Structural basis of centrosome assembly in flies

A thesis submitted to the **University of Oxford** by

ANDREAS F. M. HAENSELE

in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in Pathology

Lincoln College, Oxford.

Trinity Term 2020

Copyright © Haensele 2020.

Abstract

Structural basis of centrosome assembly in flies

Andreas Florian Matthias Haensele, Lincoln College

Submitted for the degree of Doctor of Philosophy in Pathology, Trinity Term 2020

Centrosomes are the main microtubule-organising centres in many animal cells. They consist of a pair of centrioles surrounded by a dense mass of protein known as the pericentriolar material. As the cell enters mitosis, this pericentriolar material expands dramatically in a process called centrosome maturation. In flies, the protein Centrosomin (Cnn) assembles into a network that acts as a scaffold for the recruitment of the hundreds of other proteins making up a mature centrosome. Cnn network formation is known to depend on phosphorylation of a central region within Cnn. However, it is unclear how Cnn assembles into a network and how this process is controlled.

In this thesis it is shown that Cnn network formation *in vitro* depends on an interaction between the conserved C-terminus of Cnn and the phospho-regulated central region. The crystal structure of part of this central region is solved, revealing that it forms a dimer of α-helical hairpins that are sterically incompatible with binding of the Cnn C-terminus. The structure is therefore proposed to constitute an "inactive" state. Furthermore, the structure suggests a mechanism wherein phosphorylation acts by disrupting the hairpin fold, thereby exposing the Cnn C-terminus binding site and hence allowing formation of an "assembly-competent" state where the Cnn C-terminus is bound. In addition, the calcium-binding protein Calmodulin is shown to crosslink the Cnn C-terminus *in crystallo*. However, Calmodulin inhibits network assembly of Cnn constructs in solution, possibly by preventing the assembly-competent state from forming a network. Taken together, this thesis draws on structural and functional data to provide important atomic-level insights into the mechanism of mitotic centrosome assembly in flies.

Publication resulting from this study

Feng, Z.*, Caballe, A.*, Wainman, A.*, Johnson, S.*, Haensele, A.F.M., Cottee, M.A., Conduit, P.T., Lea, S.M. and Raff, J.W., *Structural Basis for Mitotic Centrosome Assembly in Flies.* **Cell**, 2017. 169(6): p. 1078-1089.

*these authors contributed equally

The author of this thesis is underlined.

Table of contents

Abstract ..... **i**
Publication resulting from this study ..... **ii**
Table of contents ..... **iii**
Chapter 1 - Introduction **1**
1.1 Centrosome function in health and disease 1
1.1.1 Function and conservation *1*
1.1.2 The centrosome cycle *3*
1.1.3 Disease *6*
1.2 Centriole structure and assembly 9
1.2.1 Centriole ultrastructure *9*
1.2.2 Centriole assembly pathway *10*
1.3 The interphase pericentriolar material 13
1.4 The mitotic pericentriolar material 14
1.4.1 Centrosome maturation in flies and vertebrates *14*
1.4.2 Network assembly dynamics *18*
1.4.3 Phosphorylation-dependent network assembly *18*
1.4.4 Cnn flares *22*
1.4.5 Microtubule nucleation *24*
1.4.6 Centrosome maturation in C. elegans *26*
1.4.7 Centrosome size control *30*
1.5 Calmodulin and the centrosome 33
1.6 Thesis aims 35
Chapter 2 - Materials and methods **36**
2.1 Molecular cloning 36
2.1.1 Primers and polymerase chain reaction *36*
2.1.2 Restriction enzyme digestion and DNA visualisation *39*
2.1.3 Gibson assembly and site-directed mutagenesis *40*
2.1.4 Making chemically competent cells *41*
2.1.5 Bacterial transformation *42*
2.1.6 Plasmids used in this work *43*
2.2 Protein production 46

2.2.1 Protein expression in bacteria ... *46*
2.2.2 Standard protein purification ... *46*
2.2.3 Protease production ... *49*
2.2.4 Protein gel electrophoresis ... *50*
2.3 SEC-MALS ... 52
2.4 Analytical SEC ... 52
2.5 Phosphorylation assay ... 53
2.6 Fluorescence microscopy ... 53
2.7 X-ray crystallography ... 54
2.7.1 Diffraction of X-rays by crystals ... *54*
2.7.2 Data reduction and quality indicators ... *55*
2.7.3 Solving the phase problem ... *56*
2.7.4 Model refinement and validation ... *57*
2.7.5 Practice ... *58*
Chapter 3 - $Cnn^{Q403\text{-}H608}$ assembles into micron-scale networks upon binding to the Cnn C-terminus via its zipper-like motif ... 60
3.1 Introduction ... 60
3.2 $Cnn^{D490\text{-}K544}$:$Cnn^{H1082\text{-}L1148}$ interaction can be disrupted via site-directed mutagenesis ... 64
3.2.1 Purification of wild-type $Cnn^{H1082\text{-}L1148}$ and $Cnn^{D490\text{-}K544}$ ... *64*
3.2.2 Purification and characterisation of $Cnn^{D490\text{-}K544}$ mutants ... *68*
3.3 $Cnn^{Q403\text{-}H608}$ forms micron-scale assemblies upon binding to $Cnn^{H1082\text{-}L1148}$ ... 75
3.4 Cnn C-terminus interaction with $Cnn^{Q403\text{-}H608}$ appears to be specific ... 79
3.4.1 Purification of constructs almost completely covering Cnn ... *79*
3.4.2 Functional characterisation of novel Cnn constructs ... *88*
3.5 Discussion ... 92
Chapter 4 - Calmodulin crosslinks $Cnn^{H1082\text{-}L1148}$ *in crystallo* but prevents network assembly in solution ... 97
4.1 Introduction ... 97
4.2 Purification of Calmodulin ... 101
4.3 $Cnn^{S136\text{-}S205}$ interacts with Calmodulin *in vitro* ... 103
4.3.1 $Cnn^{S136\text{-}S205}$ purification ... *103*
4.3.2 Functional analysis of $Cnn^{S136\text{-}S205}$ ... *105*
4.4 $Cnn^{S136\text{-}S205}$ forms a dimeric coiled-coil *in crystallo* ... 109
4.4.1 $Cnn^{S136\text{-}S205}$ structure determination ... *109*
4.4.2 $Cnn^{S136\text{-}S205}$ structure analysis ... *114*

4.5 Calmodulin:Cnn$^{H1082-L1148}$ interaction is calcium-sensitive 118
4.6 Calmodulin crosslinks Cnn$^{H1082-L1148}$ dimers *in crystallo* 121
4.6.1 Cnn$^{H1082-V1140}$ structure determination 121
4.6.2 Calmodulin:Cnn$^{H1082-L1148}$ crystal screening 125
4.6.3 Calmodulin:Cnn$^{H1082-L1148}$ structure determination 132
4.6.4 Calmodulin:Cnn$^{H1082-L1148}$ structure analysis 136
4.7 Calmodulin prevents network formation of Cnn constructs *in vitro* 145
4.8 Discussion 149
Chapter 5 - A Cnn$^{D490-N579}$ fold that is incompatible with phosphorylation of S567 and binding of the Cnn C-terminus 156
5.1 Introduction 156
5.2 A library of mNG-Cnn^{D490-x} constructs 160
5.2.1 Design and purification of construct library 160
5.2.2 Functional analysis of construct library 167
5.3 mNG-Cnn$^{D490-N579}$ forms 2:2 complexes with the Cnn C-terminus 169
5.4 Cnn$^{D490-N579}$ forms a dimer of α-helical hairpins *in crystallo* 174
5.4.1 Cnn$^{D490-N579}$ purification and crystallisation 174
5.4.2 Cnn$^{D490-N579}$ structure determination 176
5.4.3 Cnn$^{D490-N579}$ structure analysis 179
5.5 Residues S567 and S571 can be phosphorylated *in vitro* in the absence of the Cnn C-terminus 187
5.6 Phospho-mimicking mutations of S567/S571 disrupt the double hairpin fold in solution 189
5.7 Phospho-null mutations of S567/S571/S573 do not affect the structure of Cnn$^{D490-N579}$ *in crystallo* 196
5.7.1 Phospho-null Cnn$^{D490-N579}$ purification and crystallisation 196
5.7.2 Phospho-null Cnn$^{D490-N579}$ structure determination 200
5.7.3 Phospho-null Cnn$^{D490-N579}$ structure analysis 204
5.8 Calmodulin interacts with dilipoyl-Cnn$^{A553-H608}$ 208
5.9 Discussion 214
Chapter 6 - Conclusions 226
Appendix 230
Abbreviations 240
Acknowledgments 243
Bibliography 245

Chapter 1 - Introduction

1.1 Centrosome function in health and disease

1.1.1 Function and conservation

Centrosomes (Latin and Greek for "central body") organise the microtubule cytoskeleton in many dividing animal cells and have been recognised as such for over a hundred years [1]. These structures are not present in bacteria, archaea or viruses, making them a distinctly eukaryotic organelle. Some centrosomal proteins share a degree of conservation with proteins involved in microtubule-organisation in higher plants [2, 3] and higher fungi [4]. However, only the centrosomes of animal cells (and some other organisms such as brown algae [5], some lower fungi [6] and slime moulds [7]) consist of a pair of centrioles surrounded by a pericentriolar material [7-10] (Fig. 1.1).

Mitotic spindles are synergistically formed by centrosome-, chromatin- and microtubule-mediated pathways [11]. Centrosomes are therefore not absolutely required for the formation of a bipolar spindle [12]. Cultured mammalian cells without centrosomes can form mitotic spindles [13], and flies that lose centrosomes during late embryogenesis continue to develop into morphologically normal adults [14]. Indeed, many animals naturally lose centrosomes before the oogenic meiotic divisions. In these species, a centrosome is donated by the sperm during fertilisation [15].

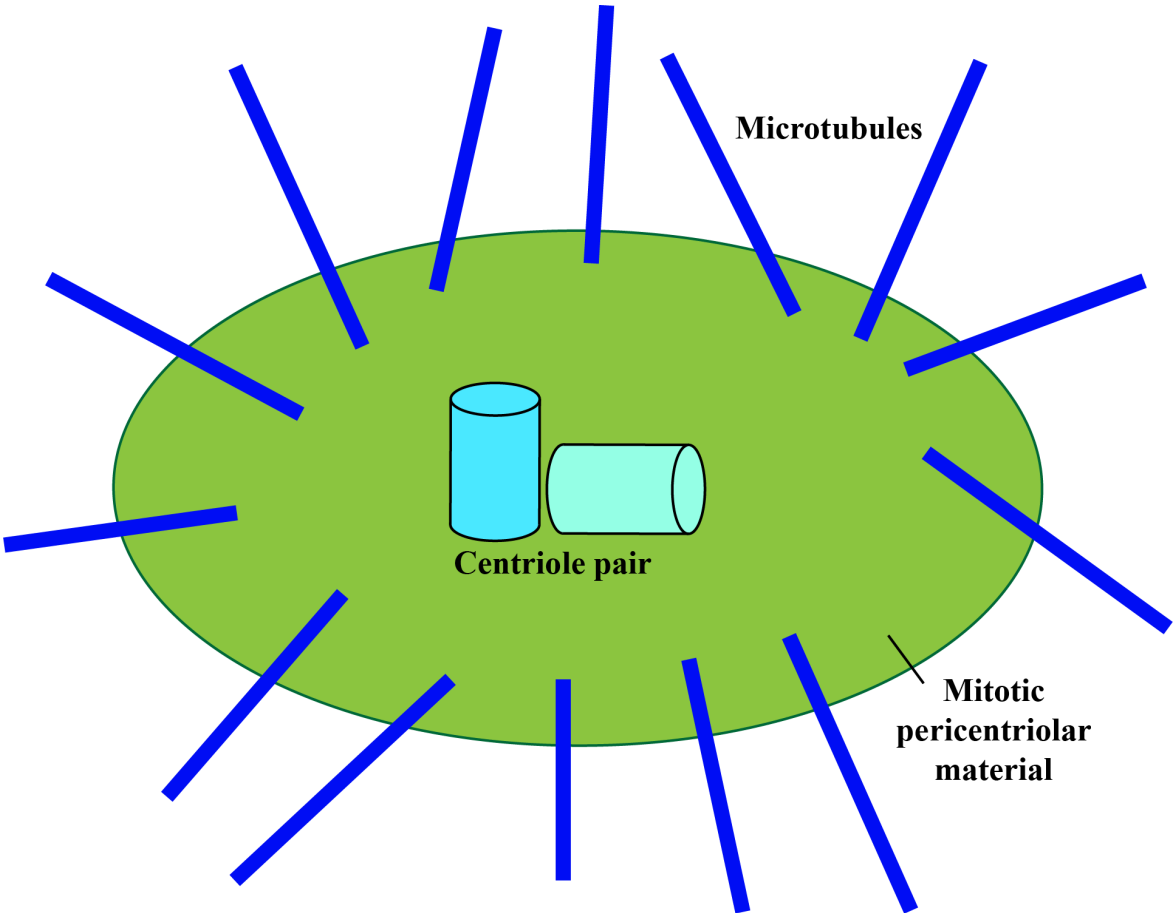


Figure 1.1 | Schematic representation of a mitotic centrosome. A centrosome consists of a pair of centrioles (light blue and turquoise) and a surrounding pericentriolar material (green). Centrioles are barrel-shaped, nine-fold symmetrical objects, while the pericentriolar material is a more amorphous, dense mass of protein from which microtubules (dark blue) are nucleated. Not drawn to scale.

When present, however, centrosomes are considered the dominant sites of spindle pole assembly [16] and they appear to have a crucial role in allowing fast, geometrically correct assembly of mitotic spindles [17, 18]. They are therefore especially important in invertebrates during syncytial embryonic divisions [19-21] and during asymmetric cell divisions [22-26]. Asymmetric centrosome inheritance is also thought to have an important role during mammalian neocortex development [27, 28]. Furthermore, lack of centrosomes can lead to an increased number of errors during mitosis and a p53-dependent cell cycle arrest in cultured vertebrate cells [17, 29-31] and p53-dependent apoptosis in mouse embryos [18].

Microtubules are nucleated within the pericentriolar material, not at the centriole [32]. Polymerisation of α/β-tubulin subunits, which appear to be enriched in the pericentriolar material [33, 34], into microtubules is thought to occur via nucleation by ring structures consisting of γ-tubulin and other accessory proteins [35-46]. In addition to microtubule organisation, centrosomes also appear to serve other functions [47], such as in the organisation of an actin network [48, 49] and by acting as a signalling hub for a range of other pathways [50].

1.1.2 The centrosome cycle

Centrosomes normally duplicate once, and only once, per cell cycle (Fig. 1.2). During G1, a mother and a daughter centriole are connected at their proximal ends via a flexible linker. Upon entry into S phase, both centrioles start assembling a cartwheel-containing procentriole in a tight orthogonal engagement which will continue to be elongated throughout S and G2 (ref [51]). Tight engagement of the procentriole appears to block reduplication of the parental centriole [52-54].

In preparation for mitosis, the two parental centrioles, which during interphase only associate with a limited amount of pericentriolar material around their proximal end, significantly increase the amount of pericentriolar material around them in a process called centrosome maturation or mitotic pericentriolar material assembly [55]. Furthermore, the intercentriolar linker is broken and the two centrosomes move from the nuclear envelope, which starts to break down, to opposite sides in the cell. There they associate with the mitotic spindle poles, ensuring that each daughter cell inherits one centriole pair only. These events are crucially dependent on a range of microtubule motor proteins [51, 56, 57].

The pericentriolar material is important for maintaining the structural integrity of centrioles [58]. Furthermore, centrosome maturation appears to be important for maintaining the association of centrosomes with spindle poles and therefore correct centriole segregation. In embryos from female flies with mutations in Cnn, the scaffold protein that allows centrosome maturation (see below), centrioles cannot stably associate with mitotic spindle poles [59]. These embryos develop lethal mitotic defects during early development [20, 21]. Flies with mutations in Cnn that survive early embryogenesis due to maternally-supplied Cnn are viable [60, 61], but their somatic cells show defects in centriole segregation [59].

As cells exit mitosis, the parental centriole and its procentriole (now called daughter centriole) disengage, and the daughter centriole starts assembling its own pericentriolar material [62-65]. This may lead to an initial asymmetry in centrosome size between the mother and the daughter, with the mother associating with a larger pericentriolar material [24, 62]. Such centrosome asymmetry has an important role in the asymmetric cell divisions of the *Drosophila* male germline, where the mother centrosome is normally retained in the germline stem cell [22, 23]. However, in *Drosophila* neuroblasts, Cnn incorporation into the mother centrosome is downregulated after the daughter centriole separates from the mother soon after mitosis. This results in a reduction of mother centrosome size and retention of the then larger daughter centrosome in the stem cell after the next division [24-26]. Mutating Cnn leads to severe defects of these asymmetric cell divisions in fly germline and neuronal stem cells [22, 24].

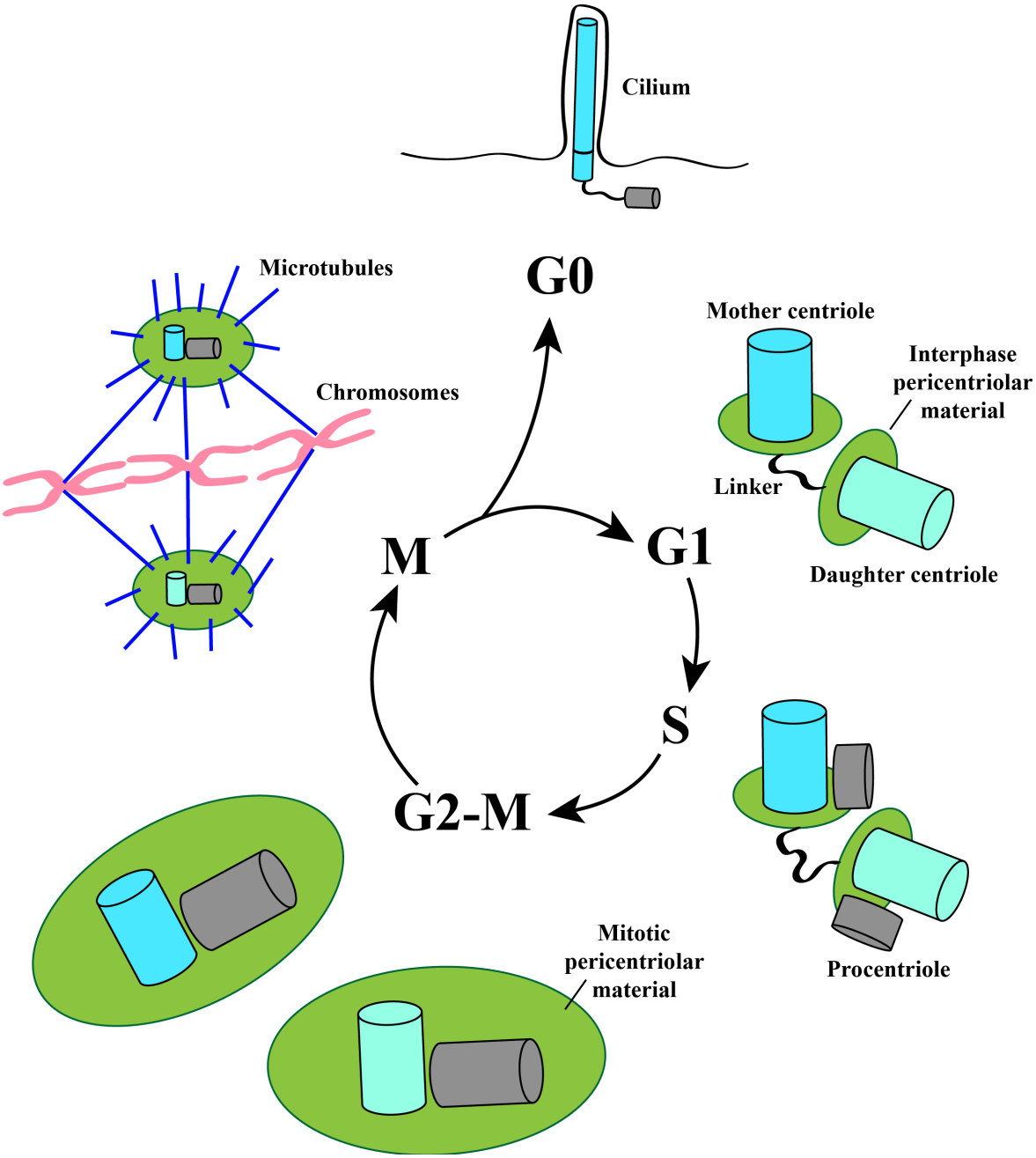


Figure 1.2 | Schematic representation of the centrosome cycle. During G1, a mother and a daughter centriole (which had previously been its procentriole) are connected by a linker and organise a limited pericentriolar material. During S phase, both centrioles duplicate, forming orthogonally associated procentrioles. As the cell prepares to enter mitosis (G2-M), centrosomes mature by forming an expanded mitotic pericentriolar material; the linker is broken and the centrosomes move apart. During M phase, the mature centrosome nucleates microtubules and associates with the spindle poles, leading to each daughter cell inheriting one centrosome only. In cells that exit the cell cycle (G0), centrioles can migrate to the plasma membrane and nucleate cilia. Not drawn to scale. For references, see main body.

In cells that stop dividing and exit the cell cycle, centrioles do not assemble centrosomes but often form cilia instead [66, 67], an organelle with important roles in e.g. cell signalling [68], fluid flow [69] and sperm motility [70]. Flies that lack centrioles cannot form cilia in their mechanosensory neurons and so are uncoordinated, cannot feed and consequently die [14].

In addition to the canonical centriole duplication pathway described above, centrioles can also arise *de novo*, suggesting that the mother centriole is best considered as a regulatory site rather than a strict template for procentriole formation. For example, centrioles can be induced to form *de novo* in unfertilised *Drosophila* eggs which lack centrioles [71]. Furthermore, some organisms that naturally lack centrosomes form centrioles *de novo* during ciliogenesis [72]. Centrioles are also formed *de novo* during mouse embryogenesis where microtubules during the early divisions are organised via non-centrosomal sites [73, 74]. In addition, cultured human cells lacking centrosomes have also been observed to assemble centrioles *de novo* [31, 75], although this process appears to be error-prone [76]. Interestingly, in cultured vertebrate cells derived from multiciliated cells depleted of centrioles before differentiation, the correct number of centrioles is assembled during multiciliogenesis. Strikingly, this *de novo* biogenesis of centrioles occurs within a cloud of pericentriolar material protein, emphasising the potential importance of the pericentriolar material microenvironment for centriole assembly or maintenance [77].

1.1.3 Disease

Centrosome defects have been implicated in disease since Theodor Boveri first suggested their involvement in tumorigenesis over a hundred years ago [78-81]. In flies, it has been observed that cells with an increased number of centrosomes have defective asymmetric cell divisions and form tumors upon transplantation into a wild-type host [82]. In addition, a direct

causal link has been suggested in mice between centrosome overduplication and increased aneuploidy, i.e. an abnormal chromosome number, and tumorigenesis [83]. Furthermore, in human cultured cells and mouse models infected with Epstein-Barr virus, which can cause cancer [84], centrosome number is amplified and cells show increased chromosomal instability [85]. Finally, centrosome amplification and the resulting changes in the cellular cytoskeleton have also been observed to promote invasive behaviour in human cultured cells and could thereby aid metastasis [86].

In mice models, there is also evidence that aneuploidy due to centrosome amplification can result in increased brain cell death and thereby cause microcephaly, a condition characterised by an abnormally small brain [87]. In humans, mutations in the gene for CEP215 (also called CDK5RAP2; the vertebrate homologue of fly Cnn) have been implicated in causing microcephaly and Seckel syndrome, which combines microcephalic primordial dwarfism and neurosensory defects [88-93]. As in flies (see above), cultured vertebrate cells that lack functional CEP215 show defects in centrosome attachment to spindle poles [94-96], and in mouse models and cultured cells, loss of functional CEP215 can lead to multipolar spindles, errors in centrosome numbers, and defects in neural development [97-99]. Mutations in many other key human centrosome components have also been linked with causing microcephaly and primordial dwarfism [100-107], demonstrating the important physiological role of this organelle.

Whether centrosome amplification causes cell death or premature cell cycle arrest (and hence microcephaly or primordial dwarfism) or induces cell proliferation (and hence tumorigenesis) might depend on how many extra centrosomes there are and how well the cell is able to cluster supernumerary centrosomes into two spindle poles [55, 83, 94]. However,

even if the cell is able to cluster extra centrosomes into a bipolar spindle, chromosome missegregation and chromosomal instability can occur due to increased merotelic anchoring, i.e. chromatid attachment to both spindle poles, resulting in lagging chromosomes during cell division [108].

1.2 Centriole structure and assembly

1.2.1 Centriole ultrastructure

Centrioles (size range: nm-μm) consist of a nine-fold symmetrically arranged ring of microtubules bound to accessory proteins (Fig. 1.3). Depending on the organism, these microtubules can be present as singlets, doublets or triplets [109]. In some organisms, the mother centriole also has additional structures at its distal end called appendages [110]. Furthermore, in some or all stages of their duplication cycle, centrioles also contain a structure called the cartwheel [111]. This structure is thought to primarily consist of stacked rings of the dimeric protein Sas6. Importantly, mutations that disrupt Sas6 self-oligomerisation into a cartwheel also perturb centriole duplication [112-114]. It has been suggested that *C. elegans* assembles a cartwheel that consists of a Sas6 spiral rather than stacked rings [115].

As mentioned above, centrioles not only form centrosomes but also cilia. The single-celled *Trichonympha* can have thousands of cilia, each one originating from a centriole that can be an order of magnitude longer than the ones found in animals [116]. The ultrastructure of the proximal region of *Trichonympha* centrioles has been determined via cryo-electron tomography. This revealed that the *Trichonympha* cartwheel consists of stacked rings 8.5 nm apart that could accommodate nine dimers of Sas6 each. The central hub of each ring has a diameter of 22 nm, and the spokes of two successive rings fuse at their tips, resulting in a 17 nm periodicity at the edge of the ring [117, 118]. Although some details might differ, many features of the mature *Trichonympha* centriole appear to be conserved in the structure of the *Chlamydomonas* procentriole, suggesting a common architecture underlying centrioles and procentrioles in different organisms [117, 119].

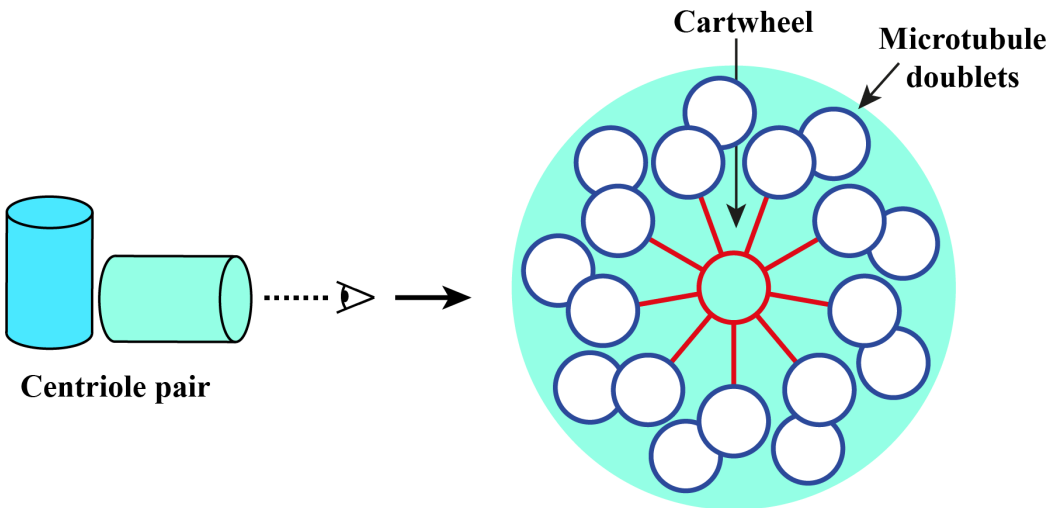


Figure 1.3 | Schematic of a centriole. End-on view of a *Drosophila* centriole, with microtubule doublets shown as blue circles and the cartwheel shown in red. Note the nine-fold symmetry.

1.2.2 Centriole assembly pathway

How the centriole is assembled was first determined by making use of the intricacies of *C. elegans* fertilisation. In this organism, the sperm donates one centriole pair to the oocyte (which lacks a centrosome), and after a round of centriole duplication using maternally-derived proteins, each daughter cells inherits a pair of centrioles at the two-cell stage capable of bipolar spindle assembly. Loss of proteins essential for centriole duplication will lead to daughter cells inheriting only a single centriole and therefore organisation of a monopolar spindle at the two-cell stage. Screens for this phenotype as well as subsequent functional studies have allowed the identification of five proteins essential for centriole duplication and have revealed their functions and recruitment hierarchy [120-122] (Fig. 1.4): ZYG-1 is the master kinase of centriole duplication [123], and it depends on SPD-2 for its recruitment, which is also essential for assembly of the pericentriolar material [124, 125]; ZYG-1 in turn is required for the recruitment of the cartwheel-building proteins SAS-5 and SAS-6 (refs [126-130]); these then promote localisation of SAS-4 (refs [131, 132]), which is required for addition of singlet microtubules to the cartwheel [121].

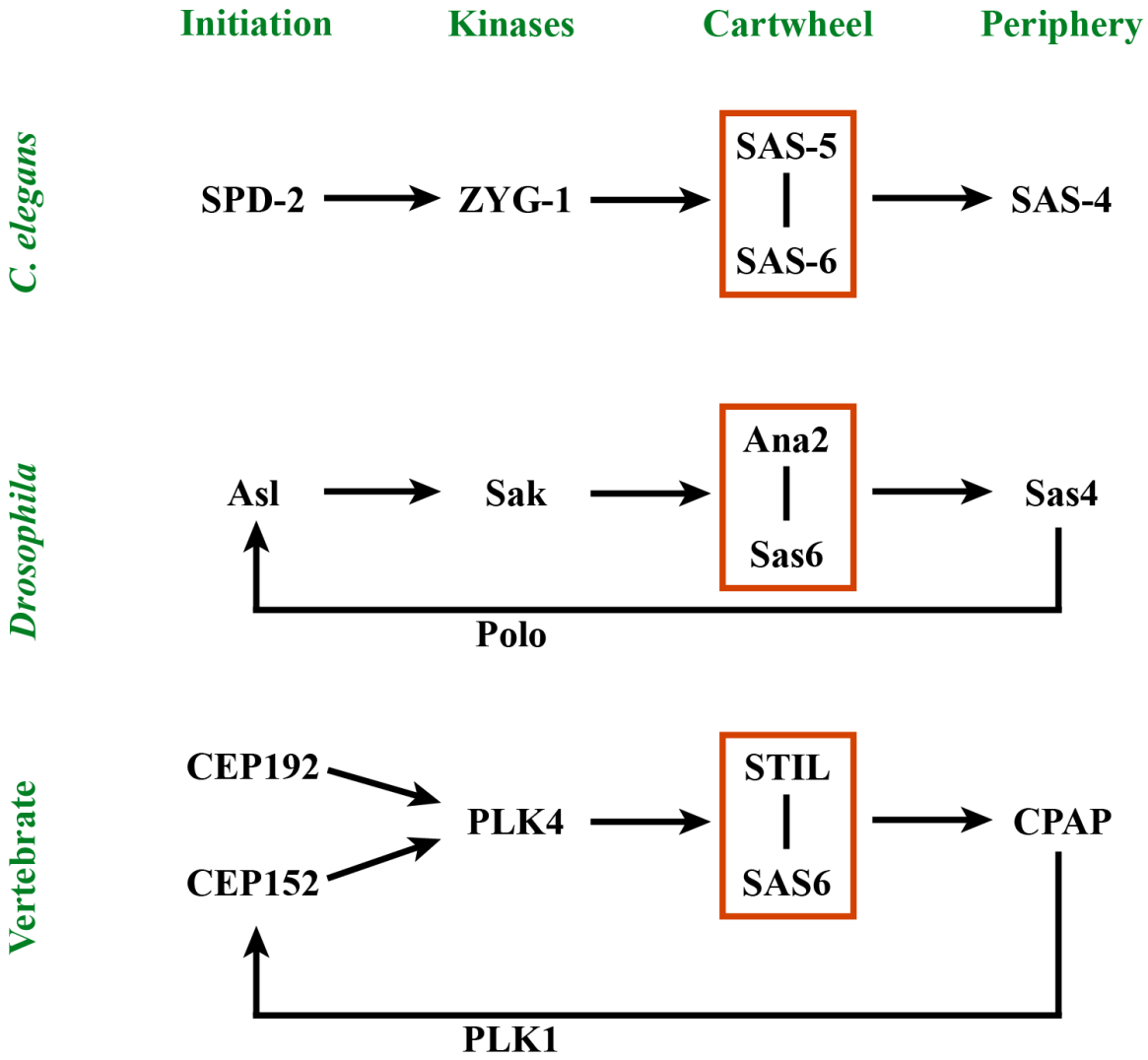


Figure 1.4 | Centriole assembly pathway. The master kinase ZYG-1/Sak/PLK4 (*C. elegans*/*Drosophila*/vertebrates) is recruited by SPD-2/CEP192 (*C. elegans*/vertebrates) or Asl/CEP152 (*Drosophila*/vertebrates). The master kinase then recruits the conserved cartwheel components SAS-5/Ana2/STIL and SAS-6/Sas6/SAS6, which in turn recruit SAS-4/Sas4/CPAP to the growing centriole. Sas4/CPAP then recruits Asl/CEP152 to the centriole during its first mitosis, probably with the help of the kinase Polo/PLK1, in a process called mitotic centriole conversion. For references, see main body.

This overall pathway is largely conserved in *Drosophila* and vertebrates, but how the master kinase (i.e. ZYG-1 in *C. elegans* [123] and its functional homologues Sak in *Drosophila* [133, 134] and PLK4 (Polo-like kinase 4) in vertebrates [135, 136]) is recruited appears to vary. As described above, in *C. elegans* the master kinase is recruited by SPD-2. However, in *Drosophila* the protein Asterless, which is not present in *C. elegans*, recruits the master kinase [137-139]. In vertebrates the master kinase is recruited by both CEP152, the vertebrate homologue of *Drosophila* Asterless, and CEP192, the vertebrate homologue of *C. elegans* SPD-2 (refs [139-143]). After recruitment to the centriole, PLK4/Sak binds and phosphorylates STIL/Ana2 (vertebrate/*Drosophila*), the functional homologue of *C. elegans* SAS-5 (ref [144]), thereby promoting STIL/Ana2 recruitment to centrioles and allowing STIL/Ana2 to bind and recruit SAS6/Sas6 (refs [145-150]). Furthermore, STIL/Ana2 also interacts with CPAP/Sas4 (refs [150, 151]), which in turn binds to CEP152/Asterless (refs [139, 143]).

However, Asterless is only recruited to the new daughter centriole during its first mitosis, rather than during S phase as for Sas4 (refs [64, 152]). This is because during mitosis in flies, the previously incorporated Sas4 is phosphorylated by the kinase Cdk1, which allows Sas4 to bind and recruit the kinase Polo to the new daughter centriole. This process is in turn required for Asterless recruitment and therefore mitotic centriole conversion, i.e. licensing of the daughter centriole to assemble a procentriole itself and to organise a pericentriolar material [63]. In vertebrates, mitotic centriole conversion also appears to be mediated by Polo-like kinase 1 (PLK1, the vertebrate homologue of Polo) [65].

1.3 The interphase pericentriolar material

The pericentriolar material has long been thought to lack a clear organisation [153]. However, it has now been shown via subdiffraction microscopy that during interphase in vertebrates and flies, the pericentriolar material assembled around the proximal region of parental centrioles shows a defined toroidal organisation, with different proteins occupying distinct regions [154-157].

Strikingly, it was found that Pericentrin/Pericentrin-like-protein (Plp) (vertebrate/*Drosophila*) adopts an extended conformation, in which the C-terminus of the molecule localises to the centriolar wall, but the N-terminus extends outwards. Within the maximal dimensions delimited by the resulting Pericentrin/Plp toroid, all other key pericentriolar material proteins such as CEP152/Asterless, CEP192/Spd2, CEP215/Cnn and γ-tubulin were observed to form concentric layers. This was found to result in an overall toroidal architecture of the interphase pericentriolar material over 100 nm in width around the proximal region of the mother centriole [154-157]. Indeed, yeast-two-hybrid studies suggest that *Drosophila* Plp can directly interact with Cnn, Spd2 and Asterless (refs [158, 159]).

Interestingly, in both vertebrate and *Drosophila* cells, procentrioles were observed to form within a ~150-200 nm "gap" within the Pericentrin/Plp toroid that was not detectable in the pericentriolar material around unduplicated centrioles [155, 156].

1.4 The mitotic pericentriolar material

1.4.1 Centrosome maturation in flies and vertebrates

During interphase, the centrosome often does not have a major role in microtubule organisation [160]. However, as the cell prepares to enter mitosis, centrioles organise a pericentriolar material that is not organised in a toroid as described above, but instead forms an extended matrix containing key proteins such as CEP215/Cnn, CEP192/Spd2 and γ-tubulin (refs [154-157]). The process during which centrosomes assemble this extended pericentriolar material is called centrosome maturation [161] (Fig. 1.5).

The mature centrosome consists of hundreds of proteins [162-165] but only a limited number are required for its assembly [165-167]. Crucially, the presence of Polo kinase activity, Spd2 and Cnn was found to be absolutely essential for recruitment of other pericentriolar material proteins in flies and hence mitotic pericentriolar material assembly [157, 167-170]. In *Drosophila* larval brain cells lacking both Cnn and Spd2, centrosome maturation is abolished and centrosomes essentially cease functioning as mitotic microtubule-organising centres, suggesting that the two proteins cooperate to form the scaffold upon which a functioning mitotic centrosome is assembled [168].

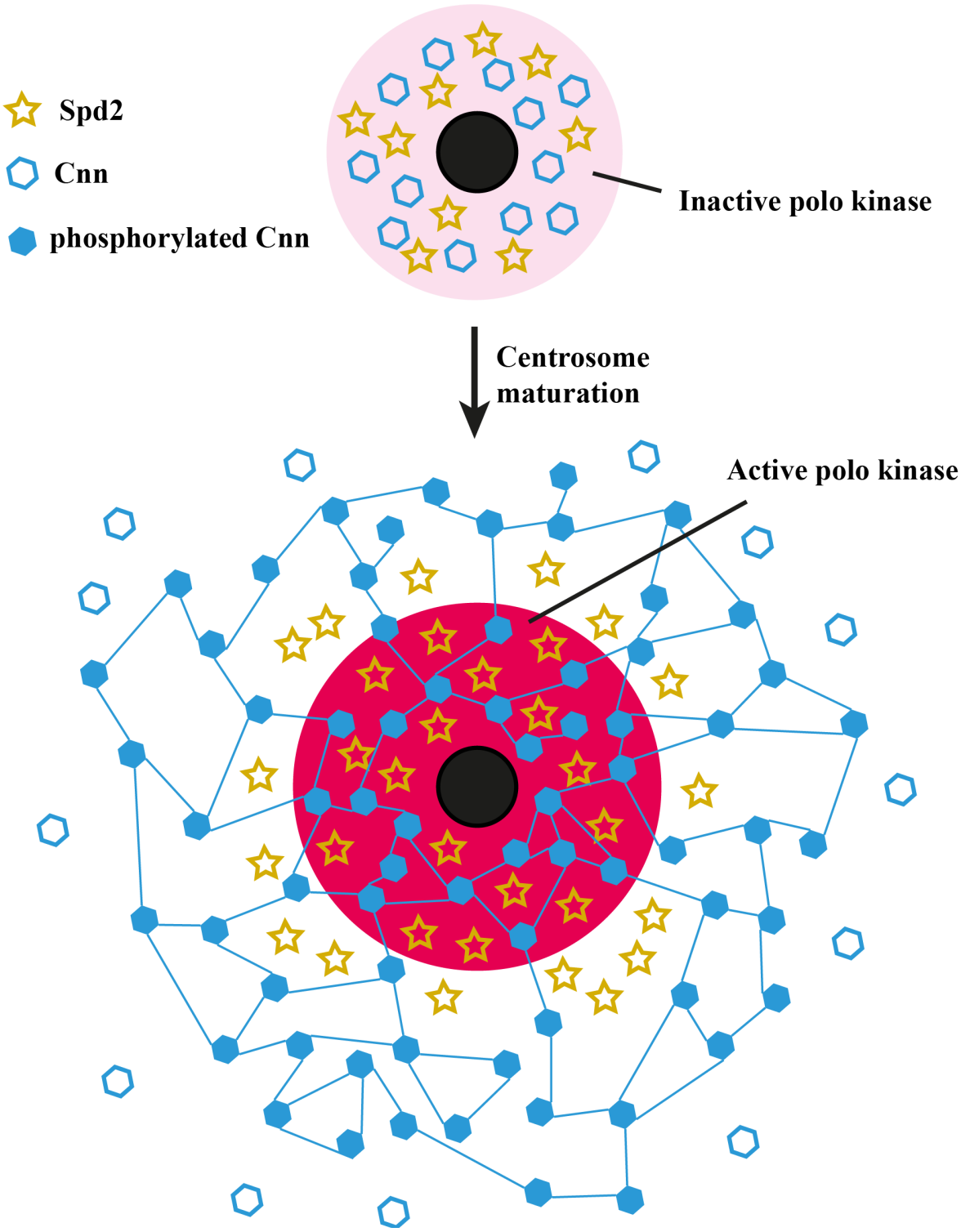


Figure 1.5 | Schematic showing centrosome maturation in *Drosophila*. Cnn (blue hexagons) does not assemble into a network in the absence of phosphorylation (i.e. when Polo kinase is inactive). Upon mitotic entry, active polo kinase phosphorylates Cnn, allowing it to form networks with the help of Spd2 (golden stars). The Cnn network expands beyond Spd2 and acts as a scaffold for the recruitment of client proteins (not shown). Towards the edge of the pericentriolar material, an unknown phosphatase might dephosphorylate Cnn again, resulting in network disassembly. The centriole is shown as a black circle (not drawn to scale). See main body for references.

The molecular pathways allowing centrosome maturation appear to involve related proteins in vertebrates and flies, i.e. PLK1, CEP192 and CEP215, although some of the details, such as the role of Pericentrin during centrosome maturation may differ [171-177]. At least in cultured vertebrate cells, CEP215 depends on Pericentrin for its centrosomal localisation, which is mediated by the extreme C-terminus of CEP215, and the interaction between these two proteins is crucially important for centrosome maturation [96, 176-178]. Vertebrate CEP215 has been shown to associate with a wide range of centrosomal, microtubule- and actin-localised proteins, including motor proteins such as dynein and HSET (ref [94]).

Fly Cnn (Fig. 1.6) contains two highly conserved regions referred to as Cnn motif 1 (CM1, i.e. $Cnn^{R98\text{-}D167}$) and Cnn motif 2 (CM2, i.e. $Cnn^{L1089\text{-}L1148}$). These motifs respectively show 40 % and 29 % identity with their corresponding regions in human CEP215. In addition, although the precise amino acid sequence of the rest of Cnn does not appear to be very conserved, both fly Cnn and human CEP215 separate these two conserved motifs by extensive predicted coiled-coil regions [179-181]. Such coiled-coils are structural motifs that form when an α-helix wraps around one or more other parallel α-helices. Coiled-coils show distinct heptad repeats, with hydrophobic residues usually present in positions a and d, and charged residues in positions e and g (Fig. 1.7) [182-186]. Leucine zippers are a special case of coiled-coils which only have leucine in position d, and the motif is commonly found in DNA-binding proteins [187-189]. *Drosophila* Cnn contains three predicted leucine zipper-like motifs, located towards the N-terminus ($Cnn^{R101\text{-}E126}$), in the centre of the molecule ($Cnn^{V527\text{-}L552}$) and towards the C-terminus ($Cnn^{K946\text{-}Q972}$), and many further predicted coiled-coils [181].

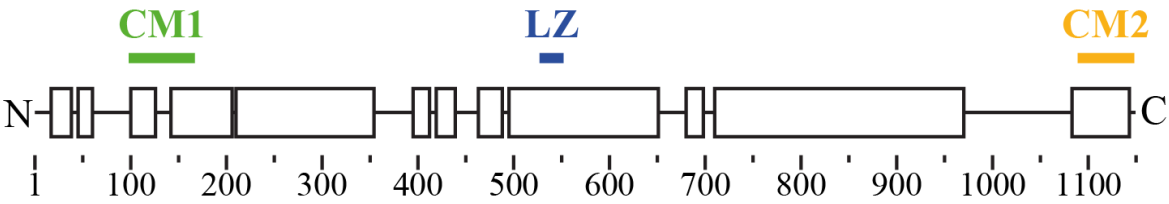


Figure 1.6 | Schematic representation of *D. melanogaster* Cnn. The N-terminus is on the left, the C-terminus on the right. Numbers indicate the amino acid counting from the N-terminus (M1) to the C-terminus (L1148). Location of Cnn motif 1 (CM1, i.e. $Cnn^{R98\text{-}D167}$), Cnn motif 2 (CM2, i.e. $Cnn^{L1089\text{-}L1148}$) and the central zipper-like motif (LZ, $Cnn^{V527\text{-}L552}$) is shown with green, blue and orange bars, respectively. Boxes indicate areas predicted to be largely helical, the other areas (horizontal black lines) are predicted to be largely disordered. Predictions were made using the XtalPred web server [190], which uses PSIPRED [191] and DISOPRED2 [192] for prediction of structural features. See Appendix for full protein sequence.

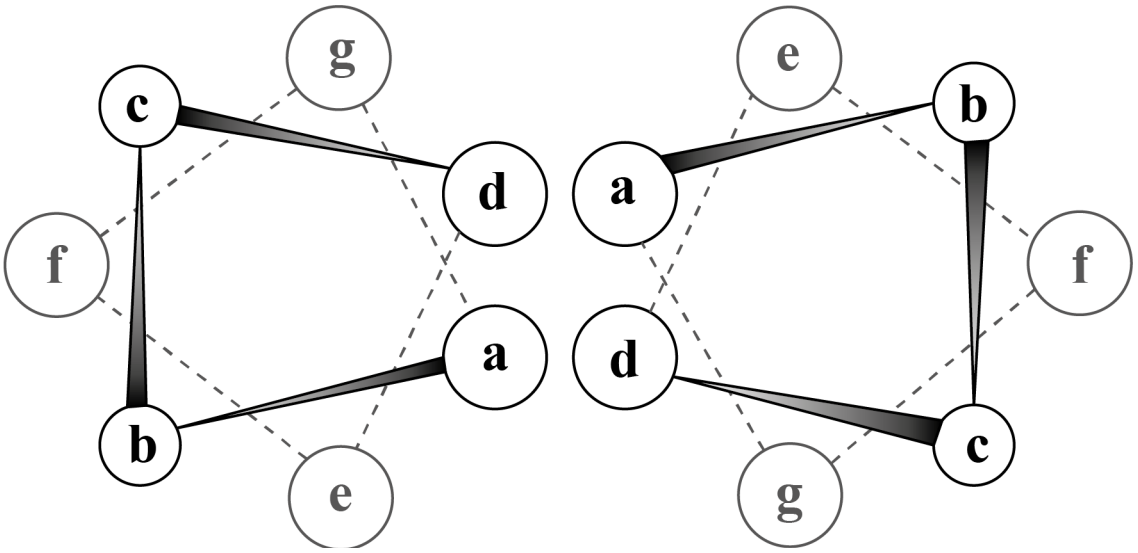


Figure 1.7 | Helical wheel representation of a parallel dimeric coiled-coil. Hydrophobic residues are preferred at positions a and d, residues with opposite charge are preferred at positions e and g of the heptad repeat abcdefg. Coiled-coils with leucines in position d are commonly called leucine zippers. See main body for references. Helical wheel representations were first used in [193].

1.4.2 Network assembly dynamics

In fly embryos, which have rapid cell cycles and where centrosomes therefore never fully disassemble their mitotic pericentriolar material [159, 194], both Cnn and Spd2 are recruited to centrosomes at an approximately linear rate during S phase. Spd2 levels at the centrosome peak just before mitosis and then decline, while Cnn continues to be recruited until nuclear envelope breakdown and then its levels plateau during mitosis [62, 195].

Strikingly, when fluorescently-labelled Spd2 is photobleached in syncytial *Drosophila* embryos, Spd2 first recovers around the centriole in an area that co-localises with Asterless, and subsequently spreads out to form a fibrous network [62, 168, 196]. Fluorescently-labelled Cnn recovers around the centriole and the wider area around the centriole overlapping with Spd2 when photobleached, and then spreads outwards to form a partially overlapping but even larger network [62, 168, 196]. This recruitment behaviour is in contrast to other proteins tested (e.g. γ-tubulin), which were observed to be recruited throughout the pericentriolar material and do not flux outwards [168].

In line with these localisation studies, Asterless was shown to help the recruitment of Spd2 to centrosomes, and Spd2 (and probably also Asterless) appears to strongly aid recruitment of Cnn, while Cnn in turn helps stabilise centrosomal Spd2, thereby establishing a positive-feedback loop [62, 168, 170, 195]. Indeed, yeast-two-hybrid studies suggest that Cnn, Spd2 and Asterless can directly interact with each other [168, 197].

1.4.3 Phosphorylation-dependent network assembly

The results described above suggest a mechanism in flies wherein Asterless recruits Spd2 and Cnn around the centriole; Spd2 also strongly recruits Cnn, which in turn stabilises Spd2;

both proteins then assemble into networks while cooperating to recruit other pericentriolar material proteins. As will be described here, this process also depends on phosphorylation (Fig. 1.8).

The highly conserved kinase Polo has a key role during centrosome maturation and is required for the recruitment of a range of pericentriolar material proteins [167, 171, 198-200]. In addition, the kinase has important roles in mitotic centriole conversion (see above) and centriole disengagement and separation [54, 201-204], as well as various other functions during cell division [205]. In cultured human cells, PLK1 is activated via phosphorylation by Aurora A as the cell prepares to enter mitosis [206, 207], and the activated kinase is present at centrosomes in cultured animal cells [208]. PLK1 binds to S-pS/pT-P motifs [209, 210] and phosphorylates N/E/D-x-S/T or related motifs within a wide range of mitotic substrates [211, 212].

In *Drosophila* syncytial embryos, Polo localises to an area around the centriole that overlaps with Spd2 (refs [168, 195]) and it has also been shown to localise around centrioles (as well as cytoplasmic microtubules) during interphase in cultured *Drosophila* cells, while extending more into the pericentriolar material during mitosis [157]. Polo appears to be initially recruited to the centriole by Sas4 (ref [63]). In addition, Spd2 in fly embryos is specifically phosphorylated at centrosomes which creates Polo binding sites and helps recruitment of the kinase to the expanded pericentriolar material [195].

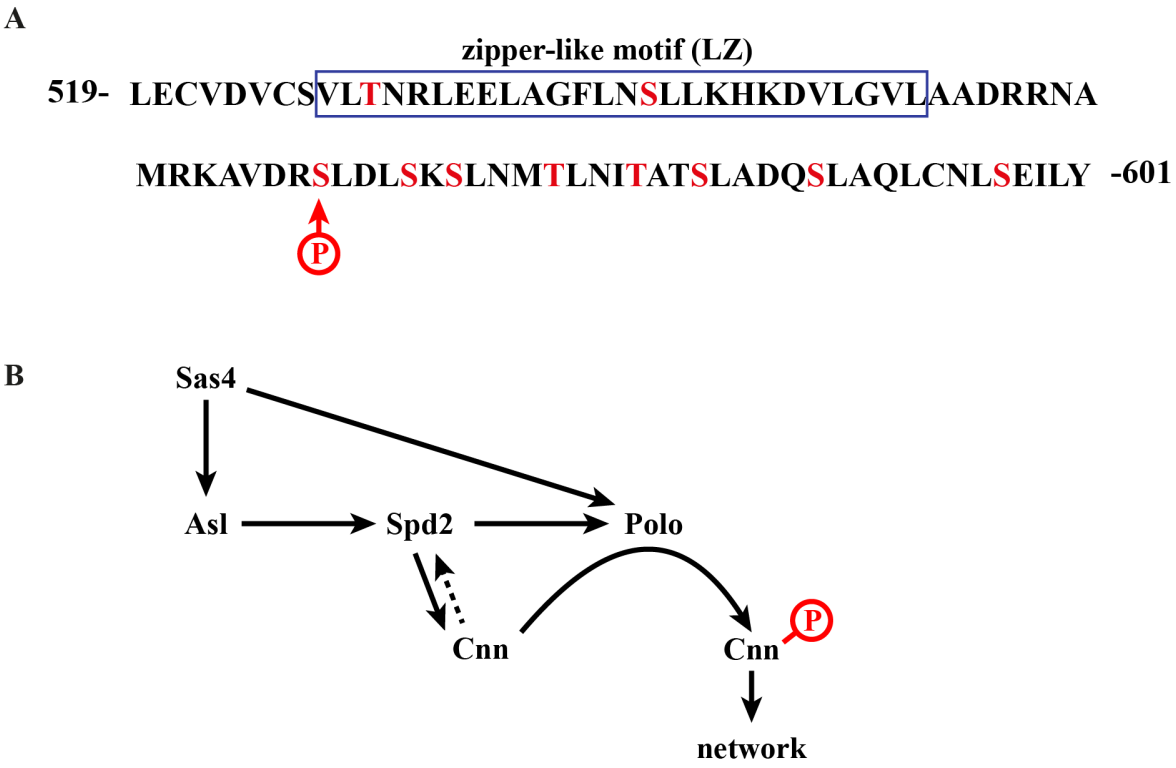


Figure 1.8 | Phosphorylation-dependent centrosome maturation in flies. A) The sequence of *D. melanogaster* $Cnn^{L519\text{-}Y601}$ is shown. The predicted zipper-like motif (also called LZ) is shown as a blue box. Ser/Thr residues identified as conserved among *Drosophila* species in [196] are shown in red. S567, which appears to be specifically phosphorylated at the centrosome, is highlighted with a red circled P. **B)** Schematic illustration of the centrosome maturation pathway in flies. Sas4 recruits Asterless (Asl) and Polo kinase; Asterless in turn recruits Spd2, which helps recruit Polo and Cnn to the pericentriolar material; Cnn stabilises Spd2 (dashed arrow) and is phosphorylated by Polo, allowing it to form a network that acts as a scaffold for the recruitment of effector proteins to the pericentriolar material. For references, see main body.

During mitosis, Cnn is specifically phosphorylated at centrosomes in a manner that depends on Polo kinase but not on other centrosomal kinases such as Sak or Aurora A (refs [167, 196]). Ten conserved Ser/Thr residues are found within a central conserved region within Cnn stretching residues K516-Y601 (Fig. 1.8 A). Some of these sites are Polo consensus sites, and indeed this region can be phosphorylated by recombinant human PLK1 *in vitro*, but not when these ten sites are mutated to alanine (10A), a residue that cannot be phosphorylated [196].

At least one of these ten sites, S567, was shown via mass spectrometry to be specifically phosphorylated at centrosomes in fly embryos, and mutation of this site to alanine disturbs network assembly *in vivo*. However, the Cnn 10A mutant disrupts network assembly even more, suggesting that phosphorylation of S567 is important, but network assembly also relies on phosphorylation of some or all of the other residues. Perhaps unexpectedly, the full-length Cnn 10D/E mutant also disrupts centrosomal network assembly *in vivo*. However, this might be because these mutants were also observed to form large cytosolic Cnn networks independently of centrioles, which might compete with centrioles for Cnn recruitment. In *Drosophila* unfertilised eggs, which lack centrioles, these cytosolic foci were even observed to act as scaffolds for the assembly of microtubules [196].

A slightly larger region of Cnn, i.e. $Cnn^{Q403\text{-}H608}$ (called phosphoregulated multimerization or PReM domain), is dimeric *in vitro*, but forms higher order oligomers when the ten potential phosphorylation sites are mutated to aspartate or glutamate (10D/E). Note that the zipper-like motif $Cnn^{V527\text{-}L552}$ (see above) is found within this PReM domain and appears to mediate phosphorylation-independent dimerisation *in vitro* and is required for phosphorylation-dependent network assembly *in vivo*. Taken together, these studies suggest a key role of Polo-mediated phosphorylation of $Cnn^{Q403\text{-}H608}$ in Cnn network assembly and therefore centrosome maturation [196].

Indeed, when a series of conserved potential Polo binding sites in Spd2 are mutated, Polo is still recruited around centrioles and can phosphorylate Cnn there, but Spd2 and Cnn cannot assemble a larger network and centrosome maturation is therefore inhibited. Network assembly probably fails because Spd2 cannot recruit Polo to the Spd2 scaffold as it attempts to expand from around the centriole, and therefore Polo is not positioned to support further

phosphorylation-dependent expansion of the Cnn network, which then in turn cannot stabilise the Spd2 scaffold, leading to a breakdown of the proposed positive-feedback loop [195].

Strikingly, overexpression of CEP215, the human homologue of Cnn, in human interphase cells can drive expansion of the pericentriolar material independently of kinase activity, although such non-physiological assemblies fail to recruit the same amount of effector proteins such as γ-tubulin when compared to mitotic centrosomes [156]. It has therefore been hypothesised that pericentriolar material proteins have a propensity to self-assemble in the absence of kinase activity if their concentrations are sufficiently high, but this might not be the case physiologically, hence the requirement for kinase activity [213].

1.4.4 Cnn flares

In *Drosophila* embryos, microtubules appear to tear Cnn-containing fragments (so called "satellites") out of the pericentriolar material, which are then seen to radiate away from the centrosomes. This "flaring" of the Cnn network (the Spd2 network does not obviously flare) is especially prominent during telophase and interphase, but it can be abolished upon loss of microtubules via colchicine treatment, which leads to a collapse of the extended Cnn network [168, 196, 214].

In *Drosophila* larval brain cells, the Spd2 network is similar in size to that in embryos and also fluxes outwards, independently of microtubules. However, no microtubule "flaring" is observed in these cells and Cnn does not flux outwards from centrioles upon recruitment, and indeed does not extend beyond the Spd2 network. Thus, the Cnn network can only flux

away from the Spd2 network if it can flare out along the centrosomal microtubules, while the Spd2 network can flux outwards independently of microtubules and the Cnn flux [215].

Plp is important for the recruitment of Cnn around the centriole in the interphase pericentriolar material and proper organisation of a Cnn network. However, the absence of Plp does not completely abolish Cnn network formation [155, 158, 159, 216]. Instead, Plp appears to have an additional role in maintaining the outer regions of the pericentriolar material, in particular Cnn flares. In contrast to the stably incorporated Plp around the centriole, the pool of Plp found in the outer pericentriolar material shows a more rapid exchange with the cytosol [158, 159, 168, 216].

Via yeast-two hybrid studies, it has been shown that the Cnn C-terminus ($Cnn^{V1087\text{-}L1148}$) can interact with Plp (ref [159]). An R1141H mutation within Cnn does not prevent localisation of the protein to the centrosome [21, 217], but it disrupts binding of $Cnn^{V1087\text{-}L1148}$ to Plp in yeast two-hybrid studies, and severely disrupts the assembly of Cnn satellites in *Drosophila* embryos [159]. While this mutation within Cnn does not prevent Plp recruitment to the centriole, which is known to depend on its C-terminal PACT (Pericentrin/AKAP450 centrosomal targeting) motif [158, 218], it does prevent Plp recruitment to the few Cnn satellites that remain [159]. However, Cnn appears to be able to bind to several proteins via its C-terminus, and the R1141H mutation also disrupts binding of $Cnn^{H1082\text{-}L1148}$ to Centrocortin in yeast-two hybrid studies, a protein involved in cleavage furrow formation [217], making any detailed analysis of the effect of mutations within the conserved C-terminal region of Cnn difficult.

1.4.5 Microtubule nucleation

The conserved motif found towards the N-terminus of Cnn (i.e. CM1) can be considered to consist of two parts: a predicted leucine zipper motif ($Cnn^{R101-E126}$, see above) and a motif which shows some sequence similarity to this motif but is less conserved ($Cnn^{E138-V158}$) (ref [219]). Cnn that lacks residues K112-L124, i.e. most of the N-terminal motif, still localises to the centrosome in *Drosophila* embryos, but centrosomes cannot efficiently recruit γ-tubulin, show no "flaring" and have impaired centrosome separation [179], a process that depends on microtubules [220].

Interestingly, while MBP-tagged canonical $Cnn^{M1-L255}$ cannot pull down γ-tubulin from embryonic extracts, the equivalent region within testes Cnn can. The extreme N-termini of the testes and canonical isoforms of Cnn differ, with the canonical isoform having an additional 77 residues at the N-terminus where the testes isoform has a much shorter (19 residues), unrelated N-terminus. This might suggest that the extreme N-terminus (i.e. Cnn^{M1-Y77}) of the canonical Cnn isoform could be involved in regulating γ-tubulin association [221]. However, His-tagged canonical $Cnn^{M1-E705}$ was observed to pull down γ-tubulin from cultured *Drosophila* cells [222], suggesting that perhaps MBP-tagged $Cnn^{M1-L255}$ cannot associate with γ-tubulin due to misfolding or interference by the tag. Ubiquitous overexpression of $Cnn^{A78-N290}$, the complete region shared between canonical and testes Cnn, causes severe developmental defects in flies that can be rescued when the construct is targeted to centrosomes via a PACT motif, demonstrating the importance of targeting the microtubule-organising capacity of Cnn to the appropriate cellular site [223].

Strikingly, fly Cnn ectopically overexpressed in mammalian cells forms acentriolar cytosolic assemblies that recruit γ-tubulin and form microtubule asters. Indeed, Cnn can pull

down γ-tubulin from cultured mammalian cells [222]. This suggests that the mechanism of γ-tubulin recruitment is probably largely conserved between vertebrates and *Drosophila*.

Human CEP215, the homologue of *Drosophila* Cnn, is also required for the centrosomal recruitment of γ-tubulin (ref [224]). CEP215 residues R58-Q90 (corresponding to residues R98-A131 in *Drosophila* Cnn) can pull down γ-tubulin from cultured human cells, but residues K101-S200 (corresponding to residues K142-A260 in *Drosophila* Cnn) cannot, suggesting that the predicted leucine zipper motif within CM1 is necessary and sufficient to associate with γ-tubulin (ref [224]). A F75A mutation in CEP215 (corresponding to *Drosophila* Cnn F115) severely disrupts the association between CEP215 and γ-tubulin (ref [225]). However, CEP215 probably does not directly bind to γ-tubulin, but rather interacts with the assembled γ-tubulin ring complex, which consists of γ-tubulin and further accessory proteins [225]. Furthermore, the region C-terminal to the predicted leucine zipper motif might also have a role, as phosphorylation on S140 is thought to enhance the association of CEP215 with γ-tubulin in cultured human cells [226]. *In vitro*, addition of $CEP215^{T51\text{-}S200}$, which contains CM1, enhances the microtubule-nucleating activity of γ-tubulin ring complexes immunoprecipitated from cultured human cells via a tagged CM1-containing fragment that is lost during the purification protocol [225]. However, functional and low-resolution structural evidence from *X. laevis* γ-tubulin ring complexes purified via a γ-tubulin antibody suggests that addition of a CM1-containing, N-terminal fragment of CEP215 does not lead to large-scale structural changes or significant increases in microtubule-nucleation capacity of the γ-tubulin ring complex in this organism, suggesting that factors other than CM1 are required, too [46].

In some animal cells, the Golgi apparatus acts as a microtubule-organising centre, and this process appears to rely on some of the same proteins as centrosome-mediated microtubule nucleation [227]. Indeed, CEP215 localises to the Golgi in cultured human interphase cells, and this appears to be mediated by AKAP450 and Pericentrin, both of which can associate with the CEP215 C-terminus (i.e. CEP215$^{H1726-S1893}$) in pull-down studies [178]. Note that in *Drosophila* post-mitotic sensory neurons, Cnn is also recruited to the Golgi, where it is involved in nucleating microtubules and thereby functions in restriction of dendrite branching [228].

Furthermore, note that there are several isoforms of Cnn that fall into two broad classes [229, 230]. The first class of Cnn is the one discussed in this thesis. Within this class, Cnn isoform A is considered the canonical embryonic form, as this isoform rescues early embryonic development in flies that have a mutation in the gene for Cnn (ref [229]). The other class of Cnn isoforms is largely absent in the early embryo and instead mainly produced in *Drosophila* testes via alternative splicing [223, 231-233]. The resulting Cnn isoform has a different central and C-terminal domain that targets the protein to mitochondria instead of centrosomes. This Cnn isoform then converts mitochondria in *Drosophila* spermatids into microtubule-organising centres, suggesting that the capacity to nucleate microtubules resides within the N-terminal region shared between the two Cnn isoforms, while cellular localisation depends on the type of C-terminal domain [223].

1.4.6 Centrosome maturation in *C. elegans*

As in vertebrates and flies, SPD-2 and the kinase PLK-1 are also essential for centrosome maturation in *C. elegans* [124, 125, 208, 234]. *C. elegans* SPD-2 localises to the centriole during

interphase and extends into the pericentriolar material during mitosis, where it recruits PLK-1 (refs [124, 125]).

However, *C. elegans* lacks obvious homologues of Pericentrin/Plp, CEP152/Asterless and CEP215/Cnn (ref [180]). Instead, centrosome maturation is dependent on the large coiled-coil protein SPD-5 (refs [130, 235]), which in the cytosol of *C. elegans* embryos probably exists as a mixture of monomers and dimers [236, 237]. Despite a lack in sequence similarity, centrosomal recruitment and function of SPD-5 are strikingly similar to CEP215/Cnn, and the proteins are therefore putative functional homologues [55]. In yeast-two-hybrid screens, SPD-5 was observed to interact with itself as well as SPD-2 and a range of other pericentriolar material proteins, and was therefore suggested to have a role in pericentriolar material scaffolding [120]. Furthermore, SPD-2 and SPD-5 are mutually dependent for their localisation to the pericentriolar material [124, 125, 235].

Interestingly, purified SPD-5 was shown to form micron-sized porous networks *in vitro* in a reversible, time- and concentration-dependent manner. Addition of SPD-2 or PLK-1, which can phosphorylate SPD-5 *in vitro*, to SPD-5 was observed to increase the rate of network formation independently and cooperatively, and the two proteins colocalised with the SPD-5 networks. SPD-5 phosphorylation on residues S653 and S658, probably by PLK-1, also appeared to be important for SPD-5 network assembly *in vivo* [236].

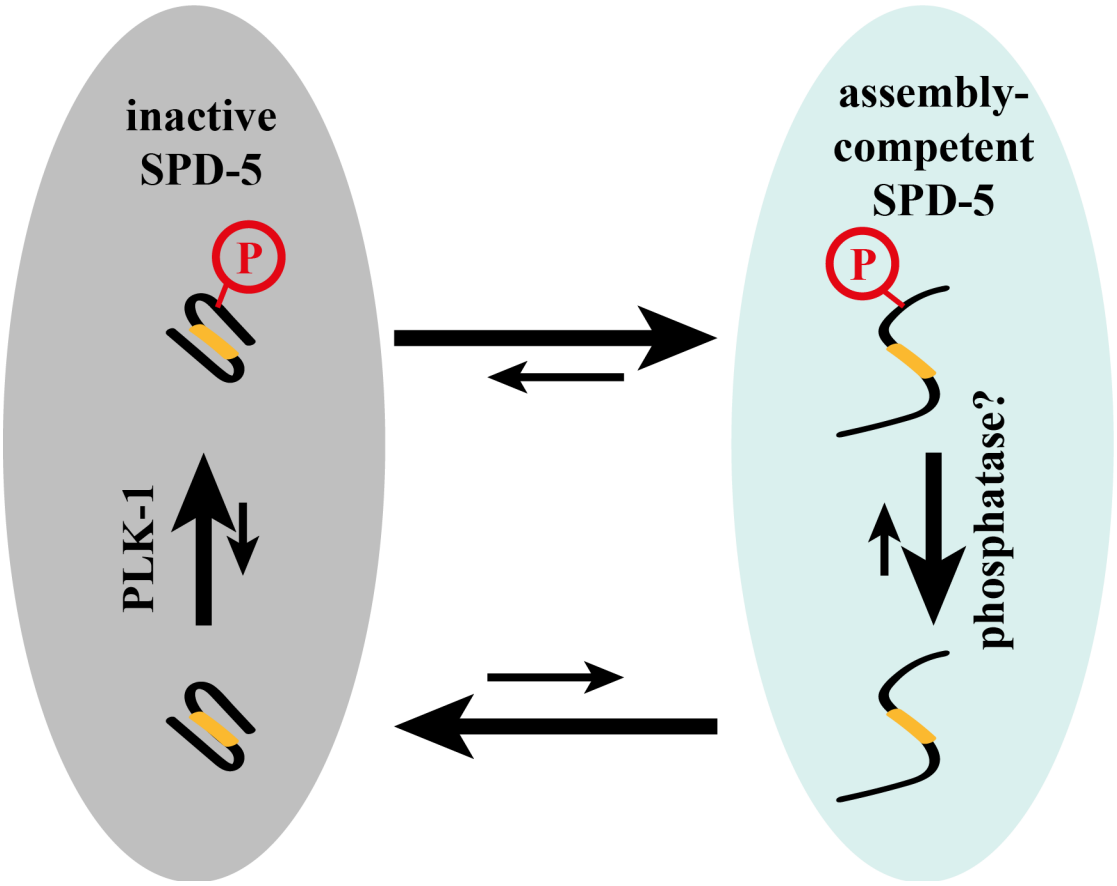


Figure 1.9 | Proposed model illustrating the interconversion of SPD-5 between an inactive state and an assembly-competent state. SPD-5 has been proposed to exist in an inactive form (cytosol, grey) and an assembly-competent form (centrosome, turquoise). PLK-1-mediated phosphorylation (red P) would bias the equilibrium between these two states towards the assembly-competent form, perhaps by inducing a conformational change that exposes binding sites (orange). These exposed bindings sites are proposed to then allow network assembly by mediating binding to other SPD-5 molecules (not shown). Upon dephosphorylation, SPD-5 would tend towards the inactive state again. Figure adapted from [238].

SPD-5 has therefore been proposed to exist in an inactive and an assembly-competent state (Fig. 1.9). In this model, the two states would be in equilibrium with each other, and SPD-5 could spontaneously isomerise between the two states. Phosphorylation by PLK-1 around the centriole would bias the equilibrium heavily towards the assembly-competent state [238, 239]. However, no evidence on an atomic level exists for such a model.

The putative assembly-competent state of SPD-5 is further hypothesised to form a liquid droplet around the centriole via phase separation in an auto-catalytic manner, i.e. assembly-competent SPD-5 would convert more inactive SPD-5 into the assembly-competent state [239]. Phase separation can physically be described as the co-existence of two locally different concentrations of a solute albeit with the same chemical potential and hence no diffusion gradient, as occurs when solutes interact with each other and not just with the solvent [240-242]. However, *C. elegans* SPD-5 assemblies appear to be in a gel-like state, as partial photo-bleaching experiments showed that SPD-5 does not rearrange within the centrosomes of metaphase-arrested *C. elegans* embryos *in vivo* [243] or within SPD-5 "condensates" *in vitro* [33]. These condensates, i.e. spherical assemblies of SPD-5, form *in vitro* when polyethylene glycol, a crowding agent, is added to purified SPD-5 in an attempt to more closely mimic the crowded environment in the cell. Interestingly, these condensates can selectively recruit so called "client proteins", i.e. effector proteins of the pericentriolar material such as PLK-1 and SPD-2. These client proteins show a much faster turnover than SPD-5 on centrosomes *in vivo*, and partial photo-bleaching experiments suggest that at least some of these client proteins can rearrange within the SPD-5 condensates *in vitro* [33].

The emerging picture of the mitotic pericentriolar material is therefore that of an underlying permeable gel-like or solid-like scaffold which can selectively concentrate client proteins [244]. This is in line with early evidence showing that clam centrosomes purified in high salt concentrations retain a fibrous network that can regain its microtubule nucleation capacity upon incubation with cytosolic embryo extracts [245].

It has been suggested that *in vitro*, SPD-5 condensates are initially in a liquid-like state before "ageing" into a gel-like state [33]. However, there is currently no evidence for a liquid-

like state of either SPD-5 in *C. elegans* or Cnn in flies [244]. Furthermore, Cnn does not appear to form networks auto-catalytically, as Cnn cannot expand into a larger network in the absence of a Polo-recruiting Spd2 scaffold *in vivo*, even if Cnn is phosphorylated around the centriole [195]. It is therefore dubious whether *C. elegans* SPD-5 and fly Cnn assemblies are ever in a liquid-like state *in vivo*, and unclear whether both proteins rely on the same mechanisms to assemble into networks.

1.4.7 Centrosome size control

In *Drosophila* embryos, Cnn and Spd2 are recruited to the centrosome exclusively around the centriole and then flux outwards, thereby forming the mature pericentriolar material (see above). When Cnn expression levels are increased in fly embryos, the Cnn incorporation rate also increases, and centrosomes grow to a larger size [62]. Furthermore, downregulation of the Cnn incorporation rate in *Drosophila* neuroblasts appears to result in a reduction of centrosomal size [24]. It therefore appears that centrosome size might be intricately linked with the recruitment rate of key pericentriolar material proteins, which in turn depends on these proteins being recruited around the centriole. If this centriolar recruitment occurs at a constant rate, centrosomes will grow in size until the recruitment rate is equal to the removal rate of these proteins, which is likely to be dependent on the total surface area of the pericentriolar material and hence on centrosome size. For a given recruitment rate this steady state will therefore always occur at the same centrosome size [62].

In *C. elegans*, a reduction of SPD-5 levels leads to a reduction in centrosome size [238], while overexpression of SPD-2 has been shown to lead to an increased incorporation rate into centrosomes and an increased amount of pericentriolar material [234]. However, SPD-5 and perhaps also SPD-2 do not appear to flux outwards from centrioles in *C. elegans* [208, 243].

Indeed, in *Drosophila* larval brain cells Cnn does not flux outwards, either [215]. In *C. elegans* early embryos there is a limited supply of maternal protein and the two daughter centrosomes assemble a pericentriolar material of precisely half the size of the mother centrosome. It has therefore been suggested that centrosomes might grow until the cytoplasmic pool of pericentriolar material proteins is sufficiently depleted [234], but direct evidence for such a model, especially in other organisms, is lacking.

In *C. elegans* embryos with defective centrioles due to partial depletion of SAS-4, centrosomes formed that had a proportionately smaller size depending on the degree of SAS-4 depletion, suggesting that the amount of pericentriolar material recruited depends on centriole size and duplication [132]. However, while centrioles are important for mitotic pericentriolar material recruitment in *C. elegans* embryos, they appear to be dispensable for its maintenance [208].

The mechanisms that allow centrosome maintenance and subsequent disassembly are largely unclear. It appears that continuous Polo-like kinase 1 activity is required to maintain the pericentriolar material in a mature state in mitotically arrested mammalian cells and *C. elegans* embryos [208, 246]. During fly oogenesis, disassembly of the pericentriolar material and therefore centriole elimination is dependent on loss of Polo kinase from the centrosome [58]. Furthermore, PLK1 is degraded in late mitosis in mammalian cells [247, 248], although it is unlikely that this is a universal mechanism to regulate centrosome disassembly, as Polo persists at centrosomes in some *Drosophila* cells [157, 198]. There might also be a competing dephosphorylation reaction. At least in *C. elegans*, the phosphatase PP2A might be involved in dephosphorylating SPD-5 and thereby aiding centrosome disassembly [249].

Microtubules might also have a role in pericentriolar material disassembly. In *C. elegans* embryos, microtubule-driven pulling forces have been shown to be important for centrosome disassembly [249], and centrioles are required for the pericentriolar material to retain its structural integrity and withstand such forces [208]. The Cnn flaring in *D. melanogaster* embryos, which is very prominent at the end of mitosis, might have an analogues function in pericentriolar material disassembly in flies [214].

1.5 Calmodulin and the centrosome

Calmodulin is a highly conserved protein consisting of two globular domains, which can bind target peptides, connected by a flexible linker [250-253]. The two domains can each bind two calcium ions, and this can increase the affinity of Calmodulin for target peptides, which are typically positively-charged, amphipathic helices. Calcium is bound via the so-called EF-hand motif, i.e. two helices perpendicular to each other binding calcium via the loop connecting them [254-258]. However, calcium-binding is not essential for at least some of its functions [259]. Note that during mitosis in cultured human cells, the centrosome is highly enriched in calcium [260], and indeed the structural integrity of the pericentriolar material in cultured mammalian cells appears to be sensitive to divalent cations such as calcium [261, 262].

In cultured fly cells, Calmodulin localises to spindle poles during mitosis, while during interphase it diffusely localises to the cytosol and perhaps also to the centrosome. Knock-down of Calmodulin leads to centrosome detachment from spindle poles and unfocussed spindles [166, 167, 263, 264]. In *Drosophila*, Calmodulin and Abnormal spindle protein (Asp), which is also involved in centrosome-pole cohesion and spindle pole focussing, are mutually dependent for their localisation to the minus ends of the mitotic spindle fibres and might act in a complex [166, 263, 265, 266], a process likely to be conserved in mammals [267, 268]. In addition, Calmodulin might also have a conserved role in centriole duplication [167, 269] and appears to interact with the centriolar protein CP110 in cultured mammalian cells, which might have a role during cytokinesis [270, 271].

Furthermore, the C-terminal region (i.e. CM2) of both human CEP215 and *Drosophila* Cnn has been shown to bind Calmodulin *in vitro*, and at least in *Drosophila* this interaction appears to be calcium-sensitive [178, 272]. Just like the association of CEP215 with Pericentrin

described above, the interaction between CEP215 and Calmodulin depends on CEP215 residues V1861-A1870 (corresponding to residues L1131-V1140 in *Drosophila* Cnn). *In vitro*, binding can be abolished by mutating the non-conserved residues K1865 and K1869 to alanine. CEP215 carrying these two mutations still targets to centrosomes in cultured human cells, while CEP215 that lacks residues V1861-A1870 cannot, suggesting that this region but not the interaction with Calmodulin is important for centrosomal localisation of CEP215 (ref [178]).

With its two globular domains, Calmodulin can bind target sites on the same or two different peptides. Especially the latter function allows Calmodulin to act as a protein linker or adaptor [273]. In yeast, which do not assemble centrosomes, the microtubule-organising protein Spc110 contains highly diverged CM1-like sequences towards its N-terminus. This protein lacks CM2 and has a C-terminal PACT-like motif instead [42, 219, 274, 275]. Calmodulin binds Spc110 within its PACT-like motif and thereby potentially crosslinks two Spc110 proteins. This binding is important for the function of the spindle pole body, the yeast microtubule-organising centre, perhaps by conferring mechanical stability [276-282].

Calmodulin can also interact with the PACT motif of the centrosomal vertebrate proteins AKAP450 and Pericentrin (refs [218, 283-285]) and in flies, Calmodulin is important for the centrosomal localisation of Plp (ref [264]). Interestingly, vertebrate AKAP450 and Pericentrin can also associate with γ-tubulin via their N-terminal regions [284, 286, 287]. There appears to be a common theme within microtubule-organising centres that involves highly coiled-coil proteins that can associate with γ-tubulin via their N-terminal regions while being recruited to the microtubule-organising centre via C-terminal motifs (i.e. PACT or CM2) that can also interact with Calmodulin.

1.6 Thesis aims

There is currently no structural information available on how centrosomes mature and we therefore lack mechanistic insights into this process. This thesis aims to:

- functionally characterise the regions within fly Cnn that are required for network assembly (chapter 3),

- investigate the interactions between Cnn and Calmodulin from a structural and functional perspective (chapter 4),

- provide a structural mechanism of how Cnn is biased towards an assembly-competent state via phosphorylation (chapter 5).

Chapter 2 - Materials and methods

2.1 Molecular cloning

2.1.1 Primers and polymerase chain reaction

Inserts and backbone vectors were usually amplified via polymerase chain reaction (PCR). PCR primers were designed using the NEBuilder Assembly Tool for standard cloning and the NEBaseChanger for site-directed mutagenesis (online tools provided by New England Biolabs, NEB). Where necessary, homology to the target sequence was extended to achieve a primer melting temperature of > 60°C. Tables 2.1 and 2.2 show the primers used in this work. Primers were stored at a concentration of 100 µM in 10 mM Tris-HCl (pH 8.5).

Primer identifier	Sequence	Used in this work for cloning of...
nG-F	TAATGAGGTACCGGATCC	pETM14-mNG
nG-R	ACCCTGGAAGTACAGGTTTTC	pETM14-mNG
37-F-nG	GAAGTTCTGTTCCAGGGGCCCATGGTGAGCAAGGGCGAG	pAH37
37-R-nG	GGAAGTACAGGTTTTCCTTGTACAGCTCGTCCATGC	pAH37
37-F-CM2	AGCTGTACAAGGAAAACCTGTACTTCCAGGGTGCAGCCACAGTACACGACTGTGCCAAGGTAG	pAH37
37-R-CM2	GATCCGGTACCTCATTAGTCCTATAACTCATTCTCCATGTTTGAGC	pAH37
90_F	AGGAAAACCTGTACTTCCAGGGTGATCAACAGAACAGCGCTGTG	pAH90, pAH91, pAH93, pAH95
90_R	TCGGATCCGGTACCTCATTAGCTGCGATCCACCGC	pAH90
91_R	TCGGATCCGGTACCTCATTAACTCTTGGAAAGATCCAAGCTG	pAH91
911_F	TCTTAATATGTAACTGAATATAACAGCTACATC	pAH911
911_R	CTCTTGGAAAGATCCAAG	pAH911
912_F	GACTCTGAATTAAACAGCTACATC	pAH912
912_R	ATATTAAGACTCTTGGAAAGATC	pAH912

Table 2.1 | Primers used in this work.

Primer identifier	Sequence	Used in this work for cloning of...
93_R	TCGGATCCGGTACCTCATTACAAGGATG TAGCTGTTATATTCAGAGTC	pAH93
95_R	TCGGATCCGGTACCTCATTAGTGGCTTG CATCACCTTCG	pAH95
159_F	TGCCAAGGCGCTTAATATGACTCTGAAT TAAACAGCTACATCC	pAH159
159_R	AGATCCAAGGCGCGATCCACCGCCTTAC G	pAH159
160_F	TCCAGGGTGGATCCGGTGGTGCCGCTGA TCGACGCAATG	pAH160
160_R	TGGAAGTACAGGTTTTCGCCGTGGCTTG CATCACCTTCG	pAH160
136_F	TCTTGAAAAGAGTCTTAATATGACTCTG AATTAAACAGC	pAH136
136_R	TCCAATTCGCGATCCACCGCCTTACG	pAH136
141_F	TATGGAACTGAATTAAACAGCTACATCC TTGGCTGC	pAH141
141_R	TTAAGTTCCTTGGAAAGATCCAAGCTGC	pAH141
DmCM1b _F2	TCCAGGGTGGATCCGGTGGTTCCACAGA ATCCTTAAGCAAACAGCTC	pAH28-6
DmCM1b _R3	TGGAAGTACAGGTTTTCGCCTTATCAAG ACTCCGCCATCTGATAGGCG	pAH28-6
49-F	AGGAAAACCTGTACTTCCAGGGTATGGA CCAGTCTAAACAGGTTTTG	pAH62, pAH82
53-R	TCGGATCCGGTACCTCATTACTTAGCGC AGTCGTTCTCCTTG	pAH62
63_F	AGGAAAACCTGTACTTCCAGGGTGATGC AAGCCACAAAACTTTC	pAH64
64_R	TCGGATCCGGTACCTCATTACTCTACAG TCTTAAGCATCGCAC	pAH64, pAH82
F1_Cal	GAAGTTCTGTTCCAGGGGCCCATGGCCG ATCAACTGACGGAAG	pAH33
R149_Cal	GATCCGGTACCTCATTAGTCTCATTTAG AGGTCATCATGGTAACAAATTCTTCATA GTTTACC	pAH33

Table 2.2 | Further primers used in this work.

Q5 Hot Start High-Fidelity 2x Master Mix (NEB) was used to prepare 25 µl PCR reactions as shown in Table 2.3. Reactions were typically incubated in a thermocycler as outlined in Table 2.4. Where necessary, primer annealing temperature and/or extension time were adjusted experimentally to achieve successful amplification.

Component	Volume	Final concentration
Forward primer (10 µM)	1.25 µl	0.5 µM
Reverse primer (10 µM)	1.25 µl	0.5 µM
DNA template (~ 1 ng/µl)	1 µl	~ 40 pg/µl
2x NEB Master Mix	12.5 µl	1x
Nuclease-free water	9 µl	/

Table 2.3 | Standard PCR reaction mixture. The 2x NEB Master Mix used was Q5 Hot Start High-Fidelity 2x Master Mix.

Step	Temperature	Time
Initial denaturation	98°C	30 s
20 cycles	98°C	10 s
	65°C	20 s
	72°C	2 min 30 s
Final extension	72°C	2 min
Hold	4°C	∞

Table 2.4 | Standard PCR thermocycling conditions.

2.1.2 Restriction enzyme digestion and DNA visualisation

Alternatively, backbone vectors were linearised by digestion with the restriction enzymes EcoRI or NcoI (both NEB). Typically, a reaction mixture as outlined in Table 2.5 was made up. Reactions were incubated at 37°C for 1 hr. EcoRI reactions were then heat-inactivated at 65°C for 20 min, while NcoI reactions were heat-inactivated at 80°C for 20 min. In order to prevent religation, 2 µl of shrimp alkaline phosphatase (NEB) were then added to the reaction mixture, and the reaction incubated for a further 30 min at 37°C before heat-inactivation of the phosphatase at 65°C for 5 min.

Component	Final amount or volume
Restriction enzyme (NEB)	2 µl
DNA template	40 µl (~ 10 µg)
10x NEBuffer (NEB)	5 µl (1x)
Nuclease-free water	1 µl
Shrimp alkaline phosphatase (NEB)	2 µl

Table 2.5 | Typical restriction enzyme reaction. The reaction was incubated for 1 hr at 37°C. Either EcoRI or NcoI (both NEB) was used as restriction enzyme, which were heat inactivated at 65°C and 80°C for 20 min, respectively. At this point, the phosphatase was added, and the reaction incubated for another 30 min at 37°C. The phosphatase was inactivated at 65°C for 5 min.

10 µl of 6x DNA loading dye (ThermoFisher) were then added to the 50 µl reaction containing the linearised backbone vector. The solution was separated on a 1 % (w/v) agarose gel using a PowerPac™ Basic Power Supply (Bio-Rad) at 80 V for 30 min and 40 mM Tris-Acetate (pH 8.5), 1 mM EDTA as running buffer. DNA was visualised with SYBR safe DNA gel stain (Invitrogen) added to the agarose gel from a 10,000x stock concentrate. GeneRuler 1 kb DNA ladder (ThermoFisher) was used as a molecular weight marker. Based on its size, the band corresponding to the linearised backbone vector was then excised and purified using the QIAquick Gel Extraction kit (Qiagen).

2.1.3 Gibson assembly and site-directed mutagenesis

Assembly of the insert into the backbone vector was carried out via Gibson assembly using the NEBuilder HiFi DNA Assembly cloning kit (NEB) as detailed in Table 2.6. For both insert and backbone DNA, 1 µl of PCR product (or gel extraction product) was generally used, and incubation of the 20 µl reaction occurred for 15 min at 50°C in a thermocycler. For site-directed mutagenesis, the KLD Enzyme Mix (NEB) was used with 1 µl of PCR product according to Table 2.7, and the 10 µl reaction was incubated for 5 min at room temperature. In both cases, the assembled constructs were then either directly transformed into *E. coli* DH5αT cells or stored at -20°C.

