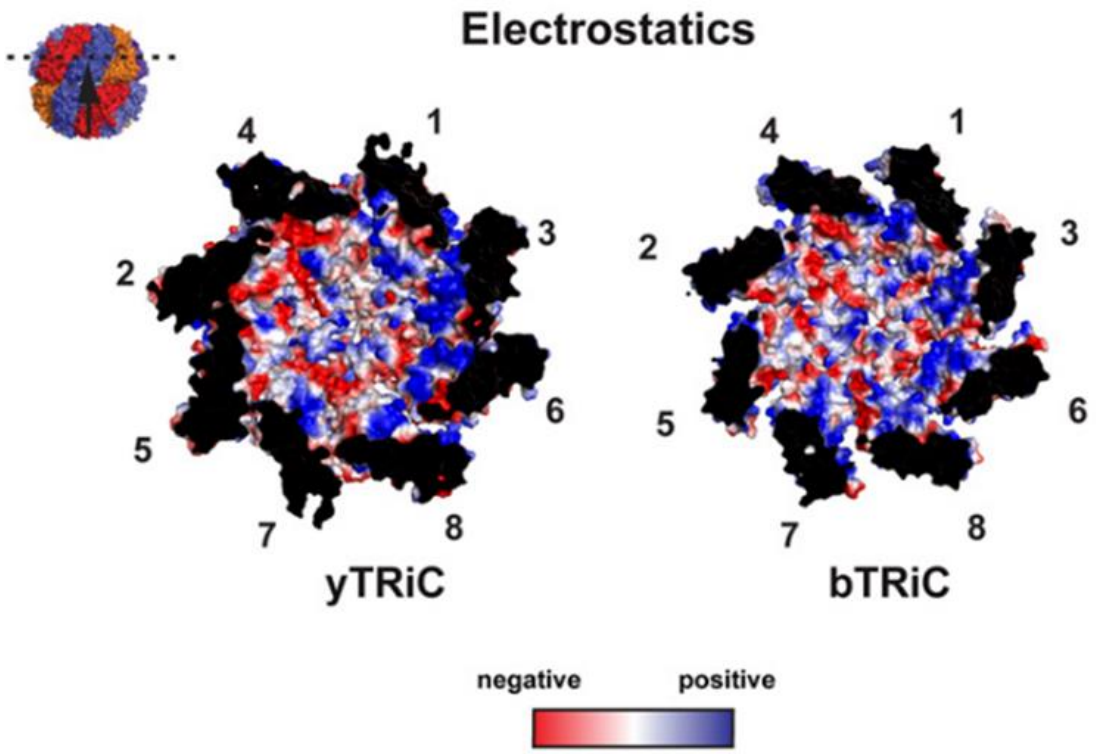


Figure 1.9: TRiC folding chamber is separated into positive and negatively charged hemispheres. Electrostatic charge distribution of positive (blue), neutral (white), and negative (red) residues within the folding cavity of yeast TRiC (yTRiC) and bovine TRiC (bTRiC). The folding cavity shows one mostly positively charged hemisphere including CCT1-CCT3-CCT6-CCT8 opposite a hemisphere containing residues with mostly negative and neutral charges including CCT7-CCT5-CCT2-CCT4. Figure borrowed from Leitner A., *et al* 2012[132]. with permissions from Elsevier.

1.4 Role of ATP binding and hydrolysis

1.4.1 ATP binding and hydrolysis plays a key functional role in TRiC-mediated substrate folding

Early studies of chaperonins described the importance of ATP binding and hydrolysis for the proper folding of substrates. In the case of TRiC and homologues from lower organisms, the ATP binding site is located in a surface exposed pocket formed in the equatorial domain of each subunit (Figure 1.6). This pocket is formed by several conserved motifs that bind ATP and coordinate ATP hydrolysis (Figure 1.5,

Figure 1.6, Table 5.2). Binding of ATP has been shown to be important for rearrangement of subunit conformation and allows the complex to engage substrate in a pre-hydrolysis state, in which subunits are still open and the cavity is accessible to the cellular environment[138]. Subsequent hydrolysis of ATP causes a more dramatic conformational change that promotes movement of the subunits to a fully closed state where substrates are trapped inside the complex and the lid protrusions of the apical domains close off access to the cellular environment[27]. These structural changes are detailed below.

1.4.2 Structural studies reveal consequences of ATP binding and hydrolysis

The greatest understanding of the TRiC mechanism from a structural basis has come from recent ensembles of cryo-EM structures from yeast TRiC treated with different nucleotide concentrations[27,138]. The TRiC conformational cycle begins with all CCT subunits in their fully open, nucleotide-free states poised to engage with unfolded polypeptides, with the apical and intermediate domains of CCT2 bending a full 120° and tucking behind the adjacent CCT5 apical domain (Figure 1.10A). This state was termed the nucleotide partially preloaded (NPP) state due to the observation that residual adenosine diphosphate (ADP) remains bound in the subunits CCT3/6/8 after hydrolysis has occurred from one TRiC cycle. This unique conformational state was achieved by studying TRiC in the absence of nucleotide, and is the only such example shown in the literature to date[138]. Upon addition of the nucleotide adenylyl-imidodiphosphate (AMP-PNP), an ATP analogue that promotes the open state, TRiC undergoes a slight rearrangement with respect to the apical domains (Figure 1.10B). This feature, seen in multiple chaperonin small angle X-ray scattering (SAXS) and cryo-EM

experiments[42,129,138,146], results in an arrangement of a ‘tetramer of dimers’, in which the CCT subunits are grouped together as four groups of two subunits, based on proximity of the apical domains. These four groups include CCT1-CCT3, CCT2-CCT4, CCT5-CCT7, and CCT6-CCT8 pairs (Figure 1.10C). Additionally, as compared with the NPP state the CCT5 apical domain closes towards the folding cavity by 14° rotation and CCT7 closes by 60°, indicating these subunits, along with CCT2, undergo significantly more conformational change than the rest of the CCT subunits[138] (Figure 1.10D).

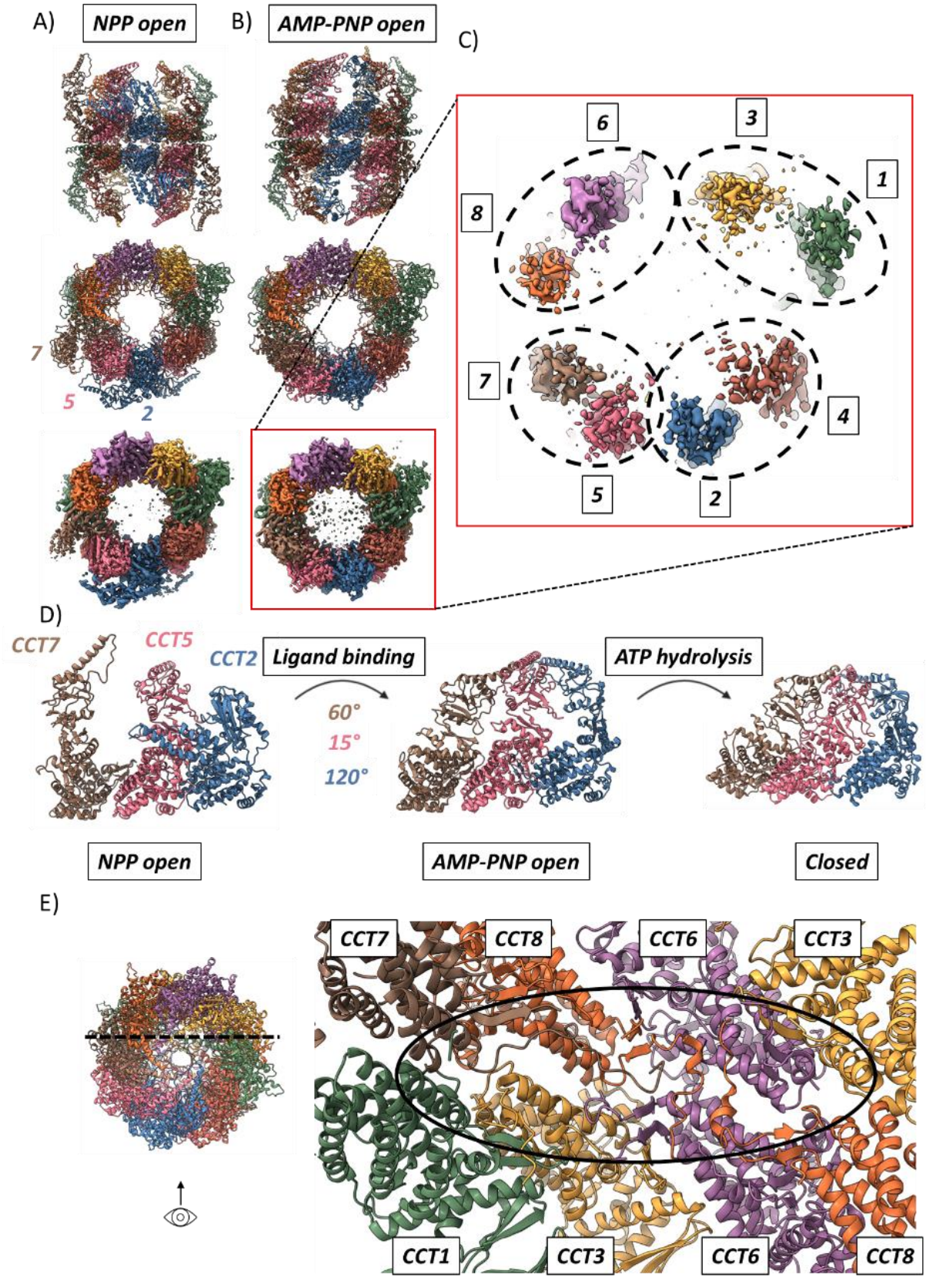

A)
NPP open
B)
AMP-PNP open
C)
6
3
1
8
7
4
5
2
7
5
2
D)
CCT7
CCT5
CCT2
Ligand binding
60°
15°
120°
ATP hydrolysis
NPP open
AMP-PNP open
Closed
E)
CCT7
CCT8
CCT6
CCT3
CCT1
CCT3
CCT6
CCT8

Figure 1.10: An ensemble of yeast TRiC structures reveals details of subunit movements throughout cycle of ATP hydrolysis. A) Yeast TRiC in the nucleotide partially preloaded (NPP) state (PDB: 5GW4)[138] shown as ribbon model in side view (left), top-down view (centre), and top-down map (right). B) Yeast TRiC in AMP-PNP bound state (PDB: 5GW5)[138] shown in the same manner as panel A. C) Slice of apical domains in AMP-PNP state in top-down view shows CCT subunits in a 'tetramer of dimers' (dashed circles). D) NPP state (left) and AMP-PNP state (centre) models showing rotational movement of CCT7, CCT5, and CCT2 upon ligand binding followed by closed state (right). Angles of apical domain rotation from open to closed are coloured by subunit. E) TRiC in the closed state (post ATP hydrolysis, PDB: 6KS6)[27] showing the inter-ring interaction network (black circle) represented by the N-termini of CCT3, CCT6, CCT7, and CCT8 to form tight stacking with the trans ring

.

Additional intermediary structures of the conformational cycle were resolved in this yeast TRiC study by titrating ATP and the components $Al(NO_3)_3$ and NaF to mimic the post-hydrolysis state. It is hypothesised that ADP remains bound in the closed conformation along with aluminium fluoride (AlF_X) in the γ-phosphate position, thus inducing a closed conformation. This study led to a proposed mechanism whereby individual subunits in the open state bind ATP at different stages to trigger hydrolysis. In this model, subunits begin to shift from the NPP state to the partially open AMP-PNP state in a stepwise manner as separate CCT subunits undergo conformational changes at different nucleotide concentrations. These show a sequential mechanism whereby CCT7 initiates ring closure, followed by CCT2, and a final step in which CCT4 binds nucleotide and acts as a sensor to initiate hydrolysis of ATP and subsequent ring closure. Of note, CCT8 had little movement between the open and closed states and did not exchange nucleotide, suggesting it does not play a large role in ring closure and may function more in substrate binding or TRiC assembly[27]. Upon ATP hydrolysis the equatorial domains sit in the centre of the folding chamber such that the septum area is constricted. This conformational change is mediated by the N-termini of each subunit (except CCT4) to lock with the opposite ring and induce a tightly closed conformation, thus confining the substrate to promote proper folding[27,132] (Figure 1.10E).

1.4.3 Kinetic characterisation of ATP hydrolysis in CCT subunits

Investigation into ATP binding kinetics of individual subunits showed differences in ATP affinity that can be divided into two hemispheres, consistent with the electrostatic hemispheres previously described[132], in which the CCT2 hemisphere subunits (CCT1/2/4/5) have high affinity for ATP, with CCT5 and CCT4 having the highest affinities, while the CCT6 hemisphere subunits (CCT3/6/7/8) have low affinities[147]. This was observed in both open and closed states, indicating that ATP hydrolysis by the low-affinity subunits is not essential for TRiC function. Additionally, on average only four subunits per ring are occupied in TRiC under saturating ATP conditions, providing further evidence that not all subunits contribute to ATP binding equally[147,148]. The variation in nucleotide binding can be attributed to divergence in residues that interact with the nucleoside region of ATP rather than those that interact with the triphosphate region including the nucleotide-sensing lysine and catalytic aspartic acid (Figure 1.5 and Table 5.1). Further to this point, nucleoside contacts are more conserved among high-affinity subunits with the highest conservation to other chaperonins seen in CCT4, suggesting CCT4 plays an important role in ATP activity, while CCT6 showed the highest sequence diversity consistent with a lack of detectable ATP binding[147] (Figure 1.11A). One possible explanation for why low affinity subunits are less likely to exchange nucleotide[138] could be due to steric clashes in residues created near the nucleoside region, including those seen at positions Asp499 (compared to conserved Glu/Gln residues) in CCT6 and CCT8 and Tyr47 (compared to conserved Gly-Pro loop) in CCT8 (Figure 1.11B)[149].

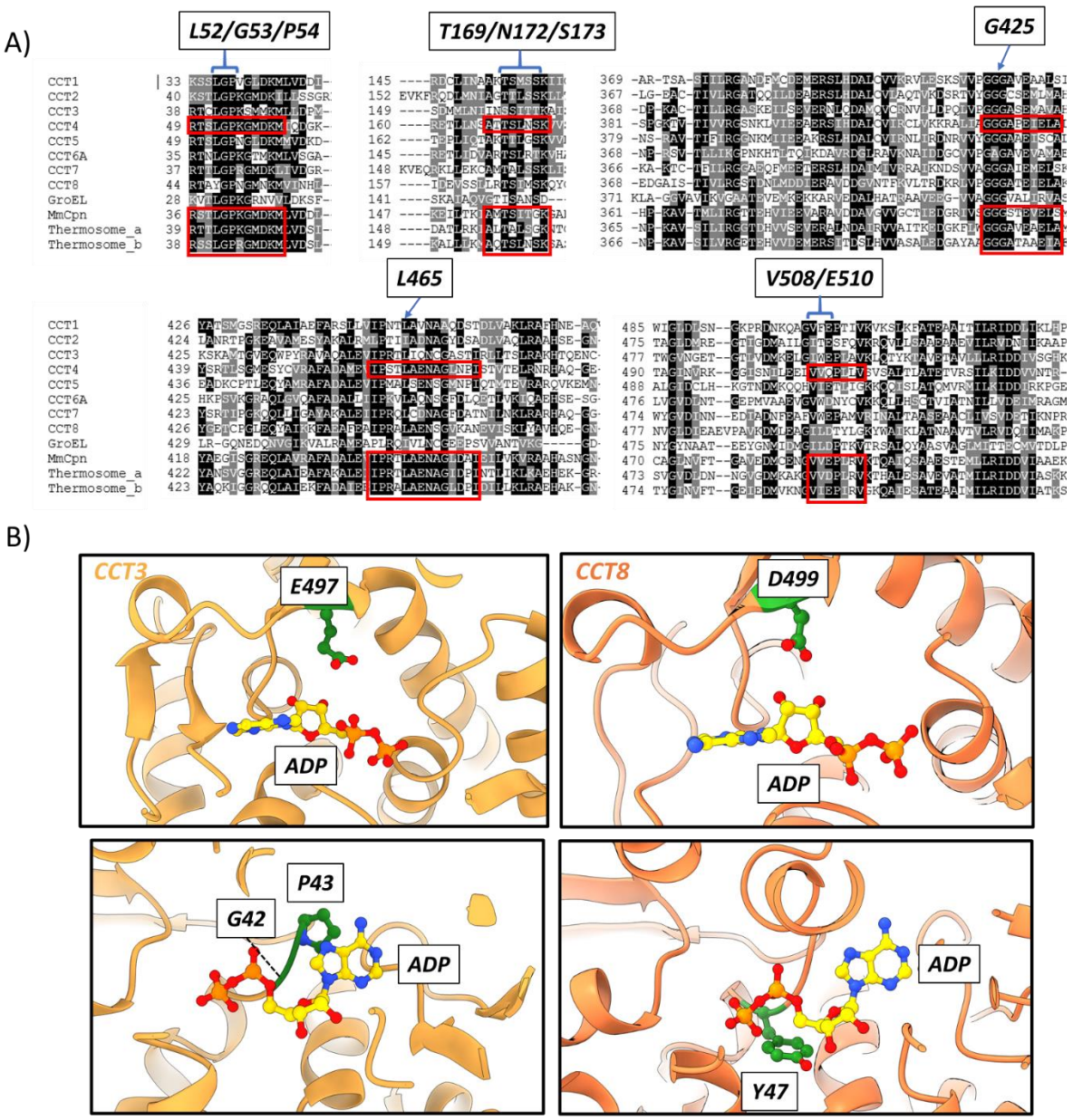


Figure 1.11: Sequence conservation of nucleoside binding regions of chaperonin subunits. A) Focused sequence alignment of group I and group II chaperonin subunits labelled with conserved nucleoside/TRiC subunit residues (black boxes and blue arrows/brackets). Red boxes show conserved regions between lower order group II chaperonins and CCT4, the most conserved of the CCT subunits to other chaperonins. B) Ribbon diagram of nucleotide pocket from yeast TRiC in the closed state at 3.0 Å resolution (PDB: 6KS6)[27] showing specific nucleoside binding residues (green) around ADP (gold). Conserved residues are represented by CCT3 (yellow) and are compared to divergent residues causing potential steric clashes in CCT8 (orange). Top panels compare conserved Glu/Gln to divergent Asp seen in CCT6 and CCT8 (position E510 in panel A); bottom panels compare conserved Gly-Pro loop present in all subunits to CCT8, which contains a tyrosine at this position (position L52/G53/P54 in panel A).

On a macromolecular level, ATP hydrolysis assays in bovine TRiC exhibited concentration-dependent affinities for ATP characterised by positive intra-ring cooperativity and negative inter-ring cooperativity, meaning affinities for ATP are higher when both rings are unoccupied with nucleotide and affinity decreases as ATP sites are bound[150]. This allosteric cooperativity profile is thought to encourage ATP binding within a single ring so the chaperonin molecule only utilises one ring at a time to act as a two-stroke motor[150]. The biphasic hydrolysis kinetics of TRiC reveal an initial fast phase of ATP hydrolysis activity that includes a dependence on ATP concentration based on initial velocities of ATP hydrolysis. As ATP concentration is increased these initial velocities demonstrate a slow phase with hyperbolic dependence on ATP concentration that plateaus at 0.2 mM[151]. Taken together the ATP kinetics of TRiC reveal a sequential mechanism of ATP turnover suggesting lid closure does not take place in a concerted fashion as is the case for group I chaperonins[152]. Further investigation into the ATP-induced lid mechanism showed that binding of ATP is not sufficient to induce TRiC lid closure and instead requires the ATP hydrolysis reaction[146]. This was further demonstrated by mutation of the catalytic aspartic acid in the intermediate domain of CCT4 to alanine, which inhibited TRiC closure[147]. The lid domain was also shown to play an important role in allosteric cooperativity as it is not required for ATP binding but is necessary for lid closure[153].

1.4.4 ATP activity of some subunits is essential for cellular health

Mutational analyses of CCT subunits in *Saccharomyces cerevisiae* have also provided better understanding of the importance of each subunit relative to its ATP hydrolysis activity. Mutation of the P-loop residues (GDGTT) to five alanine residues in

each subunit exhibit lethal phenotypes in yeast in the four high-affinity subunits, while those in the low-affinity subunits had no effect on growth[147]. A milder mutation of the P-loop in which the aspartic acid was changed to an alanine also causes a lethal phenotype in CCT4, further suggesting an important role for CCT4 in proper functioning of the conformational cycle[147]. As previously mentioned, a separate mutation of the conserved catalytic aspartic acid to alanine has provided a useful tool to study the necessity of ATP activity for each chaperonin subunit, as this mutation allows subunits to bind, but not hydrolyse ATP, as shown for Mm-Cpn where lid closure could no longer occur when this mutation was present[153,154]. Replicating this mutation in yeast TRiC has showed that there were no growth effects in low-affinity subunits, while high affinity subunits have showed a range of phenotypes including lethality in CCT4, severe growth defects in CCT5, and mild growth defects in CCT1[147], altogether highlighting the importance of this hemisphere to ATP activity and cell viability. Further to this, mutations that introduced a bulky negatively charged glutamic acid in place of the conserved catalytic aspartic acid also produced varying phenotypes in yeast among the different CCT subunits including actin accumulation for CCT2 and aggregation for CCT7[80]. Similarly, mutations in the P-loop motif differ between high and low affinity subunits as well. For example, mutation of this loop in CCT6 did not cause serious growth effects[155], while mutation of the P-loop in CCT4 exhibited growth defects when a single G345D variant was introduced that also abolished intra- and inter-ring cooperativity in ATP binding[151]. These observations together show the importance of ATP activity is different among the CCT subunits.

1.5 Mechanism of substrate binding

1.5.1 Substrate recognition at the surface apical domain region of TRiC

The importance of ATP-dependent substrate folding by TRiC was established early on, however the definitive mechanism of substrate binding remained elusive for some time. Early cryo-EM studies of actin,[156] tubulin,[157] and prefoldin-actin[158] bound to TRiC revealed a general mechanism whereby nascent polypeptides synthesized by the ribosome are delivered *via* the helical protrusions of the prefoldin co-chaperone to TRiC. Upon engagement with TRiC by prefoldin, these unfolded substrates bind near the apical domains of specific CCT subunits. In the case of tubulin, three subunits were shown to interact with the apical domains of TRiC[156], while five subunits were shown to interact with one tubulin dimer[157]. However, these models were of low resolution and were subsequently found to be based on an incorrect subunit orientation. Therefore, higher resolution snapshots of the substrate binding apical regions were necessary to understand the basis of substrate recognition in TRiC. Apart from structures of the entire TRiC chaperonin, crystallisation of the human CCT3 apical domain provided the first opportunity to reveal the organisation of the polar and electrostatic interactions of the apical loop protrusion that confers subunit specificity to CCT3, particularly as it applies to tubulin[159].

The study of subunit apical domains through a combination of biophysical, computational, and XL-MS experiments provided an enhanced understanding of their roles in substrate recognition[160]. In this study several known apical domain–substrate partners were used, including the CCT1 apical domain (ApiCCT1) with the hydrophobic Box1 motif of von Hippel Lindau tumour suppressor (VHL)[161], and the CCT3 apical

domain (ApiCCT3) with the human immunodeficiency virus (HIV) protein p6[121]. Nuclear magnetic resonance (NMR) perturbation experiments revealed a mechanism in which sequence divergence in CCT apical domains leads to different polar and hydrophobic residues in helix 11 and the proximal loop apical region among different CCT apical domains (Figure 1.5). This pattern displayed a dual mode of substrate recognition in which both polar and hydrophobic residues play a role in substrate recognition, expanding the complexity and number of substrates that can engage the eight different CCT apical domains[123]. XL-MS experiments further located the contact points between TRiC and three full-length substrates: actin, tubulin, and HIV Gag. These experiments showed a multivalent contact pattern between TRiC and its substrates, in which one substrate binds to more than one CCT subunit, shedding light on how TRiC can bind a diverse array of substrates. These data suggest that binding to CCT apical domains orients the substrate in a specific manner to allow for proper folding while associated with the chaperonin. When comparing chaperonin complexity and proteome size, a trend emerges whereby an increased number of eukaryotic chaperonin subunits also corresponds to an enhancement of protein diversity over an evolutionary scale[160]. The binding specificity of the eukaryotic chaperonin apical domains with substrates, and the observation that GroEL does not fold actin[162] and tubulin[163], also suggest that these, and possibly other substrates, co-evolved with TRiC during the course of evolution. This evidence provides further rationale for how the apical domain diversity between different CCT subunits plays an important role in the recognition and folding of a wide range of substrates[123].

1.5.2 Kinetic and biochemical studies involving actin

To bridge gaps in structural information, biochemical assays that measure folding kinetics are useful to further understand the substrate folding mechanism of TRiC. Folding assays using yeast TRiC revealed that native actin folds very slowly *in vitro*[164]. Recent studies of β-actin-TRiC interactions including actin folding kinetics, hydrogen/deuterium exchange-mass spectrometry (H/DX-MS), and cryo-EM investigated various events throughout cycle of actin folding mediated by TRiC ring closure and ATP hydrolysis[71,165]. Kinetics of actin folding involves two phases, one initial burst of folding on a short timescale, followed by a longer phase that allows unfolded actin to engage in the TRiC folding cycle multiple times[71,165]. In contrast, the half-life for binding of mLST8 (40 min) and Raptor (37 min) are significantly slower than those of actin (7 min) in the same study[165], suggesting that a slow folding mechanism is common for WD40 proteins due to their complex topologies compared to actin. Addition of the prefoldin co-chaperone accelerates this slow second phase and increases the amount of natively folded actin present *in vitro*[71], further establishing the role of prefoldin in TRiC-mediated protein folding.

In H/DX-MS experiments, the degree of deuteration correlates with structural flexibility and can be used to elucidate residues that do not have secondary structure, are not buried within the fold of a protein domain, and are not part of a protein-protein interaction[166,167]. Understanding of substrate regions containing secondary structure can in turn help infer which regions have achieved their native folds compared to those that are still disordered. This technique was employed to determine the folding state of the four subdomains of actin[162] (Figure 1.3A). In the H/DX-MS experiment using actin-TRiC-

bound sample, actin was mostly folded except for residues 8-31 in subdomain 1, which act to stabilise the nucleotide binding site. Additionally, loop regions of actin were more protected from deuterium and are thus likely to bind CCT subunits (peptides 2–8, 54–66, 105–117, 141–150, 198–205, and 228–239)[162]. These regions provide possible anchor points to ensure the regions that are important for TRiC-mediated folding are correctly positioned within the folding chamber. Repeating this experiment with ADP-beryllium fluoride (BeF_x) and ADP-AlF_x revealed sequential deuteration of actin in which the peptide 17-35 in subdomain 1 was the last to fold[162]. This region forms the ATP binding cleft and is likely stabilised by ATP upon release into the cytosol.

1.5.3 Interaction of substrates inside the folding chamber revealed through structural studies

Early biochemical studies have revealed that two molecules of tubulin or actin can be bound by TRiC at once[168], however these data were not structurally validated. Endogenous purification and crystallisation of bovine TRiC in the open state led to a 5.5 Å resolution structure showing the approximate location of tubulin within the folding chamber of TRiC[136]. Tubulin was shown to localise near the intermediate and apical domains of one cavity of TRiC and appeared to contact the helical protrusion in the apical domain and β-hairpin stem loop in the equatorial domain. However, in this model only 200-300 residues of tubulin (out of ~450) could be accounted for in the substrate density, and building a model was unfeasible due to limited resolution. This study was performed prior to the discovery of the correct TRiC subunit orientation, however a subsequent re-analysis of the tubulin-bound crystal structure with the determined

orientation revealed the placement of tubulin near the negatively charged subunits CCT3-6-8[132].

Low resolution cryo-EM experiments on actin–chaperonin complexes led to GroEL-actin and TRiC-actin reconstructions at 10.3 Å and 9.1 Å resolutions, respectively, that showed actin localised to the apical domains in the open state (Figure 1.12A)[162]. The substrate density in the GroEL-actin map only corresponded to 28% of the total volume occupied by native actin. This suggests that actin is disordered and unable to achieve its native fold through the GroEL chaperonin, providing further evidence that actin and tubulin co-evolved with TRiC. Conversely, in the map with TRiC-actin in the open state, the substrate density corresponded to 76% of the maximum volume. Additionally, the substrate density in the TRiC-actin map was expanded compared to native actin implying a large degree of the actin molecule was unfolded upon engaging TRiC. Taken together with the H/DX-MS data, these suggest that actin is largely unfolded when it engages with TRiC in the open state, and subsequently achieves a near-native fold when within the TRiC folding cavity in the closed state[162]. The cryo-EM data also showed that actin binds subunits CCT4, CCT2, CCT5, CCT7, and CCT8 in the open state, corresponding mostly to the positively charged hemisphere of the TRiC chamber. H/DX-MS data showed that the CCT2 and CCT5 apical domains disengage actin at the RLS loops in the ADP-BeF_x state suggesting that ATP binding induces substrate release from these subunits, altering the conformational dynamics of bound actin. However, TRiC-actin structures are still hindered by low resolution reconstructions and better structural data is needed to confirm these observations.

Further to actin and tubulin structures, a range of other substrate bound TRiC complexes have recently been investigated that shed additional light on substrate localisation within the TRiC complex. One such example is that of G proteins, which require TRiC and PhLP1 to form a stable ternary complex[74,89]. A low resolution cryo-EM reconstruction of TRiC-Gβ showed substrate density that correctly fit the near-native size and shape of the WD40 Gβ β-propeller domain (low resolution limited an atomic model), and XL-MS studies determined the N-terminal α-helix of Gβ form contacts with CCT3, demonstrating that this substrate localises to the CCT6 hemisphere of TRiC. A separate low resolution TRiC-Gβ-PhLP1 reconstruction showed Gβ within the folding chamber with an extra region of density corresponding to PhLP1 positioned above the Gβ. This extra density was located closer to the tips of the apical domains (density map unpublished), suggesting that PhLP1 plays a role in interacting with Gβ after it has achieved its native fold and stabilises the Gγ-binding domain until the Gβγ dimer can form[169]. This is consistent with another low resolution cryo-EM structure of PhLP that localised to the top of the apical domains of TRiC in the open state, likely inhibiting substrate folding either by limiting access to the folding cavity from the outside or by trapping substrates within the complex[73].

Consistent with the actin study above, low resolution cryo-EM studies of the transcription factor AML1-ETO (Section 1.2.3) in complex with human TRiC in the open state placed the location of the β-sheet rich DNA binding domain of AML (1-175) at the apical domains of CCT6 and CCT8 (Figure 1.12B)[170], also on the CCT6 hemisphere.

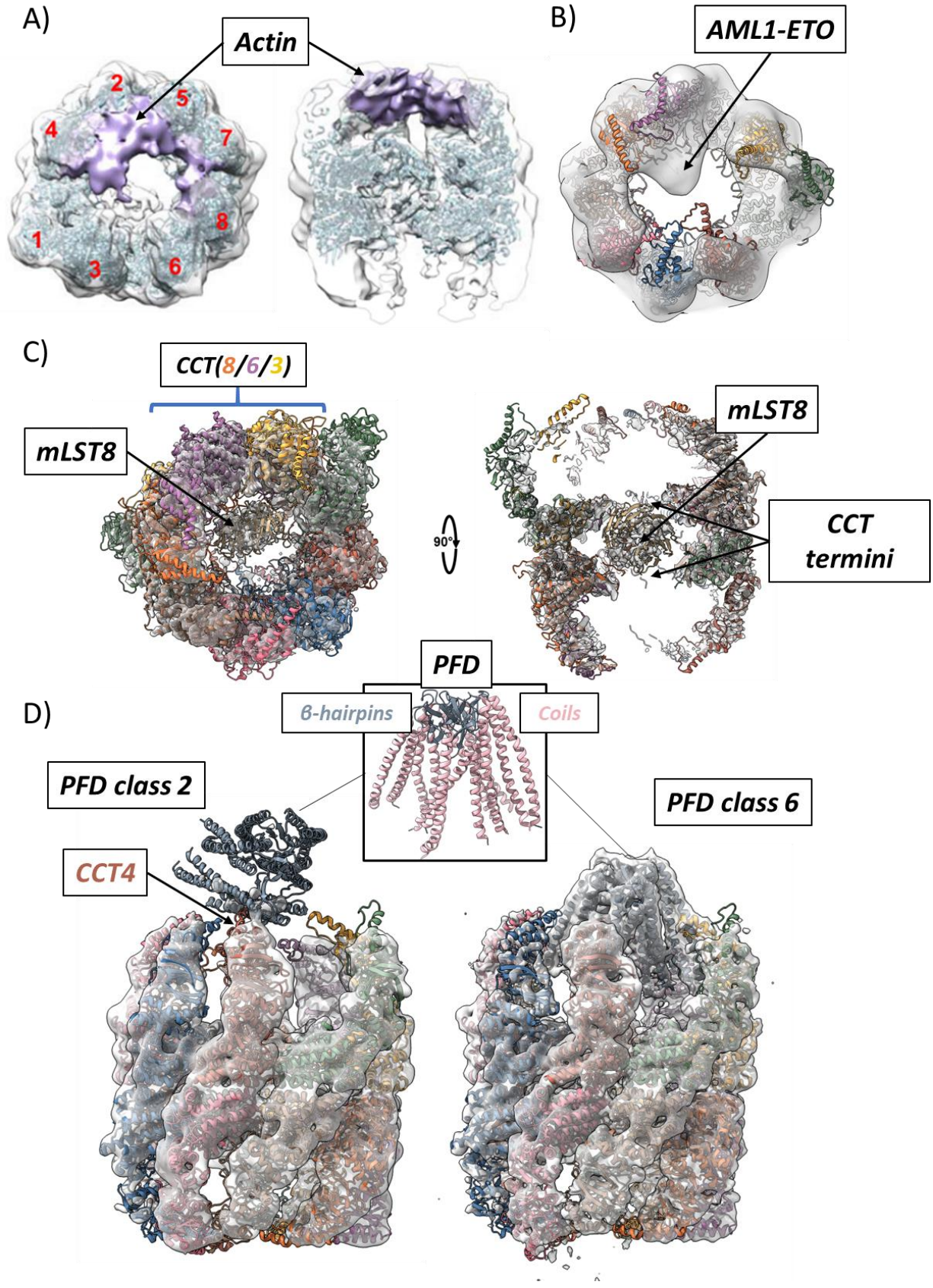

A)
Actin
2
5
4
7
8
1
3
6
B)
AML1-ETO
C)
CCT(8/6/3)
mLST8
mLST8
90°
CCT
termini
PFD
D)
β-hairpins
Coils
PFD class 2
PFD class 6
CCT4

Figure 1.12: Recent cryo-EM analysis reveals a pattern of substrate binding within human TRiC. A) Top-down (left) and side slice (right) views of crystal structure of apo-TRiC (PDB: 4B2T)[133] fit into 16.7Å TRiC-actin map (EMD-0018)[165] with actin density in purple. Figure borrowed from Balchin D., *et al* 2019[165] with permissions from Elsevier. B) 26Å density map (EMD-6227)[170] of TRiC containing AML1-ETO subunit density near CCT6 and CCT8 apical domains (TRiC PDB: 5GW5). C) 4.0Å density map (EMD-4489)[171] of mLST8 (beige) bound within the septum formed by TRiC rings in top-down view (left) and side slice (right). D) Left: Side view of 8.7Å density map (EMD-0493, class 2)[71] and model (PDB: 6NRB)[71] of prefoldin bound to CCT4 subunit. Right: Side view of 7.8Å density map (EMD-0490, class 6)[71] and model (PDB: 6NR8)[71] of prefoldin fully engaged with CCT apical domains.

The binding of the mTOR pathway component mLST8 to TRiC was validated by overexpression in human cells followed by co-immunoprecipitation with TRiC. A structure of mLST8 bound in the open conformation of TRiC was solved at 4.0 Å resolution using cryo-EM and showed for the first time the localisation of substrate density in the septum. The density for mLST8 showed engagement with the equatorial domains of the CCT subunits instead of the apical domains as shown in previous examples (Figure 1.12C). The density also localised to subunits CCT3/6/8 and formed interactions with the termini of CCT5, CCT7, CCT8, CCT6, and CCT3 in one of the rings, and CCT5′, CCT7′, CCT8′, CCT6′, and to a lesser extent CCT1′ in the other ring. As the density was not high enough to build a model of mLST8 bound to TRiC, complementary XL-MS experiments revealed the orientation of mLST8 and confirmed it is in a near-native state upon interaction with TRiC[171]. These examples show a trend in subunit localisation to the CCT6 hemisphere and shed light on the initial substrate engagement with TRiC during the folding process, however more structural understanding of the intermediate and final steps of substrate folding is necessary.

1.5.4 Structures of prefoldin co-chaperone bound to TRiC

The prefoldin co-chaperone (Section 1.2.2) was originally shown to bind actin and tubulin. Similar to TRiC, PFD has been more recently shown to interact with an array

proteins including those involved in cancer, neurodegeneration, and other cellular pathways[65–70], and deletion of PFD causes widespread protein aggregation *in vivo*[71]. The crystal structure of archaeal prefoldin from *Methanobacterium thermoautotrophicum* revealed a jellyfish-like structure containing two α and four β subunits, known in human as PFD1-6, that contain two or one central β-hairpin, respectively (Figure 1.12D: inset). The β-sheets play a role in complex assembly and the α-helical coiled coil regions extend down like tentacles from the β-sheets and create a hydrophobic pocket for substrates to bind[172]. The structural basis of prefoldin engagement with TRiC was detailed through a mixture of cryo-EM structures solved with PFD bound to human TRiC[71]. One cryo-EM experiment in which PFD was incubated with TRiC revealed a class of TRiC without PFD bound and five classes of TRiC with PFD bound at different points within the PFD-TRiC engagement mechanism[71]. In one class (Figure 1.12D: class 2) PFD latches onto TRiC through interactions between CCT4 and PFD subunits 4 and 6 through electrostatic properties. Three intermediate classes then showed PFD shifts to position itself closer to the TRiC apical domains with CCT3 engaged with PFD subunit 1. The final class (Figure 1.12D, class 6) showed PFD fully engaged with the apical domains of TRiC, poised to deliver substrate to the folding cavity[71].

1.6 Project aims and thesis scope

1.6.1 Unanswered questions from structure and biophysical studies to date

Structural and biophysical studies of TRiC have enhanced the understanding of how nascent substrates achieve their native folds. With the evidence gathered more recently, these investigations reveal a clearer mechanism for the initial stages of substrate folding, whereby nascent polypeptides are synthesized and delivered to the

apical domains of the CCT subunits by the co-chaperone prefoldin (Figure 1.13: steps 1 and 2). This interaction is also mediated by phosducin-like proteins that restrict the access of substrates between the folding cavity and the cellular environment, however details surrounding this mechanism remain unclear (Figure 1.13: step 2a). Upon both release of prefoldin and ATP binding to TRiC, the subunits then undergo a slight structural rearrangement to position substrates either near the apical domains or in the septum formed by the equatorial domains (Figure 1.13: step 3), however more understanding of this intermediary stage is needed. Upon ATP hydrolysis the TRiC rings close, trapping substrates in a compressed chamber that is shielded from the cellular environment to promote folding (Figure 1.13: step 4). The details of this step are also unclear as only a glimpse from H/DX-MS has provided details into how substrate domains fold, and the degree to which secondary structures are present at different stages of ATP hydrolysis. Finally, release of ADP and subsequent opening of the ring releases substrates to the cytosol that have achieved their native fold (Figure 1.13: step 5). TRiC-mediated substrate recognition has been well characterised through XL-MS and low resolution cryo-EM structures, however more detailed views into substrate interactions during intermediary folding states are required to understand the exact mechanism of substrate folding.

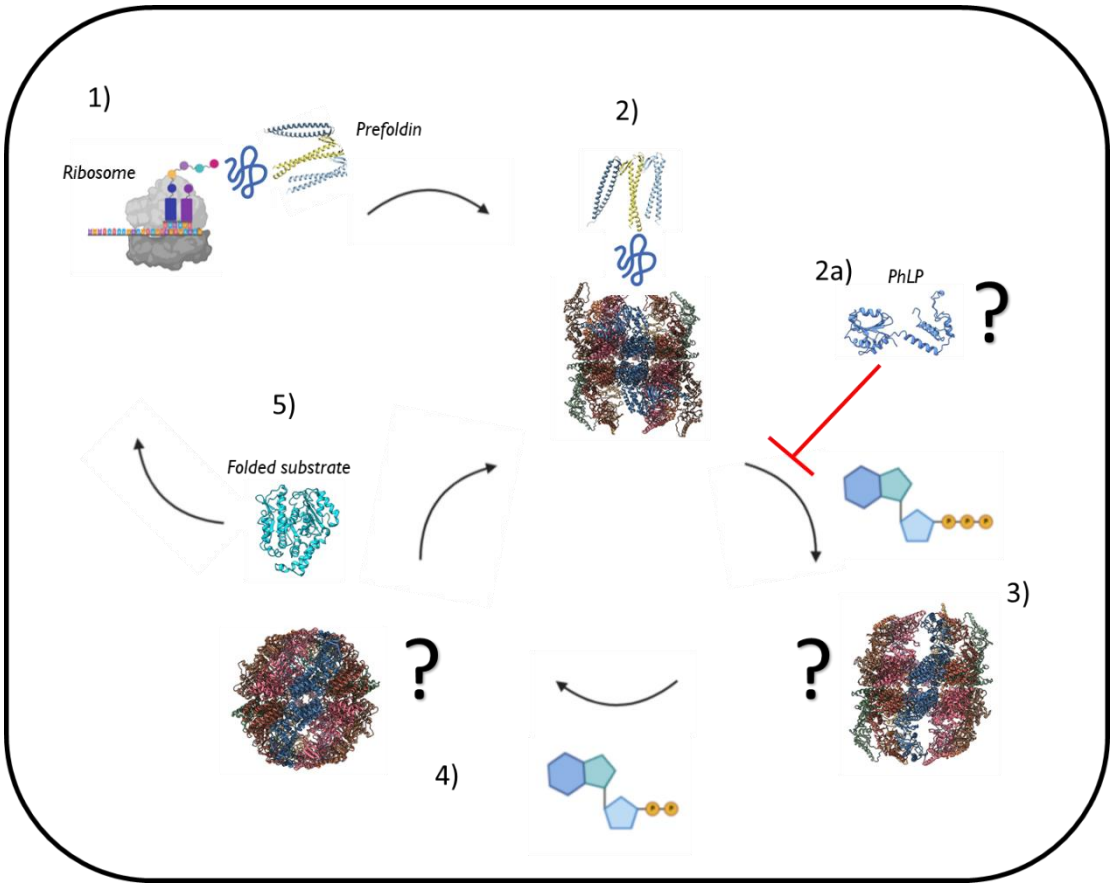


Figure 1.13: Updated mechanism of TRiC-mediated substrate folding. 1) Nascent polypeptides are synthesized by the ribosome and 2) delivered to TRiC in the open state by prefoldin, where 2a) PhLP co-chaperones act as modulators of TRiC-mediated substrate folding. 3) ATP then binds and causes a conformational re-arrangement to a partially open state. 4) ATP hydrolyses and one or both TRiC rings close, causing substrate folding. 5) ADP and substrate are released and TRiC re-opens to begin the cycle anew.

1.6.2 Thesis objectives

This thesis describes investigations into the structure and function of group II chaperonins through the course of a four-year DPhil project. It aims to fill in gaps in understanding of the group II chaperonin by further characterising individual CCT subunits as well as the native TRiC complex. The role of individual subunits has been studied for some including CCT4 and CCT5, but not all subunits, and investigations into whether each subunit can be individually isolated and characterised are reported. This

thesis also explores methods to isolate native human TRiC for structural and biochemical analysis. These analyses aim to better understand how TRiC mediates substrate folding in an endogenous setting, to fill in gaps in mechanistic understanding using a biologically relevant system.

1.6.3 Thesis chapter layout

Chapter 2 describes the methods used to undertake these studies including purification of individual subunits and native TRiC, biochemical investigations using ATP activity, stability, and substrate folding assays, and structural methods to observe chaperonin complexes including SAXS and cryo-EM.

Chapter 3 details the expression, purification, and subsequent biochemical and biophysical characteristics of recombinantly expressed homo-oligomeric subunits. Here it is reported that all subunits can be individually isolated and that some additional subunits form homo-oligomeric complexes, and that CCT4 and CCT5 retain the highest level of biological function.

Chapter 4 highlights strategies for isolating recombinant/endogenous large multi-protein complexes. Both genome editing and nanobody approaches successfully isolated native TRiC that could subsequently be used for structural and biochemical characterisation.

Chapter 5 describes the structure and function of the natively isolated human chaperonin CCT/TRiC. Several cryo-EM reconstructions showed endogenous human TRiC in both open and closed conformations, including several substrate-bound classes in each. Resolution was also sufficient to identify intermediary steps in tubulin and actin

binding while bound within TRiC in the closed state. The co-chaperone PhLP2A was also present in the actin bound reconstruction and exhibited details of its role in mediating substrate folding. Chapter 6 summarises the work completed in this thesis and outlines future perspectives.

2. **Methods**

2.1 Recombinant protein production of CCT4 and CCT5 from *E. coli*

DNA fragments encoding full-length human CCT4 and CCT5 were subcloned into the pNIC28-Bsa4 vector incorporating an N-terminal TEV-cleavable His6-tag. Rosetta *E. coli* cells were transfected with DNA plasmid, and a 50ml starter culture was incubated overnight in the presence of kanamycin and chloramphenicol. Cells were transferred to 6L Terrific Broth with kanamycin and incubated at 37°C until reaching OD600 2.0, at which point the cells were induced with 0.1 M IPTG and incubated overnight at 18°C. Cells were harvested and stored at -80°C. Cells were thawed and resuspended 1:1 in binding buffer (BB, 500 mM NaCl, 5% glycerol, 50 mM 4-(2-Hydroxyethyl)-1-piperazine ethanesulfonic acid (HEPES, pH 7.5), 0.5 mM TCEP, 20 mM Imidazole) with the addition of ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor. The mixture was sonicated for 15 minutes and lysate centrifuged using a JA25.5 rotor for 45 minutes at 35,000 x g at 4°C to remove insoluble debris. The soluble mixture was purified using affinity column (Talon Ni Resin; GE Healthcare), eluting with elution buffer (BB with 250 mM Imidazole) in a stepwise fashion, followed by size-exclusion chromatography (Superose 6; GE Healthcare). The His6 tag was optionally removed by addition of His-tagged Tobacco Etch Virus (TEV) protease and cleavage was confirmed by intact mass spectrometry after incubation overnight at 4°C. TEV was removed from purified protein by reverse affinity column (Talon Ni Resin; GE Healthcare) and protein concentration was determined using Bradford assay (Bio-Rad) and flash frozen at -80C. Frozen protein was thawed and buffer exchanged into low-salt

buffer (150 mM NaCl, 10 mM HEPES pH 7.5, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP)) and incubated with nucleotide for 1hr at 30°C prior to grid preparation.

2.2 Recombinant protein production of CCT1-CCT8 in *Sf9* insect cells

Constructs were cloned as described in section 4.3.1 and transformed into DH10Bac cells before purification of bacmid. *Spodoptera frugiperda* (Sf9) insect cells were transfected with bacmid DNA that initiates expression of viral proteins that produce the baculovirus for infection and protein expression. Each baculovirus was harvested from the medium and used to infect Sf9 cells where the virus allows expression of the protein of interest first at small scale, and then at a larger scale by expanding the supply of virus for subsequent infections. Sf9 cells were transfected and grown for 74 hours at 27°C. All steps were performed by staff in the lab of Dr. Nicola Burgess-Brown in the Structural Genomics Consortium.

Cells were harvested and pellets were resuspended 1:5 in gel filtration buffer (GF, 50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 0.5 mM TCEP) and sonicated for 15 minutes total with a cycle of 5 seconds on and 10 seconds off. Lysed cells were centrifuged in a JA-25.5 rotor for 60 minutes at 50,000 x g. Clarified lysate was filtered using an 0.45 µm syringe filter and applied on a HisPur™ Cobalt Resin (Thermo Scientific™) affinity column. Columns were washed with 100 column volumes (CVs) of GF buffer, followed by 10 CVs of GF buffer with 5- and 10-mM imidazole. Proteins were eluted in GF buffer with 150 mM imidazole. Elution fractions containing target protein were pooled, concentrated, and applied on a Superose 6 size exclusion chromatography column. Fractions containing target protein were pooled, concentrated, and frozen at -80°C.

2.3 Endogenous protein production

2.3.1 Endogenous TRiC purification from wild-type human HEK293 cells

Cultured wild-type human embryonic kidney 293 (HEK293) cells were harvested by centrifugation. Cell pellets were resuspended in PBS and centrifuged at 3000 rpm for 10 minutes. Supernatant was decanted and the cell pellet was resuspended in 5x pellet volume with PBS and centrifuged again. The cell pellet was resuspended in 5x pellet volume with hypotonic buffer A (10 mM Tris pH 7.9 (at 4°C), 1.5 mM $MgCl_2$, 10 mM KCl, and 0.5 mM dithiothreitol (DTT)) and centrifuged at 3000 rpm for 5 minutes. Supernatant was decanted and cell pellet was resuspended in 3x pellet volume hypotonic buffer A and let sit at 4°C for 10 minutes to begin gentle lysis of cells without perturbing the nuclear envelope. Cells were dounced 35 times using a pestle B dounce homogeniser. Lysates were then centrifuged for 15 minutes at 2500 rpm. The top layer containing the cytosolic fraction was collected and separated from cell debris and nuclear layer and 1 mM ATP was added to the cytosolic fraction. Ammonium sulfate was added to the cytosolic fraction in two steps, the first at 25% in which the supernatant was collected, and the second at 55% in which the supernatant was discarded. Pellet was dissolved in MQ-A buffer (20 mM HEPES/KOH pH 7.4, 50 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, 1 mM DTT, 0.1 mM protease inhibitor, 0.1 mM EDTA, 1 mM ATP) and dialysed overnight in 4 L MQ-A at 4°C. The dialysed sample was clarified by centrifugation at 15,000 x g to remove aggregates and was then passed over a HiLoad 26/10 Q sepharose column (GE Healthcare). TRiC was eluted by 40 % MQ-B (MQ-A with 1 M NaCl). The fractions containing TRiC were pooled, diluted in half by MQ-A, and applied to a Heparin HiTrap HP column (GE Healthcare) and was eluted

using a gradient of 20 to 65 % MQ-B (MQ-A with 1M NaCl). The fractions containing TRiC were pooled, concentrated, and run on a Superose 6 10/300 GL size exclusion column (GE Healthcare). Elution fractions containing TRiC were pooled, concentrated, and stored at -80°C.

2.3.2 CRISPR/Cas9-mediated genome editing to introduce affinity tag into CCT5 genetic locus

To insert a purification tag to the endogenous TRiC/CCT, the following protocol was completed by Dr. Dale Tranter at the University of Helsinki in the lab of Dr. Ville Paavilainen. HEK293T cells were transfected with Cas9 and gRNA targeting the gene encoding for the CCT5 subunit (Addgene eSpCas9(1.1) vector) along with a double stranded DNA fragment for a donor template containing a 3xFLAG/SpyTag[173] to be inserted at the C-terminus of CCT5. CCT5 was targeted for tagging due to previous success in expression and purification of recombinant CCT5 that is His-tagged at the C-terminus[39]. Insertion of a tag to the endogenous TRiC/CCT was achieved by a marker-free co-selection strategy[174]. HEK293T cells were transfected with plasmid encoding eSpCas9(1.1) and gRNAs targeting CCT5 and ATP1A1 (modified from Addgene #86613) along with linear ssDNAs to act as HDR templates by which to insert 3xFLAG/Spytag at the C-terminus of CCT5, as well as the Q118R and N129D ouabain resistance conferring point mutations to ATP1A1. Following ouabain selection, the polyclonal cell pool was assessed for FLAG-tagged CCT5 via immunoblotting. Single cells from confirmed pools were seeded into 96-well plates by FACS and resultant monoclonal lines were again screened *via* in-cell Western assay. Three positive monoclonal lines were identified, and each was further verified by PCR, sequencing,

and Western Blots. Primers and guides used to crease CCT5 C-terminally 3xFLAG/SpyTagged line include HDR template primers HAL F: GCGCTAGGATCCCAAAAAGATATGGTGGCTGTAGCTT, HAL R adaptor: TAGCTCCCTGAAAATAGAGGTTCTCTTCTTCAGATTCTCCAGGCTTACGA, HAR F adaptor: TGCGTATAAGCGGTACAAGTAGTAATGAAGACATTGAGAAAACTATGTAG, HAR R: CGCGTAGCGGCCGCAGGAAATCTGACTAGTTAAGGTGGC, and gRNA target sequence: CTTCATTCTTCAGATTCTCC.

2.3.3 Purification of native TRiC

Monoclonal HEK293T cells confirmed to contain FLAG-tagged CCT5 were grown in adherent culture plates and subsequently adapted to suspension. Cell pellets were harvested by freeze-thaw of cell pellet and resuspended in lysis buffer (100 mM HEPES pH 7.5, 100 mM NaCl, protease inhibitors, 1 mM DTT, benzonase, and 1% Triton X-100) and incubated at 4°C for one hour. Lysate was centrifuged at 35,000 x g and supernatant was passed through a 0.4 µm filter and was incubated at 4°C for one hour with 5 mL of anti-FLAG resin (Anti-DYKDDDK Tag (L5) Affinity Gel Protocol BioLegend cat no. 651503). Resin/lysate mixture was passed through a gravity column, washed with 100 CVs of wash buffer (20 mM HEPES pH 7.5, 100 mM NaCl) and eluted with wash buffer containing 0.15 mg/mL 3x FLAG peptide (Sigma cat no. F4799). The elution fractions were concentrated and run through a Superose 6 10/300 size exclusion chromatography column. Fractions containing purified TRiC were concentrated, flash frozen and stored at -80°C.

2.3.4 Nanobody mediated pull-down

CCT5-specific nanobodies were raised by immunisation of a llama[175], using as antigen the recombinant E. coli-expressed human CCT5, which was previously shown to exist as a TRiC-like homo-hexadecamer ($CCT5_{16}$)[39]. The llama received six injections of the antigen over a span of six weeks, after which point nanobody binders were isolated by extracting blood, separating out peripheral blood lymphocytes, and isolating RNA for use in creating cDNA libraries. *In vitro* selection of conformationally selective nanobodies was performed using phage display and binders were evaluated by enzyme-linked immunosorbent assay (ELISA) detecting presence of nanobody in the sample. Nanobodies were constructed in a pMESy4 vector incorporating C-terminal His6 and CaptureSelect C-tag (or EPEA-tag), which allows recombinant expression in E. coli, and purification by nickel-nitrilotriacetic acid (Ni-NTA) affinity and size-exclusion chromatography. 28 nanobodies from 18 families were identified along with the work stated above by Dr. Els Pardon in the lab of Dr. Jan Steyaert at the VIB-VUB University of Brussels. In this thesis 17 nanobodies from 15 families were determined to interact with CCT5 either by pulling down endogenous TRiC confirmed by gel excised liquid chromatography tandem mass spectrometry (LC-MS/MS) or by bio-layer interferometry (BLI), or both.

2.4 Mass spectrometry

2.4.1 Intact mass spectrometry

Denaturing intact mass spectrometry (MS) was used to check for expected masses of purified samples and to identify the presence of posttranslational modifications, mutations etc. Intact protein mass determination was run using 1 µl of protein sample resuspended in 59 µl of 0.1% formic acid, which was then analysed

using LC-QTOF instruments (Agilent). Protein mass was measured to within +/- 0.5 Da and compared to the expected mass derived from the construct sequence of the target protein.

2.4.2 Peptide identification by mass spectrometry

Quadruplicate samples of WT HEK293T and CCT5-FLAG tagged cells were lysed and purified using anti-FLAG affinity matrix (Biolegend cat no. 651501) using 20 mM HEPES pH 7.5 and 100 mM NaCl wash buffer and 200 mM glycine/150 mM NaCl, pH 2.2 elution buffer. Affinity purified mass spectrometry (AP-MS) elution fractions were reduced in volume in a speedvac (2.5 hours, 56°C) and re-suspended in 50 µL 7.2 M urea, 100 mM NH_4HCO_3 and incubated for 15 minutes at 25°C. Cysteines were reduced with 10 mM DTT (0.5 µL of 1M stock) for one hour at 51°C and then protected with 30 mM iodoacetamide (Sigma, I1149-25G; 4.2 µL of 0.36 M stock) for 45 minutes in the dark at 25°C. The reaction was quenched with 25 mM DTT and urea was diluted to < 1 M with 50 mM NH_4HCO_3. MS-grade trypsin (Thermo, MS grade) was added in a ratio of 1:20 (trypsin:protein, w:w) to solution and incubated for 16 hours overnight at 37°C. Trypsinized peptides were desalted using C18 desalting pipette tips (ThermoFisher, 87782).

Peptides were injected (2.0 µL) and separated using an Ultimate 3000 RSLC nano liquid chromatography system (Thermo Scientific) coupled to a LTQ Orbitrap Velos mass spectrometer (Thermo Scientific) via an EASY-Spray source. Sample volumes were loaded onto a trap column (Thermo Scientific, cat no. 164564) at 8 µl/min in 2% acetonitrile, 0.1% trifluoroacetic acid). Peptides were eluted on-line to a 50 cm analytical column (Thermo Scientific, cat no. ES803). Separations were carried out using a

ramped 120 min gradient from 1-90% buffer B (buffer A: 5% dimethyl sulfoxide (DMSO), 0.1% formic acid; buffer B: 75% acetonitrile, 5% DMSO, 0.1% formic acid). The mass spectrometer was operated in positive polarity using a data-dependent acquisition mode. Ions for fragmentation were determined from an initial MS1 survey scan at 30,000 resolution (at m/z 200) followed by Collision-induced dissociation (CID) of the top 10 most abundant ions with a normalised collision energy of 35. A survey scan m/z range of 350 – 1500 m/z was used, and charge state exclusion enabled for unassigned, +1, +8 and >+8 ions. Lock mass correction was enabled using the following ions: 401.92272 and 445.12003. Data was processed using the MaxQuant[176,177] software platform (v1.6.2.3) with database searches carried out by the in-built Andromeda search engine against the Uniprot homo sapiens database (v20180104; 161,549 entries). A reverse decoy database was used at a 1% false discovery rate (FDR) for peptide spectrum matches and protein identification. Data was visualised and analysed further in Perseus (version 1.6.2.2)[178].

For chemical crosslinking coupled with mass spectrometry (XL-MS) the experiment was repeated as above after the following initial steps: A 5 mg : 5 mg aliquot of isotopically-coded bis(sulfosuccinimidyl)suberate (BS3) d0/d4crosslinker (Thermo) was reconstituted to 25 mM in water and immediately added to 25ug TRiC in the optimal ratio determined experimentally (between equimolar amounts to the number of moles of lysine residues to 10X the number of lysines). The crosslinking reaction was incubated for 30 min at 25°C with mild agitation and the reaction was quenched with 50 mM NH_4HCO_3 (1:20 dilution from 1 M stock, or 5 µL) for 20 min at 25°C.

2.5 Biophysical and biochemical assays

2.5.1 Bio-layer interferometry

BLI experiments were performed on a 16-channel ForteBio Octet RED384 instrument at 25°C, in buffer containing 20 mM HEPES pH 7.5, 150 mM NaCl, and 0.1% bovine serum albumin. 50 µL of 100 ng/µL of biotinylated CCT5 (expressed and purified in *E. coli* pNIC-Bio3 vector, as above) was loaded to the streptavidin coated sensors. The concentration for nanobodies used ranged from 39 nM to 40 µM. Measurements were performed using a 300 second association step followed by a 300 second dissociation step in a black 384-well plate with tilted bottom (ForteBio) using a serial dip method from low to high Nb concentration. The baseline was stabilised for 30 sec prior to association and signal from the reference sensors was subtracted. Equilibrium constants were determined using steady-state affinity kinetics based on signal level and concentration using Octet Data Analysis software (ForteBio). Equilibrium dissociation constants are calculated using the equation: $R_{eq} = \frac{CR_{max}}{K_d+C} + RI$, where R_{eq} is the response at equilibrium, C is analyte concentration, R_{max} is the maximum binding response, and RI is the refractive index contribution.

2.5.2 Thermal stability measurements by differential scanning fluorimetry

Protein samples were diluted to 0.5–2 mg/ml 10 µl volume in the SEC buffer used for a given sample. 8 µl of each was diluted in 64 µl sample buffer. 8 µl of 100x SyproOrange dye (ThermoFisher cat no. S6650) was added to each sample. 20 µl of each sample was then added to a white 96-well PCR plate in triplicate and differential scanning fluorimetry (DSF) was run using MxPro QPCR system (Agilent Technologies).

Data were analysed in Prism software (GraphPad) to determine thermal melting temperatures (T_ms) using nonlinear Boltzmann equation.

2.5.3 Malachite green ATPase activity assay

Protein samples were prepared in ATPase assay buffer (20 mM HEPES pH 7.5, 150 mM KCl, 5 mM $MgCl_2$). Chaperonin samples were diluted to 0.5 µM and incubated for 30 minutes in the presence of 250 µM ATP in a final volume of 50 µL. 100 µL of malachite green reagent was added to the sample wells and incubated for an additional 30 minutes. Absorbance at 620 nm was read and the amount of free phosphate released (ρmol) was quantified using a phosphate standard curve.

2.5.4 HγD aggregation suppression assay

A pQE.1 vector containing the sequence for human γD-crystallin (HγD) was obtained as a gift from the Jonathan King lab at Massachusetts Institute of Technology. 55 µM HγD was unfolded overnight at 37°C in unfolding buffer (5.5 M guanidine hydrochloride, 50 mM Tris-HCl, pH 7.5, and 5 mM DTT). Chaperonin samples were added in a 1:1 molar ratio in refolding buffer (50 mM Tris-HCl pH 7.5, 1 mM DTT, 50 mM KCl, 5 mM $MgCl_2$, and 1 mM ATP) that initiates HγD aggregation in the absence of chaperonin. Aggregation kinetics were measured at 350 nm on a SpectraMax plate reader (Molecular Devices). Half-lives were determined using exponential decay formula $N(t) = N_0(\frac{1}{2})^{(\frac{t}{t_{\frac{1}{2}}})}$, where *N(t)* is quantity of substrate remaining, N_0 is initial quantity of substrate, *t* is time elapsed, and $t_{1/2}$ is the half life.

2.6 Small angle X-ray scattering

CCT subunits were subjected to size exclusion chromatography coupled with small angle X-ray scattering (SEC-SAXS) on a KW403-4F column (Shodex) at 10 mg/mL in buffer consisting of 500 mM NaCl, 50 mM HEPES pH 7.5, 5% glycerol, 0.5 mM TCEP, by flowing sample through an in-vacuum quartz capillary of 1.6mm diameter at 0.16 mL/min by staff at the Diamond Light Source BL21 beamline. Scattering data were collected using a Pilatus2M detector (Dectris, CH) with a sample-detector distance of 3914 mm and a wavelength of 1 Å. The data covering the peak were radially averaged and the scattering of the buffer was subtracted before analysis using Scatter[179] and the ATSAS suite of programs[180] in this thesis.

2.7 Negative stain electron microscopy

Copper grids with Formvar carbon coating (300 mesh, Quantifoil Micro Tools GmbH) were glow discharged for 30 seconds. 4 µL of purified protein was placed on the grids for 30 seconds. Excess sample was blotted off using filter paper (Whatman), and the grids were washed once by placing grid in a bead of deionised water and the grids were blotted again. Grids were then applied to a drop of filtered 1.5 % uranyl acetate (Sigma-Aldrich) for 30 seconds. Grids were visualised under a T12 or F20 transmission electron microscope (TEM), and digital micrographs were acquired.

2.8 Cryo-EM

2.8.1 Cryo-EM of CCT5 homo-oligomer

A 3 µl aliquot of purified CCT5 at 1 mg/ml was applied on a glow-discharged, holey-carbon coated grid (Cflat CF-1.2/1.3-2C; Protochips). The grids were blotted at ~90% relative humidity at 4°C and plunge frozen to liquid ethane using a vitrification apparatus (Vitrobot; FEI, Thermo Fisher Scientific). Data were acquired using a 300 kV

transmission electron microscope (Titan Krios; FEI, Thermo Fisher Scientific) at the Electron Bio-Imaging Centre (eBIC, Diamond Light Source, UK) equipped with a direct electron detector (K2 Summit; Gatan). Movies of 25 frames each were collected at in electron counting mode at calibrated magnification of 47,170× resulting in a total dose of ~50 e^- per Å2 and pixel size of 1.06 Å. Data were processed in the Scipion software framework[181].

Movie frames were aligned to account for drift and electron beam induced damage in MotionCor2[182]. The resulting drift corrected, dose-weighted micrographs were used to pick particles in Xmipp in a semi-automated fashion[183] (de la Rosa-Trevín et al. 2013). Contrast transfer function (CTF) parameters were estimated using CTFFIND4[184] from the micrographs that had not been dose-weighted. Particles were extracted for reference-free 2D classification in Relion[185] and Xmipp[183] (de la Rosa-Trevín et al. 2013). Class averages from Xmipp were used in EMAN2[186] to create initial models with D8 symmetry. Particles from RELION 2D classification were selected for model-based 3D multi-reference refinement in cryoSPARC[187].

2.8.2 Cryo-EM of native TRiC

Purified TRiC (1.5 mg/ml) was premixed with Nb18 at a 1:1 molar ratio, with 1 mM ATP, 5 mM MgCl2 and AlFx (5 mM $Al(NO_3)_3$ and 30 mM NaF)[129]. A 3 µl aliquot was pipetted on a holey carbon cryo-EM grid (1.3/1.2 Cu 200 mesh, Quantifoil) and vitrified by plunge-freezing in liquid nitrogen-cooled liquid ethane (Vitrobot Mark IV, Thermo Fisher Scientific). The data were collected using a 300 kV transmission electron microscope (Titan Krios, TFS) on a direct electron detector (Gatan K3).

Cryo-EM data were processed in Scipion software framework. Movie frames were aligned, dose weighted and averaged using MotionCor2[182]. CTF parameters were estimated using CTFfind4[184]. Particles were picked using crYOLO and subjected to several rounds of classification in RELION[185,188]. The particles were classified to closed and open states first using 2D classification. The initial 3D model for each state was calculated *ab initio*. Particles were further 3D classified in four separate sets. The closed particles were further classified to different substrate-bound states using 3D classification and option –relax_sym C2 in RELION[189]. This revealed actin–co-chaperone class which was not captured using conventional 3D classification in RELION. Particles in the selected classes, including the consensus map combining all classes with C2 symmetry imposed, were refined using the gold-standard refinement protocol in RELION. The resolution of the final cryo-EM maps was estimated by Fourier shell correlation (FSC) combined with phase randomisation to account for the effects of masking.

2.8.3 Model building of TRiC and substrates

To model the atomic structure of TRiC in the closed state, a homology model was created for each chain in SWISS-MODEL[190] using the yeast structure (PDB: 6KS6) as a template[27]. A model for Nb18 was created in a similar manner (PDB: 6QX4 chain C)[191]. These models were fitted to the 2.5 Å resolution cryo-EM map of the closed state maximising real space cross-correlation in UCSF Chimera (Table 2.1). The tubulin[192] (PDB:6I2I chain B) and actin[62] (PDB:2Q1N) structures was were fitted in a similar way using as a backbone the 2.5 Å resolution closed model. Models were manually adjusted in COOT[193] and subjected to several rounds of real space refinement in PHENIX[194]. To

model the open state, the previously solved open state (PDB:6NRA) model was first fitted as a rigid body to the 3.5-Å resolution cryo-EM map of the open state and then manually built in COOT[71] (Table 2.1). Molecular graphics and analyses performed with UCSF ChimeraX[28,29], developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.

