Structural Studies of PCM Scaffold Assembly

in *Drosophila melanogaster*

Zhe Feng

Sir William Dunn School of Pathology
Lincoln College



A thesis submitted to the University of Oxford
for the degree of Doctor of Philosophy

Trinity Term 2015

Acknowledgements

Time flies... At the beginning of my PhD study I would never have thought I could actually make this far. The four years in Raff lab/Lea lab and in Oxford are certainly one of the best moments in my life. I feel so lucky that I met loads of nice people, and they all contribute significantly to my good memories in Oxford.

First and foremost, I would like to express my gratitude to Jordan and Susan. Thank you Jordan for choosing me to join the Raff lab at the first place. You are the best supervisor I could ever ask for. You are always so enthusiastic, encouraging and supportive. I will not forget about all the embarrassing tears in front of you in your office. Oh man, the first year of my PhD was tough. Despite nothing worked at all, yet Jordan somehow believed we would get somewhere in the end. Thank you so much for giving me great advice at all times, and all the freedom (especially never worry about any resources and the cost) in the lab. I really enjoyed the free discussion with you at any time. Thank you Susan for tolerating my mistakes and standing up for me in many cases. I am truly grateful for your consistent guidance and spending your precious time on me. Without either of you, I could not have done it.

Thanks to all the previous and current lab members in both Raff lab and Lea lab. I would like to give my special thanks to Matthew (or “the Cottee man” as I called him). You taught me everything from construct design to structure determination. I very much appreciate your time and effort to sit down and go through everything with me step-by-step especially coming down to structure determination. Thank you for being extremely patient to such a stupid and stubborn student. I had a lovely time working with you, and your “madness” made the lab life much more fun. Thanks for all the great suggestions and discussions during my PhD, and this is certainly one of the things I miss a lot since you left the lab. Also thanks to Steve for helping me with structure determination and carrying out all the MALS experiments. Thank you Paul for our close collaborative work, and that was the beginning of my time things finally started to go uphill.

Thank you Vicente for being a great colleague and a lovely BFF (like the cool kids always say). It is so nice to meet you, and later on make friends with you. Thank you for listening to all my worries and non-sense. You might not believe this, but I truly enjoyed sharing my chocolate with you (mainly because I can gain less weight.... kidding!). Thank you Alan for our recent collaborative work and all the interesting scientific discussions in the lab. Thanks to Kasia and Nadine for teaching me how to cycle. Although I only bicycled for half a year, I am sure this is a useful life skill! It was very fun hanging out with you guys and feeding our “food babies”. Thanks to Zsofi for being very supportive in my late years of PhD, and I enjoyed our conversations inside and outside the lab. Thanks Matthijs, Justin, Rich, Ivan and Alice for the lab discussions and organizing social events in the Lea lab.

Huge thanks to my parents for supporting me all the time in my life. I could not have made all these achievements without you two. Although you didn’t understand much about my scientific work (and probably still do not), you always listened to me and tried your best to support me through my PhD. Thank you for providing me a shelter, and make me feel I can try anything without any fear. Sorry being a self-fish daughter, and for the past ten years I have barely spent enough time with you. I really appreciate you giving me the opportunity to study abroad, to broaden my horizon and to enrich my life experience.

非常感谢我的爸爸妈妈一直以来都很支持我。如果没有你们，我也无法达到今天的这些成就。尽管你们对我的学科并不是很了解，但你们也一直尽力给我建议并对我给予最大的鼓励。谢谢你们给我搭建了一个很温暖的港湾，让我可以勇敢出去闯荡不惧怕。对不起你们的女儿一直很自私，在过去的十年里面都没有好好和你们呆在一起孝顺你们。真的很感谢你们给了我一个出去深造的机会，给了我一个看世界的机会。谢谢你们，我亲爱的爸爸妈妈！

In the end, I would like to give a big hug to my beloved Jeffrey. I feel lucky to meet my Mr. Right at the right time. Thanks for being there for me all the time. You made me believe I can be myself in front of my partner. I know the long distance relationship has been challenging, but we made it! Honestly, I never really thought we would fail though. Sorry I am a boring person, and I guess you must have suffered quite a bit especially when I moan-moaned about my “exciting” scientific frustrations. Thanks for supporting me through all my emotional roller coasters and the boring

thesis writing time. The months we spent together in Oxford are the most beautiful moments I had in my PhD. We faced our problems and fight through them. Now we are heading into our new page of life, I am ready to face all the moments with you and grow better with you.

Since this might be the only part of my thesis that people would actually read, I wrote it a bit longer than normally should be (like I always do...). Hope you were not yet bored, because there are still 300 pages of words to enjoy!

THESIS SUMMARY

Structural Studies of PCM Scaffold Assembly In *Drosophila melanogaster*

Zhe Feng, Lincoln College
Submitted for degree of Doctor of Philosophy, Trinity Term 2015

Centrosomes are the major microtubule (MT) organising centres in most animal cells, and they play important roles in many cellular processes. Centrosomes comprise a pair of centrioles surrounded by a matrix of proteins known as the pericentriolar material (PCM). Centrosomes assemble around the mother centriole when the PCM is recruited to nucleate and organise the MTs. For decades, the PCM has been viewed as an amorphous cloud of proteins. Although there are hundreds of PCM proteins, a small number of these are thought to assemble an underlying PCM scaffold that extends around the mitotic centriole and subsequently recruits other mitotic PCM proteins. In flies, Centrosomin (Cnn) and Spindle-defective protein 2 (Spd-2) are such essential proteins, and Asterless (Asl) helps recruit these proteins to the mother centriole. However, it is unclear how these proteins assemble into a scaffold at the centrosome, and none of these core PCM assembly proteins have been characterised at a structural level.

In this thesis, I have identified and recombinantly expressed functional domains of Cnn, Spd-2 and Asl in order to characterise them *in vitro*. In collaboration with colleagues in the laboratory, the biological significance of these domains has been investigated *in vivo*.

I define a **P**hospho-**Re**gulated **M**ultimerisation (PReM) domain within Cnn that is phosphorylated by Polo kinase *in vitro*, that dimerises via a leucine zipper, and that can form higher order oligomers when phospho-mimicking mutations are incorporated into the protein *in vitro*. *In vivo*, forms of Cnn that cannot be phosphorylated in the PReM domain cannot efficiently assemble a Cnn scaffold, while phospho-mimicking forms can spontaneously assemble into scaffold structures. I show that the **C**entrosomin-**M**otif-**2** (CM2) domain of Cnn is also required to promote Cnn scaffold assembly. The crystal structure of CM2 reveals an unusual asymmetric dimer that further assembles into a tetramer *in crystallo*. Strikingly, Cnn-CM2 demonstrates a zinc-binding activity that stabilises the dimer, and preliminary *in vivo* studies indicate that the tetramer interface is essential for Cnn scaffold assembly *in vivo*. Finally, I show that Cnn-CM2 exhibits a weak interaction with the first of three **A**SPM, **S**pd-2, **H**ydin (ASH) domains in Spd-2, and that this domain facilitates Spd-2 expansion into the PCM region. The structure of two of the Spd-2 ASH domains has been investigated using X-ray crystallography and Nuclear Magnetic Resonance spectroscopy, revealing an Immunoglobulin (Ig)-like protein fold. These studies provide the first atomic insights into the molecular mechanism underlying mitotic PCM assembly, and reveal how the assembly of a Cnn scaffold may be controlled to ensure robust centrosome assembly *in vivo*.

Publication resulting from this study

The centrosome-specific phosphorylation of Cnn by Polo/Plk1 drives Cnn scaffold assembly and centrosome maturation

Conduit, P.T.[1], Feng, Z.[1], Richens, J.H., Baumbach, J., Wainman, A., Bakshi, S.D., Dobbelaere, J., Johnson, S., Lea, S.M. and Raff, J.W. (2014). Developmental Cell **28**, 659-669

[1] These authors contribute equally to this work.

Abbreviations

Abs	Absorbance
Ana2	Anastral Spindle 2
ASH	Asp, SPD-2, Hydin, domain
Asl	Asterless
ASPM	Abnormal spindle-like microcephaly-associated protein
AurA	Aurora-A Kinase
BME	β-mercaptoethanol
$CaCl_2$	Calcium Chloride
CaM	Calmodulin
CC	Coiled Coil
CC+	Coilcheck+
CD	Circular Dichroism
CDK	Cyclin-dependent kinase
CDK5RAP2	CDK5 regulatory subunit-associated protein 2
CeSPD-2	*Caenorhabditis elegans* SPD-2
CM1	Common Motif 1
CM2	Common Motif 2
Cnn	Centrosomin
CPAP	Centrosomal P4.1-associated protein
CPB	Cryptic Polo Box
CR	Conserved Region
CSS	Complex Significance Score
CV	Column Volume
Da	Dalton
DMAB	Dimethylaminoborane
DmCnn	*Drosophila melanogaster* Cnn
DmSpd-2	*Drosophila melanogaster* Spd-2
DmAsl	*Drosophila melanogaster* Asl
DNA	Deoxyribonucleic Acid
DOSY	Diffusion Ordered NMR Spectroscopy
DSC	Differential Scanning Calorimetry
DTT	Dithiothreitol
FRAP	Fluorescence Recovery After Photobleaching
FT	Flow Through
FWHM	Full Width Half Maximum
EDTA	Ethylenediaminetetraacetic acid
EG	Ethylene Glycol
EGTA	Ethylene glycol tetraacetic acid
EM	Electron Microscopy
ET	Electron Tomography

GFP	Green Fluorescent Protein
γTuSC	gamma-tubulin small complex
γTuRC	gamma-tubulin ring complex
H-bond	Hydrogen Bond
HCl	Hydrogen Chloride
H_2O	Water
HsCDK5RAP2	*Homo sapiens* CDK5RAP2
HsCep192	*Homo sapiens* Cep192
HsCep152	*Homo sapiens* Cep152
HSQC	Heteronuclear Single Quantum Coherence Spectroscopy
IEX	Ion Exchange Chromatography
Ig	Immunoglobulin
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
IR	Infra-Red
Kan	Kanamycin
KCl	Potassium Chloride
Kd	Dissociation Constant
LB	Luria Bertani Broth
LRRK	Leucine Rich Repeat Kinase
LZ	Leucine Zipper motif
MAD	Multiwavelength Anomalous Dispersion
MALDI	Matrix Assisted Laser Desorption/Ionization
MALS	Multi Angle Light Scattering
MBP	Maltose Binding Protein
MCPH	Microcephaly Primary Hereditary (Autosomal Recessive Primary Microcephaly)
$MgCl_2$	Magnesium Chloride
MR	Molecular Replacement
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MSA	Multiple Sequence Alignment
MST	Microscale Thermophoresis
MT	Microtubule
MTOC	Microtubule Organising Centre
MSP	Major Sperm Protein
MW	Molecular Weight
MWCO	Molecular Weight Cut Off
NB	Neuroblast
NCS	Non-Crystallographic Symmetry
NMR	Nuclear Magnetic Resonance
Ni	Nickel
Ni-NTA	Nickel Nitrilotriacetic Acid
No.	Number

OD_{600}	Optical Density at 600 nm
PCM	Pericentriolar Material
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
pI	Isoelectric point
PISA	Proteins, Interfaces, Structures and Assemblies server
PLK1	Polo-Like Kinase 1
PLK4	Polo-Like Kinase 4
PLP	Pericentrin-Like Protein
PReM	Phosphorylation-Regulated Multimerisation domain
PSI-BLAST	Protein-Specific Iterative Basic Local Alignment Search Tool
RMSD	Root Mean Square Deviation
RNA	Ribonucleic Acid
RNAi	RNA interference
RT	Room Temperature
SAD	Singlewavelength Anomalous Dispersion
SAS-4/5/6	Spindle Assembly Abnormal 4/5/6
SD	Standard Deviation
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SEM	Standard Error of the Mean
SPD-2	Spindle-defective protein 2
SPD-5	Spindle-defective protein 5
SRF	Self Rotation Function
SSP-19	Sperm-Specific Protein-19
STAN	STIL/Ana2 Domain
STIL	SCL TAL1 Interrupting Locus
S2	Schneider 2 Cells
TB	Terrific Broth
T-jump	Temperature-Jump
Ubq	Ubiquitin
VAPB	Vesicle-associated membrane protein-associated protein B
WB	Working Buffer
WT	Wild Type
Y2H	Yeast Two Hybrid
$ZnCl_2$	Zinc Chloride
ZYG-1	Zygotic Defective 1
3D-SIM	Three Dimensional Structured Illumination Microscopy

Table of Contents

Thesis Summary ..... **I**

Publication resulting from this study ..... **VII**

Abbreviations ..... **VIII**

Table of Contents ..... **XI**

Chapter 1: Introduction **1**
1.1 Centrosome function **1**
1.1.1 Microtubule composition and nucleation 2
1.1.2 Mitotic spindle assembly 5
1.1.3 Centrioles in cilia 6
1.1.4 Centrosomes in development 7
1.2 Centrosomes in disease **9**
1.2.1 Centrosome and cancer 9
1.2.2 Centrosomes and microcephaly 10
1.3 The centrosome cycle **12**
1.4 Centriole biogenesis **14**
1.4.1 Centriole structure 14
1.4.2 Centriole assembly 17
1.5 Centrosome maturation **20**
1.6 PCM structure and composition **22**
1.6.1 PCM organisation in interphase 22
1.6.2 PCM organisation in mitosis 22
1.6.3 Centrosomin (Cnn) and its related family proteins 24
1.6.4 Spindle defective protein-2 (Spd-2) and its related proteins 28
1.6.5 Asterless (Asl) and its family member proteins 31
1.6.6 PCM dynamics 33
1.7 Thesis Aims **35**
Prefatory Note **35**

Chapter 2: Identification and characterisation of a PhosphoRegulated-Multimerisation (PReM) domain in Cnn **37**
2.0 Introduction **37**
2.0.1 Centrosomin (Cnn) 37
2.0.2 Aims 38
2.0.3 Key findings 38
2.1 Bioinformatic analysis of Cnn **39**
2.1.1 Cnn and CDK5RAP2 contain some conserved features 39
2.1.2 Identification of a highly conserved Cnn middle segment in flies 41
2.2 Expression and purification of the Cnn middle segment **42**
2.2.1 Screening for soluble fragments covering the Cnn middle segment 42
2.2.2 Attempt to purify HisCnn$^{455-622}$ on a large scale 43
2.2.3 Large-scale purification of MBPCnn fusion constructs for *in vitro* assays 45
2.3 Radioactive kinase assay of the Cnn middle segment **51**
2.4 Investigating the potential significance of phosphorylation and the LZ on the self interaction of the Cnn middle segment **52**
2.4.1 SEC-MALS analysis of MBPCnn fusion constructs 52
2.5 *In vivo* analysis of PReM domain function **54**
2.5.1 An mRNA injection assay to assess centrosomal localisation 54
2.5.2 Construct cloning and RNA preparation 55
2.5.3 Phosphorylation of PReM domain is essential for efficient Cnn scaffold formation and PCM assembly 56
2.5.4 Multisite phosphorylation is required for efficient Cnn scaffold assembly 58

2.5.5 The LZ within the PReM domain is required for efficient Cnn scaffold formation 60
2.5.6 Phosphomimetic mutations within the PReM domain allow Cnn to form spontaneous scaffolds independently of centrosomes....60
2.5.7 Perspectives on *in vivo* experiments....62
2.6 Purification and characterisation of the phosphomimetic PReM domain constructs....64
2.6.1 Purification and characterisation of $Cnn^{403-608}$ 10D/E....64
2.6.2 Limited proteolysis of $Cnn^{403-608}$ 10D/E....66
2.6.3 Purification of $Cnn^{443-608}$ 10D/E and $Cnn^{490-608}$ 10D/E....68
2.6.4 SEC-MALS analysis of truncated Cnn 10D/E constructs....72
2.6.5 Circular Dichroism analysis of Cnn 10D/E constructs....73
2.7 Crystallisation trials with the phosphomimetic PReM domain constructs....74
2.7.1 Crystallisation of $Cnn^{443-608}$ 10D/E and $Cnn^{490-608}$ 10D/E....74
2.7.2 Perspectives on crystallisation trials....74
2.8 Perspectives....76

Chapter 3: Structure and function analysis of Cnn Common Motif 2 (CM2)....79
3.0 Introduction....79
3.0.1 Common Motif 2 (CM2)....79
3.0.2 Aims....79
3.0.3 Key findings....80
3.1 Designing a construct for structural characterisation....81
3.2 Purification and characterisation of Cnn-CM2....82
3.2.1 Purification of Cnn-CM2....82
3.2.2 Biophysical characterisation of Cnn-CM2....84
3.3 Crystallisation of Cnn-CM2....86
3.4 Structural determination of Cnn-CM2....87
3.4.1 'Native' data collection and processing....87
3.4.2 Anomalous data collection and processing....89
3.4.3 Solving the phase problem....92
3.4.4 Structure refinement....94
3.4.5 Improving crystallisation efficiency....97
3.4.6 Improved density map....99
3.5 Analysis of the Cnn-CM2 structure....103
3.5.1 Cnn-CM2 forms an unusual parallel, asymmetric dimer....103
3.5.2 Two dimers of Cnn-CM2 interlocked into an antiparallel tetramer....108
3.5.3 Conservation of the Cnn-CM2 structure....111
3.5.4 The Zn^{2+} binding sites within Cnn-CM2 may be evolutionarily conserved....113
3.6 Further characterisation of Cnn-CM2....115
3.6.1 Characterisation of Zn^{2+} binding *in vitro*....115
3.6.2 Can Cnn-CM2 form a tetramer in solution?....116
3.6.3 Does the T1133E mutation disrupt the formation of higher-order Cnn-CM2 oligomers in solution?....121
3.7 *In vivo* studies of Cnn-CM2 structural features....124
3.7.1 Design of constructs....124
3.7.2 Dual function of Cnn-CM2 in centrosomal localisation and scaffold stability....125
3.7.3 Importance of Zn^{2+} binding *in vivo*....127
3.7.4 Planned experiments with transgenic *Drosophila*....128
3.8 Perspectives....130

Chapter 4: Function and structure characterisation of Spd-2....135
4.0 Introduction....135
4.0.1 Spindle-defective protein 2 (Spd-2)....135
4.0.2 Aims....136
4.0.3 Key findings....136
4.1 Bioinformatic analysis of Spd-2....137

4.1.1 Sequence alignments reveal conserved regions within Spd-2 137
4.1.2 Identification of conserved domain folds 137
4.2 Analysis of Spd-2 scaffolding activity *in vivo* 141
4.2.1 Construct design and RNA preparation 141
4.2.2 Full-length Spd-2 localises strongly to centrosomes 143
4.2.3 The N-terminal region of Spd-2 is required for centriolar targeting, and the C-terminal region is needed for PCM localisation 144
4.2.4 Spd-2 scaffold assembly is facilitated by the ASH1 and ASH2 domains 146
4.2.5 Attempts to identify phosphorylation sites within Spd-2 148
4.2.6 Construct design and cloning for transgenic *Drosophila* lines 150
4.2.7 Perspectives on *in vivo* experiments 151
4.3 Purification and characterisation of the Spd-2 ASH1 domain 153
4.3.1 Expression and purification of the Spd-2 ASH1 domain 153
4.3.2 Biophysical characterisation of Spd-2 ASH1 domain 156
4.4 Structural characterisation of the Spd-2 ASH1 domain 158
4.4.1 Crystallisation trials with Spd-$2^{697-805}$ 158
4.4.2 Sample preparation for NMR spectroscopy 162
4.4.3 NMR data collection for assignment of Spd-$2^{697-805}$ 163
4.5 Purification and characterisation of the Spd-2 ASH2 domain 164
4.5.1 High-throughput screen for s soluble fragment of the Spd-2 ASH2 domain 164
4.5.2 Large-scale purification of MBPSpd-$2^{820-942}$ 167
4.5.3 Large-scale purification and characterisation of the phosphomimetic Spd-2 ASH2 domain 169
4.6 Purification and characterisation of the Spd-2 ASH3 domain 173
4.6.1 Expression and purification of the Spd-2 ASH3 domain 173
4.6.2 Biophysical characterisation of the Spd-2 ASH3 domain 175
4.6.3 Analysis of the ASH1: ASH3 domain interaction 177
4.7 Structure characterisation of the Spd-2 ASH3 domain 179
4.7.1 Crystallisation trials with the Spd-2 ASH3 domain 179
4.7.2 Data collection and processing 180
4.7.3 Structure determination by molecular replacement 184
4.7.4 Refinement of the Spd-$2^{1047-1146}$ structure 186
4.8 Analysis of the Spd-2 ASH3 structure 188
4.8.1 The Spd-2 ASH3 domain adopts an immunoglobulin (Ig)-like fold 188
4.8.2 Sequence and structure conservation of the Spd-2 ASH3 domain 190
4.8.3 Comparison to proteins with similar folds 192
4.9 Perspectives 196

Chapter 5: Interaction study between the 2 conserved domains in Cnn and Spd-2 201
5.0 Introduction 201
5.0.1 Evidence for Cnn: Spd-2 interaction 201
5.0.2 Aims 201
5.0.3 Key findings 202
5.1 Probing a putative interaction between the ASH1 and CM2 domains in solution 203
5.1.1 Interaction characterisation via MST measurements 203
5.1.2 Interaction characterisation via ^{1}H-^{15}N HSQC NMR measurements 206
5.1.3 T1133E mutation does not affect the CM2: ASH1 interaction 209
5.2 Design of mutations to dissect the interaction between the CM2 and ASH1 domains 211
5.3 Future experiments 213

Chapter 6: Purification and characterisation of Asterless (Asl) 215
6.0 Introduction 215
6.0.1 Asterless (Asl) 215
6.0.2 Aims 216

6.0.3 Key findings....216
6.1 Bioinformatic analysis of Asl....217
6.1.1 Asl proteins are highly coiled-coil, showing little sequence conservation....217
6.1.2 Rationalisation of the Asl constructs designed for *in vitro* characterisation....219
6.2 Purification and characterisation of Asl N-terminal fragments....221
6.2.1 Screening for soluble N-terminal fragments....221
6.2.2 Purification of HisAsl$^{1-86}$....222
6.2.3 Biophysical characterisation of HisAsl$^{1-86}$....223
6.2.4 Purification of HisAsl$^{1-356}$....226
6.2.5 Limited proteolysis of HisAsl$^{1-356}$....226
6.2.6 Purification of HisAsl$^{47-356}$....229
6.2.7 Biophysical characterisation of HisAsl$^{47-180/356}$....231
6.2.8 Attempts in structural characterisation of Asl$^{47-180/356}$....233
6.3 Purification and characterisation of the Asl C-terminal fragments....237
6.3.1 Purification of HisAsl$^{648/733-994}$....237
6.3.2 Purification of MBPAsl$^{648/733-994}$....238
6.3.3 Biophysical characterisation of Asl$^{733-994}$....241
6.4 Perspectives....243

Chapter 7: Discussion....247
7.1 Summary....247
7.2 PCM expansion during mitosis....249
7.2.1 Temporal and spatial control....249
7.2.2 Feedback mechanism....250
7.2.3 A unified model of centrosome maturation?....250
7.3 Future directions....252
7.4 Conclusion....254

8.0: Appendix....255
8.1 Standard laboratory methods....255
8.1.1 Primer design, PCR amplification and ligation....255
8.1.2 Gateway cloning....256
8.1.3 Site-directed mutagenesis....257
8.1.4 Bacterial transformation....257
8.1.5 Small-scale expression and purification trials....258
8.1.6 Large-scale protein expression and purification....259
8.1.7 SDS-PAGE gel electrophoresis....262
8.1.8 Multi Angle Light Scattering (MALS)....263
8.1.9 Differential Scanning Calorimetry (DSC)....263
8.1.10 Crystallisation trials....264
8.1.11 Protein labelling and purification for use with MicroScale Thermophoresis (MST)....264
8.1.12 mRNA preparation....265
8.1.13 Image presentation and analysis....266
8.2 X-ray crystallography....267
8.2.1 Data collection....267
8.2.2 Data processing....268
8.2.3 The Phase problem....270
8.2.4 Phase determination by Molecular Replacement (MR)....271
8.2.4 Anomalous diffraction in Phase solution....273
8.2.5 Structure model refinement....275
8.3 MSA of Spd-2 homologs reveal conservation within the SPD-2 domain....277
8.4 Purification and characterisation of HisCnn$^{93-185}$....280

Bibliography....283

Chapter 1: Introduction

1.1 Centrosome function

Centrosomes comprise a pair of barrel-shaped centrioles embedded in an amorphous matrix of proteins known as the pericentriolar material (PCM) [Berns and Richardson, 1997; Gould and Borisy, 1977], (**Figure 1.1**). The centrosome is the major microtubule-organising centre (MTOC) found in animal cells, and it has been implicated in a variety of cellular processes including cell polarity, cell division and cell motility [Bornens, 2002; Bornens, 2012; Kellogg et al., 1994]. In interphase, the centrosome organises microtubules (MTs) that facilitate the positioning of cellular organelles and the transport of molecules within the cell [Kapitein and Hoogenraad, 2011]. The centrosome duplicates during S phase, so that during mitosis one centrosome migrates to each pole of the nucleus and organises the MT-based mitotic spindles [Walczak and Heald, 2008]. This ensures the correct segregation of the chromosomes to the daughter cells. Centrioles can also become basal bodies that nucleate MTs for the formation of cilia and flagella [Dawe et al., 2007; Ringo, 1967; Snell et al., 1974].

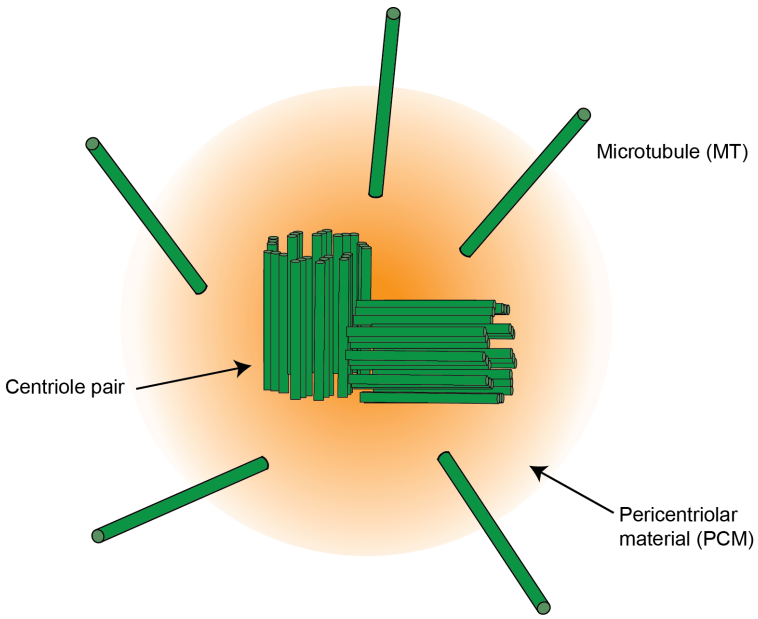


Figure 1.1 Schematic representation of a centrosome containing a pair of orthogonally arranged centrioles embedded in an amorphous matrix of proteins known as the pericentriolar material (PCM). The centrosome is the major microtubule-organising center (MTOC) in animal cells. The centriole is a cylinder like structure surrounded by MT blades. MTs are nucleated and anchored in the PCM.

1.1.1 Microtubule composition and nucleation

MTs are hollow cylindrical tubes made from α/β-tubulins with a diameter of ~25 nm [Kollman et al., 2011], (**Figure 1.2**). Heterodimers of α- and β-tubulins polymerise in a GTP-dependent manner into protofilaments through longitudinal contacts [Nogales et al., 1998; Nogales et al., 1999]. *In vivo*, 13 protofilaments associate laterally to form a MT [Evans et al., 1985; Ledbetter and Porter, 1964; Tilney et al., 1974]. The longitudinal orientation of tubulin heterodimers leads to an establishment of intrinsic polarity within MTs. In particular, the α-tubulin exposed end is the so-called minus end and the β-tubulin exposed end is the so-called plus end [Fan et al., 1996; Mitchison, 1993], (**Figure 1.2**). While the minus end is relatively stable and embedded in the centrosome, the plus end is highly dynamic and exposed to the cytoplasm [Jiang and Aknmanova, 2010]. Thus, polymerisation and de-polymerisation primarily occur at the plus end contributing to the "dynamic instability" of MTs [Howard and Hyman, 2003].

In vitro, MT polymerisation can occur spontaneously in the presence of α- and β-tubulins at sufficiently high concentrations [Herzog and Weber, 1977; Lee and Timasheff, 1975]. However, *in vivo* this process usually requires a nucleator such as γ-tubulin [Oakley et al., 1990; Spang et al., 1996; Stearns et al., 1991; Stearns and Kirschner, 1994; Zheng et al., 1995]. Two molecules of γ-tubulin associate with GCP2 and GCP3 (Dgrip84 and 91 in *Drosophila*) in a γ-tubulin small complex (γTuSC), [Murphy et al., 1998; Oegema et al., 1999]. Multiple γTuSCs further assemble into a γ-tubulin ring complex (γTuRC), in which GCP4, 5 and 6 (Dgrip75, 128 and 163 in *Drosophila*) together function as a cap-like scaffold at the base of MT protofilaments [Gunawardane et al., 2000; Murphy et al., 1998; Zheng et al., 1995], (**Figure 1.3A,B**). How does the γTuRC nucleate MT assembly? Recent studies *in vitro* have demonstrated the formation of *Saccharomyces cerevisiae* γTuSC oligomers with MT-like symmetry in the absence of γTuRC [Kollman et al., 2010; Lin et al., 2014]. Cryo-EM study of those γTuSC rings further showed that the plus ends of γ-tubulin are fully exposed, suggesting a longitudinal contact with the minus ends of α- and β-tubulins. This supports a template model for MT nucleation, where γ-tubulin subunits self associate into a ring like structure of γTuRC and make longitudinal contacts with α/β-tubulin dimers to scaffold the MT assembly [Keating and Borisy, 2000; Moritz et al., 2000], (**Figure 1.3C**).

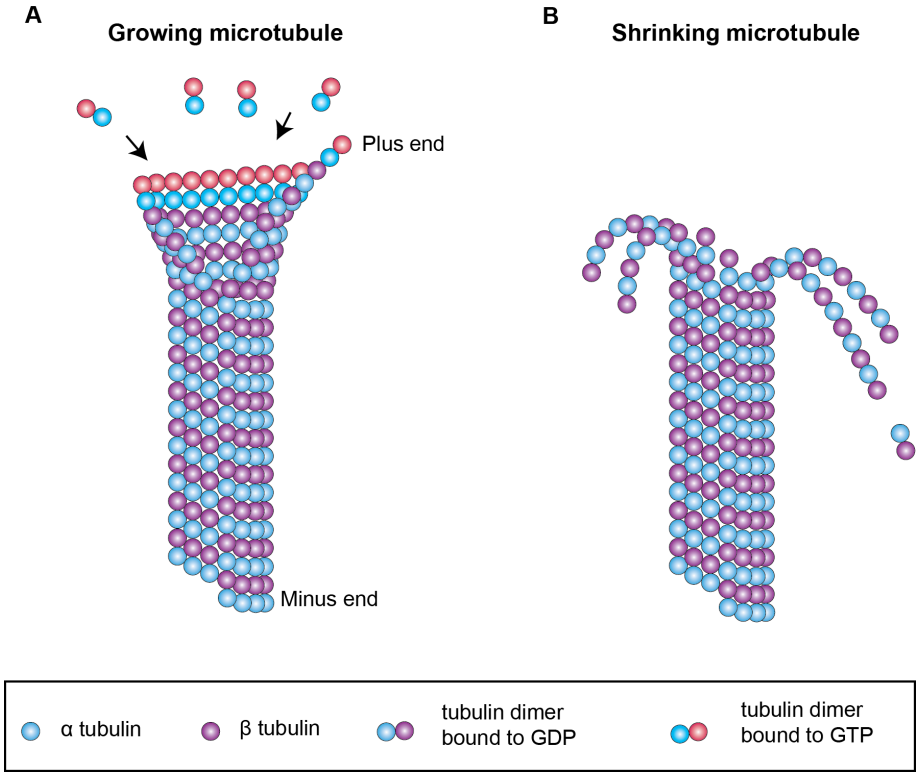


Figure 1.2 Microtubule structure. Microtubules are composed of α- and β-tubulin heterodimers, which are represented by spheres of different colors, stacking vertically to form protofilaments. Each MT typically consists of 13 protofilaments that assemble laterally into a 25 nm wide hollow, cylindrical structure. **(A)** Diagram of a growing MT showing outward curvature at the plus end possibly reflecting the conformational changes that occur within the subunit during polymerisation. New GTP-bound subunits are added to the growing MT, which in turn completes the active site for GTP hydrolysis, and consequently the main lattice is composed of GDP-bound αβ-tubulins. The GDP lattice is unstable, but is protected by a stabilising GTP cap at the plus end. **(B)** Diagram of a shrinking MT. The loss of GTP cap leads to rapid depolymerisation that is characterised by outward curling of protofilaments. Diagram modified from [Howard and Hyman, 2003].

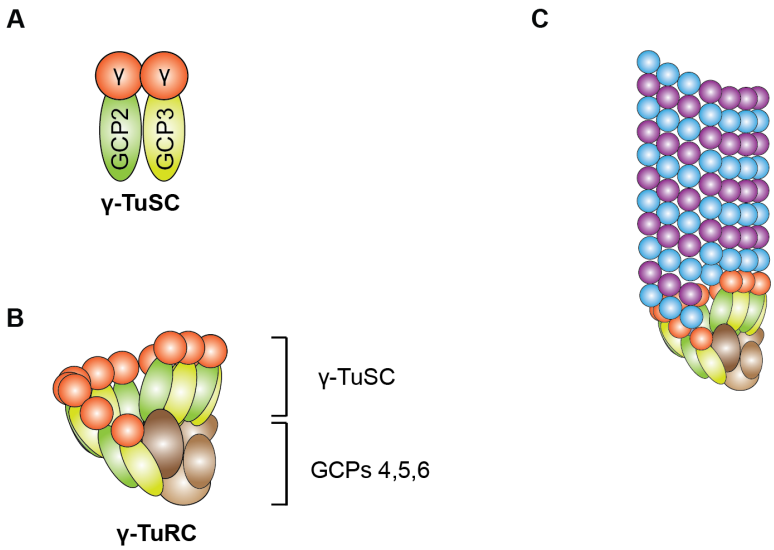


Figure 1.3 γ-tubulin complexes and the model of MT nucleation. (A) Schematic representation of γ-tubulin small complex (γ-TuSC) that contains two copies of γ-tubulin and one copy of each γ-tubulin complex protein (GCP2/Dgrip84 and GCP3/DGrip91). **(B)** Composition of γ-tubulin ring complex (γ-TuRC) in which multiple γ-TuSCs assemble side-to-side with a cap-like scaffold at the base. GCP4 (DGrip75), GCP5 (DGrip128) and GCP6 (DGrip163) together function as the cap-like structure. **(C)** Diagram illustrating the "template model" for MT nucleation. γ-TuRC, at the base of the MT, presents a ring of γ-tubulin subunits for head-to-tail interaction with α/β-tubulin heterodimers and thus promotes MT polymerisation. Diagram modified from [Kollman et al., 2011].

Although γTuRCs are present throughout the cytoplasm [Moudjou et al., 1996], their nucleating activity is limited to specific locations in the cell such as the centrosome or within the mitotic spindle [Kollman et al., 2011]. It is likely the direct binding of γ-tubulin in the γTuSC by attachment factors induces its nucleating activity possibly via a conformational change in GCP3 (DGrip91 in *Drosophila*), [Kollman et al., 2010; Kollman et al., 2011]. The PCM contains several factors that recruit and anchor γ-tubulin complexes to the centrosome, allowing it to function as the major MTOC [Delgehyr et al., 2005; Fong et al., 2008; Gomez-Ferreria et al., 2007; Sawin et al., 2004; Takahashi et al., 2002; Zhu et al., 2008; Zimmerman et al., 2004]. Once the γTuRCs are embedded in the PCM, they provide a cap to prevent the growth or shrinkage of MTs at the minus end [Wiese and Zheng, 2000].

Nonetheless, MT nucleation also occurs via acentrosomal pathways to ensure the integrity of spindle formation and to increase the efficiency of chromosome segregation. In the chromatin pathway, MT nucleation occurs proximal to the chromosome [Heald et al., 1996]. At least in some systems, it is guided by a gradient of RanGTP that in turn triggers the release of spindle assembly factors to promote the nucleation and stabilisation of MTs around the mitotic chromatin [Clarke and Zhang, 2008; Kalab et al., 2006; O'Connell and Khodjakov, 2007]. MT nucleation can also occur on pre-existing MTs via the Augmin complex. This pathway was initially identified in *Drosophila* cultured cells, where the knock down of Augmin subunits led to the impaired recruitment of γ-tubulin to spindles and aberrant mitotic spindle formation [Goshima et al., 2008; Wainman et al., 2009]. Later on, the same defects have also been demonstrated in *Xenopus* egg extracts and human cells [Lawo et al., 2009; Petry et al., 2013; Uehara and Goshima, 2010]. In addition, in acentriolar MTOCs (aMTOCs), some PCM components can self organise into foci that nucleate and anchor MTs although the underlying molecular mechanism is less studied [Baumbach et al., 2015].

1.1.2 Mitotic spindle assembly

During mitosis, a pair of centrosomes migrates to opposing sides of the nucleus and forms the spindle poles. The mitotic spindle is essential for chromosome segregation during cell division. During mitotic spindle formation, MTs are nucleated via various pathways as outlined above (**Figure 1.4**). In particular, MTs originated from centrosomes make contacts with kinetochores to form K-fibres, and with MTs from the opposite spindle pole [Rieder, 2005]. The kinetochore is a specialised macromolecular complex on the chromosome, and it links the centrosome to the chromosome via associated MT bundles. The K-fibres provide pulling forces to separate the sister chromatids during segregation. The overlapping MTs, known as polar MTs, create forces to push apart the centrosomes and chromosomes [Lodish et al., 2000]. In addition, MTs in contact with the cell cortex, named as astral MTs, are thought to be involved in correct positioning of the mitotic spindle apparatus and subsequently to determine the site of cleavage furrow formation during cytokinesis [Lodish et al., 2000].

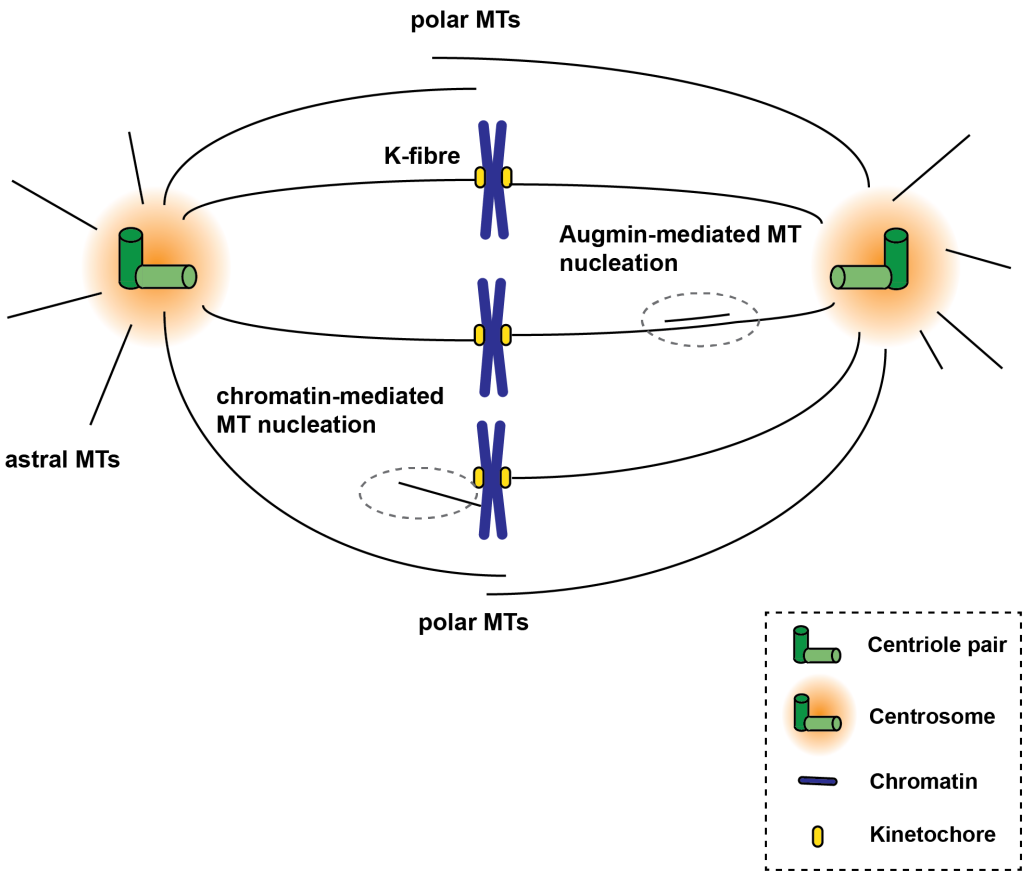


Figure 1.4 Schematic representation of mitotic spindle formation. The centrosome at either pole of the spindle nucleates MTs that are involved in the following activities: 1) astral MTs, form contacts with the cell cortex and help to position the mitotic spindle; 2) K-fibres, interact with the kinetochore; 3) polar MTs, form between MTs emanated from opposite poles. Both K-fibres and polar MTs generate pulling forces during the chromosome segregation. MTs can also be nucleated via acentrosomal pathways. Shown in the figure above are two examples, where MTs are nucleated within the existing spindle via a multi-protein complex known as Augmin complex and around the chromatin driven by a Ran-GTP gradient. Diagram modified from [Conduit, 2010].

1.1.3 Centrioles in cilia

A differentiated cell enters G0/quiescent phase after mitosis. In most vertebrate quiescent cells, the pair of centrioles migrates to the plasma membrane where a cilium is formed. The mother centriole being closer to the cell membrane differentiates into the basal body that nucleates the cilium [Broekhuis et al., 2013], (**Figure 1.5**). The cilium contains a MT-based cytoskeleton (called an axoneme) surrounded by a cell membrane [Breslow et al., 2013]. Molecules are transported along the cilia in both directions via a mechanism known as intraflagellar transport (IFT) [Ke and Yang, 2014; Malicki and Avidor-Reiss, 2014; Sung and Leroux, 2013]. There are two types of cilia: motile and non-motile, depending on the MT organisation pattern within the axoneme. The axoneme of motile cilia contains a central pair of MT singlets and nine peripheral pairs of MT doublets, termed as a "9+2" arrangement; in the axoneme of non-motile cilia, the pair of central MTs is missing, hence containing a "9+0" arrangement (**Figure 1.5B**). Motile cilia are usually found on the cell surface in large numbers and beat for locomotion or to move liquid over the cell [Roy, 2009]. They are structurally similar to flagella in eukaryotic cells. Non-motile cilia are present in almost every human tissue [Veland et al., 2009], and they are required for transducing signals from the extracellular environment to intracellular signalling cascades including Hedgehog (Hh) signalling [Huangfu et al., 2003; Wong and Reiter, 2008]. Note in *Drosophila* Hh signalling does not require cilia, although it is essential for the fly development [Roy, 2011].

Defects in cilia assembly and organisation can lead to a variety of complex pathologies collectively known as ciliopathies, such as polycystic kidney disease and Meckel-Gruber syndrome, and these have been discussed extensively in several recent reviews [Basten and Giles, 2013; Bettencourt-Dias et al., 2011; Chavali et al., 2014; Debec et al., 2010; Mahjoub, 2013; Nigg and Raff, 2009; Nigg et al., 2014; Valente et al., 2014].

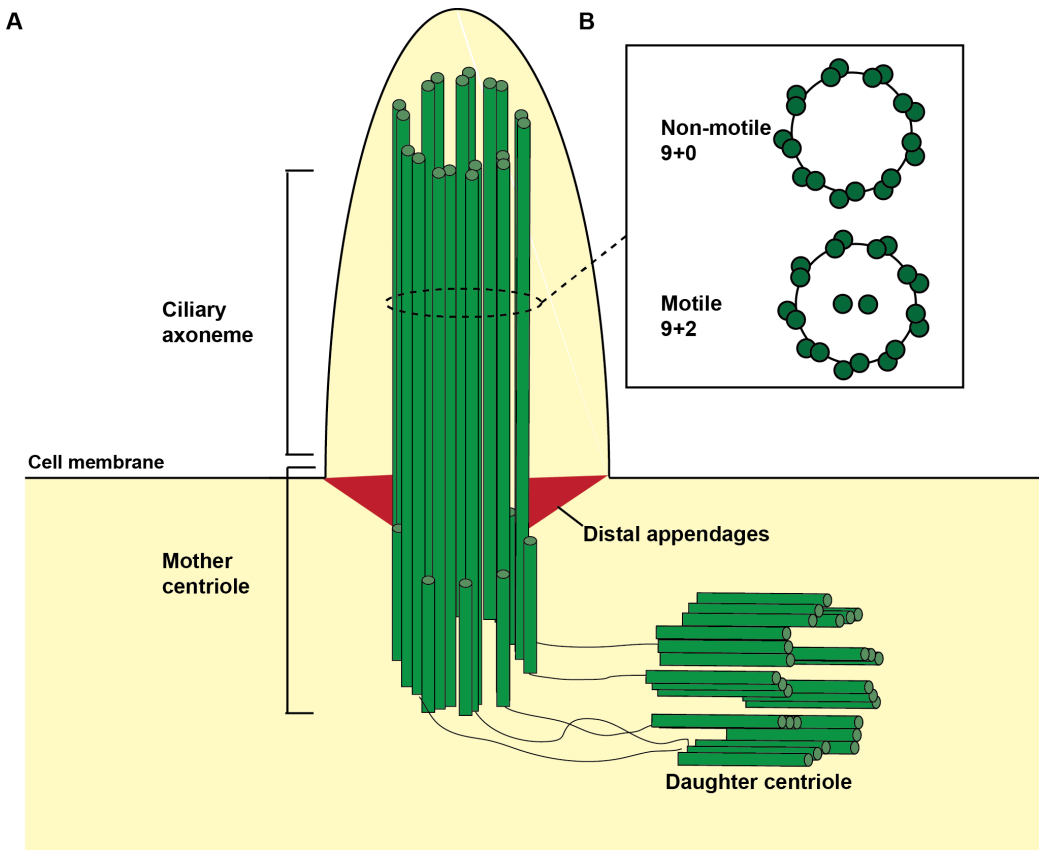


Figure 1.5 Diagram representing a cilium. (A) The mother centriole docks at the plasma membrane via its distal appendages that are missing in the daughter centriole. It then becomes the basal body of a cilium, from which the ciliary axoneme is nucleated. The axoneme consists of MT doublets arranged in different patterns, as shown in (**B**). **(B)** Top view of a horizontal section across the axoneme. Non-motile/primary cilia contain nine doublets with no central pair (9+0 arrangement). Motile cilia have a central MT pair (9+2 arrangement).

1.1.4 Centrosomes in development

Although the centrosome is a major MTOC in many animal cells, it is not essential for spindle assembly. As mentioned above, there are alternative mechanisms to nucleate and stabilise MTs for the spindle formation. But the presence of centrosome ensures more efficient spindle assembly [Heald et al., 1997; Khodjakov et al., 2000; Sir et al., 2013]. For instance, in cultured cells, the removal of centrosomes by laser ablation or microsurgery did not lead to an increase in non-bipolar spindle formation. The bipolar spindles were nucleated by mitotic chromosomes instead [Berns and Richardson, 1997; Khodjakov et al., 2000; Khodjakov et al., 2002; Loncarek et al., 2008; Piel et al., 2001; Uetake et al., 2007]. However, these acentrosomal cells showed a higher frequency of failure in cytokinesis. And even if cytokinesis was completed, cells tended to arrest in the following G1/S phase.

Centrosomes might be dispensable for normal cell division, but they are thought to be more important in asymmetric cell division [Basto et al., 2006], a process that gives rise to two daughter cells of different developmental potential. This is because the astral MTs are essential for direct contact with the cell-fate determinants in the cell cortex for correct positioning and consequently the appropriate segregation [Knoblich, 2010; Li et al., 2014]. However, studies of *asl* and *cnn* mutants in *Drosophila* indicate that centrosomes are not essential for asymmetric cell divisions in larval neuroblasts (NBs) since minor defects were observed in spindle orientation during prophase [Bonaccorsi et al., 2000; Giansanti et al., 2001; Lucas and Raff, 2007; Megraw et al., 2001]. In addition, during telophase a mutual MT-cortex interaction restores the orientation of mitotic spindle along the apical basal axis, known as the “telophase rescue” [Siegrist and Doe, 2007].

Indeed, centrosomes are not essential for adult development in *Drosophila*. Evidence is provided from studies of the fly embryos that carry homozygous mutation for the *Sas-4* gene, which is a crucial regulator of centriole duplication [Basto et al., 2006; Stevens et al., 2007]. Centrosomes are indispensable for early fly embryogenesis where extremely rapid centrosome assembly occurs [Stevens et al., 2007; Varmark et al., 2007]. In this case, the mutant fly embryo could pass through the early stages due to its maternally contributed Sas-4 protein by the heterozygous mother. The maternal pool of Sas-4 protein was then exhausted after the early stages of embryogenesis, leading to a gradual loss of centrioles and centrosomes. Surprisingly, these mutant flies lacking centrosomes hatched at near normal rates, and did not show any morphological defects. However, they were uncoordinated due to failure in cilia formation. Interestingly, recent studies suggest that most vertebrate cells can also undergo relatively normal chromosome segregation without centrosomes, although a p53-dependent pathway is activated in response to centrosome loss and triggers either cell apoptosis or cell cycle arrest at G1 phase [Bazzi and Anderson, 2014; David et al., 2014; Insolera et al., 2014; Lambrus et al., 2015; Wong et al., 2015]. In flies, no active cell stress pathways are detected, at the transcriptional level, in the presence of too many or too few centrosomes [Baumbach et al., 2012].

1.2 Centrosomes in disease

1.2.1 Centrosome and cancer

More than 100 years ago, Theodor Boveri first proposed the importance of centrosomes in cancer development [Boveri, 1914]. He suggested that extra centrosomes in adult cells could lead to abnormal mitotic spindle assembly and consequently to chromosomal instability (CIN) that can lead to tumour formation. Indeed, 90% of human solid tumours carry abnormal numbers of chromosomes (a condition known as aneuploidy), [Mitelman et al., 2015; Vitre and Cleveland, 2012].

Centrosome defects can be numerical or structural, and both have been linked to tumorigenesis [Godinho and Pellman, 2014]. Overduplication of centrosomes, for example by overexpression of the centriolar duplication protein Sak, can lead to the formation of supernumerary centrosomes [Basto et al., 2008]. Failure in normal cell division including the defective cytokinesis and cell-cell fusion can also cause centrosome amplification [Nigg, 2006]. These subsequently lead to the formation of multipolar spindles and multipolar cell divisions. This model, however, is not validated from recent studies in somatic cells [Godinho and Pellman, 2014]. It has been demonstrated that the extra centrosomes often eventually cluster into bipolar spindles and chromosomes segregate relatively normally during mitosis [Gadde and Heald, 2004; Gergely and Basto, 2008; Kwon et al., 2008; Quintyne et al., 2005; Ring et al., 1982]. In addition, cells arising from multipolar divisions often undergo apoptosis [Basto et al., 2006; Ganem et al., 2009; Kwon et al., 2008; Silkworth et al., 2009]. Nonetheless, multipolar intermediates often form before centrosome clustering, and these transient intermediates can show abnormal attachment to the kinetochore and MTs [Gordon et al., 2012]. This promotes the occurrence of lagging chromosomes during anaphase, which in turn increases the frequency of chromosomal mis-segregation [Ganem et al., 2009; Gordon et al., 2012; Kwon et al., 2008; Silkworth et al., 2009]. Tumours can also carry structural centrosome aberrations, although this is much less studied. Such defects may result from misregulation of the expression and/or post-translational modification of centrosomal proteins, which could lead to aberrant cell shape, mobility or polarity [Nigg, 2006]. Alternatively, centrosome defects may interfere with asymmetric division of stem

cells, which is also suggested to be related with tumorigenesis [Basto et al., 2008; Castellanos et al., 2008].

Very recently, another mechanism has been proposed for the link between centrosome amplification and cancer. It has been demonstrated that invasive protrusions are formed from cells with centrosome amplification induced by PLK4 overexpression in 3D cell culture assay [Godinho et al., 2014]. The formation of such protrusions is not directly associated with extra centrosomes, but due to the increased PCM and MT nucleation that activates Rac1, a small GTPase associated with oncogenic signalling and induction of metastasis. The inappropriate activation of Rac1 leads to defects in cell-cell contacts and hence cells with extra centrosomes appear more migratory, which may be important in metastasis.

1.2.2 Centrosomes and microcephaly

Centrosomes are also linked to defects in brain development, one of which is known as autosomal recessive primary microcephaly (MCPH). MCPH patients are characterised by smaller brain sizes with impairment in intellectual development [Faheem et al., 2015; Thorton and Woods, 2009]. However, the architecture of the brain is not largely affected. The patient is often of reduced stature and underweight, and the development of motor functions and speech may be delayed. At least 12 genes are genetically implicated in MCPH (**Table 1.1**), and the majority of reported mutations are loss-of-function mutations [Bond et al., 2002; Bond et al., 2005; Genin et al., 2012; Guernsey et al., 2010; Hussain et al., 2012; Hussain et al., 2013; Jackson et al., 2002; Kumar et al., 2009; Nicholas et al., 2010; O'Driscoll et al., 2006; Sir et al., 2011; Venkatesh and Suresh, 2014; Yu et al., 2010]. Almost all of the MCPH related proteins are associated with centrosomes, suggesting the importance of this organelle in human brain development. Brain size at birth is largely dependent on the ability of neuroprogenitor cells to proliferate and self-renew [Barbelanne and Tsang, 2014; Huttner and Kosodo, 2005; Peyre and Morin, 2012; Thorton and Woods, 2009]. Initially, the symmetric cell division of neuroprogenitor cells lead to an increase in the progenitor pool. This is followed by asymmetric cell division, producing a progenitor cell and a precursor cell that ultimately differentiates [Fish et al., 2008; Gotz and Huttner, 2005]. The latter

process ensures the maintenance of a progenitor pool. The balance between symmetric and asymmetric divisions is thought to be disrupted in the presence of centrosomal abnormalities and hence the mis-orientated MTs; the disrupted division pattern might result in reduced number of progenitor cells and thus lead to the smaller brain size [Chavali et al., 2014; Megraw et al., 2011; Sun and Hevner, 2014]. Moreover, some of the identified MCPH proteins are involved in cell cycle checkpoints and DNA damage response, suggesting DNA damage repair may be impaired in MCPH patients. Indeed mitotic delay and induced cell apoptosis are observed when centrosomes are defective, and both can contribute to the reduced cell numbers [Chavali et al., 2014; Marthiens et al., 2013; Uetake and Sluder, 2010; Yang et al., 2008].

MCPH protein	Function
Microcephalin	Cell cycle control, DNA damage response and DNA repair
WDR62	Spindle pole associated
CDK5RAP2	Key centrosomal protein
CASC5	Kinetochore scaffold protein
ASPM	Spindle pole associated and DNA damage repair
CENPJ/CPAP	Key centriolar protein
STIL	Key centriolar protein
Cep135	Key centriolar protein
Cep152	Key centriolar protein
ZNF335	Nuclear protein
PHC1	DNA damage repair
CDK6	Cell cycle control

Table 1.1 Proteins found to be mutated in MCPH patients.

1.3 The centrosome cycle

Centrosome duplication is coupled with the cell cycle, which contains 4 stages: G1, S, G2 and M (**Figure 1.6**). The first 3 stages constitute interphase, and M phase consists of mitosis and cytokinesis. During G1 phase, also known as the growth phase, cells grow in size to accumulate protein stocks and increase the number of organelles preparing for DNA replication during S phase. Cells can grow more in G2 phase before entering mitosis where chromosome segregation occurs. At the end of mitosis, cytokinesis takes place to split a cell into two daughter cells, each with one nucleus. In differentiated or quiescent cells, a G0 phase is reached after G1. The centrosome is directed away from the centre of a cell, but is often involved in other specialised cellular activities such as cell motility and polarity.

Using electron microscopic analysis, the centrosome cycle has been divided into three main stages: centriole disengagement, centriole duplication and elongation, and centrosome maturation [Bettencourt-Dias and Glover, 2007; Nigg and Stearns, 2011], (**Figure 1.6**). It is well established that centrosome duplication starts in S phase and completes before mitosis such that two centriole pairs can form the bipolar spindle poles. The tightly associated centriole pair within each spindle pole is disengaged at the end of mitosis. This disengagement appears to play a role in licensing the centriole for the subsequent cycle of duplication [Novak et al., 2014; Tsou and Stearns, 2006]. During G1, the amount of PCM is reduced, and the daughter centriole starts to mature into a mother. At this stage, both centrioles often still remain loosely attached via flexible linkers at proximal ends [Bornens et al., 1987]. At the transition from G1 to S phase, a procentriole is formed adjacent to the pre-existing mother centriole, which elongates to its full length in G2 phase. At the G2 to M transition, the linkers between two centrioles are lost, and subsequently the two centrosomes start to separate [Piel et al., 2000]. Each centrosome extensively recruits PCM in a process known as centrosome maturation [Palazzo et al., 2000]. It is thought that the recruitment of PCM components from cytoplasmic stores facilitates the further binding of γ-tubulin and other protein complexes. Consequently, the centrosome increases in its MT-nucleating capacity [Palazzo et al., 2000] hence preparing itself for entering M phase. During mitosis, each centrosome migrates to the opposite side of the nucleus, and ultimately forms the bipolar poles of the mitotic spindles.

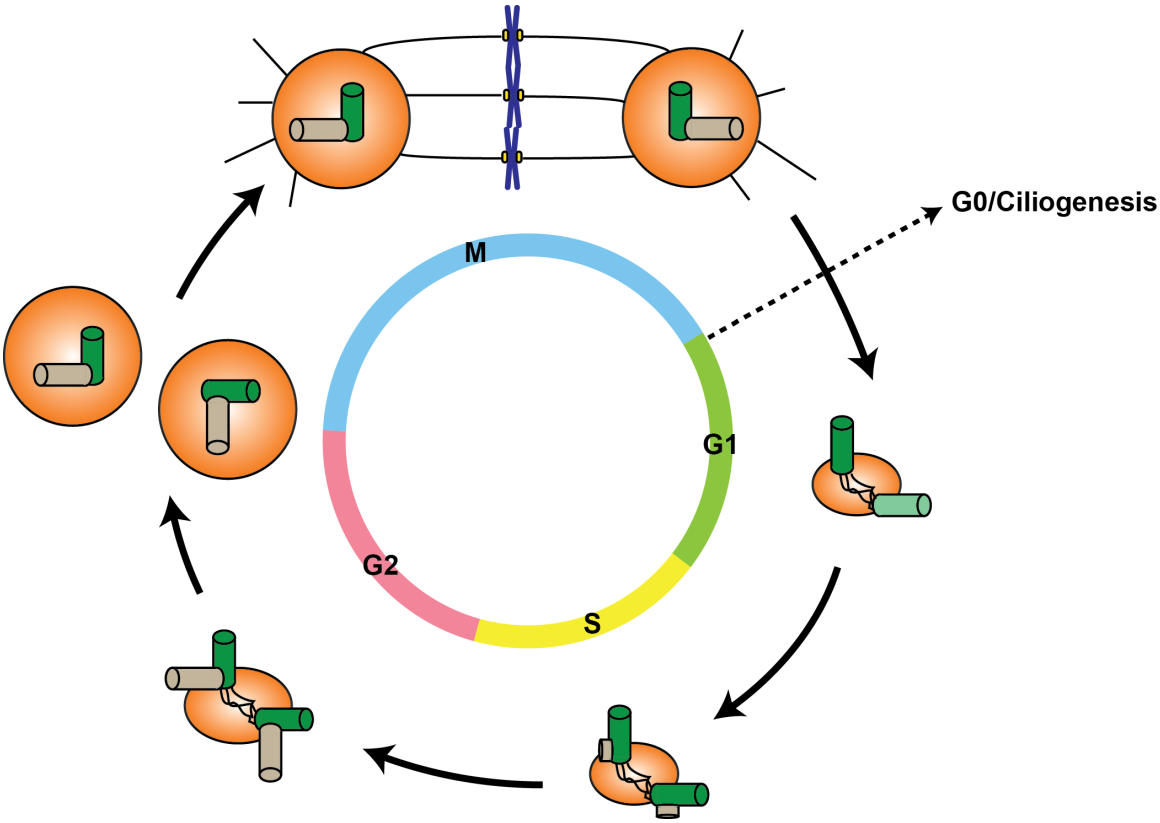


Figure 1.6 Schematic representation of the centrosome cycle. In G1 phase, a pair of disengaged centrioles is present, composed of a fully matured mother centriole and a daughter centriole that starts to mature. A fibrous linker is present to connect the proximal ends of two centrioles, and it is ultimately lost at the onset of M phase. At the G1 to S transition, both centrioles act as mothers, and start to template the formation of procentrioles. The newly formed centrioles continue to grow to their full lengths during G2. At the G2 to M transition, each centrosome recruits extensive amounts of PCM, and the two centrosomes separate in preparation to enter mitosis. During mitosis, the two separate centrosomes migrate to opposite sides of the nucleus forming biopolar spindle poles. At the end of mitosis, each centrosome is segregated into a single daughter cell. Non-dividing cells enter G0 phase where centrioles form the basal body of cilia and flagella. Diagram modified from [Cottee, 2014].