Component	Volume
Insert DNA	1 µl
Backbone DNA	1 µl
2x NEB Master Mix	10 µl
Nuclease-free water	8 µl

Table 2.6 | HiFi DNA Assembly reaction mixture. Insert and backbone DNA were usually untreated 25 µl PCR reaction products (see Tables 2.3 and 2.4). The 2x NEB Master Mix used was NEBuilder HiFi DNA Assembly 2x Master Mix. The reaction was incubated at 50°C for 15 min.

Component	Volume	Final concentration
PCR product	1 µl	/
2x KLD reaction buffer	5 µl	1x
10x KLD enzyme mix	1 µl	1x
Nuclease-free water	3 µl	/

Table 2.7 | KLD reaction mixture. The reaction was incubated for 5 min at room temperature.

2.1.4 Making chemically competent cells

E. coli cells were made chemically competent via the protocol outlined below. Buffers used are shown in Table 2.8.

Buffer	Components
RF1	15 % (v/v) glycerol 100 mM RbCl 50 mM $MnCl_2$ 30 mM potassium acetate (pH 5.8) 10 mM $CaCl_2$
RF2	15 % (v/v) glycerol 10 mM RbCl 10 mM 3-(*N*-morpholino)propanesulfonic acid (MOPS, pH 6.8) 75 mM $CaCl_2$

Table 2.8 | Buffers used for making chemically competent cells.

Using sterile technique, cells were stroke out from a glycerol stock onto an LB agar plate lacking antibiotics. The plate was left to incubate overnight at 37°C. A single colony was then picked from the plate and used to inoculate 10 ml of 2x YT broth in a 100 ml flask. This was left shaking at 37°C overnight. 1.5 ml of this overnight culture were then used to inoculate 300 ml of 2x YT broth in a 1 L flask. This was left incubating at 37°C until, usually after 2-3 hr, an OD_{600} of 0.4-0.6 was reached. At this point, the cell culture was chilled on ice for at least 10 min. All further steps were carried out at 4°C.

The cell culture was pelleted by centrifugation at 2,400 x g for 10 min in 6 x 50 ml tubes. Pellets were gently resuspended in 17 ml ice-cold buffer RF1 each and left on ice for at least 15 min. Cells were then pelleted again at 2,400 x g for 10 min. This time, pellets were gently resuspended in 4 ml ice-cold buffer RF2 each and left on ice for 15 min. 110 µl aliquots of the now chemically competent cells were prepared and frozen in sterile 96-well blocks on dry ice. Chemically competent cells were stored at -80°C.

2.1.5 Bacterial transformation

For transformation, 2 µl of HiFi DNA Assembly reaction product**,** KLD reaction product or previously cloned plasmid DNA (at a concentration of ~100 ng/µl) were added to 20 µl of chemically competent *E. coli* cells in a 1.2 ml Eppendorf tube. Unless otherwise stated, strain BL21 (DE3) was used for recombinant protein expression (other strains occasionally used include C41 (DE3) and B834 (DE3)) and strain DH5αT was used for cloning. Chemically competent cells were produced in-house (see below). Cells were left on ice for 30 min and then heat shocked at 42°C in a water bath for 45 s. Cells were then incubated on ice again for 5 min. 80 µl of 2x YT medium were then added to the cells, followed by incubation at 37°C with shaking. After 1 hr, the entire volume of cells was plated on agar plates containing 50 µg/ml kanamycin or 100 µg/ml ampicillin depending on the plasmid used. Plates were left at 37°C overnight.

Single colonies of *E. coli* DH5αT cells transformed with the desired plasmid were picked and used to inoculate 5 ml of LB containing 50 µg/ml kanamycin (Melford) or 100 µg/ml ampicillin (Cayman Chemical Company) depending on the plasmid used. Cells were incubated at 37°C with shaking overnight. The next morning, cells were pelleted and DNA extracted using the QIAprep Spin Miniprep kit (Qiagen) according to the manufacturer's protocol. All plasmid inserts were then sequenced to confirm correct assembly.

2.1.6 Plasmids used in this work

Tables 2.9 and 2.10 show the plasmids used in this work, plasmid maps are shown in the corresponding results chapters.

pETM14 is a vector that confers kanamycin resistance and encodes a T7/lac promoter site and an N-terminal, 3C protease cleavable His_6 tag [288]. **pETM44** is a vector that confers kanamycin resistance and encodes a T7/lac promoter site and an N-terminal, 3C protease cleavable His_6-MBP tag [288]. **pLip** is a vector that confers ampicillin resistance and encodes a T7/lac promoter site and N- and C-terminal, TEV protease cleavable, His_6-tagged lipoyl domains derived from *Bacillus stearothermophilus* dihydrolipoamide acetyltransferase [151]. **pLip-mono** is a custom-made vector based on pLip but containing a premature STOP codon that results in constructs only containing the N-terminal tag. **pWaldo-GFP-A206K** is a vector that confers kanamycin resistance and encodes a T7/lac promoter site and a C-terminal, TEV protease cleavable monomeric GFP tag followed by a His_6 tag [289, 290]. **pETM14-mNG** is a pETM14-derived vector (generated in this thesis) that confers kanamycin resistance and encodes a T7/lac promoter site and an N-terminal, 3C protease cleavable His_6 tag followed by a TEV protease cleavable monomeric NeonGreen (mNG) tag (see Appendix for protein sequence).

Plasmid identifier	Insert	Backbone vector	Tag	Source
pLip-Vlt-CM2	Cnn 1082-1148	pLip-mono	N-terminal lipoyl(His)	Violet Feng
CM2 I1126E	Cnn 1082-1148 (I1126E)	pLip-mono	N-terminal lipoyl(His)	Violet Feng
CM2 T1133E	Cnn 1082-1148 (T1133E)	pLip-mono	N-terminal lipoyl(His)	Violet Feng
CM2 L1137E	Cnn 1082-1148 (L1137E)	pLip-mono	N-terminal lipoyl(His)	Violet Feng
pVlt-LZ	Cnn 490-544	pETM44	N-terminal MBP(His)	Violet Feng
pVlt-LZ-L528E	Cnn 490-544 (L528E)	pETM44	N-terminal MBP(His)	Violet Feng
pVlt-LZ-L532E	Cnn 490-544 (L532E)	pETM44	N-terminal MBP(His)	Violet Feng
pVlt-LZ-L535E	Cnn 490-544 (L535E)	pETM44	N-terminal MBP(His)	Violet Feng
pVlt-LZ-L539E	Cnn 490-544 (L539E)	pETM44	N-terminal MBP(His)	Violet Feng
pVlt-LZ-L542E	Cnn 490-544 (L542E)	pETM44	N-terminal MBP(His)	Violet Feng
MBP-PReM	Cnn 403-608	pETM44	N-terminal MBP(His)	Violet Feng
PReM-GFP	Cnn 403-608	pWaldo-GFP-A206K	C-terminal mGFP(His)	Violet Feng

Table 2.9 | Plasmids used in this work. These plasmids were produced by Dr Violet Feng.

Plasmid identifier	Insert	Backbone vector	Tag	Source
pAH37	Cnn 1077-1148	pETM14-mNG	N-terminal mNG(His)	This work
pAH90	Cnn 490-567	pETM14-mNG	N-terminal mNG(His)	This work
pAH91	Cnn 490-573	pETM14-mNG	N-terminal mNG(His)	This work
pAH911	Cnn 490-576	pETM14-mNG	N-terminal mNG(His)	This work
pAH912	Cnn 490-579	pETM14-mNG	N-terminal mNG(His)	This work
pAH136	Cnn 490-579 (S567E, S571E)	pETM14-mNG	N-terminal mNG(His)	This work
pAH141	Cnn 490-579 (S573E, T577E)	pETM14-mNG	N-terminal mNG(His)	This work
pAH159	Cnn 490-579 (S567A, S571A, S573A)	pETM14-mNG	N-terminal mNG(His)	This work
pAH93	Cnn 490-585	pETM14-mNG	N-terminal mNG(His)	This work
pAH95	Cnn 490-608	pETM14-mNG	N-terminal mNG(His)	This work
pAH160	Cnn 553-608	pLip	dilipoyl(His)	This work
pAH28-6	Cnn 136-205	pLip-mono	N-terminal lipoyl(His)	This work
pAH62	Cnn 1-360	pETM14-mNG	N-terminal mNG(His)	This work
pAH64	Cnn 604-1049	pETM14-mNG	N-terminal mNG(His)	This work
pAH82	Cnn 1-1049	pETM14-mNG	N-terminal mNG(His)	This work
pAH33	Calmodulin (FL)	pETM14	N-terminal His	This work

Table 2.10 | Further plasmids used in this work. These plasmids were produced in this work.

2.2 Protein production

2.2.1 Protein expression in bacteria

10 ml of LB or TB medium (both Fisher Chemical) were used to resuspend an inoculated agar plate. 5 ml each of this culture were added to 2 x 1 L of LB (10 g tryptone, 5 g yeast extract, 10 g NaCl per 1 L) or TB (12 g tryptone, 24 g yeast extract, 9.4 g di-potassium phosphate, 2.2 g mono-potassium phosphate, per 1 L supplemented with 0.4 % (v/v) glycerol) medium supplemented with either 50 μg/ml kanamycin or 100 μg/ml ampicillin depending on the plasmid used. Cells were incubated at 37°C with shaking until an optical density at 600 nm of 0.6-0.8 was reached. At this point, protein expression was induced via addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG, VWR Life Science), and cells were left to grow at 20-21°C overnight.

The following day, cells were pelleted by centrifugation at 5,000 x g and 4°C for 15 min. Cell pellets from 2 x 1 L cultures were pooled and resuspended in 30 ml of phosphate-buffered saline (Dulbecco's, Sigma-Aldrich) and pelleted again by centrifugation at 2,400 x g for 15 min. The supernatant was discarded and the pellet retained. These steps are known as "harvesting".

2.2.2 Standard protein purification

Unless otherwise stated, the following protocol was used during purification of proteins of interest. For typical buffer compositions, see Table 2.11.

Cell pellets were resuspended in 30 ml buffer A. Cells were homogenised using a hand-held homogeniser and lysed using an Emulsiflex C5 homogeniser (Avestin) at 15,000 psi and

4°C. The soluble fraction containing the protein of interest was separated from the insoluble fraction by centrifugation at 46,000 x g and 4°C for 30 min.

Buffer	Composition
Buffer A	50 mM Tris-HCl (pH 8.0) 300 mM NaCl 5 mM β-mercaptoethanol (BME)
Buffer B	50 mM Tris-HCl (pH 8.0) 300 mM NaCl 5 mM BME 25 mM imidazole
Buffer C	50 mM Tris-HCl (pH 8.0) 300 mM NaCl 5 mM BME 250 mM imidazole
Buffer D	50 mM Tris-HCl (pH 8.0) 300 mM NaCl 1 mM dithiothreitol (DTT)
Buffer E	50 mM Tris-HCl (pH 8.0) 300 mM NaCl 1 M imidazole
Standard TBS	20 mM Tris-HCl (pH 8.0) 150 mM NaCl 0.5 mM tris(2-carboxethyl)phosphine (TCEP)

Table 2.11 | Typical composition of buffers used during protein purification. The following suppliers were used: Sigma-Aldrich for Tris and BME, VWR Chemicals for NaCl, Acros Organics for imidazole, Iris Biotech for TCEP and GE Healthcare for DTT. TBS is Tris-buffered saline.

Affinity chromatography was carried out using a Ni-NTA Superflow Cartridge (Qiagen) with a column volume (CV) of 2 x 5 ml. The column was pre-equilibrated in buffer A. The supernatant containing the protein of interest was applied to the column using a peristaltic pump at a flow speed of 5 ml/min. The column was washed with 10 CV of buffer B. The His_6-tagged protein of interest was eluted with 3 CV of buffer C. The concentration of protein was measured using a NanoDrop Spectrophotometer ND-1000 (ThermoFisher).

If tag cleavage was desired, recombinant TEV or 3C protease (produced in-house, see below) was added to the sample. Typically, 1 mg of protease was added per 50 mg of protein. The sample was then left dialysing overnight at 4°C against 2 L of buffer D. SnakeSkin™ Dialysis Tubing (ThermoFisher) with an appropriate molecular weight cut-off of either 3,500 or 10,000 Da was used for dialysis. The next day, the dialysed sample was applied at 5 ml/min to a Ni-NTA column (CV of 2 x 5 ml) pre-equilibrated in buffer D. The flow-through containing the cleaved protein of interest was collected. Any uncleaved protein or tag bound to the column was eluted with 3 CV of buffer E and the column equilibrated again in buffer D. If desired, the cleaved protein was further purified by re-applying it to the column as above. These steps were skipped if tag cleavage was not required.

For further purification, the fraction containing the protein of interest was concentrated to 5 ml using centrifugal concentrators (Amicon) of the appropriate molecular weight cut-off, and applied to a HiLoad 16/600 Superdex prep grade column (GE Healthcare) with an appropriate bead pore size (i.e. 75 or 200 depending on the expected size of the protein of interest). The column was pre-equilibrated in standard TBS (Tris-buffered saline). Size exclusion chromatography (SEC) was carried out using an ÄKTA pure (GE Healthcare) at a flow speed of 1 ml/min. The peak containing the protein of interest was pooled and concentrated as above, then flash-frozen in liquid nitrogen and stored at -80°C.

Predictions of molecular weight, pI and extinction coefficient (ε, reported as Abs 0.1 %) were made using ProtParam [291].

2.2.3 Protease production

Recombinant 3C protease was produced from a pGEX plasmid, resulting in GST-tagged protein. After harvesting of a 1 L LB culture, cell pellets were resuspended in 45 ml of buffer containing 20 mM HEPES (pH 7.5), 150 mM NaCl and 1 protease inhibitor tablet (Sigma).

After lysis of cells and centrifugation of the lysate as detailed above, affinity chromatography was carried out using a GSTrap column (GE Healthcare) with CV = 5 ml pre-equilibrated in 20 mM HEPES (pH 7.5), 150 mM NaCl. Supernatant was applied to the column at 1 ml/min, after which the column was washed for 5 x CV using the equilibration buffer. 3C protease was eluted from the GSTrap column using 5-10 ml of elution buffer, which contained 20 mM HEPES (pH 8.0), 150 mM NaCl, 5 % (v/v) glycerol and 10 mM reduced glutathione. During elution of the protein, fractions were taken by successively letting 2 ml of the eluate drip into 2 ml of elution buffer in order to immediately dilute the protein after elution.

10 ml of the fractions containing the highest concentration of 3C protease were pooled and loaded on a Superdex 200 26/600 column (GE Healthcare) for size-exclusion chromatography. The column was pre-equilibrated in 50 mM Tris-HCl (pH 8.0), 150 mM NaCl. The peak containing recombinant 3C protease was pooled and concentrated to 2 mg/ml. The solution was then mixed with the same volume of 50 % (v/v) glycerol, resulting in a final concentration of 1 mg/ml of 3C protein with 25 % (v/v) glycerol. The protease was then flash-frozen in 500 µl aliquots using liquid nitrogen.

Recombinant TEV protease was provided by Dr Violet Feng and Dr Alice Ahn.

2.2.4 Protein gel electrophoresis

Unless otherwise stated, the following samples were taken during protein expression and purification in preparation for sodium dodecyl sulphate – polyacrylamide gel electrophoresis (SDS-PAGE). Buffers are shown in Table 2.12.

Buffer	Composition
3x SDS loading dye	150 mM Tris-HCl (pH 6.8) 30 % (v/v) glycerol 6 % (w/v) SDS 0.3 % (w/v) bromophenol blue 150 mM DTT
1x SDS-PAGE running buffer	25 mM Tris-HCl (pH 8.3) 192 mM glycine 0.1 % (w/v) SDS
15 % SDS-PAGE running gel	15 % (w/v) acrylamide 375 mM Tris-HCl (pH 8.8) 0.1 % (w/v) SDS 0.1 % (w/v) ammonium persulphate (APS) 0.04 % (v/v) tetramethylethylenediamine (TEMED)
12 % SDS-PAGE running gel	12 % (w/v) acrylamide 375 mM Tris-HCl (pH 8.8) 0.1 % (w/v) SDS 0.1 % (w/v) APS 0.04 % (v/v) TEMED
SDS-PAGE stacking gel	5 % (w/v) acrylamide 125 mM Tris-HCl (pH 6.8) 0.1% (w/v) SDS 0.1 % (w/v) APS 0.1 % (v/v) TEMED

Table 2.12 | Composition of buffers used in protein gel electrophoresis. Dilution occurred with H_2O. The 3x SDS loading dye for non-reducing gels lacks DTT. The following suppliers were used: Sigma-Aldrich for TEMED and APS, Bio-Rad for bromophenol blue, VWR Chemicals for SDS, National Diagnostics ProtoGel for the 30 % (w/v) acrylamide stock (37.5:1 ratio to bis-acrylamide). If appropriate, 3x SDS loading dye was diluted to 1x SDS loading dye with H_2O.

1 ml of a 1 L cell culture was taken before induction with IPTG, pelleted and resuspended using 98 µl of 1x SDS loading dye (pre-induction sample). 150 µl of a 1 L cell culture were taken the following day after induction with IPTG, pelleted and resuspended using 98 µl of

1x SDS loading dye (post-induction sample). After homogenisation of the cells and subsequent centrifugation, 98 µl of 1x SDS loading dye were added to 2 µl of supernatant and a touch of the pellet. 98 µl of 1x SDS loading dye were also added to 2 µl of flow-through during Ni-NTA chromatography. 36 µl of 1x SDS loading dye were added to 4 µl of wash and elute during Ni-NTA chromatography, as well as any SEC fraction. For concentrations below ~ 1 mg/ml, 10 µl of 3x SDS loading dye were added to 20 µl of sample.

Samples were incubated at 95°C for 10 min, then 10 µl were loaded on to a 4-20 % SDS-PAGE gel (Bio-Rad Mini-PROTEAN TGX stain-free gels) or a 15 % or 12 % SDS-PAGE gel (produced in-house, Table 2.12). 7 µl of PageRuler Prestained Protein Ladder (ThermoFisher) were used as markers. The gel was run at 200 V using a PowerPac™ Basic Power Supply (Bio-Rad) and with 1x SDS-PAGE running buffer. Staining occurred with Coomassie InstantBlue (Expedeon).

2.3 SEC-MALS

For size-exclusion chromatography multi-angle light scattering (SEC-MALS) experiments, samples were prepared to have a final volume of 110 µl at the desired concentration using standard TBS as a diluent. Samples were separated according to size and shape via size exclusion chromatography using GE Healthcare columns Superdex 75 10/300 (increase), Superdex 200 10/300 (increase) or Superose 6 10/300 (increase). This was followed in line by a Dawn Heleos-II light scattering detector and an Optilab-TrEX refractive index monitor (both Wyatt Technologies). Molecular mass calculations were performed using ASTRA 6.1.1.17 (Wyatt Technologies) assuming a dn/dc value of 0.186 ml/g. Experiments were carried out with the help of Dr Steven Johnson (Lea laboratory) or Dr David Staunton (Molecular Biophysics Suite, Department of Biochemistry, University of Oxford).

2.4 Analytical SEC

Unless otherwise stated, the following protocol was used for analytical size-exclusion chromatography (SEC). Samples were prepared to have a final volume of 100 µl at the desired concentration. After an incubation period of 25-30 min at room temperature, samples were loaded on a GE Healthcare Superdex 75 10/300 (increase) or Superdex 200 10/300 (increase) depending on the size of protein(s) to be analysed. The column was pre-equilibrated in standard TBS (Table 2.6) or a derivative thereof, e.g. containing further components such as $CaCl_2$. SEC was carried out using an ÄKTA pure (GE Healthcare) at a flow speed of 0.8 ml/min. UV spectra were measured at 280 nm or also at 215 nm. 0.5 ml fractions were collected and, if desired, run on an SDS-PAGE gel as described above. Note that separation on a gel filtration column leads to further dilution of the protein of interest.

2.5 Phosphorylation assay

Unless otherwise stated, the following protocol was used for *in vitro* phosphorylation assays. Table 2.13 shows the reaction composition. Recombinant human Polo-like kinase 1 (PLK1, ProQinase) was stored in 50 mM HEPES (pH 7.5), 100 mM NaCl, 5 mM DTT, 15 mM reduced glutathione, 20 % (v/v) glycerol. In control experiments, an equal volume of kinase storage buffer was added instead of PLK1. Reactions were made up to 100 µl or 110 µl with H_2O and incubated at 30°C for 30 min. 1x kinase buffer (Cell Signaling Technology) contained 25 mM Tris-HCl (pH 7.5), 5 mM β-glycerophosphate, 2 mM DTT, 0.1 mM Na_3VO_4, 10 mM $MgCl_2$. ATP was obtained from Cell Signaling Technology.

Component	Stock concentration	Final concentration
Protein of interest	Variable	Variable
PLK1	0.11 mg/ml	0.044 mg/ml
ATP	10 mM	1 mM
Kinase buffer	10x	1x

Table 2.13 | Sample composition for *in vitro* phosphorylation assays

2.6 Fluorescence microscopy

Reactions containing 20 µM of $Cnn^{Q403\text{-}H608}$-mGFP ("PReM-GFP") alone or with an equimolar concentration of WT $Cnn^{H1082\text{-}L1148}$ ("CM2") or a mutant thereof (R1103E, I1126E, T1133E or L1137E) were set up. Red fluorescent polystyrene beads (Invitrogen, 0.2 mm) at 10,000x dilution were added and the reactions made up to 50 µl using 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM BME as buffer. Reactions were incubated at room temperature for 30 min, then 2 µl were pipetted onto a microscope slide and immediately covered with an 18 mm x 18 mm cover glass (VWR). Samples were imaged by Dr Alan Wainman with a Zeiss 880 microscope equipped with an Airyscan detector using a 63x 1.4NA lens.

2.7 X-ray crystallography

2.7.1 Diffraction of X-rays by crystals

Crystals are translationally periodic assemblies of a finite number of unit cells. These unit cells are a combination of a molecular motif and a unit lattice with cell edge lengths a, b, c and angles α, β, γ. There are a limited number of translational and rotational symmetry operations allowable in order to generate a crystal from a chiral protein, giving rise to the 65 chiral space groups. The minimum information required to assemble a unit cell with a certain set of symmetry operations is contained within the so-called asymmetric unit [292].

The unit cell can mathematically be described as being sliced by sets of parallel planes separated by a distance d_{hkl}. Conventionally, these lattice planes are described with so-called Miller indices h, k, l corresponding to the number of times these planes intercept the unit cell edges. From this, a reciprocal lattice can be constructed wherein a given point (h, k, l) is separated from the lattice origin by the reciprocal value of distance d_{hkl} along a vector that is normal to the Miller lattice planes. Where a reciprocal lattice point is found therefore depends on the unit lattice, but the intensity of a given reciprocal lattice point depends on the molecular motif [292].

During a diffraction experiment, X-rays can be conceptualised as being reflected on Miller lattice planes. Sir William Lawrence Bragg showed that maximum constructive interference occurs when:

$$n\lambda = 2d_{hkl} \sin\theta$$

where n is an integer, λ is the X-ray wavelength and θ is the angle between a partial X-ray and a lattice plane. Rotation of the crystal (i.e. varying θ) within an X-ray beam allows

different Miller lattice planes to meet this reflection condition. This can be conceptually visualised by considering that a reflection will occur when a reciprocal lattice point intersects a sphere (the so-called Ewald's sphere) of radius 1/λ around the crystal. Rotation of the crystal is equivalent to rotation of the reciprocal lattice around its origin at a distance of 1/λ from the crystal along the direction of an unreflected X-ray. This will allow other reciprocal lattice points to intersect with the Ewald's sphere, and a reflection will occur [292].

The amplitude and phase of the wave diffracted from the Miller lattice planes h, k, l is described by the structure factor:

$$F_{hkl} = |F_{hkl}|e^{i\alpha_{hkl}}$$

where $|F_{hkl}|$ is the amplitude and $e^{i\alpha_{hkl}}$ represents the phase. The structure factor of a given reflection contains contributions from every atom within the unit cell, i.e. F_{hkl} is a summation of all individual atomic scattering factors. The amplitude of a given atomic scattering factor is proportional to the number of electrons of that atom, and the relative phase depends on its distance from the Miller lattice planes. Atoms that do not lie on a given set of Miller lattice planes h, k, l have a phase contribution to F_{hkl} that is non-zero [292].

2.7.2 Data reduction and quality indicators

After data collection via rotation of the crystal in the X-ray beam, the data are first indexed, i.e. Miller indices h, k, l are assigned to each spot, then integrated, i.e. the intensity of each reflection is calculated. The square root of the intensity of a reflection h, k, l is proportional to the amplitude of F_{hkl}. Intensities are converted to amplitudes in a process called Wilson scaling, which also merges multiple observations of the same reflection while taking into account experimental errors [292].

A number of indicators can be used to determine the quality of diffraction data. The precision of unmerged data is indicated by R_{merge} and R_{meas}, which factors in data multiplicity. The precision of merged data is indicated by R_{pim}. As these factors are statistically not well established, the preferred quality indicator of merged data is $CC_{1/2}$, which is a measure of the correlation between two random halves of the data. $I/\sigma(I)$ is an indicator for the signal-to-noise ratio [293-295].

2.7.3 Solving the phase problem

Unfortunately, we cannot measure the phase of F_{hkl} during a diffraction experiment (the so-called "phase problem") and therefore need to obtain phases via other means.

In this thesis, phases are mostly determined computationally via a method called molecular replacement (MR). In MR, phases from a model with a known structure and likely structural similarity are used to obtain an estimate of initial phases for the unknown structure. Specifically, the model is first rotated and then translated within the unit cell until a good match between observed data from the unknown structure and calculated data from the model is found. For computational ease, this search is sometimes carried out using the so-called Patterson function, which is the Fourier transformation of the structure factor intensities with the phases set to zero and can be represented as a centrosymmetric map of the interatomic vectors of the structure [292]. Coiled-coil proteins can be particularly hard to solve via MR due to their repetitive nature along the helical axis, which might make the translational search ambiguous, or also due to their internal non-crystallographic symmetry, i.e. the intertwined α-helices they consist of often pack in parallel within the asymmetric

unit or with other crystallographic copies, resulting in many nearly identical Patterson vectors [296].

Phases can also be determined experimentally. In this thesis, this is done through a process called anomalous diffraction. Anomalous diffraction occurs when X-rays with an energy close to the absorption edge of an atom (e.g. a heavy metal) are used. In such an experiment, the scattered wave gains an imaginary scattering contribution that is shifted by + 90°. Crucially, this leads to a breakdown of Friedel's law which states that the amplitudes of a Friedel pair (i.e. reflections with Miller indices h, k, l and -h, -k, -l) are equal. The location of the anomalously scattering atom can thus be determined by direct methods [292].

2.7.4 Model refinement and validation

Once amplitudes and initial phases have been determined, structure factors (F_{hkl}), i.e. information in reciprocal space, can be converted to electron density ($\rho(x,y,z)$), i.e. information in real space, via a mathematical operation called Fourier transformation [292]:

$$F_{hkl} = V \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(x,y,z) e^{2\pi i(hx+ky+lz)} dxdydz$$

$$\rho(x,y,z) = \frac{1}{V} \sum_{hkl} |F_{hkl}| e^{-2\pi i(hx+ky+lz)+i\alpha_{hkl}}$$

A model is built into the electron density (commonly presented as a contour map at a standard deviation of 1.0 σ) and the initial phases are improved in successive cycles of refinement. Prior knowledge of typical protein secondary structure restraints and protein stereochemistry can be used to aid refinement. After each cycle of refinement, models are

usually improved using $\mathbf{F_{obs} - F_{calc}}$ and $\mathbf{2F_{obs} - F_{calc}}$ electron density difference maps, wherein F_{obs} are the observed structure factors and F_{calc} are the structure factors calculated from the model. The overall match between F_{obs} and F_{calc}, and therefore the quality of the refined structure, is quantified by the reliability factor R_{work}. In order to avoid over-fitting and for cross-validation, a small set of reflections (usually 5 %) is never used during refinement. R_{free} estimates how well the refined structure predicts this free data set [292]:

$$R = \frac{\sum_{hkl}\left||F_{obs}| - |F_{calc}|\right|}{\sum_{hkl}|F_{obs}|}$$

Model validation is commonly carried out by investigating whether the model closely matches the usual stereochemistry found in protein structures, e.g. by looking at the Ramachandran plot of Φ and Ψ angles as well as common side chain rotamer libraries, and whether there are any clashes of atoms. In the final structure, each atom is described by several parameters such as real space coordinates x, y, z, a temperature B factor which is an estimate of the mobility of the atom, and occupancy which describes what fraction of a given atom in the crystal occupies this position [292].

2.7.5 Practice

In this thesis, crystals were sought using sitting drop vapour diffusion at 295 K. Protein of interest at the desired concentration was mixed with the mother liquor at ratios of 1:1 and 3:1 in 300 nl drops and deposited on MRC Crystallisation Plates (SWISSCI) using a Mosquito crystallisation robot (TTP Labtech). Drops equilibrated against 75 µl of mother liquor. Plates were sealed with ClearVue adhesive film (Molecular Dimensions). A range of commercially available screens (Molecular Dimensions) was used as detailed in the

corresponding chapters. Custom screens were set up using a Dragonfly robot (TTP Labtech). Crystals were fished and frozen in liquid nitrogen with the help of Prof Susan Lea using an appropriate cryoprotectant solution [297].

X-ray diffraction data were collected remotely by Prof Susan Lea using the Diamond Light Source beamline i04 (proposals mx18069 and mx23459) or automatically at the European Synchrotron Radiation Facility (ESRF) beamline MASSIF1 [298-301]. Diffraction data were reduced using Xia2 [302-305], autoPROC [294, 304-307] or XDSAPP [308]. Software suite CCP4 [309, 310], including programs matthews_coef [311, 312], sftools [Hazes, B., unpublished results], phaser [313], molrep [314, 315], chainsaw [316], buccaneer [317, 318] and refmac [319-328], and software suite PHENIX [329], including programs AutoSol [330], phenix.xtriage [331] and phenix.refine [332-335], were used for data analysis and processing. Coot [336, 337] was used for model building and PyMOL [338] for visualisation.

Chapter 3 - Cnn$^{Q403\text{-}H608}$ assembles into micron-scale networks upon binding to the Cnn C-terminus via its zipper-like motif

3.1 Introduction

The *Drosophila* protein Cnn is thought to form networks that act as scaffolds for the recruitment of client proteins to the mature pericentriolar material. This process appears to depend on phosphorylation of Cnn (Fig. 1.8). More specifically, the Cnn region K516-Y601 contains ten conserved Ser/Thr sites, and Polo-mediated phosphorylation on one of these sites, S567, and possibly also on some of the other sites appears to be critically important for the formation of an expanded Cnn network *in vivo*. An N-terminally MBP-tagged Cnn construct covering this region (MBP-Cnn$^{Q403\text{-}H608}$ or MBP-PReM) was found to be dimeric *in vitro* but formed higher order oligomers upon introduction of phospho-mimicking mutations (Glu/Asp) at these ten sites. It was therefore hypothesised that this region has a crucial role in phosphorylation-dependent network assembly [196].

Interestingly, Dr Violet Feng (Lea & Raff laboratory) discovered that the zipper-like motif found within this region (Cnn$^{D490\text{-}K544}$) could bind to the highly conserved C-terminus of Cnn (Cnn$^{H1082\text{-}L1148}$). The conserved Cnn C-terminus, also called CM2, had previously been suggested to be involved in protein-protein interactions with e.g. Plp and Centrocortin (refs [159, 217]), and in humans also Calmodulin (ref [178]).

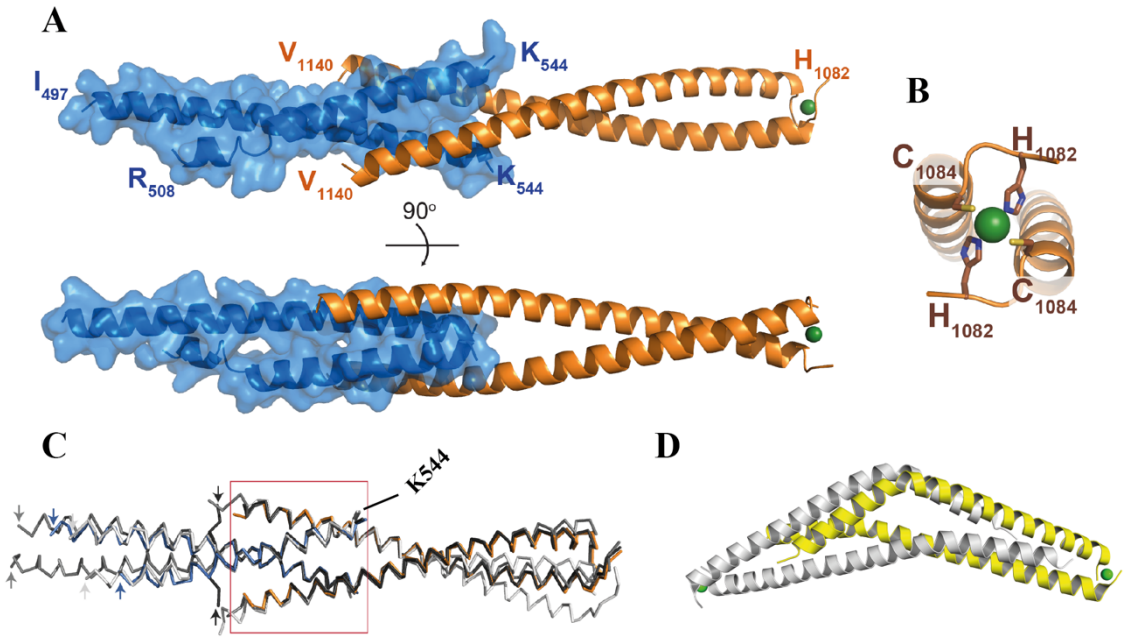


Figure 3.1 | Structure of the complex between Cnn$^{H1082-L1148}$ and Cnn$^{D490-K544}$. **A)** The structure of the complex between Cnn$^{D490-K544}$ (blue) and Cnn$^{H1082-L1148}$ (orange) is shown in cartoon representation (PDB code: 5MVW). The surface of Cnn$^{D490-K544}$ is also shown in blue with reduced opacity. The zinc ion coordinated at the N-terminus of Cnn$^{H1082-L1148}$ is shown as a green sphere. The last residues at the N- and C-termini of the two constructs are indicated. The structure is shown from two orientations rotated by 90°. Note the two constructs come together to form a head-to-head dimer of dimers. **B)** Detail showing the zinc-coordination site at the N-terminus of Cnn$^{H1082-L1148}$. Side chains of zinc-coordinating residues H1082 and C1084 are shown as sticks. **C)** Overlay of the structure shown in A (blue/orange, PDB code: 5MVW) with three other structures of the same complex in grey, i.e. Cnn$^{H1082-L1148}$:Cnn$^{D490-K544(L535E)}$ (PDB code: 5MW0), Cnn$^{H1082-L1148}$:Cnn$^{D490-K544(L535E)}$ (PDB code: 5MW9) and Cnn$^{H1082-L1148}$:Cnn$^{D490-K567}$ (PDB code: 5MWE). All structures are shown in ribbon representation. Note there is little structural variation at the interaction site (red box), but considerable variation towards the N-termini of the constructs, demonstrated by different conformations of the Cnn$^{H1082-L1148}$ N-terminus and variations in the observed N-termini of the different Cnn^{D490-x} constructs (arrows). Further note that despite the increased C-terminal length of the construct Cnn$^{D490-K567}$, no residues beyond K544 (indicated) are observed. **D)** Structure of the Cnn$^{H1082-L1148}$ dimer of dimers (PDB code: 5i7C) in cartoon representation, with one dimer in grey and the other in yellow; zinc ions are shown as green spheres. All structures were solved by Dr Violet Feng and Dr Steven Johnson. Figure modified from [339] under Creative Commons licence CC-BY (see Acknowledgements).

In collaboration with Dr Steven Johnson, Dr Violet Feng then managed to solve several structures of this complex between the zipper-like motif and the Cnn C-terminus shortly before the author joined the lab. In the crystal structure, $Cnn^{D490\text{-}K544}$ and $Cnn^{H1082\text{-}L1148}$ both form parallel dimers, with binding between the two dimers occurring towards their C-termini in an antiparallel fashion, leading to the formation of a 2:2 complex (Fig. 3.1). Note that in this complex structure, $Cnn^{D490\text{-}K544}$ does not form a canonical leucine zipper. Instead, the leucine-rich region previously predicted to form a leucine zipper splays open and binds the Cnn C-terminus, and the motif is therefore referred to as a zipper-like motif. Dr Steven Johnson and Dr Violet Feng also successfully solved the structure of $Cnn^{H1082\text{-}L1148}$ alone, which crystallised as a dimer of dimers [339] (Fig. 3.1 D).

It therefore appears possible that the Cnn C-terminus, by binding to the zipper-like motif found within $Cnn^{Q403\text{-}H608}$, might have a role in Cnn network assembly. This chapter aims to investigate this hypothesis. It will be shown that the interaction between $Cnn^{H1082\text{-}L1148}$ and $Cnn^{D490\text{-}K544}$ can be disrupted via site-directed mutagenesis. Furthermore, evidence from two different techniques will be provided showing that $Cnn^{Q403\text{-}H608}$ assembles into micron-scale assemblies upon incubation with $Cnn^{H1082\text{-}L1148}$, but not with mutants of $Cnn^{H1082\text{-}L1148}$ that cannot bind the zipper-like motif. In addition, constructs covering virtually the entire length of Cnn will be purified, and it will be shown that the Cnn C-terminus only interacts with $Cnn^{Q403\text{-}H608}$ and no other region within Cnn under the conditions tested. Finally, conclusive evidence will be provided showing that $Cnn^{Q403\text{-}H608}$ network assembly is not an artefact due to short construct lengths or the tag system being used.

The work described in this chapter was a collaboration between several researchers in the Lea and Raff laboratories. Unless otherwise stated, the author primarily carried out the

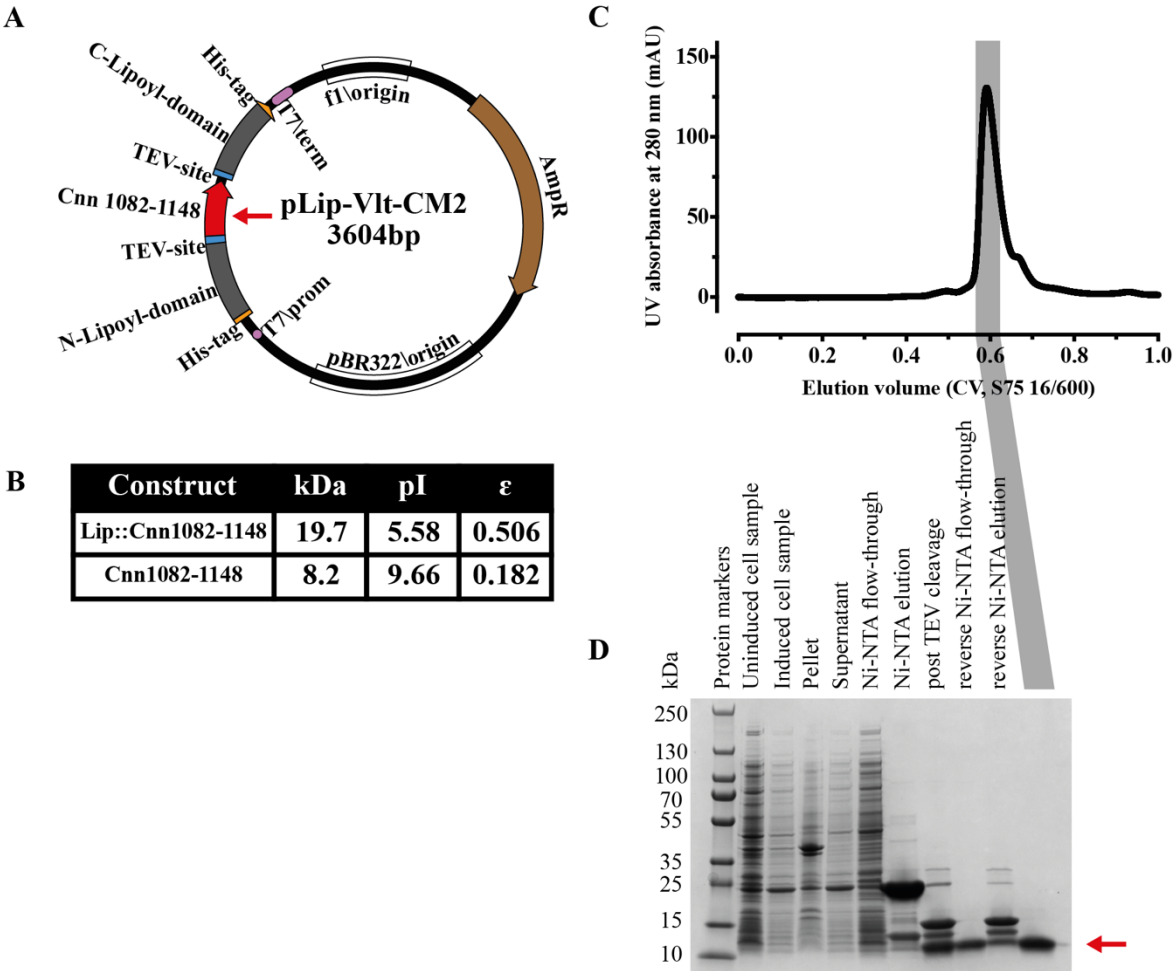
experiments described. For a detailed account of contributions, see main body, figure legends and Acknowledgments section. Much of the work in this chapter has been published in [339], where the author was a middle author.

3.2 $Cnn^{D490\text{-}K544}$:$Cnn^{H1082\text{-}L1148}$ interaction can be disrupted via site-directed mutagenesis

3.2.1 Purification of wild-type $Cnn^{H1082\text{-}L1148}$ and $Cnn^{D490\text{-}K544}$

Dr Violet Feng had previously designed plasmids that encode $Cnn^{H1082\text{-}L1148}$ and $Cnn^{D490\text{-}K544}$ on a pLip-mono and a pETM44 vector, respectively. This allows production of $Cnn^{H1082\text{-}L1148}$ containing an N-terminal, TEV-cleavable, His-Lipoyl-domain (Fig. 3.2 A) and production of $Cnn^{D490\text{-}K544}$ with an N-terminal, 3C-cleavable, His-MBP-domain (Fig. 3.3 A). Furthermore, Dr Violet Feng had previously purified these proteins for structural studies, but in order for further functional studies to be carried out, more protein was produced by the author on several occasions. The protocols used by the author resulting in the highest yield of protein will therefore be described below.

$Cnn^{H1082\text{-}L1148}$ was purified from *E. coli* C41 (DE3) cells grown in 8 L of LB according to the usual protocol but using buffers A-E at pH 7.3 (see Materials and methods). In addition, before homogenisation of the harvested cells, 40 µl of $ZnCl_2$ (from a 100 mM stock) were added to the resuspended cell pellet in order to aid zinc-dependent dimerisation of $Cnn^{H1082\text{-}L1148}$ (Fig. 3.1 B). After Ni-affinity chromatography, 1.5 ml of TEV protease (from a 1.25 mg/ml stock) were added and the solution dialysed overnight. The next day, reverse Ni-affinity chromatography was carried out and the first flow-through collected, concentrated to 5 ml and applied to a S75 16/600 column for SEC using standard TBS. The SEC profile had a single peak (CV = 0.57-0.62) containing $Cnn^{H1082\text{-}L1148}$ as judged by the corresponding SDS-PAGE gel (Fig. 3.2). The peak was therefore pooled and concentrated to 16.8 mg/ml and flash-frozen in 50 µl aliquots, with a final yield of ~ 2 mg/L.

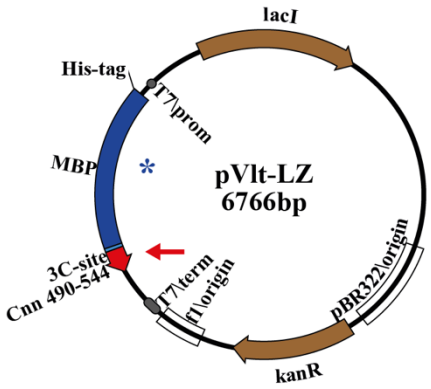


Construct	kDa	pI	ε
Lip::Cnn1082-1148	19.7	5.58	0.506
Cnn1082-1148	8.2	9.66	0.182

Figure 3.2 | Purification of Cnn$^{H1082-L1148}$. A) Overview of plasmid pLip-Vlt-CM2, which encodes Cnn$^{H1082-L1148}$ with an N-terminal, TEV-cleavable, His-Lipoyl-domain. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged protein and the protein after cleavage with TEV. **C)** SEC (S75 16/600) elution profile of Cnn$^{H1082-L1148}$ after Ni-affinity chromatography, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.57-0.62) indicates the fractions pooled after SEC and run on the SDS-PAGE gel in D, containing Cnn$^{H1082-L1148}$. SEC was carried out using standard TBS. **D)** 4-20 % reducing SDS-PAGE gel showing successive purification steps; the band corresponding to Cnn$^{H1082-L1148}$ is highlighted with a red arrow. The plasmid was cloned by Dr Violet Feng and the protein had previously been purified by her.

$Cnn^{D490\text{-}K544}$ was produced by growing an 8 L culture of *E. coli* BL21 (DE3) cells in LB and purifying the protein of interest via a variation of the usual protocol with buffers A-E at pH 7.3. Ni-affinity chromatography was carried out using a Ni-NTA column with CV = 5 ml pre-equilibrated in buffer A containing an additional 10 mM imidazole to prevent non-specific binding. In order to maximise binding to the column, the post-homogenisation supernatant containing MBP-$Cnn^{D490\text{-}K544}$ was applied to the column, but the flow-through was at first allowed to flow back into the supernatant, leading to the supernatant being looped around the column. After 30 min, the flow-through was finally collected and a more stringent wash with 15 CV of buffer B was carried out before elution of MBP-$Cnn^{D490\text{-}K544}$. Dialysis occurred according to the usual protocol after the addition of 2 mg of 3C protease (2 ml from a 1 mg/ml stock). Reverse Ni-affinity chromatography was carried out twice the next day, and the second flow-through concentrated to 10 ml and applied to a S75 26/600 column pre-equilibrated in standard TBS at pH 7.3. The prominent peak on the SEC elution profile (CV = 0.46-0.50) was confirmed to contain $Cnn^{D490\text{-}K544}$ via SDS-PAGE, pooled and concentrated to 9.3 mg/ml (50 µl aliquots) before flash-freezing (Fig. 3.3). The final yield was ~ 0.8 mg/L.

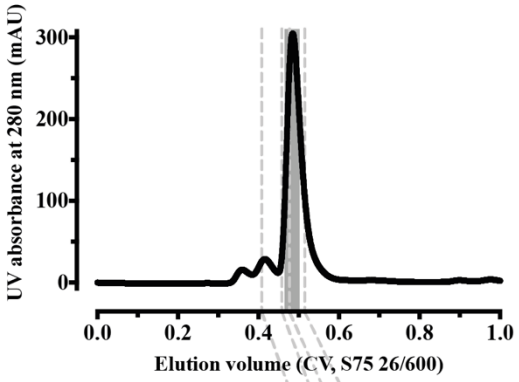
A



B

Construct	kDa	pI	ε
MBP::Cnn490-544	49.0	5.31	1.466
Cnn490-544	6.5	4.65	0.842

C



D

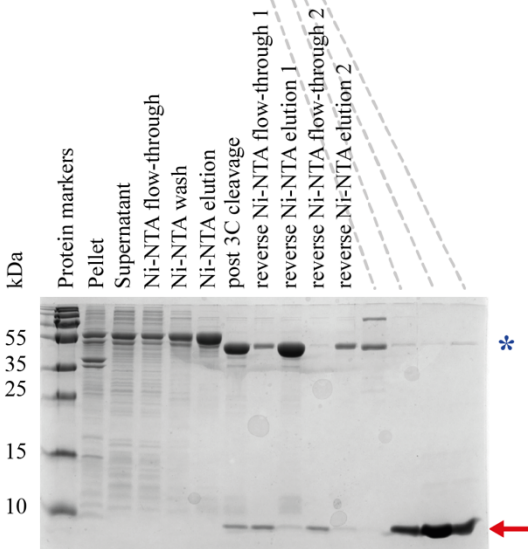


Figure 3.3 | Purification of Cnn$^{D490-K544}$. A) Overview of plasmid pVlt-LZ, which encodes Cnn$^{D490-K544}$ with an N-terminal, 3C-cleavable, His-MBP-domain. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of tagged protein and the protein after cleavage with 3C. **C)** SEC (S75 26/600) elution profile of Cnn$^{D490-K544}$ after Ni-affinity chromatography, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 320 ml) on the x-axis. Grey area (CV = 0.46-0.50) indicates the fractions pooled after SEC, containing Cnn$^{D490-K544}$; fractions run on the SDS-PAGE gel in D are shown with dashed grey lines. SEC was carried out using standard TBS at pH 7.3. **D)** 15 % reducing SDS-PAGE gel showing successive purification steps; the band corresponding to Cnn$^{D490-K544}$ is highlighted with a red arrow, the band corresponding to the MBP-tag is highlighted with a blue asterisk. The plasmid was cloned by Dr Violet Feng and the protein had previously been purified by her.

3.2.2 Purification and characterisation of $Cnn^{D490\text{-}K544}$ mutants

With the structure of the complex between $Cnn^{D490\text{-}K544}$ and $Cnn^{H1082\text{-}L1148}$ having been solved, it was possible for Dr Violet Feng to design single point mutants of both constructs targeting the interface between them, in an attempt to disrupt this interaction. First, she designed and purified several mutants of $Cnn^{H1082\text{-}L1148}$ targeting conserved residues found in the interface, i.e. I1126E, T1133E, and L1137E. In addition, she also designed several point mutants of $Cnn^{D490\text{-}K544}$ targeting the leucines of the zipper-like motif, which were also buried in the 2:2 interface, i.e. L528E, L532E, L535E, L539E, and L542E. Purification of all $Cnn^{D490\text{-}K544}$ mutants was successfully carried out in a joint effort by Dr Violet Feng and the author and will therefore be recorded here. Purification of $Cnn^{D490\text{-}K544(L539E)}$ was carried out by the author alone and will therefore be described in more detail.

Each plasmid of mutant $Cnn^{D490\text{-}K544}$ was transformed into *E. coli* B834 (DE3) cells by the author and grown up in 2 L of TB each. IPTG induction and protein production occurred according to the usual protocol, and cell pellets were frozen at -20°C after harvesting. Over-production of tagged $Cnn^{D490\text{-}K544}$ mutants was successful in each case (Fig. 3.4).

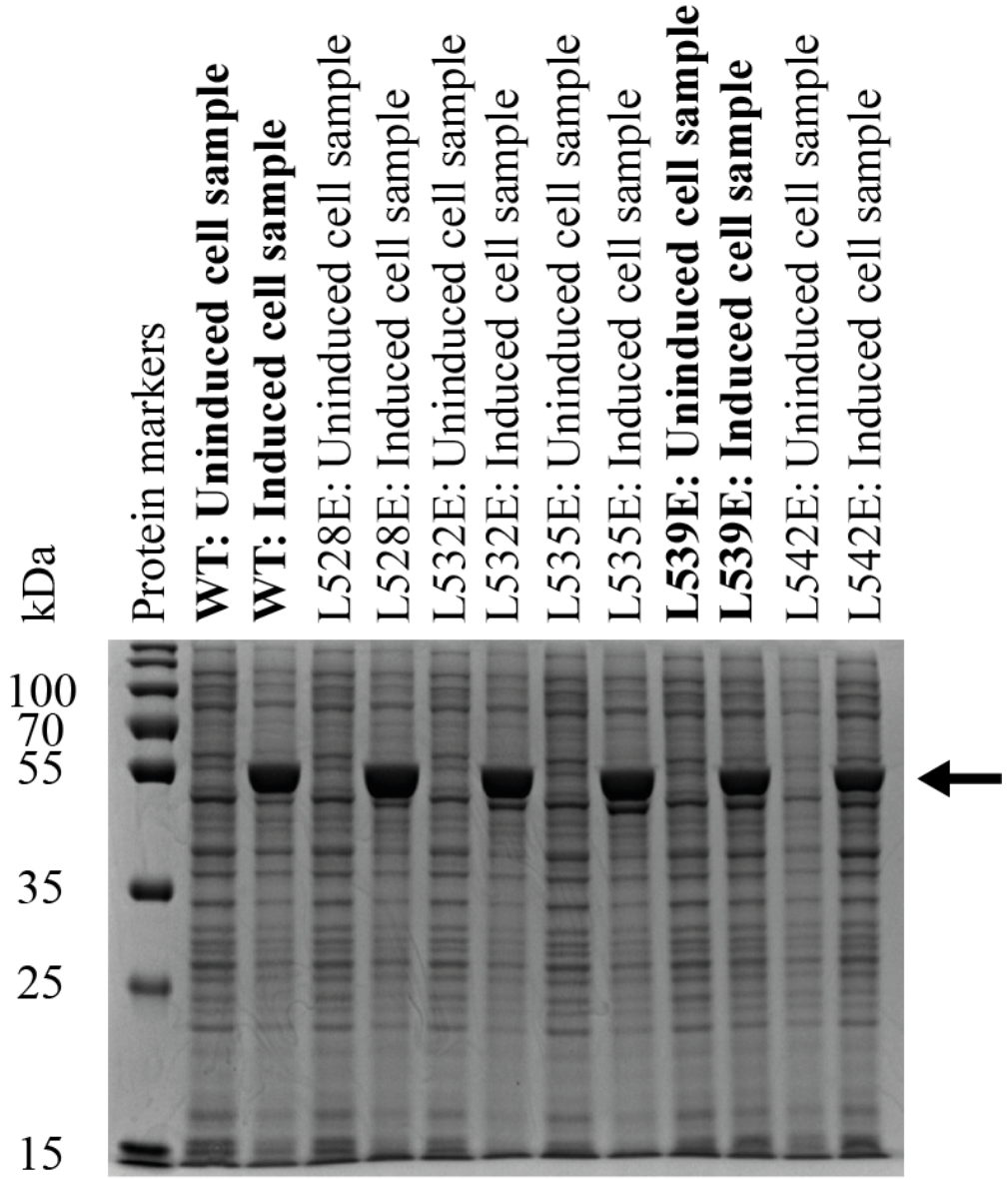
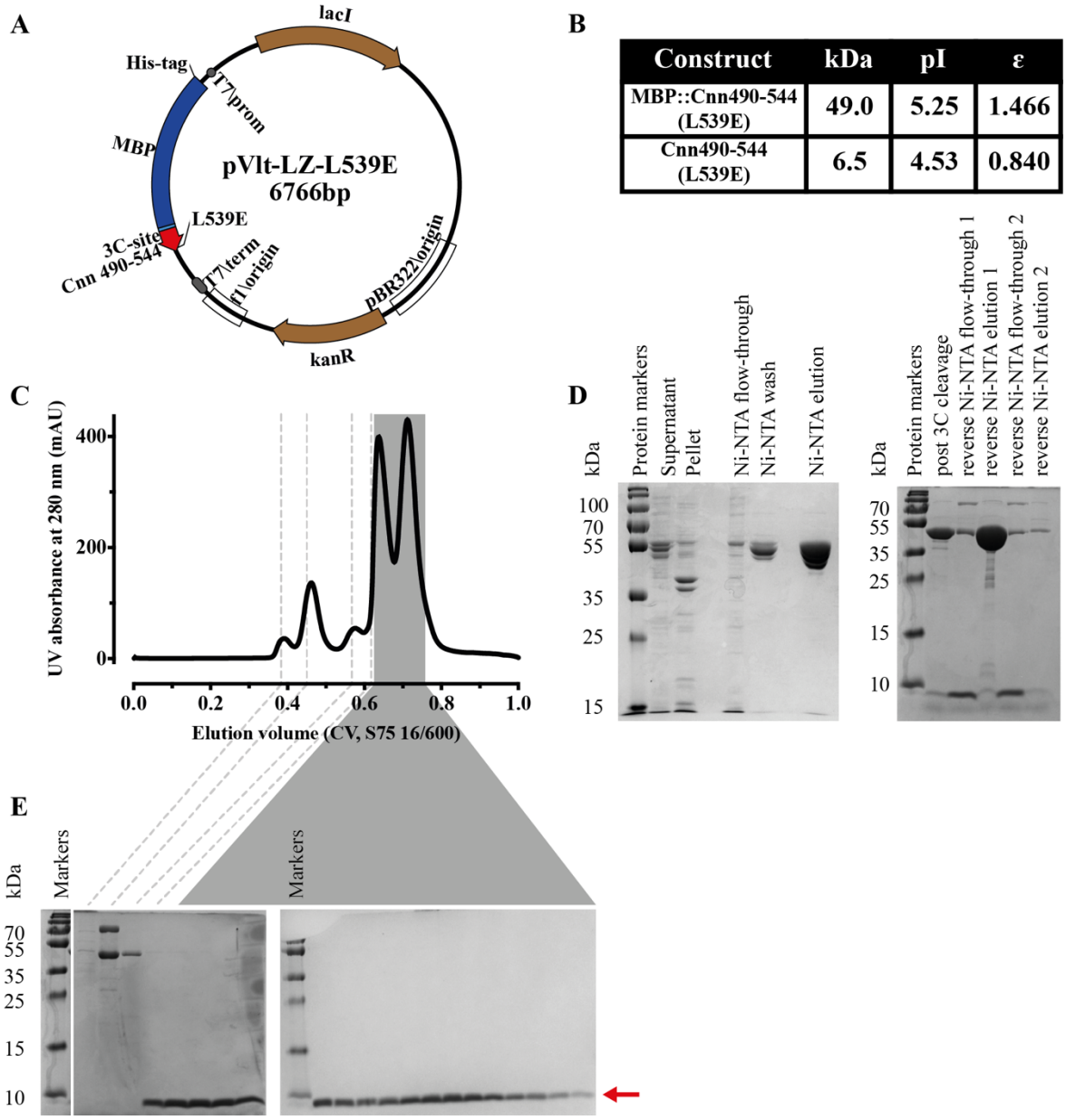


Figure 3.4 | Over-expression of WT Cnn$^{D490\text{-}K544}$ and its mutants. 12 % reducing SDS-PAGE gel showing successful overproduction of His-MBP-Cnn$^{D490\text{-}K544}$ and mutants thereof in B834 (DE3) cells. Black arrow indicates the tagged protein of interest. Subsequent purification of the constructs highlighted in bold (i.e. WT and L539E) was completely carried out by the author, all other constructs were purified in collaboration with Dr Violet Feng. All plasmids were cloned by Dr Violet Feng.

$Cnn^{D490\text{-}K544(L539E)}$ was purified using a protocol that differs from the standard one described in Chapter 2 in order to maintain consistency with the other mutants of $Cnn^{D490\text{-}K544}$ purified in collaboration with Dr Violet Feng. Specifically, rather than treatment with a hand-held homogeniser, cell pellets were resuspended in buffer containing 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 5 mM BME, 1 protease inhibitor tablet (Sigma), 1 mg/ml lysozyme (Sigma) and 50 μg/ml DNase (Sigma). This was left stirring at 4°C for 30 min before homogenisation with an Emulsiflex C5 homogeniser (Avestin). Ni-affinity chromatography was carried out using 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 5 mM BME buffer supplemented with 10 mM imidazole for sample loading, 20 mM imidazole for column washing (15 CV in total) and 300 mM imidazole for sample elution. Note the supernatant after homogenisation was also supplemented with 10 mM imidazole before application to the Ni-NTA column in an effort to reduce non-specific binding.

After elution of MBP-$Cnn^{D490\text{-}K544(L539E)}$ from the Ni-NTA column with 18 ml of the elution buffer described above, 2.5 mg of 3C protease (1 mg/ml stock) were added and the solution dialysed against buffer containing 50 mM Tris-HCl (pH 7.5), 300 mM NaCl, 1 mM DTT and 5 % (v/v) glycerol. Glycerol was added to the dialysis buffer in an effort to increase solubility of the mutant construct.



Construct	kDa	pI	ε
MBP::Cnn490-544 (L539E)	49.0	5.25	1.466
Cnn490-544 (L539E)	6.5	4.53	0.840

Figure 3.5 | Purification of Cnn$^{D490\text{-}K544(L539E)}$. **A)** Overview of plasmid pVlt-LZ-L539E, which encodes Cnn$^{D490\text{-}K544(L539E)}$ with an N-terminal, 3C-cleavable, His-MBP-domain. The plasmid was cloned by Dr Violet Feng. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged protein and the protein after cleavage with 3C. **C)** SEC (S75 16/600) elution profile of Cnn$^{D490\text{-}K544(L539E)}$ after Ni-affinity chromatography, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis. Grey area (CV = 0.63-0.76) indicates the fractions pooled after SEC, containing Cnn$^{D490\text{-}K544(L539E)}$, and run on the SDS-PAGE gel in E; further fractions run on the SDS-PAGE gel in E are shown with dashed grey lines. SEC was carried out using buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM BME, 5 % glycerol. **D)** 12 % (left) and 15 % (right) reducing SDS-PAGE gels showing successive purification steps. **E)** Two 15 % reducing SDS-PAGE gels showing various fractions taken during SEC; the band corresponding to Cnn$^{D490\text{-}K544(L539E)}$ is highlighted with a red arrow. Note the protein elutes in two peaks during SEC, possibly suggesting different oligomeric states.

The next day, reverse Ni-affinity chromatography was carried out with a Ni-NTA column (CV = 10 ml) pre-equilibrated in 50 mM Tris-HCl (pH 7.5), 300 mM NaCl, 30 mM imidazole. The dialysed sample was also made up to contain 30 mM imidazole before application to the column, taking into account the reduced imidazole concentration after dialysis the night before. Reverse Ni-affinity chromatography was carried out twice, with column elution between sample applications occurring with 50 mM Tris-HCl (pH 7.5), 300 mM NaCl, 1 M imidazole. The second flow-through was collected, concentrated to 5 ml and applied to a S75 16/600 column pre-equilibrated in 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 % (v/v) glycerol. Interestingly, the SEC elution profile had two prominent peaks, both of which contained $Cnn^{D490\text{-}K544(L539E)}$ as determined via SDS-PAGE (Fig. 3.5). Both peaks were therefore pooled together (CV = 0.63-0.76), concentrated to 6.9 mg/ml and flash-frozen in 55 µl aliquots, with a final yield of ~ 2-3 mg/L.

Having purified $Cnn^{H1082\text{-}L1148}$, $Cnn^{D490\text{-}K544}$ and mutants of both constructs, functional characterisation of the interaction was now possible. In the SEC-MALS experiments described below, samples were prepared by Dr Violet Feng, SEC-MALS carried out by Dr Steven Johnson and data analysis and figures done by the author. For sample preparation and SEC-MALS conditions, see Fig. 3.6 legend and [339].

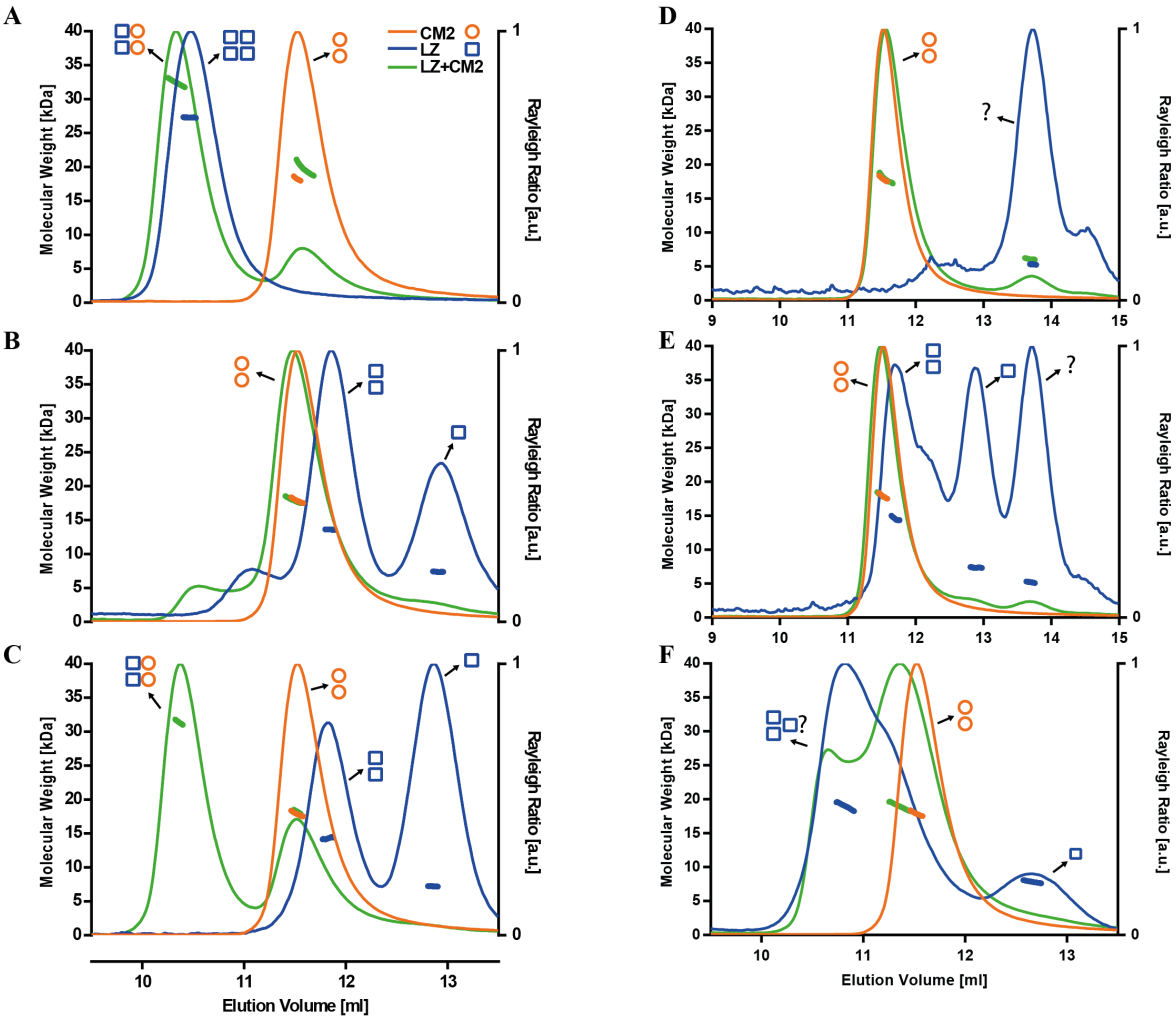


Figure 3.6 | SEC-MALS analysis of WT Cnn$^{H1082-L1148}$ and Cnn$^{D490-K544}$ or mutants thereof. A) WT Cnn$^{H1082-L1148}$ and WT Cnn$^{D490-K544}$. **B)** WT Cnn$^{H1082-L1148}$ and Cnn$^{D490-K544(L528E)}$. **C)** WT Cnn$^{H1082-L1148}$ and Cnn$^{D490-K544(L535E)}$. **D)** WT Cnn$^{H1082-L1148}$ and Cnn$^{D490-K544(L532E)}$. **E)** WT Cnn$^{H1082-L1148}$ and Cnn$^{D490-K544(L539E)}$. **F)** WT Cnn$^{H1082-L1148}$ and Cnn$^{D490-K544(L542E)}$. Cnn$^{H1082-L1148}$ ("CM2") is shown in orange, Cnn$^{D490-K544}$ and its mutants ("LZ") are shown in blue, both mixed together shown in green ("LZ+CM2"). The most likely oligomeric state, as based on the measured molecular weight, is shown as orange circles (CM2) and blue squares (LZ). Unclear oligomeric states are indicated with a question mark. Note binding and therefore 2:2 formation between the two constructs only occurs with WT Cnn$^{D490-K544}$ (A) and Cnn$^{D490-K544(L535E)}$ (C); in all cases, mutation of Cnn$^{D490-K544}$ disrupts tetramerisation of the construct. Molecular weight (in kDa) is plotted on the left y-axis, Rayleigh ratio (relative scale) plotted on the right y-axis, elution volume (in ml) of column S75 10/300 plotted on the x-axis. All proteins tested were purified with the author's help; SEC-MALS experiments were carried out by Dr Violet Feng and Dr Steve Johnson after dialysis of all constructs into 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 10 µM $ZnCl_2$, 5 mM BME. The figure was produced by the author and has been published in a modified form in [339].

WT $Cnn^{H1082\text{-}L1148}$ was observed to form dimers and WT $Cnn^{D490\text{-}K544}$ formed tetramers, but upon complexation the two proteins formed species with an average mass most closely corresponding to a 2:2 complex (Fig. 3.6 A), as had previously been observed by Dr Violet Feng. Interestingly, all point mutants of $Cnn^{D490\text{-}K544}$ led to a disruption of the tetrameric fold. L528E (Fig. 3.6 B), L535E (Fig. 3.6 C) and L539E (Fig. 3.6 E) point mutants led to the breakdown of $Cnn^{D490\text{-}K544}$ tetramers into mixtures of dimers and monomers. $Cnn^{D490\text{-}K544(L542E)}$ also failed to efficiently form tetramers, and instead a peak with an average mass most closely resembling that of a trimer was observed, in addition to a much smaller peak corresponding to monomers (Fig. 3.6 F). Finally, $Cnn^{D490\text{-}K544(L532E)}$ and $Cnn^{D490\text{-}K544(L542E)}$ also formed species the mass of which was even below that expected for a monomer (Fig. 3.6 D/E). This could be the result of limited proteolysis of the monomeric mutant constructs.

Importantly, none of the mutant $Cnn^{D490\text{-}K544}$ constructs could bind $Cnn^{H1082\text{-}L1148}$ with the exception of $Cnn^{D490\text{-}K544(L535E)}$. The interaction between the two constructs was therefore successfully disrupted. Furthermore, these results would suggest that tetramerisation of $Cnn^{D490\text{-}K544}$ and binding of $Cnn^{H1082\text{-}L1148}$ occurs via the same interface on $Cnn^{D490\text{-}K544}$, as both are disrupted in the $Cnn^{D490\text{-}K544}$ point mutants purified here. By solving the crystal structure of the $Cnn^{D490\text{-}K544(L535E)}$:$Cnn^{H1082\text{-}L1148}$ complex, Dr Violet Feng was later able to show that the L535E mutation does not disrupt interaction with $Cnn^{H1082\text{-}L1148}$ because the mutant Glu residue is accommodated within the binding interface by hydrogen bonding to the conserved residue T1133 (ref [339], PDB codes 5MW0 and 5MW9). Furthermore, Dr Violet Feng also confirmed that the mutations within $Cnn^{H1082\text{-}L1148}$ described above (i.e. I1126E, T1133E, and L1137E) also led to a disruption of the interaction between these two regions within Cnn (ref [339]).

3.3 $Cnn^{Q403\text{-}H608}$ forms micron-scale assemblies upon binding to $Cnn^{H1082\text{-}L1148}$

In order to investigate whether binding of $Cnn^{H1082\text{-}L1148}$ to the zipper-like motif has a role in the proposed function of $Cnn^{Q403\text{-}H608}$ in network assembly, Dr Violet Feng designed a construct that results in $Cnn^{Q403\text{-}H608}$ with a C-terminal monomeric GFP-tag. Purification of $Cnn^{Q403\text{-}H608}$-mGFP was carried out by the author using the same purification protocol described above for $Cnn^{D490\text{-}K544(L539E)}$, although no tag cleavage occurred and SEC was carried out using a S200 16/600 column pre-equilibrated in 50 mM Tris-HCl (pH8.0), 150 mM NaCl, 5 mM BME. The peak predominantly containing $Cnn^{Q403\text{-}H608}$-mGFP as based on its green colour and elution position on an SDS-PAGE gel was pooled (CV = 0.60-0.66), concentrated to 2.8 mg/ml and flash frozen in 50 µl aliquots, resulting in a yield of ~ 2 mg/L (Fig. 3.7).

Fluorescence microscopy experiments were carried out to investigate the interaction between $Cnn^{H1082\text{-}L1148}$ and $Cnn^{D490\text{-}K544}$ in the context of the larger $Cnn^{Q403\text{-}H608}$-mGFP construct (see Materials and methods). Strikingly, micron-scale network-like assemblies were observed under the fluorescent microscope when $Cnn^{Q403\text{-}H608}$-mGFP was incubated with $Cnn^{H1082\text{-}L1148}$, but no such assemblies appeared in the absence of $Cnn^{H1082\text{-}L1148}$. Importantly, these assemblies also did not occur when $Cnn^{Q403\text{-}H608}$-mGFP was incubated with one of the mutants of $Cnn^{H1082\text{-}L1148}$ that disrupt interaction with the zipper-like motif (Fig. 3.8 A and B). These observations suggest that $Cnn^{Q403\text{-}H608}$-mGFP assemblies are not non-specific aggregates but rather depend on the precise interaction between $Cnn^{H1082\text{-}L1148}$ and the zipper-like motif found within $Cnn^{Q403\text{-}H608}$-mGFP.

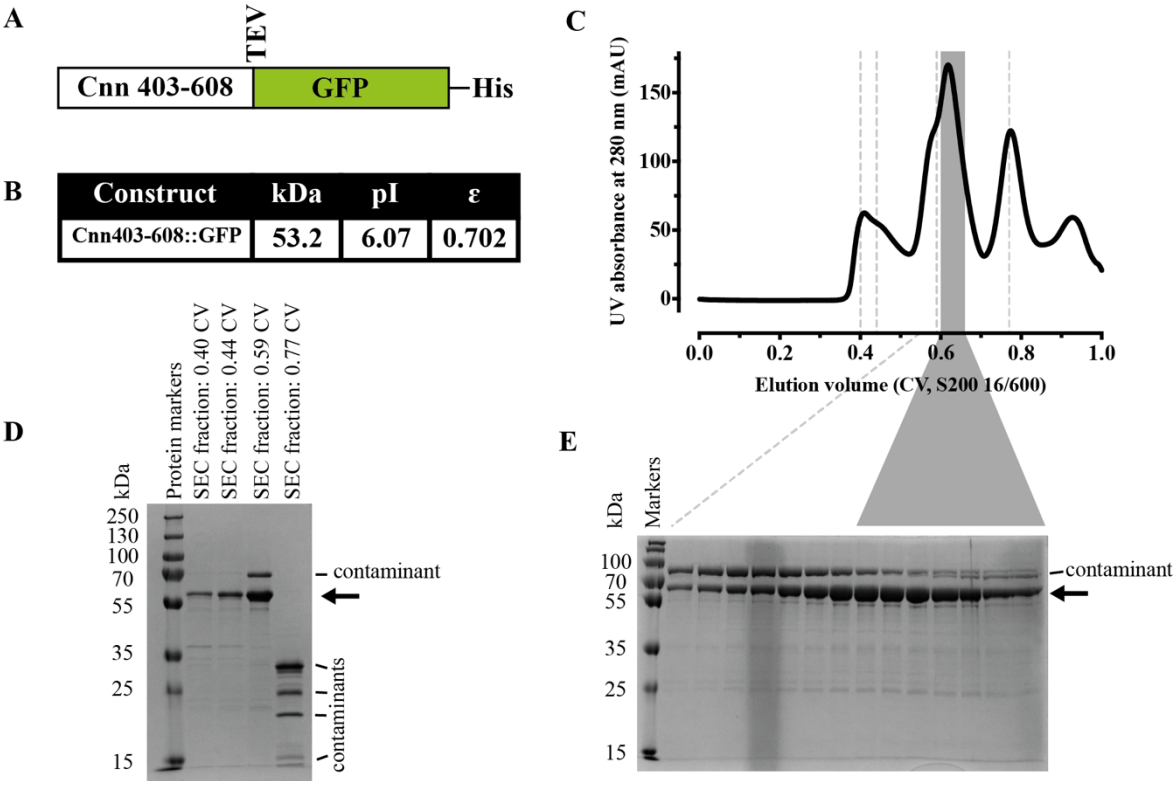


Figure 3.7 | Purification of C-terminally GFP-tagged Cnn$^{Q403\text{-}H608}$. A) Schematic overview of the construct used, i.e. Cnn$^{Q403\text{-}H608}$ with a C-terminal, TEV-cleavable, mGFP-His-domain. The corresponding plasmid (pWaldo-GFP-A206K backbone) was cloned by Dr Violet Feng. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged protein. **C)** SEC (S200 16/600) elution profile of Cnn$^{Q403\text{-}H608}$-mGFP after Ni-affinity chromatography, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis. Grey area (CV = 0.60-0.66) indicates the fractions pooled after SEC, containing Cnn$^{Q403\text{-}H608}$-mGFP, and run on the SDS-PAGE gel in E; further fractions run on the SDS-PAGE gels in D and E are shown with dashed grey lines. SEC was carried out using 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 5 mM BME as buffer. **D)** 12 % reducing SDS-PAGE gel showing several SEC elution fractions. E) 12 % reducing SDS-PAGE gel showing various fractions taken across the main SEC peak containing Cnn$^{Q403\text{-}H608}$-mGFP, as determined by its green colour. In both (D) and (E), Cnn$^{Q403\text{-}H608}$-mGFP is indicated with an arrow; contaminants are also labelled. The final pool contained protein of an acceptable purity for downstream applications.

In order to exclude the possibility that the type or location of tag being used has an impact on the observations made, Dr Violet Feng also purified a $Cnn^{Q403\text{-}H608}$ construct with an N-terminal MBP-tag [339]. With the help of Dr Steven Johnson, the author carried out SEC-MALS experiments on this construct on a Superose 6 column pre-equilibrated in buffer 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 10 μM $ZnCl_2$, 5 mM BME. As expected, incubation of MBP-$Cnn^{Q403\text{-}H608}$ (at 200 μM) with $Cnn^{H1082\text{-}L1148}$ (at 300 μM) resulted in the formation of large assemblies, characterised by a clear peak shift and the formation of species at least an order of magnitude larger than dimeric MBP-$Cnn^{Q403\text{-}H608}$ on its own. As in the fluorescence experiments described above, $Cnn^{H1082\text{-}L1148}$ with a single point mutant (T1133E) that disrupts binding to the zipper-like motif could not induce formation of these large-scale assemblies (Fig. 3.8 C and D).

Taken together, these experiments demonstrate that $Cnn^{Q403\text{-}H608}$ forms micron-scale networks upon binding of the Cnn C-terminus to the zipper-like motif, and that these assemblies are not an artefact due to the tag system being used, as they are observed to form both with an N-terminal MBP tag and a C-terminal mGFP tag on $Cnn^{Q403\text{-}H608}$ (Fig. 3.8).

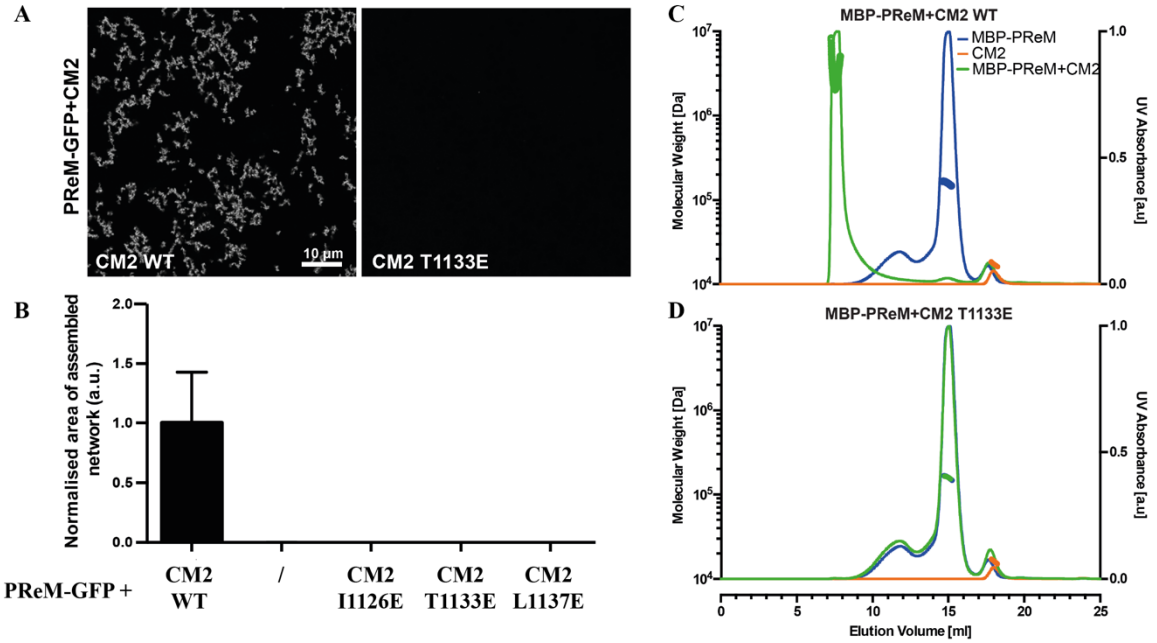


Figure 3.8 | $Cnn^{Q403\text{-}H608}$ forms micron-scale assemblies upon mixing with $Cnn^{H1082\text{-}L1148}$ but not with any of its point mutants. A/B) Fluorescence microscopy analysis of $Cnn^{Q403\text{-}H608}$-mGFP ("PReM-GFP") alone, mixed with WT $Cnn^{H1082\text{-}L1148}$ ("CM2") or mutants thereof. Representative micrographs for two conditions are shown in (A), quantifications for all conditions shown in (B). Note $Cnn^{Q403\text{-}H608}$-mGFP only assembles into micron-scaled assemblies in the presence of WT $Cnn^{H1082\text{-}L1148}$. $Cnn^{Q403\text{-}H608}$-mGFP was purified by the author; CM2 mutants were purified by Dr Violet Feng; samples were prepared by the author, imaging and data analysis was done by Dr Alan Wainman. **C/D)** SEC-MALS analysis of N-terminally His-MBP-tagged $Cnn^{Q403\text{-}H608}$ (MBP-PReM) and WT $Cnn^{H1082\text{-}L1148}$ (C) or $Cnn^{H1082\text{-}L1148(T1133E)}$ (D). $Cnn^{H1082\text{-}L1148}$ ("CM2") and $Cnn^{H1082\text{-}L1148(T1133E)}$ ("CM2 T1133E") are shown in orange, MBP-$Cnn^{Q403\text{-}H608}$ is shown in blue, both mixed together shown in green ("MBP-PReM + CM2"). Note how in the presence of WT $Cnn^{H1082\text{-}L1148}$, but not $Cnn^{H1082\text{-}L1148(T1133E)}$, the peak containing dimeric MBP-$Cnn^{Q403\text{-}H608}$ disappears, and a new void peak containing species with a molecular weight at least an order of magnitude higher than that of dimeric MBP-$Cnn^{Q403\text{-}H608}$ appears. Molecular weight (in Da) is plotted on the left y-axis, UV absorbance (arbitrary units) plotted on the right y-axis, elution volume (in ml) of column Superose 6 10/300 plotted on the x-axis. MBP-$Cnn^{Q403\text{-}H608}$ was at 200 µM and WT or mutant $Cnn^{H1082\text{-}L1148}$ at 300 µM. SEC-MALS was carried out using 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 10 µM $ZnCl_2$, 5 mM BME as buffer. Figure modified from [339] under Creative Commons licence CC-BY (see Acknowledgements).

3.4 Cnn C-terminus interaction with $Cnn^{Q403\text{-}H608}$ appears to be specific

3.4.1 Purification of constructs almost completely covering Cnn

It remained to be demonstrated that $Cnn^{Q403\text{-}H608}$ network assembly upon Cnn C-terminus binding is specific to this region with Cnn, and that $Cnn^{Q403\text{-}H608}$ network assembly is a genuine function of this region of Cnn rather than a result of having used construct boundaries that are too small and therefore potentially result in inherently unstable protein *in vitro*.

To address these two questions, constructs covering virtually the entire length of Cnn were purified using a novel tag system the author generated by genetically modifying a pETM14 vector to provide an N-terminal, TEV-cleavable His_6-3C-mNG tag (monomeric neon-green protein with an N-terminal, 3C-cleavable hexa-His tag). Protein purification systems using fluorescently-tagged protein have previously been used, such as in the production of high quantities of membrane proteins [289, 340]. As will be shown in this sub-chapter, this mNG-tag system often results in high over-production of protein in *E. coli* BL21 (DE3), good protein solubility, the ability to visually observe the tagged protein of interest during purification, as well as the flexibility to selectively cleave off the His_6-tag only (with 3C protease) or the entire mNG-tag (with TEV protease).

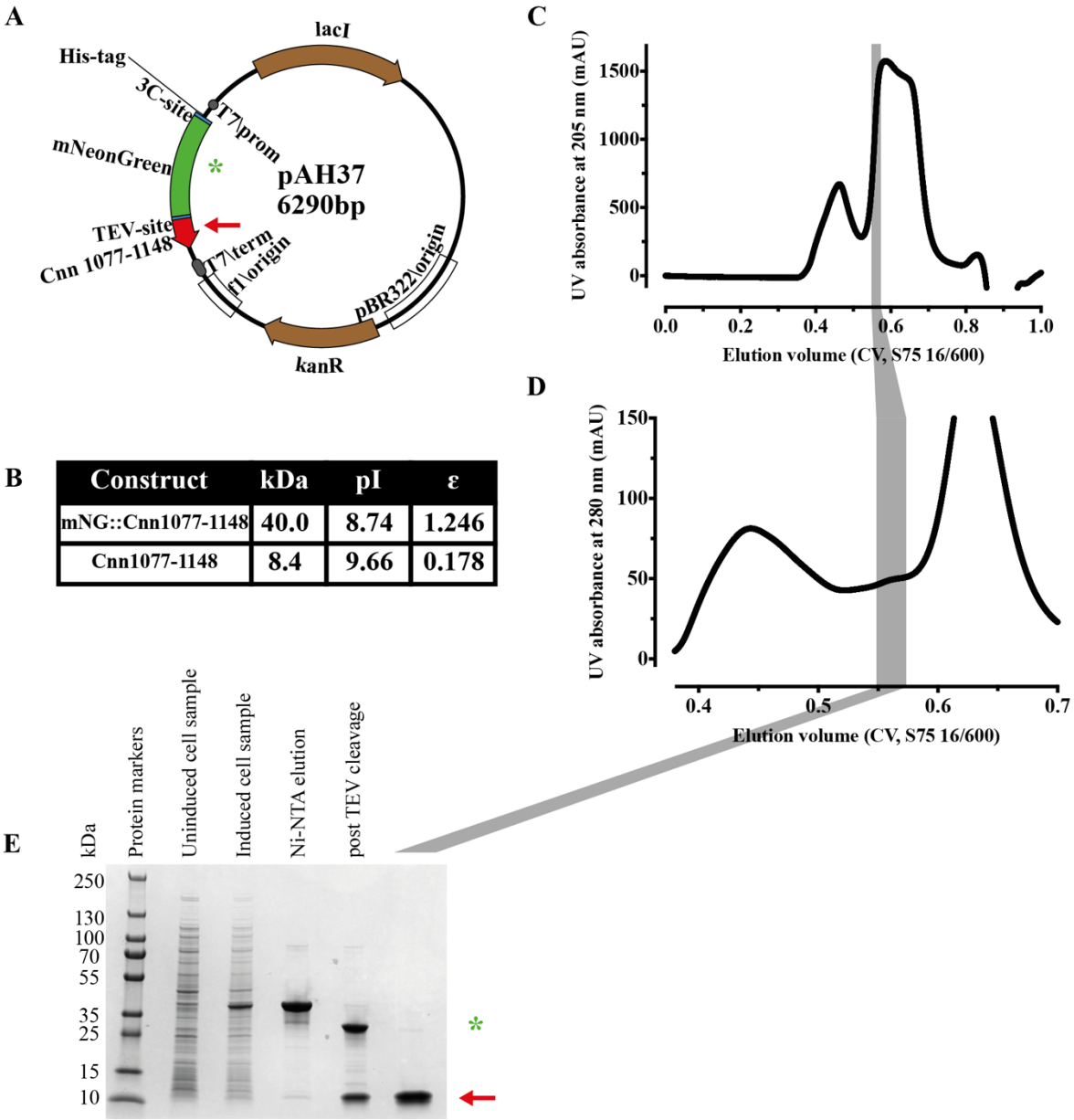


Figure 3.9 | Purification of Cnn$^{G1077-L1148}$. **A)** Overview of plasmid pAH37, which encodes Cnn$^{G1077-L1148}$ with an N-terminal, TEV-cleavable, His-3C-mNG-domain. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged protein and the protein after cleavage with TEV. **C)** SEC (S75 16/600) elution profile of Cnn$^{G1077-L1148}$ after Ni-affinity chromatography, with UV absorbance (205 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.55-0.57) indicates the Cnn$^{G1077-L1148}$-containing fractions pooled after SEC and run on the SDS-PAGE gel in E. The peak eluting immediately afterwards (around CV = 0.65) was determined to contain mNG as based on its green colour. SEC was carried out using standard TBS. **D)** The same SEC experiment as in (C), but with differently scaled x- and y-axes and plotting UV absorbance at 280 nm on the y-axis; note Cnn$^{G1077-L1148}$ barely absorbs at this wavelength (see (B)), hence the very small peak. E) 4-20 % reducing SDS-PAGE gel showing successive purification steps; the band corresponding to Cnn$^{G1077-L1148}$ is highlighted with a red arrow; the band corresponding to the mNG-tag is highlighted with a green asterisk.