Table 2.1. Cryo-EM data collection, refinement, and validation statistics.

	Closed, consensus	Closed, tubulin-bound	Closed, Actin/PhLP2A-bound	Open
Data collection and processing				
Magnification	81,000	81,000	81,000	81,000
Voltage (kV)	300	300	300	300
Electron exposure (e–/Å^2)	62	62	62	62
Defocus range (μm)	0.75 to 2.5	0.75 to 2.5	0.75 to 2.5	0.75 to 2.5
Pixel size (Å)	1.06	1.06	1.06	1.06
Symmetry imposed	C2	C1	C1	C1
Initial particle images (no.)	3,856,544	3,856,544	3,856,544	3,856,544
Final particle images (no.)	316,195	93,758	63,082	144,903
Map resolution (Å)	2.5	3.0	3.1	3.5
FSC threshold	0.143	0.143	0.143	0.143
Map resolution range (Å)	2.5–4.4	2.8–4.8	2.9–6.6	3.1–9.9
Map sharpening *B* factor (Å^2)	-63.0	-65.3	-57.7	-70.2
Refinement				
Initial model used (PDB code)	6KS6	6KS6	6KS6	6NRA
Model-to-map resolution (Å)	2.64	3.05	3.27	3.51
FSC threshold	0.5	0.5	0.5	0.5
Model-to-map correlation (Phenix)	0.73	0.86	0.74	0.81
Model resolution range (Å)				
Model composition				
Non-hydrogen atoms	66,477	69,139	71.009	34,246
Protein residues	8626	8970	9219	4480
Ligands	48	48	48	44
B factors (Å^2)				
Protein	72.62	72.55	70.77	64.35
Ligand	171.45	171.45	158.28	136.15
R.m.s. deviations				
Bond lengths (Å)	0.007	0.004	0.008	0.005
Bond angles (°)	0.827	0.607	0.863	0.816
Validation				
MolProbity score	1.64	1.77	1.85	2.07
Clash score	4.63	6.80	7.32	6.81
Poor rotamers (%)	1.73	1.75	1.55	1.87
Ramachandran plot				
Favored (%)	96.57	96.69	95.6	92.01
Allowed (%)	3.33	3.22	4.29	7.81
Disallowed (%)	0.09	0.09	0.11	0.18

3. Individual TRiC subunits enable chaperonin characterisation

3.1 Background

Until recently much of what is known about the structure and function of group II chaperonins has relied on the study of simpler chaperonin systems from lower organisms. These contain one or two constituent subunits/proteins that are arranged in double stacked rings each containing 7-9 copies[11,17,20,21]. These model systems enable characterisation of chaperonins without the need for isolating all eight paralogous subunits of higher eukaryotic TRiC, a task that is difficult but has been made possible from bovine and yeast purification strategies[143,195]. The chaperonin from the archaeal species *Methanococcus maripaludis* (Mm-Cpn) is one such group II chaperonin that is composed of one subunit forming a 16-subunit homo-oligomeric complex similar in architecture to the eukaryotic TRiC[21]. Mm-Cpn has been studied in structural experiments that helped identify the general architecture of group II chaperonins, including the manner in which they stack into two rings each of equal subunit number and contain three distinct equatorial, intermediate, and apical domains[41,42]. Mm-Cpn biochemical studies also shed light on the group II chaperonin ATP hydrolysis and substrate folding activities[40], proving a basis of comparison to native TRiC.

These systems subsequently led to the discovery that several human CCT subunits exhibit similar self-assembly upon over-expression of individual subunits in *E.coli*, including CCT4 and CCT5[39]. The CCT4 and CCT5 homo-oligomeric complexes (sometimes written as CCTX_{16}) were shown to be structurally homologous to TRiC, with similar 16-mer arrangement into two stacked rings, by negative stain electron-microscopy[39]. They are both also active in ATP activity assays and can suppress

aggregation of the TRiC substrate human γ-D crystallin (HγD)[39]. The functional relevance of these individual subunits and synthetic homo-oligomeric complexes remains unknown. This chapter details efforts to further characterise the homo-oligomeric CCT complexes and explores a wider range of CCT subunits through over-expression in the *Spodoptera frugiperda* (Sf9) host system, a higher organism that is better able to produce soluble CCT subunits.

The aim of this chapter is to investigate individual TRiC subunits to establish techniques as proof-of-concept to study TRiC in its native form. Purifications of overexpressed CCT4 and CCT5 were repeated based on previous literature and used to recapitulate biochemical assays including ATPase activity, substrate refolding, and thermal stability. These assays agreed with published results that CCT4 and CCT5 can form stable homo-oligomeric complex that resemble the biological activities of TRiC, and structural investigation using cryo-EM showed that these complexes form 16-subunit assemblies similar to TRiC. Further biochemical and structural investigations were performed on other TRiC subunits, which are found to also form homo-oligomers. These include the biochemical assays described above, along with small-angle x-ray scattering (SAXS) to determine the oligomeric state of the purified individual CCTs.

3.2 Expression and purification of CCT4 and CCT5 in *E. coli*

The first efforts in this project involved reproducing the human CCT4 and CCT5 model homo-oligomeric complexes, following previous work[39], with the goal of enabling biochemical assays and structural methods to characterise chaperonin complexes. Full-length CCT4 (N-terminal His6 tag) and CCT5 (C-terminal His6-FLAG tag) were over-expressed in the *E. coli* BL21 expression system and purified by Ni-NTA affinity

chromatography followed by Superose 6 size exclusion chromatography and another reverse Ni-NTA step (Figure 3.1).

Both CCT4 and CCT5 monomers are ~60kDa in molecular weight (MW) and were found during recombinant expression to co-purify with the *E. coli* bifunctional polymyxin resistance protein ArnA (labelled in Figure 3.1), a contaminant of ~70 kDa that is difficult to separate in affinity and size exclusion chromatography. Interestingly, bands corresponding to CCT4 (Figure 3.1A) and CCT5 (Figure 3.1B) exist in two separate elution peaks in size exclusion, one at elution volume (V_e = ~70 mL) that corresponds to high MW around a 1MDa complex, and one (V_e = ~100 mL) that corresponds to a smaller MW closer in size to a CCT monomer. This observation is especially pronounced for the CCT5 full-length construct (Figure 3.1B) and could be explained by excess CCT subunits that were not incorporated into ring-like complexes or are present due to dissociation from the chaperonin complexes. The second peak often contains contaminating proteins given that the elution volume is around the same size as many monomeric proteins. In summary, pure homo-oligomeric CCT4 and CCT5 complexes were obtained in sufficient quantities, and separated from their monomeric or low MW species, for follow-up biochemical and structural studies (Figure 3.1, green boxes).

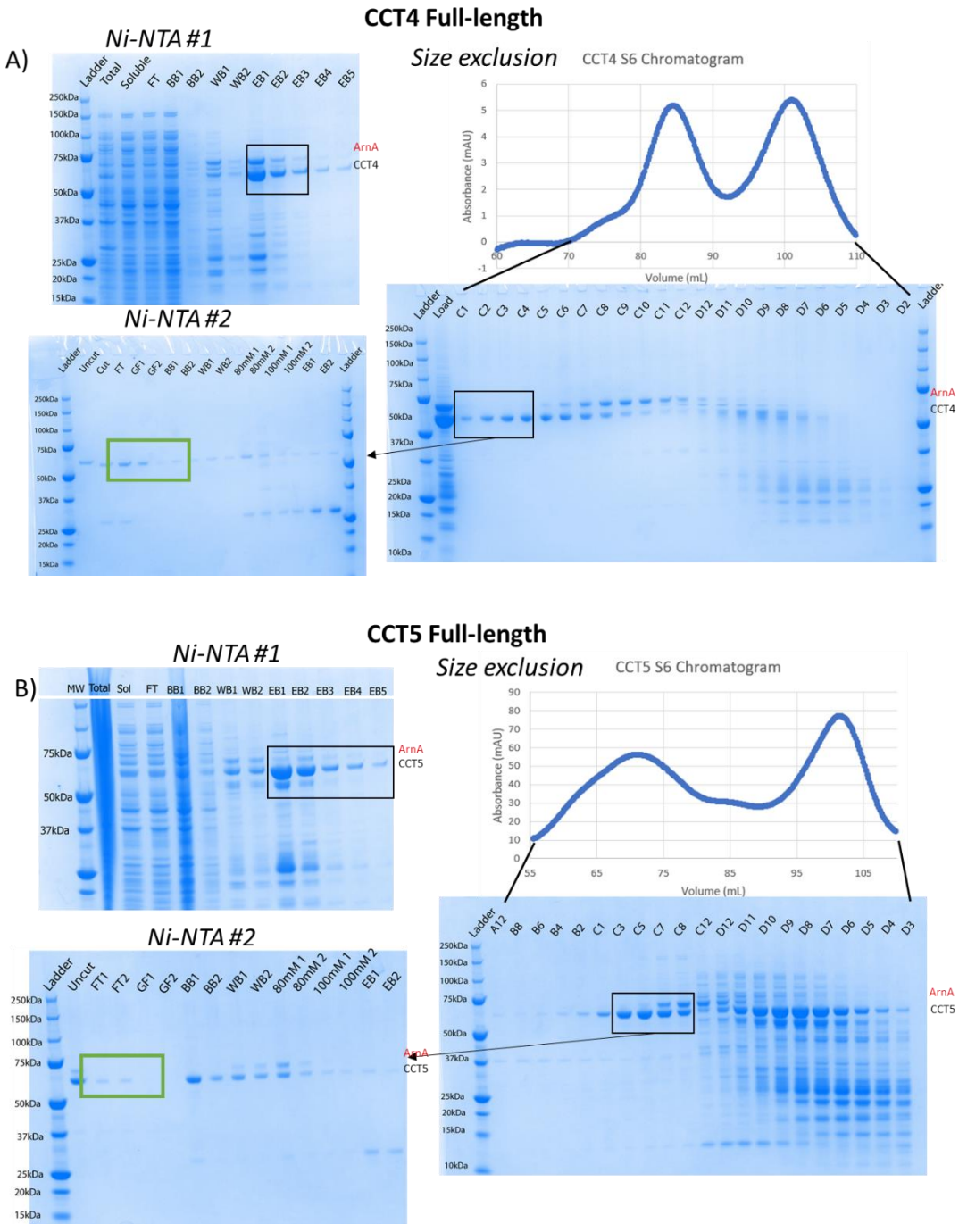


Figure 3.1: Purification of full-length CCT4 and CCT5. Purification of A) CCT4 and B) CCT5 full-length constructs by Ni-NTA affinity chromatography (Ni-NTA #1), Superose 6 size exclusion chromatography, and reverse Ni-NTA affinity chromatography (Ni-NTA #2). Black boxes on gels indicate pooled protein for next steps. Black lines on chromatograms indicate size exclusion fractions analysed on SDS-PAGE. ArnA contaminant (red) and CCT4/5 (black) are labelled next to gels that contain both bands. Ni-NTA gels are labelled with total, soluble, and flow through fractions, followed by binding (BB), wash (WB), and elution (EB) buffer fractions that contain increasing imidazole. Green boxes indicate fractions pooled for final purified protein.

3.3 Biophysical investigations of CCT4 and CCT5 chaperonins

3.3.1 CCT5 ATPase activity

Native TRiC relies on ATP hydrolysis to induce ring closure and promote substrates to achieve their native fold. The $CCT5_{16}$ complex was used as a model chaperonin in this chapter, to assess whether CCT5 could recapitulate the ATP hydrolysis activity of native TRiC *in vitro*, with the goal of optimising assay conditions prior to successful isolation of TRiC. CCT5 activity was first measured using a malachite green ATP hydrolysis assay that uses a dye that changes in absorbance upon exposure to free phosphate[196] (Figure 3.2A). Thus, the amount of ATP hydrolysed can be quantified using a standard curve based on the amount of free phosphate present in the sample after the hydrolysis experiment. These experiments were adapted from a different ATP hydrolysis method using radiolabelled [α-^{32}P] ATP, showing that CCT4 and CCT5 have similar hydrolysis activities to TRiC (rates not reported)[39].

Briefly, $CCT5_{16}$ was incubated with 250 µM ATP for 30 minutes and quenched by the addition of working reagent that contains malachite green. An absorbance reading was measured at 620 nm (Figure 3.2B), and the amount of free phosphate generated was correlated to a free phosphate standard curve (Figure 3.2C). Two concentrations of $CCT5_{16}$ complex, 20 and 100 nM, were tested and showed dose-dependent ATP hydrolysis as compared to the blank sample (Figure 3.2D). Thus, this ATPase assay was enabled for future study of group II chaperonins.

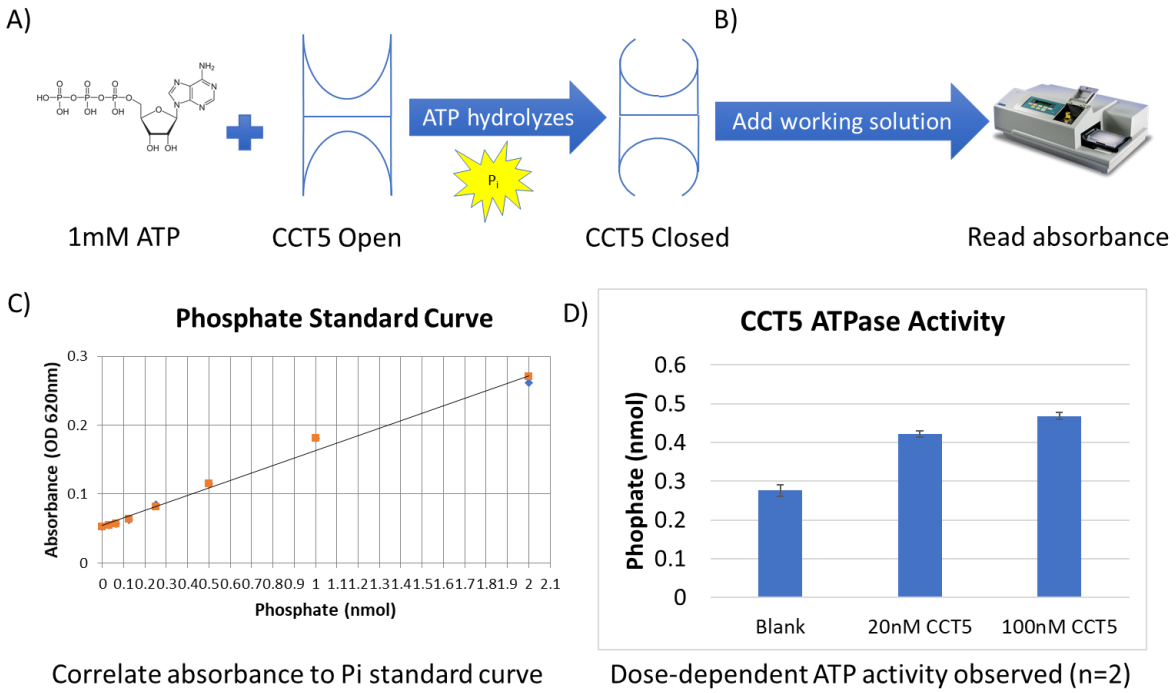


Figure 3.2: CCT5 ATPase activity assay. Workflow of malachite green ATPase assay: A) ATP is incubated with CCT5 chaperonin and undergoes hydrolysis to transition CCT5 from the open to closed state, B) after the desired incubation period, the working reagent is added to the well containing CCT5 and ATP and an absorbance reading is measured, C) free phosphate concentration is correlated to a free phosphate standard curve, and D) ATP activity is plotted as a function of free phosphate released by the reaction. Error bars indicate standard deviation of n=2 samples.

3.3.2 CCT5 aggregation suppression of human γD crystallin

Another chaperonin assay used to characterise CCT5 function involved aggregation suppression of the known TRiC substrate human γD crystallin (HγD). HγD is a structural protein in the lens region of the human eye composed of domains with the double Greek key fold, and is known to aggregate under ageing conditions such as oxidation, deamidation, or cleavage causing cataract formation[197]. Chaperonins including Mm-Cpn, $CCT4_{16}$, $CCT5_{16}$, and TRiC in their oligomeric complexes have previously been shown to suppress the aggregation of HγD as a measure of chaperonin activity[39,40,198].

In this assay (Figure 3.3: top) HγD is unfolded in guanidine hydrochloride (Gn-HCl) prior to the experiment. Upon rapid dilution to a Gn-HCl free buffer suitable for refolding, HγD undergoes aggregation that can be tracked by measuring turbidity absorbance (350 nm) over time. When incubated with chaperonin, the amount of HγD aggregation is reduced. This suppression of aggregation by the chaperonin can be tracked over time as compared to a sample with HγD only[199]. Additionally, this sample can be filtered to remove aggregates and run on a sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) gel or applied to a size exclusion column to quantify the amount of refolded HγD.

The HγD-only sample run in this experiment exhibited a steep curve in the plot of Abs_{350nm} over time, indicative of aggregation (Figure 3.3B: red line). Upon addition of recombinant $CCT5_{16}$, this aggregation profile was reduced in comparison to the HγD-only sample, as evident by a slower aggregation profile and longer half-life (Figure 3.3B: blue line). The reduction in aggregation was not as pronounced as expected however, likely because the synthetic CCT5 homo-oligomer is less active than native TRiC. The samples run in the experiment were then filtered and analysed on SDS-PAGE to check for refolded HγD (Figure 3.3:C), however none was detected in reaction lanes containing CCT5 and five-fold excess of CCT5. This could be due to the small amount of material used, as this assay was adapted from cuvette to a 96-well format. However, due to the reduction in overall aggregation it is assumed that refolding would have occurred in the experiment and that these parameters could be optimised during experiments with native TRiC.

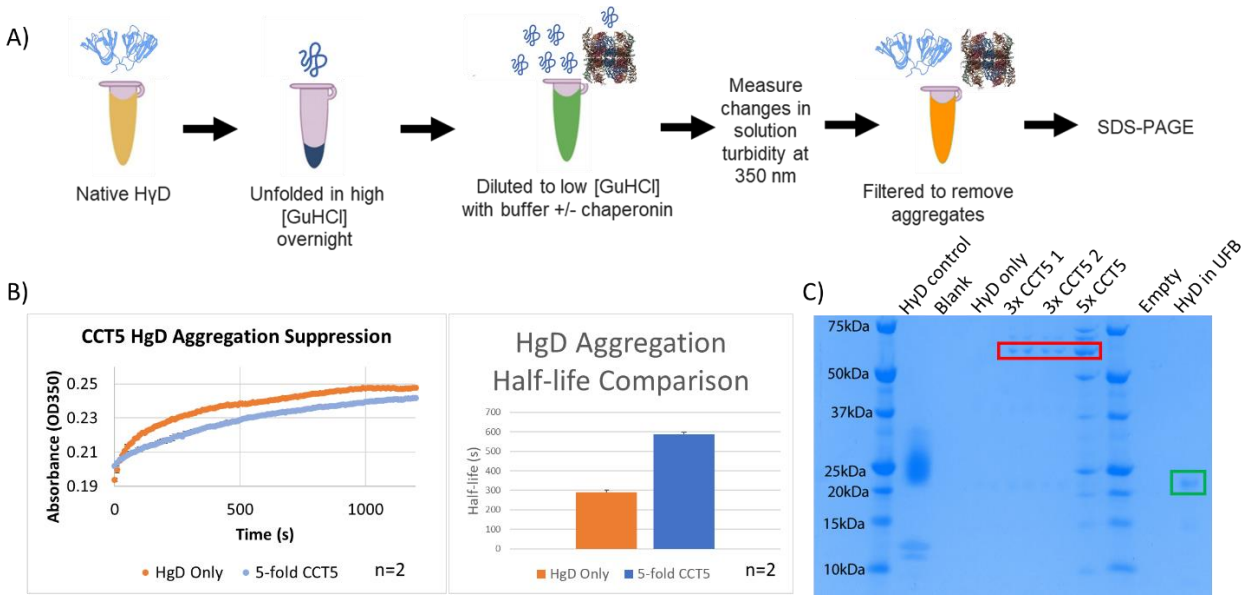


Figure 3.3: CCT5 aggregation suppression assay. A) Schematic of chaperonin-mediated suppression of HγD aggregation. B) Aggregation kinetics (left) and half-life analysis (right) of CCT5-mediated aggregation suppression. Half-life calculated using exponential decay formula. Error bars indicated standard deviation of n=2 replicates. C) SDS-PAGE analysis of experimental samples from the aggregation assay, checking for the presence of refolded HγD. Lanes include native HγD control, blank, aggregated HγD-only sample, samples containing CCT5, and HγD still in unfolding buffer (UFB). Gel bands were excised and run on LC-MS/MS: red box indicates CCT5 and green box indicates HγD.

3.3.3 CCT5 thermal stability

Thermal stability of a protein or complex is often a useful indicator of protein quality from recombinant preparation, and can be useful in determining optimal conditions for downstream structural applications including cryo-EM[200]. Differential scanning fluorimetry (DSF) is a common technique for measuring thermal stability, and involves a simple protocol in which proteins are incubated with a specialised fluorogenic dye that binds to hydrophobic surfaces upon heat denaturation, and the amount of dye detected increases with higher temperature as proteins are denatured[201]. Recombinant $CCT5_{16}$ as-purified has a relatively stable thermal melt (T_m) profile with an average T_m of 56°C +/- 0.7°C, slightly lower than the 60°C melt temperature observed previously[39].

Next, the T_m profile of $CCT5_{16}$ was determined in the presence of different ATP-like ligands, to assess whether its stability is different between ligand-induced open and closed conformations. These ligands include the hydrolysis product ADP, non-hydrolytic analogues AMP-PNP and ADP-AlF_x, the latter of which is created by incubating ATP with $Al(NO_3)_3$ and NaF. In solution, it is hypothesised that ATP is hydrolysed and AlF_3 occupies the position of the cleaved γ-phosphate, trapping ADP in an ATP-like conformation. As shown in Figure 3.4, the melt temperature of $CCT5_{16}$ is unaffected in the presence of ADP, AMP-PNP, and ADP-AlF_x. These ligands, aside from ADP-AlF_x, all promote chaperonins to be in their open states based on previous work. Upon addition of ATP and ATP-γ-S, an ATP mimic that 'traps' group II chaperonins in the open pre-hydrolysis state, $CCT5_{16}$ is stabilised by 3 and 5°C, respectively.

The stabilisation by ATP seen here with $CCT5_{16}$ could be explained in that ATP promotes the closed state of chaperonins and likely confers additional stability compared to the more flexible open conformation. However, this explanation does not hold up for the stabilisation by ATP-γ-S, as this analogue has been shown to promote the open state[146]. The lack of effect seen in ADP-AlF_x is also surprising, as this ATP analogue should trap $CCT5_{16}$ in the closed conformation and have a similar effect to ATP. As demonstrated later in this chapter, the ADP-AlF_x ligand did not seem to consistently induce the closed state of CCT5 during this study, possibly due to chemical instability of reagents, improper handling, or inherent differences between the CCT5 chaperonin and others.

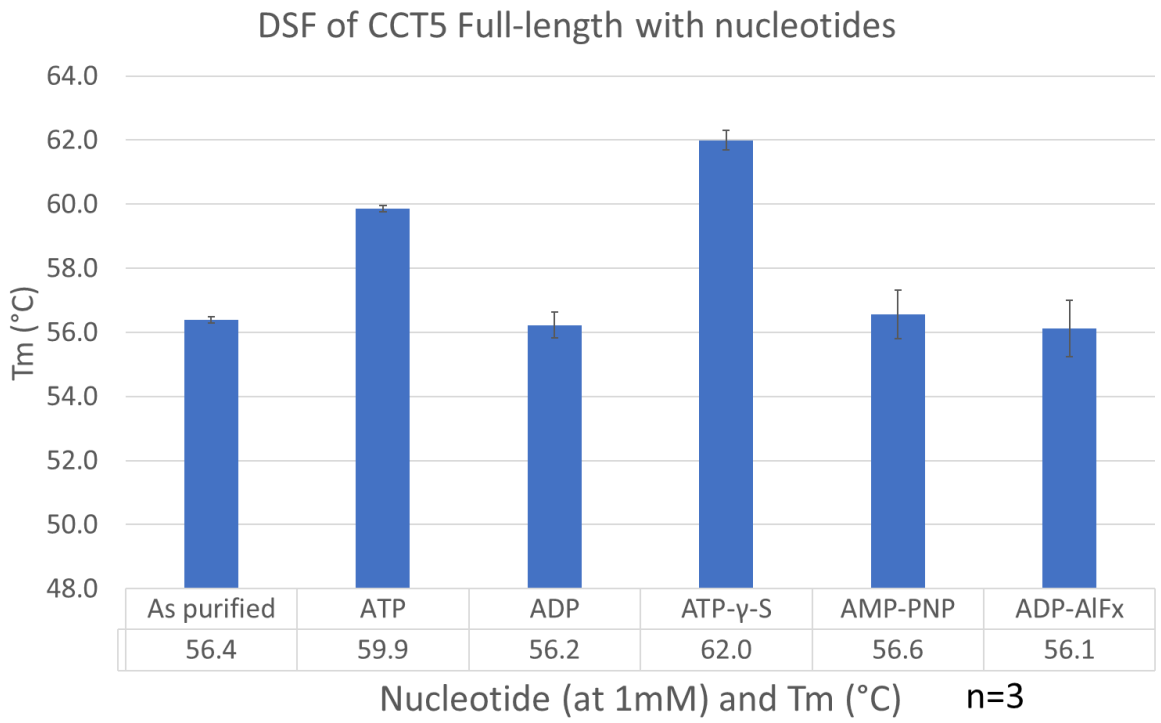


Figure 3.4: Thermal stability of CCT5 with ATP analogues. Bar graph representing thermal melt temperature of $(CCT5)_{16}$ incubated with and without ATP and ATP analogues. Average T_m is reported in table format below. Error bars indicate standard deviation of n=3 replicates.

3.4 Structural study of CCT4 and CCT5 using cryo-EM

3.4.1 Background of negative stain and cryo-EM

The $CCT4_{16}$ and $CCT5_{16}$ homo-oligomers had previously been observed to contain eight-membered double rings in negative stain electron microscopy (EM)[39]. Negative stain EM is a structural technique in which protein samples are visualised by electron scattering in a transmission electron microscope (TEM). To perform this technique, a specimen is adsorbed to an electron transparent sample support, known as a grid, that is a metal mesh that provides support upon which samples sit. These adsorbed samples, or particles, sit in a film of heavy metal salts (usually uranyl formate or uranyl acetate) that provide contrast between the particle and its surrounding. The

stain settles around the sample, filling in between any gaps or spaces within the sample, providing a dark background in contrast to the biological sample, which remains white. Grids are then imaged using a TEM that uses an accelerated beam of electrons that pass through the sample specimen, resulting in scattering of the electrons that is detected using an electron detector. These images can be combined into 2D projections of the sample that can be further used to assemble a 3D structure[202]. The technique is referred to as 'negative' staining as the sample appears white with a dark background.

This contrasts with a related technique known as cryogenic electron microscopy (cryo-EM), in which sample is suspended in vitreous ice without stain on a grid (made of copper or gold typically) with a foil of holey carbon on top. Cryo-EM has the advantage of achieving higher resolution of macromolecular structures as the stain present in negative stain EM masks high-resolution features due to the grain size of the heavy atom salt[203]. However, the increased contrast from negative stain allows for easier viewing of a protein sample[204]. Thus, negative stain EM is often used as a precursor to cryo-EM to check sample quality and observe low resolution features prior to more time consuming and expensive cryo-EM experiments.

Cryo-EM has emerged as an alternative structural method for projects that are not amenable to X-ray crystallography, including samples that do not crystallise, or that form large multicomponent complexes. Samples are adsorbed on a carbon grid (Figure 3.5: step 1) and undergo cryogenic freezing to allow for measurement of high-resolution features. This technique involves plunge-freezing sample grids into liquid nitrogen-cooled liquid ethane to vitrify them (Figure 3.5: step 2), so that particles are suspended in the holey carbon foil holes[205,206]. Images are collected using a TEM and electron

detector (Figure 3.5: step 3a) and particles are grouped based in their orientations to calculate 3D models (Figure 3.5: step 3b).

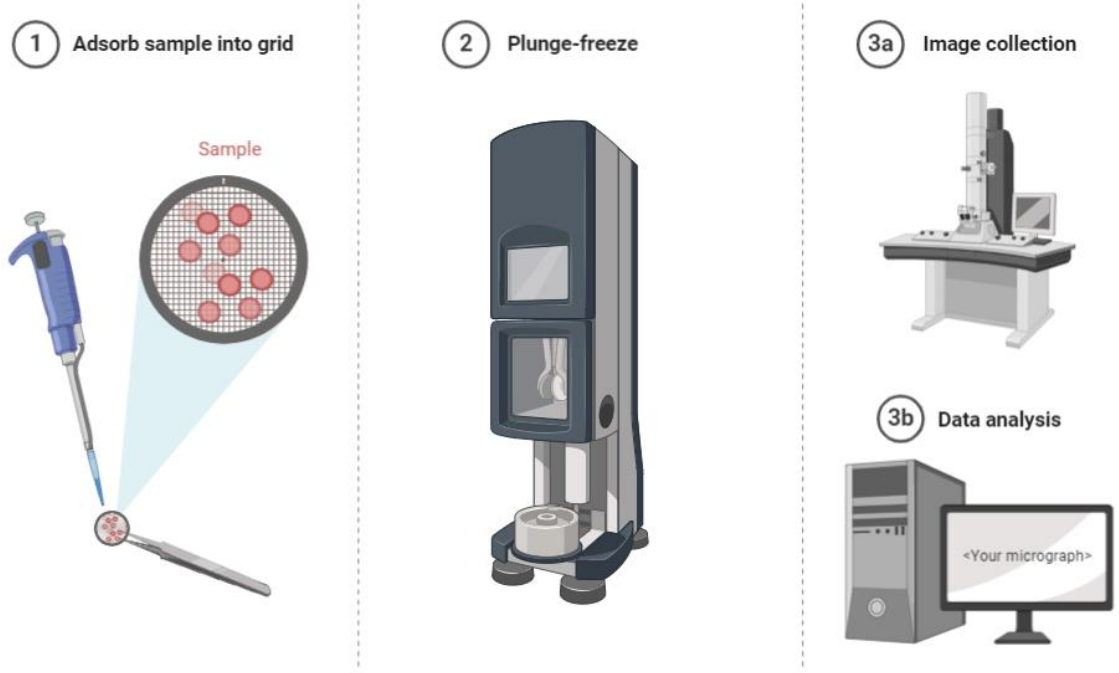


Figure 3.5: Cryo-EM workflow. Diagram shows workflow of a cryo-EM experiment in which samples are applied to an EM grid (step 1), are plunge-frozen in liquid nitrogen-cooled liquid ethane (step 2), and data is collected on a TEM and electron detector (step 3a) and subsequently processed (step 3b) to create a 3D map of the sample. Diagram created using BioRender online server.

In the data processing arm of cryo-EM, a computational technique called "projection matching" is used to match each 2D particle with an initial model that is similar to the expected final structure of the given specimen and assigns angles to the particles based on cross-correlation[207]. The relative orientation of particles can be determined using the Fourier transform of the 2D particle images, as projections with the same orientation have common lines in the 3D Fourier transform[208,209]. The power spectrum generated from the Fourier transform can also be used to estimate defocus parameters using the amplitudes from the power spectrum in an algorithm known as the

contrast transfer function (CTF)[210]. These defocus estimates are important for extracting high resolution information from the power spectrum as the CTF oscillations become more rapid towards higher resolution.

Software packages for cryo-EM data processing, including the popular RELION framework, combine the latest techniques to construct high resolution density maps that can be used to build atomic models of target proteins[185,211,212]. These packages typically employ a standard framework (Figure 3.6) including: 1) import movie frames and correct for drift and beam-induced motion using motion correction[213], 2) particle picking[188], 3) contrast transfer function (CTF) correction estimation[184], 4) extract particles and classify into 2D images, 5) create initial model using class averages[186], 6) perform 3D classification and refinement of particles that produce electron density maps.

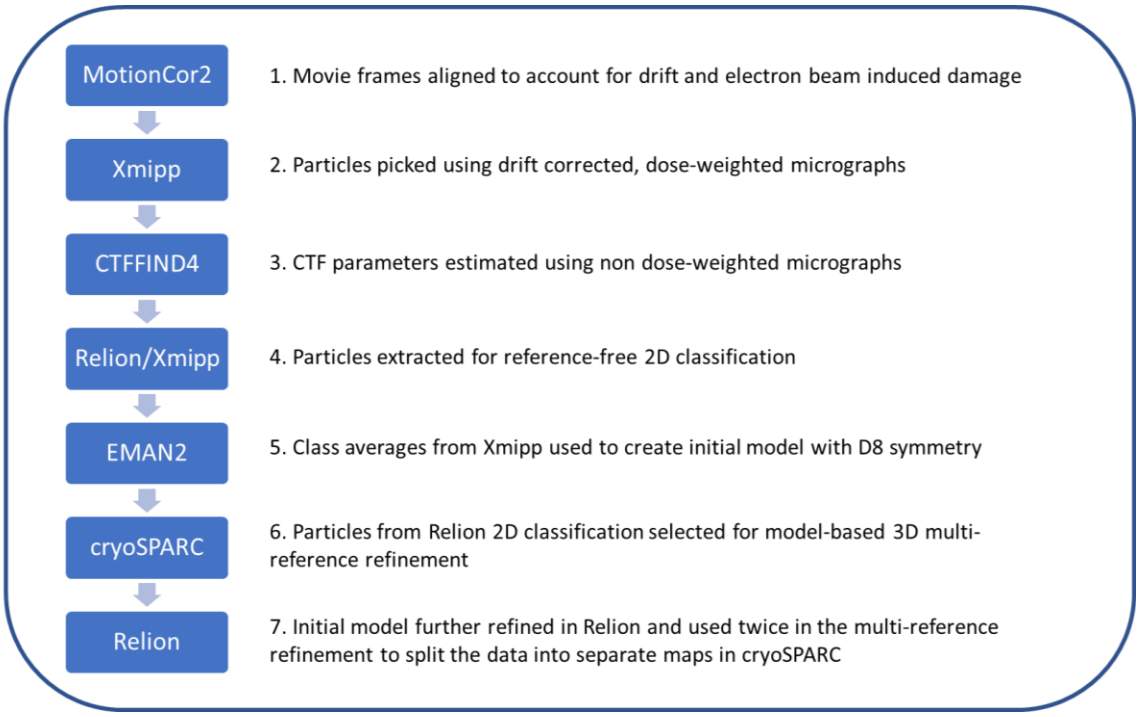


Figure 3.6: Typical cryo-EM data processing workflow. Data processing involves movie alignment, particle picking, CTF estimation, particle extraction, 2D and 3D classification, and iterative refinement.

As both recombinant $CCT4_{16}$ and $CCT5_{16}$ homo-oligomers had previously been observed in negative stain, indicating proof of concept for structural studies, each was directly screened in this study using cryo-EM to obtain high resolution structures of each. The goal of these experiments was to determine the structures of CCT4 and CCT5 and would have resulted in the first full-length human CCT subunits to be structurally characterised at the time of the study.

3.4.2 Initial screening of grids and small data collection

To carry out this workflow for the CCT homo-oligomeric complexes, samples containing $CCT4_{16}$ and $CCT5_{16}$ were applied to cryo-EM grids and vitrified using a Vitrobot. Data were initially collected for each sample on a 300 kV Polara electron microscope fitted with a K2 detector at the Oxford Particle Imaging Centre (OPIC), to assess whether grids were of high enough quality for data collection on a high-end Titan Krios equipped with a newer K3 detector. Micrographs from the Polara collection revealed monodisperse samples of $CCT5_{16}$, while $CCT4_{16}$ contained a mixture of single particles and polymers containing CCT4 stacked on itself in filamentous rods (Figure 3.7). Data were collected and images were processed according to the workflow mentioned in the previous section. 2D classification revealed oblong CCT4 projections that appeared to be in the open conformation (Figure 3.7: left inset). The CCT4 polymerisation effect observed in the micrographs hindered further data processing to high resolution due to a lack of different orientations and was not pursued further. On the contrary, the 2D class averages for the CCT5 data (Figure 3.7: right inset) contained various top-down and side views of $CCT5_{16}$ that showed eight-membered rings.

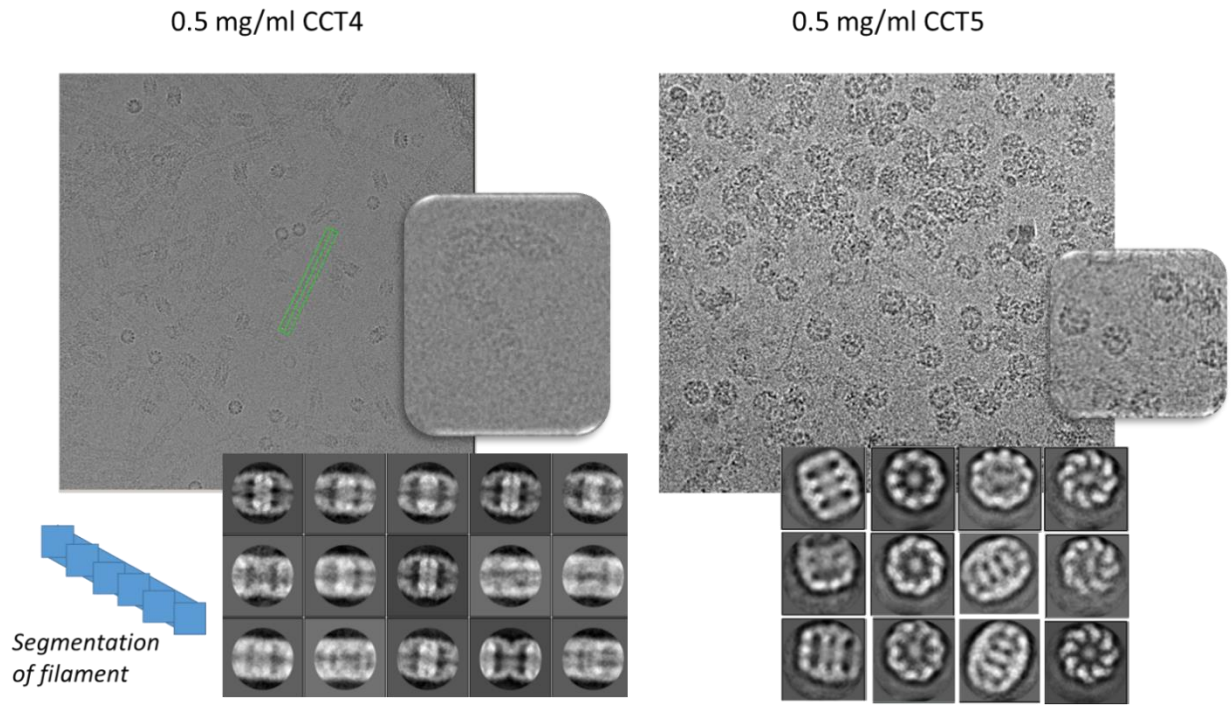


Figure 3.7: Preliminary grid screening and data collection of CCT4 and CCT5 homo-oligomeric complexes. Top: Cryo-EM micrographs of CCT4 (left) and CCT5 (right) showing particle distribution of CCT complexes. Insets: zoom-in of electron micrographs. Bottom: representative 2D classifications. Green box and blue diagram denote filamentous polymers of CCT4.

3.4.3 Large scale data collection of CCT5 homo-oligomeric complexes in open and closed conformations

The $CCT5_{16}$ data obtained on a Polara TEM at the Oxford Particle Imaging Centre (OPIC) was deemed sufficient for collection on the more powerful Titan Krios to achieve higher resolution final maps than were possible on the Polara TEM. Samples were prepared in which $CCT5_{16}$ was incubated with either ATP-γ-S or ADP-AlF_x to induce the open and closed conformations, respectively. Grids were screened and confirmed to contain monodisperse particles with optimal ice thickness (Figure 3.8). The influence of ice thickness on cryo-EM samples will be further elaborated in Section 5.2.2 of Chapter 5.

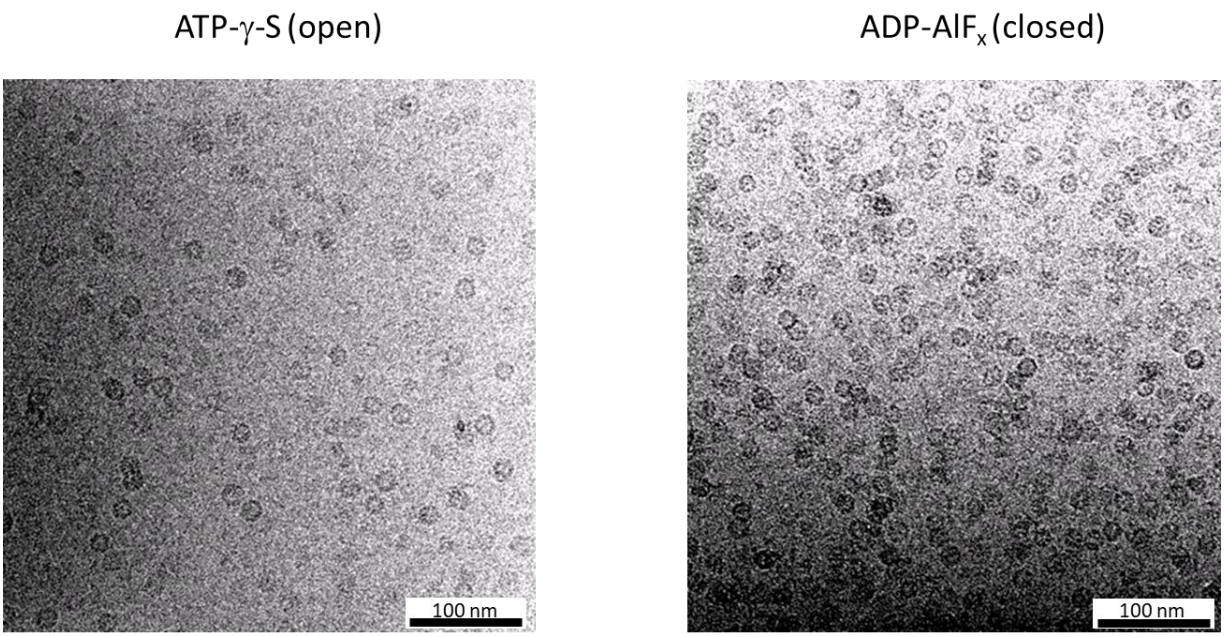


Figure 3.8: Representative cryo-EM micrographs from CCT5$_{16}$ grid screening. Left: CCT5$_{16}$ incubated with ATP-γ-S to induce open conformation. Right: CCT5$_{16}$ incubated with ADP-AlF$_x$ to induce closed conformation.

Data were collected for each set and were further processed using the workflow mentioned above (Figure 3.5). The ADP-AlF$_x$ data set achieved higher resolution and was pursued in favour of the ATP-γ-S set, as the latter data set likely exhibited lower resolution due to the increased flexibility of the apical domains in the open conformation. The ADP-AlF$_x$ data set contained approximately 150k particles and resulted in a cryo-EM density map with global resolution 4.0 Å after imposing D8 symmetry (Figure 3.9).

Imaging conditions		
Data Acquisition	Electron detector	K2 Summit direct
	Movies collected	25 frames each
	Magnification	47,170×
	e^- dose	~50 e^- per Å^2
	Defocus range	0.8–3.0µm
	Micrographs	1,085
Pre-processing	Pixel size	1.06 Å
	Particles (after 2D classification)	191,782 (157,777)
	Box size	300 pixels



CCT5-ADP-AlFx	
Symmetry	D8
Particles (3D)	66,516 (40,494)
Resolution (Å)	4.0
B-factor (Å^2)	178

Figure 3.9: $CCT5_{16}$-ADP-AlF_x data processing statistics. Parameters for imaging conditions (left) and resulting map information (right).

The $CCT5_{16}$-ADP-AlF_x data set achieved a final resolution of around 4.0 Å and was intended to induce the closed state. However, the map generated clearly did not reflect a closed state, as resolution of the density decreased gradually from the equatorial domains to the apical domains (Figure 3.10A). Unfortunately, the density around the nucleotide binding pocket in these maps was not of sufficient resolution to identify whether a nucleotide was bound. It is therefore possible that the $CCT5_{16}$ in the sample prepared prior to grid freezing did not bind ADP-AlF_x properly, leading to a large portion of molecules in the open state. The lack of apical domain resolution and overall average density for the equatorial domains made building an atomic model a superfluous exercise.

A model of yeast CCT5[138] could be fit into the density of a single subunit (Figure 3.10B) and fitting in repeating units of this subunit revealed that $CCT5_{16}$ indeed forms double-stacked rings each comprising eight identical CCT5 subunits (Figure 3.10C). While this fitted model reveals a seemingly flexible open state of CCT5, no atom-level information was revealed for human CCT5 due to a lack of side chain density (Figure 3.10B). While this work was ongoing, a crystal structure of human $CCT5_{16}$ at 3.6 Å resolution was published in which the synthetic chaperonin was incubated with ADP to promote the closed state[214]. While atomic models of homo-oligomeric CCT4 and CCT5 were not completed due to the reasons stated, a clear path to structural characterisation of chaperonins was established during the course of the study, whereupon successful isolation of native TRiC (see Chapter 4) could be further used to determine the structure of human TRiC (see Chapter 5).

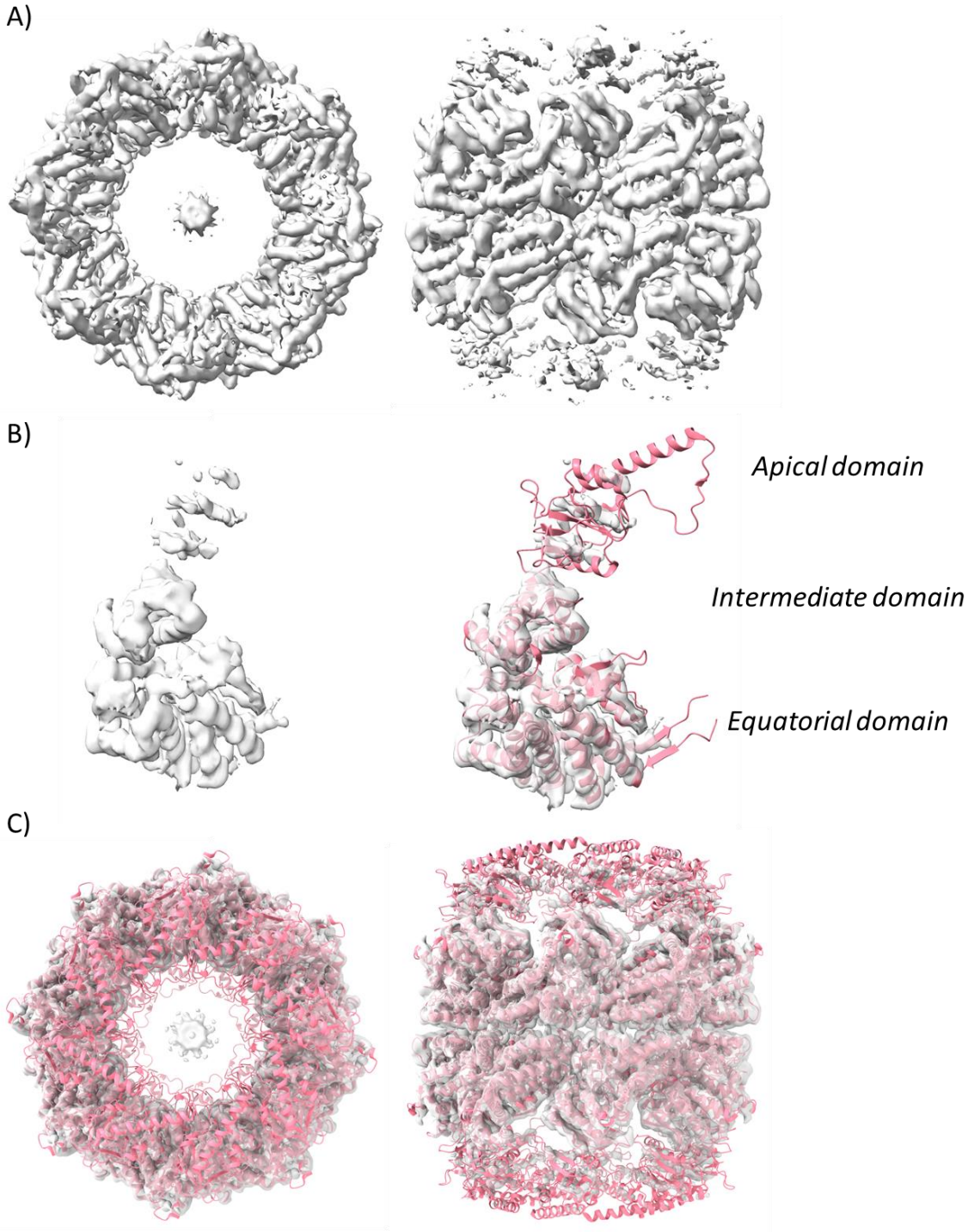


Figure 3.10: Density map of $CCT5_{16}$-ADP-AlF_x. A) Side (left) and top-down (right) views of 4.0 Å reconstruction of CCT5 homo-oligomeric complex processed using D8 symmetry. B) Side view of map of a single CCT5 subunit without (left) and with (right) yeast CCT5 fit into the density (PDB: 5GW5, pink cartoon)[138]. C) Overall $CCT5_{16}$ homo-oligomer assembly.

3.5 Expression and purification of individual subunits from insect cells

While investigations were ongoing into the previously studied CCT4 and CCT5 homo-oligomeric complexes overexpressed in *E. coli*, the other TRiC subunits (CCT1, CCT2, CCT3, CCT6, CCT7, and CCT8) were also individually sub-cloned for expression in baculovirus-mediated Sf9 insect cells. The goal was to further investigate what made CCT4 and CCT5 assemble into homo-oligomeric complexes while the other subunits were seemingly insoluble based on previous reported attempts at *E. coli* over-expression[39]. It is possible that other subunits could also form these synthetic homo-oligomers when over-expressed in a different host. To this effect, all CCT subunits were ligation independent cloning (LIC) cloned into the pFL MultiBac acceptor vector (see section 4.3.1 for more details) incorporating an N-terminal His6 tag. Each plasmid encoding a single subunit was transiently transfected, expressed in 50 mL Sf9 cultures, and purified through Ni-NTA affinity columns to assess whether expressions yielded soluble protein (Figure 3.11A).

The expression test clearly showed purified CCT1-CCT8 subunits from one-step affinity purification, with no observable degradation and of sufficient quantities to further characterise using size exclusion chromatography (SEC). The total, unnormalised elution fractions from each individually expressed subunit were passed through an S200 SEC column and compared side-by-side (Figure 3.11B). As expected from studies involving *E. coli* expression, $CCT4_{16}$ and $CCT5_{16}$ expressed from insect cells both eluted as large high molecular weight (HMW) complexes with elution volume (V_e) ~70 mL, close to a 1 MDa species as revealed from size exclusion based on molecular weight standards. However, other individually expressed subunits exhibited a range of

elution profiles. CCT2 and CCT4 appeared to only exist as large HMW species, suggesting that CCT2 forms homo-oligomeric complexes. In contrast, other subunits seemed to elute in two separate peaks to varying degrees. Among them, CCT6, CCT7, and CCT8 eluted predominantly as low molecular weight (LMW) species with V_e ~90 mL, while CCT1 and CCT5 eluted predominantly as HMW species. CCT3 appeared to contain peaks at both the HMW and LMW species.

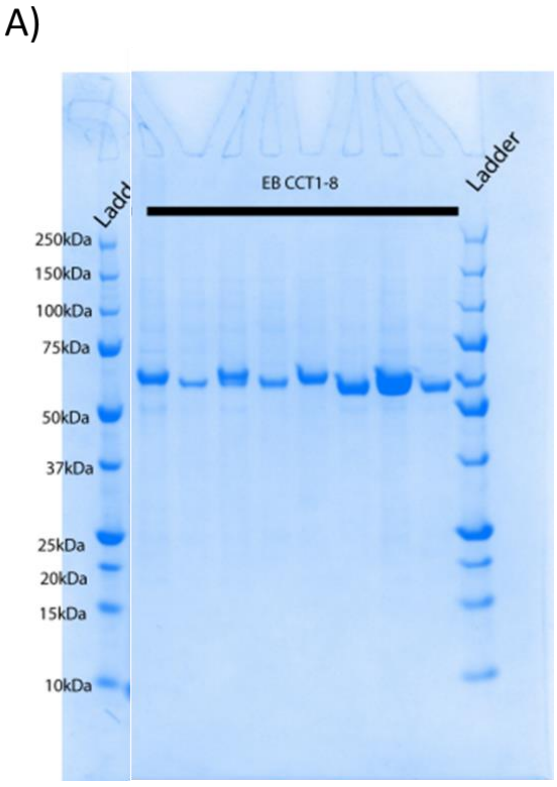
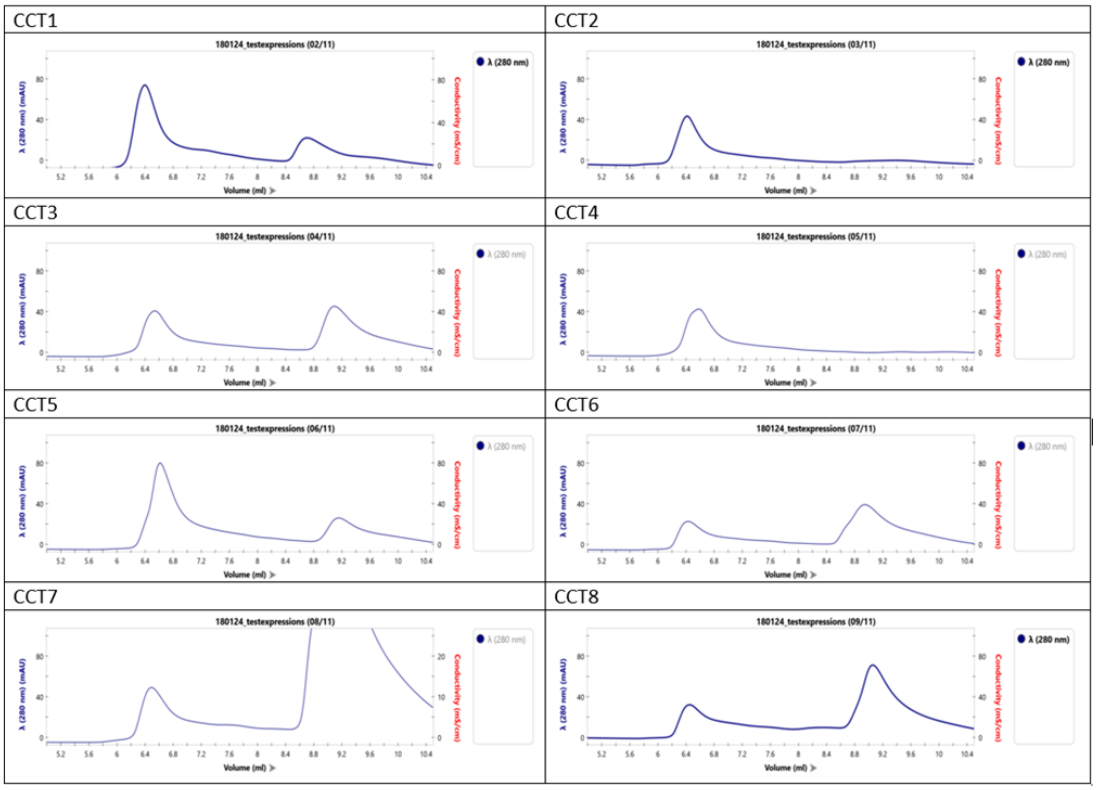


Figure 3.11: All CCT subunits can be solubly over-expressed individually. A) SDS-PAGE of Ni-NTA elution fractions for all eight CCT subunits individually expressed in Sf9 cells. B) Size exclusion chromatograms of individual subunits from SEC fractions, displaying varying elution profiles between high and low molecular weight species.

B)



To recapitulate the results observed in the small-scale expression and prepare sufficient quantities for further biochemical study, each individual subunit was scaled up to 1–6 L expressions and the purifications were repeated. After affinity purification, size exclusion chromatography was performed in a HiLoad 16/600 Superose 6 prep grade column. It was observed that the HMW peaks and LMW peaks for all CCT1-CCT8 subunits eluted between 60-80 mL and 80-100 mL, respectively (Figure 3.12). In the large-scale expression, the elution profile was slightly different from small scale, albeit consistently among three replicate experiments for each subunit. Subunits CCT1, CCT3, CCT6, CCT7, and CCT8 all mainly eluted as LMW species, whereas CCT2, CCT4, and CCT5 all eluted as HMW species (Figure 3.12). However, CCT4 and CCT5 eluted as two equal peaks containing both HMW and LMW species.

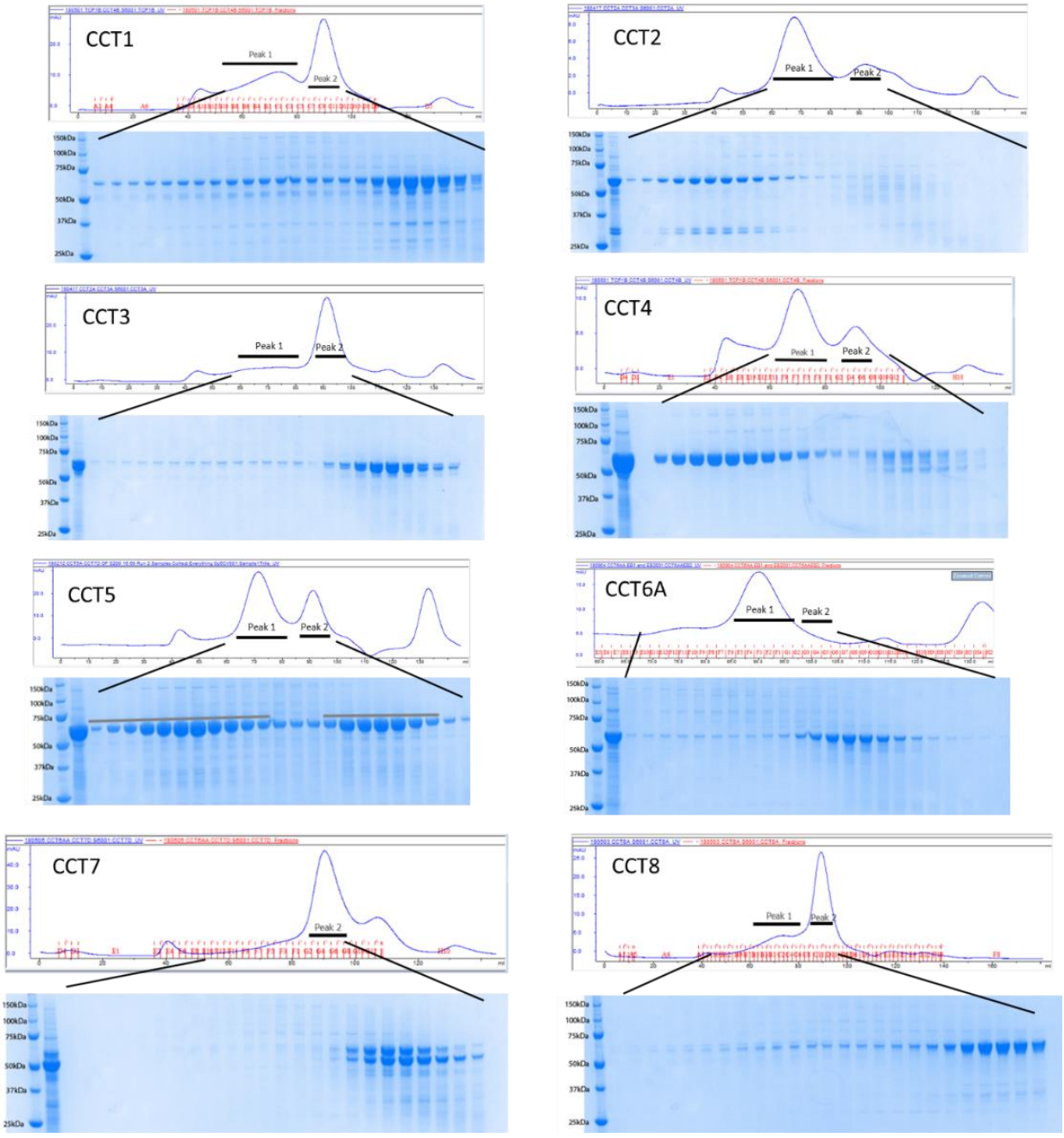


Figure 3.12: Large scale purifications of CCT subunits. Superose 6 SEC chromatogram (top) and SDS-PAGE gels (bottom) are shown for 6L expressions of each CCT subunit except CCT6, in which 12L were expressed. Black lines indicate regions of SEC chromatogram that are represented on gel.

3.6 SAXS reveals different oligomeric states

The HMW species observed for the Sf9-expressed subunits would imply that these samples are also ~1 MDa in molecular weight and could contain 16 copies, as observed with $CCT5_{16}$. To confirm whether this formation of homo-oligomeric complexes was indeed the case for other subunits, each purified subunit protein was investigated using size exclusion coupled with small angle X-ray scattering (SEC-SAXS) by staff at the Diamond Light Source BL21 beamline. SAXS uses high-flux X-ray scattering at small angles to identify differences in the size and shape of monodisperse macromolecular samples[215–217]. A range of intensities (I) are measured over these scattering angles that give a one-dimensional intensity profile as a function of the momentum transfer vector *q*. This vector is used to estimate several features of the sample including molecular weight, volume, and overall size and shape *via* the radius of gyration (R_g) and maximum dimension (D_{max}). These features can subsequently be used to build *ab initio* bead models of the sample to estimate its oligomeric state. Therefore, based on MW estimations, the profile derived through SAXS can estimate the approximate number of subunits within a sample to confirm whether each CCT subunit purified in HMW and LMW peaks are homo-oligomeric, monomeric, or a mixture of micro-complexes (dimers, trimers, etc.) in between.

Each purified CCT subunit was run at the Diamond Light Source beamline B21 with CCT5 HMW and LMW samples initially characterised as reference points to compare the oligomeric state of other subunits (Figure 3.13). For both CCT5 HMW and LMW samples the scattering curves (Figure 3.13A) showed a linear Guinier fit (ln I(*q*) vs. q^2; Figure 3.13B), indicative of a monodispersed sample. The calculated R_g of CCT5 HMW

was ~70 Å, while the R_g of the CCT5 LMW sample was ~35 Å, suggesting the LMW species was significantly smaller. The Kratky plot ($q^2I(q)$ vs q) indicated a folded, multi-domain species for HMW (Figure 3.13C: left), while the LMW solution was estimated to be closer to a monomeric state (Figure 3.13C: right).

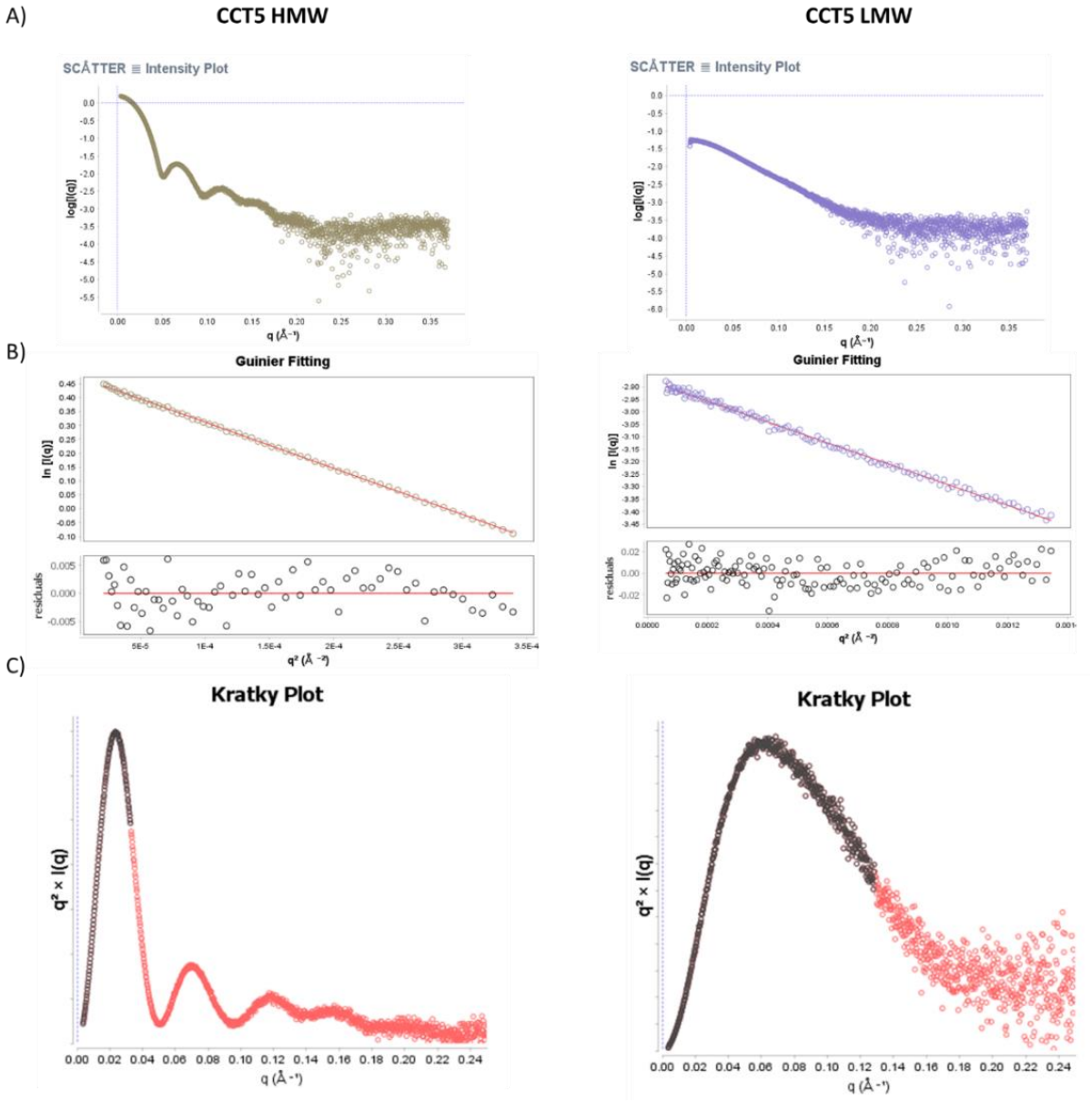


Figure 3.13: Representative SAXS analysis of CCT5 HMW and LMW samples. A) Scattering curve of HMW and LMW samples of CCT5 gathered at the B21 beamline of Diamond Light Source synchrotron. Data are plotted to show log[I(q)] against q. B) Gurnier plot of ln[I(q)] against q^2 exhibiting a linear fit for both HMW and LMW samples. C) Kratky plot of q^2 x I(q) against q showing folded species of each sample, while the HMW species exhibited a profile indicative of a large oligomeric species, while the LMW species appeared to correspond to low molecular weight similar in size to a monomer/dimer.