1.4 Centriole biogenesis

1.4.1 Centriole structure

A canonical centriole is ~450 nm long and 250 nm wide, which comprises a nine-fold symmetric array of MT triplets organised around a central cartwheel at the proximal end [Gonczy, 2012; Jana et al., 2014], (**Figure 1.7A**). Although MT triplets are normally observed in vertebrates and several unicellular organisms such as *Chlamydomonas reinhardtii*, MT doublets are detected in most *Drosophila* cells (**Figure 1.7B**). Each MT triplet contains an A-tubule, B-tubule and C-tubule [Guichard et al., 2010; Guichard et al., 2013; Li et al., 2012], whereas in *Drosophila* somatic cells the C-tubule is missing [Callaini et al., 1997]. Only the A-tubule is complete and contains 13 protofilaments (as seen in a canonical MT), while the other two are incomplete (**Figure 1.7A**). In addition, a linker is formed between the A-tubule of one triplet and the C-tubule of the next triplet, hence being referred to as A-C linker. Interestingly, such linker is not observed between MT doublets in *Drosophila* somatic cells and alternatively, the space appears to be filled in by an electron dense material [Callaini et al., 1997], (**Figure 1.7B**). In a mature mother centriole, the A- and B-tubules often extend further than the C-tubule, leaving MT doublets organised in 9-fold symmetry at the distal end (**Figure 1.7A**). This region is where the subdistal and distal appendages are acquired, which are important for plasma membrane anchoring and nucleation of the MT cytoskeleton within the cilium [Azimzadeh and Marshall, 2010; Bettencourt-Dias and Glover, 2007; Bornens, 2012]. However, the mother and daughter centrioles in *Drosophila* are morphologically indistinguishable due to the absence of appendages (**Figure 1.7B**). Note that *Drosophila* centrioles tend to be shorter than vertebrate centrioles.

The MT doublets or triplets are organised around a central cartwheel in centrioles. Structural features of this cartwheel have started to emerge from cryo-electron tomography (cryo-ET) studies of the extended cartwheel structures found in *Trichonymopha*, a parasite that lives in the gut of termites [Guichard et al., 2012; Guichard et al., 2013]. The cartwheel contains a central hub of 22 nm in diameter, and nine long, extended spokes radiate out from the hub connecting to the MT blades that are also organised in a nine-fold symmetry. Several layers of cartwheels stack on top of each other at the proximal region (**Figure 1.8A**). Intriguingly, the radial spokes are spaced at ~8.5 nm periodicity, and at 20 nm from the central hub the two spokes are combined into a

single layer hence leaving a 17 nm gap between the two layers (**Figure 1.8B**). This nearly 20 nm gap between the cartwheel stacks is also observed in other systems using conventional microscopy, suggesting the merging density between adjacent spokes seen in *Trichonymopha* might be a common feature of the cartwheel [Geimer and Melkonian, 2004; Hirono, 2014].

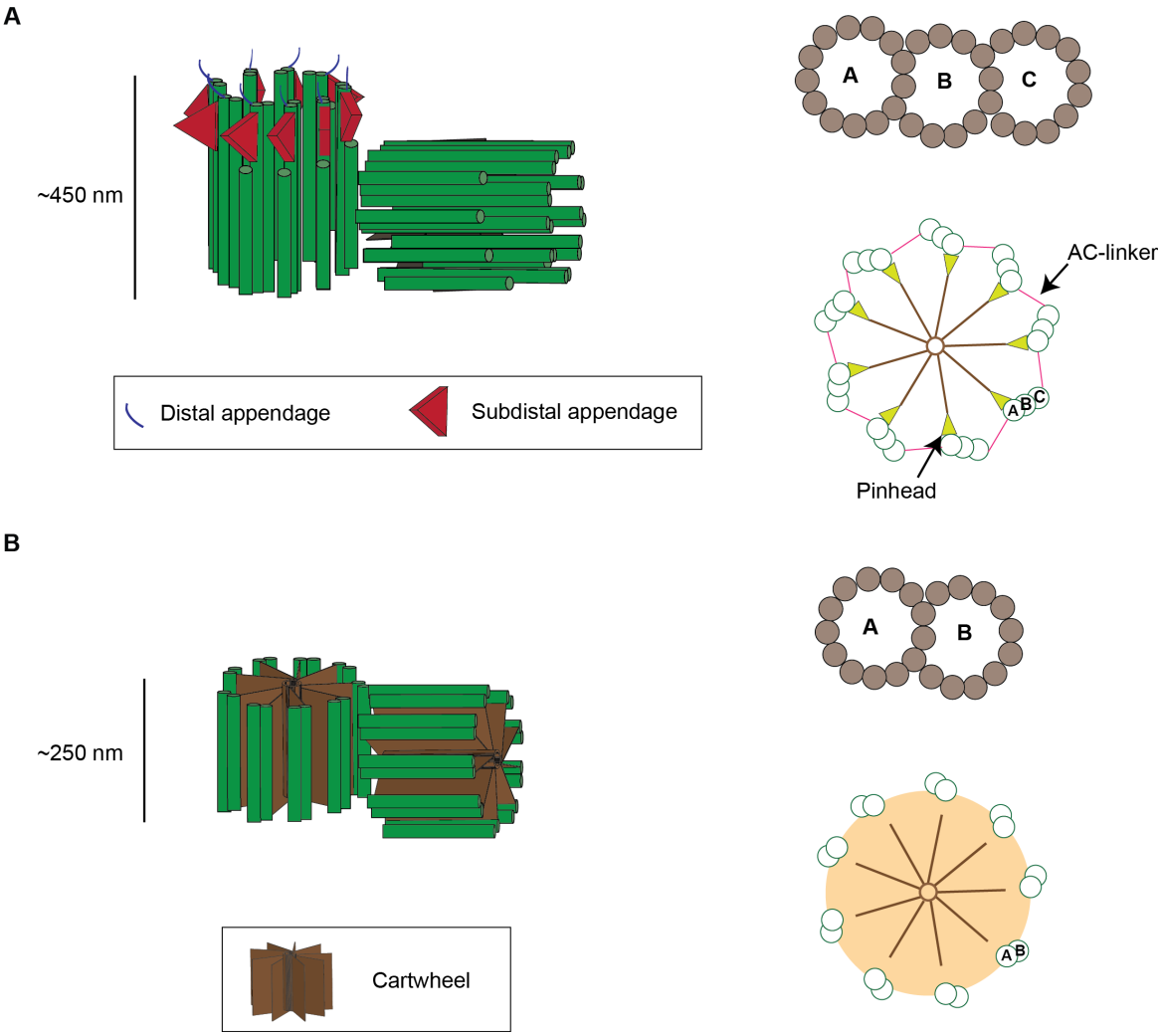


Figure 1.7 Structure of the centriole, by schematic representations. (A) Left, diagram showing the detailed structure of vertebrate centrioles. Right above, protofilament arrangement in the MT triplet according to the EM studies carried out by Guichard et al., 2013. Right bottom, the centriole is viewed proximal to distal. A-C linker is formed between A- and C-tubules. The pinhead connects the cartwheel spoke to the MT blades. Note the fully matured mother centriole does not contain a cartwheel structure at its proximal end, whereas the daughter centriole does. **(B)** Schematics of *Drosophila* centrioles that both contain a cartwheel structure and are morphologically indistinguishable. MT doublets are observed, lacking the C-tubule. No A-C linker is found, and the space around MT doublets is filled up with an electron dense material. All the schematics are not drawn to scale.

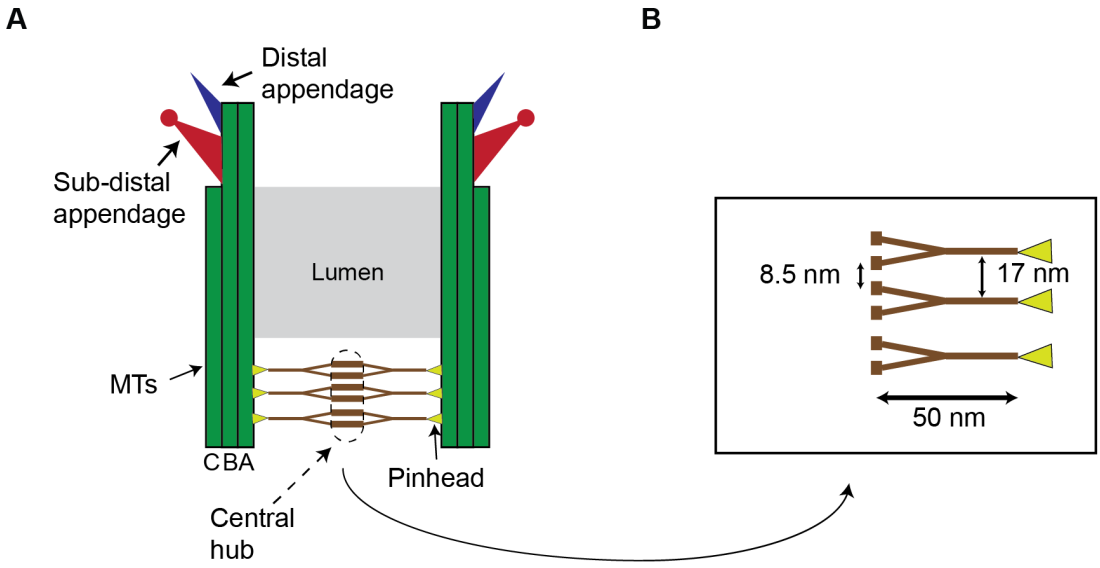


Figure 1.8 Schematics of cartwheel stacks in a canonical centriole, as seen in vertebrates and several unicellular organisms. Model is drawn according to the cryo-electron tomography analysis of *Trichonympha* basal body [Guichard et al., 2012; Guichard et al., 2013]. (**A**) Side-on view of the stacked cartwheels located at proximal end of a centriole. (**B**) Focussed view of the cartwheel spoke showing an 8.5 nm separation between the central hubs (left hand side). Spokes in adjacent layers of the cartwheel merge at ~20 nm from the hub, resulting in a 17nm vertical periodicity. The distance from the edge of the central hub to the pinhead is ~50 nm.

X-ray crystal structure of SAS-6 protein, a core structural component of the central cartwheel [Culver et al., 2009; Gopalakrishnan et al., 2010; Kleylein-Sohn et al., 2007; Nakazawa et al., 2007], has been solved for its homologs in several species, which all showed an evolutionarily conserved N-terminal globular domain followed by an extended coiled-coil (CC) [Cottee et al., 2015; Hilbert et al., 2013; Kitagawa et al., 2011; Qiao et al., 2012; van Breugel et al., 2011; van Breugel et al., 2014]. Full-length SAS-6 dimerises via its CC domain, and the dimers further associate into oligomers via hydrophobic interaction between the N-terminal head groups. Such oligomers are observed to form ring-like structure of 22 nm diameter with 9-fold symmetry *in vitro* [Kitagawa et al., 2011; van Breugel et al., 2011]. This is consistent with the *in vivo* observation via cryo-ET. The radial spokes are connected to the MT triplets via a complex structure called the pinhead [Guichard et al., 2013], (**Figure 1.8**). Note *Caenorhadbitis elegans* centrioles do not contain cartwheel hub or spokes, but alternatively have a central tube formed via spiral arrangement of CeSAS-6 while retaining the nine-fold symmetry consistent with other systems [Hilbert et al., 2013]. The nine-fold symmetry of the cartwheel is not only determined by the nine-fold symmetric SAS-6 oligomer, but Bld10/Cep135 protein also appears to be important for establishing the symmetry at least in some organisms [Hirono, 2014]. Bld10 protein localises at the region of the spoke close to the MT triplet, and its null mutation has led to the complete loss of

centriolar MTs in *Chlamydomonas* [Matsuura et al., 2004]. When a truncated form of Bld10 protein was expressed in a *bld10* null mutant, the cartwheel contained spokes of shorter lengths, and hence its diameter was overall smaller leading to the assembly of eight MT triplets rather than nine [Hiraki et al., 2007]. However, *Drosophila* Cep135 protein (homolog of Bld10) appears not essential for cartwheel formation, although cartwheel stability is affected in the absence of Cep135 [Roque et al., 2012]. Similarly in human cells, the Cep135 depletion also leads to morphologically perturbed cartwheel formation with aberrant number of MT triplets [Lin et al., 2013]. Cep135 has been shown to directly interact with both SAS-6 and MTs, and is essential for centriole elongation [Lin et al., 2013]. Its MT binding activity is also observed with *Drosophila* Cep135, suggesting a common role of this protein in connecting the inner and outer cartwheel components [Carvalho-Santos et al., 2010].

Apart from SAS-6 and Cep135/Bld10, other protein components may also link the cartwheel spoke and the MTs. For instance, SAS-4/CPAP provides such a molecular linker [Jana et al., 2014], and it localises at the inner centriolar wall showing a ring-like distribution in *Drosophila* and human cells [Fu and Glover, 2012; Lawo et al., 2012; Mennella et al., 2012]. This protein can bind to MTs and Bld10/Cep135 [Lin et al., 2013], and is essential for centriole duplication. SAS-4/CPAP also interacts with Ana2/STIL/SAS-5, which is another key component of the centriole duplication pathway [Cottee et al., 2013; Hatzopoulos et al., 2013; Tang et al., 2011]. The latter protein has been shown to interact with Sas6 [Dzhindzhev et al., 2014; Hilbert et al., 2013; Kratz et al., 2015; Ohta et al., 2014; Qiao et al., 2012], and both proteins depend on each other for centriolar recruitment [Delattre et al., 2006; Pelletier et al., 2006]. In addition, the overexpression of Ana2 and SAS-6 leads to the formation of cartwheel-like structures in *Drosophila* spermatocytes [Stevens et al., 2010].

1.4.2 Centriole assembly

In recent years, substantial progresses have been made to identify proteins associated with centrioles via proteomics analysis, mass spectrometry analysis and comparative genomic studies in a wide range of species [Andersen et al., 2003; Keller et al., 2005; Li et al., 2004]. Through

comparison of proteins identified in different organisms, a core set of proteins is determined to be essential for centriole duplication. The early genome-wide RNAi screen carried out in *C. elegans* has defined 5 key proteins: SPD-2, ZYG-1, SAS-6, SAS-5 and SAS-4 [Delattre et al., 2006; Pelletier et al., 2006]. Latter studies in *Drosophila* and human cells have confirmed the conserved significance of these proteins in centriole duplication [Balestra et al., 2013; Dobbelaere et al., 2008]. Nonetheless, in flies, Spd-2 (Spindle-defective protein 2) is not required for centriole assembly and instead, it is only essential for PCM recruitment. Centriole assembly has been discussed extensively in several recent reviews [Firat-Karalar and Stearns, 2014; Gonczy, 2012; Hirono, 2014; Jana et al., 2014], and I will therefore briefly summarise this process here.

ZYG-1 is a functional homolog of PLK4 and Sak in human and flies, respectively [Bettencourt-Dias et al., 2005; Habedanck et al., 2005; O'Connell et al., 2001]. This kinase appears as a master regulator in the centriole assembly pathway [Kleylein-Sohn et al., 2007; Pelletier et al., 2006]. Its recruitment depends on Asterless (Asl) in *Drosophila* [Blachon et al., 2008; Dzhindzhev et al., 2010; Klebba et al., 2015] and on SPD-2 in *C. elegans* [Kemp et al., 2004; Pelletier et al., 2004]. In human cells, however, homologs of both proteins can interact with PLK4 and are involved in a cooperative recruitment of the kinase to the centriole [Cizmecioglu et al., 2010; Hatch et al., 2010; Kim et al., 2013; Sonnen et al., 2013]. PLK4 contains a conserved N-terminal kinase domain, followed by three polo box (PB) domains. The first two PB domains are known as the cryptic polo box (CPB), and are required for interaction with Asl/Cep152 and SPD-2/Cep192. Recently, X-ray crystal structures have been solved for the CPB of PLK4/Sak/ZYG-1 in different systems [Shimanovskaya et al., 2014; Slevin et al., 2012], and the crystal structure of protein complexes containing PLK4 CPB and the acidic N-terminal region of Cep152 (human homolog of Asl) or Cep192 (human homolog of Spd-2) has also been solved revealing the protein binding interface at atomic resolution [Park et al., 2014]. In *C. elegans*, direct interaction is observed between ZYG-1 and SAS-6, and SAS-6 also interacts with SAS-5. Both interactions are required for the centriolar recruitment of SAS-6 [Lettman et al., 2013]. Recently, three independent studies in *Drosophila* and human cells have consistently shown that Ana2/STIL is phosphorylated by Sak/PLK4 at the centriole within the highly conserved STIL/Ana2 (STAN) domain [Dzhindzhev et al., 2014; Kratz et al., 2015; Ohta et al., 2014]. The phosphorylated STAN domain in turn binds SAS-6 protein for its recruitment to the newborn centriole. SAS-6 then oligomerises to form the

central cartwheel structure. The atomic structures of Ana2/SAS-5 oligomerisation domains have also been characterised very recently [Cottee et al., 2015; Rogala et al., 2015]. In both *Drosophila* and *C. elegans*, mutations at the oligomerisation interface have led to significantly reduced localisation of Ana2/SAS-5 at centrioles, suggesting self-interaction is essential for centriole assembly. The localised SAS-6 and SAS-5/Ana2/STIL proteins subsequently incorporate SAS-4/CPAP to the outer region of the cartwheel to help assemble the centriolar MTs [Pelletier et al., 2006; Tang et al., 2009; Tang et al., 2011]. In flies, SAS-4 then recruits Asl to the newly formed daughter centriole, which then becomes a new mother for the next cell cycle. This recruitment only occurs when the mother and daughter centrioles are disengaged, and subsequently provides a reduplication licensing for the new mother centriole [Dzhindzhev et al., 2010; Novak et al., 2014]. In addition, the new mother centriole only recruits PCM once separated from the old mother centriole in the next cell cycle.

1.5 Centrosome maturation

From G1 to G2 centrioles often organise only small amounts of PCM, and this appears to be organised exclusively by the mother centriole [Fu and Glover, 2012; Lawo et al., 2012; Mennella et al., 2012; Sir et al., 2011; Sonnen et al., 2012]. As cells prepare to enter mitosis, the amount of PCM recruited around the mother centriole significantly increases in a process known as centrosome maturation [Conduit et al., 2015]. Due to the accumulation of γ-tubulin and other PCM components, the MT nucleating capacity of centrosomes also increases, demonstrated by the increased number of MTs emanated from the PCM [Palazzo et al., 2000; Piel et al., 2004]. Phosphorylation has long been recognised to play an important part in centrosome maturation. Evidence come from several early studies including an increased level of phosphorylation detected at the mitotic centrosome [Vandre et al., 1991] and the observation that the addition of protein kinase inhibitors to an interphase *Xenopus* egg extract incubated with isolated centrosomes blocked PCM accumulation and the ability to nucleate MTs [Ohta et al., 1993; Verde et al., 1991].

Two mitotic protein kinases, PLK1/Polo and Aurora-A, play a particularly important part during centrosome maturation. The kinase activity of both proteins increases during the G2/M transition and peaks at mitosis [Bruinsma et al., 2012; Daub et al., 2008], which can be achieved either via primed phosphorylation in the kinase domain or by the suppression of its inhibitors [Archambault and Glover, 2009; Zitouni et al., 2014].

The importance of Polo kinase in centrosome maturation was first demonstrated by the *polo* mutation in *Drosophila melanogaster*, which showed the presence of monopolar spindles [Sunkel and Glover, 1988]. This was also observed in human cells (HeLa and Hs68 cell lines) after the injection of anti-PLK1 antibodies and in the *S. pombe plo1* mutant (a Polo homolog in fission yeast), [Archambault and Glover, 2009; Lane and Nigg, 1996; Ohkura et al., 1995]. Several PLK1/Polo phosphorylation substrates have been identified. For instance, in human cells PLK1 specifically phosphorylates Pericentrin, a key PCM protein in vertebrates, during mitosis. The Pericentrin phosphomutant (carrying point mutations resistant to PLK1 phosphorylation) could not recruit other PCM proteins, such as Cep192 and γ-tubulin, into the centrosome and hence failed

in centrosome maturation [Lee and Rhee, 2011]. In addition, expression of a PLK1-Pericentrin fusion protein in human cells can promote the accumulation of Cep192 and γ-tubulin even at the interphase centrosome [Lee and Rhee, 2011]. Centrosomin (Cnn), another essential PCM protein required for the recruitment of many other PCM components (also see **Section 1.6.3**), is specifically phosphorylated at the centrosome in a Polo-dependent manner during mitosis [Dobbelaere et al., 2008]. In S2 cells, the depletion of Polo led to a dramatic reduction in the amount of Cnn and γ-tubulin recruited into the centrosome [Dobbelaere et al., 2008]. Subsequently, the recruitment of other PCM proteins was also severely disrupted.

Aurora-A kinase also plays an important part in centrosome maturation. In vertebrate cells, Aurora-A is involved in AurA/PLK1 phosphorylation cascade involving a crucial PCM protein Cep192, a human homolog of Spd-2 (will be discussed in more details in **Section 1.6.4**). This cascade is important for the recruitment of phosphorylated and thus activated γ-TuRC and subsequently the assembly of mitotic MTs [Joukov et al., 2014]. Aurora-A kinase and its co-activator, Bora, are required for the efficient centrosomal recruitment of Cnn during mitosis in *Drosophila* sensory organ precursor cells [Berdnik and Knoblich, 2002; Terada et al., 2003]. A direct interaction was mapped between the kinase and Cnn in S2 cells [Terada et al., 2003]. Later studies, however, have revealed that Aurora-A and Bora together mediate the activation of PLK1/Polo kinase at the onset of mitosis [Bruinsma et al., 2014; Jang et al., 2002; Seki et al., 2008]. Therefore, it is possible that the diminished Cnn localisation observed in the absence of Aurora-A kinase was due to a lower Polo activity. In addition, a genome-wide RNAi screen conducted in S2 cells has demonstrated that in the absence of Aurora-A kinase the amount of PCM localised at the centrosome was reduced, but not as severe as in the absence of Polo kinase [Dobbelaere et al., 2008].

1.6 PCM structure and composition

1.6.1 PCM organisation in interphase

For decades, the PCM has been referred to as an amorphous matrix of proteins surrounding the pair of orthogonally arranged centrioles. This was due to the lack of architectural details seen within the PCM when viewed via different methodologies including conventional electron microscopy and electron tomography [Guichard et al., 2010; Ibrahim et al., 2009; Moritz et al., 1995a], but also to the fact that most proteins within the PCM are apparently not very ordered [Dos Santos et al., 2013]. In addition, the PCM organisation differs depending on the cell cycle stage i.e. interphase or mitosis. This is because interphase centrioles recruit much less PCM than mitotic centrioles.

Recently, four independent studies, using super resolution microscopy techniques such as three-dimensional structured illumination (3D-SIM) and stochastic optical reconstruction microscopy (STORM) in *Drosophila* and human cells, have revealed the appearance of highly ordered subdomains within the PCM in interphase [Fu and Glover, 2012; Lawo et al., 2012; Mennella et al., 2012; Sonnen et al., 2012]. The PCM can be partitioned into two regions: a proximal layer occupied by Pericentrin/PLP and Cep152/Asl, is localised close to the centriole wall; and a distal layer containing CDK5RAP2/Cnn, Cep192/Spd-2 and γ-tubulin, that shows a matrix-like distribution extending further into the outer PCM region. In addition, using antibodies raised against non-overlapping regions one could detect specific orientations for proteins within the proximal layer [Lawo et al., 2012; Mennella et al., 2012]. Intriguingly, the C-terminus of Pericentrin/PLP, which contains a PACT domain required for centrosomal targeting, localises close to the centriole wall, whereas its N-terminal region expands into the second PCM layer [Lawo et al., 2012; Mennella et al., 2012]. In *Drosophila*, PLP forms molecular fibrils that recruit and organise the interphase PCM, at least in S2 cells [Mennella et al., 2012].

1.6.2 PCM organisation in mitosis

As mentioned above, the amount of PCM recruited to centrioles increases dramatically during mitosis in a process known as centrosome maturation. It has long been thought that a subset of

key proteins may scaffold the assembly of other PCM components during mitosis, and the PCM expands largely to form a fully nucleating mitotic centrosome [Conduit et al., 2015]. This is supported by the observation that the ability of the centrosome to nucleate MTs can be removed from the isolated centrosome with the treatment of urea, proteases and high salt. For example, isolated centrosomes purified from the surf clam *Spisula solidissima* were treated with 1 M potassium iodide, after which many PCM components including γ-tubulin and the ring structures of γ-TuRC were stripped away [Schnackenberg et al., 1998]. This, however, left an intact centriole surrounded by a matrix-like network built from 12-15 nm wide filaments, termed the "centromatrix". On its own, the centromatrix was unable to nucleate MTs when incubated with purified tubulin. After incubation in a mitotic extract, however, the centromatrix rebound PCM proteins, such as the γ-TuRC, and this restored the MT nucleation activity. Therefore, it was proposed that the centromatrix serves as an underlying scaffold onto which other elements, such as γ-TuRC, are recruited from the cytoplasm [Conduit et al., 2015; Woodruff et al., 2014]. Similar observations of a salt-resistant centrosomal scaffold were also made in *Drosophila* [Moritz et al., 1995b; Moritz et al., 1998].

Studies have been carried out in order to identify and characterise the proteins that build such a structural scaffold within the mitotic PCM. In vertebrate cells, Pericentrin is important for the mitotic PCM recruitment, and it interacts with other key PCM components such as CDK5RAP2 [Buchman et al., 2010; Chen et al., 2014; Delaval and Doxsey, 2010; Kim and Rhee, 2014]. Pericentrin, in human cells, undergoes spatial re-organisation from the toroid-like to a matrix-like pattern [Lawo et al., 2012; Mennella et al., 2012; Sonnen et al., 2012]. *Drosophila* PLP, however, does not undergo such mitotic expansion and it remains tightly associated with the centriole wall [Fu and Glover, 2012; Mennella et al., 2012]. This might explain the less significant role of PLP during mitosis in *Drosophila* [Lerit et al., 2015; Martinez-Campos et al., 2004; Richens et al., 2015]. Instead, two other core centrosomal proteins, Cnn and Spd-2, are essential for the PCM recruitment during mitosis in flies [Conduit et al., 2014b; Hamill et al., 2002; Lucas and Raff, 2007; Megraw et al., 1999; Woodruff et al., 2015]. Interestingly, these two proteins (SPD5, a functional ortholog of Cnn, and SPD2, a homologue of Spd-2) are also important for the mitotic PCM organisation in worms. The recent super resolution microscopy studies have shown that upon mitotic entry, CDK5RAP2/Cnn and Cep192/Spd-2 undergo considerable expansion in the PCM. A

proportion of Cep192/Spd-2, however, associates with the centriole wall overlapping with the distribution of Asl N-terminal antibody [Fu and Glover, 2012; Lawo et al., 2012; Mennella et al., 2012]. In addition, *Drosophila* Spd-2 does not spread as far out into the peripheral PCM region as Cnn, and both proteins show no directionality within the mitotic PCM [Lawo et al., 2012; Mennella et al., 2012]. In flies, Asl may provide a molecular platform linking centriole formation and PCM recruitment [Conduit et al., 2010; Dzhindzhev et al., 2010; Novak et al., 2014]. I will therefore introduce each of the major protein families that appear to be essential for the PCM assembly during both interphase and mitosis (see **Sections 1.6.3-1.6.5**).

1.6.3 Centrosomin (Cnn) and its related family proteins

Drosophila Cnn is a 150 kDa centrosomal protein, and related proteins are found in most eukaryotes from the fission yeast *S. pombe* to humans [Megraw et al., 1999; Sawin et al., 2004; Verde et al., 2001; Wang et al., 2000]. Recently, *C. elegans* SPD-5 has been suggested to be a functional ortholog of Cnn, although it lacks detectable sequence conservation [Hamill et al., 2002].

Drosophila Cnn is required for the efficient recruitment of many other PCM proteins including D-TACC, Polo, Aurora-A, γ-tubulin, CP60 and CP190 [Dobbelaere et al., 2008; Lucas and Raff, 2007; Megraw et al., 1999; Megraw et al., 2001; Megraw et al., 2002; Vaizel-Ohayon and Schejter, 1999; Zhang and Megraw, 2007]. In both embryos and somatic cells, Cnn is required for the efficient assembly of the mitotic centrosome and subsequently a functional MTOC at the spindle poles [Mahoney et al., 2006; Megraw et al., 1999; Megraw et al., 2001; Moutinho-Pereira et al., 2013; Vaizel-Ohayon and Schejter, 1999]. In $cnn^{-/-}$ mutant embryos, γ-tubulin was mislocalised at interphase, and was missing from the spindle poles during mitosis [Vaizel-Ohayon and Schejter, 1999]. In addition, embryonic development arrested before cellularisation with major mitotic defects due to nuclear fusion and nuclear fall-out as well as the failure to organise pseudocleavage furrows [Megraw et al., 1999; Vaizel-Ohayon and Schejter, 1999]. Nonetheless, depletion of Cnn led to only relatively subtle mitotic defects in somatic cells [Megraw et al., 2001]. This is expected since flies can develop relatively normally without centrosomes in their somatic

cells [Basto et al., 2006]. In particular, asymmetric cell divisions were slightly defective in *cnn*$^{-/-}$ mutant flies. In *cnn*$^{-/-}$ neuroblasts, metaphase spindle mis-orientation occurred ~22% of the time [Lucas and Raff, 2007; Megraw et al., 2001]. Although 20-30% of brain cells had abnormal numbers of centrioles (other than 2), ~95% of the NB cells ultimately divided asymmetrically since spindle orientation was corrected in late mitosis ("telophase rescue"), [Lucas and Raff, 2007]. Cnn depletion had a more severe effect in male GSC asymmetric division [Yamashita et al., 2007]. The centrosomes were randomly positioned in male GSCs, and consequently the spindles were not properly oriented towards the hub that contains the stem cell niche. As a result, the stem cell number was greatly increased by 20% to 30% [Yamashita et al., 2007]. Surprisingly, the asymmetric division of male GSCs does not appear to be dramatically perturbed in the absence of centrosomes [RIparbelli and Callaini, 2011].

Moreover, in *cnn*$^{-/-}$ embryos, PCM components, although recruited, could not be maintained properly around the centrioles. As a result, centrioles moved randomly in the embryo, leaving a trail of PCM behind them, and frequently detached from the nuclear envelope in interphase and from the spindle poles in mitosis [Lucas and Raff, 2007]. This phenotype is referred to as "centriole rocketing". In addition, this rocketing was MT-dependent. As the MTs were depolymerised following colchicine treatment, centrioles retained their connection to the PCM although they were still not localised at the center of the PCM [Lucas and Raff, 2007].

In fly embryos, Cnn exhibits an unusual dynamic behavior where large particles of Cnn release from centrosomes, so-called "flares", and these particles contain D-TACC, Msp and PLP [Lerit et al., 2015; Megraw et al., 2002]. The back and forth movement of flare particles is MT-dependent; and the amount of flaring alters with the cell cycle where the highest intensity is observed at interphase centrosomes and lowest at mitotic centrosomes [Lerit et al., 2015; Megraw et al., 2002].

In humans, there are two homologs of Cnn, Myomegalin [Verde et al., 2001] and CDK5RAP2/Cep215 [Fong et al., 2008; Megraw et al., 2011]. Myomegalin is highly expressed in muscle cells, whereas CDK5RAP2 is expressed in many other cell types. CDK5RAP2 mutations in human cause microcephaly, but the underlying mechanism is still poorly understood [Bond et

al., 2005; Buchman et al., 2010; Faheem et al., 2015; Lizarraga et al., 2010; Megraw et al., 2011]. The function of Myomegalin is not well characterised, although the protein is implicated in anchoring components of the cAMP-dependent pathway to the Golgi and to centrosomes [Verde et al., 2001]. Recently, this protein has also been implicated in the formation of MTs nucleated from centrosomes and Golgi [Graser et al., 2007; Roubin et al., 2013].

CDK5RAP2 associates with GCP3, a γTuSC component, and γ-tubulin in order to attach the γTuRC to centrosomes and to stimulate its microtubule nucleating activity [Choi et al., 2010; Fong et al., 2008]. In interphase and mitotic cells depleted of CDK5RAP2, the amount of γ-tubulin was significantly reduced at the centrosome and its distribution was more dispersed into the cytoplasm [Fong et al., 2008; Lee and Rhee, 2010]. Overexpression of CDK5RAP2 led to an accumulation of γ-tubulin and Pericentrin at the centrosome, and also to the formation of protein aggregates in the cytoplasm, which associated with MTs [Fong et al., 2008]. The region of CDK5RAP2 for interaction with γ-tubulin is mapped to a highly conserved domain denoted as Common Motif 1 (CM1) that is identified across different species [Choi et al., 2010; Fong et al., 2008; Zhang and Megraw, 2007], and the CM1 domain in *Drosophila* Cnn has also been demonstrated to immunoprecipitate γ-tubulin [Terada et al., 2003]. Expression of this motif in U2OS cells resulted in the removal of centrosomal γ-tubulin and the impaired MT nucleation and organisation ability of the centrosome, in both interphase and mitosis [Choi et al., 2010; Fong et al., 2008]. Very recently, an LRRK-mediated phosphorylation site has been mapped close to the CM1 domain, and phosphorylation was demonstrated to promote CDK5RAP2-dependent MT nucleation during mitosis [Hanafusa et al., 2015]. This kinase (LRRK) is activated via PLK1 and CDK1 phosphorylation. Nevertheless, the importance of the CM1 domain in MT nucleation is contradictory. Deletion of this domain (*cnn1*$^{-/-}$ mutant in the chicken cell line DT40) showed no detectable mislocalisation of γ-tubulin, and centrosomes retained the capacity to nucleate MTs at a similar level to the wild type [Barr et al., 2010]. Intriguingly, centriole detachments were frequently observed in *cnn1*$^{-/-}$ mutants likely due to the mislocalisation of a downstream PCM protein AKAP450, which interacts with the p150^{glued} subunit of dynactin to maintain a stable link between the centrosome and spindle pole [Barr et al., 2010; Lee and Rhee, 2010]. CDK5RAP2 has also been implicated in maintaining cohesion between the centrosomes until the late G2 phase, so centrosomes can separate in prophase and generate a bipolar spindle for correct

chromosome segregation during mitosis [Barr and Gergely, 2007; Barrera et al., 2010; Fry, 2015; Pagan et al., 2015].

Drosophila Cnn contains two evolutionarily conserved motifs, CM1 close to the N-terminus and CM2 at the extreme C-terminus [Zhang and Megraw, 2007]. As mentioned above, the CM1 domain is present in all homolog from yeast to human. The CM2 domain, however, appears conserved only among metazoan homologs. A CM1 deletion mutant showed reduced γ-tubulin recruitment to the centrosome in fly embryos [Zhang and Megraw, 2007]. Centrosomes failed to separate after duplication, which is likely due to the impaired MT assembly and consequently the insufficient pushing forces [Zhang and Megraw, 2007]. The CM2 domain was shown to interact with Centrocortin, and a cnn^{B4} mutation (Arg1141 mutated to His) that knocked out the Cnn-Centrocortin interaction led to embryonic arrest prior to cellularisation due to failure in cleavage furrow assembly [Kao and Megraw, 2009]. Very recently, another role of Arg1141 was identified as the cnn^{B4} mutant also failed to interact with PLP in *Drosophila* [Lerit et al., 2015]. The Cnn-PLP interaction is involved in maintaining the scaffold strength at the outermost region of Cnn at the PCM [Lerit et al., 2015; Richens et al., 2015].

The CM2 domain has been implicated in CDK5RAP2 recruitment to the centrosome in human cells [Kim and Rhee, 2014; Wang et al., 2010]. In HeLa cells, a C-terminal fragment of CDK5RAP2 containing the CM2 region showed strong localisation to the centrosomes when it was ectopically expressed [Wang et al., 2010]. In addition, a calmodulin (CaM) binding domain was identified within CM2, and deletion of this domain led to a diffused localisation pattern in the cytoplasm [Wang et al., 2010]. However, centrosomal targeting was not dependent on CaM binding as the point mutants (that lost the ability to bind CaM *in vitro*) displayed normal localisation at the centrosomes. Immunoprecipitation experiments successfully pulled down Pericentrin using CM2 WT fragment but not the CaM domain mutant [Wang et al., 2010]. In addition, depletion of Pericentrin blocked the localisation of CDK5RAP2 to the centrosome [Buchman et al., 2010; Graser et al., 2007; Haren et al., 2009; Kim and Rhee, 2014; Wang et al., 2010]. These results suggested that CDK5RAP2 requires its interaction with Pericentrin for centrosomal localisation, likely via the CM2 domain. Nonetheless, recent studies indicated another dynein-dependent recruitment pathway for CDK5RAP2 in human cells [Jia et al., 2013;

Lee and Rhee, 2010]. In flies, the recruitment of Cnn to centrosomes is MT-independent, since FRAP analysis showed normal signal recovery before and after the colchicine treatment [Conduit et al., 2014a]. The role of the CM2 domain in *Drosophila* Cnn has not been directly characterised. However, the *mfs* mutants of Cnn, which have truncated deletions of various lengths at the C-terminus (the shortest deletion included the last 300 amino acids), failed to localise to the centrosome [Megraw et al., 1999]. It would therefore be interesting to investigate if the role of the CM2 domain in centrosomal localisation is conserved through evolution. In addition, it is unlikely that PLP plays an essential role in centrosomal targeting of Cnn since antibody injection of anti-PLP into *cnn*$^{-/-}$ mutant embryos expressing GFP-Cnn did not block its localisation at the centrosome [Conduit et al., 2010], and the amount of GFP-Cnn localisation at centrosomes was not abolished, although was reduced, in the absence of PLP [Lerit et al., 2015; Richens et al., 2015]. Instead, Cnn recruitment was rapidly decreased in the presence of anti-Asl and/or anti-Spd-2 antibody [Conduit et al., 2010].

1.6.4 Spindle defective protein-2 (Spd-2) and its related proteins

Spd-2 is another key conserved centrosomal protein. This protein was originally identified in *C. elegans* where it is involved in both centriole duplication and centrosome maturation [Kemp et al., 2004; Leidel and Gonczy, 2005; O'Connell et al., 2000; Pelletier et al., 2004; Pelletier et al., 2006], and this also seems to be the case in human cells [Gomez-Ferreria et al., 2007; Zhu et al., 2008]. In contrast, *Drosophila* Spd-2 is not essential for centriole duplication (its role in duplication appears to be fulfilled by Asl instead), but is required for the PCM recruitment [Dix and Raff, 2007]. In all species, Spd-2/Cep192 accumulates at the centrosome with a fraction tightly associated with the centriole and another with the PCM [Fu and Glover, 2012; Lawo et al., 2012; Mennella et al., 2012; Sonnen et al., 2013].

C. elegans SPD-2 is the first protein binding onto a pre-exisiting mother centriole, which is needed for the centriolar recruitment of ZYG-1, the homolog of PLK4. ZYG-1 then recruits the remaining three proteins, SAS-6, SAS-5 and SAS-4, to centrioles [Delattre et al., 2006; Pelletier et al., 2004; Pelletier et al., 2006]. In *C. elegans*, SPD-2 also acts early in the centrosome

maturation pathway [Kemp et al., 2004; Pelletier et al., 2004]. RNAi assays confirmed the requirement of SPD-2 for the localisation of all centrosomal proteins tested so far, such as SPD-5, PLK1, Aurora-A and γ-tubulin [Greenan et al., 2010; Hamill et al., 2002; O'Connell et al., 2000; Pelletier et al., 2004]. It has recently been demonstrated that recombinantly purified SPD-5 (a putative functional ortholog of fly Cnn) could self assemble into a mesh-like structure that recruited SPD-2 and PLK1 *in vitro* and this polymerisation process was more efficient in the presence of purified SPD-2 and PLK1 phosphorylation [Woodruff et al., 2015].

Like in worm embryos, Cep192, the human homolog of Spd-2, also plays dual functions in centrosome biogenesis [Gomez-Ferreria et al., 2007; Zhu et al., 2008]. From the X-ray crystal structure, it is proposed that the N-terminal acidic region of Cep192 interacts with the CPB domain of PLK4 and this interaction is required for the initial centriolar targeting of PLK4 [Park et al., 2014]. Subsequently, Cep152 (the human homolog of fly Asl) recruitment induces the relocation of PLK4 from the outer edge of the Cep192 ring to that of the Cep152 ring, and Cep152 binds to the CPB with a higher affinity hence excluding the bound Cep192 [Park et al., 2014]. Binding of Cep152 to PLK4 then likely promotes the recruitment of downstream effectors to induce the assembly of a procentriole next to the mother centriole [Kim et al., 2013; Kleylein-Sohn et al., 2007; Sonnen et al., 2013]. Indeed, co-depletion of both proteins resulted in a strong impairment of centriole duplication in U2OS cells, in which up to 60% of the cells showed zero to one centriole [Sonnen et al., 2013].

Cep192 is also implicated in PCM recruitment and MT nucleation in vertebrate cells. In the absence of Cep192, centrioles did not localise to the opposite spindle poles but rather dispersed randomly forming non-bipolar structures [Gomez-Ferreria et al., 2007]. Consistent with its role in PCM recruitment, mitotic centrosomes were not present in the absence of Cep192, and MTs were organised at sites of chromosomes in mitotic-arrested HeLa cells [Gomez-Ferreria et al., 2007; Zhu et al., 2008]. Noticeably, Cep192 siRNA-treated interphase cells have also shown a reduced amount of centrosomal γ-tubulin (30%), but this reduction was much less severe than in the mitotic cells (nearly 100% depletion), [Gomez-Ferreria et al., 2007; O'Rourke et al., 2014]. Therefore, Cep192 is required for the MT nucleation particularly during mitosis. Its requirement for γ-tubulin recruitment appears to be mediated by NEDD1, a γ-TuRC-interacting protein that binds

directly to γ-tubulin [Gomez-Ferreria et al., 2012; Manning et al., 2010]. In addition, Cep192 is found to help recruit several other key PCM components including Aurora-A and Pericentrin [Gomez-Ferreria et al., 2007; Joukov et al., 2010; Joukov et al., 2014; Meng et al., 2015]. Interestingly, phosphorylated Pericentrin is also required to localise Cep192 to the centrosome during mitosis, creating a positive feedback loop [Lee and Rhee, 2011]. The recruitment of Cep192 might then trigger the downstream Aurora-A-PLK1 cascade. Aurora-A kinase interacts with Cep192 at the centrosome, which in turn activates the kinase by self-primed phosphorylation [Joukov et al., 2010; Joukov et al., 2014]. The activated Aurora-A then phosphorylates and activates PLK1 that self primes its binding to the N-terminal region of Cep192 [Joukov et al., 2014; Meng et al., 2015]. The bound PLK1 can then phosphorylate other sites within Cep192 or other proteins in close vicinity to generate binding sites for γ-TuRC and possibly other spindle assembly factors, which subsequently leads to the formation of a large and extended PCM matrix [Meng et al., 2015].

Spd-2 is not the most upstream element of the core duplication module in *Drosophila*. Strikingly, it is not even essential to centriole duplication in somatic cells, but is required for PCM recruitment [Dix and Raff, 2007; Giansanti et al., 2008]. For instance, Giansanti et al., have examined brain preparations from two different *spd-2* mutant alleles that carried premature stop codons and hence truncated the full-length protein to shorter lengths. In both cases, the astral MTs were defective or completely absent in neuroblasts. Spd-2 is also essential for PCM recruitment to sperm centrioles [Dix and Raff, 2007]. In mutant spermatocytes, Cnn incorporation to the sperm centrioles was strongly impaired and centrosomes have lost the ability to nucleate astral MTs. As a result, pronuclear fusion between the male and female nuclei failed at fertilisation, and the embryos arrested prior to the first mitotic division [Dix and Raff, 2007]. Further analysis in *spd-2* mutant brains revealed that the protein acts early in the PCM recruitment pathway. In particular, the localisation of many PCM proteins including Cnn, γ-tubulin, Grip91, and D-TACC was significantly reduced in *spd-2* mutant cells [Dix and Raff, 2007]. Nonetheless, depletion of PLP showed normal Spd-2 signals in brain cells suggesting PLP is not required for Spd-2 localisation in flies, in contrast to human cells [Martinez-Campos et al., 2004]. Antibody injection study in *Drosophila* syncytial embryos has suggested the requirement of Asl in Spd-2 targeting to the centrosome [Conduit et al., 2010; Conduit et al., 2014b]. Following the injection of anti-Asl

antibody, the rate of Spd-2 incorporation to the centrosome was reduced by ~75%. Consistently, in *asl* mutant brain cells, the recruitment of Spd-2 was disrupted. Moreover, Cnn appears to help maintain/stabilise Spd-2 in the PCM. This is supported by the observation that the PCM fraction of Spd-2 moved rapidly away from the centriole in flares in the absence of Cnn [Conduit et al., 2014b]. Spd-2 and Cnn may therefore form a positive feedback loop, where Spd-2 helps recruit Cnn and Cnn then helps maintain Spd-2, and this allows efficient PCM expansion during mitosis. In brain cells, mitotic PCM assembly is strongly diminished in the absence of either protein, but it appears to be abolished in the absence of both proteins [Conduit et al., 2014b]. A possible model is that Cnn and Spd-2 proteins cooperate to form an underlying PCM scaffold, which then recruits many other centrosomal proteins. A similar mechanism is seen in the worm system [Woodruff et al., 2015].

1.6.5 Asterless (Asl) and its family member proteins

B. Wakimoto et al., first identified the asterless gene via a cytological screen of male sterile mutants in *Drosophila* [Bonaccorsi et al., 1998]. Asterless (Asl) is important for astral MT formation during male meiosis. In these mutants, γ-tubulin and Cnn lost their centrosomal localisations and were dispersed in small aggregates without MT-nucleating activity [Bonaccorsi et al., 1998].

Later the asl gene was cloned and found to encode a protein that functions in PCM recruitment [Varmark et al., 2007]. In asl^{1} (the first *asl* mutant allele) spermatocytes, the PCM assembly/stability was dramatically impaired, which led to defects in astral MT nucleation and meiotic spindle assembly [Bonaccorsi et al., 1998; Varmark et al., 2007]. It has also been shown that, in $cnn^{-/-}$ null mutant embryos, the incorporation of GFP-Cnn into the PCM was significantly reduced after injection of anti-Asl antibodies [Conduit et al., 2010]. While these studies demonstrated an important role for Asl in PCM recruitment, Blachon et al., generated a different *asl* mutation (asl^{mecD}, for mechanosensory mutant D), which led to a stronger loss of Asl protein; surprisingly PCM recruitment appeared largely unperturbed in mutant spermatocytes and instead the protein appeared to be essential for centriole assembly [Blachon et al., 2008]. Moreover, a

genome-wide RNAi screen in S2 cells identified Asl as being required for both centriole duplication and PCM recruitment [Dobbelaere et al., 2008]. These conflicting data may be explained by the presence of an alternative PCM recruitment pathway in the absence of Asl in spermatocytes.

More recently, an *in vitro* binding assay has linked Asl to both centriole duplication and PCM recruitment [Dzhindzhev et al., 2010]. Various constructs of ^{35}S-labelled Asl protein were expressed by coupled transcription-translation *in vitro*. Using these fragments, a minimal binding region for PLK4 was mapped to 300 residues at the extreme N-terminus of Asl. In particular, residues 24 to 56 were essential for this direct interaction (see **Figure 6.2** in **Chapter 6** for diagram showing the interaction regions). Mutational studies of negatively charged residues within this region to Ala showed defective PLK4 localisation at the centrosome [Dzhindzhev et al., 2010; Slevin et al., 2012]. This indicates a specific role for Asl in recruitment/stabilisation of Sak/PLK4, a likely upstream component in *Drosophila* centriole assembly pathway, and hence ensures the duplication of centrioles. Additionally, overexpression of an *asl* mutant lacking the ability to bind PLK4 in unfertilised eggs promoted the formation of multiple MTOCs [Dzhindzhev et al., 2010]. These MTOCs did not have centrioles in the core, but instead showed amorphous, electron dense material similar to the PCM and frequently localised γ-tubulin.

A direct interaction was also found between the human homolog of Asl (Cep152) and PLK4 [Cizmecioglu et al., 2010; Hatch et al., 2010]. Nevertheless, Asl and Cep152 seem to differ in their recruitment dependencies for PLK4. After Cep152 RNAi treatment of U2OS cells, little effect was observed on the centrosomal level of PLK4 and *vice versa* [Hatch et al., 2010]. This is now better understood with the involvement of Cep192 in the initial recruitment of PLK4 to the centriole as mentioned above (**Section 1.6.4**). Depletion of Cep152 still allowed the association of Cep192 with PLK4 for centrosomal localisation [Kim et al., 2013; Sonnen et al., 2013]. Depletion of Cep192, however, led to a severe reduction in the centrosomal level of PLK4 and Cep152 [Kim et al., 2013; Klebba et al., 2015; Sonnen et al., 2013]. In flies, Sak/PLK4 only binds to Asl due to the absence of an acidic region in Spd-2. However, the C-terminal fragment of Asl has recently been shown to also interact with the CPB domain of Sak/PLK4 [Klebba et al., 2015], which

conflicts with the data from a binding assay carried out by Dzhiendzhev et al., 2010. It remains unclear why these results differ.

Asl/Cep152 also directly interacts with Sas-4/CPAP in flies and humans shown by two independent studies [Cizmecioglu et al., 2010; Dzhindzhev et al., 2010]. Residues 648 to 994 of Asl bind to Sas-4, a protein implicated in centriole duplication. Injection of anti-Sas-4 antibody, which specifically targets the region involved in Asl interaction, can block the recruitment of Asl to the newly formed daughter centrioles in fly embryos, suggesting a role for Sas-4 in Asl centriolar targeting [Novak et al., 2014].

In summary, Asl/Cep152 proteins are involved in multiple interactions with other core centriole duplication proteins and play an important part in centriole duplication in humans and flies. In flies, Asl also plays an important part in PCM assembly, at least in embryos and some somatic cells, but it may have a more minor role in this process in spermatocytes. It is currently unclear whether Cep152 plays an important part in mitotic PCM assembly in vertebrate cells.

1.6.6 PCM dynamics

Many proteins are concentrated at centrosomes [Andersen et al., 2003; Muller et al., 2010], but it is unclear how these factors assemble at the centrosome. Because the centrosome is a non-membrane bound organelle, proteins can freely assemble through diffusion [Mahen and Venkitaraman, 2012]. How are PCM proteins exchanged between the cytoplasm and centrosome? The recruitment and turnover of PCM proteins at the centrosome can be measured by fluorescence recovery after photobleaching (FRAP). A high-powered laser selectively bleaches the centrosomal fluorescence signal: this does not perturb the behaviour of the tagged protein but irreversibly bleaches the fluorescent molecules. One can then monitor the recovery of fluorescent signal, driven either by the addition of new unbleached proteins or the turnover of the bleached proteins as they are replaced by new, unbleached, proteins from the cytosol.

Recently, a detailed dynamics analysis has been carried out for the key PCM proteins in *Drosophila* embryos [Conduit et al., 2014b]. Of the nine proteins analysed, Asl-GFP, Sas-4-GFP, AurA-GFP, Grip71-WD-GFP, γ-tubulin-GFP and Polo-GFP fluorescence all recovered evenly across the centrosomal region that they originally occupied, suggesting their binding sites are distributed throughout the PCM [Conduit et al., 2014b]. PLP has two centrosomal fractions, one fraction tightly associated with the centriole that turns over very slowly, whereas the other fraction localises to the PCM and exhibits a rapid turnover rate [Martinez-Campos et al., 2004]. Analysis of the mobile PCM fraction of PLP also showed an even fluorescence recovery throughout the centrosomal region [Conduit et al., 2014b].

In contrast, Cnn and Spd-2 proteins exhibit distinct dynamic behaviors at the centrosome [Conduit et al., 2010; Conduit et al., 2014a; Conduit et al., 2014b]. Both proteins are initially incorporated into the region of PCM close to the centrioles and then spread gradually outwards into the PCM. These results strongly suggest that their binding sites are specifically located close to the centrioles. The outward movement of Cnn is MT-dependent, since Cnn distribution collapses into a much narrower region after colchicine treatment [Conduit et al., 2014b]. This is distinct from Spd-2 whose outward flux is not MT-dependent [Conduit et al., 2014b; Conduit and Raff, 2015]. During the centrosome cycle, the two mitotic centrosomes are of similar sizes. If the molecules were only added to the cytoplasmic surface, the larger centrosome would be preferred over the smaller one because of its bigger surface area [Zwicker et al., 2014]. The incorporation profiles observed with Cnn and Spd-2 ensure the addition only around the mother centrioles hence maintaining the equal size throughout the assembly processes. Although the outward movement of Spd-2 is detected in both embryos and somatic cells, the MT-dependent centrosomal flux of Cnn is only observed in fly embryos [Conduit and Raff, 2015]. In worm embryos, SPD-5 also does not exhibit this outward movement at the centrosome [Laos et al., 2015]. Thus, the general significance of the “centrosomal flux” of Cnn and Spd-2 molecules observed in fly embryos remains uncertain.

1.7 Thesis Aims

Significant progress towards understanding the mechanism of centrosome assembly has been achieved during the past few years [Conduit et al., 2015]. In particular, it now seems clear that two proteins, Cnn and Spd-2, play a particularly important role in mitotic PCM assembly in flies, and that Asl is involved in recruiting these proteins around the mother centriole. However many questions remain unanswered. Most importantly, it is unclear how Cnn and Spd-2 proteins cooperate to form an underlying PCM scaffold that then recruits many other PCM components to centrosomes during mitosis? No high-resolution structures have been published for any of these essential PCM scaffold proteins. I therefore wanted to target the core PCM proteins, Cnn, Spd-2 and Asl, for structural characterisation, in order to gain insights into the mechanism underlying PCM assembly in flies at the atomic level.

Due to the large molecular weights of all three proteins to be studied, I took the approach of recombinantly expressing and purifying fragments of these proteins using a bacterial system. I planned to characterise these proteins using biophysical techniques, and to solve their structures using X-Ray crystallography or NMR spectroscopy. The biological relevance of these structures to protein function could then be investigated using the powerful *in vivo* techniques available in *Drosophila*.

Prefatory Note

In this thesis, a Perspective section is added at the end of each Result chapter to discuss the significance of documented experimental results. In some chapters, additional Perspective sections are included to discuss the importance of *in vivo* and/or *in vitro* experimental results. This is to help rationalise the subsequent experimental approaches.

Chapter 2: Identification and characterisation of a PhosphoRegulated-Multimerisation (PReM) domain in Cnn

2.0 Introduction

2.0.1 Centrosomin (Cnn)

Drosophila Cnn is a large coiled coil protein of about 150 kDa. Proteins related to Cnn appear to be present in all organisms that form a centrosome [Hamill et al., 2002; Megraw et al., 2001; Sawin et al., 2004; Wang et al., 2000].

In flies, Cnn is essential for proper centrosome assembly, since its depletion significantly reduces the amount of PCM recruited around centrioles [Dobbelaere et al., 2008; Lucas and Raff, 2007; Megraw et al., 1999; Megraw et al., 2001; Vaizel-Ohayon and Schejter, 1999; Zhang and Megraw, 2007]. As outlined in **Introduction**, in early fly embryos Cnn is incorporated into the PCM close to the centrioles and then fluxes outwards along the centrosomal MTs [Conduit et al., 2010; Conduit et al., 2014b; Conduit et al., 2014a]. Live three-dimensional structured illumination (3D-SIM) super-resolution microscopy analysis revealed that GFP-Cnn forms an extended, scaffold-like structure around the mother centriole, but how Cnn forms such a scaffold is unclear [Conduit et al., 2014b].

A potential clue came from the observation that Cnn is phosphorylated exclusively at centrosomes during mitosis, in a Polo-dependent manner [Conduit et al., 2014a; Dobbelaere et al., 2008]. A simple model would therefore be that PCM maturation is triggered when Cnn molecules are recruited to the centrioles and phosphorylated there by Polo kinase during mitosis, allowing Cnn to assemble into a scaffold that then recruits other PCM components. Prior to the start of my DPhil our lab carried out a tandem mass spectrometry analysis (MS/MS) of Cnn, and

identified several sites that are only phosphorylated at the centrosome [Baumbach, 2014]. This information was used for the design of my experiments.

2.0.2 Aims

- Identify conserved regions in Cnn that contain potential Polo phosphorylation sites.
- Biophysical and structural characterisation of a conserved region of Cnn (the PReM domain) that is a strong candidate for Polo phosphorylation.
- Test the potential significance of phosphorylation of the PReM domain *in vitro* and, in collaboration with Dr Paul T. Conduit (a postdoctoral fellow in the lab), *in vivo*.

2.0.3 Key findings

Cnn contains a phosphorylation-driven multimerisation domain (PReM) that is phosphorylated by recombinant PLK1 kinase *in vitro* and that is essential for efficient Cnn- and PCM-scaffold assembly at the centrosome *in vivo*. The multisite phosphorylation of Cnn seems to ensure that efficient PCM assembly normally occurs only at centrosomes. The PReM domain contains a leucine zipper motif that facilitates the dimerisation of Cnn *in vitro* and phospho-mimicking mutations allow the PReM domain to multimerise into higher order oligomers, potentially providing an important clue as to how a Cnn-scaffold is assembled around mother centrioles.

2.1 Bioinformatic analysis of Cnn

2.1.1 Cnn and CDK5RAP2 contain some conserved features

In order to define conserved domains of Cnn, a Multiple Sequence Alignment (MSA) of a range of species from flies to vertebrates was carried out using ClustalX [Larkin et al., 2007]. The primary sequence conservation of Cnn proteins is generally low. For instance, the overall sequence identity is 33% between *Drosophila melanogaster* and *Aedes aegypti* (mosquito) and only 19% between *D. melanogaster* and *Homo sapiens*. Two domains, denoted as Common Motif 1 (CM1) and Common Motif 2 (CM2), show the most enriched conservation (**Figure 2.1**). A PSIPRED secondary structural prediction [Jones, 1999] was carried out on several Cnn homologs, suggesting that this protein is largely alpha helical. The COILs sever [Lupas et al., 1991] was used to analyse DmCnn and HsCDK5RAP2, and the results suggested that both proteins contain a large number of coiled-coil domains (**Figure 2.2A,B,D,E**). In addition, there are three leucine zipper (LZ) motifs predicted within Cnn (as previously described by [Heuer et al., 1995] and two within CDK5RAP2 (**Figure 2.2C,F**). A LZ motif consists of multiple Leu residues at 7-residue intervals, forming a hydrophobic surface that facilitates protein self-interaction [Hakoshima, 2005]. Interestingly, one of the motifs (LZ1) is always located within CM1, while another (LZ2) is close to the middle of the protein (**Figure 2.2A,D**).