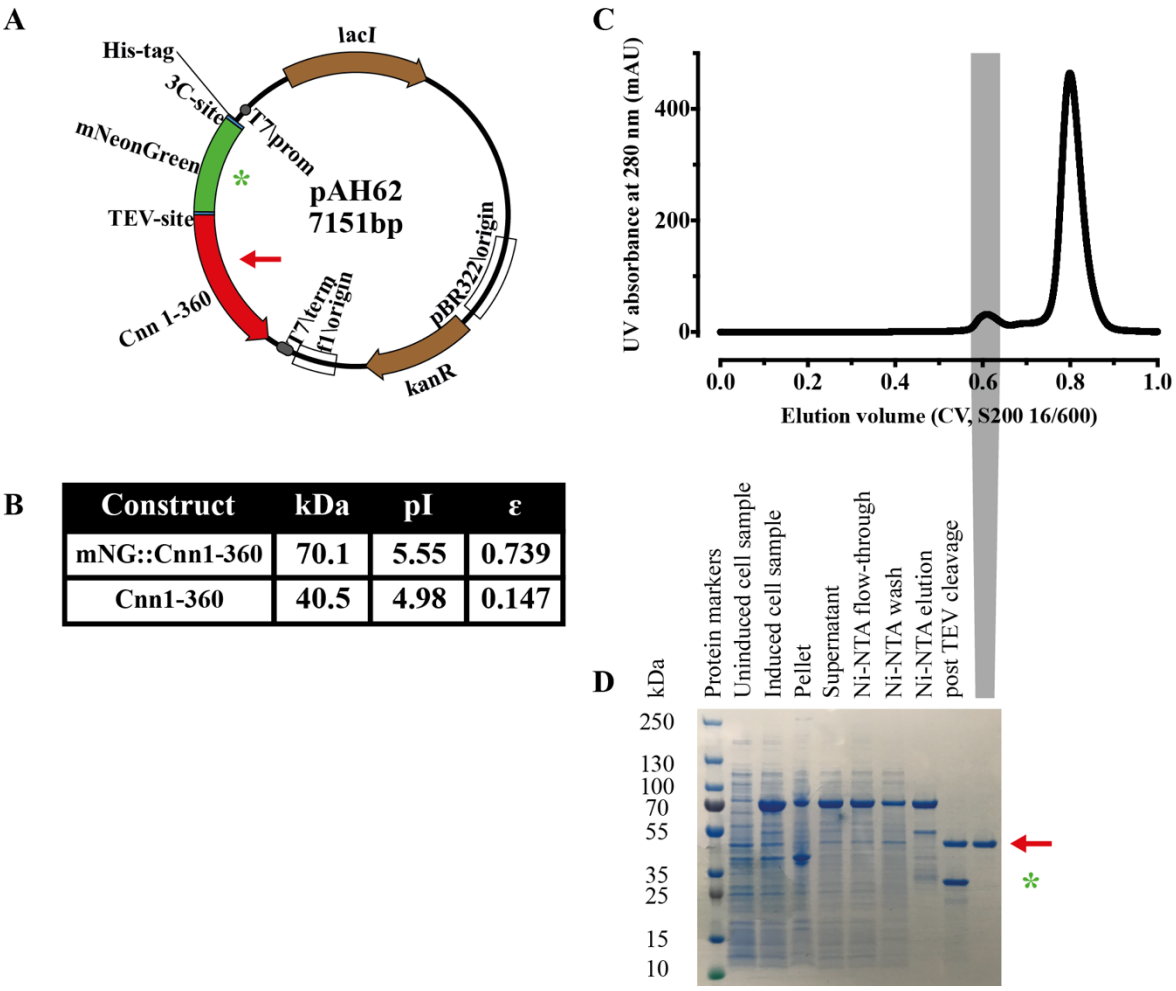
The original plasmid of this system, pAH37, was designed in a three-component NEBuilder reaction containing the original backbone vector (NcoI-cleaved pETM14), an insert encoding mNG with a C-terminal TEV-cleavage site and an insert encoding $Cnn^{G1077\text{-}L1148}$, i.e. a slightly longer version of the Cnn C-terminus (Fig. 3.9 A).

mNG-$Cnn^{G1077\text{-}L1148}$ was produced according to the usual protocol in *E. coli* BL21 (DE3) cells grown in 6 L of LB. As with purification of $Cnn^{H1082\text{-}L1148}$ described above, $ZnCl_2$ (20 µl from a 100 mM stock) was added to the resuspended cell pellet before homogenisation to allow efficient dimerisation. During Ni-affinity chromatography, a reduced volume of wash buffer (5 CV) was used to prevent excess sample loss during this step. After elution of mNG-$Cnn^{G1077\text{-}L1148}$ from the Ni-NTA-column, the solution containing the protein was diluted to 35 ml with buffer A in order to reduce the protein concentration and thereby avoid potential aggregation during the subsequent overnight dialysis. 1.5 ml TEV protease (1.25 mg/ml stock) were used for tag cleavage.

Three rounds of reverse Ni-affinity chromatography were carried out the next day, and the third flow-through concentrated to 5 ml and applied on a S75 16/600 column pre-equilibrated in standard TBS. Unfortunately, however, even such stringent reverse Ni-affinity chromatography appeared to result in some cleaved mNG-tag being retained in the flow-through. During size-exclusion chromatography, the elution profile was therefore dominated by the much larger, strongly absorbing mNG as opposed to $Cnn^{G1077\text{-}L1148}$ (Fig. 3.9). However, the peak containing mNG was easily identified by its green colour (CV = 0.65) and could therefore be avoided when pooling the region containing $Cnn^{G1077\text{-}L1148}$ (CV = 0.55-0.57), as confirmed by SDS-PAGE analysis (Fig. 3.9). $Cnn^{G1077\text{-}L1148}$ was concentrated to 12.5 mg/ml and flash-frozen in 30 µl aliquots, resulting in a yield of ~ 0.6

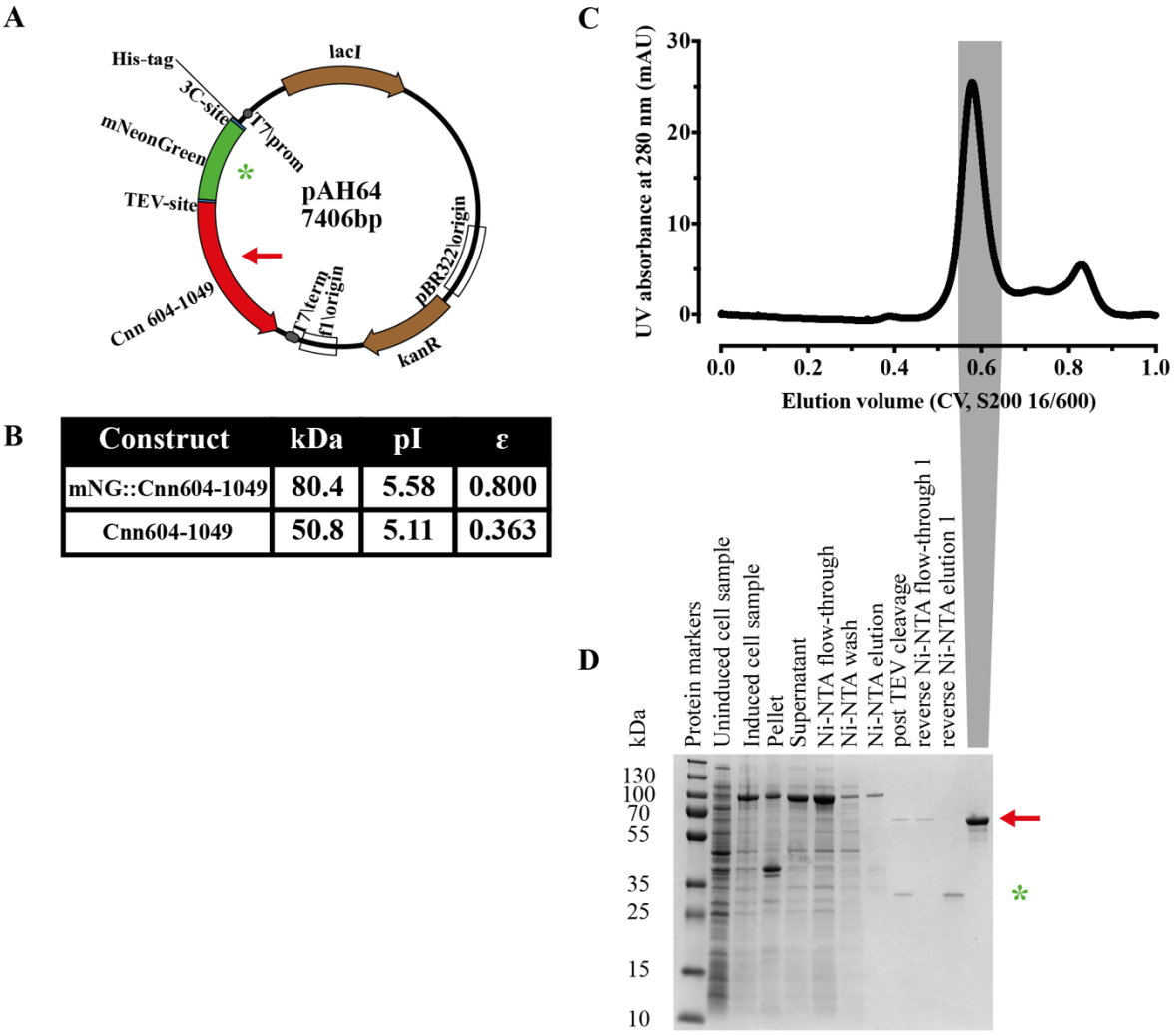
mg/L. As this yield was much lower than that of $Cnn^{H1082-L1148}$ purified via the pLip-mono system (2 mg/L, see above), it was decided that this novel mNG-tag system instead would be trialled for purification of larger constructs that would elute on a size-exclusion column well away from the residual mNG-peak.

A construct was designed covering the N-terminal region of Cnn, i.e. M1-K360 (Fig. 3.10 A). Purification occurred from *E. coli* BL21 (DE3) cells grown in 2 L of LB according to the usual protocol. During Ni-affinity chromatography, a 5 ml Ni-NTA column was used and the flow-through re-applied to the Ni-NTA column for 30 min to maximise binding before a more stringent wash of 15 CV buffer B was carried out. For tag cleavage, 1 mg TEV protease (1 mg/ml stock) was added. After dialysis overnight, no reverse Ni-affinity chromatography was carried out, and instead the sample was concentrated to 5 ml and directly applied to a S200 16/600 column pre-equilibrated in standard TBS. The SEC profile had two distinct peaks, one corresponding to the mNG-tag as determined by its green colour (CV = 0.8), the other corresponding to $Cnn^{M1-K360}$ (CV = 0.57-0.64) as based on SDS-PAGE analysis (Fig. 3.10). The $Cnn^{M1-K360}$-containing peak was pooled and concentrated to 10.7 mg/ml, then flash-frozen in 45 µl aliquots, resulting in a yield of ~ 3 mg/L.



Construct	kDa	pI	ε
mNG::Cnn1-360	70.1	5.55	0.739
Cnn1-360	40.5	4.98	0.147

Figure 3.10 | Purification of Cnn$^{M1-K360}$. A) Overview of plasmid pAH62, which encodes Cnn$^{M1-K360}$ with an N-terminal, TEV-cleavable, His-3C-mNG-domain. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged protein and the protein after cleavage with TEV. **C)** SEC (S200 16/600) elution profile of Cnn$^{M1-K360}$ after Ni-affinity chromatography, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.57-0.64) indicates the Cnn$^{M1-K360}$-containing fractions pooled after SEC and run on the SDS-PAGE gel in D. The later-eluting peak (around CV = 0.8) was determined to contain mNG as based on its green colour. SEC was carried out using standard TBS. **D)** 4-20 % reducing SDS-PAGE gel showing successive purification steps; the band corresponding to Cnn$^{M1-K360}$ is highlighted with a red arrow; the band corresponding to the mNG-tag is highlighted with a green asterisk.



Construct	kDa	pI	ε
mNG::Cnn604-1049	80.4	5.58	0.800
Cnn604-1049	50.8	5.11	0.363

Figure 3.11 | Purification of Cnn$^{G604-E1049}$. A) Overview of plasmid pAH64, which encodes Cnn$^{G604-E1049}$ with an N-terminal, TEV-cleavable, His-3C-mNG-domain. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged protein and the protein after cleavage with TEV. **C)** SEC (S200 16/600) elution profile of Cnn$^{G604-E1049}$ after Ni-affinity chromatography, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.55-0.65) indicates the Cnn$^{G604-E1049}$-containing fractions pooled after SEC and run on the SDS-PAGE gel in D. SEC was carried out using standard TBS. **D)** 4-20 % reducing SDS-PAGE gel showing successive purification steps; the band corresponding to Cnn$^{G604-E1049}$ is highlighted with a red arrow; the band corresponding to the mNG-tag is highlighted with a green asterisk.

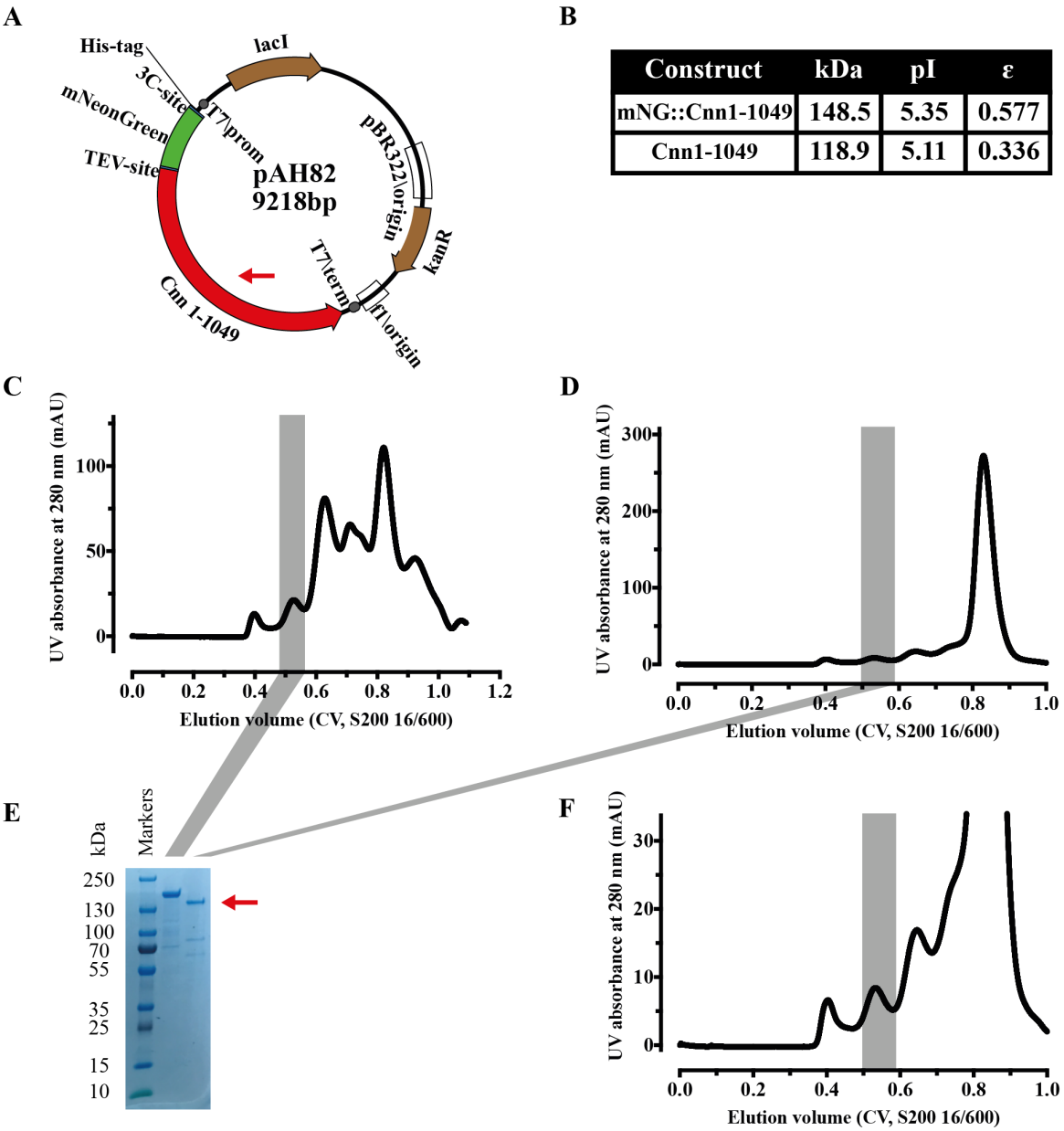
Furthermore, a construct was designed covering the large majority of the region between the central phosphorylated region (PReM) and the Cnn C-terminus (CM2), i.e. G604-E1049 (Fig. 3.11 A). This construct was purified from a 2 L LB culture of *E. coli* BL21 (DE3) cells. For Ni-affinity chromatography, the 5 ml Ni-NTA column was pre-equilibrated in buffer A supplemented with 10 mM imidazole to reduce non-specific binding, and the column was washed with 15 CV of buffer B. 0.5 mg TEV protease (1 mg/ml stock) were used for tag cleavage. The first flow-through after reverse Ni-affinity chromatography was collected and concentrated to 5 ml before loading on a S200 16/600 pre-equilibrated in standard TBS. The SEC profile had one clear peak (CV = 0.55-0.65), which was confirmed to contain $Cnn^{G604-E1049}$ via SDS-PAGE analysis (Fig. 3.11). This peak was therefore pooled and concentrated to 5.6 mg/ml before flash-freezing in 50 µl aliquots. The final yield was ~ 1 mg/L.

Finally, a construct was designed that covered almost the entire length of Cnn, i.e. M1-E1049 (Fig. 3.12 A). This construct only lacks the extreme C-terminus of the protein, i.e. CM2. Purification of this construct was carried out by Joseph Fisher (Part II student, Lea laboratory) under the direct supervision of the author. The protein was purified from *E. coli* BL21 (DE3) cells grown in 2 L of LB and using four 5 ml Ni-NTA columns during Ni-affinity purification according to the usual protocol. After elution from the Ni-NTA columns, the sample was split in half.

The first half was concentrated to 5 ml and applied to a S200 16/600 column pre-equilibrated in standard TBS. This resulted in a complex elution profile, probably due to proteolytic degradation of mNG-$Cnn^{M1-E1049}$ (Fig. 3.12 C). However, the earliest eluting peak after the void volume was green in colour and contained protein of a mass that roughly corresponds to mNG-$Cnn^{M1-E1049}$ as determined by SDS-PAGE analysis (Fig. 3.12 E). It was therefore

concluded that this peak contained mNG-Cnn$^{M1\text{-}E1049}$, and the peak was therefore pooled (CV = 0.48-0.56), concentrated to 2.3 mg/ml and flash-frozen in 40 µl. The yield was ~ 0.4 mg/L.

To the second half of the sample after Ni-affinity chromatography, TEV protease was added for tag cleavage (1 ml from 1 mg/ml stock), and the sample dialysed according to the usual protocol. After dialysis, the sample was also concentrated to 5 ml and applied to a S200 16/600 column pre-equilibrated in standard TBS. This resulted in an elution profile dominated by the strongly absorbing mNG-peak around CV = 0.83, identifiable by its green colour (Fig. 3.12 D). However, the earliest eluting peak after the void volume eluted at a position (CV = 0.50-0.59) that was slightly after where mNG-Cnn$^{M1\text{-}E1049}$ was observed to elute (Fig. 3.12 F), and SDS-PAGE analysis demonstrated that it predominantly contained a species with a mass perhaps 20-30 kDa lower than that of mNG-Cnn$^{M1\text{-}E1049}$ (Fig. 3.12 E). It was therefore concluded that this peak contained Cnn$^{M1\text{-}E1049}$, which was therefore concentrated to 3.2 mg/ml and flash-frozen in 30 µl aliquots, resulting in a yield of ~ 0.4 mg/L. Note how the yield for mNG-Cnn$^{M1\text{-}E1049}$ and Cnn$^{M1\text{-}E1049}$ were both roughly the same, suggesting little sample loss during tag cleavage and overnight dialysis.



Construct	kDa	pI	ε
mNG::Cnn1-1049	148.5	5.35	0.577
Cnn1-1049	118.9	5.11	0.336

Figure 3.12 | Purification of Cnn$^{M1-E1049}$. A) Overview of plasmid pAH82, which encodes Cnn$^{M1-E1049}$ with an N-terminal, TEV-cleavable, His-3C-mNG-domain. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged protein and the protein after cleavage with TEV. **C)** SEC (S200 16/600) elution profile of mNG-Cnn$^{M1-E1049}$ after Ni-affinity chromatography, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.48-0.56) indicates the fractions containing mNG-Cnn$^{M1-E1049}$ as judged by their green colour; these fractions were pooled after SEC and run on the SDS-PAGE gel in E. SEC was carried out using standard TBS. **D)** SEC (S200 16/600) elution profile of Cnn$^{M1-E1049}$ after cleavage with TEV, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.50-0.59) indicates the Cnn$^{M1-E1049}$-containing fractions pooled after SEC and run on the SDS-PAGE gel in E. SEC was carried out using standard TBS. **E)** 4-20 % reducing SDS-PAGE gel showing the final pooled fractions of mNG-Cnn$^{M1-E1049}$ and Cnn$^{M1-E1049}$ (indicated with a red arrow). **F)** The same elution profile as in D but with a differently scaled y-axis, thereby highlighting the pooled peak. The author cloned the plasmid and supervised the purification of the proteins by Joseph Fisher.

3.4.2 Functional characterisation of novel Cnn constructs

As MBP-Cnn$^{Q403-H608}$ and Cnn$^{H1082-L1148}$ had previously been shown to be dimeric in solution (ref [196] and Figs. 3.6 and 3.8 C), it remained to be determined whether Cnn$^{M1-K360}$ and Cnn$^{G604-E1049}$ are also dimeric in solution. SEC-MALS was therefore carried out on Cnn$^{M1-K360}$ and Cnn$^{G604-E1049}$ at a range of concentrations. Cnn$^{M1-K360}$ was tested at 20, 50 and 90 μM on a S200 10/300 pre-equilibrated in standard TBS. Molecular weights of 84.1 kDa (+/- 0.2 %), 84.6 kDa (+/- 0.3 %) and 85.8 kDa (+/- 0.2 %) were measured, respectively. These correspond closely to the expected mass for dimeric Cnn$^{M1-K360}$ (80.9 kDa), and Cnn$^{M1-K360}$ is therefore mostly dimeric under these conditions (Fig. 3.13 A). Cnn$^{G604-E1049}$ was tested at 55 and 110 μM under the same conditions, and molecular weights of 101.4 kDa (+/- 0.2 %) and 102.7 kDa (+/- 0.1 %) were measured, respectively, which also closely correspond to the mass expected for a dimer of Cnn$^{G604-E1049}$, i.e. 101.5 kDa (Fig. 3.13 B). Both constructs therefore appear to be mostly dimeric under the conditions tested. Indeed, under the supervision of the author, Joseph Fisher also tested the oligomeric state of Cnn$^{M1-E1049}$ at 20 μM in standard TBS, concluding that this construct also is mostly dimeric in solution (Fig. 3.14 C), with the measured molecular weight of 286.6 kDa (+/- 0.4 %) most closely corresponding to the theoretical mass of a dimer (237.7 kDa). The discrepancy between the measured and theoretical values might be due to experimental error or a slight tendency of Cnn$^{M1-E1049}$ to form oligomers larger than dimers.

As described above, Cnn$^{Q403-H608}$ forms micron-scale assemblies upon binding to the Cnn C-terminus. In order to investigate whether such network assembly is specific to Cnn$^{Q403-H608}$, or if e.g. Cnn$^{M1-K360}$ and Cnn$^{G604-E1049}$ could also oligomerise in the presence of the Cnn C-terminus, a binding experiment was conducted in which Cnn$^{M1-K360}$, MBP-Cnn$^{Q403-H608}$, and Cnn$^{G604-E1049}$ at roughly equimolar concentrations (respectively 30, 20 and 30 μM) were

mixed in the presence of a fivefold molar excess of Cnn$^{G1077\text{-}L1148}$ in standard TBS, and the resulting species separated via analytical SEC (Fig. 3.14 A and B). When all four constructs were mixed together, a novel peak was observed to elute in the void volume. SDS-PAGE analysis revealed that this peak only contained MBP-Cnn$^{Q403\text{-}H608}$ and Cnn$^{G1077\text{-}L1148}$, corresponding to the previously observed assemblies (Fig. 3.8). Strikingly, no other peak shift or co-elution of bands on the gel was observed, suggesting that under these conditions, the only interaction occurring between Cnn fragments is that of MBP-Cnn$^{Q403\text{-}H608}$ with the Cnn C-terminus. Cnn C-terminus-induced network assembly is therefore specific to the central MBP-Cnn$^{Q403\text{-}H608}$ fragment.

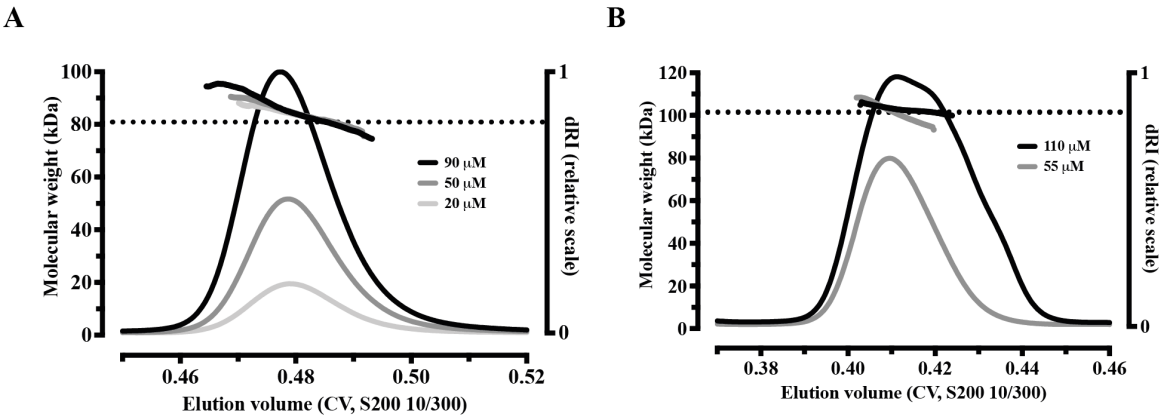


Figure 3.13 | SEC-MALS analysis showing that Cnn$^{M1\text{-}K360}$ and Cnn$^{G604\text{-}E1049}$ are both dimeric in solution. **A)** The average molecular weight in solution of Cnn$^{M1\text{-}K360}$ was measured at 20 µM (light grey: 84.1 kDa +/- 0.2 %), 50 µM (grey: 84.6 kDa +/- 0.3 %) and 90 µM (black: 85.8 kDa +/- 0.2 %). Dotted line indicates theoretical molecular weight expected for dimeric Cnn$^{M1\text{-}K360}$ (i.e. 80.9 kDa). **B)** The average molecular weight in solution of Cnn$^{G604\text{-}E1049}$ was measured at 55 µM (grey: 101.4 kDa +/- 0.2 %) and 110 µM (black: 102.7 kDa +/- 0.1 %). Dotted line indicates theoretical molecular weight expected for dimeric Cnn$^{G604\text{-}E1049}$ (i.e. 101.5 kDa). Molecular weight (in kDa) is plotted on the left y-axis, change in refractive index (dRI, relative scale) plotted on the right y-axis, column elution volume (CV, 24 ml) of column S200 10/300 plotted on the x-axis. Standard TBS was used as buffer.

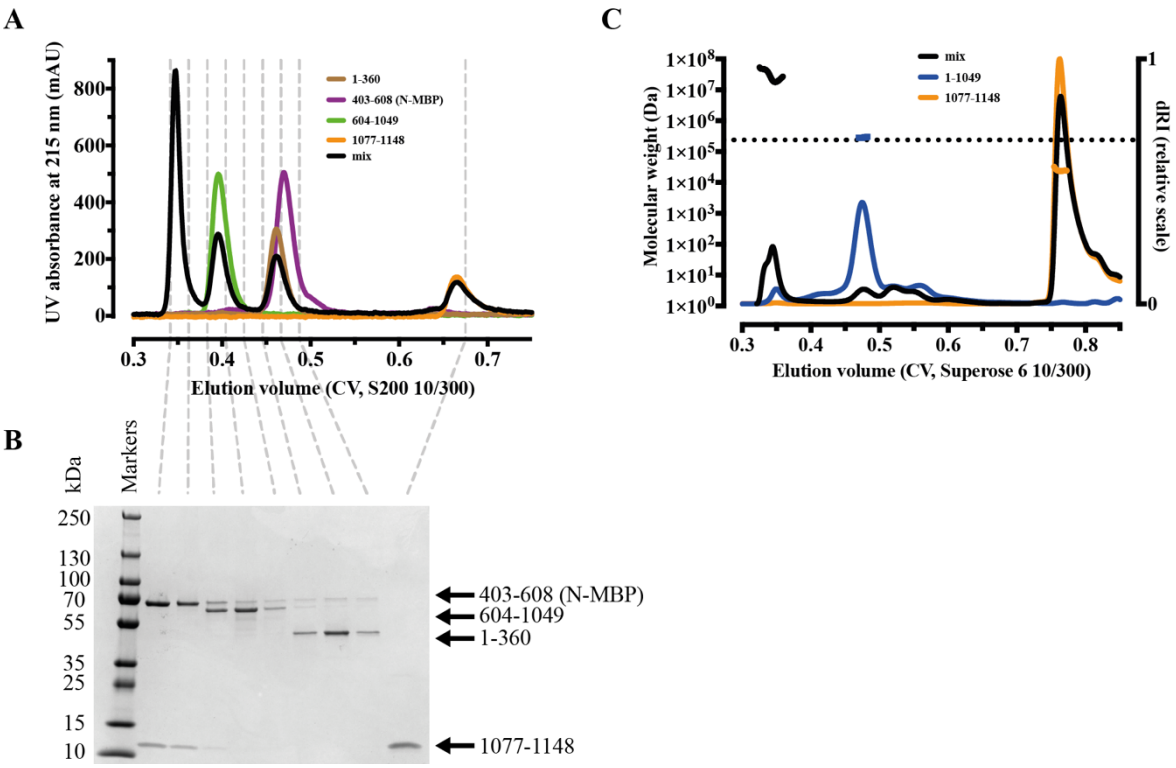


Figure 3.14 | Analytical SEC and SEC-MALS studies of constructs covering almost the entire length of Cnn. **A)** Analytical SEC experiment of Cnn$^{M1-K360}$ (brown, 30 µM), MBP-Cnn$^{Q403-H608}$ (purple, 20 µM), Cnn$^{G604-E1049}$ (green, 30 µM), Cnn$^{G1077-L1148}$ (orange, fivefold molar excess) and all constructs mixed (black, concentrations as before). Dashed grey lines indicate fractions taken and run on the SDS-PAGE gel in B; UV absorbance at 215 nm (in mAU) is plotted on the y-axis, column elution volume (CV, 24 ml) of column S200 10/300 plotted on the x-axis. Standard TBS was used as buffer. Note that a novel peak appears around CV = 0.35 (void peak) upon incubation of all constructs together, suggesting the formation of a large species. **B)** 4-20 % reducing SDS-PAGE gel showing fractions taken during the experiment with all constructs mixed (see A, black). The bands correspond to the constructs indicated. Note the void peak (first two fractions) contains only MBP-Cnn$^{Q403-H608}$ and Cnn$^{G1077-L1148}$. No other bands appear to co-elute, and the interaction between MBP-Cnn$^{Q403-H608}$ and Cnn$^{G1077-L1148}$ therefore appears to be the only one occurring under these conditions. **C)** SEC-MALS experiment of Cnn$^{M1-E1049}$ (blue, 20 µM), Cnn$^{G1077-L1148}$ (orange, tenfold molar excess) and both constructs mixed (black, concentrations as before). Standard TBS was used as buffer. Dotted line indicates theoretical molecular weight expected for dimeric Cnn$^{M1-E1049}$ (i.e. 237.7 kDa). Molecular weight (in Da, logarithmic scale) is plotted on the left y-axis, change in refractive index (dRI, relative scale) plotted on the right y-axis, column elution volume (CV, 24 ml) of column Superose 6 10/300 plotted on the x-axis. Standard TBS was used as buffer. Note that Cnn$^{M1-E1049}$ is predominantly dimeric under these conditions (measured molecular weight: 286.6 kDa +/- 0.4 %), but in the presence of Cnn$^{G1077-L1148}$ the peak corresponding to dimeric Cnn$^{M1-E1049}$ disappears and instead a void peak appears containing species several orders of magnitude larger. SEC-MALS samples in (C) were prepared by Joseph Fisher under the supervision of the author.

Finally, it remained to be determined whether network assembly is an artefact due to the relatively small size of the MBP-Cnn$^{Q403\text{-}H608}$ construct, which might have resulted in inherently unstable protein. To that end, Joseph Fisher carried out a SEC-MALS experiment under the direct supervision of the author, in which Cnn$^{M1\text{-}E1049}$ (at 20 μM) was mixed with a tenfold molar excess of Cnn$^{G1077\text{-}L1148}$. Note that these two constructs together cover ~ 98 % of Cnn, only missing 27 amino acids located in a region predicted to be largely disordered (Fig. 1.6). Strikingly, when the two constructs were mixed, the peak corresponding to dimeric Cnn$^{M1\text{-}E1049}$ disappeared, and a new peak eluting in the void volume was observed, containing species that were many orders of magnitude larger than dimeric Cnn$^{M1\text{-}E1049}$, suggesting that Cnn network assembly had occurred (Fig. 3.14 C).

3.5 Discussion

In this chapter, dimeric $Cnn^{M1\text{-}E1049}$ is shown to assemble into large assemblies in a manner that depends on the Cnn C-terminus (i.e. CM2). More specifically, the ability to assemble micron-scale networks *in vitro* is shown to reside within the region Q403-H608 of Cnn. Mutations that prevent the Cnn C-terminus from binding to the zipper-like motif found within this region also prevent $Cnn^{Q403\text{-}H608}$ from assembling into a network, suggesting that the interaction of the Cnn C-terminus with this zipper-like motif induces network assembly. However, the zipper-like motif itself ($Cnn^{D490\text{-}K544}$) does not assemble into networks but rather forms 2:2 complexes with the Cnn C-terminus. This suggests that the ability to bind the Cnn C-terminus resides within the zipper-like motif, but additional sequences within $Cnn^{Q403\text{-}H608}$ are required to allow network assembly after binding.

$Cnn^{Q403\text{-}H608}$ contains ten conserved Ser/Thr sites, and phosphorylation on some or all of these sites is thought to have an important role in Cnn network assembly *in vivo* [196]. As demonstrated in this chapter, however, it is possible to observe network assembly *in vitro* in the absence of any phosphorylation events. This is probably due to the relatively long timescales and high concentrations used in solution. Indeed, Dr Violet Feng has since shown that, in the presence of PLK1 kinase *in vitro*, the $Cnn^{Q403\text{-}H608}$-mGFP:$Cnn^{H1082\text{-}L1148}$ networks shown in Fig. 3.8 A form more efficiently [339].

The *in vitro* studies described in this chapter have since been further complemented by *in vivo* assays that have shown that disruption of the interaction between the Cnn C-terminus and the zipper-like motif via introduction of the mutants described above also perturbs Cnn network assembly in the *Drosophila* embryo in the context of the full-length protein. Furthermore, it has been shown that cytosolic Cnn networks, which form in unfertilised

Drosophila eggs upon the introduction of phospho-mimicking mutants of the ten potential phosphorylation sites (i.e. Cnn-10D/E) found within the centre of the protein, do not form when a mutant Cnn-10D/E construct is used that has a truncated C-terminus. These *in vivo* studies are therefore in line with a role of the Cnn C-terminus in network assembly [339]. Furthermore, using an antibody that recognises phosphorylated S567, collaborators have confirmed that Cnn is specifically phosphorylated on S567 at the centrosome [339], as was previously described via mass spectrometry [196]. Interestingly, the outer edges of the Cnn network were not recognised by this antibody, suggesting that Cnn is dephosphorylated on S567 towards the periphery of the centrosome, a process that might have a role in Cnn network disassembly [339]. Finally, these *in vivo* studies also confirmed that the Cnn C-terminus is involved in Cnn targeting to the centrosome [339], a function that had previously been shown for human CEP215 (ref [178]), and which explains why the testes isoform of Cnn, which has a different C-terminal region, is targeted to mitochondria rather than centrosomes [223].

Other groups have independently confirmed the interaction between the Cnn C-terminus and a central region within Cnn via NMR and large-scale yeast-two-hybrid studies [197, 272]. These NMR studies also suggest that Plp (amino acids 583-740) and the zipper-like motif interact with the same patch of residues on the Cnn C-terminus, and mutation of Cnn Leu1131 and Leu1137 to Ala also disrupted interaction of the Cnn C-terminus with the zipper-like motif as well as Plp in a yeast-two hybrid experiment [272]. Furthermore, GFP-labelled $Cnn^{E1064-L1148}$ appeared to be recruited around the centriole during interphase in cultured *Drosophila* cells before expanding into a wider area around the centriole during mitosis in the presence of full-length Cnn, possibly by integrating into the expanded Cnn network or by binding Plp (ref [272]).

SPD-5, the putative functional homologue of Cnn in *C. elegans*, was previously shown to form micron-scale assemblies, too. The formation of these networks is enhanced by PLK-1, probably via phosphorylation of SPD-5 on S653 and S658 (ref [236]). This chapter conclusively shows that *C. elegans* SPD-5 and *D. melanogaster* Cnn have a shared capacity to assemble into micron-scale networks. However, this chapter provides a more detailed analysis of the Cnn regions and interactions required for such assemblies to occur. Future work will have to show whether such regions can also be identified in *C. elegans* SPD-5. If the mechanism of network assembly is conserved between the two proteins, it is possible that the SPD-5 C-terminus might interact with the region around S653/S658 to allow network formation. Preliminary studies carried out by the author using purified *C. elegans* SPD-5 constructs were unable to confirm such an interaction (Appendix Figure 1), but further investigations using different construct boundaries, concentrations and experimental conditions will be necessary before definitive conclusions can be drawn.

Furthermore, in the presence of polyethylene glycol, a crowding agent, *C. elegans* SPD-5 collapses into spherical "condensates" that can recruit effector proteins of the pericentriolar material and even nucleate microtubules *in vitro* [33]. It therefore needs to be investigated whether *Drosophila* Cnn can also form such "condensates" and whether the networks described in this chapter also act as genuine scaffolds for the recruitment of client proteins. The point mutations characterised in this chapter *in vitro* and confirmed to disrupt centrosome assembly *in vivo* by collaborators [339] would be expected to also disrupt formation of these "condensates" if they do indeed relate to the biologically relevant assembly.

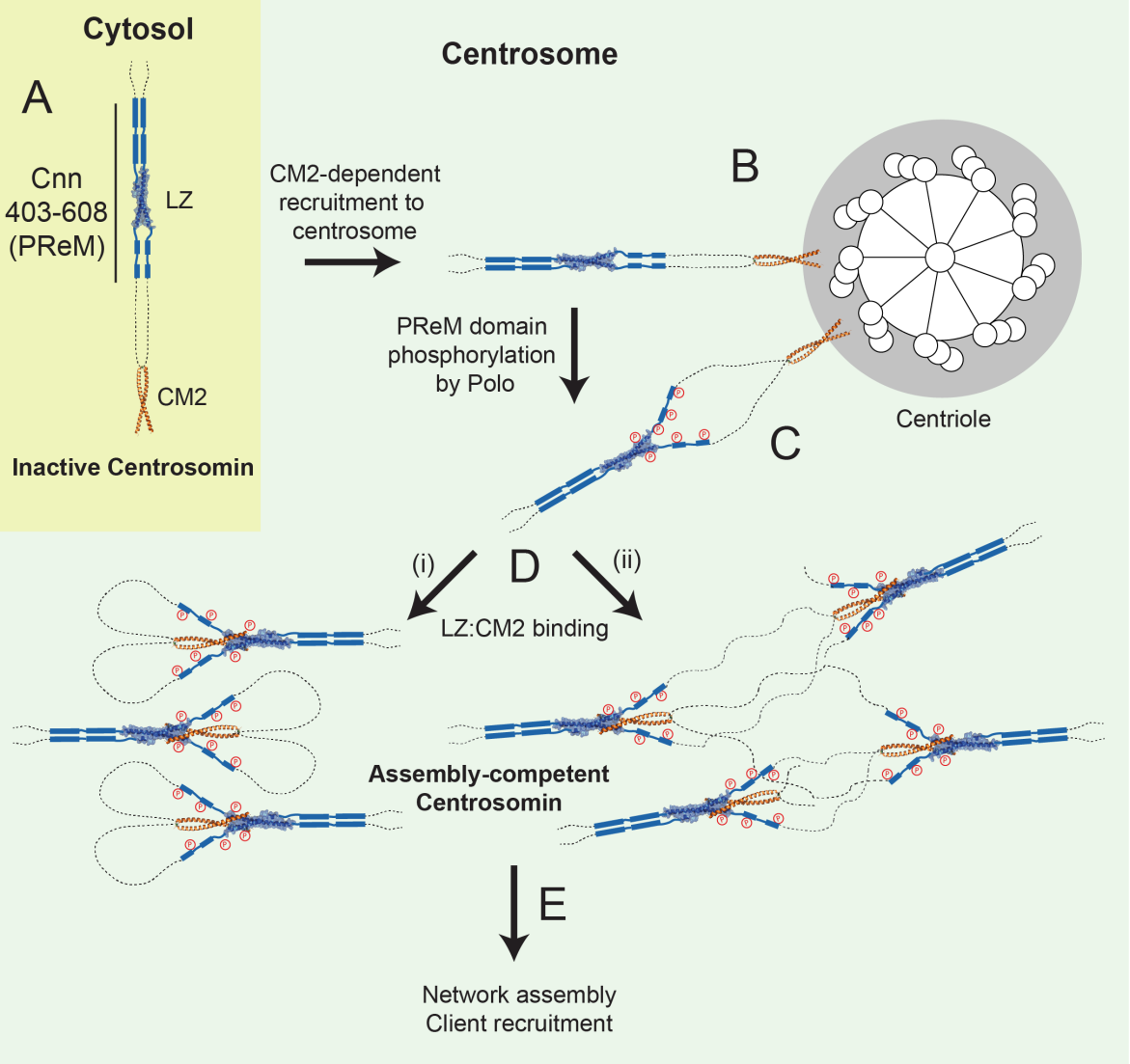


Figure 3.15 | Model illustrating Cnn network formation around the centriole. Models of CM2 ($Cnn^{H1082\text{-}L1148}$, Cnn C-terminus) and LZ ($Cnn^{D490\text{-}K544}$, zipper-like motif) are shown; PReM ($Cnn^{Q403\text{-}H608}$) is shown as blue boxes; remaining regions within Cnn are shown as dotted lines. **A)** Cytosolic Cnn is dimeric and not phosphorylated. **B)** Cnn is recruited to the centrosome in a manner that depends on its C-terminus. **C)** At the centrosome, Polo-mediated phosphorylation of Cnn destabilises the central region $Cnn^{Q403\text{-}H608}$, resulting in a conformational change in the vicinity of the zipper-like motif. **D)** The Cnn C-terminus can now bind the zipper-like motif, thereby converting Cnn into an assembly-competent state; note binding of the Cnn C-terminus could occur intra-molecularly (i) or inter-molecularly (ii) or both. **E)** Assembly-competent Cnn forms large-scale networks that can recruit client proteins, thereby forming the mature centrosome. Figure modified from [339] under Creative Commons licence CC-BY (see Acknowledgements). Not drawn to scale.

In summary, the data presented and discussed in this chapter allow for a molecular model of *Drosophila* Cnn network assembly around the centriole to be formulated (Fig. 3.15), which incorporates aspects of a previously described model for *C. elegans* SPD-5 network assembly (Fig. 1.9). In this model, Cnn is not phosphorylated in the cytosol on S567 and probably other key sites within $Cnn^{Q403\text{-}H608}$, and this disfavours binding of the Cnn C-terminus to the zipper-like motif. This state of Cnn can be described as "inactive" as it does not assemble networks unless e.g. phospho-mimicking residues are introduced (Cnn-10D/E mutant described above). Upon recruitment of Cnn to the centrosome via its C-terminus, however, Cnn is phosphorylated by Polo kinase on S567 and probably other residues. The model postulates that phosphorylation results in a conformational change that now favours binding of the Cnn C-terminus to the zipper-like motif. Upon formation of a complex between the two motifs, Cnn is in an "assembly-competent" state that can now form larger networks via interactions mediated by further regions within $Cnn^{Q403\text{-}H608}$ (Fig. 3.15).

This model provides a conceptual framework that allows further questions about the exact nature of Cnn network formation to be asked. For example, what interactions allow recruitment of Cnn to the centrosome via its C-terminus? What is the nature of the conformational change induced by phosphorylation? Does the Cnn C-terminus preferentially bind to the zipper-like motif located on the same dimer (intra-molecular) or a different dimer (inter-molecular)? What exactly are the $Cnn^{Q403\text{-}H608}$-mediated interactions that allow network formation after binding of the Cnn C-terminus to the zipper-like motif? Some of these questions will be addressed in the following chapters.

Chapter 4 - Calmodulin crosslinks Cnn$^{H1082\text{-}L1148}$ *in crystallo* but prevents network assembly in solution

4.1 Introduction

Calmodulin is a highly conserved calcium-binding protein which has an important role in a variety of cellular processes [250-253, 258]. In cultured fly cells, the protein is predominantly found at the spindle poles during mitosis, where it aids centrosome to spindle pole cohesion and helps focus spindles by acting together with Asp (refs [166, 167, 263-266]).

Calmodulin has also been shown via NMR to bind to the Cnn C-terminus in a calcium-sensitive manner [272]. This interaction is conserved in humans, where it depends on CEP215$^{V1861\text{-}A1870}$ (corresponding to fly Cnn$^{L1131\text{-}V1140}$, Fig. 4.1), thereby falling within the highly conserved CM2-like region of human CEP215 (ref [178]). The C-termini of both human CEP215 and *Drosophila* Cnn are important for centrosomal localisation of the corresponding proteins [178, 339]. However, at least in cultured human cells, centrosomal localisation does not depend on its interaction with Calmodulin (ref [178]). The precise function of this interaction therefore remains unclear.

Interestingly, Calmodulin can also interact with the microtubule-organising centre-targeting PACT motif found towards the C-terminus of the centrosomal vertebrate proteins AKAP450 and Pericentrin (refs [218, 283-285]) and the yeast spindle pole body protein Spc110 (refs [278-282]). It has been suggested that Calmodulin might act as a chaperone by binding to PACT in the cytosol [218]. Another model suggests that Calmodulin might confer mechanical stability to the yeast spindle pole body by crosslinking neighbouring Spc110 molecules [276, 277]. Indeed, the structure of the complex between a *Saccharomyces cerevisiae* Spc110 peptide (amino

acids 891-944) and *Kluyveromyces lactis* Calmodulin has been solved (Fig. 4.2) and *in crystallo*, two molecules of Calmodulin interlock and crosslink two monomeric Spc110 peptides [276].

Using the Calmodulin target database [341], two further, previously unreported Calmodulin binding sites can be predicted towards the N-terminus of Cnn, within CM1. The first site overlaps with the conserved predicted leucine zipper motif (Fig. 4.3). There is strong evidence suggesting that this region has a crucial role in γ-tubulin recruitment to the pericentriolar material in flies and vertebrates [179, 222, 224, 225]. This is important, as microtubules are known to be nucleated within the pericentriolar material [32], and this probably occurs by the so called γ-tubulin ring complex (γ-TuRC), i.e. assemblies of γ-tubulin and other accessory proteins [35-45]. It is unclear how Calmodulin might have a role in this.

Interestingly, though, Cnn is predicted to contain a further Calmodulin binding site within CM1, in the region immediately C-terminal to this predicted leucine zipper motif. This area is also relatively well conserved (Fig. 4.3), but less is known about its function. $CEP215^{K101-S200}$ (equivalent to $Cnn^{K142-A260}$) cannot pull down γ-tubulin from cultured human cells [224], but phosphorylation of CEP215 on Ser140 is thought to enhance the association of CEP215 with γ-tubulin in cultured human cells [226]. However, sequence alignment with fly Cnn is poor around S140, and indeed only fly Cnn is predicted to bind Calmodulin within CM1 (Fig. 4.3). These observations could be indicative of functional divergence between the two proteins in this less conserved C-terminal region of CM1, and it would therefore be interesting to study this region further.

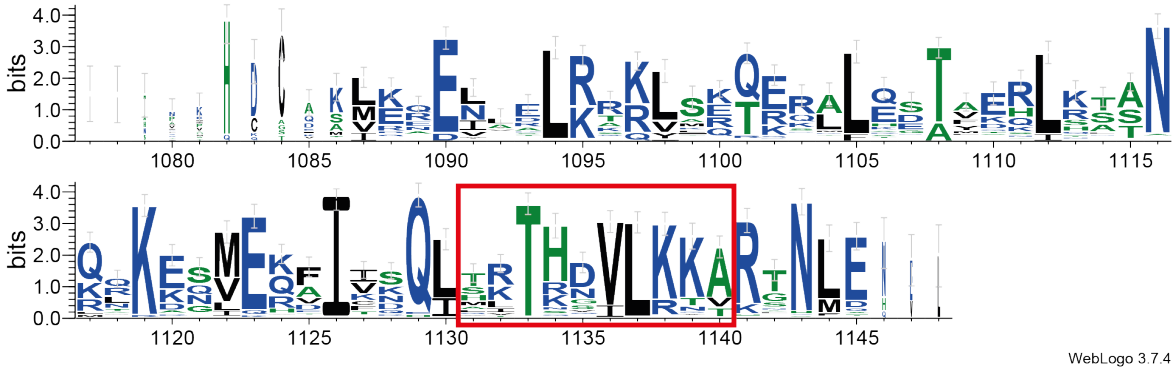


Figure 4.1 | Conservation of the Cnn C-terminus. Conservation of Cnn$^{G1077\text{-}L1148}$ is shown in WebLogo format with hydrophilic amino acids in blue, neutral amino acids in green, and hydrophobic amino acids in black. The Calmodulin binding site identified experimentally in CEP215 (see main body) is shown in red. The highly conserved region at the C-terminus of Cnn is also referred to as CM2. For multiple-sequence alignment, see Appendix. Sequence logo created using WebLogo [342].

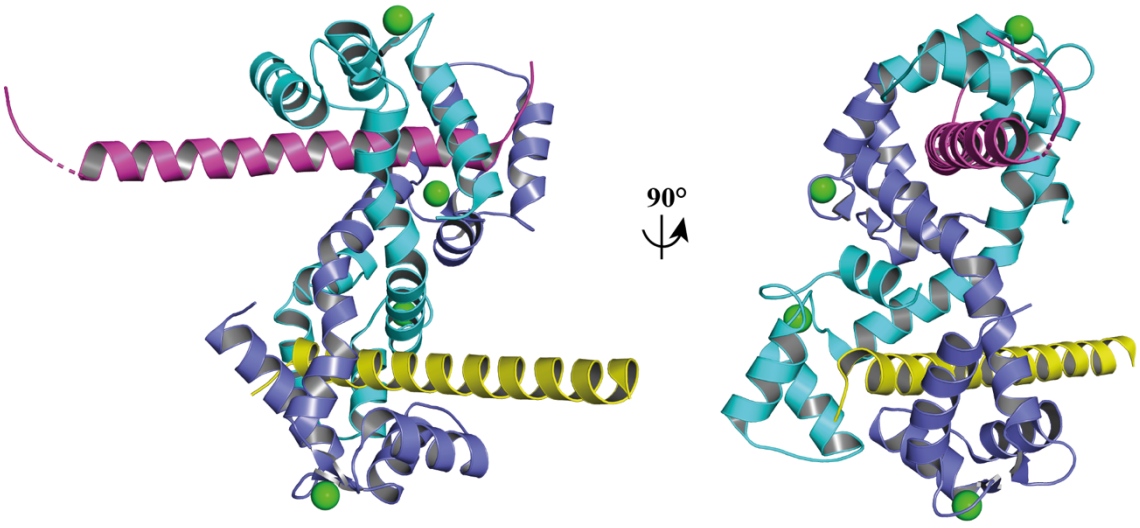


Figure 4.2 | Complex structure of yeast Calmodulin and Spc110 (amino acids 891-944). Cartoon model generated via PyMOL based on the structure (PDB code: 4DS7) of the complex between Calmodulin and Spc110 (amino acids 891-944). The two Calmodulin molecules are shown in blue and turquoise; the two Spc110 fragments are shown in pink and yellow. Calcium ions are shown as green spheres; note each Calmodulin only binds two ions in this crystal. Note Calmodulin crosslinks two Spc110 fragments [276].

```
Cnn       98 RSVRELEEQMSALRKENFNLKLRIYFLEEGQPGARADSSTESLSKQLIDAKIEIA
             |:::::.|.|::.|:|||||||||||||||| :........||.:.|..|:.|:|:.
CEP215    58 RNMKDFENQITELKKENFNLKLRIYFLEE-RMQQEFHGPTEHIYKTNIELKVEVE

Cnn      153 TLRKTVDVKMELLKDAARAISHHEELQRKADIDSQAIIDELQEQIHAYQMAESGG
             :|:::.:..:.:||..|::|:   |.|.......:.|.:.::.::::
CEP215   112 SLKRELQEREQLLIKASKAV---ESLAEAGGSEIQRVKEDARKKV----------

Cnn      208 QPVENIAKTRKMLRLESEVQRLEEELVNIEARNVAARNELEFMLAERLESLTA
             |.||::. |:::|.||.:|...:.||     ....|..|.|..|..||||..:
Cep215   154 QQVEDLL-TKRILLLEKDVTAAQAEL-----EKAFAGTETEKALRLRLESKLS
```

Figure 4.3 | Alignment of *D. melanogaster* Cnn$^{R98\text{-}A260}$ and *H. sapiens* CEP215$^{R58\text{-}S200}$. The two sequences show 31.3 % identity, 54.0 % similarity and 12.3 % gaps. CM1 (ref [179]) is shown in red. The highly conserved previously predicted leucine zipper motif (ref [181]) is highlighted in bold. The Calmodulin binding sites predicted by the author are highlighted in grey. Note the corresponding regions in CEP215 are not predicted to bind Calmodulin. Pairwise alignment was performed by the author using EMBOSS Needle [343], Calmodulin target database was used by the author for binding site prediction [341].

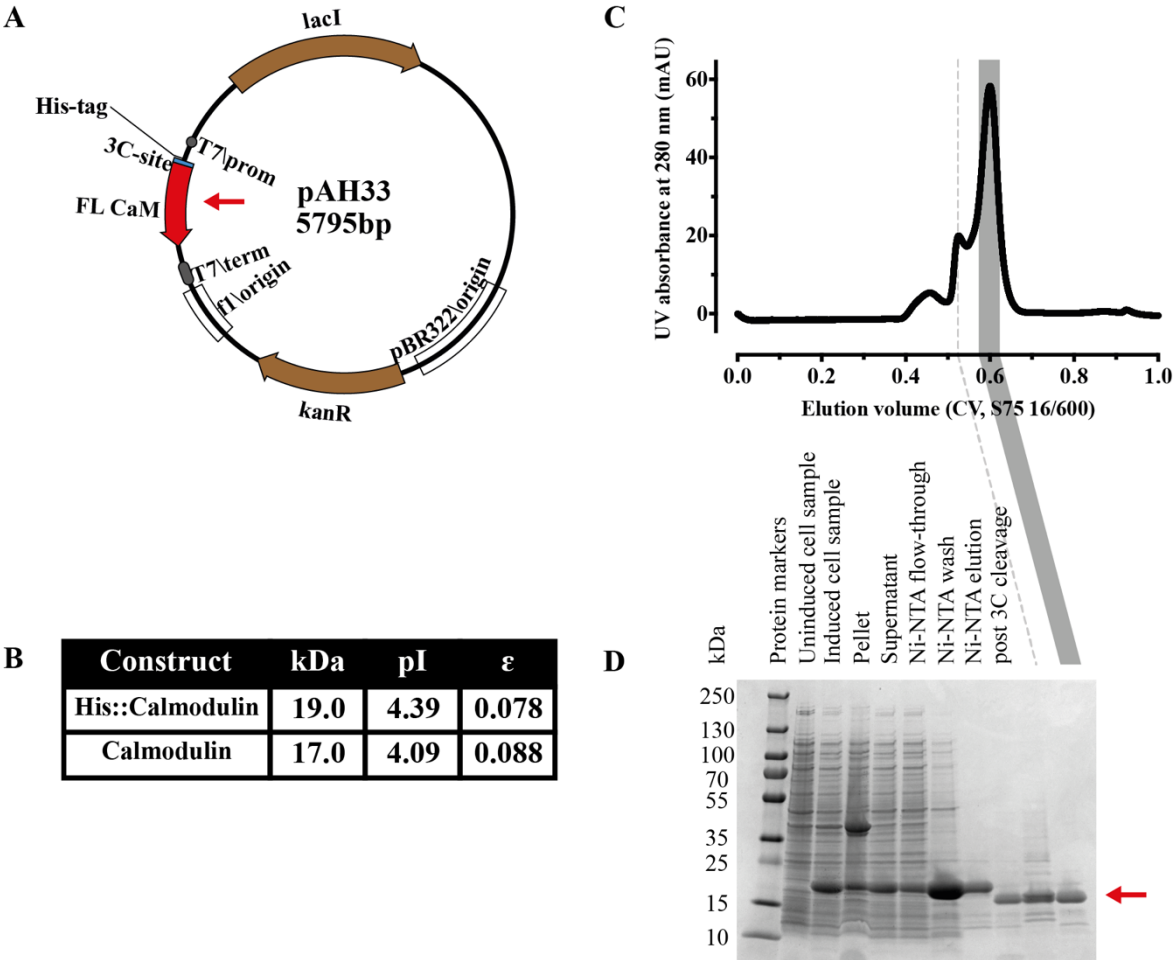
This chapter aims to investigate the partly conserved C-terminal region of CM1 from a structural perspective. Furthermore, this chapter aims to investigate whether binding of Calmodulin to the Cnn C-terminus could have a role in chaperoning or crosslinking Cnn, two competing models that have been suggested in yeast.

It will be shown that Cnn$^{S136\text{-}S205}$ does indeed bind Calmodulin *in vitro*, and that it forms a dimeric coiled-coil *in crystallo* that shows a striking similarity with the conformation of Cnn$^{H1082\text{-}L1148}$ in the previously known structure of the Cnn$^{H1082\text{-}L1148}$:Cnn$^{D490\text{-}K544}$ complex. Furthermore, it will be shown that Calmodulin crosslinks dimers of Cnn$^{H1082\text{-}L1148}$ *in crystallo*, resulting in the formation of an interlocking crystallographic lattice. Finally, it will also be shown that, surprisingly, Calmodulin can prevent network formation of Cnn constructs *in vitro* via formation of a tertiary complex between Calmodulin, Cnn$^{D490\text{-}H608}$ and Cnn$^{G1077\text{-}L1148}$.

4.2 Purification of Calmodulin

The full-length sequence of *Drosophila melanogaster* Calmodulin was cloned into the vector pETM14, allowing production of N-terminally His_6-tagged protein (Fig. 4.4 A). This construct was purified after overexpression in *E. coli* BL21 (DE3) cells grown in 3 L of LB according to the usual protocol. Note, however, that due to the lack of cysteine residues in Calmodulin, purification was carried out in the absence of any reducing agents. Furthermore, Ni-affinity chromatography was carried out with a reduced column volume of CV = 5 ml and with the presence of 10 mM imidazole in the loading buffer in order to reduce non-specific binding. The column was washed with 15 CV buffer B before elution of His_6-tagged Calmodulin.

1 mg of 3C protease (1 mg/ml stock) was added to the eluate before overnight dialysis. The dialysed sample was directly concentrated to 5 ml and applied to a S75 16/600 column pre-equilibrated in standard TBS lacking TCEP. The elution profile showed one clear peak as well as a slightly earlier eluting peak with reduced absorbance (Fig. 4.4 C). SDS-PAGE analysis confirmed that the main peak contained Calmodulin (Fig. 4.4 D). The peak (CV = 0.57-0.62) was therefore pooled, concentrated to 13.6 mg/ml and aliquoted into 55 µl fractions before flash-freezing. The final yield was ~ 4.7 mg/L.



Construct	kDa	pI	ε
His::Calmodulin	19.0	4.39	0.078
Calmodulin	17.0	4.09	0.088

Figure 4.4 | Purification of Calmodulin. A) Overview of plasmid pAH33, which encodes full-length Calmodulin (FL CaM) with an N-terminal, 3C-cleavable, His_6-tag. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged protein and the protein after cleavage with 3C. **C)** SEC (S75 16/600) elution profile of Calmodulin after Ni-affinity chromatography, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.57-0.62) indicates the fractions pooled after SEC and run on the SDS-PAGE gel in D, containing Calmodulin. A further fraction run on the gel is shown with a dashed line. SEC was carried out using standard TBS lacking reducing agent. **D)** 4-20 % non-reducing SDS-PAGE gel showing successive purification steps; the band corresponding to Calmodulin is highlighted with a red arrow.

4.3 Cnn$^{S136\text{-}S205}$ interacts with Calmodulin *in vitro*

4.3.1 Cnn$^{S136\text{-}S205}$ purification

The pLip vector has proven useful as a tool for the purification of short peptides in the lab. It was therefore decided to also use this system to purify Cnn$^{S136\text{-}S205}$, which contains the putative binding site of Calmodulin within the C-terminal region of CM1. However, the presence of both an N-terminal and a C-terminal TEV-cleavage site leads to a substantial overhang of six amino acids (ENLYFQ|G, where | is the TEV-cleavage site) at the C-terminus of the protein of interest after protease cleavage. This could interfere with crystallisation of the construct. A plasmid was therefore designed (pAH28-6) using the pLip-mono vector as a backbone, which had previously been used for the purification of the Cnn C-terminus (Fig. 3.2). In this vector, a STOP codon is inserted after the protein of interest (in this case Cnn$^{S136\text{-}S205}$), resulting in a construct that lacks the C-terminal tag and therefore has fewer exogenous amino acids (Fig. 4.5 A and B).

Protein expression was carried out using *E. coli* C41 (DE3) cells grown in 6 L of LB for 5-7 hours after IPTG induction at 20°C. After harvesting, cells were frozen at -20°C. Cells were homogenised and Ni-affinity chromatography was carried out using a column volume of 5 ml. The column was pre-equilibrated in buffer A supplemented with 10 mM imidazole, and washed with 15 CV of buffer B after sample application. After elution of Lip-Cnn$^{S136\text{-}S205}$ from the column, 1 mg of TEV protease (1 ml from a 1 mg/ml stock) was added and the solution dialysed overnight. The next day, reverse Ni-affinity chromatography (CV = 5 ml) was carried out twice, and the second flow-through was concentrated and further purified using SEC, with the column pre-equilibrated in standard TBS lacking reducing agent.

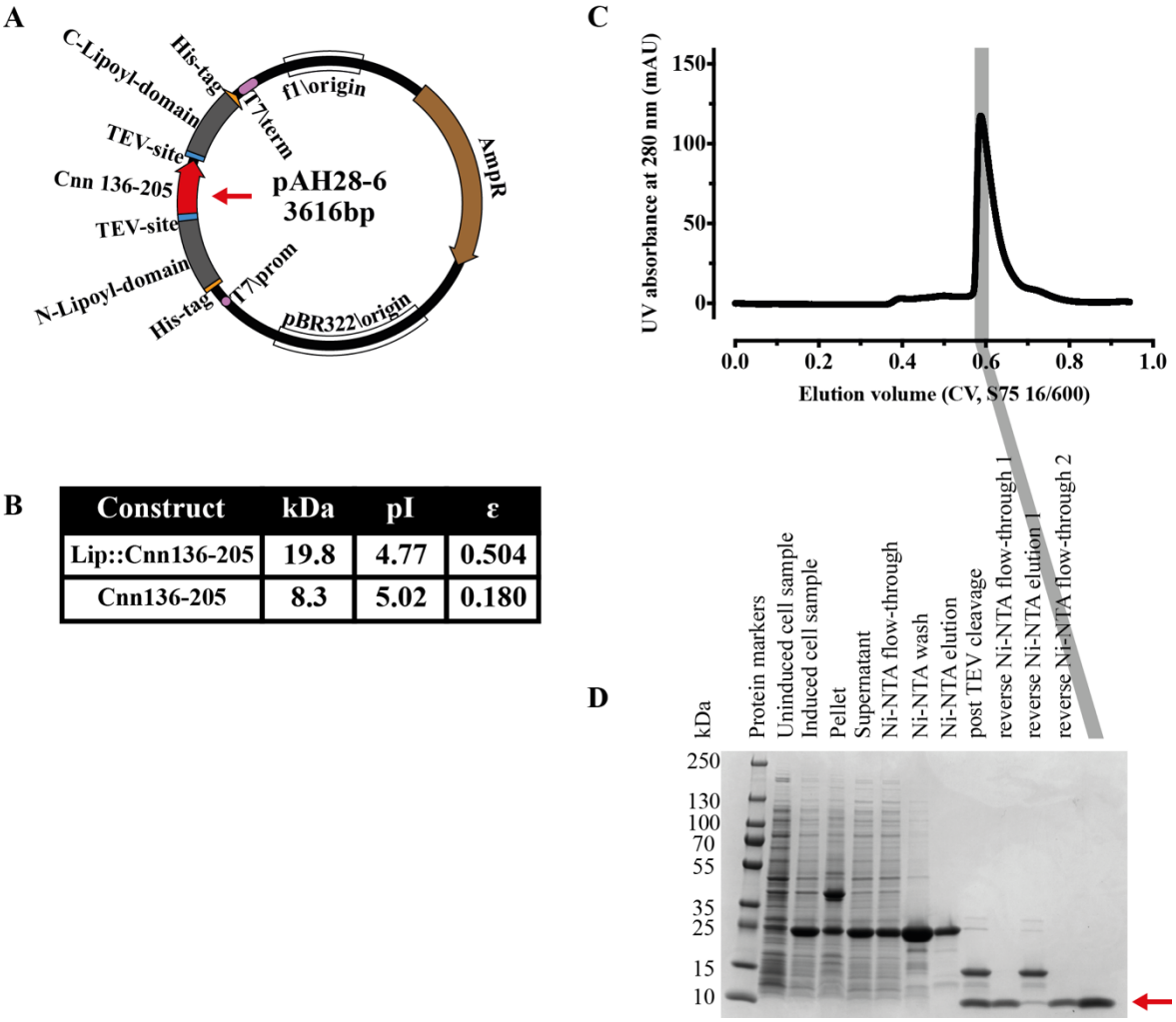


Figure 4.5 | Purification of Cnn$^{S136-S205}$. A) Overview of plasmid pAH28-6, which encodes Cnn$^{S136-S205}$ with an N-terminal, TEV-cleavable, His-Lipoyl-domain. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged protein and the protein after cleavage with TEV. **C)** SEC (S75 16/600) elution profile of Cnn$^{S136-S205}$ after Ni-affinity chromatography, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.57-0.61) indicates the fractions pooled after SEC and run on the SDS-PAGE gel in D, containing Cnn$^{S136-S205}$. SEC was carried out using standard TBS. **D)** 4-20 % reducing SDS-PAGE gel showing successive purification steps; the band corresponding to Cnn$^{S136-S205}$ is highlighted with a red arrow.

The SEC profile showed one single peak with a pronounced tail (Fig. 4.5 C). The peak (excluding the tail) was pooled (i.e. CV = 0.57-0.61) and concentrated to ~ 10 mg/ml. A gel was run, showing that the pooled peak only contained $Cnn^{S136-S205}$ and no contaminants (Fig. 4.5 D). $Cnn^{S136-S205}$ was therefore aliquoted into 50 µl fractions and flash-frozen. The final yield was ~ 1.8 mg/L.

4.3.2 Functional analysis of $Cnn^{S136-S205}$

In order to investigate the oligomeric state of $Cnn^{S136-S205}$ in solution, SEC-MALS was carried out with the help of Dr David Staunton (Molecular Biophysics Suite, Department of Biochemistry, University of Oxford) using a S200 10/300 column pre-equilibrated in standard TBS. $Cnn^{S136-S205}$ was tested at three different concentrations (Fig. 4.6). Dilution to the appropriate concentration used standard TBS. A slight shift in elution position towards an earlier-eluting peak and a minor increase in the average molecular weight measured was observed upon increasing the concentration from 100µM to 200 µM and finally 400 µM. Specifically, the average molecular weight increased from 8.4 kDa to 8.7 kDa and finally 8.9 kDa (Fig. 4.6). As the theoretical monomeric molecular weight of $Cnn^{S136-S205}$ is 8.3 kDa, this suggests that the construct is mostly monomeric in solution under these conditions. This was surprising, considering that $Cnn^{M1-K360}$, which contains this motif, is clearly dimeric (Fig. 3.13 A). Indeed, there appears to be a very slight tendency to oligomerise upon increasing the concentration of $Cnn^{S136-S205}$. Such concentration-dependent oligomerisation could also explain the pronounced tail observed during SEC purification (Fig. 4.5 C).

A further SEC-MALS experiment was carried out in which Calmodulin was mixed with $Cnn^{S136-S205}$ using standard TBS as buffer supplemented with an excess of $CaCl_2$, as the predicted interaction between the two constructs might be calcium-dependent. The control

experiments confirmed that Calmodulin (at 200 μM) and $Cnn^{S136\text{-}S205}$ (300 μM) were mostly monomeric under the conditions tested, with measured molecular weights in solution of 17.6 kDa (+/- 1.2 %) and 10.4 kDa (+/- 2.4 %), respectively. However, when both constructs were mixed together under the same conditions, a novel, early-eluting peak appeared that contained species with an average mass of 20.5 kDa +/- 1.2 % (Fig. 4.7 A). This value is below that expected for a 1:1 complex of Calmodulin and $Cnn^{S136\text{-}S205}$ with a theoretical mass of 25.3 kDa. As MALS reports the average mass of species in a given peak, transient interactions that are not fully occupied can appear to be lower in mass than expected. The observed peak with an average mass of 20.5 kDa +/- 1.2 % therefore probably does correspond to Calmodulin binding $Cnn^{S136\text{-}S205}$, but the interaction appears to be weak under the conditions tested, resulting in complexes with an apparent mass slightly below that of a 1:1 complex.

In order to test whether Calmodulin could also bind this region within Cnn in the context of a larger construct, an analytical SEC experiment was carried out using the same buffer as before, but this time using $Cnn^{M1\text{-}K360}$ as a probe for Calmodulin interaction. Surprisingly, when this Cnn construct (at 50 μM) was mixed with Calmodulin (at 180 μM), no significant peak shifts could be observed upon separation of the species on a S200 10/300 column, and SDS-PAGE analysis confirmed that the two constructs did not co-elute, suggesting that no binding occurred under the conditions tested (Fig. 4.7 B and C). Taken together, these experiments suggest that Calmodulin can weakly interact with monomeric $Cnn^{S136\text{-}S205}$ in solution, but not when this motif is found within the context of the larger, dimeric $Cnn^{M1\text{-}K360}$ construct under the conditions tested. This might be due to dimerisation burying the binding site or sterically hindering Calmodulin binding.

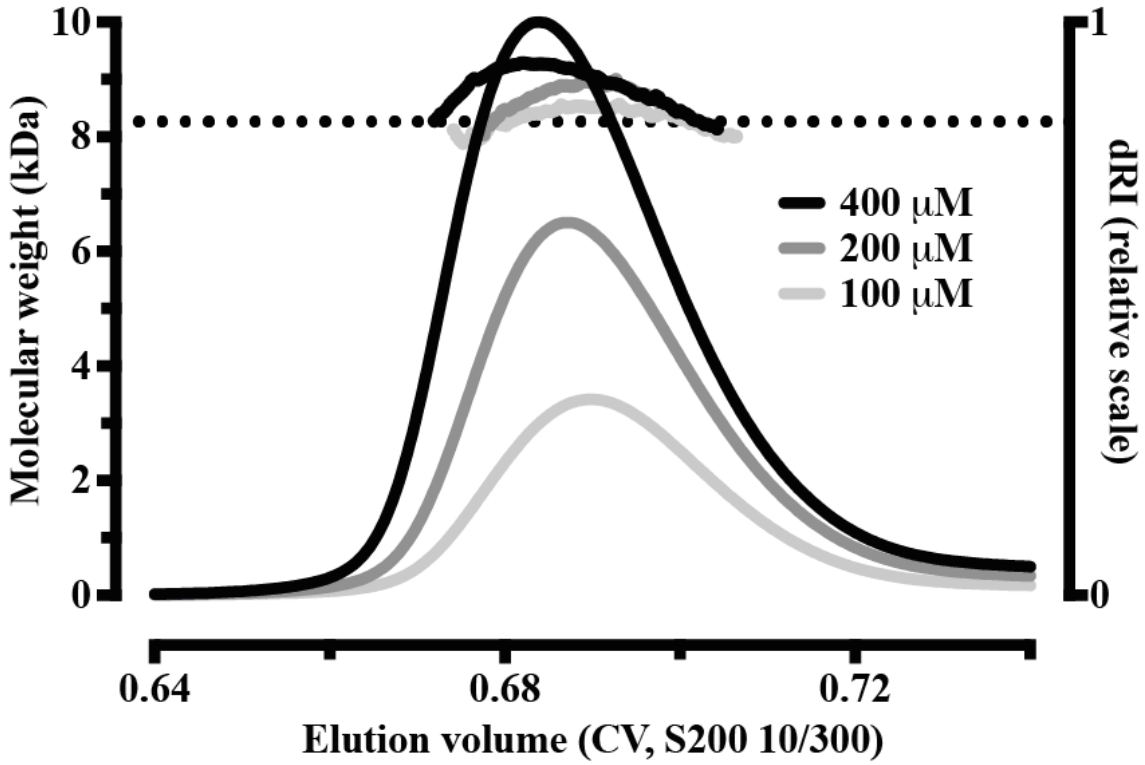


Figure 4.6 | SEC-MALS analysis of Cnn$^{S136-S205}$. The average molecular weight in solution of Cnn$^{S136-S205}$ was measured at 100 µM (light grey: 8.4 kDa +/- 0.7 %), 200 µM (grey: 8.7 kDa +/- 0.8 %) and 400 µM (black: 8.9 kDa +/- 0.8 %). Dotted line indicates theoretical molecular weight expected for monomeric Cnn$^{S136-S205}$ (i.e. 8.3 kDa). Note that there is a slight tendency towards an earlier-eluting peak and a slightly increased molecular weight upon increasing the concentration, but overall the construct is largely monomeric within this concentration range. Molecular weight (in kDa) is plotted on the left y-axis, change in refractive index (dRI, relative scale) plotted on the right y-axis, column elution volume (CV, 24 ml) of column S200 10/300 plotted on the x-axis. Standard TBS was used as buffer. Experiments were carried out with the help of Dr David Staunton.

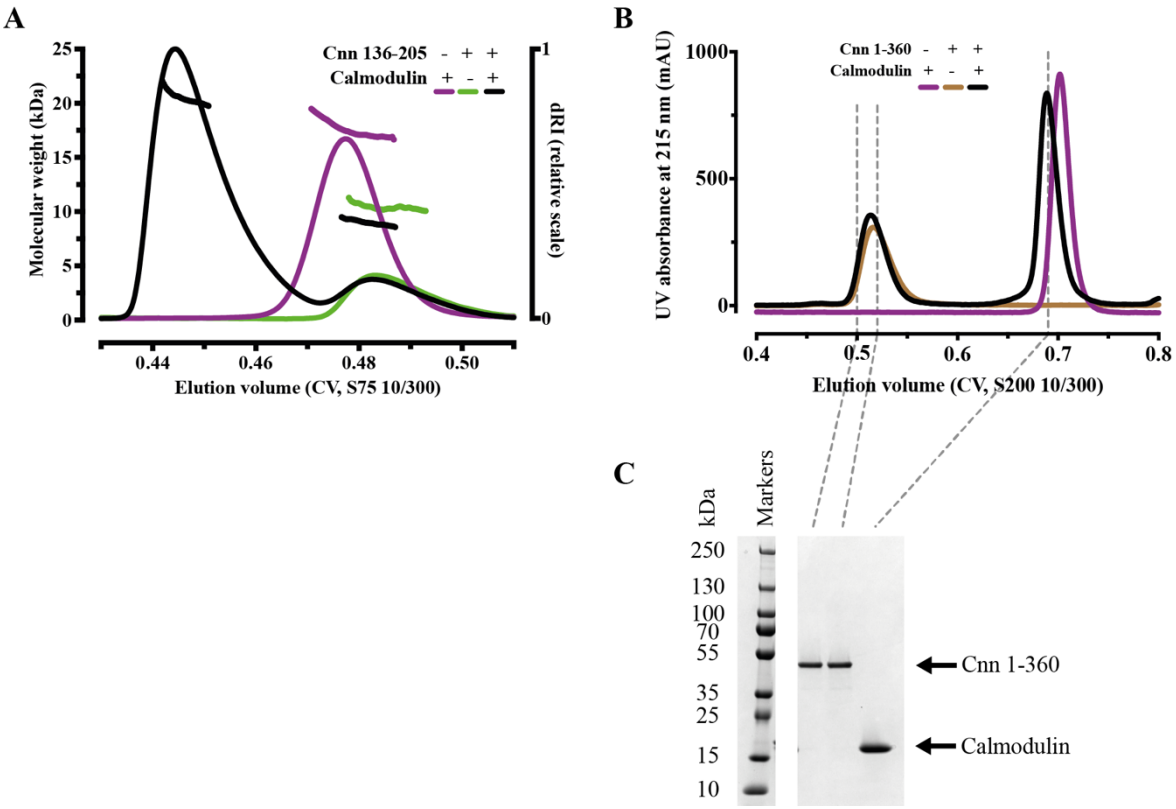


Figure 4.7 | Calmodulin can bind Cnn$^{S136-S205}$ *in vitro* but not in the context of a larger Cnn fragment. A) SEC-MALS analysis of binding interaction between Cnn$^{S136-S205}$ and Calmodulin. The average molecular weight in solution of Calmodulin at 200 µM (purple: 17.6 kDa +/- 1.2 %), Cnn$^{S136-S205}$ at 300 µM (green: 10.4 kDa +/- 2.4 %), and both mixed at these concentrations (black: 20.5 kDa +/- 1.2 % and 9.0 kDa +/- 1.7 %) was measured. Note the clear peak shift and increased mass upon mixing the two constructs, suggesting that complex formation occurred. Molecular weight (in kDa) is plotted on the left y-axis, change in refractive index (dRI, relative scale) plotted on the right y-axis, column elution volume (CV, 24 ml) of column S75 10/300 plotted on the x-axis. Standard TBS supplemented with 5 mM $CaCl_2$ was used as buffer. **B)** Analytical SEC experiment of Calmodulin alone (purple, 180 µM), Cnn$^{M1-K360}$ alone (brown, 50 µM), and both mixed together (black, same concentrations as before). The same buffer as in A was used. **C)** 4-20 % reducing SDS-PAGE gel showing fractions taken during the analytical SEC experiment with both constructs mixed (see B, black). The bands correspond to the constructs indicated. Note Cnn$^{M1-K360}$ and Calmodulin are not observed to co-elute, indicating that under these conditions the two constructs do not bind.

4.4 $Cnn^{S136-S205}$ forms a dimeric coiled-coil *in crystallo*

4.4.1 $Cnn^{S136-S205}$ structure determination

Crystals for the Calmodulin:$Cnn^{S136-S205}$ complex were sought under the conditions detailed in Table 4.1 in the presence of $CaCl_2$ in an attempt to aid complex formation. 42 crystals in total from a range of conditions and morphologies were fished and frozen using 25 % (v/v) ethylene glycol as a cryoprotectant in each case. Data collection was carried out automatically at the ESRF beamline MASSIF1.

Number	Screen type	[$Cnn^{S136-S205}$]	[CaM]	[$CaCl_2$]	Crystals?
97	Stura	0.3 mM	0.6 mM	10 mM	Yes
98	JCSG+	0.3 mM	0.6 mM	10 mM	Yes
99	ProPlex	0.3 mM	0.6 mM	10 mM	Yes
100	Structure Screen 1+2	0.3 mM	0.6 mM	10 mM	Yes
101	Clear Strategy I	0.3 mM	0.6 mM	10 mM	Yes
102	Clear Strategy II	0.3 mM	0.6 mM	10 mM	Yes
103	Stura	0.6 mM	0.6 mM	10 mM	Yes
104	MIDAS+	0.6 mM	0.6 mM	10 mM	NO
105	Morpheus	0.6 mM	0.6 mM	10 mM	Yes
106	Clear Strategy II	0.6 mM	0.6 mM	10 mM	Yes
107	Structure Screen 1+2	0.6 mM	0.6 mM	10 mM	Yes
108	PACT	0.6 mM	0.6 mM	10 mM	NO

Table 4.1 | Crystallisation screens of Calmodulin:$Cnn^{S136-S205}$. The first column indicates the unique tray number used to identify the screen. The second column shows the screen type used, commercially available from Molecular Dimensions. The next three columns show the respective concentrations of $Cnn^{S136-S205}$, Calmodulin (CaM) and $CaCl_2$ (all in mM) used in the final stock solution. The last column indicates whether the screen gave crystals that were used for data collection at the synchrotron.

All crystals that could be indexed were determined to have the same space group ($P4_122$) and almost identical unit cell dimensions, suggesting that they all contained the same protein composition. The crystal that diffracted to the highest resolution, which had grown in tray 103 condition B3 (i.e. 1.5 M ammonium sulphate, 0.15 M sodium citrate, pH 5.5) was therefore used for structure determination (Table 4.2).

As mentioned above, Calmodulin was present in the crystallisation conditions, but trying to successively place the two domains of Calmodulin proved unsuccessful, suggesting that the crystal only contained $Cnn^{S136\text{-}S205}$. This was not surprising, considering that the interaction between the two constructs appeared to be relatively weak in solution (see above).

Since no obvious homology model for $Cnn^{S136\text{-}S205}$ was available, a 30 amino acid stretch of a generic helix (PDB code: 5AL6) was used. Phaser was instructed to look for eight copies of this helix and successfully managed to place three, with translational function Z-scores of 9.3, 8.8 and 10.9. Z-scores indicate the number of standard deviations above the mean, with Z-scores over 8 often corresponding to correct solutions in Phaser [344]. Inspection of the electron density confirmed that the helices were placed into connecting stretches of electron density. Interestingly, two of the helices were placed immediately after one another, and indeed the two clashed at their termini, suggesting that they might form part of a longer helical stretch (Fig. 4.8 A). The two helices were therefore connected into one in Coot, and Phaser was then successfully used to place another copy of this elongated helix. Chainsaw was then used to prune all side chains. Iterative cycles of model building (Coot), refinement (refmac) and molecular replacement (Molrep) to place more short helices were then tried in order to build up the unit cell contents.

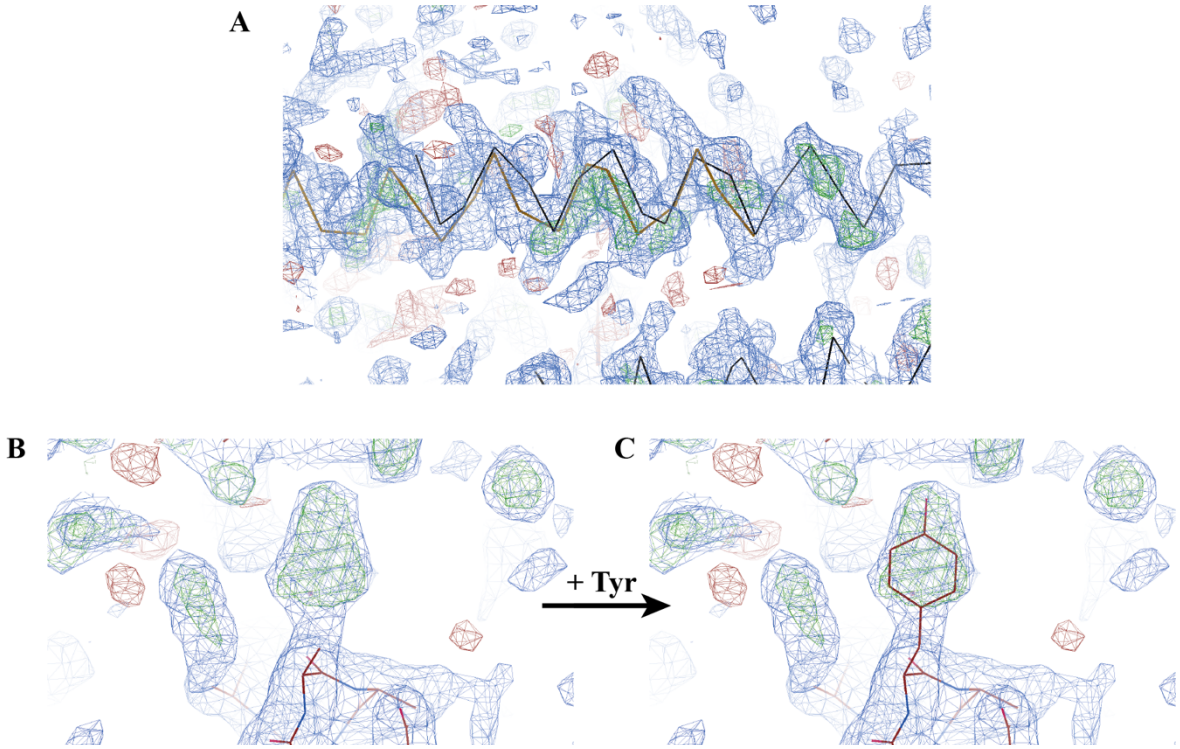


Figure 4.8 | Cnn$^{S136-S205}$ model building and refinement. A) Map and backbone model showing two of the three short helices (in orange and black) after initial molecular replacement; note the clash between the helices at their termini, suggesting that they form part of one longer helix; electron density map contoured at 1.5 σ. **B/C)** Detail showing side chain density of one residue towards the C-terminus of chain A; note the density can be modelled by introduction of a Tyr side chain (i.e. Cnn^{Tyr200}), which therefore allows assignment of the correct amino acid register to the rest of the chain; electron density map contoured at 1.0 σ.