Comparing SAXS results from all the subunits showed a general trend, whereby HMW species had larger R_g and D_{max} values as compared to LMW species (Table 3.1). This trend agreed with the expectation that the HMW species represent a larger homo-oligomeric species, and the experimentally derived size estimates of HMW peaks suggested complexes of larger than 1 MDa. The CCT4 and CCT5 samples were already known to form 16-mer complexes and thus it is assumed the other HMW species form similar complexes of 60 kDa species copied 16 times, equalling ~1 MDa. The LMW species by contrast had estimated sizes closer to the monomer/dimer range (e.g., the MW observed for CCT5 LMW = 72 kDa, where the theoretical MW of one monomer = 60 kDa; Table 3.1). Table 3.1 also shows that the estimated sizes were consistently larger than expected for both HMW and LMW species. This could be because the oblong shape of CCT subunits in the open state creates flexibility in movement of the apical domains. This increased range of flexibility would artificially increase the size estimates as the diameter of the sample would appear larger than the tight sphere created in the closed state.

Table 3.1: SAXS data analysis parameters for all CCT subunits. HMW samples are shown in purple text while LMW are in brown text. Radius of gyration (R_g), D_{max}, estimated experimental size (Est Size), theoretical monomer size (Exp Size monomer), and χ^2 values fitted with C1 and D8 symmetry are reported with poor fit denoted (n/a) as appropriate.

Sample ID	R_g (Å)	D_{max} (Å)	Est Size	Exp Size (monomer)	χ^2 1mer	χ^2 16mer
CCT1 LMW	41.6	136	103kDa	60.3 kDa	n/a	n/a
CCT2 HMW	79.1	320	1.5MDa	57.5 kDa	n/a	55.823
CCT3 HMW	88.3	320	1.6MDa	60.5 kDa	94.821	4.92
CCT3 LMW	39.6	120	92kDa	60.5 kDa	61.454	n/a
CCT4 HMW	81	268	1.9MDa	57.9 kDa	n/a	n/a
CCT4 LMW	72	220	1.3MDa	57.9 kDa	n/a	n/a
CCT5 HMW	70.2	198	1.25MDa	59.7 kDa	n/a	n/a
CCT5 HMW (split)	71.2	220	1.29MDa	59.7 kDa	n/a	28.579
CCT5 LMW (split)	35.2	106	72kDa	59.7 kDa	1.774	n/a
CCT6 HMW	102	367	1.5	58 Da	57.447	7.013
CCT7 LMW	44.8	151	83kDa	59.3 kDa	n/a	n/a
CCT8 LMW	35.4	105	76kDa	59.6 kDa	n/a	n/a

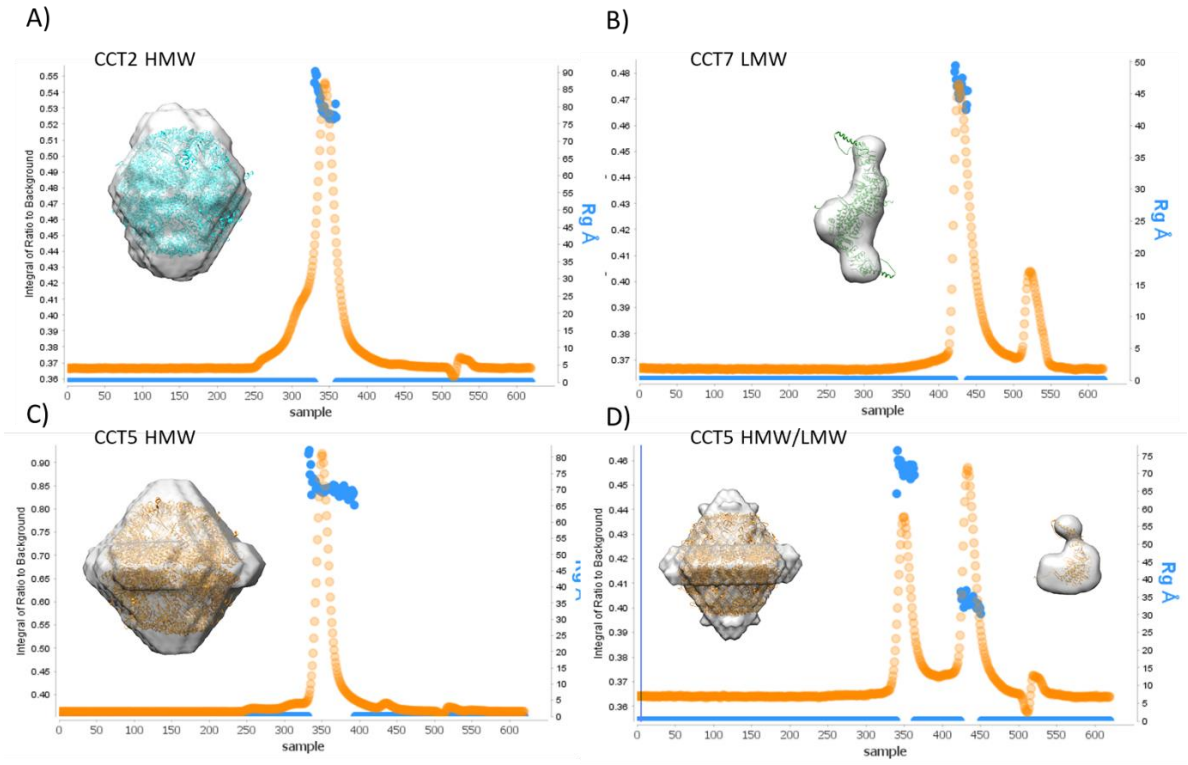
The SAXS data from several subunits were further analysed including CCT2, CCT5, and CCT7. CCT2 and CCT7 each only elute into one species from size exclusion column, namely HMW for CCT2 and LMW for CCT7 (Figure 3.14A and B). CCT5 consistently elutes as an equal distribution of HMW and LMW species, and both peaks were individually analysed in SEC-SAXS (Figure 3.14C and D). *Ab initio* bead models were generated with the expected symmetry for each of these SAXS data sets, to analyse the overall shape of the SAXS data.

The *ab initio* model of CCT2 HMW is consistent with a 16-mer complex that was representative of its large R_g values in the SEC trace (Figure 3.14A). Interestingly, CCT7 LMW did not produce reasonable bead models when C1 symmetry was imposed, however a model was generated with C2 symmetry (Figure 3.14B). This model could

accommodate a dimer that stacks at the equatorial domain and could explain the estimated size being larger than 60 kDa as suggested by values derived from the Kratky plots, being unable to fit a monomer or 16-mer. The CCT5 HMW sample behaved in a similar manner as the CCT2 HMW sample, and the bead model also fit a large complex, likely the 16-mer (Figure 3.14C). Surprisingly, the CCT5 LMW sample re-distributed into two peaks when run on SEC-SAXS, hence forming an equilibrium between the HMW and LMW species (Figure 3.14D; CCT5 HMW/LMW). These two peaks were analysed for R_g and D_{max} determination, and the HMW peak was similar in R_g and D_{max} to the already processed CCT5 data, while the LMW peak was similar to CCT7 in size. In shape, the CCT5 LMW sample appeared to fit into a bead model resembling a monomer of a CCT subunit.

Bead models are inherently untrustworthy as symmetry is applied at the beginning of the run. To provide statistical validation for bead models, each of the subunits analysed by bead model were fit with a χ^2 value that describes the quality of fit of the experimentally derived data as compared to homology models of monomers and homo-oligomeric complexes from PDB coordinates (Figure 3.14E; also shown in Table 3.1). The HMW species from CCT2 and CCT5 could not be fitted to a monomer but did fit to a certain degree to 16-mer complexes (χ^2 = 56 and 29, respectively). By contrast, the CCT5 LMW peak expectedly fit a monomer (χ^2 = 1.8) but not a 16-mer. Interestingly, the fit of CCT7 supports the bead model as it was able to be fitted to a dimer model (χ^2 = 5.6), but not a monomer or 16-mer. Overall, the SAXS data provides further evidence that the various samples of individual CCT subunits seen eluting at HMW are homo-

oligomeric complexes, while the LMW species are likely monomers, dimers, or a mixture of such species.



E)

Sample ID	R_g (Å)	Dmax (Å)	χ^2 1mer	χ^2 2mer	X^2 16mer
CCT2 HMW	79.1	320	n/a		55.823
CCT5 HMW	70.2	198	n/a		n/a
CCT5 HMW (split)	71.2	220	n/a		28.579
CCT5 LMW (split)	35.2	106	1.774		n/a
CCT7 LMW	44.8	151	n/a	5.578	n/a

Figure 3.14. CCT subunit size and shape estimates using SEC-SAXS. SEC traces, R_g values, and bead models (inset) of A) CCT2 HMW, B) CCT7 LMW, C) CCT5 HMW, and D) CCT5 HMW/LMW. Blue dots indicate radius of gyration while orange dots indicate UV spectrum from SEC run. Bead models are shown in grey surface and are fitted with CCT subunits. E) Table listing values of R_g, Dmax, as well as χ^2 fitted using C1, C2, and D8 symmetry (1mer, 2mer, and 16mer). Purple text denotes HMW profiles and brown text denotes LMW profiles.

3.7 CCT subunits characterised by stability and activity assays.

3.7.1 DSF stability measured for all eight CCT subunits

While the SEC (Section 3.5) and SEC-SAXS (Section 3.6) data suggested the individual CCT subunits isolated are soluble protein that form either homo-oligomeric or lower species, it still was not known whether the protein being isolated was of functional relevance. To assess this in a biophysical context, all individual subunits from each peak formed at SEC-SAXS were studied using DSF at the same final concentration (Figure 3.15). The samples showed a trend in which HMW samples for a given subunit were more thermostable than the LMW counterpart of the same subunit, exhibiting T_m differences between HMW and LMW in the range of 2 to 7°C. The two samples with the highest T_m values and T_m difference included CCT4 and CCT5 homo-oligomers, suggesting these subunits are the most thermostable among all the CCT samples. While not all subunits were able to be measured by DSF, this trend is expected to be consistent due to the stability provided by homo-oligomeric assemblies.

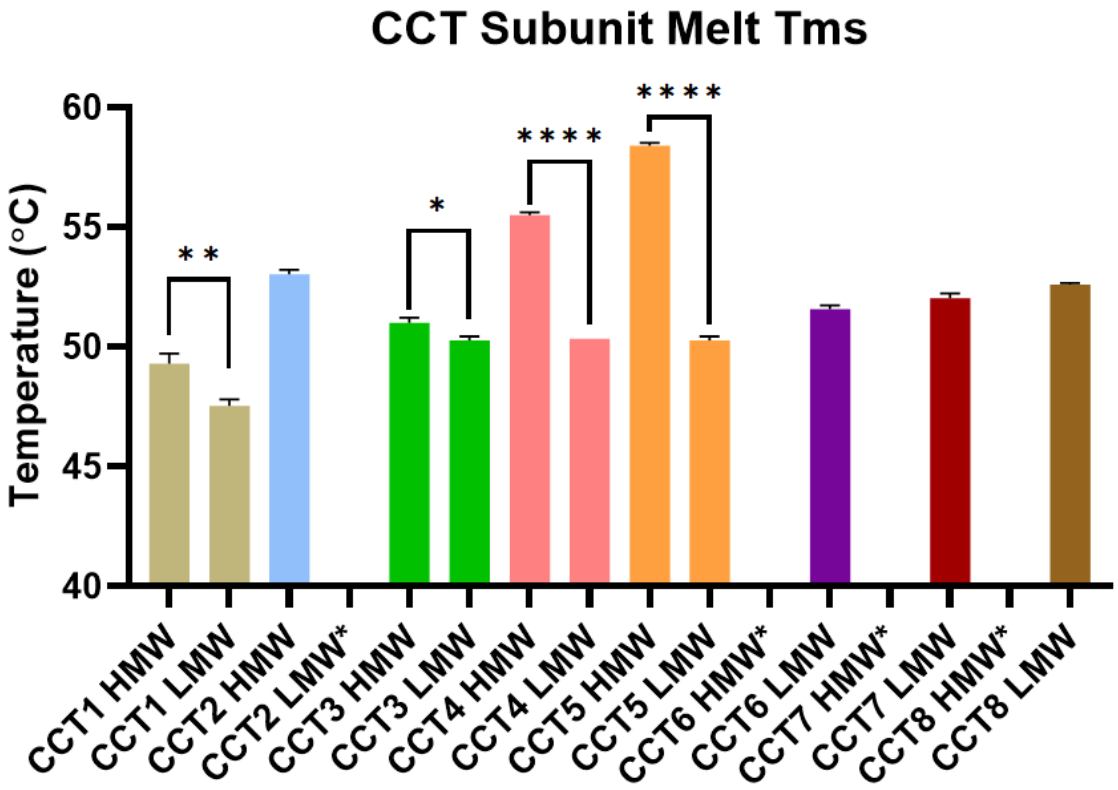


Figure 3.15: Melt temperatures (T_ms) of all eight CCT subunits. Samples include HMW and LMW samples as indicated. Samples that could not be run in DSF, either due to a lack of their formation in the first place or due to noisy data that could not be processed, are indicated (X). Errors bars indicate standard deviation from n=3, *** p<0.0005, **** p<0.0001, statistics based on one-way ANOVA.

3.7.2 ATPase activity assay measured for all subunits

To evaluate whether the additional CCT subunits outside of CCT4 and CCT5 are functionally active, each individual subunit protein, in available HMW and/or LMW species, was tested using the ATPase assay (Figure 3.3A) to check for ATP hydrolysis activity (Figure 3.16). To run HMW and LMW samples in parallel for comparison, protein concentrations were normalised to contain the same molar amount of a given subunit to accurately assess differences in ATP hydrolysis between these two species. The subunit with the highest absolute ATPase activity was CCT5 and, interestingly, both

HMW and LMW peaks exhibited robust hydrolysis activity, with HMW activity significantly higher. Of additional interest was CCT4, which exhibited the second highest activity but only for the LMW species. One explanation for the low ATP hydrolysis activity for the CCT4 HMW peak could be due to polymerisation tendency of the homo-oligomeric $CCT4_{16}$ complexes, as seen in cryo-EM micrographs (Figure 3.5: left). The remaining subunits all showed low or no ATPase activity, regardless of their HMW or LMW status. There were significant differences in activity between HMW and LMW species of the same subunit, though no trends emerged.

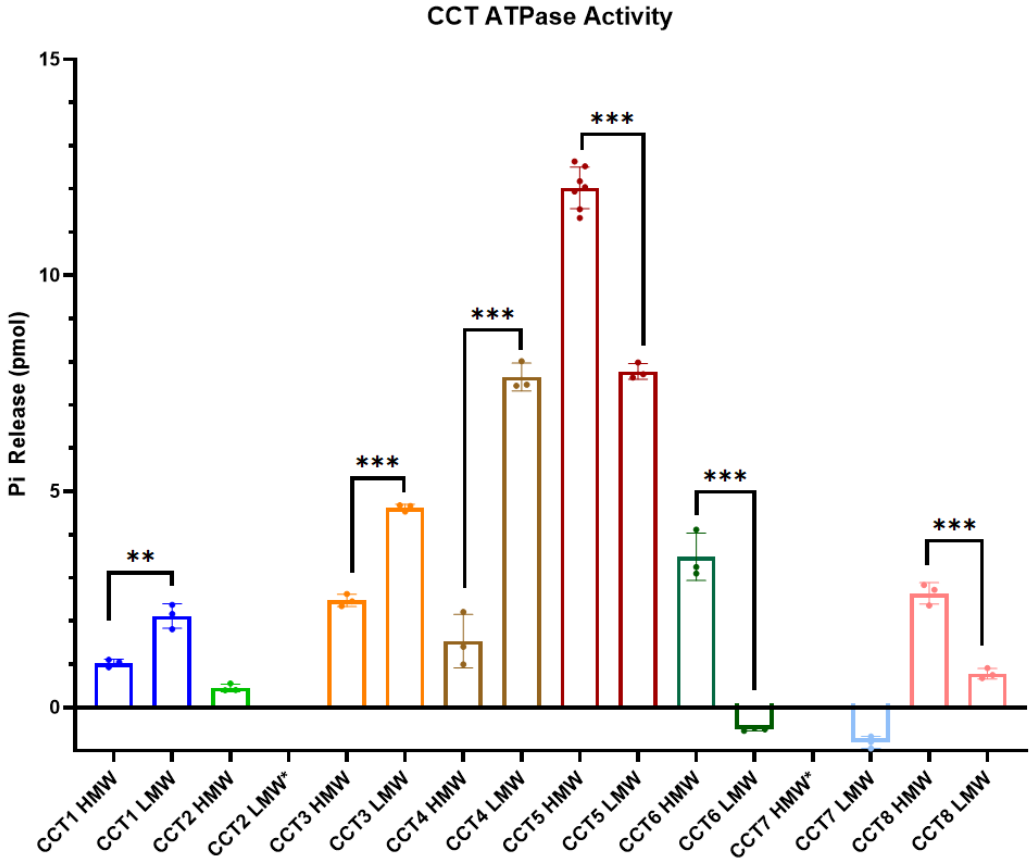


Figure 3.16: CCT subunits measured for ATP hydrolysis activity. Free phosphate Pi release is plotted with averages (bar graph) and individual data points (dots). Asterisks mark samples that could not be run due to lack of formation of that species. Errors bars indicate standard deviation from n=3, ** p<0.002, *** p<0.001, statistics based on one-way ANOVA.

3.8 Negative stain EM reveals homo-oligomeric CCT complexes

The DSF and ATPase data begin to show a trend in which the previously well characterised CCT4 and CCT5 subunits are the only hydrolytically active individual subunits, and that while the other CCT homo-oligomeric subunits can be purified they might not be stable and functional complexes. To directly observe whether these HMW samples assemble into homo-oligomeric complexes, each was applied to negative stain EM grids and imaged (Figure 3.17). From the electron micrographs, CCT5 contained the most uniform ring-like particles as expected. CCT4 also contained ring-like particles though some polymerisation was also observed as seen in CCT4 cryo-EM studies (Figure 3.7). All remaining subunits showed evidence of homo-oligomeric complexes for their HMW samples, including CCT1 in which the eight-subunit ring arrangement could be directly discerned from the particles. CCT2 particles were of the same size and shape as those seen in the other subunits, although the number of subunits present could not be determined. CCT3, CCT6, and CCT8 all showed particles that resembled homo-oligomeric CCTs, however each had high background contamination (CCT7 not shown due to lack of HMW species).

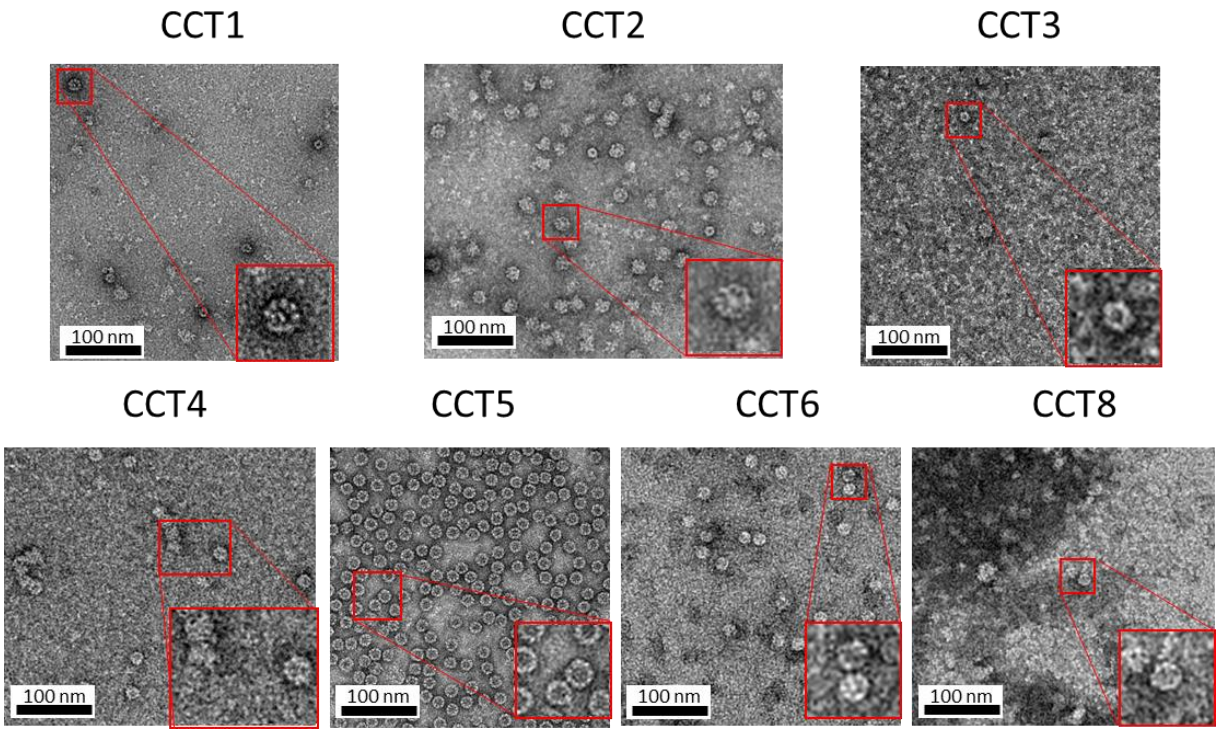


Figure 3.17: Negative stain micrographs of CCT HMW species. Insets show zoom-in of particles resembling homo-oligomeric complexes.

3.9 Summary

The data presented in this chapter detail efforts to enable the study of architecture and mechanism of group II chaperonins, using surrogates that bypass the need for complex purification strategies to isolate native TRiC. The first section uses the previously studied $CCT4_{16}$ and $CCT5_{16}$ model systems to establishes biochemical and structural methods to characterise chaperonins including thermal stability, ATP activity, substrate aggregation suppression, and EM. These investigations showed that the observations previously reported for these homo-oligomeric subunits could be recapitulated and these assays were established to enable downstream characterisation of additional synthetic and native chaperonins.

The second section explores the expression of individual subunits and characterises each using these techniques. The individual CCT subunits emerged as interesting subjects for study as CCT subunits outside of CCT4 and CCT5 have not been well characterised. This work is the first to date showing that all CCT subunits can be expressed recombinantly and purified. Therefore, previous observations that only CCT4 and CCT5 can be expressed recombinantly is likely a function of the expression system. However, it appears from this study that CCT4 and CCT5 are the only subunits that retain a significant degree of activity outside of being in the TRiC complex. While CCT subunits might have activity outside of the TRiC complex, this thesis aimed to characterise chaperonin complexes and thus further study of the biologically relevant TRiC complex containing all eight subunits is preferable to further characterisation of individual CCT subunits. Therefore, the need arose to explore methods of isolating native human TRiC and several avenues to this end are explained in the following Chapter 4.

4. Strategies for native protein complex isolation

4.1 Previous attempts to purify eukaryotic TRiC

Expression and purification of TRiC has been a challenge since its discovery due to its composition of eight paralogous subunits. As a result, much of what is understood about chaperonins comes from characterisation of chaperonins with simpler organisation, including GroEL, Mm-Cpn, the thermosome, and individual CCT homo-oligomers described in the previous Chapter 3. Several strategies have been employed in the past to isolate the full TRiC complex, the earliest and most successful to date of which includes endogenous isolation primarily from bovine testes[24,195,218]. This purification procedure is time-consuming, contains many steps, and requires a large amount of starting material as it does not utilise affinity tagging for isolation and purification. Instead, basic biochemical techniques coupled with several chromatography steps are needed to isolate TRiC to reasonable levels of purity. This approach has also been performed in HeLa cells to generate endogenous human TRiC, but required a large starting sample (100 gram cell pellet) and has not seen widespread use[198]. Recombinant strategies have not been commonly employed, however one study showed success in co-expressing all eight CCT subunits, but this resulted in low yields[219]. More recently exogenously tagged subunits in yeast[131,220] and human[149] have improved purifications of TRiC and are used in a more widespread manner. This chapter explores isolation and purification strategies for protein complexes, using recombinant expression, nanobody pull-down, and CRISPR/Cas9 mediated genome editing. These techniques are employed using TRiC as a model system to illustrate broad applications to a wide range of protein complexes.

4.2 Endogenous purification from wild-type mammalian cells

The route to isolate TRiC with the least amount of prior effort remains endogenous purification from wild-type cells. As the projects in this thesis are wholly dedicated to understanding human TRiC, the first purification strategy employed was to mimic the purification protocol used in isolating human TRiC from mammalian cell cultures[198]. While the published protocol began with a large cell pellet from a 20 L HeLa culture, obtaining this large cell mass is not a sustainable method for isolating usable quantities of TRiC for structural and biochemical characterisation. Therefore, this strategy was undertaken using ExpiHEK293 cells in suspension, a cell line with similar CCT expression levels to HeLa (Figure 4.1) and using smaller pellet sizes to assess the feasibility of this method.

Protein	H.sapiens - Cell line, Hela, SC (Geiger,MCP,2012)	H.sapiens - Cell line, HEK293, SC (Geiger,MCP,2012)
TCP1	722 ppm	1142 ppm
CCT2	270 ppm	41.3 ppm
CCT3	762 ppm	1747 ppm
CCT4	1038 ppm	997 ppm
CCT5	708 ppm	815 ppm
CCT6A	433 ppm	375 ppm
CCT7	339 ppm	289 ppm
CCT8	1111 ppm	1314 ppm

Figure 4.1: Expression levels of CCT subunits in HeLa and HEK293 cell lines. Data from Pax-DB[221] shows abundance of TRiC subunits in parts per million (ppm).

Briefly, this purification strategy includes: 1) gentle cell lysis and separation of the nuclear fraction to remove DNA by ammonium sulfate precipitation, 2) ion exchange chromatography, 3) heparin affinity chromatography, 4) size exclusion chromatography, and, when necessary, 5) a second ion exchange step (Figure 4.2). Five attempts with this purification protocol were undertaken, each taking >5 days from start to finish and varying in protein yields. The most successful preps yielded only trace amounts of TRiC (~10-100 µg) that was not sufficient to carry forward for further characterisation. As CCT subunits were not observed in SDS-PAGE gels due to low abundance, the presence of TRiC was tracked by anti-CCT5 western blots (Figure 4.2: purple stars) as well as gel excised tryptic digest MS/MS.

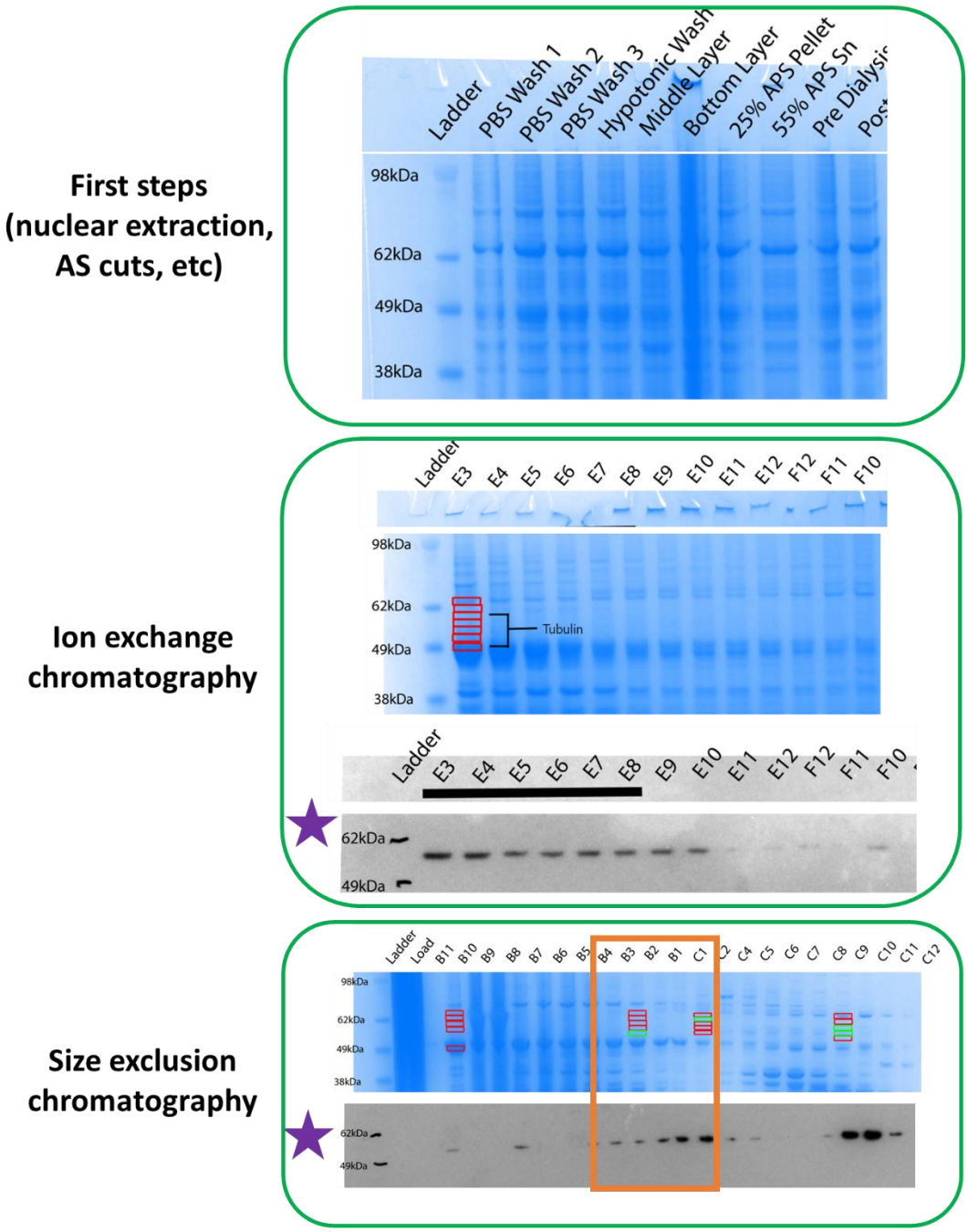


Figure 4.2: Endogenous TRiC purification from WT-HEK293 cells. Representative purification #5 of endogenous TRiC including biochemical isolation, ion exchange, and size exclusion chromatography steps. Anti-CCT5 western blots are included as indicated by purple stars. Orange box depicts gel region with purified TRiC and corresponding fractions in anti-CCT5 western blot. Black bars in SDS-PAGE gels denote fractions pooled for further steps. Gel-band sized boxes indicate results from gel excised tryptic digest mass spectrometry including CCT5 (green), other CCT subunit (purple) and other protein (red).

Attempts #1-4 did not yield enough TRiC for downstream experiments. A larger cell pellet was used during the fifth attempt (Figure 4.2) that yielded a relatively pure sample at quantities sufficient for analysis by electron microscopy (120 µg). This sample was confirmed during the size exclusion chromatography step by gel excised MS/MS and anti-CCT5 western blot (Figure 4.2: size exclusion step, orange box). This sample was analysed using negative stain EM for the presence of ring-like megadalton complexes similar to TRiC. Several ring-like particles were observed (Figure 5.3, red arrows), however high background contributed to a lack of contrast and it was difficult to determine whether these particles contained eight subunits.

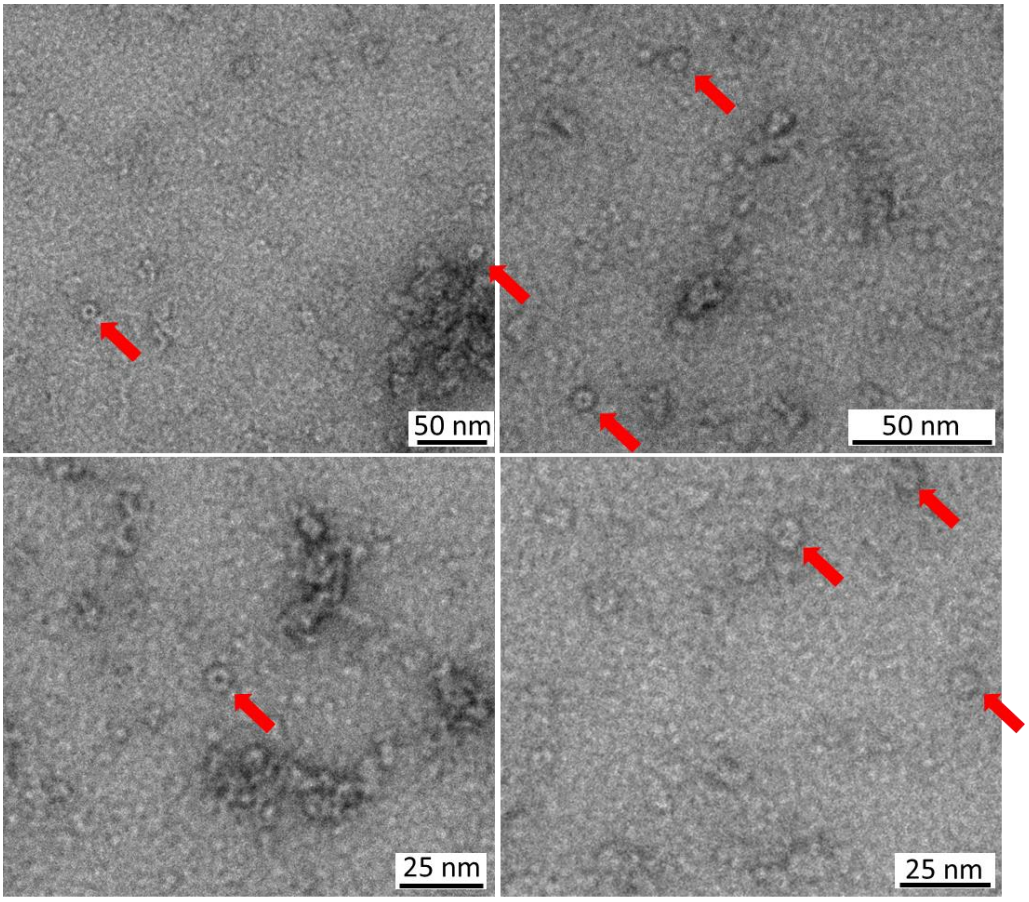


Figure 4.3: Negative stain EM of endogenously purified TRiC from WT-HEK293 cells. Electron micrographs of purified TRiC from endogenous purification #5. TRiC-like particles are indicated by red arrows.

The endogenous purification scheme from wild-type mammalian cells has the advantage of being easily employed, as the only starting requirement is a cell pellet and basic laboratory equipment. This did yield samples that possibly contained TRiC, although the time-consuming method meant that samples were kept at 4°C for >5 days, possibly reducing the stability of the complex. Additionally, a large cell mass is required and average yields between the five purifications came out to a meagre ~15 µg/L of culture volume. Taken together these data demonstrate proof of concept for isolating native human TRiC, however more advanced methods need to be employed to improve quality and quantity of the purification product for larger scale studies. Three more advanced techniques were instead pursued and are detailed in this chapter: recombinant expression using Multibac vectors (Section 4.3), CRISPR/Cas9-mediated genome editing (Sections 4.4 and 4.5), and nanobody-mediated endogenous isolation (Section 4.6).

4.3 Recombinant purification through Multibac vectors

Using recombinant protein expression to obtain TRiC has not been an attractive strategy for several reasons. The most tenable recombinant strategy until recently was to co-express eight plasmids, each containing a different CCT subunit, in BHK-21 cells infected with human adenovirus D expressing T7 RNA polymerase[219]. This was not a method pursued by many groups, and when achieved produced low yields of the TRiC complex. Another disadvantage of this technique is that expression levels of each subunit cannot be easily regulated, and as a result this could lead to conditions that are not optimal for proper assembly of the TRiC complex. Additionally, as described in the

previous Chapter 3, CCT subunits can form homo-oligomers, and unregulated recombinant expression could result in purification of a mixture of homo/hetero-oligomers that would be extremely difficult to separate from one-another. To tackle these challenges, the MultiBac vector system was investigated as a way to co-express all eight subunits in a controlled manner in insect cells.

4.3.1 MultiBac donors and acceptor vectors

MultiBac vectors have emerged relatively recently as an option for expressing protein complexes, by combining multiple genes into one large expression cassette using a baculovirus/insect cell system[222,223]. Using this technology to express all eight TRiC genes at once required several intermediate sub-cloning steps that involved donor and acceptor vectors. Donor and acceptor vectors are needed for assembly of multi-protein complexes as they diversify antibiotic resistance for selection of new gene inserts. This is achieved by combining donor DNA containing antibiotic resistance along with the new genes that are inserted into the acceptor vector.

For this work, the acceptor plasmid pFL has ampicillin resistance and contains the affinity tagged CCT subunit. Donor vectors contain one untagged CCT gene such that multiple subunits are not affinity purified at the same time, an important feature in distinguishing recombinantly produced native TRiC from isolation of several individual CCT subunits. There are two different donor plasmids pUCDM and pSPL that have chloramphenicol and spectinomycin resistance, respectively. These also contain a Multiplication Module (Figure 4.4: ‘MM’) that allows for iterative assembly of genes within a vector. Each vector contains a promotor (p10 or polh) and uses

restriction/ligation compatible sites that are removed upon successful recombination. This feature enables the pUCDM and pSPL vectors to be incorporated with multiple sets of genes, thus expanding the total number that can be incorporated into the final expression vector (hence the ‘donor’ plasmid nomenclature). Once assembled, these donor plasmids can be fused with the acceptor (pFL) vector by Cre-LoxP (for pUCDM) or Tn7 (for pSPL) recombination *in vitro*, leading to assembly of the final expression vector.

4.3.2 Strategies for the MultiBac vector

To use the MultiBac technology for expression of the eight-subunit TRiC, the donor/acceptor plasmids were designed such that four two-gene plasmids were created. To achieve flexibility in which subunits can be paired all CCT subunits were individually LIC cloned into the pFL acceptor vector, while subunits CCT1-CCT4 were individually cloned into pUCDM and CCT5-CCT8 were cloned into pSPL donor vectors (Figure 4.4A). The pUCDM plasmids are then ready for subsequent recombination using the Multiplication Module to create one four-gene donor (Figure 4.4B). Thus, the final Cre-LoxP/Tn7 recombination would include the two-gene pFL acceptor, the two-gene pSPL donor, and the four-gene pUCDM donor (Figure 4.4C). Upon selection for all three antibiotics, one eight-gene plasmid could then be carried forward for regulated co-expression of TRiC in which all subunits were regulated by the same p10 promotor (Figure 4.4D).

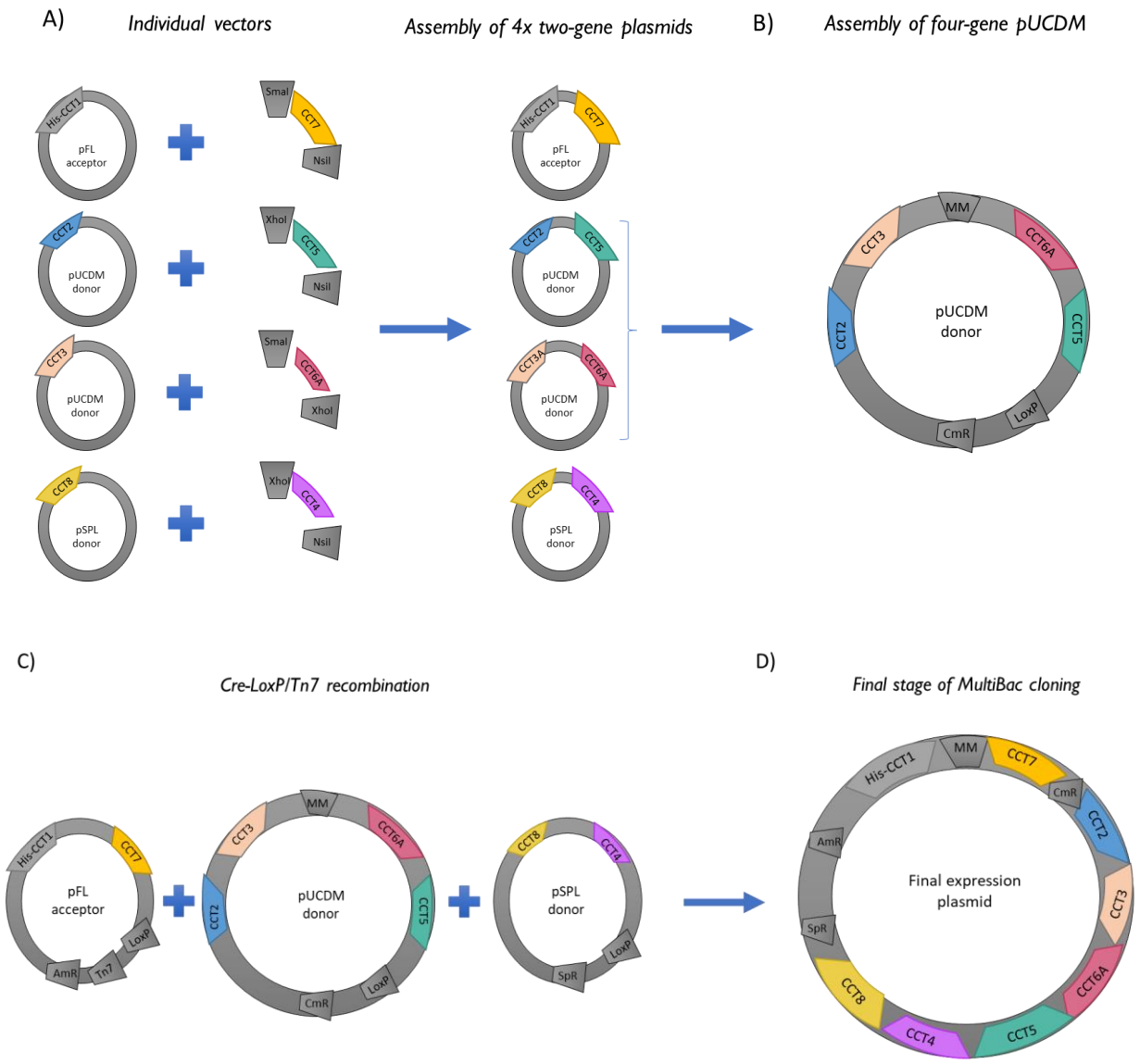


Figure 4.4: MultiBac cloning strategy for assembling eight-gene expression plasmid containing CCT1-CCT8. A) Strategy begins by created two-gene vectors assembled by restriction/ligation that produce CCT1/CCT7, CCT2/CCT5, CCT3/CCT6, and CCT8/CCT4 vectors. B) Second step involves creation of one four-gene cassette within the pUCDM donor vector. C) Final step includes recombination of the two-gene pFL acceptor, the four-gene pUCDM donor, and the two-gene pSPL donor to create D) a final expression plasmid containing all eight CCT subunits. Cassettes are labelled by CCT genes, LoxP/Tn7 sites, multiplication module (MM), and antibiotic resistance markers (Am = ampicillin, Cm = chloramphenicol, Sp = spectinomycin).

Interestingly, during transfection of the pFL vectors into Sf9 cells in preparation of this workflow, it was observed that each of all eight subunits could be solubly expressed

individually, leading directly to the results discussed in the previous results chapter (Section 3.5, Figure 3.11). CCT1 was chosen as the single tagged subunit, as it belongs to the subset that does not tend to form homo-oligomers upon overexpression, though other subunits could have been selected for the same reasons (Section 3.6).

Upon cross-referencing restriction enzyme cut sites in the multiple cloning sites of pFL, pUCDM, and pSPL, the following strategy for generating the two-gene pFL acceptors and two-gene pSPL donors was developed: CCT7 would be sub-cloned into pFL (already containing CCT1) using SmaI/NsiI restriction enzymes to create an acceptor plasmid containing His-CCT1 and CCT7. CCT5 would be restriction sub-cloned into pUCDM containing CCT2 using XhoI/NsiI restriction enzymes to create a donor plasmid containing CCT2 and CCT5. CCT6 would be sub-cloned into pUCDM containing CCT3 using SmaI/XhoI restriction enzymes to create a donor plasmid containing CCT3 and CCT6. CCT4 would be sub-cloned into pSPL containing CCT8 using XhoI/NsiI restriction enzymes to create a donor plasmid containing CCT8 and CCT4 (Figure 4.4A). Then, using the Multiplication Module either the CCT3/CCT6 cassette would be inserted into the pUCDM vector containing CCT2/CCT5 (or the other way around; Figure 4.4B). The last step would entail Cre-LoxP recombination of the CCT2/3/5/6 pUCDM cassette and Tn7 recombination of the CCT8/CCT4 pSPL cassette to create one eight-gene plasmid containing all CCT subunits (Figure 4.4C).

4.3.3 Sub-cloning to generate constructs

To begin the MultiBac cloning experiment DNA inserts encoding CCT7, CCT5, CCT6, and CCT4 were prepared through polymerase chain reaction (PCR) amplification

that produced products approximately 1.6 kb (Figure 4.5A). The following vectors were subjected to double digestion using the restriction enzymes previously mentioned: CCT1 in pFL, CCT2 in pUDCM, CCT3 in pUCDM, and CCT8 in pSPL. Inserts and vectors were combined and ligated to create four two-gene plasmids as illustrated in Figure 4.4A. These were subjected to colony PCR to select for single clones containing successful insertions of the desired CCT gene into the vector backbone (Figure 4.5B: green boxes). Positive colonies were seen for all combinations except CCT3 and CCT6. While the CCT3/CCT6 restriction cloning was repeated the successful insertion of each clone was checked by digesting the recombined vector with the restriction enzymes used in the cloning step. If successful, each insert should ‘drop out’ and be seen on an agarose gel. This method provided further evidence that CCT1/CCT7, CCT2/CCT5, and CCT8/CCT4 insertions were successful, as each ~1.6kb product was observed (Figure 4.5C: purple labels). The CCT3/CCT6 insertion was repeated after the CCT6 insertion was gel-purified to increase chances of successful insertion. This led to successful insertion of CCT6 into the pUCDM vector containing CCT3 upon colony PCR (Figure 4.5D: left) and subsequent digestion check (Figure 4.5D: right). Each plasmid was then sequenced to provide certainty that inserts were properly integrated into their respective vectors.

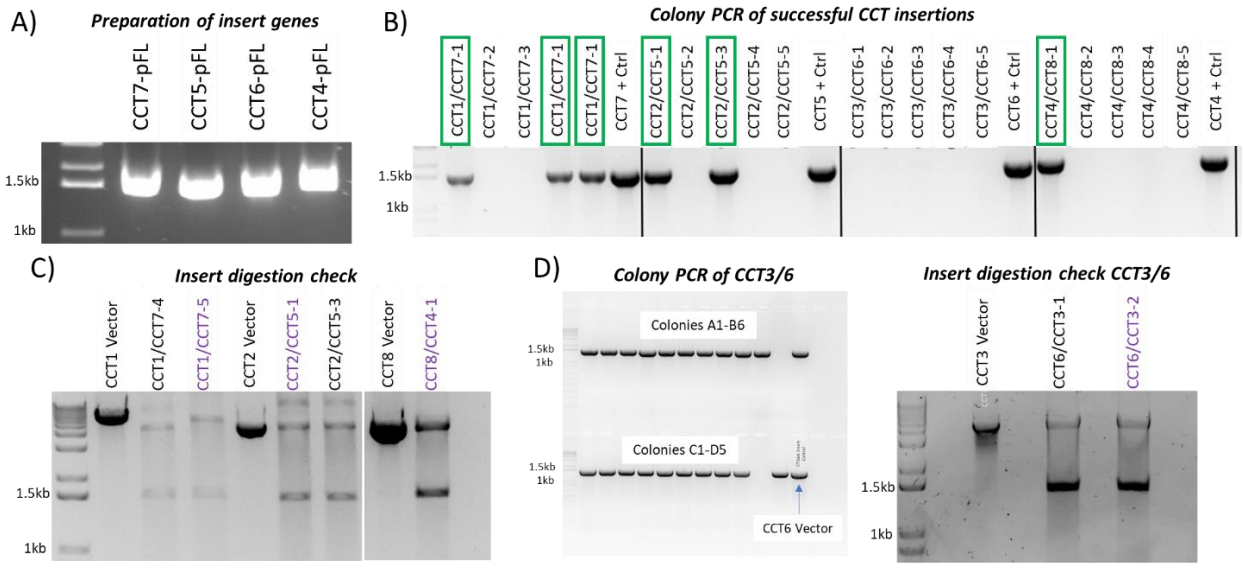


Figure 4.5: Initial cloning phase of TRiC MultiBac vector assembly. A) Preparation of insert fragments containing single genes to be restriction cloned into respective vectors. B) Colony PCR showing colonies with successful insertion (green boxes) including CCT1/CCT7, CCT2/CCT5, and CCT8/CCT4 as compared to vector-only positive controls. CCT3/CCT6 did not show successful insertion. C) Digestion check of inserts to check for dropout of gene. Clones labelled in purple were taken forward for additional cloning steps. D) Re-attempt of colony PCR and digestion check for CCT3/CCT6 insertion.

The next phase of cloning involved combining the two pUCDM vectors containing CCT2/CCT5 and CCT3/CCT6, as detailed in Figure 4.4B, to generate the four-gene vector. Due to the Multiplication Module this step could be done either way; in this case the attempt was made to insert CCT3/CCT6 into the vector containing CCT2/CCT5 based on availability of restriction enzymes. The CCT2/CCT5 was linearised using restriction enzymes BstZ17I and SpeI to produce a product of ~6.3 kb (Figure 4.6A, lane 3), while the CCT3/CCT6 cassette was separated from the vector backbone by addition of AvrII and PmeI to produce a fragment ~1.9 kb (remaining vector backbone 4.4 kb; Figure 4.6A, lane 5). The main band corresponding to the CCT2/CCT5 vector and the lower band corresponding to the CCT3/CCT6 insert were gel excised, ligated, and transformed into PirA cells and incubated to produce colonies with four-gene insert.

Only a few colonies grew, and each were subjected to colony PCR using primers for CCT6, however no CCT6 was detected (Figure 4.6B, lanes 2-7). The lack of CCT6 present in the colonies demonstrates that the ligation attempt was unsuccessful and the expected four-gene pUCDM plasmid was not produced.

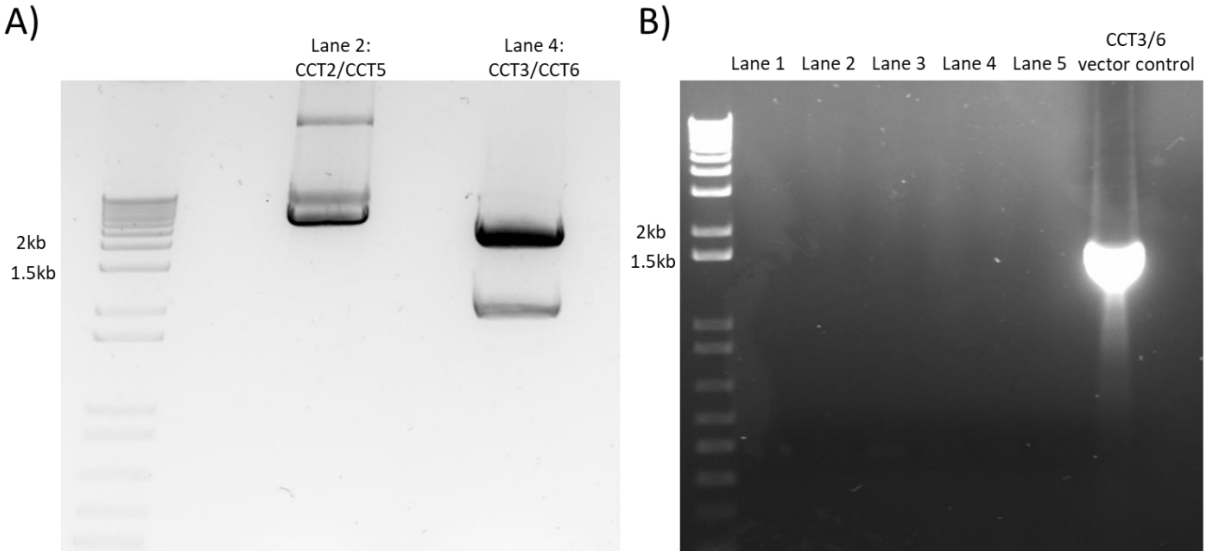


Figure 4.6: Recombination attempt of CCT2/CCT5 and CCT3/CCT6 vectors to form four-gene CCT2/3/5/6 pUCDM plasmid. A) Digestion of CCT2/CCT5 vector backbone and CCT3/CCT6 insert (bottom band of lane 4). B) Colony PCR showing lack of CCT3/6A positive colonies in lanes 1-5.

At this stage more work could have been done to produce a four-gene pUCDM plasmid for recombination with the pFL and pSPL vectors, to construct the final eight-gene plasmid. The four two-gene plasmids incorporated the correct target insert with high confidence, and thus it was expected that the CCT3/CCT6 insertion into CCT2/CCT5 be successful. However, this step could have failed due to poor restriction enzyme cutting efficiency leading to incomplete digest of the insertion fragment. Additionally, the restriction enzymes available to insert of CCT2/CCT5 into the CCT3/CCT6 pUCDM vector could have also been attempted. However, these efforts

were paused indefinitely during the course of the DPhil. The primary reason is that other methods to isolate TRiC were progressing at a more rapid pace, and these will be discussed in the remainder of this chapter. Additionally, the recombinant method of expression and purification was at the time disfavoured compared to endogenous methods, as there were still questions as to whether homo-oligomeric complexes would be produced, and concerns about whether they could be differentiated from native TRiC at the purification level. It was later shown in the literature, after these efforts were halted in this study, that the Multibac technique was successfully employed to purify native TRiC, and no evidence of homo-oligomeric complexes were reported[71]. This was done by a much simpler method than the one undertaken here: in this newer study four two-gene plasmids were generated and co-infected, bypassing the complex assembly of an eight-subunit plasmid. In summary, the recombinant method provided promise for isolation of recombinant TRiC but gave way in this DPhil study to more desirable methods that had come to fruition.

4.4 CRISPR/Cas9 mediated genome editing of CCT3

CRISPR/Cas9 mediated genome editing has emerged over the last decade as a powerful tool applicable to a wide range of biological applications[224,225]. CRISPR, short for clustered regularly interspaced short palindromic repeats, was first identified as a mechanism used by microbes to defend against foreign nucleic acids. This process is mediated by CRISPR guide RNA (gRNA) that provides a template for the enzyme CRISPR associated protein 9 (Cas9; or other Cas enzymes) to perform a specific double-stranded DNA cleavage. This occurs at the recognised site of the foreign genome and thus provides immunity from the foreign pathogen.

The CRISPR/Cas9 system has since been converted to a laboratory format broadly useful for genome editing in biological research. One such application, relevant to this DPhil research, includes knock-in of affinity tags into specific genomic loci, providing a method for direct isolation of endogenous complexes[174,226]. This workflow was shown previously to allow for purification of large complexes in sufficient quantities for downstream characterisation by mass spectrometry, structural methods, and enzymatic assays[226].

The CRISPR/Cas9 knock-in protocol can be broken down into the following steps: 1) identify target regions within a gene locus to incorporate tag insertion, 2) assemble DNA fragments necessary for genome editing including homology arms, desired insert oligo, and gRNA (Figure 4.7), 3) perform transfection of all components, and 4) select successfully edited cell lines. The following section is split into two projects that together detail a full tagging CRISPR/Cas9 tagging pipeline. The first project (Sections 4.4.1-4.4.5) describes a strategy to incorporate an affinity tag into an internal loop within the CCT3 subunit of TRiC, which was unsuccessful at the selection stage to identify positively edited cells containing the FLAG tag of interest. However, this workflow was the same for the insertion of an affinity tag at the C-terminus of CCT5 completed through a collaboration that did prove successful and is described as a second project (Section 4.5) from the point of transfection.

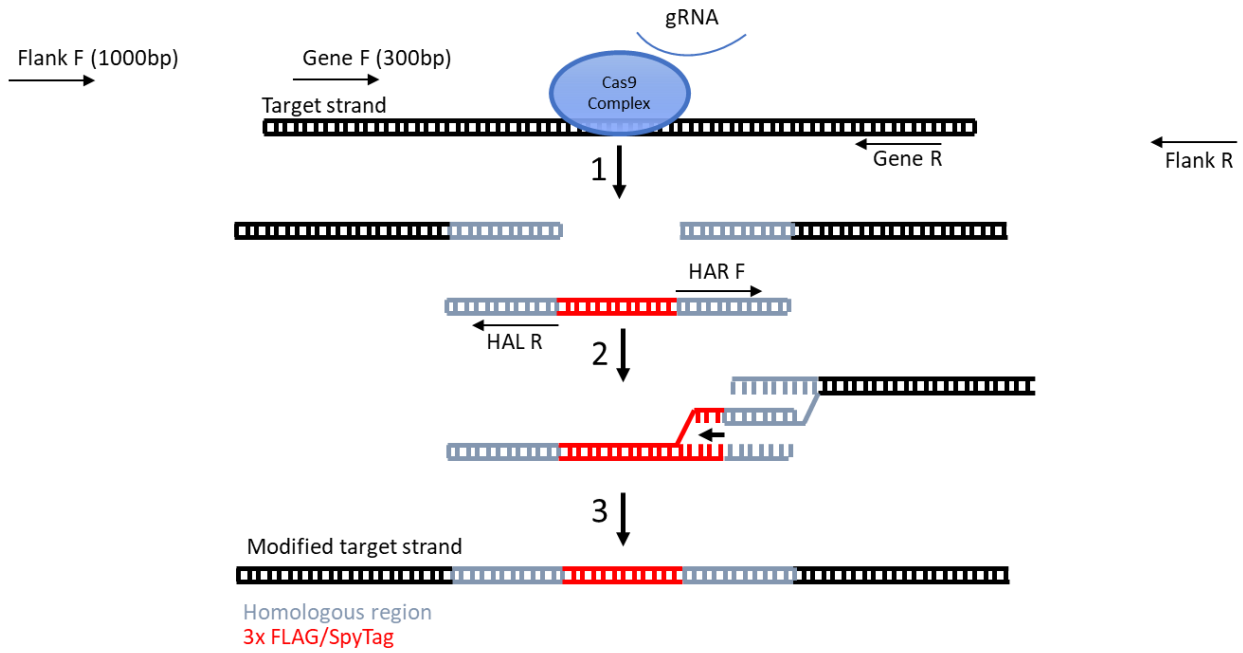


Figure 4.7: CRISPR/Cas9 mediated genome editing workflow to incorporate knock-in affinity tag at target site. Knock-in tagging entails 1) genotyping forward/reverse primers (Gene) to sequence the target loci and flanking forward/reverse primers (Flank) to PCR amplify template region, 2) homology arm left (HAL) reverse and homology arm right (HAR) forward primers to create 3) homology-directed repair template containing desired insert (red) and complimentary regions to the target site (blue).

4.4.1 Design of CCT3 tagging strategy

The first step in developing the tagging strategy for CCT3 (and CCT5) was to identify potential tag locations within the genomic loci of *CCT3*. At the time this workflow was designed, there was a lack of structural information for all human CCT subunits. However, based on the structures of homologues including Mm-Cpn[21] and the archaeal thermosome[19], it has been shown that the N- and C-termini of each subunit point inwards towards the folding chamber and stack with the stem-loop of the adjacent subunit (Figure 1.7). Due to the apparent importance of the termini in chaperonin function these regions were avoided in favour of surface-exposed loop regions. Homology models of CCT3 from other organisms including yeast CCT3[138] and the

thermosome α subunit[227] helped deduce four potential loop regions tenable for tagging, however the exact location of these residues in the human sequence could not be confirmed prior to the start of the experiment. Therefore, residues from these loops of CCT3 were selected as sites for tag insertion C-terminally from the given target residue. Each originate from different regions of the CCT3 polypeptide, including one equatorial domain loop (Gly485), one intermediate domain loop (Asp158), and two apical domain loops (Arg339 and Met229; Figure 4.8). Multiple regions were selected for cloning to increase the chance of successful insertion, as these sites were designed based on guessing where the loop regions are in the human sequence. Additionally, the N-terminus was also targeted as evidence existed from the previous chapter and from the literature[39] that all subunits could be individually purified using N- and C-terminal His6 tags (Figure 3.11).

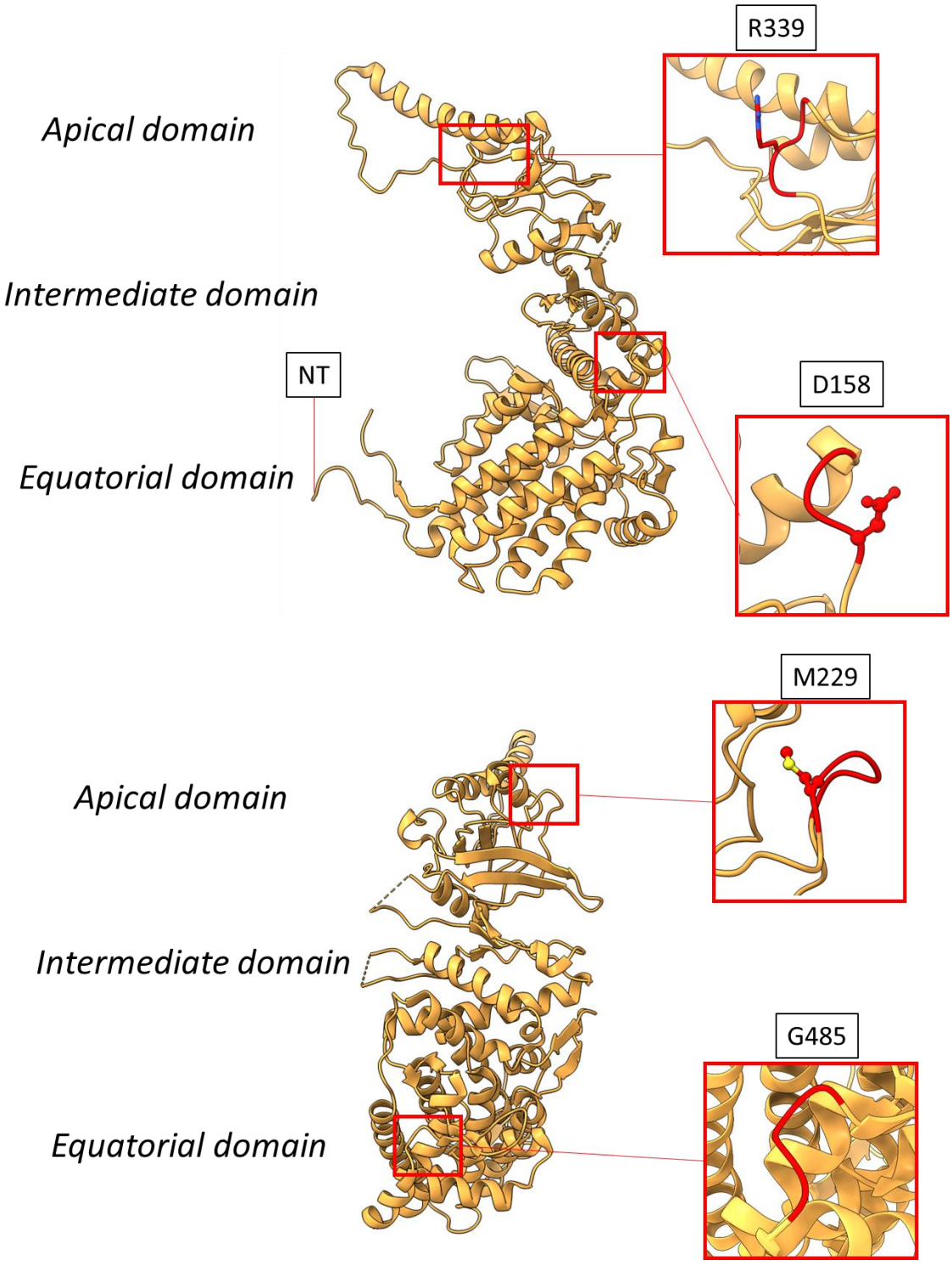


Figure 4.8: Target regions for tag insertions in CCT3. Orthogonal side (top) and end-on (bottom) views of CCT3 (PDB: 6KS6)[27] including insets showing surface exposed loops where affinity tag could be inserted. These sites are referred to in subsequent text and figures as “N-terminus”, “intermediate D158”, “apical M229”, “apical R339” and “equatorial G485”.

4.4.2 Assembly of CCT3 homology-directed repair templates

For the knock-in experiment, three components were required: a DNA oligo containing the desired insert, homology arms that overlap the insert to promote successful incorporation into the target genome, and a Cas9 vector containing gRNA to direct Cas9 to the target cleavage site (Figure 4.7: step 1). To provide flexibility in affinity purification, two different tags were inserted in tandem. The first was a 3xFLAG tag, which is commonly used for expression and purification of human proteins from lysate[228]. The second is a SpyTag, a short peptide with high affinity for its counterpart SpyCatcher that was more recently developed as a shorter and higher affinity alternative to the FLAG and other affinity tags[173]. The DNA oligo synthesized for this workflow consisted of a single-stranded DNA fragment containing a glycine-serine linker (3x GGGS), 3xFLAG tag, enterokinase cleavage site, SpyTag, and ended with another glycine-serine linker. Three different guide RNAs (gRNA4, gRNA7, and gRNA9) were designed using the CCtop online prediction interface[229]. Multiple guide RNAs were used to increase the likelihood of successful insertion. Both a 'top' and 'bottom' guide were ordered that were complimentary to each other as these are later annealed and ligated into the eSpCas9(1.1) vector. This left the homology arms to be created and this step was the primary bottleneck for attempting the genome editing.

To begin assembly of the homology arms (Figure 4.7, step 2), genomic DNA was extracted from the desired cell line; here both HEK293T adherent and K562 suspension cell lines were used. Both were chosen as HEK293T cells were shown to successfully incorporate the same tag in a CCT5 C-term tagged line, while K562 cells were already suspension adapted and scaling up the edited cell lines would be expedited for large

scale purification if successful. The target sites were sequenced using ‘genotyping’ primers flanking each site ~300 bp from the target cut site, to ensure these regions did not contain any mutations in the host line used for the genome editing experiment. Another set of ‘flanking’ primers were designed for each target ~1 kb away from the target site, to assemble the homology-directed repair (HDR) template (Figure 4.7, HAL and HAR). These flanking primers were used to PCR amplify the ~2 kb desired target region from genomic DNA as templates for HDR (Figure 4.9). This process proved to be the most difficult and became the point at which many insertion sites designated at the beginning of this study were no longer pursued.

In brief, the ‘N-terminal’ (template was ~1.5 kb instead of 2 kb due to shorter upstream primer) and ‘R339 apical’ insert regions (Figure 4.8) were successfully amplified (Figure 4.9A), while the remaining sites were unable to be amplified (Figure 4.9B). These likely could have been amplified upon further optimisation of temperature, DMSO concentration, and primer chosen (genotyping primers could also be used). However, this was deemed unnecessary at the time of study, as the ‘N-terminal’ and ‘R339 apical’ sites were sufficient for the scope of the experiment. Both target templates were sequenced and confirmed, and the ‘R339 apical’ site was carried forward to the next step, as an internal loop region was preferable to termini for targeting.