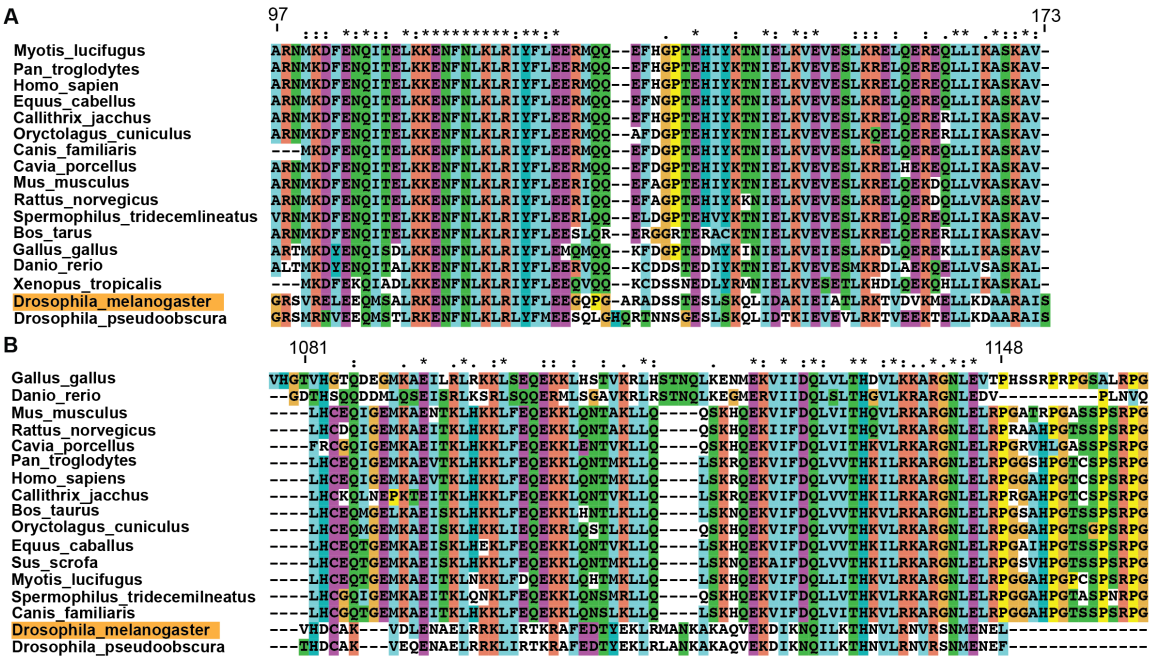


Figure 2.1 MSA analysis of Cnn showing the enriched conservation in CM1 (DmCnn$^{97-172}$) (A) and CM2 (DmCnn$^{1082-1148}$) (B) domains. Residue numbers for *D. melanogaster* are indicated above the sequence alignment. Invariant residues are highlighted with asterisks, conserved residues with colons and semi-conserved residues with full stops. Residues are coloured according to the ClustalX colour scheme.

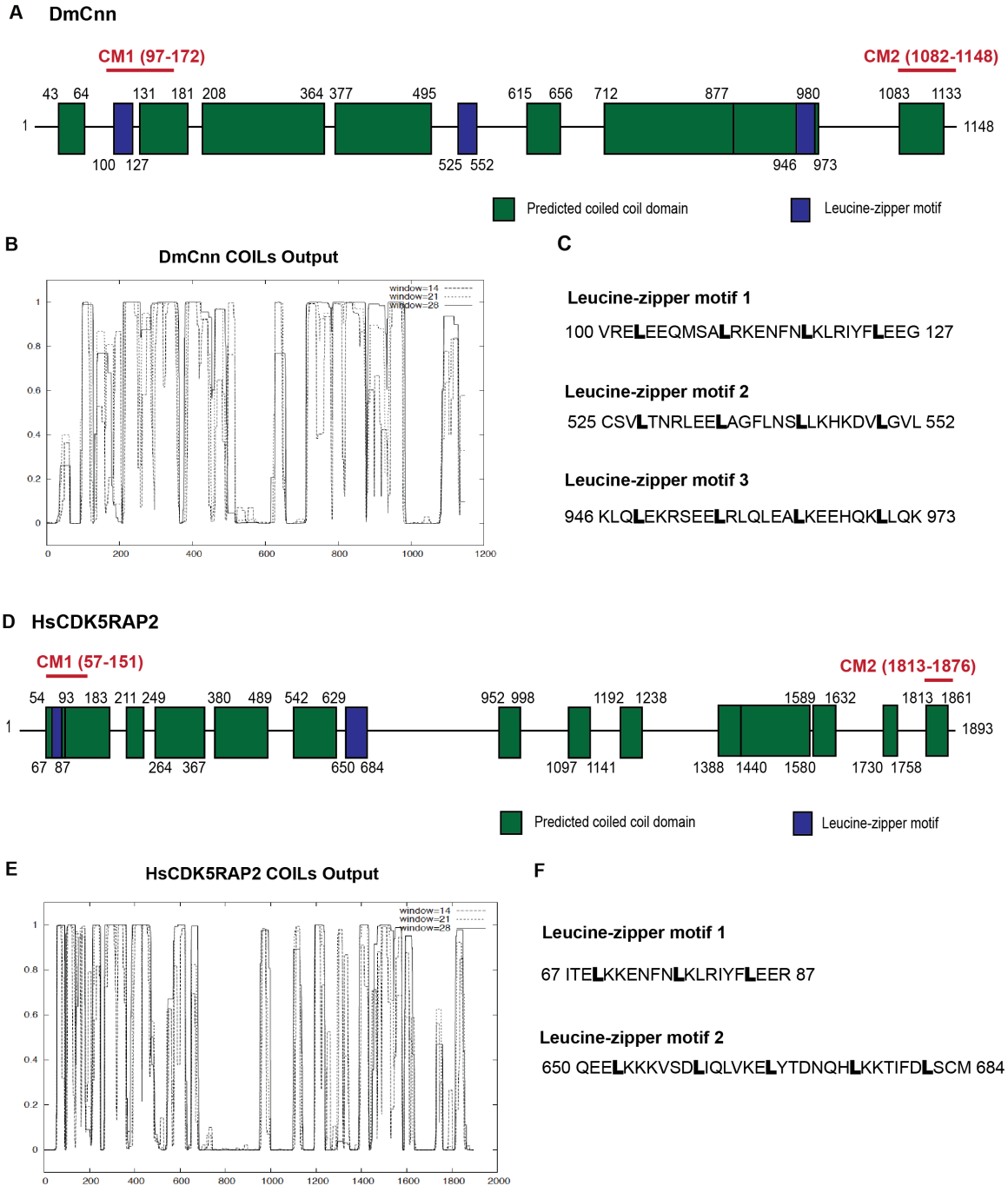


Figure 2.2 Bioinformatic analysis of Cnn. (**A,D**) Schematics showing the overall domain arrangement of *Drosophila* Cnn (**A**) and Human CDK5RAP2 (**D**). Both proteins comprise a large number of coiled-coil domains, and no globular domain was identified. The positions of CM1 and CM2 are also highlighted. (**B,E**) Analysis of Cnn homologs by the COILs server showing the predicted positions of coiled-coil domains. (**C,F**) Protein sequences of the identified leucine zipper motifs (LZ) in DmCnn (**C**) and HsCDK5RAP2 (**F**). Leu residues at *d* position of a heptad repeat of LZ are highlighted. Note that LZ1 is positioned in the CM1 and LZ2 is close to the middle segment of Cnn or CDK5RAP2.

2.1.2 Identification of a highly conserved Cnn middle segment in flies

The lab previously carried out a MS/MS analysis to probe the phosphorylated sites within Cnn, and one of the phospho-peptides detected contained a phosphorylated Ser residue at position 567 (**Figure 2.3A**). Importantly, this phospho-peptide was only found in the centrosomal fraction of Cnn. A consensus PLK1 motif was previously described as a phosphorylated Ser or Thr residue surrounded by an acidic amino acid at +3 position, and a hydrophobic amino acid at position +1 [Santamaria et al., 2011], (**Figure 2.3B**). Due to the poor sequence conservation between Cnn homologs (**Section 2.1.1**), and hence the low confidence of alignment, only Cnn sequences from *Drosophila* species were aligned for the search of potential Polo phosphorylation sites. In particular, the middle segment of Cnn protein (residues 516 to 601) is highly conserved across fly species, and includes Ser567 (**Figure 2.3C**). Intriguingly, this region also contains nine other highly conserved Ser/Thr, several of which closely match the PLK1 consensus sequence (**Figure 2.3C**). In addition, the conserved LZ2 motif is also contained in this region. This domain was therefore selected for further characterisation.

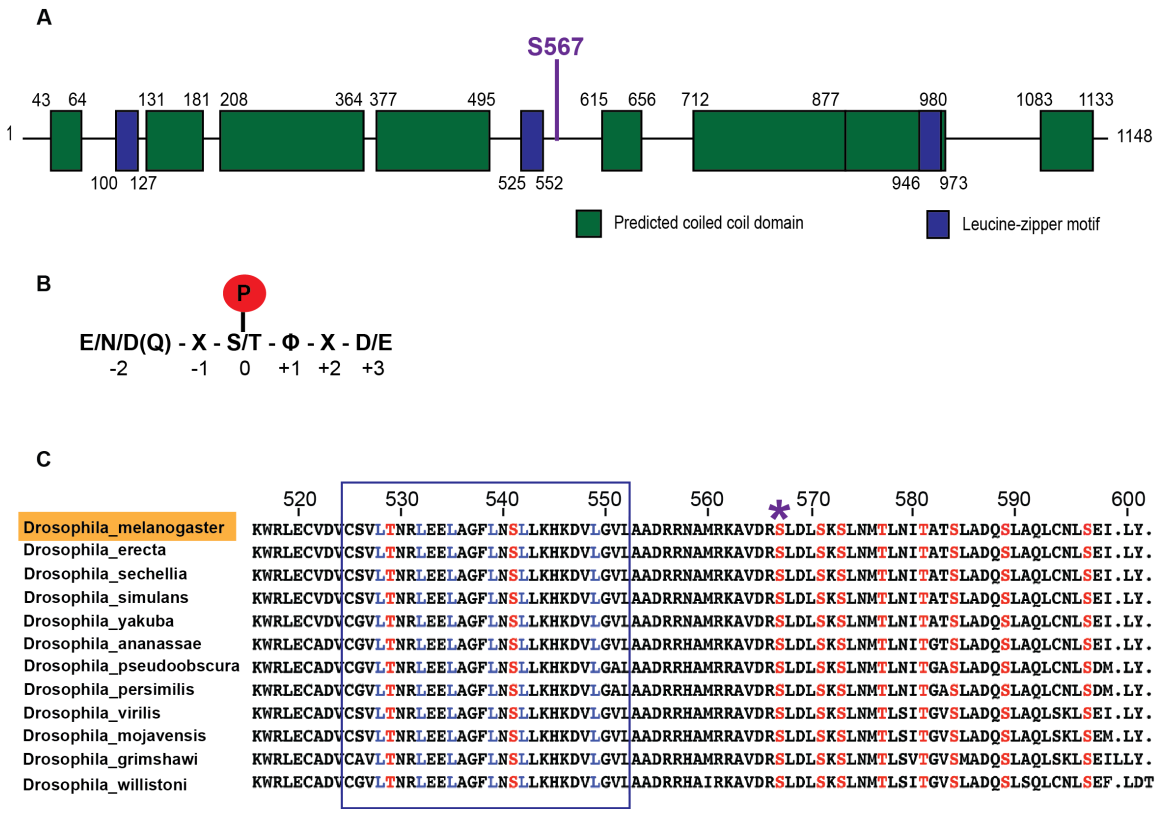


Figure 2.3 Identification of potential Polo phosphorylation sites within Cnn. (**A**) Schematic showing the overall topology of *Drosophila* Cnn. The MS/MS-identified Ser567 phosphorylation site is indicated. (**B**) Diagram showing the PLK1 consensus motif. (**C**) MSA of Cnn from *Drosophila* species, indicating the position of the highly conserved middle segment ($Cnn^{516-601}$). Ser567 is highlighted with an asterisk. LZ2 is boxed. Conserved Ser/Thr residues are highlighted in *red*. Conserved Leu residues within LZ2 are highlighted in *blue*.

2.2 Expression and purification of the Cnn middle segment

2.2.1 Screening for soluble fragments covering the Cnn middle segment

In order to obtain large amounts of protein for subsequent biophysical and structural characterisation, I overexpressed various constructs in a bacterial system. Based on the secondary structure prediction, I designed six constructs of different lengths, which all cover the highly conserved middle segment of Cnn (**Figure 2.4**). The N- and C-termini were chosen not to end in any predicted alpha helices. The constructs were PCR amplified from a cDNA encoding the full length *D. melanogaster* Cnn and then cloned into a pETM-14 vector with an N-terminal polyhistidine tag (His6). A small scale high-throughput screen was carried out to assess the solubility and optimal growth condition for each fragment (see **Appendix 8.1.6** for detailed methods). Details of the screened Cnn constructs are summarised in **Table 2.1**.

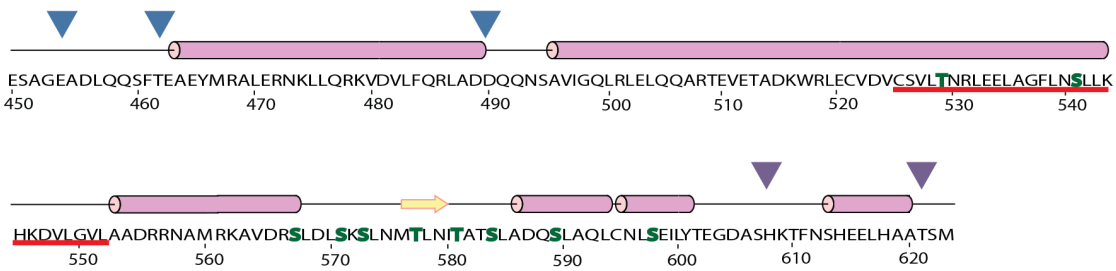


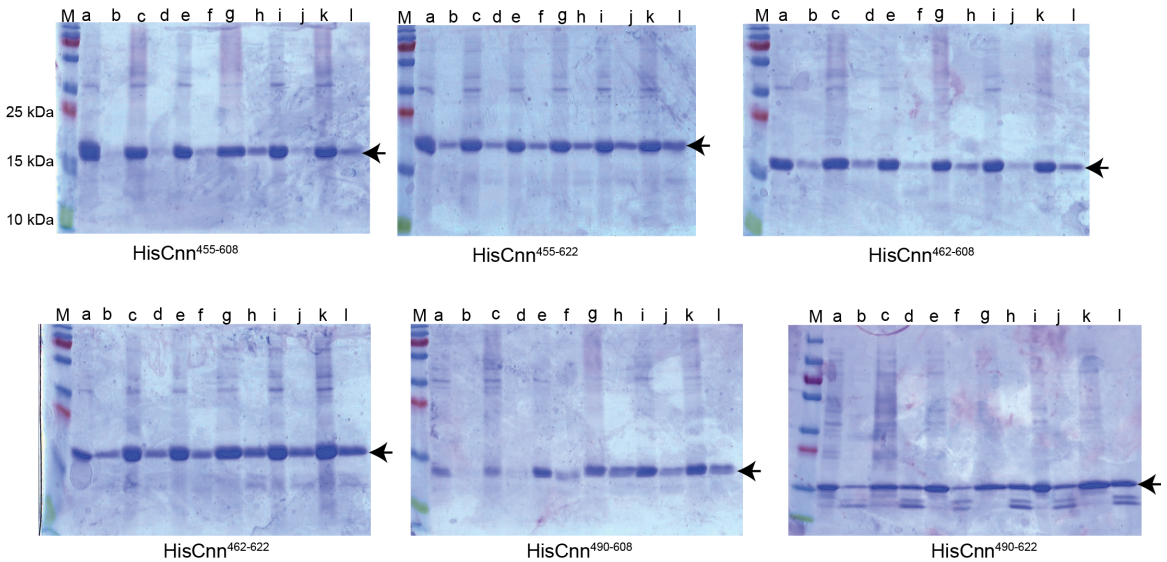
Figure 2.4 Primary sequence of the Cnn middle segment with the predicted secondary structural elements. The constructs cloned for small scale high-throughput screen are marked by *blue* (N-terminus) and *purple* (C-terminus) arrowheads. LZ is underlined. The conserved Ser/Thr residues are highlighted in *green*.

Construct	MW (kDa)	pI	Abs 0.1%	No. of Cys
HisCnn$^{455-608}$	19.7	6.05	0.431	3
HisCnn$^{455-622}$	21.2	6.04	0.399	3
HisCnn$^{462-608}$	18.9	6.23	0.448	3
HisCnn$^{462-622}$	20.5	6.19	0.414	3
HisCnn$^{490-608}$	15.5	6.03	0.451	3
HisCnn$^{490-622}$	17.1	6.03	0.41	3

Table 2.1 Details of the Cnn constructs included in the small scale high-throughput screen. The calculations were carried out using Protparam [Wilkins et al., 1999]. Due to the presence of cysteine residues, 5 mM β-mercaptoethanol (BME) was included in working buffer (WB) to prevent disulfide bonds.

Plasmids encoding the constructs were transformed into 3 different *Escherichia coli* strains, BL21 (DE3), B834 (DE3) and Tuner (DE3). Inoculated small-scale cell cultures (5 ml; in LB or TB medium) were grown at 37 °C until OD_{600} reached 0.6. Protein expression was then induced overnight at 21 °C with the addition of 1 mM IPTG. Proteins were purified using magnetic Ni-NTA

beads (detailed protocol in **Appendix 8.1.6**). All proteins were overexpressed, and some soluble protein could be detected for all constructs, although all were degraded to some extent (**Figure 2.5**). Cnn$^{455-622}$, Cnn$^{462-622}$ and Cnn$^{490-622}$ appeared to be the most soluble, especially when expressed in Tuner (DE3) cells (**Figure 2.5**), so these were used for large-scale purification.



Induced cell sample and Ni beads-eluted protein sample were loaded in a pair, next to each other.The growth conditions are as following:
(a-b) BL21 cell + LB medium (c-d) BL21 cells + TB medium (e-f) B834 cells + LB medium (g-h) B834 cells + TB medium
(i-j) Tuner cells + LB medium (k-l) Tuner cells + TB medium

Figure 2.5 15% reducing SDS-PAGE gel analysis of the Cnn constructs expressed and purified in a small scale high-throughput screen. All cell cultures were induced overnight with 1 mM IPTG at 21 °C. Gel samples were taken for induced cell cultures (left lane). The Ni-NTA eluates were also analysed on gel (right lane). M: PageRuler Plus Prestained Protein Ladder (Thermo Scientific). Arrows indicate the bands corresponding to HisCnn constructs. Constructs ending with C-terminal Ser622 appeared to be more soluble. Tuner (DE3) cells plus TB medium (lanes k-l) seemed to be the optimal growth condition for all the protein fragments. Unfortunately the purified bands did show degradation, especially for HisCnn$^{490-622}$.

2.2.2 Attempt to purify HisCnn$^{455-622}$ on a large scale

Based on the results of the high-throughput solubility screen, HisCnn$^{455-622}$ was first chosen for a large-scale production. A detailed general protocol for recombinant protein expression and purification can be found in **Appendix 8.1.7**, but it will be described once here briefly. The working buffers (WB) for all constructs are listed in **Appendix Table 8.8**. The inoculated cell culture was grown at 37 °C until OD_{600} reached 0.6. 1 mM IPTG was added to induce protein expression at 21 °C overnight. Cells were then harvested by centrifugation. The cell pellet was resuspended in lysis buffer that was prepared from working buffer (WB; For HisCnn$^{455-622}$ purification, WB is 50 mM Tris (pH 7.5), 150 mM NaCl and 5 mM BME.) with the addition of

protease inhibitor tablet (Roche), Lysozyme (Sigma) and DNase (Sigma). Cells were lysed via five passes through an Emulsiflex C5 homogeniser (Avestin) at 4 °C. The cell lysate was clarified by centrifugation at 19,500 rpm for 30 mins, and the supernatant was collected for purification via Ni-NTA affinity chromatography. The lysate supernatant was supplemented to 20 mM imidazole before loading onto a 1 ml Ni-NTA column (Qiagen). The column was then washed with 50 CV of WB containing 50 mM imidazole, and HisCnn$^{455-622}$ was eluted from the Ni column using WB plus 300 mM imidazole. SDS-PAGE analysis of Ni-NTA purification fractions showed that although HisCnn$^{455-622}$ was overexpressed, most of the protein was found in the insoluble fraction and only a small proportion was present in the lysate supernatant (**Figure 2.6A**).

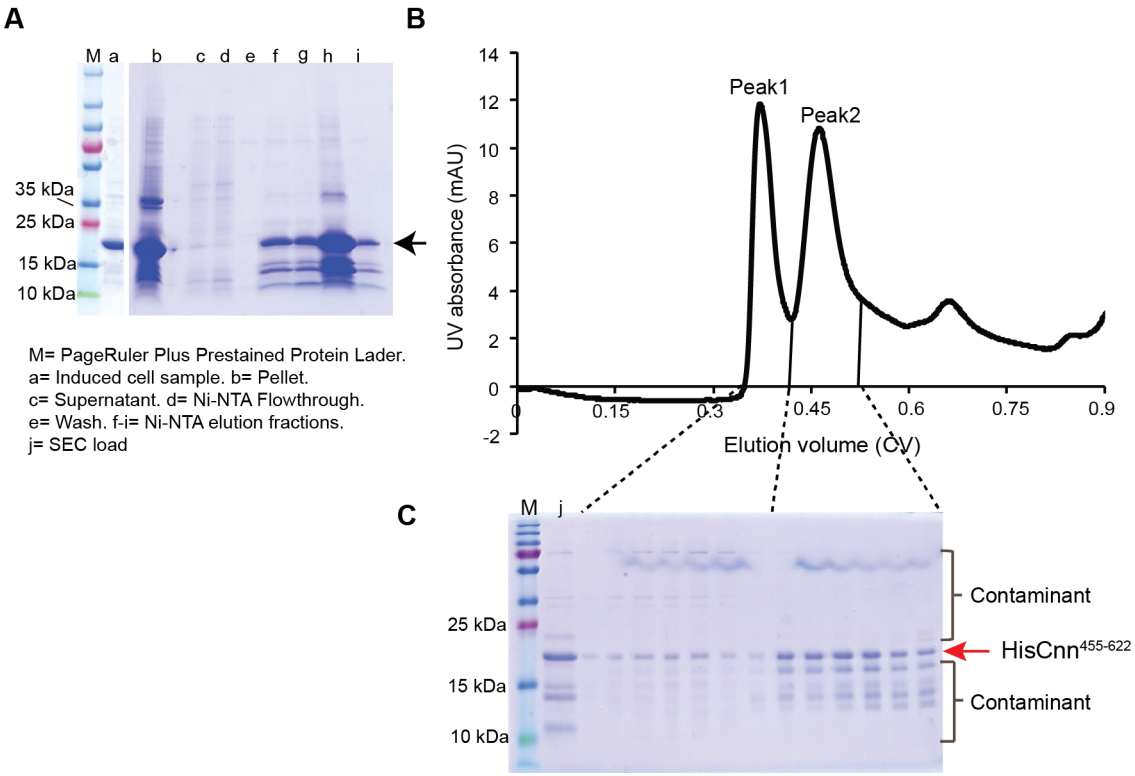


Figure 2.6 Purification of HisCnn$^{455-622}$. (**A**) 15% reducing SDS-PAGE gel analysis of Ni-NTA purification fractions. The band corresponding to HisCnn$^{455-622}$ is highlighted. (**B**) SEC trace from an S75 16/60 column. Elution with 50 mM Tris (pH7.5), 150 mM NaCl and 5 mM BME. Two major elution peaks were observed. Note that the UV absorbance reading was fairly low. (**C**) 15% reducing SDS-PAGE gel analysis showed elution of the full-length protein, as well as lower molecular weight contaminants, within the second peak. Peak1 contained bacterial contaminants of higher molecular weight.

The Ni elution fractions contained about 5 mg of protein in total, although the full-length protein was co-eluted with other contaminants. These fractions were then pooled for size exclusion chromatography (SEC; HiLoad 16/60 Superdex 75 column (GE Healthcare)). The SEC trace of HisCnn$^{455-622}$ showed very low UV absorbance readings for the eluted protein fractions,

suggesting a low yield of purified protein (**Figure 2.6B**). Gel analysis of peak fractions indicated that the second peak contained mostly protein of the expected size, but the lower molecular weight degradation products and/or contaminants were still present (**Figure 2.6C**). Due to the relatively poor yield, I decided to clone the protein constructs into a pMAL-C2X gateway vector with an MBP tag at the N-terminus. The high solubility of MBP often improves the yield of the fusion protein.

2.2.3 Large-scale purification of MBPCnn fusion constructs for *in vitro* assays

MSA of Cnn proteins from *Drosophila* species showed that the protein sequences are very well conserved from position 516 to 601, but other blocks of relatively high sequence conservation could also be detected at $Cnn^{403-438}$ and $Cnn^{462-516}$ (**Figure 2.7**). Therefore, I decided to also attempt to purify a longer construct ($MBPCnn^{403-608}$), which includes more conserved residues that could be of functional importance, for biophysical characterisation.

The MBP fusion proteins, $MBPCnn^{403/462-608}$, (construct details in **Table 2.2**) were expressed using Tuner (DE3) cells in 1 l of TB medium, and purified using MBPTrap HP affinity column (GE Healthcare, detailed protocol in **Appendix 8.1.7**; **Figure 2.8**), followed by size exclusion chromatography. For both constructs, an S200 16/60 column (GE Healthcare) was used for SEC purification with WB as the running buffer. The yield of MBP-tagged protein was significantly greater than the His6-tagged construct. Two protein peaks were observed on the SEC profile, one of which was close to the void volume (void volume of an S200 16/60 column= ~0.33 CV; **Figure 2.9A,C**). Gel analysis indicated that full-length fusion protein was present in both peaks (**Figure 2.9B,D**). I also noticed that $MBPCnn^{403-608}$ eluted earlier on the S200 16/60 column (GE Healthcare) compared to $MBPCnn^{462-608}$, and the two elution peaks therefore significantly overlapped (**Figure 2.9C**). For both constructs I only pooled fractions well away from the void volume to avoid the presence of any aggregated protein (**Figure 2.9B,D**).

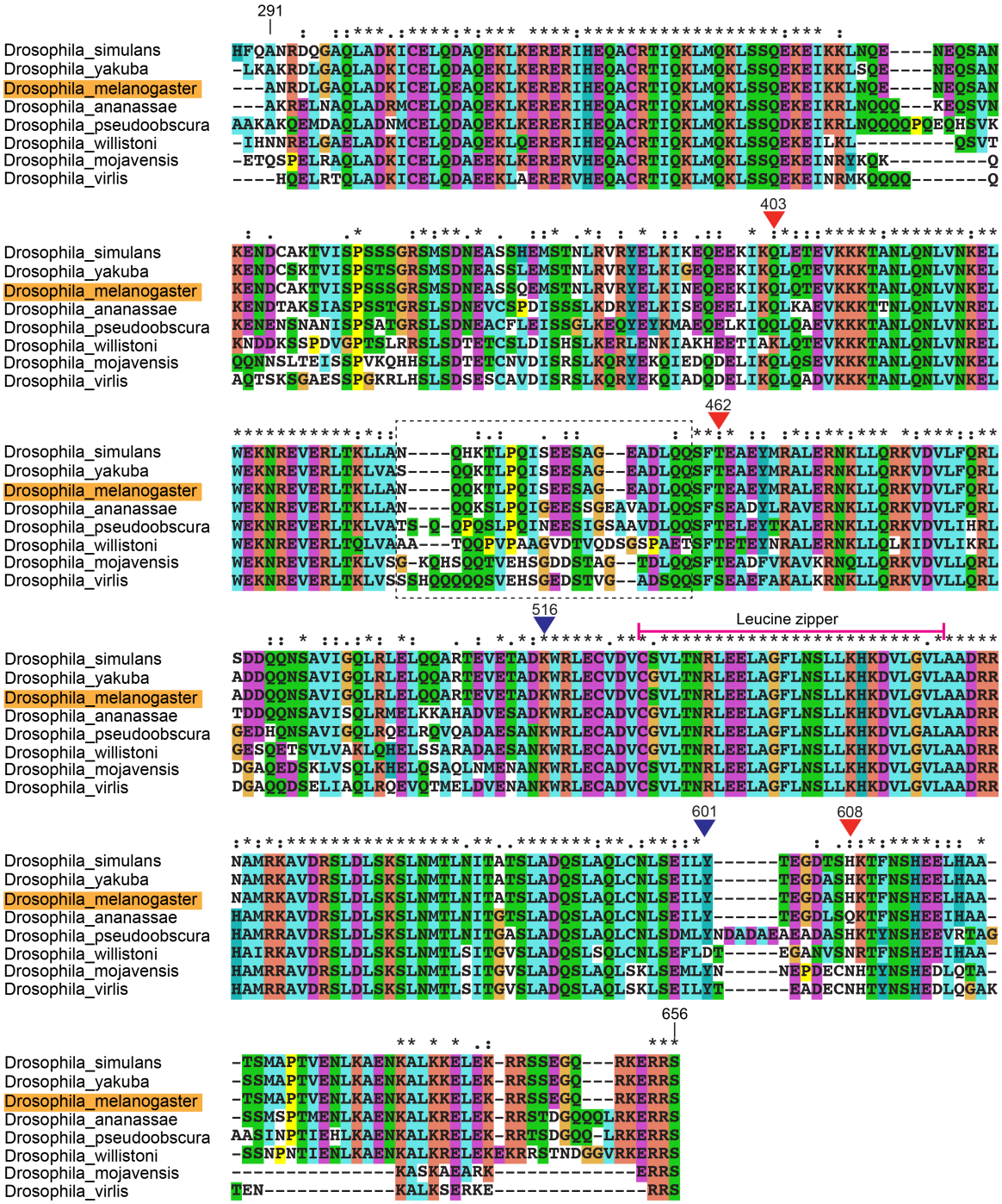
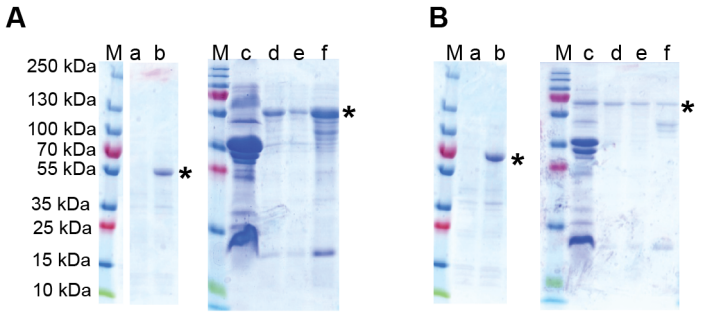


Figure 2.7 MSA (*Drosophila* species only) of the Cnn middle segment showing the enriched sequence conservation between residues 403 and 608, with a gap from Asn439 to Gln459 (highlighted in *grey* box). Residue numbers for *D. melanogaster* are indicated. Invariant residues are highlighted with asterisks, conserved residues with colons and semi-conserved residues with full stops. Residues are coloured according to the ClustalX colour scheme. The constructs purified for *in vitro* assays are indicated by the *red* arrowheads ($Cnn^{403-608}$ and $Cnn^{462-608}$). $Cnn^{516-601}$ (highlighted with *blue* arrowheads) shows high sequence conservation among *Drosophila* species.

Construct	MW (kDa)	pI	Abs 0.1%	No. of Cys
MBPCnn$^{462-608}$	60.58	5.47	1.26	3
MBPCnn$^{403-608}$	67.48	5.52	1.213	3

Table 2.2 Details of the MBPCnn fusion proteins purified on a large scale. Calculations were carried out using Protparam [Wilkins et al., 1999]. Due to the presence of cysteine residues, 5 mM β-mercaptoethanol (BME) was included in working buffer (WB) to prevent disulfide bonds.



a= Uninduced cell sample. b= Induced cell sample. c= Pellet. d= Supernatant.
e= MBPTrap Flowthrough. f= Wash.

Figure 2.8 Purification of MBPCnn fusion proteins by MBPTrap HP affinity chromatography. 15% reducing SDS-PAGE analysis of purification fractions for MBPCnn$^{462-608}$ (**A**) and MBPCnn$^{403-608}$ (**B**). Both constructs were overexpressed and highly soluble. M: PageRuler Plus Prestained Protein Ladder (Thermo Scientific). Asterisks indicate the bands corresponding to MBPCnn fusion proteins.

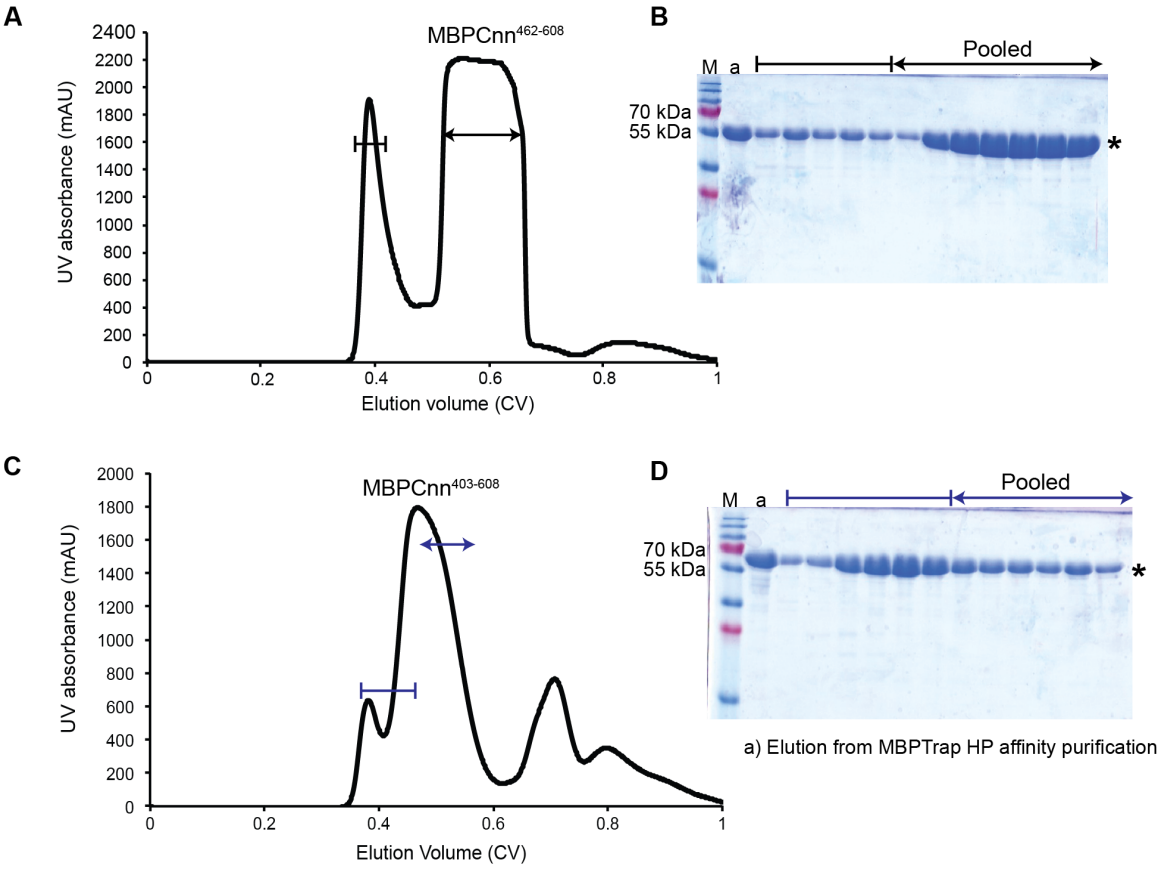


Figure 2.9 Purification of MBPCnn fusion proteins by size exclusion chromatography (SEC). (**A,B**) S200 16/60 SEC trace of MBPCnn$^{462-608}$ showed two elution peaks. Elution with 50 mM Tris (pH7.5), 150 mM NaCl and 5 mM BME. 15% reducing SDS-PAGE gel analysis of peak fractions showed that they contained only pure MBPCnn$^{462-608}$. The majority of the full-length protein eluted around 0.6 CV. The second peak was very broad, suggesting the column was overloaded with protein. (**C,D**) A similar SEC profile was observed for MBPCnn$^{403-608}$. Note that MBPCnn$^{403-608}$ eluted earlier into the column when compared to the shorter construct (**A**). The peak close to void volume (i.e. ~0.33 CV) overlapped with the major elution peak around 0.5 CV, so only fractions well away from the void peak were pooled for *in vitro* assays. M: PageRuler Plus Prestained Protein Ladder (Thermo Scientific). Asterisks indicate the bands corresponding to MBPCnn constructs.

For *in vitro* assays (described in **Chapter 2.3 and 2.4**) I designed the following mutant constructs (**Figure 2.10A,B; Table 2.3**): 1) MBPCnn$^{462-608}$ 10A, where the 10 conserved potential phosphorylation sites, identified by MSA and MS/MS (**Figure 2.3C;Figure 2.10A**), were mutated

to Ala; 2) MBPCnn$^{403-608}$ LZA, where six Leu and one Lys (Lys546) residues in the LZ2 sequence at the *a* and *d* positions of the LZ heptad (**Figure 2.10B**) were mutated to Ala. The replacement of heptad Leu with less hydrophobic Ala is expected to disrupt the coiled-coil interaction and hence LZ formation; 3) MBPCnn$^{403-608}$ 10D/E, where all 10 potential phosphorylation sites were mutated to Asp or Glu. Phosphorylation of Ser or Thr amino acid adds a covalently bound, negatively charged phosphate group, so the substitution of Ser/Thr residue with acidic Asp/Glu residue can mimic the effect of phosphorylation; 4) MBPCnn$^{403-608}$ 10D/E + LZA, where the phosphomimetic mutations and LZ mutations were combined in one mutant construct. In order to eliminate the effect of the MBP tag, maltose binding protein alone was also purified as a control for subsequent *in vitro* assays.

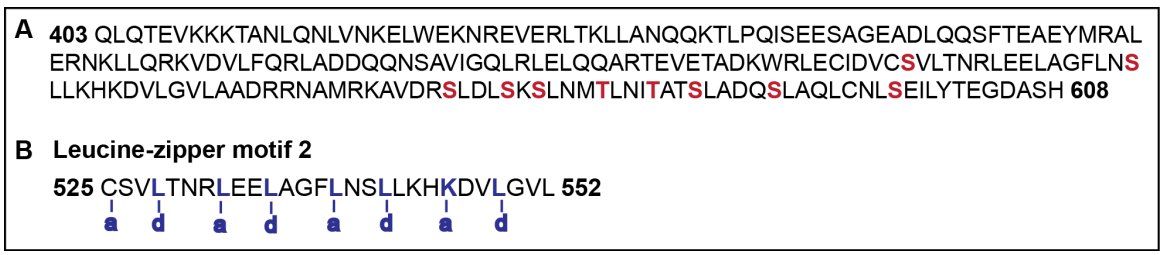


Figure 2.10 MBPCnn mutant constructs designed for *in vitro* assays. (**A**) The sequence of Cnn$^{403-608}$ showing the 10 potential phosphorylation sites in *red*. These highlighted Ser/Thr residues are highly conserved in *Drosophila* species based on the MSA. MBPCnn$^{403-608}$ 10A had the conserved Ser/Thr mutated to Ala. MBPCnn$^{403-608}$ 10D/E had Ser/Thr residues mutated to Asp/Glu. (**B**) The sequence of LZ2 within Cnn showing the mapped heptad repeats. The Leu residues at *a* and *d* positions (in *blue*) were relatively well conserved in MSA (*Drosophila* species only) and were mutated to Ala in MBPCnn$^{403-608}$ LZA. Lys546 (in *blue*) was also mutated to Ala. Phosphomimetic mutations (10D/E) and LZ mutations (LZA) were combined in the MBPCnn$^{403-608}$ 10D/E+LZA mutant construct.

Constructs	MW (kDa)	pI	Abs 0.1%	No. of Cys
MBPCnn$^{462-608}$ 10A	60.39	5.24	1.264	3
MBPCnn$^{403-608}$ LZA	67.09	5.25	1.22	3
MBPCnn$^{403-608}$ 10D/E	67.68	4.94	1.209	3
MBPCnn$^{403-608}$ 10D/E + LZA	67.37	4.91	1.215	3
MBP	44.01	5.17	1.542	0

Table 2.3 Details of the MBPCnn mutant constructs purified for *in vitro* characterisation. The calculations were carried out using Protparam [Wilkins et al., 1999]. Due to the presence of cysteine residues, 5 mM β-mercaptoethanol (BME) was included in working buffer (WB) to prevent disulfide bonds.

To clone these mutant constructs, point mutations were introduced into the *cnn* coding sequence by gene synthesis (Genewiz). Constructs were then amplified by PCR and cloned into a pMAL-C2X destination vector using attLR reaction (Life Technologies, **Appendix 8.1.2**). All the mutants were expressed using Tuner (DE3) cells in 1 l of TB medium, and then purified in the same way as the WT construct (see above). SDS-PAGE gel of SEC purified MBPCnn$^{462-608}$ 10A and MBPCnn$^{403-608}$ 10D/E with or without LZA is shown in **Figure 2.11**.

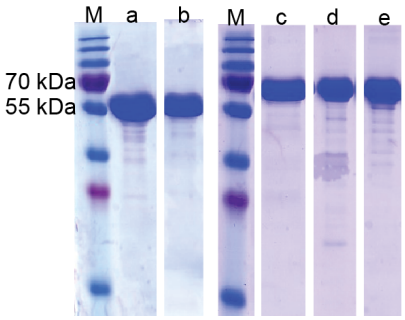
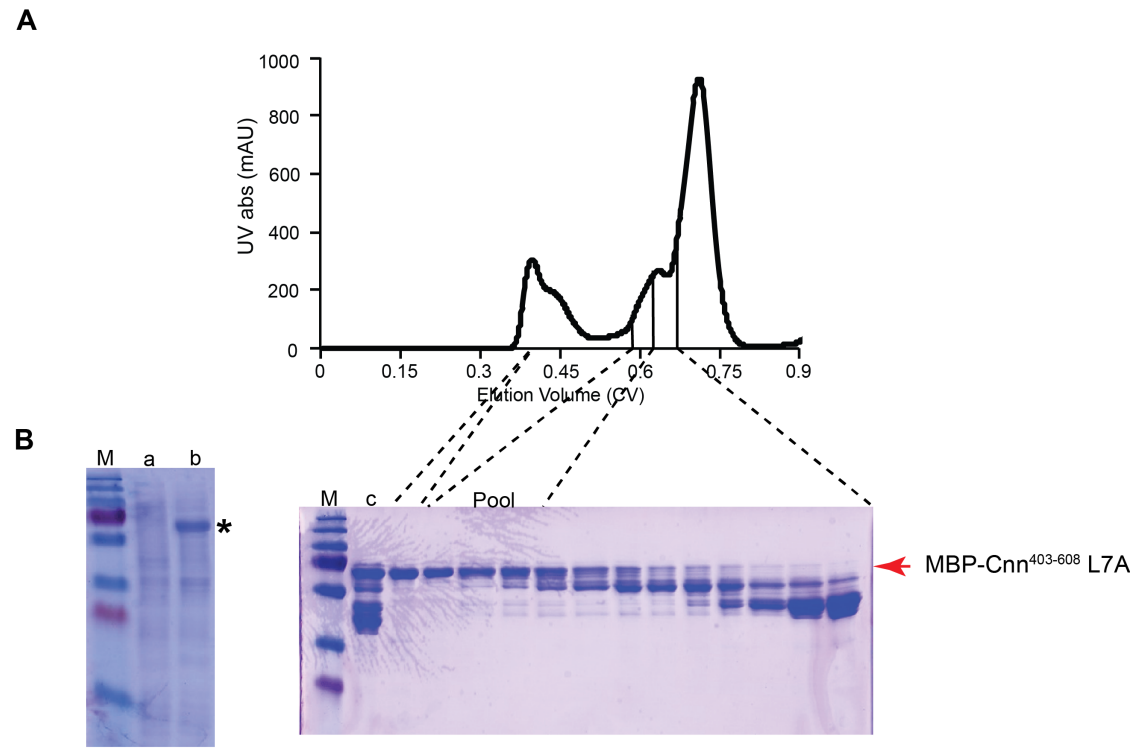


Figure 2.11 15% reducing SDS-PAGE gel showing the SEC purified MBPCnn mutant constructs, which were used in the subsequent *in vitro* assays. a) MBPCnn$^{462-608}$ WT. b) MBPCnn$^{462-608}$ 10A. c) MBPCnn$^{403-608}$ WT. d) MBPCnn$^{403-608}$ 10D/E. e) MBPCnn$^{403-608}$ 10D/E+LZA. M: PageRuler Plus Prestained Protein Ladder (Thermo Scientific).

However, the SEC profile of MBPCnn$^{403-608}$ LZA eluted off an S200 16/60 column (GE Healthcare) showed that most protein was eluted close to the void, suggesting the formation of protein aggregates (**Figure 2.12A**). A small peak at ~0.6 CV was observed, preceding the major peak at ~0.7 CV. Gel analysis revealed that lower molecular weight contaminants, including MBP tag alone, were eluted within the major peak and full-length MBPCnn$^{403-608}$ LZA was largely eluted in the small preceding peak (**Figure 2.12B**). Therefore, only the first three fractions within the preceding peak were pooled for subsequent analysis.



a) Uninduced cell sample.
b) Induced cell sample.
c) Pooled elution from MBPTrap affinity column

Figure 2.12 Purification of the MBPCnn$^{403-608}$ LZA mutant. (**A,B**) SEC trace from an S200 16/60 column (**A**) and the subsequent 15% reducing SDS-PAGE gel analysis (**B**) showing the elution of the LZA mutant construct. Elution with 50 mM Tris (pH7.5), 150 mM NaCl and 5 mM BME. The yield of protein in the non-void peak was very low, and the major elution peak around 0.7 CV contained lower molecular weight contaminants. The three fractions across the small preceding peak, around 0.6 CV, were pooled. The fractions within the void peak were not pooled for *in vitro* analysis. M: PageRuler Plus Prestained Protein Ladder (Thermo Scientific). Asterisk indicates the band corresponding to overexpressed MBPCnn$^{403-608}$ LZA.

In addition, I expressed and purified MBP, which was used as a negative control for my *in vitro* assays. I introduced a TAA stop codon, by site-directed mutagenesis (Agilent), after the MBP encoding sequence, in an empty pMAL-C2X plasmid. The vector was transformed into Tuner (DE3) cells, which were grown in 4 l of TB medium. The protein was expressed and purified using the standard protocol (**Appendix 8.1.7**).

2.3 Radioactive kinase assay of the Cnn middle segment

To assess whether the highly conserved Cnn$^{462-608}$ region contained potential Polo phosphorylation sites, I carried out a radioactive kinase assay using recombinantly expressed and purified human PLK1 kinase (ProQinase). PLK1 is the human homolog of Polo kinase in flies. MBP protein was included as a negative control.

I used 1 µl of MBPCnn$^{462-608}$ and 1 µl of PLK1, both at 1 mg/ml, in a total reaction volume of 20 µl. The phosphorylation was performed in 50 mM Tris-HCl (pH 7.3), 50 mM KCl, 10 mM $MgCl_2.6H_2O$, 20 mM sodium β-glycerophosphate, 15 mM EGTA, 1 mM DTT, 0.1 mM cold ATP, and 1 µCi of [γ-^{32}P]-ATP (PerkinElmer). The reaction was then left at 30 °C for 30 mins, and was stopped by the addition of 5x SDS loading dye. The protein was denatured at 95 °C for 5 mins. For SDS-PAGE analysis, I loaded 15 µl of the total reaction onto a 10% Tris-Glycine gel. The gel was stained with Coomassie Brilliant Blue R-250, and was vacuum dried for autoradiography. The blot showed a phosphorylated band for the wild-type (WT) fusion protein, but not for MBP alone (**Figure 2.13**). To test if the phosphorylation was dependent on the 10 highly conserved Ser/Thr residues, I conducted the same kinase assay with MBPCnn$^{462-608}$ 10A. This mutant was not phosphorylated efficiently by PLK1 kinase *in vitro* (**Figure 2.13**). These results demonstrated that at least some of the 10 potential phosphorylation sites identified by MSA and MS/MS analysis can be phosphorylated by recombinant PLK1 *in vitro*.

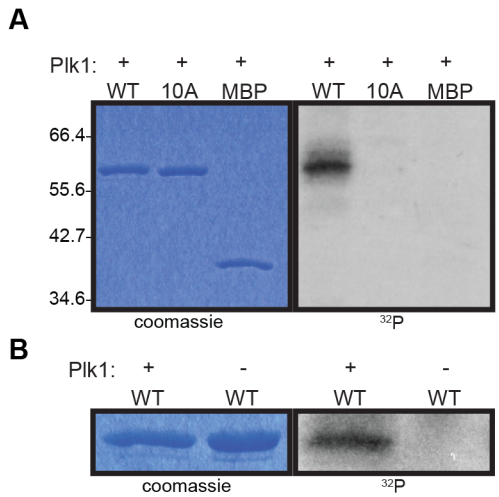


Figure 2.13 Western blot showing results of the radioactive kinase assay. (**A**) Coomassie blue stained 12% reducing SDS-PAGE gel (left panel) shows the MBPCnn$^{462-608}$ WT and 10A proteins and MBP protein. The corresponding autoradiograph (right panel) shows that Plk1 can phosphorylate the Cnn WT sequence but not the 10A mutant sequence. (**B**) Coomassie blue stained SDS-PAGE gel (left panel) and corresponding autoradiograph (right panel) confirm that the phosphorylation of Cnn WT is dependent on the exogenously added PLK1.

2.4 Investigating the potential significance of phosphorylation and the LZ on the self interaction of the Cnn middle segment

As mentioned in **Introduction**, Cnn is specifically phosphorylated at centrosomes, where Cnn scaffolds are formed and extend broadly throughout the PCM recruiting many other PCM components. Phosphorylation can regulate protein function by inducing conformational changes, by creating additional protein interaction surfaces, or both [Nishi et al., 2014]. In the case of Cnn, it seemed plausible that phosphorylation could change the oligomeric state of the protein, perhaps allowing Cnn to assemble into a scaffold structure. Intriguingly, the highly conserved middle segment of Cnn ($Cnn^{516-601}$) contains a LZ motif (**Figure 2.3**). I therefore decided to investigate the oligomeric state of $MBPCnn^{403-608}$ *in vitro*, by conducting Multi Angle Light-Scattering (MALS) experiments on $MBPCnn^{403-608}$ WT and various mutant constructs.

2.4.1 SEC-MALS analysis of MBPCnn fusion constructs

Purified $MBPCnn^{403-608}$ WT and 10D/E were prepared at a range of concentrations: 15 µM, 75 µM, 150 µM and 300 µM, using a 3 kDa MWCO centrifugal concentrator (Amicon). $MBPCnn^{403-608}$ LZA with or without 10D/E were only prepared at 15 µM and 60 µM due to the much lower protein yield of the LZ mutant (**Section 2.2.3**). Samples were injected over a Superose6 column (GE Healthcare), which has a broad protein molecular weight range for high-resolution separation, using 50 mM Tris-HCl (pH7.5), 150 mM NaCl, 5% glycerol and 5mM BME as the running buffer. I added 5% glycerol in the MALS buffer for stabilisation of the LZA mutant construct. $MBPCnn^{403-608}$ WT eluted predominantly as a dimer at all concentrations tested (**Figure 2.14A**). The LZ2 appeared to facilitate dimerisation of the wild-type protein since $MBPCnn^{403-608}$ LZA was predominantly monomeric in solution (**Figure 2.14C**). Remarkably, $MBPCnn^{403-608}$ 10D/E eluted with mass estimations consistent with a pentameric species at all concentrations (**Figure 2.14B**). Most of these higher order oligomers reverted to a monomeric state when LZ2 was mutated, although a small peak with an estimated mass close to a dimer was also present ($MBPCnn^{403-608}$ 10D/E+LZA, **Figure 2.14D**). The LZ2 missense mutations appeared to significantly affect protein stability as shown by the formation of large aggregates on SEC-MALS (*red* arrow, **Figure 2.14C**). Therefore, the observation of dimeric $MBPCnn^{403-608}$ 10D/E+LZA could be explained by the

aggregation-prone LZ mutations. In summary, phosphomimetic mutations within the Cnn middle segment allowed MBPCnn$^{403-608}$ to self assemble into a larger and stable pentameric structure, which was LZ-dependent. Therefore I hereafter refer to Cnn$^{516-601}$ (the minimal region containing the LZ and the 10 conserved Ser/Thr residues) as the phosphoregulated-multimerisation (PReM) domain.

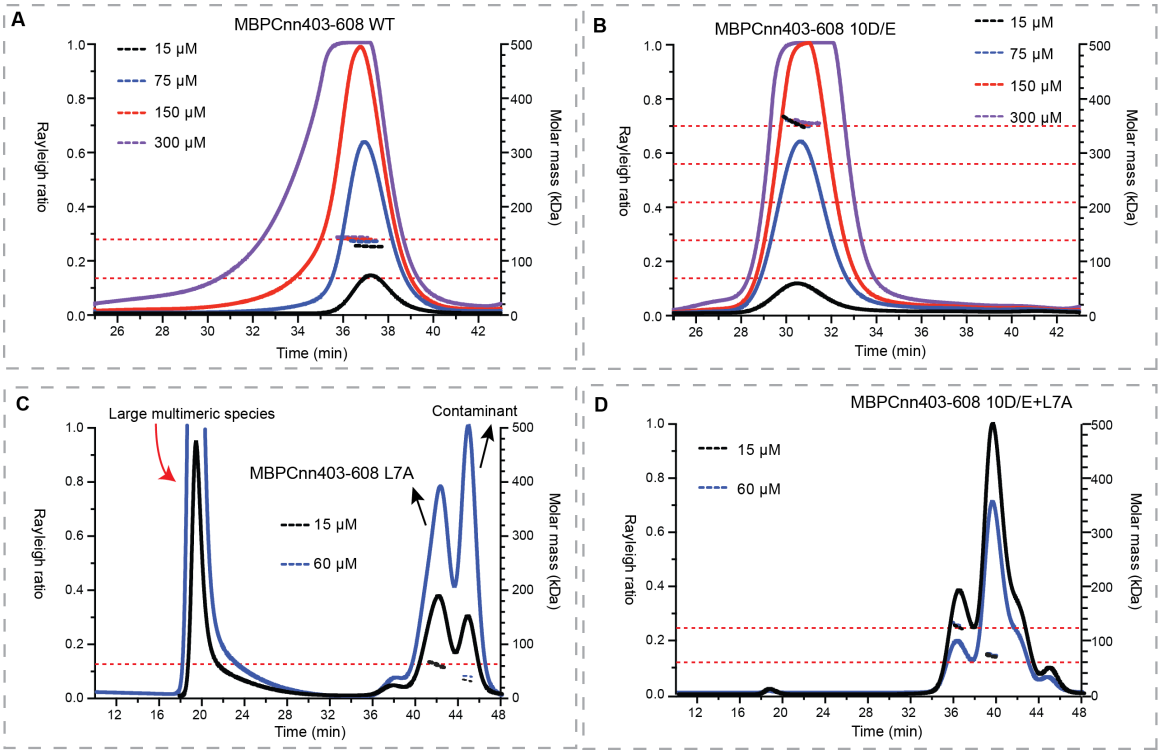


Figure 2.14 SEC-MALS analysis of MBPCnn$^{403-608}$ WT and its mutant constructs. Elution with 50 mM Tris (pH7.5), 150 mM NaCl, 5% Glycerol and 5 mM BME. MBPCnn$^{403-608}$ WT eluted predominantly as a dimer at various concentrations (**A**). Solid lines show the relative Rayleigh ratio while dashed lines show the observed average mass across the peak. Expected masses of a monomer and successive multimers are indicated with *red* dotted lines across the graph. (**B**) MBPCnn$^{403-608}$ 10D/E formed pentameric structures at a range of concentrations from 15 µM to 300 µM. (**C**) The LZA mutant eluted in two peaks, one of which was the void (*red* arrow) possibly due to the unspecific aggregation. The third peak furthest into the elution column corresponded to the elution of small molecular weight contaminants. LZ mutation disrupted the dimerisation of Cnn, as the LZA mutant eluted as a monomer, although the majority of the protein aggregated in the void. (**D**) MBPCnn$^{403-608}$ 10D/E+LZA eluted predominately as a monomer at tested concentrations. This was in big contrast to MBPCnn$^{403-608}$ 10D/E mutant, which showed self-assembly into pentamers even at 15 µM. Note that there was a small preceding peak with an estimated mass close to a dimer. Unlike the LZA mutant, no large void peak was observed for 10D/E+LZA mutant, suggesting the phosphomimetic mutation could have provided stability to the protein. However, the presence of minor dimeric species could also be due to the aggregation-prone LZ mutation.

2.5 *In vivo* analysis of PReM domain function

2.5.1 An mRNA injection assay to assess centrosomal localisation

To characterise the functional importance of the PReM domain in flies, RNA injection experiments were designed. *Drosophila* embryos provide an ideal system to analyse the localisation of centrosomal proteins. During the syncytial stage of embryogenesis, all nuclei undergo rapid mitosis without cytokinesis (alternating between S and M phases). After cycle 10, most nuclei have moved to the surface of the egg, allowing multiple nuclei to be visualised as a monolayer as they synchronously undergo several rounds of nuclear division [Foe and Alberts, 1983]. *In vitro* transcribed mRNAs, encoding GFP-fusion constructs of interest, can be injected into the embryo blastoderm as previously described [Oliveira et al., 2010], (**Figure 2.15**). The GFP-fusion constructs are translated into proteins and their localisation can be tracked using fluorescence microscopy. Since many centrosomes can be visualised at the same time, quantitative analysis can be carried out. In addition, this assay can be conducted in a relatively high-throughput manner, without the need to generate stable recombinant lines for each construct.

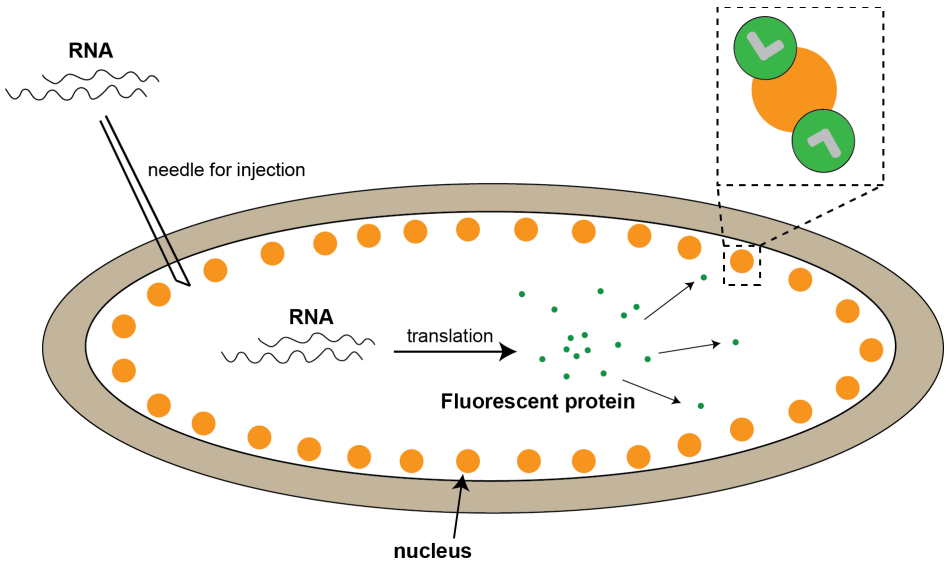


Figure 2.15 A schematic illustration of mRNA injection assay in fly embryos. GFP-Cnn constructs were cloned and transcribed into mRNA *in vitro,* using the T3 mMESSAGE mMACHINE® kit. mRNA was injected into the syncytial cytoplasm, where it was translated into fluorescent Cnn molecules. The GFP-fusion proteins were then free to diffuse throughout the cytoplasm and be incorporated into centrioles. Note that the centrioles in close up view are greyed out because Cnn is a PCM protein and its GFP signal will not label the centriolar wall. Diagram modified from [Cottee, 2014].

2.5.2 Construct cloning and RNA preparation

A PCR-amplified Cnn WT construct was cloned into a pRNA_GFPNT vector using T4 ligation (**Appendix 8.1.1**). Different mutations were introduced into the wild-type cDNA sequence using the gene synthesis service (Genewiz). Each mutant construct was originally cloned into a pUC57-Kanamycin resistant vector (Genewiz; **Figure 2.16A**). The Cnn insert was then re-ligated into a pRNA_GFPNT vector by T4 ligation (NEB, **Figure 2.16B,C**). RNA was obtained via *in vitro* transcription, and all constructs were injected at 2 mg/ml (**Appendix 8.1.12**). However, it should be noted that the amount of RNA injected into fly embryos, the rate and efficiency of translation *in vivo,* and the stability of RNA *in vivo* cannot be controlled. Dr Paul T. Conduit, a post-doctoral fellow in the lab, carried out all the fly husbandry, mRNA injection and microscopy work for these experiments. All RNA injections were conducted in a *cnn*$^{-/-}$ null mutant background.