Beamline	ESRF ID30A-1, MASSIF1
Wavelength (Å)	0.966
Resolution limits (Å)	8.21-1.84 (1.88-1.84)
Space group	$P4_122$
Unit cell dimensions (Å, º)	59.23, 59.23, 174.53 90, 90, 90
Unique reflections	27987 (1962)
Multiplicity	5.6 (5.1)
Completeness (%)	99.3 (96.7)
I/ σ(I)	9.1 (0.41)
R_{merge} (%)	11.5 (493.1)
R_{meas} (%)	12.7 (546.7)
$CC_{1/2}$	1.0 (0.3)
Processing pipeline	XDSAPP (ESRF)

Table 4.2 | $Cnn^{S136-S205}$ data collection and processing statistics. Where applicable, values for the inner resolution shell are given in brackets. Note that XDSAPP did not output a value for R_{pim}.

Resolution (Å)	41.89-1.84
No. of reflections:	
working set	26392
test set	1390
R_{work} / R_{free} (%)	25.45 / 27.34
Asymmetric unit composition	dimer
Observed residues	Chain A: E138-S205 Chain B: E138-E204
No. of atoms:	
Overall	1154
Protein	1078
Water	76
Average B factors (Å^2):	
Overall	52.6
Protein	52.6
Water	52.7
Ramachandran statistics (%):	
favoured	100.0
allowed	0
outliers	0
Rotamer outliers (%)	2.56
RMS deviation:	
bond lengths (Å)	0.011
bond angles (º)	0.958
Clashscore	3.22
MolProbity score	1.42

Table 4.3 | $Cnn^{S136-S205}$ refinement statistics.

The clear side chain density of Tyr200 was then used as an anchor to build in all side chains in the correct register (Fig. 4.8 B and C). Further cycles of model building and refinement followed until an R_{free} of 27.34 % and an R_{work} of 25.45 % was reached (Table 4.3).

At a resolution of 1.84 Å, lower values for R_{free} and R_{work} could be expected. However, note that the crystal contained an unusually large solvent content. Matthews coefficient suggested that for an estimated molecular weight of 16.6 kDa (corresponding to a dimer of $Cnn^{S136-S205}$) the solvent content is 73 %, and so the probability of having only one dimer of $Cnn^{S136-S205}$ per asymmetric unit was determined to be 1 %. Indeed, the probability of having such a crystal diffract to 1.84 Å was deemed to be zero by Matthews coefficient. In contrast, the probability of having a tetramer of $Cnn^{S136-S205}$ in the asymmetric unit at this resolution was deemed to be 100 %, with a solvent content of 47 %. However, no significant areas of unexplained density were observed, and the crystal packing observed would be enough to explain lattice formation. It therefore appears that the crystal is indeed highly unusual in containing a high solvent content while still diffracting to a very high resolution. This large solvent content, in combination with a range of disordered side chains in the model, can explain such a comparatively high R_{free} at this resolution.

4.4.2 Cnn$^{S136-S205}$ structure analysis

The asymmetric unit contains two copies of Cnn$^{S136-S205}$. Dimerisation occurs via the N-terminus, which forms a parallel, 22 amino acid long canonical coiled-coil covering residues L140-K161. Note that this corresponds closely to the predicted Calmodulin binding site around residues S139-V158. Residue K161 probably breaks the coiled-coil, as charged residues in position d are usually disallowed, and the dimer splays open after this residue (Fig. 4.9 A/B/E). The construct is probably mostly monomeric in solution and only shows a weak tendency to oligomerise at the concentrations tested (Fig. 4.6) because of the short length of the coiled-coil region not being able to drive full dimerisation in solution.

Analysis of the distribution of charged and polar residues across the dimer shows that towards the N-terminus, where the construct forms a coiled-coil, there are both patches of positive and negatively charged regions. In contrast, however, where the dimer splays apart, negatively charged areas clearly dominate and indeed are also found facing inwards towards the other chain, possibly explaining why these regions do not form a coiled-coil anymore (Fig. 4.9 C). Furthermore, the variation in B-factor value across the two chains clearly shows that flexibility is minimised where the construct dimerises to form a coiled-coil, but is increased where it splays open. This could suggest that, in the context of the rest of Cnn, the region L140-K161 forms a stable dimer, while the region immediately C-terminal to it shows a degree of flexibility and could possibly accommodate different conformations, one of which is captured in this crystal structure (Fig. 4.9 D).

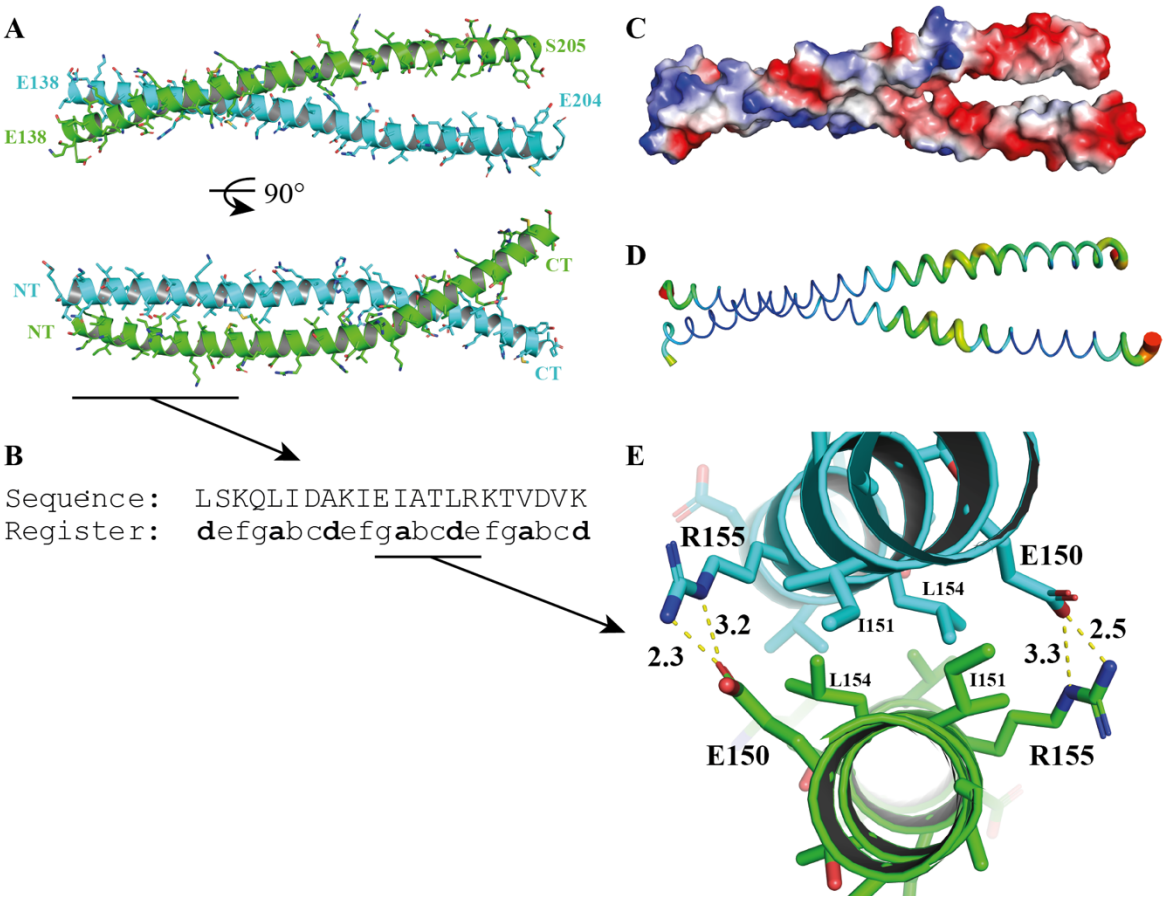


Figure 4.9 | Cnn$^{S136-S205}$ forms a dimer with a coiled-coil towards its N-terminus. A) The dimeric structure is shown as a cartoon representation, with side chains as sticks. One chain is in green, the other in turquoise. The structure is shown from two angles rotated by 90°. The N-termini (NT) and C-termini (CT) and the corresponding residues of both chains are indicated. **B)** Coiled-coil register of Cnn$^{L140-K161}$, determined using SOCKET [345]. Register positions d and a, which face inside, are highlighted in bold. Note the coiled-coil area in the crystal overlaps almost exactly with the predicted Calmodulin binding site. **C)** Vacuum electrostatics surface view of Cnn$^{S136-S205}$ generated via PyMOL. Negatively charged regions are red, positively charged regions are blue. **D)** B-factor putty model of Cnn$^{S136-S205}$ generated via PyMOL. Residues with a lower B-factor, i.e. little flexibility, are highlighted as a thin tube in blue; residues with a higher B-factor, i.e. greater flexibility, are highlighted as a thick tube in green-yellow-red at increasing flexibility. Note B-factors are minimal where the coiled-coil forms and increase as it splays open, with crystal contacts stabilising the more C-terminal regions again. **E)** Close up view of one heptad repeat of the coiled-coil region towards the N-terminus of Cnn$^{S136-S205}$. Note the typical coiled-coil conformation of hydrophobic residues at heptad positions a (I151) and d (L154) facing inwards, while charged residues at positions g (E150) and e (R155) form salt bridges with each other across the two chains. Inter-atomic distances are given in Å.

Interestingly, the splayed C-terminus of the $Cnn^{S136\text{-}S205}$ dimer packs against another crystallographic symmetry copy to form a crystallographic dimer of dimers (Fig. 4.10 A/B). This stabilising packing likely explains why part of the C-terminal region of $Cnn^{S136\text{-}S205}$ has reduced B-factor values (Fig. 4.9 D).

Strikingly, when this crystallographic dimer of $Cnn^{S136\text{-}S205}$ dimers is overlaid with the known structure of the $Cnn^{H1082\text{-}L1148}$:$Cnn^{D490\text{-}K544}$ 2:2 complex, the $Cnn^{H1082\text{-}L1148}$ dimer overlays almost perfectly with a dimer of $Cnn^{S136\text{-}S205}$. $Cnn^{D490\text{-}K544}$ overlays with parts of the crystallographic copy of $Cnn^{S136\text{-}S205}$, but diverges significantly from it towards the N-terminus (Fig. 4.10 C). Such a head-to-head association of dimeric coiled-coils also occurs in other proteins. A structure-guided search using the search engine TopSearch and 5MVW as a search model identified several hits, such as domains of MuRF1 (PDB code 4M3L), α-tropomyosin (1KQL, 2D3E), Homer1 (3CVE) and Homer3 (3CVF). However, it is surprising to find two regions with such a similar conformation within the same protein. Whether this similarity in conformation hints at a potential binding event is unclear, as the Cnn C-terminus did not appear to interact with $Cnn^{M1\text{-}K360}$, which contains $Cnn^{S136\text{-}S205}$, under the conditions tested (Fig. 3.14).

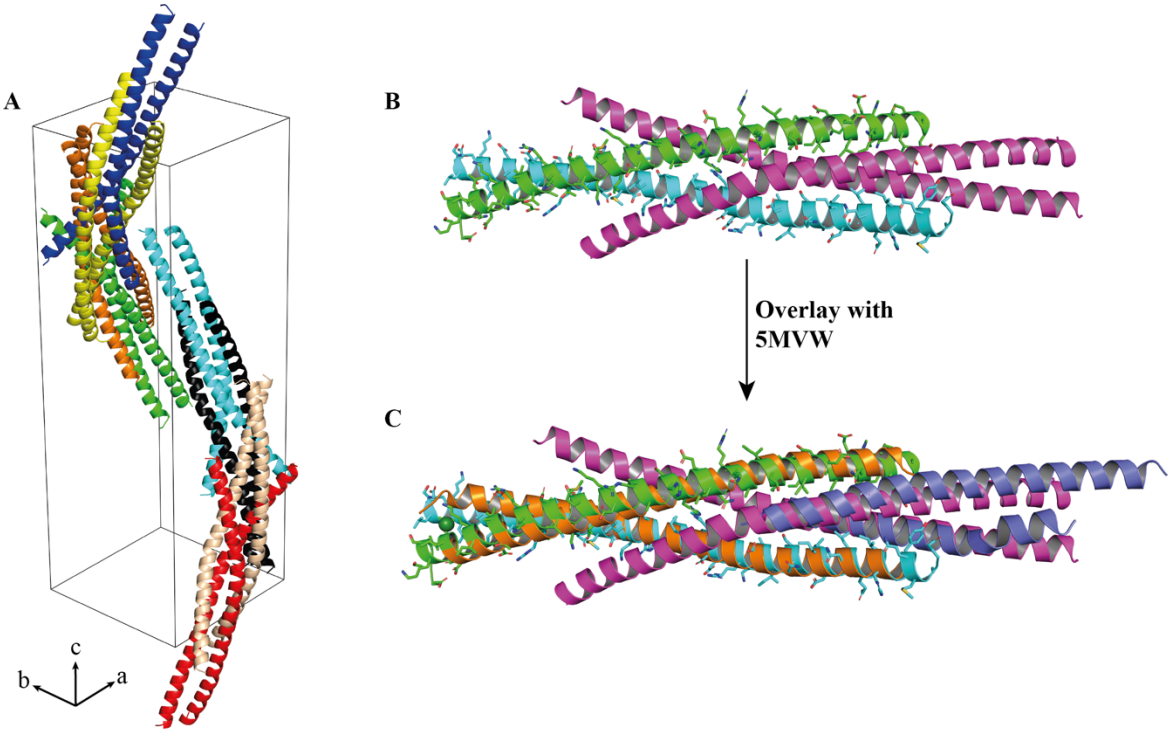


Figure 4.10 | Cnn$^{S136\text{-}S205}$ structure shows striking similarity with Cnn$^{H1082\text{-}L1148}$:Cnn$^{D490\text{-}K544}$ complex structure. A) Unit cell and crystal lattice arrangement (space group $P4_122$) of Cnn$^{S136\text{-}S205}$ in cartoon representation, with each asymmetric unit in a different colour. Unit cell axes are indicated. The unit cell contains eight dimers of Cnn$^{S136\text{-}S205}$ which appear to form four dimers of dimers via crystallographic contacts. **B)** The structure of Cnn$^{S136\text{-}S205}$ is shown as in Fig. 4.9 A, but a further crystallographic copy is shown as a cartoon representation in pink, illustrating the formation of a dimer of dimers. **C)** Overlay of the Cnn$^{S136\text{-}S205}$ structure shown in A with the 2:2 complex structure of Cnn$^{H1082\text{-}L1148}$:Cnn$^{D490\text{-}K544}$ (PDB code: 5MVW, cf. Fig. 3.1, [339]). Cnn$^{H1082\text{-}L1148}$ is shown as an orange cartoon, Cnn$^{D490\text{-}K544}$ is shown as a blue cartoon, the zinc ion at the N-terminus of Cnn$^{H1082\text{-}L1148}$ is shown as a green sphere.

4.5 Calmodulin:Cnn$^{H1082\text{-}L1148}$ interaction is calcium-sensitive

An interaction between Calmodulin and the Cnn C-terminus (i.e. CM2) had previously been reported [272]. This interaction was confirmed via analytical SEC. A peak shift occurred when Calmodulin and Cnn$^{H1082\text{-}L1148}$ were mixed at 240 μM and 450 μM, respectively, before loading on a S75 10/300 column pre-equilibrated in buffer containing 20 mM Tris-HCl (pH 7.3), 150 mM NaCl, 5 mM BME, and 10 mM $CaCl_2$. SDS-PAGE analysis showed that this early eluting peak contained both constructs, thereby confirming that complexation between the two constructs had occurred (Fig. 4.11 A and B).

In order to characterise the stoichiometry of this interaction, Calmodulin and Cnn$^{H1082\text{-}L1148}$ were mixed at increasing concentrations of 250 μM, 500 μM and 1 mM, with Cnn$^{H1082\text{-}L1148}$ at a twofold molar excess. The samples were analysed via SEC-MALS (Fig. 4.11 C). This showed that, interestingly, the mass of the complex appeared to increase from 29.8 kDa (+/- 0.3 %) to 32.8 kDa (+/- 0.2 %) and finally 36.6 kDa (+/- 0.4 %) as the concentration was increased, and the position of the peak concomitantly shifted to an earlier-eluting position. Furthermore, in all three cases, the peak corresponding to the complex had a pronounced tail, suggesting that it contained a heterogeneous mix of species. The measured masses closely match the expected mass of a 1:2 complex of Calmodulin:Cnn$^{H1082\text{-}L1148}$ at 33.3 kDa. However, the pronounced tail, the change in elution position and the increase in mass upon increasing the concentration suggest that a dynamic range of complexes can be formed between these two constructs. Indeed, some complexes might even include stoichiometries higher than 1:2, considering that at 1 mM concentration, the measured mass was slightly higher than that expected for a 1:2 complex (Fig. 4.11 C).

While the interaction between Calmodulin and the Cnn C-terminus had previously been characterised to be calcium-sensitive [272], the interaction between Calmodulin and the CEP215 C-terminus was suggested to lack a calcium-dependency [178]. Using SEC-MALS it could be confirmed that an interaction between Calmodulin and $Cnn^{H1082\text{-}L1148}$ still occurs even in the presence of 1 mM EGTA (Fig. 4.11 D). However, the apparent mass of the complex (19.7 kDa +/- 0.3 %) was drastically lower than the mass observed at the same concentrations of Calmodulin and $Cnn^{H1082\text{-}L1148}$ in the presence of 10 mM $CaCl_2$ (Fig. 4.11 C). This suggests that complex formation is indeed calcium-sensitive. However, note that $Cnn^{H1082\text{-}L1148}$ dimerisation was affected in the presence of EGTA, indicated by the presence of a pronounced tail, resulting in a noisy apparent mass (Fig. 4.11 D). $Cnn^{H1082\text{-}L1148}$ dimerisation is known to be dependent on the binding of zinc at its N-terminus, and it is therefore possible that complexation between Calmodulin and $Cnn^{H1082\text{-}L1148}$ was reduced not due to the lack of calcium but rather due to a disruption of the quaternary structure of $Cnn^{H1082\text{-}L1148}$.

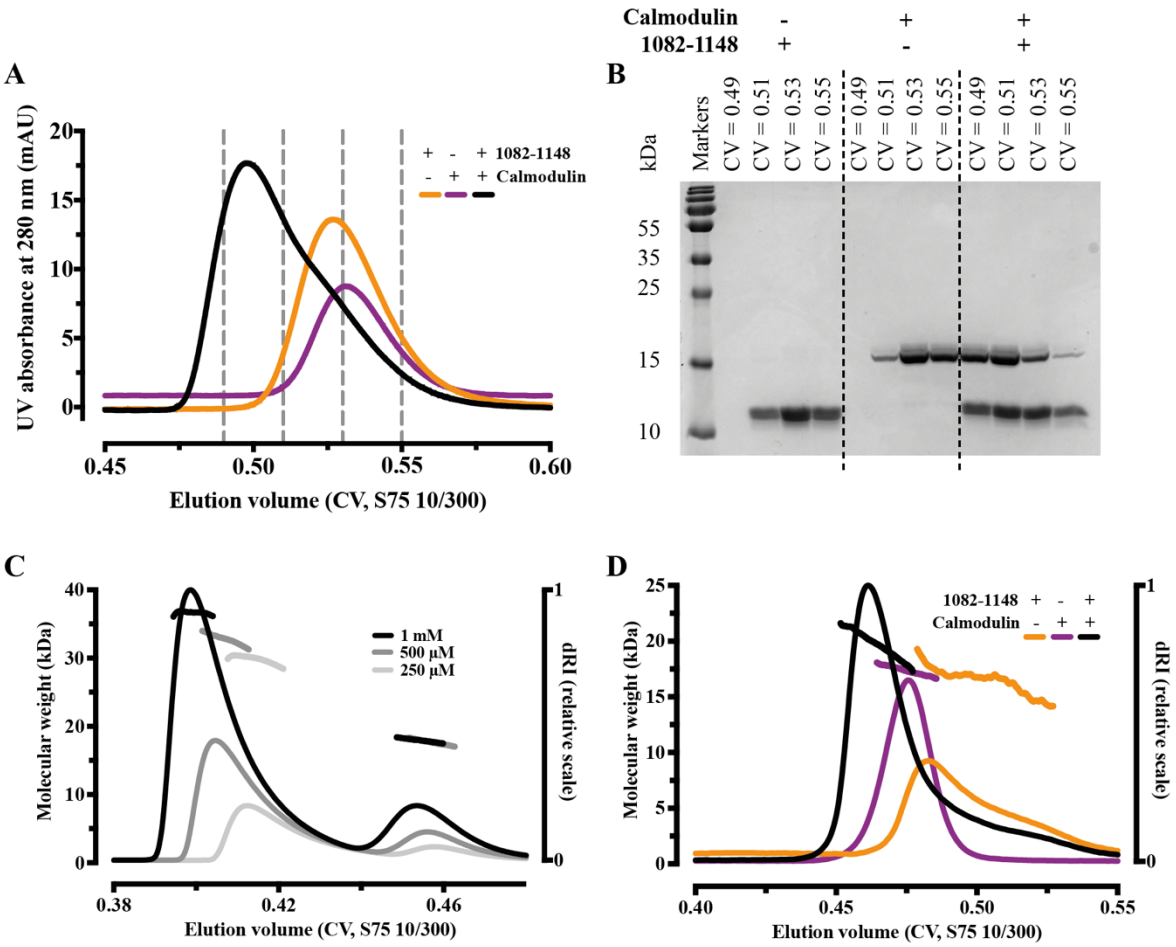


Figure 4.11 | Calmodulin interacts with the Cnn C-terminus in a calcium-sensitive manner. **A/B)** Analytical SEC elution profiles and 15 % SDS-PAGE gel looking at the interaction between Calmodulin and $Cnn^{H1082\text{-}L1148}$. 20 mM Tris-HCl (pH 7.3), 150 mM NaCl, 5 mM BME, 10 mM $CaCl_2$ was used as buffer. UV absorbance (280 nm) plotted on the y-axis, column elution volume (CV, 24 ml) of column S75 10/300 plotted on the x-axis. Calmodulin (240 µM) in purple, $Cnn^{H1082\text{-}L1148}$ (450 µM) in orange, both mixed (same concentrations) in black. Dashed grey lines correspond to samples run on the gel in B. Note the peak shift and appearance of both constructs in fraction CV = 0.49 upon mixing, indicating binding. **C)** SEC-MALS experiment measuring the average mass of the Calmodulin:$Cnn^{H1082\text{-}L1148}$ complex at increasing concentrations. Concentration of Calmodulin is indicated, $Cnn^{H1082\text{-}L1148}$ is in twofold molar excess. The same buffer as in A was used, but with 0.5 mM TCEP instead of BME. The measured mass of the complex rises from 29.8 kDa (+/- 0.3 %) to 32.8 kDa (+/- 0.2 %) and finally 36.6 kDa (+/- 0.4 %). This matches the theoretical mass of a 1:2 complex between Calmodulin:$Cnn^{H1082\text{-}L1148}$ (33.3 kDa). **D)** SEC-MALS experiment using the same buffer as in C but in the presence of 1 mM EGTA instead of $CaCl_2$. Calmodulin (250 µM, 17.3 kDa +/- 0.2 %) in purple, $Cnn^{H1082\text{-}L1148}$ (500 µM, 16.7 kDa +/- 1.7 %) in orange, both mixed (same concentrations, 19.7 kDa +/- 0.3 %) in black. Note the reduced apparent mass of the complex compared to C, indicating that in the absence of calcium the complex still forms but does so less efficiently. In both C and D, molecular weight (in kDa) is plotted on the left y-axis, change in refractive index (dRI, relative scale) plotted on the right y-axis, column elution volume (CV, 24 ml) of column S75 10/300 increase (C) and S75 10/300 (D) plotted on the x-axis.

4.6 Calmodulin crosslinks Cnn$^{H1082\text{-}L1148}$ dimers *in crystallo*

4.6.1 Cnn$^{H1082\text{-}V1140}$ structure determination

Crystals for the complex between Calmodulin and CM2 were initially sought using Cnn$^{H1082\text{-}V1140}$ as a binding partner for Calmodulin. This construct, which was initially cloned and purified by Dr Violet Feng for use in other experiments, contains the site where Calmodulin is predicted to bind (Fig. 4.1). However, it lacks eight amino acids at the extreme C-terminus which could not be observed in the structure of Cnn$^{H1082\text{-}L1148}$ previously solved by Dr Violet Feng (PDB code: 5I7C, [339]). These residues are therefore likely to be disordered, which could impede crystallisation of the complex. Crystals were sought under the conditions detailed in Table 4.4.

Data collection was carried out by Prof Susan Lea at the DIAMOND beamline i04 (proposal mx18069). Only one crystal could be indexed ($P4_12_12$) with acceptable statistics (Table 4.5). This crystal had grown in tray 49 condition A3 (i.e. 0.2 M imidazole malate, pH 5.5, 33 % (v/v) PEG 600), and the mother liquor had been used for cryoprotection due to its high PEG 600 content.

Number	Screen type	[$Cnn^{H1082-V1140}$]	[CaM]	[$CaCl_2$]	Crystals?
49	Stura	0.6 mM	0.9 mM	0.2 mM	Yes
51	ProPlex	0.6 mM	0.9 mM	0.2 mM	No
53	JCSG+	0.6 mM	0.9 mM	0.2 mM	Yes
55	PACT	0.6 mM	0.9 mM	0.2 mM	No
57	Clear Strategy I	0.6 mM	0.9 mM	0.2 mM	No
59	Clear Strategy II	0.6 mM	0.9 mM	0.2 mM	No
61	PGA	0.6 mM	0.9 mM	0.2 mM	No

Table 4.4 | Crystallisation screens of Calmodulin:$Cnn^{H1082-V1140}$. The first column indicates the unique tray number used to identify the screen. The second column shows the screen type used, commercially available from Molecular Dimensions. The next three columns show the respective concentrations of $Cnn^{H1082-V1140}$, Calmodulin (CaM) and $CaCl_2$ (all in mM) used in the final stock solution. The last column indicates whether the screen gave crystals that were used for data collection at the synchrotron. Note all final stock solutions were also made up to contain 0.5 mM TCEP and further contained 2.5 % (v/v) glycerol.

Beamline	DIAMOND i04
Wavelength (Å)	1.0719
Resolution limits (Å)	43.70-2.54 (2.58-2.54)
Space group	$P4_12_12$
Unit cell dimensions (Å, º)	43.70, 43.70, 175.14 90, 90, 90
Unique reflections	6130 (297)
Multiplicity	15.9 (16.4)
Completeness (%)	99.4 (98.7)
I/ σ(I)	16.6 (1.1)
R_{merge} (%)	8.7 (225.6)
R_{meas} (%)	9.0 (232.7)
R_{pim} (%)	2.3 (56.3)
$CC_{1/2}$	1.0 (0.9)
Processing pipeline	Xia2 3dii (DIAMOND)

Table 4.5 | $Cnn^{H1082-V1140}$ data collection and processing statistics. Where applicable, values for the inner resolution shell are given in brackets.

Matthews coefficient calculated that a species of 16.4 kDa would result in a solvent content of 51.90 % and assigned a probability of 98 % to this composition. This suggested that the asymmetric unit therefore probably contained either Calmodulin or a dimer of $Cnn^{H1082-V1140}$, but not the complex. In the published structure of $Cnn^{H1082-L1148}$, the Cnn C-terminus crystallised as a dimer of dimers in $P6_1$ (Fig. 3.1 D, [339]). One of the two dimers (PDB code: 5I7C) was therefore deleted and the resulting model used for molecular replacement via Phaser. This resulted in a single solution with a translational function Z-score of 14.1. As the model was longer than the construct that had crystallised, the termini were subsequently trimmed and the model was refined via refmac and phenix.refine for a few cycles. This resulted in an R_{free} of 35.18 % and an R_{work} of 31.49 % (Table 4.6). These reasonably good statistics, together with the lack of any further large unexplained density and the initial Matthews coefficient prediction led to the conclusion that this crystal did indeed not contain the complex between Calmodulin and $Cnn^{H1082-V1140}$ and it was therefore not refined any further.

$Cnn^{H1082-V1140}$ crystallised as a dimer in the asymmetric unit. However, in the crystal the dimer closely packs against another crystallographic copy of itself to form a dimer of dimers that is strikingly similar to the structure of the $Cnn^{H1082-L1148}$ tetramer, which had crystallised in a different space group (Fig. 4.12). This suggests that the tetramer formation first observed *in crystallo* for $Cnn^{H1082-L1148}$ is probably not an artefact due to the crystal packing in that crystal (PDB code: 5I7C), but rather reflects a genuine propensity for the Cnn C-terminus to splay open and receive a binding partner such as another dimer of itself (as observed in the structures for $Cnn^{H1082-V1140}$ solved here and $Cnn^{H1082-L1148}$ solved previously by Dr Violet Feng), or a dimer of the zipper-like motif (Fig. 3.1).

Resolution (Å)	42.41-2.54
No. of reflections:	
working set	5732
test set	302
R_{work} / R_{free} (%)	31.49 / 35.18
Asymmetric unit composition	dimer
Observed residues	Chain A: C1084-N1139 Chain B: A1085-V1140
No. of atoms:	
Overall	931
Protein	931
Water	0
Average B factor (Å^2)	111.9
Ramachandran statistics (%):	
favoured	99.07
allowed	0.93
outliers	0
Rotamer outliers (%)	4.00
RMS deviation:	
bond lengths (Å)	0.003
bond angles (°)	0.509
Clashscore	3.11
MolProbity score	1.56

Table 4.6 | Cnn$^{H1082\text{-}V1140}$ refinement statistics.

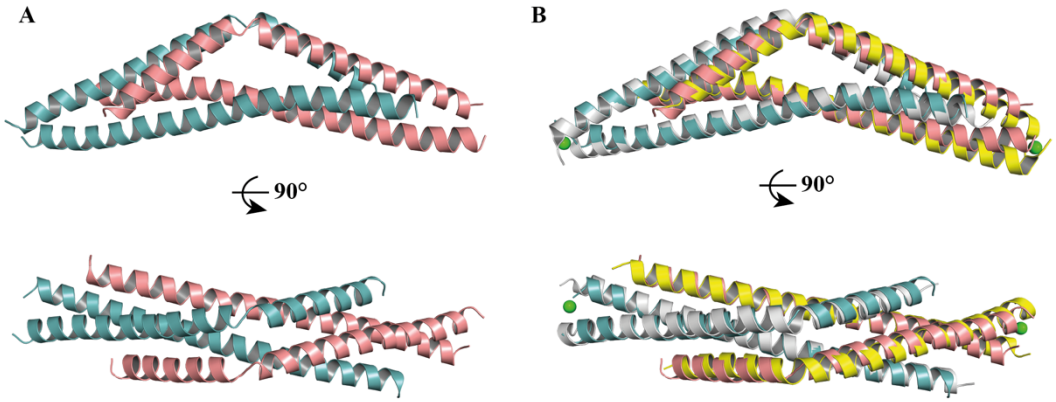


Figure 4.12 | Cnn$^{H1082\text{-}V1140}$ adopts the same overall structure as Cnn$^{H1082\text{-}L1148}$. A) The structure of Cnn$^{H1082\text{-}V1140}$ is shown in cartoon representation. The asymmetric unit contains a dimer of Cnn$^{H1082\text{-}V1140}$ (turquoise), which in the crystal lattice packs against another dimer (red). **B)** Overlay of the Cnn$^{H1082\text{-}V1140}$ dimer and its crystallographic copy as shown in A with the Cnn$^{H1082\text{-}L1148}$ tetramer (PDB code: 5I7C, [339]) as shown in Fig. 3.1 D. Note the two structures adopt the same overall conformation.

4.6.2 Calmodulin:Cnn$^{H1082\text{-}L1148}$ crystal screening

The crystallisation trials described above were carried out in the presence of 0.2 mM $CaCl_2$ to avoid formation of salt crystals. However, considering that the concentration of Calmodulin was 0.9 mM, and therefore the concentration of calcium-binding sites on Calmodulin 3.6 mM, a concentration of 0.2 mM $CaCl_2$ was probably too low to drive formation of the calcium-sensitive interaction between Calmodulin and the Cnn C-terminus. Going forward, crystals of the complex were therefore sought at a higher concentration of $CaCl_2$. Furthermore, the slightly longer construct Cnn$^{H1082\text{-}L1148}$ was used for further crystal trials. Crystals were sought under the conditions shown in Table 4.7.

Tables 4.8-4.10 detail the 43 crystals that were ultimately fished and frozen with the help of Prof Susan Lea and sent for automatic data collection at the ESRF beamline MASSIF1. A total of 16 out of 43 crystals were successively indexed. The space group and unit cell dimensions of these crystals are shown in Table 4.11. Note that almost all of these crystals were indexed in the same point group (622) and with similar unit cell dimensions. Indeed, these were all similar to the space group ($P6_1$) and unit cell dimensions (a = 49.2 Å, b = 49.2, c = 211.2; $\alpha = 90$, $\beta = 90$, $\gamma = 120$) observed for the previously published structure of Cnn$^{H1082\text{-}L1148}$ (PDB code: 5I7C, [339]). It is therefore likely that these crystals also all contained just Cnn$^{H1082\text{-}L1148}$ and not the Calmodulin:Cnn$^{H1082\text{-}L1148}$ complex, and structure determination of these crystals was therefore not attempted.

Number	Screen type	[$Cnn^{H1082-L1148}$]	[CaM]	[$CaCl_2$]	Crystals?
62	Stura	0.75 mM	0.75 mM	5 mM	Yes
63	Stura	0.75 mM	0.75 mM	10 mM	No
64	ProPlex	0.75 mM	0.75 mM	5 mM	Yes
65	ProPlex	0.75 mM	0.75 mM	10 mM	No
66	ProPlex	0.75 mM	0.75 mM	10 mM	No
67	JCSG+	1.4 mM	0.7 mM	10 mM	Yes
68	PACT	1.4 mM	0.7 mM	10 mM	Yes
69	Clear Strategy I	1.4 mM	0.7 mM	10 mM	Yes
70	Clear Strategy II	1.4 mM	0.7 mM	10 mM	Yes
71	PGA	1.4 mM	0.7 mM	10 mM	Yes
72	Structure Screen 1+2	1.4 mM	0.7 mM	10 mM	Yes
73	Morpheus	1.4 mM	0.7 mM	10 mM	Yes
74	MIDAS	1.4 mM	0.7 mM	10 mM	Yes
75	LMB	1.4 mM	0.7 mM	10 mM	Yes
76	ProPlex	1.4 mM	0.7 mM	10 mM	Yes

Table 4.7 | Crystallisation screens of Calmodulin:$Cnn^{H1082-L1148}$. The first column indicates the unique tray number used to identify the screen. The second column shows the screen type used, commercially available from Molecular Dimensions. The next three columns show the respective concentrations of $Cnn^{H1082-L1148}$, Calmodulin (CaM) and $CaCl_2$ (all in mM) used in the final stock solution. The last column indicates whether the screen gave crystals that were used for data collection at the synchrotron. Note all final stock solutions were also made up to contain 0.5 mM TCEP.

ID	Screen	Condition	Cryoprotectant
VLT1	62_B9	0.75 M sodium citrate tribasic dihydrate 0.01 M sodium borate pH 8.5	20 % (v/v) glycerol
VLT2	62_G1	1.0 M lithium sulphate 2 % (w/v) PEG 8000	20 % (v/v) glycerol
VLT3	62_H4	2.0 M ammonium sulphate 0.1 M sodium HEPES pH 7.5 5 % (v/v) PEG 400	20 % (v/v) glycerol
VLT4	62_H3	2.0 M ammonium sulphate 0.1 M sodium HEPES pH 7.5 2 % (v/v) PEG 400	20 % (v/v) glycerol
VLT5	64_H4	1.4 M sodium malonate dibasic monohydrate pH 6.0	15 % (v/v) glycerol
VLT6	64_A6	0.2 M sodium chloride 0.1 M sodium/potassium phosphate pH 6.5 25 % (w/v) PEG 1000	25 % (v/v) PEG 300
VLT7	64_H5	1.2 M potassium sodium tartrate tetrahydrate 0.1 M Tris pH 8.0	15 % (v/v) glycerol
VLT8	76_G9	1.3 M sodium/potassium phosphate pH 7.0	25 % (v/v) glycerol
VLT9	76_G9	1.3 M sodium/potassium phosphate pH 7.0	25 % (v/v) glycerol
VLT10	76_E10	0.2 M ammonium sulphate 0.1 M Tris pH 8.5 12 % (w/v) PEG 8000	25 % (v/v) PEG 300
VLT11	70_G8	1.8 M lithium sulphate 0.1 M Tris pH 8.5	25 % (v/v) glycerol
VLT12	70_E3	2.0 M sodium formate 0.1 M Tris pH 7.5	20 % (v/v) PEG 300
VLT13	70_G7	2.7 M ammonium sulphate 0.1 M Tris pH 8.5	25 % (v/v) glycerol
VLT14	70_C9	4.0 M sodium formate 0.1 M sodium cacodylate pH 6.5	5 % (v/v) PEG 300
VLT15	72_A7	0.2 M ammonium acetate 0.1 M sodium citrate pH 5.6 30 % (w/v) PEG 4000	10 % (v/v) PEG 300
VLT16	72_B1	0.2 M sodium citrate tribasic dihydrate 0.1 M sodium cacodylate pH 6.5 30 % (v/v) 2-propanol	25 % (v/v) PEG 300
VLT17	71_E4	0.2 M potassium bromide 0.2 M potassium thiocyanate 0.1 M MES pH 6.5 3 % (w/v) PGA, 0 % (w/v) PEG 2000 MME	25 % (v/v) PEG 300

Table 4.8 | Potential Calmodulin:Cnn$^{H1082\text{-}L1148}$ crystals sent for data collection. First column indicates the unique crystal ID; second column shows the screen and well where the crystal was found; third column shows the composition of the mother liquor of that well; fourth column shows the cryoprotectant used.

ID	Screen	Condition	Cryoprotectant
VLT18	71_E6	0.2 M potassium bromide 0.2 M potassium thiocyanate 0.1 M MES pH 6.5 3 % (w/v) PGA, 5 % (w/v) PEG 4000	25 % (v/v) PEG 300
VLT19	71_D6	0.1 M MES pH 6.5 5 % (w/v) PGA, 30 % (v/v) PEG 500 MME	Mother liquor
VLT20	71_C2	0.1 M ammonium sulphate 0.3 M sodium formate 0.1 M sodium acetate pH 5.0 3 % (w/v) PGA, 20 % (v/v) PEG 500 MME	Mother liquor
VLT21	73_A10	0.06 M magnesium/calcium chloride 0.1 M Tris pH 8.5 20 % (v/v) PEG 500 MME 10 % (w/v) PEG 20000	Mother liquor
VLT22	73_B7	0.09 M halogens 0.1 M HEPES pH 7.5 20 % (v/v) glycerol, 10 % (w/v) PEG 4000	Mother liquor
VLT23	73_F10	0.12 M monosaccharides 0.1 M Tris pH 8.5 20 % (v/v) ethylene glycol 10 % (w/v) PEG 8000	Mother liquor
VLT24	73_G7	0.1 M carboxylic acids 0.1 M HEPES pH 7.5 20 % (v/v) glycerol, 10 % (w/v) PEG 4000	Mother liquor
VLT25	73_D7	0.12 M alcohols 0.1 M HEPES pH 7.5 20 % v/v glycerol, 10 % (w/v) PEG 4000	Mother liquor
VLT26	73_E6	0.12 M ethylene glycols 0.1 M HEPES pH 7.5 20 % (v/v) ethylene glycol 10 % (w/v) PEG 8000	Mother liquor
VLT27	68_D12	0.002 M zinc chloride 0.1 M Tris pH 8.0 20 % (w/v) PEG 6000	25 % (v/v) PEG 300
VLT28	68_C9	0.2 M lithium chloride 0.1 M HEPES pH 7.0 20 % (w/v) PEG 6000	25 % (v/v) PEG 300
VLT29	68_C7	0.2 M sodium chloride 0.1 M HEPES pH 7.0 20 % (w/v) PEG 6000	25 % (v/v) PEG 300
VLT30	68_H1	0.2 M sodium fluoride 0.1 M Bis-Tris propane pH 8.5 20 % (w/v) PEG 3350	25 % (v/v) PEG 300

Table 4.9 | Further potential Calmodulin:Cnn$^{H1082\text{-}L1148}$ crystals sent for data collection. See Table 4.8 for legend. VLT27 is marked in green.

ID	Screen	Condition	Cryoprotectant
VLT31	67_A2	0.1 M sodium citrate pH 5.5 20 % (w/v) PEG 3000 20 % (w/v) PEG 3350	25 % (v/v) PEG 300
VLT32	67_E1	1.0 M sodium citrate tribasic dihydrate 0.1 M sodium cacodylate pH6.5	15 % (v/v) glycerol
VLT33	74_G12	0.1 M Bis-Tris pH 6.0 15 % (w/v) Sokalan CP 5	25 % (v/v) PEG 300
VLT34	75_F2	0.1 M ammonium sulphate 0.1 M sodium citrate pH 5.8 16 % (w/v) PEG 4000, 20 % (w/v) glycerol	Mother liquor
VLT35	69_C8_1	0.2 M lithium sulphate 0.1 M sodium cacodylate pH 6.5 15 % (w/v) PEG 4000	25 % (v/v) PEG 300
VLT36	69_C8_2	0.2 M lithium sulphate 0.1 M sodium cacodylate pH6 .5 15 % (w/v) PEG 4000	25 % (v/v) PEG 300
VLT37	69_B8_1	0.2 M Li sulphate 0.1 M sodium acetate pH 5.5 8 % (w/v) PEG 20000 8 % (v/v) PEG 500 MME	15 % (v/v) PEG 300
VLT38	69_B8_2	0.2 M lithium sulphate 0.1 M sodium acetate pH 5.5 8 % (w/v) PEG 20000 8 % (v/v) PEG 500 MME	15 % (v/v) PEG 300
VLT39	69_C3_1	0.2 M magnesium chloride hexahydrate 0.1 M sodium cacodylate pH 6.5 25 % (w/v) PEG 2000 MME	20 % (v/v) PEG 300
VLT40	69_C3_2	0.2 M magnesium chloride hexahydrate 0.1 M sodium cacodylate pH 6.5 25 % (w/v) PEG 2000 MME	20 % (v/v) PEG 300
VLT41	69_C3_3	0.2 M magnesium chloride hexahydrate 0.1 M sodium cacodylate pH 6.5 25 % (w/v) PEG 2000 MME	20 % (v/v) PEG 300
VLT42	69_F12_1	0.8 M sodium formate 0.1 M Tris pH 7.5 8 % (w/v) PEG 20000 8 % (v/v) PEG5 00 MME	15 % (v/v) PEG 300
VLT43	69_F12_2	0.8 M sodium formate 0.1 M Tris pH 7.5 8 % (w/v) PEG 20000 8 % (v/v) PEG 500 MME	15 % (v/v) PEG 300

Table 4.10 | Further potential Calmodulin:Cnn$^{H1082\text{-}L1148}$ crystals sent for data collection. See Table 4.8 for legend.

ID	Space group	Unit cell dimensions (Å, º)
VLT1	$P6_222$	44.7, 44.7, 129.6 90, 90, 120
VLT2	$P6_222$	45.0, 45.0, 129.7 90, 90, 120
VLT5	$P6_322$	48.3, 48.3, 205.5 90, 90, 120
VLT7	$P6_222$	44.5, 44.5, 128.7 90, 90, 120
VLT15	$P6_322$	48.1, 48.1, 202.4 90, 90, 120
VLT20	$P6_122$	49.2, 49.2, 207.1 90, 90, 120
VLT21	$P6_122$	49.3, 49.3, 206.7 90, 90, 120
VLT22	$P6_122$	49.2, 49.2, 217.1 90, 90, 120
VLT23	$P6_122$	49.2, 49.2, 206.9 90, 90, 120
VLT24	$P6_122$	49.6, 49.6, 207.4 90, 90, 120
VLT27	$P2_12_12_1$	37.4, 84.9, 95.7 90, 90, 90
VLT28	$P6_222$	49.0, 49.0, 100.2 90, 90, 120
VLT34	$P6_122$	49.4, 49.4, 209.9 90, 90, 120
VLT35	$P6_322$	49.0, 49.0, 207.4 90, 90, 120
VLT36	$P6_122$	49.0, 49.0, 207.3 90, 90, 120
VLT37	$P6_322$	49.1, 49.1, 206.2 90, 90, 120

Table 4.11 | Space group and unit cell dimensions of potential Calmodulin:Cnn$^{H1082-L1148}$ crystals that could be indexed. VLT27 is marked in green.

One single crystal (VLT27, see Fig. 4.13 A and B), however, showed a strikingly different space group and cell unit dimensions. Indeed, Matthews coefficient suggested that if the asymmetric unit was composed of one Calmodulin and a dimer of $Cnn^{H1082\text{-}L1148}$ (i.e. 33.3 kDa), the solvent content of VLT27 would be 46.07 %, and this was assigned a 100 % probability. This interesting lead was therefore pursued further, and structure determination was attempted on this crystal. Data collection and processing statistics of VLT27 are shown in Table 4.12.

Beamline	ESRF ID30A-1, MASSIF1		
Wavelength (Å)	0.966		
Resolution limits (Å)	47.91-2.576 (2.621-2.576)	9.44-2.11 (2.17-2.11)	9.44-2.11 (2.17-2.11)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions (Å, º)	37.47, 85.0, 95.82 90, 90, 90	37.42, 84.86, 95.68 90, 90, 90	37.42, 84.86, 95.68 90, 90, 90
Unique reflections	10037 (480)	32620 (2200)	17901 (1192)
Multiplicity	4.3 (4.6)	2.3 (1.9)	4.3 (3.6)
Completeness (%)	98.4 (99.6)	96.3 (87.8)	98.3 (90.5)
I/ σ(I)	11.4 (2.2)	5.05 (0.38)	6.78 (0.52)
R_{merge} (%)	7.9 (68.5)	10.7 (217.0)	12.1 (243.3)
R_{meas} (%)	9.0 (77.5)	13.6 (278.4)	13.8 (279.0)
R_{pim} (%)	4.2 (35.3)	see legend	see legend
$CC_{1/2}$	1.0 (0.7)	1.0 (0.1)	1.0 (0.2)
ccAnom (%)		2 (0)	
sigAnom		0.753 (0.594)	
Processing pipeline	autoPROC (ESRF)	XDSAPP anom (ESRF)	XDSAPP (ESRF)

Table 4.12 | VLT27 data collection and processing statistics. The same diffraction dataset was processed via three different indexing strategies. In the XDSAPP anom dataset, Friedel pairs are unmerged, and the anomalous signal is indicated. Note that XDSAPP did not output a value for R_{pim}. Where applicable, values for the inner resolution shell are given in brackets.

4.6.3 Calmodulin:Cnn$^{H1082\text{-}L1148}$ structure determination

Calmodulin consists of two domains connected by a flexible linker, so the relative orientation between the two domains can vary substantially from structure to structure. Consequently, a search model was generated that consisted of just the Calmodulin N-terminal domain (based on PDB code: 4CLN). Furthermore, this model was truncated at its N-terminus by a few amino acids, as termini are often disordered in crystals. This truncated N-terminal domain was used as a search model for molecular replacement via phaser using the autoPROC dataset (Table 4.12). This resulted in a single solution with a translational function Z-score of 7.2. As the Calmodulin binding site on the Cnn C-terminus was predicted to be towards the extreme C-terminus (Fig. 4.1), a short stretch of α-helix towards the N-terminal side of Cnn$^{H1082\text{-}L1148}$ (based on PDB code: 5I7C) was used as a search model next, in order to avoid any potential clashes with the placed Calmodulin N-terminal domain. The top solution gave a comparatively low translational function Z-score of 4.5, and this partial model was refined via refmac.

However, further model building proved difficult using this dataset and further work was carried out using the XDSAPP (anom) dataset. This dataset had slightly better statistics (Table 4.12) and in addition a weak anomalous diffraction signal which could be used to provide better estimates for phases. The partial solution with the Calmodulin N-terminal domain and a short stretch of N-terminal Cnn$^{H1082\text{-}L1148}$ α-helix was used as a search model in phaser for this dataset. This resulted in a single solution with a translational function Z-score of 15.2, and this solution was refined using refmac. Note any potential model bias resulting from using a partial model built in a different dataset would subsequently be lost as substantial further model building was required. Using this dataset, a further short Cnn$^{H1082\text{-}L1148}$ α-helix could be placed via phaser with a translational function Z-score of 6.2.

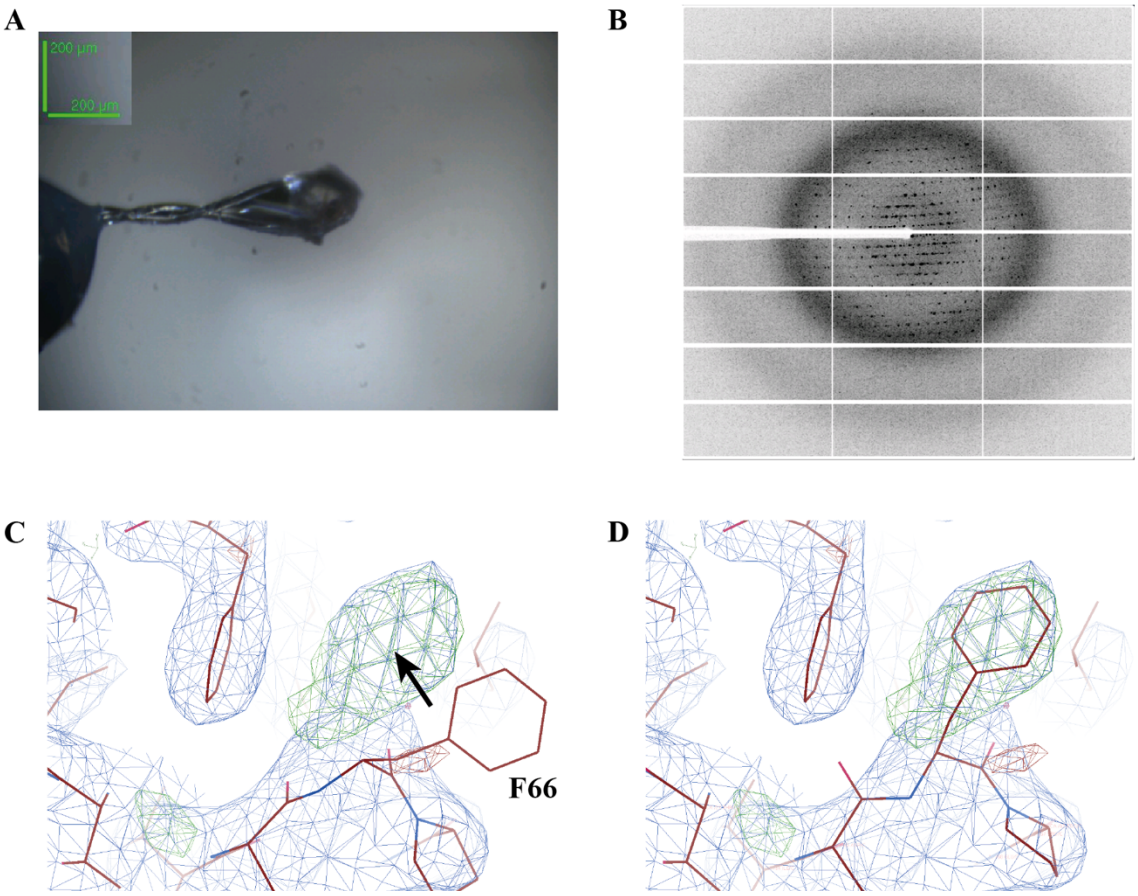


Figure 4.13 | Data collection and structure determination of Calmodulin:Cnn$^{H1082\text{-}L1148}$. **A)** VLT27 mounted on the cryo-loop right before data collection. The crystal displays a sheet-like morphology. **B)** Representative diffraction image of VLT27. Each dot is a reflection with Miller indices *hkl*. The beam stop is seen as a thick white line. **C/D)** Manual model refinement in Coot. Note the side chain of Calmodulin^{F66} is moved into the positive electron density shown as a green mesh. Electron density map contoured at 1.0 σ.

The model that had been built until so far was then used as a partial model and fed into phenix.AutoSol for experimental phasing. Furthermore, the buccaneer pipeline was used, which can automatically build a model when provided with the expected protein sequence. Crucially, these experiments suggested that where the Calmodulin N-terminal domain had been placed was probably the location of the C-terminal domain. Calmodulin^{M1-K76} and Calmodulin$^{M77-K149}$ show 47.4 % sequence identity when aligned using EMBOSS Needle [343]. Positioning the Calmodulin C-terminal domain at the position where the N-terminal domain was allowed placement of the Calmodulin N-terminal domain at a different position after further cycles of refinement and processing in buccaneer. Despite the similarity between the two domains, some clear side chain densities allowed the two domains to be unambiguously distinguished. For example, the density observed for F66 could fit a phenylalanine side chain after it was moved in place manually (Fig. 4.13 C and D) but not the larger tyrosine side chain of Y139, with which this residue aligns in the other domain.

The experimental phasing via phenix.AutoSol in combination with buccaneer merely helped to resolve the ambiguity between the Calmodulin N- and C-terminal domains but did not help determine the phases in any further way. It was therefore decided that further model building and refinement should be carried out in the XDSAPP dataset with merged Friedel pairs, as this dataset had better statistics than the XDSAPP anom dataset and also was indexed to a higher resolution than the autoPROC dataset (Table 4.12). In order to avoid any model bias upon transitioning to the XDSAPP dataset with merged Friedel pairs, pdbset was used via the command terminal to randomly displace all atoms by up to 0.5 Å per atom, and all B-factors were reset to 50.

Several further cycles of model building, refinement via phenix.refine and validation by inspection and correction of Ramachandran and rotamer outliers in Coot followed, until an R_{free} of 26.82 % and an R_{work} of 23.76 % was reached and structure determination was concluded (Table 4.13).

Resolution (Å)	47.85-2.11
No. of reflections:	
working set	17007
test set	896
R_{work} / R_{free} (%)	23.76 / 26.82
Asymmetric unit composition	2 x $Cnn^{H1082-L1148}$ 1 x Calmodulin
Observed residues	Cnn chain A: H1082-N1143 Cnn chain B: H1082-R1141 Calmodulin: Q4-S148
No. of atoms:	
Overall	2266
Protein	2157
Ligand	6
Water	103
Average B factors (Å^2):	
Overall	60.5
Protein	60.7
Ligand	60.9
Water	57.4
Ramachandran statistics (%):	
favoured	98.85
allowed	1.15
outliers	0
Rotamer outliers (%)	0.85
RMS deviation:	
bond lengths (Å)	0.002
bond angles (°)	0.397
Clashscore	4.88
MolProbity score	1.25

Table 4.13 | Calmodulin:$Cnn^{H1082-L1148}$ refinement statistics.

4.6.4 Calmodulin:Cnn$^{H1082\text{-}L1148}$ structure analysis

The asymmetric unit contains one copy of Calmodulin and two copies of Cnn$^{H1082\text{-}L1148}$ (Fig. 4.14). The two Cnn$^{H1082\text{-}L1148}$ chains dimerise at their N-termini via the zinc ion previously observed in other structures of Cnn$^{H1082\text{-}L1148}$ (Fig. 4.15 D). Calmodulin, which has all four of its calcium-binding sites occupied (Fig. 4.15 A), binds to Cnn$^{H1082\text{-}L1148}$ towards its C-terminus. In the asymmetric unit, the Calmodulin C-terminal domain binds to Cnn$^{H1082\text{-}L1148}$ chain B. In contrast, Cnn$^{H1082\text{-}L1148}$ chain A adopts a conformation that is distinctly different from Cnn$^{H1082\text{-}L1148}$ chain B. Specifically, Cnn$^{H1082\text{-}L1148}$ chain A shows a 90° kink centred around K1117, suggesting that Cnn$^{H1082\text{-}L1148}$ displays a large degree of structural plasticity (Fig. 4.14).

Interestingly, an ion is observed to be coordinated in a roughly tetrahedral coordination by Cnn^{H1134} from both chains, Calmodulin^{E115} and probably a water molecule. Considering the type of electron donors, the coordination geometry, and that the crystal grew in the presence of calcium, zinc and chloride as the only ions, the observed ion is most likely another zinc (Fig. 4.15 C). Through its coordination by all three polypeptides present in the asymmetric unit, the zinc ion links all three. Indeed, it is interesting to note that of the 43 crystals sent for data collection, the crystal that resulted in this complex structure was the only one that grew in the presence of zinc (Tables 4.8-4.10). It is therefore possible that this zinc ion is important for stabilising the 90° kink of Cnn$^{H1082\text{-}L1148}$ chain A as well as the overall binding between the two proteins in this conformation, thereby allowing for a stable crystal to form.

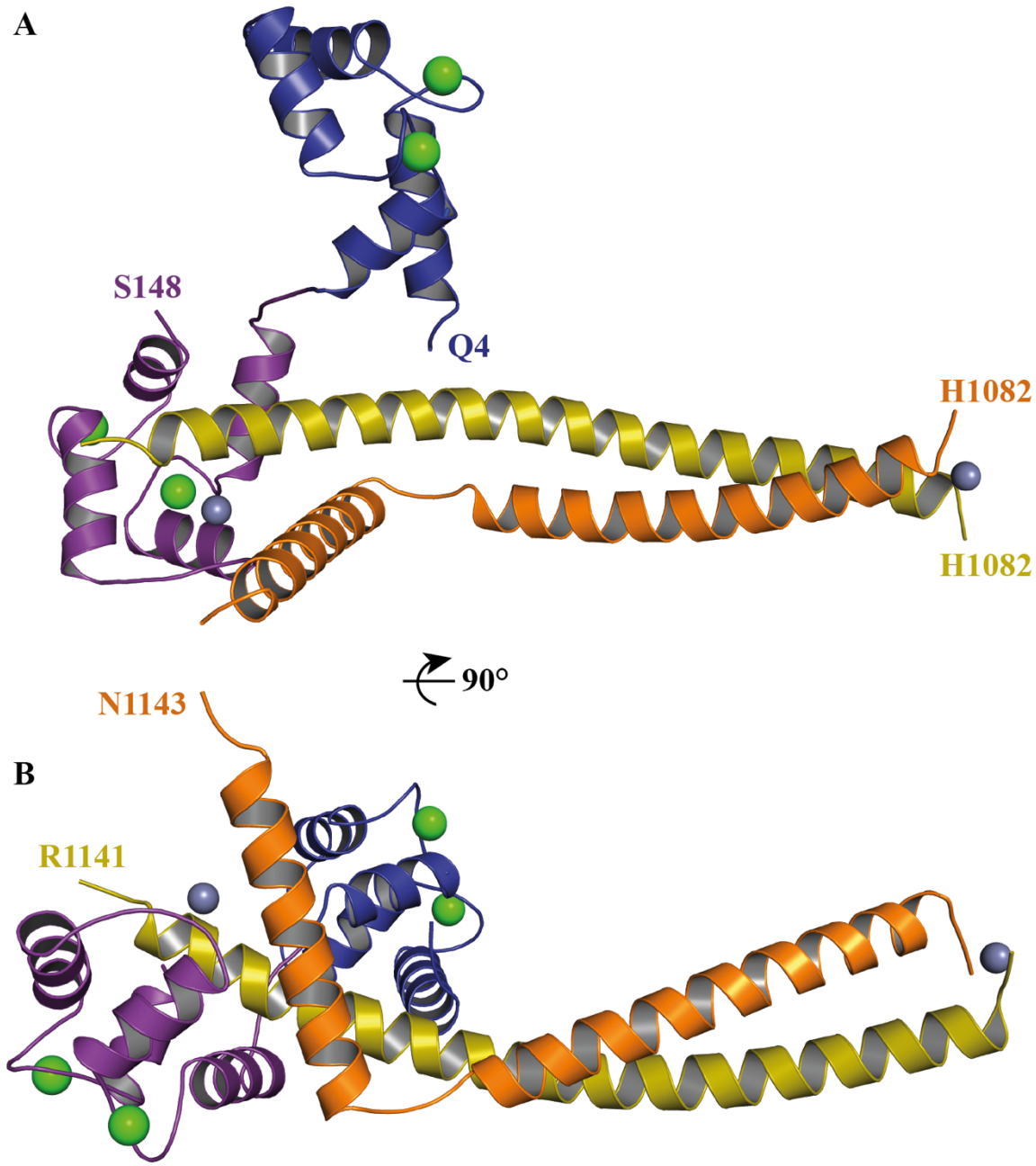


Figure 4.14 | Structure of the complex between Calmodulin and Cnn$^{H1082-L1148}$. The structure is shown in cartoon representation; zinc (grey) and calcium (green) ions are shown as spheres. Cnn$^{H1082-L1148}$ chain A is shown in orange, Cnn$^{H1082-L1148}$ chain B in yellow; the Calmodulin N-terminal domain is shown in blue, the C-terminal domain is shown in purple. N- and C-terminal residues on all chains are indicated. The structure is shown from two orientations rotated by 90° (A/B). Note the asymmetric unit consists of a dimer of Cnn$^{H1082-L1148}$ bound to one Calmodulin towards the C-terminus of one of its chains. Chain A of Cnn$^{H1082-L1148}$ shows a 90° bent, indicating great flexibility of the protein chain within this region. All four calcium-binding sites on Calmodulin are occupied, and two zinc ions are observed.

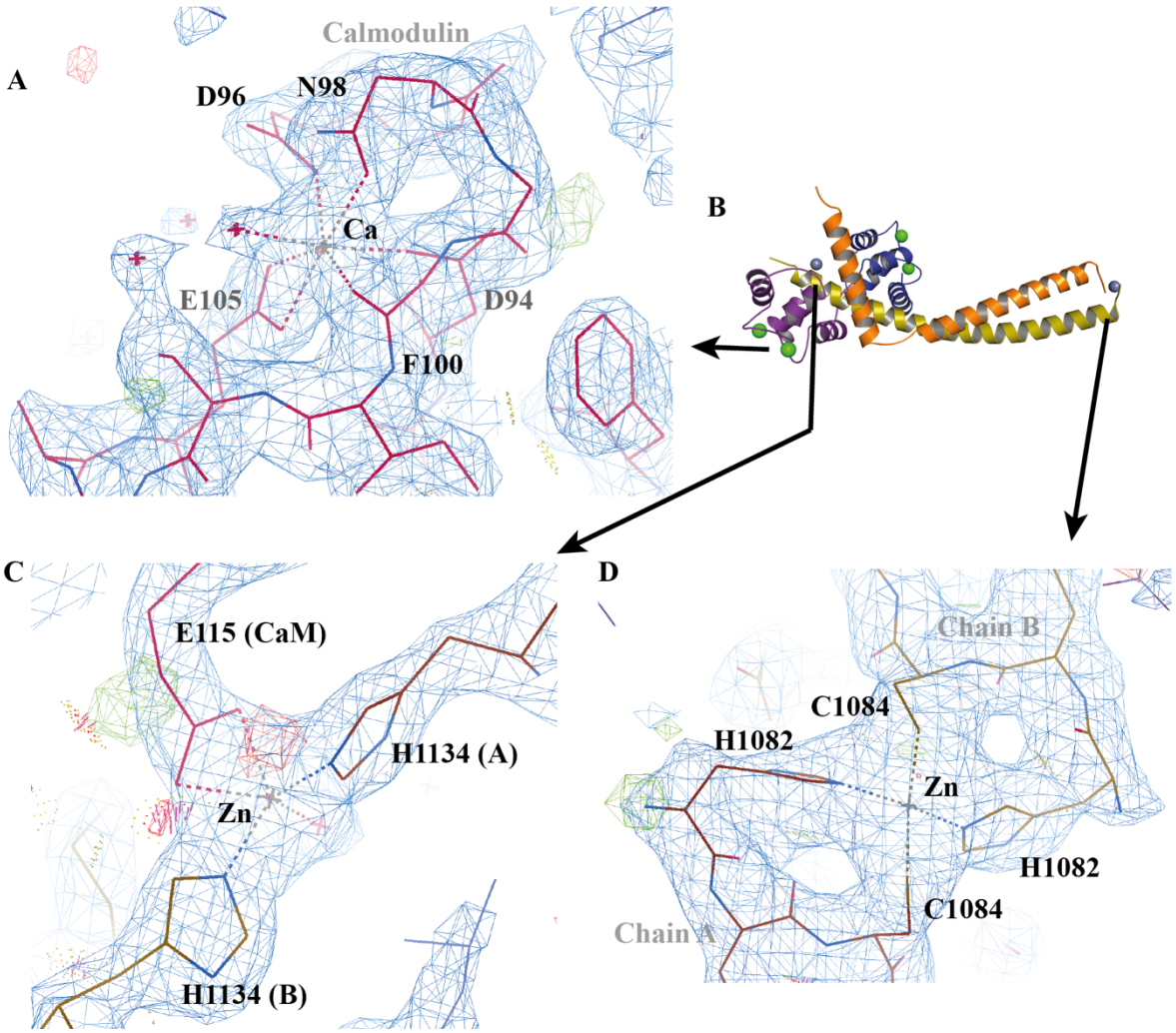


Figure 4.15 | Metal ion coordination in the Calmodulin:Cnn$^{H1082\text{-}L1148}$ structure. **A)** One of four calcium ions bound to Calmodulin. Note the ion is coordinated by the side chains of residues D94, D96, N98 and E105, the peptide bond carbonyl group of F100, and a water molecule. The electron density map is contoured at 1.5 σ. **B)** Calmodulin:Cnn$^{H1082\text{-}L1148}$ structure as seen in Fig. 4.14 B. **C)** Zinc coordination by side chains from residues of all three polypeptides found in the asymmetric unit, i.e. Calmodulin^{E115} and Cnn^{H1134} from both chains of Cnn$^{H1082\text{-}L1148}$; the fourth coordination position is probably occupied by a water molecule. The electron density map is contoured at 2.0 σ. **D)** Zinc coordination at the N-terminus of the two chains of Cnn$^{H1082\text{-}L1148}$ by H1082 and C1084 on both chains. The electron density map is contoured at 1.0 σ. In A, C and D, ions are shown as grey crosses, water molecules as red crosses.

In the asymmetric unit, the Calmodulin N-terminal domain does not appear to participate in the binding of the $Cnn^{H1082\text{-}L1148}$ dimer. However, observation of the crystal packing reveals that, strikingly, the Calmodulin N-terminal domain binds to $Cnn^{H1082\text{-}L1148}$ chain A from another dimer (Fig. 4.16). Specifically, the Calmodulin N-terminal domain binds to the region of $Cnn^{H1082\text{-}L1148}$ chain A C-terminal to the 90° kink. This means that the N- and C-terminal domains of one molecule of Calmodulin bind to different chains on distinct $Cnn^{H1082\text{-}L1148}$ dimers, and this pattern is repeated throughout the crystal. Calmodulin therefore does not only crosslink two $Cnn^{H1082\text{-}L1148}$ dimers but rather allows the assembly of a whole chain of crosslinked $Cnn^{H1082\text{-}L1148}$ dimers *in crystallo* (Fig. 4.16).

Fig. 4.17 B shows a view of a large section of the crystal along the short a-axis. What appears as blurry stacks of $Cnn^{H1082\text{-}L1148}$ dimers are in fact the long, Calmodulin-crosslinked chains described above, extending along the a-axis. Solvent channels can be observed crossing through the crystal along the same axis. The Calmodulin N-terminal domain is more exposed to these solvent channels than the Calmodulin C-terminal domain, which is more tightly packed in the crystal (Fig. 4.17 C). Indeed, analysis of the distribution of B-factor values across the protein shows that the Calmodulin N-terminal domain has higher B-factors on average than the Calmodulin C-terminal domain, suggesting higher flexibility, probably as a result of the increased exposure to the solvent channel (Fig. 4.17 A).

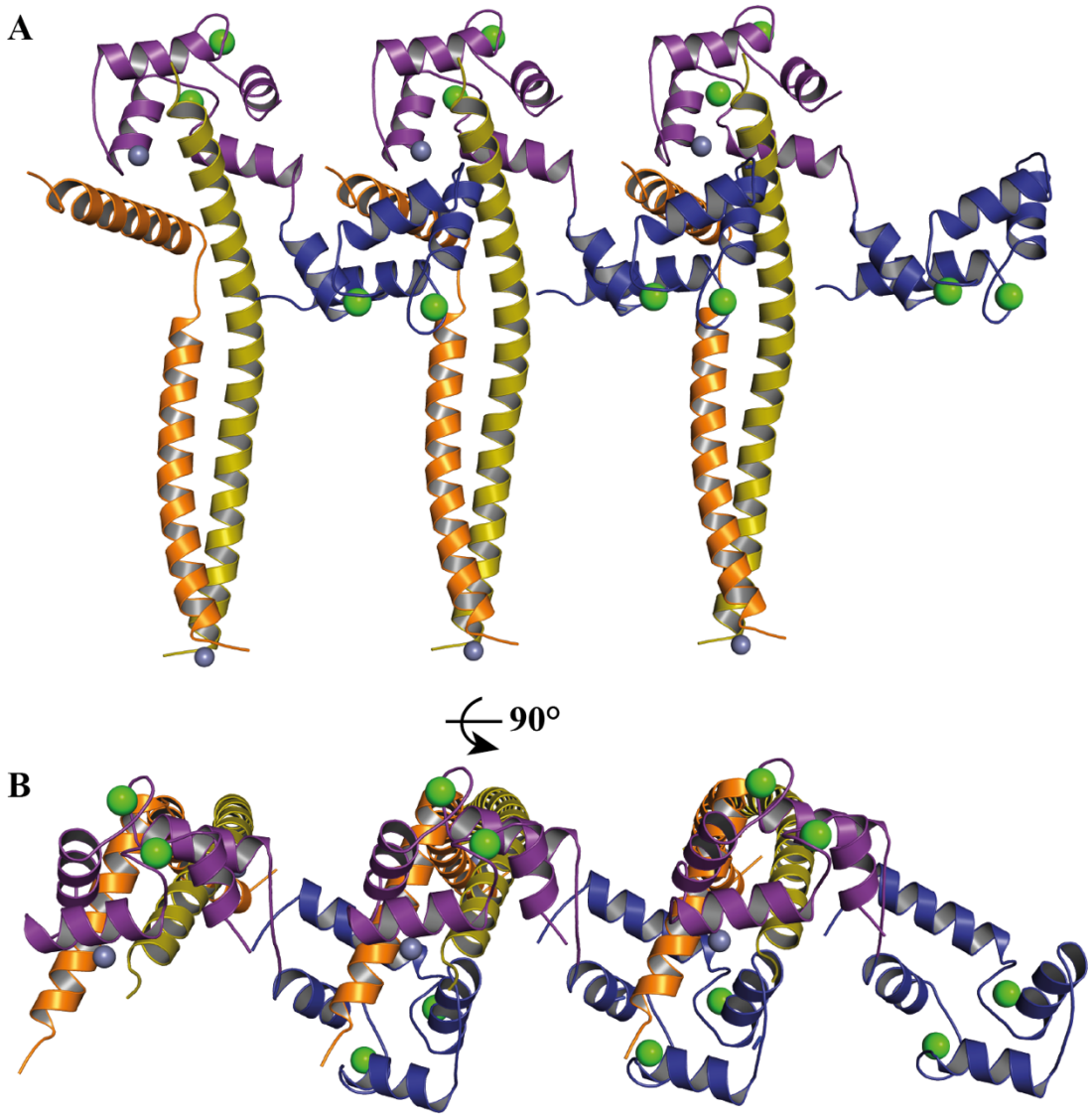


Figure 4.16 | Investigation of the crystal packing shows that Calmodulin crosslinks Cnn$^{H1082-L1148}$ dimers. **A)** The complex structure is shown as in Fig. 4.14 A but rotated by 90° around an axis orthogonal to the reader. Furthermore, two symmetry copies of the complex are shown. **B)** The same model as in A but rotated by 90° around the axis indicated. Note how the Calmodulin N-terminal domain (blue), which in the asymmetric unit does not appear to bind another protein (see Calmodulin N-terminal domain on the right) in the crystal packing always reaches over to another dimer of Cnn$^{H1082-L1148}$ and binds its chain A (orange). The C-terminal domain (purple) from another Calmodulin always binds Cnn$^{H1082-L1148}$ chain B (yellow). As Cnn$^{H1082-L1148}$ chains A and B form a stable dimer via zinc coordination at their N-termini, Calmodulin crosslinking chains A and B of Cnn$^{H1082-L1148}$ creates an unbroken string of protein in the crystal.

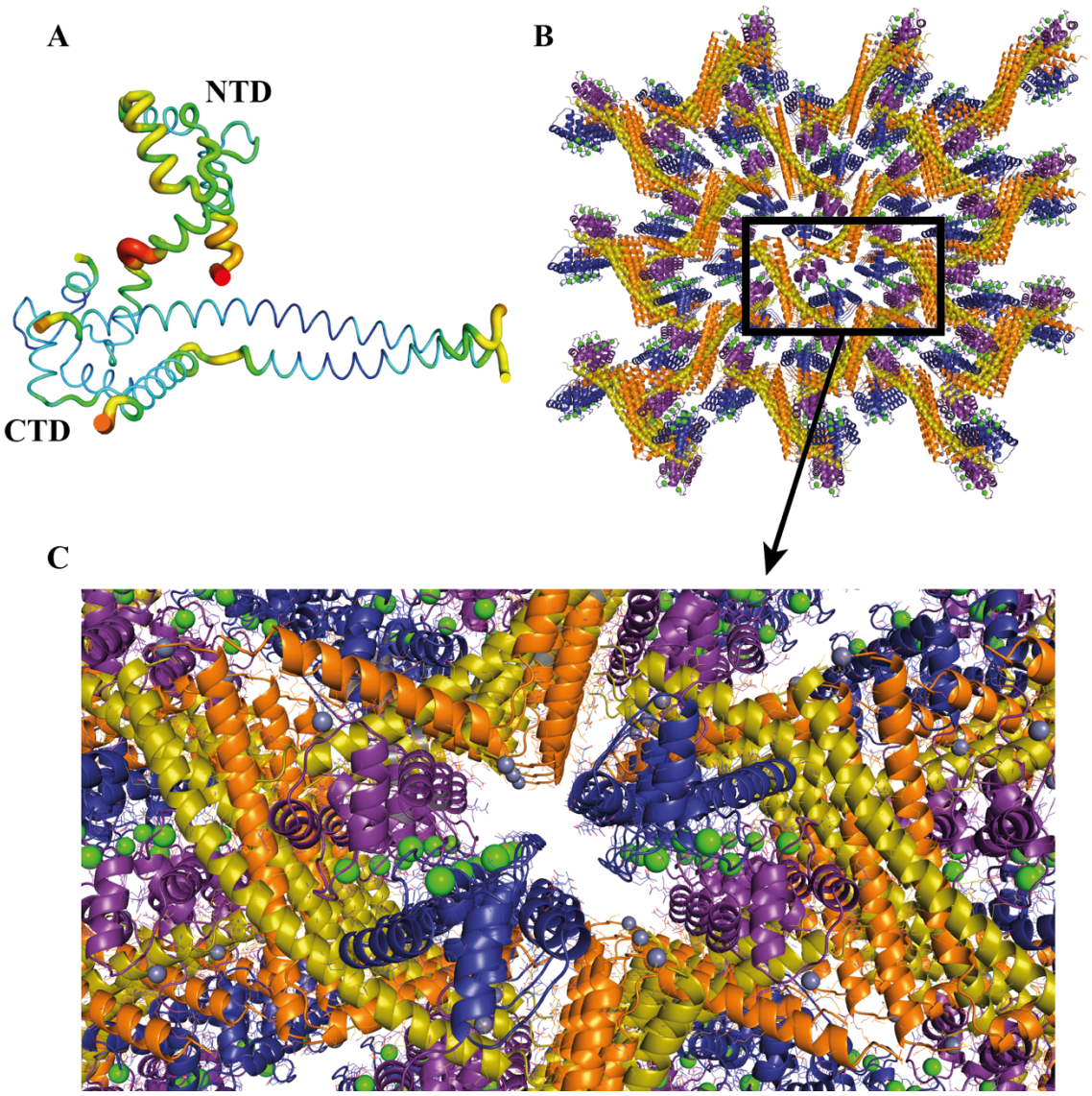


Figure 4.17 | The Calmodulin N-terminal domain shows increased flexibility. **A)** B-factor putty model of Calmodulin:Cnn$^{H1082\text{-}L1148}$ generated via PyMOL and shown in the same orientation as in Fig. 4.14 A. Residues with a lower B-factor, i.e. little flexibility, are highlighted as a thin blue tube; residues with a higher B-factor, i.e. greater flexibility, are highlighted as a thick tube in green-yellow-red at increasing flexibility. Unsurprisingly, flexibility is increased at the termini of all three polypeptide chains, the 90° kink in Cnn$^{H1082\text{-}L1148}$ chain A and the linker between the N- and C-terminal domains of Calmodulin. However, note that the Calmodulin N-terminal domain (NTD) also shows a greater overall flexibility than the Calmodulin C-terminal domain (CTD). **B)** The Calmodulin:Cnn$^{H1082\text{-}L1148}$ complex structure is shown in the same representation as in Fig. 4.14, but all symmetry copies in a 250 Å radius are shown. **C)** Detail showing one of the solvent channels crossing through the crystal; side chains are shown as lines. Note the Calmodulin N-terminal domain (blue) is exposed to the solvent channel, while the Calmodulin C-terminal domain (purple) is not, explaining the increased flexibility of the N-terminal domain.

The binding site of Calmodulin on $Cnn^{H1082\text{-}L1148}$ roughly corresponds to where Calmodulin is thought to bind human CEP215 (Fig. 4.1). Interestingly, however, the N- and C-terminal domains of Calmodulin do not bind to exactly the same sites on the two chains. While the C-terminal domain binds to residues Q1129-R1138 on Cnn, the binding site of the N-terminal domain is shifted forward by three amino acids to K1132-R1141. This means that the binding sites of the two domains are separated by almost but not quite one turn of a helix (Fig. 4.18). This suggests that binding of Calmodulin to the Cnn C-terminus is not sequence specific but rather a result of the general characteristics of this region within Cnn. Indeed, the site around which Calmodulin binds is clearly amphipathic, with one face of the helix predominantly having hydrophobic residues such as V1136 and L1137, while the other face of the helix has a number of positively charged residues such as R1138 and R1141 (Fig. 4.18). Calmodulin is known to preferentially bind such basic amphipathic α-helices [254].

As was shown in Chapter 3, binding of $Cnn^{H1082\text{-}L1148}$ to $Cnn^{Q403\text{-}H608}$, and therefore network assembly *in vitro*, depends on residues I1126, T1133 and L1137 of $Cnn^{H1082\text{-}L1148}$ (Fig. 3.8). This binding site is partially overlapping with the Calmodulin binding sites described above. Indeed, when the Calmodulin:$Cnn^{H1082\text{-}L1148}$ structure solved in this chapter is overlaid with the previously solved structure of the $Cnn^{D490\text{-}K544}$:$Cnn^{H1082\text{-}L1148}$ complex (PDB code: 5MVW, [339]), not only does the large conformational change required of $Cnn^{H1082\text{-}L1148}$ to bind one or the other protein become apparent, but it also becomes evident that concomitant binding of both $Cnn^{D490\text{-}K544}$ and Calmodulin to $Cnn^{H1082\text{-}L1148}$ would result in a steric clash (Fig. 4.19). Binding of $Cnn^{H1082\text{-}L1148}$ to $Cnn^{D490\text{-}K544}$ or Calmodulin is therefore mutually exclusive.

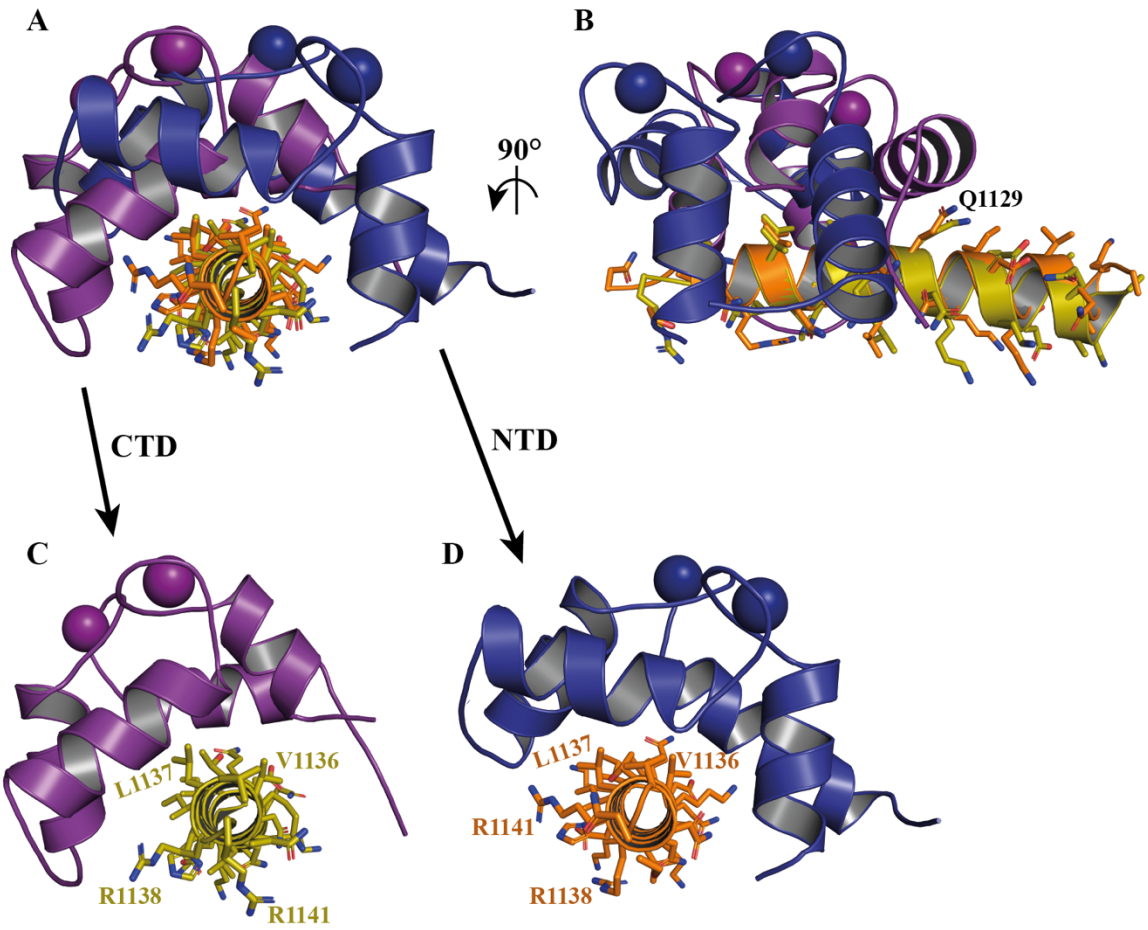


Figure 4.18 | The Calmodulin N- and C-terminal domains bind to the Cnn C-terminus in a different register. **A/B)** Overlaid model of the Calmodulin N-terminal domain (NTD, blue) and the Calmodulin C-terminal domain (CTD, purple) binding to the respective C-termini of chains A and B of $Cnn^{H1082\text{-}L1148}$, and viewed from two angles rotated by 90°. The two Cnn chains are overlaid. **C)** The Calmodulin CTD bound to $Cnn^{H1082\text{-}L1148}$ chain B (yellow) shown in the same orientation as in A. **D)** The Calmodulin NTD bound to $Cnn^{H1082\text{-}L1148}$ chain A (orange) shown in the same orientation as in A. Calmodulin and $Cnn^{H1082\text{-}L1148}$ are shown in cartoon representation, with side chains on $Cnn^{H1082\text{-}L1148}$ shown as sticks. Calcium ions are shown as spheres in the same colour as the Calmodulin domain they are bound to. Residues of interest on $Cnn^{H1082\text{-}L1148}$ are indicated. The Calmodulin NTD binds residues K1132-R1141, while the Calmodulin CTD binds residues Q1129-R1138 on Cnn, i.e. a difference in register of three amino acids. Note the amphipathic nature of the Cnn C-terminus, with both Calmodulin domains mostly burying the hydrophobic side (e.g. residues V1136 or L1137) rather than the positively charged side (e.g. residues R1138 or R1141) in their binding clefts.

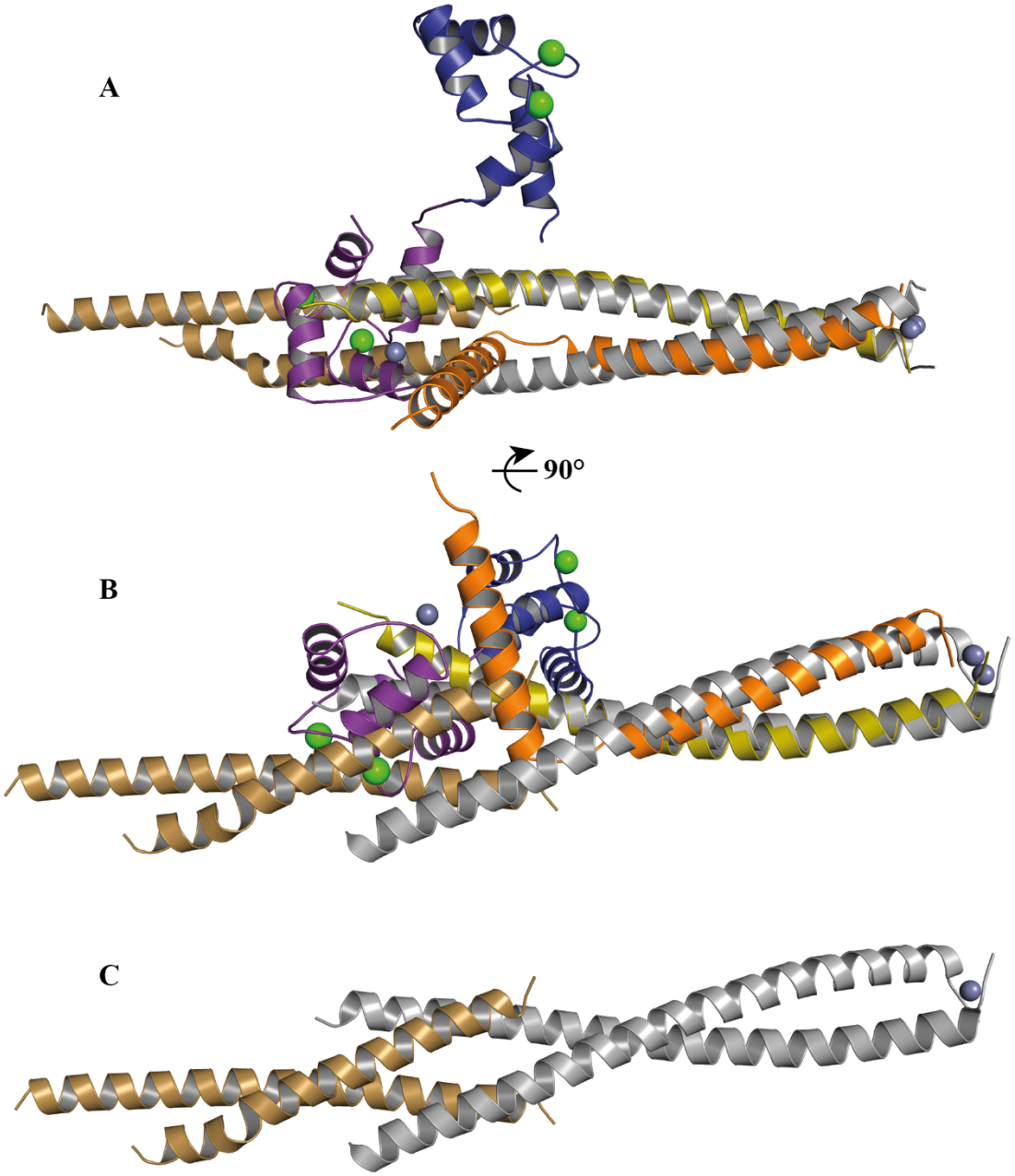


Figure 4.19 | Binding of Cnn$^{H1082-L1148}$ to Calmodulin or Cnn$^{D490-K544}$ is mutually exclusive. **A/B)** The structure of the Calmodulin:Cnn$^{H1082-L1148}$ complex shown in the same style and orientations as in Fig. 4.14, but in addition overlaid with the structure of the Cnn$^{D490-K544}$:Cnn$^{H1082-L1148}$ complex (PDB code: 5MVW, [339]). The Cnn$^{D490-K544}$ dimer is in brown, the Cnn$^{H1082-L1148}$ dimer from 5MVW is in grey; both are shown in cartoon representation. Note that Cnn$^{D490-K544}$ and Calmodulin bind to the same site on Cnn$^{H1082-L1148}$, resulting in a steric clash when the structures for both complexes are overlaid. This suggests that binding is mutually exclusive. Further note that, while Cnn$^{H1082-L1148}$ chain B (yellow) adopts a similar conformation in both complex structures, Cnn$^{H1082-L1148}$ chain A (orange) shows a large conformational change, with the observed 90° kink suggesting great flexibility within the polypeptide chain. **C)** The Cnn$^{D490-K544}$:Cnn$^{H1082-L1148}$ complex structure alone (PDB code: 5MVW, [339]) shown in the same orientation as in B.