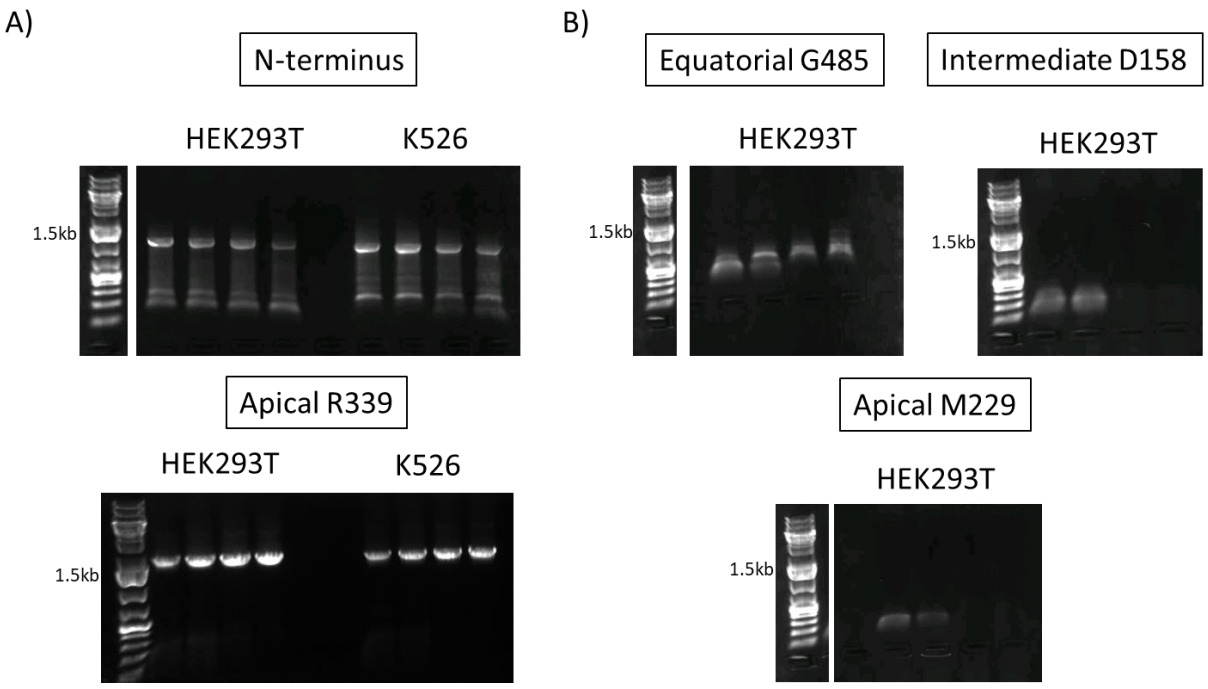


Figure 4.9: PCR amplification of CCT3 target regions. A) Successful target PCR amplifications in HEK293T and K562 genomes of N-terminal (top) and R339 apical (bottom) target loops. B) Unsuccessful amplifications of remaining target sites using HEK293T genomic DNA. Replicates of four represent increasing temperature of PCR reaction from left to right.

To create left and right homology arms of the template region, homology arm left (HAL) and homology arm right (HAR) primers were ordered that go in the reverse direction from the target site to amplify the HAL, while the HAR primer is in the forward direction from the end of the target site (Figure 4.7, step 2). These also have a complementary region to the single stranded 3xFLAG SpyTag insert DNA, allowing these amplified homology arms to act as primers in the next round of PCR. The HAL and HAR fragments were successfully amplified using these primers (Figure 4.10A), meaning the full assembly of the HDR template could now be generated by combining the HAL, HAR, and single-stranded DNA (ssDNA) insert components.

The first attempts at assembling the HDR template did not work, likely because the flanking primers spanned a longer distance from the target site and thus were more prone to amplification errors, or due to UV-induced DNA damage as DNA was imaged using UV prior to excising gel bands. To address these issues genotyping primers were used instead and gels were excised prior to imaging, causing a subsequent lack of bands in the images below (Figure 4.10B). The 2 kb template, HAL and HAR, were easily amplified using genotyping primers instead of flanking primers and the gel purified products were used to assemble the complete HDR template for the CCT3 apical R339 site (Figure 4.10C). This step concludes the assembly of the HDR template.

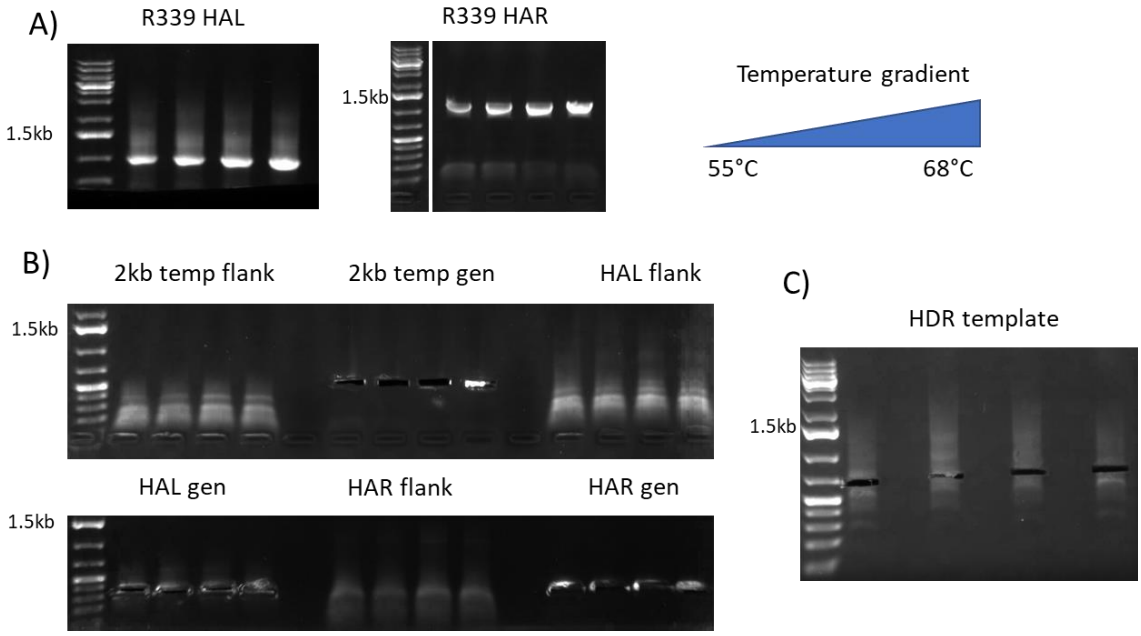


Figure 4.10: Assembly of homology-directed repair template for CCT3 R339 target site. A) PCR amplification of homology arm left (HAL) and homology arm right (HAR) with increasing temperature conditions from 55°C to 68°C. B) PCR amplification of 2 kb template (temp), HAL, and HAR using flanking (flank) and genotyping (gen) primers. Gel bands were excised prior to imaging to reduce UV radiation damage. C) Full HDR template assembled using HAL, HAR, and ssDNA insert. Lanes show increasing temperature gradient during annealing step of PCR cycle.

4.4.3 Transfection of CRISPR/Cas9 components for CCT3 genome editing

Upon assembly of the HDR template the eSpCas9 vector was digested using BbsI endonuclease, and the three guide RNAs designed to target the CCT3 R339 site were annealed and ligated into the eSpCas9 vector. These were transfected into DH5α cells and miniprepped to isolate vector DNA for sequencing. Each guide insertion was confirmed and thus all components were generated to perform the genome editing experiment.

The HDR template and eSpCas9 vector containing gRNA were transfected by electroporation into K562 cells and were left undisturbed for several days. The eSpCas9(v1.1) vector was designed to be resistant to ouabain, a potent inhibitor of the essential Na^+/K^+ ATPase, by also transfecting gRNA for the ATP1A1 gene that encodes the Na+/K+ ATPase. Upon ouabain selection, only cells that have successfully incorporated the eSpyCas9(v1.1) vector are able to survive. This method was used to select for positively edited K562 cells with an insert in the CCT3 apical R339 locus. Only two of the three cell populations (those treated with gRNA7 and gRNA9) survived and these two populations were checked for the presence of FLAG tag by anti-FLAG western blotting by Dr. Dale Tranter. Unfortunately, no FLAG tag was observed for either population, and this workflow was subsequently stopped (Figure 4.11).

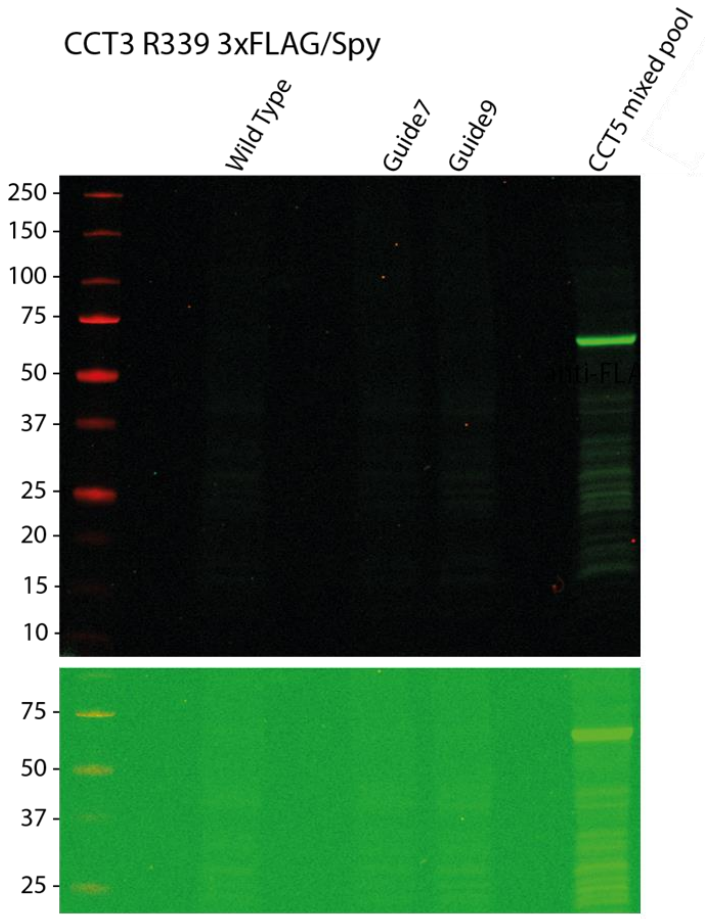


Figure 4.11: Assessment of CCT3-affinity tag insertion. Top: Anti-FLAG western blot run with wild-type HEK293T parental cell line, transfected cell populations with Guides gRNA7 and gRNA9, and a cell line expressing FLAG-tagged CCT5 as a positive control. No band was observed for CCT3 as compared to the CCT5 positive control cell line. Bottom: anti-FLAG western blot as above but visualised with higher contrast. Image provided courtesy of Dr. Dale Tranter.

Upon reflection of the data, this insert might not have been tolerated as the target site (R339 at apical domain) is part of the release-loop of substrate motif that plays an important role in substrate folding (Figure 1.6). Additionally, it was observed that another group had previously inserted a tag between Pro374 and Lys375 of yeast CCT3[220]. This was achieved by replacing the genetic locus of *CCT3* with a copy containing the tag insertion, resulting in successful isolation and purification of yeast TRiC. Therefore, this loop should have been targeted for insertion as well. Fortunately,

this workflow was successfully undertaken for a CCT5 C-terminally tagged cell line with the same insert and protocol and will be described in the next section 4.5.

4.5 Selection and validation of CCT5 C-terminal affinity tag knock-in, as a viable strategy for isolating native TRiC

In a different workflow, the N- and C-termini of CCT5 were chosen for tagging as these regions were previously purified with N- (this thesis) and C-terminal affinity tags[39]. Additionally, a loop near Ser154 that was expected to be surface exposed was targeted for tagging in case the buried position of the termini could not tolerate affinity tags. This CCT5 cell line tagging approach was carried out in the same manner as described for CCT3. Only the C-terminus of CCT5 was successfully tagged, and this cell line was carried forward for selection after transfection. The HDR assembly, transfection, and early selection of the C-terminal CCT5-FLAG tagged cell line was carried out by Dr. Dale Tranter in the lab of Dr. Ville Paavilainen at the University of Helsinki, while the initial design of the CCT5 tagging strategy and subsequent characterisation of the CCT5-FLAG line is described below.

4.5.1 Validation of C-terminal affinity tagged CCT5 cell line

The CCT5 C-terminal 3xFLAG/SpyTag insert was transfected along with the eSpCas9(v1.1) vector in HEK293T adherent cells. The mixed pool of cells was checked for the presence of FLAG tag by anti-FLAG western blot both before and after ouabain selection (Figure 4.12A). A strong band was present for the CCT5-FLAG tagged cell pool before ouabain selection as compared to the parental wild-type HEK293T cell line. This band was further enriched after ouabain selection and provided confidence that a

significant portion of the mixed pool of cells contained successfully incorporated inserts at the C-terminus of CCT5, however this band appears to be higher either due to overloading of sample or slicing of gels. A small cell pellet was harvested from these adherent cells and subjected to purification using anti-FLAG affinity gel matrix. The elution fractions contained both CCT5 and FLAG tag, as detected by anti-CCT5 and anti-FLAG western blots, suggesting the genome edited cell line could be useful for purification of endogenous TRiC (Figure 4.12B).

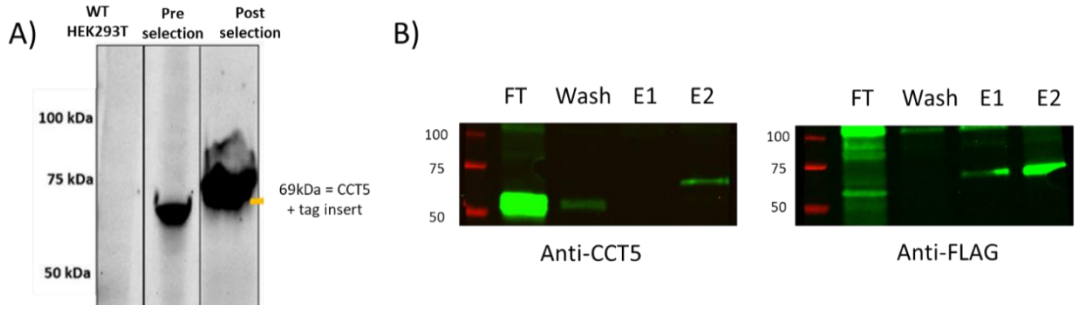


Figure 4.12: Identification of CCT5 C-term 3xFLAG/SpyTag knock in. A) Anti-FLAG western blot of parental HEK293T cell line including CCT5-FLAG edited cell lines pre and post ouabain selection. Image provided courtesy of Dr. Dale Tranter. B) Anti-CCT5 (left) and anti-FLAG (right) western blots of samples from small scale FLAG affinity purification including flow-through (FT), wash fraction, and elution fractions 1 and 2 (E1 and E2).

4.5.2 Small scale purification of CCT5-FLAG cell line successfully isolates native TRiC

To check for evidence of purification of native TRiC and not just FLAG-tagged CCT5, lysates from the CCT5 mixed pool and FLAG affinity elution fractions were run in western blots with eight different subunit-specific antibodies to verify the presence of all eight CCT subunits (Figure 4.13A). CCT subunits were detected in the elution fraction for all samples except for CCT6 and CCT8, likely due to poor quality of anti-CCT6 and

anti-CCT8 primary antibodies. This data demonstrates that most, if not all, CCT subunits are also present in the elution in addition to FLAG-tagged CCT5. A higher band of an additional ~9 kB was observed in the anti-CCT5 and anti-FLAG western blots, suggesting two copies of CCT5 existed simultaneously including a wild-type copy and a copy containing successful incorporation of the 3x FLAG/Spy tag .

To assess whether a TRiC-like ring structure could be observed from these small-scale preparations, the elution sample was concentrated and applied to negative stain EM grids (Figure 4.13B). Indeed, TRiC-like particles could be seen but a high degree of background made it difficult to identify the number of subunits present. Nevertheless, a small data collection was carried out at the Helsinki Institute of Life Science HiLIFE on an F20 microscope. A data set with 150 negative stain images of the elution sample was processed, containing 15,000 particles, and subsequent 2D classification showed that the TRiC-like rings indeed contained eight subunits.

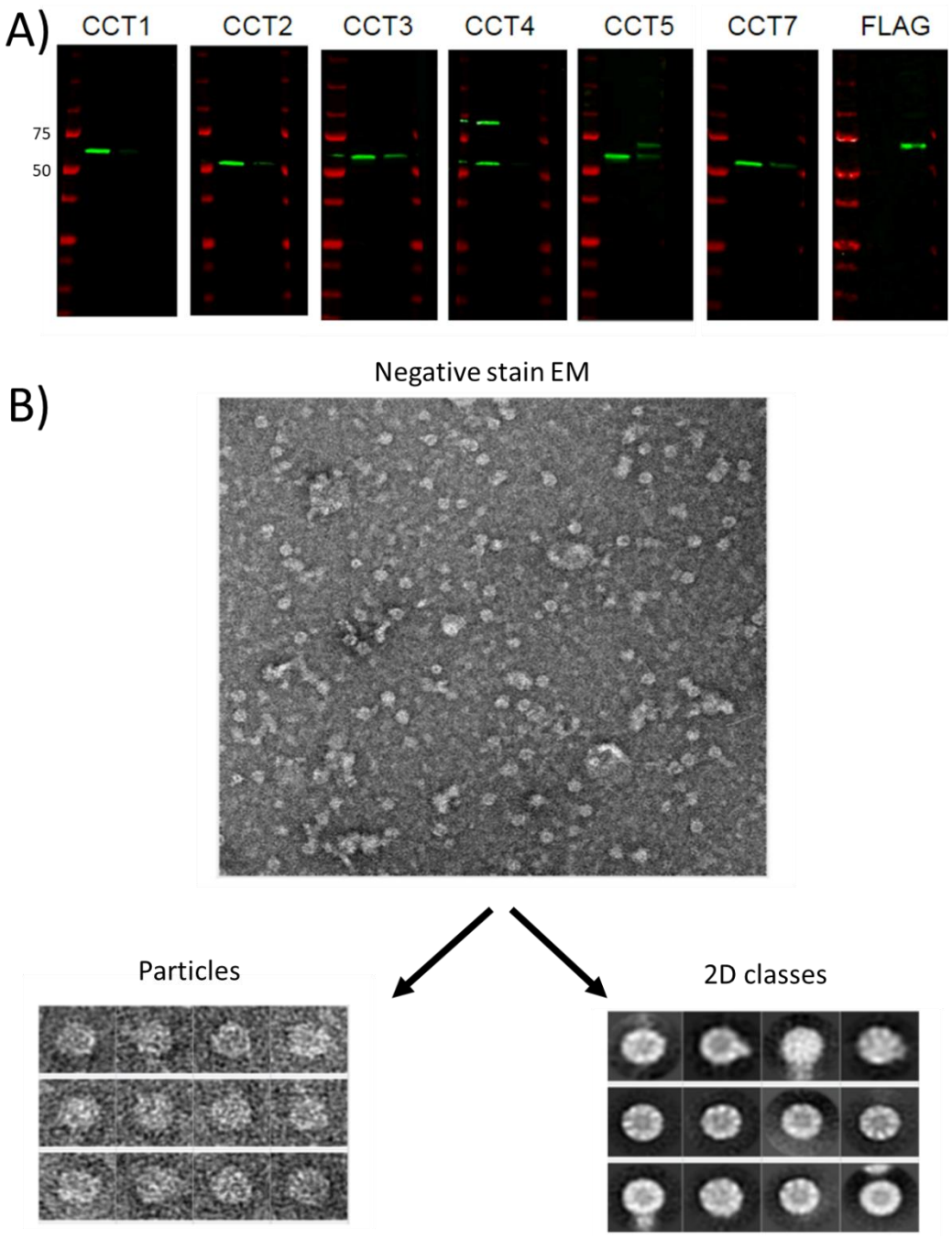


Figure 4.13: Small scale purification of CCT5 C-term 3xFLAG/SpyTag knock in. A) Western blots of lysate (left lane after marker) and elution (second lane after marker) using antibodies against CCT subunits. Higher band in CCT5 and FLAG blot show presence of CCT5 with 9 kB insert. B) Negative stain EM of purified TRiC including representative micrograph (top), grouped particles for classification (bottom left) and 2D classifications (bottom right).

4.5.3 Large scale purification from CCT5-FLAG line further confirms isolation of native TRiC

The negative stain EM data collection provided evidence that TRiC-like complexes were purified, and western blots suggested these complexes contain multiple CCT subunits. To improve the utility of this cell line for purification of endogenous TRiC, the mixed pool of CCT5-FLAG edited cells was sorted using flow cytometry into monoclonal cell lines by Dr. Dale Tranter. Five monoclonal cell lines were isolated for further study and subjected to anti-FLAG western blotting (Figure 4.14A). Three of the five monoclonal lines were positive for containing a FLAG tag (Figure 4.14A: lanes 1.E4, 3.B5, and 4.D7). Each contained approximately 100x the amount of FLAG-tagged CCT5, as compared to the mixed pool from which they were sorted. The cell line 3.B5 (Figure 4.14A: green box) was selected and grown at a larger scale (ten million cells total) to observe whether native TRiC could be isolated from the monoclonal cell line.

The scaled-up pellet was lysed and passed into anti-FLAG affinity resin, and eluted with competing 3xFLAG peptide. The elution fractions from this sample could be visualised on SDS-PAGE and bands were gel-excised and submitted for tryptic digest LC-MS/MS (Figure 4.14B: top). Gel bands were excised, and unique TRiC subunit peptides were identified by mass spectrometry in the double-banded region between the 50 kDa and 75 kDa markers, an expected observation based on the CCT subunit molecular weights ranging ~57-60 kDa. Western blots using antibodies against CCT1, CCT2, CCT3, CCT5, and CCT6 also confirmed these subunits were present in the

sample (Figure 4.13B: bottom; antibodies against CCT4, CCT7, and CCT8 were unavailable at the time of experiment).

Interestingly, there are two bands seen in the anti-CCT5 western blot: one at 60 kDa and one at 70 kDa (Figure 4.13B: bottom). These bands appear to be in equal molar ratio, suggesting the complexes isolated from this sample contain both a wild-type and FLAG-tagged copy of CCT5. This would suggest that not all the genomic loci of *CCT5* were edited, and that native untagged CCT5 was also present in the sample assembling TRiC subunits along with tagged copies of CCT5 similar to what was observed in the mixed pool purification.

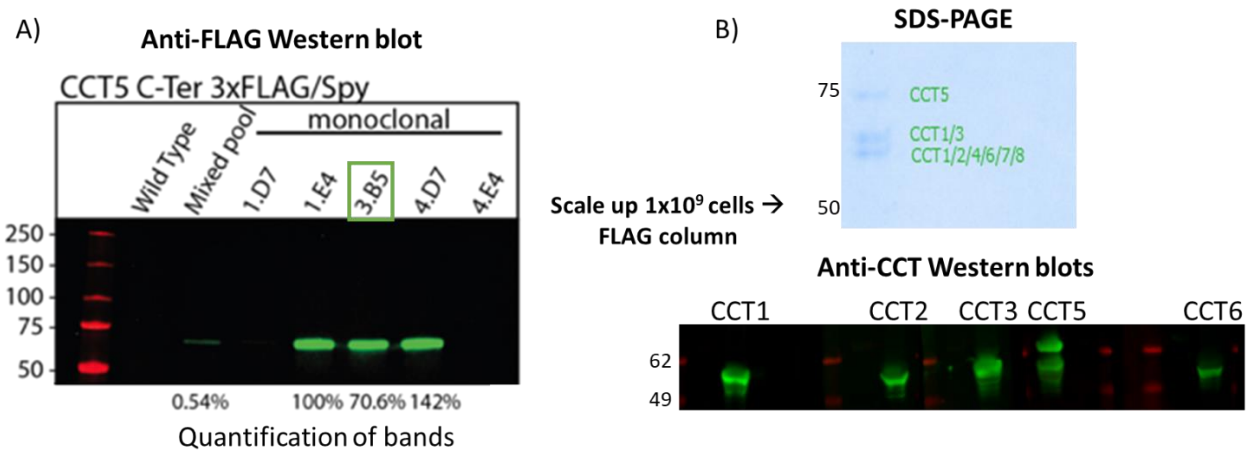
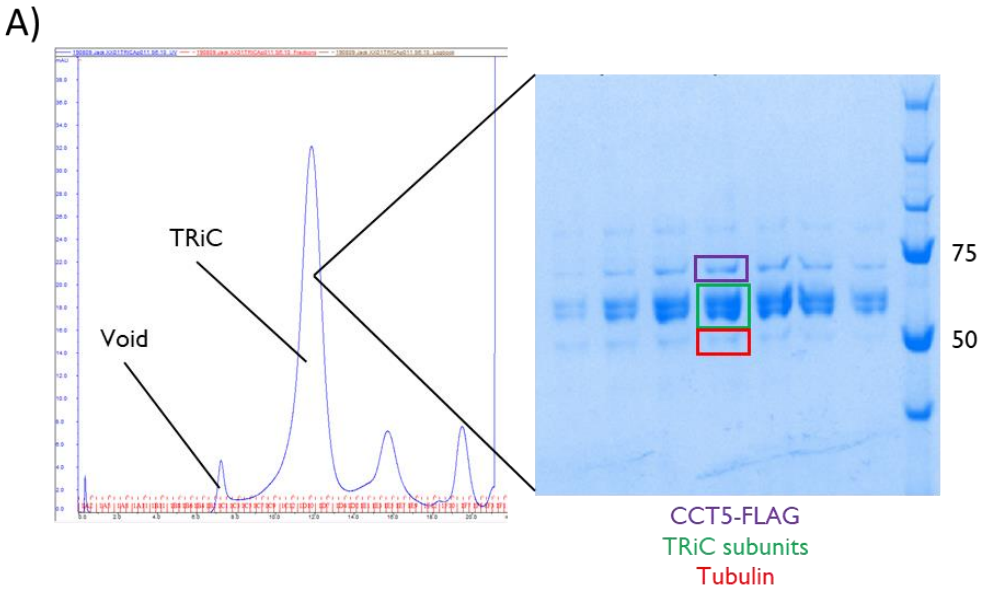


Figure 4.14: Selection and large-scale purification of monoclonal CCT5 C-term edited cell lines. A) Anti-FLAG western blot of wild type HEK293T, mixed pool, and five monoclonal cell lines. Quantification of bands included below. Green box indicates 3.B5 monoclonal cell line selected for further study. B) SDS-PAGE including tryptic digest MS/MS results (green text) and anti-CCT western blots of mixed pool 3.B5 together showing presence of all CCT subunits.

Additionally, to expand the amount of material that could be generated for downstream applications, a sample with larger cell mass was obtained by adapting the adherent monoclonal cell line into suspension so that more cells could be generated for harvest and subsequent purification. This process involved weaning the cell culture from

adherent to suspension media and was successful in generating larger amounts of protein. Through this workflow enough sample was obtained during purification to be applied on a size exclusion chromatography column (Figure 4.15A), in an attempt to separate out native TRiC from low molecular weight contaminants that were hindering cryo-EM studies (see section 5.2.3). The TRiC complex eluted around the size of a megadalton complex (slightly larger than 670 kDa molecular weight standard). Interestingly, tubulin was found to co-elute with the TRiC complex, as observed in the SDS-PAGE and verified by gel excised LC-MS/MS. The elution from size exclusion was applied on negative stain grids (Figure 4.15B). The particles seen on the negative stain micrographs contained less background than the elution from the mixed pool, showing eight-membered ring particles with much higher contrast.



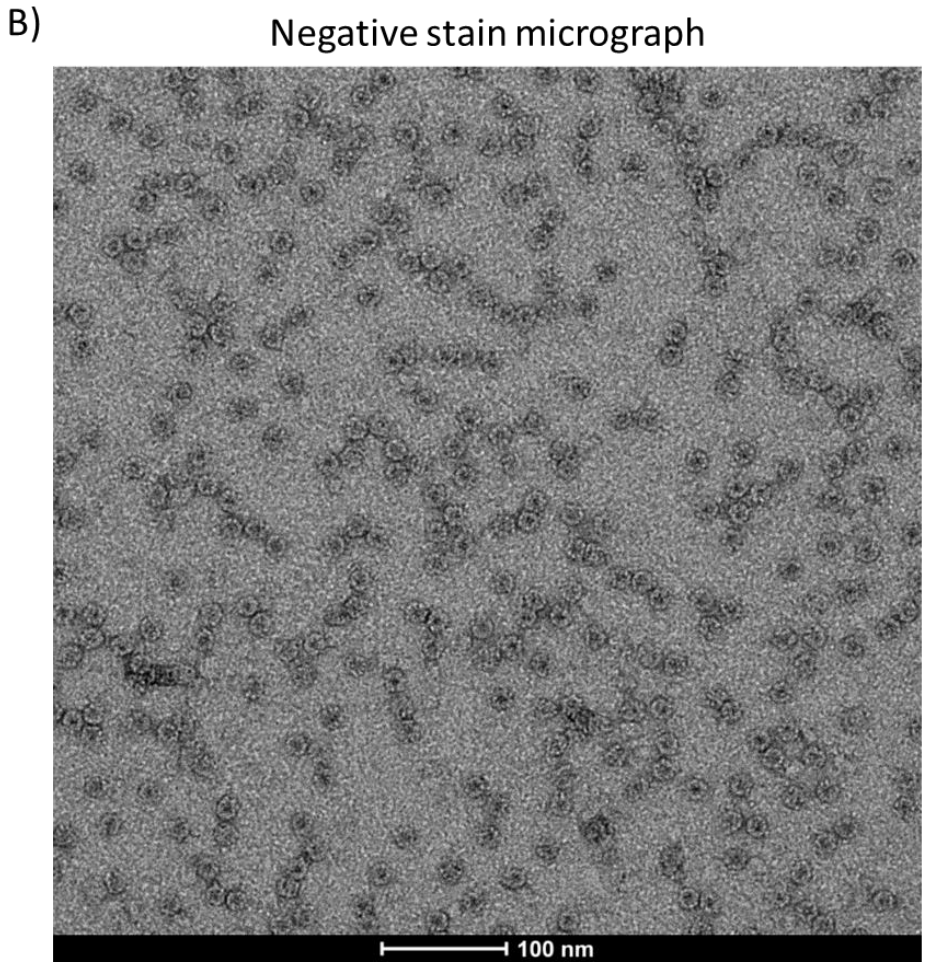


Figure 4.15: Large-scale purification of TRiC from monoclonal CCT5 C-term edited cell lines. A) Size exclusion profile of large-scale TRiC purification including tryptic digest MS/MS showing CCT5-FLAG band (purple), double-banded CCT subunit pattern (green), and tubulin (red) co-eluting with TRiC. B) Negative stain micrograph showing monodisperse TRiC particles.

4.5.4 Final purification procedure for TRiC

The final purification protocol of large-scale isolation from CCT5-FLAG edited cell lines (Section 2.3), which can be used routinely for structural studies, includes gentle cell lysis with detergent followed by an anti-FLAG affinity gel matrix in which TRiC was eluted with 3x FLAG peptide (Figure 4.16A). Elution fractions were concentrated and run on a Superose 6 size exclusion chromatography column and the first peak representing high molecular weight species was collected (Figure 4.16B and C).

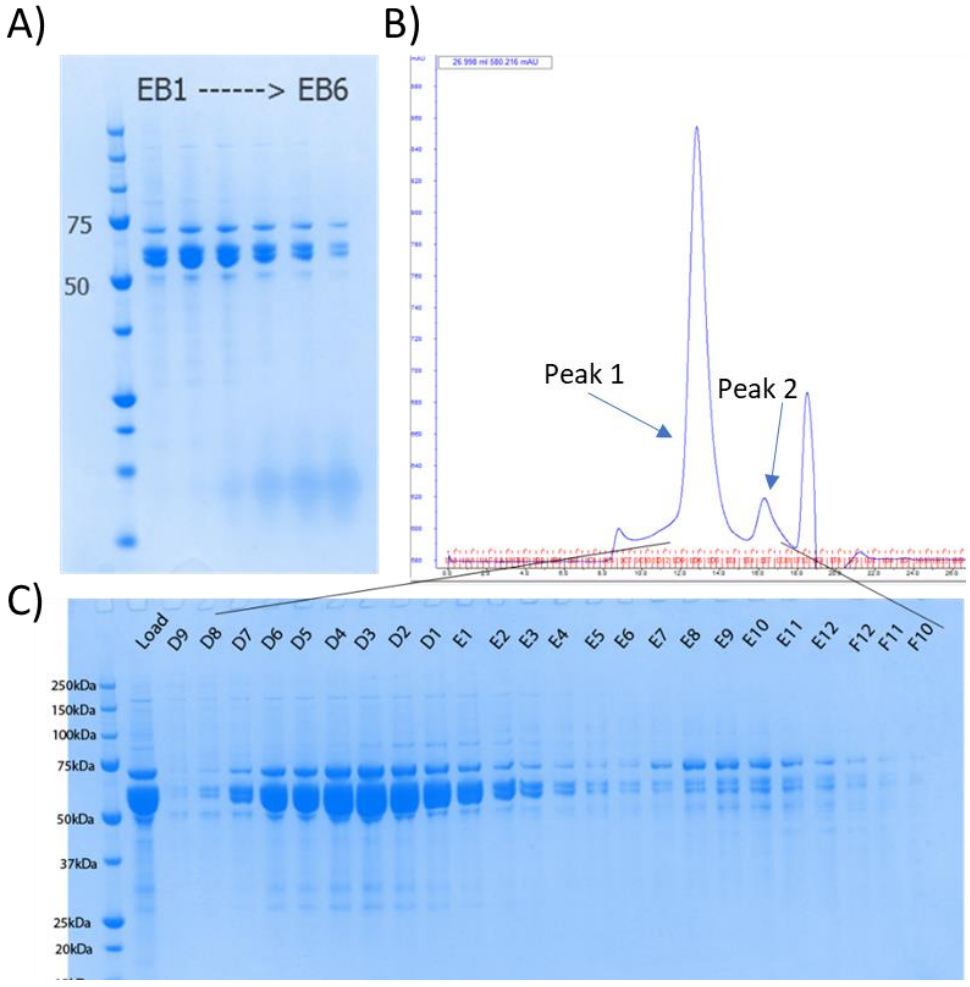


Figure 4.16: Standard purification of large scale TRiC. A) SDS-PAGE analysis of elution fractions (EB1-EB6) from an Anti-FLAG column. B) Profile of Superose 6 size exclusion chromatography. C) SDS-PAGE of SEC fractions corresponding to two peaks containing TRiC subunits.

Interestingly, the TRiC banding pattern is also seen outside the main peak at low molecular weight, similar to what was observed for LMW individual subunits in the previous chapter (Figure 3.12). These bands were identified as CCT subunits and could represent unstable complexes of lower molecular weight that have fallen apart or micro-complexes of only a few CCT subunits; these were not studied further in this work. From this method, 1.2 mg of pure homogenous TRiC was isolated from a 6 L culture of endogenous cells. This yield is thus sufficient for biochemical and structural follow-up studies as detailed in Chapter 5.

4.6 Nanobody-mediated pull-down

The monoclonal cell line showed promise in its ability to isolate endogenous human TRiC, making data collection using cryo-EM the logical and achievable next step for this sample, as the goal for this project was to obtain high-resolution structural information for human TRiC. However, the subunit arrangement of the TRiC complexes was still not known and differentiating the eight structurally homologous subunits in cryo-EM reconstructions could be a limiting factor in achieving high resolution structural data. While the CRISPR/Cas9-mediated genome editing pipeline was in progress, nanobody technology was also being pursued in this study, both as a possible way to isolate native TRiC and for use in identifying a single subunit within the TRiC complex during cryo-EM data processing.

Nanobodies (Nbs) have recently emerged as important tools in structural biology, as they can stabilise antigen conformations and assist in crystal packing, modulate enzymatic activity, and can also be used as affinity bait to isolate endogenous protein complexes[230–232]. Nbs are single heavy-chain only antibody fragments that are derived

from immunisation in camelid organisms including llamas, and contain three variable loops that aid antigen-specific binding, known as complementarity-determining regions (CDR), and are numbered as CDR1, CDR2, and CDR3[175]. In this study, CCT5 was chosen as an antigen for which nanobodies could be raised, to identify CCT5 within the TRiC ring and assist alignment of particles during EM studies.

4.6.1 Development of anti-CCT5 nanobody binders

As introduced in Chapter 3, full-length CCT5 was expressed in *E. coli* as a homo-oligomer $CCT5_{16}$, isolated by purification to the purest extent possible, to ensure that nanobodies were raised against CCT5 and not contaminating proteins. The homo-oligomeric peak from size exclusion was isolated (Figure 4.17) and sent for immunisation in llamas at the Steyaert lab at VIB-VUB, Brussels, as a collaboration through the INSTRUCT Nanobodies for Structural Biology programme. In this way, using a sample expected to form the $CCT5_{16}$ homo-oligomer, antibodies should only be raised on surface-exposed regions similar to those present when CCT5 is assembled within the native TRiC complex.

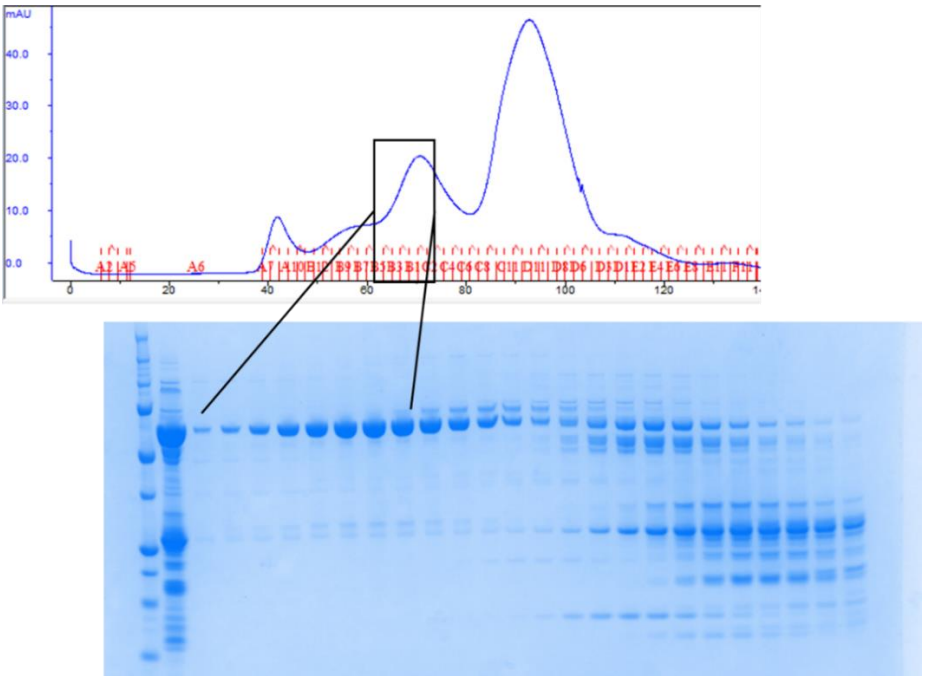


Figure 4.17: CCT5$_{16}$ purified for immunisation in llamas. Size exclusion chromatogram (top) and SDS-PAGE (bottom) of CCT5 purified for immunisation. Black box indicates fractions pooled for immunisation.

The immunisation procedure outlined in this paragraph was carried out by staff in the lab of Jan Steyaert at the Vlaams Instituut voor Biotechnologie. In brief, Llamas received several injections of the CCT5$_{16}$ antigen over a span of four months after which point nanobodies were isolated by extracting blood, separating out peripheral blood lymphocytes, and isolating RNA for use in creating cDNA libraries[175]. *In vitro* selection (panning) of conformationally selective nanobodies was performed using phage display and binders were evaluated by an ELISA that detected presence of nanobody in the sample. ELISAs were run with CCT5 antigen coated on the surface of a 96-well plate and nanobody binders were identified. In the CCT5 immunisation experiment, 28 different Nb binders (Nb1-Nb28) were identified that could be grouped into sequence families based on whether the CDR3 regions contain identical length and share >80% sequence identity[175] (Figure 4.18).

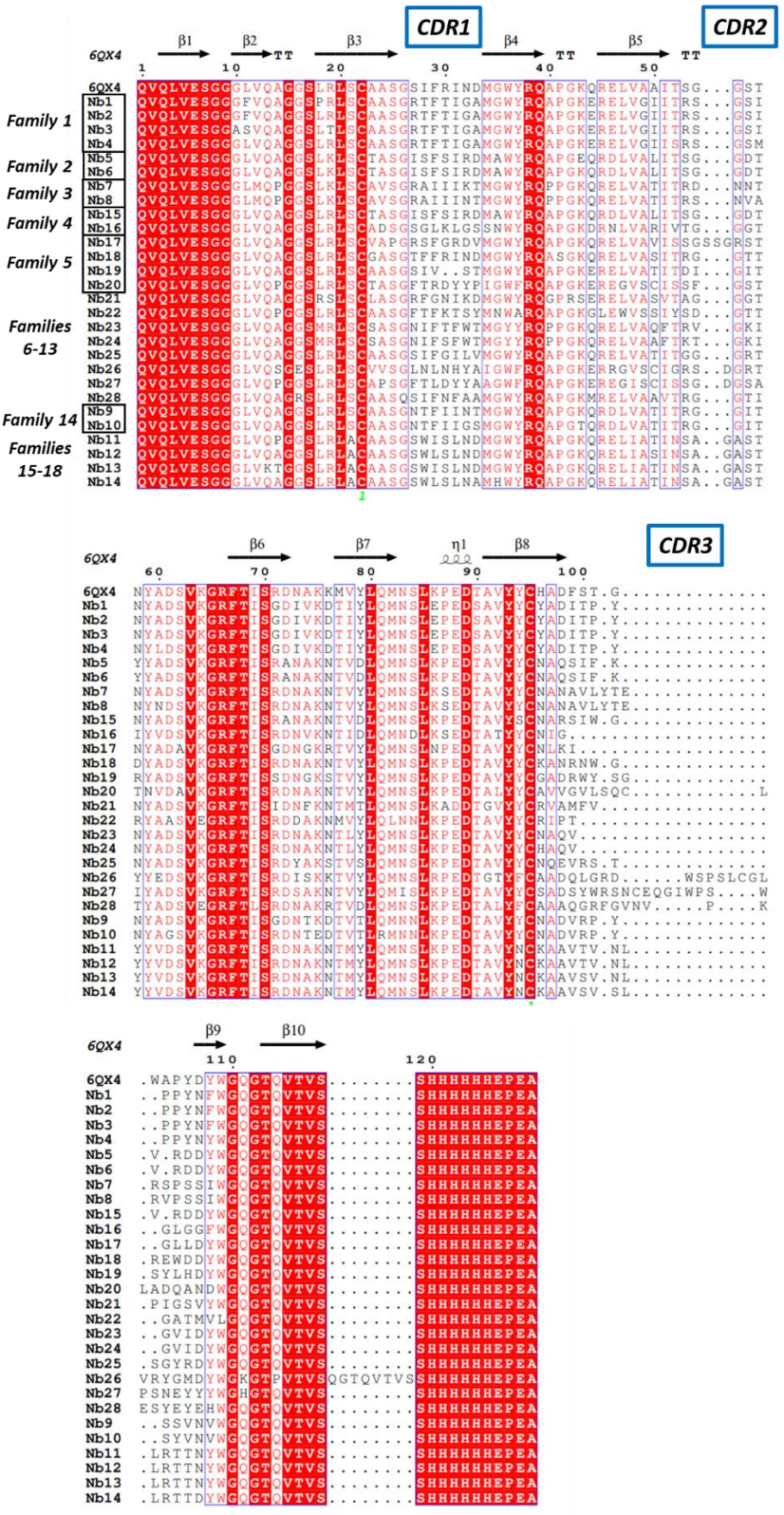


Figure 4.18: Summary of 28 anti-CCT5$_{16}$ nanobody binders. Sequence alignment of 28 nanobodies Nb1-Nb28, overlayed with secondary structures from a nanobody homology model (PDB: 6QX4 chain C) with the three CDR loops labelled.

4.6.2 Anti-CCT5 nanobodies can isolate native TRiC

All 28 nanobodies were constructed in a pMESy4 vector allowing recombinant expression in *E. coli*. The nanobody constructs were designed to incorporate both a His6 tag and a C-tag, the latter being a four-residue peptide designed to have high affinity to a commercial anti-C-tag nanobody that can be incorporated at the C-terminus of a target protein[233]. From expression in 1 L cultures, 24 Nb proteins were successfully purified *via* Ni-NTA affinity column followed by size-exclusion chromatography (Figure 4.19A). The presence of C-tag allowed pull-down experiments to be conducted with the His6-tagged CCT5 construct sent for immunisation, to rank and triage nanobodies by their ability to isolate CCT5 from over-expressed lysate using C-tag affinity resin.

A pull-down experiment was carried out whereby the purified nanobodies were mixed with overexpressed lysate from the same full-length CCT5 construct submitted for immunisation. After incubation, the nanobody-lysate mixtures were purified using C-tag affinity gel matrix and visualised on an SDS-PAGE gel (Figure 4.19B). The eluant from the C-tag column showed a gel band corresponding to CCT5 for 18/24 nanobodies, indicating that these nanobodies could bind CCT5 by pulling it out from lysate, and therefore have potential to also bind CCT5 within the native TRiC complex for isolation from endogenous source. To test this, the same lysate pulldown experiment was repeated with nanobodies Nb3, Nb16, Nb18, Nb22, Nb24, and Nb25 that had co-eluted with the largest observable CCT5 bands (Figure 4.19B), but this time wild-type HEK293 lysate was used for the pull-down experiment (Figure 4.19C). Three of the six nanobodies (Nb18, Nb22, and Nb24) successfully isolated endogenous TRiC from HEK293 lysate, based on verification by gel excised LC-MS/MS.

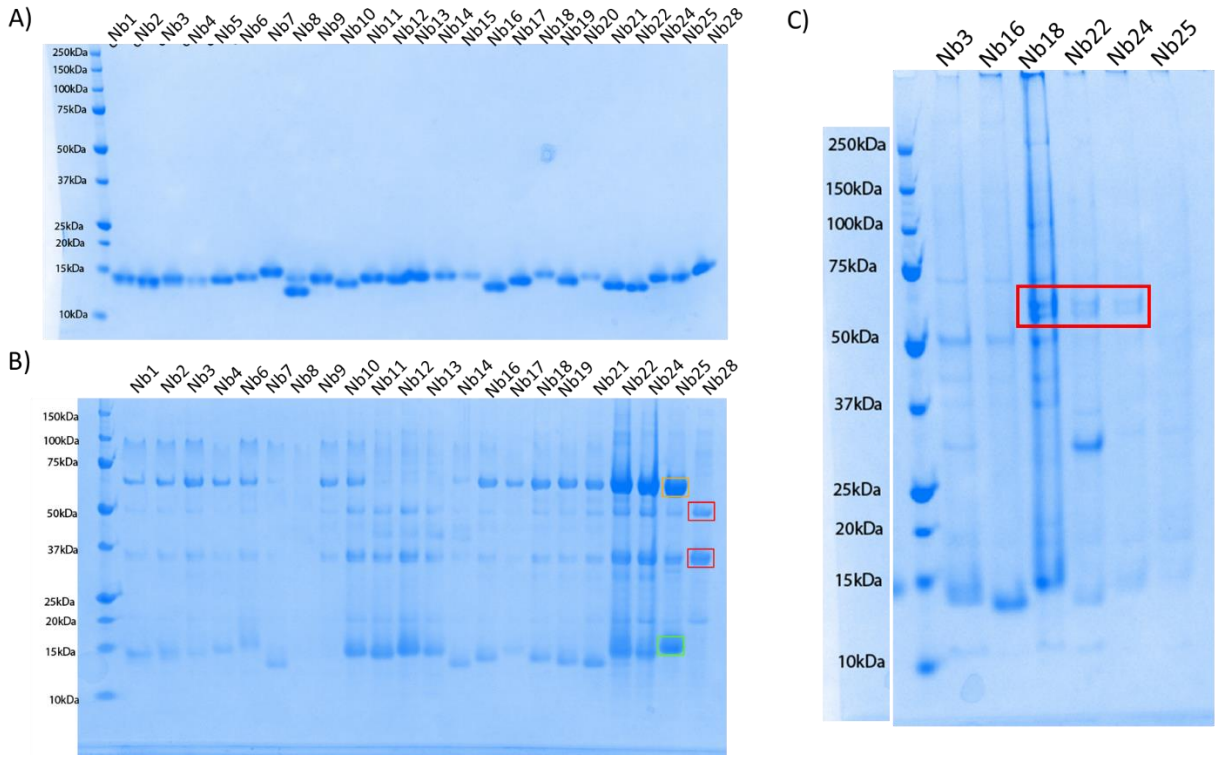


Figure 4.19: Nanobody purification and pull-down experiments. A) Purification of 24 anti-CCT5 nanobody constructs overexpressed in *E. coli*. Gel shows elution after size exclusion column for each nanobody. B) Nanobody-mediated pull-down of CCT5 from full-length CCT5 over-expressed *E. coli* lysate. Tryptic digest MS/MS identification denoted by boxed regions confirm presence of CCT5 (orange), nanobody (green), and contaminating proteins (red). C) Nanobody-mediated pull-down of TRiC from wild-type HEK293 lysate. Red box indicates CCT subunits identified by tryptic digest MS/MS indicative of purified TRiC.

Due to the success of the nanobody-mediated isolation of native human TRiC, the pull-down experiments were next repeated with additional nanobodies (n=19) from the original set, that could be readily purified in large-scale (Figure 4.20). In all, 17 out of the 19 tested nanobodies pulled down TRiC from wild-type HEK293 lysate and only elution fractions from constructs Nb11 and Nb13 did not contain CCT subunits, as verified by tryptic digest LC-MS/MS. Altogether, these data suggest that the anti-CCT5 nanobody binders developed for isolation of endogenous TRiC work in proof-of-concept. However, these were not pursued for the isolation and purification of TRiC at the time of

study, as the contaminating proteins in each sample were significant, and the alternative approach of purification from the genome edited CCT5 C-terminal line (Section 4.5) was already yielding highly pure TRiC at moderate yields. Instead, the utility of these nanobodies was directed towards cryo-EM experiments and thus nanobodies that bind tightly to CCT5 were of the highest interest to carry forward in Chapter 5.

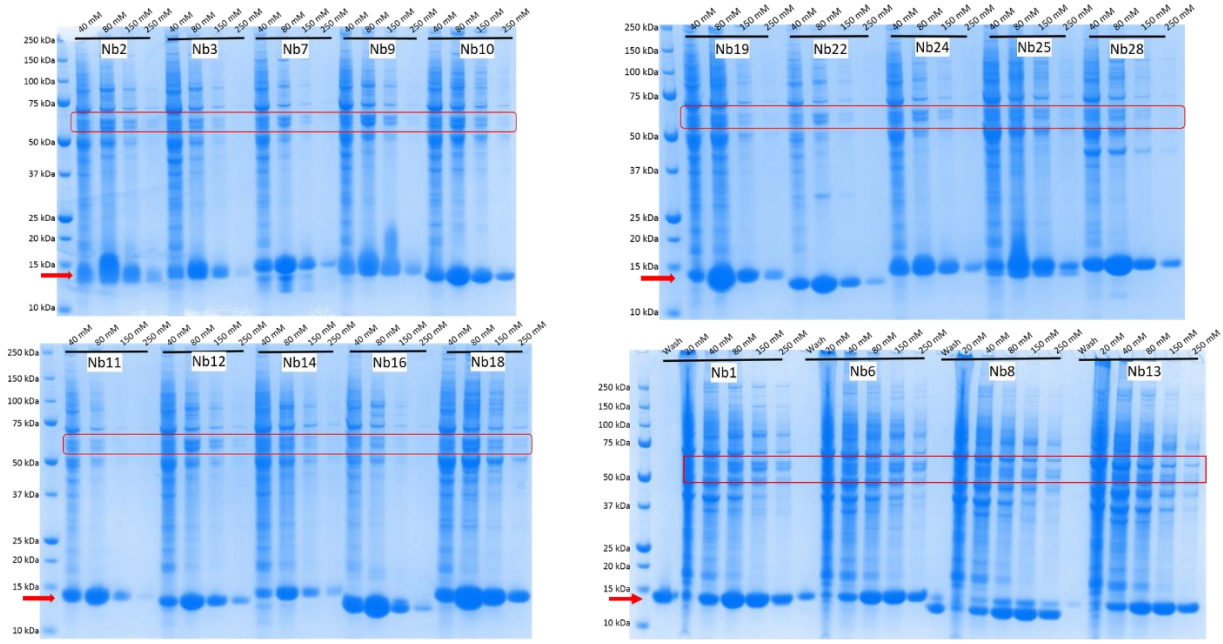


Figure 4.20: Follow-up of nanobody-mediated TRiC isolation. SDS-PAGE gels of wash step and elution fractions containing a gradient of imidazole. Nanobody constructs are identified by black bars. Red rectangles indicate region of gel with CCT subunits representing endogenously isolated TRiC. Red arrows indicate nanobodies.

4.6.3 Biophysical assessment of binding affinities for anti-CCT5 nanobodies

To further characterise the interactions between CCT5 and the series of CCT5-specific nanobodies, bio-layer interferometry (BLI) was used to determine binding kinetics of nanobody-CCT5 complexes. BLI is a biophysical method that measures the change in angle of reflected white light to determine binding kinetics based on

interference caused by protein-ligand interactions[234]. For these experiments, a biotinylated $CCT5_{16}$ construct was created for immobilisation on a streptavidin sensor. Each expressed and purified nanobody was incubated with the CCT5-bound sensors at concentrations of nanobody ranging from low (39 nM) to high (40 µM, Figure 4.21). In total, 14 nanobody-CCT5 complexes were tested, and steady-state binding kinetics were fitted for 10 Nb-CCT5 interactions (Table 4.1). Overall affinities tended to be in the nanomolar range, with a few nanobodies (Nb17 and Nb19) displaying weaker, low micromolar affinities. The BLI data also allow analysis of the on-rate (association) and off-rate (dissociation), by inspecting the slope and shape of the binding curves. Slow kinetics are characterised by shallow curves, while steep curves indicate fast kinetics. The on-rates of nanobodies plateaued at the highest concentration by 100 seconds of association, and dissociation rates varied from slow (Nbs 4 and 22) to medium (Nbs 6, 8, 10, and 22) to fast (Nbs 17, 18, 19, 21, and 25). Nbs 11, 12, 13, and 28 did not show evidence of binding, which agrees with the TRiC pull-down data (Figure 4.21) in which these nanobodies did not pull down significant amount of TRiC from wild-type HEK293 lysate as compared to other nanobodies with higher affinities (e.g. Nbs 8, 10, and 18).

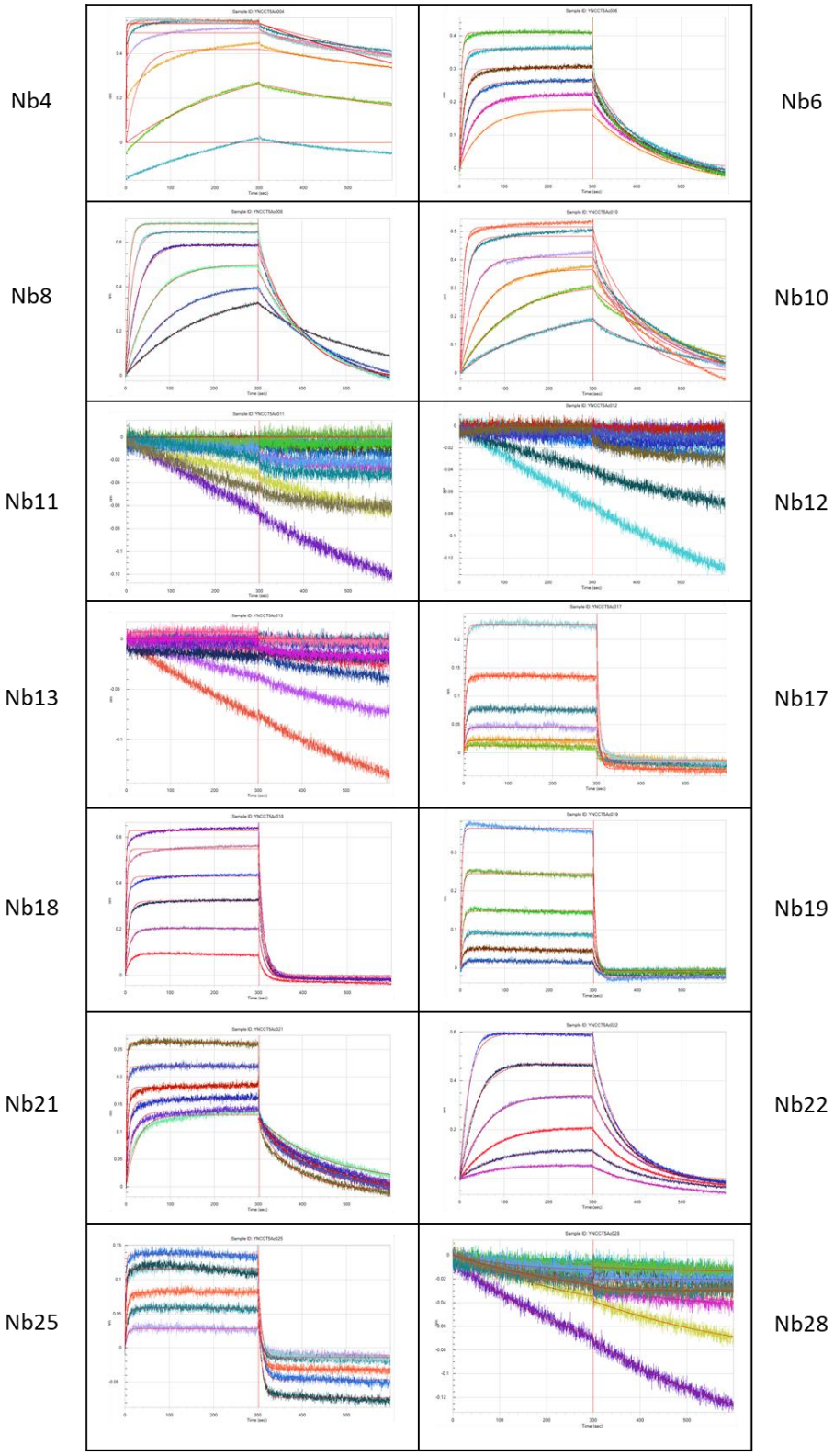


Figure 4.21: Binding kinetics of Nb-CCT5 complexes measured by BLI. Raw curves of CCT5-loaded sensor titrated into increasing concentration of nanobody. Curves were fitted to steady-state kinetics.

Table 4.1: Steady-state kinetics of Nb-CCT5 complexes measured by BLI. Equilibrium dissociation constant (K_D), maximum response determined from the fit of the binding data (Rmax) and fit of curve (R^2) are included.

Nanobody construct	Steady-state binding K_D	Rmax	R^2
c004	170nM	0.5912	0.7629
c006	130nM	0.4007	0.9247
c008	38nM	0.6755	0.9031
c010	29nM	0.5195	0.9712
c017	2.2µM	0.6217	0.999
c018	86nM	0.7377	0.9967
c019	2.2µM	0.6882	0.9993
c021	430nM	0.2418	0.7792
c022	110nM	0.803	0.9992
c025	510nM	0.1491	0.9783

4.6.4 Functional analysis of TRiC through ATP activity

The isolated TRiC sample was run in an ATPase assay (described in Section 3.3.1) to test the biological relevance of the purified complex (Section 4.5.4) along with different anti-CCT5 nanobodies that were isolated for native isolation and cryo-EM experiments (described in Chapter 4). This helps to determine whether the nanobodies can stabilise different conformations of TRiC, as inhibition of ATPase activity would imply the chaperonin exists in a trapped conformation under such conditions. Nanobodies were first run in the presence of the $CCT5_{16}$ homo-oligomeric chaperonin, to observe whether any had inhibitory or activating properties on $CCT5_{16}$ function. $CCT5_{16}$ was used to screen nanobodies, as recombinant $CCT5_{16}$ is easier to produce in high abundance and has a higher stoichiometry of binding sites per nanobody used,

and thus any effects due to nanobody binding would be amplified using $CCT5_{16}$. Interestingly, there were examples of nanobodies such as Nb18 that did not show any effect on CCT5 ATPase activity even at high nanobody concentration (Figure 4.22A). By contrast, another anti-CCT5 nanobody, Nb6, did exhibit inhibitory effects on ATPase activity (Figure 4.22A).

Next, TRiC ATPase activity was measured as compared to $CCT5_{16}$ activity, and was shown to be twice as active (Figure 4.22B). Upon incubation with Nb6 (at equal molar ratio of Nb:CCT5), TRiC activity was reduced by ~20% (Figure 4.22B), indicative of inhibition of the TRiC complex through Nb-CCT5 binding. This reduction in ATPase activity of TRiC suggests that this nanobody might be conformationally selective and could be a useful tool for further structural investigations of TRiC.

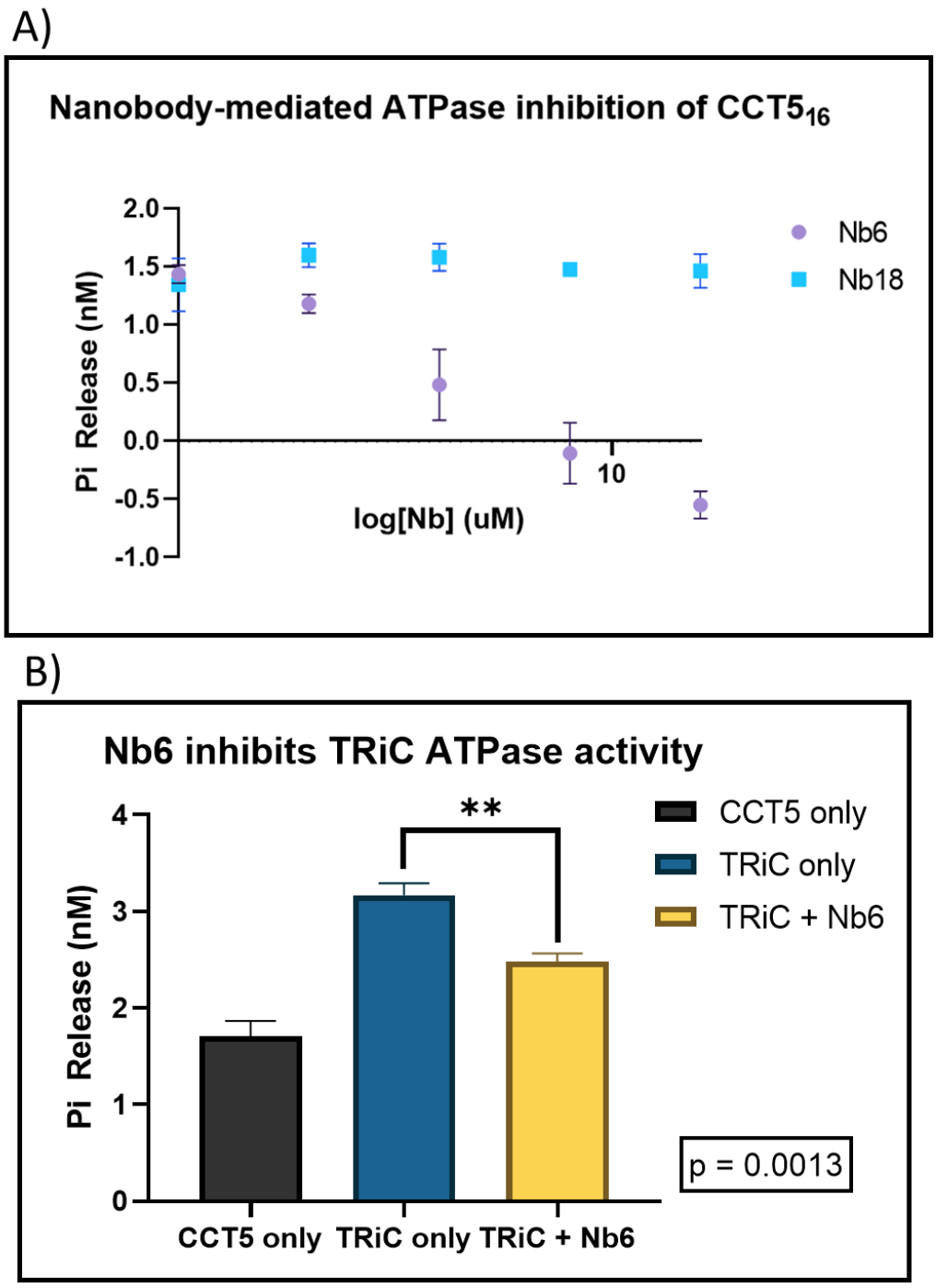


Figure 4.22: ATPase activity of CCT5$_{16}$ and TRiC with anti-CCT5 nanobody binders. A) Scatter plot of free phosphate (Pi) release over nanobody concentration in the presence of CCT5$_{16}$. Experiments contained 0.25mM ATP with reaction time of 30 minutes. B) Bar graph of free phosphate release for CCT5$_{16}$ (black), TRiC (blue), and TRiC in the presence of Nb6 (yellow). P value indicates statistical significance measured by T test.

4.7 Summary

This chapter describes four different strategies to isolate human TRiC. The first strategy involves endogenous isolation from wild-type mammalian cells. This technique shows promise as a proof-of-concept but remains challenging to achieve high enough yields for significant characterisation of the TRiC complex. The second strategy details recombinant co-expression using MultiBac, a technique that showed promise but was not fully realised in favour of alternate methods. The third strategy involved knock-in tagging within the genomic loci of a given CCT subunit for endogenous isolation. This proved the most successful as a 3xFLAG/SpyTag insert was successfully incorporated at the C-terminus of CCT5. Isolation of a monoclonal line that was subsequently suspension-adapted enables large-scale purification of endogenous TRiC especially due to the high expression of CCT subunits in mammalian cells. The final strategy involved raising nanobodies against a single subunit to isolate the TRiC complex *via* pull-down from wild-type mammalian cells. This also proved successful and will also be useful in studying the human TRiC complex structurally by enabling identification of the CCT5 subunit in cryo-EM experiments, as described in Chapter 5. Finally, the anti-CCT5 nanobodies isolated for studying the TRiC complex showed additional promise as modulators and/or conformationally selective antigens to further study the folding pathway of the eukaryotic group II chaperonin.

5. Structure and characterisation of native human TRiC

5.1 Current understanding of TRiC-mediated substrate folding

As detailed in Chapter 1, much is known about the structural basis of substrate recognition for the TRiC chaperonin. These details are highlighted by low resolution cryo-EM and X-ray structures, in combination with NMR, cross-linking MS analysis, and biochemical characterisation of substrate folding. Several structural studies have revealed the substrate location in the open conformation of TRiC. Interestingly, the location within the folding chamber differs between these substrates. Actin, tubulin and the transcription factor AML1-ETO have been found near the apical domains[156,157,165], while the mTOR component mLST8 was seen suspended in the middle of the septum between the two rings[171] (Figure 1.12). The co-chaperone prefoldin has also been shown by cryo-EM experiments to first bind CCT apical domains and then utilise the different electrostatic properties of the two hemispheres of each TRiC ring to deliver substrates in a specific orientation[71] (Figure 1.12). Members of the PhLP co-chaperone family have also been shown to interact at the apical domains of TRiC above bound substrates[169]. Additionally, XL-MS has helped identify the approximate locations of TRiC-substrate contacts[123]. These data have altogether suggested a mechanism whereby nascent folding substrates first engage with specific apical domains before they are shuttled into the folding chamber, either localising to the TRiC apical domains or the septum. Upon ATP hydrolysis the substrates fold into their native structures and are released to the cytosol. However, the intermediate substrate folding steps have yet to be identified, as no studies have shown folding intermediates at resolutions high enough to observe the atomic environment surrounding substrates. This lack of data is

often a result of low affinities for substrates that are not trapped in recombinant preparations of the complex. This chapter aims to address this limitation, and details efforts to utilise cryo-EM experiments with endogenously isolated TRiC to observe high resolution substrate folding intermediates bound within the TRiC folding chamber.