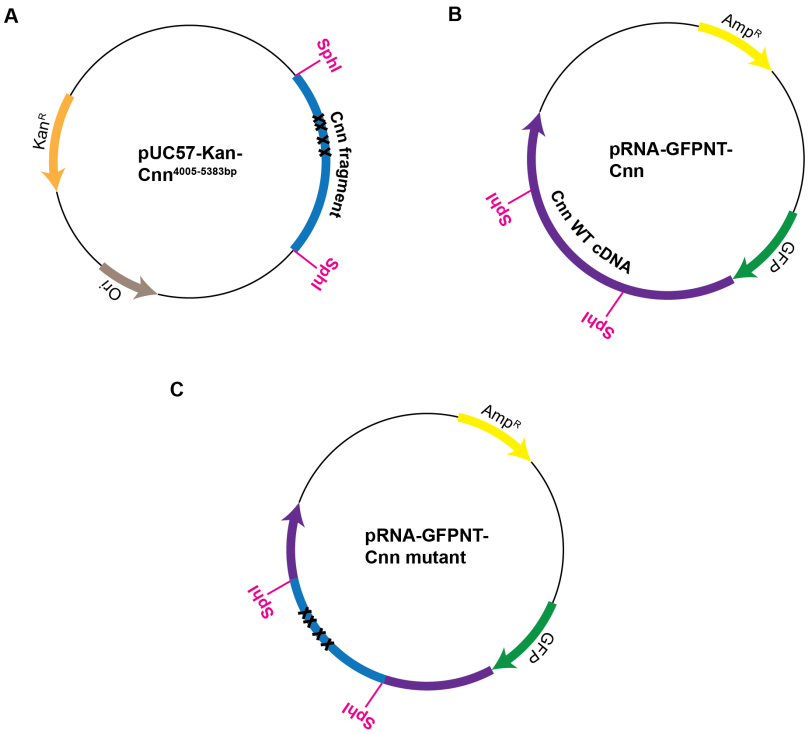


Figure 2.16 Cloning of pRNA-GFPNT-Cnn mutant constructs. (**A**) The cDNA sequence of Cnn, comprising nucleotides 4005 to 5383, with different mutations indicated by X, (for *in vivo* and *in vitro* analysis) was synthesised using the Genewiz service. The pUC57-Cnn vector was digested with SphI restriction enzyme (Roche), and the Cnn insert was purified by agarose gel extraction (Qiagen). (**B**) The pRNA-GFPNT-Cnn vector was also digested with SphI and treated with Antartic Phosphatase (NEB) to prevent self-ligation of linearised vector. (**C**) The SphI-digested insert in (**A**) was then ligated into the linearised vector in (**B**) by T4 ligation. The correct plasmid contained the cDNA sequence encoding the full-length Cnn, and any desired mutations.

2.5.3 Phosphorylation of PReM domain is essential for efficient Cnn scaffold formation and PCM assembly

As a control, the wild-type Cnn fusion protein was first tested in *cnn*$^{-/-}$ null mutant embryos. Confocal images were taken to assay protein behaviour 1-2 hrs after mRNA injection. As expected, GFP-Cnn WT formed an extended PCM scaffold, as observed in transgenic flies expressing GFP-Cnn WT (**Figure 2.17A**). This indicated that the GFP-Cnn WT fusion protein could rescue the *cnn*$^{-/-}$ null mutant phenotype. *In vitro* MBPCnn$^{462\text{-}608}$ containing the 10 potential phosphorylation sites mutated to Ala could not be phosphorylated by PLK1 kinase (**Section 2.3**). *In vivo*, GFP-Cnn 10A localised at centrosomes only very weakly, and no extensive scaffold structure appeared to assemble around the centrioles (**Figure 2.17B**). This was confirmed by live 3D-SIM analysis (**Figure 2.17C,D**). Interestingly, however, FRAP analysis showed that GFP-Cnn 10A was continuously recruited around the centrioles at near normal levels, but the recruited protein did not assemble into a scaffold that spread outward, as was the case for the wild-type protein (**Figure 2.17E,F**).

Moreover, the localisation of another key PCM protein, Spd-2, was also severely affected in embryos injected with GFP-Cnn 10A mutant. A much lower level of Spd-2-RFP signal was detected at centrosomes, and its PCM expansion was significantly reduced (**Figure 2.18**).

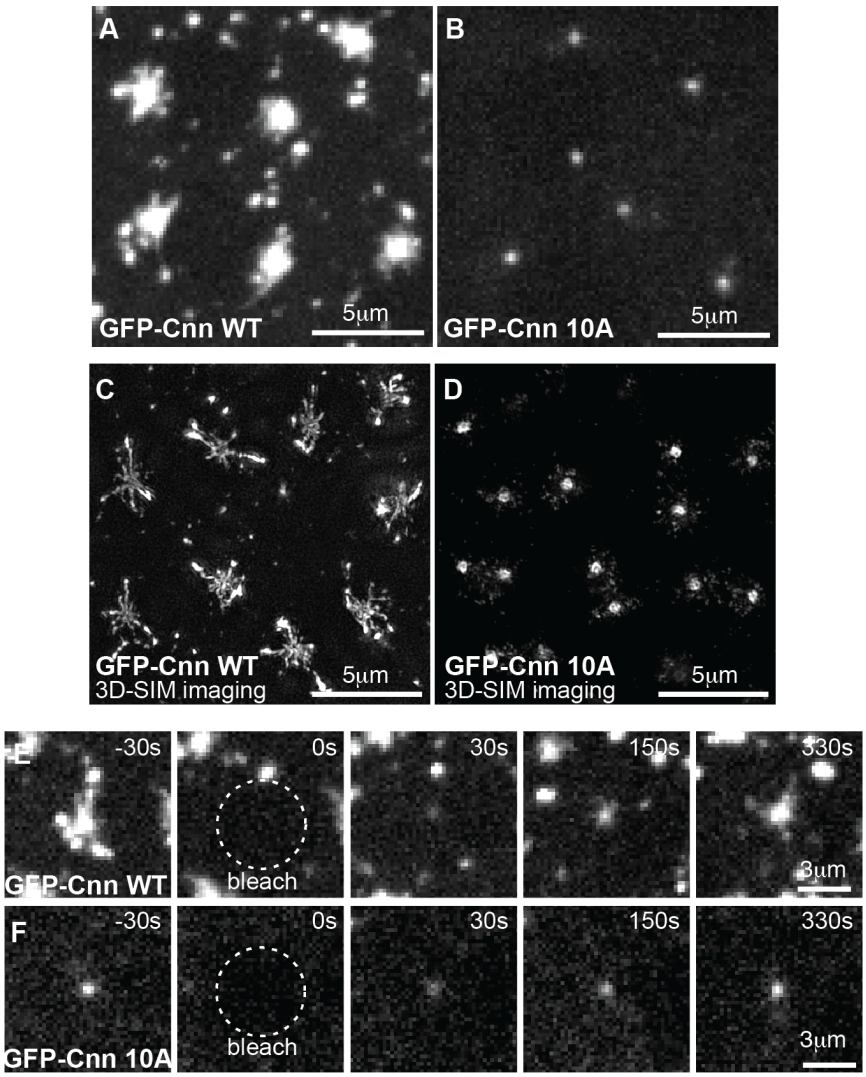


Figure 2.17 Phosphorylation within the Cnn PReM domain is essential for efficient Cnn scaffold formation. (**A,B**) Confocal microscopy images showing that GFP-Cnn WT (**A**) localised to centrioles normally in a *cnn-/-* null mutant background, forming extended scaffold-like structures. The 10A mutant (**B**) could not form Cnn scaffolds around centrioles. Images were taken using a Perkin Elmer ERS spinning disk (Volocity software) mounted on a Zeiss Axiovert microscope using a 63X/ 1.4NA oil immersion objective and on Orca ER CCD camera (Hamamatsu Photonics, Japan). Images shown are maximum intensity projections of several z-slices. (**C,D**) Super-resolution 3D-SIM images showing the centrosomal localisation of GFP-Cnn WT and GFP-Cnn 10A. Note that the extended Cnn scaffolds were not observed in 10A mutant. Images were processed using SoftWorx software (Applied Precision). Images shown are maximum intensity projections of several z-slices. (**E,F**) FRAP analysis showing the fluorescence recovery of GFP-Cnn WT and GFP-Cnn 10A after photobleaching. GFP signals were bleached using a focussed 440 nm laser, during S phase of cycle 11 or 12. 0.5 μm thick confocal sections through the center of a selected centrosome were collected. Both proteins showed a fast recovery at the centre of the centrosome suggesting normal recruitment around centrioles. However, the 10A mutant did not spread outwards like the wild-type construct. All images were taken and processed by Dr Paul T. Conduit.

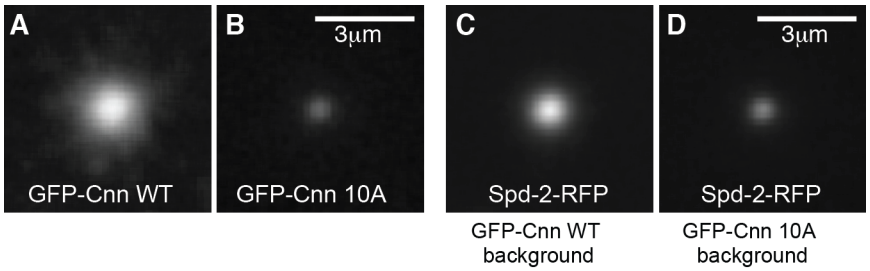


Figure 2.18 Phosphorylation of the Cnn PReM domain is essential for PCM assembly in flies. Confocal images show the average centrosomal fluorescence of GFP-Cnn WT (**A**) and GFP-Cnn 10A (**B**). Images were taken using a Perkin Elmer ERS spinning disk (Volocity software) mounted on a Zeiss Axiovert microscope using a 63X/ 1.4NA oil immersion objective and on Orca ER CCD camera (Hamamatsu Photonics, Japan). Images shown are maximum intensity projections of several z-slices. The failure of the 10A mutant to assemble a proper Cnn scaffold leads to a reduction in the amount of Spd-2-RFP recruited to centrosomes (**C,D**). All images were taken and processed by Dr Paul T. Conduit.

2.5.4 Multisite phosphorylation is required for efficient Cnn scaffold assembly

Results thus far suggested that the 10 phosphorylation sites identified within the Cnn PReM domain are essential for the efficient assembly of extended scaffold structures around centrioles. We wanted to test whether all 10 potential sites contributed equally to scaffold assembly. In order to compare the behaviour of mutant constructs in a more quantitative way, we calculated the average fluorescence intensity across centrosomes using radial profiling (see **Appendix 8.1.13** for detailed methods; **Figure 2.19**). The single point mutation of Ser567 (identified by MS/MS) to Ala was first analysed. Its radial profile was narrower than WT, but was not as narrow as the 10A mutant. Two more sites (Ser571; Ser573) close to Ser567 were then mutated, on their own and in combination with Ser567. The single point mutation of Ser567, Ser571 or Ser573 did not show as severe reduction in PCM expansion as mutating all three sites together (**Figure 2.19**). We tested many other combinations of Ser/Thr sites and the results suggested that no single phosphorylation site within the PReM domain is essential for scaffold assembly, although some sites appear more important than others. Our results also showed that the more Ser/Thr sites present, the greater the ability of Cnn to assemble into a scaffold structure.

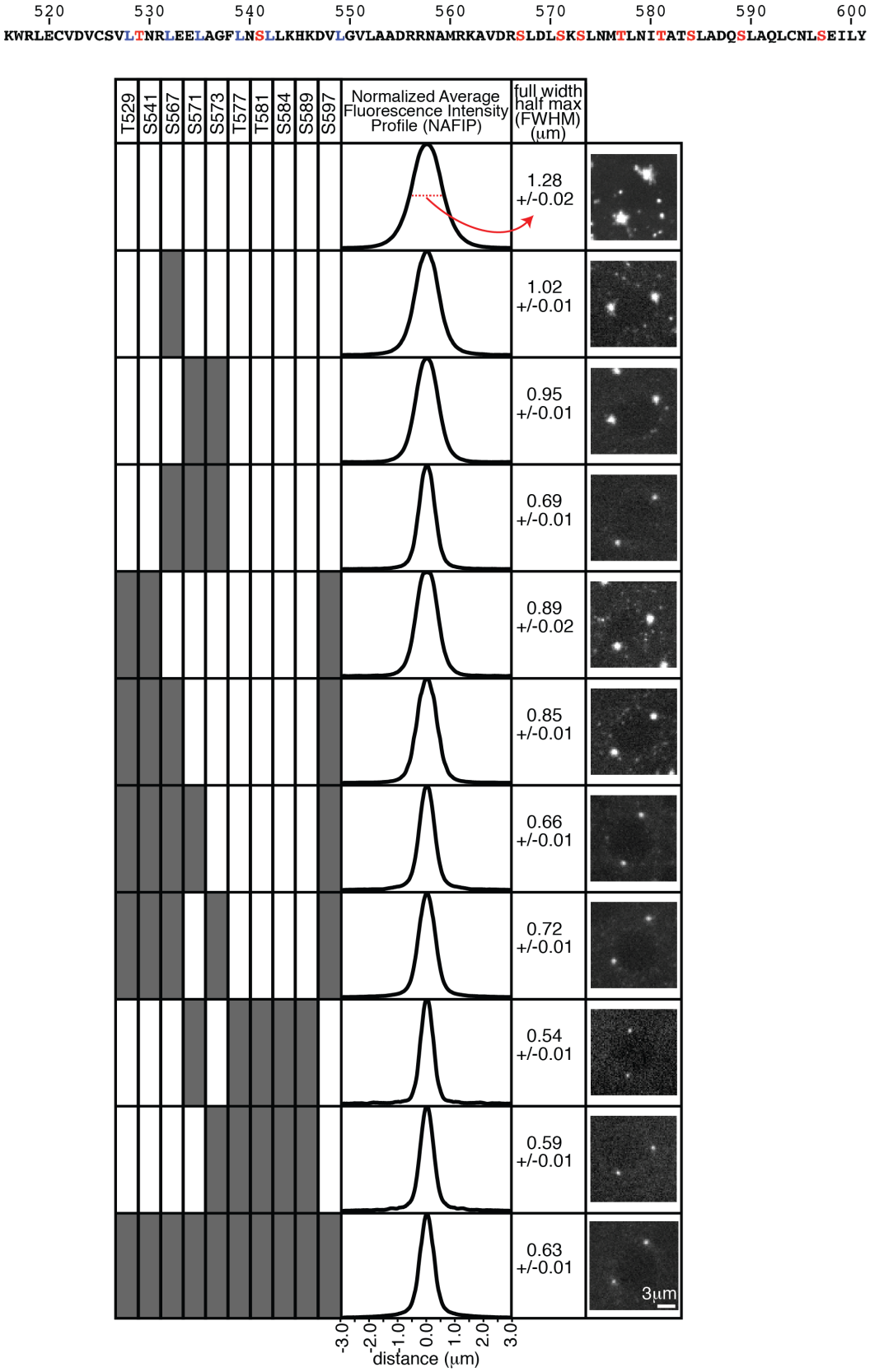


Figure 2.19 The PReM domain is phosphorylated at multiple sites, regulating Cnn scaffold assembly. The primary sequence of the Cnn PReM domain is displayed above the diagram. The potential phosphorylation sites are highlighted in *red*. Different Ser/Thr residues were mutated to Ala, and its effect on centrosomal localisation of GFP-Cnn is displayed. A normalised average fluorescence intensity profile of centrosomes was plotted for each mutant (28 centrosomes on average from five embryos; calculations were carried out in ImageJ as stated in [Conduit et al., 2014a]), and a representative confocal image is also shown. The extent of PCM expansion for each mutant protein was quantified by measuring FWHM ± SE. All images were taken and processed by Dr Paul T. Conduit, who carried out the subsequent image analysis.

2.5.5 The LZ within the PReM domain is required for efficient Cnn scaffold formation

In vitro, the LZ within the PReM domain facilitated the dimerisation of MBPCnn$^{403-608}$ fusion protein (**Figure 2.13**). To test the importance of the LZ *in vivo*, we investigated the behaviour of GFP-Cnn LZA in fly embryos. The LZA mutant had a similar phenotype to the 10A mutant (**Figure 2.20**). Both showed impaired PCM scaffold assembly, as well as a diminished centrosomal localization. Therefore, the LZ within the PReM domain is essential for efficient phosphorylation-driven Cnn scaffold formation *in vivo*.

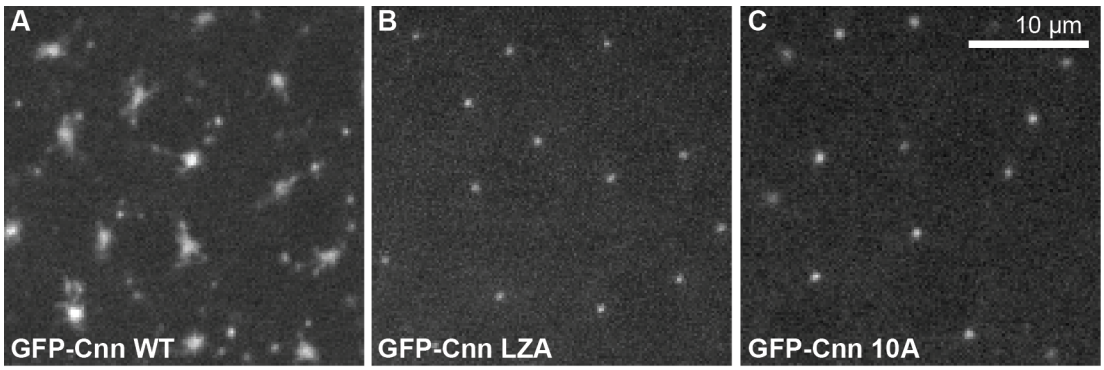


Figure 2.20 Centrosomal localisation of GFP-Cnn LZA and GFP-Cnn 10A in *cnn*$^{-/-}$ null mutant embryos. (**A**) Centrosomal localisation of the wild-type Cnn showing the assembly of scaffold-like structures. Mutation of the LZ within the PReM domain (**B**) severely disrupted the assembly of Cnn scaffolds around centrioles, a phenotype similar to the 10A mutant (**C**). Images were taken using a Perkin Elmer ERS spinning disk (Volocity software) mounted on a Zeiss Axiovert microscope using a 63X/ 1.4NA oil immersion objective and on Orca ER CCD camera (Hamamatsu Photonics, Japan). Images shown are maximum intensity projections of several z-slices. All images were taken and processed by Dr Paul T. Conduit.

2.5.6 Phosphomimetic mutations within the PReM domain allow Cnn to form spontaneous scaffolds independently of centrosomes

I have shown that phospho-mimicking mutations in the PReM domain can promote the multimerisation of MBPCnn$^{403-608}$ *in vitro* (**Figure 2.14**). We wondered whether this multimerisation might provide the basis of Cnn scaffold assembly in flies. We injected mRNA encoding GFP-Cnn 10D/E into *cnn*$^{-/-}$ null mutant embryos. Strikingly, large protein aggregates formed spontaneously in the cytoplasm (**Figure 2.21A,B**). This strongly suggested that phospho-mimicking mutations in Cnn allow the protein to spontaneously assemble into scaffold-like structures in embryos.

My *in vitro* data demonstrated that the higher order protein complexes formed upon phosphomimetic mutations were LZ-dependent, and this was also true of the large protein aggregates formed in the cytoplasm: embryos expressing a form of GFP-Cnn 10D/E in which the LZ was mutated (GFP-Cnn 10D/E+LZA) no longer formed large aggregates (**Figure 2.21C**). Since the LZ was also essential for the formation of centrosomal Cnn scaffolds (**Figure 2.20**), this observation supported the hypothesis that these cytoplasmic foci of GFP-Cnn 10D/E assemble by the same LZ-dependent mechanism as the centrosomal scaffolds.

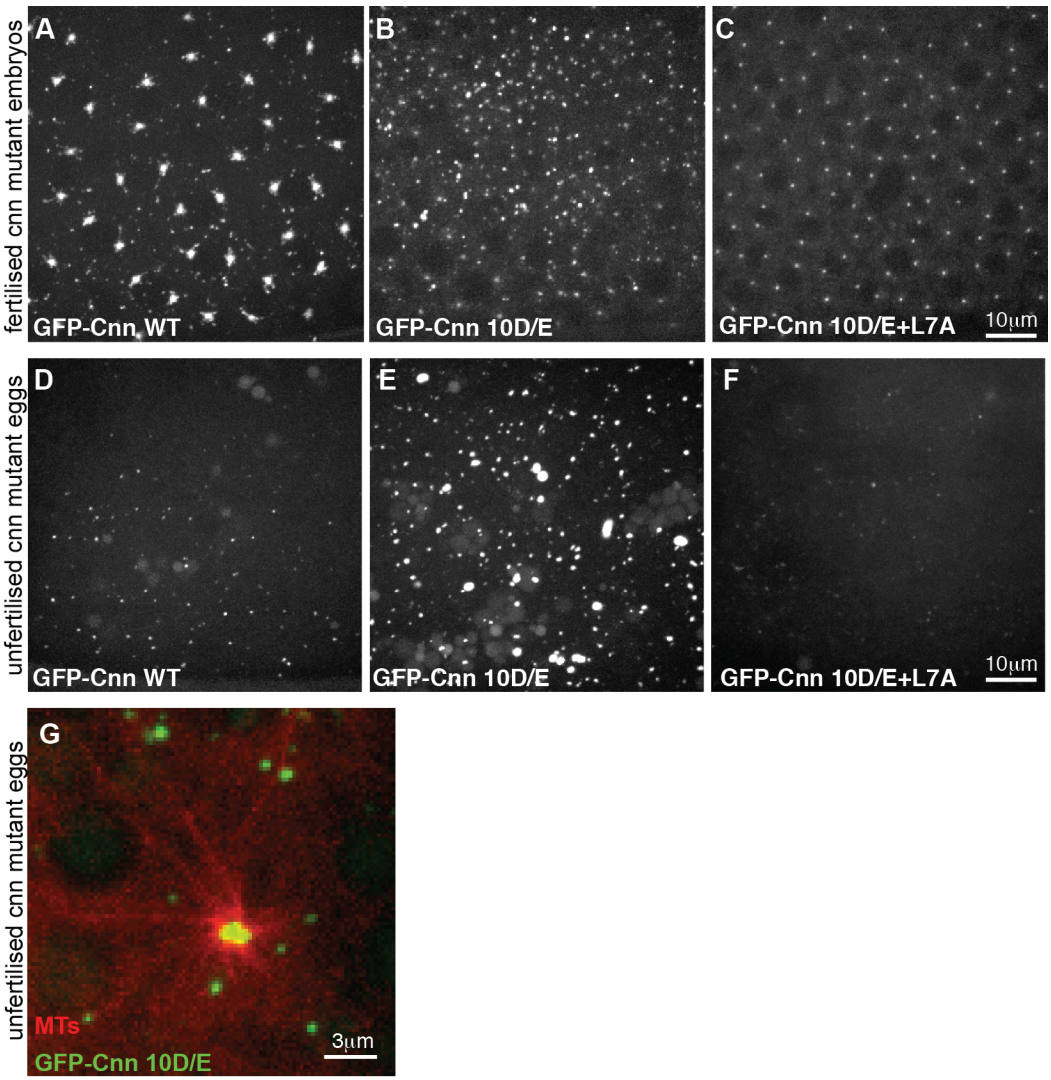


Figure 2.21. Phosphorylation within the Cnn PReM domain promotes Cnn scaffold formation in a LZ-dependent manner. (**A-C**) Confocal images showing cytosolic foci formed by GFP-Cnn 10D/E, which were not observed when the LZ was mutated. Note that some cytosolic foci of GFP-Cnn WT were seen in fertilised embryos. These are flares that have broken away from the PCM periphery (**Chapter 1.6.3**). In order to distinguish between the many cytoplasmic foci formed by GFP-Cnn10D/E and the centrosomes, mRNAs were injected into unfertilised *cnn*$^{-/-}$ null mutant eggs that do not have any centrosome (**D-F**). Whereas GFP-Cnn WT formed small foci in the cytoplasm, GFP-Cnn 10D/E formed larger cytosolic aggregates even in the absence of centrosomes. GFP-Cnn 10D/E+LZA formed fewer foci, which were of much weaker fluorescence intensity. These GFP-Cnn 10D/E foci could also organise MT asters (**G**). Images were taken using a Perkin Elmer ERS spinning disk (Volocity software) mounted on a Zeiss Axiovert microscope using a 63X/ 1.4NA oil immersion objective and on Orca ER CCD camera (Hamamatsu Photonics, Japan). Images shown are maximum intensity projections of several z-slices. All images were taken and processed by Dr Paul T. Conduit.

To confirm that GFP-Cnn 10D/E could form scaffold-like structures independently of centrosomes, we injected the mRNA into unfertilised eggs, which lack centrosomes. Unlike GFP-Cnn WT, that occasionally formed small foci in unfertilised eggs, GFP-Cnn 10D/E often formed large aggregates whose formation was strongly reduced in the presence of the LZ mutations (**Figure 2.21D-F**). Surprisingly, those large cytoplasmic foci formed by GFP-Cnn 10D/E also organised MT asters in unfertilised eggs, indicating that these phosphorylation-driven protein multimers can organise MTs even in the absence of centrosomes (**Figure 2.21G**).

2.5.7 Perspectives on *in vivo* experiments

Taken together, our *in vivo* data provide important insights into how Cnn scaffolds might assemble around centrioles. Phosphorylation of the PReM domain regulates the formation of Cnn scaffolds that extend throughout the PCM. Mutating the 10 conserved Ser/Thr sites within the PReM domain strongly impaired the ability of Cnn to assemble into scaffold-like structures (**Figure 2.17**). This in turn resulted in a severely disrupted PCM assembly, as shown by the reduced amount of Spd-2-RFP incorporation to centrioles (**Figure 2.18**) and the insufficient scaffold formation. In contrast, mutating the 10 Ser/Thr sites to phosphomimetic residues (10D/E) significantly increased the efficiency of Cnn scaffold formation, as GFP-Cnn 10D/E could form spontaneous aggregates in the cytoplasm (**Figure 2.21**). Such protein aggregates could be assembled even without centrosomes and can organise MT asters (**Figure 2.21G**). These observations indicated that the phosphomimetic mutant protein no longer has to localise to centrioles where Cnn normally gets phosphorylated. Phosphorylation could be considered as a switch that coverts Cnn into an assembly-competent form, and this assembly is LZ-dependent.

In flies, Polo kinase is highly concentrated around the mother centriole throughout the cell cycle [Fu and Glover, 2012]. This could help limit phosphorylation of Cnn to the area around the mother centrioles, and so explain why Cnn scaffold assembly normally only occurs at this site. In human cells, PLK1 kinase is activated at G2/M transition [Petronczki et al., 2008; Seki et al., 2008]. Polo kinase could also be activated at the beginning of mitosis in flies, thus phosphorylating Cnn to initiate scaffold assembly and centrosome maturation.

Nonetheless, there are several questions from this *in vivo* study remaining to be answered. The extent of Spd-2 spread into the PCM is always smaller than that observed for Cnn because Cnn scaffolds are pulled further away by MTs [Conduit et al., 2014b]. Interestingly, the fluorescence signals from GFP-Cnn 10A and Spd-2-RFP seemed to largely co-localise around centrioles. This result hinted that Cnn might be involved in regulating the extent of Spd-2 assembly at centrosomes, but the underlying mechanism is unclear.

How does phosphorylation trigger Cnn oligomerisation? Without any structural information it is difficult to predict a detailed mechanism, and therefore I attempted to structurally study the PReM domain using X-ray crystallography.

2.6 Purification and characterisation of the phosphomimetic PReM domain constructs

2.6.1 Purification and characterisation of Cnn$^{403-608}$ 10D/E

I first tried to purify His6-tagged Cnn$^{455-622}$ WT protein, but its yield was very low (**Figure 2.6**). The use of an N-terminal MBP tag significantly increased the protein yield (**Figure 2.8; Figure 2.9**) and SEC-MALS indicated that MBPCnn$^{403-608}$ and MBPCnn$^{403-608}$ 10D/E formed stable dimers and pentameric structures at various concentrations. I wanted to structurally characterise the PReM domain with or without the 10D/E mutation, in order to understand the role of phosphorylation in Cnn multimerisation. The wild-type construct, however, was insoluble after the MBP tag cleavage and was not pursued for structural studies. As a result, I only document the *in vitro* characterisation of Cnn$^{403-608}$ 10D/E in the following sections.

The cDNA sequence encoding Cnn$^{403-608}$ 10D/E was initially cloned into a pMAL-C2X gateway vector (**Section 2.2.3**), which allowed an easy transfer of DNA inserts between plasmids using Gateway cloning. However, the linker between the N-terminal MBP tag and Cnn$^{403-608}$ 10D/E contains a Factor Xa protease cleavage site (**Figure 2.22A**). Because this protease has to be purchased commercially and a large amount would be required, I decided to clone this construct into a pETM-44 vector. The expressed protein would then contain a linker between MBP tag and Cnn$^{403-608}$ 10D/E encompassing a 3C protease cleavage site (**Figure 2.22A**; **Table 2.4**). I expressed and purified 3C protease in house.

Construct	MW (kDa)	pI	Abs 0.1%	No. of Cys
MBPCnn$^{403-608}$ 10D/E	66.50	5.07	1.208	3
Cnn$^{403-608}$ 10D/E	24.03	4.71	0.582	3

Table 2.4 Details of the MBPCnn$^{403-608}$ 10DE construct before and after the tag cleavage. Calculations were carried out using Protparam [Wilkins et al., 1999]. Due to the presence of cysteine residues, 5 mM β-mercaptoethanol (BME) was included in working buffer (WB) to prevent disulfide bonds.

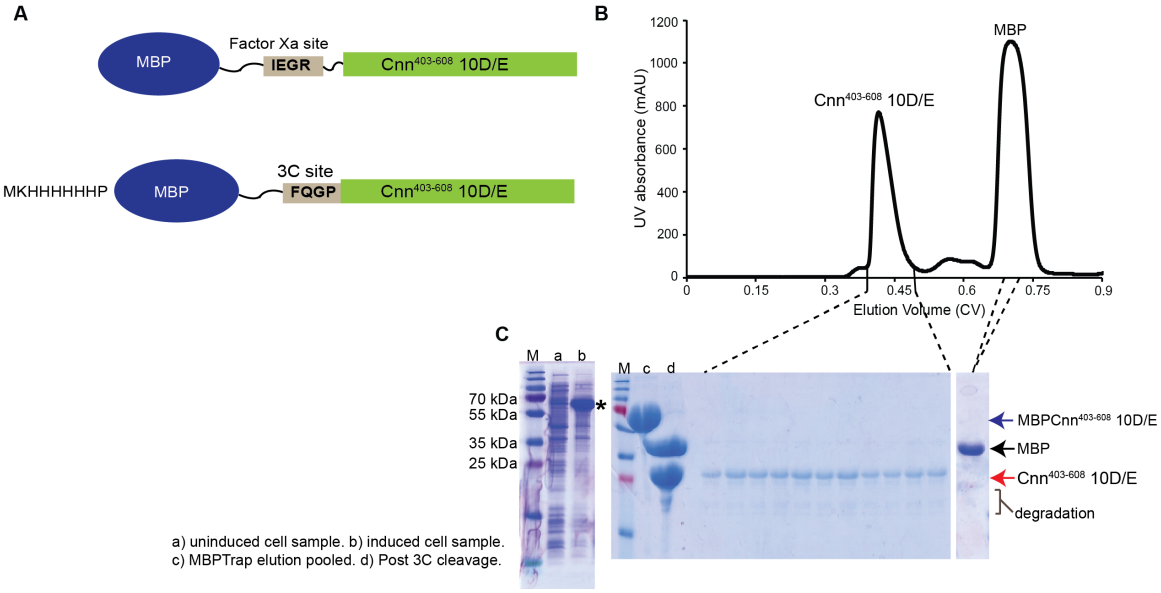


Figure 2.22 Purification of Cnn$^{403-608}$ 10D/E. (**A**) Schematics of MBPCnn$^{403-608}$ 10D/E expressed from pMAL-C2X vector (top) or pETM-44 vector (bottom). Factor Xa protease cleavage leaves a linker of 11 residues N-terminal to the Cnn protein sequence. (**B**) SEC profile from an S200 26/60 column showing the elution of Cnn$^{403-608}$ 10D/E around 0.42 CV. Elution with 50 mM Tris (pH7.5), 150 mM NaCl and 5 mM BME. The cleaved MBP tag eluted at ~0.7 CV and so was well separated. (**C**) 15% reducing SDS-PAGE gel analysis of SEC fractions showing the elution of Cnn$^{403-608}$ 10D/E within the first major peak. A ladder of lower molecular weight bands was also observed. M: PageRuler Plus Prestained Protein Ladder (Thermo Scientific). Asterisk indicates the band corresponding to overexpressed MBPCnn$^{403-608}$ 10D/E.

I expressed MBPCnn$^{403-608}$ 10D/E (cloned into pETM-44 vector) using Tuner (DE3) cells in 4 l of TB medium. Protein was purified using an MBPTrap HP affinity column (GE Healthcare, **Appendix 8.1.7**). 3C protease was added to MBPTrap HP affinity purified eluate at a protease to target ratio of 1 mg: 1.48 µmol, and the reaction was dialysed against fresh WB buffer overnight at 4 °C (**Appendix Table 8.8**). The untagged Cnn 10D/E protein was further purified by size exclusion chromatography. The cleavage reaction was concentrated to 5 ml, using a 3 kDa MWCO centrifugal concentrator (Amicon), before loading onto an S200 26/60 column (GE Healthcare). Size exclusion chromatography was performed using WB as the running buffer. Cnn$^{403-608}$ 10D/E was well resolved from the MBP peak (**Figure 2.22B**). The 3C protease was eluted close to the void volume (void volume of an S200 26/60 column= ~0.33 CV). Gel samples were taken for each elution peak, and fractions containing purified Cnn$^{403-608}$ 10D/E were pooled for MALS analysis (**Figure 2.22C**). A range of protein concentrations was prepared: 40 µM, 200 µM, 400 µM and 800 µM, using a 3 kDa MWCO centrifugal concentrator (Amicon). Samples were then injected onto a Superose6 column (GE Healthcare) using the same running buffer as before on MBPCnn$^{403-608}$ 10D/E. Whereas the MBPCnn$^{403-608}$ 10D/E fusion protein eluted consistently as a pentamer (**Figure 2.13B**), Cnn$^{403-608}$ 10D/E alone appeared to assemble into oligomers that increased in size with the increasing protein concentration (**Figure 2.23**). As the protein

concentration doubled, one more copy of protein seemed to be added onto the existing protein complex. For example, the average mass estimation increased from an octamer to a nonamer when the protein concentration doubled from 200 µm to 400 µM. This behaviour was unexpected as Cnn WT was shown to dimerise via the LZ in solution. It should be noted, however, that the estimated mass across each elution peak is an average of all species eluted within that peak, so a range of multimers may be present in different ratios within the same peak. Doubling the protein concentration could have changed the equilibrium state of the oligomers. As a result, an increase in molecular weight corresponding to a monomeric mass does not necessarily mean only a monomer was added to an existing protein complex.

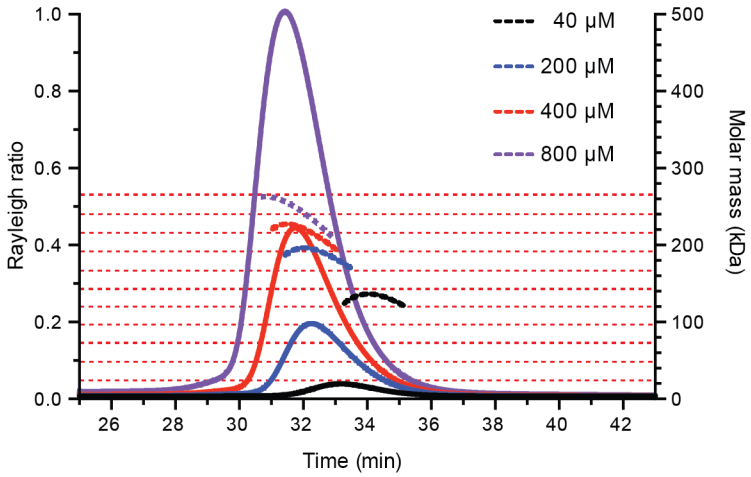


Figure 2.23 SEC-MALS analysis of Cnn$^{403-608}$ 10D/E. Elution with 50 mM Tris (pH7.5), 150 mM NaCl and 5 mM BME. Solid lines show the relative Rayleigh ratio; dashed lines show the observed average mass across the peak. Expected masses of a monomer and successive multimers are indicated with *red* dotted lines across the graph. Unlike MBPCnn$^{403-608}$ 10D/E, which stayed consistently as a pentamer at various concentrations, Cnn$^{403-608}$ 10D/E alone assembled into larger protein complexes (ranging from pentameric to decameric) with the increasing protein concentration.

2.6.2 Limited proteolysis of Cnn$^{403-608}$ 10D/E

Gel analysis of SEC eluted fractions indicated minor contaminant bands for Cnn$^{403-608}$ 10D/E, and the intensity of these lower bands increased over time suggesting they were degradation products (**Figure 2.22C**). In order to obtain a core structure for the protein, Cnn$^{403-608}$ 10D/E was subjected to limited proteolysis with trypsin. A trypsin digestion trial was first set up to determine the optimal *w/w* ratio of protease to target protein. After SEC, purified Cnn$^{403-608}$ 10D/E was concentrated to 6.7 mg/ml using a 3 kDa MWCO centrifugal concentrator (Amicon). A 40 µl reaction was then set up as listed in **Table 2.5** with an increasing gradient of trypsin: target protein ratio. Reactions were left at room temperature (RT) for 2 hrs. 3x SDS loading dye was then added to stop the reaction.

w/w trypsin:Cnn ratio	Trypsin concentration	Cnn concentration
Undigested control	/	1 mg/ml
1 to 10,000	100 ng/ml	1 mg/ml
1 to 5,000	200 ng/ml	1 mg/ml
1 to 1,000	1 µg/ml	1 mg/ml
1 to 500	2 µg/ml	1 mg/ml
1 to 250	4 µg/ml	1 mg/ml

Table 2.5 Set up of limited proteolysis reactions for Cnn$^{403-608}$ 10D/E.

Trypsin digested samples were then analysed on a 15% reducing SDS-PAGE gel (**Figure 2.24A**). Undigested protein had already shown a ladder of degraded bands even without the addition of any protease. As more trypsin was added to the reaction, further bands of lower molecular weight appeared, and I could not detect a single band for the core protein structure. Nevertheless, two dominant bands were observed when trypsin was added at a 1 to 500 protease: target protein ratio. I therefore repeated the limited proteolysis (at 1: 500 trypsin: target protein *w/w* ratio) on a larger scale in order to obtain enough protein for mass spectrometry (MS) analysis. The digested products were separated on an S200 16/60 column (GE Healthcare, **Figure 2.24B**). Gel analysis showed that two major protein species plus a minor band in between were eluted within the peak around 0.5 CV (**Figure 2.24C**).

10 µl of 1 mg/ml SEC peak elution sample was submitted for MS at the Department of Chemistry, University of Oxford. The two major digested products were detected with MS, and the masses were 17902 Da and 13112 Da. Upon comparison to the sequence of Cnn$^{403-608}$ 10D/E, I calculated that these species could comprise residues 403 or 443 to 557 of Cnn, indicating that the C-terminal 52 residues containing the 10 potential phosphorylation sites and the N-terminal 40 residues have been removed (**Figure 2.24D**). Despite the phosphomimetic region of Cnn being removed by limited proteolysis, I decided to include this region for further study of the phosphorylation mechanism.

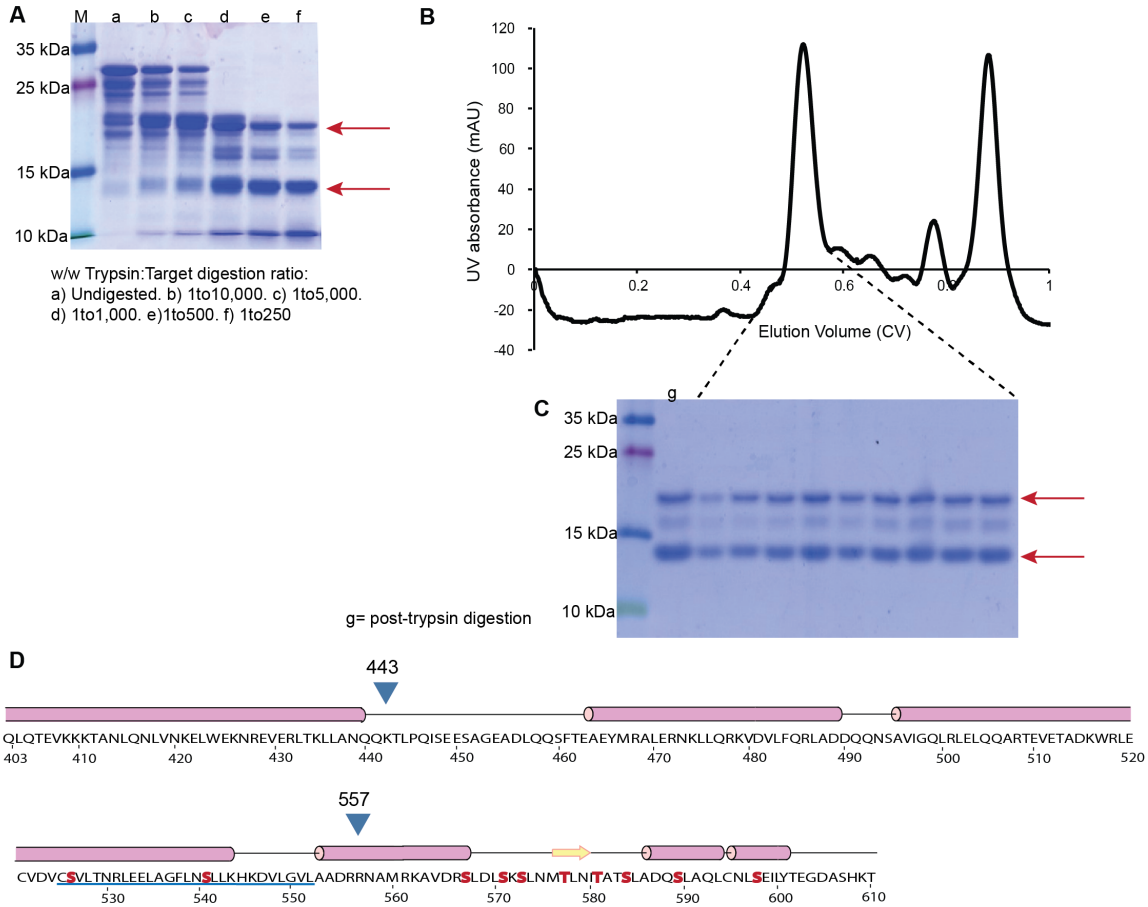


Figure 2.24 Limited proteolysis analysis of Cnn$^{403-608}$ 10D/E. (**A**) 15% reducing SDS-PAGE gel analysis showing limited proteolysis of Cnn$^{403-608}$ 10D/E with varying amounts of trypsin in 50 mM Tris (pH7.5), 150 mM NaCl and 5 mM BME. A ladder of digestion bands was observed with increasing amount of trypsin. The full-length protein was not trimmed to a single dominant species representing the core structure, but rather two major bands were observed. The optimal trypsin: protein ratio was 1:500 (lane e). This digestion ratio was used for a large-scale limited proteolysis reaction to obtain enough protein for MS. (**B**) An S200 16/60 column was used to purify the digested sample. Elution with 50 mM Tris (pH7.5), 150 mM NaCl and 5 mM BME. 15% reducing SDS-PAGE gel analysis (**C**) of SEC fractions showed elution of major digestion products (highlighted with arrows) within the first peak. (**D**) Primary sequence of Cnn$^{403-608}$ showing the trypsin cleavage sites (by arrowheads) used during limited proteolysis. The LZ is underlined and the 10 potential phosphorylation sites are highlighted in *red*. M: PageRuler Plus Prestained Protein Ladder (Thermo Scientific).

2.6.3 Purification of Cnn$^{443-608}$ 10D/E and Cnn$^{490-608}$ 10D/E

Based on the result of the limited proteolysis I designed a shorter Cnn construct, Cnn$^{443-608}$ 10D/E. I also cloned a protein fragment with minimal length to cover the PReM domain and not break predicted alpha helices (Cnn$^{490-608}$ 10D/E). Both constructs were expressed using Tuner (DE3) cells in 4 l of TB medium (construct details in **Table 2.6**).

Construct	MW (kDa)	pI	Abs 0.1%	No. of Cys
Cnn$^{443-608}$ 10D/E	19.23	4.38	0.441	3
Cnn$^{490-608}$ 10D/E	13.81	4.31	0.506	3

Table 2.6 Details of the Cnn 10D/E constructs expressed and purified for structural characterisation. Calculations were carried out using Protparam [Wilkins et al., 1999]. Due to the presence of cysteine residues, 5 mM β-mercaptoethanol (BME) was included in working buffer (WB) to prevent disulfide bonds.

Proteins were purified using an MBPTrap HP affinity column (GE Healthcare, **Appendix 8.1.7**) followed by 3C protease cleavage overnight to cleave off the MBP tag, as described above (**Section 2.6.1**). The cleavage reactions were concentrated to 5 ml, using a 3 kDa MWCO centrifugal concentrator (Amicon), for loading onto an S200 26/60 SEC column (GE Healthcare, **Figure 2.25A,D**). For $Cnn^{490-608}$ 10D/E, I passed the cleavage reaction through a 1 ml GST column followed by a 5 ml Ni-NTA column to remove any 3C protease (which is GST-tagged) and free His6-MBP tag before size exclusion chromatography (**Figure 2.25C**). SDS-PAGE analysis of SEC eluted fractions indicated that both samples still contained lower molecular weight contaminants or degradation products (**Figure 2.25B,E**). In order to obtain protein of higher purity, the pooled SEC fractions were subjected to ion exchange chromatography (IEX, **Figure 2.26**).

According to the calculated pI for $Cnn^{443-608}$ 10D/E and $Cnn^{490-608}$ 10D/E (**Table 2.6**), proteins should be negatively charged at pH 7.5. The protein was dialysed overnight into 50 mM Tris-HCl (pH7.5), 50 mM NaCl, 5 mM BME at 4 °C. A 6 ml Resource^{TM} Q (GE Healthcare) or 1mL MonoQ 5/50 column (GE Healthcare) was used to further purify the protein. A very fine elution gradient, 0.6 to 1.0 mM NaCl per ml, was applied to ensure the separation of contaminants/degradation products (**Figure 2.26A,C**). Both IEX profiles showed a very broad elution peak. Gel analysis showed that Cnn 10D/E was eluted across the whole peak with different ratios of full-length protein and lower molecular weight contaminants, which could be degraded protein products (**Figure 2.26B,D**). This suggested that protein of different lengths, due to degradation, might have self associated, and the patch of concentrated negative charges could allow protein to bind the anion exchange column in different ways. Nevertheless, those fractions eluted with a higher salt concentration seemed more enriched with the full-length protein, and were therefore pooled for further characterisation. Gel analysis of the pooled protein samples showed that $Cnn^{490-608}$ 10D/E was at a higher purity than the longer construct (**Figure 2.26E**).

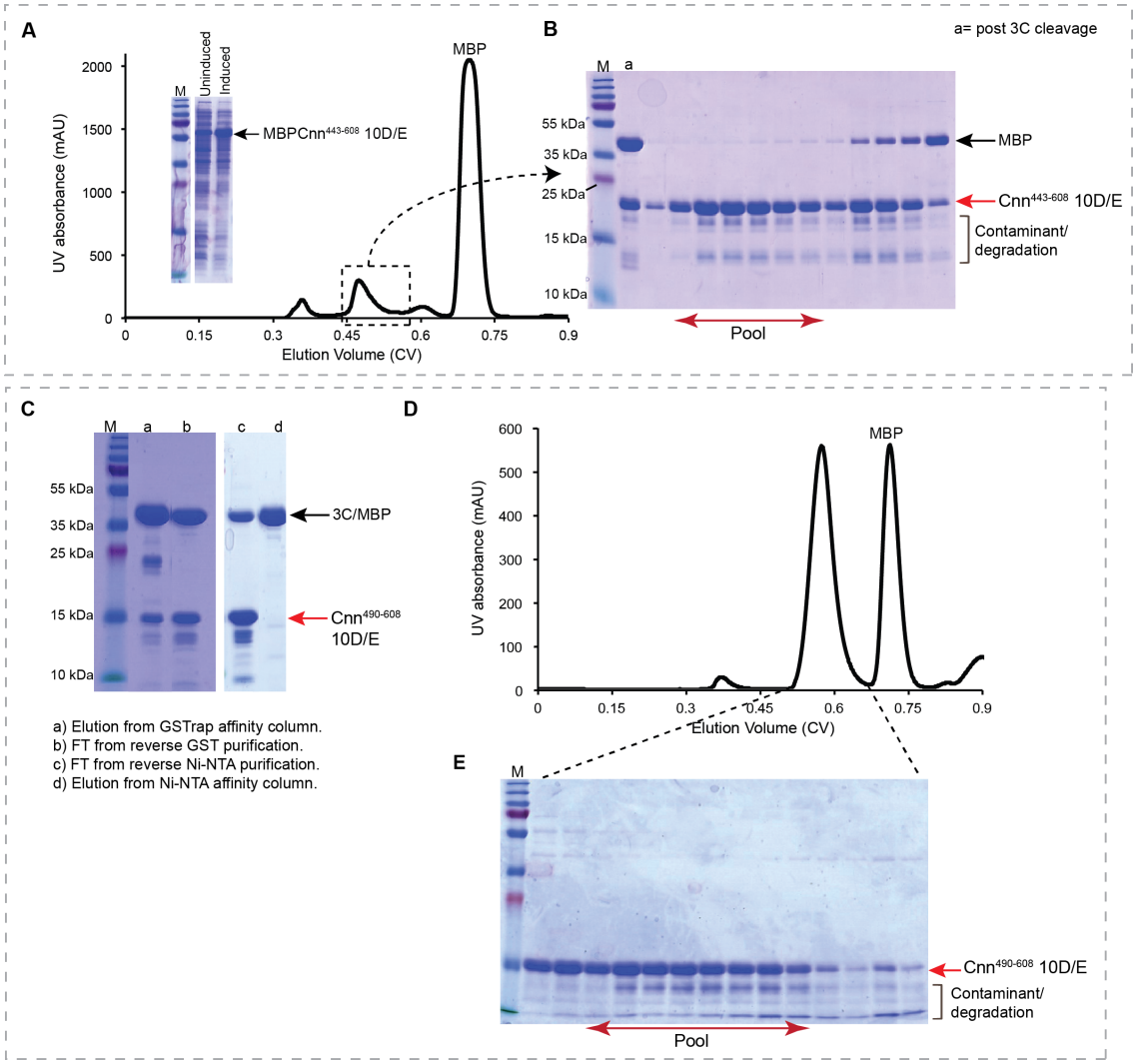


Figure 2.25 Purification of Cnn$^{443/490-608}$ 10D/E. (**A,B**) SEC purification of Cnn$^{443-608}$ 10D/E. Elution with 50 mM Tris (pH7.5), 150 mM NaCl and 5 mM BME. Left, 15% reducing SDS-PAGE gel showing the overexpression of MBPCnn$^{443-608}$ 10D/E (**A**). Right, 15% reducing SDS-PAGE gel analysis of SEC peak fractions (**B**) showing the elution of full-length protein (untagged Cnn$^{443-608}$ 10D/E) across the peak ~0.47 CV (**A,** highlighted with box). However, a ladder of degradation was observed. (**C**) Cnn$^{490-608}$ 10D/E, after MBP cleavage, was purified using a tandem GSTrap-NiNTA affinity column to remove GST-3C protease and free His-MBP tag. 15% reducing SDS-PAGE gel analysis showed the depletion of GST-3C and partial removal of MBP, although some Cnn 10D/E protein was lost due to unspecific binding onto the column. Note that GST-3C is about 46 kDa, similar to the molecular weight of the MBP tag (MW= 44 kDa). (**D,E**) SEC purification for Cnn$^{490-608}$ 10D/E. Elution with 50 mM Tris (pH7.5), 150 mM NaCl and 5 mM BME. SEC trace (**D**) and the subsequent 15% reducing SDS-PAGE analysis (**E**) showed that Cnn$^{490-608}$ 10D/E could be separated out from MBP, although the two elution peaks were very close. Few degradation bands were observed on the gel. For both constructs, highlighted SEC fractions were pooled for further purification by ion exchange chromatography. M: PageRuler Plus Prestained Protein Ladder (Thermo Scientific).

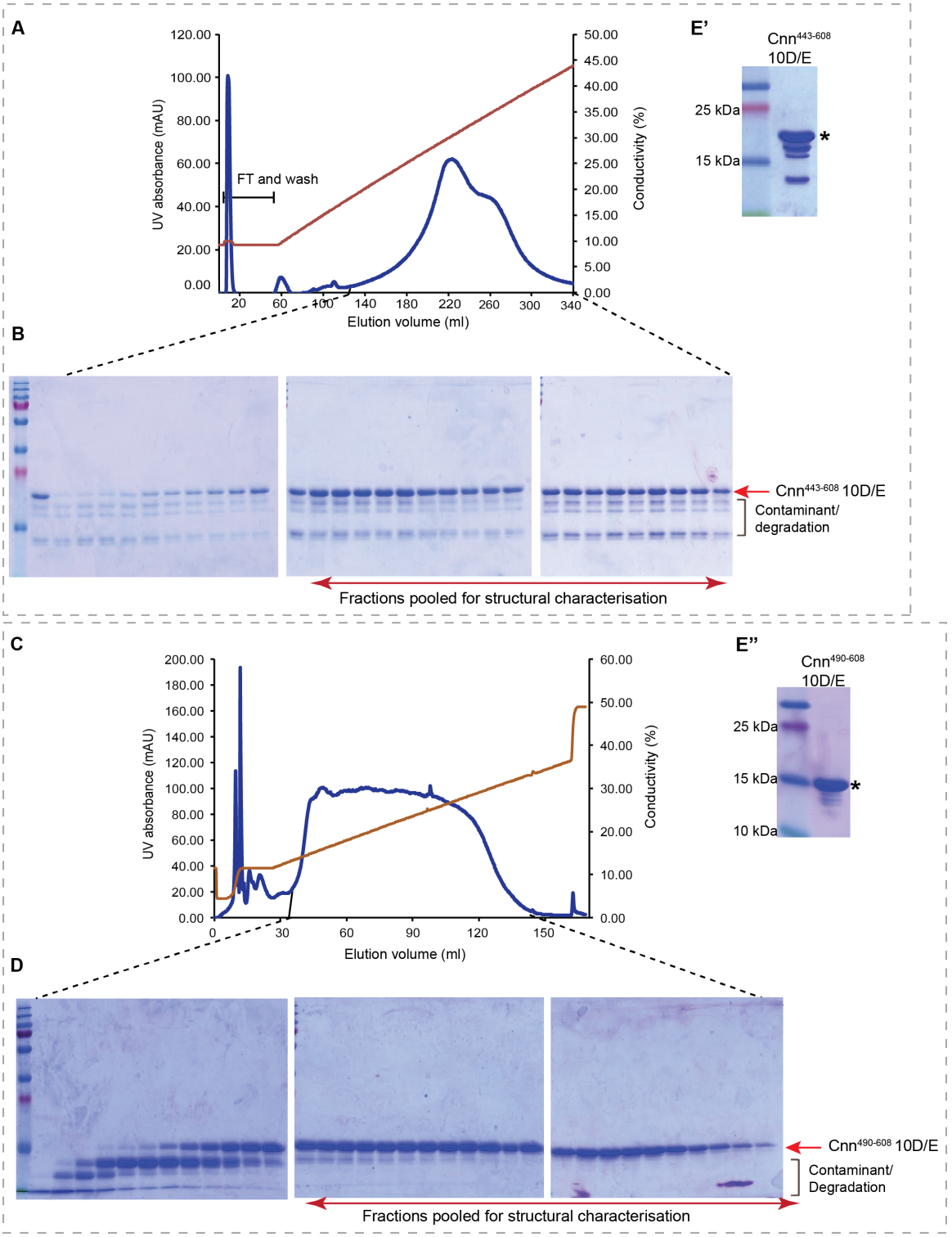


Figure 2.26 Purification of Cnn$^{443/490-608}$ 10D/E by IEX. (**A,B**) A 6ml Resource^{TM} Q column (GE Healthcare) was used to further purity Cnn$^{443-608}$ 10D/E from **Figure 2.24A,B**. A very fine elution gradient, 50 to 225 mM NaCl over 300 ml (~0.6 mM per ml), was applied to ensure a better separation of contaminants (**A**). 15% reducing SDS-PAGE gel analysis (**B**) of fractions across the IEX peak showed the separation of some contaminants/degradation products with a lower salt concentration. (**C,D**) IEX purification for Cnn$^{490-608}$ 10D/E from **Figure 2.25C,D**. A 1 ml monoQ 5/50 column (GE Healthcare) was used. Sample was loaded at 50 mM Tris (pH 7.5), 60 mM NaCl and 5 mM BME, and an elution gradient between 60 to 190mM NaCl over 135 ml (~1 mM per ml) was applied (**C**). Subsequent 15% reducing SDS-PAGE gel analysis (**D**) of elutions across the IEX peak suggested that Cnn$^{490-608}$ 10D/E could be further purified, although a lower band was still seen. (**E**) Purities of the final Cnn 10D/E samples were compared on a 15% reducing SDS-PAGE gel. Asterisks indicate the bands corresponding to Cnn$^{443/490-608}$ 10D/E.

2.6.4 SEC-MALS analysis of truncated Cnn 10D/E constructs

In order to assess the effect of truncation on protein oligomerisation, I prepared samples at 250 μM and 500 μM for MALS experiments that were carried out using the column and running buffer described above (**Section 2.6.1**). SEC-MALS showed that both Cnn$^{443-608}$ 10D/E and Cnn$^{490-608}$ 10D/E were eluted with an estimated mass of a largely tetrameric species (**Figure 2.27**). This was not consistent with the self-interaction state of Cnn$^{403-608}$ 10D/E, which showed formation of a range of multimers with the increasing protein concentration.

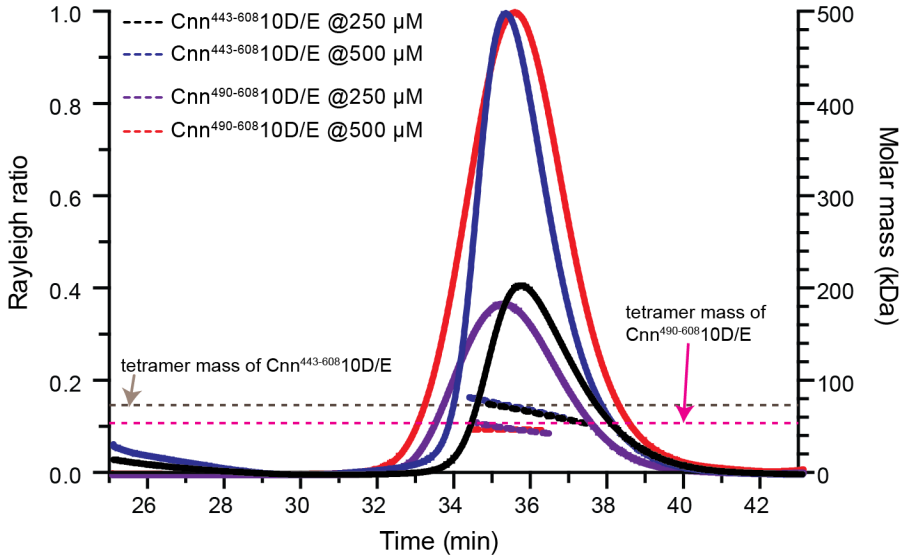


Figure 2.27 SEC-MALS analysis of the truncated Cnn 10D/E constructs. Elution with 50 mM Tris (pH7.5), 150 mM NaCl and 5 mM BME. Cnn$^{443-608}$ 10D/E and Cnn$^{490-608}$ 10D/E both eluted as tetrameric species at tested concentrations. Solid lines represent elution of protein over a Superose 6 column. Dot lines show the average mass estimation across the peak. The expected mass of a tetramer is indicated by the *red* dotted line across the graph.