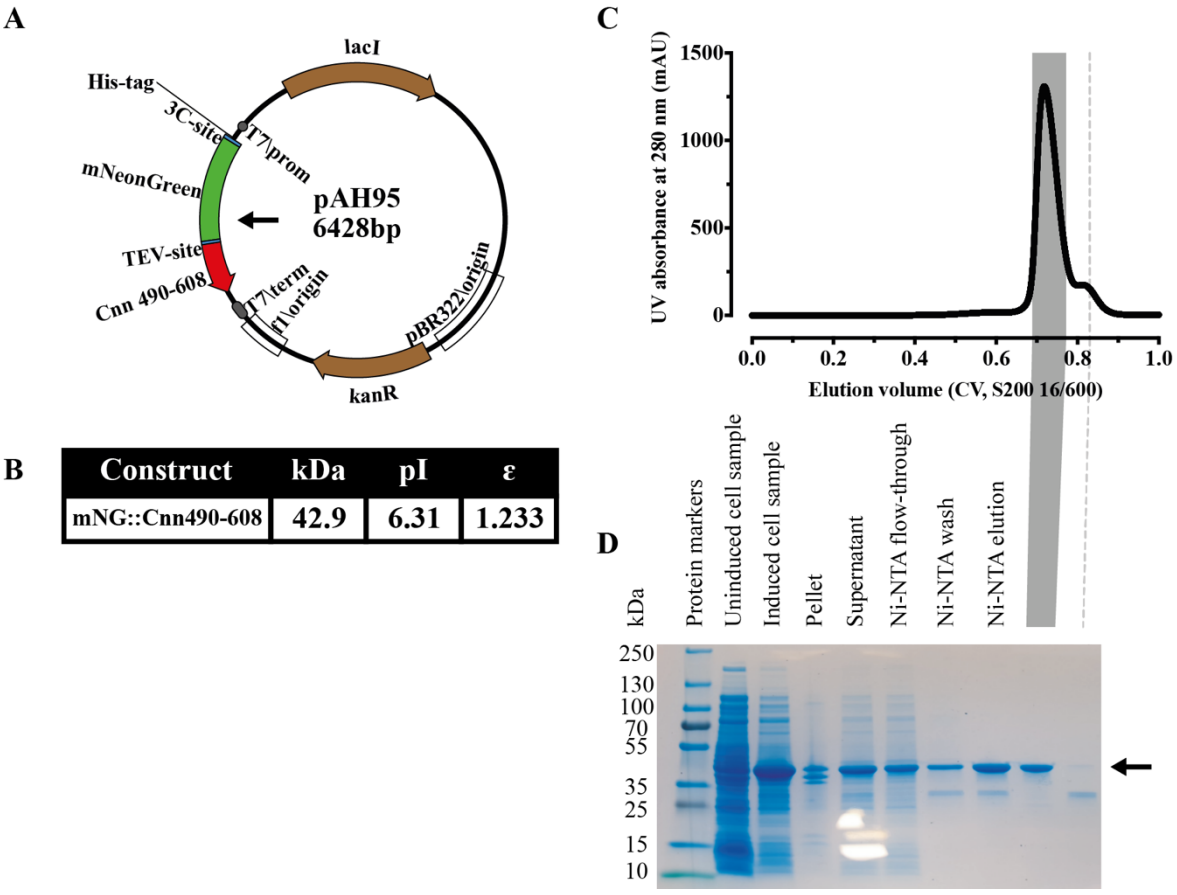
4.7 Calmodulin prevents network formation of Cnn constructs *in vitro*

The crystal structure of the Calmodulin:Cnn$^{H1082\text{-}L1148}$ complex allowed two competing hypotheses to be formulated for the role of Calmodulin: one in which Calmodulin aids network assembly by crosslinking Cnn$^{H1082\text{-}L1148}$, another in which it prevents network assembly by inhibiting binding of Cnn$^{H1082\text{-}L1148}$ to the zipper-like motif.

In order to test these hypotheses, a new construct was cloned and purified by Joseph Fisher (Part II student, Lea laboratory) under the direct supervision of the author (Fig. 4.20). This construct, mNG-Cnn$^{D490\text{-}H608}$, had the same N-terminus as the zipper-like motif from the Cnn$^{D490\text{-}K544}$:Cnn$^{H1082\text{-}L1148}$ structure (Fig. 3.1) but had an extended C-terminus matching the longer Cnn$^{Q403\text{-}H608}$-mGFP construct, which was shown in Chapter 3 to assemble into networks in the presence of Cnn$^{H1082\text{-}L1148}$ (Fig. 3.8). It was reasoned that the region N-terminal to D490 was probably redundant for network assembly, as all potential phosphorylation sites which might have a role in network assembly are found C-terminal to it (Fig. 1.8 A). Indeed, the region Q403-D489 within Cnn appears to be largely dispensable for Cnn network formation *in vivo* [personal communication, Dr Anna Caballe, Raff laboratory].

Purification of mNG-Cnn$^{D490\text{-}H608}$ was carried out by Joseph Fisher according to the usual protocol from a 2 L LB culture of *E. coli* BL21 (DE3) cells. After Ni-affinity chromatography, the eluate could not be concentrated further than ~ 10 ml due to what appeared to be a very high concentration of protein. A 5 ml fraction was therefore taken and applied to a S200 16/600 column pre-equilibrated in standard TBS; the rest was discarded.

The elution profile had a single clear peak, which was confirmed to contain mNG-Cnn$^{D490-H608}$ via SDS-PAGE (Fig. 4.20). The peak was pooled (CV = 0.69-0.77), concentrated to 15.8 mg/ml and flash-frozen in 50 µl aliquots, resulting in a final yield of ~ 9.1 mg/L.



Construct	kDa	pI	ε
mNG::Cnn490-608	42.9	6.31	1.233

Figure 4.20 | Purification of N-terminally mNG-tagged Cnn$^{D490-H608}$. **A)** Overview of plasmid pAH95, which encodes Cnn$^{D490-H608}$ with an N-terminal, TEV-cleavable, His-3C-mNG-domain. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged protein. **C)** SEC (S200 16/600) elution profile of mNG-Cnn$^{D490-H608}$ after Ni-affinity chromatography, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.69-0.77) indicates the fractions containing mNG-Cnn$^{D490-H608}$ as judged by their green colour; these fractions were pooled after SEC and run on the SDS-PAGE gel in D. A further fraction run on the gel in D is indicated by a dashed grey line. SEC was carried out using standard TBS. **D)** 4-20 % reducing SDS-PAGE gel showing successive purification steps; the band corresponding to mNG-Cnn$^{D490-H608}$ is highlighted with a black arrow. The plasmid was cloned and protein was purified by Joseph Fisher under the author's direct supervision.

It was first tested whether mNG-Cnn$^{D490\text{-}H608}$ also assembled into networks upon binding to Cnn$^{G1077\text{-}L1148}$. To that end, mNG-Cnn$^{D490\text{-}H608}$ (at 200 μM) was mixed with Cnn$^{G1077\text{-}L1148}$ (at 100 μM), and the mixture separated on a S200 10/300 column for analytical SEC using standard TBS supplemented with 10 mM $CaCl_2$ as buffer. Indeed, a void peak was observed that, according to SDS-PAGE analysis, contained both mNG-Cnn$^{D490\text{-}H608}$ and Cnn$^{G1077\text{-}L1148}$. (Fig. 4.21). Residues N-terminal to D490 therefore appear to be redundant for network assembly *in vitro*. Note that as Cnn$^{G1077\text{-}L1148}$ was at a limiting concentration, not all of mNG-Cnn$^{D490\text{-}H608}$ was shifted to the void.

Next, it was investigated via analytical SEC whether Calmodulin could bind to mNG-Cnn$^{D490\text{-}H608}$. However, when the two constructs were respectively mixed at 250 μM and 200 μM, using the same buffer conditions as above, they were not observed to co-elute but rather eluted in two separate peaks (Fig. 4.21). This suggests that Calmodulin cannot bind mNG-Cnn$^{D490\text{-}H608}$ under these conditions.

Finally, an experiment was carried out in which all three constructs, i.e. Calmodulin (at 250 μM), mNG-Cnn$^{D490\text{-}H608}$ (at 200 μM) and Cnn$^{G1077\text{-}L1148}$ (at 100 μM) were mixed and analysed via analytical SEC using the same buffer as above. Strikingly, in the presence of Calmodulin, the mNG-Cnn$^{D490\text{-}H608}$:Cnn$^{G1077\text{-}L1148}$ void peak disappeared, and instead the mNG-Cnn$^{D490\text{-}H608}$ peak had a pronounced, earlier-eluting shoulder. Surprisingly, this early-eluting shoulder contained all three constructs, suggesting that they formed a complex (Fig. 4.21). Calmodulin can therefore inhibit network assembly of Cnn constructs under the conditions tested, and this appears to occur via formation of a tertiary complex between Calmodulin, mNG-Cnn$^{D490\text{-}H608}$ and Cnn$^{G1077\text{-}L1148}$.

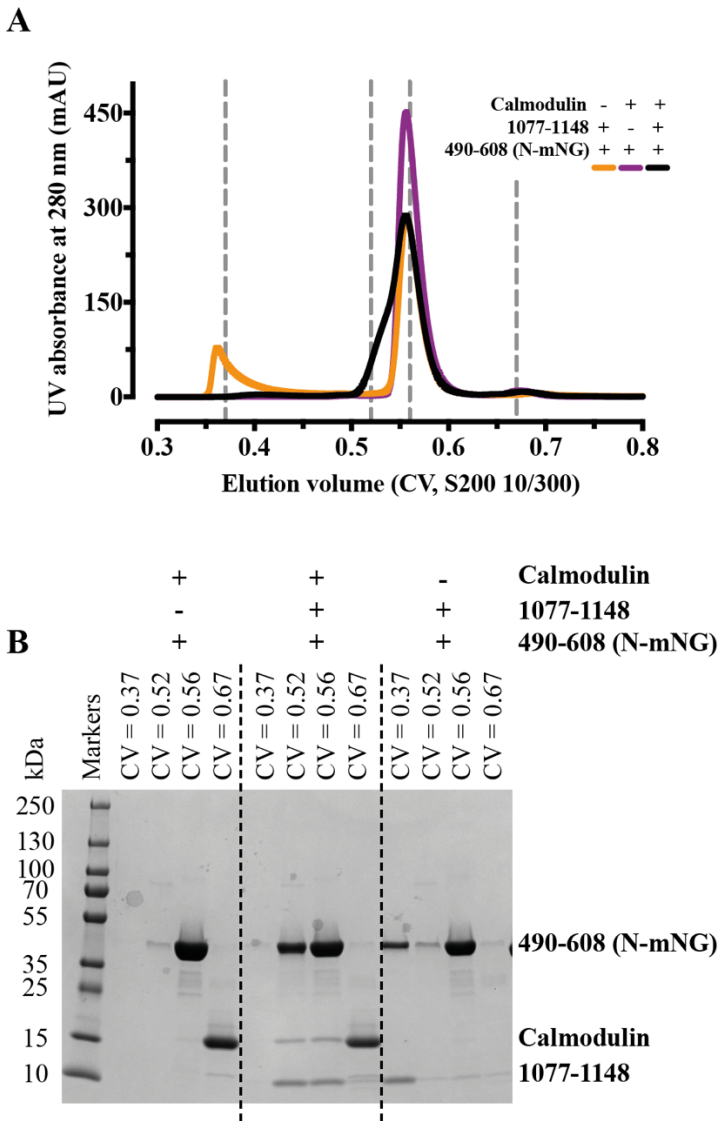


Figure 4.21 | Calmodulin inhibits network assembly of Cnn constructs *in vitro*. A) Analytical SEC experiment with mNG-Cnn$^{D490\text{-}H608}$ (200 µM), Calmodulin (250 µM) and Cnn$^{G1077\text{-}L1148}$ (100 µM). The orange trace shows the experiment where mNG-Cnn$^{D490\text{-}H608}$ and Cnn$^{G1077\text{-}L1148}$ are mixed; the purple trace shows mNG-Cnn$^{D490\text{-}H608}$ and Calmodulin; the black trace shows all three together. Dashed grey lines indicate fractions taken and run on the SDS-PAGE gel in B; the same fractions were taken for each experiment. UV absorbance at 280 nm (in mAU) is plotted on the y-axis, column elution volume (CV, 24 ml) of column S200 10/300 increase plotted on the x-axis. Standard TBS supplemented with 10 mM $CaCl_2$ was used as buffer. **B)** 4-20 % reducing SDS-PAGE gel showing fractions taken during the three experiments. The bands correspond to the constructs indicated. Note that when mNG-Cnn$^{D490\text{-}H608}$ and Cnn$^{G1077\text{-}L1148}$ are mixed (orange, right side of gel), a void peak appears that contains both constructs, i.e. corresponding to mNG-Cnn$^{D490\text{-}H608}$:Cnn$^{G1077\text{-}L1148}$ networks (CV = 0.37). As Cnn$^{G1077\text{-}L1148}$ is at a limiting concentration, not all of mNG-Cnn$^{D490\text{-}H608}$ is shifted to the void. Strikingly, when Calmodulin is also added to the mix (black, centre of gel), the void peak disappears, and instead an early-eluting shoulder appears that contains all three constructs (CV = 0.52). This suggests that Calmodulin inhibits network assembly via formation of a tertiary complex. Calmodulin and mNG-Cnn$^{D490\text{-}H608}$ are not observed to co-elute, and therefore do not appear to bind to each other, in the absence of Cnn$^{G1077\text{-}L1148}$ (purple, left side of gel).

4.8 Discussion

In this chapter, the crystal structure of the complex between Calmodulin and Cnn$^{H1082\text{-}L1148}$ is solved, showing that Calmodulin binds to the region around Q1129-R1141 in a manner that does not appear to be strictly sequence specific. *In crystallo*, Calmodulin crosslinks successive dimers of Cnn$^{H1082\text{-}L1148}$. In addition, the region within Cnn required to allow network formation *in vitro* upon binding to the Cnn C-terminus is further narrowed down to D490-H608. Calmodulin is shown to prevent network formation of this minimal construct, probably via the formation of a tertiary construct consisting of Calmodulin, mNG-Cnn$^{D490\text{-}H608}$ and Cnn$^{G1077\text{-}L1148}$. Furthermore, Calmodulin is shown to bind to Cnn$^{S136\text{-}S205}$ *in vitro*, although this interaction appears to be prevented in context of the larger Cnn$^{M1\text{-}K360}$ construct. The structure of Cnn$^{S136\text{-}S205}$ is solved, showing that it forms a typical dimeric coiled-coil towards its N-terminus.

Although attempts to crystallise the structure of the complex between Cnn$^{S136\text{-}S205}$ and Calmodulin were unsuccessful, the structure of Cnn$^{S136\text{-}S205}$ presented here is the first atomic resolution structure of this conserved region in Cnn. While it does not provide any functional insights *per se*, in the future it can be used to guide our understanding of this important region of the protein. Specifically, an investigation of the conservation within this region can be informative. *Drosophila* Cnn$^{T137\text{-}I172}$ and human CEP215$^{T96\text{-}V131}$ share a high degree of sequence similarity. However, C-terminal to that, the evolutionary conservation between the two proteins breaks down and they diverge significantly (Fig. 4.3). This corresponds well with the region within the structure of Cnn$^{S136\text{-}S205}$ presented in this chapter which forms a canonical coiled-coil, i.e. residues L140-K161. This suggests that formation of this coiled-coil might be evolutionary conserved.

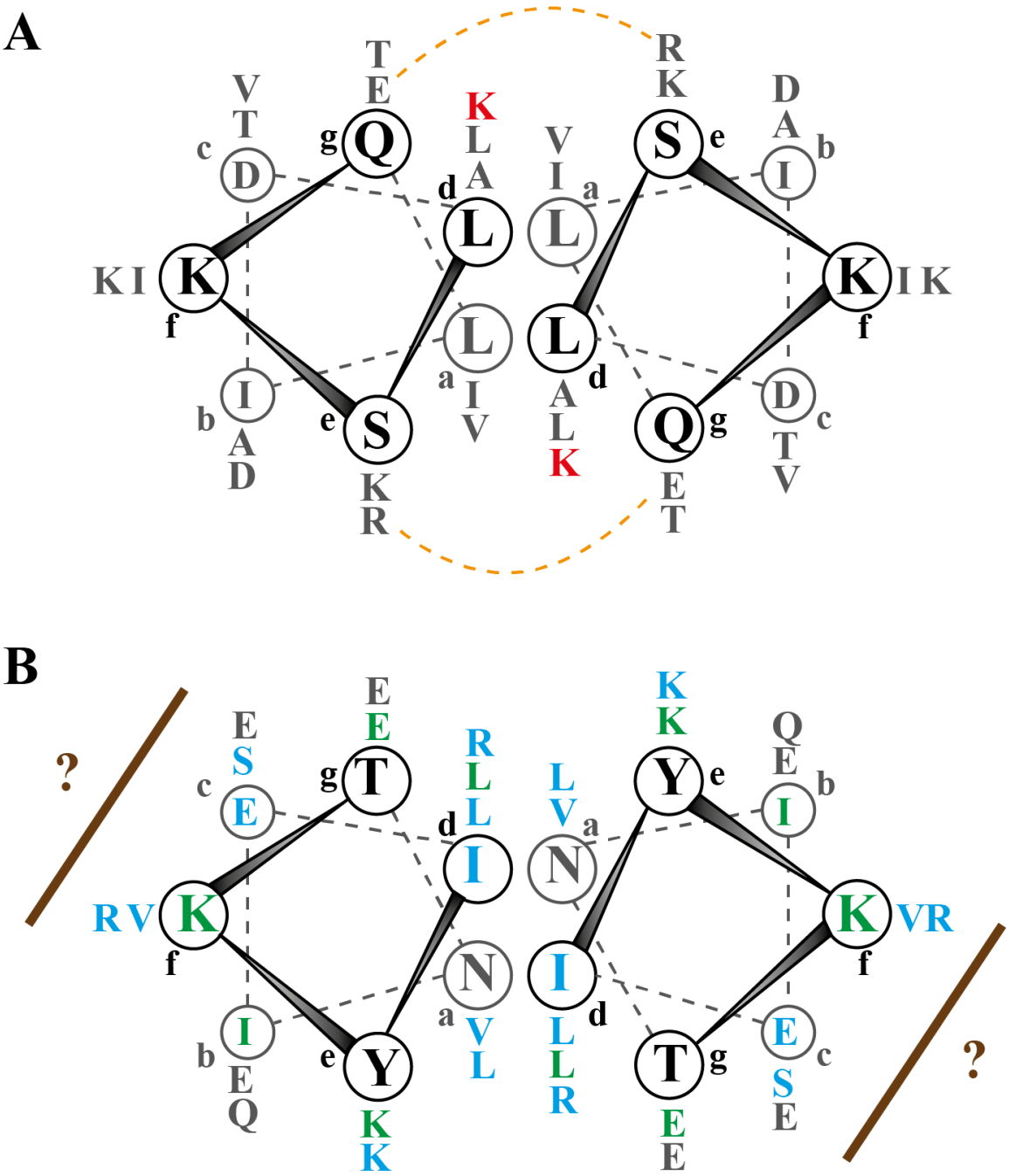


Figure 4.22 | Helical wheel representations of Cnn$^{L140\text{-}K161}$ and CEP215$^{I99\text{-}R120}$. **A)** Helical wheel representation of Cnn$^{L140\text{-}K161}$ based on the solved structure of Cnn$^{S136\text{-}S205}$. The salt bridges between E150 and R155 are shown as orange dashed lines. K161, the residue that breaks the coiled-coil, is shown in red. **B)** Sequence of CEP215$^{I99\text{-}R120}$ superimposed on the helical wheel representation of Cnn$^{L140\text{-}K161}$. Identical residues are shown in green, similar residues in blue. Note conservation is clustered at positions a and d (hence possible conservation of coiled-coil formation) and on the side of the helix formed by positions c and f (denoted by brown lines and question marks; possibly suggesting a surface involved in a conserved binding interaction).

Indeed, generation of a helical wheel plot for CEP215$^{I99\text{-}R120}$ using the solved structure of the Cnn$^{L140\text{-}K161}$ coiled-coil as a model suggests that at heptad repeat positions a and d, which are crucial for the formation of a coiled-coil, almost all residues are similar between Cnn and CEP215. Furthermore, one side of the helix facing away from the other chain also shows a high degree of conservation, possibly suggesting that this side of the helix might be involved in binding interactions (Fig. 4.22). In contrast, the C-terminal end of Cnn$^{S136\text{-}S205}$ is not highly conserved and in the structure is seen to splay open, packing against another crystallographic copy of itself. It cannot be excluded that some residues within this region have an important role. Indeed, it has already been mentioned that phosphorylation of CEP215 on Ser140 appears to have a role in enhancing γ-tubulin association in cultured human cells [226]. However, it is reasonable to assume that the region is probably not structurally conserved as a whole, and the specific packing observed *in crystallo* might be artefactual.

The observation that Calmodulin can bind Cnn$^{S136\text{-}S205}$ but cannot bind Cnn$^{M1\text{-}K360}$ under similar conditions suggests that the Calmodulin binding site within Cnn$^{S136\text{-}S205}$ is inaccessible in the context of this larger construct. Importantly, this means that under these conditions, Calmodulin also cannot bind the predicted leucine zipper motif within CM1, the other predicted interaction site found within Cnn$^{M1\text{-}K360}$ immediately N-terminal to the site studied in this chapter (Fig. 4.3). The inability of Calmodulin to bind these two sites under the conditions tested might be due to the presence of a regulatory mechanism that only allows access to these sites upon a conformational change within Cnn$^{M1\text{-}K360}$ induced by e.g. post-translational modification. As CM1 has a crucial role in γ-tubulin recruitment (see Introduction), it is interesting to speculate that the same regulatory mechanism might be involved in controlling association with γ-tubulin *in vivo*. Indeed, it has been hypothesised

that the extreme N-terminus might have a role in regulating γ-tubulin access to CM1 in *Drosophila* [221].

The striking similarity between the crystallographic packing of two $Cnn^{S136-S205}$ dimers and the $Cnn^{H1082-L1148}$ dimer from the previously solved $Cnn^{H1082-L1148}$:$Cnn^{D490-K544}$ 2:2 complex is shared with other structures of coiled-coils and could therefore reflect a propensity of such structural motifs to crystallise in this way. However, the structural similarity could also suggest that, under the right conditions, CM1 and the region C-terminal to it might interact with CM2. Indeed, if the regulatory mechanism controlling CM1 accessibility proposed above is correct, it is not surprising that the Cnn C-terminus is unable to bind $Cnn^{S136-S205}$ in the context of the larger $Cnn^{M1-K360}$ construct (Fig. 3.14) under similar conditions in which Calmodulin cannot bind CM1 within $Cnn^{M1-K360}$ (Fig. 4.7). It therefore remains to be tested whether CM2 can bind this region under different conditions where CM1 is accessible. What role such an interaction might have is unclear, but it is interesting to speculate that the two conserved motifs mediating the two key functions of Cnn, i.e. network assembly (CM2) and microtubule nucleation (CM1), might interact and thereby perhaps coordinate their functions.

The structure of the complex between Calmodulin and $Cnn^{H1082-L1148}$ solved in this chapter is reminiscent of the known structure between yeast Calmodulin and a Spc110 (amino acids 891-944) peptide (Fig. 4.2). In both cases, Calmodulin crosslinks target peptides. Indeed, this mode of binding is not uncommon for Calmodulin (ref [273]). However, in all cases the author is aware of, this results in formation of a closed complex. In the case of Spc110 (amino acids 891-944) this means that Calmodulin bridges two monomeric fragments of the protein [276]. The Calmodulin:$Cnn^{H1082-L1148}$ structure presented here therefore is unique, as

Calmodulin bridges neighbouring $Cnn^{H1082\text{-}L1148}$ peptides which themselves are dimeric, leading to the assembly of an interlocking chain of protein. Indeed, such chains of protein would normally not be expected to form into an orderly crystal, but rather get entangled and precipitate. This is probably reflected in the fact that hundreds of crystals had to be grown and dozens sent for data collection, with only a single crystal consisting of the Calmodulin:$Cnn^{H1082\text{-}L1148}$ complex, likely stabilised by the presence of zinc in the crystallisation condition.

In yeast, Calmodulin is thought to enhance the mechanical stability of the spindle pole body by bridging neighbouring Spc110 molecules as seen in the crystal structure [276, 277]. Indeed, the Calmodulin:$Cnn^{H1082\text{-}L1148}$ structure presented here offers the intriguing possibility that such a function of Calmodulin might be conserved in animal centrosomes. The centrosomes of dividing cultured human cells are highly enriched in calcium [260] and the pericentriolar material of mammalian centrosomes shows a sensitivity to calcium [261, 262] . In this chapter, the interaction between Calmodulin and $Cnn^{H1082\text{-}L1148}$ is confirmed to be calcium-sensitive. While it appears to be clear that a main function of Calmodulin during mitosis involves helping Abnormal spindle protein during centrosome attachment to the spindle pole and spindle focussing [166, 263], this does not exclude the possibility that some Calmodulin might be involved in perhaps fine-tuning the mechanical stability of the Cnn network. This might be especially important in the outer regions of the network, where Cnn becomes dephosphorylated, and the Cnn C-terminus might therefore not bind to the zipper-like motif as readily anymore [339].

However, perhaps unrealistically high local concentrations of Calmodulin would be required for Calmodulin to crosslink the Cnn C-terminus, as crosslinking is observed *in crystallo*, but

in solution even at millimolar concentrations of protein, complexes are observed that have a mass that is only slightly higher than that expected for one Calmodulin bound to one dimer of $Cnn^{H1082\text{-}L1148}$ (Fig. 4.11 C). Furthermore, Chapter 3 clearly shows that the Cnn C-terminus allows Cnn network assembly by binding to the zipper-like motif. As is shown in this chapter, binding of the Cnn C-terminus to the zipper-like motif and to Calmodulin must be mutually exclusive due to the steric clash resulting from the overlap in binding sites. A role for Calmodulin as a chaperone that prevents undue network assembly in the cytosol therefore appears more likely based on the structural information.

Indeed, this chapter shows that Calmodulin does indeed prevent network assembly of Cnn constructs *in vitro*. Unexpectedly, however, Calmodulin does not do so by preventing the Cnn C-terminus from binding to mNG-$Cnn^{D490\text{-}H608}$, but it appears to prevent network assembly via formation of a tertiary complex with these two constructs. One possibility is that in this tertiary complex, the Cnn C-terminus binds to mNG-$Cnn^{D490\text{-}H608}$ with only one of its chains, while the other one is displaced and binds one domain of Calmodulin. Considering the high degree of structural flexibility that $Cnn^{H1082\text{-}L1148}$ exhibits, this might be plausible. Indeed, this flexibility in adopting different conformations is probably the reason why the Cnn C-terminus can act as a multi-functional site that binds to a range of targets [159, 272]. However, it appears unclear how this would prevent network formation.

An alternative explanation might therefore be that Calmodulin binds to the site within mNG-$Cnn^{D490\text{-}H608}$ that is responsible for network formation and thereby blocks it from doing so, but only once it has undergone the previously proposed conformational change induced or stabilised by the Cnn C-terminus (Fig. 3.15). This would explain why Calmodulin cannot bind mNG-$Cnn^{D490\text{-}H608}$ in the absence of the Cnn C-terminus, which in this model would

first have to bind mNG-Cnn$^{D490\text{-}H608}$ before Calmodulin can then bind mNG-Cnn$^{D490\text{-}H608}$ as well on a different site. This might provide a mechanism for the cell to prevent accidental Cnn network formation due to association of the Cnn C-terminus with the zipper-like motif in the cytosol, in the absence of phosphorylation. It therefore needs to be tested whether Calmodulin fails to prevent network formation of Cnn constructs *in vitro* in the presence of phosphorylation, thereby mimicking the centrosomal environment, where network assembly is desirable.

Most importantly, however, further structural elucidation of Cnn$^{D490\text{-}H608}$, i.e. the region within Cnn that allows network formation, is required in order to understand the proposed conformational change within this region that underpins the findings of this and the previous chapter.

Chapter 5 - A Cnn$^{D490\text{-}N579}$ fold that is incompatible with phosphorylation of S567 and binding of the Cnn C-terminus

5.1 Introduction

Polo kinase phosphorylates Cnn at the centrosome during mitosis [167, 196]. The conserved kinase phosphorylates N/E/D-x-S/T consensus motifs, with hydrophobic residues often located at positions -3 and +1 as counted from the phosphorylated Ser/Thr residue [211, 212]. Although Cnn constructs can assemble networks in the absence of phosphorylation *in vitro* (see previous chapters), Polo kinase phosphorylation enhances network assembly *in vitro* and is essential *in vivo* [196, 339].

The Cnn region D490-H608 is shown in the previous chapter to be the minimal region so far identified to still form networks *in vitro* upon binding to the Cnn C-terminus. In *Drosophila melanogaster*, this region contains six conserved Ser/Thr residues that are similar to the Polo consensus motif. These residues are S567, S571, T577, T581, S589 and S597 (ref [196]). However, only four of these sites (S567, S571, T577, S589) have consensus motifs which are also conserved across a range of species (Fig. 5.1). These four residues are therefore most likely to be conserved targets for Polo-dependent phosphorylation. Indeed, S567 has been shown via mass spectrometry and antibody-labelling to be specifically phosphorylated at the centrosome in fly embryos [196, 339].

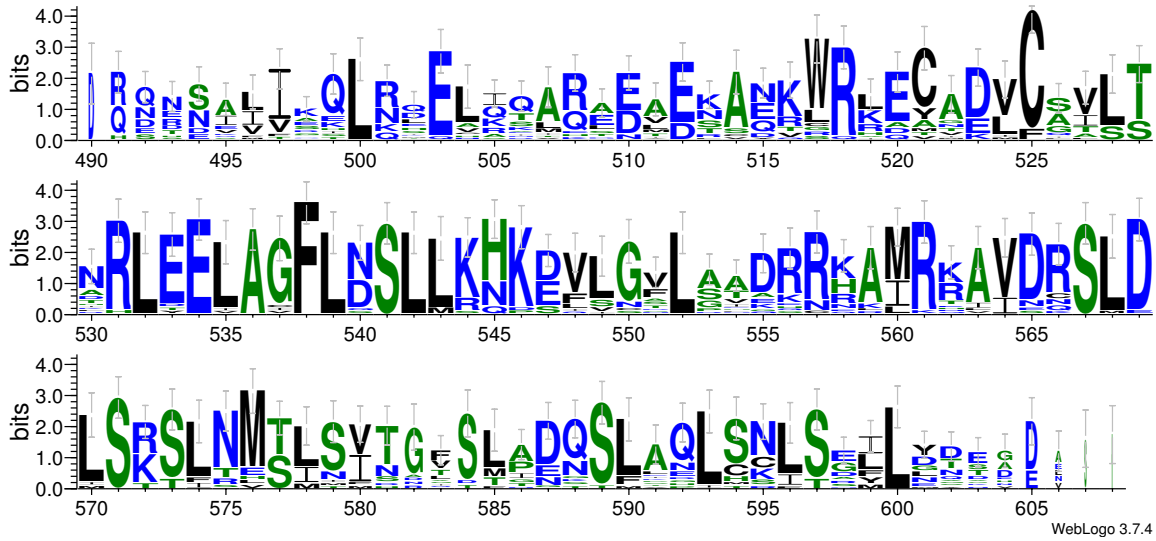
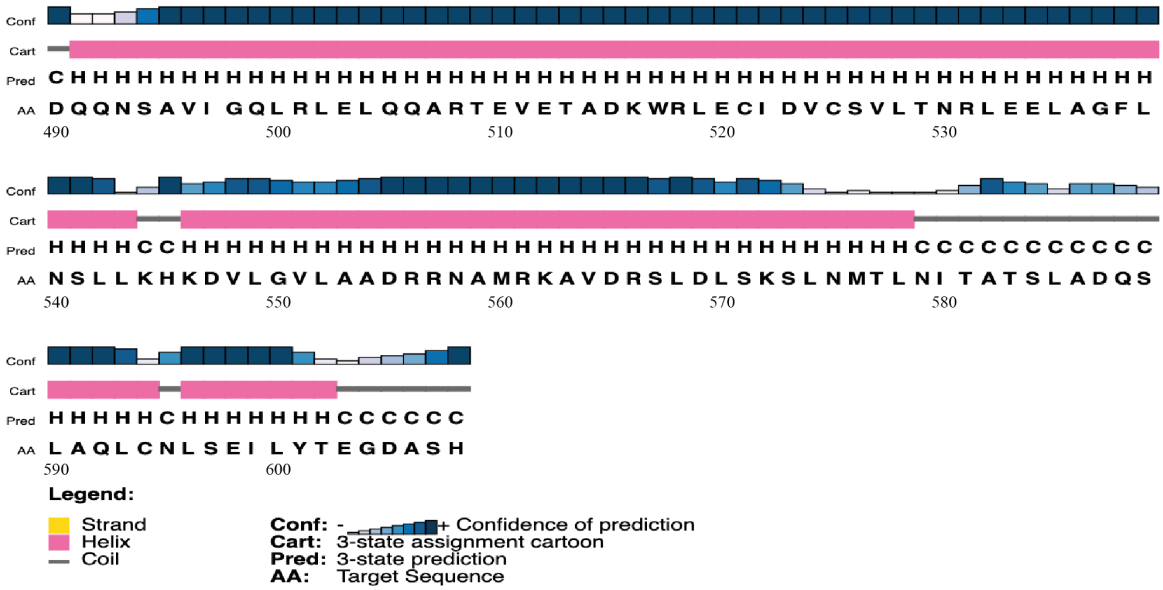


Figure 5.1 | Conservation of Cnn$^{D490\text{-}H608}$. Conservation across a range of insect species is shown in WebLogo format with hydrophilic amino acids in blue, neutral amino acids in green, and hydrophobic amino acids in black. For multiple-sequence alignment, see Appendix. Sequence logo created using WebLogo [342].


Conf
Cart
Pred
AA
C H H H

A mutagenesis screen has been carried out in which one or a combination of ten conserved Ser/Thr residues found within D490-H608 (see red residues in Fig. 1.8 A), including the four Polo consensus sites identified above, are mutated to alanine and the effect of these mutations on Cnn network assembly analysed in fly embryos. The strongest phenotypes are observed when at least two of the four conserved Polo-consensus targets are mutated to alanine. For example, the mutant $Cnn^{S567A/S571A/S573A}$ shows a much stronger phenotype than the mutants Cnn^{S567A} or $Cnn^{S571A/S573A}$ (ref [196]). However, this study does not control for possible effects these alanine-substitutions might have on the folding or structure of Cnn, which may be adversely affected by such point mutations.

The same study also showed that when all of these ten sites are mutated to glutamate or aspartate (10D/E mutant), Cnn forms networks in the cytosol rather than at the centrosome. In unfertilised eggs, these networks can even act as scaffolds for the assembly of microtubules [196].

A key part of the model for Cnn network assembly proposed in this thesis (Fig. 3.15) is that phosphorylation of the helices surrounding the zipper-like motif results in a conformational change that favours binding of the Cnn C-terminus, and this conformational change is then in turn stabilised upon complexation between the two motifs, thereby locking the protein into an "assembly-competent" state until dephosphorylation towards the periphery of the centrosome shifts the balance again towards dissociation of the two motifs and adoption of an "inactive" state (see also [339]). In light of this model, the Cnn 10D/E mutant described above would be in a constitutively assembly-competent state and therefore also forming networks in the cytosol, while Cnn with Polo phosphorylation sites mutated to alanine would be in a constitutively inactive state and therefore not able to form a network at all.

However, it is unclear what the nature of this phosphorylation-dependent conformational change is and therefore what role phosphorylation might have in the transition from an inactive to an assembly-competent state. Furthermore, the relative contribution or precise function of the potential phosphorylation target sites are also unknown. This chapter aims to address these questions.

It will be shown that $Cnn^{D490\text{-}N579}$ forms a dimer of helical hairpins *in crystallo*. This conformation is sterically incompatible with binding of the Cnn C-terminus, and is therefore proposed to represent the inactive state. Furthermore, it will be shown that mutating S567, S571 and S573 to alanine does not alter the conformation of $Cnn^{D490\text{-}N579}$ *in crystallo*, suggesting that the previously reported effect of these mutations *in vivo* is due to functional, not structural reasons. Based on these structures, phosphorylation on S567 and S571, but not S573 and T577, is predicted to have a role in conversion of Cnn to an assembly-competent state by necessitating a disruption of the hairpin structure.

5.2 A library of mNG-Cnn$^{D490\text{-}x}$ constructs

5.2.1 Design and purification of construct library

In preparation for structural studies, a library of constructs was cloned and purified by Joseph Fisher (Master's student) under the direct supervision of the author. In this library, all constructs had the same N-terminus (D490), but their C-termini varied between S567 and H608. These constructs were cloned into pETM14-mNG vectors, i.e. all constructs had an N-terminal, TEV-cleavable monomeric NeonGreen tag, as this system had been useful in the purification of other constructs in this thesis.

The smallest construct of the library, Cnn$^{D490\text{-}S567}$, has previously been characterised, and indeed the structure of Cnn$^{D490\text{-}S567}$ bound to Cnn$^{H1082\text{-}L1148}$ has been solved (PDB code: 5MWE). The two motifs are observed to form a 2:2 complex [339]. However, no residues beyond K544 can be observed in this structure (Fig. 3.1). Interestingly, secondary structure prediction suggests that around K544 there might be a break in helicity, and the highest confidence in helical prediction is only achieved again around D555 (Fig. 5.2). This region might therefore be intrinsically disordered, which would explain why it is not observed in the crystal structure. The largest construct in the library, Cnn$^{D490\text{-}H608}$, was characterised in the previous chapter, where mNG-Cnn$^{D490\text{-}H608}$ was shown to form networks upon binding to Cnn$^{G1077\text{-}L1148}$ (Fig. 4.21). By varying the C-terminus between these two extremes, it was hoped that a construct could be produced that could be crystallised.

Considering sequence conservation (Fig. 5.1) and secondary structure prediction (Fig. 5.2) of Cnn$^{D490\text{-}H608}$, domain boundaries were chosen based on a "frayed-ends" approach, in which the C-terminus, starting from S567, was successively extended by three or six amino

acids. Cnn$^{D490\text{-}x}$ constructs with the C-termini S567, S573, M576, N579 and L585 were thus produced.

All five proteins were purified by Joseph Fisher under the direct supervision of the author using the same protocol as described for mNG-Cnn$^{D490\text{-}H608}$ in the previous chapter. In each case, SDS-PAGE analysis was carried out to confirm successful purification of the protein.

The elution peak containing mNG-Cnn$^{D490\text{-}S567}$ (CV = 0.61-0.72) was concentrated to 12.2 mg/ml and frozen in 50 µl aliquots, with a final yield of ~ 7.9 mg/L (Fig. 5.3). The elution peak containing mNG-Cnn$^{D490\text{-}S573}$ (CV = 0.59-0.66) was concentrated to 11.5 mg/ml and frozen in 50 µl and 100 µl aliquots, with a final yield of ~ 11.5 mg/L (Fig. 5.4). The elution peak containing mNG-Cnn$^{D490\text{-}M576}$ (CV = 0.67-0.76) was concentrated to 8.7 mg/ml and frozen in 65 µl and 100 µl aliquots, with a final yield of ~ 12.8 mg/L (Fig. 5.5 C and D). The elution peak containing mNG-Cnn$^{D490\text{-}N579}$ (CV = 0.66-0.75) was concentrated to 24.1 mg/ml and frozen in 50 µl aliquots, with a final yield of ~ 15.1 mg/L (Fig. 5.6 C and D). The elution peak containing mNG-Cnn$^{D490\text{-}L585}$ (CV = 0.71-0.77) was concentrated to 9.2 mg/ml and frozen in 50 µl aliquots, with a final yield of ~ 6.0 mg/L (Fig. 5.7).

Note that in protein preparations involving the pETM14-mNG vector as described in this chapter, a residual amount of just the mNG-tag was always co-purified and observed to elute around CV = 0.8 on a S200 16/600. However, as this mNG peak was well removed from the mNG-Cnn$^{D490\text{-}x}$ peaks it did not affect purification of the proteins.

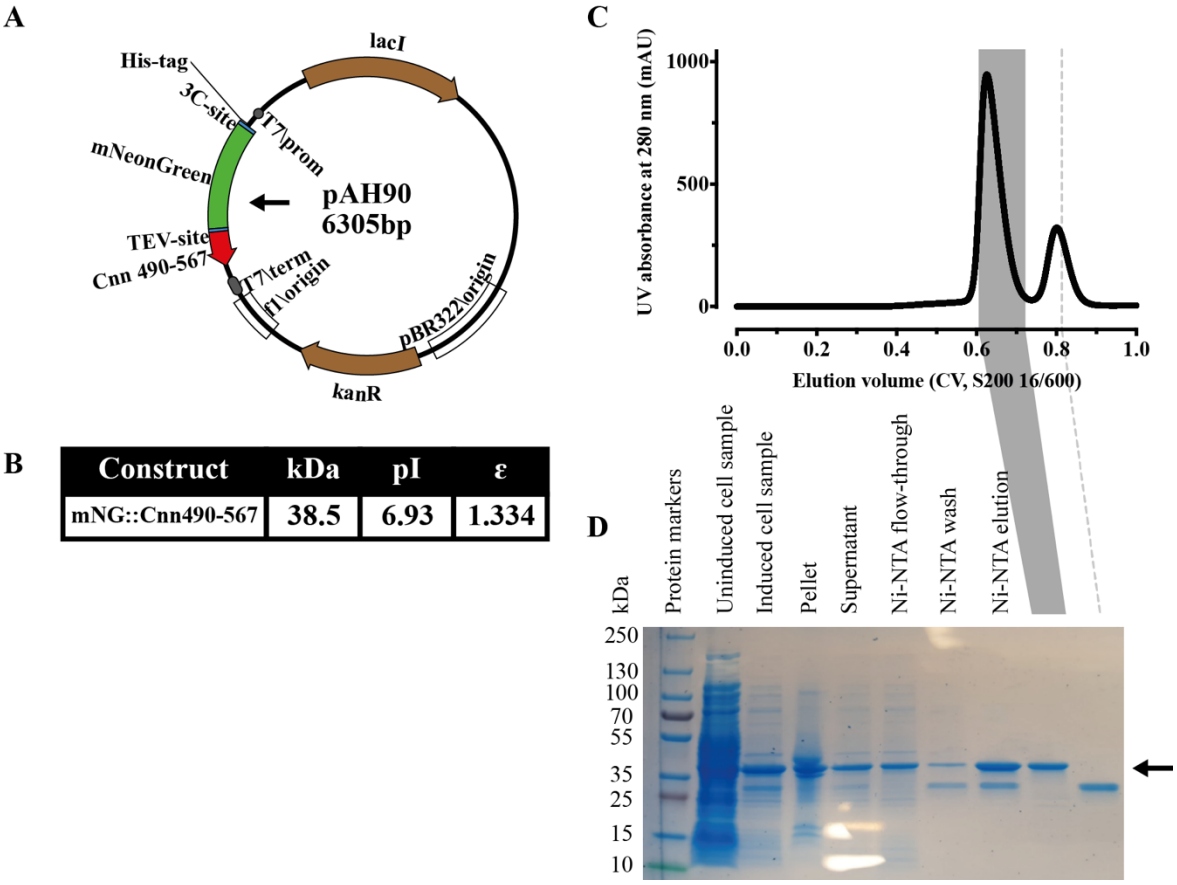


Figure 5.3 | Purification of N-terminally mNG-tagged Cnn$^{D490-S567}$. **A)** Overview of plasmid pAH90, which encodes Cnn$^{D490-S567}$ with an N-terminal, TEV-cleavable, His-3C-mNG-domain. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged protein. **C)** SEC (S200 16/600) elution profile of mNG-Cnn$^{D490-S567}$ after Ni-affinity chromatography, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.61-0.72) indicates the fractions pooled after SEC and run on the SDS-PAGE gel in D, containing mNG-Cnn$^{D490-S567}$. A further fraction run on the gel in D is indicated by a dashed grey line, likely corresponding to the mNG tag alone. SEC was carried out using standard TBS. **D)** 4-20 % reducing SDS-PAGE gel showing successive purification steps; the band corresponding to mNG-Cnn$^{D490-S567}$ is highlighted with a black arrow. The plasmid was cloned and protein was purified by Joseph Fisher under the author's direct supervision.

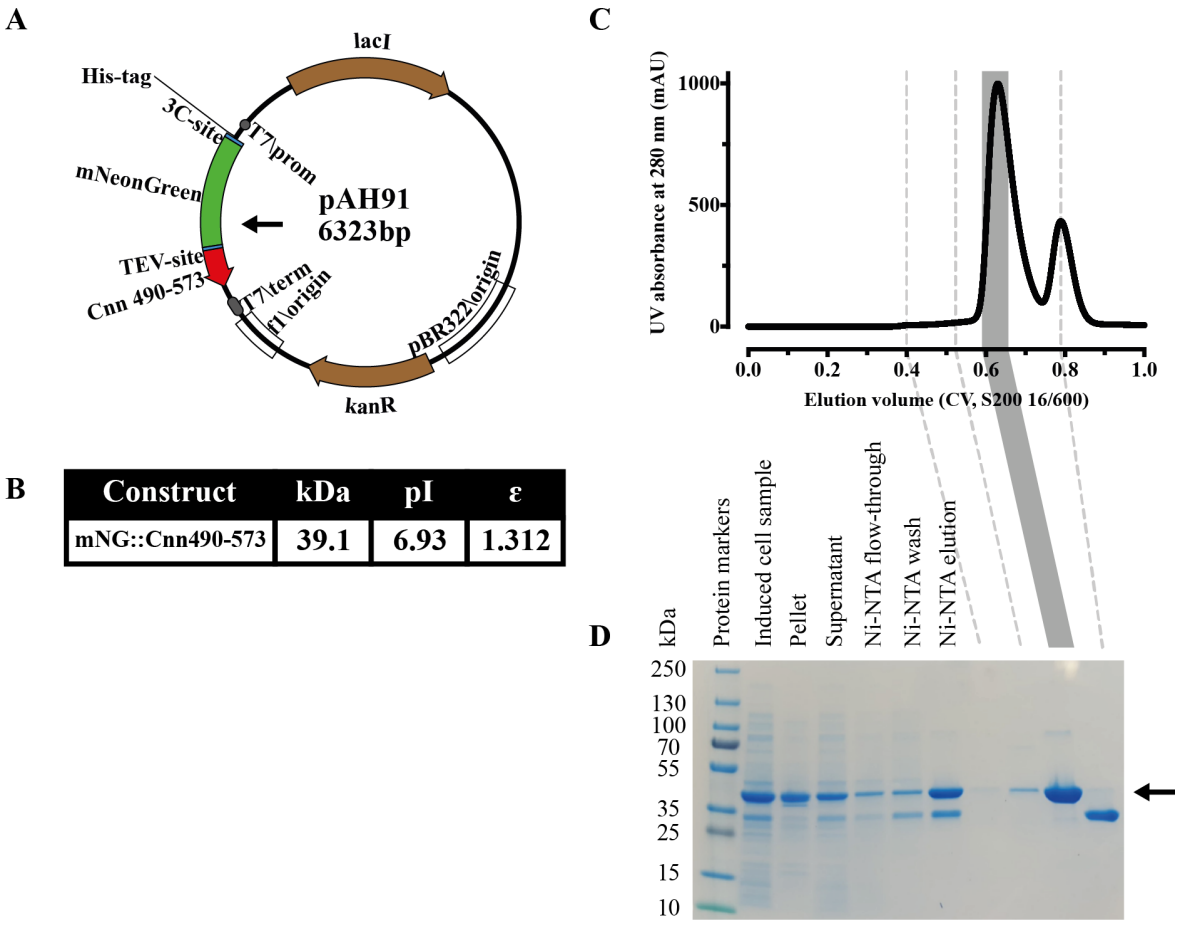


Figure 5.4 | Purification of N-terminally mNG-tagged Cnn$^{D490-S573}$. **A)** Overview of plasmid pAH91, which encodes Cnn$^{D490-S573}$ with an N-terminal, TEV-cleavable, His-3C-mNG-domain. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged protein. **C)** SEC (S200 16/600) elution profile of mNG-Cnn$^{D490-S573}$ after Ni-affinity chromatography, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.59-0.66) indicates the fractions pooled after SEC and run on the SDS-PAGE gel in D, containing mNG-Cnn$^{D490-S573}$. Further fractions run on the gel in D are indicated by dashed grey lines. SEC was carried out using standard TBS. **D)** 4-20 % reducing SDS-PAGE gel showing successive purification steps; the band corresponding to mNG-Cnn$^{D490-S573}$ is highlighted with a black arrow. The plasmid was cloned and protein was purified by Joseph Fisher under the author's direct supervision.

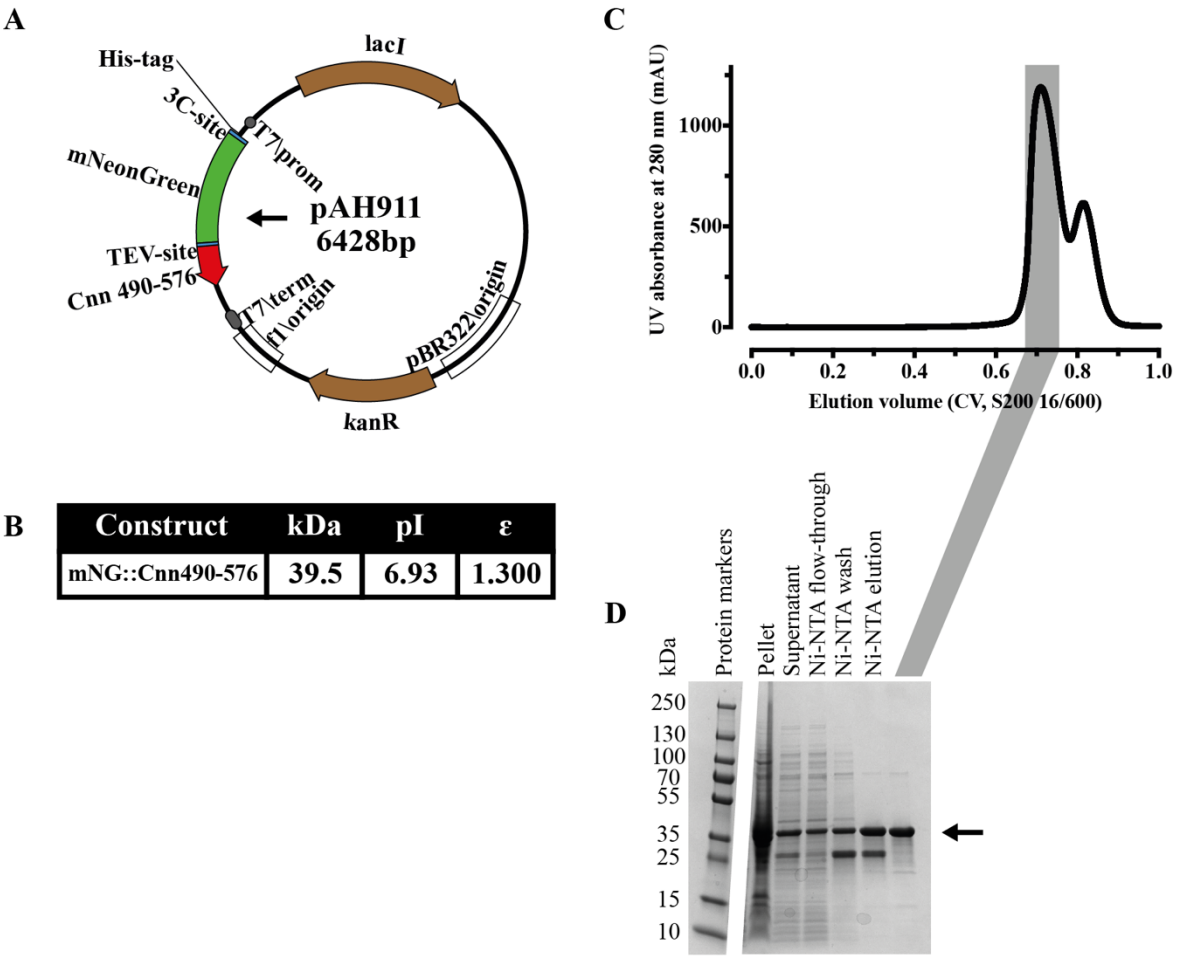
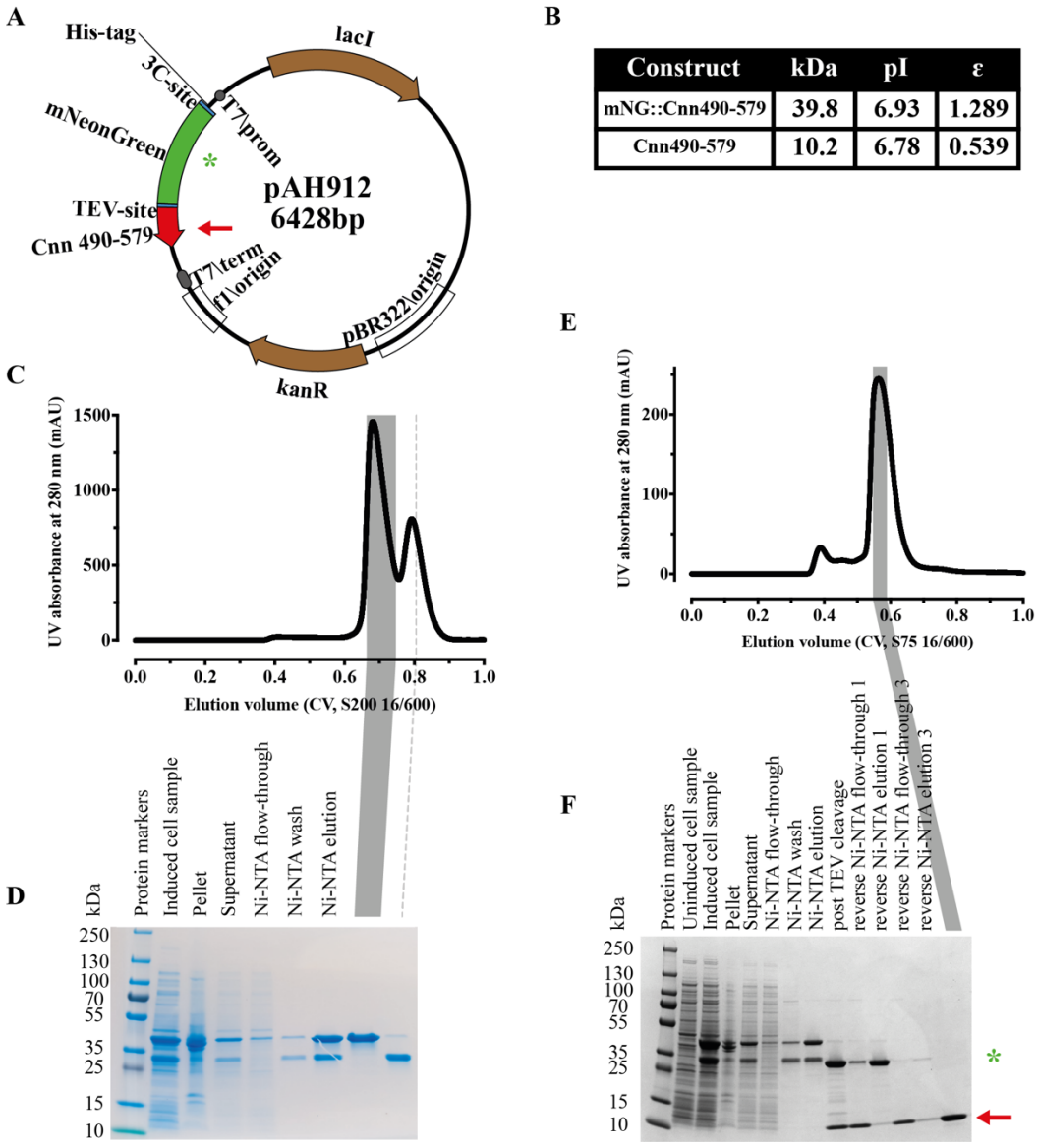
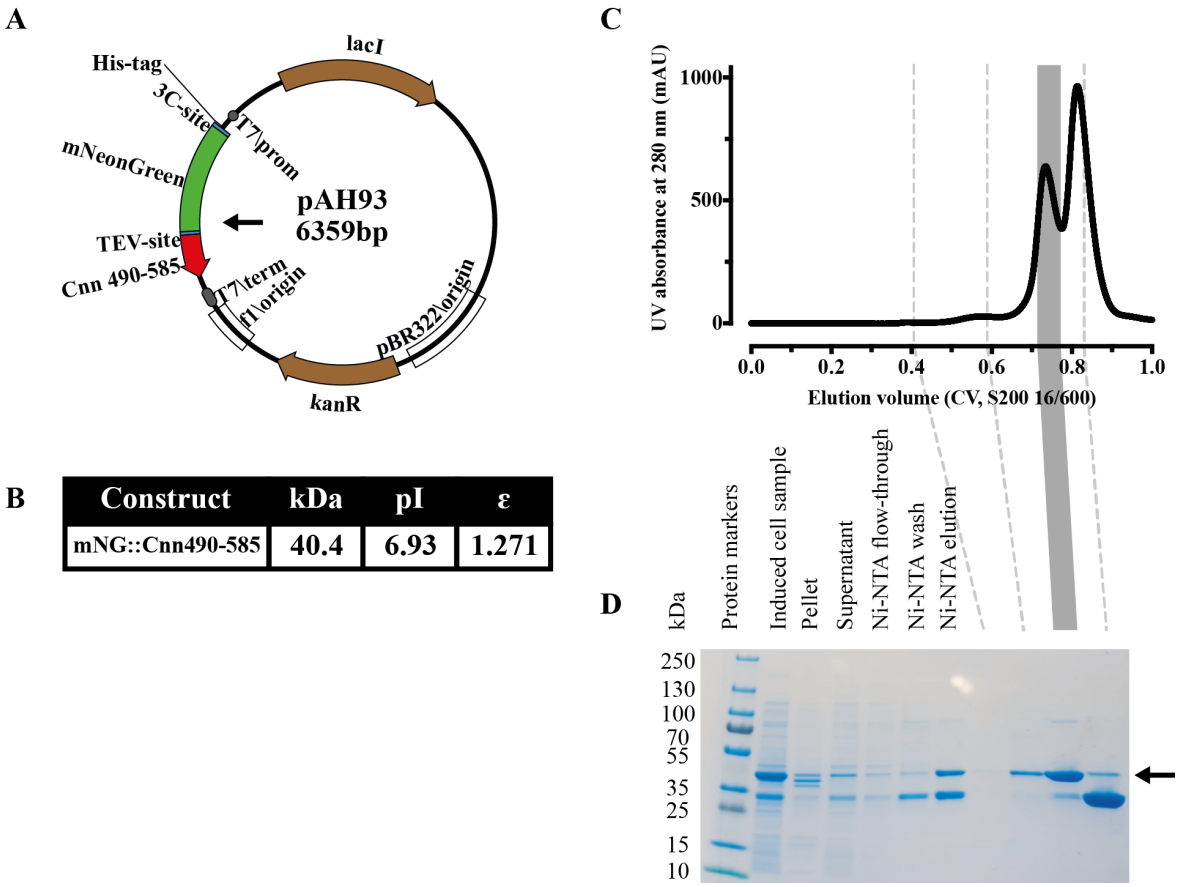


Figure 5.5 | Purification of N-terminally mNG-tagged Cnn$^{D490-M576}$. **A)** Overview of plasmid pAH911, which encodes Cnn$^{D490-M576}$ with an N-terminal, TEV-cleavable, His-3C-mNG-domain. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged protein. **C)** SEC (S200 16/600) elution profile of mNG-Cnn$^{D490-M576}$ after Ni-affinity chromatography, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.67-0.76) indicates the fractions pooled after SEC and run on the SDS-PAGE gel in D, containing mNG-Cnn$^{D490-M576}$. SEC was carried out using standard TBS. **D)** 4-20 % reducing SDS-PAGE gel showing successive purification steps. The band corresponding to Cnn$^{D490-M576}$ is marked with a black arrow. The plasmid was cloned and protein was purified by Joseph Fisher under the author's direct supervision.



Construct	kDa	pI	ε
mNG::Cnn490-579	39.8	6.93	1.289
Cnn490-579	10.2	6.78	0.539

Figure 5.6 | Purification of Cnn$^{D490-N579}$. **A)** Overview of plasmid pAH912, which encodes Cnn$^{D490-N579}$ with an N-terminal, TEV-cleavable, His-3C-mNG-domain. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged and untagged protein. **C)** SEC (S200 16/600) elution profile of mNG-Cnn$^{D490-N579}$ after Ni-affinity chromatography, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.66-0.75) indicates the fractions pooled after SEC and run on the SDS-PAGE gel in D, containing mNG-Cnn$^{D490-M576}$. A further fraction run on the gel in D is indicated by a grey dashed line and contains the mNG-tag alone. SEC was carried out using standard TBS. **D)** 4-20 % reducing SDS-PAGE gel showing successive purification steps. **E)** SEC (S75 16/600) elution profile of Cnn$^{D490-N579}$ after tag cleavage, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.55-0.59) indicates the fractions pooled after SEC and run on the SDS-PAGE gel in F, containing Cnn$^{D490-N579}$. SEC was carried out using standard TBS. **F)** 4-20 % reducing SDS-PAGE gel showing successive purification steps. The band corresponding to Cnn$^{D490-N579}$ is marked with a red arrow, the band corresponding to the mNG-tag is indicated with a green asterisk. The plasmid was cloned and protein was purified by Joseph Fisher under the author's direct supervision.



Construct	kDa	pI	ε
mNG::Cnn490-585	40.4	6.93	1.271

Figure 5.7 | Purification of N-terminally mNG-tagged Cnn$^{D490-L585}$. **A)** Overview of plasmid pAH93, which encodes Cnn$^{D490-L585}$ with an N-terminal, TEV-cleavable, His-3C-mNG-domain. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged protein. **C)** SEC (S200 16/600) elution profile of mNG-Cnn$^{D490-L585}$ after Ni-affinity chromatography, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.71-0.77) indicates the fractions pooled after SEC and run on the SDS-PAGE gel in D, containing mNG-Cnn$^{D490-L585}$; further fraction run on the gel in D are indicated by dashed grey lines. SEC was carried out using standard TBS. **D)** 4-20 % reducing SDS-PAGE gel showing successive purification steps; the band corresponding to mNG-Cnn$^{D490-H608}$ is highlighted with a black arrow. The plasmid was cloned and protein was purified by Joseph Fisher under the author's direct supervision.

5.2.2 Functional analysis of construct library

SEC-MALS experiments were carried out by Joseph Fisher under the direct supervision of the author in order to determine the oligomeric state of the constructs within the mNG-Cnn$^{D490\text{-}x}$ library. In all cases, mNG-Cnn$^{D490\text{-}x}$ was tested on a S200 10/300 at 50 µM using standard TBS as buffer (Fig. 5.8).

For mNG-Cnn$^{D490\text{-}S567}$, an average molecular weight of 154.4 kDa (+/- 0.6 %) was measured, which closely matched the theoretical mass expected for a tetramer of this construct at 153.9 kDa (Fig. 5.8 A). In contrast, an average molecular weight of 91.4 kDa (+/- 0.7 %) was measured for mNG-Cnn$^{D490\text{-}H608}$, thereby closely corresponding to 85.7 kDa, the theoretical mass expected for a dimer (Fig. 5.8 F). These results were therefore in line with those obtained for Cnn$^{D490\text{-}K544}$ (tetrameric, Fig. 3.6 A) and MBP-Cnn$^{Q403\text{-}H608}$ (dimeric, Fig. 3.8 C), suggesting that dimerisation is driven by the C-terminal residues.

Indeed, mNG-Cnn$^{D490\text{-}S573}$ was observed to form both tetramers (140.1 kDa +/- 0.3 %) and dimers (92.0 kDa +/- 0.4 %) under the same conditions, at roughly equal proportions (Fig. 5.8 B). Extension of the construct by another three amino acids then led to complete dimerisation, as mNG-Cnn$^{D490\text{-}M576}$ was observed to form dimeric species with an average molecular weight of 88.4 kDa +/- 0.8 % (Fig. 5.8 C). Surprisingly, the addition of just nine amino acids from S567 to M576 was therefore sufficient to drive complete dimerisation of the construct.

As expected, the constructs with an even longer C-terminus (mNG-Cnn$^{D490\text{-}N579}$ and mNG-mNG-Cnn$^{D490\text{-}L585}$) were also fully dimeric at 86.2 kDa (+/- 0.7 %) and 87.7 kDa (+/- 0.2 %), respectively (Fig. 5.8 D and E).

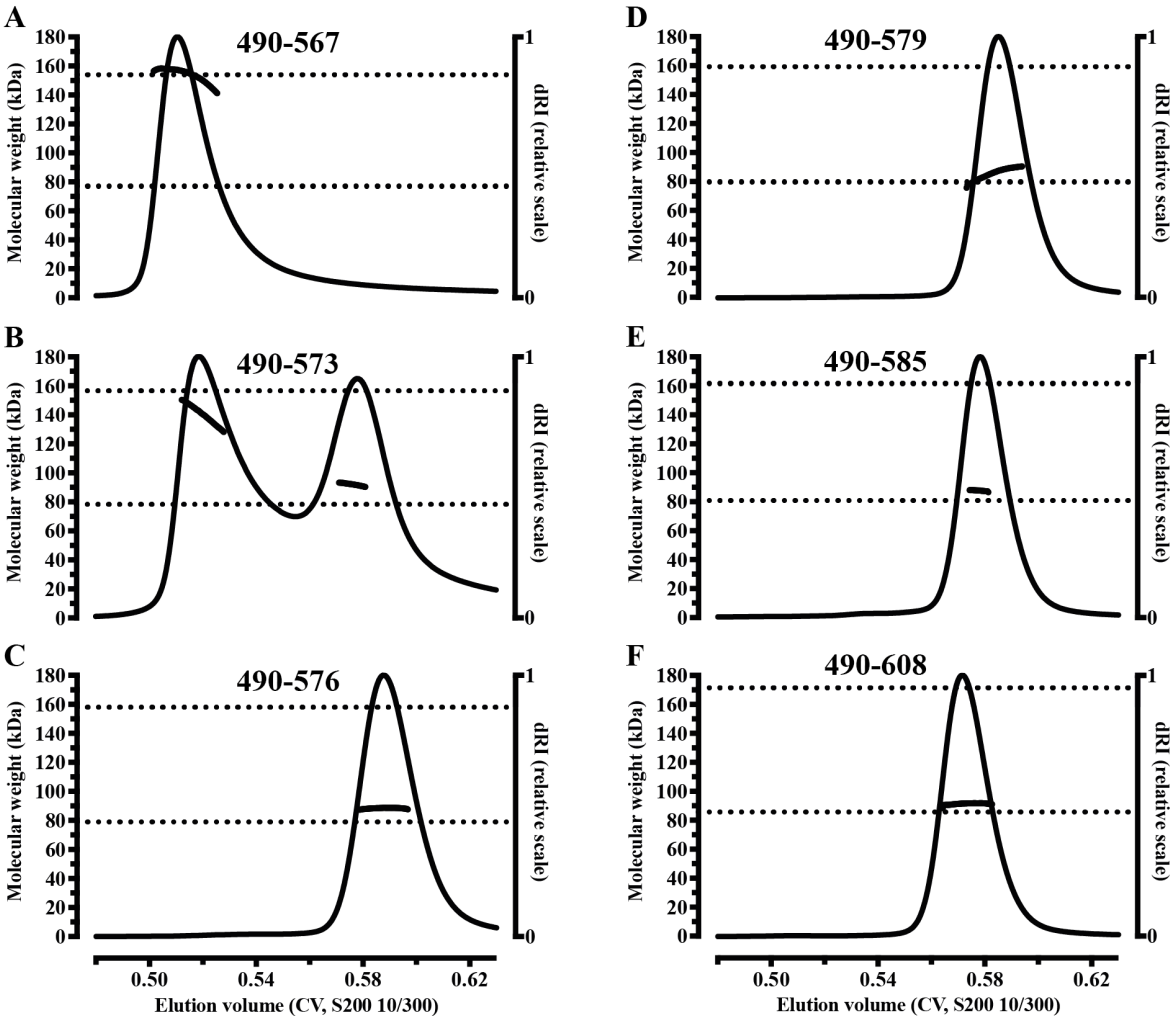


Figure 5.8 | SEC-MALS analysis of Cnn$^{D490\text{-}x}$ oligomerisation. The average molecular weight in solution of different mNG-Cnn$^{D490\text{-}x}$ constructs was measured at 50 µM. In each case, the dotted lines indicate the theoretical molecular weight expected for dimeric and tetrameric mNG-Cnn$^{D490\text{-}x}$. Molecular weight (in kDa) is plotted on the left y-axis, change in refractive index (dRI, relative scale) plotted on the right y-axis, column elution volume (CV, 24 ml) of column S200 10/300 plotted on the x-axis. Standard TBS was used as buffer. **A)** mNG-Cnn$^{D490\text{-}S567}$ is tetrameric in solution (154.4 kDa +/- 0.6 %). **B)** mNG-Cnn$^{D490\text{-}S573}$ is roughly equally divided between tetramers (140.1 kDa +/- 0.3 %) and dimers (92.0 kDa +/- 0.4 %) in solution. **C)** mNG-Cnn$^{D490\text{-}M576}$ is dimeric in solution (88.4 kDa +/- 0.8 %). **D)** mNG-Cnn$^{D490\text{-}N579}$ is dimeric in solution (86.2 kDa +/- 0.7 %). **E)** mNG-Cnn$^{D490\text{-}L585}$ is dimeric in solution (87.7 kDa +/- 0.2 %). **F)** mNG-Cnn$^{D490\text{-}H608}$ is dimeric in solution (91.4 kDa +/- 0.7 %). Note how extension of mNG-Cnn$^{D490\text{-}S567}$ by nine amino acids at its C-terminus induces complete dimerisation of the construct. Experiments were carried out by Joseph Fisher under the direct supervision of the author.

5.3 mNG-Cnn$^{D490\text{-}N579}$ forms 2:2 complexes with the Cnn C-terminus

As mentioned above, Cnn$^{D490\text{-}S567}$ forms 2:2 complexes with the Cnn C-terminus [339], while mNG-Cnn$^{D490\text{-}H608}$ forms large networks upon binding the Cnn C-terminus (Fig. 4.21). In order to investigate whether full dimerisation and network assembly rely on the same residues, SEC-MALS was used to study the interaction between mNG-Cnn$^{D490\text{-}N579}$ and Cnn$^{G1077\text{-}L1148}$. This construct was used, rather than mNG-Cnn$^{D490\text{-}M576}$, which is the shortest construct observed to fully dimerise (Fig. 5.8), as mNG-Cnn$^{D490\text{-}N579}$ not only formed complete dimers in solution but also incorporates the entire region predicted to still be helical (Fig. 5.2), and was therefore deemed to have more appropriate domain boundaries.

As a control, 110 µl reactions of mNG-Cnn$^{D490\text{-}N579}$ alone were set up at half (25 µM) and four times (200 µM) the concentration used in the previous SEC-MALS experiment with this construct (Fig. 5.8) using standard TBS as buffer. The average mass in solution measured for the peaks observed in these experiments was 84.7 kDa (+/- 0.8 %) and 94.3 kDa (+/- 0.9 %), respectively (Fig 5.9). As the theoretical dimeric mass of this construct is 79.6 kDa, this suggests that mNG-Cnn$^{D490\text{-}N579}$ is mostly dimeric in solution but shows a tendency to also transiently form higher-order oligomers in a concentration-dependent manner, resulting in a measured average mass slightly higher than that expected for a dimer. This is not surprising considering the previous analysis demonstrating that constructs shorter than Cnn$^{D490\text{-}N579}$ can form tetramers (Fig. 5.8).

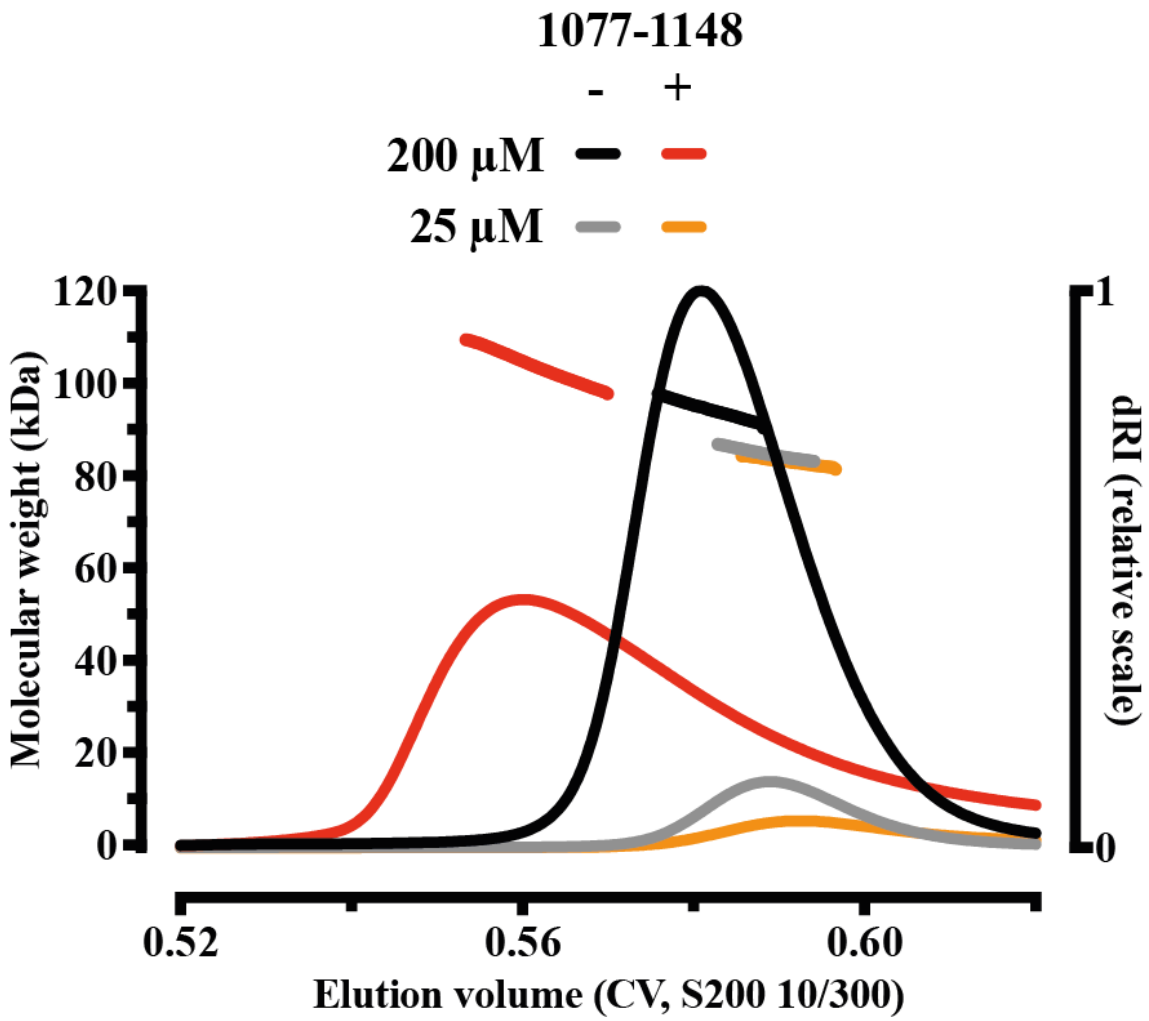


Figure 5.9 | SEC-MALS analysis of interaction between mNG-Cnn$^{D490\text{-}N579}$ and Cnn$^{G1077\text{-}L1148}$. The average molecular weight in solution of mNG-Cnn$^{D490\text{-}N579}$ alone was measured at 25 µM (light grey: 84.7 kDa +/- 0.8 %) and 200 µM (black: 94.3 kDa +/- 0.9 %), as well as mixed with a roughly equimolar concentration of Cnn$^{G1077\text{-}L1148}$ shown in orange (82.9 kDa +/- 0.8 %) and red (103.7 kDa +/- 1.0 %), respectively. Molecular weight (in kDa) is plotted on the left y-axis, change in refractive index (dRI, relative scale) plotted on the right y-axis, column elution volume (CV, 24 ml) of column S200 10/300 increase plotted on the x-axis. Standard TBS was used as buffer. mNG-Cnn$^{D490\text{-}N579}$ shows concentration-dependent oligomerisation but mostly forms dimers in the absence of Cnn$^{G1077\text{-}L1148}$. At the higher concentration, a peak shift and increase in mass can be observed upon addition of Cnn$^{G1077\text{-}L1148}$, indicating binding and probably 2:2 formation. Note the long tail observed under these conditions (red), suggesting that the complex is not fully occupied or perhaps interacts with the column resin, which might explain the apparent disappearance of the mNG-Cnn$^{D490\text{-}N579}$ peak at the lower concentration upon addition of Cnn$^{G1077\text{-}L1148}$ (orange).

Two further SEC-MALS experiments were then carried out under the same conditions, but in addition a roughly equimolar concentration of $Cnn^{G1077\text{-}L1148}$ (i.e. the Cnn C-terminus) was added. At the lower concentration (25 µM of mNG-$Cnn^{D490\text{-}N579}$), no peak shift could be observed and the measured mass for mNG-$Cnn^{D490\text{-}N579}$ (82.9 kDa +/- 0.8 %) was consistent with the previously measured mass of this construct in the absence of $Cnn^{G1077\text{-}L1148}$. However, the mNG-$Cnn^{D490\text{-}N579}$ peak was reduced in size, suggesting that addition of $Cnn^{G1077\text{-}L1148}$ did have an effect on mNG-$Cnn^{D490\text{-}N579}$ (Fig. 5.9). Indeed, at the higher concentration (200 µM of mNG-$Cnn^{D490\text{-}N579}$), addition of $Cnn^{G1077\text{-}L1148}$ clearly led to the appearance of an earlier-eluting peak with an average mass of 103.7 kDa +/- 1.0 % (Fig. 5.9), therefore probably corresponding to a 2:2 complex between mNG-$Cnn^{D490\text{-}N579}$ and $Cnn^{G1077\text{-}L1148}$, which has a theoretical molecular weight of 96.4 kDa. The earlier-eluting peak had a long tail, which might be due to only partial occupancy of the complex or due to an interaction between the complex and the column resin. Importantly, no larger assemblies could be observed, suggesting that under these conditions, network assembly *in vitro* also requires residues I580-H608, which mNG-$Cnn^{D490\text{-}N579}$ lacks.

As phosphorylation is important for efficient network assembly *in vitro* [339] it was tested whether addition of commercial human PLK1 kinase affects complex formation between mNG-$Cnn^{D490\text{-}N579}$ and $Cnn^{G1077\text{-}L1148}$. Human PLK1 is highly conserved (> 40 % sequence identity between *D. melanogaster* Polo, human PLK1 and *C. elegans* PLK-1) and therefore also expected to phosphorylate *Drosophila* targets. mNG-$Cnn^{D490\text{-}N579}$ contains three of the four consensus Polo kinase target sites identified above (S567, S571, T577).

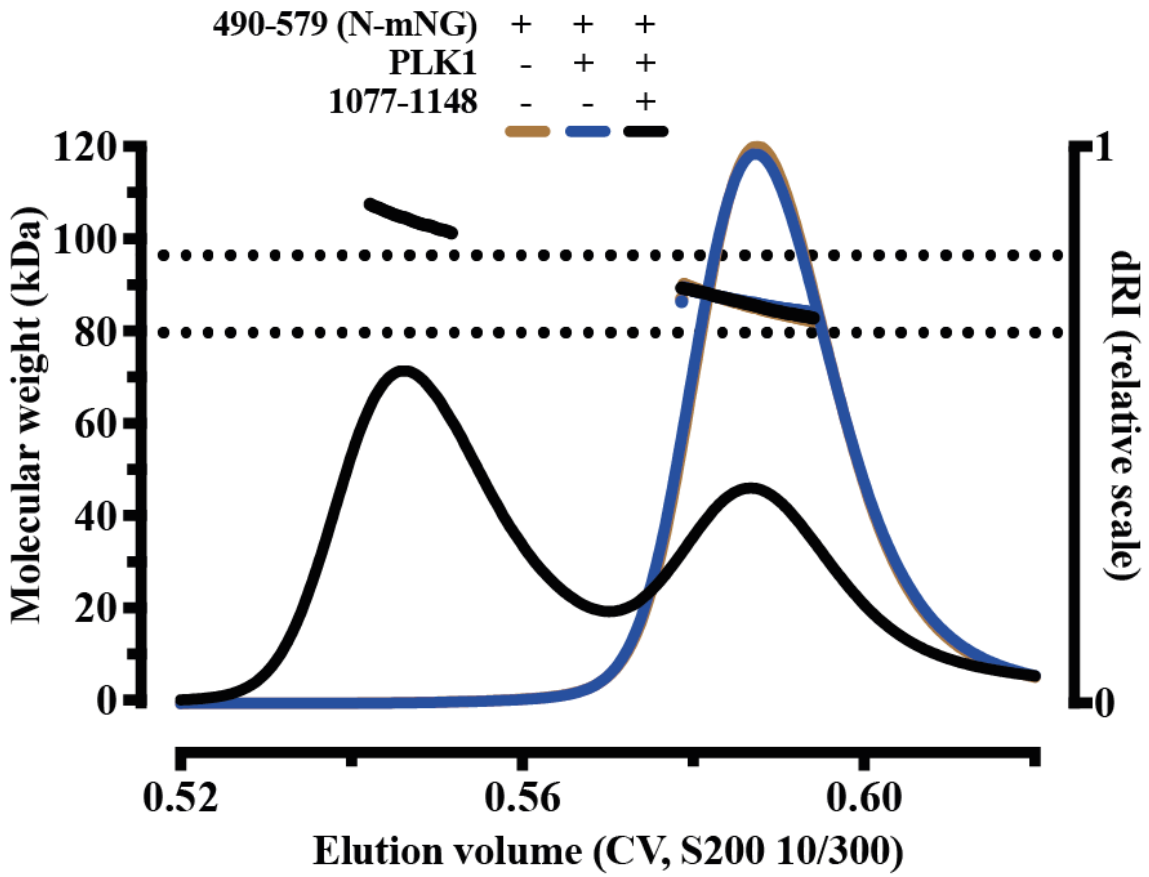


Figure 5.10 | mNG-Cnn$^{D490-N579}$ forms 2:2 complexes with Cnn$^{G1077-L1148}$ in the presence of kinase. SEC-MALS was used to measure the average molecular weight in solution of mNG-Cnn$^{D490-N579}$ alone at 25 µM (brown, 85.5 kDa +/- 0.7 %), mNG-Cnn$^{D490-N579}$ in the presence of commercial PLK1 kinase (blue, 86.2 kDa +/- 0.7 %) and mNG-Cnn$^{D490-N579}$ mixed with an equimolar concentration of Cnn$^{G1077-L1148}$ in the presence of PLK1 (black, 104.2 kDa +/- 0.7 % and 85.9 kDa +/- 0.7 %). Dotted lines indicate the theoretical molecular weight expected for dimeric mNG-Cnn$^{D490-N579}$ and a 2:2 complex of mNG-Cnn$^{D490-N579}$ with Cnn$^{G1077-L1148}$. Molecular weight (in kDa) is plotted on the left y-axis, change in refractive index (dRI, relative scale) plotted on the right y-axis, column elution volume (CV, 24 ml) of column S200 10/300 plotted on the x-axis. 25 mM Tris-HCl (pH 7.5), 150 mM NaCl, 2 mM DTT was used as buffer. Note how mNG-Cnn$^{D490-N579}$ forms a 2:2 complex with Cnn$^{G1077-L1148}$ in the presence of kinase. Experiments were carried out by Joseph Fisher under the direct supervision of the author.

Under the direct supervision of the author, 110 µl reactions were set up by Joseph Fisher, containing 25 µM of mNG-Cnn$^{D490-N579}$ and 0.2 mM ATP. mNG-Cnn$^{D490-N579}$ remained dimeric at this concentration both in the presence and absence of PLK1, with measured molecular weights of 86.2 kDa (+/- 0.7 %) and 85.5 kDa (+/- 0.7 %), respectively. However, when Cnn$^{G1077-L1148}$ was added at a roughly equimolar concentration, a new peak appeared that contained species with an average molecular weight of 104.2 kDa (+/- 0.7 %), closely corresponding to the expected mass for a 2:2 complex between the two constructs (Fig. 5.10). No larger assemblies could be observed, suggesting that addition of kinase was not sufficient to induce formation of networks by these Cnn constructs.

Note that a distinct peak shift could not previously be observed at 25 µM of mNG-Cnn$^{D490-N579}$ in the absence of kinase, and the 2:2 complex peak in the presence of kinase was more symmetrical than the one observed at 200 µM in the absence of phosphorylation (Figs 5.9 and 5.10). This might be due to phosphorylation allowing more efficient complex formation between mNG-Cnn$^{D490-N579}$ and Cnn$^{G1077-L1148}$. Alternatively, this might be a limitation of the assay since, as described above, the apparent disappearance of the mNG-Cnn$^{D490-N579}$ peak at 25 µM and the long tail at 200 µM in the absence of kinase (Fig 5.9) might be due to an interaction between the unphosphorylated 2:2 complex and the column resin.