Cryo-EM has emerged in the last decade as a tractable method for determining structures of large protein complexes at atomic resolution that are intractable for X-ray crystallography. The “resolution revolution” era of cryo-EM arrived as a combination of the advent of direct electron detectors and faster data processing algorithms, leading to exponential growth in the number of deposited EM maps in public repositories (Figure 5.1)[235,236]. Due to the size and flexibility of the TRiC complex, its intrinsic pseudo C2 symmetry, and the associated low yields from recombinant systems, cryo-EM was chosen over crystallography as the preferred structural method for this study. To obtain high resolution structures of endogenous human TRiC, the CCT5-FLAG tagged HEK293 cell line described in the previous chapter 4 was used to generate material for cryo-EM.

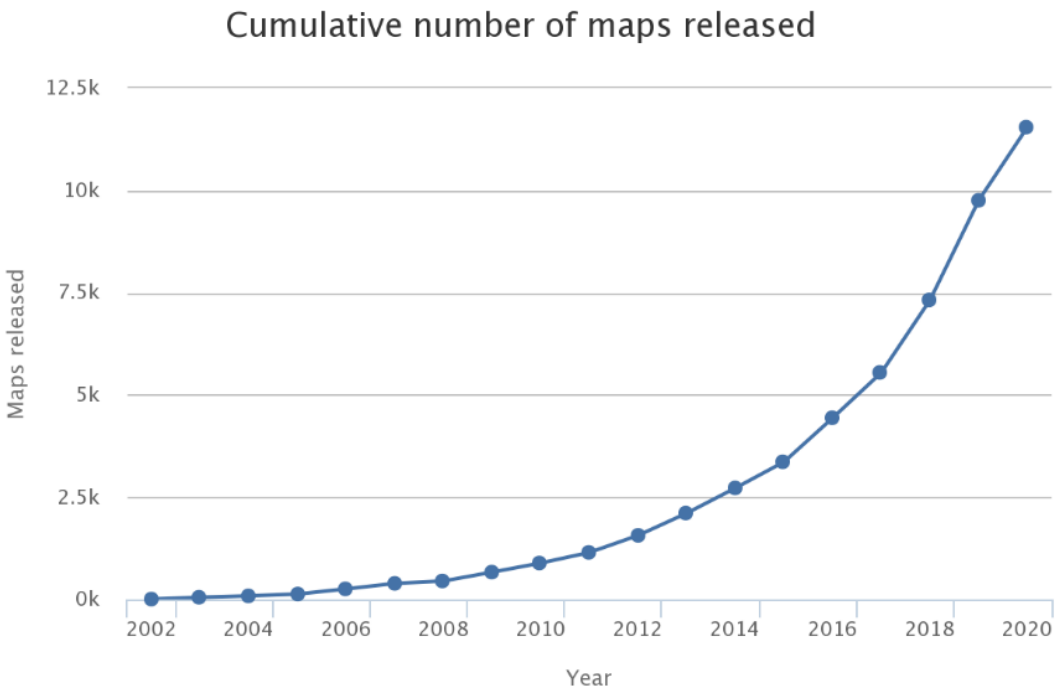


Figure 5.1: Plot of EM maps released in the public domain from 2001-2019. The number of EM maps deposited show exponential growth over the last decade beginning with the "resolution revolution" ~2013-2014 (data from EMDB).

5.2 Considerations for grid preparation and optimisation

Several bottlenecks can occur during sample preparation and data collection that can influence the quality of data during a cryo-EM experiment, including sample heterogeneity, ice thickness, and conformational flexibility[237,238].

5.2.1 Sample heterogeneity

Sample heterogeneity can lead to poor cryo-EM reconstructions and is usually caused by a given protein existing in multiple different conformations within the same sample, however it can also be caused by contaminating proteins in the background of the sample. The presence of these contributions reduces the accuracy of identification of the particle locations in the micrographs, which is then propagated through cryo-EM

data processing and ultimately reduces the resolution in the calculated reconstructions. Heterogeneity caused by contaminating proteins is dealt with during the protein purification stage, and thus is a crucial step prior to grid making (Figure 5.2A). However, if the overall number of particles collected is sufficient, conformational heterogeneity can become an advantage rather than a burden to the overall experiment, as classifications can be performed during the data processing that can lead to multiple structures from a single cryo-EM data set (as shown later in this chapter).

5.2.2 Ice thickness

Ice thickness is often a major parameter that impacts high quality data collection and typically requires iterative optimisation to find ideal conditions that work on a consistent basis[237]. Fortunately for TRiC, several cryo-EM studies[27,71,138,171] have previously determined a range of optimal protein concentration that provided a reasonable starting point, however further optimisation is always necessary due to variation in buffer solutions, humidity, sample heterogeneity, and blotting parameters.

5.2.3 Orientation bias

Additionally, some proteins studied using EM also suffer from orientation bias in which certain regions preferentially bind to the copper grids or the air-water interface[237]. In the case of TRiC, orientation bias was not a major problem, however individual TRiC complexes (particles) did preferentially bind the carbon foil on the EM grid and the region around the edge of the holes in this foil. This effect contributed to inconsistent particle concentration leading to grids that tended to be too sparse for collection (Figure 5.2B). To overcome this problem, a two-blotting system was used in which sample was

first applied to the grid and incubated in a similar manner to negative stain EM preparation[239]. After incubation, grids were manually blotted and sample was reapplied, followed by the normal vitrification process performed by a Vitrobot. It was observed that longer incubation times upon reapplication of the sample during the Vitrobot cycle tended to cause particles to aggregate on grids due to overcrowding (Figure 5.2C). Eventually, the correct timing for this step was determined and this led to consistent high-quality grids of the TRiC sample ready for data collection (Figure 5.2D).

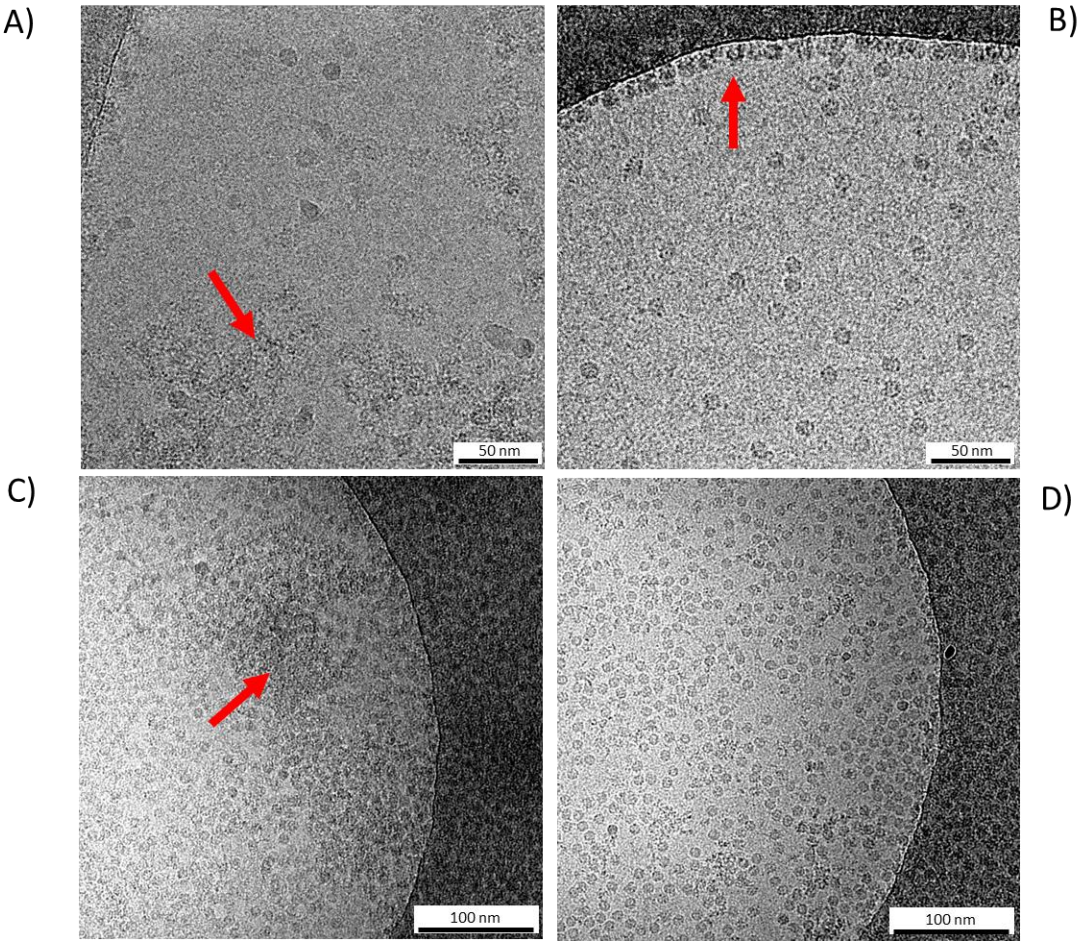


Figure 5.2: Optimisation of cryo grid conditions for TRiC. A) Cryo grid with TRiC made without size exclusion chromatography in which aggregation is present (red arrow). Cryo grids with TRiC in which B) particles preferentially bind carbon (red arrow) and hole distribution is sparse, C) particles are too concentrated and aggregation is present (red arrow), and D) particles are well distributed and ready for data collection.

5.2.4 Conformational flexibility

The final bottleneck to high resolution cryo-EM structure determination is conformational flexibility[237]. Proteins with flexible domains are more difficult to align than rigid molecules and averaging of particles often leads to missing density for these regions. Previous cryo-EM studies of TRiC utilised ATP analogues to “lock” ATP-driven proteins in various states. For example, upon incubation with the ATP analogue ADP-AlF_x (a mixture of ATP, $Al(NO_3)_3$, and NaF) the TRiC complex is “locked” in the post-hydrolysis state. Under these conditions ADP is bound within the CCT subunits along with an AlF_x moiety that mimics the γ-phosphate and prevents opening of the complex. When cryo grids are prepared under these conditions, the portion of open particles is reduced overall, leading to more particles that can be aligned in the more rigid closed state.

5.2.5 Structural homology of TRiC subunits

In the case of TRiC, an additional challenge lies in the structural homology between the subunits, which complicates data processing. While the CCT subunits range in sequence identity from 25–40%[25], each subunit is highly structurally similar in size and shape. In the closed state this problem is particularly pronounced as the complex appears perfectly symmetric at low resolution making subunit identification during model building tricky. This is opposed to what is seen in some open models of TRiC where some subunits, particularly CCT1, open to a larger degree than others[138] which can help in the identification of these subunits. To get around this, the nanobodies raised against CCT5 described in the previous Chapter 4 were used to pick

out an individual subunit within the TRiC complex during cryo-EM data processing. For this purpose, nanobody 18 (Nb18) was selected due to its ability to pull down large amounts of $CCT5_{16}$ and TRiC (Figure 4.14B and C) and because of its favourable binding kinetics (Figure 4.17, Table 4.2).

5.3 Cryo-EM data collection and processing of TRiC-Nb18-ATP·AlF_x

Ultimately, cryo-EM grids were prepared with purified endogenous TRiC in the presence of Nb18 and ATP-AlF_x and collected on a Titan Krios electron microscope from eBIC at Diamond Light Source. Data were processed in the Scipion software framework[181] as detailed in Chapter 2. Briefly, 11,682 movie frames from the data collection were aligned, dose weighted and averaged using MotionCor2[213]. Contrast transfer function (CTF) parameters were estimated using CTFfind4[210]. 3,856,544 particles were picked using crYOLO[188] and subjected to several rounds of classification using RELION[185]. The initial model was calculated *ab initio* using experimental data and the RELION 3D initial model. Particles in the best classes were refined using the gold-standard refinement protocol in RELION. The resolution of the final cryo-EM maps was estimated by FSC combined with phase randomisation to account for the effects of masking.

During the 2D classification step it was revealed that ~90% of the total particles were in the closed conformation (Figure 5.3A: top row), as expected from the ligand incubated with the sample. Surprisingly, the other 10% were in the open state (Figure 5.3A: bottom row). Particles from the closed and open states were then subjected separately to 3D classification. This separation included classifications in which copies of Nb18 were aligned such that the final maps generated reconstructions that “broke”

the pseudo symmetry caused by the structural homology of TRiC subunits (Figure 5.3B). The presence of nanobody in only two locations within the complex (one per copy of CCT5) at this stage suggested that particles were homogenous and that subsequent reconstructions could achieve high resolution.

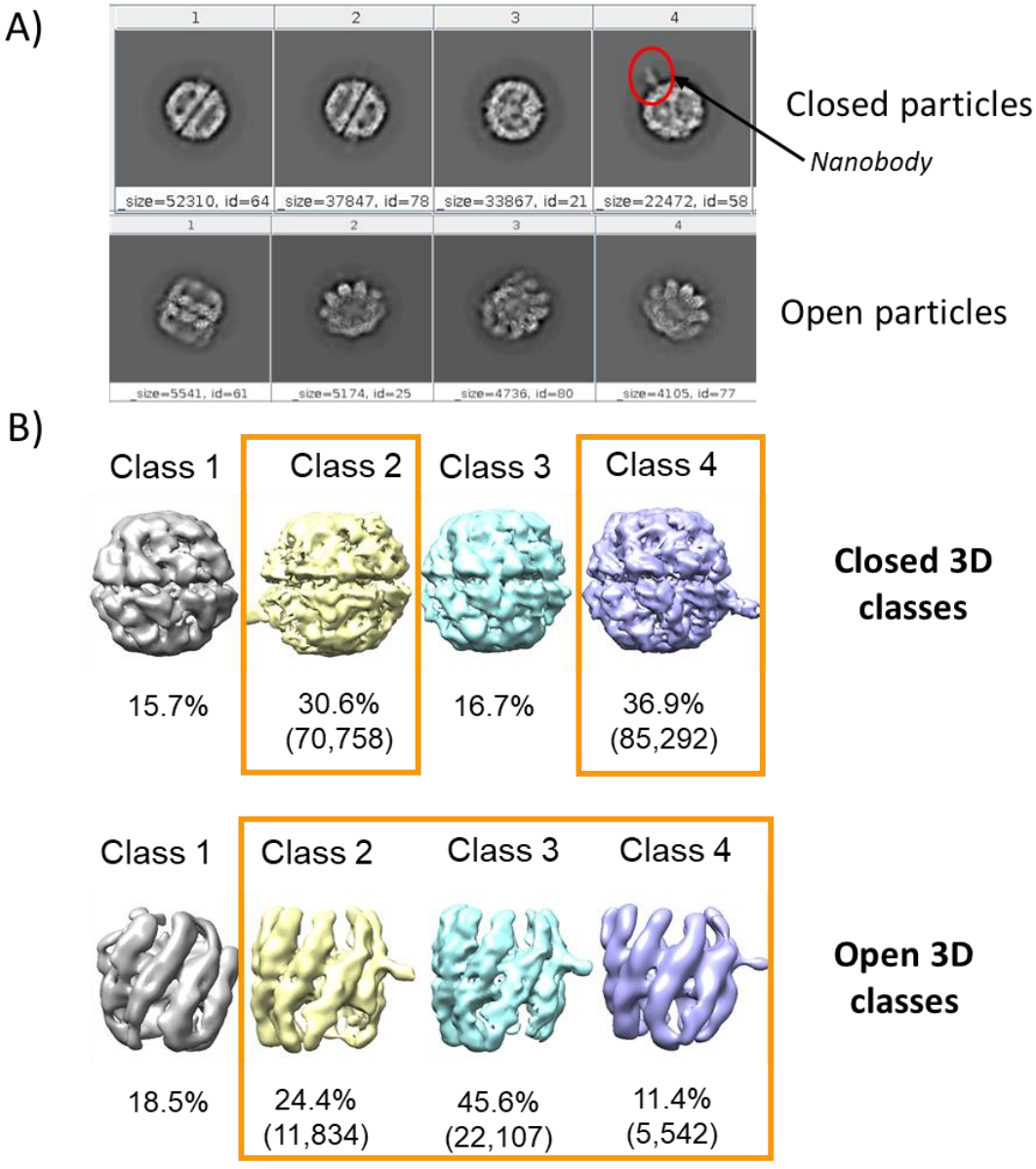


Figure 5.3: Data processing of TRiC-Nb18-ATP·AlFx complex. A) Representative 2D classes of TRiC particles picked using crYOLO particle picker. Classes represent particles in the closed (top row) and open (bottom row) conformations. Nanobody can be seen in several frames and is indicated by red circle. B) Representative 3D classifications of TRiC in the closed (top) and open (bottom) conformations. Orange boxes indicated particles selected for refinement.

After refinement, a consensus map of TRiC in the closed state was generated at 2.5 Å resolution by imposing pseudo-C2 symmetry along the axis created by CCT2 and CCT6 stacking on themselves (Figure 5.4A: left). This C2 symmetry was performed by aligning the placement of the nanobodies along the two-fold axis, with the assumption that each nanobody was bound to CCT5. Similarly, a consensus map of TRiC in the open state was generated at 4.0 Å resolution, containing density within the centre of the TRiC folding chamber that could not be attributed to any CCT subunit (Figure 5.4A: right). Local resolution estimates (Figure 5.4B) show the closed map to be globally around 2.5 Å resolution, except for the most distant regions of each nanobody where density is of lower resolution and appear noisy as these regions have not been filtered to sufficiently low resolution. Similar estimations show the open state to have a gradient of resolution, as expected from previous open maps of yeast and human TRiC[138,171], in which the rigid equatorial domains are close to 3.0 Å, while resolution decreases along the flexible apical domains, to as low as ~8 Å (Figure 5.4C). Both closed and open states also show presence of substrate density, which will be further addressed in subsequent sections (see section 5.6).

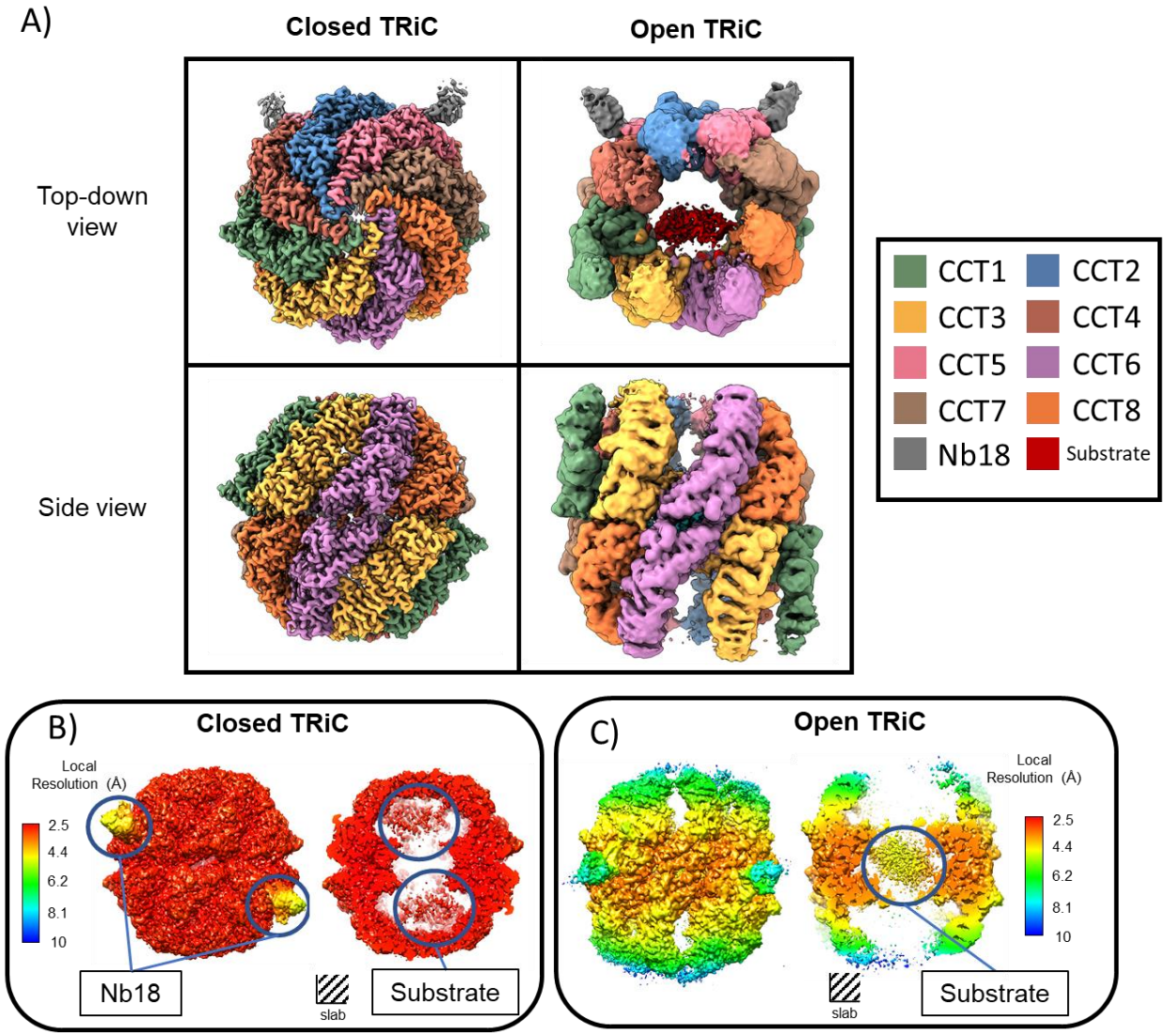


Figure 5.4: Reconstructions of TRiC-Nb18-ATP-AlFx cryo-EM data collection. A) 2.5 Å reconstruction of TRiC in the closed state represented in top-down (top left) and side (bottom left) views, and 4.0 Å reconstruction of TRiC in the open state represented in top-down (top right) and side (bottom right) views. Colours for each subunit are labelled in panel. Local resolution maps of B) closed (left) and C) open (right) states including side view and slice showing presence of substrate.

5.4 Global structure of human TRiC in the closed state

The original aim of the cryo-EM experiment was to characterise native human TRiC at high resolution by inducing the closed conformation, an exercise that has not yet been reported in the literature. This goal was achieved upon reconstruction of the

2.5 Å map of the consensus closed particles, that is, the group of all closed particles after refinement and before any further classifications (Figure 5.4A). This map revealed well-ordered density for the two sets of eight TRiC subunits and two copies of Nb18 that allowed building and refining an atomic model of the entire complex. The eight paralogous subunits are assembled in the order of CCT2-CCT5-CCT7-CCT8-CCT6-CCT3-CCT1-CCT4 when viewed clockwise down the vertical axis for each of the two stacked rings, as previously reported[142] (Figure 5.5A). Upon stacking of the two rings, CCT2 and CCT6 each sit on top of their respective opposite copy (CCT2' and CCT6') at the ring interface, with remaining subunit stacking including CCT8-CCT3', CCT7-CCT1', and CCT5-CCT4' (Figure 5.5B).

5.4.1 Structural details of TRiC in the closed state

The closed-state model revealed well-defined features of TRiC seen in previous structural studies[27], including: the apical lid helices as well as the proximal and release loops from the apical domain, the P-loop motif and NSL loop from the intermediate domain, and the stem loop in the equatorial domain (Figure 5.5C). Clear nucleotide density can be seen in all 16 ATP-binding sites that fits well to one of ADP, Mg^{2+} and AlF_3 each (Figure 5.6A). This corresponds to the expected post-hydrolysis nucleotide that captures the closed state by "locking" the ligand in place, as the unhydrolysable AlF_3 mimics the γ-phosphate of ATP. Similar hydrogen bonding patterns are observed between each ligand and the surrounding nucleotide region along with coordinating water molecules involved in ATP hydrolysis (Figure 5.6B). The specific residues involved in ligand binding are conserved in structural location around the ATP pocket, but vary in sequence numbering among the CCT subunits (Table 5.1).

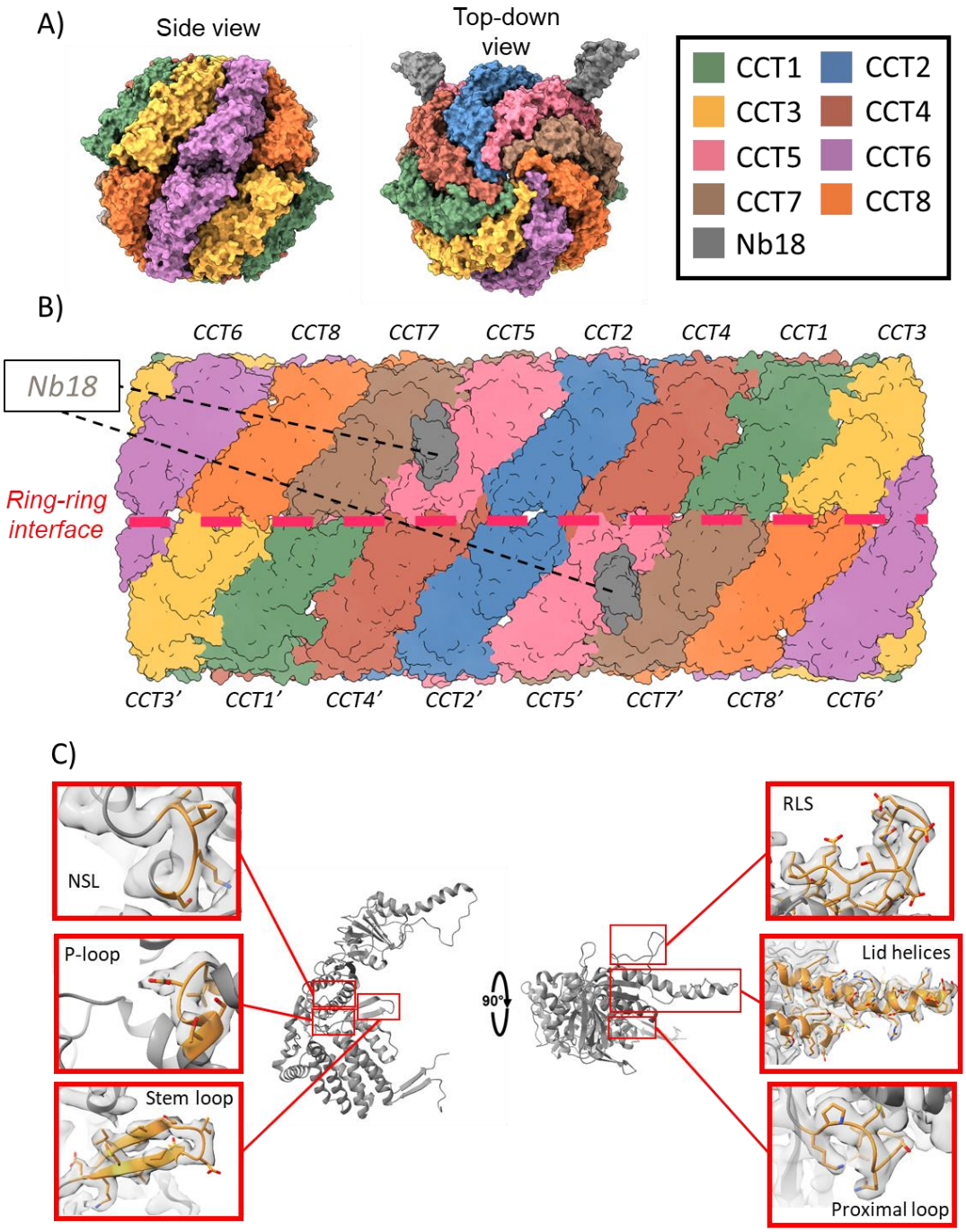


Figure 5.5: Atomic models of TRiC-Nb18 in the closed state. A) Atomic model of TRiC in the closed conformation rendered as surfaces in top-down (left) and side (right) views. Nanobody 18 is indicated in grey (left nanobody is binding to CCT5 (pink) in trans ring that is out of view). B) Side view of polar transformation of TRiC atomic coordinates in the closed conformation rendered as surfaces. View shows TRiC from the outside surface exposed region. Nanobodies are indicated in grey and ring-ring interface is labelled with a dotted red line. CCT2 and CCT6 stack on top of their *trans* counterpart. C) Conserved chaperonin structural motifs within each CCT subunit (here represented by CCT1) including apical lid helices, proximal loop, release loop of substrate (RLS), P-loop, nucleotide-sensing loop (NSL), and stem loop. Motifs are coloured orange in zoom-in panels and density map is shown in grey.

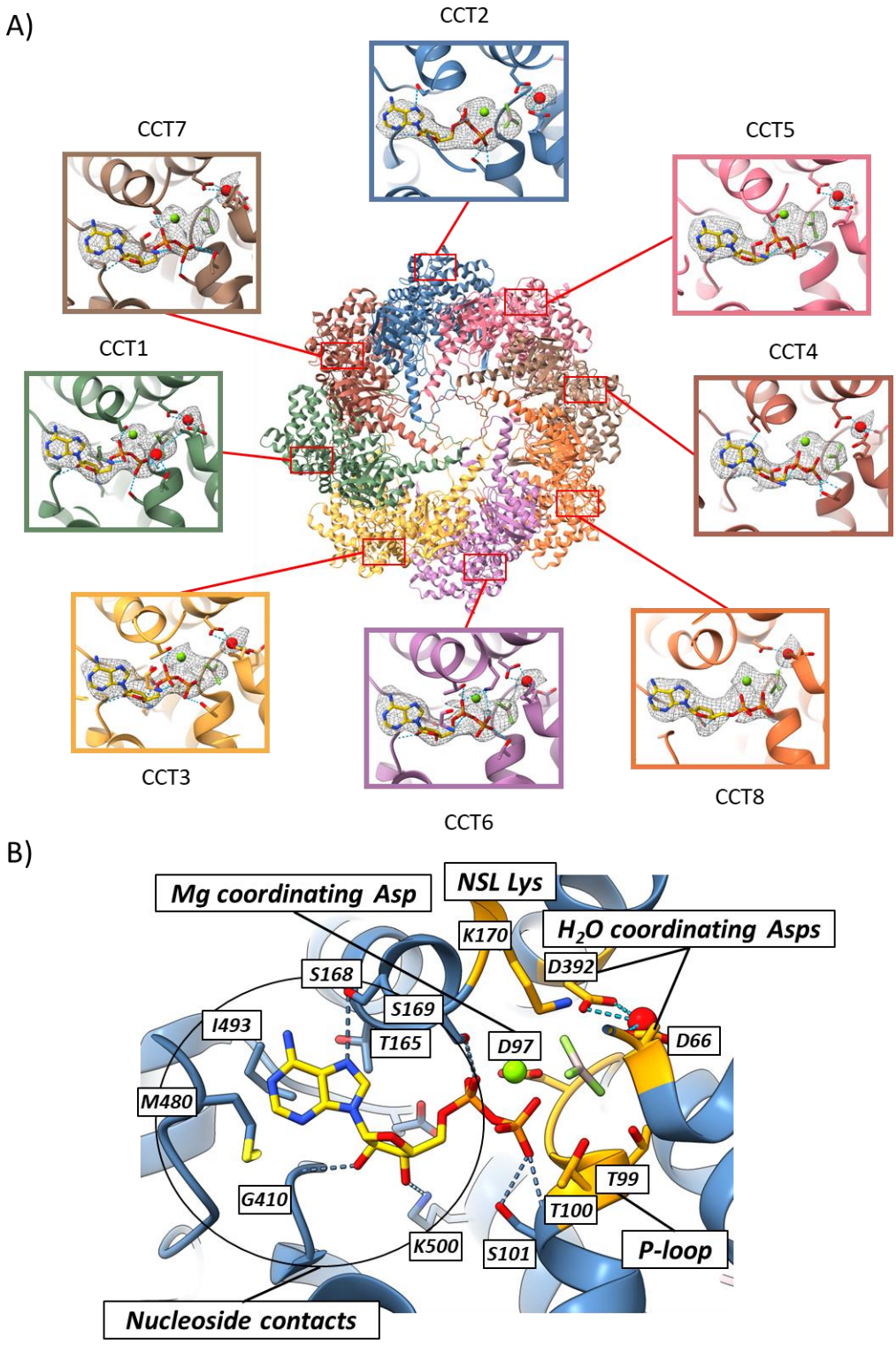


Figure 5.6: Ligand binding within TRiC in the closed state. A) Snapshot of ligand density for each CCT subunit in the closed model. ADP molecules are shown in yellow with Mg^{2+} ion in green and water molecule in red. Dashed lines represent hydrogen bonding. B) Zoomed in view of CCT2 ligand pocket (same view as panel A) including residue contacts listed in Table 5.1. Conserved residues are coloured orange.

Table 5.1: Key residues around CCT nucleotide binding pockets. Numbering follows human subunit sequences.

	P-loop (coordinates γ-phosphate)	NSL (senses γ-phosphate)	Mg coordinating residue	Catalytic Asp	H_20 coordinating residues	Nucleoside contacts
CCT1	G87-D88-G89-T90-T91	K159	D88	D394	D57, D394	P38, T154, S157, G412, L452, R496 V503, E505
CCT2	G96-D97-G98-T99-T100	K170	D97	D392	D66, D392	P45, S101, T165, S168, S169, T185, G410 M480, I493, E495, K500
CCT3	G92-D93-G94-T95-T96	K163	D93	D393	D62, D393	P43, T162, G411, L451, I495, E497
CCT4	G103-D104-G105-T106-T107	K174	D104	D407	D73, D407	P54, T169, N172, S173, G425, L465, V508,
CCT5	G103-D104-G105-T106-T107	K176	D104	D404	D73, D404	P54, T171, G422, L462, C493, M501, V506, E508
CCT6	G89-D90-G91-T92-T93	K159	D90	D393	D59, D393	L38, P40, T158, A411, L541, L481, V494, D496
CCT7	G91-D92-G93-T94-T95	K166	D92	D391	D61, D391	L40, P42, T161, S164, S165, G409, L449, V492, E494
CCT8	G98-D99-G100-T101-N102	K171	D99	D394	D68, D394	Y47, P49, M169, S170, G412, L452, I482, M492, I497, D499

5.4.2 Inter-subunit interactions

The model of TRiC in the closed state showed the same motifs involved in subunit assembly as previously reported[27]. These include β-sheet interactions consisting of the β1-β2 hairpin stem-loop of one subunit and the N-/C-terminal β-strands of the adjacent subunit within the same ring (Figure 5.7A). This stacking assists the formation of a rigid septum between rings, constricting the interior into two largely separate cavities. This is the case for all subunit interactions except for CCT4, in which the N-terminus kinks outward towards the exterior of the TRiC complex (Figure 5.7B). The interaction between the C-terminus of CCT4 and the stem-loop of CCT2 appears to be less stable, as density is not as strong in the stem-loop of CCT2 compared to other stem-loops, though the relevance of this feature remains unknown.

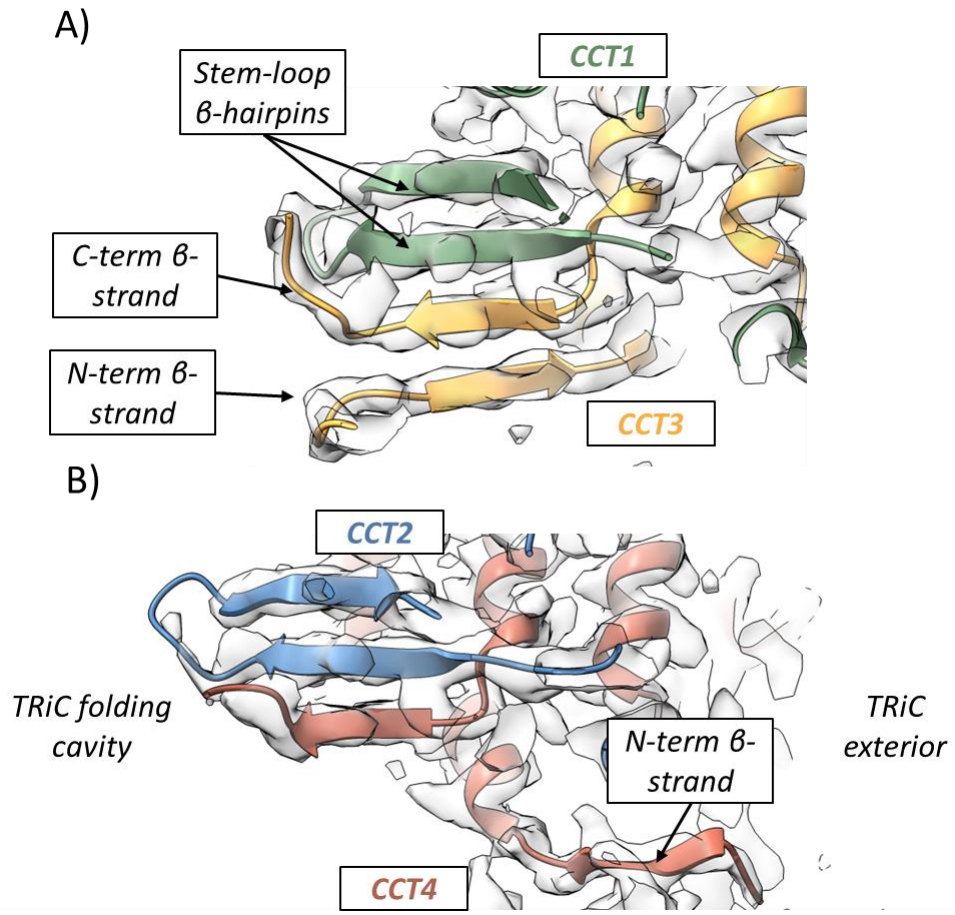


Figure 5.7: Subunit assembly is mediated by β-sheet stacking between subunits. A) Representative stacking of β-strands including the stem-loop of CCT1 (green) and N- and C-terminal β-strands of CCT3 (yellow). B) Unique assembly between CCT2 (blue) and CCT4 (red), showing the N-terminus extending towards the exterior of the TRiC complex.

Additionally, almost all N-terminal residues beyond the first β-strand could be modelled in the closed map. These show a network of interaction between the *cis* and *trans* rings that is regulated by the N-termini of specific CCT subunits. In this network, the N-terminal extensions appear to interact not with the subunit immediately below in the ring, but to the +2 subunit (N' + 2) in the anticlockwise direction (Figure 5.8). Due to the stacking of CCT2-CCT2' and CCT6-CCT6', the CCT4 and CCT8 N-termini interact with their pairs in the opposite ring, although CCT4 does not form any interaction as it protrudes towards the exterior of the TRiC complex. The exact function of this network

is not fully understood, though these interactions likely stabilise the tight closing of the septum to promote substrate folding.

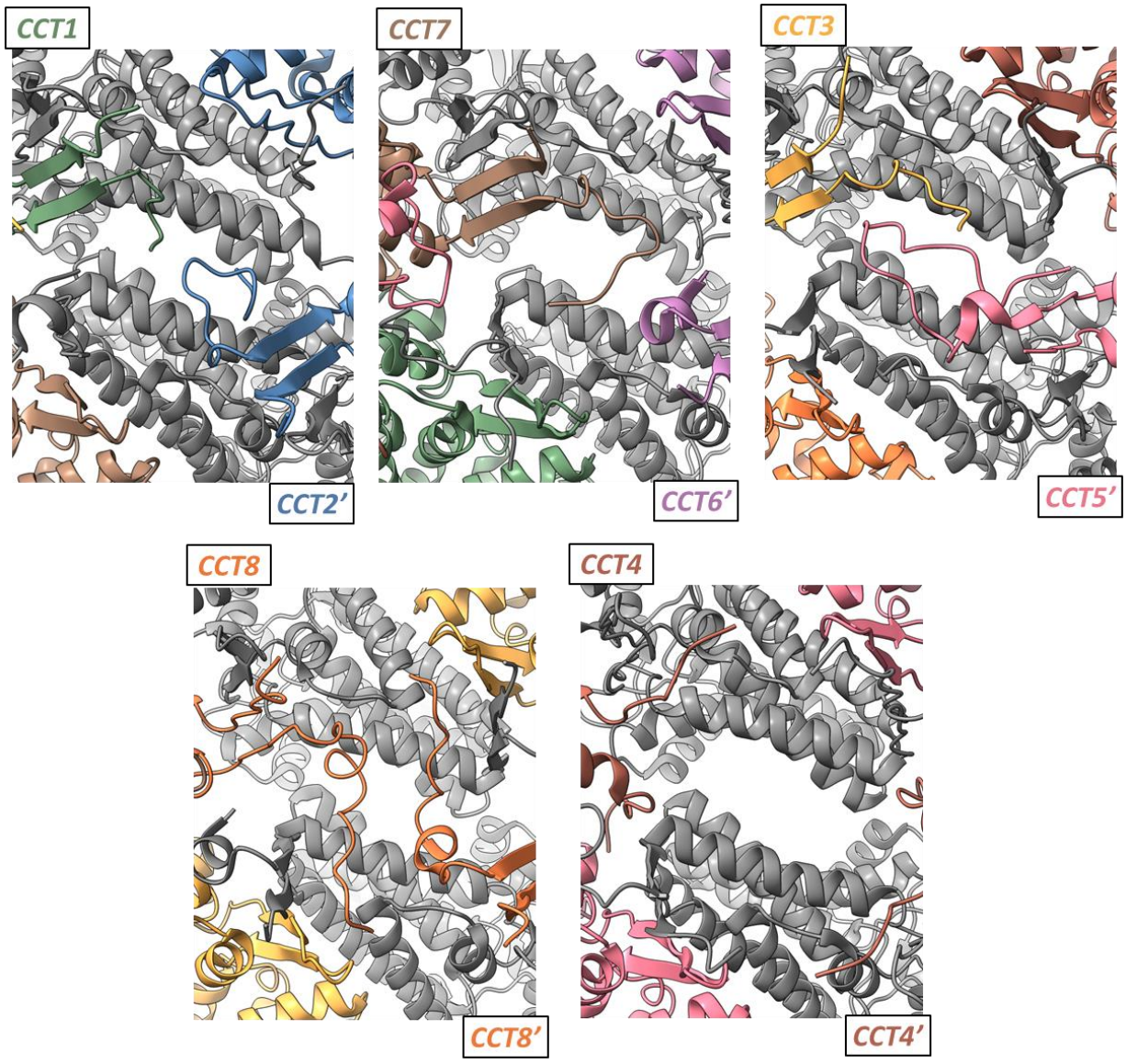


Figure 5.8: N-terminal interaction network between *cis* and *trans* rings. Panels show pairs of N-terminal extensions from TRiC subunit N' + 2 pairs. CCT8 N-terminus interacts with its counterpart in the opposite ring due to stacking of CCT2-CCT2' and CCT6-CCT6'. CCT4 N-termini do not interact with opposite ring due to its exterior protrusion outside the TRiC complex. Stacked subunits in end-on view are greyed out to highlight N-terminal extensions of adjacent subunits.

5.4.3 Nanobody interaction with CCT5

Two copies of Nb18 were observed in the map density, binding the two copies of the CCT5 subunit present in the TRiC complex. Nb18 binding to CCT5 takes place on a hydrophobic patch on the CCT5 equatorial domain, proximal to the ATP binding site (Figure 5.9A and B). The binding interface of Nb18 consists mainly of interactions from the variable CD1 and CD3 loops (residues Phe29/Arg30 and Asn100/Trp101/Arg103/Trp105, respectively). While the density observed for Nb18 was of high enough resolution to observe these binding interactions, further validation was performed using mass spectrometry analysis (next section) in tandem with the cryo-EM studies to observe the Nb18-CCT5 binding interactions, in the event that resolution limitations prevented these structural observations.

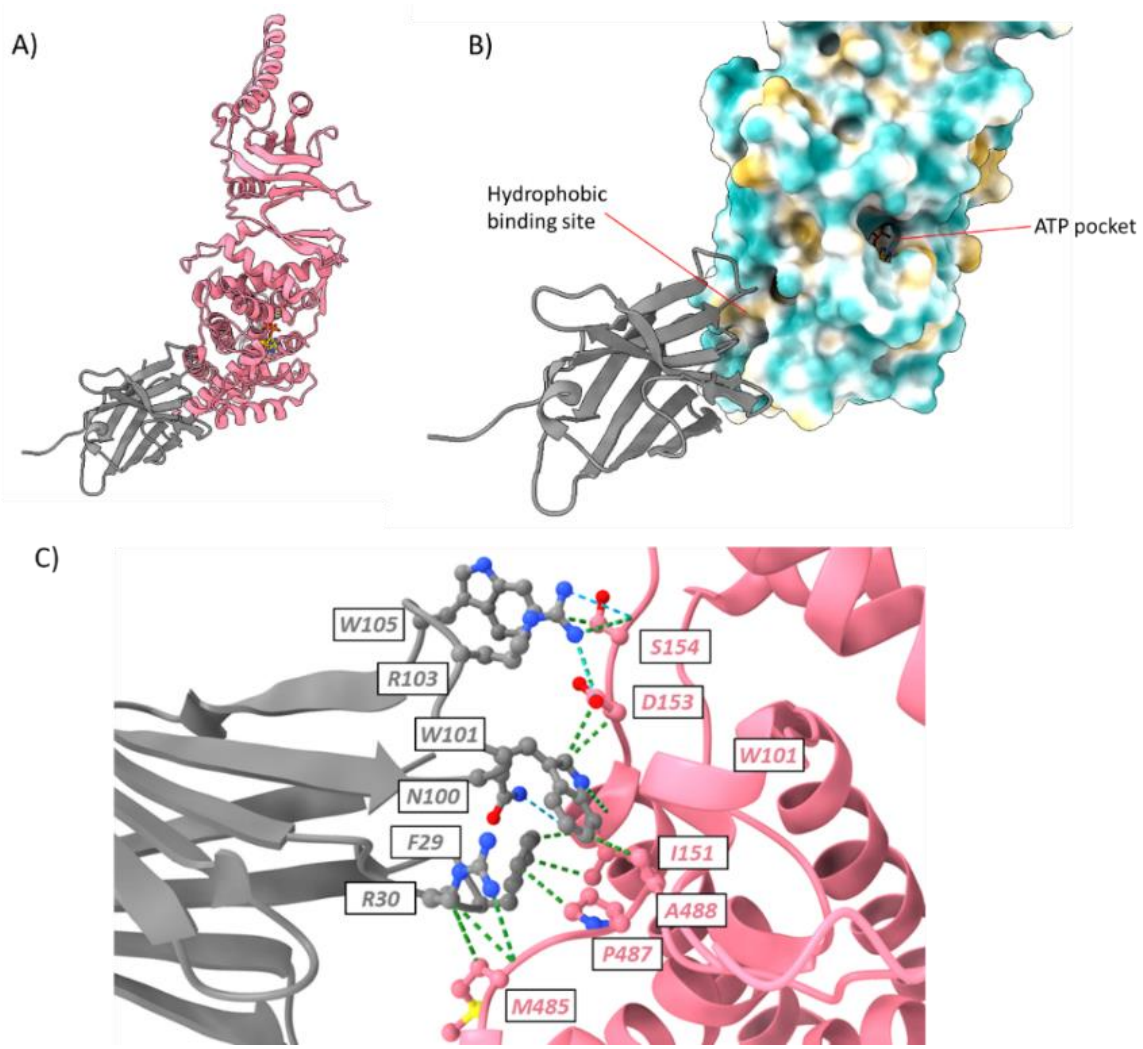


Figure 5.9: Nanobody binding to CCT5. A) Zoomed out cartoon representation of nanobody (grey) binding to the equatorial domain of CCT5 (pink). B) Zoomed in view showing nanobody (cartoon) binding to hydrophobic interface of CCT5 (surface representation) in proximity to ATP binding pocket. C) Interacting residues between nanobody and CCT5 in the closed state. Green dashes indicate van der Waals contacts; blue dashes indicate hydrogen bonds.

5.5 Solution digest MSMS and crosslinking identified potential substrates bound within TRiC folding cavity

The endogenous purification used for this cryo-EM experiment allowed for the possibility to capture native substrates in their intermediate folding states while bound inside the TRiC complex. Indeed, the open state map of TRiC, without further

classification, clearly revealed additional density inside the substrate folding cavity that could not be attributed to TRiC subunits (Figure 5.3C). Relevant to this, during purification a ~50 kDa protein was observed that co-purified with TRiC through size exclusion chromatography, suggesting the protein could be bound within the chaperonin (Figure 4.13C). With this observation in mind, the as-purified sample was evaluated using affinity purified mass spectrometry (AP-MS) by solution digesting peptides and analysing them using LC-MS/MS prior to cryo-EM data collection, to identify potential TRiC substrates within the sample. Additionally, samples were also treated with crosslinking agents before LC-MS/MS, to confirm observations from cryo-EM experiments, especially those of subunit orientation and location of Nb18-CCT5 binding.

5.5.1 Crosslinking condition optimisation

The ideal crosslinking conditions for XL-MS strike a balance between too little crosslinker, where not enough peptides are cross-linked making enrichment difficult, and too much crosslinker, where many peptides are crosslinked leading to higher levels of nonspecific background. TRiC crosslinking conditions were first optimised by comparing samples of recombinant $CCT5_{16}$, endogenous TRiC, and recombinant Nb18, with and without varying concentrations of the amine-to-amine crosslinker bis(sulfosuccinimidyl)suberate (BS3). BS3 was chosen as a crosslinker because it is water soluble and widely used, having a crosslinking distance of ~12 Å[240]. The BS3 used was in the deuterated heavy/light d0:d4 form, so that crosslinked peptides were exactly four hydrogen mass units heavier, a technique used to enable unambiguous identification of crosslinked peptides.

Samples were incubated and visualised using SDS-PAGE to observe the effect of BS3 on the samples alone versus in complex with nanobody (Figure 5.10). Control lanes of input samples ($CCT5_{16}$, TRiC, Nb18) without BS3 (lanes 1-3) showed no crosslinking as expected. Samples treated with BS3 in 10-fold excess (moles of lysine residues in sample) showed a clear reduction in band intensity corresponding to its SDS-resolved monomer for both chaperonins ($CCT5_{16}$ and TRiC), while dimer bands appeared around 120 kDa and further higher order species could be seen as aggregates at the top of the gel (Figure 5.10: lanes 4 and 5). Nb18 also exhibited a small dimer crosslink band with 10-fold excess BS3 but did not show a reduction in intensity in the main band (Figure 5.10: lane 6). Additionally, no higher order species were observed, suggesting 10-fold BS3 did not have a large effect on Nb18 crosslinking. A 1:1 ratio of BS3 to chaperonin-Nb18 complex was not sufficient to produce evidence of crosslinking (Figure 5.10: lanes 7 and 10). BS3 at 10-fold excess showed a similar crosslinking pattern to the control lanes including a sharp reduction of band intensity at 60 kDa, indicating most of the sample crosslinked into higher order species (Figure 5.10: lanes 9 and 12). BS3 excess of only 5-fold provided a similar banding pattern without significantly decreasing the intensity of the main 60 kDa TRiC subunit band (Figure 5.10: lanes 8 and 11). Therefore, 5-fold excess BS3 concentration was determined as the optimum condition for the solution-digest LC-MS/MS experiment, as it was likely to provide sufficient crosslinks while not contributing excess background.

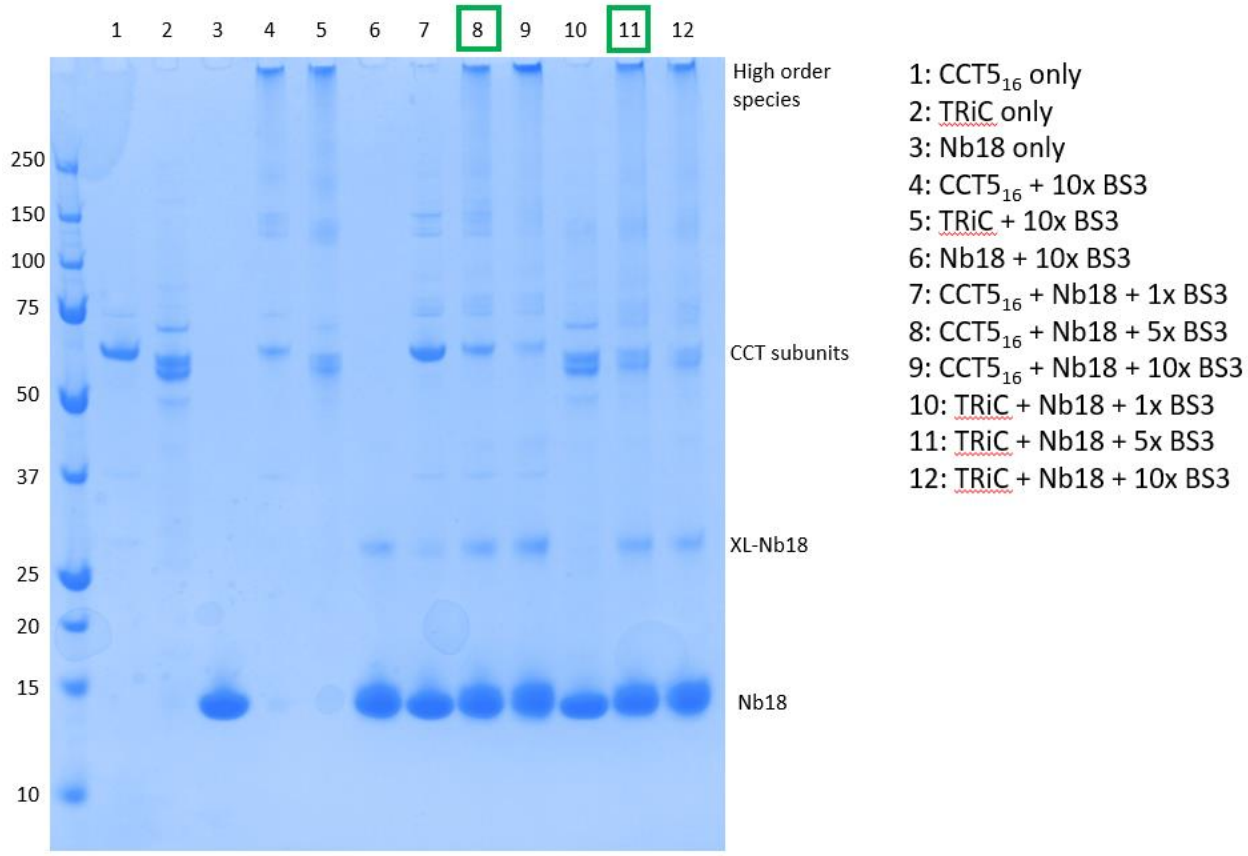


Figure 5.10: Optimisation of crosslinking conditions to determine optimal ratio of crosslinker to chaperonin:nanobody complex. Chaperonin (CCT5$_{16}$ or TRiC) and Nb18 were incubated without (lanes 1-3) or with crosslinker (lanes 4-6) and compared with CCT5$_{16}$-Nb18 (lanes 7-9) or TRiC-Nb18 (lanes 10-12) complexes with varying excess of BS3 crosslinker (1x, 5x, 10x). Lanes 8 and 11 (green boxes) indicate optimum BS3 concentration determined from this experiment.

5.5.2 Cross-linking analysis of Nb18 and TRiC

Samples containing TRiC with or without Nb18 were incubated with 5-fold BS3, digested using trypsin, and peptides were run through LC-MS/MS by Dr. Holger Kramer at the MRC London Institute of Medical Sciences. Samples were analysed first for total peptides present in the sample and second for crosslinks. For the first analysis, all CCT subunits were present in abundance in samples with TRiC alone as expected (Table 5.2, green text). Many other peptides were also identified with a high degree of

confidence that correspond to a range of potential TRiC substrates (Table 5.2, black text). The most abundant peptides originate from known TRiC substrates including tubulin and actin, followed by proteins containing domains that are common to known TRiC substrates including seven-bladed β propeller structures (e.g. RBBP4/7), WD40 repeat containing proteins (e.g. TLE3), histone deacetylases (e.g. HDAC1/2), and cell cycle regulators (e.g. CDH4). From the peptides observed, the AP-MS analysis appeared successful in isolating TRiC along with many client substrates.

Table 5.2: List of peptides identified in purified TRiC sample. Peptides ranked by signal intensity show top putative substrates including CCT subunits (green), tubulin, actin, and others (black). Protein identification numbers, gene names, and sequence coverage are also included.

Protein IDs	Protein names	Gene names	Sequence coverage [%]	TRiC Only_avg_intensity
P78371	T-complex protein 1 subunit beta	CCT2	99.8	2.7679E+11
P48643	T-complex protein 1 subunit epsilon	CCT5	99.8	2.22295E+11
P49368	T-complex protein 1 subunit gamma	CCT3	89.2	2.157E+11
P40227	T-complex protein 1 subunit zeta	CCT6A	98.3	1.86965E+11
Q99832	T-complex protein 1 subunit eta	CCT7	98	1.86355E+11
P17987	T-complex protein 1 subunit alpha	TCP1	98.7	1.762E+11
P50990	T-complex protein 1 subunit theta	CCT8	93.2	1.73365E+11
P50991	T-complex protein 1 subunit delta	CCT4	91.3	1.70355E+11
P68371	Tubulin beta-4B chain	TUBB4B	79.3	39336000000
Q16576	Histone-binding protein RBBP7	RBBP7	87.3	28767000000
O94776	Metastasis-associated protein MTA2	MTA2	87.6	18727500000
Q09028	Histone-binding protein RBBP4	RBBP4	84.7	16552500000
Q04726	Transducin-like enhancer protein 3	TLE3	74.9	16550500000
P07437	Tubulin beta chain	TUBB	75.2	15939500000
Q13330	Metastasis-associated protein MTA1	MTA1	91.3	14605000000
Q13547	Histone deacetylase 1	HDAC1	84.9	12805000000
Q14839	Chromodomain-helicase-DNA-binding protein 4	CHD4	62.5	12127000000
P60709	Actin, cytoplasmic 1	ACTB	77.9	10533500000
Q92769	Histone deacetylase 2	HDAC2	75.4	9147300000
P68363	Tubulin alpha-1B chain	TUBA1B	78	8520250000
P0DMV9	Heat shock 70 kDa protein	HSPA1B	75.4	7986500000
Q92526	T-complex protein 1 subunit zeta-2	CCT6B	24.5	7559050000
Q86YP4	Transcriptional repressor p66-alpha	GATAD2A	76.5	7383700000
O95983	Methyl-CpG-binding domain protein 3	MBD3	82.1	7156550000
Q3ZCM7	Tubulin beta-8 chain	TUBB8	25.9	7133700000

For the second analysis, crosslinking data was processed using XiSearch developed by the Rappsilber lab[241], however crosslinks were sparsely identified using this method. Data were then processed using XlinkX in Proteome Discoverer[242] (Thermo Scientific). This method yielded more crosslinks by focusing on the mass shift difference in heavy/light d0:d4 BS3 crosslinker, achieved through preferentially selecting those features for fragmentation using a 5% false discovery rate (Figure 5.11A). Using this method, several intra-ring crosslinks were identified between TRiC subunits that are known to be adjacent based on the previously determined TRiC orientation (also corroborated in this study), including CCT1-CCT4 and CCT6-CCT8 (Figure 5.11B-C). CCT5-Nb18 crosslinks were also identified that confirmed the location of nanobody binding to CCT5 (Figure 5.11D-E). The crosslinks identified here agree with the structural observations from the cryo-EM data, providing further validation of interactions between Nb18 and CCT5.

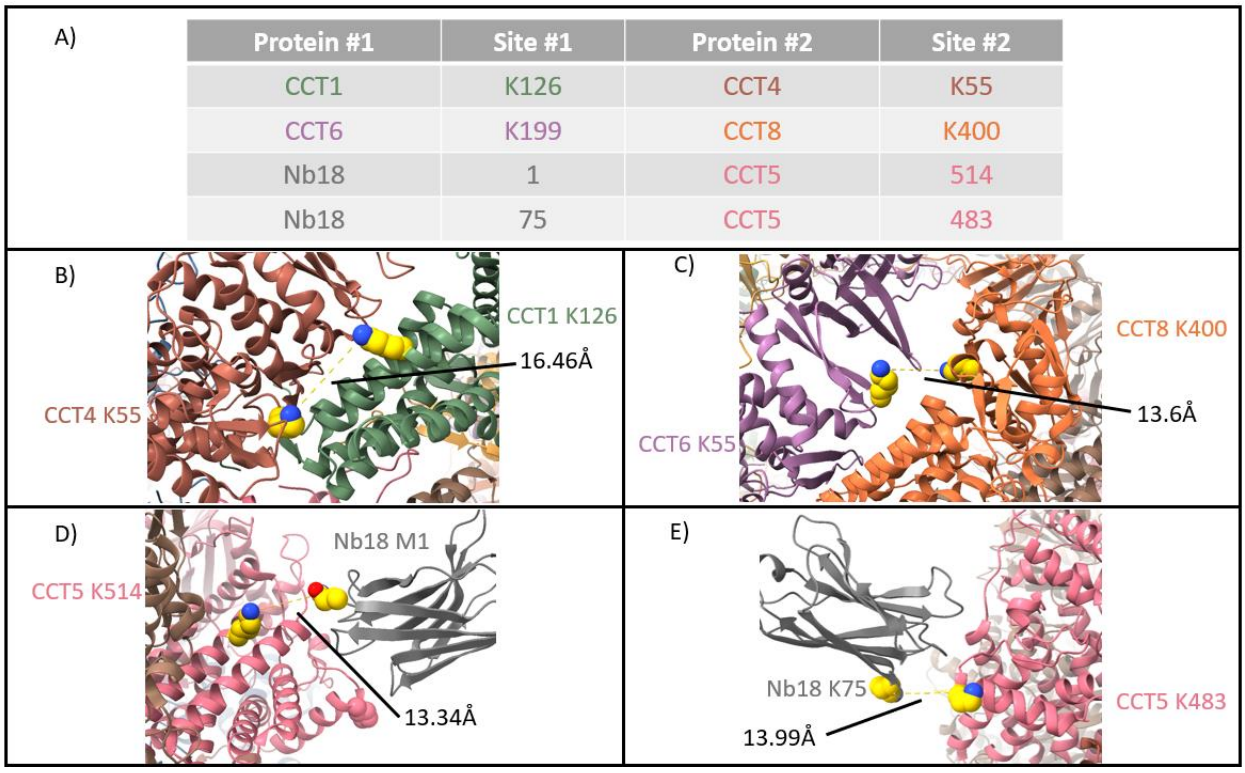


Figure 5.11: Crosslinks identified by d0/d4 BS3 LC/MSMS. A) List of crosslinked residues identified with statistical and biological confidence. Colours correspond to CCT subunits. Intra-ring TRiC crosslinks between B) CCT1 and CCT4 and C) CCT6 and CCT8. CCT5-Nb18 crosslinks between D) N-terminus of Nb18 and CCT5 and E) K75 of Nb18 and K483 of CCT5. Crosslinked residues are highlighted as spheres in gold. Dotted lines indicate distances between amines.

5.5.3 Peptide enrichment in endogenous TRiC affinity purification

To further expand on the list of peptides identified in the TRiC sample without crosslinker, and to validate any potential substrates bound within cryo-EM map densities, AP-MS was again performed on a larger and more robust scale. Two sets of samples were prepared for solution digest LC-MS/MS, including one set from the CCT5-FLAG cell line (described in Chapter 4) and one set from wild-type HEK293T cells (as a negative control). Four biological replicates of each were affinity purified using anti-FLAG affinity gel matrix and eluted samples were digested with trypsin and run through LC-MS/MS.

A volcano plot displaying signal intensity and fold change of CCT5-FLAG vs control HEK293T cell lines revealed enrichment of several peptides beyond the TRiC subunits. These include immunoglobulin-binding protein 1 (IGBP1), programmed cell death protein 5 (PDCD5), phosducin-like protein 3 (PDCL3), many isoforms of β-tubulin, among others (Figure 5.12A). Interestingly, actin, a canonical TRiC substrate, was not significantly enriched in this plot. Actin was slightly enriched in two peptide identifications: one that could not distinguish between α-skeletal muscle, α-cardiac muscle 1, aortic smooth muscle, and γ-enteric smooth muscle, and cytoplasmic 2 isoforms (ACTA1;ACTC1;ACTA2;ACTG2;ACTB), and one that identified cytoplasmic 2 actin (Figure 5.12B).

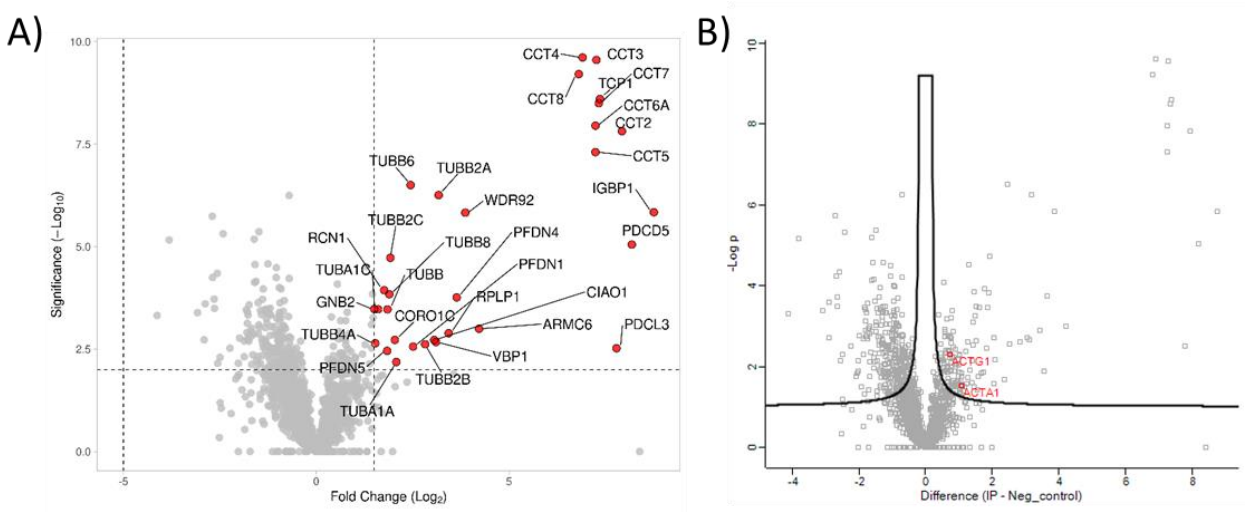


Figure 5.12: Volcano plot of peptides identified in solution digest LC-MS/MS. A) Peptides plotted according to intensity significance vs fold change over control sample. Fold change threshold set to >1.5 and significance threshold set to 2. B) Peptides plotted according to intensity significance vs fold change over control sample with statistical significant threshold depicted by black line. Actin isoforms (labelled red) ACTA1;ACTC1;ACTA2;ACTG2;ACTB (shown as ACTA1) and ACTG1 are slightly enriched according to log2 difference.

In the LC-MS/MS data (Appendix 2) it was also observed that several peptides were present in all four samples isolated from the CCT5-FLAG cell line, while not being

observed in negative control samples isolated from WT HEK293T cells. These peptides are derived from proteins including: PCNA interacting partner (PARPBP), target of rapamycin complex subunit LST8 (mLST8), estradiol 17-beta-dehydrogenase 8 (HSD17B8), actin-related protein 2/3 complex subunit 1B (ARPC1B), cullin-associated NEDD8-dissociated protein 1 (CAND1), cyclic AMP-dependent transcription factor ATF-7 (ATF7), NEDD8 (NEDD8), denticleless protein homolog (DTL), tubulin-folding cofactor B (TBCB), regulator of MON1-CCZ1 complex (C18orf8), and phosducin-like protein (PDCL). These proteins are of potential interest to the TRiC folding pathway, and of note mLST8–TRiC interactions have previously been characterised by cryo-EM[171]. These analyses identify peptides that could be of interest to structural studies of TRiC, as these proteins are enriched upon purification of TRiC and could potentially be observed in structural investigations.