2.6.5 Circular Dichroism analysis of Cnn 10D/E constructs

Circular Dichroism (CD) analysis was carried out on purified Cnn 10D/E constructs to compare their secondary structure content. The protein was diluted to about 0.1 mg/ml and dialysed overnight into 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.5) at 4 °C. CD spectra were measured at 20 °C on a J-720 machine (Jasco), measuring between 190 to 260 nm, at 0.5 nm intervals. Spectra were averaged over four accumulations, and normalised over a blank measurement from the dialysis buffer. As shown in **Figure 2.28**, all three proteins showed an alpha helical enriched profile consistent with the secondary structural prediction.

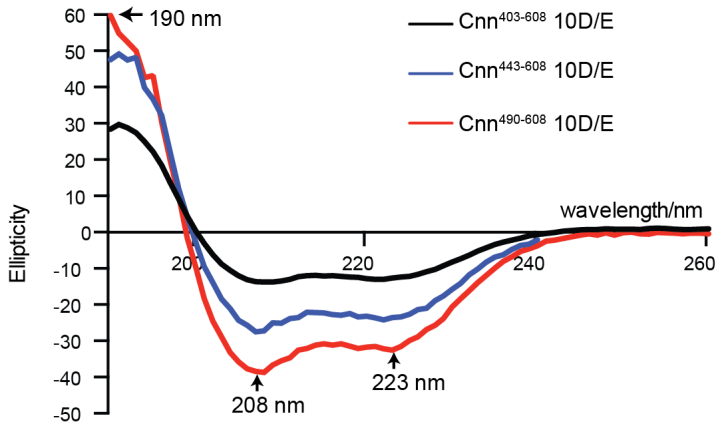


Figure 2.28 Background-subtracted CD spectra of Cnn 10D/E constructs. Analysis was carried out in 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.5) at 4 °C. All three constructs showed high alpha helical content, characterised by two negative bands, at 222 nm and 208 nm, and a positive one at 190 nm. This result agrees with the secondary structure prediction.

2.7 Crystallisation trials with the phosphomimetic PReM domain constructs

2.7.1 Crystallisation of Cnn$^{443-608}$ 10D/E and Cnn$^{490-608}$ 10D/E

Following ion exchange chromatography, Cnn$^{443-608}$ 10D/E and Cnn$^{490-608}$ 10D/E were purer, and both proteins behaved as a homogenous tetramer in solution as shown by MALS. Crystallisation trials were then carried out for both constructs.

Samples were first dialysed into 20 mM Tris-HCl (pH 7.5), 50 mM NaCl and 2 mM TCEP, and then concentrated using a 3 kDa MWCO centrifugal concentrator (Amicon). In order to determine the starting protein concentration for crystallisation trials, a Stura test was conducted. An appropriate concentration is indicated by the presence of precipitation 1 to 2 mins after mixing 1 µL protein solution with 1 µL condition A7 from the Stura Footprint Screen (Molecular Dimensions). The test indicated an initial Cnn 10D/E concentration of 48 mg/ml should be used for crystallisation trials. I then set up broad screens to explore multiple crystallisation conditions (**Appendix 8.1.10**). Crystallisation was carried out using the sitting drop vapour diffusion method. All drops were set up by mixing 0.1 µl or 0.15 µl of protein stock (48 mg/ml) with 0.1 µl or 0.05 µl of mother liquor (0.2 µl drop size), and allowed to equilibrate at 20 °C over a well containing 75 µl of mother liquor (see **Appendix 8.1.10** for detailed procedure). Unfortunately no crystals grew despite some precipitation and phase separation occurring.

2.7.2 Perspectives on crystallisation trials

Crystallisation trials carried out thus far have been unsuccessful, suggesting other crystallisation parameters could be modified. Such variables include lowering the growth temperature, changing the mother liquor: protein solution ratios or increasing the protein concentration. In addition, I could try reductive methylation of Cnn 10D/E protein. Methylation of Lys side chains has been successfully applied to encourage the crystallisation of proteins that were otherwise unable to crystallise despite the high sample homogeneity and purity [Walter et al., 2006]. The Lys side chain is long and flexible, which can impede the packing of ordered crystal lattices. Improving the

sample purity could also facilitate crystallisation, for example by removing the small amount of lower molecular weight contaminants in Cnn$^{443/490-608}$ 10D/E samples (**Figure 2.26E**). In addition, my limited proteolysis analysis showed that the region containing most of the potential phosphorylation sites (Cnn$^{558-608}$ 10D/E) was removed by trypsin digestion, suggesting that Cnn$^{558-608}$ 10D/E may be largely unstructured and therefore inhibit crystallisation. I have previously tried to purify HisCnn$^{455-608}$ WT, but the protein yield was relatively poor. In addition, Cnn$^{403-608}$ WT precipitated out of solution after the MBP tag cleavage (data not shown). It seems likely that phospho-mimicking mutations increased protein stability in solution. Loss of Cnn$^{558-608}$ 10D/E over time in crystallisation trials (due to further degradation) could therefore lead to a resulting protein fragment of poor stability hence inhibiting the formation of an ordered crystal lattice. Nevertheless, it is also possible that the PReM domain alone may not have an intrinsic propensity to crystallise. Cnn is a large coiled coil protein that forms scaffold-like structures *in vivo*. It may need high structural flexibility to accommodate the docking of many other PCM components, but this property could discourage the crystal packing. Moreover, the full-length Cnn protein is 150 kDa, and other regions of Cnn might be required for the stability of the PReM domain, which are missing in my constructs.

2.8 Perspectives

Prior to my project, the significance of *Drosophila* Cnn in PCM assembly had been well documented [Dobbelaere et al., 2008; Lucas and Raff, 2007; Megraw et al., 1999; Megraw et al., 2001]. Previous studies had showed that Cnn seemed to be exclusively phosphorylated at the centrosome in a Polo/PLK1-depedent manner during mitosis. However, the molecular mechanism underlying how Cnn assembles into PCM scaffolds remained unclear.

In this chapter, I used bioinformatics to search for likely functional domains in Cnn that contained potential phosphorylation sites. Secondary structural prediction [Jones, 1999] and COILs prediction [Lupas et al., 1991] both showed DmCnn and HsCDK5RAP2 are coil-rich. MSA of Cnn and its homologs failed to detect highly conserved domains other than CM1 and CM2. Nonetheless, an alignment of Cnn sequences from *Drosophila* species showed a highly conserved region across the middle segment of Cnn ($Cnn^{516-601}$) that contains 10 conserved Ser/Thr sites and a LZ motif (referred to as LZ2). At least 6 of the 10 identified Ser/Thr sites at least partially conform to a consensus PLK1 phosphorylation site [Santamaria et al., 2011].

For *in vitro* and *in vivo* analysis, I generated 10A and 10D/E mutants in which the 10 conserved Ser/Thr sites were mutated to either Ala, or Glu/Asp, respectively. Using an *in vitro* kinase assay, I showed that at least some of the 10 Ser/Thr sites identified with MSA are potential Polo/PLK1 phosphorylation sites. In $cnn^{-/-}$ null mutant embryos, GFP-Cnn 10A failed to assemble into scaffold-like structures at centrosomes, and mutants with different combinations of Ser/Thr sites mutated to Ala showed inefficient formation of Cnn scaffolds to various extents. It seems likely that multisite phosphorylation of Cnn is required for efficient PCM assembly, and this in turn provides a spatial restriction because Polo kinase is only highly concentrated around centrioles.

$MBPCnn^{403-608}$ was shown to form dimers in solution, and this dimerisation is LZ2-dependent. *In vivo* this LZ2 was also essential for building the Cnn scaffolds. SEC-MALS analysis further demonstrated that $MBPCnn^{403-608}$ 10D/E could assemble into stable pentameric structures in an LZ2-dependent manner. Interestingly, once the tag was cleaved, the 10D/E protein could drive formation of complexes even bigger than a pentamer, suggesting steric hindrance from the MBP

tag may be limiting the extent to which it can oligomerise. The phosphorylation-driven multimerisation of the PReM domain is likely important for Cnn scaffold assembly *in vivo*. Mimicking the phosphorylation effect (10D/E) *in vivo* allowed Cnn to spontaneously form prominent cytosolic foci even in the absence of centrosomes. Importantly, these cytoplasmic foci can organise MT asters. In my opinion, this provides the strongest evidence that Cnn can genuinely form a scaffold structure that can recruit PCM and organise MTs, and that the assembly of this scaffold is likely regulated by phosphorylation.

To better understand how Cnn might form a scaffold I attempted to use X-ray crystallography to solve the structure of the PReM domain. Unfortunately, the wild-type PReM domain was expressed poorly and was largely insoluble. After extensive optimisation, however, I was able to obtain large amounts of highly purified phosphomimetic PReM (10D/E) domain. Surprisingly, this domain ($Cnn^{443/490-608}$ 10D/E) only formed a tetramer. This is, however, consistent with the reported oligomeric state of LZ being dimer, trimer or tetramer. Despite the behaviour of this protein as a stable tetramer *in vitro*, I was unable to grow any crystals, so the nature of the PReM domain oligomers remains mysterious.

How might phosphorylation of the PReM domain trigger oligomerisation? First, phosphorylation within the PReM domain might induce local conformational changes leading to activation of intrinsically unstructured protein. I was not able to produce enough quantity of the wild-type PReM domain construct for biophysical characterisation. Therefore, I could not directly compare the folding status of Cnn WT and 10D/E. Second, phosphorylation might 'push open' intertwined Cnn regions. As shown by Y2H results [Conduit et al., 2014b], different regions of Cnn can cross interact (for example, $Cnn^{766-1148}$ interacts with $Cnn^{383-765}$). Introduction of negatively charged phosphate groups into the PReM domain (upon Polo phosphorylation) might disrupt intramolecular contacts converting Cnn from a tight knob-like bundle to an extended 'fibre-like' structure freeing more residues for multimerisation. Finally, phosphorylation within the PReM domain might introduce additional interface for self-interaction with other Cnn regions. A potential clue came with the observation that in the absence of residues 403 to 442, $Cnn^{443-608}$ 10D/E formed a stable tetramer in solution, whereas $Cnn^{403-608}$ 10D/E could oligomerise into complexes of increasing sizes with the higher protein concentration. $Cnn^{403-442}$ contains eight highly

conserved, positively charged residues (Lys/Arg, **Figure 2.7**), which have the potential to form additional salt bridges with the phosphomimetic PReM domain. For future experiments, a truncation mutant with a deletion of this region should be investigated *in vivo* to analyse its potential importance in Cnn scaffold assembly.

Very recently, a similar mechanism of scaffold formation has been observed with SPD-5, a putative functional ortholog of Cnn in *C. elegans* [Woodruff et al., 2015]. Recombinantly purified SPD-5 protein alone can form a porous network *in vitro*, which then recruits SPD-2 and PLK1. SPD-5 is phosphorylated by PLK1 kinase at four central residues, and this phosphorylation is essential for SPD-5 polymerisation *in vitro* and its PCM expansion *in vivo* [Woodruff et al., 2015].

Taken together, studies presented in this chapter and published work in *C. elegans* suggest that Cnn/SPD-5 and Polo/PLK1 contribute to a conserved module in mitotic PCM assembly in flies and worms. We have also attempted to search for any obvious PReM domains in vertebrate Cnn homologs, but could not identify any. Nevertheless, as shown by my bioinformatic analysis, Cnn and its homologs are highly coiled-coil and therefore many of these proteins would fulfill the minimum requirement of a PReM domain– a potential coiled-coil interaction domain and a region encompassing multiple PLK1 target sites.

Conduit et al., measured the dynamic behaviour of Cnn in *Drosophila* embryos, and showed an inside-out movement of Cnn molecules at the centrosome [Conduit et al., 2010], (see **Chapter 1.6.6**). In particular, Cnn molecules are incorporated to the center of the centrosome, closest to centrioles, and then gradually spread throughout the PCM. Results documented in this chapter have provided further understanding into how Cnn scaffolds are built around centrioles. However, the mechanism underlying the centrosomal incorporation of Cnn remains poorly understood and I have attempted to address this topic in **Chapter 3**.

Chapter 3: Structure and function analysis of Cnn Common Motif 2 (CM2)

3.0 Introduction

3.0.1 Common Motif 2 (CM2)

As mentioned in **Introduction** and **Chapter 2.1**, there is an evolutionarily conserved domain at the C-terminus of Cnn known as the Common Motif 2 (CM2). Missense mutations in CM2 of HsCDK5RAP2 significantly reduced its centrosomal localisation in HeLa cells [Wang et al., 2010]. A C-terminal fragment of CDK5RAP2 containing the CM2 region showed strong recruitment to centrioles [Wang et al., 2010]. In flies, an Arg to His mutation within Cnn-CM2 (cnn^{B4} mutation) led to embryonic arrest prior to cellularisation [Kao and Megraw, 2009]. This was likely because centrocortin, a protein essential for cleavage furrow assembly, failed to interact with Cnn. The *B4* mutation has also disrupted Cnn scaffold integrity at the outermost region due to failure in PLP interaction [Lerit et al., 2015]. Recently another single point mutation, T1133E, within CM2 has been shown to severely affect the centrosomal localisation of Cnn by increasing the rate at which Cnn molecules leave the centrosome (Dr Paul T. Conduit, personal communication). The presence of endogenous Cnn partially rescued the phenotype suggesting that wild-type Cnn either directly or indirectly facilitated the accumulation of GFP-Cnn T1133E at the centrosome.

3.0.2 Aims

- Biophysical and structural characterisation of the Cnn-CM2 domain, to provide insight into how CM2 contributes to Cnn function.
- Test the *in vivo* significance of any features identified by these structural studies.

3.0.3 Key findings

Cnn-CM2 forms an unusual asymmetric parallel dimer that appears to further self associate into an antiparallel tetramer. As seen in the crystal structure, the invariant Thr1133 is involved in the tetramer interface, which may provide an explanation for its significance in Cnn scaffold assembly. Interestingly, the CM2 domain exhibits strong binding to zinc (Zn^{2+}) ions, which is structurally important for efficient dimerisation. In fly embryos, CM2 is required both for localisation and stabilisation of the Cnn scaffolds at centrosomes.

3.1 Designing a construct for structural characterisation

As outlined in **Chapter 2.1**, the MSA of Cnn and its homologs showed enriched conservation in the CM1 and CM2 domains (**Figure 2.1**). Due to the low confidence of sequence alignment, I performed another MSA on Cnn sequences only from *Drosophila* species. In both alignments, sequence conservation was observed from position 1082 to 1148 (**Figure2.1B**; **Figure 3.1A**). Based on the MSA, PSIPRED [Jones, 1999], (**Figure 3.1B**) and COILs [Lupas et al., 1991], (**Figure 2.2**) predictions, I have designed Cnn$^{1082\text{-}1148}$ as a construct for biophysical and structural characterisations, and I refer to this as the Cnn-CM2 domain.

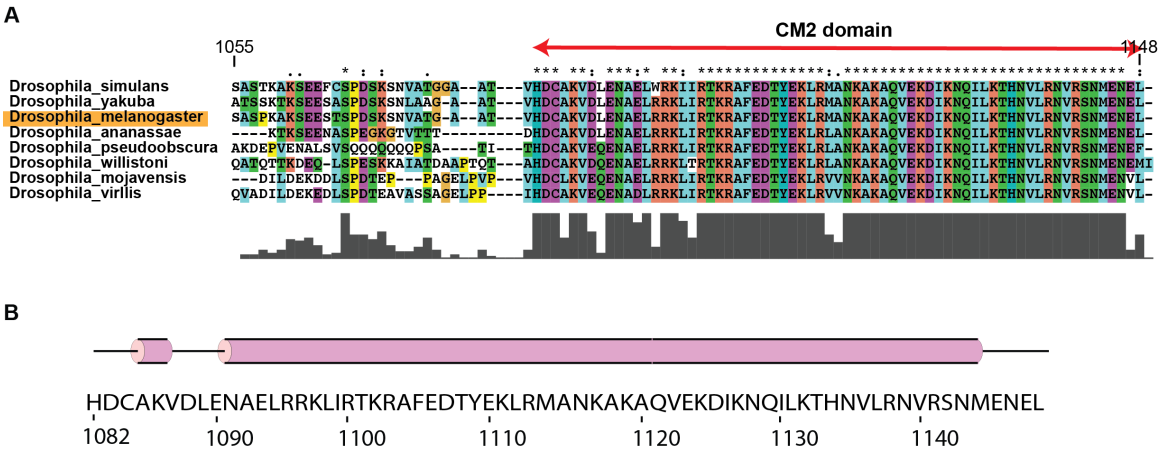


Figure 3.1 Bioinformatic analysis of Cnn-CM2. (**A**) MSA of Cnn sequences from *Drosophila* species showing the highly conserved CM2 domain comprising residues 1082 to 1148. Residue numbers for *D. melanogaster* are indicated. Invariant residues are highlighted with asterisks, conserved residues with colons and semi-conserved residues with full stops. Residues are coloured according to the ClustalX colour scheme. (**B**) PSIPRED secondary structural prediction for the Cnn-CM2 domain.

3.2 Purification and characterisation of Cnn-CM2

3.2.1 Purification of Cnn-CM2

Due to its small size, the Cnn$^{1082-1148}$ construct was cloned into a custom vector (pLip vector) to facilitate protein expression (The pLip vector was obtained from Mark Allen (LMB-MRC)). Recombinantly expressed protein will contain a TEV-cleavable His-tagged Lipoyl domain (from *Bacillus stearothermophilus dihydrolipoamide acetyltransferase*) fused to both the N- and C-termini of the target protein (**Figure 3.2**). The highly soluble Lipoyl domains allow a higher yield of fusion protein and, once purified, they can be cleaved off by TEV protease to isolate the target protein. The Lipoyl domains on both termini of protein also act as protection against proteolytic cleavage. A fragment encoding Cnn$^{1082-1148}$ was PCR amplified from cDNA, double digested with BamHI (Roche) and EcoRI (Roche), and inserted into the pLip vector backbone using T4 ligation (NEB). The construct details are summarised in **Table 3.1**.

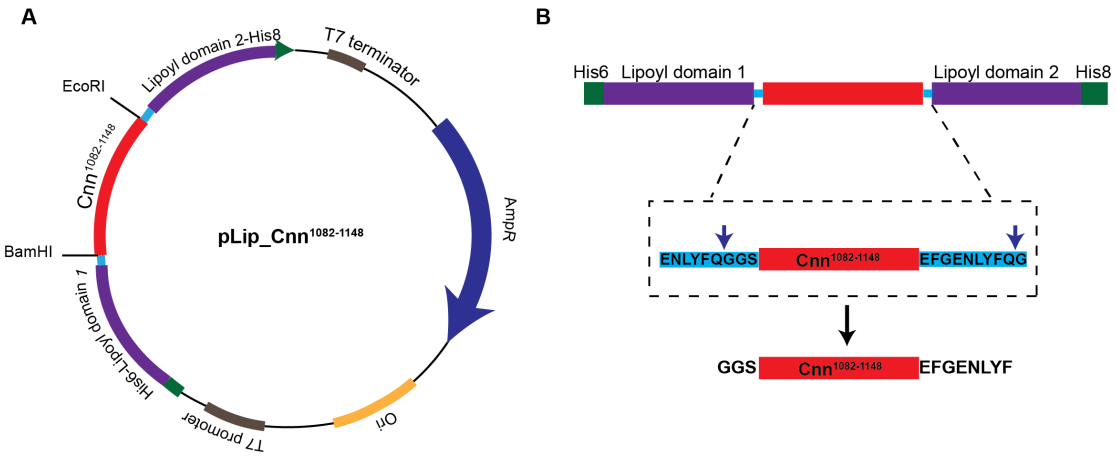
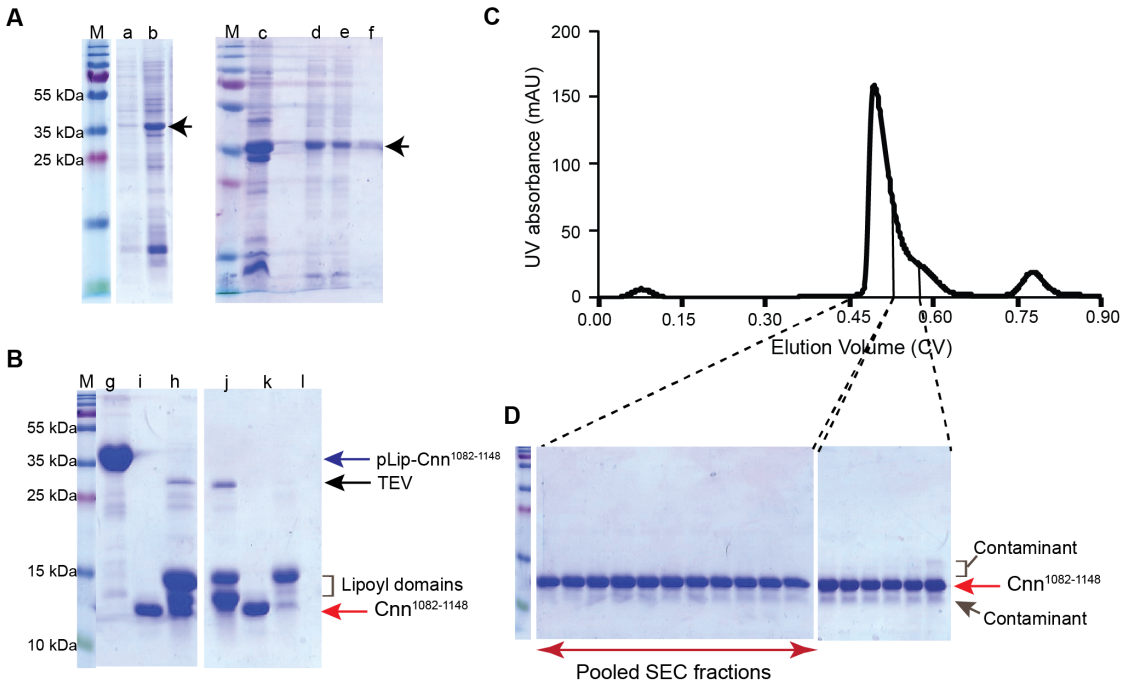


Figure 3.2 Cloning and expression of pLip-Cnn$^{1082-1148}$. (**A**) The pLip plasmid map showing the ORF for Cnn$^{1082-1148}$ cloned in frame between the 2 Lipoyl domains. (**B**) The expression product from the pLip plasmid, target protein flanked by cleavable His-containing Lipoyl domains. TEV recognition sequences are displayed with the position of cleavage pointed by *blue* arrows. The isolated target protein carries residual sequences from the cleavage sites. Diagram modified from [Cottee, 2014].

Construct	MW (kDa)	pI	Abs 0.1%	No. of Cys
diLipoyl-Cnn$^{1082-1148}$	31.43	5.22	0.587	3
Cnn$^{1082-1148}$	9.31	9.17	0.32	1

Table 3.1 Details of Cnn$^{1082-1148}$ with and without Lipoyl domains calculated using Protparam [Wilkins et al., 1999]. Due to the presence of cysteine residues, 5 mM β-mercaptoethanol (BME) was included in working buffer (WB) to prevent disulfide bonds.

The pLip_Cnn$^{1082-1148}$ plasmid was transformed into *E. coli* C41 cells, which were then grown in TB medium. The fusion protein was expressed and purified according to the standard protocol (**Appendix 8.1.7**), using Ni-NTA affinity chromatography (**Figure 3.3A**). Recombinant TEV protease was added to the Ni-NTA purified diLipoyl-Cnn$^{1082-1148}$ at a ratio of 1 mg TEV to 60 mg of fusion protein. The reaction was incubated at 4 °C overnight dialysing into 50 mM Tris-HCl (pH 7.0), 300 mM NaCl, 5 mM BME. Since the TEV protease and Lipoyl domains are both His-tagged, Cnn$^{1082-1148}$ was isolated by a reverse Ni-NTA purification (i.e. the cleavage reaction was run back over a Ni-NTA column; **Figure 3.3B**). I added 30 mM imidazole to the cleavage reaction in order to reduce the non-specific binding of Cnn. The cleaved sample was then passed twice through the Ni-NTA column for separation of untagged Cnn-CM2 protein (which was collected in the FT).



M: Pre-stained PageRuler molecular weight marker.
a) Uninduced cells b) Induced cells. c) Pellet. d) Supernatant.
e) Ni-NTA Flowthrough. f) Wash. g) Pooled Ni-NTA elution fractions. h) Post TEV cleavage.
i) FT1 from reverse Ni purification. j) First elution from reverse Ni purification. k) FT2 from reverse Ni purification.
l) Second elution from reverse Ni purification

Figure 3.3 Purification of Cnn$^{1082-1148}$. (**A**) 15% reducing SDS-PAGE gel analysis of Ni-NTA purification fractions. Band corresponding to pLip-Cnn$^{1082-1148}$ is indicated by arrow. (**B**) 15% reducing SDS-PAGE gel analysis of reverse Ni-NTA purification demonstrating the removal of TEV protease and Lipoyl domains by passing the cleavage reaction through the column twice. Untagged Cnn$^{1082-1148}$ was in the FT, although a minor proportion unspecifically bound to the Ni column. (**C,D**) SEC trace of an S75 26/60 column and 15% reducing SDS-PAGE gel analysis showing the elution of purified Cnn$^{1082-1148}$ across the peak around 0.5 CV with minor contaminant bands seen towards the tail of the peak. Elution with 20 mM Tris (pH7.0), 150 mM NaCl and 1 mM TCEP. The pooled fractions are highlighted on the gel.

The reverse Ni purified protein was concentrated using a 3 kDa MWCO centrifugal concentrator (Amicon) to 30 ml for SEC using 20 mM Tris-HCl (pH 7.0), 150 mM NaCl, 1 mM TCEP as the

running buffer. Protein precipitation was noticed during concentration indicative of its low solubility (~300 µM). As a result, I had to carry out three SEC runs using an S75 26/60 column (GE Healthcare, 10 ml injection per run). Gel analysis of fractions across the SEC peak indicated that Cnn$^{1082-1148}$ was purified within the major peak around 0.5 CV (**Figure 3.3C,D**). A small peak around 0.57 CV was observed overlapping with the major elution. This peak contained Cnn$^{1082-1148}$ with minor contaminants. I therefore only pooled SEC fractions within the main peak trying to avoid any contamination (**Figure 3.3C,D**). 10 µl of the final protein sample was sent for MS, which confirmed that only the expected protein fragment was detectable. The typical yield of final purified Cnn$^{1082-1148}$ was 30 mg per 2 l culture.

3.2.2 Biophysical characterisation of Cnn-CM2

To assess the oligomeric state of purified Cnn$^{1082-1148}$, I concentrated the protein to 220 µM for MALS analysis using WB as the running buffer. A SEC-MALS trace from an S200 10/30 column (GE Healthcare) suggested Cnn-CM2 formed a dimer in solution (**Figure 3.4**).

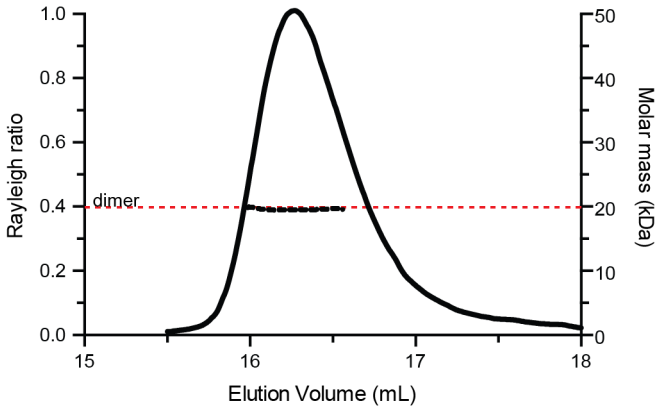


Figure 3.4 SEC-MALS analysis of Cnn$^{1082-1148}$ showing the elution of a dimer from an S200 10/30 column. Elution with 50 mM Tris (pH7.5), 150 mM NaCl and 5 mM BME. The solid line represents the relative Rayleigh ratio, and the dashed line represents the observed mass across the elution peak. The expected mass of a dimer is indicated.

I then carried out a CD experiment to confirm the predicted secondary structure content of Cnn$^{1082-1148}$. The sample was diluted to 0.4 mg/ml and dialysed overnight into 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.0) at 4 °C. CD spectra were collected in four accumulations from 260 to 190 nm at 20 °C, and a blank measurement was taken with the dialysis buffer for background subtraction. Two negative bands, at 208 nm and 222 nm, were observed on the spectrum

suggesting the protein was α-helical (**Figure 3.5A**). In order to further investigate the stability of this structure, I performed a melting CD analysis with an increasing temperature gradient from 4 to 80 °C (**Figure 3.5B**). CD spectra were collected at 2 °C temperature increments scanning between 195 and 240 nm. With the increasing temperature, the spectrum characteristic of an α-helical structure was shown to shift towards one representing a mixture of α-helix and random coil. This change suggested a gradual unfolding and that Cnn$^{1082\text{-}1148}$ might not have completely unfolded even at 80 °C. The plot of ellipticity measured at 222 nm against temperature further indicated that the sample has not reached its full thermal denaturation (**Figure 3.5C**). The sample was then allowed to cool down to 20 °C and another CD spectrum was measured. The trace closely matched to the pre-melt experiment (**Figure 3.5A**). Overall, these results suggest that Cnn$^{1082\text{-}1148}$ is α-helical and extremely stable in solution.

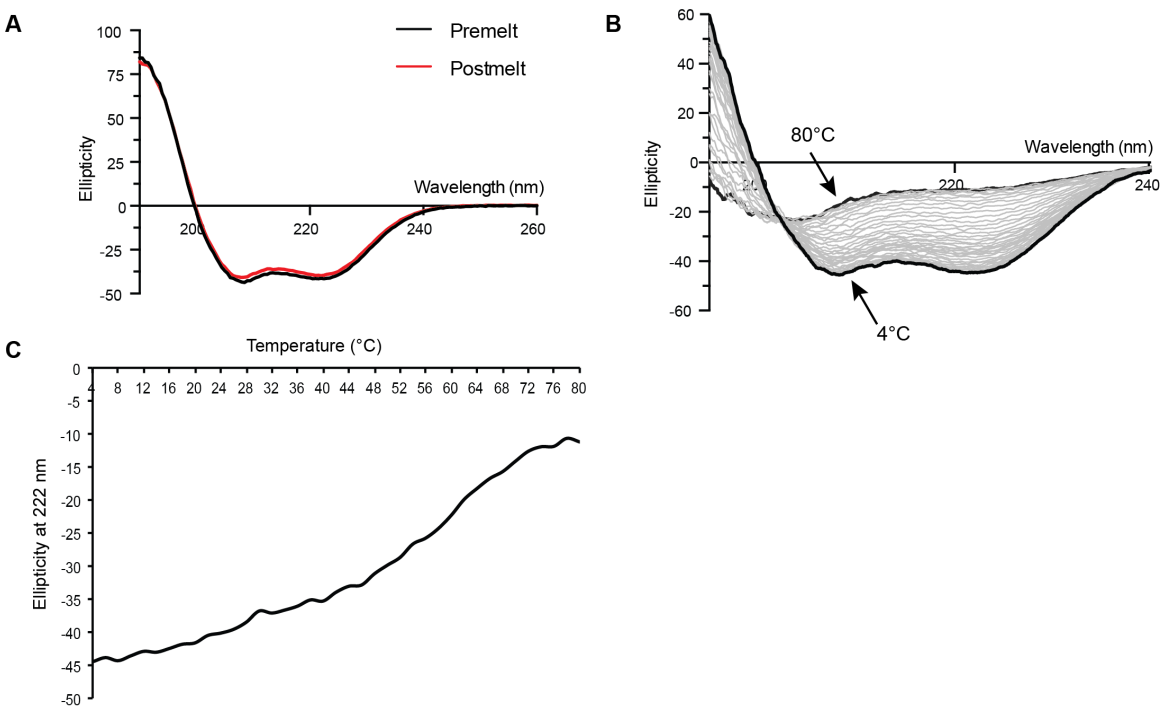


Figure 3.5 CD analysis of Cnn$^{1082\text{-}1148}$ in 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.0)**.** (**A**) Premelt and postmelt CD spectra demonstrating high α-helical content within Cnn-CM2. The Cnn sample, after melting CD analysis, has cooled down to 20 °C for another CD scan that showed the same trace as the premelt sample. (**B**) CD spectra collected at 2 °C temperature increments from 4 to 80 °C showed gradual shift towards a spectrum representing a mixture of α-helix and random coil. (**C**) Plot of ellipticity measured at 222nm in (**B**) against temperature. The unfolding transition has not reached a plateau even at 80 °C, although the ellipcility reading starts to level off around 72 °C.

3.3 Crystallisation of Cnn-CM2

I carried out crystallisation trials on the SEC purified Cnn$^{1082-1148}$. This construct was not highly soluble in solution, and I managed to concentrate it maximally to ~3 mg/ml. However, some fine precipitate was seen at this concentration that could be resolubilised upon dilution. The protein sample was centrifuged at 15,000x g for 2 mins to pellet any precipitate, and only the supernatant was carefully pipetted out for crystallisation trials. The final protein concentration, after removal of precipitate, was measured with a nanodrop (Thermo Scientific). Crystallisation trays were set up with a drop size of 0.3 µl using a Mosquito LCP robot (TTPLabtech). The starting concentration of Cnn$^{1082-1148}$ was 3 mg/ml, and 2 protein: mother liquor ratios of 75% and 50% *v/v* were applied. Diamond-shaped or plate-like crystals grew in several conditions overnight, and more were observed after 2-3 days (**Table 3.2**). Crystals of different appearances (**Figure 3.6**) were harvested for data collection, with the indicated cryoprotectants (**Table 3.2**).

Crystal	Mother liquor	µl protein: µl mother liquor	[EG] in cryoprotectant
VLT_C10	100 mM HEPES (pH 7.0), 200 mM MgCl2, 20% *w/v* PEG6K	0.15:0.15	20% *v/v*
VLT_C11	100 mM HEPES (pH 7.0), 200 mM CaCl2, 20% *w/v* PEG6K	0.15:0.15	20% *v/v*
VLT_C4	100 mM HEPES (pH 7.0), 10% *w/v* PEG6K	0.15:0.15; 0.225:0.075	35% *v/v*
VLT_H7	35% *v/v* dioxane	0.15:0.15; 0.225:0.075	20% *v/v*
VLT_A2	100 mM Buffer 1*, pH6.5, 60 mM divalents*, 30% *w/v* EDO-P8K*	0.225:0.075	Self-protected

Table 3.2 Crystallisation conditions of Cnn$^{1082-1148}$ for crystals fished and frozen for data collection. Buffer System 1 from Morphues® contains 1 M MES/Imidazole mix at pH 6.5; Divalents stock contains $MgCl_2$, $CaCl_2$; EDO-P8K precipitant stock contains 40% EG, 20% PEG8K.

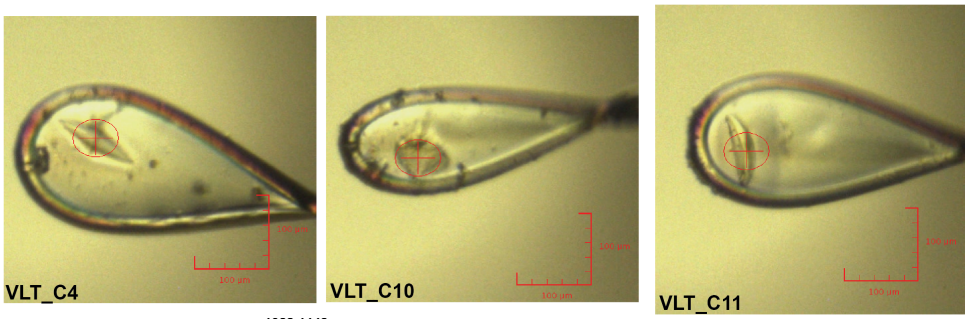


Figure 3.6 Typical crystals of Cnn$^{1082-1148}$ after fishing with a cryo loop. Crystals were grown in conditions listed in **Table3.2**.

3.4 Structural determination of Cnn-CM2

3.4.1 'Native' data collection and processing

Data collection was performed by Prof Susan M. Lea at I04-1 beamline of the Diamond Light Source. Crystals were centred and exposed to an X-ray beam of 0.9174 Å wavelength. Two indexing images were acquired for each crystal, which all showed reflections characteristic of a protein crystal. Automatic strategy calculation was then carried out using Mosflm [Leslie and Powell, 2007] and EDNA [Incardona et al., 2009]. Only the crystal (Crystal VLT_C11, grown in 100 mM HEPES (pH 7.0), 200 mM $CaCl_2$ and 20% *w/v* PEG6K) that diffracted to a higher resolution was then subjected to full data collection (213° of data was collected), and two datasets were collected with exposure times of 0.034 sec and 0.1 sec respectively.

Data were processed using Xia2 in the 3daii mode [Winter, 2010], which used XDS for indexing and integration [Kabsch, 2010] and AIMLESS for scaling and merging [Evans and Murshudov, 2013b]. This produced two datasets, and the high resolution processing (Cnn-CM2_1) is described below. The data quality and resolution limit were assessed using averaged values of R_{pim}, R_{merge}, I/σI and $CC_{1/2}$ (Pearson correlation coefficient). The statistics are summarised in **Table 3.3**. Indexing followed by analysis of the integration in Pointless [Evans, 2006b] illustrated that Cnn-CM2 was crystallised in a primitive hexagonal system with a point group of 6_122 or 6_522. The high-resolution shell data for Cnn-CM2_1 were still reliable, as indicated by $CC_{1/2}$ value of 0.725, I/σI of 2.2 and multiplicity of 22.5 with 100% completeness. A detailed description of the statistics metrics used in this thesis can be found in **Appendix 8.2.2**.

Due to the lack of an obvious homology model, I could not use the molecular replacement method for structure solution in this case. Unexpectedly, a strong anomalous signal was detected in the data as illustrated by the high anomalous correlations shown in **Figure 3.7**, even though the crystal was grown from unlabelled protein sample with no heavy atom included at any stage of protein purification and crystallisation. Additionally, the high-resolution shell exhibited 100% anomalous completeness with high multiplicity of 12.5 and CC_{anom} of 0.103.

Dataset	Cnn-CM2_1 (2.90 Å)
Space group	$P6_122$
Unit cell dimensions (Å)	a = 50.25, b = 50.25, c = 215.97
Unit cell angles	α = β = 90°, γ = 120°
Resolution (Å)	43.52–2.90
No. unique observations (overall/inner/outer)	4130/75/288
Completeness (overall/inner/outer)	99.9/97.0/100.0
Multiplicity (overall/inner/outer)	20.6/11.8/22.5
R_{merge} (overall/inner/outer)	0.108/0.041/1.603
R_{pim} (overall/inner/outer)	0.025/0.012/0.348
I/σI (overall/inner/outer)	21.0/49.3/2.2
Anomalous completeness (overall/inner/outer)	99.9/90.0/100.0
Anomalous multiplicity (overall/inner/outer)	12.4/12.9/12.5
Anomalous correlation (overall/inner/outer)	0.271/0.503/0.103
Mid-slope of anomalous normal probability	1.050

Table 3.3 Statistics from processing of Cnn-CM2 'native' dataset. Note that the R_{merge} and R_{pim} values listed above are for reflections within I+/I- due to the presence of anomalous differences.

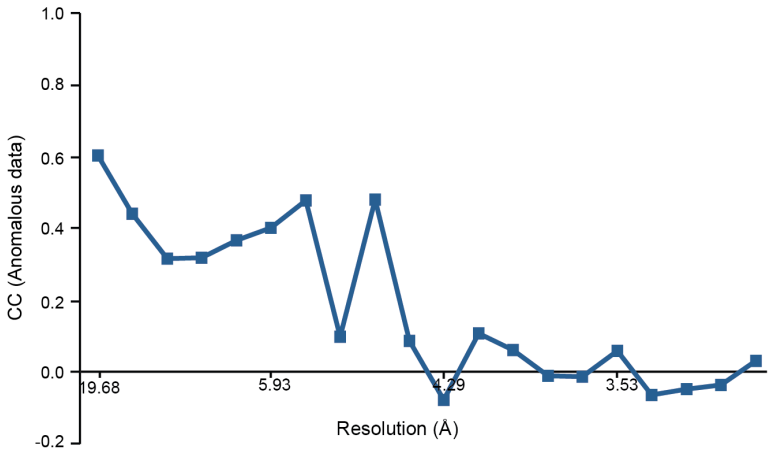


Figure 3.7 The correlation of anomalous signal, between random halves of the data, across different resolution shells for the Cnn-CM2_1 dataset.

In order to identify the anomalous scatter in the crystal, I put down more crystallisation trays at the optimised conditions (**Figure 3.8**) such that the same crystals could be reproduced for identification of the anomalous scatter and for singlewavelength anomalous dispersion (SAD) data collection. Note that Ca^{2+} is present in the crystallisation condition, but its anomalous signal is unlikely to be detected at this wavelength. At 0.9174 Å, the anomalous component (f") for Ca^{2+} is only 0.5046e (calculated by CROSSEC [Cromer, 1983]).

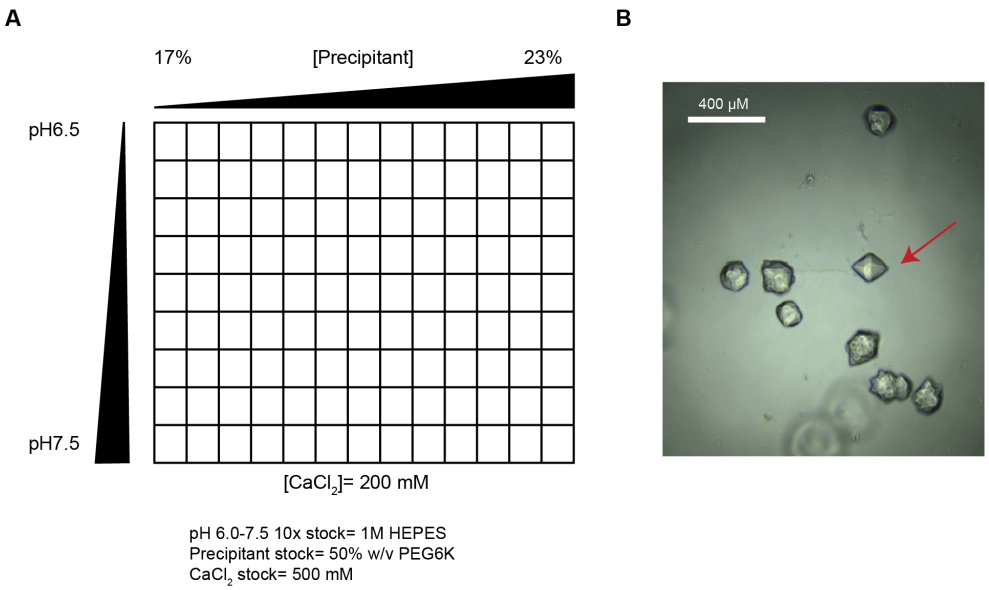


Figure 3.8 Production of more crystals at the optimised conditions. (**A**) Optimisation screen used to obtain crystals for anomalous data collection. Crystal were originally grown in 100 mM HEPES (pH 7.0), 200 mM $CaCl_2$, 20% *w/v* PEG6K. (**B**) Example of crystals grown from the optimisation screen. A single crystal in that drop, indicated by a *red* arrow, was fished for data collection.

3.4.2 Anomalous data collection and processing

An X-ray fluorescence spectrum was first collected to give a broad indication of elements in the sample. Using both 100% and 12.8% beamline transmission, an energy peak between 8,000 and 10,000 eV was detected suggesting it was Zn-K_α (**Figure 3.9A**). This is surprising because no Zn^{2+} ion was supplemented at any stage of protein purification and it must have come from the bacterial cell culture. As the optimal wavelength range for Zn-K_α anomalous diffraction is narrow, an edge scan was then performed to fine tune the wavelength for data collection. The X-ray fluorescence was measured from the sample as a function of the beam energy (**Figure 3.9B**)**.** An increase in fluorescence was observed as the heavy atom in the crystal sample absorbed the beam energy and re-emitted it in the form of fluorescence. The imaginary scattering component, f'', is proportional to this measurable quantity. And the real scattering component, f', can be calculated from f'' using the Kramers-Kronig equation. The maximum anomalous signal is collected when values of the two components differ the most, also known as the absorption edge. Based on the fluorescence scan, the peak wavelength where f'' reached its maximum was 1.2822 Å (E= 9670.01 eV) and the inflection point wavelength where f' is highest was 1.2832 Å (E= 9662 eV). Data were then collected at 1.2822 Å where the maximal difference in f'' and f' values was achieved (**Figure 3.9B**). SAD data collection was carried out in this case to minimise the radiation

damage on crystals that is normally the outcome from collecting data at multiple wavelengths (MAD).

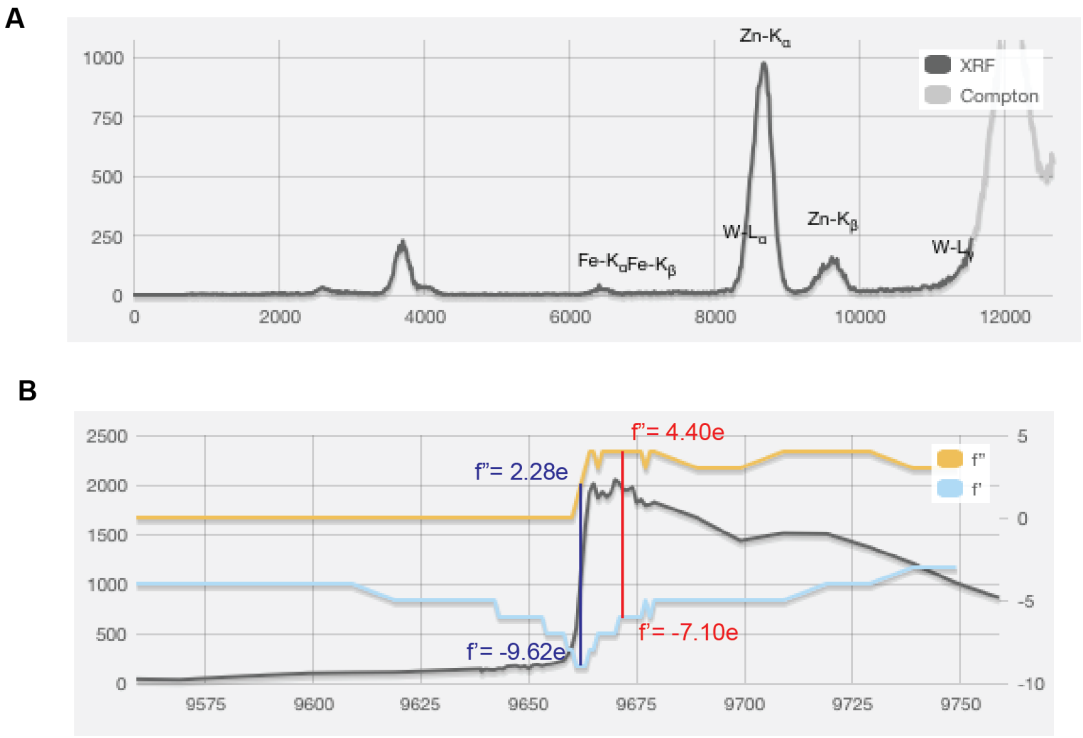


Figure 3.9 (**A**) MCA (Multichannel Analyser) spectrum collected for the Cnn-CM2 crystal on I04 at Diamond in order to identify the anomalous scatter. The incident X-ray beam was at 12,657.99 eV with 100% beamline transmission. The sample contained zinc that was subsequently used for phasing. (**B**) X-ray fluorescence scan conducted on the same crystal at 1.60% transmission. The peak wavelength was determined at which the f' and f" components differ maximally.

For native data collection, each reflection (h,k,l) in a dataset has an associated reflection (-h,-k,-l) collected from opposite sides of the crystal lattice plane. According to Friedel's law, these two reflections are equal in intensity but opposite in phase, and are known as a Friedel pair. Thus, the number of reflections to be collected can be halved to obtain a complete native dataset. However, in the presence of anomalous scattering, a 90° anticlockwise phase shift occurs changing the phases and amplitudes of reflections thus breaking the centrosymmetry (**Appendix 8.2.4**). The structure factors of the original Friedel pair are no longer equivalent. Differences in their amplitudes then contribute to the anomalous content of a dataset. Because the Freidel pairs are broken, all reflections need to be collected, which means that a crystal needs to be rotated through 360° for a complete anomalous dataset in the absence of other symmetry. Fortunately, the Cnn-CM2 crystal showed a high symmetry ($P6_122$) and thus each reflection pair would be observed multiple times in a complete dataset. Anomalous differences from symmetry-related Bijvoet pairs can thus be averaged to minimise random errors and increase the precision.

Dataset	Cnn-CM2_2	Cnn-CM2_3	Cnn-CM2_4
Space group	$P6_122$	$P6_122$	$P6_122$
Unit cell dimensions (Å)	a= 49.60, b= 49.60, c= 213.27	a= 49.88, b= 49.88, c= 214.22	a= 49.49, b= 49.49, c= 212.28
Unit cell angles	α= β = 90°, γ = 120°	α= β = 90°, γ = 120°	α= β = 90°, γ = 120°
Resolution (Å)	42.65–3.62	43.20–3.60	42.46–3.89
No. unique observations (overall/inner/outer)	2148/42/145	2203/45/139	1745/36/113
Completeness (overall/inner/outer)	99.9/98.0/100.0	99.9/99.5/100.0	100.0/97.6/100.0
Multiplicity (overall/inner/outer)	19.5/9.4/21.1	19.1/8.8/20.5	18.3/8.2/20.6
R_{merge} (overall/inner/outer)	0.081/0.030/0.76	0.074/0.033/0.94	0.105/0.046/0.51
R_{pim} (overall/inner/outer)	0.024/0.010/0.224	0.022/0.011/0.278	0.032/0.019/0.134
I/σI (overall/inner/outer)	25.7/53.4/4.9	26.3/43.5/4.3	17.8/25.0/6.3
Anomalous completeness	99.8/100.0/100.0	99.9/100.0/100.0	100.0/0.0/100.0
Anomalous multiplicity	12.5/17.7/12.1	12.2/17.5/11.8	11.6/0.0/11.8
Anomalous correlation	0.742/0.000/-0.075	0.802/0.000/0.073	0.463/0.000/0.144
Mid-slope of anomalous normal probability	1.241	1.333	1.196

Table 3.4 Statistics from processing of Cnn-CM2 anomalous datasets. Note that the R_{merge} and R_{pim} values listed above are for reflections within I+/I- due to the presence of anomalous differences.

Three anomalous datasets (Cnn-CM2_2, Cnn-CM2_3 and Cnn-CM2_4 from three different crystals) were collected and processed using Xia2 in 3daii mode [Winter, 2010], (**Table 3.4**). The data showed strong anomalous signal indicated by a good correlation between the random halves of anomalous differences (**Figure 3.10**). The Cnn-CM2_4 dataset (*green* line in **Figure 3.10**) contained slightly weaker anomalous signal when compared to the other two datasets. As expected, the anomalous correlation decreases with the resolution due to the lower anomalous multiplicity and subsequently fewer observations. The Cnn-CM2_2 dataset (*blue* line in **Figure 3.10**) was used for deriving the phase information.

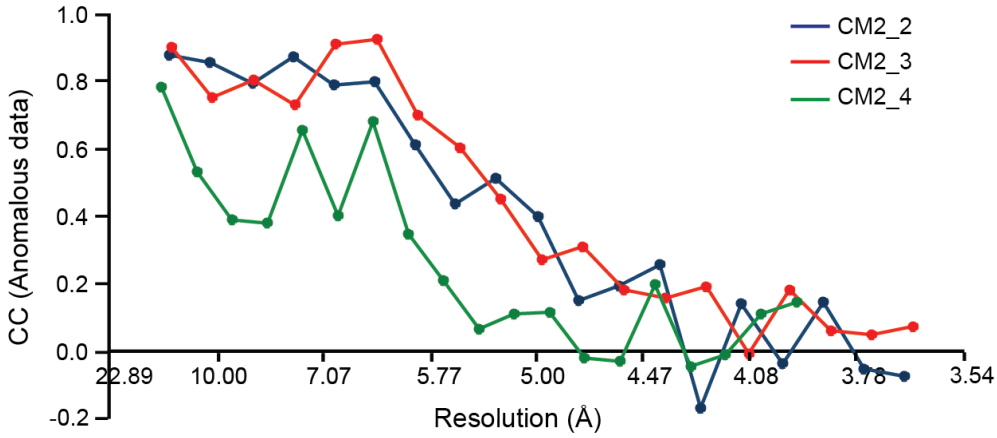


Figure 3.10 The correlation of anomalous differences for the three anomalous datasets. Note that both Cnn-CM2_2 and Cnn-CM2_3 datasets show stronger anomalous signals, and the CC_{anom} values stay above zero at nearly all resolution shells. Cnn-CM2_2 dataset was chosen for heavy atom detection.

3.4.3 Solving the phase problem

The structure factors, later to be Fourier transformed into electron density, are described by both an amplitude and a phase. From the diffraction pattern of a crystal, the intensity of each reflection can be obtained, which is used to calculate the amplitude of its structure factor. But, the phase problem still needs to be solved. In this project, I used the SAD method taking advantage of the protein bound Zn atom.

Initial attempts using autoSHARP [Bricogne et al., 2003] did not give interpretable electron density maps, I therefore used the Phenix Autosol [Adams et al., 2010] pipeline of programs for experimental phasing into which the Cnn-CM2_2 dataset was fed with an initial search for two Zn atom sites. Autosol runs a data analysis program *phenix.xtriage* to assess the data quality and rejects any reflection outliers [Zwart et al., 2005]. The anomalous data is normally truncated to a lower resolution that provides the greatest anomalous difference. HySS (Hybrid Substructure Search) is then used to locate the heavy atom substructure [Grosse-Kunstleve and Adams, 2003]. The anomalous difference is used in both Patterson and direct methods to calculate the position of heavy atoms. Eight potential two site, Zn atom substructures were found. These sites were further refined using Phaser [McCoy et al., 2007]. The parameters for each substructure are then optimised, which include the occupancy, position, B factor, anomalous scattering components (f' and f'') and variances indicating errors in the parameters [McCoy and Read, 2010], using a probabilistic approach. Using the structure factors derived from the anomalous

contribution, a log-likelihood gradient (LLG) map is computed to check if any heavy atom site should be added or removed. The substructure is considered as completed if all the parameters converge. For the Cnn-CM2_2 dataset, after the LLG completion, each substructure was confirmed to contain two Zn atom sites.

As the substructures are refined, their finalised locations are used to derive the phase information for each reflection (**Appendix 8.2.4**). In SAD, there are two possible phase solutions but only one of them is correct (**Appendix 8.2.4**). In order to avoid analysing maps from all possible centroid combinations, the average of the two phases is used [McCoy and Read, 2010]. The calculated map would therefore contain both the real electron density, from the true phases, and the noise, from the wrong phases. The noise, however, can be removed by density modification such as solvent flattening. During this process, the handedness of heavy atom substructure is also identified. The heavy atom model is derived from interatomic vectors that are centrosymmetric. As a result, the enantiomorph of a substructure would also be recognised as a satisfying solution, and its phase information is calculated separately from the original hand solution. Nonetheless, the density map from the inverted hand solution would not improve during density modification such that the correct hand can be determined. In Autosol wizard, this procedure is conducted by RESOLVE [Terwilliger, 2002], which uses a *priori* knowledge about electron density map. For example, the disordered solvent regions are normally flat whereas the protein regions are highly variable [Sherwood and Cooper, 2011]. Therefore, flattening the solvent density around suspected protein regions is carried out to sharpen the density of protein molecules. Based on the symmetries in heavy atom sites, RESOLVE [Terwilliger, 2002] also identifies the presence of non-crystallographic symmetry (NCS) that can be used as a source of prior knowledge as the NCS-related regions should be similar. Based on this information, density of NCS-related protein regions can be averaged and further refined.

Four phase solutions from the same two site, Zn atom substructure were determined before automatic model building and refinement, and the maps in $P6_122$ space group gave more favourable statistics than in $P6_522$ space group. Three structure models were built using Autobuild [Terwilliger et al., 2008] in Phenix. The top hit had 51 residues (out of 79 total expected residues) modelled in with an R_{free} of 52.54% (**Table 3.5**). Two Zn sites were detected in the

asymmetric unit, one of which was not located within any density and hence was not a real site. The map, although of poor quality, showed clear density for several short helices (**Figure 3.11**).

No. residues	Map-model CC	R_{work}	R_{free}
51	0.46	47.73%	52.54%

Table 3.5 Statistics for the model built using AutoBuild in AutoSol after substructure search, phasing and density modification.

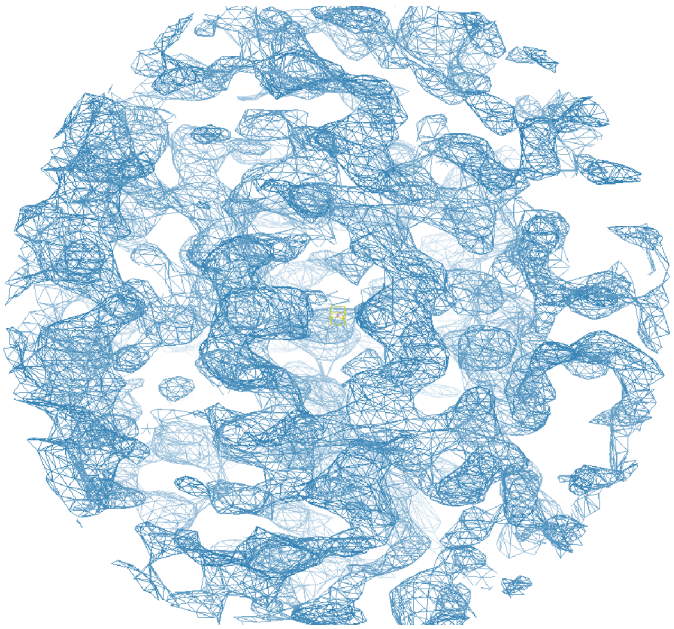


Figure 3.11 2Fo-Fc map from the top hit of Autosol solutions showing densities for short but continuous helices into which the Cnn-CM2 protein can fit. The map is contoured at 1.0σ.

3.4.4 Structure refinement

As the resolution of the phasing dataset was low (<3.5 Å), attempts were made to take the model into the higher resolution 'native' data (Cnn-CM2_1). The model built by Autosol was placed in the 'native' data using Phaser Molecular Replacement (MR; the best solution had a strong TFZ score of 10.5, standing out from all other solutions), [McCoy et al., 2007; Winn et al., 2011]. With the model positioned in the unit cell, I could now refine the structure against the higher resolution 'native' dataset Cnn-CM2_1 producing more detailed maps. Phases from autoBuster refinement [Bricogne et al., 2011] were used for automated model building with Buccaneer in CCP4 [Cowtan, 2006; Winn et al., 2011]. This program carries out 5 cycles of model building and 10 cycles of Refmac refinement after each build. The atomic model from Buccaneer refined in autoBuster was

manually modified in Coot [Emsley et al., 2010] if a large positive residual was evident in the Fo-Fc difference map.

After three runs of such refinement, densities for two continuous helices, a bent one and a straighter one, and a Zn^{2+} ion became clearer. However, I could not see many well-defined pockets of density into which side chains could fit. Density for the main protein chain and the Zn^{2+} coordination sphere was unconnected making it difficult to manually build in further residues at the N-terminus (**Figure 3.12**). Although no residual density of more than 4 r.m.s.d was observed in the Fo-Fc map, the R factors (R_{work} = 33.41%, R_{free}= 38.86%) were still not in a reasonable range further indicating the model was not largely correct in explaining the observed data.