5.4 Cnn$^{D490\text{-}N579}$ forms a dimer of α-helical hairpins *in crystallo*

5.4.1 Cnn$^{D490\text{-}N579}$ purification and crystallisation

Cnn$^{D490\text{-}N579}$ was purified by Joseph Fisher, under the direct supervision of the author, as described above. In preparation for crystallisation trials, however, the mNG tag was cleaved off by addition of 2 mg TEV protease (from a 1 mg/ml stock) and dialysis overnight, followed by three rounds of reverse Ni-affinity chromatography the next day to remove the tag. After concentration to 5 ml and application to a S75 16/600 for SEC, a single peak was obtained (CV = 0.55-0.59), which was confirmed to contain Cnn$^{D490\text{-}N579}$ via SDS-PAGE analysis (Fig. 5.6 E and F). The protein was concentrated to 18.0 mg/ml and frozen in 45 µl aliquots. The final yield was ~ 4.1 mg/L.

In order to find the ideal concentration of Cnn$^{D490\text{-}N579}$ for crystallisation studies, three Stura screens (Molecular Dimensions) were set up by the author with increasing concentrations of Cnn$^{D490\text{-}N579}$ at 6 mg/ml (0.6 mM), 12 mg/ml (1.2 mM) and 18 mg/ml (1.8 mM). Surprisingly, after four days, crystals were observed on all three trays in the same two conditions:

B5: 0.8 M potassium/sodium phosphate (pH 7.0)

H4: 2.0 M ammonium sulphate, 0.1 M sodium HEPES (pH 7.5), 5 % (v/v) PEG 400

No other conditions were observed to contain crystals. Crystals from both of these conditions and all three concentrations were therefore fished and frozen using 25 % (v/v) glycerol as cryoprotectant for crystals grown in B5 and 15 % (v/v) glycerol for crystals grown in H4,

which were already partially cryoprotected through the presence of 5 % (v/v) PEG 400 in the mother liquor. Data collection was carried out by Prof Susan Lea at the DIAMOND beamline i04 (proposal mx18069). The crystals diffracted to a varying quality, and the crystal with the highest quality dataset after indexing from each condition is shown in Table 5.1. Note two diffraction datasets of the best crystal from condition H4 were collected.

Beamline	DIAMOND i04		
Wavelength (Å)	0.9795		
Crystal	1.2 mM Condition H4 (1)	1.2 mM Condition H4 (2)	1.8 mM Condition B5
Resolution limits (Å)	51.13-1.99 (2.03-1.99)	54.11-1.96 (1.99-1.96)	107.84-2.04 (2.07-2.04)
Space group	$P6_1$	$P6_1$	$P6_1$
Unit cell dimensions (Å, º)	59.04, 59.04, 108.07 90, 90, 120	59.14, 59.14, 108.22 90, 90, 120	58.98, 58.98, 107.84 90, 90, 120
Unique reflections	14576 (746)	15302 (739)	13590 (634)
Multiplicity	11.7 (11.1)	11.2 (9.8)	11.2 (11.4)
Completeness (%)	100.0 (100.0)	99.5 (99.6)	100.0 (96.8)
I/ σ(I)	19.6 (0.6)	16.3 (1.0)	8.1 (1.4)
R_{merge} (%)	5.4 (452.8)	6.5 (258.9)	15.8 (175.4)
R_{meas} (%)	5.7 (474.8)	6.8 (273.2)	16.6 (183.7)
R_{pim} (%)	1.7 (142.1)	2.0 (87.0)	5.0 (54.3)
$CC_{1/2}$	1.0 (0.4)	1.0 (0.4)	1.0 (0.7)
Processing pipeline	autoPROC (DIAMOND)	Xia2 dials (DIAMOND)	Xia2 dials (DIAMOND)

Table 5.1 | $Cnn^{D490\text{-}N579}$ data collection and processing statistics. The crystal row refers to the concentration of $Cnn^{D490\text{-}N579}$ and condition under which the crystal grew; Condition H4 (1) and Condition H4 (2) are two different diffraction datasets collected from the same crystal. Where applicable, values for the inner resolution shell are given in brackets.

5.4.2 $Cnn^{D490\text{-}N579}$ structure determination

Using Matthews coefficient, it was estimated that the most likely composition of the asymmetric unit for the two crystals grown in H4 (second dataset) and B5 was a dimer of $Cnn^{D490\text{-}N579}$, with a respective solvent content of 54.07 % and 53.65 % and a probability of 100 % for an expected mass of 20.4 kDa.

As a homology model for phaser, a truncated version of $Cnn^{D490\text{-}K544(L535E)}$ based on its complex structure with the Cnn C-terminus (PDB code: 5MW9) was used, which only retained the core dimer and lacked the termini (as well as its binding partner in the structure, the Cnn C-terminus). It was hypothesised that the dimeric core might be structurally conserved in $Cnn^{D490\text{-}N579}$. Indeed, a single solution with a translation function Z-score of 9.6 was obtained when using this search model on the first H4 dataset, indexed with autoPROC. After some minor refinement involving changing a few side chain orientations, the model was fixed and phaser instructed to search for a straight, 20 amino acids long, poly-alanine α-helix (adapted from PDB code: 5AL6). This resulted in 14 solutions, with the top hit having a translational function Z-score of 11.2. This cycle of placing a short stretch of a poly-alanine α-helix via phaser or molrep followed by refinement via Coot and refmac was repeated several times until no satisfactory placements were possible anymore. At this point, buccaneer was used with this partial solution and the expected protein sequence in order to build in the correct sequence of amino acids. Several cycles of refinement via refmac and Coot followed.

However, considering that the second dataset collected on this crystal had better statistics after indexing (Table 5.1), further structure determination was decided to be carried out using the second H4 dataset, indexed with Xia2 dials. In order to debias the model, all atoms

were randomly displaced by up to 0.5 Å per atom via pdbset on the command terminal, and all B-factors were reset to 60. This noisy model was then used as a search model in phaser and placed into this new map with a translational function Z-score of 28.9. Iterative cycles of model building and refinement via refmac, phenix.refine and Coot followed and the model thereby slowly built up by the addition of further residues as well as solvent molecules. Furthermore, Ramachandran and side chain rotamer outliers were eliminated where possible and plausible until a final R_{free} of 21.79 % and R_{work} of 20.34 % was reached (Table 5.2).

In addition, structure determination on the B5 dataset was carried out. The final result from the second H4 dataset was used as a starting model, but all solvent molecules were deleted, B-factors reset to 60 and a noise of up to 0.5 Å per atom introduced as above. This noisy model was directly used for refinement in phenix.refine including rigid body refinement to account for minor differences in the unit cell dimensions. This resulted in an initial R_{free} of 31.01 % and R_{work} of 28.62 %. The model was then refined and built up by adding solvent molecules as well as further residues at the termini, which were resolved in this dataset, and by changing a few side chains that were in a different conformation. After model validation, structure determination was concluded when a final R_{free} of 21.82 % and R_{work} of 20.55 % was reached (Table 5.2).

Crystal	1.2 mM Condition H4 (2)	1.8 mM Condition B5
Resolution (Å)	51.23-1.96	51.09-2.04
No. of reflections:		
working set	14575	12920
test set	701	606
R_{work} / R_{free} (%)	20.34 / 21.79	20.55 / 21.82
Asymmetric unit composition	dimer	dimer
Observed residues	Chain A: S494-N575 Chain B: A495-N575	Chain A: N493-M576 Chain B: Q492-T577
No. of atoms:		
Overall	1361	1452
Protein	1280	1334
Ligand	20	30
Water	61	88
Average B factors (Å^2):		
Overall	68.0	59.0
Protein	68.4	59.8
Ligand	55.6	46.0
Water	63.5	51.0
Ramachandran statistics (%):		
favoured	100.0	100.0
allowed	0	0
outliers	0	0
Rotamer outliers (%)	0.71	0
RMS deviation:		
bond lengths (Å)	0.003	0.003
bond angles (°)	0.453	0.465
Clashscore	3.03	3.64
MolProbity score	1.09	1.15

Table 5.2 | Cnn$^{D490\text{-}N579}$ refinement statistics. The crystal row refers to the concentration of Cnn$^{D490\text{-}N579}$ and condition under which the crystal grew (cf. Table 5.1).

5.4.3 $Cnn^{D490\text{-}N579}$ structure analysis

The asymmetric unit contains a dimer of $Cnn^{D490\text{-}N579}$ (Fig. 5.11 A/B), in line with its oligomeric state in solution (Fig. 5.8 D). Both chains are almost entirely α-helical, except for two loop regions centred around K544 and L552, with a short, distorted α-helix in between. Note this closely matches the secondary structure previously predicted, which predicted K544 to be involved in a random coil and L552 to be part of a helix but with low confidence (Fig. 5.2). These secondary structure features allow both chains to adopt a hairpin fold as a tertiary structure, with the C-terminal helix A553-M576 folding back against the N-terminal helix N493-L542. The quaternary structure of $Cnn^{D490\text{-}N579}$ therefore consists of a dimer of α-helical hairpins. Note that the two chains of $Cnn^{D490\text{-}N579}$ are not completely identical. For example, more residues can be observed on chain B, including the C-terminal T577 (Fig. 5.11 A/B).

The polypeptide chains of the two structures of $Cnn^{D490\text{-}N579}$ solved in this thesis (from the two datasets H4 and B5) are almost identical, except for minor differences in some side chain conformations and the number of residues that are resolved in the two structures (Fig. 5.11 C). Importantly, four sulphates can be observed in the structure from the H4 dataset. Two of these sulphates are each coordinated by the same residues on both chains (Fig. 5.12 A/B). In the structure from the B5 dataset, which crystallised in the presence of phosphate rather than sulphate, densities corresponding to phosphate can be observed at the same locations where the sulphates in H4 are found. In addition, however, two further clear densities corresponding to two additional phosphates are also present (Fig. 5.12 C/D), bringing the total number of phosphates per asymmetric unit to six.

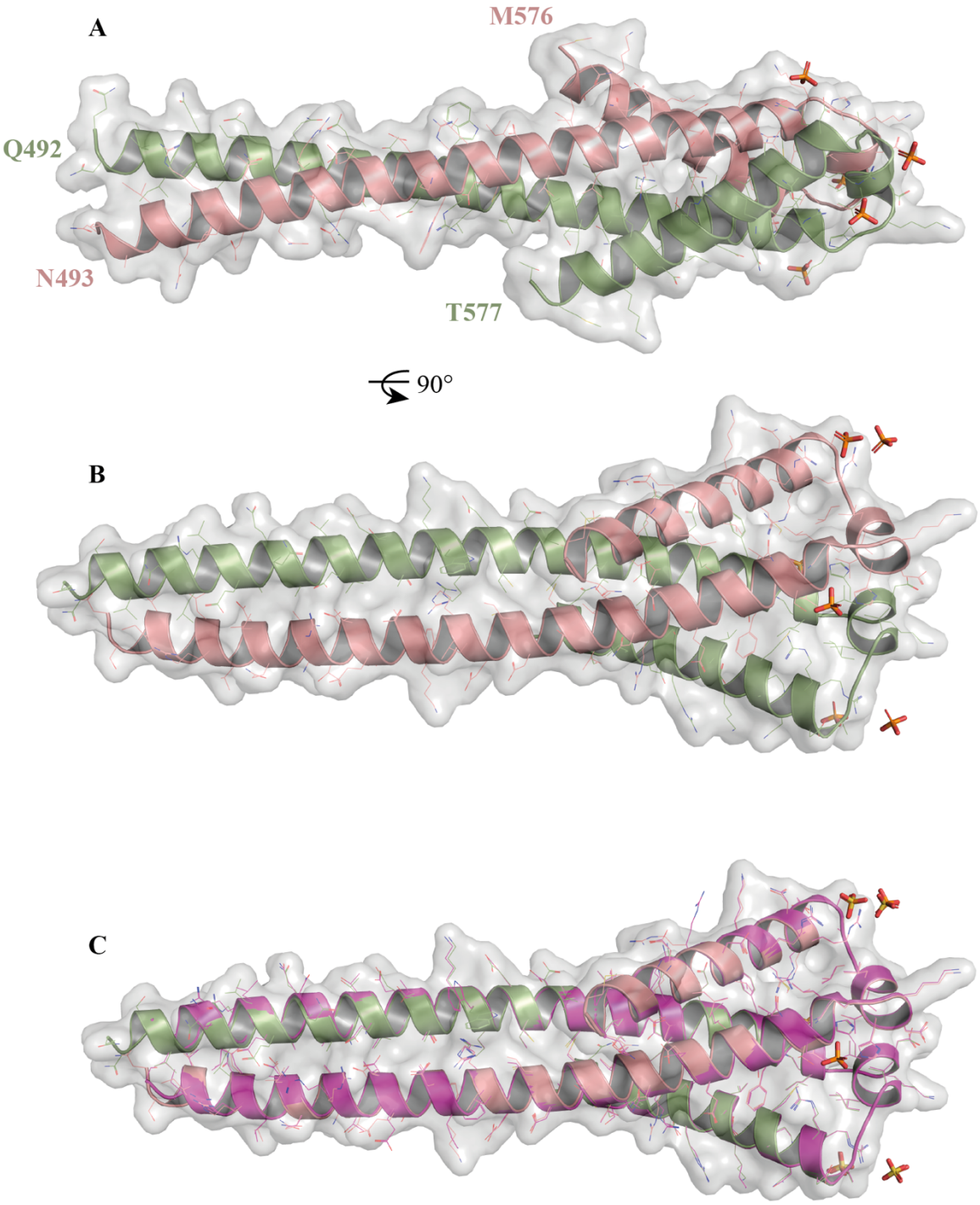


Figure 5.11 | Cnn$^{D490-N579}$ forms a dimer of helical hairpins *in crystallo*. A/B) The structure (B5 dataset) is shown at two orientations in cartoon representation with side chains shown as lines and surface shown in grey at 60 % transparency; phosphates are shown as sticks. Chain A is shown in salmon, chain B is shown in green. Terminal residues of both chains are indicated **C)** The B5 structure as shown in B is overlaid with the structure of Cnn$^{D490-N579}$ from the H4 dataset in pink cartoon representation with side chains shown as lines and sulphates shown as sticks. Note Cnn$^{D490-N579}$ adopts the same fold in both datasets.

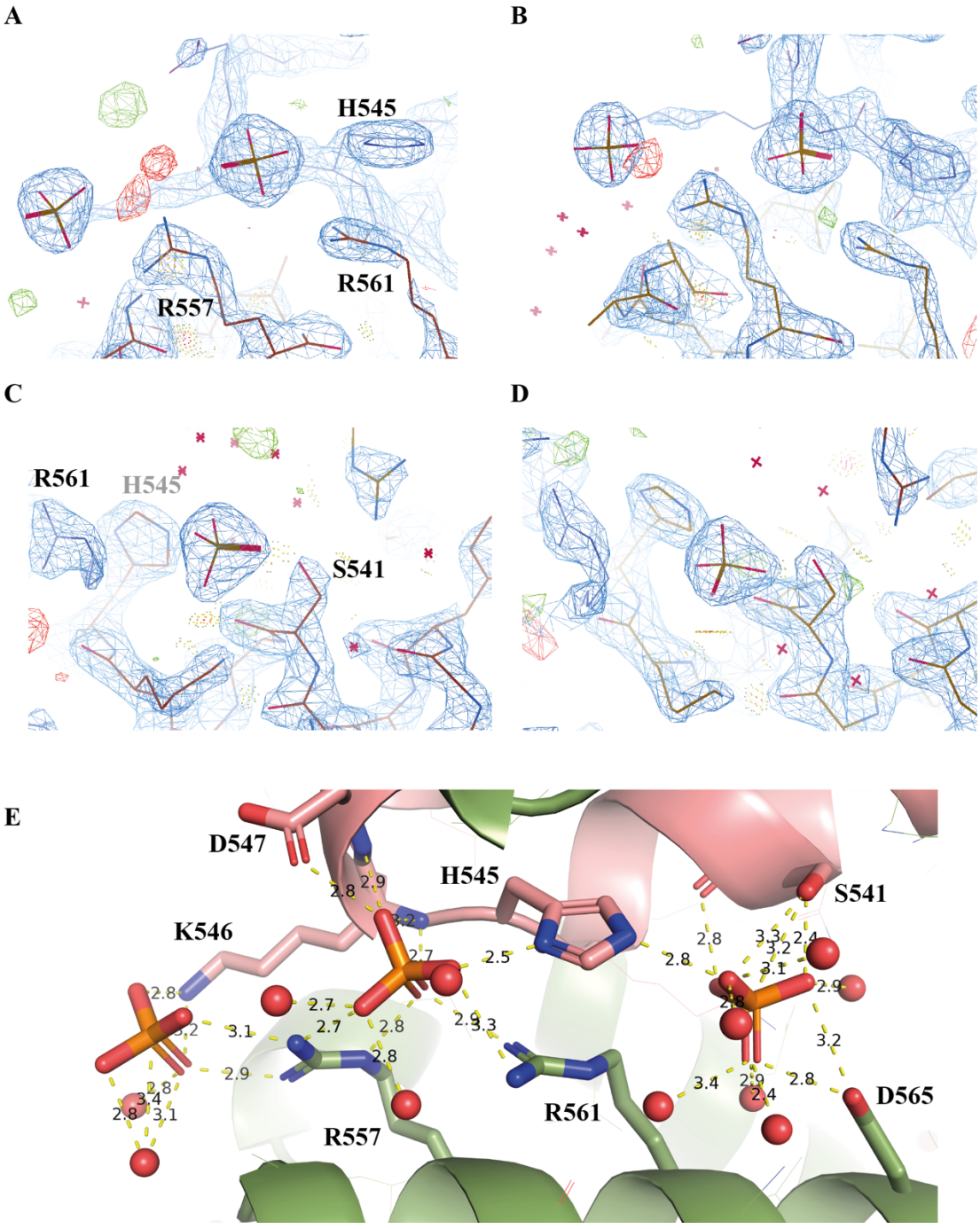


Figure 5.12 | Phosphate/sulphate coordination in the Cnn$^{D490-N579}$ crystal structure. A/B) The four sulphates observed in the Cnn$^{D490-N579}$ structure (H4 dataset) are shown. (A) shows R557 and R561 from chain A, (B) shows R557 and R561 from chain B from a similar angle. **C/D)** The two additional phosphates observed in the Cnn$^{D490-N579}$ structure (B5 dataset) are shown. (C) shows S541 from chain A, (D) shows S541 from chain B. In (A-D), the electron density map is contoured at 3.0 σ. **E)** Detail showing polar interaction network of three of the six phosphates observed in the Cnn$^{D490-N579}$ structure (B5 dataset). Cnn$^{D490-N579}$ Chain A is shown in salmon, Cnn$^{D490-N579}$ chain B from another crystallographic copy is shown in green. Main chains are shown in cartoon representation or sticks, with side chains shown as sticks or lines, phosphates shown as sticks, and phosphate-coordinating water molecules as balls. Interatomic distances are shown in Å. Residues involved in phosphate coordination are highlighted.

These six phosphates sit at the interface between two neighbouring crystallographic copies of the $Cnn^{D490\text{-}N579}$ dimer, with three phosphates each coordinated by the two chains. An extensive polar interaction network directly connects these three phosphates with eleven water molecules, the side chains of R557 and R561 from one chain of a $Cnn^{D490\text{-}N579}$ dimer, and side chains as well as backbone peptide bonds of residues S541, H545 and K546 from another chain of another $Cnn^{D490\text{-}N579}$ dimer. The negatively charged side chains of D565 from one chain and D547 from another chain delineate this polar, positively charged, phosphate binding pocket between two dimers of $Cnn^{D490\text{-}N579}$ (Fig. 5.12 E). Note that $Cnn^{D490\text{-}N579}$ only crystallised under conditions that contained either sulphate or phosphate, and it is therefore reasonable to assume that the crystallographic contacts observed must necessarily be stabilised by phosphate or sulphate (Fig. 5.13).

Overlaying the structure of $Cnn^{D490\text{-}N579}$ with the previously solved structure of the complex between $Cnn^{D490\text{-}K544}$ and $Cnn^{H1082\text{-}L1148}$ (PDB code: 5MVW, [339]) shows that the hairpin fold adopted by $Cnn^{D490\text{-}N579}$ sterically impedes binding of $Cnn^{H1082\text{-}L1148}$. The linker and the C-terminal helix of $Cnn^{D490\text{-}N579}$ occlude the binding site of $Cnn^{H1082\text{-}L1148}$ by folding against the N-terminal helix of $Cnn^{D490\text{-}N579}$ in almost exactly the position where $Cnn^{H1082\text{-}L1148}$ is observed to bind in the complex structure (Fig. 5.14). $Cnn^{H1082\text{-}L1148}$ therefore cannot bind to the zipper-like motif as observed in the complex structure unless the α-helical hairpin regions C-terminal to the zipper-like motif undergo a conformational change that opens the hairpin.

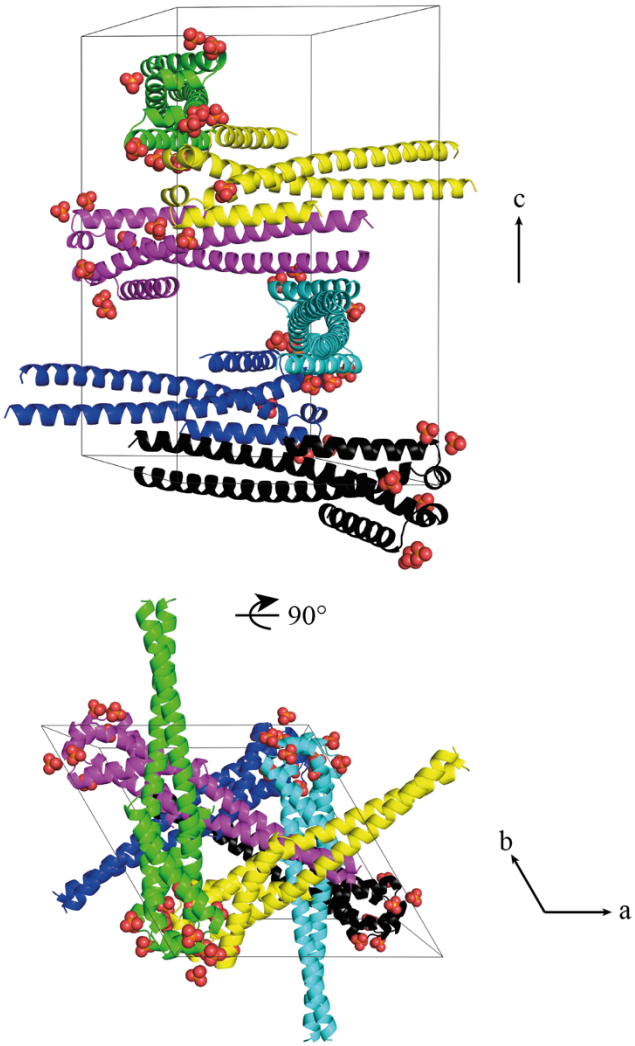


Figure 5.13 | Cnn$^{D490\text{-}N579}$ crystal lattice arrangement. Unit cell and crystal lattice (space group $P6_1$) of Cnn$^{D490\text{-}N579}$ (B5 dataset) from two different orientations rotated by 90°, with each asymmetric unit in a different colour. Phosphates are shown as spheres; protein chains are shown in cartoon representation. Unit cell axes are indicated. The unit cell contains six dimers of Cnn$^{D490\text{-}N579}$.

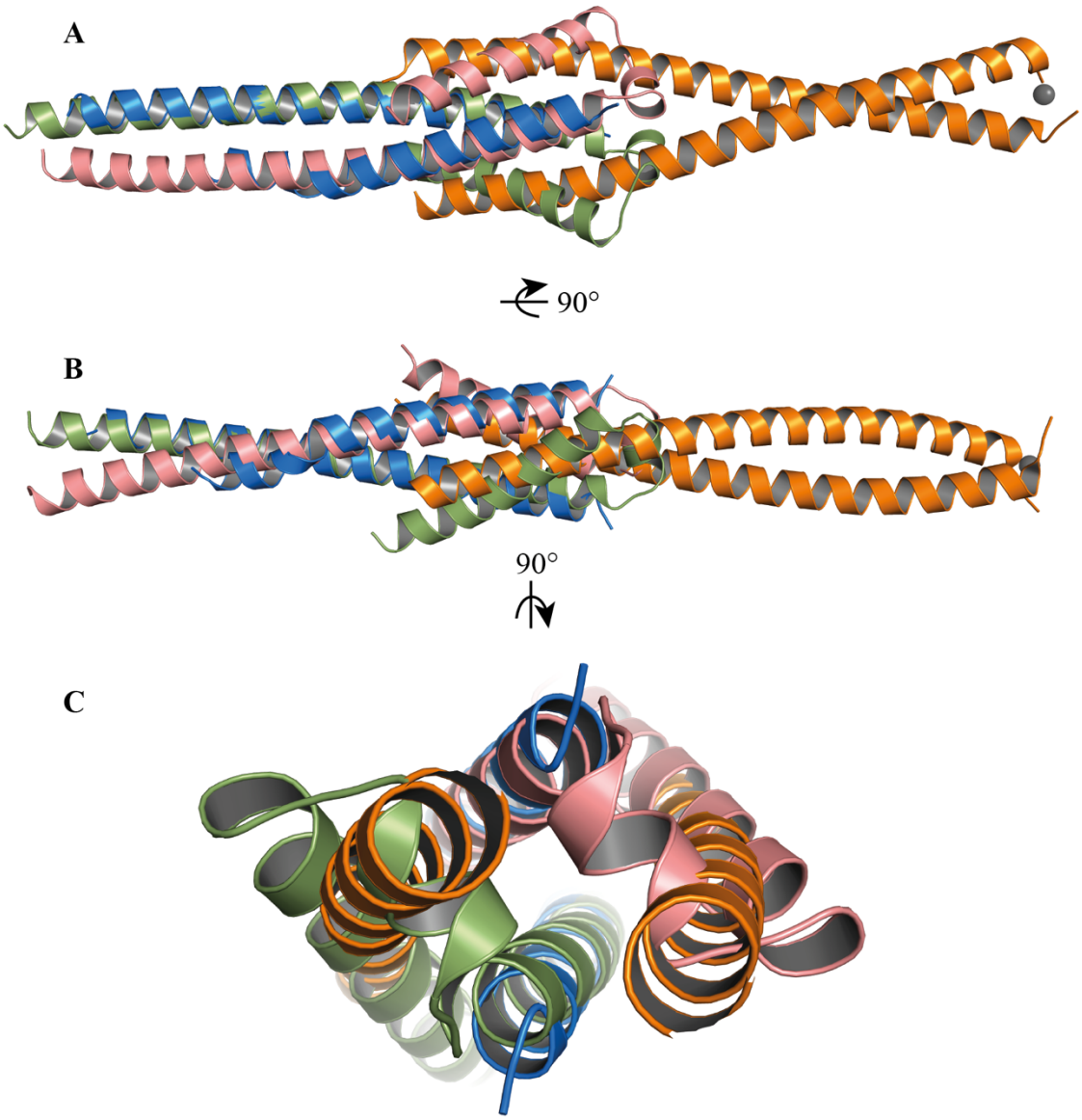


Figure 5.14 | The Cnn$^{D490\text{-}N579}$ hairpin fold sterically impedes binding of the Cnn C-terminus. Cnn$^{D490\text{-}N579}$ (B5 dataset) chains A and B are shown in salmon and green, respectively. The structure is overlaid with the Cnn$^{D490\text{-}K544}$:Cnn$^{H1082\text{-}L1148}$ complex structure (PDB code: 5MVW, [339]), with Cnn$^{D490\text{-}K544}$ in blue and Cnn$^{H1082\text{-}L1148}$ in orange. All chains are shown in cartoon representation. The zinc ion is shown as a grey sphere. The overlay is shown from three different orientations (A-C).

The C-terminal helix of $Cnn^{D490\text{-}N579}$ contains four conserved Ser/Thr residues, i.e. S567, S571, S573 and T577 (Fig. 5.1). Three of these (S567, S571 and T577) closely match the Polo consensus phosphorylation target sequence (see above). Inspection of the structure of $Cnn^{D490\text{-}N579}$ reveals that S567 and S571 are located on the side of the C-terminal helix that is buried in the hairpin fold, while S573 and T577 are on the other, surface-exposed side of the helix (Fig. 5.15 A-C). This suggests that phosphorylation on S567 or S571 would necessarily open up the hairpin fold due to steric and probably also charge repulsion, while phosphorylation on S573 or T577 would not disrupt the $Cnn^{D490\text{-}N579}$ fold as observed *in crystallo*. These observations suggest that the double hairpin fold observed *in crystallo* cannot possibly exist at the centrosome, where S567 is known to be phosphorylated [196, 339]. It is therefore likely that the structure of $Cnn^{D490\text{-}N579}$ presented here represents the cytosolic inactive form of Cnn, where the protein is not phosphorylated on S567.

Polo kinase is likely able to phosphorylate S567 and potentially also S571 despite these residues being buried in the hairpin structure because the C-terminus of $Cnn^{D490\text{-}N579}$ shows a high degree of flexibility and might therefore temporarily expose parts of the usually buried side of the C-terminal helix to the solvent (Fig. 5.15 D). This is probably also the case in context of the entire Cnn protein *in vivo*, as the region immediately C-terminal to what is observed in the crystal structure is predicted to form a random coil (Fig. 5.2) and therefore also likely to show a high degree of flexibility.

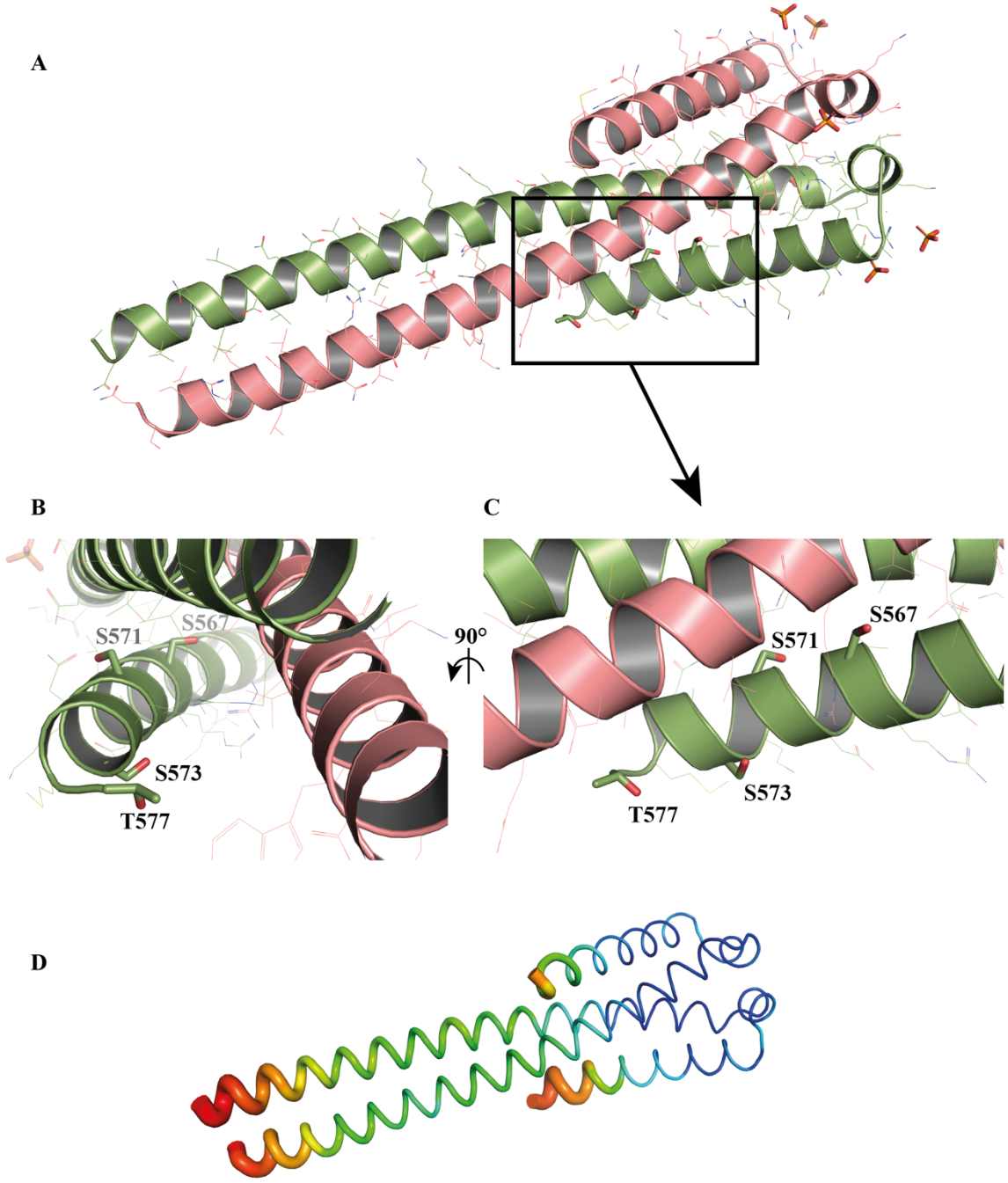


Figure 5.15 | Phosphorylation target site S567 is buried in the hairpin structure. **A)** $Cnn^{D490\text{-}N579}$ (B5 dataset) structure is shown as in Fig. 5.11 A/B but without surface. In addition, side chains of residues S567, S571, S573 and T577 from chain B are shown as sticks. **B/C)** Detail from two orientations, showing how S567 and S571 are buried in the hairpin structure, while S573 and T577 are exposed at the surface. **D)** B-factor putty model generated via PyMOL, showing $Cnn^{D490\text{-}N579}$ (B5 dataset) in the same orientation as in (A). Residues with a lower B-factor, i.e. little flexibility, are highlighted as a thin blue tube; residues with a higher B-factor, i.e. greater flexibility, are highlighted as a thick tube in green-yellow-red at increasing flexibility. Note flexibility is maximal at the chain termini.

5.5 Residues S567 and S571 can be phosphorylated *in vitro* in the absence of the Cnn C-terminus

The structure of $Cnn^{D490\text{-}N579}$ suggests a mechanism wherein phosphorylation on S567 (and possibly also S571) destabilises the observed double hairpin fold, thereby facilitating binding of the Cnn C-terminus and hence formation of assembly-competent Cnn. If phosphorylation is to precede binding of the Cnn C-terminus, Polo kinase must be able to phosphorylate $Cnn^{D490\text{-}N579}$ on S567 in the absence of the Cnn C-terminus. A structural mechanism for how this is possible is given above, wherein the extreme C-terminus of $Cnn^{D490\text{-}N579}$ shows high flexibility that might confer a degree of accessibility to S567 in context of the double hairpin fold.

In order to determine whether $Cnn^{D490\text{-}N579}$ could indeed be phosphorylated on residue S567 in the absence of the Cnn C-terminus, a 100 µl phosphorylation reaction was set up in which $Cnn^{D490\text{-}N579}$ at a concentration of 100 µM was incubated with PLK1 according to the usual protocol. As a control, $Cnn^{D490\text{-}N579}$ was diluted to 100 µM with buffer containing 25 mM Tris-HCl (pH 7.5), 150mM NaCl, 2mM DTT and also incubated for 30 min at 30°C. Samples from both reactions were then run on a 4-20 % reducing SDS-PAGE gel and the appropriate gel bands excised and sent to the Oxford Proteomics Facility, where further experiments were carried out by Dr Svenja Hester. These involved digestion of $Cnn^{D490\text{-}N579}$ by chymotrypsin or trypsin followed by fragmentation and mass spectrometry analysis.

Fragments from all Ser and Thr residues in $Cnn^{D490\text{-}N579}$ were observed in these experiments. Indeed, it was shown that the three consensus targets S567, S571, T577 as well as, surprisingly, the non-consensus target S573 could be phosphorylated under these conditions

by human recombinant PLK1 *in vitro* (Appendix Figure 2). This demonstrates that not only are the surface-exposed residues S573 and T577 accessible to kinase-mediated phosphorylation in solution but also the less exposed residues S567 and S571.

5.6 Phospho-mimicking mutations of S567/S571 disrupt the double hairpin fold in solution

The effect of phosphorylation of individual residues on the overall structure of $Cnn^{D490\text{-}N579}$ cannot be investigated via the phosphorylation assay described above, as it cannot be controlled which residues are phosphorylated in a given experiment. It was therefore decided to design and purify mutants instead which have phospho-mimicking mutations (i.e. glutamate) either of the inside-facing residues S567 and S571 or the outwards-facing residues S573 and T577. It would be expected that S567E/S571E mutation disrupts the hairpin fold, while S573E/T577E has no effect.

The two proteins were cloned and purified according to the usual protocol, with a Ni-NTA Superflow Cartridge CV of 5 ml. In order to avoid having to concentrate the protein using a spin filter and therefore risking precipitation, the most highly concentrated 5 ml of elute from Ni-affinity chromatography were directly loaded on a S200 16/600 column for SEC. For both proteins, two peaks could be observed, with the first peak corresponding to the mNG-$Cnn^{D490\text{-}N579}$ mutant and the second peak containing the mNG-tag alone, according to SDS-PAGE analysis (Figs. 5.16 and 5.17).

The elution peak corresponding to mNG-$Cnn^{D490\text{-}N579\ (S567E/S571E)}$ (Fig. 5.16) was pooled (CV = 0.60-0.65), concentrated to 16.7 mg/ml and frozen in 55 µl aliquots, resulting in a final yield of ~ 10.1 mg/L. The elution peak corresponding to the other mutant, i.e. mNG-$Cnn^{D490\text{-}N579\ (S573E/T577E)}$ (Fig. 5.17) was also pooled (CV = 0.66-0.71). This construct was concentrated to 11.6 mg/ml and frozen in aliquots of the same volume as above, giving a final yield of ~ 6.1 mg/L.

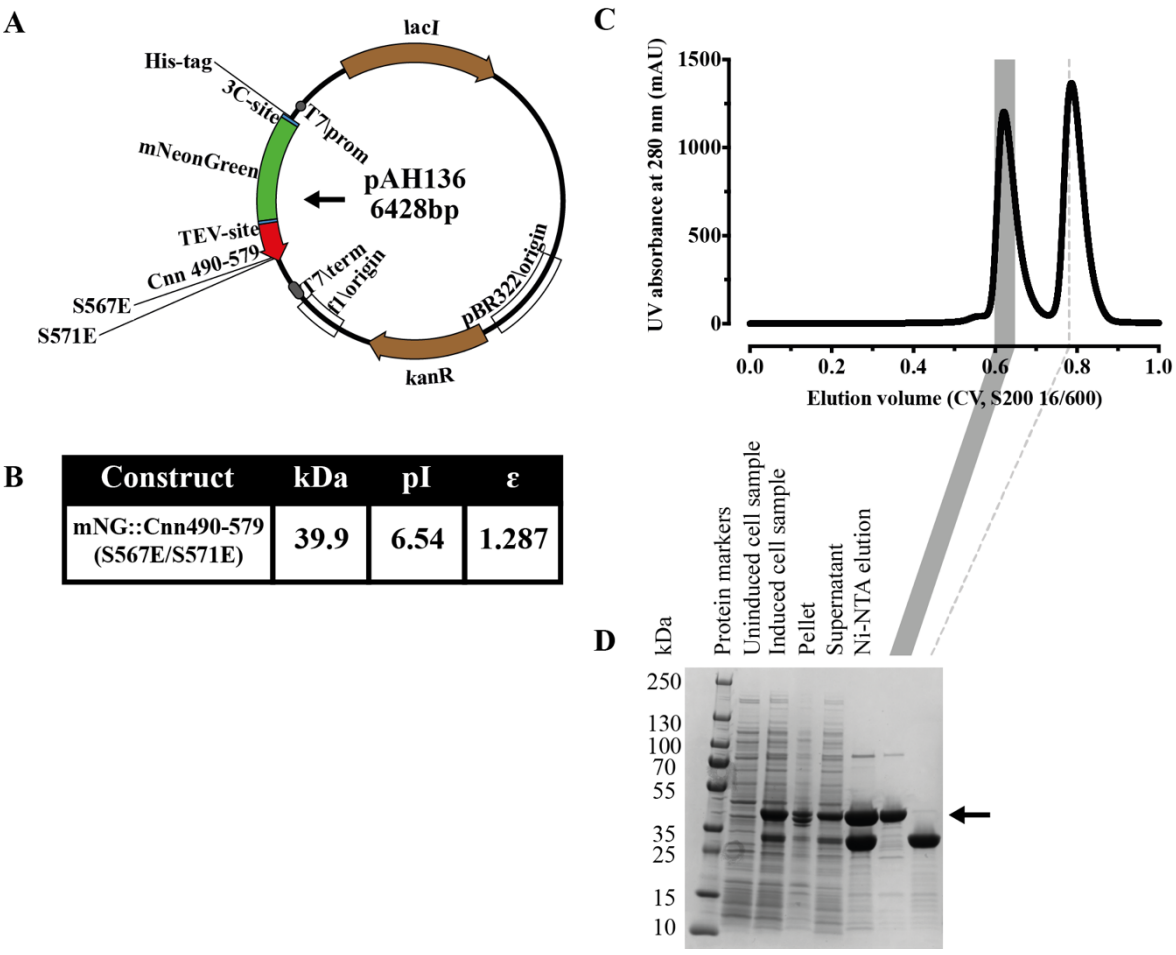


Figure 5.16 | Purification of N-terminally mNG-tagged Cnn$^{D490\text{-}N579\ (S567E/S571E)}$. **A)** Overview of plasmid pAH136, which encodes Cnn$^{D490\text{-}N579\ (S567E/S571E)}$ with an N-terminal, TEV-cleavable, His-3C-mNG-domain. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged protein. **C)** SEC (S200 16/600) elution profile of mNG-Cnn$^{D490\text{-}N579\ (S567E/S571E)}$ after Ni-affinity chromatography, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.60-0.65) indicates the fractions pooled after SEC and run on the SDS-PAGE gel in D, containing mNG-Cnn$^{D490\text{-}N579\ (S567E/S571E)}$. A further fraction run on the gel in D is indicated by a dashed grey line, likely corresponding to the mNG tag alone. SEC was carried out using standard TBS. **D)** 4-20 % reducing SDS-PAGE gel showing successive purification steps; the band corresponding to mNG-Cnn$^{D490\text{-}N579\ (S567E/S571E)}$ is highlighted with a black arrow.

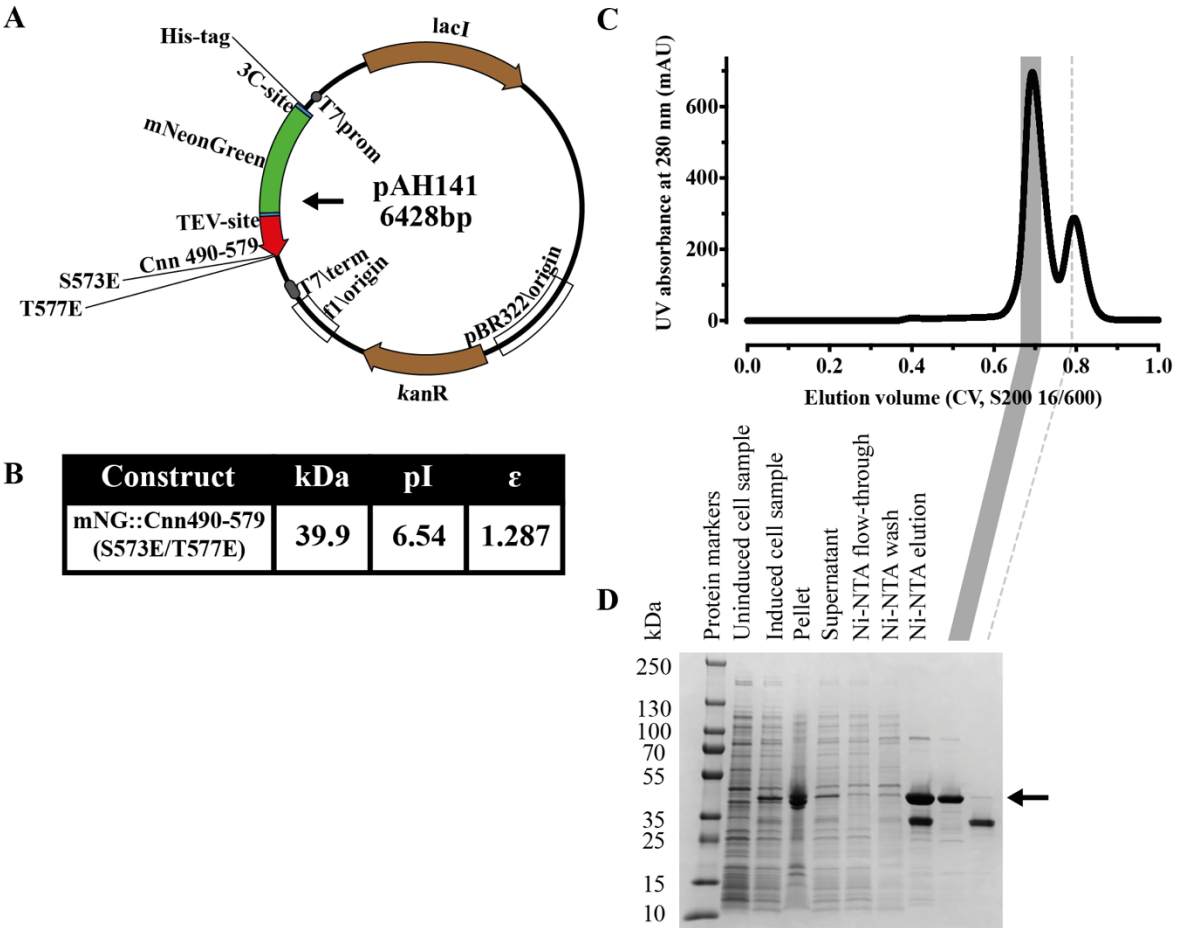


Figure 5.17 | Purification of N-terminally mNG-tagged Cnn$^{D490\text{-}N579\ (S573E/T577E)}$. **A)** Overview of plasmid pAH141, which encodes Cnn$^{D490\text{-}N579\ (S573E/T577E)}$ with an N-terminal, TEV-cleavable, His-3C-mNG-domain. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged protein. **C)** SEC (S200 16/600) elution profile of mNG-Cnn$^{D490\text{-}N579\ (S573E/T577E)}$ after Ni-affinity chromatography, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.66-0.71) indicates the fractions pooled after SEC and run on the SDS-PAGE gel in D, containing mNG-Cnn$^{D490\text{-}N579\ (S573E/T577E)}$. A further fraction run on the gel in D is indicated by a dashed grey line, likely corresponding to the mNG tag alone. SEC was carried out using standard TBS. **D)** 4-20 % reducing SDS-PAGE gel showing successive purification steps; the band corresponding to mNG-Cnn$^{D490\text{-}N579\ (S573E/T577E)}$ is highlighted with a black arrow.

Hairpin disruption can be scored by investigating the oligomeric state of the $Cnn^{D490\text{-}N579}$ mutants in solution. As has been shown above, mNG-$Cnn^{D490\text{-}N579}$ is dimeric, while the shorter construct mNG-$Cnn^{D490\text{-}S567}$, which probably cannot fully form a hairpin due to its truncated C-terminus, is tetrameric in solution (Fig. 5.8). Tetramerisation probably occurs via the same interface as binding of the Cnn C-terminus, as mutations that disrupt $Cnn^{D490\text{-}K544}$ tetramerisation also disrupt binding of $Cnn^{H1082\text{-}L1148}$ (Fig. 3.6). Hairpin formation therefore probably prevents tetramerisation via the same steric clash described above for the binding of the Cnn C-terminus (Fig. 5.14). Disruption of the hairpin structure should hence lead to tetramerisation of the construct. SEC-MALS was therefore carried out on both mutant mNG-$Cnn^{D490\text{-}N579}$ constructs to determine their oligomeric state in standard TBS (Fig. 5.18).

Both constructs were tested at two different concentrations, i.e. 25 and 75 µM, in order to account for any possible concentration-dependent effects. At both concentrations mNG-$Cnn^{D490\text{-}N579\ (S573E/T577E)}$ appeared to be mostly dimeric in solution, with apparent masses of 85.3 kDa (+/- 0.9 %) and 90.7 kDa (+/- 0.9 %), respectively. Only a minor earlier-eluting shoulder could be observed at the higher concentration (Fig. 5.18 A). As expected, mutating the solvent-facing residues S573 and T577 to phospho-mimicking residues therefore does not affect the oligomeric state of $Cnn^{D490\text{-}N579}$.

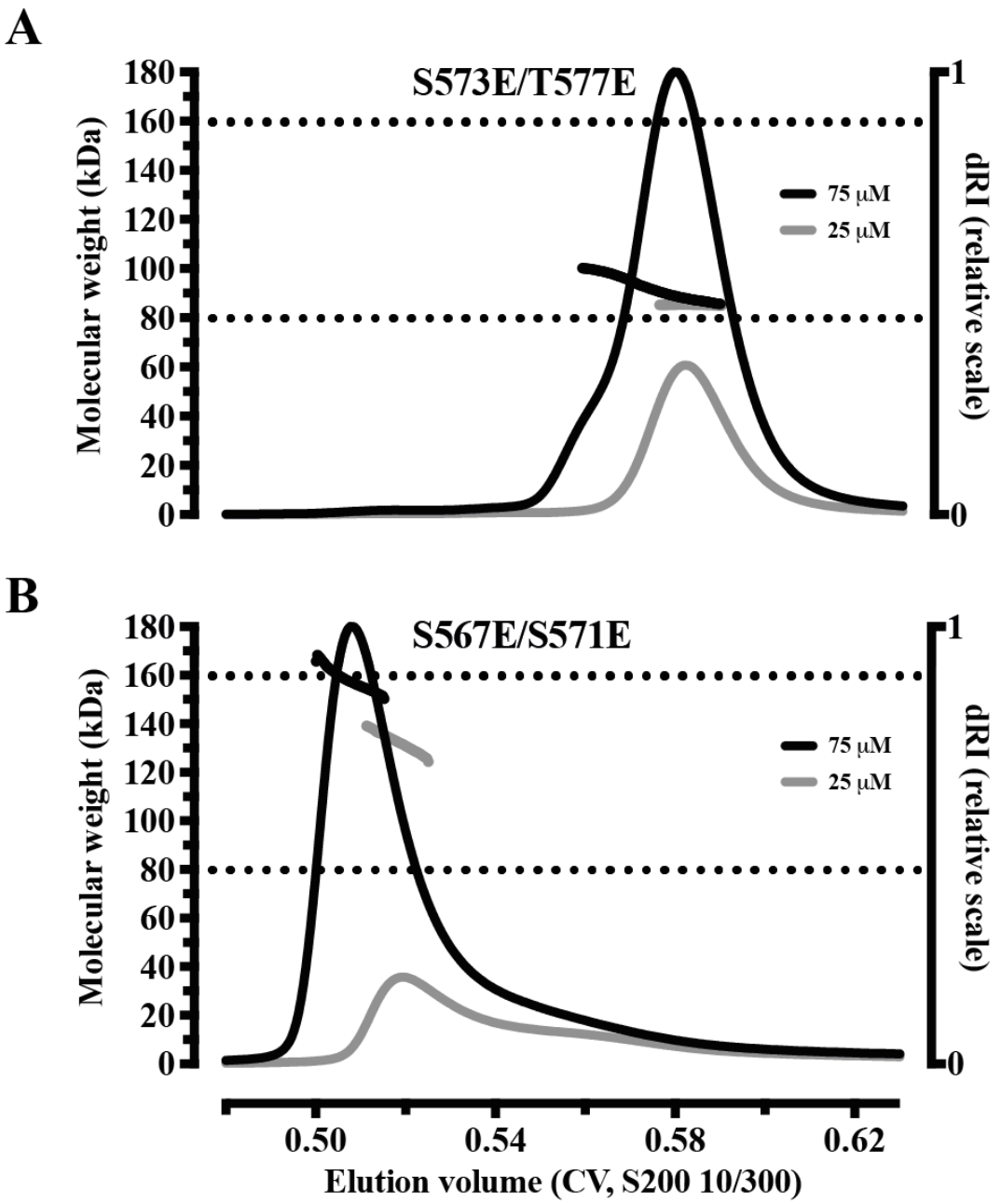


Figure 5.18 | SEC-MALS analysis of two mutants of mNG-Cnn$^{D490\text{-}N579}$. A) The average molecular weight in solution of mNG-Cnn$^{D490\text{-}N579\ (S573E/T577E)}$ was measured at 25 µM (light grey: 85.3 kDa +/- 0.9 %) and 75 µM (black: 90.7 kDa +/- 0.9 %). Note the construct is mostly dimeric. **B)** The average molecular weight in solution of mNG-Cnn$^{D490\text{-}N579\ (S567E/S571E)}$ was measured at 25 µM (light grey: 132.0 kDa +/- 0.8 %) and 75 µM (black: 157.6 kDa +/- 0.9 %). Note the construct tends towards a tetramer in a concentration-dependent manner. In both (A) and (B), dotted lines indicate theoretical molecular weight expected for tetrameric or dimeric mNG-Cnn$^{D490\text{-}N579}$ mutants. Molecular weight (in kDa) is plotted on the left y-axis, change in refractive index (dRI, relative scale) plotted on the right y-axis, column elution volume (CV, 24 ml) of column S200 10/300 increase plotted on the x-axis. Standard TBS was used as buffer.

In contrast, at 25 µM mNG-Cnn$^{D490\text{-}N579\ (S567E/S571E)}$ formed oligomers with an average molecular weight of 132.0 kDa (+/- 0.8 %), i.e. slightly higher than that expected for a trimer (Fig. 5.18 B). As these are average masses, it is likely that at this concentration, mNG-Cnn$^{D490\text{-}N579\ (S567E/S571E)}$ consists of a mixture of dimers and tetramers, thereby resulting in an average mass most closely resembling that of a trimer. Indeed, at the higher concentration (75 µM), mNG-Cnn$^{D490\text{-}N579\ (S567E/S571E)}$ was completely shifted to a tetrameric state with an average mass of 157.6 kDa (+/- 0.9 %). Introducing phospho-mimicking mutations at the inward-facing residues S567 and S571 therefore induces tetramerisation of Cnn$^{D490\text{-}N579}$, probably due to a disruption of the hairpin fold.

In order to test binding of these mNG-Cnn$^{D490\text{-}N579}$ mutant constructs to the Cnn C-terminus, a further SEC-MALS experiment was carried out under the same conditions as above, with the corresponding mNG-Cnn$^{D490\text{-}N579}$ mutant at 25 µM. But in addition, Cnn$^{G1077\text{-}L1148}$ was added at a fivefold molar excess (Fig. 5.19). No difference between the two mNG-Cnn$^{D490\text{-}N579}$ mutants could be observed under the conditions tested, with both constructs forming 2:2 complexes with Cnn$^{G1077\text{-}L1148}$. The measured molecular weights were shown to closely match each other, with a complex mass of 104.3 kDa (+/- 0.9 %) for mNG-Cnn$^{D490\text{-}N579\ (S567E/S571E)}$ and Cnn$^{G1077\text{-}L1148}$ and a complex mass of 104.6 kDa (+/- 1.0 %) for mNG-Cnn$^{D490\text{-}N579\ (S573E/T577E)}$ and Cnn$^{G1077\text{-}L1148}$ (Fig. 5.19). Cnn$^{G1077\text{-}L1148}$ therefore forms 2:2 complexes with both mNG-Cnn$^{D490\text{-}N579}$ mutants, further confirming that, even in the presence of phospho-mimicking mutations, the region Cnn$^{D490\text{-}N579}$ is not sufficient to allow network assembly, i.e. residues C-terminal to N579 are required for network assembly upon complexation between the Cnn C-terminus and the zipper-like motif.

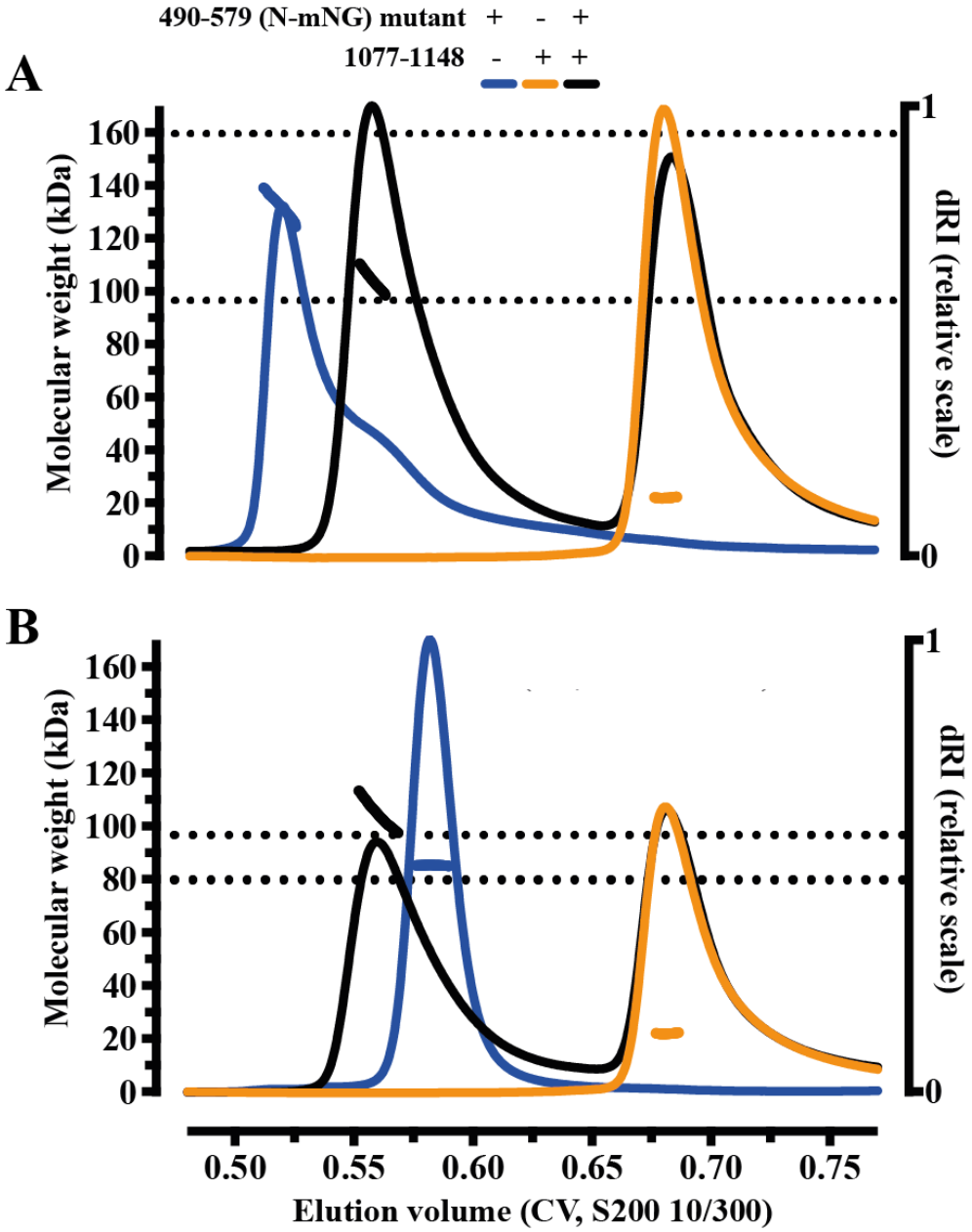


Figure 5.19 | Both mNG-Cnn$^{D490\text{-}N579}$ mutants bind the Cnn C-terminus under *in vitro* conditions. **A)** The average molecular weight in solution of mNG-Cnn$^{D490\text{-}N579\ (S567E/S571E)}$ at 25 µM mixed with a fivefold molar excess of Cnn$^{G1077\text{-}L1148}$ was measured (black: 104.3 kDa +/- 0.9 %). Dotted lines indicate theoretical molecular weight expected for a tetramer of mNG-Cnn$^{D490\text{-}N579\ (S567E/S571E)}$ (top) and a 2:2 complex between the two constructs (bottom). **B)** The average molecular weight in solution of mNG-Cnn$^{D490\text{-}N579\ (S573E/T577E)}$ at 25 µM mixed with a fivefold molar excess of Cnn$^{G1077\text{-}L1148}$ was measured (black: 104.6 kDa +/- 1.0 %). Dotted lines indicate theoretical molecular weight expected for a 2:2 complex between the two constructs (top) and a dimer of mNG-Cnn$^{D490\text{-}N579\ (S573E/T577E)}$ (bottom). (A) and (B) also show the corresponding controls, i.e. either mNG-Cnn$^{D490\text{-}N579}$ mutant alone (blue) from Fig. 5.18, or Cnn$^{G1077\text{-}L1148}$ alone (orange). Molecular weight (in kDa) is plotted on the left y-axis, change in refractive index (dRI, relative scale) plotted on the right y-axis, column elution volume (CV, 24 ml) of column S200 10/300 increase plotted on the x-axis. Standard TBS was used as buffer. Note both mNG-Cnn$^{D490\text{-}N579}$ mutants form 2:2 complexes with Cnn$^{G1077\text{-}L1148}$.

5.7 Phospho-null mutations of S567/S571/S573 do not affect the structure of Cnn$^{D490\text{-}N579}$ *in crystallo*

5.7.1 Phospho-null Cnn$^{D490\text{-}N579}$ purification and crystallisation

Drosophila embryos with Cnn carrying mutations that convert the three residues S567, S571 and S573 to alanine have severely disrupted centrosomes [196]. However, it is unclear whether network assembly fails because these mutations have an adverse impact on the folding of Cnn, or because network assembly fails because phosphorylation on S567 (and possibly the other two sites) is essential but not possible upon introduction of these phospho-null mutations. A structural approach was therefore taken in order to exclude the possibility that these mutations adversely affect the structure of Cnn.

Site-directed mutagenesis was used to clone a Cnn$^{D490\text{-}N579}$ construct that contained the three mutations S567A, S571A and S573A (Fig. 5.20 A/B). The protein was produced and purified according to the usual protocol and using 1.5 ml TEV protease (1.25 mg/ml stock) for overnight tag cleavage. After two rounds of reverse Ni-affinity chromatography, the protein was concentrated to 5 ml and applied to a S75 16/600 column pre-equilibrated in standard TBS, resulting in a profile that had one clear peak with a trailing shoulder. According to SDS-PAGE analysis, both the peak and the shoulder contained Cnn$^{D490\text{-}N579\ (S567A/S571A/S573A)}$ (i.e. phospho-null Cnn$^{D490\text{-}N579}$). However, in order to ensure sample homogeneity prior to structural studies, the peak was pooled very conservatively (CV = 0.55-0.59) and the shoulder thereby excluded from the final protein sample (Fig. 5.20 C/D).

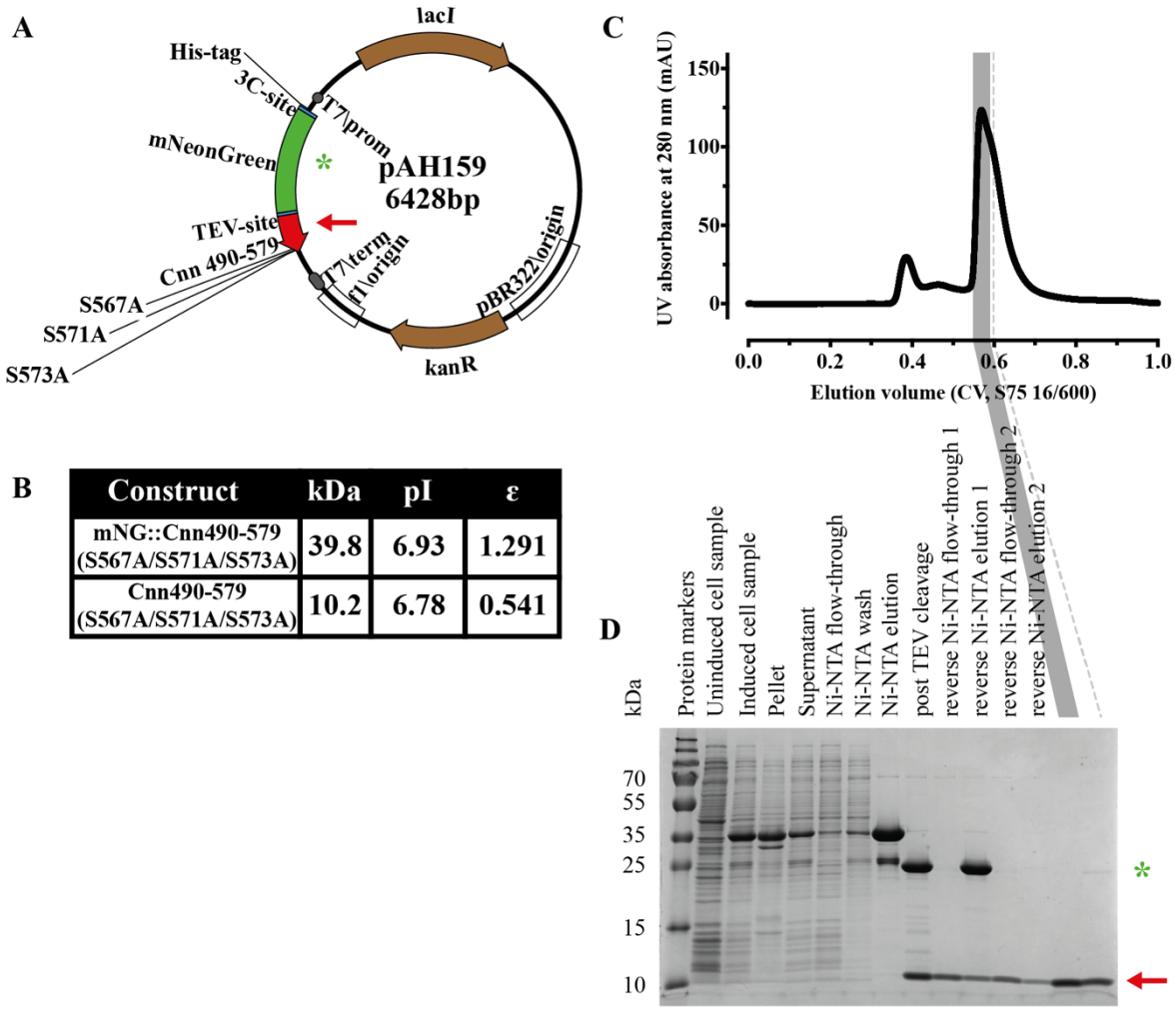


Figure 5.20 | Purification of Cnn$^{D490\text{-}N579\ (S567A/S571A/S573A)}$. **A)** Overview of plasmid pAH159, which encodes Cnn$^{D490\text{-}N579\ (S567A/S571A/S573A)}$ with an N-terminal, TEV-cleavable, His-3C-mNG-domain. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged and untagged protein. **C)** SEC (S75 16/600) elution profile of Cnn$^{D490\text{-}N579\ (S567A/S571A/S573A)}$ after tag cleavage, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.55-0.59) indicates the fractions pooled after SEC and run on the SDS-PAGE gel in D, containing Cnn$^{D490\text{-}N579\ (S567A/S571A/S573A)}$. A further fraction run on the gel is indicated with a grey dashed line. SEC was carried out using standard TBS. **D)** 4-20 % reducing SDS-PAGE gel showing successive purification steps. The band corresponding to Cnn$^{D490\text{-}N579\ (S567A/S571A/S573A)}$ is marked with a red arrow, the band corresponding to the mNG-tag is indicated with a green asterisk.

The pooled peak containing pure phospho-null $Cnn^{D490\text{-}N579}$ was concentrated to around 13-15 mg/ml when the protein started heavily precipitating. The precipitant was spun down and the protein diluted using standard TBS to a final concentration of 11.7 mg/ml to avoid further precipitation.

Without freezing the protein, two crystallisation trays (Stura and Structure Screen I +II) were immediately set up with phospho-null $Cnn^{D490\text{-}N579}$ at this stock concentration. The rest of the protein was frozen in 45 µl aliquots, with a final yield of ~ 2.1 mg/L, also counting the protein already used in crystallisation trials. After two weeks, small crystals could be observed in condition B11 of Structure Screen I + II (Molecular Dimensions), which contained 0.2 M $CaCl_2$, 0.1 M sodium HEPES (pH 7.5), 28 % (v/v) PEG 400.

In an attempt to grow larger crystals under similar conditions, a custom screen was made which had a linear horizontal gradient of PEG 400 from 20 % (v/v) to 40 % (v/v), as well as a linear vertical gradient of $CaCl_2$ from 0 to 0.4 M. The concentration and pH of the buffer (0.1 M sodium HEPES, pH 7.5) were not varied (Fig. 5.21). Crystals of phospho-null $Cnn^{D490\text{-}N579}$ were allowed to grow in this optimised crystallisation screen at the same concentration as above over a period of two months. This yielded dozens of crystals, mainly in the top left quadrant, i.e. at higher concentrations of $CaCl_2$ but lower concentrations of PEG 400 compared to the initial condition from Structure Screen I + II.

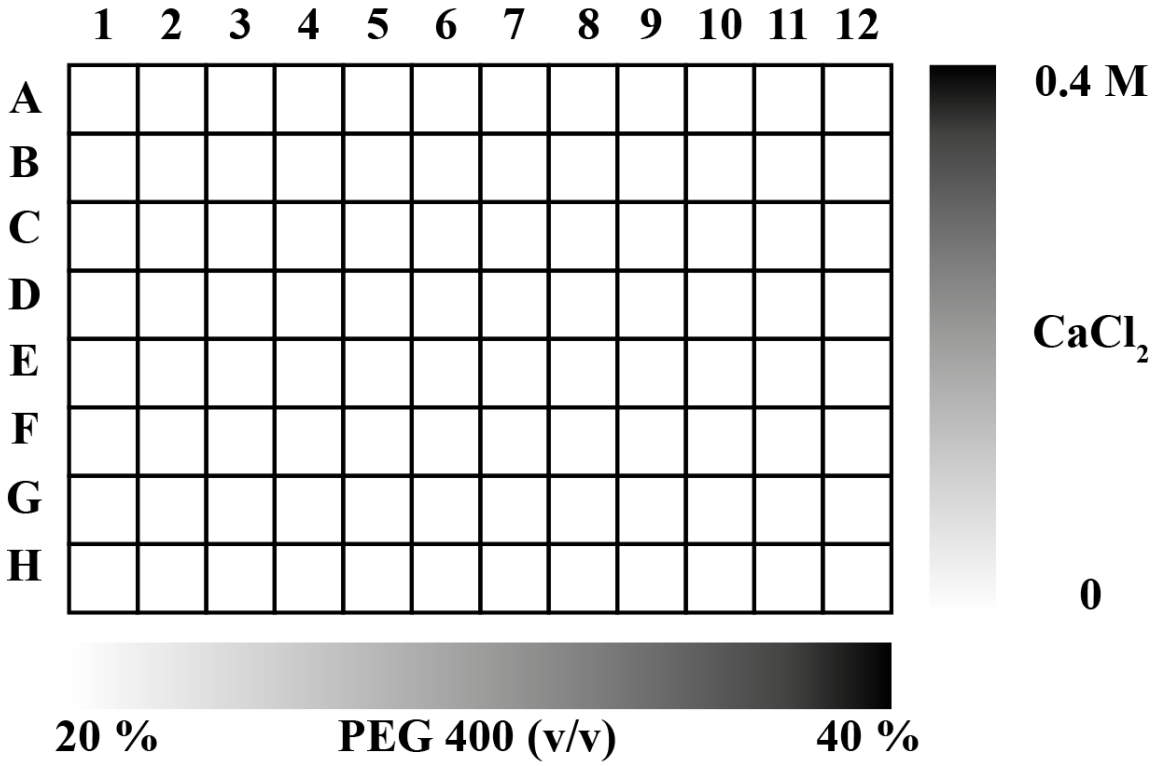


Figure 5.21 | Phospho-null custom crystallisation screen. Every well also contained 0.1 M sodium HEPES (pH 7.5). Stock concentrations were 1 M sodium HEPES (pH 7.5), 100 % (v/v) PEG 400 and 1 M $CaCl_2$. Water was used as diluent.

Beamline	DIAMOND i04	
Wavelength (Å)	0.9795	
Resolution limits (Å)	50.31-2.91 (2.96-2.91)	76.52-3.03 (3.08-3.03)
Space group	$P222_1$	P1
Unit cell dimensions (Å, º)	102.00, 115.67, 47.09 90, 90, 90	47.12, 102.00, 115.68 90, 90, 90
Unique reflections	12739 (632)	41215 (2011)
Multiplicity	9.9 (9.8)	2.7 (2.8)
Completeness (%)	99.2 (99.1)	98.6 (95.7)
I/ σ(I)	12.4 (1.1)	4.7 (1.0)
R_{merge} (%)	10.1 (177.5)	6.1 (64.2)
R_{meas} (%)	10.7 (187.3)	7.6 (79.2)
R_{pim} (%)	3.3 (59.0)	4.4 (45.6)
$CC_{1/2}$	1.0 (0.6)	1.0 (0.8)
Processing pipeline	Xia2 3dii (DIAMOND)	Xia2 dials (Dr Steven Johnson)

Table 5.3 | Phospho-null $Cnn^{D490-N579}$ data collection and processing statistics. The same diffraction dataset was indexed in $P222_1$ (automatically at the beamline) and P1 (by Dr Steven Johnson). Where applicable, values for the inner resolution shell are given in brackets.

The largest crystals were found in condition C2 of this custom screen, i.e. at 0.3 M $CaCl_2$, 0.1 M sodium HEPES (pH 7.5), 24 % (v/v) PEG 400. Four crystals from this condition were fished and frozen using an additional 5 % (v/v) PEG 400 for cryoprotection. Data collection on these crystals was carried out by Prof Susan Lea at the DIAMOND beamline i04 (proposal mx23459). The crystal with the best diffraction data was used for subsequent structure determination. This crystal was indexed in $P222_1$ by the automatic indexing pipeline at the beamline but, for reasons outlined below, the same dataset was also indexed in P1 by Dr Steven Johnson (Table 5.3).

5.7.2 Phospho-null $Cnn^{D490\text{-}N579}$ structure determination

Matthews coefficient suggested that for a P1 unit cell with the given dimensions, there was a 100 % probability that the asymmetric unit consisted of an object with a molecular mass of 203,194 Da, equivalent to 20 copies of phospho-null $Cnn^{D490\text{-}N579}$, with a solvent content of 55.07 %. At the higher symmetry of $P222_1$ this would be equivalent to five copies of phospho-null $Cnn^{D490\text{-}N579}$. Structure determination was initially trialled at this higher symmetry.

Phaser was used for molecular replacement using the previously solved structure of wild-type $Cnn^{D490\text{-}N579}$ as a homology model. A single solution was found for placing a single copy of this dimer, with a translational function Z-score of 10.6. The result after ten cycles of refinement via refmac was fixed, and phaser instructed to find another copy of the wild-type $Cnn^{D490\text{-}N579}$ dimer. Two solutions were found and the top one was retained, which had a translational function Z-score of 19.0. After some further refinement via refmac as before, phaser was used a third and final time to look for a copy of the wild-type $Cnn^{D490\text{-}N579}$ dimer, with the two previously placed dimers fixed. A single solution was found with a translational

function Z-score of 24.7. However, this time phaser only placed one of the two chains of the dimer in the asymmetric unit, consistent with the Matthews coefficient prediction that the asymmetric unit would consist of five copies of the protein in total, with dimerisation of the fifth copy occurring with its symmetry copy.

Further cycles of refinement via phenix.refine were carried out. However, R_{free} and R_{work} could not be improved beyond 32.19 % and 28.53 %, respectively, suggesting that model building was not complete. Especially the N-termini of the protein chains proved difficult to model, requiring constant adjustment after each refinement cycle without achieving a clear solution. It was therefore suspected that there might be genuine differences between the N-termini of the chains from apparent crystallographic copies, and that the diffraction data were mis-indexed at a higher symmetry. Dr Steven Johnson therefore re-indexed the diffraction dataset in P1, and these data were then used for further structure determination (Table 5.3).

Phaser was instructed to look for four copies of the best partial model obtained from the higher symmetry dataset, which it successfully did. However, in order to avoid bias during further refinement, the R_{free} columns of the P1 dataset were deleted via sftools. This modified reflection dataset, in addition to the partial model from phaser, was then used in phenix.refine, and the program was instructed to create new R_{free} flags from 5 % of the dataset. This modified reflection dataset file, containing the new R_{free} flags, was used during further refinement with phenix.refine.

Substantial model building followed, which involved modelling of all termini, correction of side chain and Ramachandran outliers, addition of solvent molecules as well as deletion and

re-building of the linker connecting the N- and C-terminal helices in four of the 20 chains. Model building was concluded when R_{free} and R_{work} values of 30.82 % and 25.99 %, respectively, were reached, which are acceptable for a structure at 3.0 Å resolution (Table 5.4).

The N-termini of the four sets of chains, i.e. ABCDE, FGHIJ, KLMNO, PQRST that were thought to constitute a higher symmetry, were all observed to be different in the final model, confirming that there are genuine differences between these sets of chains that were previously thought to be symmetry copies of each other in $P222_1$ (Table 5.5). Indeed, when the most complete set of chains (i.e. KLMNO) was used as a search model to be placed into the $P222_1$ dataset with phaser, and the resulting solution refined with precisely the same settings in phenix.refine as were used for the final refinement run in the P1 dataset, this resulted in values for R_{free} and R_{work} of 36.03 % and 30.17 %, respectively. The crystallographic packing was the same when compared to the structure in P1, however. As these values were significantly worse than those obtained in P1, it was concluded that structure determination in the lower symmetry was indeed the correct approach.

Resolution (Å)	76.54-3.03
No. of reflections:	
working set	38939
test set	1998
R_{work} / R_{free} (%)	25.99 / 30.82
Asymmetric unit composition	Ten dimers
Observed residues	See Table 5.5
No. of atoms:	
Overall	12935
Protein	12893
Water	42
Average B factors (Å^2):	
Overall	96.0
Protein	96.1
Water	72.3
Ramachandran statistics (%):	
favoured	98.45
allowed	1.24
outliers	0.31
Rotamer outliers (%)	0.51
RMS deviation:	
bond lengths (Å)	0.001
bond angles (°)	0.347
Clashscore	5.56
MolProbity score	1.30

Table 5.4 | Phospho-null Cnn$^{D490\text{-}N579}$ refinement statistics.

Chain	Observed residues		Chain	Observed residues
A	N493-N579		**K**	G(489)-N579
B	Q492-N579		**L**	Q492-N579
C	Q499-N579		**M**	S494-N579
D	Q492-N579		**N**	L502-N579
E	V511-N579		**O**	E510-N579
Chain	**Observed residues**		**Chain**	**Observed residues**
F	G(489)-N579		**P**	G(489)-N579
G	S494-N579		**Q**	Q492-N579
H	L500-N579		**R**	I497-N579
I	N493-N579		**S**	E503-N579
J	E510-N579		**T**	V511-N579

Table 5.5 | Residues observed in the structure of phospho-null Cnn$^{D490\text{-}N579}$. Note G(489) is a left-over residue from the TEV-cleavage site and not found in the original sequence of Cnn, with the residue number therefore shown in brackets.

5.7.3 Phospho-null Cnn$^{D490\text{-}N579}$ structure analysis

The asymmetric unit is highly complex, consisting of ten dimers of phospho-null Cnn$^{D490\text{-}N579}$ arranged in two non-crystallographic screw axes (Fig. 5.22 A). Overlaying all ten dimers with each other reveals that there is a degree of variation between the different dimers with regards to their N-termini and the conformation of the loop connecting the N- and C-terminal helices (Fig. 5.23). Indeed, these differences in conformation are matched by an increased flexibility of these regions as demonstrated by the variation in B-factor values, which are maximal at the N-termini and in this loop region. In contrast, the packing within the asymmetric unit results in high stabilisation of the C-termini (Fig. 5.22 B), and indeed the extreme C-terminus can be observed in every chain, while different N-termini are observed for different sets of chains (Table 5.5). Note that this is in marked contrast to wild-type Cnn$^{D490\text{-}N579}$ which packs *in crystallo* with the help of phosphates or sulphates in a way that highly stabilises the loop region but leaves the C-terminus flexible (Fig. 5.15 D).

Importantly, however, overlaying all ten phospho-null Cnn$^{D490\text{-}N579}$ dimers with the dimer of wild-type Cnn$^{D490\text{-}N579}$ shows that the overall structure of the construct, i.e. the double hairpin fold, is unaffected by introduction of the three point mutations S567A, S571A and S573A (Figs. 5.23 and 5.24). It can therefore be excluded with high confidence that disruption of centrosome assembly *in vivo* upon mutating these three residues to alanine in context of the entire Cnn (ref [196]) is due to abnormally folded protein. Instead, it seems most probable that the phenotype is solely due to the inability of Polo kinase to induce or stabilise a conformational change that disfavours the hairpin fold by phosphorylating S567 (and potentially the other residues).

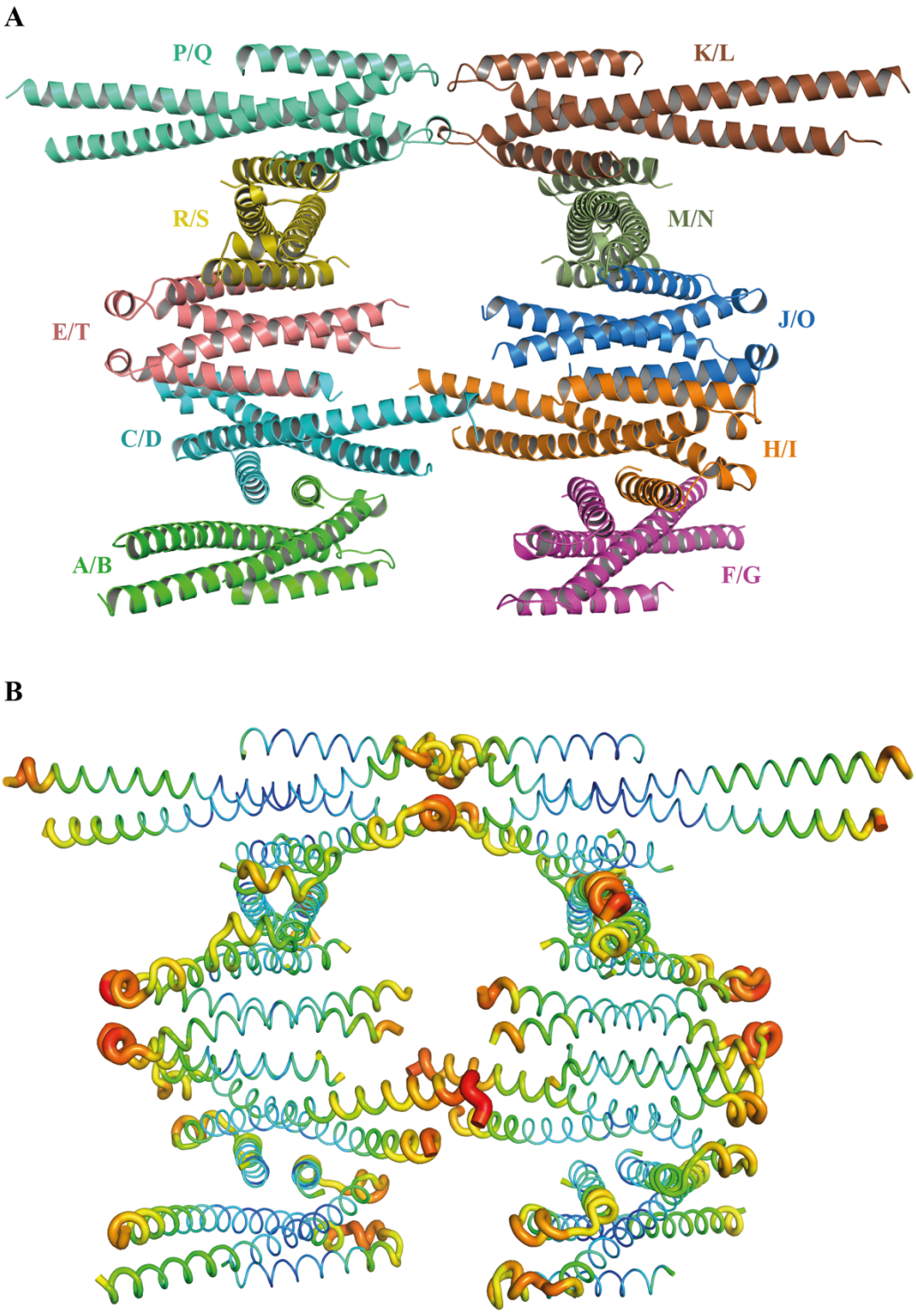


Figure 5.22 | Structure of phospho-null Cnn$^{D490-N579}$. A) The ten dimers of phospho-null making up the asymmetric unit in P1 are shown in cartoon representation with each dimer in a different colour. Different polypeptide chains are labelled. **B)** B-factor putty model generated via PyMOL with phospho-null Cnn$^{D490-N579}$ in the same orientation as in (A). Residues with a lower B-factor, i.e. little flexibility, are highlighted as a thin blue tube; residues with a higher B-factor, i.e. greater flexibility, are highlighted as a thick tube in green-yellow-red at increasing flexibility. Note flexibility is maximal at the N-termini and the loop region connecting the N- and C-terminal helices; the C-terminus shows little flexibility due to tight packing against other chains (cf. Fig. 5.15 D).

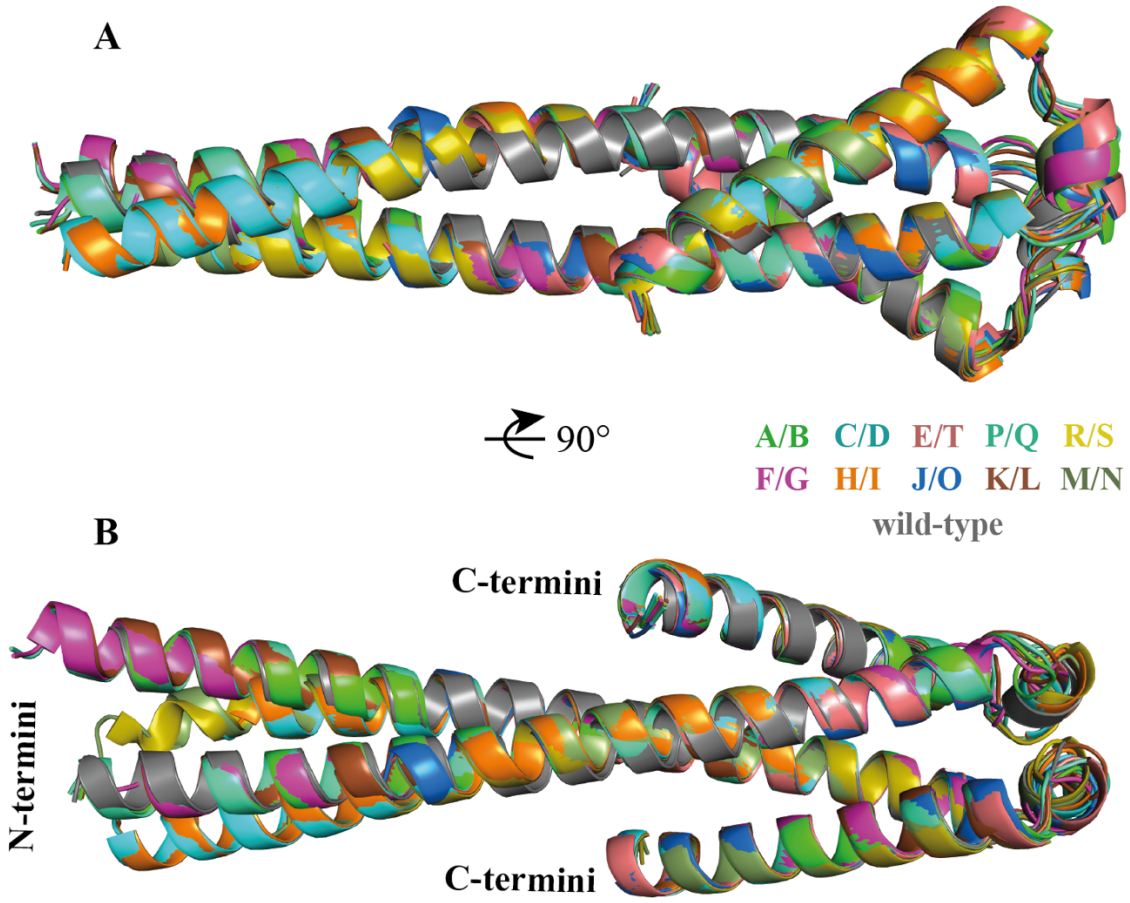


Figure 5.23 | Phospho-null mutations do not disturb the double hairpin structure of $Cnn^{D490\text{-}N579}$. The ten dimers of the phospho-null $Cnn^{D490\text{-}N579}$ asymmetric unit are shown in cartoon representation and in the same colour-code as in Fig. 5.22 A, but they are overlaid with each other as well as with wild-type $Cnn^{D490\text{-}N579}$ in grey. Note the N-termini and loop region connecting the N- and C-terminal helices show variability, but the overall double hairpin structure is unaffected by introduction of the phospho-null mutations S567A/S571A/S573A. The overlay is shown in two orientations (A/B) rotated by 90°.