5.6 Substrates bound within TRiC closed state structure

As previously discussed in this chapter, the open map revealed substrate-like density within the TRiC folding chamber (Figure 5.3C). Taken together with the observation that several peptides were highly enriched in the LC-MS/MS data (Table 5.2, Figure 5.12A), it was highly probable that substrate-bound particles should also exist in the closed data set. Indeed, density was observed in the consensus closed map, likely attributable to bound substrate (Figure 5.4B: left panel slice view). Given the high resolution of the closed state reconstruction and the large number of particles (2,278,922) in the data set, the closed state was subjected to another round of 3D classification in RELION with the goal of parsing out substrates into individual classes that might have been averaged in the consensus map.

5.6.1 3D classification of TRiC in the closed state reveals substrate densities with unique features

Ten new classes from the second round of 3D classification revealed varying substrate densities within the TRiC folding chamber that displayed different morphologies (Figure 5.13A). Four of the ten classes revealed strong density for substrate and were broken into two sets. Classes 1 and 10 formed one set, which showed density in one chamber, and classes 6 and 9 formed another set, which shows density in both chambers (Figure 5.13B).

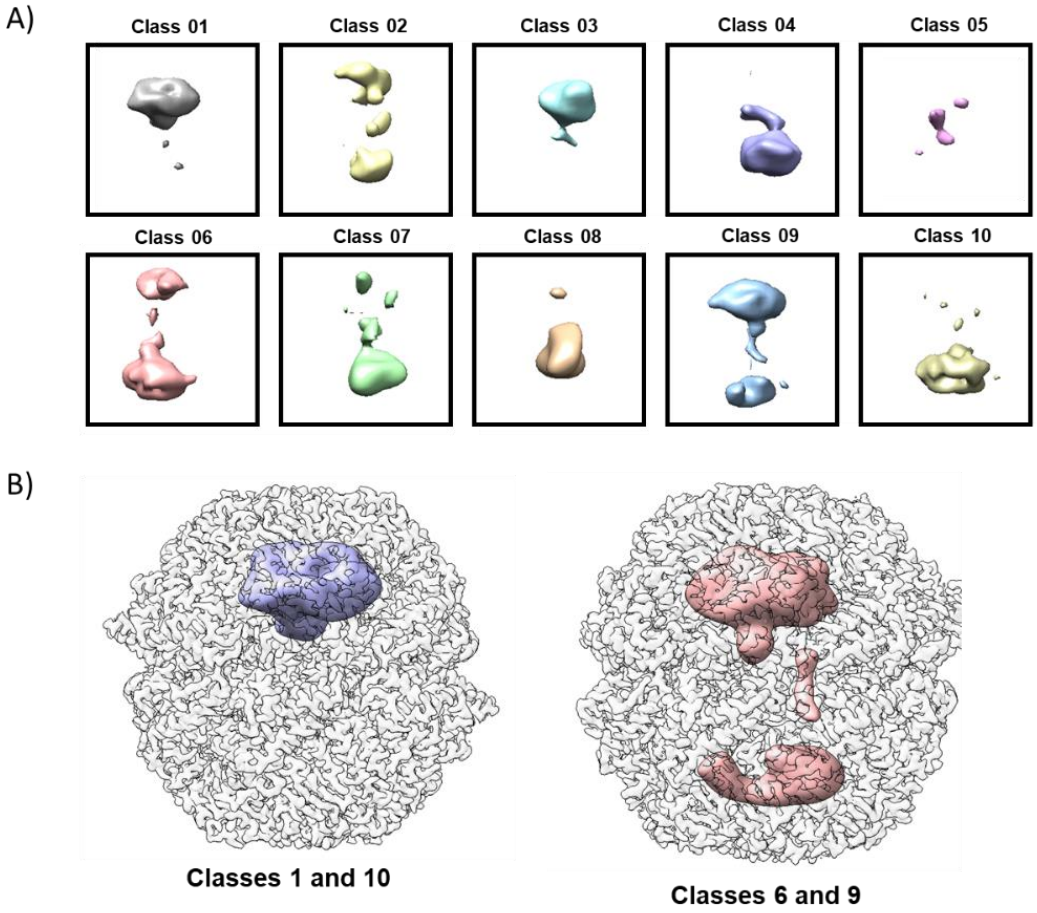


Figure 5.13: 3D classifications of closed state particles. A) Slices of each set of particles from 3D classifications. B) Volumes of each 3D classification. Density is limited to that which is found inside the cavity formed by the CCT subunits. C) Volumes of classes 1 and 10 (top: purple) and classes 6 and 9 (bottom: red) fit into TRiC density.

Classes 1 and 10 and classes 6 and 9 appeared to represent same structural states, just rotated 180 degrees (Figure 5.13A), suggesting that substrates were split between *cis* and *trans* rings. To achieve higher resolution, particles from classes 9 and 10 were rotated 180° to match their counterpart, and both new sets of particles were subjected to another round of refinement. The map generated from combined classes 6 and 9 did not yield strong enough density to model (Figure 5.14A: left). However, the map generated from combined classes 1 and 10 showed strong density within the chamber for a substrate to be potentially identified (Figure 5.14A: right). This observation was consistent with classes 1 and 10 being the two most abundant classes (35.5% of total closed particles).

Notably, this density was positioned relatively higher up in the folding chamber, as compared to the substrate density seen in the open state (Figure 5.4B: right panel slice view) and localised near the apical domains of subunits CCT8/6/3 (Figure 5.14B). Local resolution estimated the substrate density to be around 3.0 Å (Figure 5.14C), which was sufficient to unambiguously assign tubulin as the substrate bound to TRiC in these classes (Figure 5.14D). Tubulin was also the first choice considered as the bound substrate, because it was previously observed in endogenously prepared samples of bovine TRiC crystallography data[136], and was highly enriched in the LC-MS/MS data (Figure 5.12A).

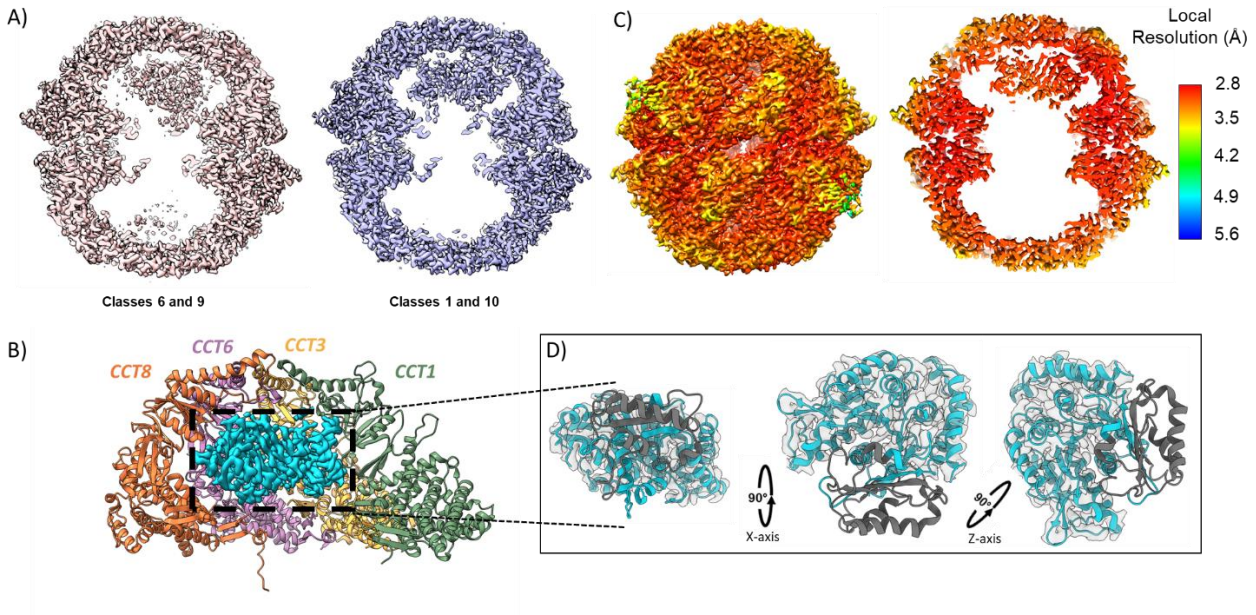


Figure 5.14: Tubulin bound within TRiC. A) Slice of post-refinement maps derived from combinations of classes 6 and 9 (left: pink) and classes 1 and 10 (right: purple). B) Localisation of substrate density to CCT subunits 3-6-8-1. C) Local resolution estimates of tubulin-bound map derived from classes 1 and 10. D) Model of N- and C-terminal regions of β-tubulin built into substrate density with TBD region present to highlight region with missing density (grey).

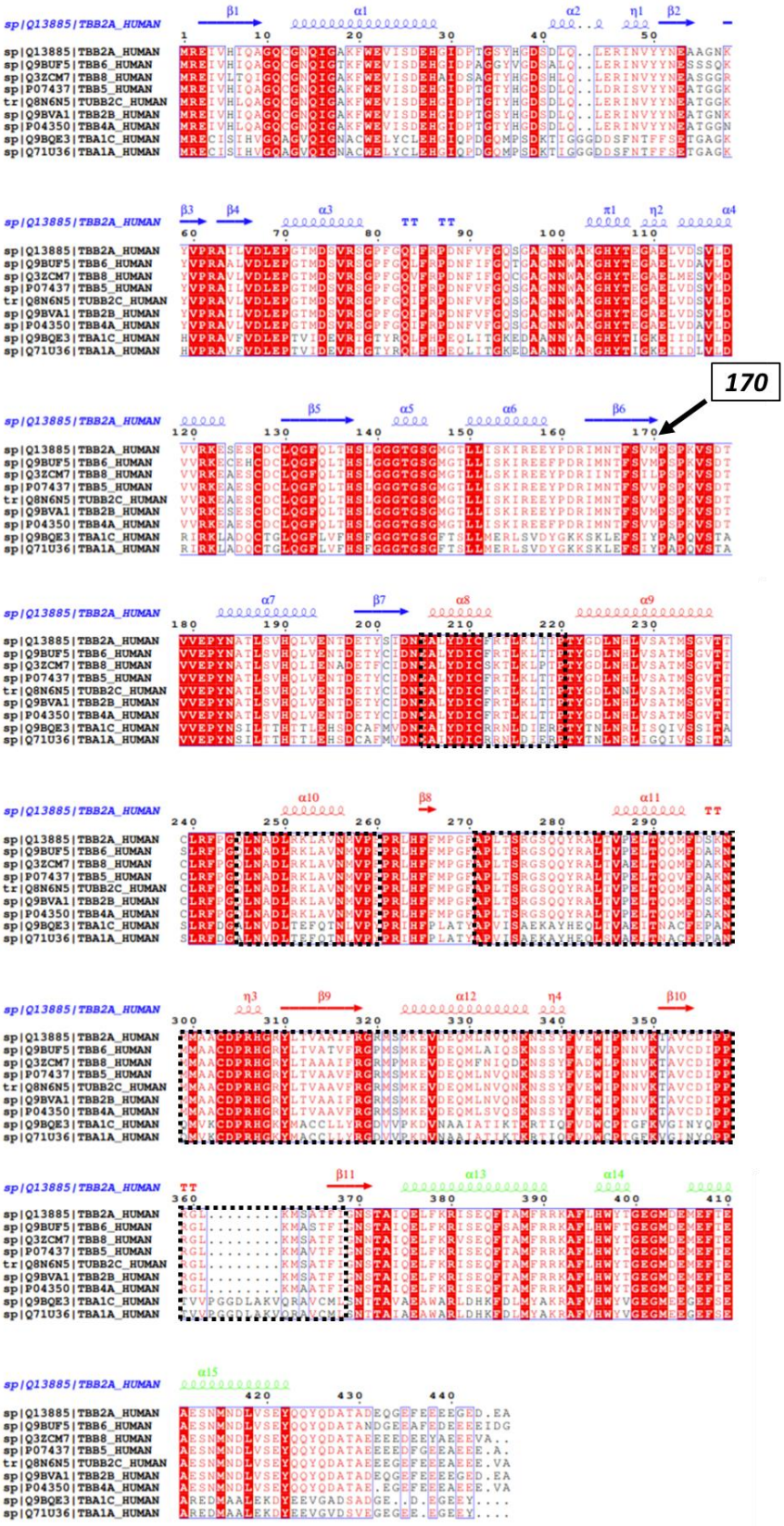
5.6.2 Identification of tubulin bound within folding cavity of TRiC in the closed state

Microtubules are assembled through the polymerisation of α- and β-tubulin dimers that bind head-to-tail into filaments. Both α- and β-tubulin isomers bind one molecule of GTP, with only the β-tubulin isoform able to exchange nucleotide, thus mediating assembly/disassembly of microtubule dynamics[243]. The γ-tubulin isoform is also abundant in eukaryotes and plays a role in α/β nucleation as well as centriole duplication[244].

The tubulin molecules contributing to map density are likely a mixture of isomers most likely consisting of α- and β-tubulin, as these were significantly enriched in LC-MS/MS analysis as opposed to γ-tubulin (Figure 5.12A, e.g. TUBB2A, TUBB6, TUBB8,

etc.). Nine of the top 25 most enriched (based on positive difference value) peptides included tubulin (seven β-tubulin and two α-tubulin), and thus it was highly likely that the tubulin mixture was enriched with β-tubulin isoforms (Figure 5.15A). A model was subsequently built based on tubulin beta-2A chain (TUBB2A), the highest enriched tubulin peptide (Figure 5.12A). TUBB2A fit well with the map density in all residue positions, except position 170 in which the valine from isoforms TUBB (listed as TBB5 on the sequence alignment, Figure 5.15A), TUBB2C, and TUBB4A was used in place of the methionine from TUBB2A (Figure 5.15B).

A)



B)

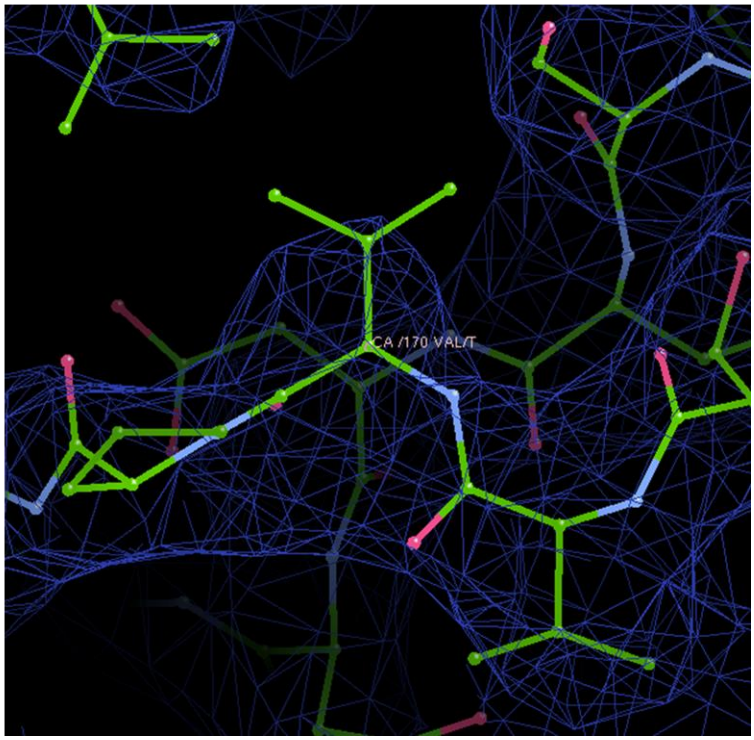


Figure 5.15: LC-MS/MS and structural analysis enable building of atomic model of tubulin. A) Sequence and secondary structure alignment between different α- and β-tubulin isoforms. Domains are colour coded by secondary structure based on α-tubulin, including N-terminal domain (blue), TBD (crimson), and C-terminal domain (dark green). Secondary structure assignment was obtained from cryo-EM structure of human β-tubulin at 3.6 Å resolution (PDB: 6I2I). Missing regions of density from TRiC-tubulin structure are indicated by dashed black boxes. Residue 170 that deviates from TUBB2A sequence in the density is labelled. B) Snapshot of tubulin residue 170 that fits better to valine of TUBB/TUBB2C/TUBB4, as opposed to methionine of TUBBA2.

Each tubulin monomer comprises an N-terminal nucleotide-binding Rossman domain (aa 1-205), an intermediate Taxol-binding domain (TBD; aa 206-381), and a C-terminal motor protein binding domain (aa 382 to C-terminus)[59]. The N- (aa 1-207) and C-terminal (aa 377-434) domains of β-tubulin are clearly defined in the density, and could be built to atomic resolution, along with a small number of residues from the TBD (Figure 5.16B: left). Compared to full-length native tubulin, most of the TBD domain is disordered (Figure 5.16B: right) including residues 208-221, 240-255, and 271-376 (Figure 5.15A). The truncated tubulin model interacts mainly with CCT3 and CCT6, and

to a lesser extent with CCT8 and CCT1 (Figure 5.16B). Of note, the CCT8/6/3 subunits were assigned in previous studies to the CCT6 hemisphere of the TRiC chamber with low rates of ATP binding and hydrolysis[147]. Superimposing native tubulin into the model shows that the missing TBD domain would be situated in the middle of the folding cavity (Figure 5.16B inset, grey ribbon). Interestingly, at a lower map isosurface threshold, density can be seen within the middle of the folding chamber expanding to adjacent subunits of the opposite hemisphere, suggesting that the TBD could interact with the C-termini of CCT1 and CCT2 (Figure 5.16B inset). Altogether, these data show the tubulin molecule has already achieved a near-native fold in the closed conformation and that the disordered TBD region is potentially loosely tethered to the opposite side of the folding chamber to stabilise its position. This could allow for specific folding of the TBD upon opening of the TRiC complex and completion of the folding cycle.

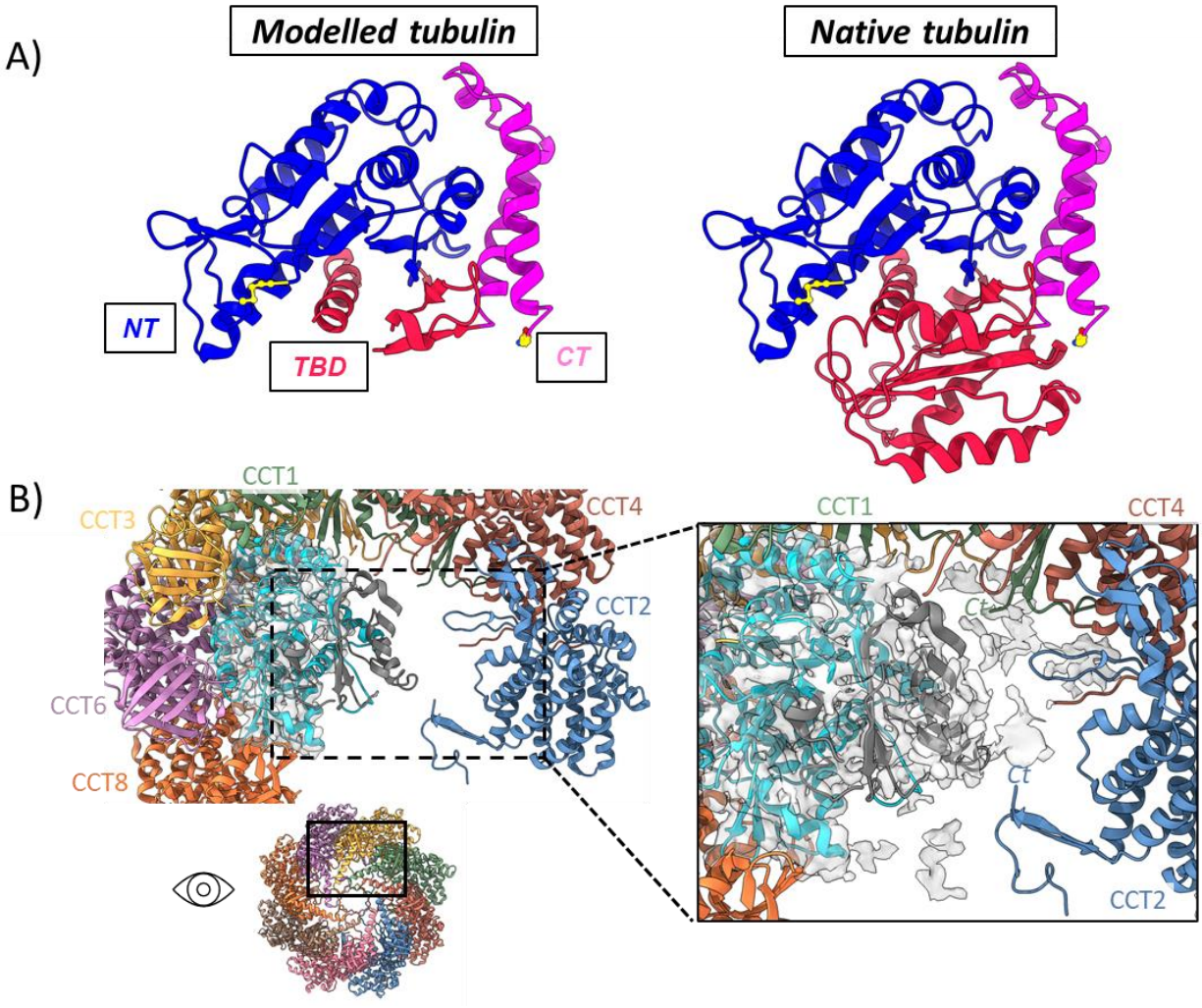


Figure 5.16: Tubulin model bound to TRiC. A) Tubulin model that can be built into the map density (left) compared to full beta-tubulin model (PDB: 6I2I; right). Domains are colour coded including N-terminal (blue), TBD (crimson), and C-terminal (magenta). N- and C-terminal residues (yellow) are rendered as ball-and-stick. B) Top-down slice view of tubulin density at high isosurface threshold within the context of TRiC with modelled tubulin (cyan) and missing TBD domain (grey). Inset: tubulin density at lower isosurface threshold showing density expanding towards the C-terminal domains of CCT1 (green) and CCT2 (blue).

5.6.3 Characterisation of tubulin bound within folding cavity of TRiC in the closed state

In general, the tubulin-CCT contacts are formed with the subunit apical and intermediate domains, with a few contributed from the stem-loops and C-termini.

Tubulin contacts several subunits, to varying degree, as described below using the *TUBB2A* sequence. Going around the TRiC ring in a clockwise manner viewed in top-down orientation, tubulin localises to CCT8, CCT6, CCT3, and CCT1 (Figure 5.17).

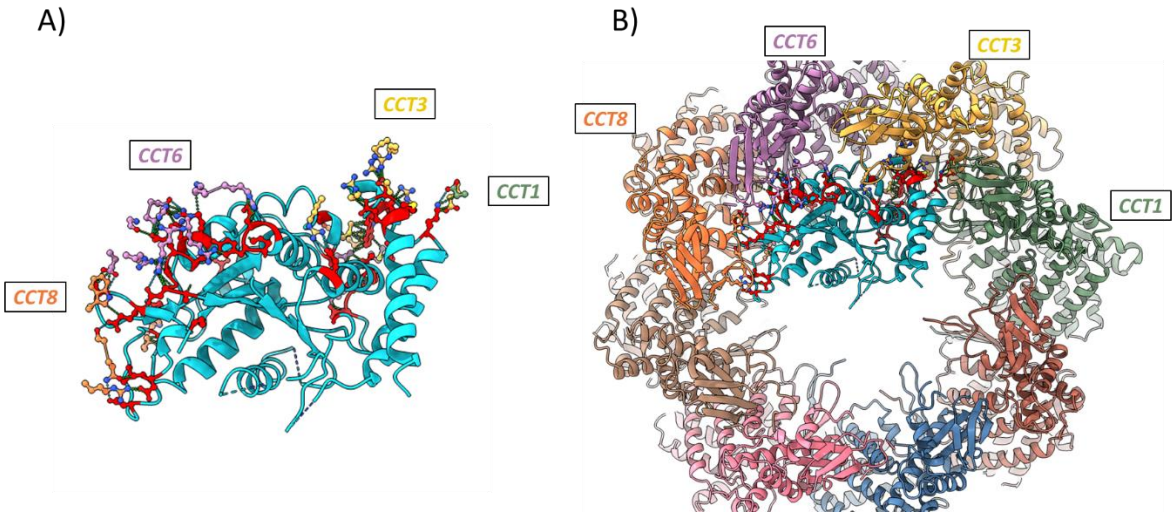


Figure 5.17: Overview of tubulin contacts with TRiC subunits. A) Truncated tubulin monomer modelled into density, with contacting residues of tubulin (red) with TRiC subunits represented as ball-and-sticks. B) Top-down view showing location of tubulin interactions within the TRiC folding chamber.

CCT8 interacts with tubulin modestly using the stem-loop (His59 and Leu60 of CCT8 with Pro32 and Glu83 of tubulin), intermediate domain (Arg314 and Arg316 of CCT8 with Asp45 and Leu44 of tubulin, respectively), and there is one contact between the apical domain helical protrusion loop (Glu252 of CCT8) and the N-terminus of tubulin (Arg2; Figure 5.18A). Of note, there is an interaction network between the intermediate domain of CCT8 (Asn316-Ser317-Trp319), the apical domain of CCT6 (Glu357), and an N-terminal domain loop of tubulin (Glu53-Ala55; Figure 5.18A: inset).

CCT6 participates the most in tubulin binding and the main interaction site is a negatively charged region of the beta-sheets of the CCT6 apical domain that sits just beneath the helical protrusion of the apical domain with helices α4 and α5 of tubulin

(Figure 5.18B: inset; Figure 5.15A). Two equatorial regions also interact with tubulin, including the stem-loop of CCT6 (Ala51) with Asp74 of tubulin, and, interestingly, the C-terminus of CCT6 (Figure 5.18B). The C-terminus of CCT6 is fully ordered in the tubulin-bound map up to Lys530 but can only be modelled up to residue Met526 in the consensus closed map, suggesting that the C-terminal extensions likely play a part in substrate binding as opposed to the role in inter-ring stacking seen for the N-terminal extensions. This stabilised C-terminal region of CCT6 interacts with N- and C-terminal tubulin domains including Met526 of CCT6 with Phe402/Trp405 of tubulin, Leu529 of CCT6 with Asn99//Val179/Val180/Trp405 of tubulin, and Lys530 of CCT6 with Asp177/Thr178 of tubulin, along with regions from CCT3 (Figure 5.18C: inset).

Tubulin contacts with CCT3 include the stem-loop (Pro53/Met54) with Trp405 of tubulin, and helix α3 of CCT3 (Arg68) with His404 of tubulin. Other interacting regions between CCT3 and tubulin include the CCT3 equatorial domain helix α3 (Glu89) with Lys400 of tubulin, and intermediate domain of CCT3 (Arg200/Arg316/Thr318/Arg322) with Asp412/Met414/Glu415 of tubulin (Figure 5.18C). CCT3 also interacts in two locations with both tubulin and CCT1: the apical lid loops of both CCT3 (Gly250 of CCT3 with Pro261 of tubulin) and CCT1 (Lys245 of CCT3 with Phe260 of tubulin) along with helix α3 of CCT3 (Glu89 of CCT3 with Lys400 of tubulin) and the stem-loop of CCT1 (Asp358 of CCT1 with Lys400 of tubulin; Figure 5.18D: insets).

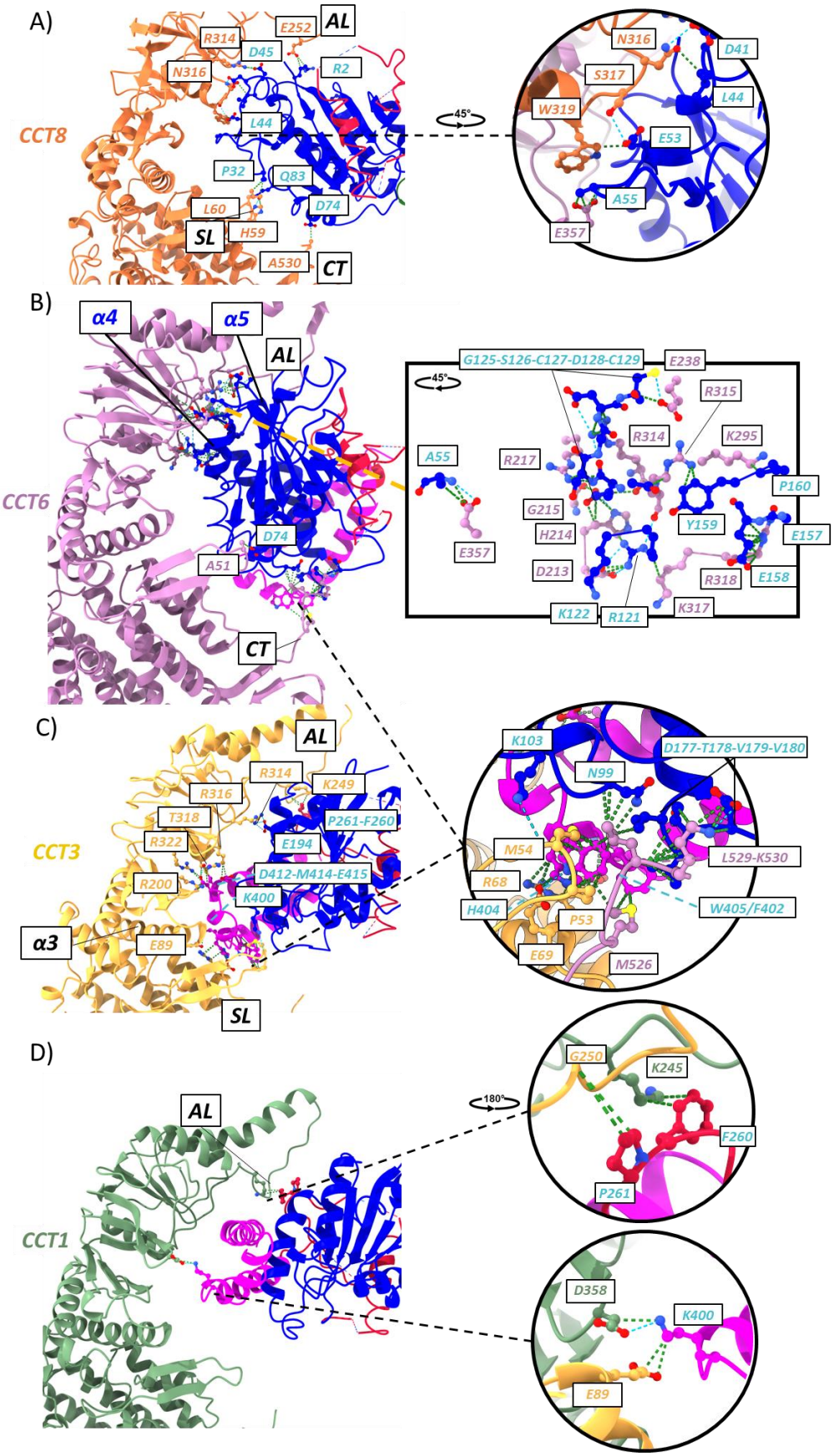

A)
CCT8
E252
AL
R314
D45
N316
R2
L44
P32
Q83
L60
D74
SL
H59
A530
CT
45°
N316
D41
S317
L44
W319
E53
A55
E357
B)
α4
α5
AL
CCT6
D74
A51
CT
G125-S126-C127-D128-C129
E238
45°
R315
R314
K295
A55
R217
P160
G215
Y159
E157
H214
E357
D213
R318
E158
K122
R121
K317
C)
AL
CCT3
R314
K249
R316
T318
P261-F260
R322
E194
D412-M414-E415
R200
K400
α3
E89
SL
K103
D177-T178-V179-V180
N99
M54
L529-K530
R68
H404
W405/F402
P53
F69
M526
D)
AL
CCT1
180°
G250
K245
F260
P261
D358
K400
E89

Figure 5.18: Contacts between CCT subunits and tubulin within closed conformation. A) Side view of CCT8 (orange) binding to tubulin. Tubulin is coloured by domain including N-term domain (blue), TBD (crimson), and C-term domain (magenta). Residues are labelled by CCT colour scheme or in cyan to represent tubulin. Inset (rotated 45°) shows network of interactions between CCT8 (orange) and CCT6 (purple). Green dashes indicate van der Waals contacts; blue dashes indicate hydrogen bonds. Structural motifs are labelled as apical loop (AL), stem-loop (SL), C-terminus (CT) and helix X (αX). B) Side view of CCT6 binding to tubulin. Inset shows interaction network between tubulin N-term domain and CCT6 apical domain residues displayed as atoms. C) Side view of CCT3 (yellow) binding to tubulin. Inset shows interaction between C-terminus of CCT6, stem-loop of CCT3, and tubulin regions ranging from the N-term domain and C-term domain. D) Side view of CCT1 (green) binding to tubulin. Inset (top; rotated 180°) shows interactions between TBD and C-term tubulin domains, CCT3 apical domain, and CCT1 apical domain. Inset (bottom) shows interactions between C-term tubulin domain with intermediate domains of CCT1 and CCT3.

5.7 Relaxed 3D classifications yield further substrate-bound reconstructions.

To improve the density of other substrate-bound maps for the TRiC closed conformation, another 3D classification was performed on the consensus particles, this time with a new option allowing to relax the symmetry and that was not available at the time of the initial 3D classification runs. Here this option was used to consider the particles in both up and down orientations at each iteration[189]. This yielded five new classes, some of which had a similar density profile original 3D classification run (Figure 5.19A). Class 1 represented the same density seen for the tubulin-bound map described in the previous section (Figure 5.19B: class 1). Class 2 contained density in both the *cis* and *trans* rings, neither of which could be fit to represent tubulin (Figure 5.19B: class 2), and thus likely represented a different substrate. Class 3 was very similar to class 1 except for a small region of extra density on the opposite/*cis* side of the folding chamber that was of insufficient resolution to be modelled (Figure 5.19B: class 3). This density lies close to the septum region created by the stacking of CCT2-CCT2’ and CCT4-CCT5’ at the ring interface and could be an anchor point for the disordered TBD in this set of particles, however this remains unconfirmed. Classes 4 and 5 contained weaker density compared to the rest of the classes; substrate density

in class 4 is almost non-existent (Figure 5.19B: class 4), while class 5 shows very weak density in the approximate region of tubulin and is likely a sum of other less abundant substrates (Figure 5.19B: class 5).

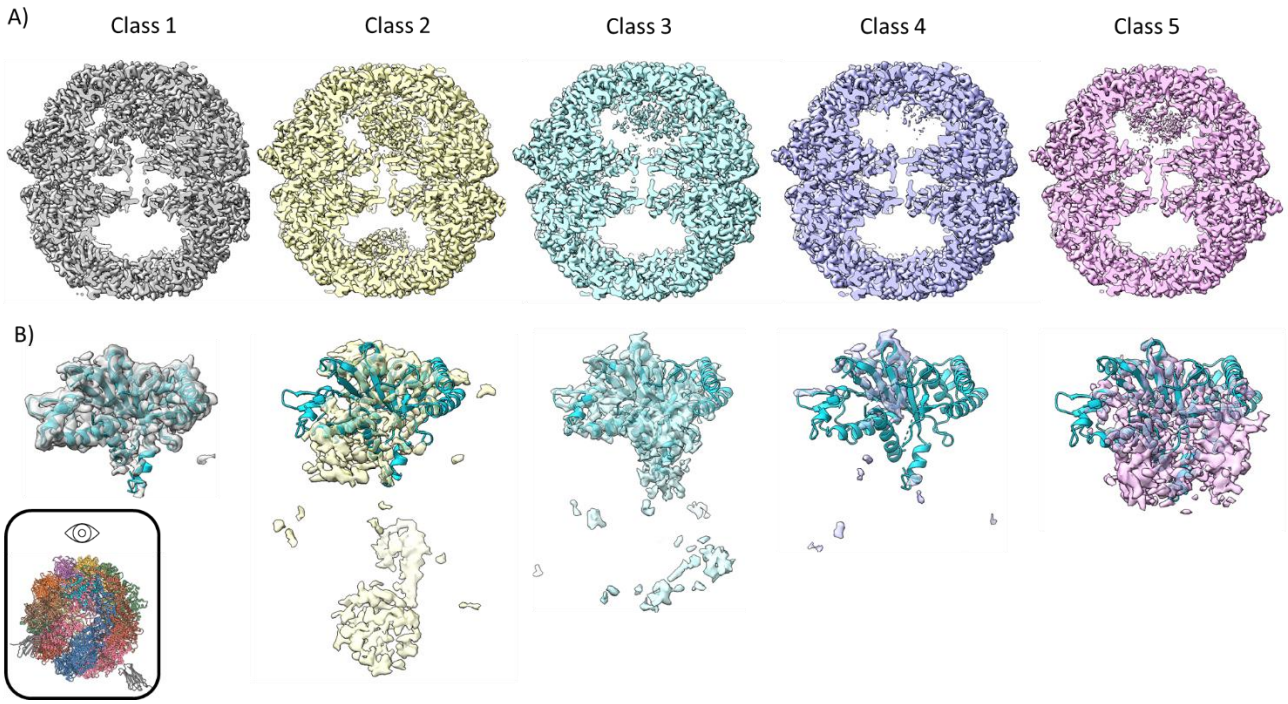


Figure 5.19: Relaxed classifications of closed particles. A) Side view slice of maps from classes 1-5 of relaxed 3D classifications, showing varying substrate densities. B) Tilted view of substrate densities (view indicated in inset) for classes 1-5, overlayed with tubulin model built from initial 3D classifications to indicate fit/misfit.

5.7.1 Actin bound to the *cis* ring of TRiC in the closed state

The density from the relaxed class 2 map was the most intriguing of the five maps generated, as there was clear evidence of substrates bound within both chambers, however neither appeared to be tubulin based on shape fitting (Figure 5.19B). In one chamber, the density showed one large substrate in the same general location as tubulin near subunits CCT8-CCT6-CCT3 (Figure 5.20A). The substrate bound was identified as actin; although actin was not highly enriched in the LC-MS/MS

data (Figure 5.12B), it is one of the canonical TRiC substrates in addition to tubulin and fit well into the map density.

LC-MS/MS data identified two actin peptides: one that could not distinguish between α-skeletal muscle, α-cardiac muscle 1, aortic smooth muscle, and γ-enteric smooth muscle, and cytoplasmic 1 isoforms (ACTA1;ACTC1;ACTA2;ACTG2;ACTB), and one that identified cytoplasmic 2 actin (ACTG1). Side chain analysis of the density map determined that cytoplasmic actin 1 and 2 (ACTB and ACTG1) were the most likely isoforms present in the sample, given that the HEK293T cell line used for endogenous purification is not likely to contain muscle actin isoforms. The sequences of β-actin (ACTB) and γ-actin (ACTG1) only differ in their first ten residues, most of which are disordered in the map density. Altogether, ACTB was therefore used to build almost an entire atomic model of actin (Figure 5.20A: right), though the sample is likely to be a mixture between both isoforms of cytoplasmic actin.

An actin monomer is divided into two domains (small and large) that in turn have two subdomains each, termed subdomains 1-4 (Figure 1.3). The small domain is divided into subdomain 1 and subdomain 2. Subdomain 1 contains the N- and C-termini including residues 1–32, 70–144, and 338–374, while subdomain 2 is the smallest consisting of residues 33–69. The large domain is divided into subdomain 3 including residues 145–180 and 270–337 and subdomain 4 ranging residues 181–269. The surface formed by subdomains 2 and 4 are polar, are referred to in the literature as the “pointed” end of the actin monomer and are separated by a cleft that leads to an ATP binding site in the middle of the monomer. The surface formed by subdomains 1 and 3 are positively charged and known as the “barbed” end[57] (Figure 5.20B).

The density seen in the actin-bound map details a nearly folded molecule of actin. All subdomains can be seen in the density, although the density for the entire subdomain 4 is weaker than the other subdomains 1-3 (Figure 5.20A). To this end, only a C-α trace is built for much of subdomain 4 from residues 177 to 272. However, the only missing residues that were not built at all into the actin model are residues 35-52 in subdomain 2 (Figure 5.20B), which corresponds to a region known as the DNase I binding loop (D-loop). This loop was shown to form an α-helix upon binding of ATP and was shown to be ordered when in the G-actin polymer, but disordered when in the F-actin monomer state[62,245].

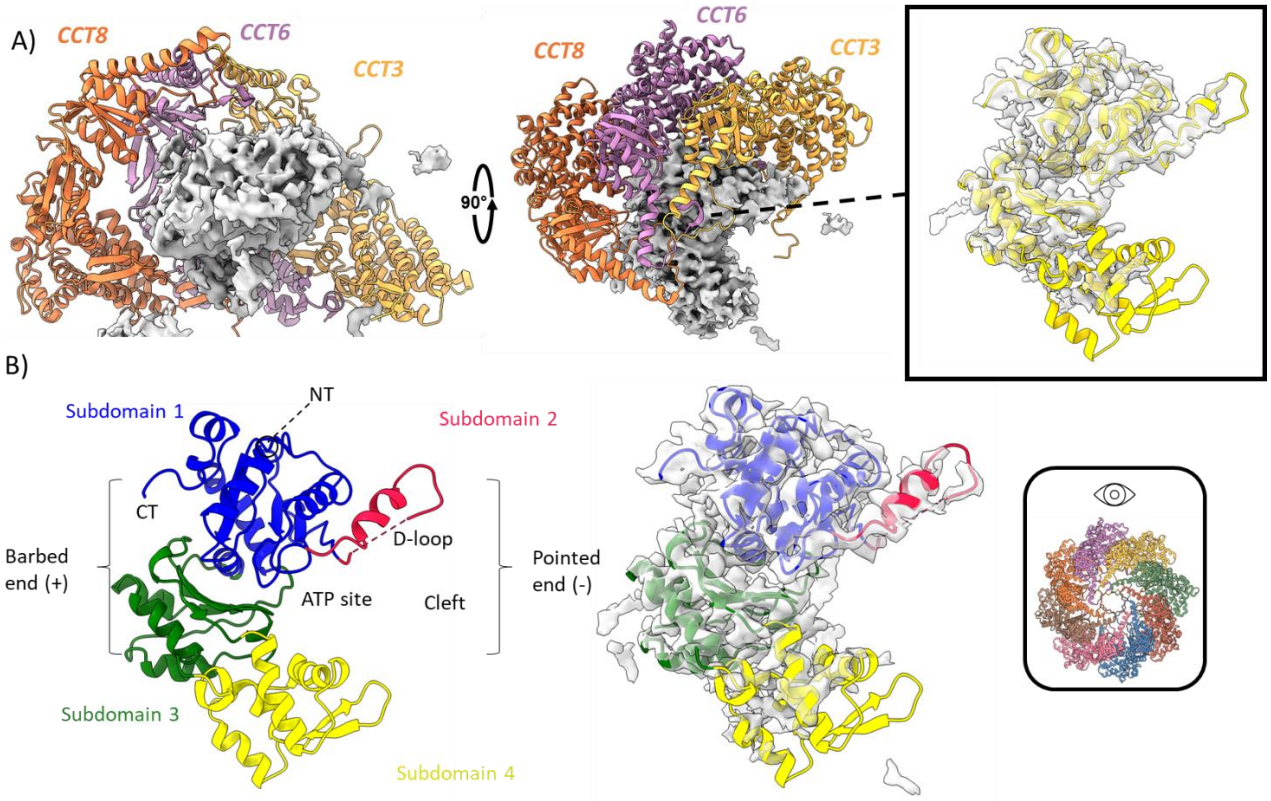


Figure 5.20: Actin-bound map in the closed conformation. A) Location of actin binding within the TRiC folding chamber in side (left) and top-down (middle) views relative to subunits CCT8/6/3. Inset: actin (gold) model fit within substrate density. B) Top-down view of actin model built from substrate labelled with actin structural motifs including subdomains 1 (blue), 2 (crimson), 3 (dark green), and 4 (yellow) without (left) and with map density as grey surface (right).

5.7.2 Actin interactions with TRiC in the closed state

Actin binds in a similar region to tubulin around the CCT6 hemisphere but differs in that the primary subunit contacts are CCT6/3/1 (Figure 5.20A), with only a few contacts from CCT8. This is unlike for tubulin where primary contacts were observed near CCT8/6/3. Additionally, actin interacts with CCT7 and CCT4, two subunits that are not involved in the tubulin binding mode (Figure 5.16; Figure 5.21). The actin–TRiC interactions are mediated in several actin locations: subdomain 1 localises to CCT3 and CCT6, subdomain 2 inserts into a groove of CCT1, subdomain 3 interacts with CCT7 and CCT8, and the poorly defined density of subdomain 4 appears to have a single contact with CCT4 though this was difficult to validate due to limitations in resolution in this region (Figure 5.21).

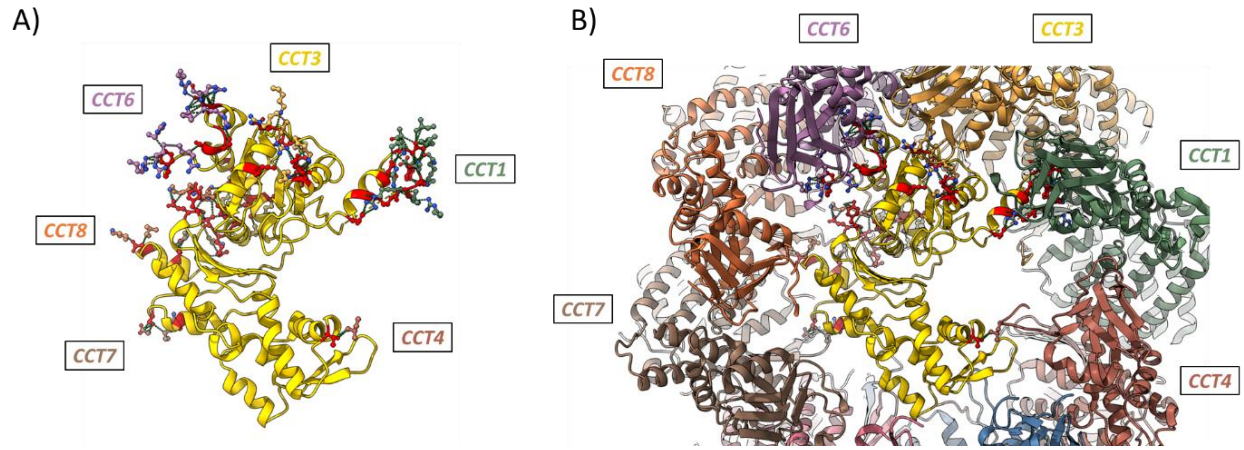


Figure 5.21: Actin-TRiC binding locations. A) Actin model (gold) in cartoon representation, displayed with actin residues (red sticks) that interact with TRiC subunit residues (sticks colour-coded with subunits). B) Top-down view showing location of actin interactions within TRiC folding chamber.

More specific interactions for individual subunits are described below in clockwise TRiC subunit orientation using the *ACTB* sequence for actin.

CCT8 interacts with subdomain 3 through its stem-loop (Leu60/Lys62) with Asp288 of actin (Figure 5.22: left), and its C-terminus (Figure 5.22: inset). The C-terminal binding includes the following contacts with actin: Gly531/Gly532/Pro533 of CCT8 contact Leu349 in subdomain 1 of actin, and Pro533 of CCT8 additionally contacts Tyr143 from subdomain 1 as well as Glu167 from subdomain 3. Lys534 of CCT8 also interacts with subdomain 3 including residues Glu167 and Tyr169. Interestingly, the C-terminus of CCT7 is also involved in coordinating actin in this region (Figure 5.22: inset, Figure 5.26).

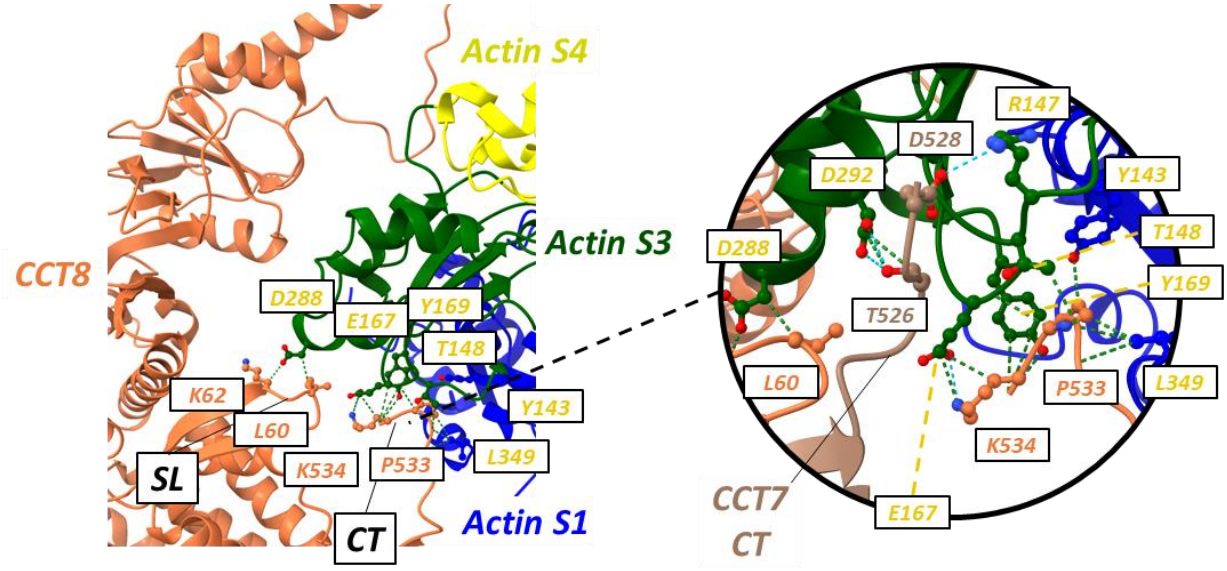


Figure 5.22: Interactions between actin and CCT8. Side view of CCT8 (orange) binding to actin. Actin is coloured by domain including subdomain 1 (S1, blue), subdomain 3 (S3, dark green), and subdomain 4 (S4, yellow). Residues are labelled by CCT colour or in gold to represent actin. Green dashes indicate van der Waals contacts; blue dashes indicate hydrogen bonds. Structural motifs are labelled as stem-loop (SL) and C-terminus (CT).

CCT6 interacts exclusively with the C-terminus of actin in subdomain 1 through the same interface in the intermediate domain as the tubulin-bound model (Figure 5.18B; Figure 5.23). This includes a positively charged region of CCT6 containing Arg314,

Arg217, and His214 that bind to the C-terminal Phe375 residue of actin. Additionally, contacts are formed between Arg315 of CCT6 and Ser368 of actin, Arg318 of CCT6 and Gly366/Pro367 of actin (Figure 5.23: inset top), and interactions are observed between Glu364 of actin and Lys317/Arg319/Arg323 of CCT6 (Figure 5.23: inset bottom).

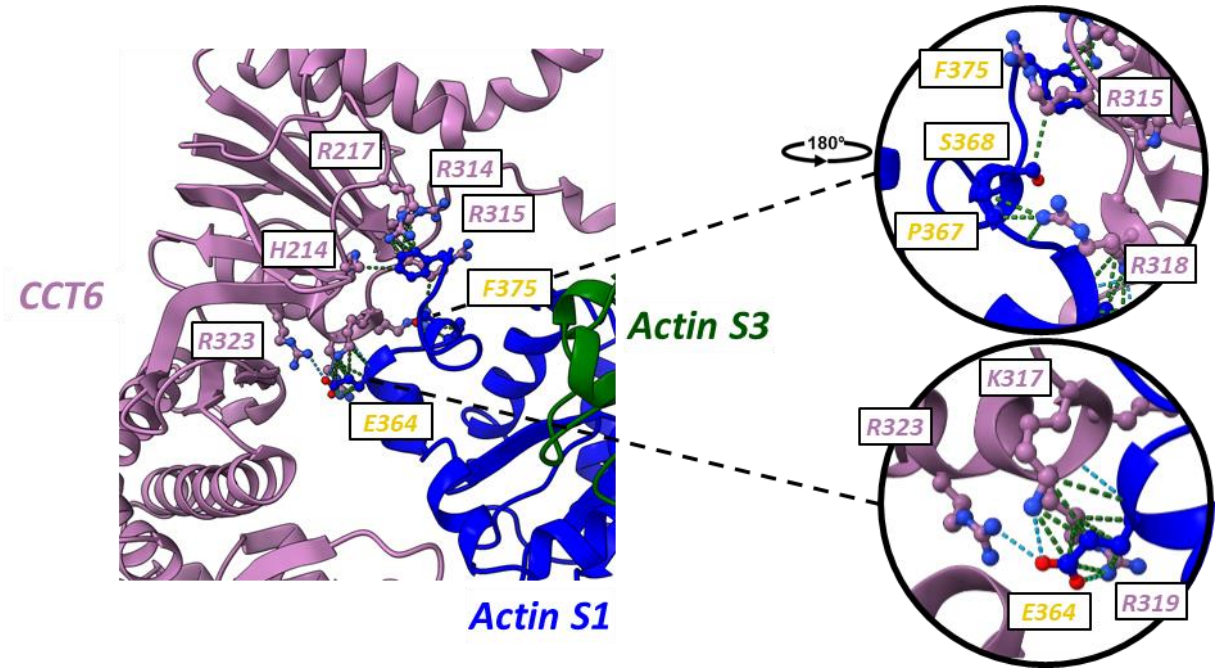


Figure 5.23: Interactions between actin and CCT6. Side view of CCT6 (purple) binding to actin. Actin is coloured by domain including subdomain 1 (S1, blue), subdomain 3 (S3, dark green). Residues are labelled by CCT colour or in gold to represent actin. Green dashes indicate van der Waals contacts; blue dashes indicate hydrogen bonds. Inset (top; rotated 180°) shows zoomed in interactions near C-terminus of actin. Inset (bottom) shows residue contacts between actin residue Glu364 and CCT6.

CCT3 interacts with actin subdomain 1, through its stem-loop (Met54 of CCT3 with Phe21/His101 of actin) and apical domain. The apical domain interactions consist of a similar positively charged interface to CCT6 and includes Lys222/Arg313 of CCT3 with Glu125 of actin, Arg314 of CCT3 with Lys118/Glu83 of actin, and Lys248/Lys294 of CCT3 with Asp80 of actin (Figure 5.24).

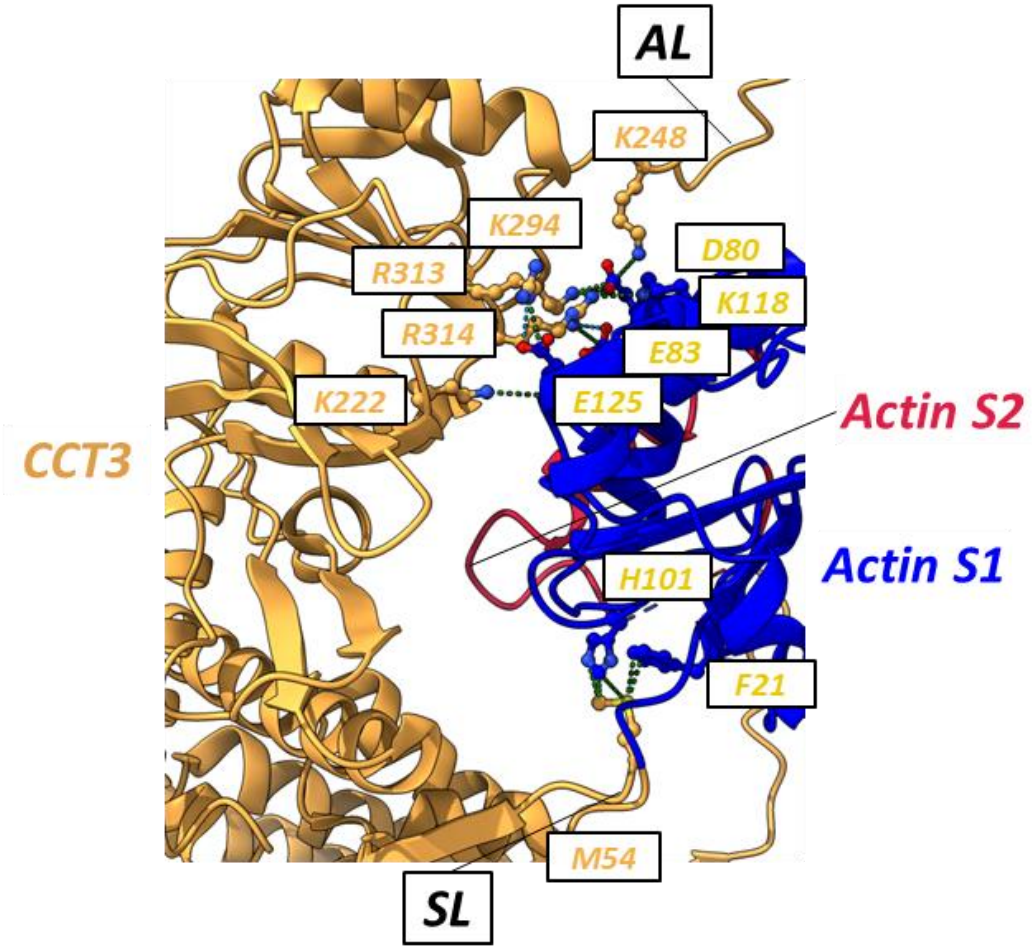


Figure 5.24: Interactions between actin and CCT3. Side view of CCT3 (yellow) binding to actin. Actin is coloured by domain including subdomain 1 (S1, blue) and subdomain 2 (S2, crimson). Residues are labelled by CCT colour or in gold to represent actin. Green dashes indicate van der Waals contacts; blue dashes indicate hydrogen bonds. Structural motifs are labelled as apical domain loop (AL) and stem-loop (SL).

CCT1 interacts with the main region of actin subdomain 2 at the site of the disordered D-loop that plays a role in ATP binding (Figure 5.25A). The interaction between CCT1 and actin spans residues Asp51 to Thr66 of actin (Figure 5.25B). This region interacts with a groove in the intermediate/lower apical domain of CCT1 and includes 12 CCT1 residue contacts and nine actin subdomain 2 residue contacts (Figure 5.25B: insets), suggesting CCT1 plays an important role in sequestering this disordered region during actin folding.

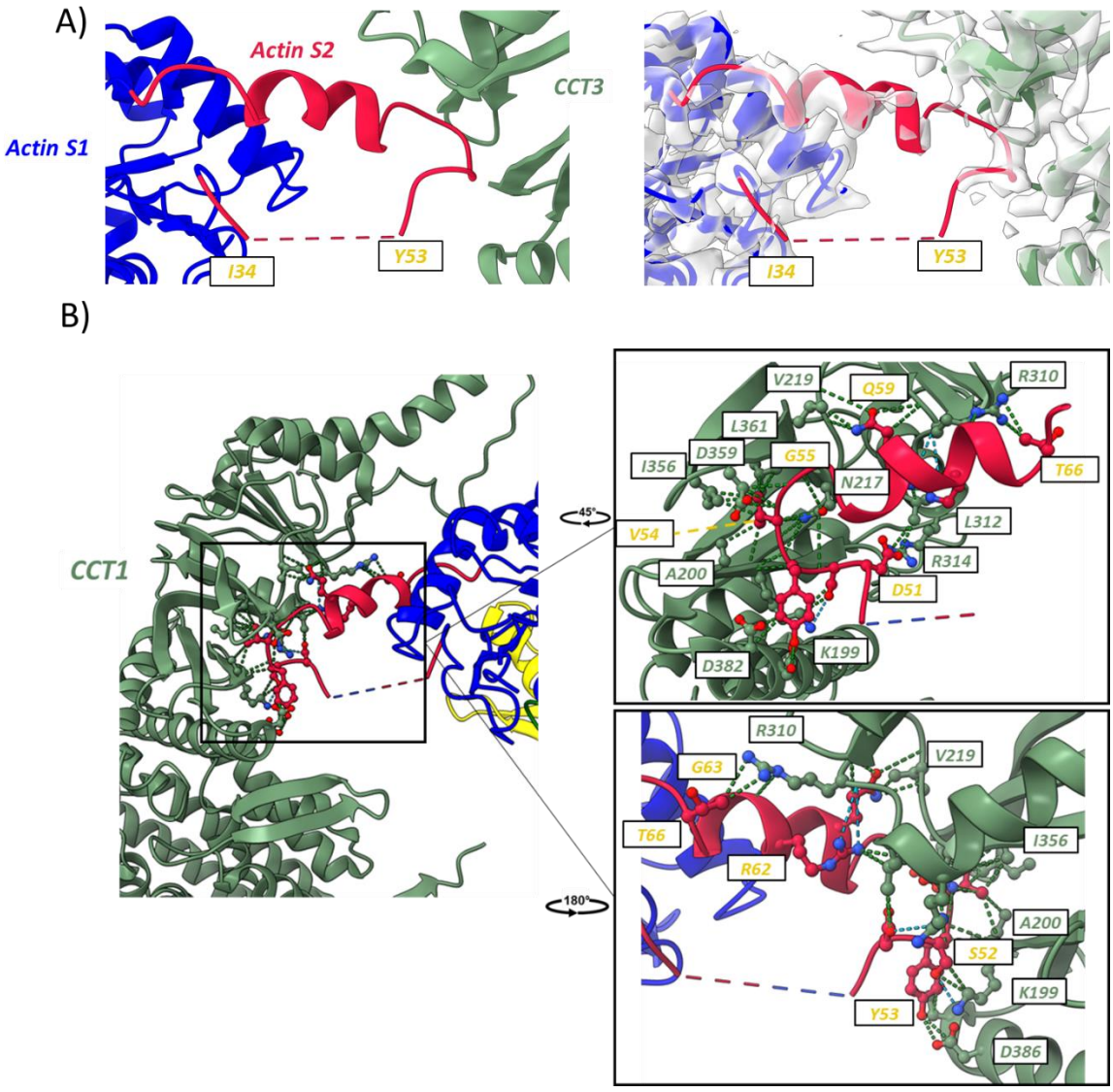


Figure 5.25: Interactions between actin and CCT1. A) Zoomed in view of subdomain 2 binding within groove formed by CCT1 (green) and missing density between residues 34-53 without (left) and with map (right). B) Side view of CCT1 (green) binding to actin. Actin is coloured by domains including subdomain 1 (S1, blue) and subdomain 2 (S2, crimson). Residues are labelled by CCT colour or in gold to represent actin. Green dashes indicate van der Waals contacts; blue dashes indicate hydrogen bonds. Insets (top rotated 45°; bottom rotated 180°) show actin subdomain 2 helix binding within groove of CCT1

While interactions between CCT4 and actin were difficult to determine due to the weak density exhibited in subdomain 4 of actin, Ile304 and Leu305 of CCT4 appeared to engage the density near subdomain 4 (Figure 5.26A). This is the only region of CCT4 that could interact with actin and suggests that subdomain 4 is not well ordered and only

associates loosely with CCT4 to position the actin monomer for subsequent folding. CCT7 makes several contacts with actin subdomain 3 including the stem-loop (Asp51 of CCT7 with Ser232 of actin, and Arg53 of CCT7 with Thr318 of actin) and through its C-terminus (Thr526 of CCT7 with Asp292 of actin, and Asp528 of CCT7 with Arg147 of actin; Figure 5.26B), in the same regions as that of CCT8 (Figure 5.22). Taken together, these interactions display an actin monomer in its near-native fold that is coordinated by specific interactions with the CCT6 hemisphere of TRiC in the closed state.

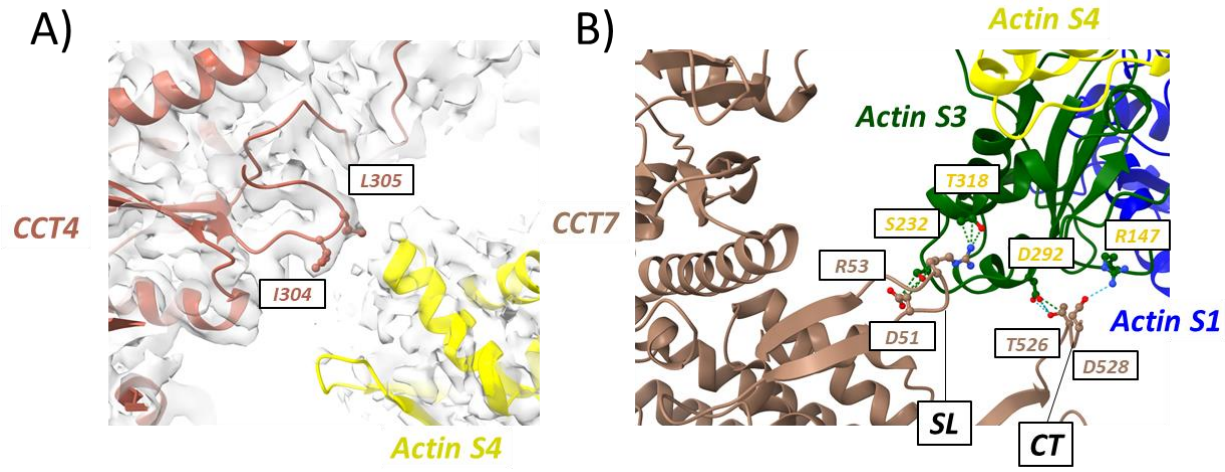


Figure 5.26: Interactions between actin and CCT4 and CCT7. Side view of A) CCT4 (dark red) and B) CCT7 (brown) binding to actin. Actin is coloured by domains including subdomain 1 (blue), subdomain 3 (dark green), and subdomain 4 (yellow). Residues are labelled by CCT colour or in gold to represent actin. Green dashes indicate van der Waals contacts; blue dashes indicate hydrogen bonds. Structural motifs are labelled as stem-loop (SL) and C-terminus (CT).

5.8 Co-chaperone PhLP2A is bound in the *trans* ring

The substrate density from relaxed class 2 was of sufficient resolution to place actin in the *cis* ring, however unmodelled density remained in the *trans* ring (Figure 5.19A and B: class 2; Figure 5.27A). Additionally, actin did not account for all the substrate density in the *cis* ring, as there remained density beneath subdomains 1 and 3

(Figure 5.27B). This density was well-ordered and appeared to be a short helix from an unidentified protein.

The density in the *trans* ring was smaller than the dimensions of actin but far too large to be accounted for by any extension of CCT termini (Figure 5.27C). This density appeared to contain several β-strands and α-helices that aided identification of a potential second protein bound within the TRiC chaperonin. At a low isosurface threshold, this density in the *trans* cavity appeared to span the entire TRiC folding cavity, passing through the septum created by the equatorial domains of each TRiC ring and ending in the *cis* ring near the location of actin (Figure 5.27D). Taken together, these observations supported the possibility of an additional substrate or binding partner of actin. The enrichment seen in LC-MS/MS data (Figure 5.12A) provided several possible substrates to investigate including immunoglobulin-binding protein 1, programmed cell death protein 5, and phosducin-like protein 3 (PhLP2A). Of these, PhLP2A was determined to be the second substrate bound within the *trans* TRiC cavity in the closed state, as discussed below.