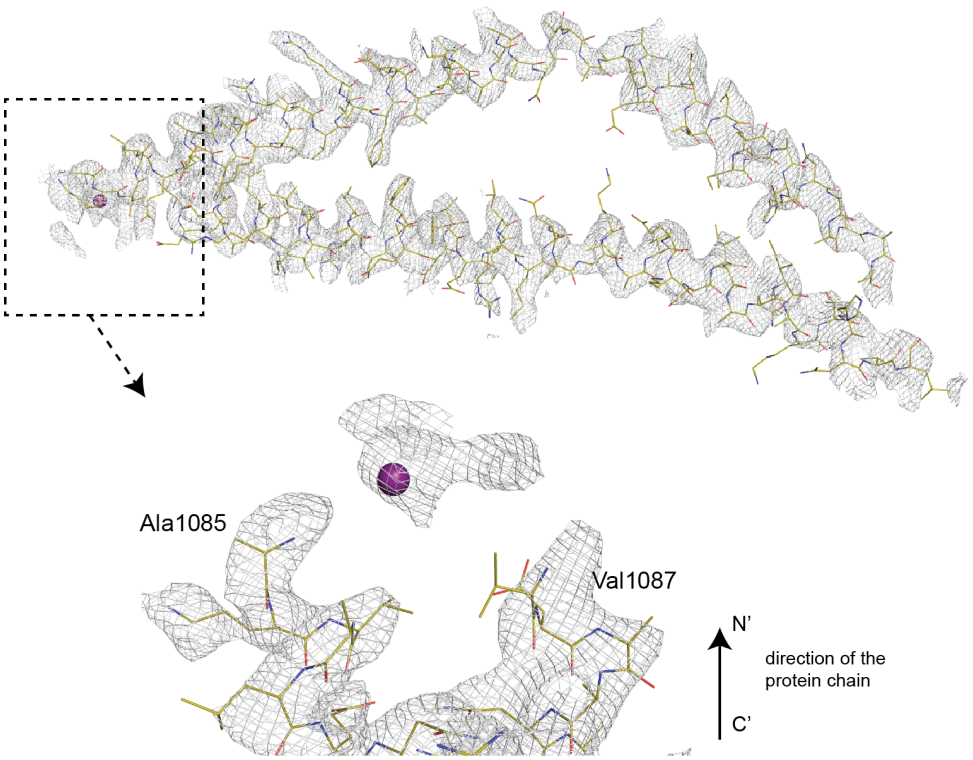


Figure 3.12 View of maps resulting from the Autobuster refinement in $P6_122$ space group. *Grey* density is the 2Fo-Fc map contoured at 1.0σ. *Green* density shows the Fo-Fc map at +4.0σ. The Zn^{2+} ion is shown as a *purple* sphere. There is no positive residual density above 4.0σ, based on which the model can be further modified. Bottom, zoomed view of the Zn coordination region showing density for the protein chain (Ala1085A and Val1087B, highlighted, are the first residues modelled in each chain at this stage) is unconnected with density for the Zn atom. The lack of positive residual density in the Fo-Fc map makes it difficult to manually build in extra residues in an unbiased manner.

At this stage of the structure refinement, I switched to use Phenix [Adams et al., 2010] because this program is better at keeping the overall geometry reasonable than autoBuster refinement [Bricogne et al., 2011], and it is difficult to retain the geometry for data at this resolution. In addition, due to the low resolution of the data, the number of observations per parameter to be refined was extremely low (4130 unique reflections observed/(4x 984 atoms in the model)= 1.05

data point per parameter; 4 parameters (x, y, z and B factor) are associated with each atom). I used TLS refinement in which the molecule is defined as groups of atoms based on, for example, secondary structures. The TLS groups are then refined with each group carrying 20 more parameters. Although the total number of refined parameters is slightly increased compared to individual atomic parameter refinement, TLS refinement provides information that helps to direct the initial treatment of thermal motion of the molecule. To start with, the TLS parameters were first refined with all B factors set constant to 80. After the first refinement run, the B factor hold was released and the grouped B factors were then refined. After 3 iterative cycles of refinement and manual building, density showing continuous helical turns started to appear at the N-terminus (*orange* density in **Figure 3.13**). More terminal residues could be modelled in. The change in R factors became smaller after another 2 cycles of refinement. I then changed to individual B factor refinement. Torsion-angle NCS restraints were used in all refinements. X-ray/stereochemistry or ADP weights were also optimised. Density became more evident around the Zn binding sites as the refinement continued (*blue* density in **Figure 3.13**). The Cys1084 side chain was clearly bound to the Zn^{2+} ion, and this residue is indeed one of the most common Zn coordinates [Laitaoja et al., 2013]. I have also manually built in Asp1083 with its side chain pointing towards the Zn^{2+} ion so that its terminal oxygen atom can be involved in Zn coordination. This, however, was not modelled with great confidence and therefore other data was sought (**Section 3.4.6**).

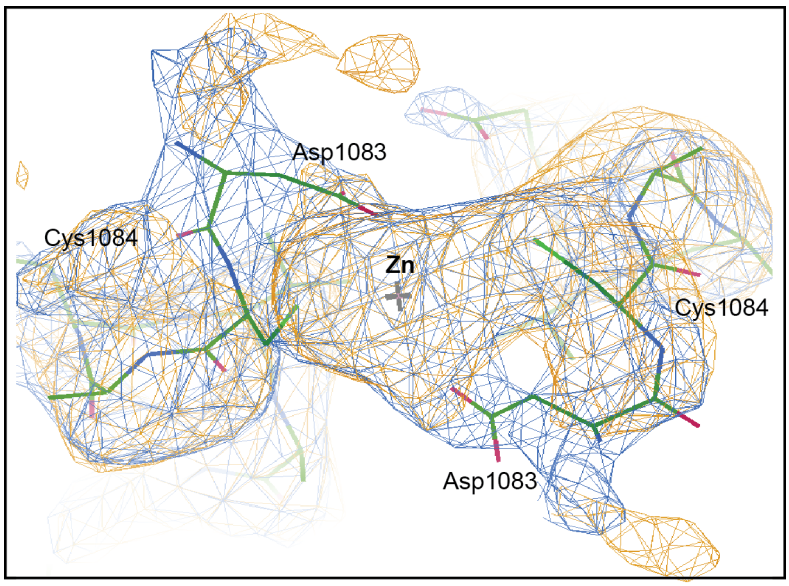


Figure 3.13 View of maps from Phenix refinements in Coot. Both maps are contoured at 1.0σ. *Orange* density is the 2Fo-Fc map obtained after 6 cycles of refinement in Phenix. TLS refinement strategy was applied, grouped B factors were refined. The Zn coordination sphere started to appear continuous with the main protein chain. *Blue* density is the 2Fo-Fc map obtained after 9 cycles of individual B factor refinement. The density became more evident around the Zn^{2+} ion allowing terminal residues to be modelled in although with low confidence.

3.4.5 Improving crystallisation efficiency

No Zn^{2+} was present in protein purification and crystallisation trials, and yet one Zn^{2+} ion has bound to a dimer of $Cnn^{1082-1148}$ as seen in the crystal structure. This suggests that the Zn must have come from the bacterial cytoplasm and its amount could be limiting. A mixture of Zn^{2+}-bound and -unbound proteins could discourage ordered packing of Cnn-CM2 molecules in the crystal lattice. I therefore carried out more crystallisation trials where the amount of precipitants and the $ZnCl_2$ concentration were optimised (**Figure 3.14A**). Trays were set up with a protein stock of 3 mg/ml at 75% and 50% *v/v* protein: mother liquor ratios. After 3 days of incubation at 20 °C, drops showing over-nucleation were observed in the absence of $ZnCl_2$. Single crystals with sharp edges were obtained when the optimisation conditions were supplemented with 20 to 40 µM $ZnCl_2$ (**Figure 3.14B**). The addition of more than 40 µM $ZnCl_2$ in the mother liquor led to precipitation in drops.

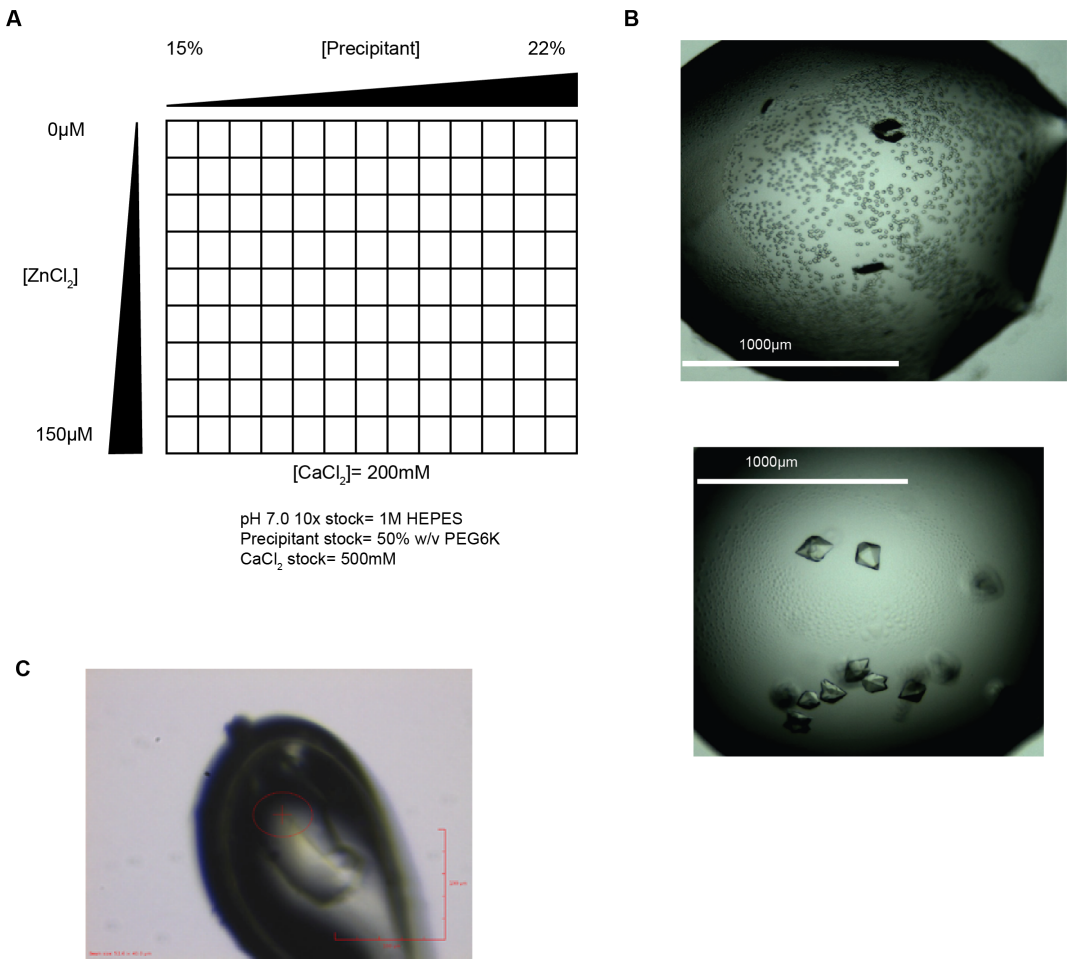


Figure 3.14 Crystallisation trays put down in the presence of varying amount of $ZnCl_2$ and with the non-TEV version (Section 3.6.2) of Cnn-CM2 protein. (**A**) Optimisation screen used to obtain crystals of better quality for data collection. The $ZnCl_2$ concentration was optimised. (**B**) Comparison of crystals grown with (top) and without $ZnCl_2$ supplemented in the mother liquor conditions. The best crystals were obtained when 20 to 40 µM $ZnCl_2$ was added. The bottom image showing an example of crystals grown in a drop containing 75% protein stock and 25% mother liquor (100 mM HEPES (pH 7.0), 200 mM $CaCl_2$, 19.5% *w/v* PEG6K, 40 µM $ZnCl_2$). (**C**) Crystal of $Cnn^{1082-1148_noTEV}$ mounted in a cryo loop at Diamond Beamline I04.

The residual TEV cleavage site could be flexible (it was not seen in the structure from **Section 3.4.4**) and could therefore introduce disorder in the crystal lattice. In an attempt to improve crystal quality I therefore removed the TEV cleavage site at the C-terminus of Cnn-CM2 protein sequence by introducing a premature stop codon in the pLip_Cnn$^{1082\text{-}1148}$ vector (discussed in **Section 3.6.2**). SEC purified Cnn$^{1082\text{-}1148\text{-}noTEV}$ was dialysed overnight at 4 °C into WB buffer (1 mM TCEP was used instead of 5 mM BME; **Appendix Table 8.8**) containing 10 µM $ZnCl_2$. Strikingly, the removal of C-terminal TEV site significantly increased the protein solubility by more than 10-fold (from 3 mg/ml to ~50 mg/ml). This might be explained by the hydrophobic nature of this sequence in particular when fused to a protein of shorter length. Broad screen trays (**Appendix 8.1.10**) were then put down using protein stock at 40 mg/ml. Remarkably, some crystals grew immediately (one example is shown in **Figure 3.14C**).

More crystals were fished and snap frozen for data collection by Prof Susan M. Lea at Diamond Light Source. All the diffraction images were processed using Xia2 in 3daii mode [Winter, 2010] and two datasets were obtained from two different crystals. Although the optimised crystals looked better morphologically, they did not diffract to significantly better resolutions. Both datasets were indexed to $P6_122$ space group with unit cells of similar dimensions to previously collected datasets (**Table 3.6**).

Dataset	**Cnn-CM2_5**	**Cnn-CM2_6**
Beamline	Diamond I04-1	Diamond I04
Wavelength	0.9174 Å	0.9793 Å
Space group	$P6_122$	$P6_122$
Unit cell dimensions (Å)	a= 49.19, b= 49.19, c= 210.48	a= 49.22, b= 49.22, c= 211.23
Unit cell angles	α = β = 90°, γ = 120°	α = β = 90°, γ = 120°
Resolution (Å)	42.10–3.07	70.41–2.97
No. unique observations (overall/inner/outer)	3288/62/221	3658/71/269
Completeness (overall/inner/outer)	99.9/95.5/100.0	99.9/99.7/100.0
Multiplicity (overall/inner/outer)	31.5/17.1/37.1	17.2/8.9/18.7
R_{merge} (overall/inner/outer)	0.089/0.043/0.903	0.065/0.027/0.720
R_{pim} (overall/inner/outer)	0.016/0.009/0.148	0.016/0.008/0.169
I/σI (overall/inner/outer)	30.7/48.5/5.1	30.5/54.5/4.4

Table 3.6 Processing statistics for the native Cnn$^{1082\text{-}1148}$ (Cnn-CM2_5) and Cnn$^{1082\text{-}1148_noTEV}$ (Cnn-CM2_6) datasets.

3.4.6 Improved density map

The Cnn-CM2_6 dataset showed a slightly higher resolution, and thus was chosen for further model refinement. However, the model could not be correctly positioned into the new dataset as indicated by high R_{free} and R_{work} values (nearly 40%). Data were run through Zanuda [Lebedev and Isupov, 2012] to double check the space group, which determined $P6_1$ could be a possible alternative (**Table 3.7**). I then reprocessed the images in $P6_1$ space group, and this produced a dataset with reliable statistics to 2.80 Å (**Table 3.8**). No alternative indexing was possible as checked by Pointless program [Evans, 2006b; Winn et al., 2011], and the model was then molecular replaced into the new data using Phaser [McCoy et al., 2007; Winn et al., 2011]. In the asymmetric unit of the $P6_1$ space group, 2 inverted pairs of parallel dimers self assembled into an antiparallel tetramer creating a two-fold pseudo-symmetry, which was averaged out in the $P6_122$ space group (**Figure 3.15**).

Space group	R factor	R_{free}
$P6_122$	0.2483	0.3983
$P6_1$	0.2485	0.3656

Table 3.7 Statistics of model refined by Zanuda to check the best space group solution. R factor in $P6_1$ subgroup is as good as in the original space group suggesting it can be an alternative indexing solution.

Dataset	Cnn-CM2_7
Space group	$P6_1$
Unit cell dimensions (Å)	a = 49.22, b = 49.22, c = 211.23
Unit cell angles	α = β = 90°, γ = 120°
Resolution (Å)	70.41–2.80
Number of unique observations (overall/inner/outer)	7171/90/548
Completeness (overall/inner/outer)	99.9/99.8/100.0
Multiplicity (overall/inner/outer)	10.4/8.1/10.1
R_{merg} (overall/inner/outer)	0.073/0.023/1.666
R_{pim} (overall/inner/outer)	0.024/0.008/0.550
$CC_{1/2}$ (overall/inner/outer)	1.000/0.999/0.602
I/σI (overall/inner/outer)	19.7/54.2/1.4

Table 3.8 Statistics for $Cnn^{1082\text{-}1148_noTEV}$ dataset processed in $P6_1$ space group.

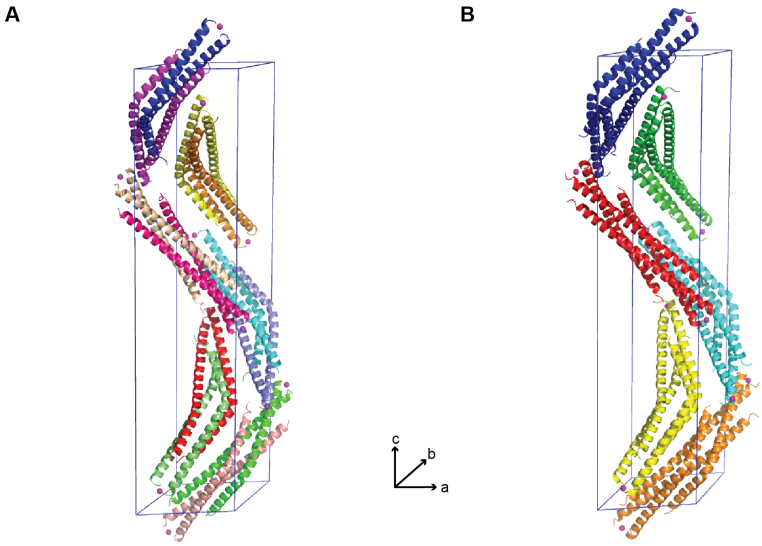


Figure 3.15 The arrangement of Cnn-CM2 crystal lattices. In each unit cell, each asymmetric unit is shown in a different colour. Left, the $P6_122$ unit cell contains 12 copies of NCS dimers related by the crystallographic symmetry. Right, the $P6_1$ unit cell contains 6 copies of NCS tetramers. Although the parallel dimers forming each tetramer are pseudo-symmetrically related, they are not crystallographic as in the $P6_122$ unit cell.

Further refinements were carried out in Phenix [Adams et al., 2010] and Coot [Emsley et al., 2010] as described above (**Section 3.4.4**), using the dataset processed in $P6_1$ space group. Additional density was brought out around the Zn coordination sites indicating an extra residue could be built in along the chain (**Figure 3.16A**). In addition, the positive residual density around the Asp1083 backbone suggested a possible positioning of its side chain. I therefore manually rotated the side chain of Asp1083 so that it became solvent exposed, and then built in His1082. His is in fact a common Zn^{2+}-binding species owing to the nitrogen atom of its imidazole ring [Laitaoja et al., 2013]. In the 2Fo-Fc map contoured at 1.0σ, only a smeared out density was observed for the Zn^{2+} coordination sphere lacking restraints (**Figure 3.16B**). As a result, the metal centre was modelled in accordance with a *priori* knowledge of Zn-S/N bond lengths and angles defined by ReadySet in Phenix [Adams et al., 2010].

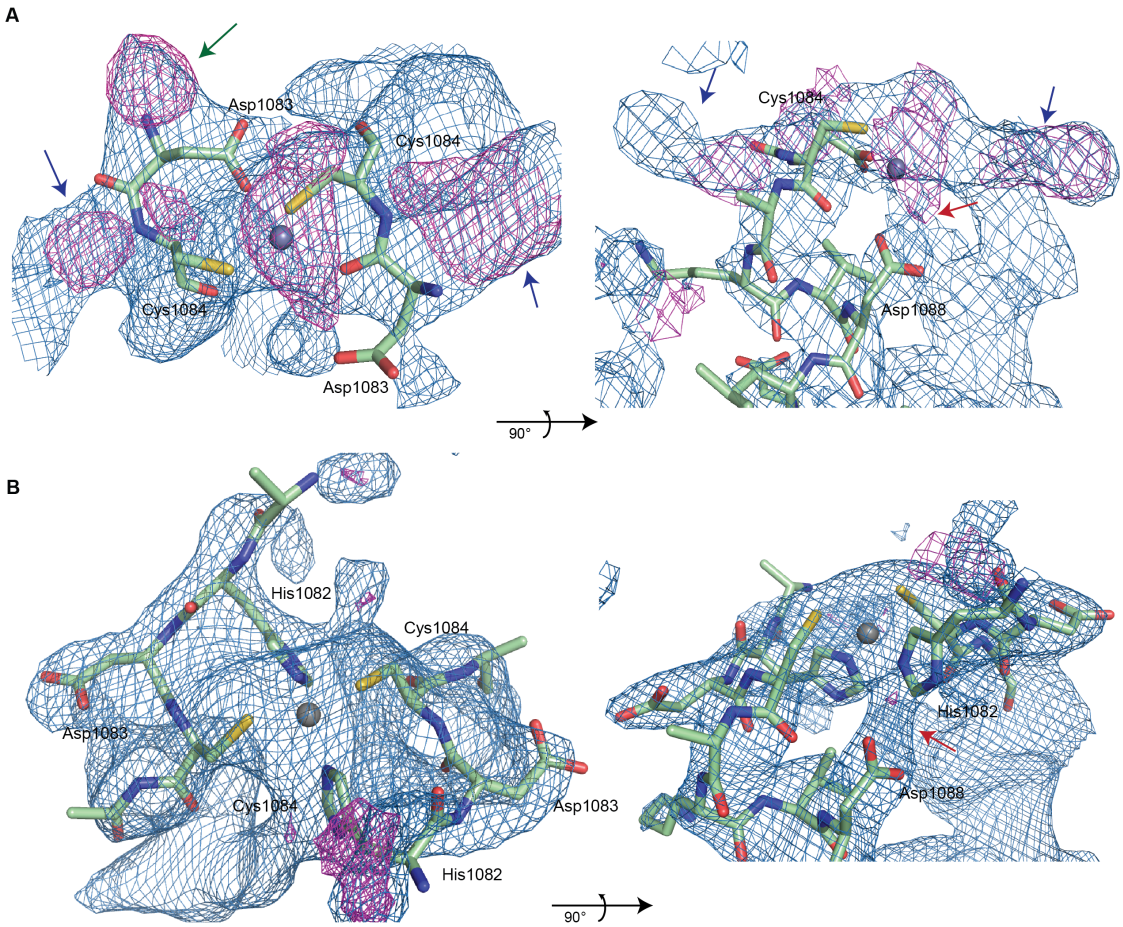


Figure 3.16 Top-down (left) and side-on (right) views of maps from refinements of the previously refined model using the 2.8 Å data (processed in $P6_1$ space group). *Blue* density shows the 2Fo-Fc map contoured at 1.0σ. *Purple* density shows the Fo-Fc map contoured at +3.0σ. The Zn^{2+} ion is shown as a *grey* sphere. In both views, chain B is shown on left and chain A on the right. For clarity, the side-on view of maps does not show the model for chain A. (**A**) Maps from placing the model that was previously refined in the $P6_122$ dataset (Cnn-CM2_1) into the $P6_1$ dataset (Cnn-CM2_7) and subsequently refined in Phenix for 3 runs. *Blue* arrows point to pockets of density into which the Asp side chains could fill. Another residue along the chain could fill the space pointed by the *green* arrow. Connecting density is evident between the potential Zn coordinate species and the side chain of Asp1088 (indicated by a *red* arrow). (**B**) Maps from the final refinement. The empty density (with positive residuals) in (**A**) has been filled in with the side chain of Asp1083. Residual density is evident at the N-terminus (highlighted) indicating the chain direction. Maps on the right showing the density linking the side chains of His1082B and Asp1088A (highlighted with a *red* arrow). A salt bridge might form between these two residues.

When the structure was refined assuming full occupancy, the Zn^{2+} ion and its potential ligands have higher B factors (around 120, calculated by Baverage in CCP4 [Winn et al., 2011]) with respect to residues in the middle of the chain indicating a higher degree of mobility and disorder (**Figure 3.17**). This could be because of its close-to-termini location. Indeed, the residues at the C-terminus also showed higher B factors (all-atoms averaged). The refinement continued until the final model and the 2Fo-Fc map were in good agreement as visualised in Coot [Emsley et al., 2010] and supported by final R_{work}/R_{free} values of 26.65%/23.99%. The final refinement statistics are summarised in **Table 3.9**.

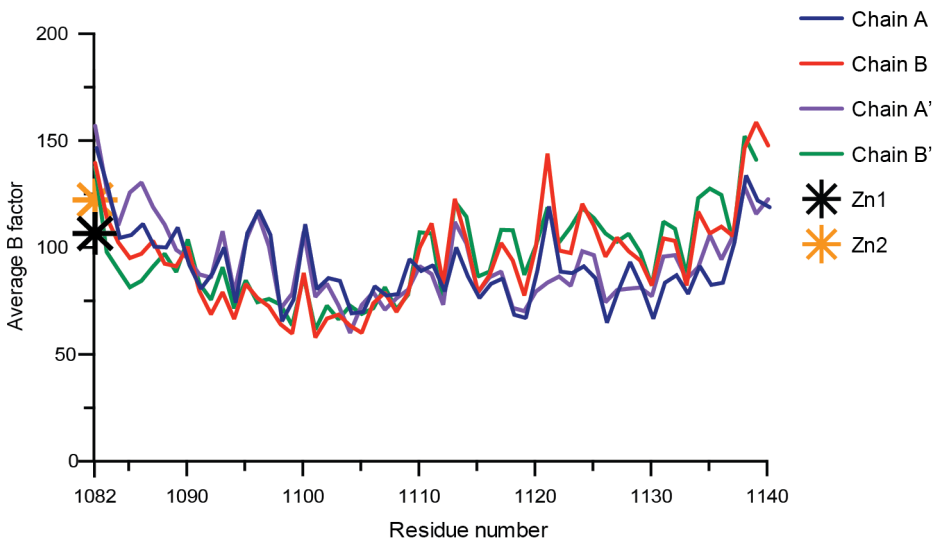


Figure 3.17 Plot of average B factors over the length of the structure in all four chains. Zn1 is the metal coordinated by chain A and chain B; Zn2 is the one between chain A' and chain B'. The all-atoms averaged B factors are shown for each residue. The structure overall has high B factors with an average of 95, possibly explaining the moderate resolution of data (2.8 Å). Both N- and C-termini are more mobile as indicated by the higher B factors.

	Cnn$^{1082\text{-}1148_noTEV}$
Residue seen in structure	His1082–Asn1140
Resolution range	42.626–2.804Å
R_{work} (%)	23.99
R_{free} (%)	26.65
Clashscore	4.47 (100th percentile)
Side chain with poor rotamers (%)	0.47
Ramachandran outliers (%)	0
Ramachandran favoured (%)	98.10
MolProbity score	1.22 (100th percentile)

Table 3.9 Refinement statistics for the Cnn$^{1082\text{-}1148_noTEV}$ structure. Note that the test size for calculation of R-values is 5%.

3.5 Analysis of the Cnn-CM2 structure

The crystal structure of Cnn-CM2 shows several striking and unexpected features including the formation of an asymmetric parallel dimer that then assembles into an antiparallel tetramer, and its Zn^{2+} binding activity, which will all be presented in this chapter.

3.5.1 Cnn-CM2 forms an unusual parallel, asymmetric dimer

A coiled coil (CC) consists of two or more α-helices wound around each other into a rope-like structure. Its sequence is usually arranged in a regular pattern known as a heptad repeat [Mason and Arndt, 2004]. Residues at *a* and *d* positions of the repeat are usually non polar and are involved in hydrophobic interactions at the dimer interface. Polar and charged residues are often found at positions *e* and *g* of the repeat, forming salt bridges and hydrogen bonds that further stabilise the CC packing. The helices supercoil in order to form the hydrophobic stripe established by successive *a*-*d'* pairings (residue at *a* position on one chain with residue at *d* position i.e. *d'* on the adjacent chain). Although the Cnn-CM2 structure comprises two parallel α-helices, its structure is highly unusual as the dimer is asymmetric (one helix is more kinked than the other) and the helices splay apart, with few contacts made between the two chains after the first 17 N-terminal residues (**Figure 3.18A,B**).

I used the CC analysis program SOCKET [Walshaw and Woolfson, 2001] to assign the heptad register for the parallel dimer structure (**Figure 3.18C**). The regular sequence pattern was only recognised from Val1087 (*d* position) to Leu1098 (*a* position). This is expected, as the SOCKET server [Walshaw and Woolfson, 2001] assigns the heptad register based on the classical knob-into-hole interaction, which was first described by Crick [Crick, 1953] and is representative of canonical CC packing. In the Cnn-CM2 structure, the two chains splay outwards, with Leu1098 defining the last parallel interaction pair (**Figure 3.18B**). The dimer interface was also analysed using PISA [Krissinel and Henrick, 2007] and Coilcheck+ (CC+) [Alva et al., 2008], and both servers identified interactions only within the first 20 amino acids of the CC. In parallel CCs, *a*-*a'* and *d*-*d'* pairings are often observed at the dimer interface. In the case of Cnn-CM2, side chain hydrogen bond formation is observed between Asn residues (Asn1091) at the *a* position of the

first heptad repeat (**Figure 3.18D**). Although Asn is a polar residue, it is often found at position *a* in parallel CC dimers [Walshaw and Woolfson, 2001]. This is because its hydrogen bonding is only possible in in-register parallel dimers, where it helps specify the CC oligomer state and orientation. Asn1091 also forms an intra-helical hydrogen bond with Arg1095 at position *e* (i), which forms another, inter-helical hydrogen bond with Glu1090 at position *g* (i'-5) in the adjacent helix (**Figure 3.18D**). Further hydrophobic interactions observed at *a* and *d* positions are shown in **Figure 3.18E**, based on which the two helices are packed with an angle of ~20°.

The two chains then move apart after 5 helical turns (**Figure 3.18B**). Scanning through the structure, this splayed conformation can be rationalised with the protein sequence. Residues at *a* and *d* positions in the more C-terminal parts of the helices are not ones frequently observed in parallel dimers (**Figure 3.19A**). For instance, Thr residues are observed at two *d* positions within the Cnn-CM2 structure (Thr1101 and Thr1108), and Thr is not favoured in parallel dimers due to its short and polar side chain (**Figure 3.19B**). Walshaw and Woolfson have investigated the amino acid preferences in two-stranded CCs using solved structures in the Protein Data Bank (PDB) database [Walshaw and Woolfson, 2001]. Phe was not seen in any of the analysed examples, and in Cnn-CM2 the two helical strands seem to spread further apart at the Phe1105 residue at position *a* of the 3rd heptad repeat, which might be to avoid potential steric clashes between the bulky residues (**Figure 3.19B**).

Strikingly, the CM2 dimer is asymmetric as Helix B is kinked at Lys1119, whereas Helix A is largely straight (**Figure 3.19C**). Lys1119 is facing inward at the dimer interface, which may cause electrostatic repulsion if the two helices are too closely packed. Instead, the CC becomes asymmetric, and the two chains are pushed up to 15 Å apart as measured by the distance between the Cα atoms of Lys1119; canonical parallel dimers are normally 6 Å apart. Around the kink, the regular helical backbone of hydrogen bonds is also broken (**Figure 3.19D**). However, a side chain hydrogen bond is formed instead between Lys1119B and Asn1116B, which presumably stabilises the hinge region of Helix B (**Figure 3.19C**). This interaction is not observed in Helix A. The MSA for Cnn-CM2 (**Figure 3.20**) indicates the presence of a Lys residue at the same position (Lys1119) in fly sequences suggesting a high possibility this kink is conserved

across fly species. However, in vertebrates a gap in sequence conservation (~1 turn of helix less) is observed around this region. In the majority of the vertebrate sequences, a polar residue, Ser, has substituted Lys1119, although there is a Lys residue next to Ser. Based on the MSA alone, it is difficult to determine whether the helix bending would also be expected in vertebrate family member proteins. Further investigation including biophysical and structural characterisations into homologous system is required.

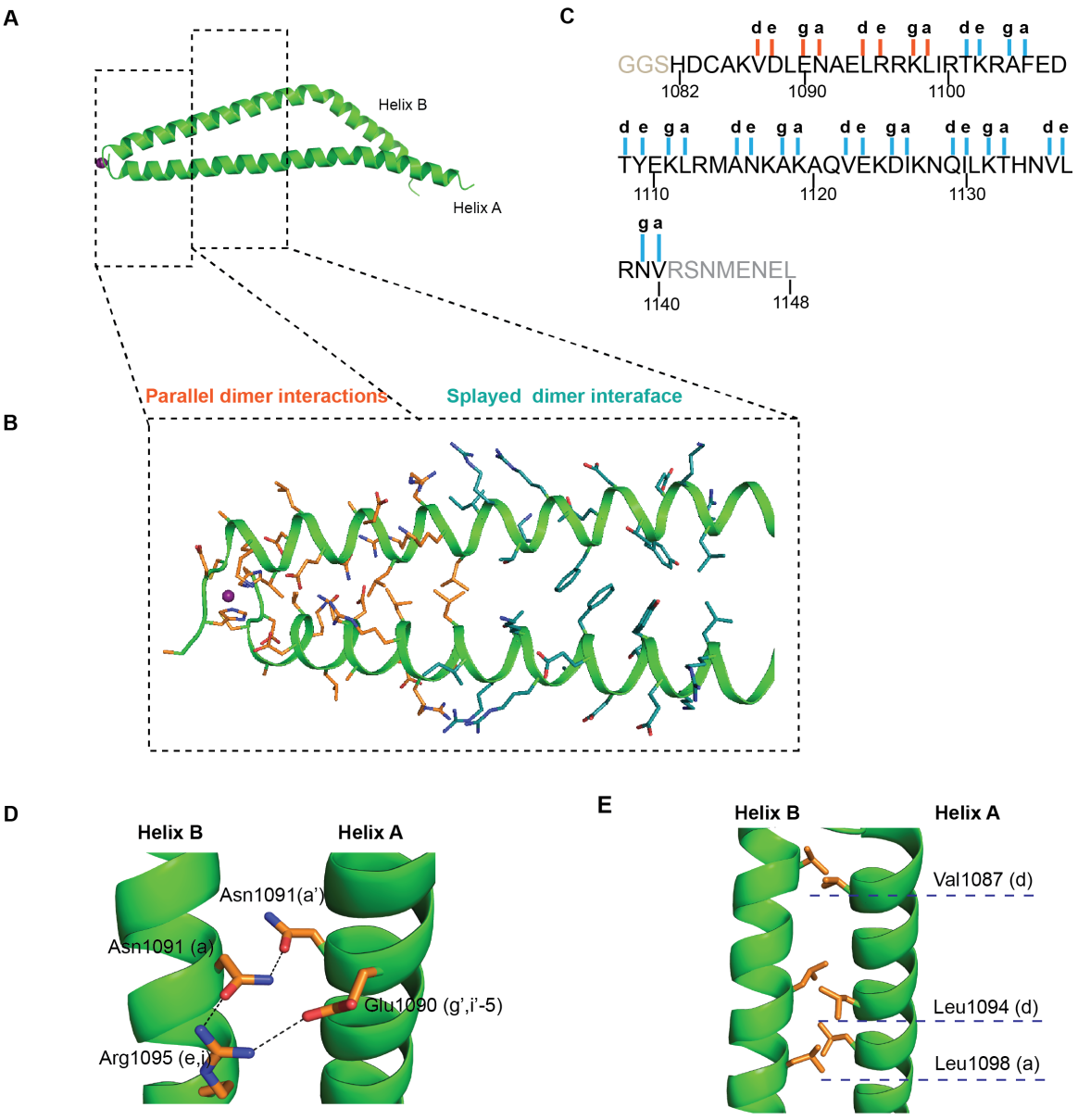


Figure 3.18 Analysis of the parallel dimer interface in the Cnn-CM2 structure. (**A**) The Cnn-CM2 parallel dimer comprises two asymmetric monomers. The first 5 helical turns of a monomer are only involved in parallel interactions as shown in (**B**), and the two chains then distant apart leaving a cavity in the parallel CC structure. Leu1098 defines the last parallel interaction pair. (**C**) The protein sequence of Cnn-CM2 is displayed with the heptad register in *orange* as assigned by the SOCKET program. The rest (in *blue*) was assigned based on the first 2 heptad repeats. The TEV sequence is shown in *beige*, and the C-terminal residues missing in the crystal lattice are shown *in grey*. (**D**) The side chain interaction is observed between Asn1091 residues at position *a* in both helices. An intra-helical hydrogen bond is formed between Asn1091 and Arg1095 in Helix B. The residue at position *e* also electrostatically interacts with Glu1090 at i'-5 position in Helix A. (**E**) Hydrophobic packing at the parallel dimer interface are demonstrated by interactions at *a* and *d* positions.

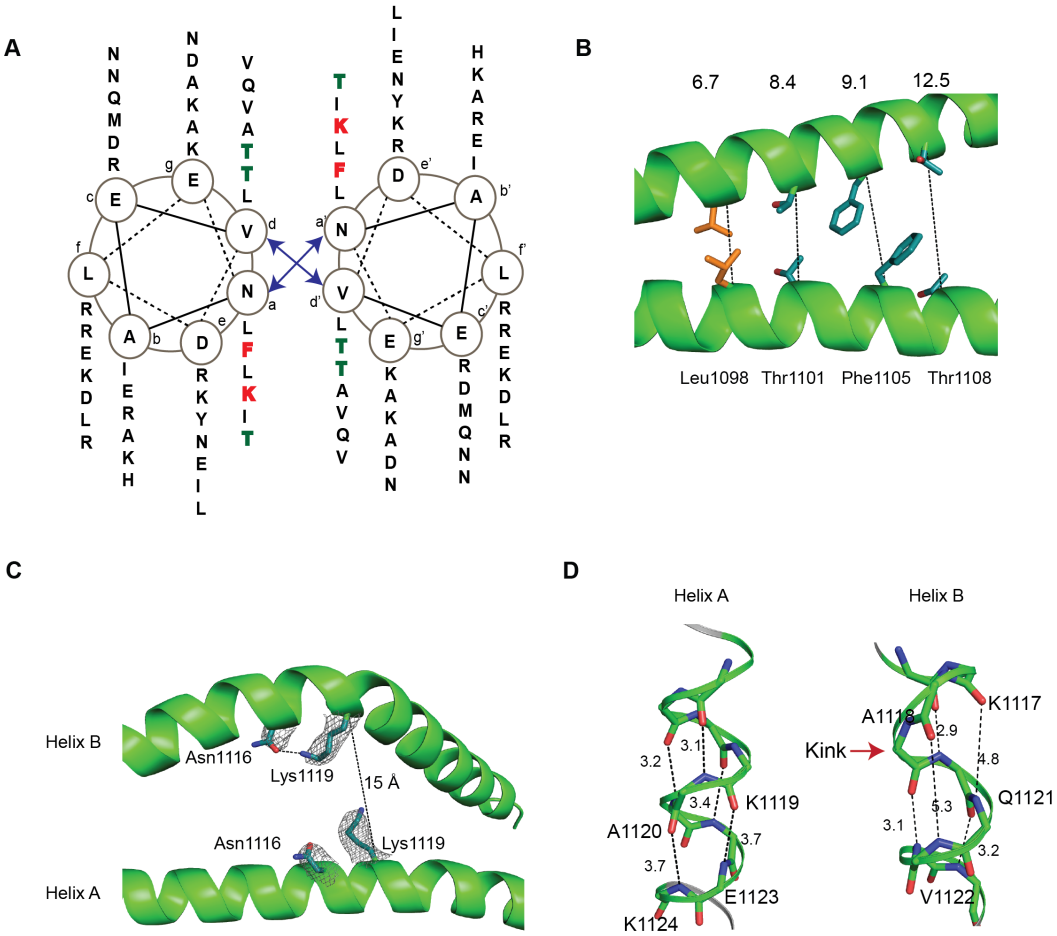


Figure 3.19 Sequence and geometrical characterisations of the kinked and open CC topology. (**A**) Cartwheel diagram showing the heptad repeats of the Cnn-CM2 structure, based on the register assigned by SOCKET (**Figure 3.16**). Amino acids that are less frequently observed at *a* and *d* positions in parallel CCs are highlighted in *green and red*. In parallel dimers, the residue at position *g* (i) often interacts with the residue at position *e'* (i'+5) in the adjacent monomer. In this case, some residues at *g* and *e'* positions carry the same charged groups that could cause electrostatic repulsions and hence the two helices are kept apart. (**B**) Distance (top, shown in Å) between the 2 splayed helices was measured between the Cα atoms of "*a*" position residues (bottom). Measurements are only shown for residues before the kinked region excluding the ones involved in parallel dimer interactions (Leu1098 is shown with its side chains highlighted in *orange*, because it is involved in parallel interaction). (**C**) A side-on view of the bent helical turns demonstrating the potential charge repulsion between Lys1119 residues if the two chains were too closely packed. This kink has satisfied the hydrogen bond formation between Asn1116 and Lys1119 providing intra-helical stability. The two chains were pushed apart by 15 Å, much more distant than in canonical CCs (6 Å). No interaction was seen on chain A between Asn and Lys at the same position. (**D**) Side-on view of hydrogen bond formation between backbone C=O (i) and H-N (i+4) around the kinked region (Asn1116 to Lys1124). The bond length can tolerate to ~3 Å. Therefore, the hydrogen bonds between Lys/K1117 and Gln/Q1121, Ala/A1118 and Val/V1122 are broken in the bent Helix B. Although Helix A is not apparently bent, it seems to curve slightly at Gln1121 as judged by the hydrogen bond lengths between Lys/K1119 and Glu/E1123, Ala/A1120 and Lys/K1124. For clarity, only those residues with the 'broken' backbone hydrogen bonds are labelled. A *red* arrow has pointed to the hinge where Lys1119 locates in Helix B.

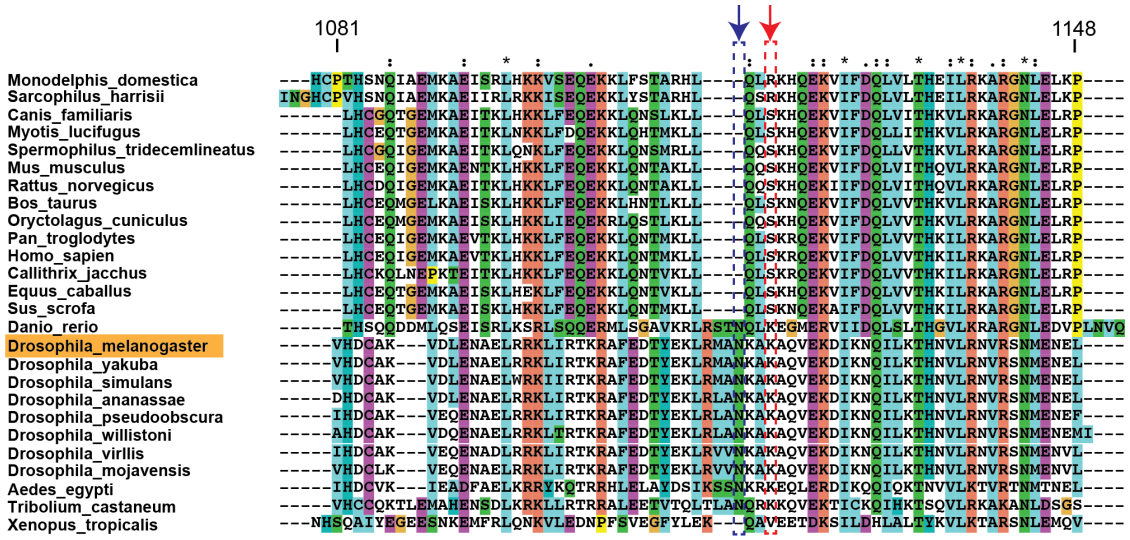


Figure 3.20 MSA of Cnn-CM2 sequences across evolution showing the gap in sequence conservation around Lys1119 (highlighted with a *red* arrow). In the majority of the vertebrate sequences, this residue is substituted by Ser at the same position despite another Lys residue is found next to Ser. Asn1116 (highlighted with a *blue* arrow), which is involved in the hydrogen bond network with Lys1119 for stabilisation of the hinge region, is also well conserved in fly sequences. Residue numbers for *D. melanogaster* are shown. Invariant residues are highlighted with asterisks, conserved residues with colons and semi-conserved residues with full stops. Residues are coloured according to the ClustalX colour scheme.

Moreover, several residues at *e* and *g* positions carry the same charged groups in the Cnn-CM2 parallel dimer structure (**Figure 3.19A**). In a canonical CC, *e* and *g* position residues are often involved in salt bridge interactions that are solvent exposed and provide additional stability for the tight packing. In the presence of both negatively or positively charged residues at these positions, as observed in the Cnn-CM2 structure, the two helices would not pack closer because of the charge repulsion. Overall, the open conformation adapted by Cnn-CM2 parallel dimer can be understood with its high occurrence of less tolerable amino acids in the heptad repeats.

As shown by CD analysis, Cnn$^{1082\text{-}1148}$ has a very stable helical structure that was not completely unfolded even at 80 °C (**Figure 3.5, Section 3.2.2**). Although its dimer structure shows only limited inter-helical contacts, and the two helices do not supercoil into a tight fold, Cnn-CM2 has a high composition of charged residues, which can help explain the structural stability of its helices. There are 11 acidic and 16 basic residues in the 67-residue domain constituting about 40% of the total sequence. Intra-helical salt bridges and hydrogen bonding are observed, which provide enhanced stabilisation of the helical structure- the hydrogen bond formation between Asn1091 and Arg1095 being one example (**Figure 3.18D**).

3.5.2 Two dimers of Cnn-CM2 interlocked into an antiparallel tetramer

PISA analysis [Krissinel and Henrick, 2007] was carried out with the Cnn-CM2 dimer and the tetramer to examine interfaces present in the crystal. Importantly, the tetrameric assembly is suggested to be the most stable structure with a predicted solvation free energy (ΔG) upon formation of -152.2 kcal/mol. This assembly buries 9640.6 Å^2, which is 65% of the total solvent accessible area, and requires 52.6 kcal/mol for dissociation. Thus, the most significant interface present is the tetramer interface between antiparallel dimers, rather than dimer interface between parallel monomers. In particular, interactions between chain B and chain B' largely contribute to tetramer formation, with a Complexation Significance Score (CSS) of 1.0 (on the scale from 0 to 1). The antiparallel dimers have on average 841.2 Å^2 buried, whereas the parallel dimer interface buries only ~600 Å^2.

The tetramer topology can be viewed as two dimers exactly opposite to each other intercalated into a tight lock. I have used the SOCKET program [Walshaw and Woolfson, 2001] to try to determine antiparallel interactions, from which it was judged the tetramer interfaces could be divided into 3 areas with region 1 centred around the hinge and two structurally equivalent region 2s on either side (**Figure 3.21A**). I attempted to use SOCKET [Walshaw and Woolfson, 2001] to assign heptad registers for the antiparallel dimers (**Figure 3.21B**). However, the new registers do not continue with the ones assigned for the parallel dimer (**Figure 3.16C**); Leu1108 adopts position *a* in the parallel dimer but adopts position *d* in the antiparallel CC, and no register can be assigned from Arg1113 to Val1122 corresponding to the most kinked region, largely because the knob-into-hole interactions are missing. The Cnn-CM2 structure indicates this protein does not adopt a canonical CC topology.

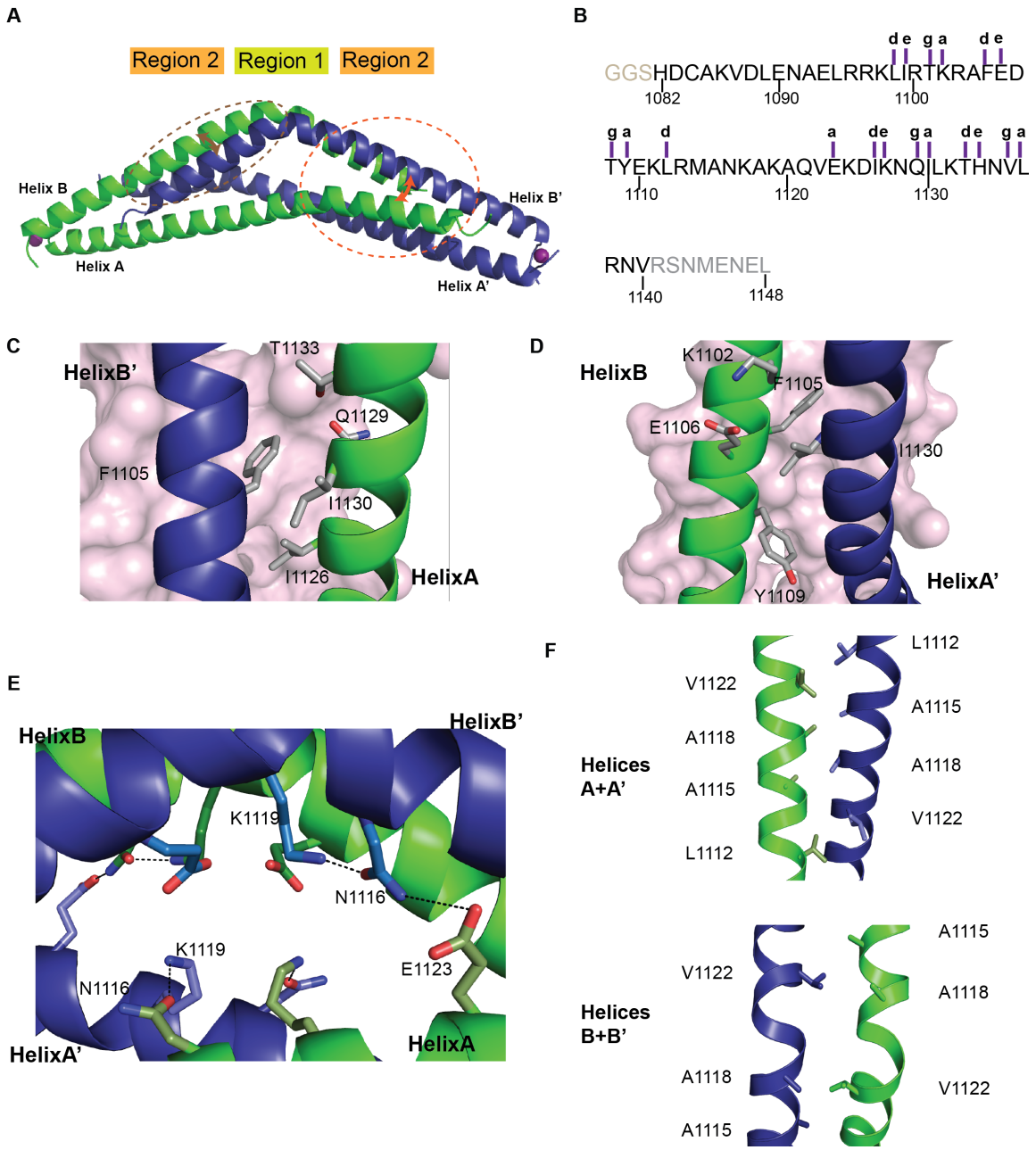


Figure 3.21 Analysis of the Cnn-CM2 antiparallel tetramer interface. (**A**) The antiparallel interactions can be divided into 3 regions centred at the hinge (region 1) with two other inverted regions. Each helix is involved in 2 sets of interactions (in circles), either with the chain of similar shape (bent or straight; indicated with a *brown* arrow) or with the other chain in the parallel dimer pair (indicated with an *orange* arrow). (**B**) Protein sequence of Cnn-CM2 with the new heptad register assigned by SOCKET based on the antiparallel dimer interactions. (**C**) Side-on view of a hydrophobic cluster. The knob, Phe/F1105, fits into the hydrophobic pocket created by side chains of the adjacent helix. (**D**) Another example showing the knob-into-hole packing at the antiparallel dimer interface. The side chain of Lys/K residue has oriented in such a way that the long carbon-containing 'arm' contributes to the hydrophobic core while the charged terminus is solvent accessible. (**E**) Side-on view of the hydrogen bond network within the kinked region. Note the interactions are mirror imaged, thus only one pair is labelled for each interaction. (**F**) Additional hydrophobic interactions can be seen around the kink.

Each tetramerisation helix is involved in two sets of antiparallel interaction with the opposite dimer pair (**Figure 3.21A**). In one set (chain A with B' and chain A' with B; highlighted with an *orange* arrow in **Figure 3.21A**), hydrophobic knob-into-hole packing is frequently observed at the dimer interface within interaction region 2s. This could be demonstrated as in the case between

Phe1105, Ile1126, Gln1129, Ile1130 and Thr1133 (**Figure 3.21C**). In some cases like Ile1130 (chain A') as a knob fitted into the hole created by Lys1102, Phe1105, Glu1106 and Try1109 (chain B). Although Lys is a charged residue, it is also considered as amphipathic with a long carbon-containing side chain that can be part of the hydrophobic core (**Figure 3.21D**). In region 1 centred at the kink, mirror imaged hydrogen bond and salt bridge formations occur with Lys1119 (**Figure 3.21E**). Lys1119 on chain A/A' is hydrogen bonded with Asn1116 on the adjacent chain (chain A'/A). Lys1119 on the bent chains (chain B/B') forms an intra-helical hydrogen bond with Asn1116 that is further hydrogen bonded with Glu1123 on the neighbouring helix (chain A'/A).

For the other set of antiparallel interactions (chain A with chain A' and chain B with chain B', highlighted with a *brown* arrow in **Figure 3.21A**), the majority of the hydrophobic interactions detected using CC+ server gathers around the kinked region. I notice that the 3 hydrophobic residues (Ala1115, Ala1118 and Val1122) within that region are all interacting with the same residues on the neighbouring chain (**Figure 3.21F**). This presumably provides further stability for the helical kink. Chain A/A' and chain B/B' are also observed to associate via hydrogen bonds. For example, the OH group of the Thr residue can make 2 hydrogen bonds with the main chain and side chain carbonyl groups, as seen in the case of Thr1108B, Thr1133B' and Gln1129B' (**Figure 3.22**).

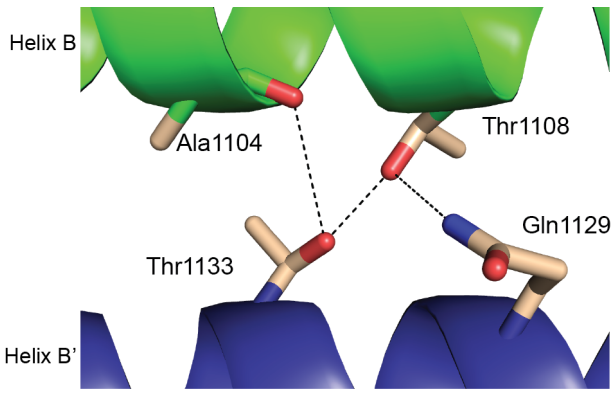


Figure 3.22 Hydrogen bond network linking adjacent helices. The OH group of Thr residues can be an acceptor as well a donor hence hydrogen bonds are formed with both the main chain (Ala) and side chain (Thr) carbonyl groups. The Thr OH group is also hydrogen bonded with the Gln amino group.

3.5.3 Conservation of the Cnn-CM2 structure

I mapped the sequence conservation onto the Cnn-CM2 structure using Consurf [Glaser et al., 2003], (**Figure 3.23**). Strikingly, the majority of the highly conserved residues are involved in antiparallel interactions that form a stripe across the C-terminal tetramerisation interface (**Figure 3.23A**). The residues involved in antiparallel interactions and located N-terminal to the kinked region show conservation in terms of the amino acid property (polar, hydrophobic or charged) (**Figure 3.23B**). In particular, the four invariant residues, Leu1094, Ile1126, Thr1133 and Leu1137, are all at *a* and *d* positions of the heptad repeat. Ile1126, Thr1133 and Leu1137 are involved in hydrophobic knob-into-hole packing for tetramer formation, as shown in **Figure 3.23C**. Leu1094 is positioned at the dimerisation interface, and is involved in hydrophobic interaction with the same residue on the adjacent helix (**Figure 3.18E**). Apart from these invariant residues, two well-conserved residues, Glu1123 and Gln1129, show sequence identity in all species included in my MSA apart from *Xenopus tropicalis*, and they are also present at the tetramerisation interface. Glu1123 locates around the hinge region and forms hydrogen bond network with the other two highly conserved residues Asn1116 and Lys1119 (**Figure 3.23D**). Gln1129 is in a *g* position that faces towards the periphery. Its side chain is orientated in such a way that the hydrophobic 'arm' contributes to the tetramer core and the terminal charged groups form a hydrogen bond with Thr1108 on the neighbouring chain hence facilitating Cnn-CM2 self-association (**Figure 3.22**).

In contrast, the solvent accessible residues show a much higher degree of variation. This strongly suggests that the tetrameric packing I observe in the Cnn-CM2 crystal structure might be biologically relevant. In order to investigate its significance, mutants were designed to disrupt tetramerisation by mutating the three invariant, buried residues (Ile1126, Thr1133 and Leu1137) to Ala or Glu. Mutation of any these residues to Glu would drastically disrupt the packing of the hydrophobic core and subsequently the tetramer formation. Substitution of these residues with Ala might weaken this tetramer interface. Fortunately, T1133A/E was one of the mutations made previously in the lab. This residue is involved in the hydrophobic knob-into-hole packing (**Figure 3.23B**) as well as creating a hydrogen bond network with residues on the neighbouring helix (**Figure 3.22**). The effect of the T1133A/E mutation on Cnn-CM2 oligomerisation state has been assessed *in vitro*, and the results are documented in **Section 3.6.5**. In addition, the *in vivo* data for this mutant are presented in **Section 3.7.2**.

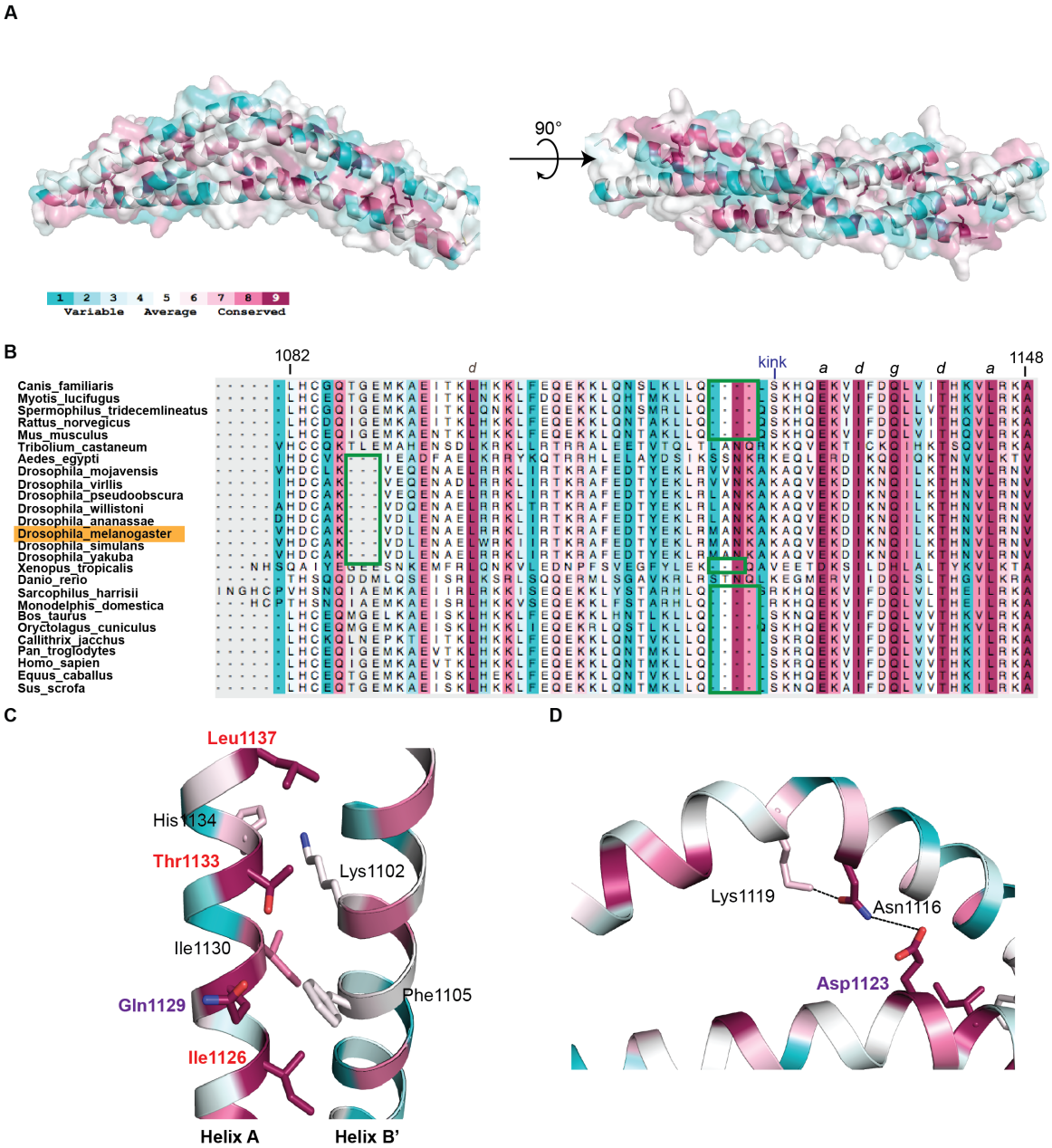


Figure 3.23 Analysis of Cnn-CM2 conservation across evolution. (**A**) Sequence conservation was mapped onto the Cnn-CM2 tetramer structure using Consurf. Residues are coloured according to the Consurf conservation score. Left, side-on view of the structure showing the side chain of the invariant residues. Right, top-on view of the structure. Note the invariant residues are buried at the dimerisation/tetramerisation interface. (**B**) MSA for Cnn-CM2 across evolution. Note that only residues seen in the structure are included in this sequence alignment. The C-terminal residues buried in the tetramer are highly conserved. Residues involved in antiparallel interactions and located at the N-terminal side of the kink show conservation in terms of amino acid property. The heptad positions assigned to invariant/well-conserved residues are displayed above the MSA. Fly homologs show a gap in sequence conservation after Lys1086 (highlighted with a *green* box), whereas vertebrate homologs miss ~1 turn of helix around the hinge region (*green* box). This presumably makes sure the heptad register is restored for those highly conserved residues at the C-terminus, but further structural characterisation into homologous systems is required. (**C**) Side-on view demonstrating how the invariant residues are involved in tetramer packing. The invariant residues Ile1126, Thr1133 and Leu1137 (highlighted in *red*) are all involved in the hydrophobic knob-into-hole interaction. The side chain arm of Gln1129 (a well-conserved residue at *g* position; highlighted in *purple*) also contributes to the hydrophobic core. (**D**) The well-conserved Asp1123 residue (identical in all species apart from *X. tropicalis*, as shown in (**B**)) is part of the hydrogen bond network formed with 2 other conserved residues Lys1119 and Asn1116 (100% identity in fly sequences).