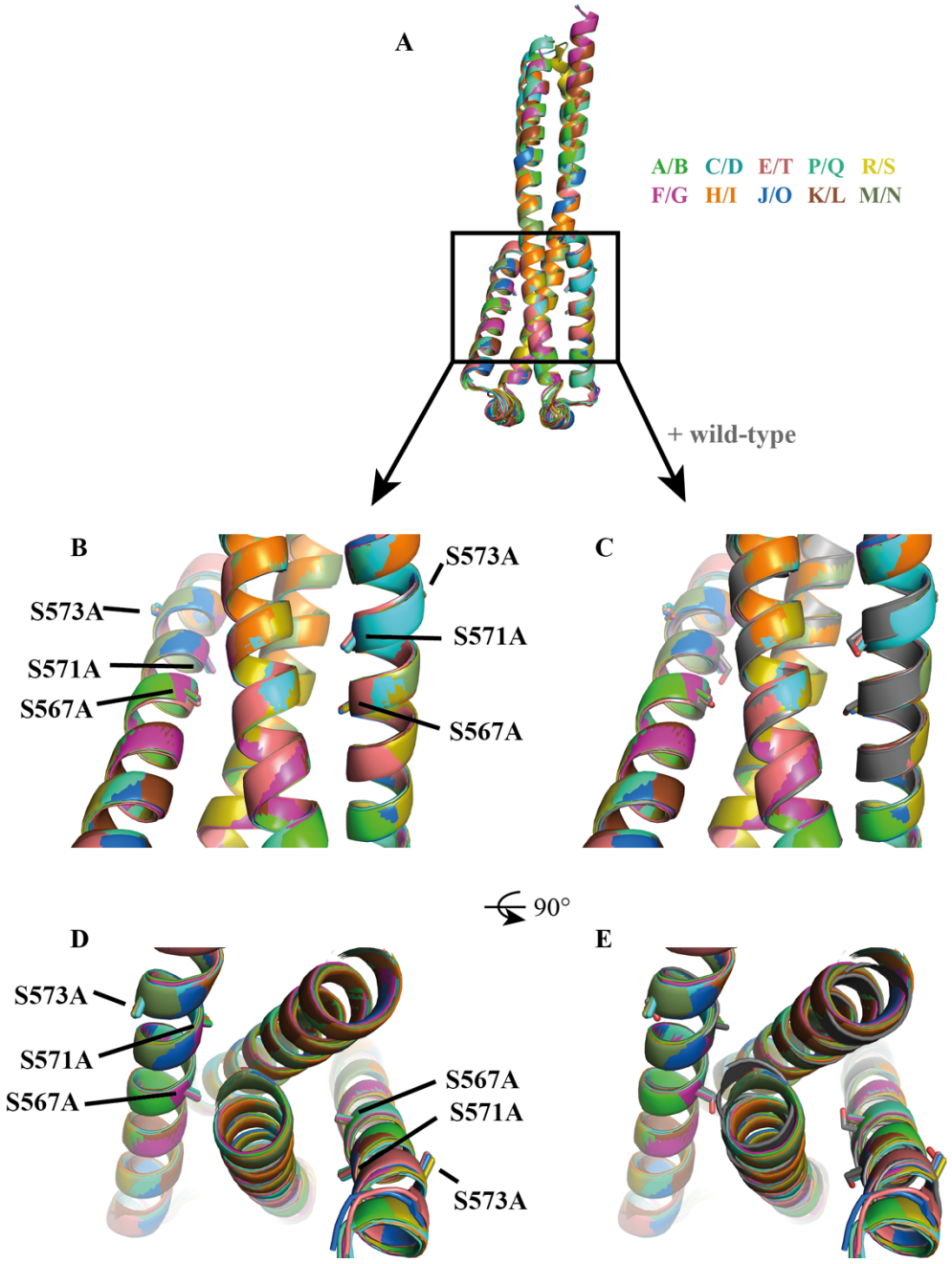


Figure 5.24 | Detail showing the three point mutations S567A, S571A and S573A. **A)** Overlay shown as in Fig. 5.23, but without wild-type $Cnn^{D490\text{-}N579}$. **B/C)** Detail focussing on the three point mutations S567A, S571A and S573A either without (B) or with (C) wild-type $Cnn^{D490\text{-}N579}$ also overlaid in grey. Side chains of these three residues are shown as sticks. **D/E)** The view from (B/C) rotated by 90°. Note introduction of the three point mutations did not affect the hairpin fold in the region around the mutations.

5.8 Calmodulin interacts with dilipoyl-Cnn$^{A553\text{-}H608}$

As shown in Chapter 4, Calmodulin cannot bind mNG-Cnn$^{D490\text{-}H608}$, but it can prevent mNG-Cnn$^{D490\text{-}H608}$:Cnn$^{G1077\text{-}L1148}$ network assembly via formation of a tertiary complex (Fig. 4.21). It was hypothesised that in this tertiary complex, Calmodulin binds to mNG-Cnn$^{D490\text{-}H608}$, but only once it has undergone a conformational change upon binding to Cnn$^{G1077\text{-}L1148}$ that would, in the absence of Calmodulin, allow it to assemble into networks. The structures of Cnn$^{D490\text{-}N579}$ presented in this chapter suggest that this conformational change is a disruption of the hairpin fold and therefore a displacement of the region C-terminal to A553. This suggests that Calmodulin might prevent network assembly *in vitro* by binding to the region A553-H608. Indeed, Calmodulin might not be able to bind this region while it adopts a hairpin fold, and could potentially only bind it once this fold has been disrupted prior to network formation.

In order to test this hypothesis, a construct was cloned that allowed Cnn$^{A553\text{-}H608}$ to be purified with a dilipoyl-tag (Fig. 5.25 A/B). As outlined in Chapter 4, this vector system is especially useful for the purification of short peptides, as the peptide is produced as the linker between two His-tagged domains which solubilise and protect the peptide from proteases.

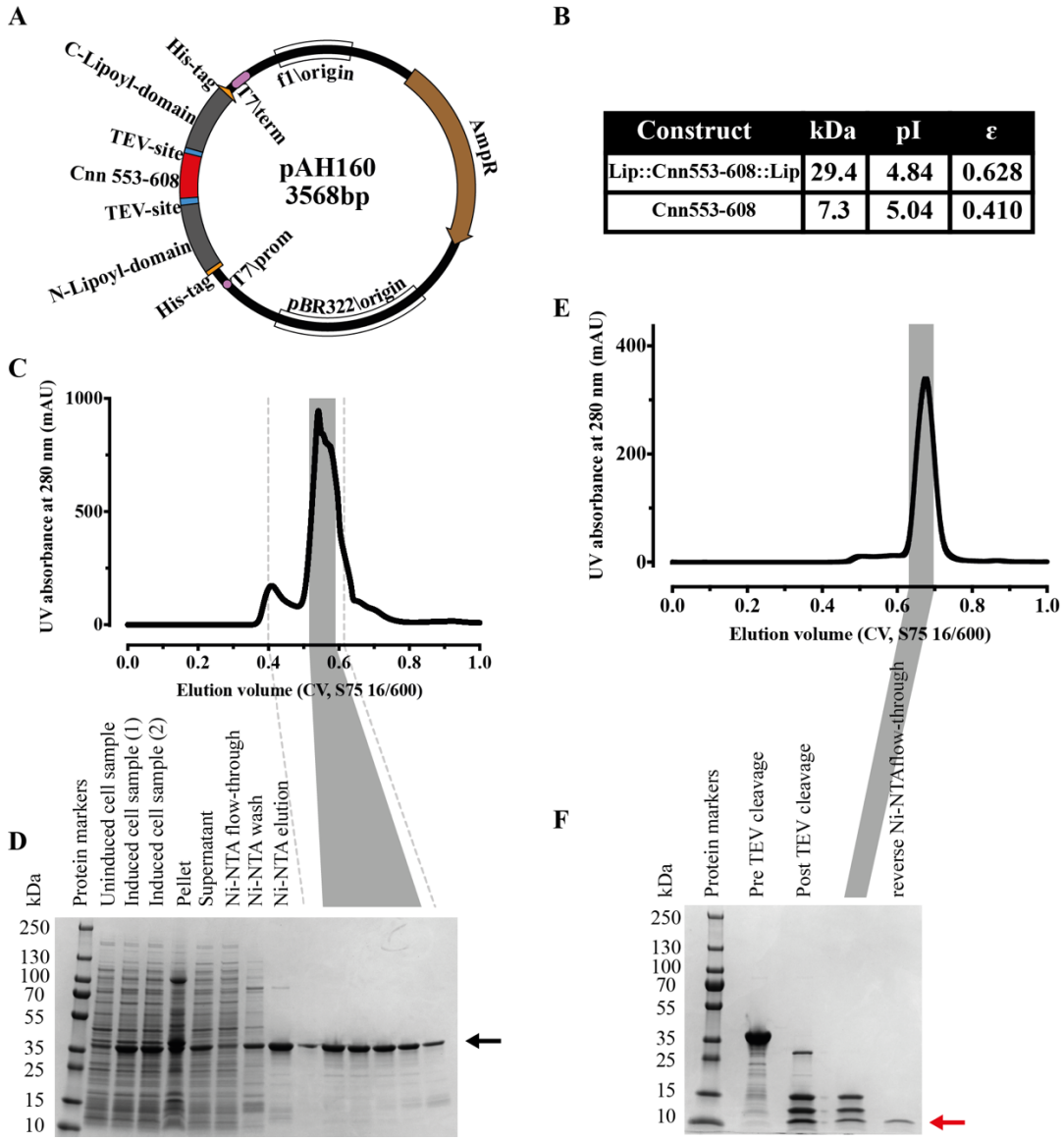


Figure 5.25 | Purification of Cnn$^{A553-H608}$. **A)** Overview of plasmid pAH160, which encodes Cnn$^{A553-H608}$ with an N- and C-terminal, TEV-cleavable, His-Lipoyl-domain. **B)** Table showing the predicted molecular weight (in kDa), pI and extinction coefficient (ε) of the tagged and untagged protein. **C)** SEC (S75 16/600) elution profile of dilipoyl-Cnn$^{A553-H608}$ after Ni-affinity chromatography, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.52-0.59) indicates the fractions pooled after SEC and run on the SDS-PAGE gel in D, containing dilipoyl-Cnn$^{A553-H608}$. Further fractions run on the gel in D are indicated by grey dashed lines. SEC was carried out using standard TBS. **D)** 4-20 % reducing SDS-PAGE gel showing successive purification steps. The band corresponding to dilipoyl-Cnn$^{A553-H608}$ is marked with a black arrow. Note that dilipoyl-Cnn$^{A553-H608}$ not only eluted in the main peak but also in the void peak, suggesting some of the protein had aggregated. **E)** SEC (S75 16/600) elution profile of Cnn$^{A553-H608}$ after tag cleavage, with UV absorbance (280 nm) plotted on the y-axis and column elution volume (CV, 120 ml) on the x-axis; grey area (CV = 0.63-0.69) indicates the fractions pooled after SEC and run on the SDS-PAGE gel in F. SEC was carried out using standard TBS. **F)** 4-20 % reducing SDS-PAGE gel showing tag cleavage and further purification steps. The band corresponding to Cnn$^{A553-H608}$ is marked with a red arrow.

Protein was produced in *E. coli* C41 (DE3) cells grown in 4 LB with IPTG induction occurring only after 7 hours due to unusually slow growth of the cells. Dilipoyl-Cnn$^{A553\text{-}H608}$ was then purified according to the usual protocol. The SEC elution profile showed one strong peak with a later-eluting shoulder, but the corresponding SDS-PAGE gel suggested that the peak mostly contained dilipoyl-Cnn$^{A553\text{-}H608}$ (Fig. 5.25 C/D). Note that some of the protein was observed to elute in the void volume (Fig. 5.25 C/D, left dashed line), suggesting that some of the protein was aggregating or forming larger assemblies during purification, perhaps a reflection of the role of this region within Cnn in network assembly. The main peak (CV = 0.52-0.59) was pooled and concentrated to 38.6 mg/ml before flash-freezing in 55 µl aliquots. The final yield of dilipoyl-Cnn$^{A553\text{-}H608}$ was ~ 11.7 mg/L.

For tag cleavage, 0.55 ml of dilipoyl-Cnn$^{A553\text{-}H608}$ at a stock concentration of 38.6 mg/ml were mixed with 1 ml of TEV protease at a stock concentration of 1.25 mg/ml. The reaction was diluted to 5 ml using standard TBS and left at 4°C overnight before application to a S75 16/600 pre-equilibrated in standard TBS for subsequent removal of TEV protease (Fig. 5.25 E). The single peak after SEC (CV = 0.63-0.69), which according to SDS-PAGE analysis contained Cnn$^{A553\text{-}H608}$ as well as the two cleaved His-tagged lipoyl-tags, was pooled and applied to a Ni-NTA column (CV = 5 ml) for removal of the two tags via reverse Ni-affinity chromatography. The flow-through was collected and confirmed to contain only Cnn$^{A553\text{-}H608}$ via SDS-PAGE analysis (Fig. 5.25 F). The protein was then concentrated to 1.32 mg/ml and frozen in 60 µl aliquots.

Dilipoyl-Cnn$^{A553\text{-}H608}$, which has a theoretical monomeric mass of 29.4 kDa, was analysed via SEC-MALS in order to determine its oligomeric state in solution. Using a S200 10/300 increase column pre-equilibrated in standard TBS, dilipoyl-Cnn$^{A553\text{-}H608}$ was tested at

200 μM and 400 μM. At both concentrations, the construct was monomeric in solution, with average molecular weights of 30.2 kDa (+/- 1.1 %) and 30.0 kDa (+/- 1.7 %), respectively (Fig. 5.26). Note that at the higher concentration, a later-eluting shoulder was observed that was not discernible at the lower concentration. Indeed, the main peak pooled during purification of this protein also had a later-eluting shoulder on the gel filtration profile and additional faint lower molecular weight bands on the SDS-PAGE gel (Fig. 5.25 C/D). These might reflect small amounts of contaminant. However, at 200 μM the protein peak appears fairly symmetric (Fig 5.26), suggesting that the protein is largely homogeneous.

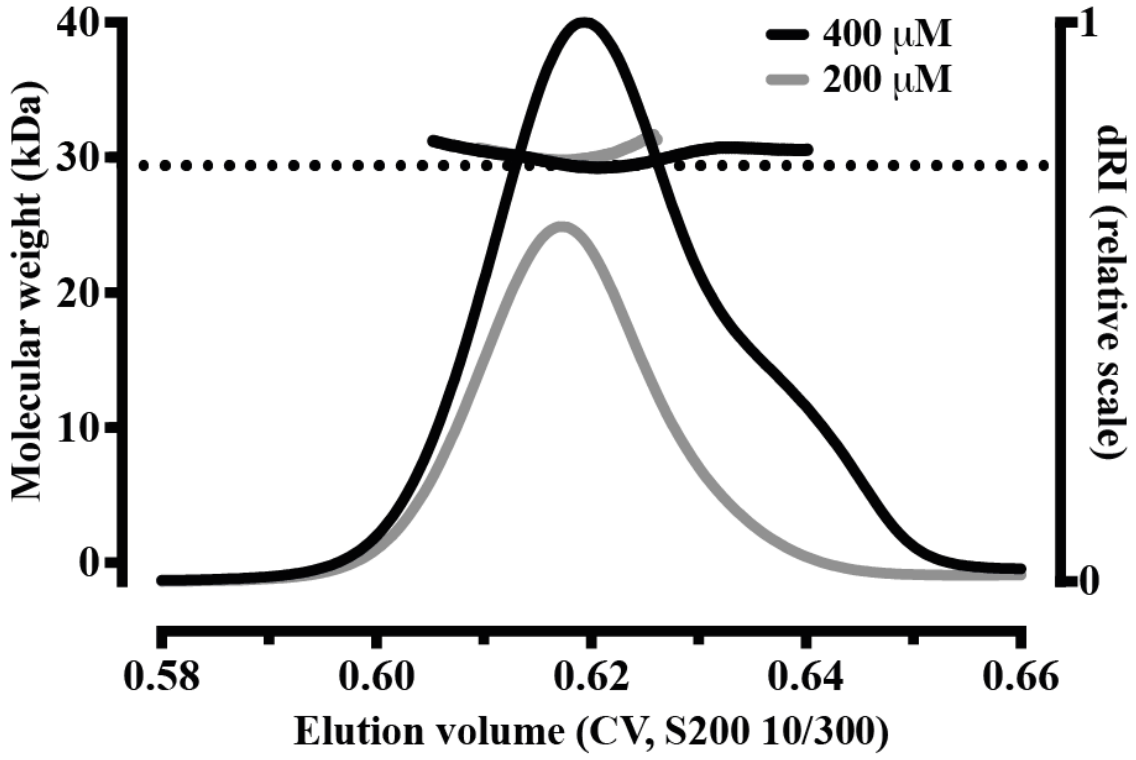


Figure 5.26 | SEC-MALS analysis of dilipoyl-Cnn$^{A553-H608}$. The average molecular weight in solution of dilipoyl-Cnn$^{A553-H608}$ was measured at 200 μM (light grey: 30.2 kDa +/- 1.1 %) and 400 μM (black: 30.0 kDa +/- 1.7 %). Dotted line indicates theoretical molecular weight expected for monomeric dilipoyl-Cnn$^{A553-H608}$ (i.e. 29.4 kDa). Note the construct is monomeric under these conditions. Molecular weight (in kDa) is plotted on the left y-axis, change in refractive index (dRI, relative scale) plotted on the right y-axis, column elution volume (CV, 24 ml) of column S200 10/300 increase plotted on the x-axis. Standard TBS was used as buffer.

Binding between Calmodulin and dilipoyl-Cnn$^{A553\text{-}H608}$ was subsequently tested via analytical SEC. Reactions were set up containing Calmodulin and dilipoyl-Cnn$^{A553\text{-}H608}$ at respective concentrations of 250 μM and 200 μM either alone or mixed together. After incubation at room temperature, these reactions were analysed via SEC on a S200 10/300 increase column. Importantly, the same set of experiments was carried out using standard TBS as well as using standard TBS with an additional 10 mM $CaCl_2$ as buffer (Fig. 5.27).

In the absence of calcium, no peak shifts could be observed when both constructs were mixed together, and the SDS-PAGE gel confirmed that Calmodulin and dilipoyl-Cnn$^{A553\text{-}H608}$ did not elute in the same fractions, suggesting that no binding had occurred (Fig. 5.27 A/B).

In contrast, when these experiments were repeated in the presence of 10 mM $CaCl_2$, a clear peak shift to an earlier-eluting species was observed after SEC, and the corresponding SDS-PAGE gel confirmed that this peak contained both Calmodulin and dilipoyl-Cnn$^{A553\text{-}H608}$, suggesting that binding had occurred (Fig. 5.27 C/D). Calmodulin therefore binds dilipoyl-Cnn$^{A553\text{-}H608}$ in a calcium-dependent manner. The tertiary complex of Calmodulin, Cnn$^{D490\text{-}H608}$ and Cnn$^{G1077\text{-}L1148}$ reported in Chapter 4 might therefore consist of Cnn$^{G1077\text{-}L1148}$ bound to the zipper-like motif within Cnn$^{D490\text{-}H608}$ and Calmodulin bound to the region A553-H608.

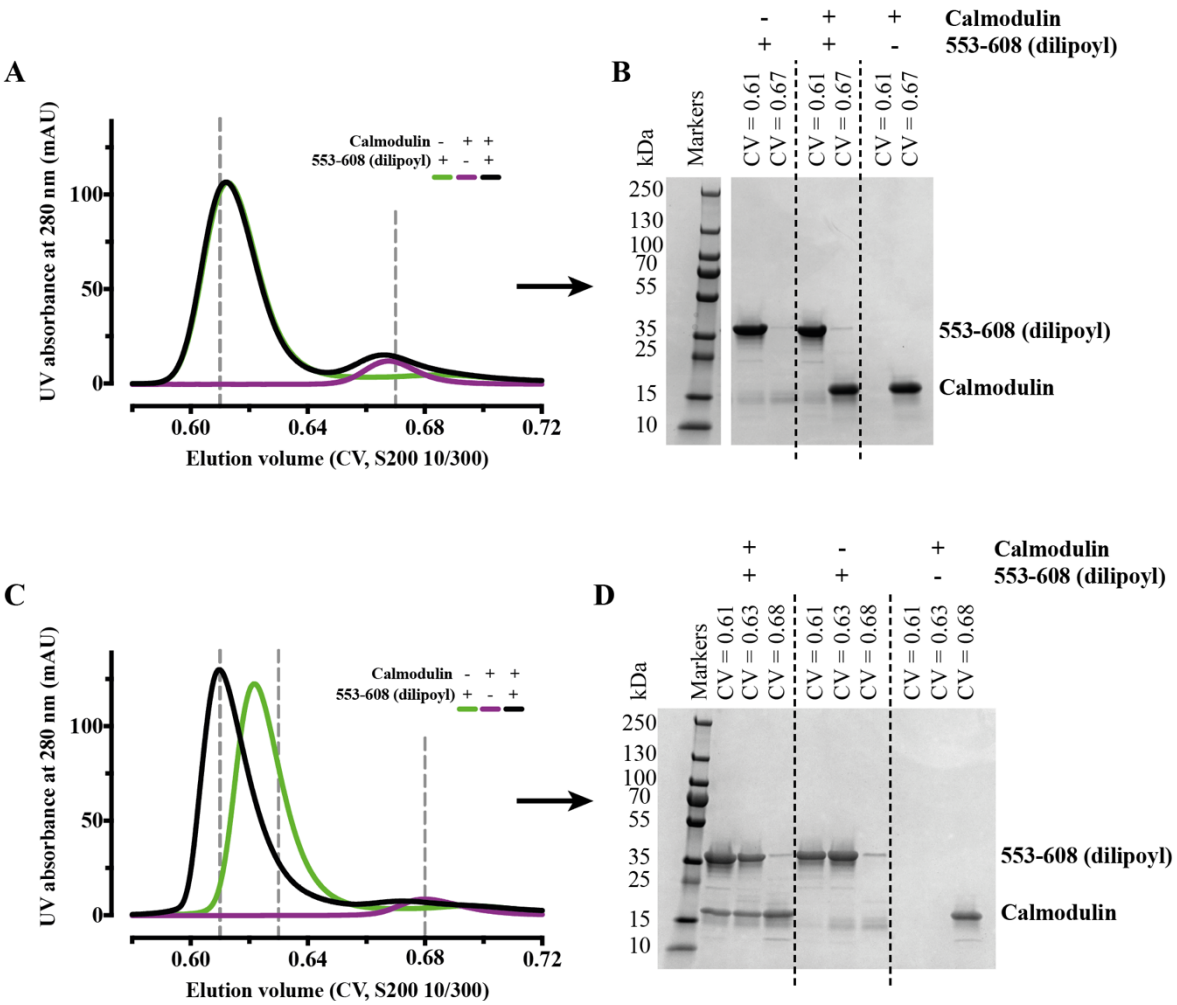


Figure 5.27 | Calmodulin binds dilipoyl-Cnn$^{A553\text{-}H608}$ in a calcium-dependent manner. **A/C)** Analytical SEC experiments with 250 µM Calmodulin alone (purple), 200 µM dilipoyl-Cnn$^{A553\text{-}H608}$ alone (green) or both constructs mixed at the same concentrations (black). Experiments were carried out using standard TBS (A) or standard TBS supplemented with 10 mM $CaCl_2$ (C) as buffer. UV absorbance at 280 nm (in mAU) is plotted on the y-axis, column elution volume (CV, 24 ml) of column S200 10/300 increase plotted on the x-axis. Dashed grey lines indicate fractions taken and run on the SDS-PAGE gel in B (corresponding to experiments in A) and the SDS-PAGE gel in D (corresponding to experiments in C). Note a peak shift upon mixing Calmodulin and dilipoyl-Cnn$^{A553\text{-}H608}$, indicating binding between the two constructs, can only be observed in the presence of calcium. **B/D)** 4-20 % reducing SDS-PAGE gels showing fractions from the analytical SEC experiments. Bands correspond to the constructs as indicated. Note no co-elution between the Calmodulin and dilipoyl-Cnn$^{A553\text{-}H608}$ bands can be observed in the absence of calcium (B), indicating no binding. In contrast, co-elution of the bands corresponding to the two constructs can be observed in CV = 0.61 and 0.63 in the presence of calcium (D), demonstrating that the early-eluting peak in (C) contains both proteins, indicating binding in the presence of calcium.

5.9 Discussion

In this chapter, the crystal structure of wild-type $Cnn^{D490-N579}$ is shown to adopt a double hairpin fold which is sterically incompatible with binding of the Cnn C-terminus. Furthermore, the hairpin fold buries the conserved residues S567 and S571 while leaving residues S573 and T577 exposed to the solvent. It is known that Cnn is specifically phosphorylated at S567 at centrosomes in fly embryos [196, 339], and S571 is also a Polo phosphorylation consensus site. Phosphorylation on either of these residues, but not on S573 or T577, is also sterically incompatible with the structure observed *in crystallo*.

Mutation of residues S567, S571 and S573 to alanine severely disrupts Cnn network assembly *in vivo* [196]. By solving the structure of $Cnn^{D490-N579}$ carrying these three mutations, it is shown that the double hairpin fold is not affected by these phospho-null mutations. The phenotype observed *in vivo* is therefore likely to be due to an essential function phosphorylation of some or all of these residues has in allowing network formation, rather than due to the production of misfolded protein. Furthermore, Chapter 3 argues that binding of the Cnn C-terminus to the zipper-like motif is essential for network formation. With phosphorylation and C-terminus binding being essential for network formation, but the structure of $Cnn^{D490-N579}$ presented here being incompatible with both, it is therefore suggested that the double hairpin fold shown here represents the inactive form of Cnn which cannot assemble networks. This form could be present for example in the cytosol, where Cnn networks are not supposed to form, or in the interphase pericentriolar material in the absence of Polo kinase activity. In order to transition into the assembly-competent state, the hairpin fold would have to be disrupted by a combination of phosphorylation and binding of the Cnn C-terminus (Fig. 5.28).

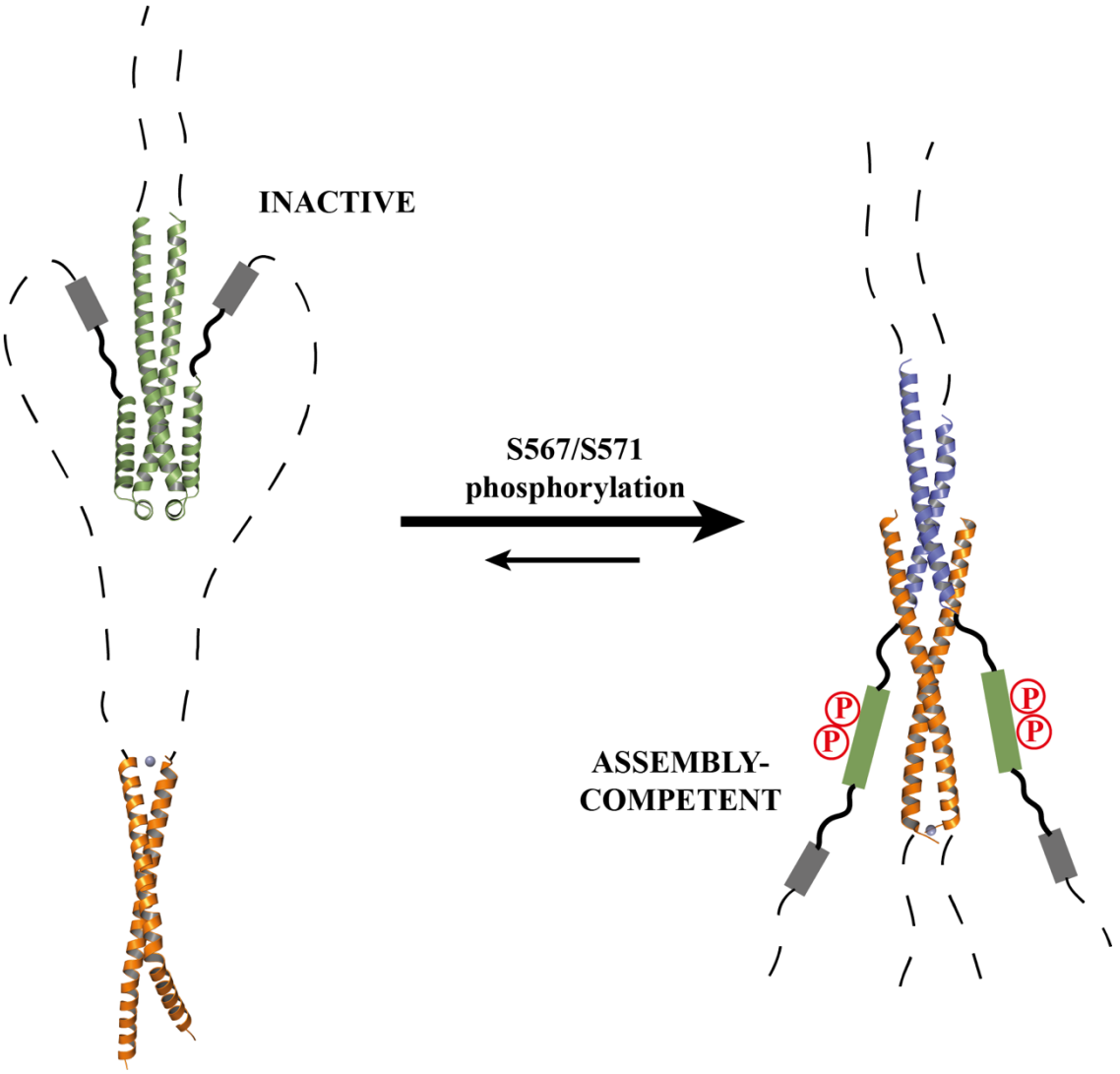


Figure 5.28 | Illustration showing conversion of Cnn from an inactive to an assembly-competent state. The inactive state is shown with the structure of $Cnn^{D490\text{-}N579}$ solved in this chapter in green, and one dimer of $Cnn^{H1082\text{-}L1148}$ (i.e. the Cnn C-terminus, CM2) from the asymmetric unit of PDB 5I7C (ref [339]) in orange. The assembly competent state is shown with the structure of the complex between $Cnn^{D490\text{-}K544}$ (blue) and $Cnn^{H1082\text{-}L1148}$ (orange) from PDB 5MVW (ref [339]). The interaction between $Cnn^{H1082\text{-}L1148}$ and $Cnn^{D490\text{-}K544}$ could be intra- or intermolecular. Models are shown in cartoon representation, zinc is shown as a grey sphere. Helical regions A553-N579 (green) and L590-T602 (grey, see Fig. 5.2 for prediction) are shown as boxes and their linkers as solid black lines, all drawn roughly to scale (note the linker conformation is unknown and could be more extended). Further regions within Cnn are indicated as dashed lines and not drawn to scale. Phosphorylation of residues S567 and S571 on both chains is indicated by a red encircled P, and is suggested to bias the equilibrium between the two states of Cnn towards the assembly-competent state.

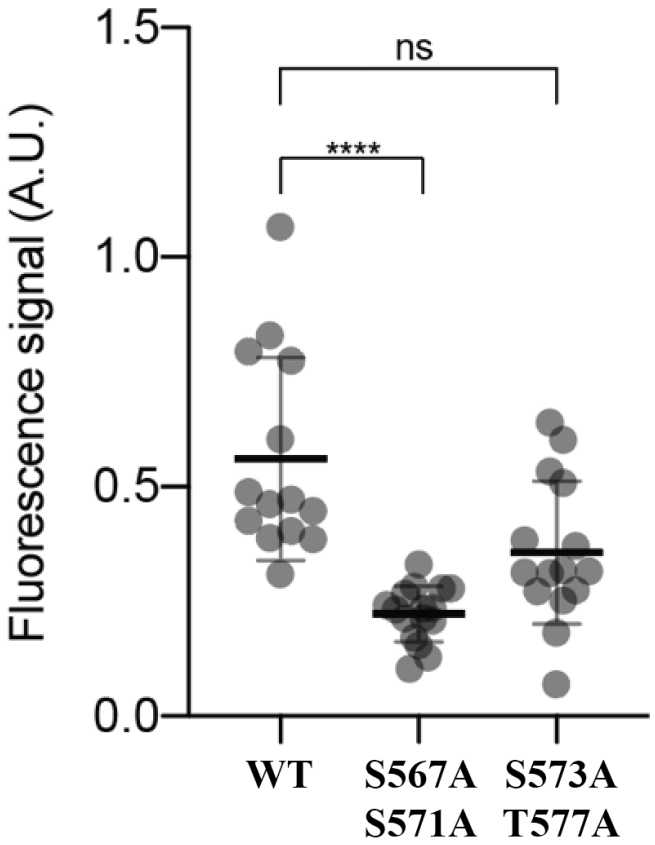


Figure 5.29 | *In vivo* analysis of Cnn phospho-null point mutations. mRNA carrying GFP-tagged Cnn with the indicated point mutations was injected into *Drosophila cnn* mutant embryos using the same method as described in [339]. For each experiment, the average of every observed centrosome per embryo is plotted as a dot. Kruskal-Wallis analysis (multiple testing) was used to calculate statistical significance, with four stars indicating a P value of < 0.0001, two stars indicating a P value of < 0.01 and ns meaning not significant. Note mutation of S567 and S571 to alanine disrupts Cnn network assembly, while mutation of S573 and T577 does not. Phosphorylation on S567 therefore appears to be necessary to allow network formation. Note this is in line with the solved structure of $Cnn^{D490\text{-}N579}$, which proposes a role for S567/S571 but not S573/T577 in disturbing the hairpin fold and thereby aiding formation of the assembly-competent state. Experiments and data analysis carried out and figure created by Isaac Wong with the help of Dr Zsofia Novak [unpublished results, reproduced here in modified form with permission].

Crucially, the structure of $Cnn^{D490\text{-}N579}$ presented here makes the clear prediction that phosphorylation of S567 and S571 has a role in 'priming' the structure for binding to the Cnn C-terminus by disrupting the hairpin fold, while phosphorylation of S573 and T577 cannot have the same effect due to these residues being exposed to the solvent. Indeed, the structure has since been used as a guide to design *in vivo* experiments by collaborators which clearly demonstrate that phospho-null (i.e. alanine) mutations of S567 and S571 disrupt Cnn network formation, while phospho-null mutations of S573 and T577 do not have a statistically significant effect (Fig. 5.29).

Considering that Cnn is known to be dephosphorylated on S567 towards the periphery of the centrosome [196, 339], this model also offers potential insight into the mechanism of centrosome size control, as dephosphorylation on S567 would be expected to shift the equilibrium between the inactive and assembly-competent state back towards dissociation of the Cnn C-terminus and therefore adoption of the inactive state (and so network disassembly) at the periphery of the centrosome. It will have to be investigated what phosphatases act on phosphorylated S567 and how their activity might be controlled.

In vitro, binding of the Cnn C-terminus and phosphorylation of S567/S571 can occur independently of each other and it is therefore unclear which event occurs first *in vivo*. Indeed, it is feasible to assume that in some instances, binding of the Cnn C-terminus disrupts the hairpin fold and thereby facilitates phosphorylation of S567/S571, which would then stabilise the assembly-competent state. In other cases, phosphorylation might occur first and disrupt the hairpin fold, thereby allowing the Cnn C-terminus to bind more readily. Note that the observation of a 2:2 complex between the Cnn C-terminus and mNG-$Cnn^{D490\text{-}N579}$ in the presence of kinase but not in the absence of kinase at the same concentrations

could mean that phosphorylation of mNG-Cnn$^{D490\text{-}N579}$ allows more efficient formation of the complex, but it could also be due to limitations of the assay such as a potential interaction of the unphosphorylated complex with the column resin (Figs 5.9 and 5.10). Furthermore, the phospho-mimicking experiments described in this chapter (Fig. 5.19) also cannot be used to directly inform about possible differences in affinity or kinetics of Cnn C-terminus binding to the zipper-like motif with the hairpin intact or disrupted upon phosphorylation. The experiments were carried out in the presence of a large molar excess of the Cnn C-terminus. Furthermore, in order to bind mNG-Cnn$^{D490\text{-}N579\ (S573E/T577E)}$ *in vitro*, the Cnn C-terminus probably needs to displace the hairpin from its binding site, while in the case of mNG-Cnn$^{D490\text{-}N579\ (S567E/S571E)}$, the Cnn C-terminus needs to first disrupt the tetramer *in vitro*. Note that tetramerisation is here shown to be concentration-dependent in solution (Fig. 5.18 B) and therefore probably does not occur *in vivo*, where the concentration of Cnn is likely to be much lower. Furthermore, binding is required to occur at different time-scales *in vivo*. Carefully controlled experiments using isothermal titration calorimetry or surface plasmon resonance could be used to investigate the binding thermodynamics and kinetics.

In *C. elegans*, it has previously been suggested that there is an equilibrium between an inactive form of SPD-5, the putative functional homologue of Cnn, and an assembly-competent form, and that phosphorylation simply biases this equilibrium towards the latter state (Fig. 1.9). A mathematical model building on these assumptions was found to accurately describe the dynamics of SPD-5 network assembly *in vivo* [238, 239]. However, there is no information on a molecular level in *C. elegans* concerning the nature of this transition. This chapter presents a molecular mechanism for the transition from an inactive to an assembly-competent state in *Drosophila*. Further experiments will have to determine

whether this mechanism is conserved in *C. elegans*, or indeed if a similar mathematical model can be expanded to the *Drosophila* system in light of the structures presented here.

The Calmodulin experiments presented in this and the previous chapter provide indirect evidence for the existence of the proposed conformational change upon binding of the Cnn C-terminus to the zipper-like motif in solution. In Chapter 4 it is shown that Calmodulin cannot bind mNG-Cnn$^{D490\text{-}H608}$ (Fig. 4.21). In contrast, this chapter shows that Calmodulin can bind dilipoyl-Cnn$^{A553\text{-}H608}$ in a calcium-dependent manner. This suggests that the region A553-H608 is inaccessible in context of the larger D490-H608 construct, probably as a result of hairpin formation. Note that it cannot be excluded that Calmodulin is actually binding to the lipoyl tag in this experiment, so it will have to be repeated using Cnn$^{A553\text{-}H608}$ lacking the two tags. This chapter shows that tag cleavage results in stable protein that can therefore be used for further experiments.

Importantly, the model presented here suggests that binding of the Cnn C-terminus to the zipper-like motif disrupts the hairpin fold in a way that exposes the region A553-H608 to the solvent. Indeed, Calmodulin is shown in the previous chapter to form a tertiary complex with mNG-Cnn$^{D490\text{-}H608}$ and Cnn$^{G1077\text{-}L1148}$ (Fig. 4.21). Presumably, this tertiary complex therefore consists of Calmodulin bound to the exposed region A553-H608 of the mNG-Cnn$^{D490\text{-}H608}$:Cnn$^{G1077\text{-}L1148}$ complex. A structural approach could definitively confirm the nature of this tertiary complex.

The notion that binding of the Cnn C-terminus to the zipper-like motif displaces the region immediately C-terminal to K544 (and therefore also A553-H608) and exposes it to the solvent is reinforced by considering that in the solved structure of the 2:2 complex between

$Cnn^{H1082\text{-}L1148}$ and $Cnn^{D490\text{-}K567}$ (PDB code: 5MWE), no residues beyond K544 can be observed *in crystallo* [339]. Note that $Cnn^{D490\text{-}K567}$ is only twelve amino acids shorter than $Cnn^{D490\text{-}N579}$, the apo-structure of which is shown in this chapter.

It is further shown in this chapter that mNG-$Cnn^{D490\text{-}N579}$ forms 2:2 complexes with $Cnn^{G1077\text{-}L1148}$ in the presence of PLK1 kinase in solution (Fig. 5.10). In contrast, it has previously been shown that mNG-$Cnn^{D490\text{-}H608}$ assembles into networks when bound to $Cnn^{G1077\text{-}L1148}$, even in the absence of kinase *in vitro* (see Chapter 4). These experiments suggest that residues I580-H608 are essential for *in vitro* network assembly. It is therefore plausible to hypothesise that Cnn networks depend on interactions mediated by the region A553-H608, and especially I580-H608, upon its displacement from the zipper-like motif (Fig. 5.30). As a consequence, the previously reported ability of Calmodulin to block network assembly *in vitro* via formation of a tertiary complex with mNG-$Cnn^{D490\text{-}H608}$ and $Cnn^{G1077\text{-}L1148}$ (see Chapter 4) might therefore be a result of Calmodulin binding to the region A553-H608 upon its displacement from the zipper-like motif and thereby blocking these potential $Cnn^{A553\text{-}H608}$-mediated interactions making up the Cnn network.

These $Cnn^{A553\text{-}H608}$-mediated interactions necessary for network formation could be with itself or with a different region within $Cnn^{D490\text{-}H608}$ (i.e. the PReM domain). Dilipoyl-tagged $Cnn^{A553\text{-}H608}$ is shown in this thesis to be monomeric in solution (Fig. 5.26), which would suggest that the proposed interaction is not with itself. However, note that during purification of the protein, a sizeable fraction was observed to elute in the void (Fig. 5.25 C), suggesting that under certain conditions the construct interacts with itself and forms bigger assemblies or aggregates.

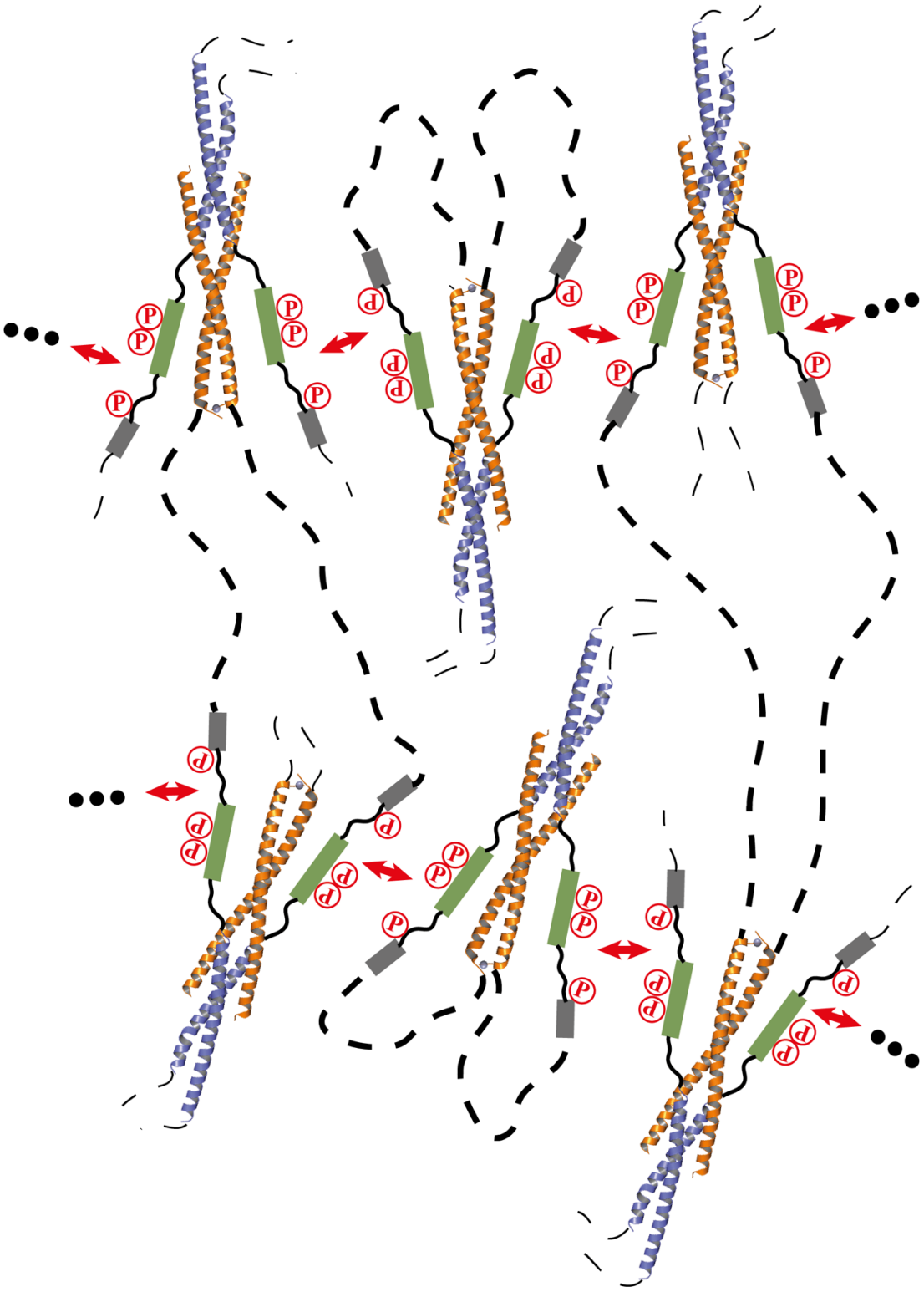


Figure 5.30 | Model depicting interactions that allow Cnn network formation. The assembly-competent state is shown as in Fig. 5.28 with an additional red encircled P indicating potential phosphorylation on S589, a conserved Polo consensus site. Such additional phosphorylation events are suggested to be required in order to allow a potential phosphorylation-mediated interaction (red arrows) between regions A553-H608 (roughly corresponding to green & grey boxes) or between A553-H608 and other regions within Cnn (not shown here, cf. Fig. 5.32). Note the model allows for both intra- and inter-molecular formation of the assembly-competent state (thick dashed lines, not drawn to scale).

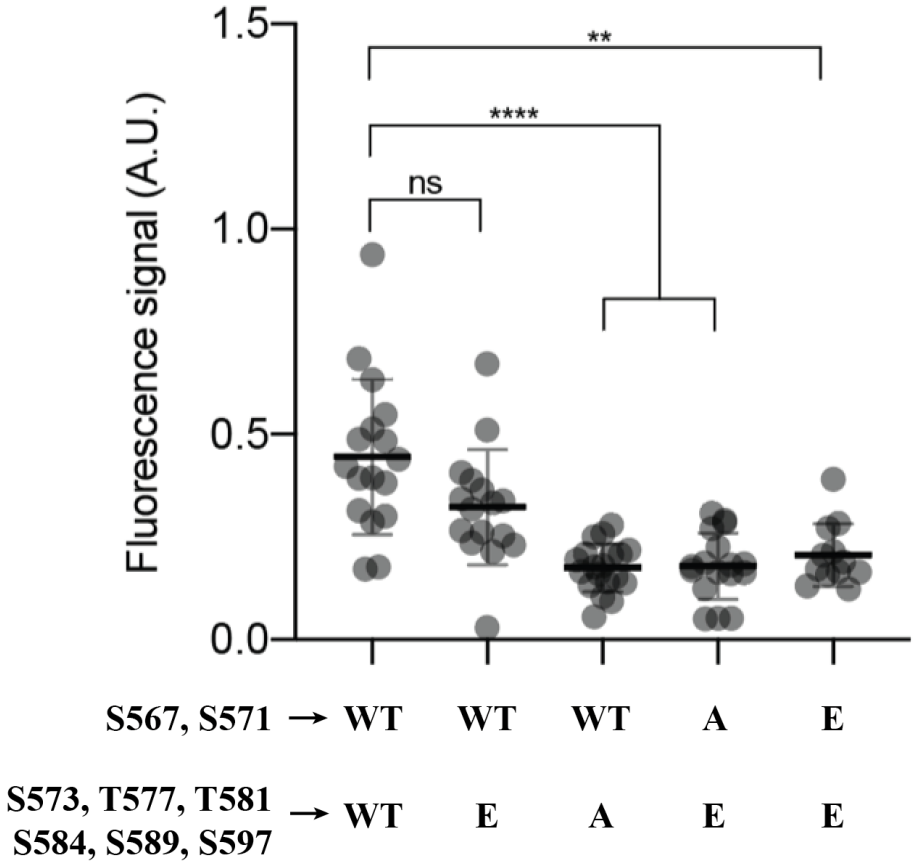


Figure 5.31 | *In vivo* analysis of further Cnn point mutations. mRNA carrying GFP-tagged Cnn with the indicated point mutations was injected into *Drosophila cnn* mutant embryos using the same method as described in [339]. For each experiment, the average of every observed centrosome per embryo is plotted as a dot. Kruskal-Wallis analysis (multiple testing) was used to calculate statistical significance, with four stars indicating a P value of < 0.0001, two stars indicating a P value of < 0.01 and ns meaning not significant. Note that hospho-null mutations of the six conserved Ser/Thr residues other than S567 and S571 disrupt Cnn network formation (WT/A), while phospho-mimicking mutations of these residues (WT/E) have no effect. This indicates that phosphorylation of at least some of these residues might also be necessary for network formation. However, potential phosphorylation of these residues is not sufficient for network formation, as phospho-null mutations of S567 and S571 disrupt network formation even when the other residues carry phospho-mimicking mutations (A/E). Surprisingly, phospho-mimicking mutations of all conserved Ser/Thr residues (E/E), as opposed to just the residues other than S567 and S571 (WT/E), lead to a disruption of network formation. As S567 is known to be exclusively phosphorylated at the centrosome [196, 339], this suggests that the correct timing and location of phosphorylation matters for S567/S571 but not the other residues. Experiments and data analysis carried out and figure created by Isaac Wong with the help of Dr Zsofia Novak [unpublished results, reproduced here in modified form with permission].

It has previously been reported that mutating almost all conserved Ser/Thr residues C-terminal to S567/S571 (i.e. S573, T577, T581, S584, S589, but excluding S597) to alanine leads to a severe disruption of Cnn network assembly [196]. This suggests that phosphorylation of S567/S571 is necessary but not sufficient to allow network assembly *in vivo*, and that phosphorylation of at least some of the other residues C-terminal to S567/S571 might have a role in network formation. These findings have been largely confirmed by collaborators (Fig. 5.31). Considering that phosphorylation of S573 and T577 is not essential for network assembly (Fig. 5.29), that T581 is only weakly conserved compared to the other sites, and that S584 and S597 do not have (conserved) consensus Polo phosphorylation sequences (Fig. 5.1), this thesis predicts that S589 is the essential phosphorylation site that allows assembly-competent Cnn to form networks *in vivo*. Indeed, those Cnn mutants with a S589A mutation consistently have the worst phenotypes in the previously published phospho-null screen [196].

Furthermore, an essential role for phosphorylation of S589 is consistent with the finding in this thesis that the region I580-H608 is essential for network assembly *in vitro*. If the function of S589 phosphorylation is to enhance the proposed $Cnn^{A553\text{-}H608}$- or $Cnn^{I580\text{-}H608}$-mediated interactions, it is reasonable to assume that this would occur via binding of phosphorylated S589 to positively charged residues. The region C-terminal to K544 has a number of highly conserved positively charged residues (Fig. 5.1) that might act as interaction partners for phosphorylated residues from another dimer of Cnn. Indeed, in the structure of wild-type $Cnn^{D490\text{-}N579}$ presented here, several residues can be observed coordinating phosphates (Fig. 5.12). This might mimic an interaction between phosphorylated S589 and these residues *in vivo*. More speculatively, phosphorylated S589 might interact with positively charged residues located in the region D490-R518, some of

which are also highly conserved (Fig. 5.1). In the inactive state of Cnn the double hairpin fold probably results in close proximity of S589 with the region D490-R518 on the same dimer. However, formation of the assembly-competent state and thereby displacement of $Cnn^{A553\text{-}H608}$ might prevent phosphorylated S589 from interacting with the region D490-R518 on the same dimer, thereby necessitating an interaction with another dimer of Cnn and hence allowing network formation (Fig. 5.32).

Irrespective of what the precise interactions are that make up the eventual network, the key aspect that requires control by the cell is when and where the network is allowed to form. This thesis answers this question, as it is phosphorylation on S567/S571, controlled by the activity of Polo kinase, which governs the conformational change that allows transition of Cnn from an inactive "hairpin" state to an assembly-competent "C-terminus-bound" state. Unpublished *in vivo* data show that mutant Cnn that has phospho-mimicking mutations in all six conserved Ser/Thr sites C-terminal to S567/S571 (i.e. including S589E) can assemble networks that are not significantly different from wild-type networks. In contrast, if S567/S571 are also mutated to glutamate, Cnn network assembly fails (Fig. 5.31). This is in line with previously published data showing that mutation of the same residues (as well as T529 and S541) to glutamate or aspartate also leads to a disruption of centrosomal network formation *in vivo*, with Cnn networks forming in the cytosol instead [196]. The timing and location of S567/S571 phosphorylation, i.e. disruption of the hairpin fold, but not the timing of phosphorylation of the other residues, therefore appears to matter tremendously during centrosome assembly.

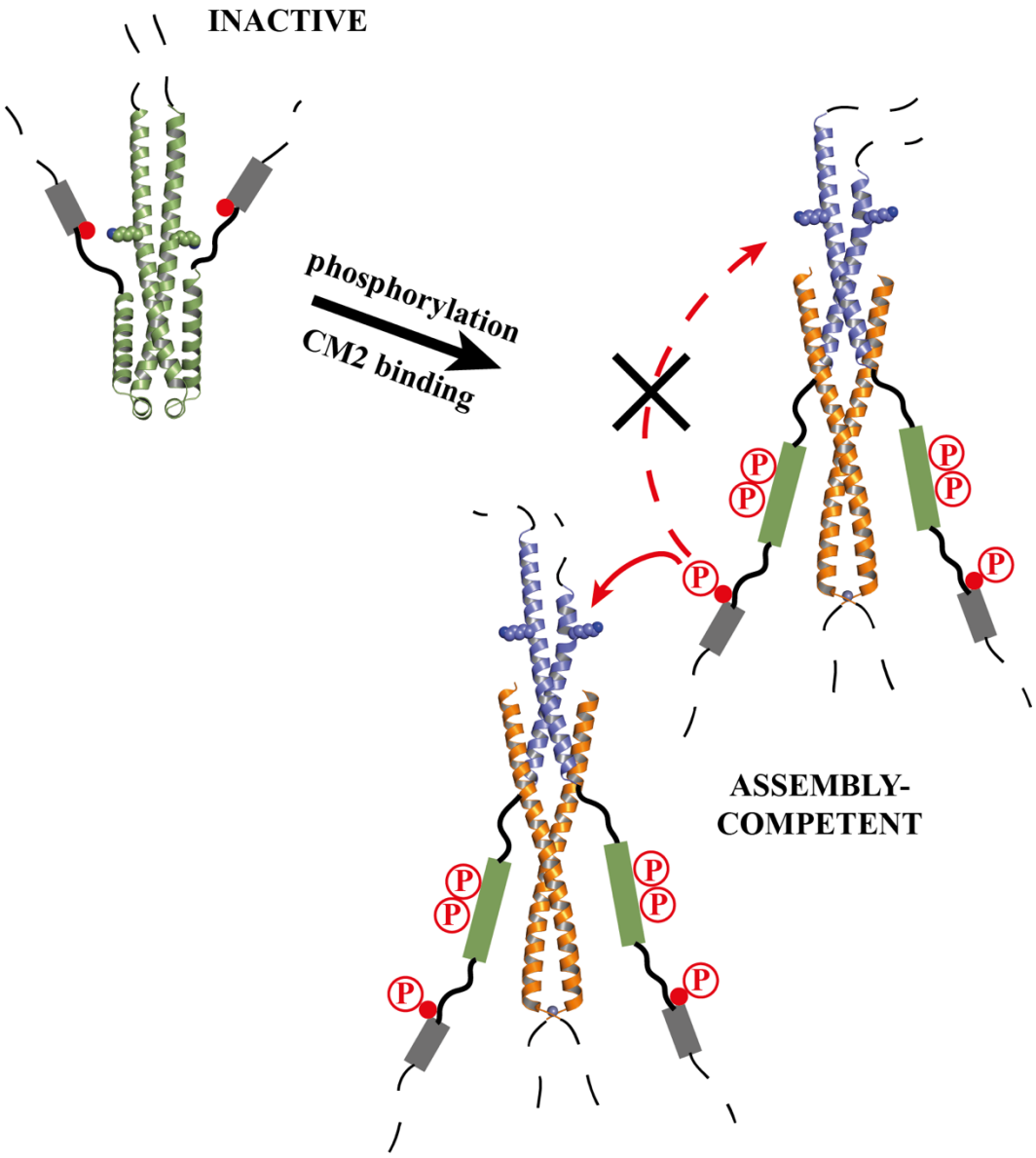


Figure 5.32 | Illustration of a possible alternative interaction allowing network formation. Inactive and assembly-competent states are shown as in Fig. 5.28, but in addition the side chain of the conserved residue K516 is shown as spheres (see Fig. 5.2 for conservation), and the putative location of residue S589 is shown as a red dot. In the inactive state, S589 is probably in close proximity to K516 on the same dimer. However, in the assembly-competent state, $Cnn^{A553\text{-}H608}$ might be displaced in a way that prevents interaction of S589 (now phosphorylated, indicated by a red encircled P) with K516 on the same dimer (dashed red arrow with a black cross). Instead, phosphorylated S589 can now only interact with K516 (or other positively-charged residues in this region) on a different Cnn dimer (solid red arrow), leading to network formation. Such displacement of S589 away from K516 on the same dimer could be ensured if in the assembly-competent state the helix A553-N579 (green box) interacts with CM2 (i.e. the Cnn C-terminus), e.g. by binding of the phosphorylated residues S567/S571 to a conserved patch of positively-charged residues towards the N-terminus of CM2 (not shown, see Fig. 4.1 for conservation). Note in this model, phosphorylation of S589 would not result in network assembly unless the assembly-competent state is adopted, as phosphorylated S589 in the inactive state would just interact with K516 on the same dimer. This is in line with the reported *in vivo* data showing that phospho-mimicking mutations of S589 but not S567/S571 are tolerated (Fig. 5.31).

Chapter 6 - Conclusions

This thesis provides some insight into why the pericentriolar material can only assemble at the centrosome and not the cytosol, revealing an important structural difference that underpins the proposed inactive and assembly-competent states of Cnn. Nevertheless, many further questions relating to centrosome assembly remain unanswered, requiring further investigation.

For example, it is currently unclear whether *in vivo*, the assembly-competent state of Cnn is formed via intra- or intermolecular interactions of the Cnn C-terminus and the zipper-like motif. In the *in vitro* experiments presented in this thesis, the interaction is necessarily inter-molecular, as two separate constructs are mixed together to allow binding and network formation. However, note that the resulting interactions would more closely resemble an intramolecular interaction *in vivo*, as the interaction between the Cnn C-terminus and the zipper-like motif *in vitro* does not per se result in linkage to a further dimer of Cnn. At least *in vitro*, therefore, networks can form that rely solely on interactions that would also be present during intramolecular formation of the assembly-competent state. The question will ultimately have to be addressed *in vivo* via careful labelling experiments.

In addition, it would be informative to obtain thermodynamic and kinetic data regarding the interaction between the zipper-like motif and the Cnn C-terminus both in the presence and absence of phosphorylation. Specifically, an *in vitro* experiment could be carried out in which $Cnn^{D490\text{-}N579}$ is pre-incubated with or without kinase before testing binding to $Cnn^{H1082\text{-}L1148}$ (or a larger construct containing the Cnn C-terminus) via e.g. isothermal titration calorimetry, surface plasmon resonance, or another suitable assay. Such

experiments could give valuable insights into the effect of phosphorylation from a thermodynamic and kinetic perspective.

The nature of the other interactions that make up the network also requires further investigation. This thesis makes an important contribution towards this by suggesting that $Cnn^{A553\text{-}H608}$-mediated interactions are involved in Cnn network formation and by providing a protocol for the successful purification of untagged $Cnn^{A553\text{-}H608}$ peptides, which can be used for further studies. Calmodulin can also be used as a tool to understand network assembly *in vitro*. For example, it will be important to determine from a structural perspective how Calmodulin can block network formation, possibly by trapping Cnn in an assembly-competent state by binding to $Cnn^{A553\text{-}H608}$. A structure of the tertiary complex between Calmodulin, $Cnn^{D490\text{-}H608}$ and $Cnn^{G1077\text{-}L1148}$ could offer valuable further information about the assembly-competent state. It will also be necessary to investigate whether the observation that Calmodulin can block network formation of Cnn constructs *in vitro* implies that Calmodulin potentially acts as a chaperone *in vivo*, and it will need to be studied, both *in vivo* and *in vitro*, whether Calmodulin can also prevent network assembly in the presence of phosphorylation.

Furthermore, it will have to be investigated which specific residues other than S567/S571 are phosphorylated and thereby have an essential role during network formation. This thesis hypothesises that phosphorylation on S589 has a critical role in this. This could be tested by introduction of a single point mutation (S589A) *in vivo* and scoring its effect on Cnn network formation. As a control, a Cnn mutant with alanine mutations in T581, S584 and S597 could be used, which should not have as strong a phenotype if S589 is the only key residue other

than S567/S571. Indeed, an antibody could be raised against phosphorylated S589 to confirm phosphorylation of this residue *in vivo*.

Next, mutant screens could be used to identify which (if any) positively-charged residues have a role in network assembly, possibly by interacting with the phosphorylated residues. Positively-charged residues are found for instance towards the N-terminus of the PReM domain (R501, R508, K516, R518) and clustered in the region K544-R562 (7/19 residues are Lys, Arg or His). Phosphorylated S589 could potentially interact with either of these clusters. Indeed, residues from the latter cluster are observed to coordinate phosphates *in crystallo* (Fig. 5.12), while an interaction of phosphorylated S589 with the former cluster would allow for the formulation of an attractive model describing network formation (Fig. 5.32). Two further positively-charged residues, i.e. R566 and K572, could primarily have a role in stabilising phosphorylation of their neighbouring residues, i.e. S567 and S571, respectively. R531 is also highly conserved, but this residue is located in the binding interface with the Cnn C-terminus and therefore probably not involved in a further interaction. Experimentally, a mutant screen could be carried out *in vivo* in which every residue from the two clusters is individually mutated to glutamate, i.e. the opposing charge. Upon identification of residues that result in disruption of Cnn network assembly when mutated, a charge-swap experiment could be carried out in which within the same mutant Cnn, S589 is also mutated to lysine. Having swapped the charges, this could restore a potential interaction between the two residues and thereby rescue the phenotype. The same charge-swap experiments could also be tested *in vitro* with the constructs used in this thesis.

In addition, it could be tested whether varying the length of the predicted linker between N579 and S589 (Fig. 5.2) affects network assembly, since a longer linker might allow

phosphorylated S589 to bind to its target site on the same dimer of Cnn even in the assembly-competent state, thereby preventing network formation. Some species of insects, such as *Ceratitis capitata* (Mediterranean fruit fly), have a Cnn homologue with a much shorter linker (see Appendix, Uniprot reference UPI00032A1AD1), which could fine-tune their ability to form mature centrosomes.

Finally, it will have to be addressed whether the structural mechanism revealed in this thesis for *Drosophila melanogaster* is also conserved in other species. The *D. melanogaster* protein Cnn was used in this thesis as any predictions based on structural findings can easily be tested *in vivo*, but in the future, it will be crucial to also study the mechanism of centrosome assembly of human CEP215, or indeed the molecular mechanism of the putative functional homologue in *C. elegans*, SPD-5. Although the exact details might differ, it is possible that the general concept of a phosphorylation-mediated conformational change, which for the first time is described at a molecular level in this thesis, is a conserved feature of centrosome assembly.

Appendix

Drosophila melanogaster Cnn isoform A (I303V, V522I)

```
        10         20         30         40         50         60
MDQSKQVLRD YCGDGNGTCA SSLKEITLIE TVTSFLEENG AAEIDRRVLR KLAEALSKSI

        70         80         90        100        110        120
DDTSPGALQD VTMENSYASF DVPRPPGGGN SPLPSQGRSV RELEEQMSAL RKENFNLKLR

       130        140        150        160        170        180
IYFLEEGQPG ARADSSTESL SKQLIDAKIE IATLRKTVDV KMELLKDAAR AISHHEELQR

       190        200        210        220        230        240
KADIDSQAII DELQEQIHAY QMAESGGQPV ENIAKTRKML RLESEVQRLE EELVNIEARN

       250        260        270        280        290        300
VAARNELEFM LAERLESLTA CEGKIQELAI KNSELVERLE KETASAESSN ANRDLGAQLA

       310        320        330        340        350        360
DKVCELQEAQ EKLKERERIH EQACRTIQKL MQKLSSQEKE IKKLNQENEQ SANKENDCAK

       370        380        390        400        410        420
TVISPSSSGR SMSDNEASSQ EMSTNLRVRY ELKINEQEEK IKQLQTEVKK KTANLQNLVN

       430        440        450        460        470        480
KELWEKNREV ERLTKLLANQ QKTLPQISEE SAGEADLQQS FTEAEYMRAL ERNKLLQRKV

       490        500        510        520        530        540
DVLFQRLADD QQNSAVIGQL RLELQQARTE VETADKWRLE CIDVCSVLTN RLEELAGFLN

       550        560        570        580        590        600
SLLKHKDVLG VLAADRRNAM RKAVDRSLDL SKSLNMTLNI TATSLADQSL AQLCNLSEIL

       610        620        630        640        650        660
YTEGDASHKT FNSHEELHAA TSMAPTVENL KAENKALKKE LEKRRSSEGQ RKERRSLPLP

       670        680        690        700        710        720
SQQFDNQSES EAWSEPDRKV SLARIGLDET SNSLAAPEQA ISESESEGRT CATRQDRNRN

       730        740        750        760        770        780
SERIAQLEEQ IAQKDERMLN VQCQMVELDN RYKQEQLRCL DITQQLEQLR AINEALTADL

       790        800        810        820        830        840
HAIGSHEEER MVELQRQLEL KNQQIDQLKL AHSTLTADSQ ITEMELQALQ QQMQEIEQLH

       850        860        870        880        890        900
ADSVETLQSQ LQKLKLDAVQ QLEEHERLHR EALERDWVAL TTYQEQAQQL LELQRSLDYH

       910        920        930        940        950        960
QENEKELKQT LVENELATRA LKKQLDESTL QASKAVMERT KAYNDKLQLE KRSEELRLQL

       970        980        990       1000       1010       1020
EALKEEHQKL LQKRSNSSDV SQSGYTSEEV AVPMGPPSGQ ATTCKQAAAA VLGQRVNTSS

      1030       1040       1050       1060       1070       1080
PDLGIESDAG RISSVEVSNA QRAMLKTVEM KTEGSASPKA KSEESTSPDS KSNVATGAAT

      1090       1100       1110       1120       1130       1140
VHDCAKVDLE NAELRRKLIR TKRAFEDTYE KLRMANKAKA QVEKDIKNQI LKTHNVLRNV

RSNMENEL
```

***Homo sapiens* CEP215**

```
        10         20         30         40         50         60
MMDLVLEEDV TVPGTLSGCS GLVPSVPDDL DGINPNAGLG NGLLPNVSEE TVSPTRARNM

        70         80         90        100        110        120
KDFENQITEL KKENFNLKLR IYFLEERMQQ EFHGPTEHIY KTNIELKVEV ESLKRELQER

       130        140        150        160        170        180
EQLLIKASKA VESLAEAGGS EIQRVKEDAR KKVQQVEDLL TKRILLLEKD VTAAQAELEK

       190        200        210        220        230        240
AFAGTETEKA LRLRLESKLS EMKKMHEGDL AMALVLDEKD RLIEELKLSL KSKEALIQCL

       250        260        270        280        290        300
KEEKSQMACP DENVSSGELR GLCAAPREEK ERETEAAQME HQKERNSFQE RIQALEEDLR

       310        320        330        340        350        360
EKEREIATEK KNSLKRDKAI QGLTMALKSK EKKVEELNSE IEKLSAAFAK AREALQKAQT

       370        380        390        400        410        420
QEFQGSEDYE TALSGKEALS AALRSQNLTK STENHRLRRS IKKITQELSD LQQERERLEK

       430        440        450        460        470        480
DLEEAHREKS KGDCTIRDLR NEVEKLRNEV NEREKAMENR YKSLLSESNK KLHNQEQVIK

       490        500        510        520        530        540
HLTESTNQKD VLLQKFNEKD LEVIQQNCYL MAAEDLELRS EGLITEKCSS QQPPGSKTIF

       550        560        570        580        590        600
SKEKKQSSDY EELIQVLKKE QDIYTHLVKS LQESDSINNL QAELNKIFAL RKQLEQDVLS

       610        620        630        640        650        660
YQNLRKTLEE QISEIRRREE ESFSLYSDQT FYLSICLEEN NRFQVEHFSQ EELKKKVSDL

       670        680        690        700        710        720
IQLVKELYTD NQHLKKTIFD LSCMGFQGNG FPDRLASTEQ TELLASKEDE DTIKIGEDDE

       730        740        750        760        770        780
INFLSDQHLQ QSNEIMKDLS KGGCKNGYLR HTESKISDCD GAHAPGCLEE GAFINLLAPL

       790        800        810        820        830        840
FNEKATLLLE SRPDLLKVVR ELLLGQLFLT EQEVSGEHLD GKTEKTPKQK GELVHFVQTN

       850        860        870        880        890        900
SFSKPHDELK LSCEAQLVKA GEVPKVGLKD ASVQTVATEG DLLRFKHEAT REAWEEKPIN

       910        920        930        940        950        960
TALSAEHRPE NLHGVPGWQA ALLSLPGITN REAKKSRLPI LIKPSRSLGN MYRLPATQEV

       970        980        990       1000       1010       1020
VTQLQSQILE LQGELKEFKT CNKQLHQKLI LAEAVMEGRP TPDKTLLNAQ PPVGAAYQDS

      1030       1040       1050       1060       1070       1080
PGEQKGIKTT SSVWRDKEMD SDQQRSYEID SEICPPDDLA SLPSCKENPE DVLSPTSVAT

      1090       1100       1110       1120       1130       1140
YLSSKSQPSA KVSVMGTDQS ESINTSNETE YLKQKIHDLE TELEGYQNFI FQLQKHSQCS

      1150       1160       1170       1180       1190       1200
EAIITVLCGT EGAQDGLSKP KNGSDGEEMT FSSLHQVRYV KHVKILGPLA PEMIDSRVLE
```

```
      1210        1220        1230        1240        1250        1260
NLKQQLEEQE YKLQKEQNLN MQLFSEIHNL QNKFRDLSPP RYDSLVQSQA RELSLQRQQI

      1270        1280        1290        1300        1310        1320
KDGHGICVIS RQHMNTMIKA FEELLQASDV DYCVAEGFQE QLNQCAELLE KLEKLFLNGK

      1330        1340        1350        1360        1370        1380
SVGVEMNTQN ELMERIEEDN LTYQHLLPES PEPSASHALS DYETSEKSFF SRDQKQDNET

      1390        1400        1410        1420        1430        1440
EKTSVMVNSF SQDLLMEHIQ EIRTLRKRLE ESIKTNEKLR KQLERQGSEF VQGSTSIFAS

      1450        1460        1470        1480        1490        1500
GSELHSSLTS EIHFLRKQNQ ALNAMLIKGS RDKQKENDKL RESLSRKTVS LEHLQREYAS

      1510        1520        1530        1540        1550        1560
VKEENERLQK EGSEKERHNQ QLIQEVRCSG QELSRVQEEL KLRQQLLSQN DKLLQSLRVE

      1570        1580        1590        1600        1610        1620
LKAYEKLDEE HRRLREASGE GWKGQDPFRD LHSLLMEIQA LRLQLERSIE TSSTLQSRLK

      1630        1640        1650        1660        1670        1680
EQLARGAEKA QEGALTLAVQ AVSIPEVPLQ PDKHDGDKYP MESDNSFDLF DSSQAVTPKS

      1690        1700        1710        1720        1730        1740
VSETPPLSGN DTDSLSCDSG SSATSTPCVS RLVTGHHLWA SKNGRHVLGL IEDYEALLKQ

      1750        1760        1770        1780        1790        1800
ISQGQRLLAE MDIQTQEAPS STSQELGTKG PHPAPLSKFV SSVSTAKLTL EEAYRRLKLL

      1810        1820        1830        1840        1850        1860
WRVSLPEDGQ CPLHCEQIGE MKAEVTKLHK KLFEQEKKLQ NTMKLLQLSK RQEKVIFDQL

      1870        1880        1890
VVTHKILRKA RGNLELRPGG AHPGTCSPSR PGS
```

Drosophila melanogaster Calmodulin

```
        10          20          30          40          50          60
MADQLTEEQI AEFKEAFSLF DKDGDGTITT KELGTVMRSL GQNPTEAELQ DMINEVDADG

        70          80          90         100         110         120
NGTIDFPEFL TMMARKMKDT DSEEEIREAF RVFDKDGNGF ISAAELRHVM TNLGEKLTDE

       130         140
EVDEMIREAD IDGDGQVNYE EFVTMMTSK
```

His_6-tagged monomeric NeonGreen as used in pETM14-mNG

(protease cleavage sites underlined)

```
         10         20         30         40         50         60
MKHHHHHHSA GLEVLFQGPM VSKGEEDNMA SLPATHELHI FGSINGVDFD MVGQGTGNPN

         70         80         90        100        110        120
DGYEELNLKS TKGDLQFSPW ILVPHIGYGF HQYLPYPDGM SPFQAAMVDG SGYQVHRTMQ

        130        140        150        160        170        180
FEDGASLTVN YRYTYEGSHI KGEAQVKGTG FPADGPVMTN SLTAADWCRS KKTYPNDKTI

        190        200        210        220        230        240
ISTFKWSYTT GNGKRYRSTA RTTYTFAKPM AANYLKNQPM YVFRKTELKH SKTELNFKEW

        250        260
QKAFTDVMGM DELYKENLYF QG
```

The following multiple-sequence alignment was used for Fig. 4.1. It was automatically created using $DmCnn^{G1077\text{-}L1148}$ as a template for 4 iterations of JACKHMMER (via the GREMLIN online tool [347, 348]) with an e value of 1E-10; sequences with less than 75 % coverage or more than 75 % gaps were excluded:

```
>seq_1_DmCnn1077-1148
GAATVHDCAKVDLENAELRRKLIRTKRAFEDTYEKLRMANKAKAQVEKDIKNQILKTHNVLRNVRSNMENEL
>UniRef90_B4JVG0/1207-1276
--TPMHDCAKVEQENADLRRKLIRTKRAFEDTYDKLRVVNKAKAQVEKDIKNQILKTHNVLRNVRSNMENVL
>UniRef90_UPI0006737806/1250-1319
--NKKHDCAKVDQENAELRRKLLRTKRAFEDTYEKLRNANQRKAQIEKDIKNQILKTHNVLKNVRSNMENEL
>UniRef90_A0A0K8TTD4/1159-1227
---RPHDCEKIEQEYSELKKKYNRTRRALEDTWNRLRSANQRKEQIERDIRQQIMKTQNVLKNVRTNMENEL
>UniRef90_A0A0M4EEH2/1105-1173
---ATHDCAKVDQEVAELRRKLIRTKRAFEDTYEKLRIVNLAKAQVERDIKNQILKTHNVLRNVRSNMENVL
>UniRef90_UPI00032A1AD1/1187-1252
---YKHDCIKVEQENAELRRKLAQTKRAFEETYKKLHLANQRKENIQKEIKSQIVKTKNVLKYARLNMA---
>UniRef90_UPI0006B7BA9D/1191-1257
--HPIHDCIKVEQENAELRRKLVQTKRAFEETYKKLHVANQRKETFQKEIKDQILKTKNVLKYARLHMA---
>UniRef90_Q0IER0/1012-1081
--SAIHDCVKIEADFAELKRRYKQTRRHLELAYDSIKSSNKRKEQLERDIKQQIQKTNVVLKTVRTNMTNEL
>UniRef90_B0WXU8/967-1036
--TPPHDCAKIEADYAELKRRYKQTRRHLELAYDSIKSANRRKERLEQDIKQQIQKTNVVLKTVRTNMTNEL
>UniRef90_N6SW03/972-1040
---NEHCCQKIEQENNELKRKLLRMKRALEETATQLNLANQRKKQVEKTICKQIHKTSQVLRKAKANLD---
>UniRef90_D6WDS7/1346-1414
---NVHCCQKMAHENSDLKRKLLRTRRALEETVTQLTLANQRKKQVEKTICKQIHKTSQVLRKAKANLD---
>UniRef90_U5ER02/851-918
--TLKHDCDKIEQELSTVKRKYHSTRRALEKAFEQIRSSNQLKEQVERDIRKQIQTTHAVLKTARTNMEH--
>UniRef90_A0A084VU56/1154-1219
---VKHDCAKIEADYAELRRRYKETRRCLTQAYDNIKQANKRKEQLEVDIKQQIHKTRDVLRTVHSNMD---
>UniRef90_E0VJP0/1600-1660
--------KKLENENAELRRRLLRTRQALEETLTQLSAANQRKKQVERAICKQLHKTHHILKKARVNFE---
>UniRef90_A0A0T6B907/1557-1625
```

---NEQCCLKIVNENNDLRKRLLRTRRALEETVSQLTIANQRKKQVEKTICKQIHKTSQVLRKAKANLD---
>UniRef90_UPI0006D50AA5/1284-1344
--------DKLELENVELRAQLARTKQALEETISQLTSANQRKRQVEKAICRQLHKTHDILKQARLNLD---
>UniRef90_W5JBS7/1134-1199
---VKHDCAKVEADFAELKRRYKVTRQSLNQAYDNIKSANKRKEQLEVEIKQQIHKTRVVLKTVHNNME---
>UniRef90_E2BTV2/1271-1328
---------ALKQENESLKRRLMKTRRALEDTFEHLSASNKNKKNVEKAIIKQLMITKNVLKKTRT-F----
>UniRef90_UPI00062365F9/1334-1390
---------ALKQENEALKRRLMKTRRALEDTFEHLSASNKNKKIVEKAIIKQLMVTRSILKKTRT------
>UniRef90_UPI00059631DF/1269-1325
---------ALKQENEALKRKLMKTKKALEDTFEHLSTSNKNKKNVEKAIIKQLMVTRSVLKKTRT------
>UniRef90_UPI0005B84E4F/1294-1349
----------LKQENKALKQKLMKTRRALEETFEHLSASNKNKKNVEKAIIKQLMTTRSILKKTRT------
>UniRef90_UPI00021A7860/1341-1397
---------ALKQENETLKRRLMKTRRALEDTFQHLSASNKNKKNVEKAITKQLQITKSILKKTRT------
>UniRef90_Q7QC40/1154-1219
---IHHDCAKIEVDYRDLMHRYNKTRHYLAVAYDKIKSSNKVKKQLEIEVKQQIHKTNVVLKTVQRNLD---
>UniRef90_K7IPX9/1311-1368
--------QSLKEENVALKRRLMKTRRALEDTFQHLSASNKNKKNVEKAITKQLMITKSVLKKTRN------
>UniRef90_A0A067REP9/2238-2298
--------KELEQENIVLKQRLTKTHRALEETMAQLTTANQRKKQVERAICRQLHKTHHILKTARVSLK---
>UniRef90_UPI0005CE9AD7/1288-1345
--------QTLKQENVVLKKRLIKTRRALEDTFQHLSASNKNKKNVEKAITKQLMITKSILKKTRN------
>UniRef90_UPI0006C9D22D/1274-1329
----------LKNENAALRRRLMRTKRALEDTFHYLSASNKNKRSVEKAITIQLMITKNILKKTRN------
>UniRef90_UPI0006793B10/660-719
----------LYEENVVLKQKLARTKATLAETLTQLNAANMRKRSVQRAICREIHKTQGVLRKARDQFEN--
>UniRef90_UPI00042C8474/266-324
----------MKRELLELRTKLSKQESLLRSTAEHLKAANQQKESMEQFIFSQLSRTHDVLKKARTNLE---
>UniRef90_UPI00033177D5/2327-2385
----------VNRELLELRTKLSKQETLLQSTVERLKTANQQKESMEKFIFSQLTRTHDVLKKARTNLE---
>UniRef90_S4PMK9/645-703
----------LYEENKILKQKLARTKATLAETLIQLNASNMRKMGVQRAICREIHKTHGVLRQVRDHVE---
>UniRef90_L8Y5Z6/2194-2252
----------MERELLDLRTKVSKQERLLQSTAERLKTANQQKESMEQFIVNQLTRTHDVLKKARTNLE---
>UniRef90_UPI0003442705/2384-2443
---------SLQRELLELRTRLSKQESVLQSTAERLKTANQQKESMEQFIFNQLTRTHDVLKKARTNLE---
>UniRef90_UPI000661C8F0/2416-2475
---------TVERELLELRAQVSKQEQLLQSTAERLKTANQHKKSMEQFIISQLTRTHDVLKKARTNLE---
>UniRef90_UPI00046B1CE0/363-420
-----------ERELLELRTKVFTQERLLQSTAERLQAANQQKEAMEQFILSQLTRTHDVLKKARTNLE---
>UniRef90_UPI0005CFFBF6/2108-2165
-----------ERELLELRTKVSKQEQLLQSTTEHLKNANQQKESMEQFIVSQLTRTHDVLKKARTNLE---
>UniRef90_UPI0003F10963/2400-2460
--------ASMKRELLELRTKLSKHETLLQSTTEQLKTANQQKESMEQFIFSQLTRTHDVLKKARTNLE---
>UniRef90_UPI0004439A33/2238-2296
----------MKRELLELKTKLSKQESLLQSTTESLKLANRQKESMEQFIFSQLTRTHDVLKKARTNLE---
>UniRef90_UPI0006253E0A/1336-1392
---------MLKQENKALKRKVQKTRKALKNTFQQLSESNKNKKDVEKAITKQLTITKNILMKTRT------
>UniRef90_UPI000642D7AA/2342-2400
----------VRRELLELRAKLSKQESLLQSTAERLKTSNRQKDSLEQFIVSQLTRTHDVLKKARSNLE---
>UniRef90_UPI00048A5FDA/2385-2445
---------SVERELVELRARVSEQQRLLQGTAERLRTANQQRESMEQFIVSQLTRTHDVLKKAKTNLEH--
>UniRef90_I3NHU7/1008-1066
----------MERELLDLRAQVSKQEKLLQTTAEHLKTANQQKENMEQFIVNQLTRTHAVLKKARTNLE---
>UniRef90_UPI00074FFC27/1531-1589
----------MKAELHQLRRKLAEQEKTLHSTMMHLHSTNQLKENMEKVIINQLALTHDVLKKARGNLE---
>UniRef90_H2ZA93/312-370
----------LRREVGKLRRRLSTQDKLLQSTVERLQTTNRMKEGIEHAIVKQLSKTHDVLRKARGNLE---
>UniRef90_V9K8X0/1435-1499
-----HKTDKVKDEIHQLRKKLTEQERLLYGTVKRLRTTNQLKEGMEKVIIDQLSLTHNVLKTARGNLE---
>UniRef90_W5M5P2/2725-2790
---QSHQGEMLKNEITRLRSKLSQQERLLTGTVKRLHTTNQLKEGMEKVIISQLSLTHGVLKKARGNLE---
>UniRef90_UPI0004ED70EF/2413-2471

----------MEKELRDLRVQVSQQEQLLQSTAERLKTANQQKESMEQFIVNQLTRTHDVLKKARTNLE---
>UniRef90_UPI00072E3C81/2252-2311
---------SLKRELLELRTKLSAQENLLQSTAERLKTANQQKQTMERIIFSQLTRTHDVLKKARTNLE---
>UniRef90_UPI0007685475/2152-2210
----------VRRELLELRTKLSRQEGLLQSTAERLQSATQQKESLEQFIFSQLTRTHDVLKKARTNLE---
>UniRef90_A4F2N3/1691-1753
------DADMLRREVGKLRKRLSAQDHLLQSTVDRLQNANKMKEGIEHAILKQLTVTHDVLKKARGNLE---
>UniRef90_UPI000769BE87/2324-2383
---------SLREEVVCLKRRLSEQEQALRDTMDSLKSSNRTKDSMEQFIVNQLSRTRDVLKKARTNLE---
>UniRef90_H9G870/527-585
----------MRTELHKLRRKLAEQEKKLHSTVKRLHSTNQLKENMERVIIDQLALTHDVLKKARGNLE---
>UniRef90_UPI000388A994/1860-1926
----AHCSQDMKAEILRLHKKLSEQEKKLHSTVKRLHSTNQLKENMEKVIIDQLALTHDVLKKARGNLE---
>UniRef90_UPI000704620D/2240-2297
-----------TQELLELRTKLSRQESLLQSTAERLRTASRQKESLEQFIFSQLTRTHDVLKKARTNLE---
>UniRef90_UPI00051E393F/121-179
----------LQGEIAALRAQLSEREDALQSTAERLRSTARLKDSMEQFIVSQLTRTHNVLRKARTNLE---
>UniRef90_A0A0G2K463/2244-2302
----------MEKELLDLRAQVSQQQQLLQSTAVRLKTANQRKKSMEQFIVSHLTRTHDVLKKARTNLE---
>UniRef90_UPI0005123087/1176-1233
-----------ERELLELRTKVSKQERLLQSTAERLKTTNQQKESMKQFIVSQQARTHDVLKKARTNLE---
>UniRef90_UPI0004A4930D/3119-3177
----------LKNEIARLKSRLSQQERMLSGTVKRLRTTNQLKEGMERVIIDQLSLTHGVLKKARGNLE---
>UniRef90_UPI0005775B56/2083-2141
----------LKEEVHTLRRRVSEQEEALRDAMETLKSSNRTKDSMEHFIVSQLSRTKDVLKKARTNLE---
>UniRef90_UPI0006C989D2/1170-1226
--------VKLRQENITLKKRLMRTRKALEETFEHLMMSNKNKENVKKAITQQLVTMKSVSKSLK-------
>UniRef90_H2VBR0/1745-1803
----------MKNEVARLKSRLSQQERMLCGAVKRLRTTNQLKEGMEKMIIDQLYVTHGVLKKARGNLE---
>UniRef90_UPI00044317BA/2350-2409
----------MKGEILELKTKLSEQENLLQSTNERLKTTNRLKESMEQFIINQLTRTHDVLKKARTNLEQ--
>UniRef90_UPI00051C09C6/1835-1893
----------MKAEILRLRKKLSEQEEKLHRTVKRLHSANMEKEHMEKGIIDQLVLTRDILKKARENLE---
>UniRef90_H2N647/1316-1373
-----------ERELLEPRTKVSKQEWLLQRTAERLKTANQQKESMEQFIISQLARTYNVLKKARTNLE---
>UniRef90_UPI000441F376/1350-1408
----------LKSELCRLRRKLAEQEQQLHSTAKCLYSTNRLKENMERVIIDQLALTHDVLKKARGNLE---
>UniRef90_UPI0004F47330/2771-2829
----------LQSEISRLKSRLSQQERMLSGAVKRLRSTNQLKEGMERVIIDQLSLTHGVLKRARGNLE---
>UniRef90_V9K849/1381-1440
---------SMKEEIFRLRSKIAEQEDLLHNAISRLKATNQVKDGMEQFIVNQLTRTHDVLKKARSNLE---
>UniRef90_UPI0007429C67/2075-2133
----------LKDEVASLRLKLSEQEKALKDAMERLKSSNLTKDSMEHFIVSQLSRTRDVLKKAKGNLQ---
>UniRef90_UPI000576A131/2936-2994
----------LKNEIARLKSRLSQQEKLLTGAVRRLRTTNQLKEGMERVIIDQLSITHGVLKKARGNLE---
>UniRef90_UPI000496FD0C/2420-2478
----------LKEELSSLRLKLSEQEQVLKDTVERLQSSNRTKDSMEHFIISQLSRTRDVLRKAKTNLQ---
>UniRef90_H3AEU2/2272-2330
----------MKEDLRKLRRRITEQESLLQNTIDRLKTTNRMKENMEQFIIGQLNRTHNVLKKARGNLE---
>UniRef90_UPI000644BC35/1755-1814
---------SLREEVVSLRLKLSEREQALKDTMETLRSSNLTKDSMEYFIVSQLSRTRDVLKKAKTNLQ---
>UniRef90_UPI000644E719/2854-2912
----------LRNEITRLKNRLSQQERMLSGAVRRLRTTNQLKEGMERVIIDQLSLTHGVLRKARGNLE---
>UniRef90_UPI000454B4D3/2340-2398
----------IEKEILDLRAQVSQQEQLLQKTNERLKTAKQKKKNMEQYIVNHLTRTHDVLKKARTNLE---
>UniRef90_F6Z0L3/2224-2284
--------QAMKEEICQLKNKLKEQEVALQEAKEHLKSTNRAKESMEHFIVSQLTRTHDVLKKTRSNLE---
>UniRef90_UPI0006443C67/1955-2012
-----------DEEITDLKKKLEEQQQALKEALENLRASNQTREGMENFITSQLSRTRDVLKKARSNLE---
>UniRef90_UPI000549A512/1046-1105
---------ALQNEITALRAQLSKREDALQSTAKQLHSMAQLKDSMEQFIVSQLTRTHNVLHKARTNLE---