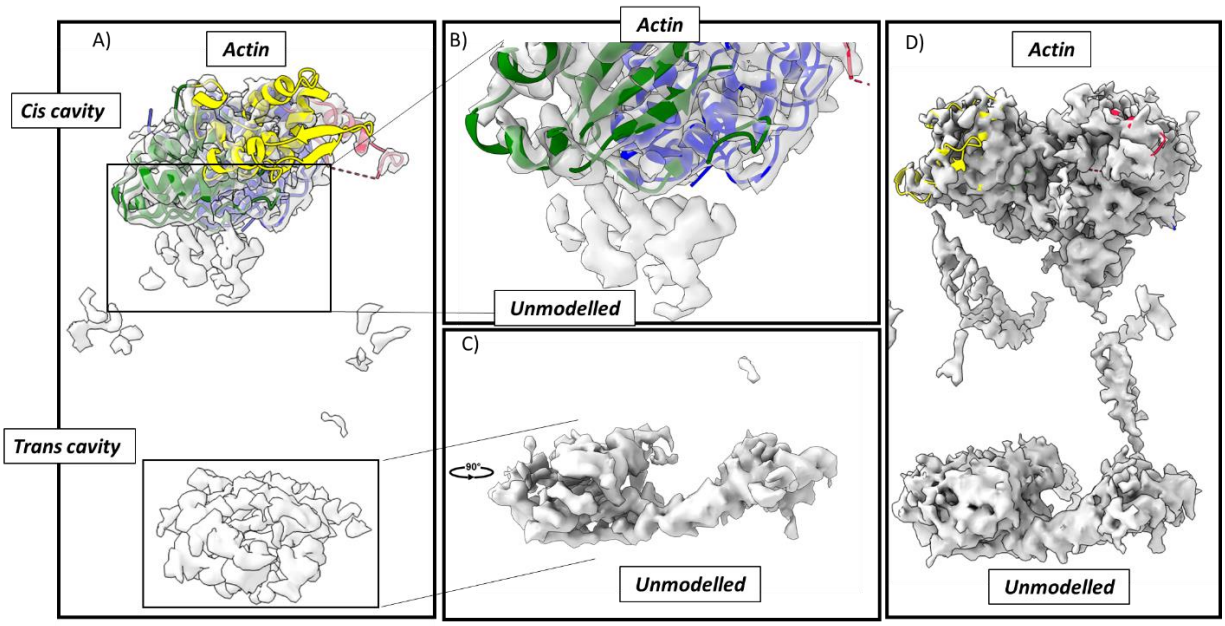


Figure 5.27: Relaxed class 2 closed map remaining unmodelled density. A) Substrate density including unmodelled density and actin model coloured by domain including subdomain 1 (blue), subdomain 2 (crimson), subdomain 3 (dark green), and subdomain 4 (yellow). B) Zoom-in view of substrate density that cannot be accounted for by actin. C) Density in trans ring that cannot be accounted for by TRiC subunits. D) Density at lower isosurface threshold reveals further density spanning the two TRiC folding chambers.

5.8.1 Identification of PhLP2A as the second substrate bound within TRiC

In the human genome, the phosducin-like protein family consists of phosducin-like protein (PhLP1), phosducin-like protein 2 (PhLP2B, *PDCL2* gene), phosducin-like protein 3 (PhLP2A, *PDCL3* gene), and thioredoxin domain-containing protein 9 (TXND9), share ~25% sequence similarity among themselves (Figure 5.28). Several of these family members have been shown to play a regulatory role in TRiC-mediated folding, including PhLP1 and PhLP2A/B, which have been shown to mediate assembly of the G protein βγ dimer[74] and actin and tubulin[75], respectively (Section 1.2.2). Importantly, PhLP1, PhLP2A, and TXND9 (but not PhLP2B) were detected in the LC-

MS/MS analysis of the as-purified endogenous TRiC sample (Table 5.2), and among them PhLP2A was significantly enriched (Figure 5.12A).

The phosducin-like family proteins contain an N-terminal domain, a conserved thioredoxin fold, and a short acidic C-terminus (Figure 5.19A). Given PhLP2A/B have been shown to bind actin[75] and that PhLP2B was not observed in the LC-MS/MS data, it was suspected that the unidentified actin binding protein in class 2 was PhLP2A.

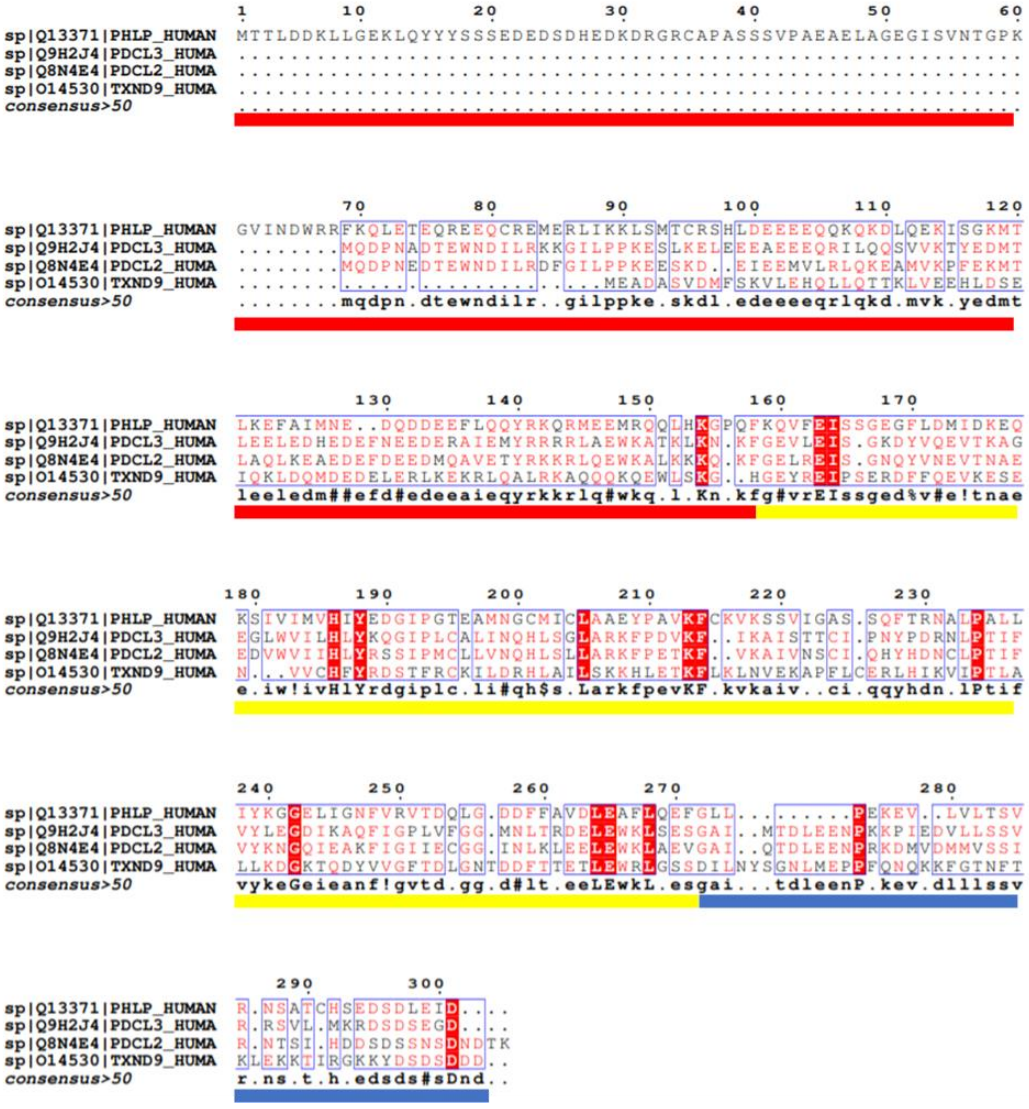


Figure 5.28: Sequence alignment of PhLP family. Domains are colour-coded by bars underneath alignment including N-terminal (red), thioredoxin (yellow), and C-terminal (blue) domains. Residue numbering is from *Phlp*.

The thioredoxin domain of PhLP2B was previously crystallised[246] and appeared to be similar in shape and fold to the unmodelled density in the trans TRiC ring in the relaxed class 2 closed map (Figure 5.29A). To date, no PhLP domains beyond the thioredoxin domain have been characterised structurally. Secondary structure predictions (Figure 5.29B) hint that the region immediately preceding the thioredoxin domain (aa 33-88) contains several long α-helical stretches.

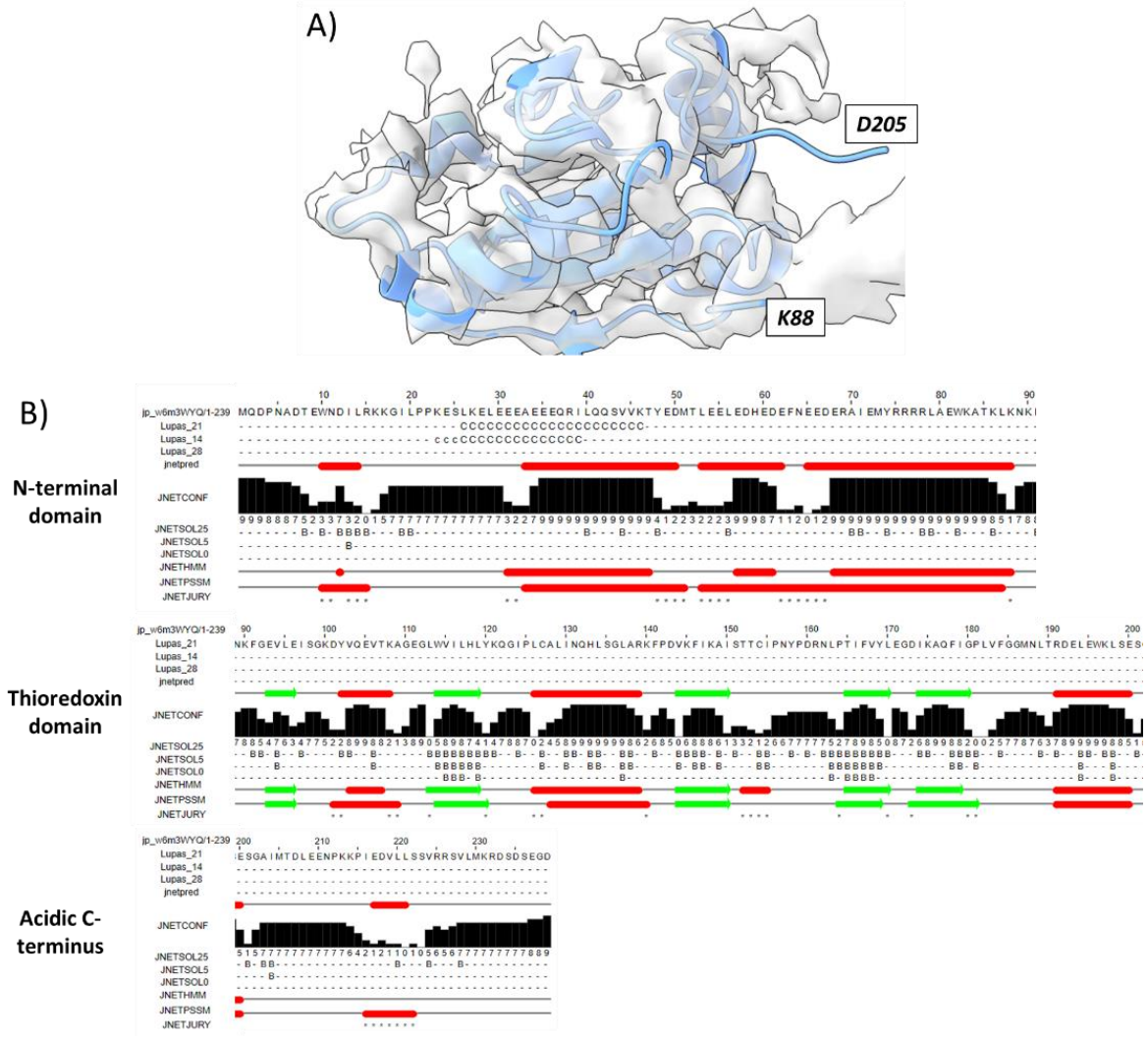


Figure 5.29: Structure-based predictions of PhLP2A as unidentified substrate bound in *trans* ring of TRiC in relaxed class 2 reconstruction. A) Crystal structure of 2.7 Å human PhLP2B[246] (PDB: 3EVI; blue) thioredoxin domain (aa 88-205) docked into relaxed class 2 *trans* ring substrate density. B) Secondary structure prediction of PhLP2A separated by domain, using JPred4[247]. Red bars indicate α-helix, green arrows indicate β-sheets.

This prediction is consistent with the residual protein density seen adjacent to the thioredoxin fold and is also the approximate shape and length to account for two long cylinders of helical density, one across the *trans* ring (H3) and the other (H2) spanning the two TRiC chambers (Figure 5.30A). Unfortunately, the density observed in these regions was too weak to definitively build an atomic model of H2 and H3, and instead helices representing these regions were roughly docked within the density. More convincingly, the N-terminus of PhLP2A (aa ~10-15) was predicted to contain a short helix (H1) that could be built into the density directly below actin (Figure 5.30B). Taken together, these data represent a ternary complex of TRiC–actin–PhLP2A, whereby PhLP2A is directly bound to actin within the TRiC chaperonin (Figure 5.30C and D).

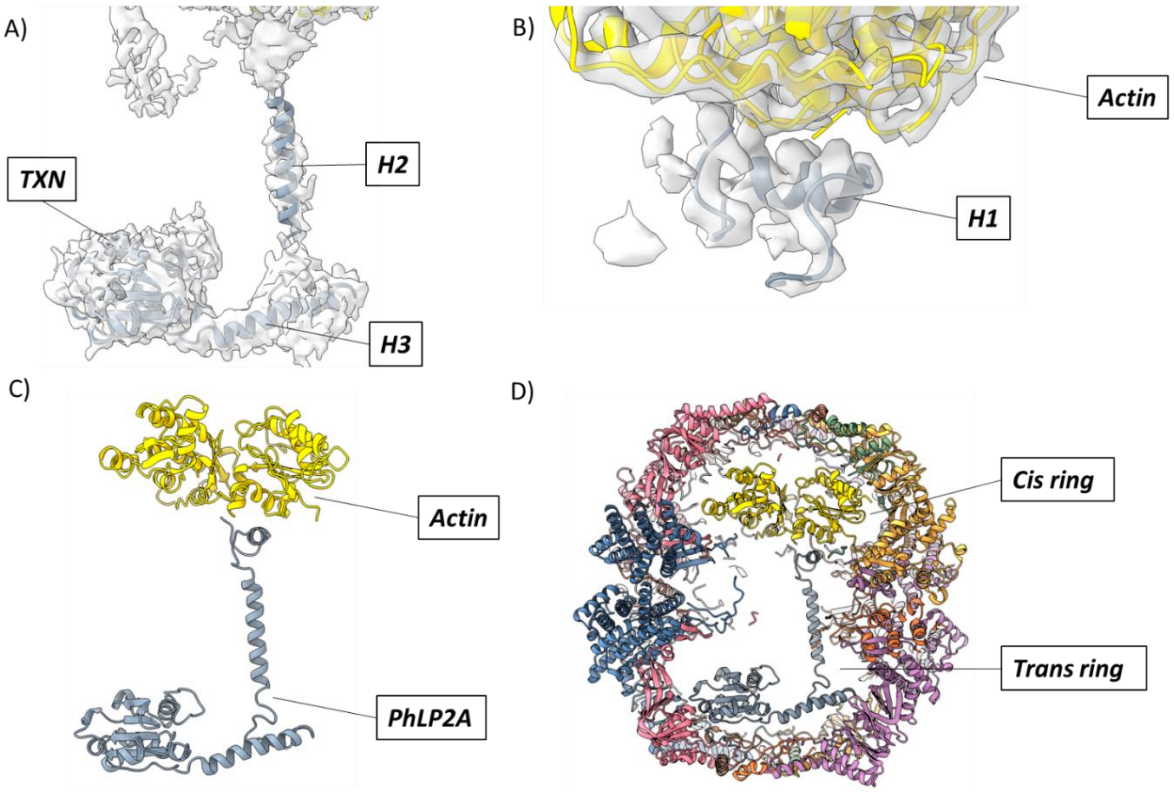


Figure 5.30: PhLP2A identified as second substrate in relaxed class 2 closed map. A) Substrate density at low isosurface threshold with representative PhLP2A helices 2 and 3 (H2 and H3) built into weak density according to secondary structure predictions along with thioredoxin (TXN) domain. B) N-terminal helix 1 (H1) of PhLP2A binding actin (gold). C) Model of actin bound to PhLP2A (grey). D) Model of actin in the *cis* ring bound to PhLP2A in the *cis* and *trans* rings within the TRiC complex.

5.8.2 Helical regions of PhLP2A hint at actin-regulating mechanism

The N-terminal helix H1 of PhLP2A interacts with actin at the interface between subdomains 1 and 3 (Figure 5.30B; Figure 5.31A). This interaction occurs primarily through Trp10 of PhLP2A with Ile345 of subdomain 1 and Thr148 of subdomain 3 (Figure 5.31A). Additional contacts from subdomain 1 of actin include Ile345 of actin with Ile13/Ile19/Trp10 of PhLP2A, Gly23/Ser344 of actin with Lys17 of PhLP2A, Ser 348 of actin with Ile13 of PhLP2A, and Leu349 of actin with Glu9/Ile13 of PhLP2A. Subdomain 2 residue contacts include Arg147 of actin with Gln2 of PhLP2A, and Thr148 of actin with Asp3/Trp10 of PhLP2A. This network of interactions between actin and PhLP2A is further stabilised by the C-termini of CCT7 and CCT8 (Figure 5.31B). The N-terminal helix 1 of PhLP2A also differs in sequence from that of the other PhLP-family member, PhLP, that has been identified as a TRiC co-chaperone (Figure 5.28), supporting previous observations that PhLP associates with Gβ and not with actin[72–75,89,169].

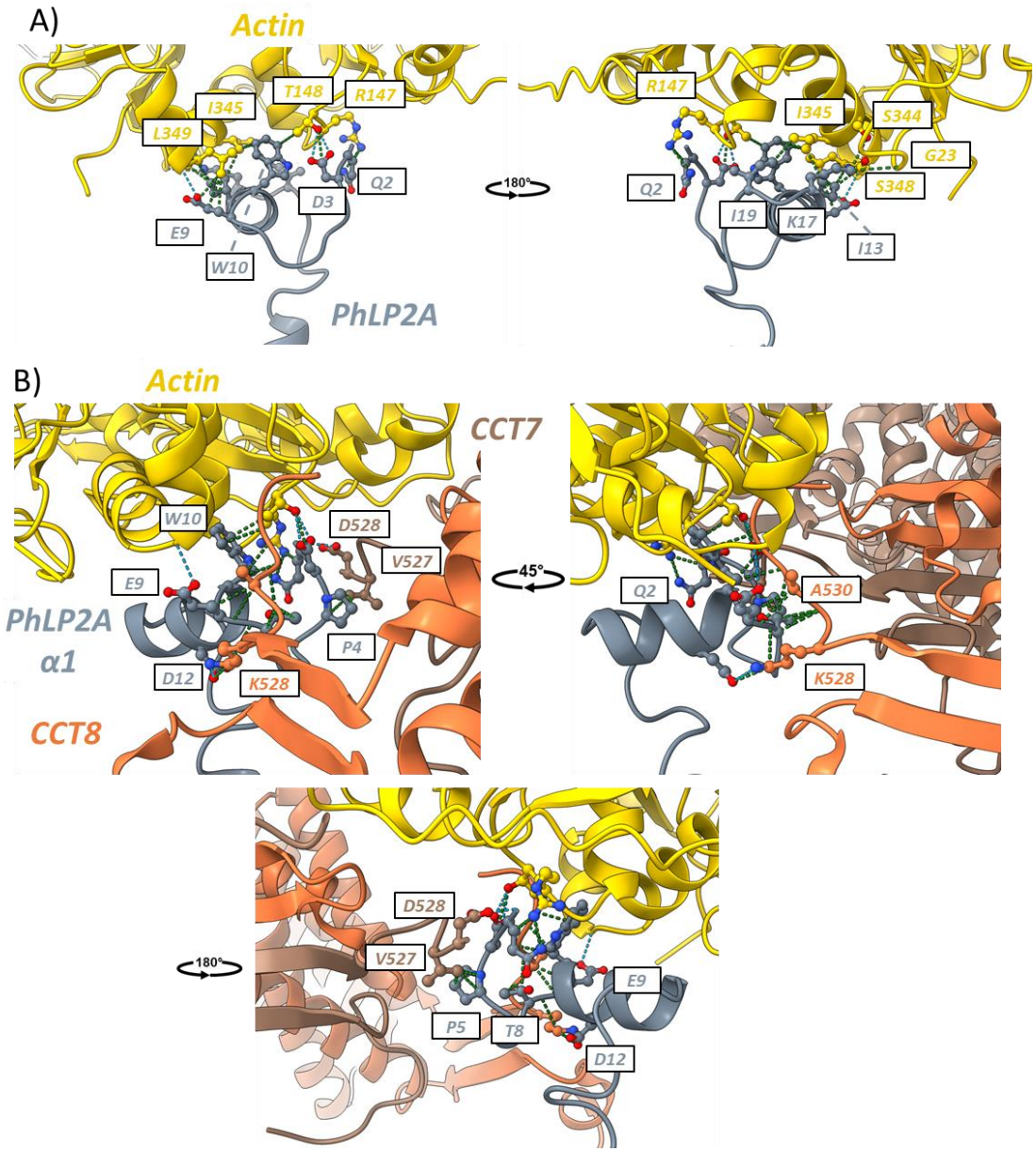


Figure 5.31: PhLP2A helix 1 interactions with actin and TRiC. A) Binding site interactions between actin (gold) and PhLP2A helix 1 (α1, slate grey). B) Binding region between actin, PhLP2A, and the C-termini of CCT7 (brown) and CCT8 (orange). Right and bottom view are rotated 45° and 180°, respectively.

The helix 2 and helix 3 (H2 and H3) regions of PhLP2A could not be reliably built at atomic resolution due to weak map density. However, these regions clearly show two long cylindrical regions of density that fit well to the helices described in secondary structure prediction (Figure 5.29B; Figure 5.30A). H2 (~aa 23-66) begins shortly after

the H1 region that directly binds actin and extends deep into the folding cavity, passing through the septum between the rings, and ends in the *trans* ring (Figure 5.32A). The junction between H2 and H3 is difficult to determine, however a large region of density is present in this area that nests near the apical domains of CCT3', CCT6', and CCT8' in the *trans* ring (Figure 5.32B). H3 (~aa 67-90) then traverses from this region across the *trans* cavity to the thioredoxin domain. These helices clearly play an important role in the regulation of substrates that are in the process of folding and could serve to function as a flexible tether that prevents substrates from exiting the folding cavity while bound within the TRiC complex.

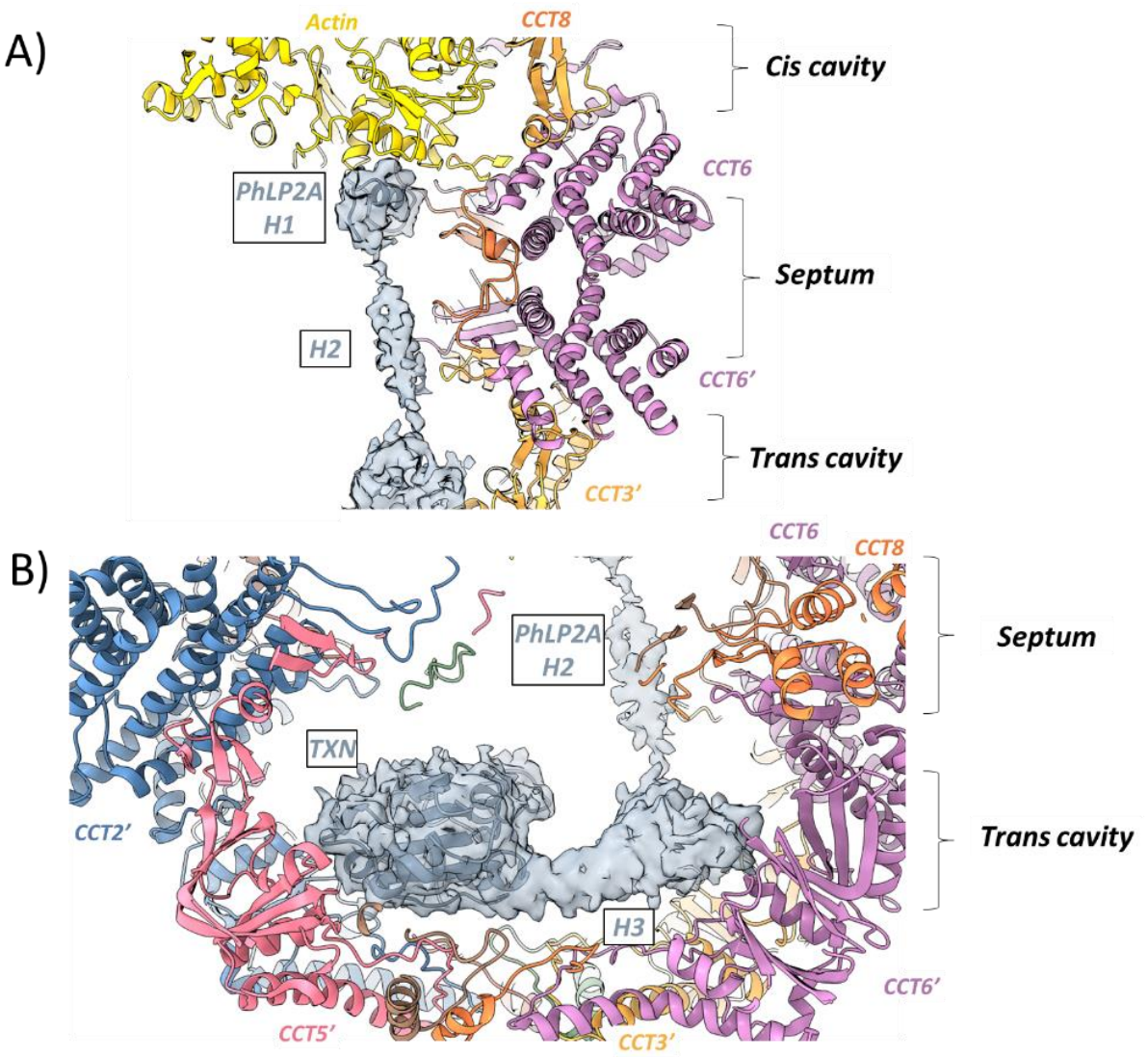


Figure 5.32: PhLP2A helix 2 and helix 3 interactions with TRiC. A) Side view slice of TRiC in the closed state interacting with actin (gold) and PhLP2A (slate grey). Helix 1 (H1) of PhLP2A is shown as ribbons and density corresponding to helix 2 (H2) is represented by slate grey density and is shown to traverse from the *cis* ring through the septum into the *trans* ring near the TRiC subunits CCT8 (orange), CCT6 (purple), and CCT3 (yellow-orange). B) Side view slice of the *trans* folding cavity showing density for helix 2 and helix 3 (H2 and H3) of PhLP2A that connects to the thioredoxin (TXN) domain of PhLP2A represented by ribbons near CCT6', CCT3', CCT5' (pink), and CCT2' (blue).

5.8.3 Thioredoxin domain and C-terminus anchor PhLP2A in *trans* ring

The PhLP2A thioredoxin domain is well-defined within the substrate map and was key in identifying PhLP2A as the additional substrate density. This domain identification was based on the stronger map density in this area that allowed for fitting of several β-

strands that were present the map density (Figure 5.29A). The thioredoxin fold is a conserved feature of enzymes that catalyse disulfide bond formation and consists of four antiparallel β-strands flanked by three α-helices[248]. The thioredoxin domain of PhLP2A contacts the apical domains of CCT7', CCT5', CCT2', and CCT4' (Figure 5.33). Lys100, Gln104, and Thr153 of PhLP2A interact with Arg313 of CCT7' and Trp304/324 of CCT5' (Figure 5.33A). It is noteworthy that CCT2 and CCT4, subunits not shown to have strong substrate interactions with tubulin in this study and for most substrates observed structurally with TRiC previously (Section 1.5.3), make contacts with the thioredoxin domain (Figure 5.33B). Three CCT2' residues including Asp245, Leu295, and Tyr297 contact Gly123 of PhlP2A. Additionally, Phe252 in the apical loop protrusion of CCT2' contacts Glu96 of PhLP2A. Residues Ile304 and Leu305 on a loop region of CCT4' interact with several PhLP2A regions including Ala128/Leu129/Gln132 and Met187 (Figure 5.33B).

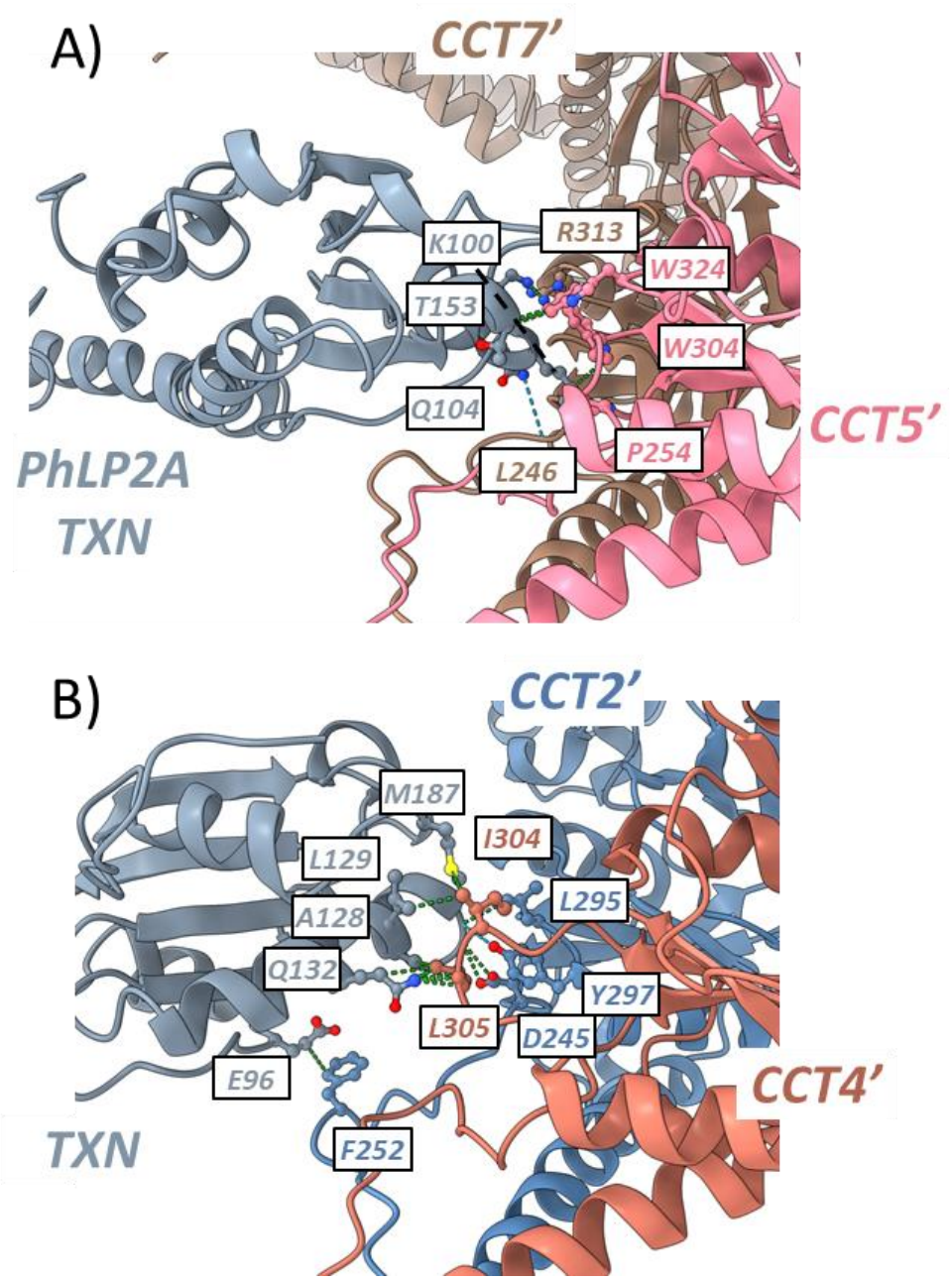


Figure 5.33: PhLP2A thioredoxin domain interactions with *trans* ring of TRiC. A) CCT7' (brown) and CCT5' (pink) interactions with thioredoxin fold of PhLP2A. B) CCT2' (blue) and CCT4' (red) interactions with thioredoxin fold of PhLP2A.

The final 27 residues of PhLP2A (aa 213-239), consisting of the acidic C-terminus (Figure 5.28), were not represented by density, and could not be built into the model. The region around the C-terminus does however contain considerable density at low

isosurface threshold (Figure 5.34: top left), that corresponded to further C-terminal extensions of CCT1' and CCT3' (Figure 5.34: top right). Indeed, 14 additional residues could be built at the C-terminus of CCT1' compared to the consensus closed model (Figure 5.34: bottom).

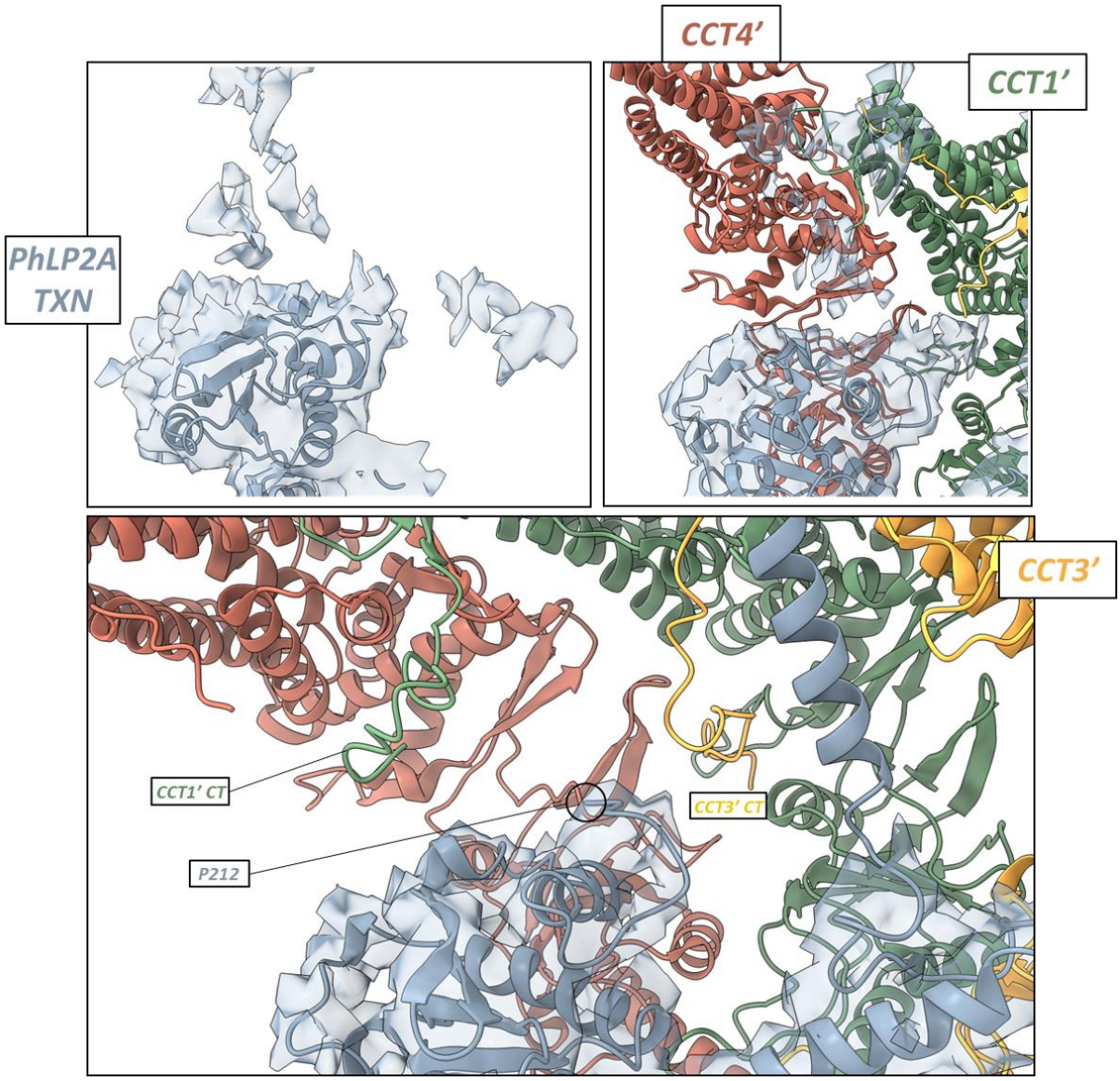


Figure 5.34: PhLP2A missing C-terminal region is proximal to TRiC subunit C-termini. C-terminal region of PhLP2A displayed with weak unmodelled density (top left) and shown proximal to CCT4', CCT1' (green), and CCT3' (top right). Bottom: View of newly built C-termini (CT) of CCT1' and CCT3' into weak density. The C-terminus of PhLP2A could not be built and the final modelled residue Pro212 is indicated for reference.

Intriguingly, the C-terminus of CCT3' could be fully built and terminates in a pocket created between the stacking of CCT1' and CCT4' (Figure 5.35A: top). This density was also observed in the *cis* ring where actin is bound in the same groove between CCT1 and CCT4 (Figure 5.35A: bottom). This density is also present in the tubulin-bound map in both the *trans* (Figure 5.35B: top) and *cis* rings (Figure 5.35B: bottom). These observations suggest a pattern by which the C-terminus of CCT3 anchors itself between CCT1 and CCT4. This interaction is likely important and could play a key role in TRiC assembly/arrangement. The C-terminal residues of PhLP2A could not be ascribed to the substrate map, however it is likely that its proximity to these CCT termini extensions plays a role in regulating the TRiC-PhLP2A interactions.

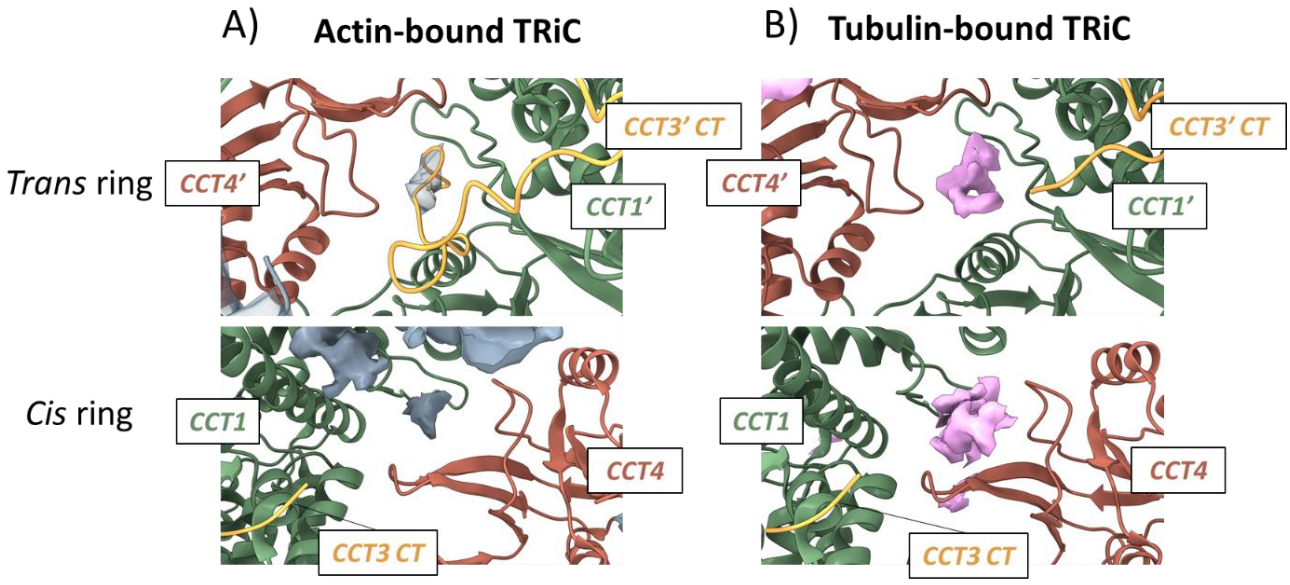


Figure 5.35: CCT3 C-terminus binds between groove of CCT4 and CCT1. A) Side view from inside *trans* TRiC ring of PhLP2A-bound model. Top: The CCT3' C-terminus (yellow) extends into density between CCT4' and CCT1'. Bottom: The density between CCT1 and CCT4 is also observed in the *cis* ring of the actin-bound model. B) Density is observed in the CCT1-CCT4 groove in the *trans* (top) and *cis* (bottom) rings of tubulin-bound map.

5.9 3D reconstruction of TRiC in the open conformation

The conformational rigidity of TRiC subunits and large number of particles collected for the closed state have allowed for many observations that are key to understanding the TRiC-mediated substrate folding mechanism. By contrast, the TRiC open form is more flexible due to movements in the apical domain protrusions, and thus it is more difficult to obtain high resolution information from cryo-EM studies in this conformation. The open map generated in this study (Figure 5.4A) was, however, of sufficient quality to derive a model of TRiC in the open state whereby the apical domains are arranged as a tetramer of dimers (CCT3-CCT1, CCT4-CCT2, CCT5-CCT7, CCT8-CCT6) as previously seen[138], indicative of a partially open ligand-bound conformation. Local resolution estimates show the equatorial domains of each subunit are around 3.5 Å, while the density fades upwards towards the tips of the apical domains, to as low as ~8 Å (Figure 5.4B: right).

From this open map, a model was built of the equatorial domains that spans the N- and C-termini of each subunit from the nucleotide binding pocket up to the intermediate domains (Figure 5.36A). Using molecular dynamics to flexibly fit the apical regions to the low-resolution density, a more complete composite model was derived that incorporates the entire apical domains in their approximate locations based on the open density (Figure 5.36B).

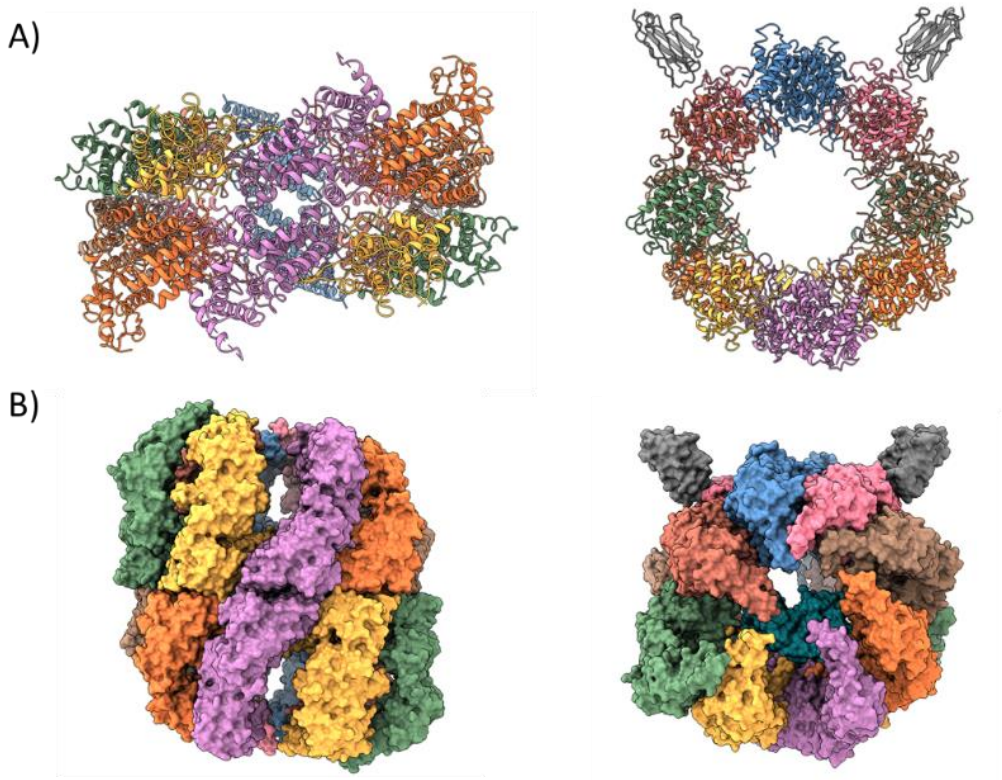


Figure 5.36: Model of TRiC in the open conformation. A) Models built of the equatorial domains of TRiC open conformation in side (left) and top-down (right) views. B) More complete models of TRiC open conformation, rendered as surfaces, including intermediate and apical domains using flexible fitting in side (left) and top-down (right) views.

All subunits contained density for ADP-AlFx within each expected ATP binding site in the equatorial domains, except CCT4 and CCT5, where the ligand density is shorter and wider than that seen for other subunits (Figure 5.37). Therefore, the ligand bound in these subunits could be ADP instead of ATP. It is possible that ADP was generated upon incubation of the TRiC sample with ATP, sodium fluoride, and aluminium nitrate before the ADP-AlF_X ligand is formed. The presence of bound ADP for these subunits, as opposed to ADP-AlF_X in the others, likely accounts for the presence of a subset of open particles in the TRiC sample, as CCT4 and CCT5 have been shown to play an important role in regulation of lid closure[27].

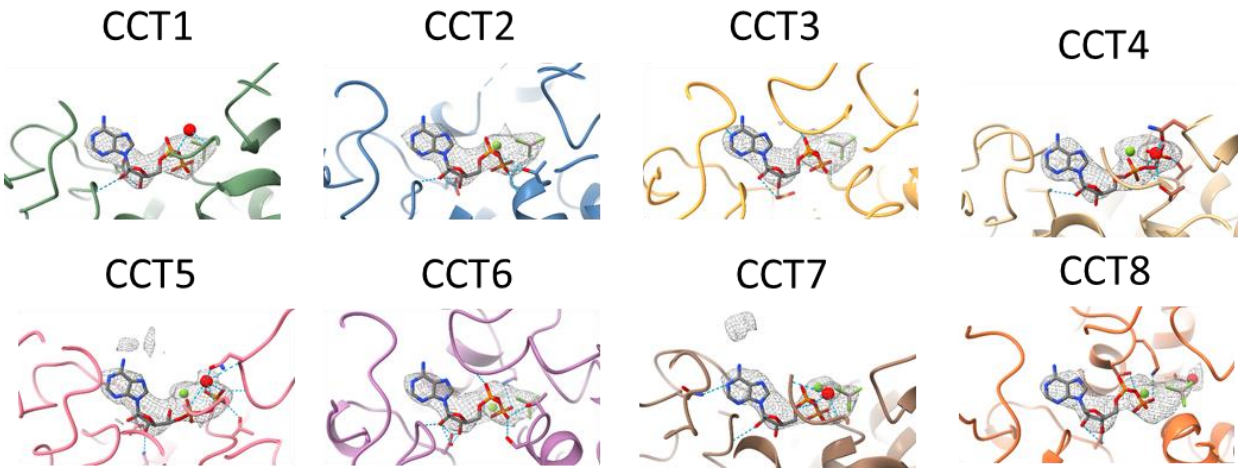


Figure 5.37: Ligand occupancy in TRiC subunits in the open conformation. Snapshots of ligand density in each CCT subunit.

5.9.1 TRiC interactions with nanobody 18 and substrate in the open state

Nanobody binding to TRiC was also observed in the open conformation although the density was much weaker than that of the closed conformation (Figure 5.4B: right). The binding interface remained similar, although interactions with CCT5 differed slightly from the closed state in which only two variable loops (CD1 and CD3) interacted with Nb18, whereas all three variable CD loops (1/2/3) interact with CCT5 in the closed conformation (Figure 5.38).

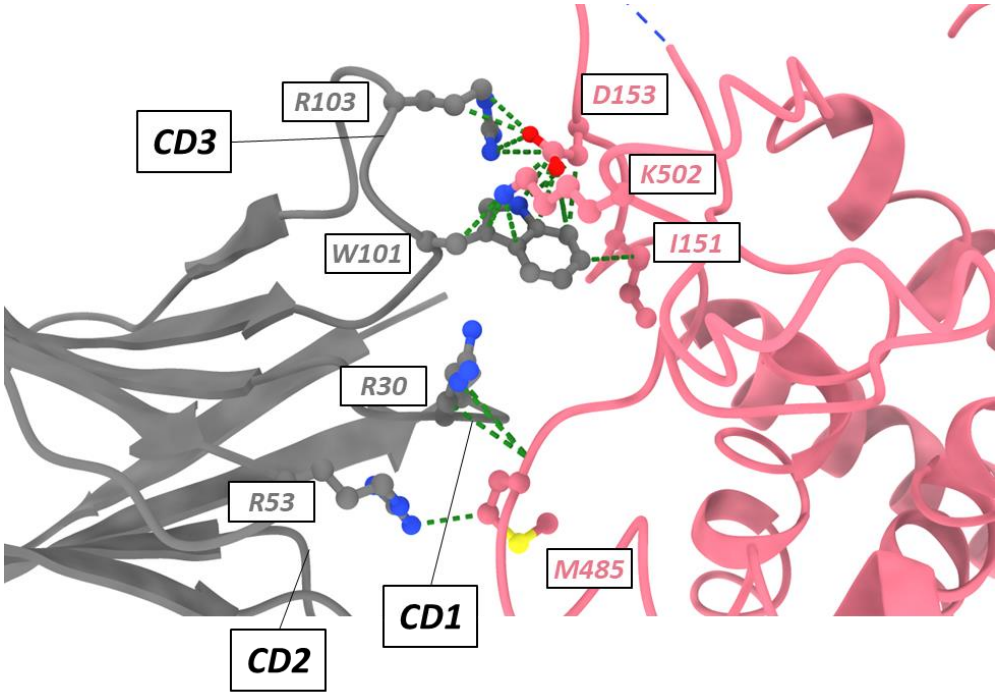


Figure 5.38: Nanobody 18 binds CCT5 in the open conformation. Interacting residues between Nb18 (grey) and CCT5 (pink) in the open state. Green dashes indicate van der Waals contacts.

As observed in the original map of TRiC in the open state (Figure 5.4A), density that could not be ascribed to TRiC was observed within the folding chamber. However, the resolution was not sufficient to unambiguously assign the identity of this protein. At high isosurface threshold the density could accommodate the thioredoxin and helix H3 region of PhLP2A instead of the other substrates identified in tubulin and actin (Figure 5.39A-D: left). However, at a lower threshold the density also appears to accommodate actin and tubulin (Figure 5.39A-D: right). As the quality of data for the open set of particles was not high enough for further classifications to reveal individual substrates, the density seen at this stage is likely a mixture of particles.

Interestingly, this density mass appears to be suspended diagonally within the septum such that the substrate is bound to the same subunits of both the *cis* and *trans*

rings (Figure 5.39). This is unlike the closed state models in which substrates localised to one of two TRiC chambers closer to the apical domains. The open state substrate is positioned >30 Å deeper into the chamber when compared to the actin and tubulin positions in closed-TRiC, occupying the same site as the mLST8 substrate in the human open-TRiC structure[171]. The main interaction seems to be with the termini of CCT7 in each ring (Figure 5.39A), however the limitations in resolution in the open map prevented further interpretations. In future, open state models could be improved either by a large increase in particle number to classify a range of open structures, though this might be computationally impractical, or by preparing the TRiC sample with a ligand that stabilises the open state such as a conformationally selective nanobody.

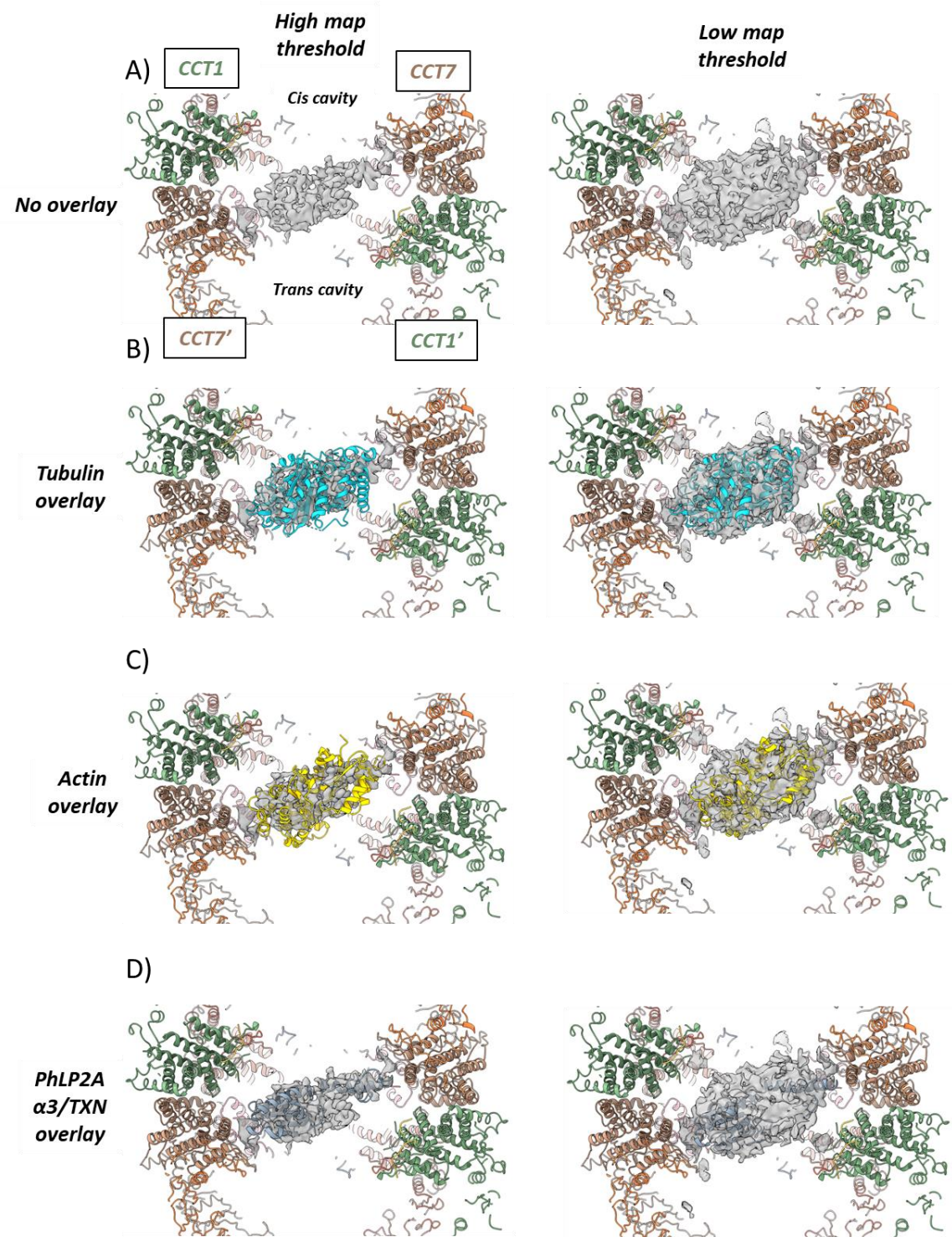


Figure 5.39: Open conformation of TRiC contains substrate bound. A) Slice of modelled substrate density in the open conformation at high (left) and low (right) isosurface threshold. CCT1 (green) and CCT7 (brown) from *cis* and *trans* rings are shown for context. Known substrates placed within density are displayed including B) tubulin (cyan), C) actin (gold), and D) the PhLP2A helix 3/thioredoxin domain (α3/TXN, slate grey).

5.9.2 Differences between TRiC in the open and closed states

When the equatorial models of open and closed conformations were compared, a stark difference was seen in the stacking of subunits between the two rings. The N-terminal β-strand and interaction networks observed for the closed model (Figure 5.8) were not observed in the open map density (Figure 5.40), possibly because these residues are no longer interacting with the opposite ring as seen in the closed conformation. As a result, the observed first residue of each subunit is positioned further from the opposite ring in the open state (Figure 5.40), making the N-terminal interactions less likely to exist as the open conformation does not require these interactions to form a tightly closed septum.

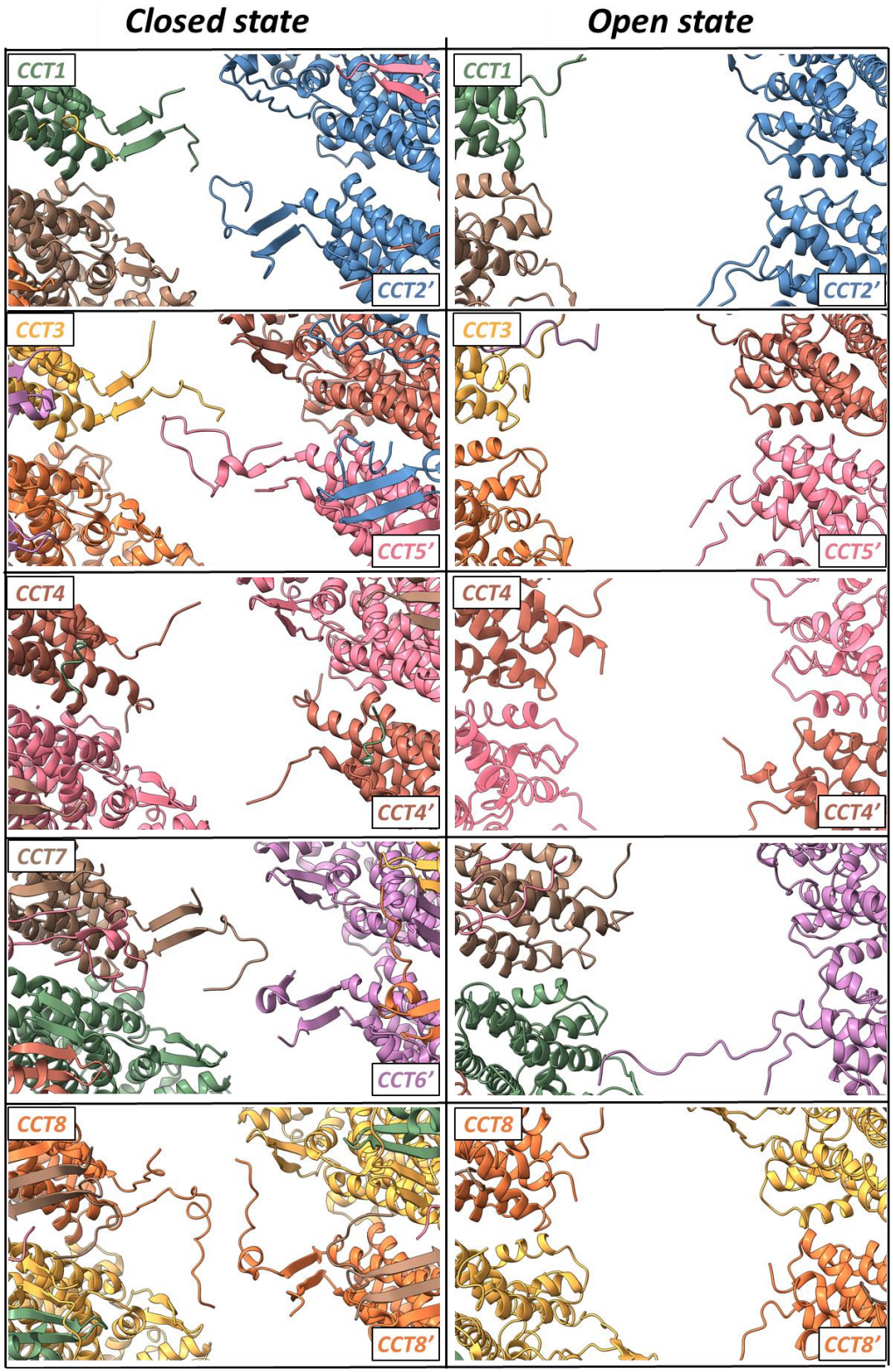


Figure 5.40: N-terminal interaction network between *cis* and *trans* rings does not exist for TRiC in the open conformation. Panels shows 'N+2 pairs from each CCT subunit of N-terminal interaction network in the closed state (left), while these regions are disordered in the open state (right). The ribbon display of the stacked subunits between each N'+2 subunit pair has been hidden for clarity.

In addition to differences in the N-terminal interactions, the open and closed states also differ in their subunit arrangement. As illustrated using the example of CCT2 (Figure 5.41), in the closed state the C-terminal region of α14 stacks perpendicular on the loop between helices α4 and α5 of its *trans* stacked partner (Figure 5.41A, orange regions). In the open state, these interactions disengage with a rearranged interface, in which helix α15 of one subunit sits parallel on top of the loop region between helices α13 and α14 of the stacked subunit (Figure 5.41B, purple regions). This region contains many interacting residues that form a broad binding surface compared to the closed state that angles the stem-loop and N- and C-termini upwards toward the TRiC folding cavity and apical domains. This parallel stacking between *cis* and *trans* ring subunits is consistent throughout each CCT subunit in the open state.

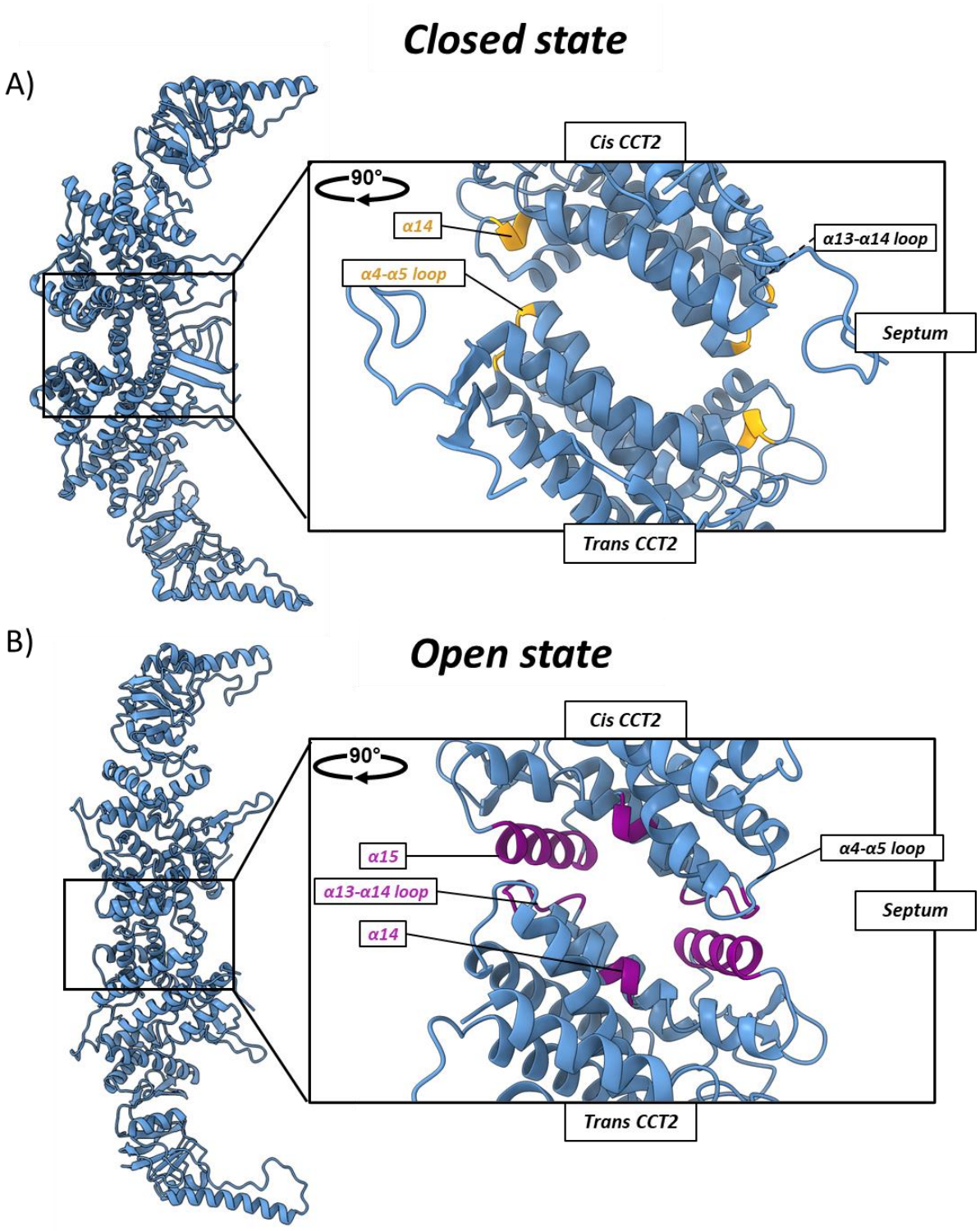


Figure 5.41: Stacking of CCT subunits differs between open and closed conformations. A) Side view of CCT2-CCT2' stacking in the closed state. Inset: regions of each subunit that contact opposite ring are shown in orange. Regions indicated in black denote binding regions in the open state. Dotted lines indicate regions that are located in back of image. B) Side view of CCT2-CCT2' stacking in the open state. Inset: regions of each subunit that contact opposite ring are shown in purple. Regions indicated in black denote binding regions in the closed state.

The rearranged intra-/inter-subunit contacts in open state TRiC indicate that the septum structure at the ring-ring interface has become more dynamic, and less constricted between the two cavities, when compared to closed state TRiC (Figure 5.8). Taken together, these observations show a mechanism whereby ATP hydrolysis catalyses a range of motion between the open and closed state.

5.10 Summary

Overall, the results in this chapter detail LC-MS/MS and cryo-EM experiments from a single TRiC sample, that yielded a wealth of insight into the mechanism of TRiC-mediated substrate binding, demonstrating the utility of proteomic and structural techniques to investigate complex biological pathways. A high resolution 2.5 Å atomic model of TRiC in the closed state revealed unique characteristics of the interplay of the stacking surface at the septum between the *cis* and *trans* TRiC rings. The TRiC-tubulin and TRiC-actin-PhLP2A complexes were described in a level of detail not previously reported and provide a novel insight into the degree and location of substrate folding while in mid-folding cycle in the closed state of TRiC. The TRiC-actin-PhLP2A complex also shed new light into the mechanism by which co-chaperones in the PhLP family regulate substrates in the TRiC pathway. Finally, the open form of TRiC was also observed in this cryo-EM experiment and provided further insight to an additional step of the TRiC folding cycle beyond what is observed for substrates bound to the fully closed chaperonin.

6. Discussion and conclusion

6.1 Overview of discoveries described in this thesis

The eukaryotic group II chaperonin TRiC is essential for cellular survival and has been implicated in a wide range of diseases due to its function in assisting the native fold of ~10% of the human proteome. The substrates of TRiC vary in cellular function and play a role in nearly all cellular functions from cell cycle regulation and structure to trafficking to transcription and translation. Therefore, understanding of the TRiC-mediated substrate folding pathway is important to developing therapies for human disease pathology. However, not much is currently known about the mechanism by which substrates fold while within the TRiC chaperonin.

This thesis details efforts to characterise the eukaryotic group II chaperonin CCT/TRiC through biochemical and structural analyses. Chapter 3 described the purification of all eight TRiC subunits and subsequent attempts to assess the levels of activity and function of individual CCT subunits. These findings confirmed previous results that CCT4 and CCT5 retain functional activity while in an over-expressed synthetic homo-oligomer of 16 subunits that are similar in assembly to native TRiC. Several other subunits showed homo-oligomeric assembly, though these did not retain the same levels of biological activity as CCT4 and CCT5 and likely do not perform chaperone activity outside of the TRiC complex. The techniques established in Chapter 3 paved the way for biochemical and structural characterisation of native TRiC, reliant on whether a method could be developed to purify native TRiC.