3.5.4 The Zn^{2+} binding sites within Cnn-CM2 may be evolutionarily conserved

Most unexpectedly, the Cnn-CM2 domain binds to a Zn atom. I have taken advantage of the anomalous scattering from Zn^{2+} for phase information in structure determination. The final structure of Cnn-CM2 reveals its Zn^{2+}-mediated dimerisation, where His1082 and Cys1084 from each monomer bind to a Zn^{2+} ion with a tetrahedral geometry (**Figure 3.24A,B**). This coordination sphere is known as C_2H_2 and is the most common arrangement seen in 82% of structures that bind Zn^{2+} [da Costa et al., 2014]. This is because the Zn^{2+} ion has a filled d-shell and it prefers coordination by four atoms. In addition, the indole-ring nitrogen of His1082 forms a cross-chain salt bridge with Asp1088 positioning the His residue for Zn^{2+} binding and also linking the adjacent chains (**Figure 3.24B**). Note that although a salt bridge would only form if the pH were below 6, the pI of the His is raised by the Zn^{2+} coordination.

Strikingly, the MSA for Cnn-CM2 across evolution indicated conservation of the His and Cys residues at the apex of this motif (**Figure 3.24C**). While these two residues are spaced by an Asp residue in fly sequences, they are adjacent to each other in several other family members. In the structure of *Drosophila* Cnn-CM2, the Zn^{2+}-binding motif interrupts the helical structure. Therefore, the neighbouring His and Cys residues seen in other CM2 sequences would likely also accommodate Zn^{2+} binding due to the flexibility of the region. To my knowledge, this is the first time a centrosomal protein has been found to bind a Zn^{2+} ion. The role of Zn^{2+} binding has been reported in various activities including gene expression, catalysis and structural stabilisation [Ganss and Jheon, 2004; Konchanczyk et al., 2015; Krishna et al., 2003; Laitaoja et al., 2013; McCall et al., 200]. It would be very interesting to investigate its role in in this case, both *in vitro* and *in vivo* (**Chapter 3.6 and 3.7**).

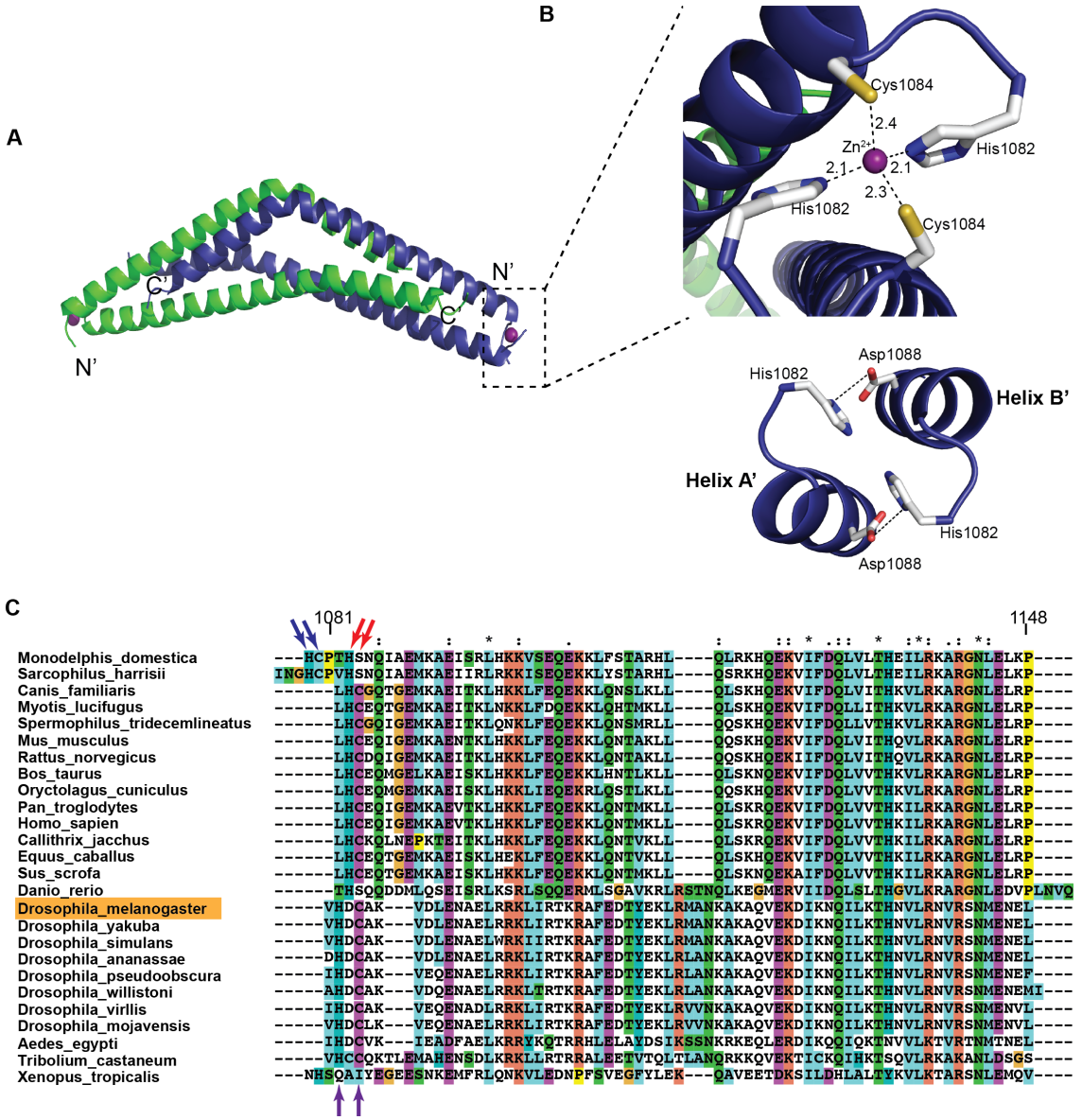


Figure 3.24 The Zn^{2+}-binding sites of Cnn-CM2. (**A**) The position of Zn^{2+} ions (*purple* sphere) in the overall tetramer structure of Cnn-CM2. The parallel dimer pair is shown in the same colour (chain A and B in *green*; chain A' and B' in *blue*). (**B**) Close-up view of Zn^{2+} coordination by His and Cys residues from Helix A' and Helix B' demonstrating a tetrahedral geometry. The Zn-N/S bond lengths are shown in Å. The imidazole nitrogen of His1082, on the side of its aromatic ring away from the Zn^{2+} binding interface, forms a salt bridge with Asp1088 on the neighbouring chain. This interaction positions the His residue for Zn^{2+} binding, as well as providing further dimer stability. (**C**) MSA analysis of CM2 domain from flies and vertebrates showing the presence of potential Zn^{2+} binding sites, as pointed, at the N-terminus. Arrows of different colours showing the position of His and Cys residues in *Sarcophilus harrisii* and *Monodelphis domestica* (*blue*), other vertebrates (*red*) and flies (*purple*). Residue numbers for *D. melanogaster* are indicated. Invariant residues are highlighted with asterisks, conserved residues with colons and semi-conserved residues with full stops. Residues are coloured according to the ClustalX colour scheme.

3.6 Further characterisation of Cnn-CM2

3.6.1 Characterisation of Zn^{2+} binding *in vitro*

As previously described in **Chapter 3.2.2**, SEC-MALS analysis of $Cnn^{1082-1148}$ showed the formation of a dimeric species in solution. In the crystal lattice, both antiparallel and parallel dimers were observed. In order to confirm which dimer was seen in solution, I diluted the SEC purified $Cnn^{1082-1148}$ to ~10 µM and then dialysed the sample overnight at 4 °C into WB buffer plus 1 mM EDTA. If the dimer seen with SEC-MALS were parallel and formed via binding to Zn^{2+} ion, the addition of EDTA would chelate the metal ion in solution and disrupt the dimerisation. $Cnn^{1082-1148}$ was prepared at ~100 µM and injected onto an S200 10/30 column (GE Healthcare) using WB buffer as the running buffer. The SEC-MALS trace showed that $Cnn^{1082-1148}$ eluted as a monomer in the presence of EDTA suggesting the Zn^{2+}-bound parallel dimer was seen in solution (**Figure 3.25A**).

I further investigated the *in vitro* significance of Zn^{2+} binding by mutating the binding sites to Ala. Unfortunately, at this stage of the project, I only had the early model refined in the Cnn-CM2_1 dataset- showing Asp1083 and Cys1084 coordinating the Zn^{2+} ion (based on the interpretable density). As a result, I designed a DCtoAA mutant that has the Asp and Cys residues mutated to Ala. In the final model, refined from the Cnn-CM2_7 data, Asp1083 is solvent exposed and His1082 is the other Zn^{2+} binding site. Thus, the DCtoAA mutant is not ideally designed to interfere with Zn^{2+} coordination, although I did not appreciate this at the time. $Cnn^{1082-1148}$ DCtoAA was expressed and purified using the same protocol as the wild-type construct. Protein was prepared at 100 µM and 300 µM for SEC-MALS analysis, using the same column and running buffer as on $Cnn^{1082-1148}$ WT. As expected, mutation of the Zn^{2+}-binding interface abrogated dimerisation in solution (**Figure 3.25B**).

I also conducted a CD experiment on the DCtoAA mutant assessing its folding state. For direct comparison, a wild-type sample was also analysed on the same day. Both proteins, diluted to 0.2 mg/ml, were dialysed into 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.0) overnight at 4 °C. CD spectra were collected in 4 accumulations scanning between 190 and 260 nm at 20 °C, and normalised using a blank measurement from the dialysis buffer. In contrast to WT, the CD spectrum of $Cnn^{1082-1148}$

DCtoAA mutant showed a profile representative of mixed alpha-helical and random-coil contents (**Figure 3.25C**). This data strongly suggests that Zn^{2+} coordination plays an essential structural role for Cnn-CM2 *in vitro*.

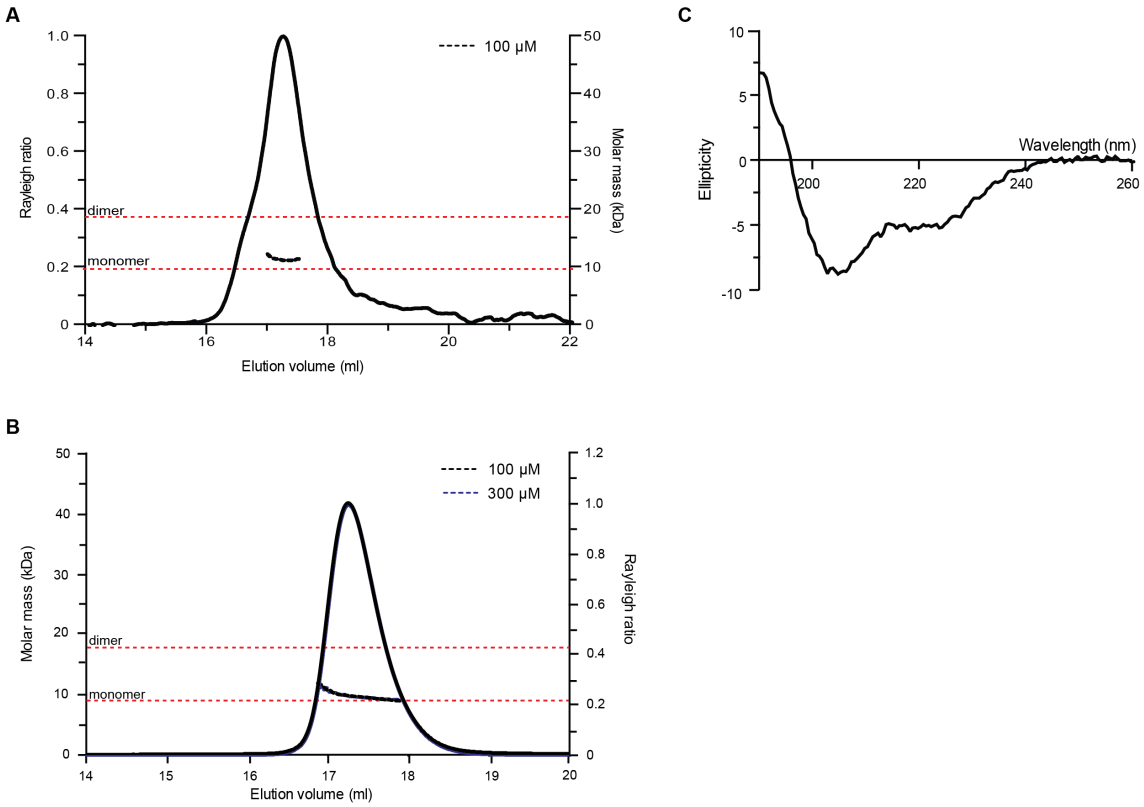


Figure 3.25 The *in vitro* significance of Zn^{2+} binding is analysed by SEC-MALS and CD experiments. (**A**) SEC-MALS analysis of $Cnn^{1082-1148}$ that has been dialysed into 50 mM Tris (pH7.0), 150 mM NaCl, 1 mM EDTA and 5 mM BME. (**B**) $Cnn^{1082-1148}$ DCtoAA mutant with or without the addition of 20 µM $ZnCl_2$. Samples were injected over an S200 10/30 column that was equilibrated into 50 mM Tris (pH7.0), 150 mM NaCl and 5 mM BME. The observed molecular mass shown as a dashed line is consistent with a highly monomeric species. The disruption of dimer formation in the addition of EDTA suggests Zn^{2+} binding is the primary contact interface. (**C**) CD spectrum of $Cnn^{1082-1148}$ DCtoAA mutant suggesting the protein is only partially folded in solution. Analysis was carried out in 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.0). This has demonstrated that Zn^{2+} plays an essential structural role.

3.6.2 Can Cnn-CM2 form a tetramer in solution?

Despite the observation of energetically favourable tetramers *in crystallo*, no *in vitro* evidence thus far has shown tetramer formation in solution. There are several possible explanations that I decided to test to see if I could detect tetramers in solution.

First, an insufficient amount of Zn^{2+} in the purification buffer could result in the formation of less stable dimers, and thus discourage the formation of tetramers. I therefore decided to supplement

the protein sample with $ZnCl_2$ and then repeat the SEC-MALS measurements. However, $ZnCl_2$ can promote protein precipitation, and an optimal concentration needs to be determined so that it is not too high to cause precipitation but high enough to saturate all the Zn^{2+} binding sites. I gradually added the 5 mM $ZnCl_2$ stock solution to SEC purified $Cnn^{1082-1148}$ sample in 10 µM increments. At 300 µM protein concentration, fine precipitate was observed upon the addition of more than 30 µM $ZnCl_2$. The sample was then prepared at 220 µM in the presence of 20 µM $ZnCl_2$ for SEC-MALS experiments. Alternatively, the sample at 220 µM was dialysed against WB plus 10 µM $ZnCl_2$ overnight at 4 °C. I kept a small amount of $ZnCl_2$ in the dialysis buffer in order to saturate all the Zn^{2+} binding sites on Cnn-CM2 while not causing any precipitation. SEC-MALS experiments were performed using the same column and running buffer as on $Cnn^{1082-1148}$ WT. Both SEC-MALS traces showed the elution of dimeric species (**Figure 3.26**).

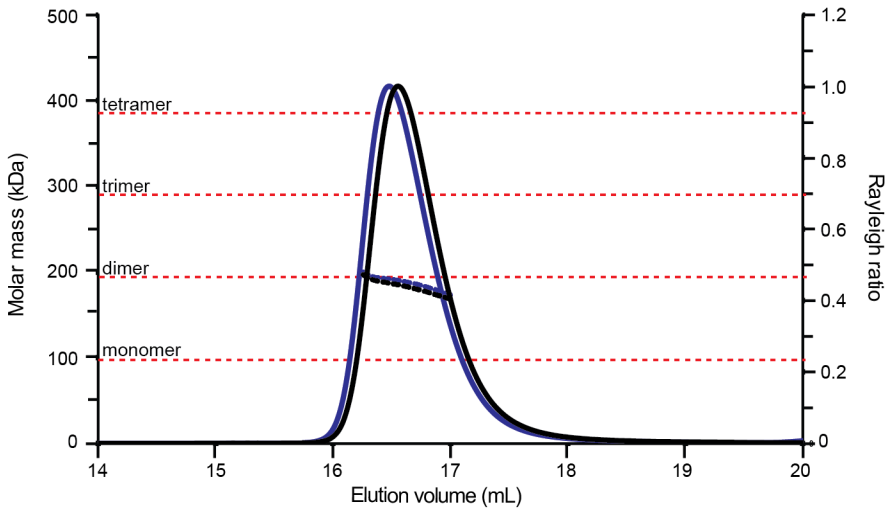


Figure 3.24 SEC-MALS analysis of $Cnn^{1082-1148}$ supplemented with $ZnCl_2$. Sample was concentrated to 220 µM before the addition of 20 µM $ZnCl_2$ (*black* curve/dashed line). Alternatively, the sample at 220 µM was dialysed overnight into 50 mM Tris (pH7.0), 200 mM NaCl and 5 mM BME plus 10 µM $ZnCl_2$ (*blue* curve/dashed line). Elution with 50 mM Tris (pH7.0), 150 mM NaCl and 5 mM BME. The solid line indicates protein elution off an S200 10/30 column. The dashed line shows the observed molecular weight across the peak. The expected mass for different oligomer states is also displayed on the diagram (red dotted line across the graph). Both protein samples eluted as dimeric species.

Second, it is possible that the tetramer assembly is of low affinity and the protein concentration is too low in solution for observation of this self-association state. Crystallisation occurs when protein slowly precipitates out of solution, driving the formation of ordered crystal lattices. Because the mother liquor concentration is higher in the reservoir well than in the crystallisation drop, vaporisation will occur in the drop until reaching equilibrium with the reservoir solution. In this vapour diffusion process, the protein concentration is increased and is therefore higher than

in solution (the estimated protein concentration *in crystallo* is ~39 mM). Moreover, in SEC-MALS experiments the protein in solution is further diluted by 5-10 fold as the protein sample migrates though the SEC column matrix. If tetramerisation has an association constant of hundred μM to mM, SEC-MALS analysis with protein at 220 μM (the effective concentration for the measurement is 44 μM**; Section 3.2.2**) would not be able to detect any tetramer formation.

In order to increase the protein concentration of Cnn-CM2 in solution, I needed to solve the problem of its low solubility. I first tested the effect of salt concentration on protein solubility. Although Cnn$^{1082\text{-}1148}$ was more soluble in the presence of a higher salt concentration, the protein was still insoluble beyond 430 μM (in 50 mM Tris (pH7.0), 220 mM NaCl, 5mM BME). The TEV cleavage of diLipoyl-Cnn$^{1082\text{-}1148}$ leaves 11 linker residues at the C-terminus of Cnn-CM2 protein sequence. These residues include Phe, Try and Gln that all have large and hydrophobic side chains. I suspected that such amino acids might affect protein solubility. In addition, the cleavage site was not observed in the crystal structure suggesting it was not within a folded region. I modified the pLip_Cnn$^{1082\text{-}1148}$ plasmid by introducing a premature stop codon after the Cnn-CM2 coding sequence using site-directed mutagenesis. The expressed fusion protein would therefore contain a single Lipoyl domain at the N-terminus of Cnn$^{1082\text{-}1148}$ (hereafter refer to Cnn$^{1082\text{-}1148_noTEV}$; construct details in **Table 3.10**). I expressed and purified the protein using the same protocol as the diLipyol-Cnn$^{1082\text{-}1148}$ (**Figure 3.27**). Interestingly, the solubility of Cnn$^{1082\text{-}1148_noTEV}$ significantly increased, and I could concentrate it to ~50 mg/ml without any precipitation. Its oligomerisation state was then investigated by two techniques: SEC-MALS and One Dimension Diffusion-ordered NMR spectroscopy (1D DOSY).

Construct	MW (kDa)	pI	Abs 0.1%	No. of Cys
Lipoyl-Cnn$^{1082\text{-}1148}$	19.70	5.58	0.506	2
Cnn$^{1082\text{-}1148_noTEV}$	8.18	9.66	0.182	1

Table 3.10 Details of Cnn$^{1082\text{-}1148_noTEV}$ with and without the N-terminal Lipoyl domain calculated using Protparam [Wilkins et al., 1999]. Due to the presence of cysteine residues, 5 mM β-mercaptoethanol (BME) was included in working buffer (WB) to prevent disulfide bonds.

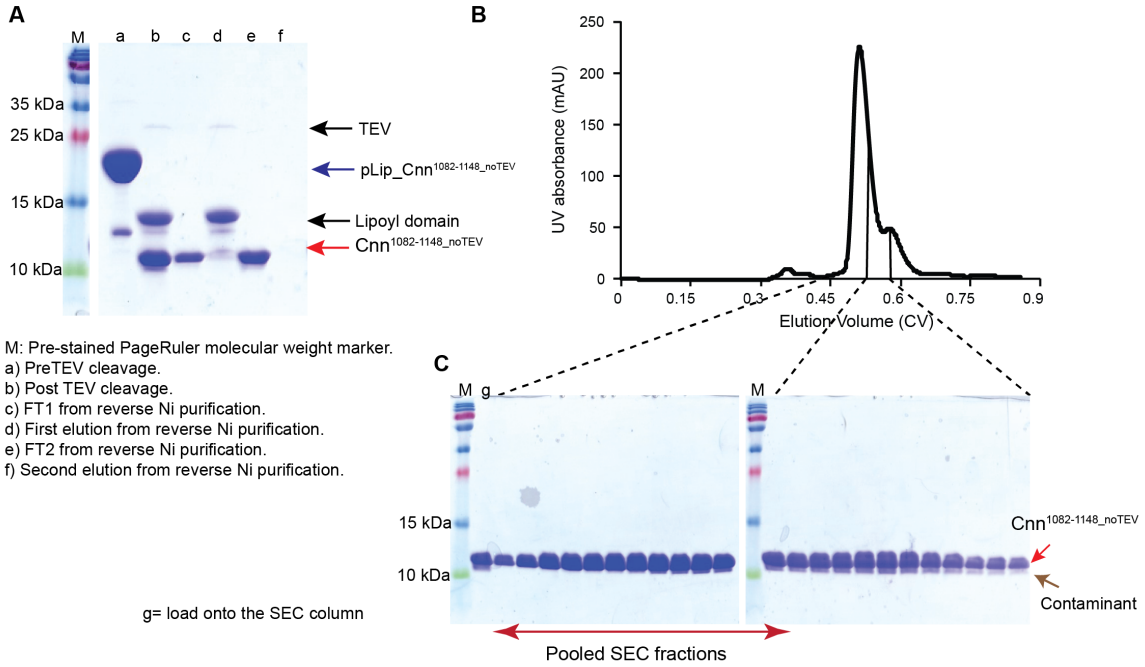


Figure 3.27 Purification of Cnn$^{1082\text{-}1148_noTEV}$ after TEV cleavage. (**A**) 15% reducing SDS-PAGE gel analysis of reverse Ni-NTA purification showing the removal of free Lipoyl domain and TEV protease. The untagged Cnn was isolated and collected in the FT. (**B,C**) S75 26/60 SEC trace and the subsequent 15% reducing SDS-PAGE gel analysis demonstrating the elution of full-length Cnn protein in the major peak around 0.5 CV. Elution with 50 mM Tris (pH 7.0), 200 mM NaCl and 5 mM BME. A minor contaminant was observed towards the second half of that peak, which overlapped with the following small peak. For purity, only the first few SEC fractions within the major elution peak, as indicated on the gel, were pooled.

For SEC-MALS experiments, I prepared Cnn$^{1082\text{-}1148_noTEV}$ at 0.4 mM, 2 mM and 4 mM. Samples were injected onto an S200 10/30 column (GE Healthcare) using 50 mM Tris (pH 7.0), 200 mM NaCl, 5 mM BME as the running buffer. The results demonstrated that protein eluted as a highly dimeric structure at all concentrations (**Figure 3.28**). This observation confirmed that dimerisation of the Cnn-CM2 domain in solution was not due to the remaining TEV site left on the C-terminus. However, I still could not detect any tetramer formation in solution. As mentioned above, the protein concentration is diluted during its migration through the SEC column (the effective concentrations for the measurements are therefore ~80 μM, ~400 μM and 800 μM) and this might explain why only dimers were observed in SEC-MALS.

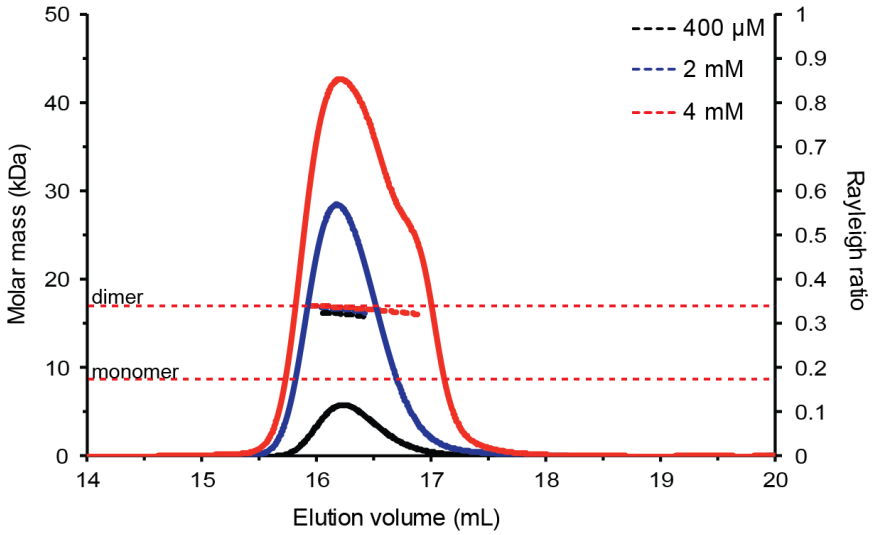


Figure 3.28 SEC-MALS analysis of Cnn$^{1082\text{-}1148_noTEV}$ showing the elution of dimeric species at all tested concentrations. Elution with 50 mM Tris (pH 7.0), 200 mM NaCl and 5 mM BME. Solid lines show the relative Rayleigh ratio, and dashed lines show the average mass across each peak. Note that the S200 10/30 column was largely saturated when the sample was loaded at 4 mM and thus the elution profile appears much broader.

Additionally, I studied the multimerisation state of Cnn$^{1082\text{-}1148_noTEV}$ by measuring its diffusion coefficient in solution using 1D DOSY NMR experiments. These measurements were carried out with the help of Dr. Devon Sheppard. The data was collected on a Bruker Avance II 500MHz spectrometer. There are several advantages associated with DOSY NMR including possible sample recovery after analysis, wide dynamic range and high sensitivity. More importantly, the protein sample is not diluted in this experiment, potentially allowing the detection of weaker interactions. The diffusion coefficient (D) that defines how fast a molecule tumbles in solution is estimated in 1D DOSY, and its value depends on the size and shape of the molecule. I compared the D values of Cnn$^{1082\text{-}1148_noTEV}$ at various concentrations to assess its self-interaction. In diffusion measurements, a gradient is applied that introduces variation in the magnetic field strength with respect to the position in the sample tube. The gradient strength, varied from 5 to 95%, defines the extent of variation in field strength. As molecules diffuse, there is a change in the magnetic field and thus an accompanying change in the precession rate of a particular spin. Having equivalent spins with different rates creates a dephasing of spins, which leads to loss of signal intensity. This change in signal intensity was then plotted against the gradient strength. I have also taken the same measurements at different diffusion delays, D200= 150 µs and 200 µs. Plots from both delay times were used for calculation of the diffusion coefficient. The D values were measured at a range of protein concentrations: 250 µM, 500 µM, 1 mM, 2 mM and 4 mM.

For NMR analysis, the buffer components have to be non-protonated. I therefore dialysed SEC purified $Cnn^{1082\text{-}1148_noTEV}$ (at 350 μM) into NMR buffer (50 mM NaH_2PO_4/Na_2HPO_4 (pH 7.0), 220 mM NaCl, 1 mM DTT and 10 μM $ZnCl_2$). The protein sample was concentrated to 4.2 mM using a 3 kDa MWCO centrifugal concentrator (Amicon), and 100% D_2O (Sigma) was added to a final concentration of 5% *v/v*. The resulting concentration of $Cnn^{1082\text{-}1148_noTEV}$ was therefore 4 mM. Cnn-CM2 samples at all other concentrations were prepared by serial dilution from 4 mM with the NMR buffer plus 5% *v/v* D_2O. The diffusion coefficient values at various concentrations are plotted in **Figure 3.29**. It seems that $Cnn^{1082\text{-}1148_noTEV}$ rotated more slowly in solution as its concentration increased. In particular, more than 10% reduction in D value was observed as the amount of $Cnn^{1082\text{-}1148_noTEV}$ doubled (from $0.66x10^{-10}$ m^2s^{-1} at 2 mM to $0.55x10^{-10}$ m^2s^{-1} at 4 mM). This decreasing trend in diffusion coefficient (slower molecular tumbling) continued, as indicated by another measurement at 5 mM protein concentration. These data suggest that $Cnn^{1082\text{-}1148_noTEV}$ can form higher-order assemblies at very high concentrations. From the solved crystal structure of Cnn-CM2, this protein is not symmetric in shape and is not globular, making the estimation of molecular weight from the calculated diffusion coefficient very difficult.

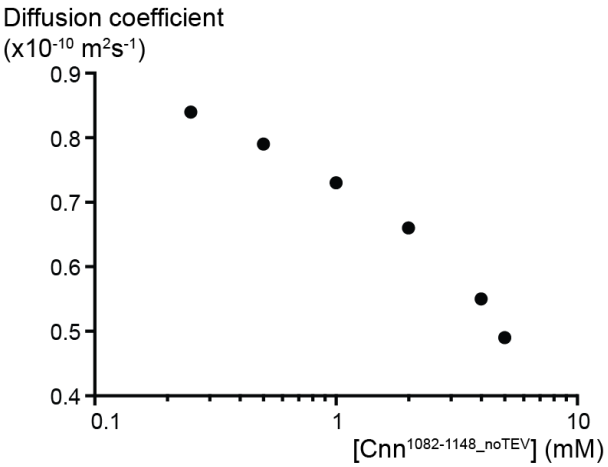


Figure 3.29 Diffusion coefficients measured for $Cnn^{1082\text{-}1148_noTEV}$ using 1D DOSY NMR spectroscopy in 50mM NaH_2PO_4/Na_2HPO_4 (pH 7.0), 220 mM NaCl, 1 mM DTT and 10 μM $ZnCl_2$. The protein appears to rotate slower in solution with the increasing concentration, suggesting the possibility of forming larger species.

3.6.3 Does the T1133E mutation disrupt the formation of higher-order Cnn-CM2 oligomers in solution?

Conservation analysis across the Cnn-CM2 structure indicated that residues at the putative tetramer interface are highly conserved in evolution, whereas the surface exposed residues are

more variable (**Figure 3.21**). This suggested the importance of tetramer formation, and that it would be worth analysing its potential functional role *in vivo*. From the MSA and Consurf analyses, Thr1133 is highly conserved from flies to vertebrates and is involved in the antiparallel tetramer interaction observed *in crystallo*. I therefore decided to investigate the effect of T1133E mutation on the multimerisation capacity of Cnn-CM2 *in vitro*.

The single point mutation (T1133E) was introduced into the pLip_Cnn$^{1082\text{-}1148}$ vector (encoding a single Lipoyl domain) by site-directed mutagenesis (Agilent). The mutant protein was expressed and purified on a large scale using the same protocol as for the wild-type Cnn$^{1082\text{-}1148_noTEV}$ construct. CD scans confirmed that the mutation had not disrupted the protein fold (**Figure 3.30A**). To analyse the self-association state of the T1133E mutant, SEC-MALS and DOSY NMR experiments were performed. For the SEC-MALS experiment, the T1133E mutant was prepared at 0.4 mM, 2 mM and 4 mM. The same column and running buffer were used as on Cnn$^{1082\text{-}1148_noTEV}$. At all tested concentrations, the mutant eluted at a similar volume compared to the wild-type protein and the observed masses were consistent with the protein being dimeric (**Figure 3.30B**). This was expected since the Thr1133 residue is not involved in parallel dimer formation as observed in the Cnn-CM2 structure.

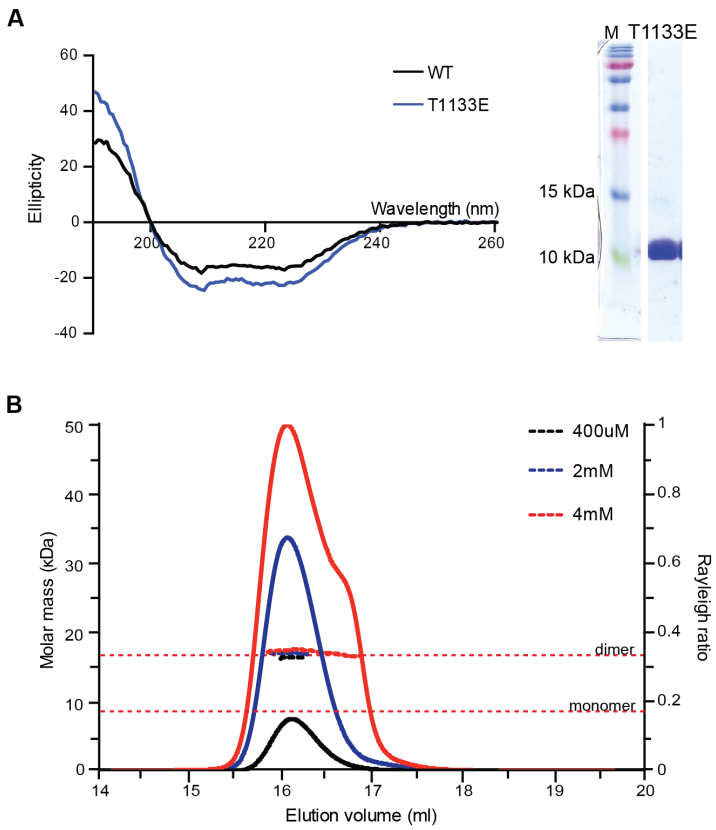


Figure 3.30 Purification and biophysical characterisation of Cnn$^{1082\text{-}1138_noTEV}$ T1133E mutant. (**A**) CD analysis of Cnn$^{1082\text{-}1148_noTEV}$ T1133E mutant in 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.0). The background-subtracted traces of WT and point mutant both indicate a highly alpha-helical fold. 15% reducing SDS-PAGE gel analysis on the right showing the purified point mutant. M: PageRuler Plus Prestained Protein Ladder (Thermo Scientific). (**B**) SEC-MALS analysis to test the effect of T1133E mutation on Cnn-CM2 self-association. Elution with 50 mM Tris (pH 7.0), 200 mM NaCl and 5 mM BME. Solid lines represent the relative Rayleigh ratio, and dashed lines show the observed molecular weight. The point mutant eluted as a dimeric species.

I next examined the diffusion rate of T1133E mutant using 1D DOSY NMR. The mutant was dialysed overnight into NMR buffer, and sample at different concentrations was prepared following the same procedure as for the wild-type construct. Measurements were carried out at two diffusion delays (D200= 150 μs and 200 μs), like in the case of WT. The diffusion coefficients were calculated and compared with $Cnn^{1082\text{-}1148_noTEV}$ WT (**Figure 3.31**). Surprisingly, the T1133E protein exhibited a higher diffusion coefficient (i.e. faster tumbling in solution) at higher protein concentrations (4 mM and above). This result suggests that the mutant protein might be more aggregative and precipitating at high concentration due to its lower stability in solution, despite no obvious precipitate formation being observed. Due to time limitations I have not repeated these experiments. Thus, I remain cautious in interpreting this data, although this approach seems worthy for further investigation.

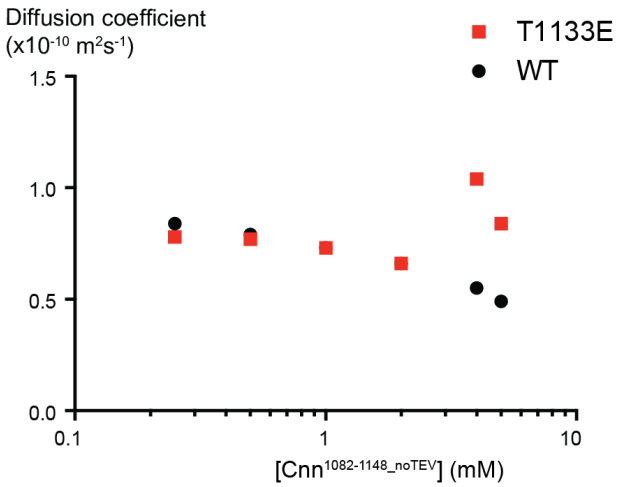


Figure 3.31 Diffusion coefficients measured for $Cnn^{1082\text{-}1148_noTEV}$ T1133E mutant protein using 1D DOSY NMR spectroscopy in 50mM NaH_2PO_4/Na_2HPO_4 (pH 7.0), 220 mM NaCl, 1 mM DTT and 10 μM $ZnCl_2$. The mutant protein appears to tumble faster at higher protein concentrations, especially at 4 mM. The lack of a continuous trend, which is seen with the wild-type protein, suggests that the mutant protein might not be well behaving in solution and could be more aggregative. Further repeats are required in order to confirm this data.

3.7 *In vivo* studies of Cnn-CM2 structural features

3.7.1 Design of constructs

In **section 3.5**, I described how the Cnn-CM2 structure reveals several unexpected features. Cnn-CM2 binds to Zn^{2+} *in vitro*, which helps to link two Cnn-CM2 helices into an asymmetric parallel dimer. In the crystal lattice, two parallel asymmetric dimers assemble into an antiparallel tetramer, in which many of the most conserved CM2 residues are buried in the hydrophobic tetramer core. In order to analyse the importance of these structural features *in vivo*, I designed mutant constructs for mRNA injections into *cnn*$^{-/-}$ null mutant embryos (as described previously for my analysis of the Cnn PReM domain, **Chapter 2**), (**Figure 3.32**).

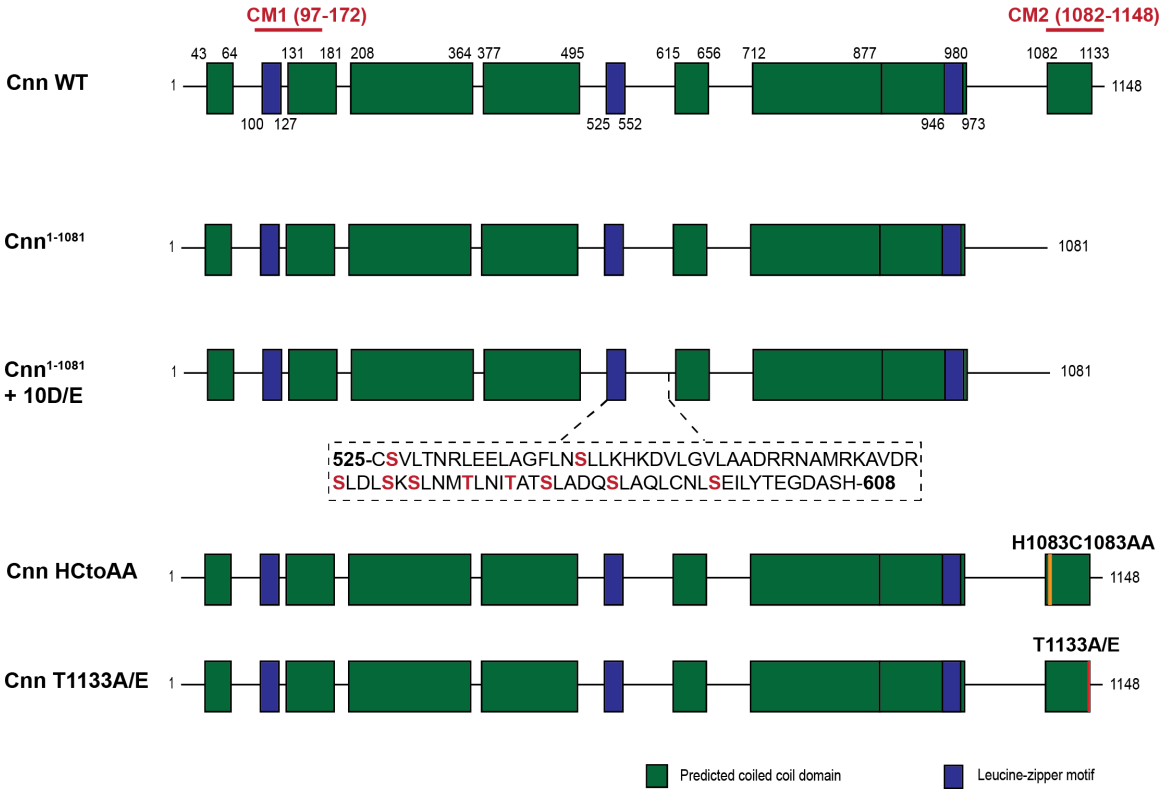


Figure 3.32 GFP-Cnn mutant constructs used for *in vivo* study in this project. Predicted CC domain boundaries are labelled for the wild-type construct. The amino acid termini for each construct are labelled. Cnn ΔCM2 + 10D/E construct contains phospho-mimiking mutations within the PReM domain (sequence shown in the box). Point mutations are marked by vertical lines.

N-terminally-fused GFP constructs carrying various mutations were generated by site-directed mutagenesis (Agilent) of the pRNA_GFPNT-Cnn WT vector. A construct lacking the entire CM2 domain (Cnn$^{1082-1148}$) was designed for analysis of its functional role in centrosome biogenesis. This truncation mutation was also combined with phosphomimetic mutations (10D/E) in the PReM

domain. As mentioned above, one of the mutations made in the lab was T1133A/E (a conserved residue at the tetramerisation interface). This allows us to examine the significance of the *in crystallo* observed tetramerisation for Cnn scaffold assembly, and the *in vivo* observations made in transgenic fly lines expressing T1133A/E mutations is presented in this section. The Zn^{2+} mutant was designed as explained in **Chapter 3.6.1** to assess the importance of metal coordination *in vivo*.

mRNA encoding different mutant constructs, as described above, were generated by *in vitro* transcription. All samples were prepared at 2 mg/ml and given to Prof Jordan Raff who carried out the fly husbandry and microscopy work for this part of the project.

3.7.2 Dual function of Cnn-CM2 in centrosomal localisation and scaffold stability

The GFP-Cnn fusion protein was first assayed for its centrosomal localisation in *cnn*$^{-/-}$ null mutant embryos expressing Spd-2-RFP. As expected, GFP-Cnn WT protein incorporated into robust scaffold-like structures at centrosomes (**Figure 3.33A**). Spd-2-RFP localised robustly to these centrosomes, demonstrating that the wild-type GFP-Cnn effectively rescued the PCM recruitment defects in the *cnn*$^{-/-}$ mutant embryos (**Figure 3.33A**).

Strikingly, the centrosomal localisation of Cnn was almost completely abolished in the absence of the CM2 domain, and the centrosomal localisation of Spd-2 remained severely disrupted, indicating that GFP-Cnn ΔCM2 could not rescue the PCM assembly defect (**Figure 3.33B**). Thus, as suggested previously in other systems [Wang et al., 2010], the CM2 domain appears to be essential for efficiently localising Cnn to centrosomes. Interestingly, the ΔCM2 phenotype was partially rescued if the PReM domain 10D/E mutation was also present, as small amounts of Cnn ΔCM2+10D/E localised diffusely in the PCM and this protein appeared to be able to associate with Spd-2-RFP (**Figure 3.33C**). Note that the levels and contrast in the images shown in **Figure 3.33B,C** have been adjusted relatively to **Figure 3.33A** to allow better visualisation of the very small amounts of the mutant Cnn proteins at the centrosomes.

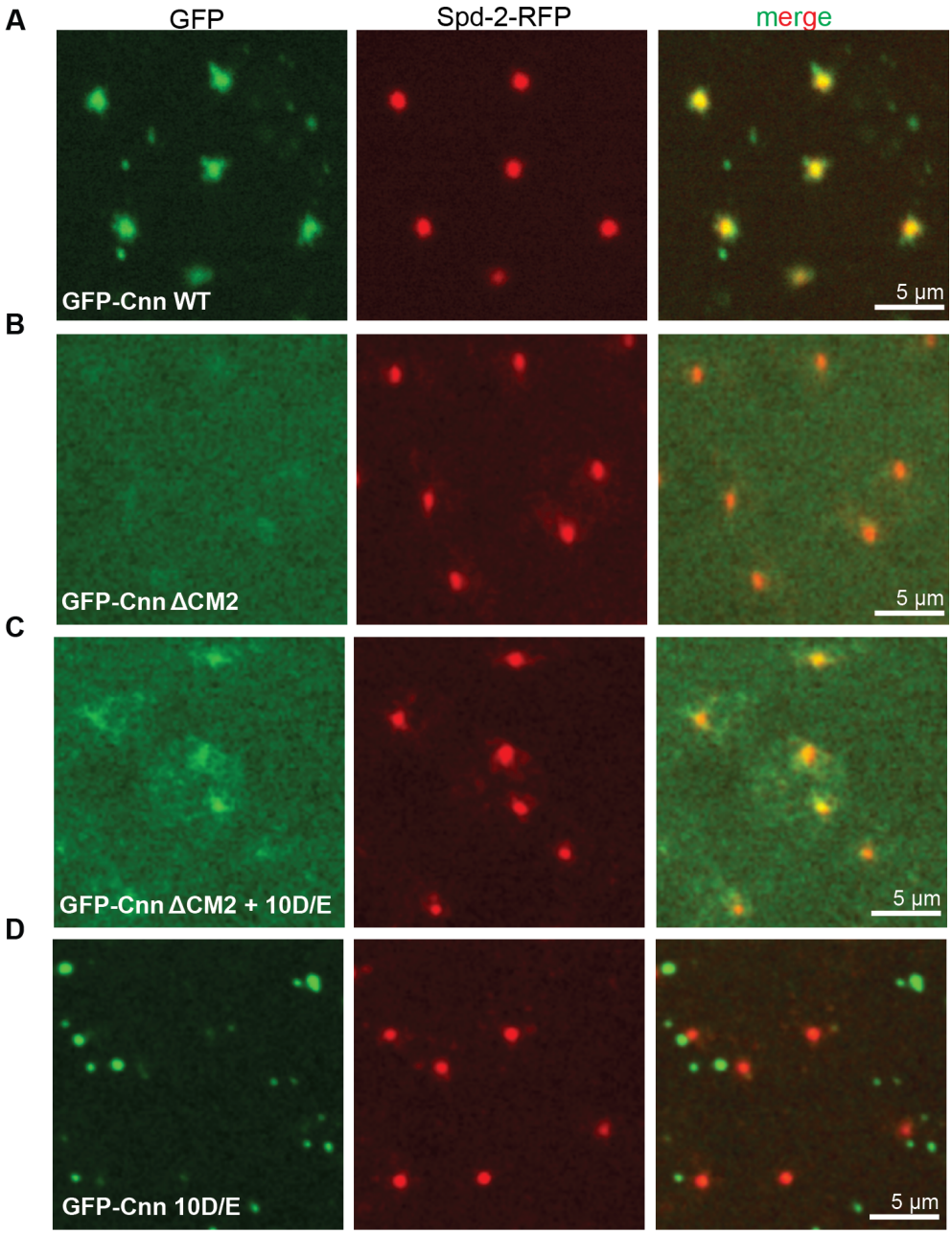


Figure 3.33 Centrosomal localisation of Cnn constructs in null mutant background. (**A**) GFP-Cnn WT has rescued the null mutant showing normal centrosomal localisation. Note GFP-Cnn has a broader PCM expansion compared to Spd-2-RFP. (**B**) GFP-Cnn ΔCM2 showed defects in centrosomal localisation. Mini flares of Spd-2-RFP were observed breaking off the centrosome. This is expected since Cnn is required for Spd-2 maintenance. (**C**) The disrupted localisation seen in (**B**) was partially rescued by phosphomimetic mutations within the PReM domain. GFP-Cnn 10D/E normally formed prominent foci in the cytoplasm, as shown in (**D**). The scaffolds assembled by GFP-CnnΔCM2 10D/E were loosely spread with significant co-localisation of Spd-2-RFP. (**D**) Aggregates of GFP-Cnn 10D/E formed spontaneously in the cytoplasm. Note Spd-2-RFP signal does not co-localise with GFP-Cnn in those cytoplasmic foci. Images were taken using a Perkin Elmer ERS spinning disk (Volocity software) mounted on a Zeiss Axiovert microscope using a 63X/ 1.4NA oil immersion objective and on Orca ER CCD camera (Hamamatsu Photonics, Japan). Images shown are maximum intensity projections of several z-slices. Note that in order to show weak GFP fluorescence signal, images have been adjusted to use the full scale of intensities for presentation purposes. All images were taken and processed by Prof Jordan Raff.

Most importantly, whereas wild-type GFP-Cnn 10D/E could spontaneously form robust protein aggregates in the cytoplasm independently of centrosomes in most *cnn*$^{-/-}$ null mutant embryos (16/20) (**Figure 3.33D;** see **Chapter 2.6.6**), such prominent aggregates were not observed (0/20 embryos) with GFP-Cnn 10D/E protein lacking the CM2 domain (**Figure 3.33C**). Thus, the CM2

domain appears to be essential to allow phospho-mimicking forms of Cnn to assemble into a robust scaffold structure.

The localisation patterns observed in experiments described above strongly suggested that the CM2 domain has two functions: first, it is essential for the efficient centrosomal localisation of Cnn; second, it is required for the efficient Cnn scaffold assembly and/or maintenance (hereafter scaffold stability). Its role in Cnn scaffold stability was further supported by the observation that many small particles of GFP-Cnn T1133A/E were rapidly released from the centrosomes, once the protein was recruited around centrioles (**Figure 3.34**). Due to the severely disrupted centrosomal localisation of the T1133E mutant, FRAP analysis was only carried out on GFP-Cnn T1133A, and it showed that this mutation has significantly increased the rate at which Cnn dissociates from the centrosome (also known as the "OFF" rate). In addition, this defect is MT-dependent since the "OFF" rate recovered after colchicine treatment.

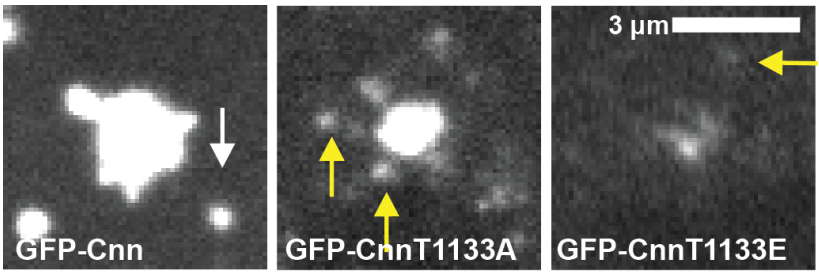


Figure 3.34 Disrupted centrosomal localisation of GFP-Cnn T1133A/E. In particular, the amount of centrosomal Cnn was significantly reduced in the presence of T1133E mutation. While the other mutants were analysed via an RNA injection approach, transgenic fly lines expressing the T1133 point mutations (crossed into a *cnn*$^{-/-}$ null mutant background) have been generated, and images shown above are taken from their embryos. Therefore, an image showing the embryo expressing GFP-Cnn WT is also shown for direct comparison. Centrosomes in embryos expressing GFP-Cnn released large flares (highlighted with a *white* arrow), whereas the centrosomes in embryos expressing GFP-Cnn T1133A/E released large amounts of mini flares (highlighted with *yellow* arrows). Images were taken using a Perkin Elmer ERS spinning disk (ERS software) mounted on a Zeiss Axiovert microscope using a 63X/ 1.4NA oil immersion objective and on Orca ER CCD camera (Hamamatsu Photonics, Japan). Images shown are maximum intensity projections of several z-slices. Images are reproduced with the permission of Dr Paul T. Conduit [Conduit, 2010].

3.7.3 Importance of Zn^{2+} binding *in vivo*

The Zn^{2+} binding mutant GFP-Cnn DCtoAA was assayed for its centrosomal localisation. Unlike the deletion of the entire CM2 domain, this mutant protein showed slightly reduced centrosomal incorporation (**Figure 3.35**). Its PCM expansion was largely unaffected, and Spd-2-RFP was seen to strongly co-localise with the mutant at the centrosome. Due to the time limitation, I could not investigate the centrosomal localisation of the proper Zn^{2+} binding mutant that has both His1082

and Cys1084 residues mutated to Ala. This would be one of the future experiments listed in the next section.

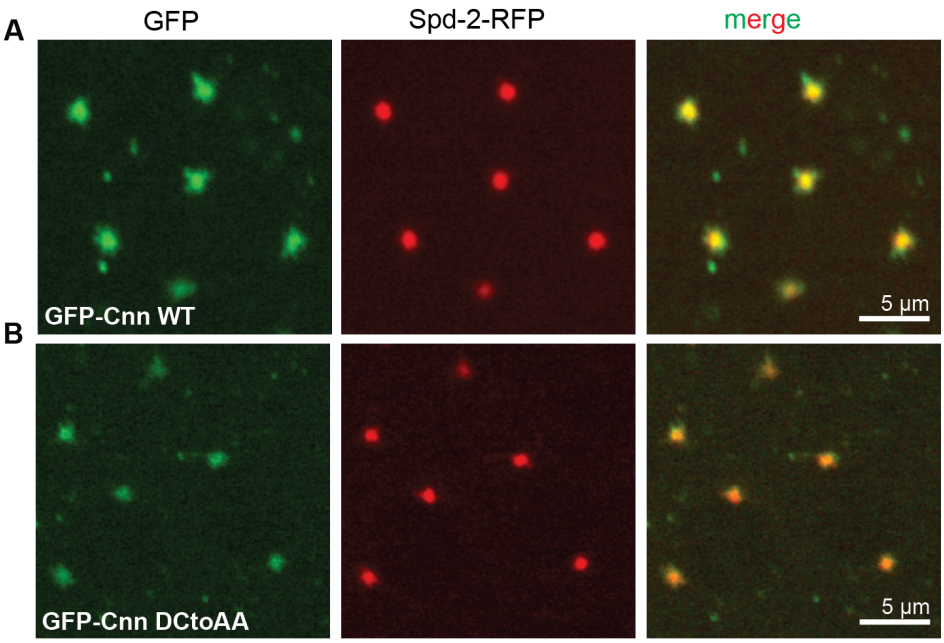


Figure 3.35 RNA injection assay to investigate the effect of Zn^{2+}-binding mutation on centrosomal localisation of Cnn. GFP-Cnn DCtoAA mutant protein showed slightly tighter and weaker centrosomal localisation compared to WT. Spd-2-RFP signal strongly co-localised with GFP-Cnn DCtoAA at the centrosomes. Images were taken using a Perkin Elmer ERS spinning disk (Volocity software) mounted on a Zeiss Axiovert microscope using a 63X/ 1.4NA oil immersion objective and on Orca ER CCD camera (Hamamatsu Photonics, Japan). Images shown are maximum intensity projections of several z-slices. All images were taken and processed by Prof Jordan Raff.

3.7.4 Planned experiments with transgenic *Drosophila*

The results from the RNA injection assay can be affected by parameters such as the amount of RNA injected and the efficiency of RNA transcription in embryos, which are difficult to control and accurately assess. As a result, I wanted to create stable transgenic *Drosophila* lines expressing these mutant proteins under the control of the Ubq promoter to allow better quantification of these phenotypes.

For all the mutations described above, I amplified the Cnn coding region by PCR from the appropriate pRNA_GFPNT-Cnn vector and cloned it into the pDonr vector via Gateway attBP cloning, and then into the pUbq_GFPNT (DEST) vector using attLR cloning (Life Technologies). Note that for the Zn^{2+}-binding mutant only GFP-Cnn HCtoAA/HCtoEE (His1082Cys1084 mutated to Ala/Glu) were cloned for transgenic fly generation. The purified plasmids were sent to the Cambridge University Injection service to generate stable transgenic lines via P-element

mediated transformation [Roberts, 1986]. At the time of writing, these *Drosophila* lines are being generated.

3.8 Perspectives

In this chapter, I have documented the first structure of the CM2 domain solved by X-ray crystallography. Unexpectedly, this motif can bind Zn^{2+} ions *in vitro*. The metal coordination appears to play a key structural role, and Zn^{2+} acts as an atomic bridge to hold together two CM2 monomers in a parallel dimer pair. What is the biological importance of this metal binding activity? My *in vitro* analysis revealed that the Zn^{2+} binding mutation (DCtoAA) strongly disrupted Cnn-CM2 dimerisation in solution. Nevertheless, *in vivo*, the GFP-Cnn DCtoAA protein could still localise to centrosomes, although at reduced levels, and showed only a relatively modest disruption in scaffold assembly and/or stability compared to deletion of the entire CM2 domain. It is clear that Cnn has at least two self-interacting domains, the PReM domain and the CM2 domain, and my analysis suggests that the CM2 domain can interact through both dimer and tetramer formation. At the centrosomes, Cnn protein is highly concentrated, so I suspect that abolishing the Zn^{2+} mediated dimerisation only subtly perturbs scaffold formation because Cnn remains able to self-interact (and presumably also to interact with other PCM proteins) through other interaction interfaces. Moreover, the Zn^{2+} binding mutation I have tested so far only has one of the two Zn binding sites mutated to Ala. It would be worth mutating both His1082 and Cys1084 residues to Ala/Glu (HCtoAA/HCtoEE mutants) and then repeating the centrosomal localisation analysis in transgenic fly lines to confirm the importance of Zn^{2+} binding.