The following multiple-sequence alignment was used for Fig. 5.1. It was automatically created using DmCnn$^{D490\text{-}H608}$ as a template for 4 iterations of JACKHMMER (via the GREMLIN online tool [347, 348]) with an e value of 1E-10; sequences with less than 75 % coverage or more than 75 % gaps were excluded:

>seq_1_CnnD490-H608_Drosophila-melanogaster
DQQNSAVIGQLRLELQQARTEVETADKWRLECVDVCSVLTNRLEELAGFLNS
LLKHKDVLGVLAADRRNAMRKAVDRSLDLSKSLNMTLNITATSLADQSLAQLCNLSEILYTEGDASH

>UniRef90_B4GGT2/499-615_Drosophila-persimilis
DHQNSAVIGQLRQELRQVQADAESANKWRLECADVCGVLTNRLEELAGFLNS
LLKHKDVLGALAADRRHAMRRAVDRSLDLSKSLNMTLNITGASLADQSLAQLCNLSDMLYNDADA—

>UniRef90_A0A0P9C0K7/761-878_Drosophila-ananassae
DQQNSAVISQLRMELKKAHADVESADKWRLECADVCGVLTNRLEELAGFLNS
LLKHKDVLGVLAADRRHAMRKAVDRSLDLSKSLNMTLNITGTSLADQSLAQLCNLSEILYTEGDLS-

>UniRef90_B4LNQ0/499-614_Drosophila-virilis
-QQDSELIAQLRQEVQTMELDVENANKWRLECADVCSVLTNRLEELAGFLNS
LLKHKDVLGVLAADRRHAMRRAVDRSLDLSKSLNMTLSITGVSLADQSLAQLSKLSEILYTEADE—

>UniRef90_A0A0Q9WXB0/492-605_Drosophila-willistoni
-QETSVLVAKLQHELSSARADAESANKWRLECADVCGVLTNRLEELAGFLNS
LLKHKDVLGVLAADRRHAIRKAVDRSLDLSKSLNMTLSITGVSLADQSLSQLCNLSEFLDTEG----

>UniRef90_B4KTP3/497-609_Drosophila-mojavensis
-QEDSKLVSQLKHELQSAQLNMENANKWRLECADVCSVLTNRLEELAGFLNS
LLKHKDVLGVLAADRRHAMRRAVDRSLDLSKSLNMTLSITGVSLADQSLAQLSKLSEMLYNN-----

>UniRef90_B0WXU8/332-447_Culex-quinquefasciatus
DRNENIIIKQLNEELIQARDETDKAQRWRKDCMDLCAISTARLEELAGFLDS
LLRNKEFSGSLPVDRRKAMRKAVDRSLDLSRSLNMSISVTGFSLPESSLAHLSCLSGYLDAGED---

>UniRef90_B4JVG0/643-755_Drosophila-grimshawi
--QSSQLVPQLRQQLQSMQTDVELANKWRLECADVCAVLTNRLEELAGFLNS
LLKHKDVLGVLAADRRHAMRRAVDRSLDLSKSLNMTLSVTGVSMADQSLAQLSKLSEILLYTD----

>UniRef90_A0A0M4EEH2/543-658_Drosophila-busckii
-QQHSSLTEHLRQELQATQADMESANKWRIECAEVCSVLTNRLDELAGFLNS
LLKHKDVLGVLAADRCRAMRRAIDRSLDLSKSLNMTLSITGTTLADQSLAQLSKLSELLYTEADN--

>UniRef90_W5JBS7/452-565_Anopheles-darlingi
--DRNIIIKQLNDELIQAREEAEKAQKWRKECTDLCAVSTIRLEELAGFLDS
LLKNKELVCNLSTDRRKAIRKAVDRSLDLSRSLNMSISVTGFSLTGNTLAQLSCLSGYLENSVD---

>UniRef90_UPI0006737806/621-732_Stomoxys-calcitrans
-RNTDAVISQLRIELKQAQANAENAEKWRRQCADLCSTLSTRLEELAGFLNS
LLKHKDVLGVLAVEKRKAMRRAVDRSLDLSKSLNMTLSVSGISMIDQSLAELHNLSDILGD------

>UniRef90_A0A0L0CM97/751-862_Lucilia-cuprina
-RQNDALIAQLQTELKQARLDAETAEKWRRQCADLCQTLSTRLEELAGFLNS
LLKHKDVLGVLAVEKRKAMRRAVDRSLDLSKSLNMTLSVSGLSMIDQSLAELHNLSDILGD------

>UniRef90_A0A084VU56/463-574_Anopheles-sinensis
--DRNIIIKQLNDELIQAREEAEKAQRWRKECADLCAISTLRLEELAGFLDS
LLKNKEFVGTLSIDRRKAMRKAVDRSLDLSRSLNMSISVTGFSTANNSLAQLSCLSGFLE-------

>UniRef90_A2I875/588-701_Aedes-aegypti
DRNENIIIKQLNEEVIQAREEAEKAQRWRKECVDLCAISTARLEELAGFLDS
LLRNDEFSGTLSIARRKAIRKAVDRSLDLSRSLNMSISVTRFSLPENSFVNLSSLTGFLNGT-----

>UniRef90_Q7QC40/461-575_Anopheles-gambiae
--DRNIIIKQLNDELIKAREEAEKAQKWRKEYADLCAISTHRLEELAGFLDS
LLRNKELIGSLSTDSRKAIRNAVDRSLDLSRSLNMSISVTGLSLINNSLAQLSCLSGYLDQS-----

>UniRef90_U5ER02/336-448_Corethrella-appendiculata
-RNENVIIKQLNDELIIARSDYETVYKRQKDCADICQTLTNHLEELAGFLNS
LLKNKEFSNDLPEGRRNAMRKAVNNSLDLSRTLNLSISVHNFSLPDTSLKNLSRLSGILDDK-----

>UniRef90_A0A0K8TTD4/549-660_Tabanus-bromius
----SEFVQKLKLELKQAREETERANVGRREVTEMCLVLTARLEEIAGFLDS
LLRHKDVLGALGADRRKAIRKAVDKSLDLSRSLNMSLSFN-GSFNENSLANLSNISALLDQSHE---

>UniRef90_A0A0A1X4T0/482-591_Zeugodacus-cucurbitae
-RSNEALIEELRQEARTARAEADNAETCRHAYAKVFGALSDRLYELAGFLNS
LMKHKEVLSFLSTDRRRAMRTAVDCSLDLSTSIRETLT-----TGDQSFEGLSNLSRLLLDEDE---

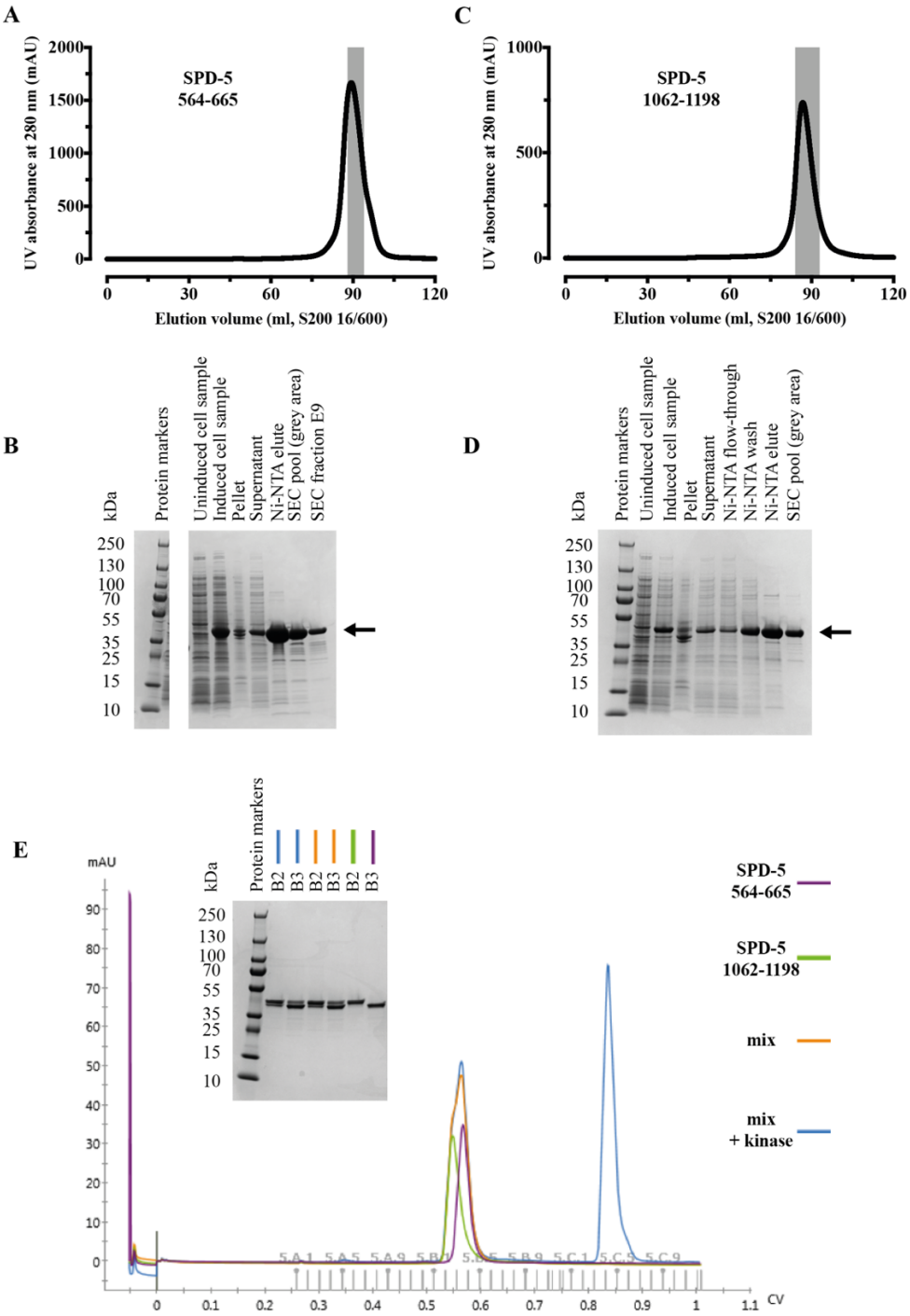
>UniRef90_UPI00032A1AD1/495-604_Ceratitis-capitata
-RSNEALIEELRLEARTARAEADNAETSRRAYIKVFSALSDRLTELAGFLQS
LMKHKEVLNFLSTDRRRAMRTAVDCSLDLSTSIRETLS-----TGDQSMIDISNLSRLLLDDDE---

>UniRef90_N6SW03/233-335_Dendroctonus-ponderosae
-----AMFDSLKSENQTIKSELDKSEQLRHEMGEVCSVLSNRLEELAIFLDS
LLKHKSVLGFLGLAKNRKLREIISNSLDMSRSFTMSLMVN----PEQSLTQLSNITALLN-------

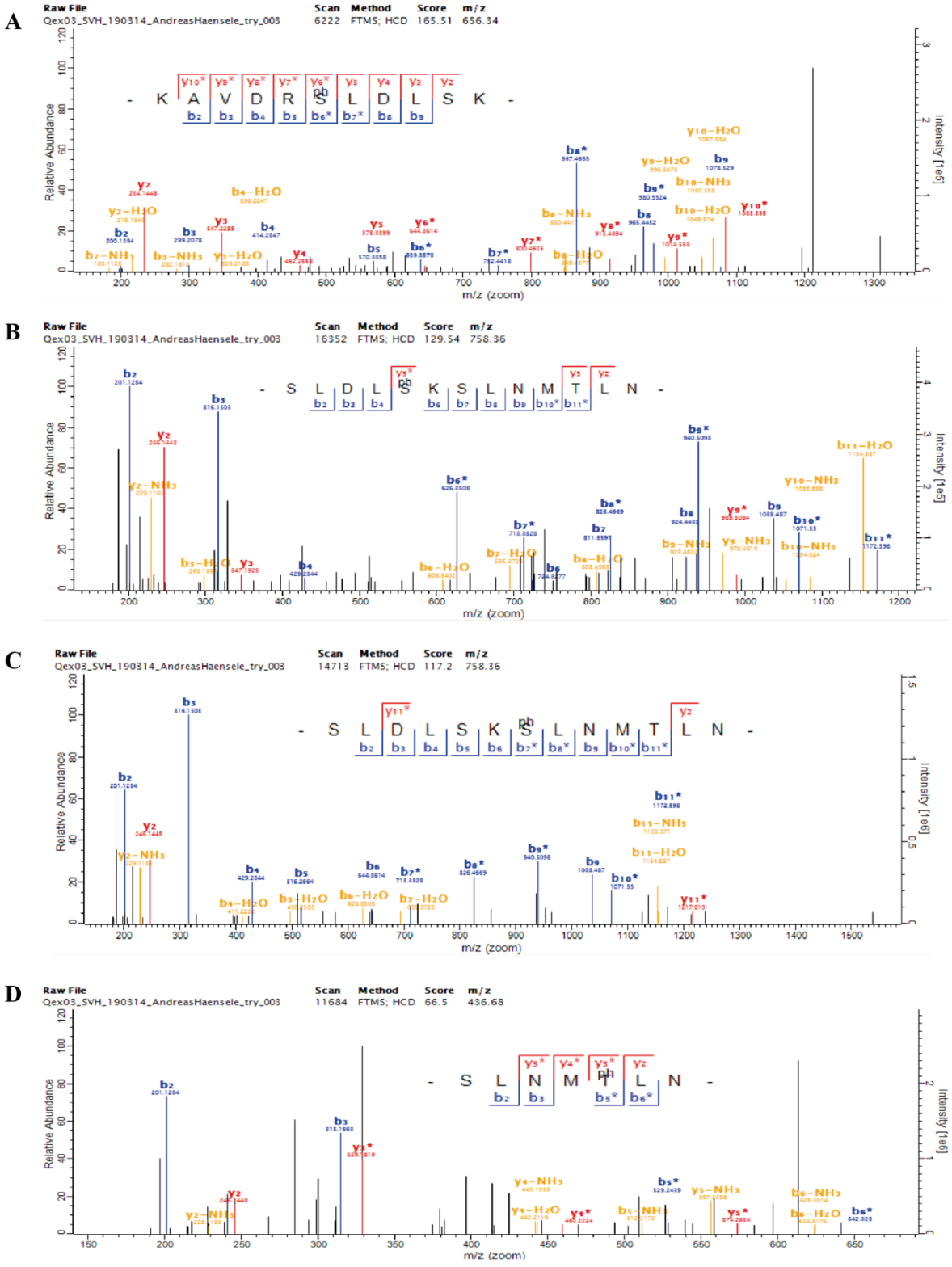
>UniRef90_A0A067REP9/921-1030_Zootermopsis-nevadensis
--------RTLRSECARLREEIDRANVWRKEAGEVCSLLTKRLEEVAGFLSS
LLSHPEQLGGLRTKHRQLLKQVIERSMELSRSLSVSLSINDQDLTNGSFPPLMSFSSLLMSAGDV--

>UniRef90_A0A0T6B907/746-846_Oryctes-borbonicus
-------IDILQKEHAELQEELEKCERLRAETNDVCIILTNRLEELACFLDS
LLKQKSILGFLGHQQNEKIRRLINQSLDISRTLTHSMSMD----PDQSLLQLSNITNILG-------

>UniRef90_D6WDS7/642-736_Tribolium-castaneum
----------NLQNELEKSEKLRLEANEVCAVLSRRLEELAVFLDS
LLKQKSVLGFLGNKENKRIRHIIDQSLDLSRTFTMSMIIN----PDKSLLQLTNISSLLNDTN----



App. Figure 1 | Purification and characterisation of two *C. elegans* SPD-5 constructs. A-D) *C. elegans* SPD-5 constructs 564-665 (A/B) and 1062-1198 (C/D) were cloned into pETM14-mNG vectors and purified via the usual protocol without tag cleavage. SEC profiles are shown in A and C, with pooled areas containing the tagged protein of interest marked in grey. Corresponding purification gels are shown in B and D, with a black arrow indicating the tagged protein of interest. Thanks to Martine Ruer (Hyman laboratory) for supplying the original full-length SPD-5 plasmid. **E)** Analytical SEC (S200 10/300) was carried out with mNG-SPD-5$^{564\text{-}665}$ alone (purple, 25 µM), mNG-SPD-5$^{1062\text{-}1198}$ alone (green, 25 µM), both mixed together (orange) and both mixed in the presence of human PLK1 kinase (blue, incubation time 1 hr at 30°C). The corresponding gel is shown as an inset. Note no peak shift can be observed, suggesting no binding under these conditions (running buffer: standard TBS).



App. Figure 2 | Mass spectrometry analysis of Cnn$^{D490-N579}$ phosphorylation. Spectra of trypsin-digested peptides with phosphorylation on S567 (A), S571 (B), S573 (C) or T577 (D) are shown. Peptides are shown as inserts, with N-terminal fragments in blue (b_x) and C-terminal fragments in red (y_x). Phosphorylated fragments (*) and phosphorylation site (ph) are marked. Experiments were carried out by Dr Svenja Hester.

Abbreviations

APS	ammonium persulphate
Asl	Asterless
Asp	Abnormal spindle protein
ATP	adenosine triphosphate
a.u.	arbitrary units
BME	β-mercaptoethanol
$CaCl_2$	calcium chloride
CaM	Calmodulin
CM1	Cnn motif 1
CM2	Cnn motif 2
Cnn	Centrosomin
CV	column volume
DTT	dithiothreitol
ESRF	European Synchrotron Radiation Facility
FL	full-length
γ-TuRC	γ-tubulin ring complex
mGFP	monomeric green florescent protein
HCl	hydrogen chloride
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
H_2O	water
IPTG	isopropyl β-D-1-thiogalactopyranoside
LB	Luria Bertani Broth
LZ	Leucine zipper-like motif
MALS	multi-angle light scattering

MBP	maltose binding protein
$MnCl_2$	manganese chloride
mNG	monomeric neon-green protein
$MgCl_2$	magnesium chloride
MR	molecular replacement
MS	mass spectrometry
NaCl	sodium chloride
Na_3VO_4	sodium orthovanadate
NEB	New England Biolabs
Ni	Nickel
NMR	nuclear magnetic resonance
NTA	nitrilotriacetic acid
PEG	polyethylene glycol
PCM	pericentriolar material
PCR	polymerase chain reaction
PDB	protein data bank
PLK1	Polo-like kinase 1
PLK4	Polo-like kinase 4
Plp	Pericentrin-like protein
PReM	Phospho-regulated multimerisation domain
RbCl	rubidium chloride
RMSD	root-mean-square deviation
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
TB	terrific broth

TEV	Tobacco Etch Virus
TBS	Tris-buffered saline
TCEP	tris(2-carboxethyl)phosphine
TEMED	tetramethylethylenediamine
WT	wild-type

Acknowledgments

I would like to thank Prof Susan Lea for her supervision, crystal fishing and X-ray data collection, Prof Jordan Raff for his supervision, and Dr Steven Johnson (Lea laboratory) for his advice and his help with SEC-MALS and structure determination. I would further like to thank all three for critical reading of the manuscript. In addition, I would like to thank all other members of the Lea and Raff laboratories, past and present, and all staff at the Sir William Dunn School of Pathology.

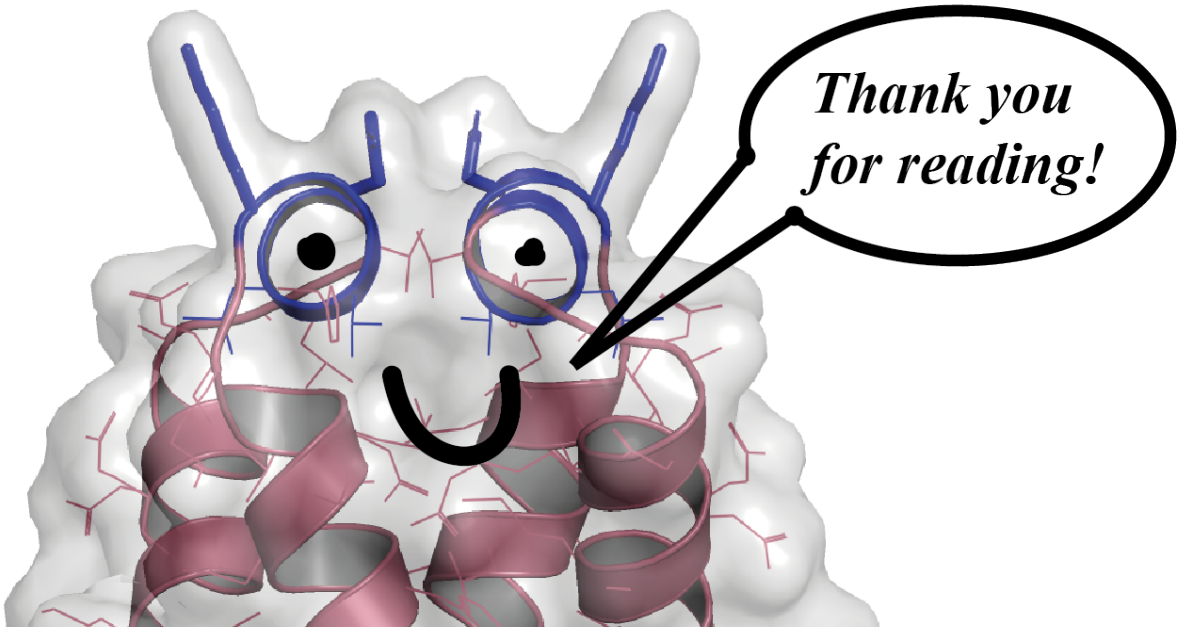
My sources of funding deserve special thanks for giving me the opportunity to pursue this degree: The Clarendon Fund (tuition fees, college fees, ~ 1/3 of living expenses), the Sir William Dunn School of Pathology (~ 2/3 of living expenses) and the Santander Group (£5,000 graduate award).

I would also like to thank the following researchers, who contributed to experiments as detailed in the thesis and described again here: Dr Violet Feng cloned constructs $Cnn^{H1082-L1148}$ (and its mutants), $Cnn^{D490-K544}$ (and its mutants), $Cnn^{Q403-H608}$-mGFP and MBP-$Cnn^{Q403-H608}$, and prepared samples for SEC-MALS of mutant $Cnn^{D490-K544}$ constructs. Dr Violet Feng (and Dr Alice Ahn) also provided TEV protease for the studies in this thesis. Joseph Fisher (Part II student, Lea laboratory) cloned and purified constructs Cnn^{D490-x}, where x is S567, S573, M576, N579, L585 or H608 under the author's direct supervision. Where indicated, SEC-MALS analysis of mNG-Cnn^{D490-x} as well as purification of and experimentation on $Cnn^{M1-E1049}$ were also carried out by Joseph Fisher under the author's direct supervision. Thomas Griffiths (summer student, Lea laboratory) aided in the cloning and purification of $Cnn^{M1-K360}$ and $Cnn^{G604-E1049}$ under the author's direct supervision. Dr Alan Wainman helped with fluorescence microscopy experiments in Chapter 3. I would

also like to thank Dr Anna Caballe, Dr Zsofia Novak and Isaac Wong, who contributed to the *in vivo* analysis, and Dr Ines Alvarez Rodrigo, who advised on PLK1 phosphorylation assays. Data produced by Isaac Wong are reported in the Discussion of Chapter 5 (Figs. 5.29 and 5.31) with his permission. Dr Svenja Hester (Oxford Proteomics Facility, University of Oxford) carried out mass spectrometry experiments, and Dr David Staunton (Molecular Biophysics Suite, Department of Biochemistry, University of Oxford) carried out SEC-MALS experiments where indicated. Sincere apologies to anyone I forgot to mention.

Where indicated, figures previously published in Feng, Z., Caballe, A., Wainman, A., Johnson, S., Haensele, A.F.M., Cottee, M.A., Conduit, P.T., Lea, S.M. and Raff, J.W., *Structural Basis for Mitotic Centrosome Assembly in Flies.* Cell, 2017. 169(6): p. 1078-1089 (DOI: https://doi.org/10.1016/j.cell.2017.05.030) are reproduced in modified form here under Creative Commons licence CC-BY.

Finally, and most importantly, thanks to all my friends and family for everything.



Bibliography

1. Boveri, T. *Zellen-Studien*. (G. Fischer, Jena; 1900).
2. Schaefer, E. *et al.* The preprophase band of microtubules controls the robustness of division orientation in plants. *Science* **356**, 186-189 (2017).
3. Drevensek, S. *et al.* The Arabidopsis TRM1-TON1 interaction reveals a recruitment network common to plant cortical microtubule arrays and eukaryotic centrosomes. *Plant Cell* **24**, 178-191 (2012).
4. Kilmartin, J.V. Lessons from yeast: the spindle pole body and the centrosome. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**, 20130456 (2014).
5. Katsaros, C., Karyophyllis, D. & Galatis, B. Cytoskeleton and morphogenesis in brown algae. *Ann Bot* **97**, 679-693 (2006).
6. Powell, M.J. Mitosis in the Aquatic Fungus Rhizophydium-Spherotheca (Chytridiales). *Am J Bot* **67**, 839-853 (1980).
7. Azimzadeh, J. Exploring the evolutionary history of centrosomes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**, 20130453 (2014).
8. Hodges, M.E., Scheumann, N., Wickstead, B., Langdale, J.A. & Gull, K. Reconstructing the evolutionary history of the centriole from protein components. *Journal of cell science* **123**, 1407-1413 (2010).
9. Carvalho-Santos, Z. *et al.* Stepwise evolution of the centriole-assembly pathway. *Journal of cell science* **123**, 1414-1426 (2010).
10. Carvalho-Santos, Z., Azimzadeh, J., Pereira-Leal, J.B. & Bettencourt-Dias, M. Evolution: Tracing the origins of centrioles, cilia, and flagella. *The Journal of cell biology* **194**, 165-175 (2011).
11. Hayward, D., Metz, J., Pellacani, C. & Wakefield, J.G. Synergy between multiple microtubule-generating pathways confers robustness to centrosome-driven mitotic spindle formation. *Developmental cell* **28**, 81-93 (2014).
12. Heald, R. *et al.* Self-organization of microtubules into bipolar spindles around artificial chromosomes in Xenopus egg extracts. *Nature* **382**, 420-425 (1996).
13. Khodjakov, A., Cole, R.W., Oakley, B.R. & Rieder, C.L. Centrosome-independent mitotic spindle formation in vertebrates. *Current biology : CB* **10**, 59-67 (2000).
14. Basto, R. *et al.* Flies without centrioles. *Cell* **125**, 1375-1386 (2006).
15. Manandhar, G., Schatten, H. & Sutovsky, P. Centrosome reduction during gametogenesis and its significance. *Biol Reprod* **72**, 2-13 (2005).
16. Heald, R., Tournebize, R., Habermann, A., Karsenti, E. & Hyman, A. Spindle assembly in Xenopus egg extracts: respective roles of centrosomes and microtubule self-organization. *The Journal of cell biology* **138**, 615-628 (1997).
17. Sir, J.H. *et al.* Loss of centrioles causes chromosomal instability in vertebrate somatic cells. *The Journal of cell biology* **203**, 747-756 (2013).
18. Bazzi, H. & Anderson, K.V. Acentriolar mitosis activates a p53-dependent apoptosis pathway in the mouse embryo. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E1491-1500 (2014).
19. de Saint Phalle, B. & Sullivan, W. Spindle assembly and mitosis without centrosomes in parthenogenetic Sciara embryos. *The Journal of cell biology* **141**, 1383-1391 (1998).

20. Megraw, T.L., Li, K., Kao, L.R. & Kaufman, T.C. The centrosomin protein is required for centrosome assembly and function during cleavage in Drosophila. *Development* **126**, 2829-2839 (1999).
21. Vaizel-Ohayon, D. & Schejter, E.D. Mutations in centrosomin reveal requirements for centrosomal function during early Drosophila embryogenesis. *Current biology : CB* **9**, 889-898 (1999).
22. Yamashita, Y.M., Mahowald, A.P., Perlin, J.R. & Fuller, M.T. Asymmetric inheritance of mother versus daughter centrosome in stem cell division. *Science* **315**, 518-521 (2007).
23. Cheng, J. *et al.* Centrosome misorientation reduces stem cell division during ageing. *Nature* **456**, 599-604 (2008).
24. Conduit, P.T. & Raff, J.W. Cnn dynamics drive centrosome size asymmetry to ensure daughter centriole retention in Drosophila neuroblasts. *Current biology : CB* **20**, 2187-2192 (2010).
25. Rebollo, E. *et al.* Functionally unequal centrosomes drive spindle orientation in asymmetrically dividing Drosophila neural stem cells. *Developmental cell* **12**, 467-474 (2007).
26. Rusan, N.M. & Peifer, M. A role for a novel centrosome cycle in asymmetric cell division. *The Journal of cell biology* **177**, 13-20 (2007).
27. Wang, X. *et al.* Asymmetric centrosome inheritance maintains neural progenitors in the neocortex. *Nature* **461**, 947-955 (2009).
28. Shao, W. *et al.* Centrosome anchoring regulates progenitor properties and cortical formation. *Nature* **580**, 106-112 (2020).
29. Hinchcliffe, E.H., Miller, F.J., Cham, M., Khodjakov, A. & Sluder, G. Requirement of a centrosomal activity for cell cycle progression through G1 into S phase. *Science* **291**, 1547-1550 (2001).
30. Piel, M., Nordberg, J., Euteneuer, U. & Bornens, M. Centrosome-dependent exit of cytokinesis in animal cells. *Science* **291**, 1550-1553 (2001).
31. Wong, Y.L. *et al.* Cell biology. Reversible centriole depletion with an inhibitor of Polo-like kinase 4. *Science* **348**, 1155-1160 (2015).
32. Gould, R.R. & Borisy, G.G. The pericentriolar material in Chinese hamster ovary cells nucleates microtubule formation. *The Journal of cell biology* **73**, 601-615 (1977).
33. Woodruff, J.B. *et al.* The Centrosome Is a Selective Condensate that Nucleates Microtubules by Concentrating Tubulin. *Cell* **169**, 1066-1077 (2017).
34. Baumgart, J. *et al.* Soluble tubulin is significantly enriched at mitotic centrosomes. *Journal of Cell Biology* **218**, 3977-3985 (2019).
35. Moritz, M., Braunfeld, M.B., Sedat, J.W., Alberts, B. & Agard, D.A. Microtubule nucleation by gamma-tubulin-containing rings in the centrosome. *Nature* **378**, 638-640 (1995).
36. Moritz, M., Braunfeld, M.B., Guenebaut, V., Heuser, J. & Agard, D.A. Structure of the gamma-tubulin ring complex: a template for microtubule nucleation. *Nature cell biology* **2**, 365-370 (2000).
37. Zheng, Y., Jung, M.K. & Oakley, B.R. Gamma-tubulin is present in Drosophila melanogaster and Homo sapiens and is associated with the centrosome. *Cell* **65**, 817-823 (1991).
38. Sunkel, C.E., Gomes, R., Sampaio, P., Perdigao, J. & Gonzalez, C. Gamma-tubulin is required for the structure and function of the microtubule organizing centre in Drosophila neuroblasts. *The EMBO journal* **14**, 28-36 (1995).

39. Hannak, E. *et al.* The kinetically dominant assembly pathway for centrosomal asters in Caenorhabditis elegans is gamma-tubulin dependent. *The Journal of cell biology* **157**, 591-602 (2002).
40. Zheng, Y., Wong, M.L., Alberts, B. & Mitchison, T. Nucleation of microtubule assembly by a gamma-tubulin-containing ring complex. *Nature* **378**, 578-583 (1995).
41. Aldaz, H., Rice, L.M., Stearns, T. & Agard, D.A. Insights into microtubule nucleation from the crystal structure of human gamma-tubulin. *Nature* **435**, 523-527 (2005).
42. Kollman, J.M., Polka, J.K., Zelter, A., Davis, T.N. & Agard, D.A. Microtubule nucleating gamma-TuSC assembles structures with 13-fold microtubule-like symmetry. *Nature* **466**, 879-882 (2010).
43. Stearns, T. & Kirschner, M. In vitro reconstitution of centrosome assembly and function: the central role of gamma-tubulin. *Cell* **76**, 623-637 (1994).
44. Oakley, B.R., Oakley, C.E., Yoon, Y. & Jung, M.K. Gamma-tubulin is a component of the spindle pole body that is essential for microtubule function in Aspergillus nidulans. *Cell* **61**, 1289-1301 (1990).
45. Stearns, T., Evans, L. & Kirschner, M. Gamma-tubulin is a highly conserved component of the centrosome. *Cell* **65**, 825-836 (1991).
46. Liu, P. *et al.* Insights into the assembly and activation of the microtubule nucleator gamma-TuRC. *Nature* **578**, 467-471 (2020).
47. Vertii, A., Hehnly, H. & Doxsey, S. The Centrosome, a Multitalented Renaissance Organelle. *Cold Spring Harbor perspectives in biology* **8**, a025049 (2016).
48. Farina, F. *et al.* The centrosome is an actin-organizing centre. *Nature cell biology* **18**, 65-75 (2016).
49. Stevenson, V.A., Kramer, J., Kuhn, J. & Theurkauf, W.E. Centrosomes and the Scrambled protein coordinate microtubule-independent actin reorganization. *Nature cell biology* **3**, 68-75 (2001).
50. Arquint, C., Gabryjonczyk, A.M. & Nigg, E.A. Centrosomes as signalling centres. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**, 20130464 (2014).
51. Firat-Karalar, E.N. & Stearns, T. The centriole duplication cycle. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**, 20130460 (2014).
52. Tsou, M.F. & Stearns, T. Mechanism limiting centrosome duplication to once per cell cycle. *Nature* **442**, 947-951 (2006).
53. Wong, C. & Stearns, T. Centrosome number is controlled by a centrosome-intrinsic block to reduplication. *Nature cell biology* **5**, 539-544 (2003).
54. Tsou, M.F. *et al.* Polo kinase and separase regulate the mitotic licensing of centriole duplication in human cells. *Developmental cell* **17**, 344-354 (2009).
55. Conduit, P.T., Wainman, A. & Raff, J.W. Centrosome function and assembly in animal cells. *Nature reviews. Molecular cell biology* **16**, 611-624 (2015).
56. Agircan, F.G., Schiebel, E. & Mardin, B.R. Separate to operate: control of centrosome positioning and separation. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**, 20130461 (2014).
57. Prosser, S.L. & Pelletier, L. Mitotic spindle assembly in animal cells: a fine balancing act. *Nature reviews. Molecular cell biology* **18**, 187-201 (2017).
58. Pimenta-Marques, A. *et al.* A mechanism for the elimination of the female gamete centrosome in Drosophila melanogaster. *Science* **353**, aaf4866 (2016).

59. Lucas, E.P. & Raff, J.W. Maintaining the proper connection between the centrioles and the pericentriolar matrix requires Drosophila centrosomin. *The Journal of cell biology* **178**, 725-732 (2007).
60. Mahoney, N.M., Goshima, G., Douglass, A.D. & Vale, R.D. Making microtubules and mitotic spindles in cells without functional centrosomes. *Current biology : CB* **16**, 564-569 (2006).
61. Megraw, T.L., Kao, L.R. & Kaufman, T.C. Zygotic development without functional mitotic centrosomes. *Current biology : CB* **11**, 116-120 (2001).
62. Conduit, P.T. *et al.* Centrioles regulate centrosome size by controlling the rate of Cnn incorporation into the PCM. *Current biology : CB* **20**, 2178-2186 (2010).
63. Novak, Z.A., Wainman, A., Gartenmann, L. & Raff, J.W. Cdk1 Phosphorylates Drosophila Sas-4 to Recruit Polo to Daughter Centrioles and Convert Them to Centrosomes. *Developmental cell* **37**, 545-557 (2016).
64. Fu, J. *et al.* Conserved molecular interactions in centriole-to-centrosome conversion. *Nature cell biology* **18**, 87-99 (2016).
65. Wang, W.J., Soni, R.K., Uryu, K. & Tsou, M.F. The conversion of centrioles to centrosomes: essential coupling of duplication with segregation. *The Journal of cell biology* **193**, 727-739 (2011).
66. Sanchez, I. & Dynlacht, B.D. Cilium assembly and disassembly. *Nature cell biology* **18**, 711-717 (2016).
67. Breslow, D.K. & Holland, A.J. Mechanism and Regulation of Centriole and Cilium Biogenesis. *Annual review of biochemistry* **88**, 691-724 (2019).
68. Nachury, M.V. How do cilia organize signalling cascades? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**, 20130465 (2014).
69. Meunier, A. & Azimzadeh, J. Multiciliated Cells in Animals. *Cold Spring Harbor perspectives in biology* **8**, a028233 (2016).
70. Escalier, D. Knockout mouse models of sperm flagellum anomalies. *Hum Reprod Update* **12**, 449-461 (2006).
71. Rodrigues-Martins, A., Riparbelli, M., Callaini, G., Glover, D.M. & Bettencourt-Dias, M. Revisiting the role of the mother centriole in centriole biogenesis. *Science* **316**, 1046-1050 (2007).
72. Azimzadeh, J., Wong, M.L., Downhour, D.M., Sanchez Alvarado, A. & Marshall, W.F. Centrosome loss in the evolution of planarians. *Science* **335**, 461-463 (2012).
73. Zenker, J. *et al.* A microtubule-organizing center directing intracellular transport in the early mouse embryo. *Science* **357**, 925-928 (2017).
74. Courtois, A., Schuh, M., Ellenberg, J. & Hiiragi, T. The transition from meiotic to mitotic spindle assembly is gradual during early mammalian development. *The Journal of cell biology* **198**, 357-370 (2012).
75. La Terra, S. *et al.* The de novo centriole assembly pathway in HeLa cells: cell cycle progression and centriole assembly/maturation. *The Journal of cell biology* **168**, 713-722 (2005).
76. Wang, W.J. *et al.* De novo centriole formation in human cells is error-prone and does not require SAS-6 self-assembly. *eLife* **4**, e10586 (2015).
77. Mercey, O. *et al.* Massive centriole production can occur in the absence of deuterosomes in multiciliated cells. *Nature cell biology* **21**, 1544-1552 (2019).
78. Boveri, T. *Zur Frage der Entstehung maligner Tumoren*. (G. Fischer, Jena; 1914).
79. Chavali, P.L., Putz, M. & Gergely, F. Small organelle, big responsibility: the role of centrosomes in development and disease. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**, 20130468 (2014).

80. Godinho, S.A. & Pellman, D. Causes and consequences of centrosome abnormalities in cancer. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**, 20130467 (2014).
81. Gonczy, P. Centrosomes and cancer: revisiting a long-standing relationship. *Nature reviews. Cancer* **15**, 639-652 (2015).
82. Basto, R. *et al.* Centrosome amplification can initiate tumorigenesis in flies. *Cell* **133**, 1032-1042 (2008).
83. Levine, M.S. *et al.* Centrosome Amplification Is Sufficient to Promote Spontaneous Tumorigenesis in Mammals. *Developmental cell* **40**, 313-322 (2017).
84. Parkin, D.M. The global health burden of infection-associated cancers in the year 2002. *Int J Cancer* **118**, 3030-3044 (2006).
85. Shumilov, A. *et al.* Epstein-Barr virus particles induce centrosome amplification and chromosomal instability. *Nature communications* **8**, 14257 (2017).
86. Godinho, S.A. *et al.* Oncogene-like induction of cellular invasion from centrosome amplification. *Nature* **510**, 167-171 (2014).
87. Marthiens, V. *et al.* Centrosome amplification causes microcephaly. *Nature cell biology* **15**, 731-740 (2013).
88. Bond, J. *et al.* A centrosomal mechanism involving CDK5RAP2 and CENPJ controls brain size. *Nature genetics* **37**, 353-355 (2005).
89. Lancaster, M.A. *et al.* Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373-379 (2013).
90. Pagnamenta, A.T. *et al.* A novel nonsense CDK5RAP2 mutation in a Somali child with primary microcephaly and sensorineural hearing loss. *Am J Med Genet A* **158A**, 2577-2582 (2012).
91. Hassan, M.J. *et al.* Previously described sequence variant in CDK5RAP2 gene in a Pakistani family with autosomal recessive primary microcephaly. *BMC Med Genet* **8**, 58 (2007).
92. Yigit, G. *et al.* Mutations in CDK5RAP2 cause Seckel syndrome. *Mol Genet Genomic Med* **3**, 467-480 (2015).
93. Nasser, H. *et al.* CDK5RAP2 primary microcephaly is associated with hypothalamic, retinal and cochlear developmental defects. *Journal of medical genetics* **57**, 389–399 (2020).
94. Chavali, P.L. *et al.* A CEP215-HSET complex links centrosomes with spindle poles and drives centrosome clustering in cancer. *Nature communications* **7**, 11005 (2016).
95. Lee, S. & Rhee, K. CEP215 is involved in the dynein-dependent accumulation of pericentriolar matrix proteins for spindle pole formation. *Cell Cycle* **9**, 774-783 (2010).
96. Barr, A.R., Kilmartin, J.V. & Gergely, F. CDK5RAP2 functions in centrosome to spindle pole attachment and DNA damage response. *The Journal of cell biology* **189**, 23-39 (2010).
97. Buchman, J.J. *et al.* Cdk5rap2 interacts with pericentrin to maintain the neural progenitor pool in the developing neocortex. *Neuron* **66**, 386-402 (2010).
98. Barrera, J.A. *et al.* CDK5RAP2 regulates centriole engagement and cohesion in mice. *Developmental cell* **18**, 913-926 (2010).
99. Lizarraga, S.B. *et al.* Cdk5rap2 regulates centrosome function and chromosome segregation in neuronal progenitors. *Development* **137**, 1907-1917 (2010).
100. Gul, A. *et al.* A novel deletion mutation in CENPJ gene in a Pakistani family with autosomal recessive primary microcephaly. *Journal of human genetics* **51**, 760-764 (2006).

101. Griffith, E. *et al.* Mutations in pericentrin cause Seckel syndrome with defective ATR-dependent DNA damage signaling. *Nature genetics* **40**, 232-236 (2008).
102. Hussain, M.S. *et al.* A truncating mutation of CEP135 causes primary microcephaly and disturbed centrosomal function. *American journal of human genetics* **90**, 871-878 (2012).
103. Farooq, M. *et al.* A novel splice site mutation in CEP135 is associated with primary microcephaly in a Pakistani family. *Journal of human genetics* **61**, 271-273 (2016).
104. Rauch, A. *et al.* Mutations in the pericentrin (PCNT) gene cause primordial dwarfism. *Science* **319**, 816-819 (2008).
105. Kalay, E. *et al.* CEP152 is a genome maintenance protein disrupted in Seckel syndrome. *Nature genetics* **43**, 23-26 (2011).
106. Al-Dosari, M.S., Shaheen, R., Colak, D. & Alkuraya, F.S. Novel CENPJ mutation causes Seckel syndrome. *Journal of medical genetics* **47**, 411-414 (2010).
107. Khan, M.A. *et al.* A missense mutation in the PISA domain of HsSAS-6 causes autosomal recessive primary microcephaly in a large consanguineous Pakistani family. *Human molecular genetics* **23**, 5940-5949 (2014).
108. Ganem, N.J., Godinho, S.A. & Pellman, D. A mechanism linking extra centrosomes to chromosomal instability. *Nature* **460**, 278-282 (2009).
109. Winey, M. & O'Toole, E. Centriole structure. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**, 20130457 (2014).
110. Gonczy, P. Towards a molecular architecture of centriole assembly. *Nature reviews. Molecular cell biology* **13**, 425-435 (2012).
111. Hirono, M. Cartwheel assembly. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**, 20130458 (2014).
112. van Breugel, M. *et al.* Structures of SAS-6 suggest its organization in centrioles. *Science* **331**, 1196-1199 (2011).
113. Kitagawa, D. *et al.* Structural basis of the 9-fold symmetry of centrioles. *Cell* **144**, 364-375 (2011).
114. Cottee, M.A. *et al.* The homo-oligomerisation of both Sas-6 and Ana2 is required for efficient centriole assembly in flies. *eLife* **4**, e07236 (2015).
115. Hilbert, M. *et al.* Caenorhabditis elegans centriolar protein SAS-6 forms a spiral that is consistent with imparting a ninefold symmetry. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 11373-11378 (2013).
116. Guichard, P. & Gonczy, P. Basal body structure in Trichonympha. *Cilia* **5**, 9 (2016).
117. Guichard, P. *et al.* Native architecture of the centriole proximal region reveals features underlying its 9-fold radial symmetry. *Current biology : CB* **23**, 1620-1628 (2013).
118. Guichard, P. *et al.* Cartwheel architecture of Trichonympha basal body. *Science* **337**, 553 (2012).
119. Li, S., Fernandez, J.J., Marshall, W.F. & Agard, D.A. Electron cryo-tomography provides insight into procentriole architecture and assembly mechanism. *eLife* **8**, e43434 (2019).
120. Boxem, M. *et al.* A protein domain-based interactome network for C. elegans early embryogenesis. *Cell* **134**, 534-545 (2008).
121. Pelletier, L., O'Toole, E., Schwager, A., Hyman, A.A. & Muller-Reichert, T. Centriole assembly in Caenorhabditis elegans. *Nature* **444**, 619-623 (2006).
122. Delattre, M., Canard, C. & Gonczy, P. Sequential protein recruitment in C. elegans centriole formation. *Current biology : CB* **16**, 1844-1849 (2006).

123. O'Connell, K.F. *et al.* The C. elegans zyg-1 gene encodes a regulator of centrosome duplication with distinct maternal and paternal roles in the embryo. *Cell* **105**, 547-558 (2001).
124. Pelletier, L. *et al.* The Caenorhabditis elegans centrosomal protein SPD-2 is required for both pericentriolar material recruitment and centriole duplication. *Current biology : CB* **14**, 863-873 (2004).
125. Kemp, C.A., Kopish, K.R., Zipperlen, P., Ahringer, J. & O'Connell, K.F. Centrosome maturation and duplication in C. elegans require the coiled-coil protein SPD-2. *Developmental cell* **6**, 511-523 (2004).
126. Lettman, M.M. *et al.* Direct binding of SAS-6 to ZYG-1 recruits SAS-6 to the mother centriole for cartwheel assembly. *Developmental cell* **25**, 284-298 (2013).
127. Qiao, R., Cabral, G., Lettman, M.M., Dammermann, A. & Dong, G. SAS-6 coiled-coil structure and interaction with SAS-5 suggest a regulatory mechanism in C. elegans centriole assembly. *The EMBO journal* **31**, 4334-4347 (2012).
128. Delattre, M. *et al.* Centriolar SAS-5 is required for centrosome duplication in C. elegans. *Nature cell biology* **6**, 656-664 (2004).
129. Leidel, S., Delattre, M., Cerutti, L., Baumer, K. & Gonczy, P. SAS-6 defines a protein family required for centrosome duplication in C. elegans and in human cells. *Nature cell biology* **7**, 115-125 (2005).
130. Dammermann, A. *et al.* Centriole assembly requires both centriolar and pericentriolar material proteins. *Developmental cell* **7**, 815-829 (2004).
131. Leidel, S. & Gonczy, P. SAS-4 is essential for centrosome duplication in C elegans and is recruited to daughter centrioles once per cell cycle. *Developmental cell* **4**, 431-439 (2003).
132. Kirkham, M., Muller-Reichert, T., Oegema, K., Grill, S. & Hyman, A.A. SAS-4 is a C. elegans centriolar protein that controls centrosome size. *Cell* **112**, 575-587 (2003).
133. Bettencourt-Dias, M. *et al.* SAK/PLK4 is required for centriole duplication and flagella development. *Current biology : CB* **15**, 2199-2207 (2005).
134. Aydogan, M.G. *et al.* A homeostatic clock sets daughter centriole size in flies. *The Journal of cell biology* **217**, 1233-1248 (2018).
135. Habedanck, R., Stierhof, Y.D., Wilkinson, C.J. & Nigg, E.A. The Polo kinase Plk4 functions in centriole duplication. *Nature cell biology* **7**, 1140-1146 (2005).
136. Kleylein-Sohn, J. *et al.* Plk4-induced centriole biogenesis in human cells. *Developmental cell* **13**, 190-202 (2007).
137. Varmark, H. *et al.* Asterless is a centriolar protein required for centrosome function and embryo development in Drosophila. *Current biology : CB* **17**, 1735-1745 (2007).
138. Bonaccorsi, S., Giansanti, M.G. & Gatti, M. Spindle self-organization and cytokinesis during male meiosis in asterless mutants of Drosophila melanogaster. *The Journal of cell biology* **142**, 751-761 (1998).
139. Dzhindzhev, N.S. *et al.* Asterless is a scaffold for the onset of centriole assembly. *Nature* **467**, 714-718 (2010).
140. Sonnen, K.F., Gabryjonczyk, A.M., Anselm, E., Stierhof, Y.D. & Nigg, E.A. Human Cep192 and Cep152 cooperate in Plk4 recruitment and centriole duplication. *Journal of cell science* **126**, 3223-3233 (2013).
141. Kim, T.S. *et al.* Hierarchical recruitment of Plk4 and regulation of centriole biogenesis by two centrosomal scaffolds, Cep192 and Cep152. *Proceedings of the National Academy of Sciences of the United States of America* **110**, E4849-4857 (2013).

142. Hatch, E.M., Kulukian, A., Holland, A.J., Cleveland, D.W. & Stearns, T. Cep152 interacts with Plk4 and is required for centriole duplication. *The Journal of cell biology* **191**, 721-729 (2010).
143. Cizmecioglu, O. *et al.* Cep152 acts as a scaffold for recruitment of Plk4 and CPAP to the centrosome. *The Journal of cell biology* **191**, 731-739 (2010).
144. Stevens, N.R., Dobbelaere, J., Brunk, K., Franz, A. & Raff, J.W. Drosophila Ana2 is a conserved centriole duplication factor. *The Journal of cell biology* **188**, 313-323 (2010).
145. Arquint, C. *et al.* STIL binding to Polo-box 3 of PLK4 regulates centriole duplication. *eLife* **4**, e07888 (2015).
146. Moyer, T.C., Clutario, K.M., Lambrus, B.G., Daggubati, V. & Holland, A.J. Binding of STIL to Plk4 activates kinase activity to promote centriole assembly. *The Journal of cell biology* **209**, 863-878 (2015).
147. Kratz, A.S., Barenz, F., Richter, K.T. & Hoffmann, I. Plk4-dependent phosphorylation of STIL is required for centriole duplication. *Biology open* **4**, 370-377 (2015).
148. Ohta, M. *et al.* Direct interaction of Plk4 with STIL ensures formation of a single procentriole per parental centriole. *Nature communications* **5**, 5267 (2014).
149. Dzhindzhev, N.S. *et al.* Plk4 phosphorylates Ana2 to trigger Sas6 recruitment and procentriole formation. *Current biology : CB* **24**, 2526-2532 (2014).
150. Vulprecht, J. *et al.* STIL is required for centriole duplication in human cells. *Journal of cell science* **125**, 1353-1362 (2012).
151. Cottee, M.A. *et al.* Crystal structures of the CPAP/STIL complex reveal its role in centriole assembly and human microcephaly. *eLife* **2**, e01071 (2013).
152. Novak, Z.A., Conduit, P.T., Wainman, A. & Raff, J.W. Asterless licenses daughter centrioles to duplicate for the first time in Drosophila embryos. *Current biology : CB* **24**, 1276-1282 (2014).
153. Robbins, E., Jentzsch, G. & Micali, A. The centriole cycle in synchronized HeLa cells. *The Journal of cell biology* **36**, 329-339 (1968).
154. Sonnen, K.F., Schermelleh, L., Leonhardt, H. & Nigg, E.A. 3D-structured illumination microscopy provides novel insight into architecture of human centrosomes. *Biology open* **1**, 965-976 (2012).
155. Mennella, V. *et al.* Subdiffraction-resolution fluorescence microscopy reveals a domain of the centrosome critical for pericentriolar material organization. *Nature cell biology* **14**, 1159-1168 (2012).
156. Lawo, S., Hasegan, M., Gupta, G.D. & Pelletier, L. Subdiffraction imaging of centrosomes reveals higher-order organizational features of pericentriolar material. *Nature cell biology* **14**, 1148-1158 (2012).
157. Fu, J. & Glover, D.M. Structured illumination of the interface between centriole and peri-centriolar material. *Open biology* **2**, 120104 (2012).
158. Richens, J.H. *et al.* The Drosophila Pericentrin-like-protein (PLP) cooperates with Cnn to maintain the integrity of the outer PCM. *Biology open* **4**, 1052-1061 (2015).
159. Lerit, D.A. *et al.* Interphase centrosome organization by the PLP-Cnn scaffold is required for centrosome function. *The Journal of cell biology* **210**, 79-97 (2015).
160. Rogers, G.C., Rusan, N.M., Peifer, M. & Rogers, S.L. A multicomponent assembly pathway contributes to the formation of acentrosomal microtubule arrays in interphase Drosophila cells. *Molecular biology of the cell* **19**, 3163-3178 (2008).
161. Palazzo, R.E., Vogel, J.M., Schnackenberg, B.J., Hull, D.R. & Wu, X. Centrosome maturation. *Curr Top Dev Biol* **49**, 449-470 (2000).

162. Andersen, J.S. *et al.* Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* **426**, 570-574 (2003).
163. Alves-Cruzeiro, J.M., Nogales-Cadenas, R. & Pascual-Montano, A.D. CentrosomeDB: a new generation of the centrosomal proteins database for Human and Drosophila melanogaster. *Nucleic acids research* **42**, D430-436 (2014).
164. Thul, P.J. *et al.* A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017).
165. Muller, H. *et al.* Proteomic and functional analysis of the mitotic Drosophila centrosome. *The EMBO journal* **29**, 3344-3357 (2010).
166. Goshima, G. *et al.* Genes required for mitotic spindle assembly in Drosophila S2 cells. *Science* **316**, 417-421 (2007).
167. Dobbelaere, J. *et al.* A genome-wide RNAi screen to dissect centriole duplication and centrosome maturation in Drosophila. *PLoS biology* **6**, e224 (2008).
168. Conduit, P.T. *et al.* A molecular mechanism of mitotic centrosome assembly in Drosophila. *eLife* **3**, e03399 (2014).
169. Dix, C.I. & Raff, J.W. Drosophila Spd-2 recruits PCM to the sperm centriole, but is dispensable for centriole duplication. *Current biology : CB* **17**, 1759-1764 (2007).
170. Giansanti, M.G., Bucciarelli, E., Bonaccorsi, S. & Gatti, M. Drosophila SPD-2 is an essential centriole component required for PCM recruitment and astral-microtubule nucleation. *Current biology : CB* **18**, 303-309 (2008).
171. Haren, L., Stearns, T. & Luders, J. Plk1-dependent recruitment of gamma-tubulin complexes to mitotic centrosomes involves multiple PCM components. *PloS one* **4**, e5976 (2009).
172. Gomez-Ferreria, M.A. *et al.* Human Cep192 is required for mitotic centrosome and spindle assembly. *Current biology : CB* **17**, 1960-1966 (2007).
173. Zhu, F. *et al.* The mammalian SPD-2 ortholog Cep192 regulates centrosome biogenesis. *Current biology : CB* **18**, 136-141 (2008).
174. Joukov, V., Walter, J.C. & De Nicolo, A. The Cep192-organized aurora A-Plk1 cascade is essential for centrosome cycle and bipolar spindle assembly. *Molecular cell* **55**, 578-591 (2014).
175. Meng, L. *et al.* Bimodal Interaction of Mammalian Polo-Like Kinase 1 and a Centrosomal Scaffold, Cep192, in the Regulation of Bipolar Spindle Formation. *Molecular and cellular biology* **35**, 2626-2640 (2015).
176. Kim, S. & Rhee, K. Importance of the CEP215-pericentrin interaction for centrosome maturation during mitosis. *PloS one* **9**, e87016 (2014).
177. Doxsey, S.J., Stein, P., Evans, L., Calarco, P.D. & Kirschner, M. Pericentrin, a highly conserved centrosome protein involved in microtubule organization. *Cell* **76**, 639-650 (1994).
178. Wang, Z. *et al.* Conserved motif of CDK5RAP2 mediates its localization to centrosomes and the Golgi complex. *The Journal of biological chemistry* **285**, 22658-22665 (2010).
179. Zhang, J. & Megraw, T.L. Proper recruitment of gamma-tubulin and D-TACC/Msps to embryonic Drosophila centrosomes requires Centrosomin Motif 1. *Molecular biology of the cell* **18**, 4037-4049 (2007).
180. Kuhn, M., Hyman, A.A. & Beyer, A. Coiled-coil proteins facilitated the functional expansion of the centrosome. *PLoS computational biology* **10**, e1003657 (2014).
181. Heuer, J.G., Li, K. & Kaufman, T.C. The Drosophila homeotic target gene centrosomin (cnn) encodes a novel centrosomal protein with leucine zippers and maps to a genomic region required for midgut morphogenesis. *Development* **121**, 3861-3876 (1995).

182. Lupas, A., Van Dyke, M. & Stock, J. Predicting coiled coils from protein sequences. *Science* **252**, 1162-1164 (1991).
183. Lupas, A.N. & Bassler, J. Coiled Coils - A Model System for the 21st Century. *Trends in biochemical sciences* **42**, 130-140 (2017).
184. Crick, F.H. Is alpha-keratin a coiled coil? *Nature* **170**, 882-883 (1952).
185. Crick, F.H.C. The Packing of Alpha-Helices - Simple Coiled-Coils. *Acta Crystallogr* **6**, 689-697 (1953).
186. Pauling, L. & Corey, R.B. Compound helical configurations of polypeptide chains: structure of proteins of the alpha-keratin type. *Nature* **171**, 59-61 (1953).
187. Landschulz, W.H., Johnson, P.F. & McKnight, S.L. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **240**, 1759-1764 (1988).
188. O'Shea, E.K., Klemm, J.D., Kim, P.S. & Alber, T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* **254**, 539-544 (1991).
189. O'Shea, E.K., Rutkowski, R. & Kim, P.S. Evidence that the leucine zipper is a coiled coil. *Science* **243**, 538-542 (1989).
190. Slabinski, L. *et al.* XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics* **23**, 3403-3405 (2007).
191. Jones, D.T. Protein secondary structure prediction based on position-specific scoring matrices. *Journal of molecular biology* **292**, 195-202 (1999).
192. Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F. & Jones, D.T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *Journal of molecular biology* **337**, 635-645 (2004).
193. Schiffer, M. & Edmundson, A.B. Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. *Biophys J* **7**, 121-135 (1967).
194. Foe, V.E. & Alberts, B.M. Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis. *Journal of cell science* **61**, 31-70 (1983).
195. Alvarez-Rodrigo, I. *et al.* Evidence that a positive feedback loop drives centrosome maturation in fly embryos. *eLife* **8**, e50130 (2019).
196. Conduit, P.T. *et al.* The centrosome-specific phosphorylation of Cnn by Polo/Plk1 drives Cnn scaffold assembly and centrosome maturation. *Developmental cell* **28**, 659-669 (2014).
197. Galletta, B.J. *et al.* A centrosome interactome provides insight into organelle assembly and reveals a non-duplication role for Plk4. *Nature communications* **7**, 12476 (2016).
198. Sunkel, C.E. & Glover, D.M. polo, a mitotic mutant of Drosophila displaying abnormal spindle poles. *Journal of cell science* **89 (Pt 1)**, 25-38 (1988).
199. Lane, H.A. & Nigg, E.A. Antibody microinjection reveals an essential role for human polo-like kinase 1 (Plk1) in the functional maturation of mitotic centrosomes. *The Journal of cell biology* **135**, 1701-1713 (1996).
200. Lee, K. & Rhee, K. PLK1 phosphorylation of pericentrin initiates centrosome maturation at the onset of mitosis. *The Journal of cell biology* **195**, 1093-1101 (2011).
201. Kim, J., Lee, K. & Rhee, K. PLK1 regulation of PCNT cleavage ensures fidelity of centriole separation during mitotic exit. *Nature communications* **6**, 10076 (2015).
202. Liu, X. & Erikson, R.L. Activation of Cdc2/cyclin B and inhibition of centrosome amplification in cells depleted of Plk1 by siRNA. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 8672-8676 (2002).

203. Loncarek, J., Hergert, P. & Khodjakov, A. Centriole reduplication during prolonged interphase requires procentriole maturation governed by Plk1. *Current biology : CB* **20**, 1277-1282 (2010).
204. Hatano, T. & Sluder, G. The interrelationship between APC/C and Plk1 activities in centriole disengagement. *Biology open* **1**, 1153-1160 (2012).
205. Barr, F.A., Sillje, H.H. & Nigg, E.A. Polo-like kinases and the orchestration of cell division. *Nature reviews. Molecular cell biology* **5**, 429-440 (2004).
206. Seki, A., Coppinger, J.A., Jang, C.Y., Yates, J.R. & Fang, G. Bora and the kinase Aurora a cooperatively activate the kinase Plk1 and control mitotic entry. *Science* **320**, 1655-1658 (2008).
207. Macurek, L. *et al.* Polo-like kinase-1 is activated by aurora A to promote checkpoint recovery. *Nature* **455**, 119-123 (2008).
208. Cabral, G., Laos, T., Dumont, J. & Dammermann, A. Differential Requirements for Centrioles in Mitotic Centrosome Growth and Maintenance. *Developmental cell* **50**, 355-366 (2019).
209. Elia, A.E. *et al.* The molecular basis for phosphodependent substrate targeting and regulation of Plks by the Polo-box domain. *Cell* **115**, 83-95 (2003).
210. Elia, A.E., Cantley, L.C. & Yaffe, M.B. Proteomic screen finds pSer/pThr-binding domain localizing Plk1 to mitotic substrates. *Science* **299**, 1228-1231 (2003).
211. Santamaria, A. *et al.* The Plk1-dependent phosphoproteome of the early mitotic spindle. *Mol Cell Proteomics* **10**, M110 004457 (2011).
212. Nakajima, H., Toyoshima-Morimoto, F., Taniguchi, E. & Nishida, E. Identification of a consensus motif for Plk (Polo-like kinase) phosphorylation reveals Myt1 as a Plk1 substrate. *The Journal of biological chemistry* **278**, 25277-25280 (2003).
213. Woodruff, J.B., Wueseke, O. & Hyman, A.A. Pericentriolar material structure and dynamics. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**, 20130459 (2014).
214. Megraw, T.L., Kilaru, S., Turner, F.R. & Kaufman, T.C. The centrosome is a dynamic structure that ejects PCM flares. *Journal of cell science* **115**, 4707-4718 (2002).
215. Conduit, P.T. & Raff, J.W. Different Drosophila cell types exhibit differences in mitotic centrosome assembly dynamics. *Current biology : CB* **25**, R650-651 (2015).
216. Martinez-Campos, M., Basto, R., Baker, J., Kernan, M. & Raff, J.W. The Drosophila pericentrin-like protein is essential for cilia/flagella function, but appears to be dispensable for mitosis. *The Journal of cell biology* **165**, 673-683 (2004).
217. Kao, L.R. & Megraw, T.L. Centrocortin cooperates with centrosomin to organize Drosophila embryonic cleavage furrows. *Current biology : CB* **19**, 937-942 (2009).
218. Gillingham, A.K. & Munro, S. The PACT domain, a conserved centrosomal targeting motif in the coiled-coil proteins AKAP450 and pericentrin. *EMBO reports* **1**, 524-529 (2000).
219. Lin, T.C. *et al.* MOZART1 and gamma-tubulin complex receptors are both required to turn gamma-TuSC into an active microtubule nucleation template. *The Journal of cell biology* **215**, 823-840 (2016).
220. Sharp, D.J., Yu, K.R., Sisson, J.C., Sullivan, W. & Scholey, J.M. Antagonistic microtubule-sliding motors position mitotic centrosomes in Drosophila early embryos. *Nature cell biology* **1**, 51-54 (1999).
221. Tovey, C.A. *et al.* gamma-TuRC Heterogeneity Revealed by Analysis of Mozart1. *Current biology : CB* **28**, 2314-2323 e2316 (2018).

222. Terada, Y., Uetake, Y. & Kuriyama, R. Interaction of Aurora-A and centrosomin at the microtubule-nucleating site in Drosophila and mammalian cells. *The Journal of cell biology* **162**, 757-763 (2003).
223. Chen, J.V., Buchwalter, R.A., Kao, L.R. & Megraw, T.L. A Splice Variant of Centrosomin Converts Mitochondria to Microtubule-Organizing Centers. *Current biology : CB* **27**, 1928-1940 e1926 (2017).
224. Fong, K.W., Choi, Y.K., Rattner, J.B. & Qi, R.Z. CDK5RAP2 is a pericentriolar protein that functions in centrosomal attachment of the gamma-tubulin ring complex. *Molecular biology of the cell* **19**, 115-125 (2008).
225. Choi, Y.K., Liu, P., Sze, S.K., Dai, C. & Qi, R.Z. CDK5RAP2 stimulates microtubule nucleation by the gamma-tubulin ring complex. *The Journal of cell biology* **191**, 1089-1095 (2010).
226. Hanafusa, H. *et al.* PLK1-dependent activation of LRRK1 regulates spindle orientation by phosphorylating CDK5RAP2. *Nature cell biology* **17**, 1024-1035 (2015).
227. Rios, R.M. The centrosome-Golgi apparatus nexus. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**, 20130462 (2014).
228. Yalgin, C. *et al.* Centrosomin represses dendrite branching by orienting microtubule nucleation. *Nature neuroscience* **18**, 1437-1445 (2015).
229. Eisman, R.C., Phelps, M.A. & Kaufman, T.C. Centrosomin: a complex mix of long and short isoforms is required for centrosome function during early development in Drosophila melanogaster. *Genetics* **182**, 979-997 (2009).
230. Adams, M.D. *et al.* The genome sequence of Drosophila melanogaster. *Science* **287**, 2185-2195 (2000).
231. Graveley, B.R. *et al.* The developmental transcriptome of Drosophila melanogaster. *Nature* **471**, 473-479 (2011).
232. Brown, J.B. *et al.* Diversity and dynamics of the Drosophila transcriptome. *Nature* **512**, 393-399 (2014).
233. Eisman, R.C., Phelps, M.A. & Kaufman, T. An Amino-Terminal Polo Kinase Interaction Motif Acts in the Regulation of Centrosome Formation and Reveals a Novel Function for centrosomin (cnn) in Drosophila. *Genetics* **201**, 685-706 (2015).
234. Decker, M. *et al.* Limiting amounts of centrosome material set centrosome size in C. elegans embryos. *Current biology : CB* **21**, 1259-1267 (2011).
235. Hamill, D.R., Severson, A.F., Carter, J.C. & Bowerman, B. Centrosome maturation and mitotic spindle assembly in C. elegans require SPD-5, a protein with multiple coiled-coil domains. *Developmental cell* **3**, 673-684 (2002).
236. Woodruff, J.B. *et al.* Centrosomes. Regulated assembly of a supramolecular centrosome scaffold in vitro. *Science* **348**, 808-812 (2015).
237. Wueseke, O. *et al.* The Caenorhabditis elegans pericentriolar material components SPD-2 and SPD-5 are monomeric in the cytoplasm before incorporation into the PCM matrix. *Molecular biology of the cell* **25**, 2984-2992 (2014).
238. Wueseke, O. *et al.* Polo kinase phosphorylation determines C. elegans centrosome size and density by biasing SPD-5 toward an assembly-competent conformation. *Biology open* **5**, 1431-1440 (2016).
239. Zwicker, D., Decker, M., Jaensch, S., Hyman, A.A. & Julicher, F. Centrosomes are autocatalytic droplets of pericentriolar material organized by centrioles. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E2636-2645 (2014).

240. Hyman, A.A., Weber, C.A. & Julicher, F. Liquid-liquid phase separation in biology. *Annual review of cell and developmental biology* **30**, 39-58 (2014).
241. Banani, S.F., Lee, H.O., Hyman, A.A. & Rosen, M.K. Biomolecular condensates: organizers of cellular biochemistry. *Nature reviews. Molecular cell biology* **18**, 285-298 (2017).
242. Boeynaems, S. *et al.* Protein Phase Separation: A New Phase in Cell Biology. *Trends in cell biology* **28**, 420-435 (2018).
243. Laos, T., Cabral, G. & Dammermann, A. Isotropic incorporation of SPD-5 underlies centrosome assembly in C. elegans. *Current biology : CB* **25**, R648-649 (2015).
244. Raff, J.W. Phase Separation and the Centrosome: A Fait Accompli? *Trends in cell biology* **29**, 612-622 (2019).
245. Schnackenberg, B.J., Khodjakov, A., Rieder, C.L. & Palazzo, R.E. The disassembly and reassembly of functional centrosomes in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 9295-9300 (1998).
246. Mahen, R., Jeyasekharan, A.D., Barry, N.P. & Venkitaraman, A.R. Continuous polo-like kinase 1 activity regulates diffusion to maintain centrosome self-organization during mitosis. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 9310-9315 (2011).
247. Ferris, D.K., Maloid, S.C. & Li, C.C. Ubiquitination and proteasome mediated degradation of polo-like kinase. *Biochem Biophys Res Commun* **252**, 340-344 (1998).
248. Lindon, C. & Pines, J. Ordered proteolysis in anaphase inactivates Plk1 to contribute to proper mitotic exit in human cells. *The Journal of cell biology* **164**, 233-241 (2004).
249. Enos, S.J., Dressler, M., Gomes, B.F., Hyman, A.A. & Woodruff, J.B. Phosphatase PP2A and microtubule-mediated pulling forces disassemble centrosomes during mitotic exit. *Biology open* **7**, bio029777 (2018).
250. Babu, Y.S. *et al.* Three-dimensional structure of calmodulin. *Nature* **315**, 37-40 (1985).
251. Ikura, M. *et al.* Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* **256**, 632-638 (1992).
252. Meador, W.E., Means, A.R. & Quiocho, F.A. Target enzyme recognition by calmodulin: 2.4 A structure of a calmodulin-peptide complex. *Science* **257**, 1251-1255 (1992).
253. Meador, W.E., Means, A.R. & Quiocho, F.A. Modulation of calmodulin plasticity in molecular recognition on the basis of x-ray structures. *Science* **262**, 1718-1721 (1993).
254. O'Neil, K.T. & DeGrado, W.F. How calmodulin binds its targets: sequence independent recognition of amphiphilic alpha-helices. *Trends in biochemical sciences* **15**, 59-64 (1990).
255. Tidow, H. & Nissen, P. Structural diversity of calmodulin binding to its target sites. *FEBS J* **280**, 5551-5565 (2013).
256. Villarroel, A. *et al.* The ever changing moods of calmodulin: how structural plasticity entails transductional adaptability. *Journal of molecular biology* **426**, 2717-2735 (2014).
257. Yamniuk, A.P. & Vogel, H.J. Calmodulin's flexibility allows for promiscuity in its interactions with target proteins and peptides. *Mol Biotechnol* **27**, 33-57 (2004).

258. Crivici, A. & Ikura, M. Molecular and structural basis of target recognition by calmodulin. *Annu Rev Biophys Biomol Struct* **24**, 85-116 (1995).
259. Geiser, J.R., Vantuinen, D., Brockerhoff, S.E., Neff, M.M. & Davis, T.N. Can Calmodulin Function without Binding Calcium. *Cell* **65**, 949-959 (1991).
260. Helassa, N., Nugues, C., Rajamanoharan, D., Burgoyne, R.D. & Haynes, L.P. A centrosome-localized calcium signal is essential for mammalian cell mitosis. *FASEB J*, fj201901662R (2019).
261. Paintrand, M., Moudjou, M., Delacroix, H. & Bornens, M. Centrosome organization and centriole architecture: their sensitivity to divalent cations. *Journal of structural biology* **108**, 107-128 (1992).
262. Baron, A.T., Suman, V.J., Nemeth, E. & Salisbury, J.L. The pericentriolar lattice of PtK2 cells exhibits temperature and calcium-modulated behavior. *Journal of cell science* **107 (Pt 11)**, 2993-3003 (1994).
263. Schoborg, T., Zajac, A.L., Fagerstrom, C.J., Guillen, R.X. & Rusan, N.M. An Asp-CaM complex is required for centrosome-pole cohesion and centrosome inheritance in neural stem cells. *The Journal of cell biology* **211**, 987-998 (2015).
264. Galletta, B.J. *et al.* Drosophila pericentrin requires interaction with calmodulin for its function at centrosomes and neuronal basal bodies but not at sperm basal bodies. *Molecular biology of the cell* **25**, 2682-2694 (2014).
265. Morales-Mulia, S. & Scholey, J.M. Spindle pole organization in Drosophila S2 cells by dynein, abnormal spindle protein (Asp), and KLP10A. *Molecular biology of the cell* **16**, 3176-3186 (2005).
266. Saunders, R.D., Avides, M.C., Howard, T., Gonzalez, C. & Glover, D.M. The Drosophila gene abnormal spindle encodes a novel microtubule-associated protein that associates with the polar regions of the mitotic spindle. *The Journal of cell biology* **137**, 881-890 (1997).
267. Xu, X.L. *et al.* The microtubule-associated protein ASPM regulates spindle assembly and meiotic progression in mouse oocytes. *PloS one* **7**, e49303 (2012).
268. Li, C.J. *et al.* Dynamic redistribution of calmodulin in HeLa cells during cell division as revealed by a GFP-calmodulin fusion protein technique. *Journal of cell science* **112 (Pt 10)**, 1567-1577 (1999).
269. Matsumoto, Y. & Maller, J.L. Calcium, calmodulin, and CaMKII requirement for initiation of centrosome duplication in Xenopus egg extracts. *Science* **295**, 499-502 (2002).
270. Tsang, W.Y. *et al.* CP110 cooperates with two calcium-binding proteins to regulate cytokinesis and genome stability. *Molecular biology of the cell* **17**, 3423-3434 (2006).
271. Chen, Z., Indjeian, V.B., McManus, M., Wang, L. & Dynlacht, B.D. CP110, a cell cycle-dependent CDK substrate, regulates centrosome duplication in human cells. *Developmental cell* **3**, 339-350 (2002).
272. Citron, Y.R. *et al.* The centrosomin CM2 domain is a multi-functional binding domain with distinct cell cycle roles. *PloS one* **13**, e0190530 (2018).
273. Villalobo, A., Ishida, H., Vogel, H.J. & Berchtold, M.W. Calmodulin as a protein linker and a regulator of adaptor/scaffold proteins. *Biochim Biophys Acta* **1865**, 507-521 (2018).
274. Lin, T.C., Neuner, A. & Schiebel, E. Targeting of gamma-tubulin complexes to microtubule organizing centers: conservation and divergence. *Trends in cell biology* **25**, 296-307 (2015).

275. Knop, M. & Schiebel, E. Receptors determine the cellular localization of a gamma-tubulin complex and thereby the site of microtubule formation. *The EMBO journal* **17**, 3952-3967 (1998).
276. Viswanath, S. *et al.* The molecular architecture of the yeast spindle pole body core determined by Bayesian integrative modeling. *Molecular biology of the cell* **28**, 3298-3314 (2017).
277. Fong, K.K. *et al.* Direct measurement of the strength of microtubule attachment to yeast centrosomes. *Molecular biology of the cell* **28**, 1853-1861 (2017).
278. Geiser, J.R., Sundberg, H.A., Chang, B.H., Muller, E.G. & Davis, T.N. The essential mitotic target of calmodulin is the 110-kilodalton component of the spindle pole body in Saccharomyces cerevisiae. *Molecular and cellular biology* **13**, 7913-7924 (1993).
279. Stirling, D.A., Welch, K.A. & Stark, M.J. Interaction with calmodulin is required for the function of Spc110p, an essential component of the yeast spindle pole body. *The EMBO journal* **13**, 4329-4342 (1994).
280. Sundberg, H.A., Goetsch, L., Byers, B. & Davis, T.N. Role of calmodulin and Spc110p interaction in the proper assembly of spindle pole body compenents. *The Journal of cell biology* **133**, 111-124 (1996).
281. Spang, A., Grein, K. & Schiebel, E. The spacer protein Spc110p targets calmodulin to the central plaque of the yeast spindle pole body. *Journal of cell science* **109 (Pt 9)**, 2229-2237 (1996).
282. Stirling, D.A., Rayner, T.F., Prescott, A.R. & Stark, M.J. Mutations which block the binding of calmodulin to Spc110p cause multiple mitotic defects. *Journal of cell science* **109 (Pt 6)**, 1297-1310 (1996).
283. Flory, M.R., Moser, M.J., Monnat, R.J., Jr. & Davis, T.N. Identification of a human centrosomal calmodulin-binding protein that shares homology with pericentrin. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5919-5923 (2000).
284. Takahashi, M., Yamagiwa, A., Nishimura, T., Mukai, H. & Ono, Y. Centrosomal proteins CG-NAP and kendrin provide microtubule nucleation sites by anchoring gamma-tubulin ring complex. *Molecular biology of the cell* **13**, 3235-3245 (2002).
285. Keryer, G. *et al.* Dissociating the centrosomal matrix protein AKAP450 from centrioles impairs centriole duplication and cell cycle progression. *Molecular biology of the cell* **14**, 2436-2446 (2003).
286. Zimmerman, W.C., Sillibourne, J., Rosa, J. & Doxsey, S.J. Mitosis-specific anchoring of gamma tubulin complexes by pericentrin controls spindle organization and mitotic entry. *Molecular biology of the cell* **15**, 3642-3657 (2004).
287. Dictenberg, J.B. *et al.* Pericentrin and gamma-tubulin form a protein complex and are organized into a novel lattice at the centrosome. *The Journal of cell biology* **141**, 163-174 (1998).
288. Dummler, A., Lawrence, A.M. & de Marco, A. Simplified screening for the detection of soluble fusion constructs expressed in E. coli using a modular set of vectors. *Microb Cell Fact* **4**, 34 (2005).
289. Drew, D.E., von Heijne, G., Nordlund, P. & de Gier, J.W. Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli. *FEBS Lett* **507**, 220-224 (2001).
290. Waldo, G.S., Standish, B.M., Berendzen, J. & Terwilliger, T.C. Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol* **17**, 691-695 (1999).
291. Gasteiger E., H.C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. Protein Identification and Analysis Tools on the ExPASy Server, in *The*

Proteomics Protocols Handbook. (ed. J.M. Walker) 571-607 (Humana Press, 2005).
292. Rupp, B. *Biomolecular crystallography : principles, practice, and application to structural biology*. (Garland Science, New York; 2010).
293. Diederichs, K. Crystallographic Data and Model Quality. *Methods Mol Biol* **1320**, 147-173 (2016).
294. Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204-1214 (2013).
295. Karplus, P.A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030-1033 (2012).
296. Dauter, Z. Solving coiled-coil protein structures. *Iucrj* **2**, 164-165 (2015).
297. McFerrin, M.B. & Snell, E.H. The development and application of a method to quantify the quality of cryoprotectant solutions using standard area-detector X-ray images. *J Appl Crystallogr* **35**, 538-545 (2002).
298. Bowler, M.W. *et al.* MASSIF-1: a beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules. *J Synchrotron Radiat* **22**, 1540-1547 (2015).
299. Bowler, M.W., Svensson, O. & Nurizzo, D. Fully automatic macromolecular crystallography: the impact of MASSIF-1 on the optimum acquisition and quality of data. *Crystallogr Rev* **22**, 233-249 (2016).
300. Nurizzo, D. *et al.* RoboDiff: combining a sample changer and goniometer for highly automated macromolecular crystallography experiments. *Acta Crystallogr D* **72**, 966-975 (2016).
301. Svensson, O., Malbet-Monaco, S., Popov, A., Nurizzo, D. & Bowler, M.W. Fully automatic characterization and data collection from crystals of biological macromolecules. *Acta Crystallogr D* **71**, 1757-1767 (2015).
302. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* **43**, 186-190 (2010).
303. Winter, G. *et al.* DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D* **74**, 85-97 (2018).
304. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72-82 (2006).
305. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010).
306. Vonrhein, C. *et al.* Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr D* **67**, 293-302 (2011).
307. Vonrhein, C. *et al.* Advances in automated data analysis and processing within autoPROC, combined with improved characterisation, mitigation and visualisation of the anisotropy of diffraction limits using STARANISO. *Acta Crystallogr A* **74**, A360-A360 (2018).
308. Sparta, K.M., Krug, M., Heinemann, U., Mueller, U. & Weiss, M.S. Xdsapp2.0. *J Appl Crystallogr* **49**, 1085-1092 (2016).
309. Winn, M.D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242 (2011).
310. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr D* **59**, 1131-1137 (2003).
311. Matthews, B.W. Solvent Content of Protein Crystals. *Journal of molecular biology* **33**, 491-& (1968).
312. Kantardjieff, K.A. & Rupp, B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Science* **12**, 1865-1871 (2003).

313. Mccoy, A.J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
314. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J Appl Crystallogr* **30**, 1022-1025 (1997).
315. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallographica Section D-Biological Crystallography* **66**, 22-25 (2010).
316. Stein, M. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J Appl Crystallogr* **41**, 641-643 (2008).
317. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D-Biological Crystallography* **62**, 1002-1011 (2006).
318. Cowtan, K. New Developments in Model Building with 'Buccaneer'. *Acta Crystallogr A* **65**, S33-S33 (2009).
319. Dodson, E.J., Murshudov, G.N. & Vagin, A.A. Description of Program Using Maximum Likelihood Residual for Macromolecular Refinement, Illustrated by Several Examples. *Acta Crystallogr A* **52**, C85-C85 (1996).
320. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* **53**, 240-255 (1997).
321. Pannu, N.S., Murshudov, G.N., Dodson, E.J. & Read, R.J. Incorporation of prior phase information strengthens maximum-likelihood structure refinement. *Acta Crystallographica Section D-Biological Crystallography* **54**, 1285-1294 (1998).
322. Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S. & Dodson, E.J. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr D* **55**, 247-255 (1999).
323. Winn, M.D., Isupov, M. & Murshudov, G.N. Use of collective variables to model anisotropic displacements; implementation in the macromolecular refinement program REFMAC. *Acta Crystallogr A* **56**, S201-S201 (2000).
324. Winn, M.D., Isupov, M.N. & Murshudov, G.N. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D* **57**, 122-133 (2001).
325. Steiner, R.A., Lebedev, A.A. & Murshudov, G.N. Fisher's information in maximum-likelihood macromolecular crystallographic refinement. *Acta Crystallogr D* **59**, 2114-2124 (2003).
326. Winn, M.D., Murshudov, G.N. & Papiz, M.Z. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Macromolecular Crystallography, Pt D* **374**, 300-321 (2003).
327. Skubak, P., Murshudov, G.N. & Pannu, N.S. Direct incorporation of experimental phase information in model refinement. *Acta Crystallogr D* **60**, 2196-2201 (2004).
328. Vagin, A.A. *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D* **60**, 2184-2195 (2004).
329. Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D* **75**, 861-877 (2019).
330. Terwilliger, T.C. *et al.* Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallographica Section D-Biological Crystallography* **65**, 582-601 (2009).
331. Zwart, P.H., Grosse-Kunstleve, R.W. & Adams, P.D. Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation. *CCP4 Newsletter No.43* **Winter** (2005).

332. Afonine, P.V., Grosse-Kunstleve, R.W., Urzhumtsev, A. & Adams, P.D. Automatic multiple-zone rigid-body refinement with a large convergence radius. *J Appl Crystallogr* **42**, 607-615 (2009).
333. Afonine, P.V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D* **68**, 352-367 (2012).
334. Headd, J.J. *et al.* Use of knowledge-based restraints in phenix.refine to improve macromolecular refinement at low resolution. *Acta Crystallogr D* **68**, 381-390 (2012).
335. Afonine, P.V., Grosse-Kunstleve, R.W., Adams, P.D. & Urzhumtsev, A. Bulk-solvent and overall scaling revisited: faster calculations, improved results. *Acta Crystallogr D* **69**, 625-634 (2013).
336. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
337. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256-2268 (2004).
338. Schrodinger, LLC The PyMOL Molecular Graphics System, Version 2.3. (2019).
339. Feng, Z. *et al.* Structural Basis for Mitotic Centrosome Assembly in Flies. *Cell* **169**, 1078-1089 (2017).
340. Drew, D. *et al.* GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae. *Nat Protoc* **3**, 784-798 (2008).
341. Yap, K.L. *et al.* Calmodulin target database. *J Struct Funct Genomics* **1**, 8-14 (2000).
342. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190 (2004).
343. Needleman, S.B. & Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology* **48**, 443-453 (1970).
344. Oeffner, R.D., Bunkoczi, G., McCoy, A.J. & Read, R.J. Improved estimates of coordinate error for molecular replacement. *Acta Crystallogr D Biol Crystallogr* **69**, 2209-2215 (2013).
345. Walshaw, J. & Woolfson, D.N. Socket: a program for identifying and analysing coiled-coil motifs within protein structures. *Journal of molecular biology* **307**, 1427-1450 (2001).
346. Buchan, D.W.A. & Jones, D.T. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic acids research* **47**, W402-W407 (2019).
347. Balakrishnan, S., Kamisetty, H., Carbonell, J.G., Lee, S.I. & Langmead, C.J. Learning generative models for protein fold families. *Proteins* **79**, 1061-1078 (2011).
348. Potter, S.C. *et al.* HMMER web server: 2018 update. *Nucleic acids research* **46**, W200-W204 (2018).