Chapter 4 described several strategies to isolate native TRiC including endogenous purification from wild-type human HEK293 cells, recombinant over-

expression of each TRiC subunit using MultiBac vectors, affinity-based pull-down of TRiC using anti-CCT5 nanobody binders, and CRISPR/Cas9-mediated genome editing of wild-type human HEK293 to insert an affinity tag in a single CCT subunit (CCT5). The latter two methods were successfully employed to purify TRiC from human cells, and the CCT5-FLAG cell line generated from the genome editing workflow was used to generate quantities of TRiC large enough to enable downstream biochemical and structural investigation into TRiC-substrate folding activity.

The structural studies described in Chapter 5 detail novel snapshots of intermediate substrate binding states for the canonical substrates actin and tubulin, as well as the less well-characterised co-chaperone PhLP2A. The two substrates were described in the closed conformation of TRiC in which they had already achieved near-native folds. The observations reported here are the first in the literature to show substrates bound with the TRiC folding cavity while TRiC is in the closed state. Importantly, these interactions show in high resolution the regions of contact between actin and tubulin with TRiC subunits, features that had previously been studied through crosslinking and hydrogen-deuterium exchange mass spectrometry[162,249].

The open conformation also showed density for substrate/s bound within the septum between the two TRiC rings but was likely a sampling of a mixture of the most abundant substrates that bind to TRiC in the purified sample. This conformation of TRiC has been previously described as the pre-hydrolysis state[42,129,138], in which a rearrangement of subunits forms a partially open state. The location of substrate was within the septum formed by the stacking of the two TRiC rings and to the subunits

CCT3, CCT6, and CCT8, an observation previously seen for actin[162] and mLST8[171] (see Chapter 1 section 1.5.3, Figure 1.12).

6.2 Relevance to chaperonin protein folding and proteostasis

Based on the data described in this thesis a model of the TRiC-substrate folding mechanism was proposed whereby a substrate first binds the apical domains in a fully open state not yet bound with nucleotide. Upon nucleotide binding, this partial rearrangement shepherds the substrate into the folding chamber and positions it between the two rings, poised for contacting specific regions of TRiC to promote native folding. Upon ATP hydrolysis the ring then closes, possibly forcing the substrate out of the septum by a rearrangement of the equatorial domains between the open and closed conformation that causes steric hinderance of the substrate. The substrate then enters the confined folding chamber of TRiC and achieves a near-native fold promoted by binding to specific regions of various CCT subunits. Upon exchange of nucleotide, the complex then re-opens, allowing the substrate to release into the cellular milieu and perform its native function (Figure 6.1). Additionally, this mechanism can be regulated by co-chaperones including PhLP2A, which could function by trapping substrates within the folding chamber to down-regulate overall substrate levels in the cellular environment, or release them only upon achieving their native folds.

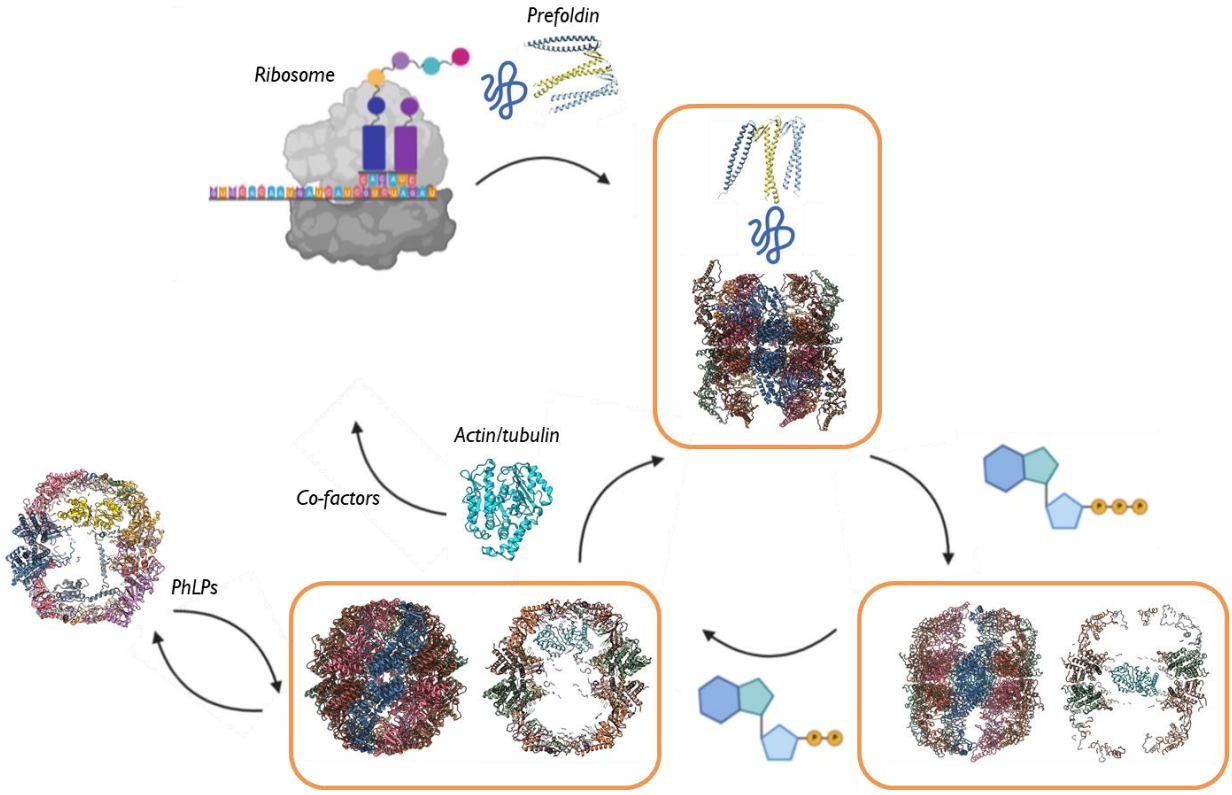


Figure 6.1: Proposed mechanism of TRiC-mediated substrate folding based on findings of this thesis. Hypothetical substrate folding pathway of TRiC in which: substrate is delivered from the ribosome to TRiC in the fully open state (PDB: 5GW4)[138] by prefoldin, binding of nucleotide induces partially open conformation whereby substrate enters the septum between the two rings, ATP is hydrolysed and substrate enters the TRiC folding chamber, and is subsequently modulated by PhLP proteins or released.

The expanded understanding of the TRiC pathway provides insights into how actin and tubulin achieve their native fold. This information could have therapeutic significance as the cytoskeleton, which largely depends on actin and tubulin dynamics, has been shown to play a role in various forms of cancer and neurodegenerative disease[250–254]. Additionally, the precise recognition motif of the N-terminal helix of PhLP2A was shown at high resolution to bind to actin subdomains 1 and 3. This information shows the mechanism by which PhLP2A acts a co-chaperone to actin and provides a greater understanding of how the PhLP family regulates the levels of TRiC-mediated substrate folding. Knowing these details could also lead to therapeutic

benefits, as this pathway could be investigated as a way to modulate levels of actin, tubulin, and possibly as yet unidentified substrates involved in this pathway.

6.3 Further areas of understanding in the TRiC folding pathway mechanism

While the results described in this thesis provide a greater understanding of how actin and tubulin achieve their native folds while trapped in the TRiC folding cavity, this information provides only a glimpse in to the overall TRiC pathway. Hundreds of potential TRiC substrates have been reported and only a few have been experimentally validated, mostly by affinity purification and mass spectrometry, to strongly interact with TRiC. Even among these, few substrates beyond actin and tubulin have been proven to require TRiC to achieve their native fold[52,218]. Therefore, it is currently unknown the degree to which modulation of the TRiC pathway can offer therapeutic benefit as the exact mechanism of substrate folding for this large set of substrates is still unknown. Further investigations similar to those described in this thesis for other substrates, especially those from a biologically relevant endogenous source, could drive TRiC from a pathway with potential therapeutic benefit to a bona fide drug target.

In addition to driving more specific therapeutic benefit, structural characterisation of additional substrates could also prove useful for understanding more specific mechanistic details of TRiC-mediated substrate folding. For example, the substrate density described for actin and tubulin within TRiC in the closed state appears to be similar on a global scale, with both roughly positioned between the apical and stem-loops of CCT3, CCT6, and CCT8. However, upon more detailed observations, these two substrates vary in many of the contacts they make with TRiC subunits. They also show similarities in certain regions including the positively charged apical region of

CCT6 that forms many tubulin and actin contacts. These observations could shed light into the process by which TRiC recognises and folds a wide topological array of substrates and provide more details into substrate specificity.

In addition to the substrates bound to TRiC in the closed state, an unidentified substrate density was observed for TRiC in the open state. This substrate was not identified because the open state could not be further parsed out using 3D classification, as was done for the closed state. This limitation meant that only broad conclusions could be drawn about the location of the substrate density within the open TRiC model, and that the density observed was likely a sampling of the highest enriched substrates within the purified sample. Therefore, detailed interactions could not be discussed for the open state. This is important, as the mechanism proposed in this thesis describes a mode in which substrates are delivered to TRiC in the open state, then localise to the septum between the two rings, and subsequently move up into the folding cavity during ATP-induced ring closure. This is the likely mechanism as the substrates observed in the closed state are in a near-folded state, suggesting that these substrates have already undergone TRiC-mediated folding activities. However, there are no comparisons to this observation, and being able to see the folding states of substrates in the open form would help confirm this mechanism. For example, if substrates have achieved a near-native fold while positioned in the septum, this could suggest that closing of the rings only acts to position substrates in the folding cavity, and that the final substrate fold is achieved during re-opening of the rings and release into the cytosol. Therefore, further structural investigations into TRiC in the open state could shed additional light on the substrate folding mechanism.

An intriguing result from ATP activity assays showed that certain anti-CCT5 nanobodies can reduce the activity of the $CCT5_{16}$ and TRiC chaperonins. Further investigation into these inhibitory nanobodies could prove to be useful tools for understanding the conformational dynamics of TRiC. Given that ATP hydrolysis is required for the transition between the open and closed states of TRiC, these nanobodies are likely conformationally selective and could stabilise either closed or open states. However, the degree to which these anti-CCT5 nanobodies inhibit TRiC was only briefly investigated, and further experiments validating these as modulators would be useful to confirm their inhibitory effects. In the experiments reported in this thesis the inhibition of anti-CCT5 nanobodies had a higher impact on the $CCT5_{16}$ homo-oligomer compared to native TRiC, an expected result as more copies of CCT5 can be inhibited in the homo-oligomer. Further nanobody development against other TRiC subunits outside of CCT5 could additionally inhibit TRiC in a stepwise manner and have potential value as therapeutic modulators of TRiC activity.

6.4 Areas of further work to understand eukaryotic group II chaperonin

Further experiments to expand understanding of TRiC-mediated substrate folding and the conformational cycle involved in the TRiC pathway could prove useful for understanding its role as a major component of proteostasis regulation. The cryo-EM study described in this thesis shows tubulin, actin, and PhLP2A isolated from a single endogenous source using HEK293 human cells. These were most likely the highest enriched substrates bound within TRiC in the closed state, or at least the ones able to be separated by 3D classification, however this expression pattern might vary from cell line to cell line. Thus, performing native isolation of TRiC from different cell lines, or from

the same cell line with different cellular conditions (e.g., cold shock, heat shock) could lead to isolation of TRiC with a wider array of substrates bound. Additionally, native TRiC could be over-expressed with or incubated with different substrates that have been shown to interact with TRiC, especially those that have already been shown to engage with TRiC in the open state such as AML1-ETO, Gβ, and mLST8. These substrates could bind within the TRiC folding cavity in the closed state in a similar manner to actin and tubulin. Some substrates, such as Gβ that has already shown interactions with PhLP, could also be modulated by co-chaperones while bound to TRiC. These studies would not require significant effort to perform and could be a suitable next step to further characterising specific substrate contacts that occur during the TRiC folding cycle.

High resolution information for TRiC in the open state has been difficult to obtain due to the flexibility of the subunit apical domains. Stabilising this conformation could be important for understanding folding dynamics including how substrates engage TRiC and enter the folding cavity. Incubation with ATP analogues that mimic the open state provided a rough picture of the stacking of the two TRiC rings near the equatorial domains but has been insufficient to visualise the apical domains. Conformationally selective nanobodies have the potential to stabilise the open state, perhaps by binding near the apical domains and acting as a wedge that reduces flexibility in these regions. These could in turn assist cryo-EM investigations by reducing motion induced limitations in density, leading to higher resolution reconstructions as was described for the TRiC closed state.

The purification strategy described in Chapter 4 was successful in providing ample quantities of pure TRiC for downstream assays. This thesis describes how this material was used for cryo-EM experiments and was briefly used to measure ATP hydrolysis activity in the presence and absence of inhibitory anti-CCT5 nanobodies. The quantities of TRiC generated by this method, however, now have the potential to be used for a wide array of downstream assays that could be used to validate TRiC as a target for human disease. For example, the role TRiC plays in aggregation could be revisited, especially to investigate whether it can suppress the aggregation of the mutant exon 1 of huntingtin (mhtt). TRiC was previously shown, through an interaction with CCT1 and the N-terminus of huntingtin, to block aggregation of huntingtin and was suggested as a potential therapeutic mechanism[109,112,255]. However, no work has since been reported to validate this interaction. One experiment that could show this interaction would be a recapitulation of the structural data that was reported in Chapter 5, in which a high-resolution snapshot is captured of a substrate bound within the TRiC folding cavity. This would provide validation of the interaction between TRiC and the N-terminus of huntingtin that could potentially establish a common mechanism whereby TRiC can suppress aggregation of toxic aggregates in neurodegenerative diseases.

A drug discovery campaign focused around TRiC would be a larger endeavour that could lead to broad therapeutic benefit in the long term. Currently there are very few known modulators of TRiC activity including the HSF1 activator[86,256], the phosducin-like protein family of co-chaperones[72], and the nanobodies reported in this thesis. A workflow could be established in which small molecule binders of TRiC are identified by a broad screening technique such as affinity selection mass spectrometry. This

technique would identify compounds that bind to TRiC that could be subsequently screened for their ability to modulate TRiC function. This thesis has described experimental workflows that would then validate these small molecules including screening modulators of ATP through malachite green ATPase activity assays, identification of strong binders by BLI, and subsequent structural analysis to observe the binding mode of these small molecules.

As the assays currently stand, this workflow is not fully suited to a high throughput screening effort and would require optimisation. For example, the ATPase assay could be adapted from an absorbance-based malachite green assay to an assay that is more sensitive and requires less protein demand such as an ADP-Glo luminescent assay[257]. Additionally, the technique used to determine the structure of human TRiC in this thesis used cryo-EM, which is currently not a high throughput structural screening method. However, sufficient quantities of TRiC can now be generated to attempt alternative structure-based methods of identifying modulators of TRiC. SAXS is one such method that has the potential to observe whether small molecule modulators induce a conformational change between the open and closed form by measuring the overall size and shape of TRiC. Another structural technique that has not been tenable previously due to protein demands is X-ray crystallography. Crystals of TRiC with substrates bound have previously been obtained[136], and with optimisation this technique has the potential to provide high-resolution structural information of small molecules bound to TRiC in a more high-throughput manner. Crystallising TRiC with nanobodies could help this workflow as nanobodies have been shown to promote crystal formation for proteins that are not readily amenable to this technique[231]. Due to the wide array of substrates

and cellular pathways TRiC is involved in (Appendix 1) both activators and inhibitors could be useful as means to modulate human disease. For example, inhibitors of TRiC could be of value in certain cancers in which expression of certain TRiC subunits or for viral infections that require TRiC for assembly. Activators of TRiC could also be useful, for example in neurodegenerative diseases where boosting the protective properties of TRiC would be advantageous.

The goal of this thesis is to provide deeper insight into the substrate-folding mechanism of the group II chaperonin CCT/TRiC. TRiC has previously been studied as a vital protein folding cellular machine for three decades, yet understanding of the role that TRiC plays in regulating proteostasis remains poor. It is widely accepted that TRiC is involved, be it in an obligate manner or otherwise, with >10% of the human proteome, however many of these interactions remain speculative. The difficulty in studying this network has largely stemmed from the challenges that arise in attempting to purify quantities large enough for downstream biochemical assays. In this thesis efforts were undertaken to validate assays that could be useful in further understanding these pathways (Chapter 3), establish methods for isolating large quantities of native TRiC (Chapter 4), and push the boundaries of what has previously been reported for the mechanism of TRiC-mediated substrate folding through cryo-EM reconstructions (Chapter 5). The therapeutic benefit of studying the TRiC pathway has yet to be fully realised, and the results described herein hopefully serve as a starting point to reach this potential.

References

1. Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M. & Ulrich Hartl, F. *Molecular Chaperone Functions in Protein Folding and Proteostasis*. *Annual Review of Biochemistry* vol. 82 (2013).

2. Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science (80-.).* **181**, 223–230 (1973).

3. Acampora, G. & Hermans, J. Reversible Denaturation of Sperm Whale Myoglobin. I. Dependence on Temperature, pH, and Composition. *J. Am. Chem. Soc.* **89**, 1543–1547 (1967).

4. Laskey, R. A., Honda, B. M., Mills, A. D. & Finch, J. T. Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature* **275**, 416–420 (1978).

5. Ellis, J. Proteins as molecular chaperones. *Nature* **328**, 378–379 (1988).

6. Hemmingsen, S. M. *et al.* Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature* **333**, 330–334 (1988).

7. Spiess, C., Meyer, A. S., Reissmann, S. & Frydman, J. Mechanism of the eukaryotic chaperonin: protein folding in the chamber of secrets. *Trends in cell biology* vol. 14 598–604 (2004).

8. Gutsche, I., Essen, L. O. & Baumeister, W. Group II chaperonins: New TRiC(k)s and turns of a protein folding machine. *J. Mol. Biol.* **293**, 295–312 (1999).

9. Bukau, B. & Horwich, A. L. The Hsp70 and Hsp60 chaperone machines. *Cell* **92**, 351–366 (1998).

10. Sternberg, N. Properties of a mutant of Escherichia coli defective in bacteriophage λ head formation (groE). II. The propagation of phage λ. *J. Mol. Biol.* **76**, 25–44 (1973).

11. Ishihama, A., Ikeuchi, T., Matsumoto, A. & Yamamoto, S. A Novel Adenosine Triphosphatase Isolated from RNA Polymerase Preparations of Escherichia coli. *J. Biochem.* **79**, 927–936 (1976).

12. Hohn, T., Hohn, B., Engel, A., Wurtz, M. & Smith, P. R. Isolation and characterization of the host protein groE involved in bacteriophage lambda assembly. *J. Mol. Biol.* **129**, 359–373 (1979).

13. Tilly, K., Murialdo, H. & Georgopoulos, C. Identification of a second Escherichia coli groE gene whose product is necessary for bacteriophage morphogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 1629–1633 (1981).

14. Cheng, M. Y. *et al.* Mitochondrial heat-shock protein hsp60 is essential for assembly of proteins imported into yeast mitochondria. *Nature* **337**, 620–625 (1989).

15. Braig, K. *et al.* The crystal structure of the bacterial chaperonin GroEL at 2.8 A. *Nature* **371**, 578–586 (1994).

16. Xu, Z., Horwich, A. L. & Sigler, P. B. The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature* **388**, 741–750 (1997).

17. Phipps, B. M., Hoffmann, A., Stetter, K. O. & Baumeister, W. A novel ATPase complex selectively accumulated upon heat shock is a major cellular component of thermophilic archaebacteria. *EMBO J.* **10**, 1711–1722 (1991).

18. Trent, J. D., Nimmesgern, E., Wall, J. S., Hartl, F. U. & Horwich, A. L. A molecular chaperone from a thermophilic archaebacterium is related to the eukaryotic protein t-complex polypeptide-1. *Nature* **354**, 490–493 (1991).

19. Ditzel, L. *et al.* Crystal structure of the thermosome, the archaeal chaperonin and homolog of CCT. *Cell* **93**, 125–138 (1998).

20. Chaston, J. J. *et al.* Structural and Functional Insights into the Evolution and Stress Adaptation of Type II Chaperonins. *Structure* **24**, 364–374 (2016).

21. Pereira, J. H. *et al.* Crystal structures of a group II chaperonin reveal the open and closed states associated with the protein folding cycle. *J. Biol. Chem.* **285**, 27958–27966 (2010).

22. Silver, L. M., Artzt, K. & Bennett, D. A major testicular cell protein specified by a mouse T/t complex gene. *Cell* **17**, 275–284 (1979).

23. Silver, L. M. A structural gene (Tcp-1) within the mouse t complex is separable from effects on tail length and lethality but may be associated with effects on spermatogenesis. *Genet. Res.* **38**, 115–123 (1981).

24. Frydman, J. *et al.* Function in protein folding of TRiC, a cytosolic ring complex containing TCP-1 and structurally related subunits. *EMBO J.* **11**, 4767–4778 (1992).

25. Archibald, J. M., Blouin, C. & Doolittle, W. F. Gene duplication and the evolution of group II chaperonins: Implications for structure and function. *J. Struct. Biol.* **135**, 157–169 (2001).

26. Kubota, H., Hynes, G. M., Kerr, S. M. & Willison, K. R. Tissue-specific subunit of the mouse cytosolic chaperonin-containing TCP-1. *FEBS Lett.* **402**, 53–56 (1997).

27. Jin, M. *et al.* An ensemble of cryo-EM structures of TRiC reveal its conformational landscape and subunit specificity. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 19513–19522 (2019).

28. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

29. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **8**, 70–82 (2020).

30. Sternlicht, H. *et al.* The t-complex polypeptide 1 complex is a chaperonin for tubulin and actin in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 9422–9426 (1993).

31. Ursic, D. & Culbertson, M. R. The yeast homolog to mouse Tcp-1 affects microtubule-mediated processes. *Mol. Cell. Biol.* **11**, 2629–2640 (1991).

32. Yaffe, M. B. *et al.* TCP1 complex is a molecular chaperone in tubulin biogenesis. *Nature* **358**, 245–248 (1992).

33. Gao, Y., Thomas, J. O., Chow, R. L., Lee, G. H. & Cowan, N. J. A cytoplasmic chaperonin that catalyzes β-actin folding. *Cell* **69**, 1043–1050 (1992).

34. Vinh, D. B. N. & Drubin, D. G. A yeast TCP-1-like protein is required for actin function in

vivo. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 9116–9120 (1994).

35. Booth, C. R. *et al.* Mechanism of lid closure in the eukaryotic chaperonin TRiC/CCT. *Nat. Struct. Mol. Biol.* **15**, 746–753 (2008).

36. Yam, A. Y. *et al.* Defining the TRiC/CCT interactome links chaperonin function to stabilization of newly made proteins with complex topologies. *Nat. Struct. Mol. Biol.* **15**, 1255–1262 (2008).

37. Thulasiraman, V., Yang, C. F. & Frydman, J. In vivo newly translated polypeptides are sequestered in a protected folding environment. *EMBO J.* **18**, 85–95 (1999).

38. Camasses, A., Bogdanova, A., Shevchenko, A. & Zachariae, W. The CCT chaperonin promotes activation of the anaphase-promoting complex through the generation of functional Cdc20. *Mol. Cell* **12**, 87–100 (2003).

39. Sergeeva, O. A. *et al.* Human CCT4 and CCT5 chaperonin subunits expressed in Escherichia coli form biologically active homo-oligomers. *J. Biol. Chem.* **288**, 17734–17744 (2013).

40. Knee, K. M. *et al.* The group II chaperonin Mm-Cpn binds and refolds human γD crystallin. *Protein Sci.* **20**, 30–41 (2011).

41. Zhang, J. *et al.* Mechanism of folding chamber closure in a group II chaperonin. *Nature* **463**, 379–383 (2010).

42. Zhang, J. *et al.* Cryo-EM structure of a group II chaperonin in the prehydrolysis ATP-bound state leading to lid closure. *Structure* **19**, 633–639 (2011).

43. Nollen, E. A. A. *et al.* Genome-wide RNA interference screen identifies previously undescribed regulators of polyglutamine aggregation. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6403–6408 (2004).

44. Grantham, J. The Molecular Chaperone CCT/TRiC: An Essential Component of Proteostasis and a Potential Modulator of Protein Aggregation. *Front. Genet.* **11**, 1–7 (2020).

45. Sergeeva, O. A., Haase-pettingell, C. & King, J. A. Co-expression of CCT subunits hints at TRiC assembly. *Cell Stress Chaperones* (2019).

46. Gavin, A.-C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).

47. Ho, Y. *et al.* Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **415**, 180–183 (2002).

48. Dekker, C. *et al.* The interaction network of the chaperonin CCT. *EMBO J.* **27**, 1827–1839 (2008).

49. Wang, D. Y., Kamuda, K., Montoya, G. & Mesa, P. The TRiC/CCT Chaperonin and Its Role in Uncontrolled Proliferation. *Adv. Exp. Med. Biol.* **1243**, 21–40 (2020).

50. Rommelaere, H., De Neve, M., Melki, R., Vandekerckhove, J. & Ampe, C. The cytosolic class II chaperonin CCT recognizes delineated hydrophobic sequences in its target proteins. *Biochemistry* **38**, 3246–3257 (1999).

51. Kubota, S., Kubota, H. & Nagata, K. Cytosolic chaperonin protects folding intermediates

of Gβ from aggregation by recognizing hydrophobic β-strands. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 8360–8365 (2006).

52. Willison, K. R. The substrate specificity of eukaryotic cytosolic chaperonin CCT. *Philos. Trans. R. Soc. B Biol. Sci.* **373**, (2018).

53. Willison, K. R. The structure and evolution of eukaryotic chaperonin-containing TCP-1 and its mechanism that folds actin into a protein spring. *Biochem. J.* **475**, 3009–3034 (2018).

54. Gestaut, D., Limatola, A., Joachimiak, L. & Frydman, J. The ATP-powered gymnastics of TRiC/CCT: an asymmetric protein folding machine with a symmetric origin story. *Curr. Opin. Struct. Biol.* **55**, 50–58 (2019).

55. Jin, M., Liu, C., Han, W. & Cong, Y. *TRiC / CCT Chaperonin : Structure and Function*. (Springer International Publishing, 2019).

56. Horwich, A. L. & Fenton, W. A. Chaperonin-assisted protein folding: A chronologue. *Q. Rev. Biophys.* (2020).

57. Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. Atomic structure of the actin: DNase I complex. *Nature* **347**, 37–44 (1990).

58. Perrin, B. J. & Ervasti, J. M. The actin gene family: Function follows isoform. *Cytoskeleton* **67**, 630–634 (2010).

59. Nogales, E., Wolf, S. G. & Downing, K. H. Structure of the αβ tubulin dimer by electron crystallography. *Nature* **391**, 199–203 (1998).

60. Lu, K. Y. *et al.* Plasmodium chaperonin TRiC/CCT identified as a target of the antihistamine clemastine using parallel chemoproteomic strategy. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 5810–5817 (2020).

61. Hannich, J. T. *et al.* 1-Deoxydihydroceramide causes anoxic death by impairing chaperonin-mediated protein folding. *Nat. Metab.* **1**, 996–1008 (2019).

62. Sawaya, M. R. *et al.* Multiple crystal structures of actin dimers and their implications for interactions in the actin filament. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64**, 454–465 (2008).

63. Vainberg, I. E. *et al.* Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell* **93**, 863–873 (1998).

64. Siegers, K. *et al.* Compartmentation of protein folding in vivo: Sequestration of non-native polypeptide by the chaperonin-GimC system. *EMBO J.* **18**, 75–84 (1999).

65. López, V. *et al.* Identification of prefoldin amplification (1q23.3-q24.1) in bladder cancer using comparative genomic hybridization (CGH) arrays of urinary DNA. *J. Transl. Med.* **11**, 1–13 (2013).

66. Xu, Y. & Her, C. VBP1 facilitates proteasome and autophagy-mediated degradation of MutS homologue hMSH4. *FASEB J.* **27**, 4799–4810 (2013).

67. Abe, A. *et al.* Prefoldin plays a role as a clearance factor in preventing proteasome inhibitor-induced protein aggregation. *J. Biol. Chem.* **288**, 27764–27776 (2013).

68. Tashiro, E. *et al.* Prefoldin protects neuronal cells from polyglutamine toxicity by

preventing aggregation formation. *J. Biol. Chem.* **288**, 19958–19972 (2013).

69. Wang, P. *et al.* PFDN1, an indicator for colorectal cancer prognosis, enhances tumor cell proliferation and motility through cytoskeletal reorganization. *Med. Oncol.* **32**, 1–13 (2015).

70. Comyn, S. A., Young, B. P., Loewen, C. J. & Mayor, T. Prefoldin Promotes Proteasomal Degradation of Cytosolic Proteins with Missense Mutations by Maintaining Substrate Solubility. *PLoS Genet.* **12**, (2016).

71. Gestaut, D. *et al.* The Chaperonin TRiC / CCT Associates with Prefoldin through a Conserved Electrostatic Interface Essential for Cellular Proteostasis Article The Chaperonin TRiC / CCT Associates with Prefoldin through a Conserved Electrostatic Interface Essential for Cellu. *Cell* **177**, 751-765.e15 (2019).

72. Hayes, N. V. L., Jossé, L., Smales, C. M. & Carden, M. J. Modulation of phosducin-like protein 3 (PhLP3) levels promotes cytoskeletal remodelling in a MAPK and RhoA-dependent manner. *PLoS One* **6**, (2011).

73. Martín-Benito, J. *et al.* Structure of the complex between the cytosolic chaperonin CCT and phosducin-like protein. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 17410–17415 (2004).

74. Lukov, G. L., Hu, T., McLaughlin, J. N., Hamm, H. E. & Willardson, B. M. Phosducin-like protein acts as a molecular chaperone for G protein βγ dimer assembly. *EMBO J.* **24**, 1965–1975 (2005).

75. Stirling, P. C. *et al.* PhLP3 modulates CCT-mediated actin and tubulin folding via ternary complexes with substrates. *J. Biol. Chem.* **281**, 7012–7021 (2006).

76. Schmidt, A., Kunz, J. & Hall, M. N. TOR2 is required for organization of the actin cytoskeleton in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 13780–13785 (1996).

77. Kim, A. R. & Choi, K. W. TRiC/CCT chaperonins are essential for organ growth by interacting with insulin/TOR signaling in Drosophila. *Oncogene* **38**, 4739–4754 (2019).

78. Aylett, C. H. S. *et al.* Architecture of human mTOR complex 1. *Science (80-.).* **351**, 48–52 (2016).

79. Anaul Kabir, M. *et al.* Physiological effects of unassembled chaperonin Cct subunits in the yeast Saccharomyces cerevisiae. *Yeast* **22**, 219–239 (2005).

80. Amit, M. *et al.* Equivalent mutations in the eight subunits of the chaperonin CCT produce dramatically different cellular and gene expression phenotypes. *J. Mol. Biol.* **401**, 532–543 (2010).

81. Huttlin, E. L. *et al.* Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505–509 (2017).

82. Saxton, R. A. & Sabatini, D. M. mTOR Signaling in Growth, Metabolism, and Disease. *Cell* **168**, 960–976 (2017).

83. Wenta, T. *et al.* HtrA3 is a cellular partner of cytoskeleton proteins and TCP1α chaperonin. *J. Proteomics* **177**, 88–111 (2018).

84. Zhao, X. *et al.* TRiC subunits enhance BDNF axonal transport and rescue striatal atrophy in Huntington's disease. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5655–E5664 (2016).

85. Chen, X. Q. *et al.* T-complex protein 1-ring complex enhances retrograde axonal transport by modulating tau phosphorylation. *Traffic* **19**, 840–853 (2018).

86. Neef, D. W. *et al.* A direct regulatory interaction between chaperonin TRiC and stress-responsive transcription factor HSF1. *Cell Rep.* **9**, 955–966 (2014).

87. Banks, C. A. S. *et al.* Differential HDAC1/2 network analysis reveals a role for prefoldin/CCT in HDAC1/2 complex assembly. *Sci. Rep.* **8**, 1–13 (2018).

88. Stein, K. C., Kriel, A. & Frydman, J. Nascent Polypeptide Domain Topology and Elongation Rate Direct the Cotranslational Hierarchy of Hsp70 and TRiC / CCT. *Mol. Cell* **75**, 1–14 (2019).

89. Lukov, G. L. *et al.* Mechanism of assembly of G protein βγ subunits by protein kinase CK2-phosphorylated phosducin-like protein and the cytosolic chaperonin complex. *J. Biol. Chem.* **281**, 22261–22274 (2006).

90. Erickson, P. *et al.* Identification of breakpoints in t(8;21) acute myelogenous leukemia and isolation of a fusion transcript, AML1/ETO, with similarity to Drosophila segmentation gene, runt. *Blood* **80**, 1825–1831 (1992).

91. Roh, S. H. *et al.* Chaperonin TRiC/CCT modulates the folding and activity of leukemogenic fusion oncoprotein AML1-ETO. *J. Biol. Chem.* **291**, 4732–4741 (2016).

92. Trinidad, A. G. *et al.* Interaction of p53 with the CCT Complex Promotes Protein Folding and Wild-Type p53 Activity. *Mol. Cell* **50**, 805–817 (2013).

93. Bassiouni, R. *et al.* Chaperonin containing TCP-1 protein level in breast cancer cells predicts therapeutic application of a cytotoxic peptide. *Clin. Cancer Res.* **22**, 4366–4379 (2016).

94. Liu, Y. J., Kumar, V., Lin, Y. F. & Liang, P. H. Disrupting CCT-β:β-tubulin selectively kills CCT-βoverexpressed cancer cells through MAPKs activation. *Cell Death Dis.* **8**, 1–10 (2017).

95. Xu, G., Bu, S., Wang, X., Zhang, H. & Ge, H. Suppression of CCT3 inhibits the proliferation and migration in breast cancer cells. *Cancer Cell Int.* **20**, 1–12 (2020).

96. Showalter, A. E. *et al.* Investigating Chaperonin-Containing TCP-1 subunit 2 as an essential component of the chaperonin complex for tumorigenesis. *Sci. Rep.* **10**, 1–14 (2020).

97. Park, S. H. *et al.* Activating CCT2 triggers Gli-1 activation during hypoxic condition in colorectal cancer. *Oncogene* **39**, 136–150 (2020).

98. Hallal, S. *et al.* Extracellular Vesicles from Neurosurgical Aspirates Identifies Chaperonin Containing TCP1 Subunit 6A as a Potential Glioblastoma Biomarker with Prognostic Significance. *Proteomics* **19**, (2019).

99. Zhang, T., Shi, W., Tian, K. & Kong, Y. Chaperonin containing t-complex polypeptide 1 subunit 6A correlates with lymph node metastasis, abnormal carcinoembryonic antigen and poor survival profiles in non-small cell lung carcinoma. *World J. Surg. Oncol.* **18**, 1–10 (2020).

100. Carr, A. C. *et al.* Targeting chaperonin containing TCP1 (CCT) as a molecular therapeutic for small cell lung cancer. *Oncotarget* **8**, 110273–110288 (2017).

101. Qian, T. *et al.* High expression of chaperonin-containing TCP1 subunit 3 may induce dismal prognosis in multiple myeloma. *Pharmacogenomics J.* (2020) doi:10.1038/s41397-019-0145-6.

102. Liu, Y. *et al.* CCT3 acts upstream of YAP and TFCP2 as a potential target and tumour biomarker in liver cancer. *Cell Death Dis.* **10**, (2019).

103. Zeng, G. *et al.* Overexpressing CCT6A contributes to cancer cell growth by affecting the G1-To-S phase transition and predicts a negative prognosis in hepatocellular carcinoma. *Onco. Targets. Ther.* **12**, 10427–10439 (2019).

104. Yao, L., Zou, X. & Liu, L. The TCP1 ring complex is associated with malignancy and poor prognosis in hepatocellular carcinoma. *Int. J. Clin. Exp. Pathol.* **12**, 3329–3343 (2019).

105. Li, L. J. *et al.* Chaperonin containing TCP-1 subunit 3 is critical for gastric cancer growth. *Oncotarget* **8**, 111470–111481 (2017).

106. Lee, M. J. *et al.* Hereditary sensory neuropathy is caused by a mutation in the delta subunit of the cytosolic chaperonin-containing t-complex peptide-1 (Cct4) gene. *Hum. Mol. Genet.* **12**, 1917–1925 (2003).

107. Bouhouche, A., Benomar, A., Bouslam, N., Chkili, T. & Yahyaoui, M. Mutation in the epsilon subunit of the cytosolic chaperonin-containing t-complex peptide-1 (Cct5) gene causes autosomal recessive mutilating sensory neuropathy with spastic paraplegia. *J. Med. Genet.* **43**, 441–443 (2006).

108. Tam, S., Geller, R., Spiess, C. & Frydman, J. The chaperonin TRiC controls polyglutamine aggregation and toxicity through subunit-specific interactions. *Nat. Cell Biol.* **8**, 1155–1162 (2006).

109. Tam, S. *et al.* The chaperonin TRiC blocks a huntingtin sequence element that promotes the conformational switch to aggregation. *Nat. Struct. Mol. Biol.* **16**, 1279–1285 (2009).

110. Sontag, E. M. *et al.* Exogenous delivery of chaperonin subunit fragment ApiCCT1 modulates mutant Huntingtin cellular phenotypes. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 3077–3082 (2013).

111. Darrow, M. C. *et al.* Structural mechanisms of mutant huntingtin aggregation suppression by the synthetic chaperonin-like CCT5 complex explained by cryoelectron tomography. *J. Biol. Chem.* **290**, 17451–17461 (2015).

112. Shahmoradian, S. H. *et al.* TRiC's tricks inhibit huntingtin aggregation. *eLife* vol. 2 (2013).

113. Sot, B. *et al.* The chaperonin CCT inhibits assembly of α-synuclein amyloid fibrils by a specific, conformation-dependent interaction. *Scientific Reports* vol. 7 (2017).

114. Khabirova, E. *et al.* The TRiC/CCT chaperone is implicated in Alzheimer's disease based on patient GWAS and an RNAi screen in Abeta-expressing Caenorhabditis elegans. *PLoS One* **9**, e102985 (2014).

115. Pavel, M. *et al.* CCT complex restricts neuropathogenic protein aggregation via autophagy. *Nat. Commun.* **7**, (2016).

116. Kudryavtseva, S. S., Stroylova, Y. Y., Kurochkina, L. P. & Muronetz, V. I. The chaperonin TRiC is blocked by native and glycated prion protein. *Arch. Biochem. Biophys.* **683**, 108319 (2020).

117. Inoue, Y. *et al.* Chaperonin TRiC/CCT participates in replication of hepatitis C virus genome via interaction with the viral NS5B protein. *Virology* **410**, 38–47 (2011).

118. Kashuba, E., Pokrovskaja, K., Klein, G. & Szekely, L. Epstein-barr virus-encoded nuclear protein EBNA-3 interacts with the ε -subunit o f the T-complex, protein 1 chaperonin complex. *J. Hum. Virol.* **2**, 33–37 (1999).

119. Wang, Y., Uraki, R., Hwang, J. & Fikrig, E. TRiC / CCT Complex, a Binding Partner of NS1 Protein, Supports the Replication of Zika Virus in Both Mammalians and Mosquitoes. *Viruses* **12**, 519 (2020).

120. Zhang, J. *et al.* Cellular Chaperonin CCT Contributes to Rabies Virus Replication during Infection. *J. Virol.* **87**, 7608–7621 (2013).

121. Hong, S. *et al.* Type D retrovirus Gag polyprotein interacts with the cytosolic chaperonin TRiC. *J. Virol.* **75**, 2526–2534 (2001).

122. Jäger, S. *et al.* Global landscape of HIV-human protein complexes. *Nature* **481**, 365–370 (2012).

123. Joachimiak, L. A., Walzthoeni, T., Liu, C. W., Aebersold, R. & Frydman, J. The structural basis of substrate recognition by the eukaryotic chaperonin TRiC/CCT. *Cell* **159**, 1042–1055 (2014).

124. Knowlton, J. J. *et al.* The TRiC chaperonin controls reovirus replication through outer-capsid folding. *Nat. Microbiol.* **3**, 481–493 (2018).

125. Wang, Q. *et al.* Cdc20 and molecular chaperone CCT2 and CCT5 are required for the Muscovy duck reovirus p10.8-induced cell cycle arrest and apoptosis. *Vet. Microbiol.* **235**, 151–163 (2019).

126. Lingappa, J. R. *et al.* A eukaryotic cytosolic chaperonin is associated with a high molecular weight intermediate in the assembly of hepatitis B virus capsid, a multimeric particle. *J. Cell Biol.* **125**, 99–112 (1994).

127. Steinemann, M., Schlosser, A., Jank, T. & Aktories, K. The chaperonin TRiC/CCT is essential for the action of bacterial glycosylating protein toxins like Clostridium difficile toxins A and B. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 9580–9585 (2018).

128. Lewis, V. A., Hynes, G. M., Zheng, D., Saibil, H. & Willison, K. T-complex polypeptide-1 is a subunit of a heteromeric particle in the eukaryotic cytosol. *Nature* **358**, 249–252 (1992).

129. Cong, Y. *et al.* Symmetry-free cryo-EM structures of the chaperonin TRiC along its ATPase-driven conformational cycle. *EMBO J.* **31**, 720–730 (2012).

130. Cong, Y. *et al.* 4.0-Å resolution cryo-EM structure of the mammalian chaperonin TRiC/CCT reveals its unique subunit arrangement. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 4967–4972 (2010).

131. Dekker, C. *et al.* The crystal structure of yeast CCT reveals intrinsic asymmetry of eukaryotic cytosolic chaperonins. *EMBO J.* **30**, 3078–3090 (2011).

132. Leitner, A. *et al.* The molecular architecture of the eukaryotic chaperonin TRiC/CCT. *Structure* **20**, 814–825 (2012).

133. Kalisman, N., Schröder, G. F. & Levitt, M. The Crystal Structures of the Eukaryotic Chaperonin CCT Reveal its Functional Partitioning. *Structure* **21**, 540–549 (2013).

134. Kim, S., Willison, K. R. & Horwich, A. L. Cystosolic chaperonin subunits have a conserved ATPase domain but diverged polypeptide-binding domains. *Trends Biochem. Sci.* **19**, 543–548 (1994).

135. Pereira, J. H. *et al.* Mechanism of nucleotide sensing in group II chaperonins. *EMBO J.* **31**, 731–740 (2012).

136. Muñoz, I. G. *et al.* Crystal structure of the open conformation of the mammalian chaperonin CCT in complex with tubulin. *Nat. Struct. Mol. Biol.* **18**, 14–20 (2011).

137. Clare, D. K., Bakkes, P. J., Van Heerikhuizen, H., Van Der Vies, S. M. & Saibil, H. R. Chaperonin complex with a newly folded protein encapsulated in the folding chamber. *Nature* **457**, 107–110 (2009).

138. Zang, Y. *et al.* Staggered ATP binding mechanism of eukaryotic chaperonin TRiC (CCT) revealed through high-resolution cryo-EM. *Nat. Struct. Mol. Biol.* **23**, 1083–1091 (2016).

139. Srikakulam, R. & Winkelmann, D. A. Myosin II folding is mediated by a molecular chaperonin. *J. Biol. Chem.* **274**, 27265–27273 (1999).

140. Rüßmann, F. *et al.* Folding of large multidomain proteins by partial encapsulation in the chaperonin TRiC/CCT. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 21208–21215 (2012).

141. Liou, A. K. F. & Willison, K. R. Elucidation of the subunit orientation in CCT (chaperonin containing TCP1) from the subunit composition of CCT micro-complexes. *EMBO J.* **16**, 4311–4316 (1997).

142. Kalisman, N., Adams, C. M. & Levitt, M. Subunit order of eukaryotic TRiC/CCT chaperonin by cross-linking, mass spectrometry, and combinatorial homology modeling. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2884–2889 (2012).

143. Zang, Y. *et al.* Development of a yeast internal-subunit eGFP labeling strategy and its application in subunit identification in eukaryotic group II chaperonin TRiC/CCT. *Sci. Rep.* **8**, 2–10 (2018).

144. Wang, H., Han, W., Takagi, J. & Cong, Y. Yeast Inner-Subunit PA–NZ-1 Labeling Strategy for Accurate Subunit Identification in a Macromolecular Complex through Cryo-EM Analysis. *J. Mol. Biol.* **430**, 1417–1425 (2018).

145. Tang, Y. C., Chang, H. C., Chakraborty, K., Hartl, F. U. & Hayer-Hartl, M. Essential role of the chaperonin folding compartment in vivo. *EMBO J.* **27**, 1458–1468 (2008).

146. Meyer, A. S. *et al.* Closing the folding chamber of the eukaryotic chaperonin requires the transition state of ATP hydrolysis. *Cell* **113**, 369–381 (2003).

147. Reissmann, S. *et al.* A Gradient of ATP Affinities Generates an Asymmetric Power Stroke Driving the Chaperonin TRIC/CCT Folding Cycle. *Cell Rep.* **2**, 866–877 (2012).

148. Jiang, Y. *et al.* Sensing cooperativity in ATP hydrolysis for single multisubunit enzymes in solution. *Proc. Natl. Acad. Sci. U. S. A.* **108**, (2011).

149. Cuéllar, J. *et al.* Structural and functional analysis of the role of the chaperonin CCT in mTOR complex assembly. *Nat. Commun.* 1–14 (2019) doi:10.1038/s41467-019-10781-1.

150. Kafri, G., Willison, K. R. & Horovitz, A. Nested allosteric interactions in the cytoplasmic chaperonin containing TCP-1. *Protein Sci.* **10**, 445–449 (2001).

151. Shimon, L., Hynes, G. M., McCormack, E. A., Willison, K. R. & Horovitz, A. ATP-Induced Allostery in the Eukaryotic Chaperonin CCT Is Abolished by the Mutation G345D in CCT4 that Renders Yeast Temperature-Sensitive for Growth. *J. Mol. Biol.* **377**, 469–477 (2008).

152. Kafri, G. & Horovitz, A. Transient kinetic analysis of ATP-induced allosteric transitions in the eukaryotic chaperonin containing TCP-1. *J. Mol. Biol.* **326**, 981–987 (2003).

153. Reissmann, S., Parnot, C., Booth, C. R., Chiu, W. & Frydman, J. Essential function of the built-in lid in the allosteric regulation of eukaryotic and archaeal chaperonins. *Nat. Struct. Mol. Biol.* **14**, 432–440 (2007).

154. Douglas, N. R. *et al.* Dual action of ATP hydrolysis couples lid closure to substrate release into the group II chaperonin chamber. *Cell* **144**, 240–252 (2011).

155. Lin, P., Cardillo, T. S., Richard, L. M., Segel, G. B. & Sherman, F. Analysis of mutationally altered forms of the Cct6 subunit of the chaperonin from Saccharomyces cerevisiae. *Genetics* **147**, 1609–1633 (1997).

156. Llorca, O. *et al.* Eukaryotic type II chaperonin CCT interacts with actin through specific subunits. *Nature* **402**, 693–696 (1999).

157. Llorca, O. *et al.* Eukaryotic chaperonin CCT stabilizes actin and tubulin folding intermediates in open quasi-native conformations. *EMBO J.* **19**, 5971–5979 (2000).

158. Martín-Benito, J. *et al.* Structure of eukaryotic prefoldin and of its complexes with unfolded actin and the cytosolic chaperonin CCT. *EMBO J.* **21**, 6377–6386 (2002).

159. Pappenberger, G. *et al.* Crystal Structure of the CCT g Apical Domain : Implications for Substrate Binding to the Eukaryotic Cytosolic Chaperonin. *J. Mol. Biol.* **318**, 1367–1379 (2002).

160. Joachimiak, L. A., Walzthoeni, T., Liu, C. W., Aebersold, R. & Frydman, J. The Structural Basis of Substrate Recognition by the Eukaryotic Chaperonin TRiC/CCT. *Cell* **159**, 1042–1055 (2014).

161. Spiess, C., Miller, E. J., McClellan, A. J. & Frydman, J. Identification of the TRiC/CCT Substrate Binding Sites Uncovers the Function of Subunit Diversity in Eukaryotic Chaperonins. *Mol. Cell* **24**, 25–37 (2006).

162. Balchin, D., Mili, G., Strauss, M., Hayer-hartl, M. & Hartl, F. U. Pathway of Actin Folding Directed by the Eukaryotic Chaperonin TRiC. *Cell* **174**, 1507–1521 (2018).

163. Tian, G., Vainberg, I. E., Tap, W. D., Lewis, S. A. & Cowan, N. J. Specificity in chaperonin-mediated protein folding. *Nature* **375**, 250–253 (1995).

164. Pappenberger, G., Mccormack, E. A. & Willison, K. R. Quantitative Actin Folding Reactions using Yeast CCT Purified via an Internal Tag in the CCT3 / γ Subunit. *J. Mol. Biol.* **360**, 484–496 (2006).

165. Balchin, D., Miličić, G., Strauss, M., Hayer-Hartl, M. & Hartl, F. U. Pathway of Actin Folding Directed by the Eukaryotic Chaperonin TRiC. *Cell* **174**, 1507-1521.e16 (2018).

166. Engen, J. R. & Smith, D. L. Investigating Protein Structure and Dynamics by Hydrogen Exchange MS. *Anal. Chem.* **73**, 256 A-265 A (2001).

167. Wales, T. E. & Engen, J. R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom. Rev.* **25**, 158–170 (2006).

168. Melki, R., Batelier, G., Soulié, S. & Williams Jr., R. C. Cytoplasmic Chaperonin Containing TCP-1: Structural and Functional Characterization. *Biochemistry* 5817–5826 (1997).

169. Plimpton, R. L. *et al.* Structures of the Gbeta-CCT and PhLP1-Gbeta-CCT complexes reveal a mechanism for G-protein beta-subunit folding and Gbetagamma dimer assembly. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 2413–2418 (2015).

170. Roh, S.-H., Kasembeli, M. M., Galaz-Montoya, J. G., Chiu, W. & Tweardy, D. J. Chaperonin TRiC/CCT Recognizes Fusion Oncoprotein AML1-ETO through Subunit-Specific Interactions. *Biophys. J.* **110**, 2377–2385 (2016).

171. Cuéllar, J. *et al.* Structural and functional analysis of the role of the chaperonin CCT in mTOR complex assembly. *Nat. Commun.* **10**, 1–14 (2019).

172. Siegert, R., Leroux, M. R., Scheufler, C., Hartl, F. U. & Moarefi, I. Structure of the Molecular Chaperone Prefoldin. *Cell* **103**, 621–632 (2000).

173. Zakeri, B. *et al.* Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. U. S. A.* **109**, (2012).

174. Agudelo, D. *et al.* Marker-free coselection for CRISPR-driven genome editing in human cells. *Nat. Methods* **14**, 615–620 (2017).

175. Pardon, E. *et al.* A general protocol for the generation of Nanobodies for structural biology. *Nat. Protoc.* **9**, 674–693 (2014).

176. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

177. Cox, J. *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).

178. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).

179. Förster, S., Apostol, L. & Bras, W. Scatter: Software for the analysis of nano-and mesoscale small-angle scattering. *J. Appl. Crystallogr.* **43**, 639–646 (2010).

180. Franke, D. *et al.* ATSAS 2.8: A comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).

181. de la Rosa-Trevin, J. M. *et al.* Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.* **195**, 93–99 (2016).

182. Zheng, S. Q. *et al.* MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).

183. De la Rosa-Trevín, J. M. *et al.* Xmipp 3.0: An improved software suite for image processing in electron microscopy. *J. Struct. Biol.* **184**, 321–328 (2013).

184. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).

185. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

186. Tang, G. *et al.* EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).

187. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

188. Wagner, T. *et al.* SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 1–13 (2019).

189. Abrishami, V. *et al.* Localized reconstruction in Scipion expedites the analysis of symmetry mismatches in cryo-EM data. *Prog. Biophys. Mol. Biol.* (2020) doi:10.1016/j.pbiomolbio.2020.05.004.

190. Waterhouse, A. *et al.* SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).

191. Fioravanti, A. *et al.* Structure of S-layer protein Sap reveals a mechanism for therapeutic intervention in anthrax. *Nat. Microbiol.* **4**, 1805–1814 (2019).

192. Adib, R. *et al.* Mitotic phosphorylation by NEK6 and NEK7 reduces the microtubule affinity of EML4 to promote chromosome congression. *Sci. Signal.* **12**, (2019).

193. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

194. Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 531–544 (2018).

195. Ferreyra, R. G. & Frydman, J. Purification of the Cytosolic Chaperonin TRiC from Bovine Testis. *Chaperonin Protoc.* **140**, 153–160 (2003).

196. Baykov, A. A., Evtushenko, O. A. & Avaeva, S. M. A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. *Anal. Biochem.* **171**, 266–270 (1988).

197. Bloemendal, H. *et al.* Ageing and vision: Structure, stability and function of lens crystallins. *Prog. Biophys. Mol. Biol.* **86**, 407–485 (2004).

198. Knee, K. M., Sergeeva, O. A. & King, J. A. Human TRiC complex purified from HeLa cells contains all eight CCT subunits and is active in vitro. *Cell Stress Chaperones* **18**, 137–144 (2013).

199. Acosta-Sampson, L. & King, J. Partially folded aggregation intermediates of human gammaD-, gammaC-, and gammaS-crystallin are recognized and bound by human alphaB-crystallin chaperone. *J. Mol. Biol.* **401**, 134–152 (2010).

200. Chari, A. *et al.* ProteoPlex: Stability optimization of macromolecular complexes by sparse-matrix screening of chemical space. *Nat. Methods* **12**, 859–865 (2015).

201. Lo, M. C. *et al.* Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal. Biochem.* **332**, 153–159 (2004).

202. De Rosier, D. J. & Klug, A. Reconstruction of three dimensional fiber structures from orthogonal projections. *Nature* **217**, 130–134 (1968).

203. Nogales, E. The development of cryo-EM into a mainstream structural biology technique. *Nat. Methods* **13**, 24–27 (2016).

204. Ohi, M., Li, Y., Cheng, Y. & Walz, T. Negative staining and image classification - Powerful tools in modern electron microscopy. *Biol. Proced. Online* **6**, 23–34 (2004).

205. Taylor, K. A. & Glaeser, R. M. Electron Diffraction of Frozen , Hydrated Protein Crystals Sunday and Workday Variations in Photochemical Air Pollutants in New Jersey and New York. *Science (80-.).* **186**, 1036–1037 (1974).

206. Dubochet, J., Lepault, J., Freeman, R., Berriman, J. A. & Homo, J.-C. Electron microscopy of frozen water and aqueous solutions. *J. Microsc.* **128**, 219–237 (1982).

207. Penczek, P., Grassucci, R. & Frank, J. The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy* **53**, 251–270 (1994).

208. Crowther, R. A. Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **261**, 221–230 (1971).

209. Van Heel, M. Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy* **21**, 111–123 (1987).

210. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).

211. Kimanius, D., Forsberg, B. O., Scheres, S. H. W. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUS in RELION-2. *Elife* **5**, 1–21 (2016).

212. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, 1–22 (2018).

213. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013).

214. Pereira, J. H. *et al.* Structure of the human TRiC/CCT Subunit 5 associated with hereditary sensory neuropathy. *Sci. Rep.* **7**, 3673 (2017).

215. Petoukhov, M. V. & Svergun, D. I. Applications of small-angle X-ray scattering to biomacromolecular solutions. *Int. J. Biochem. Cell Biol.* **45**, 429–437 (2013).

216. Malaby, A. W. *et al.* Methods for analysis of size-exclusion chromatography-small-angle X-ray scattering and reconstruction of protein scattering. *J. Appl. Crystallogr.* **48**, 1102–1113 (2015).

217. Bernadó, P., Shimizu, N., Zaccai, G., Kamikubo, H. & Sugiyama, M. Solution scattering approaches to dynamical ordering in biomolecular systems. *Biochim. Biophys. Acta - Gen. Subj.* **1862**, 253–274 (2018).

218. Feldman, D. E., Spiess, C., Howard, D. E. & Frydman, J. Tumorigenic mutations in VHL disrupt folding in vivo by interfering with chaperonin binding. *Mol. Cell* **12**, 1213–1224 (2003).

219. Machida, K. *et al.* Reconstitution of the human chaperonin CCT by co-expression of the eight distinct subunits in mammalian cells. *Protein Expr. Purif.* **82**, 61–69 (2012).

220. Pappenberger, G., McCormack, E. A. & Willison, K. R. Quantitative Actin Folding Reactions using Yeast CCT Purified via an Internal Tag in the CCT3/γ Subunit. *J. Mol. Biol.* **360**, 484–496 (2006).

221. Wang, M., Herrmann, C. J., Simonovic, M., Szklarczyk, D. & von Mering, C. Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines. *Proteomics* **15**, 3163–3168 (2015).

222. Berger, I., Fitzgerald, D. J. & Richmond, T. J. Baculovirus expression system for heterologous multiprotein complexes. *Nat. Biotechnol.* **22**, 1583–1587 (2004).

223. Sari, D. *et al.* The MultiBac Baculovirus/Insect Cell Expression Vector System for Producing Complex Protein Biologics. *Adv. Exp. Med. Biol.* **896**, 199–215 (2016).

224. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-.).* **337**, 816–821 (2012).

225. Anzalone, A. V, Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).

226. Dalvai, M. *et al.* A Scalable Genome-Editing-Based Approach for Mapping Multiprotein Complexes in Human Cells. *Cell Rep.* **13**, 621–633 (2015).

227. Paul, D. M., Beuron, F., Sessions, R. B., Brancaccio, A. & Bigotti, M. G. Internal (His)6-tagging delivers a fully functional hetero-oligomeric class II chaperonin in high yield. *Sci. Rep.* **6**, 1–10 (2016).

228. Einhauer, A. & Jungbauer, A. The FLAG™ peptide, a versatile fusion tag for the purification of recombinant proteins. *J. Biochem. Biophys. Methods* **49**, 455–465 (2001).

229. Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J. & Mateo, J. L. CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS One* **10**, 1–11 (2015).

230. Pleiner, T. *et al.* Nanobodies: Site-specific labeling for super-resolution imaging, rapid epitope- mapping and native protein complex isolation. *Elife* **4**, 1–21 (2015).

231. Kromann-Hansen, T. *et al.* A Camelid-derived Antibody Fragment Targeting the Active Site of a Serine Protease Balances between Inhibitor and Substrate Behavior. *J. Biol. Chem.* **291**, 15156–15168 (2016).

232. Uchanski, T., Pardon, E. & Steyaert, J. Nanobodies to study protein conformational states. *Curr. Opin. Struct. Biol.* **60**, 117–123 (2020).

233. Genst, E. J. De *et al.* Structure and Properties of a Complex of α -Synuclein and a Single-Domain Camelid Antibody. *J. Mol. Biol.* **402**, 326–343 (2010).

234. Petersen, R. L. Strategies using bio-layer interferometry biosensor technology for vaccine research and development. *Biosensors* **7**, (2017).

235. Kuhlbrandt, W. The Resolution Revolution. *Science (80-.).* **343**, 1443–1444 (2014).

236. Kuhlbrandt, W. Cryo-EM enters a new era. *Elife* **3**, e03665 (2014).

237. Passmore, L. A. & Russo, C. J. Specimen preparation for high-resolution cryo-EM. *Methods Enzym.* **579**, 51–86 (2016).

238. Thompson, R. F., Walker, M., Siebert, C. A., Muench, S. P. & Ranson, N. A. An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods* **100**, 3–15 (2016).

239. Snijder, J. *et al.* Vitrification after multiple rounds of sample application and blotting improves particle density on cryo-electron microscopy grids. *J. Struct. Biol.* **198**, 38–42 (2017).

240. Mattson, G. *et al.* A practical approach to crosslinking. *Mol. Biol. Rep.* **17**, 167–183 (1993).

241. Chen, Z. A. & Rappsilber, J. Quantitative cross-linking/mass spectrometry to elucidate structural changes in proteins and their complexes. *Nat. Protoc.* **14**, 171–201 (2019).

242. Liu, F., Rijkers, D. T. S., Post, H. & Heck, A. J. R. Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat. Methods* **12**, 1179–1184 (2015).

243. Löwe, J., Li, H., Downing, K. H. & Nogales, E. Refined structure of αβ-tubulin at 3.5 Å resolution. *J. Mol. Biol.* **313**, 1045–1057 (2001).

244. Alvarado-Kristensson, M. Γ-Tubulin As a Signal-Transducing Molecule and Meshwork With Therapeutic Potential. *Signal Transduct. Target. Ther.* **3**, 1–6 (2018).

245. Durer, Z. A. O. *et al.* Structural states and dynamics of the D-Loop in actin. *Biophys. J.* **103**, 930–939 (2012).

246. Lou, X., Bao, R., Zhou, C. Z. & Chen, Y. Structure of the thioredoxin-fold domain of human phosducin-like protein 2. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **65**, 67–70 (2009).

247. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: A protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394 (2015).

248. Martin, J. L. Thioredoxin -a fold for all reasons. *Structure* **3**, 245–250 (1995).

249. Joachimiak, L. A., Walzthoeni, T., Liu, C. W., Aebersold, R. & Frydman, J. The structural basis of substrate recognition by the eukaryotic chaperonin TRiC/CCT. *Cell* **159**, 1042–1055 (2014).

250. Muñoz-Lasso, D. C., Romá-Mateo, C., Pallardó, F. V. & Gonzalez-Cabo, P. Much More Than a Scaffold: Cytoskeletal Proteins in Neurological Disorders. *Cells* **9**, 358 (2020).

251. Ramaekers, F. C. S. & Bosman, F. T. The cytoskeleton and disease. *J. Pathol.* **204**, 351–354 (2004).

252. Fife, C. M., McCarroll, J. A. & Kavallaris, M. Movers and shakers: Cell cytoskeleton in cancer metastasis. *Br. J. Pharmacol.* **171**, 5507–5523 (2014).

253. Francis, S. & Antonipillai, J. Cytoskeletal molecules play a major role in cancer progression. *Insights Biomed.* **2**, 9 (2017).

254. Pasquier, E. & Kavallaris, M. Microtubules: A dynamic target in cancer therapy. *IUBMB Life* **60**, 165–170 (2008).

255. Sontag, E. M. *et al.* Exogenous delivery of chaperonin subunit fragment ApiCCT1 modulates mutant Huntingtin cellular phenotypes. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 3077–3082 (2013).

256. Neef, D. W., Turski, M. L. & Thiele, D. J. Modulation of heat shock transcription factor 1 as a therapeutic target for small molecule intervention in neurodegenerative disease. *PLoS Biol.* **8**, (2010).

257. Zegzouti, H., Zdanovskaia, M., Hsiao, K. & Goueli, S. A. ADP-Glo: A bioluminescent and homogeneous adp monitoring assay for Kinases. *Assay Drug Dev. Technol.* **7**, 560–572 (2009).

258. Bakthavatsalam, D. *et al.* Chaperonin-containing TCP-1 complex directly binds to the cytoplasmic domain of the LOX-1 receptor. *FEBS Lett.* **588**, 2133–2140 (2014).

259. Farr, G. W., Scharl, E. C., Schumacher, R. J., Sondek, S. & Horwich, A. L. Chaperonin-mediated folding in the eukaryotic cytosol proceeds through rounds of release of native and nonnative forms. *Cell* **89**, 927–937 (1997).

260. Feldman, D. E., Thulasiraman, V., Ferreyra, R. G. & Frydman, J. Formation of the VHL-elongin BC tumor suppressor complex is mediated by the chaperonin TRiC. *Mol. Cell* **4**, 1051–1061 (1999).

261. Freund, A. *et al.* Proteostatic control of telomerase function through TRiC-mediated folding of TCAB1. *Cell* **159**, 1389–1403 (2014).

262. Gallina, I. *et al.* Cmr1/WDR76 defines a nuclear genotoxic stress body linking genome integrity and protein quality control. *Nat. Commun.* **6**, 1–16 (2015).

263. Gao, X. *et al.* Splice isoforms of phosducin-like protein control the expression of heterotrimeric G proteins. *J. Biol. Chem.* **288**, 25760–25768 (2013).

264. Guenther, M. G., Yu, J., Kao, G. D., Yen, T. J. & Lazar, M. A. Assembly of the SMRT-histone deacetylase 3 repression complex requires the TCP-1 ring complex. *Genes Dev.* **16**, 3130–3135 (2002).

265. Huang, R. *et al.* New insights into the functions and localization of nuclear CCT protein complex in K562 leukemia cells. *Proteomics - Clin. Appl.* **6**, 467–475 (2012).

266. Kaisari, S., Sitry-Shevah, D., Miniowitz-Shemtov, S., Teichner, A. & Hershko, A. Role of CCT chaperonin in the disassembly of mitotic checkpoint complexes. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 956–961 (2017).

267. Kasembeli, M. *et al.* Modulation of STAT3 Folding and Function by TRiC/CCT Chaperonin. *PLoS Biol.* **12**, (2014).

268. Litterman, N. *et al.* An OBSL1-CUl7Fbxw8 ubiquitin ligase signaling mechanism regulates golgi morphology and dendrite patterning. *PLoS Biol.* **9**, 17–19 (2011).

269. Liu, X. *et al.* CCT Chaperonin Complex Is Required for the Biogenesis of Functional Plk1. *Mol. Cell. Biol.* **25**, 4993–5010 (2005).

270. Pim Pijnappel, W. W. M. *et al.* The S. cerevisiae SET3 complex includes two histone deacetylases, Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation gene program. *Genes Dev.* **15**, 2991–3004 (2001).

271. Roobol, A. *et al.* The chaperonin CCT interacts with and mediates the correct folding and activity of three subunits of translation initiation factor eIF3: B, i and h. *Biochem. J.* **458**, 213–224 (2014).

272. Scior, A. *et al.* Complete suppression of Htt fibrilization and disaggregation of Htt fibrils by

a trimeric chaperone complex. *EMBO J.* **37**, 282–299 (2018).

273. Wells, C. A., Dingus, J. & Hildebrandt, J. D. Role of the chaperonin CCT/TRiC complex in G protein βγ-dimer assembly. *J. Biol. Chem.* **281**, 20221–20232 (2006).

274. Won, K.-A., Schumacher, R. J., Farr, G. W., Horwich, A. L. & Reed, S. I. Maturation of Human Cyclin E Requires the Function of Eukaryotic Chaperonin CCT. *Mol. Cell. Biol.* **18**, 7584–7589 (1998).

275. Ying, Z. *et al.* CCT6A suppresses SMAD2 and promotes prometastatic TGF-β signaling. *J. Clin. Invest.* **127**, 1725–1740 (2017).

276. Zebol, J. R. *et al.* The CCT/TRiC chaperonin is required for maturation of sphingosine kinase 1. *Int. J. Biochem. Cell Biol.* **41**, 822–827 (2009).

Appendix (submitted as separate files)

Appendix 1: List of known CCT substrates grouped by cellular function. References used: [258] [87] [38] [100] [85] [149] [48] [259] [260] [261] [262] [263] [46] [264] [47] [121] [265] [81] [117] [122] [160] [266] [267] [118] [124] [116] [105] [126] [268] [269] [102] [73] [168] [136] [97] [270] [169] [170] [271] [272] [113] [139] [127] [75] [92] [125] [273] [83] [274] [36] [275] [276] [103] [120].

Appendix 2: Peptides identified by LC-MS/MS in purified TRiC sample. Data include significance values, protein names, sequence coverage, intensity, and other peptide information.