The crystal structure of Cnn-CM2 provides important insights into how Cnn might form a stable scaffold structure at the centrosome. The tetrameric structure has buried a large surface area and is highly stable as suggested by its crystallisation efficiency and the high dissociation energy predicted in PISA analysis. Further analysis of the Cnn-CM2 crystal structure revealed the presence of 3 (out of 4 seen in the structure) invariant residues (Ile1126, Thr1133 and Leu1137) being involved in hydrophobic packing at the tetramerisation interface (**Figure 3.23**). One of these residues was mutated to Ala or Glu (T1133A/E), and it strongly disrupted the Cnn scaffold formation (T1133E mutation had a stronger phenotype than T1133A) *in vivo*. This provides strong support to the idea that the CM2 tetramer plays an important part in scaffold assembly. At the time of writing, similar A and E mutants are being constructed for the other 2 invariant residues

(I1126A/E and L1137A/E). The centrosomal localisation of Cnn will be investigated as in the case of T1133A/E mutant for the other mutations in order to confirm the significance of tetramerisation in Cnn scaffold stability. As mentioned above, there are four invariant residues seen in the crystal structure, which are all involved in self-association. The invariant Leu1094 adopts an inward facing *d* position, and is involved in a hydrophobic interaction for dimerisation. The A and E mutants of this residue are also being constructed, and will be analysed via *in vivo* assays to confirm the importance of CM2 dimerisation interface in Cnn scaffold assembly.

Although Cnn-CM2 tetramer formation appears extremely energetically favorable *in crystallo*, my data suggests that Cnn-CM2 does not easily form tetramers in solution. I have carried out SEC-MALS and 1D DOSY NMR experiments to characterise the oligomerisation state of Cnn-CM2, and could only observe protein species potentially larger than a dimer at very high (≥4 mM) protein concentrations in DOSY NMR experiments. When the measurements were conducted for the T1133E mutant, the protein tumbled faster with the increasing protein concentration. Further experiments are required to repeat these measurements, although data at this stage indicate that the Cnn-CM2 domain may assemble into protein complexes of more than a dimer in solution and this self-association is with very low affinity. At present, it is hard to explain this discrepancy, as my mutational analysis *in vivo* strongly supports the idea that the CM2 tetramer is important for scaffold formation. I suspect, however, that this observation may be important: *in vivo* it is important that the PCM scaffold only forms around the centrioles, so mechanisms must exist to ensure that PCM scaffolding proteins like Cnn do not spontaneously start to assemble into scaffolds in the cytoplasm. Clearly the multisite phosphorylation of the PReM domain could function in this way, ensuring that Cnn only oligomerises once it is phosphorylated at multiple sites by Polo kinase, which is itself highly enriched at centrosomes. Perhaps mechanisms exist to ensure that the CM2 domain also can only efficiently assemble into a tetramer once Cnn has been recruited around the centrioles.

However, it is not clear how dimer and/or tetramer formation might be involved in localising Cnn to centrosomes. The centrosomal localisation role of the CM2 domain in *Drosophila* is consistent with the observation in HeLa cells [Wang et al., 2010]. The authors have demonstrated that a GFP fusion construct covering the CM2 region in CDK5RAP2 could localise to the centrosomes

with high affinity and specificity. My *in vivo* data supports this, as GFP-Cnn with the CM2 deletion was hardly detectable at centrosomes. Moreover, the T1133A mutant cannot form robust Cnn scaffolds, but it was recruited to centrosomes normally as suggested by FRAP analysis. This indicates that CM2 might have a dual role in centrosome localisation and Cnn scaffold assembly. Note that the significantly reduced amount of Cnn localisation at centrosomes in the absence of the CM2 domain might also be a consequence of the defective scaffold assembly and/or maintenance. Unfortunately, due to the extremely weak centrosomal signal of GFP-Cnn ΔCM2, we could not perform FRAP analysis to measure the incorporation rate of this deletion mutant. However, it is worth noting that the Cnn PReM domain mutants did not show drastically reduced centrosome localisation, although these mutants could not efficiently assemble into Cnn scaffold structures.

It might be possible that CM2 also interacts with another centriole or centrosome protein to help localise Cnn to centrosomes. The obvious candidate from vertebrate systems is Pericentrin/PLP. In HeLa cells, the CM2 domain was shown to immunoprecipitate with Pericentrin, and this interaction was abolished in the absence of a calmodulin (CaM)-binding motif that appears to be essential for centrosomal targeting [Wang et al., 2010]. This CaM-recognition sequence is mapped onto Leu1131 to Arg1141 in the *Drosophila* Cnn sequence. How can Pericentrin/PLP fit into the CM2 structure? There are two possible models. First, it is plausible that Pericentrin/PLP interacts with a dimer of CM2. Based on its crystal structure, the binding partner has to fit into the cavity created by the asymmetric dimer. In order to accommodate interactions with the ‘stripe’ of conserved residues, Pericentrin/PLP might have to adopt a specific ‘coiled’ structure. Alternatively, Pericentrin/PLP might bind to the conserved and solvent exposed residues present at the surface of the Cnn-CM2 structure. In my final structure model, the highly conserved C-terminal region (residues 1141 to 1148) is missing due to the lack of electron density. This region includes 2 invariant residues Arg1141 (*e* position) and Asn1143 (*g* position) (**Figure 3.20**). Note that Arg1141 is substituted by a Lys residue in *Tribolium casteneum*, but is otherwise identical in all other species included in my MSA. Both residues are likely facing towards the CC periphery assuming the helical nature continues. Recently, Arg1141 has been shown to interact with PLP for maintaining the scaffold integrity at the outermost region of the Cnn scaffold-like structure [Lerit et al., 2015]. In addition, this residue is also involved in centrocortin interaction as

demonstrated by the *cnn*^{B4} mutation [Kao and Megraw, 2009]. Asn1143 might also be involved in an interaction with PLP, but its exact role needs to be further investigated by the RNA injection assay with this residue mutated to Ala or Glu. However, in flies PLP has a much less severe PCM recruitment phenotype compared to its vertebrate homologs [Martinez-Campos et al., 2004]. The injection of anti-PLP antibody into fly embryos did not block the centrosomal targeting of Cnn suggesting PLP is not essential for loading Cnn to centrosomes [Conduit et al., 2010]. Instead the antibody injection experiments hinted that Spd-2 has a crucial role for Cnn recruitment [Conduit et al., 2010]. I therefore decided to investigate how Spd-2 might help localise Cnn to centrosomes, which is the subject of the next chapter, and whether this might involve an interaction within the Cnn-CM2 domain (see **Chapter 5**).

Chapter 4: Function and structure characterisation of Spd-2

4.0 Introduction

4.0.1 Spindle-defective protein 2 (Spd-2)

As outlined in **Introduction**, SPD-2 was originally identified in *C. elegans* and has a role in both centriole duplication and centrosome maturation [Kemp et al., 2004; Pelletier et al., 2006]. This is also true in humans [Gomez-Ferreria et al., 2007; Zhu et al., 2008], but in flies it seems only involved in PCM assembly [Dix and Raff, 2007; Giansanti et al., 2008]. Several recent studies have started to reveal the molecular details of SPD-2/Cep192 function in centriole duplication. In worms and humans this protein helps recruit PLK4/ZYG-1 to the mother centriole to initiate centriole duplication [Park et al., 2014; Shimanovskaya et al., 2014].

Recently, significant progress has been made towards understanding how Spd-2 contributes to PCM assembly in flies and worms. In flies, the cooperation of Spd-2 and Cnn in PCM assembly has been described [Conduit et al., 2014b]. In embryos, both proteins assemble from the inside out and so flux outwards through the PCM [Conduit et al., 2010]. Unlike Cnn, the inside-out movement of Spd-2 is not MT-dependent and its initial recruitment to centrioles is likely via an interaction with the centriolar protein Asl [Conduit et al., 2010; Conduit et al., 2014b; Conduit and Raff, 2015]. Both Cnn and Spd-2 help recruit other PCM components, and in the absence of both proteins mitotic centrosome maturation is abolished [Conduit et al., 2014b]. Interestingly, Spd-2 helps recruit Cnn to the centrosome and Cnn then helps maintain Spd-2, potentially forming a positive feedback loop that drives the expansion of the mitotic PCM [Conduit et al., 2014b]. In *C. elegans*, SPD-2 also cooperates with a protein called SPD-5 (a likely functional homolog of Cnn, though there is no sequence similarity) [Woodruff et al., 2015]. Like Cnn, SPD-5 can also form a scaffold and this process is enhanced by PLK-1 phosphorylation and by SPD-2. The cell biology

work has highlighted the importance of Spd-2 in establishing PCM, but little is understood about the molecular details of how this protein might promote PCM scaffold assembly.

4.0.2 Aims

- Identify putative functional domains of Spd-2.
- Investigate the biological importance of identified domains, in collaboration with Dr Paul T. Conduit (a postdoctoral fellow in the lab).
- Biophysical and structural characterisation of the functional domains, in order to understand the *in vivo* mechanism.

4.0.3 Key findings

Full length Spd-2 can be divided into two halves, with the N-terminal part of 696 residues required for centriolar recruitment and the C-terminal part of 450 residues required for the proper formation of the Spd-2/Cnn centrosomal scaffold. Scaffold assembly depends on two **A**sp, **S**PD-2, **H**ydin (ASH) domains [Ponting, 2006] that are predicted to have an Ig-like fold via bioinformatics and together largely encompass the conserved “SPD-2 domain” [Pelletier et al., 2006]. *Drosophila* Spd-2 also contains a third ASH domain at the extreme C-terminus, which shares similar fold topology with the other two domains but appears to have a role in limiting, rather than promoting, scaffold formation. The domain structure was confirmed via an NMR study (ASH1) and X-ray crystallography (ASH3).

4.1 Bioinformatic analysis of Spd-2

4.1.1 Sequence alignments reveal conserved regions within Spd-2

In worms, flies and vertebrates, Spd-2 proteins have a shared role in PCM assembly [Dix and Raff, 2007; Pelletier et al., 2006; Zhu et al., 2008]. To define conserved domains I carried out a multiple sequence alignment (MSA) for Spd-2/Cep192 proteins from 35 species using ClustalX [Larkin et al., 2007]. In human cells, 4 isoforms of Cep192 generated by alternative splicing were identified [Sonnen et al., 2013], one of which has an extra 603 amino acids at the N-terminus (Isoform 3) that are important for centriole duplication in humans [Kim et al., 2013; Sonnen et al., 2013]. This domain is missing in flies so Isoform 1, the canonical sequence defined in Uniprot [UniProt, 2015], was included in my alignment. MSA showed some enriched conservation in a region of 221 residues from Leu707 to Arg928 in DmSpd-2 (26.3% sequence identity/40.8% similarity between *D. melanogaster* and *H. sapiens* and 19.9%/39.8% between *D. melanogaster* and *C. elegans*). This is consistent with the results from a previous MSA analysis that identified the *Drosophila* and Human homologs of *C. elegans* SPD-2 based on a conserved "SPD-2-domain" (residues 707 to 928 in *Drosophila* Spd-2), [Pelletier et al., 2004]. A region of the MSA covering the SPD-2 domain is shown in **Appendix Figure 8.5**.

4.1.2 Identification of conserved domain folds

I carried out further bioinformatic analysis using COILs [Lupas et al., 1991] and PSIPRED [Jones, 1999] to search for structural elements that might be evolutionarily conserved. The results are summarised in **Figure 4.1**. Based on COILs prediction, Spd-2 and Cep192 are not highly coiled-coiled proteins (**Figure 4.1A**). In fact, no such domain is identified in HsCep192 Isoform 1. In CeSPD-2 and DmSpd-2, a single coiled-coil domain is predicted with high confidence in the N-terminal region. Secondary structural prediction by PSIPRED [Jones, 1999] indicates that the C-terminal region of DmSpd-2 (~460 residues), which includes the highly conserved SPD-2 domain, is rich in β strands (not shown). This high globular content is also observed in the other 2 homologs. Chris Ponting carried out PSI-BLAST search using the N-terminal 200 residues of human ASPM protein and identified an ASH domain that is conserved in **A**SPM, **S**PD-2 and

Hydin proteins [Ponting, 2006]. The ASH domain comprises eight β strands with a highly conserved Asn residue. According to this analysis, CeSPD-2 contains a single ASH domain (residues 482 to 574), whereas HsCep192 contains three ASH domains (ASH-A: 781-884; ASH-B: 1174-1284; ASH-C: 1851-1938) (**Figure 4.1A**). Interestingly, one of the ASH domains found in *C. elegans* and *H. sapiens* (ASH-B) is part of the highly conserved SPD-2 domain. I manually checked through the predicted secondary structural elements for DmSpd-2, and noticed that its C-terminal region (Spd-$2^{685-1146}$) seemed to contain similar ASH domains. To help with domain delineation, I further analysed the C-terminal region of Spd-2 using the Fold and Function Assignment System (FFAS) server [Jaroszewski et al., 2011]. This uses a psi-BLAST approach to find domains with similar folds in the PDB database. The majority of top hits were from major sperm protein (MSP), Hydin, vesicle-associated membrane protein (VAP) and that all have an immunoglobulin (Ig)-like fold comprising a seven-stranded β sandwich. Indeed, the ASH domain is found to be distantly related to MSP domains [Ponting, 2006]. Based on the models generated by structural homology (**Figure 4.1B,C**), ASH1 and ASH3 were identified in DmSpd-2: one comprised the N-terminal region of the SPD-2 domain (denoted as ASH1: residues 697 to 805), which was also identified in the human (ASH-B) and worm Spd-2 homologues (**Figure 4.1A**); the other was located at the extreme C-terminus (ASH3: residues 1047 to 1146), which was also identified in HsCep192 (ASH-C), but not CeSPD-2 (**Figure 4.2A**).

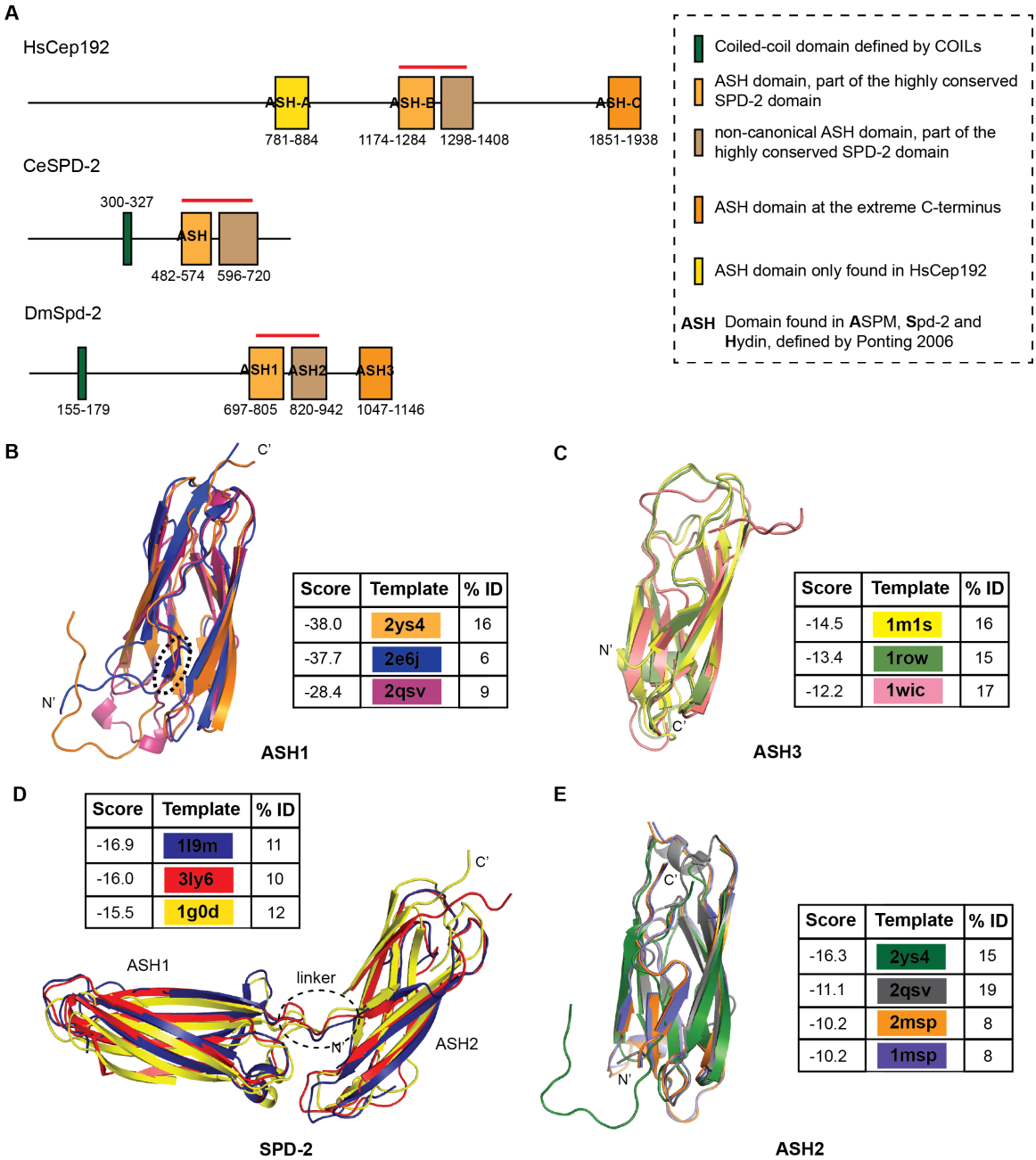


Figure 4.1 Summary of conserved domains in Spd-2/Cep192 defined by bioinformatic analysis. (**A**) Schematics showing topologies of Spd-2 in *D. melanogaster*, *H. sapiens* and *C. elegans*. The highly conserved SPD-2 domain (Spd-$2^{707\text{-}920}$) is highlighted by a *red* bar. The ASH domains highlighted for CeSPD-2 and HsCep192 are based on analysis carried out by Ponting 2006, whereas the ASH domains in DmSpd-2 are based on my bioinformatic analysis carried out in this project. (**B-E**) Homology models for ASH1 (**B**), ASH3 (**C**), SPD-2 (**D**) and ASH2 (**E**) domains from structures with similar folds predicted by FFAS server. The confidence of prediction, indicated by FFAS score, and the % identity (% ID) each model is based on are listed in a table next to the homology models. Lower FFAS scores indicate higher confidence of the prediction, and predictions with scores of lower than -9.5 contain less than 3% of false positives [Jaroszewski et al., 2005]. Models were aligned to help delineate the domain boundaries in DmSpd-2. In both ASH1 (**B**) and ASH3 (**C**) domains, 1 extra β strand was included (at N-terminus) comparing to the matching ASH domains defined in HsCep192 and CeSPD-2 by Ponting 2006. For example, in the predicted structure of ASH1, a short β strand of residues 700 to 704 was observed (in dash) in one of the models, which seemed to close up the beta sandwich structure. Therefore, the ASH1 domain in DmSpd-2 was determined as between residues 697 and 805. In the modelled structure of SPD-2 domain (**D**), a short linker was observed between the 2 subunits (ASH1 and the remaining of SPD-2 domain from residues 820 to 920). The modelled ASH2 structure (**E**) ended with Glu920. However, sequence conservation was still observed beyond this residue (**Appendix Figure 8.5**), and PSIPRED tool predicted a short helix between residues 920 and 942 that can also be found in other homologs. Therefore the ASH2 domain includes residues from 820 to 942, and the equivalent ASH2 domains are also highlighted in other homologs based on sequence conservation and PSIPRED prediction (**A**). Structures used for homology modelling are: N-terminal PapD-like domain of HYDIN protein from human (PDB entry: 2YS4); C-terminal PapD-like domain of HYDIN protein from human (PDB entry: 2E6J); *Porphyromonas gingivalis* W83 (PDB entry: 2QSV); *C. elegans* MSP protein family member WR4 (PDB entry: 1M1S); *C. elegans* MSP protein-like family member SSP-19 (PDB entry: 1ROW); MSP domain of RIKEN cDNA 6030434E15 (PDB entry: 1WIC); human transglutaminase 3 enzyme (PBD entry: 1L9M); human transglutaminase 2 (PDB entry: 3LY6); red sea bream transglutaminase (PDB entry: 1GOD); Major sperm protein β Isoform (PDB entry: 2MSP); Major sperm protein α Isoform (PDB entry: 1MSP).

Interestingly, some of the structurally homologous domains found by FFAS covered the entire SPD-2 domain (Spd-2$^{707-928}$). Aligning the modelled structures in Pymol, V-shaped subunits with enriched β strand content were observed (**Figure 4.1D**), in which the first subunit comprised the ASH1 domain and the second subunit contained the remaining SPD-2 domain (residues 820 to 942), which also resembled an Ig-like fold, and hence I termed this ASH2. The identification of this pseudo-ASH domain was further confirmed by performing an FFAS search with the sequence of Spd-2$^{800-1046}$ (**Figure 4.1E**). Unlike ASH1 and ASH3, ASH2 was predicted by PSIPRED to have a helix assembled towards the C-terminus of this domain. Interestingly, when the secondary structural predictions were compared between *D. melanogaster*, *C. elegans* and *H. sapiens*, the SPD-2 domain can also be viewed as comprising ASH1 and ASH2 like units (**Figure 4.1D**).

Overall, these results suggest that the SPD-2 domain is conserved in fly, human and worm in terms of protein sequence and fold topologies, and that Spd-2 in *Drosophila* and Cep192 in human share another domain of similar fold at the extreme C-terminus. In addition, HsCep192 seems to contain more Ig-like domains as judged by PSIPRED and FFAS analysis (**Figure 4.1A**). I decided to investigate the potential importance of the fly ASH domains both *in vivo* and *in vitro*.

4.2 Analysis of Spd-2 scaffolding activity *in vivo*

4.2.1 Construct design and RNA preparation

As mentioned in **Chapter 2**, the *Drosophila* embryo provides an ideal system for monitoring protein localisation at centrosomes. A GFP-fusion protein encoded mRNA can be injected into the early embryos and will be translated in the cytoplasm; the localisation of the GFP-tagged protein can then be analysed using fluorescence microscopy. I designed various mutations within Spd-2, as summarised in **Figure 4.2**. First, the full-length protein was truncated to contain only its N-terminal (residues 1 to 696) or C-terminal (residues 697 to 1146) region. Second, the conserved domains delineated using bioinformatics, ASH1, ASH2 and ASH3, were deleted individually to assess the effect on Spd-2 centrosomal localisation. I also deleted the SPD-2 domain (comprising ASH1 and ASH2). Ideally, this analysis would be performed in embryos lacking endogenous Spd-2. However, embryos laid by *spd-2*$^{-/-}$ mutant mothers arrest early in development due to a failure in pronuclear fusion [Dix and Raff, 2007]. Therefore, all the injections were carried out in WT background (and so in the presence of endogenous, but unlabelled, Spd-2). I generated all mutant constructs and synthesised and purified all mutant mRNAs. Dr Paul T. Conduit then carried out all the fly husbandry, RNA injections, microscopy work and data quantification for the following experiments.

Truncated constructs (Spd-2$^{1\text{-}696/697\text{-}1146}$) were PCR amplified from cDNA encoding the full-length Spd-2 before transferring to the pRNA_GFPCT vector (**Figure 4.3**) by In-Fusion cloning (Clontech). For internal deletions, site-directed mutagenesis (Agilent) was conducted on the pRNA_GFPCT-Spd-2 WT vector. Since GFP is C-terminally fused with Spd-2 protein, no stop codon was inserted at the end of the Spd-2 ORF. mRNA was generated *in vitro* using T3 transcription (Thermo Scientific) and purified by spin column chromatography (Qiagen). The concentration was adjusted to 2 mg/ml for all the constructs.

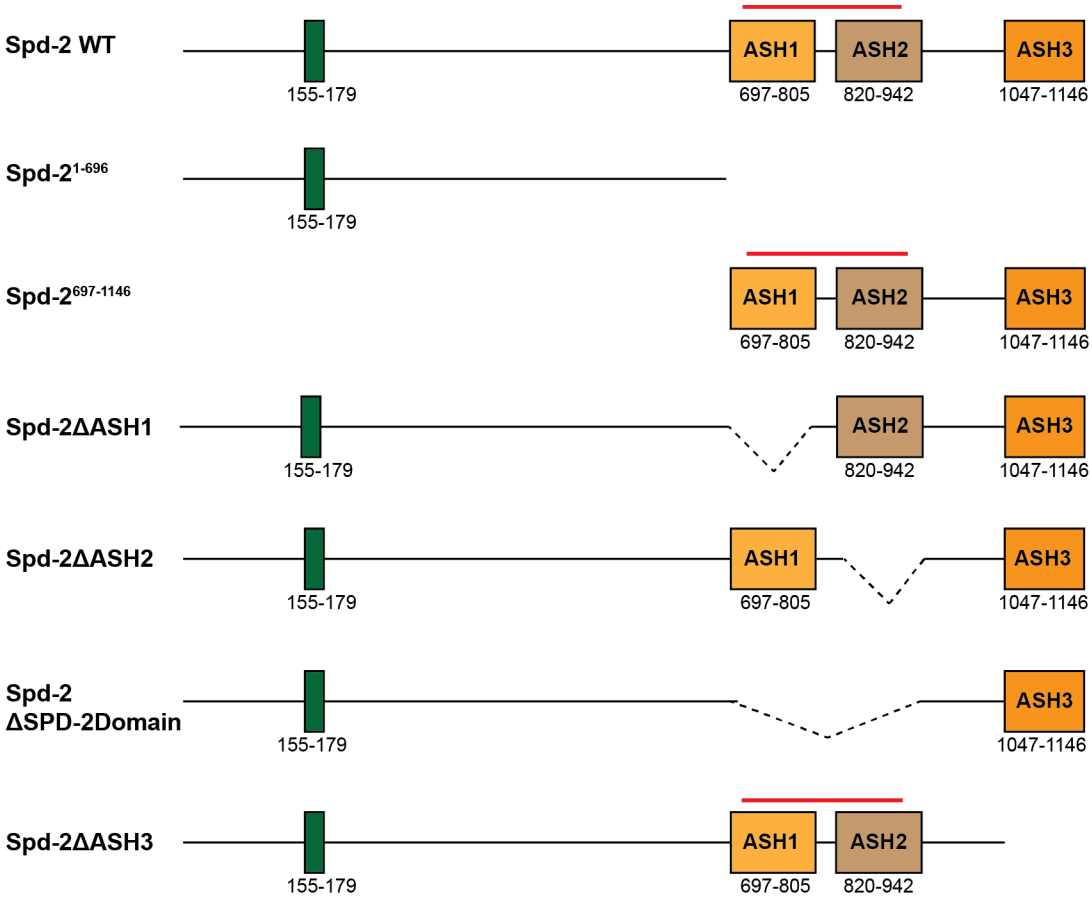


Figure 4.2 Spd-2 construct map showing the mutants analysed in this project. The *red* bar highlights the location of the SPD-2 domain.

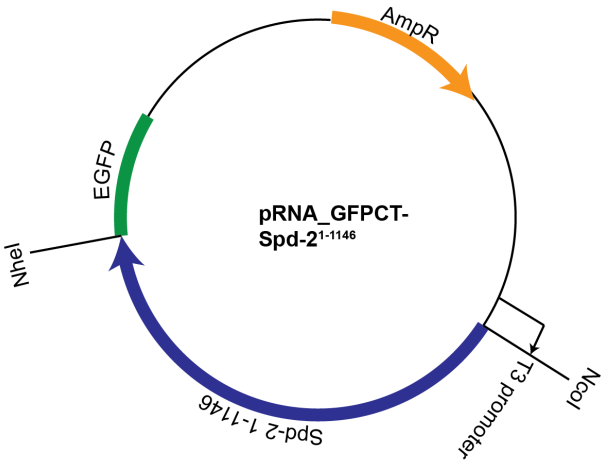


Figure 4.3 Plasmid map of pRNA_GFPCT-Spd-2. The PCR amplified construct was cloned into the NcoI/NheI linearised vector by In-Fusion reaction. Diagram modified from [Cottee, 2014].

4.2.2 Full-length Spd-2 localises strongly to centrosomes

The centrosomal localisation of the wild-type Spd-2 protein was first analysed. Spd-2-GFP was visible as bright foci at centrosomes (**Figure 4.4A**). The signal was broadly distributed rather than a tight spot. This is expected as Spd-2 has a PCM-associated fraction that assembles into scaffold-like structures, as shown by super-resolution 3D SIM [Conduit et al., 2014b; Mennella et al., 2012]. The centrosomal localisation of Spd-2 WT was further confirmed with radial profiling, which plots the average fluorescence intensity across centrosomes (detailed protocol in **Appendix 8.1.13**). As displayed in **Figure 4.4B**, Spd-2-GFP showed a broader spatial distribution at the centrosome compared to Asl-GFP, a centriolar protein. In addition, the dynamics of wild-type Spd-2 was investigated by fluorescence recovery after photobleach (FRAP) analysis, which demonstrated an initial recovery of Spd-2-GFP at centrioles and the fluorescence then slowly moved outwards into the PCM (**Figure 4.4C**) [Conduit et al., 2010].

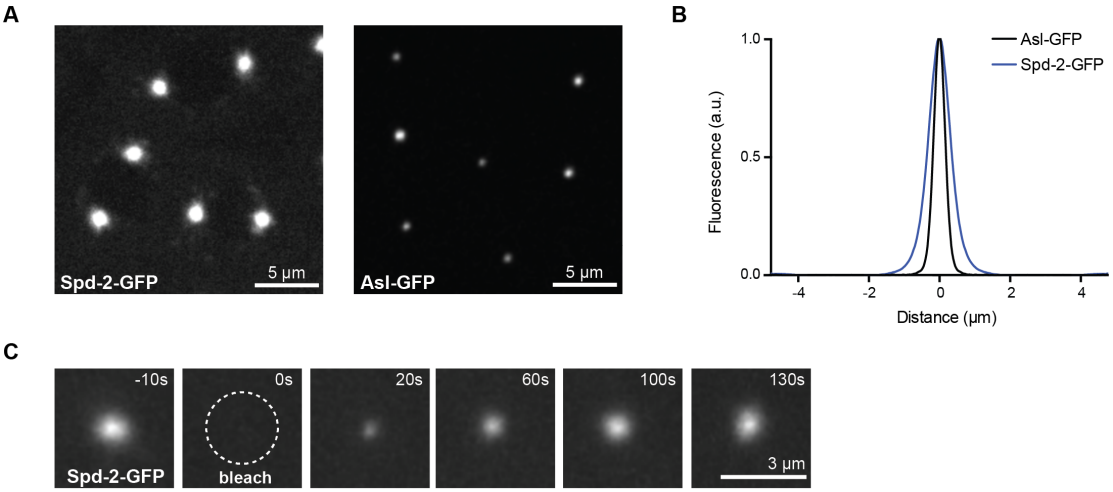


Figure 4.4 Centrosomal localisation of Spd2-GFP in WT embryos. (**A**) Spd-2-GFP exhibited a broader localisation at the centrosome comparing to Asl, a centriolar protein. Images were taken using a Perkin Elmer ERS spinning disk (Volocity software) mounted on a Zeiss Axiovert microscope using a 63X/ 1.4NA oil immersion objective and on Orca ER CCD camera (Hamamatsu Photonics, Japan). Images shown are maximum intensity projections of several z-slices. (**B**) The normalised average fluorescence intensity was measured across the centrosome for both Spd-2-GFP and Asl-GFP. The latter showed a tighter centriolar association. Measurements were carried out on 10 centrosomes from three embryos. (**C**) FRAP analysis demonstrated that Spd-2-GFP fluorescence, after photobleaching, initially recovered at the centre (20s) and then gradually moved outwards (60-130s). GFP signal was bleached using a focussed 440 nm laser, during S phase of cycle 11 or 12. 0.5 µm thick confocal sections through the center of a selected centrosome were collected. All images were taken and processed by Dr Paul T. Conduit, who carried out the subsequent image analysis.

In order to effectively compare the injection results, a centrosomal profile was calculated for each construct discussed in this chapter and will be displayed next to the confocal image showing the centrosomal localisation of each construct. The image and profile for the wild-type construct were also included in each analysis for comparison.

4.2.3 The N-terminal region of Spd-2 is required for centriolar targeting, and the C-terminal region is needed for PCM localisation

The N-terminal Spd-2^{1-696}-GFP protein was tightly focused at centrosomes; this was in contrast to the much more spread out localisation of the wild-type protein (**Figure 4.5A,B**). This was confirmed with radial profiling (**Figure 4.5D**), although the profile was not as tight as that observed for Asl-GFP. A FRAP analysis revealed that Spd-2^{1-696}-GFP fluorescence rapidly recovered throughout the central region and did not expand outwards into the PCM (**Figure 4.5E**). These observations suggest that the C-terminal region of Spd-2 (residues 697 to 1146) is not required for the centriolar targeting of Spd-2, but is required for the efficient spreading of Spd-2 molecules into the PCM.

The C-terminal Spd-$2^{697-1146}$-GFP protein was loosely localised at centrosomes and in a significantly reduced amount (**Figure 4.5A,C**). Fluorescence was mainly detected in the peripheral region, which was confirmed with radial profiling (**Figure 4.5D**). A FRAP analysis demonstrated that Spd-$2^{697-1146}$-GFP recovered at a much slower rate throughout the PCM and the initial recovery was not observed in the central region, although the extremely weak and diffused fluorescence made it difficult to visualise (**Figure 4.5F**). These observations suggest that the N-terminal region of Spd-2 (residues 1 to 696) is indispensible for the centriolar recruitment of Spd-2. Despite the fact that a small amount of Spd-2 molecules were still detected at the PCM, they could have been recruited by unspecific binding to endogenous Spd-2 protein or other PCM components.

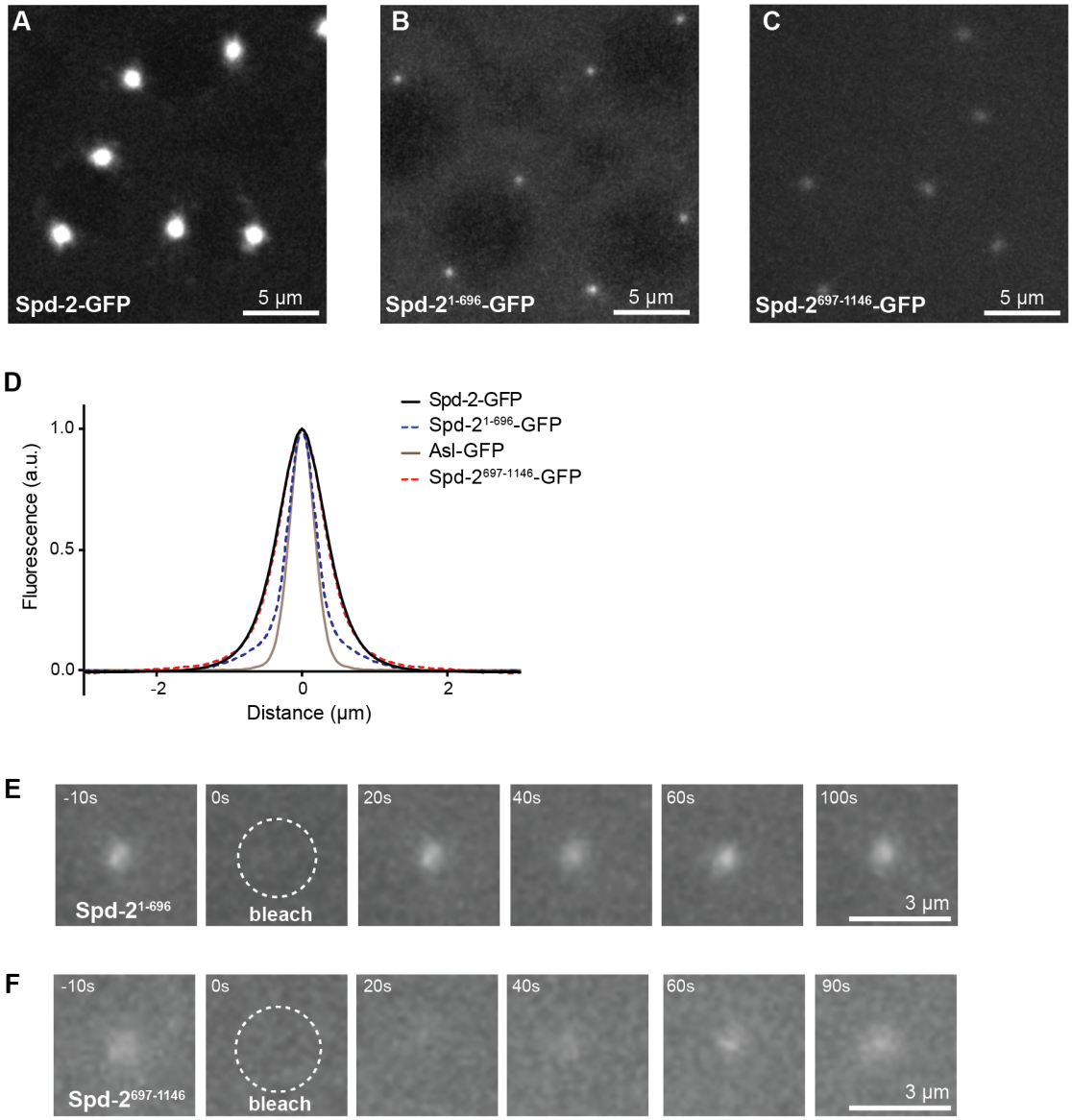


Figure 4.5 Centrosomal localisation of Spd-2 N- and C-terminal regions. (**A**) Image showing the localisation of full length Spd-2 WT construct, as a positive control. (**B**) In the absence of C-terminal region, Spd2$^{1-696}$-GFP showed a much tighter centrosomal association that was further confirmed via radial profiling (**D**). (**C**) In the absence of N-terminal region, Spd2$^{697-1146}$-GFP was visible as weak, diffused centrosomal foci. (**D**) The normalised average fluorescence intensity was measured across the centrosome for various Spd-2 mutant constructs. Measurements were carried out on 10 centrosomes from three embryos. (**E,F**) FRAP analysis of Spd-2 mutant constructs. GFP signals were bleached using a focussed 440 nm laser, during S phase of cycle 11 or 12. 0.5 µm thick confocal sections through the center of a selected centrosome were collected. After photobleaching, Spd2$^{1-696}$-GFP fluorescence rapidly recovered at the centre but little expansion in signal was detected suggesting the mutant failed to spread outwards into the PCM. In contrast, the initial fluorescence recovery from Spd-2$^{697-1146}$-GFP was observed much later (40-60s) than WT construct and the N-terminal fragment (20s). The recovery did not first appear as a tight, brighter spot in the centre as for the centriolar recruitment. Images were taken using a Perkin Elmer ERS spinning disk (Volocity software) mounted on a Zeiss Axiovert microscope using a 63X/ 1.4NA oil immersion objective and on Orca ER CCD camera (Hamamatsu Photonics, Japan). Images shown are maximum intensity projections of several z-slices. All images were taken and processed by Dr Paul T. Conduit, who carried out the subsequent image analysis.

4.2.4 Spd-2 scaffold assembly is facilitated by the ASH1 and ASH2 domains

The above experiments revealed that the C-terminal region of Spd-2, which contains all 3 ASH domains I identified bioinformatically, is required for the efficient formation of Spd-2 scaffolds. I therefore tested whether the individual ASH domains might contribute to Spd-2 scaffold formation. GFP-fusion proteins lacking either the ASH1 (Spd-2ΔASH1-GFP) or ASH2 (Spd-2ΔASH2-GFP) domains very weakly localised as a tight dot to the centrosomes (**Figure 4.6A,B**). Radial profiling confirmed that these proteins appeared to be recruited to the centrioles, but could no longer efficiently incorporate into a scaffold structure that spread away from the centrioles (and it is worth stressing that these experiments were conducted in the presence of unlabelled wild-type Spd-2, that is presumably still recruited to centrioles and is still forming such a scaffold-hence explaining why the centrosomes remain well organised and functional in these embryos) (**Figure 4.6D**). This profiling also revealed that the deletion of ASH2 had a marginally stronger phenotype than the deletion of ASH1, and that deleting both domains at the same time (so deleting the SPD-2 domain—Spd-2ΔSPD-2Domain-GFP) had an even stronger phenotype, suggesting that the domains may cooperate in PCM scaffold assembly (**Figure 4.6C,D**). Due to the very weak centrosomal signal for the ASH domain deletions, no FRAP analysis could be reliably conducted.

In contrast, the deletion of the ASH3 domain (Spd-2ΔASH3-GFP) had little effect on the ability of Spd-2 to incorporate into the PCM scaffold, and, if anything its distribution was slightly broader than the wild-type protein, perhaps suggesting that ASH3 may play an inhibitory role in promoting scaffold formation (**Figure 4.6E,F**). In support of this possibility, the Spd-2ΔSPD-2Domain-GFP protein (that deletes ASH1 and ASH2 domains, but still contains ASH3 domain) had a weaker PCM scaffold incorporation profile than the NT domain of the protein (that deletes ASH1, ASH2 and ASH3 domain altogether) (**Figure 4.6D**).

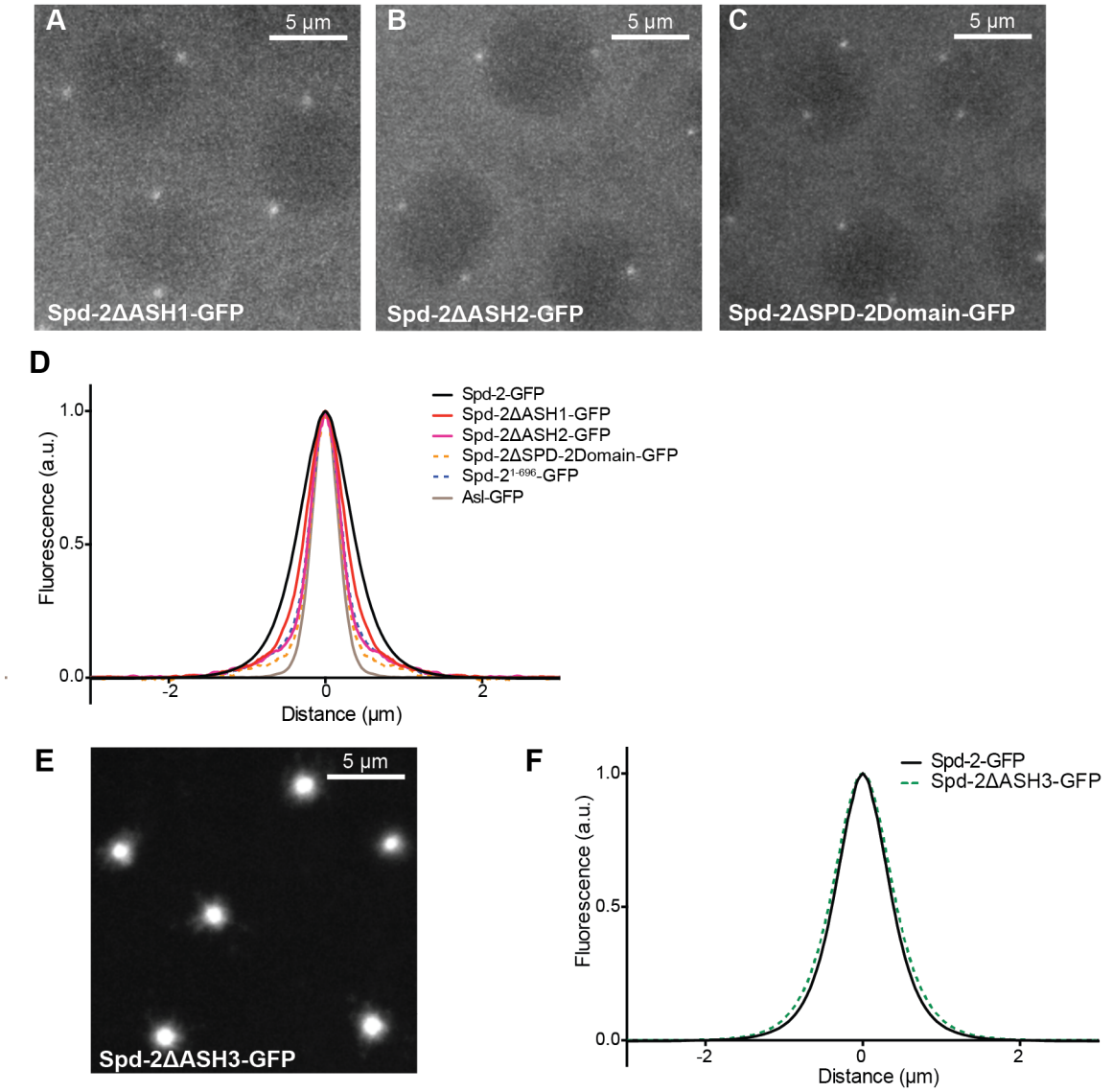


Figure 4.6 ASH domains are essential for the Spd-2 scaffold formation. (**A,B**) Spd-2-GFP showed defects in scaffold assembly in the absence of highly conserved Ig-like domains, ASH1 and ASH2. Spd-2ΔASH2-GFP exhibited a narrower spatial distribution comparing to Spd-2ΔASH1-GFP as suggested by the radial profiling data (**D**). More severe disruption was obtained when both domains were removed (**C**). (**E**) Removal of the ASH3 domain had an opposite effect, despite the influence was minor, on Spd-2 scaffold formation. Spd-2ΔASH3-GFP expanded slightly further into the peripheral region of the PCM as confirmed with radial profiling (**F**). Interestingly, as shown in (**D**), Spd2$^{1-696}$-GFP expanded slightly further into the PCM region comparing to Spd2ΔSPD-2Domain-GFP protein. This further supports the idea that the ASH3 domain might limit the amount of Spd-2 localised to the PCM. Images were taken using a Perkin Elmer ERS spinning disk (Volocity software) mounted on a Zeiss Axiovert microscope using a 63X/ 1.4NA oil immersion objective and on Orca ER CCD camera (Hamamatsu Photonics, Japan). Images shown are maximum intensity projections of several z-slices. Measurements of the average fluorescence intensity across the centrosome were carried out on 10 centrosomes from three embryos. All images were taken and processed by Dr Paul T. Conduit, who carried out the subsequent image analysis.

4.2.5 Attempts to identify phosphorylation sites within Spd-2

Prior to the start of my project, Dr Janina Baumbach, a previous PhD student in the lab, had carried out a western blot analysis on the cytoplasm- and centrosome-enriched Spd-2 fractions [Baumbach, 2014]. The result suggested that at the centrosome a mixture of phosphorylated and non-phosphorylated Spd-2 molecules was observed, whereas cytosolic Spd-2 was not phosphorylated. Polo kinase is highly concentrated at the centrosome and its role has been implicated in PCM assembly. As documented in **Chapter 2**, Cnn is phosphorylated by Polo kinase at multiple sites, which subsequently promotes the assembly of Cnn scaffolds at the centrosome. I therefore decided to search for potential Polo phosphorylation sites within Spd-2, which might provide insights into the regulatory mechanism underlying its scaffold assembly. Note that Polo kinase binding is often primed by CDK1 phosphorylation, and hence I also searched for potential CDK1 phosphorylation sites within Spd-2.

Dr Janina Baumbach had previously carried out a tandem mass spectrometry analysis (MS/MS) for Spd-2 and identified several sites specifically phosphorylated at the centrosome [Baumbach, 2014]. When these sites were mapped onto the full-length protein, I noticed they were concentrated in two regions, which I termed Phosphorylation Region 1 (PR1: residues 513 to 562) and Phosphorylation Region 2 (PR2; residues 585 to 635) that lie upstream of the conserved ASH1 domain. Due to the poor sequence conservation in MSA of Spd-2 homologs, I performed another alignment with sequences only from *Drosophila* species. The phosphorylation sites identified by MS/MS are highly conserved in the *Drosophila* sequences, and both regions also contain other conserved Ser/Thr sites, several of which closely match the CDK1 or PLK1 consensus sequence (**Figure 4.7A-C**).

In addition, the MSA highlighted another region that contains 10 highly conserved Ser/Thr sites, the majority of which at least partially match the PLK1 consensus motif (**Figure 4.7D**). This region (PR3: residues 994 to 1044) is located between the ASH2 and ASH3 domains.

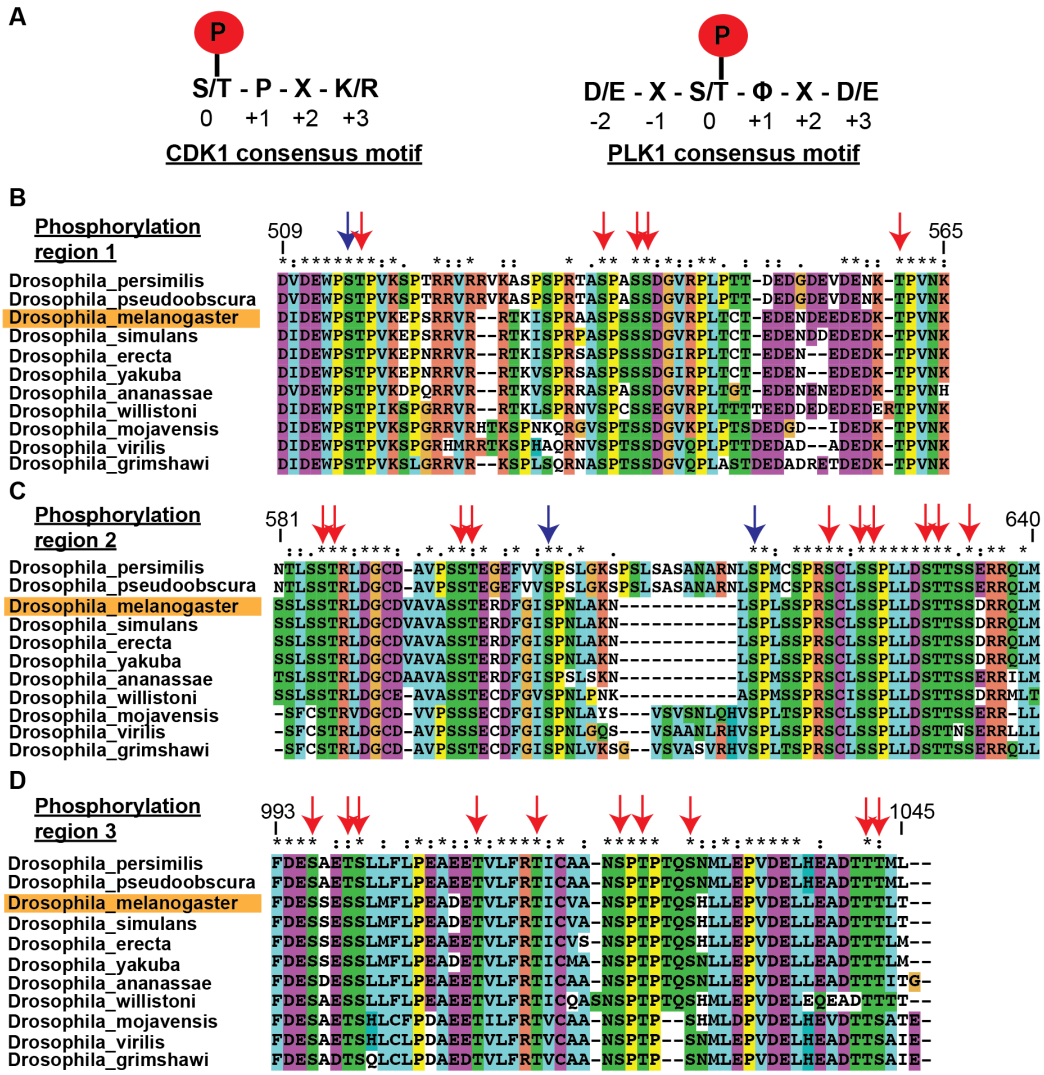


Figure 4.7 Identification of potential CDK1 or Polo phosphorylation sites within Cnn. (**A**) Diagram showing the CDK1 or PLK1 consensus motif. (**B,C**) MSA of Spd-2 from *Drosophila* species only showing the potential phosphorylation regions, PR1 (**B**) and PR2 (**C**). The MS/MS identified sites are highlighted with *blue* arrows. There are other highly conserved Ser/Thr sites in both regions, highlighted with *red* arrows, some of which at least partially resemble CDK1 or PLK1 phosphorylation sequence. (**D**) MSA of Spd-2 from *Drosophila* species only showing the PR3 region. The 10 highly conserved Ser/Thr sites are highlighted. Residue numbers for *D. melanogaster* are indicated. Invariant residues are highlighted with asterisks, conserved residues with colons and semi-conserved residues with full stops. Residues are coloured according to the ClustalX colour scheme.

I made mutant constructs that individually deleted PR1, PR2 and PR3 regions. RNA was obtained from each construct by *in vitro* transcription (Thermo Scientific) and was injected into WT embryos at 2 mg/ml. Spd-2ΔPR1-GFP and Spd-2ΔPR2-GFP both showed a strong centrosomal localisation, similar to WT (**Figure 4.8A**). However, in contrast, Spd-2ΔPR3-GFP localised weakly to centrosomes as a tight dot, indicating that the PR3 deletion affected the ability of Spd-2 to assemble into a scaffold (**Figure 4.8B**). In order to assess if this phenotype was due to the impaired phosphorylation, I further cloned a GFP-fusion protein in which all 10 conserved Ser/Thr sites (10A) in the PR3 domain were mutated to Ala. Disappointingly, the mutant protein showed a

normal centrosomal localisation, suggesting the Spd-2 localisation defect seen with ΔPR3 mutant was unlikely to be phosphorylation-dependent. This region was close to the ASH2 domain, and hence its deletion might have affected the domain fold or disrupted the specific packing between the C-terminal ASH domains. These results are still preliminary, and the experiments need to be repeated and quantified.

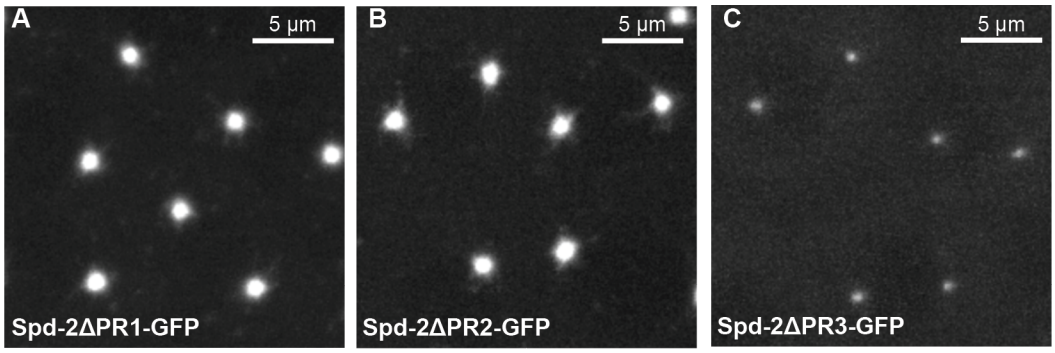


Figure 4.8 Localisation of Spd-2-GFP mutant constructs that carry internal deletion of potential phosphorylation region. (**A,B**) The centrosomal localisation of Spd-2-GFP was not affected in the absence of the two identified potential phosphorylation regions. (**C**) Deleting the potential phosphorylation region (PR3) that is located between the ASH2 and ASH3 domains had an influence on Spd-2-GFP expansion into the PCM. However, this defect might not be due to an impaired phosphorylation as the phospho-mutant (10A) showed normal centrosomal localisation. Images were taken using a Perkin Elmer ERS spinning disk (Volocity software) mounted on a Zeiss Axiovert microscope using a 63X/ 1.4NA oil immersion objective and on Orca ER CCD camera (Hamamatsu Photonics, Japan). Images shown are maximum intensity projections of several z-slices. All images were taken and processed by Dr Paul T. Conduit.

4.2.6 Construct design and cloning for transgenic *Drosophila* lines

As mentioned in **Section 4.2.1**, the RNA-injection experiments described above had to be conducted in a wild-type background because the embryos lacking Spd-2 cannot develop. In order to investigate the function of the ASH domains without any effect from the endogenous Spd-2, I wanted to generate transgenic *Drosophila* lines expressing deleted versions of Spd-2 and then cross them into the *spd-2* mutant background. I therefore cloned my mutant constructs into a pDonr vector via attBP cloning and then in to the pUbq_GFPCT destination vector by attLR cloning. Although Spd-2 10A-GFP phospho-mutant did not show obvious assembly/localisation defect at the centrosome, the presence of endogenous Spd-2 might facilitate its localisation, so I also cloned the Spd-2 10A mutant into pUbq_GFPCT vector. The purified plasmids were sent to Cambridge University Injection service for generation of transgenic lines via p-element mediated transformation. At the time of thesis writing, Dr Alan Wainman, a post doc in the lab, has been

crossing the transgenic lines into a *spd-2* mutant background and he will carry out further investigations on these flies.

4.2.7 Perspectives on *in vivo* experiments

The role of Spd-2 in PCM maturation has long been realised in flies, vertebrates and worms. Using an RNA injection approach I have shown that the conserved ASH domains are involved in the assembly of Spd-2 scaffolds at centrosomes. The ASH1 and ASH2 domains, which can also be found in other homologs, cooperatively facilitated the formation of Spd-2 scaffold-like structures. Deletion of individual domains led to severe scaffold assembly defects that were worse when both domains were absent. Interestingly, the ASH3 domain, which is found in human Cep192 and *Drosophila* Spd-2, may actually have a role in limiting the size of PCM expansion despite its predicted similar structural fold with ASH1.

When scaffold assembly was severely disrupted, the mutant constructs also demonstrated a very weak localisation at centrosomes, suggesting that ASH1 and ASH2 might be required for the centriolar recruitment of Spd-2. However, FRAP analysis on the N-terminal fragment (Spd-2^{1-696}-GFP) showed that protein recruitment was relatively normal. Thus, it seems more likely that the reduction in centrosomal fluorescence of proteins lacking ASH1 and/or ASH2 is due to a failure to properly incorporate these proteins into the PCM scaffold, rather than a failure in initial recruitment. Interestingly, however, a FRAP analysis of the C-terminal fragment (Spd-$2^{697-1146}$-GFP) indicated a slow fluorescence recovery throughout the PCM. This is surprising, as if the N-terminal region is recruiting Spd-2 to centrioles one might expect the C-terminal fragment to show no centrosomal localisation (as Spd-2 is normally only recruited into the PCM around the centrioles [Conduit et al., 2010; Conduit and Raff, 2015]). Perhaps the mutant construct can bind to the endogenous Spd-2 protein in the PCM, or to other PCM proteins in a way that the wild-type protein normally cannot. The conclusion that the Spd-2 N-terminal fragment (Spd-2^{1-696}) is necessary for centriolar recruitment is in good agreement with the strong protein interaction found between Spd-2^{1-380} and Asl$^{766-1148}$ in yeast-two-hybrid assays [Conduit et al., 2014b].

How might the Ig-like ASH domains facilitate PCM scaffold formation? These domains have been found in many proteins and implicated in various functions, including protein interaction, filament formation and extracellular spacers. For instance, MSP monomers, that are strong hits in the FFAS search with Spd-2$^{697-1146}$, can form symmetric dimers via cross-strand interactions, which then act as the building blocks of filaments as dimer-dimer interactions occur between the adjacent subfilaments [Baker et al., 2002; Bullock et al., 1996]. ASH1 and ASH2 could self-associate into oligomers at the centrosome to build Spd-2 scaffolds, a mechanism seen in the case of Cnn. However, while there is strong evidence that Cnn can homo-oligomerise to form scaffold-like structures *in vivo* (**Chapter 2**), and the same is true for the likely Cnn functional homolog SPD-5 in worms [Woodruff et al., 2015], it is less clear that Spd-2 has this property, and it remains possible that Spd-2 largely helps form a PCM scaffold by recruiting and modulating the behavior of Cnn. In order to better understand the role of ASH domains in Spd-2 scaffold assembly, I decided to structurally characterise these domains.

4.3 Purification and characterisation of the Spd-2 ASH1 domain

4.3.1 Expression and purification of the Spd-2 ASH1 domain

Due to the role of ASH domains in Spd-2 scaffold assembly, I wanted to purify and characterise these regions *in vitro*. Based on PSIPRED prediction [Jones, 1999] and using structures modelled on the FFAS hits [Jaroszewski et al., 2011], I designed a construct (Spd-2$^{697-805}$) that included the 8 predicted β strands (**Figure 4.9A**). This excluded a predicted, 2 amino acid strand at residues 690-691, since in the modelled structure this region was represented by a random coil that did not belong to part of the core β sandwich topology (dotted circle in **Figure 4.9B**). The construct was cloned using conventional T4 ligation into the bacterial expression vector, pETM-14 (**Figure 4.9C**). The expressed fusion protein contained a 3C protease-cleavable His6 tag at the N-terminus.

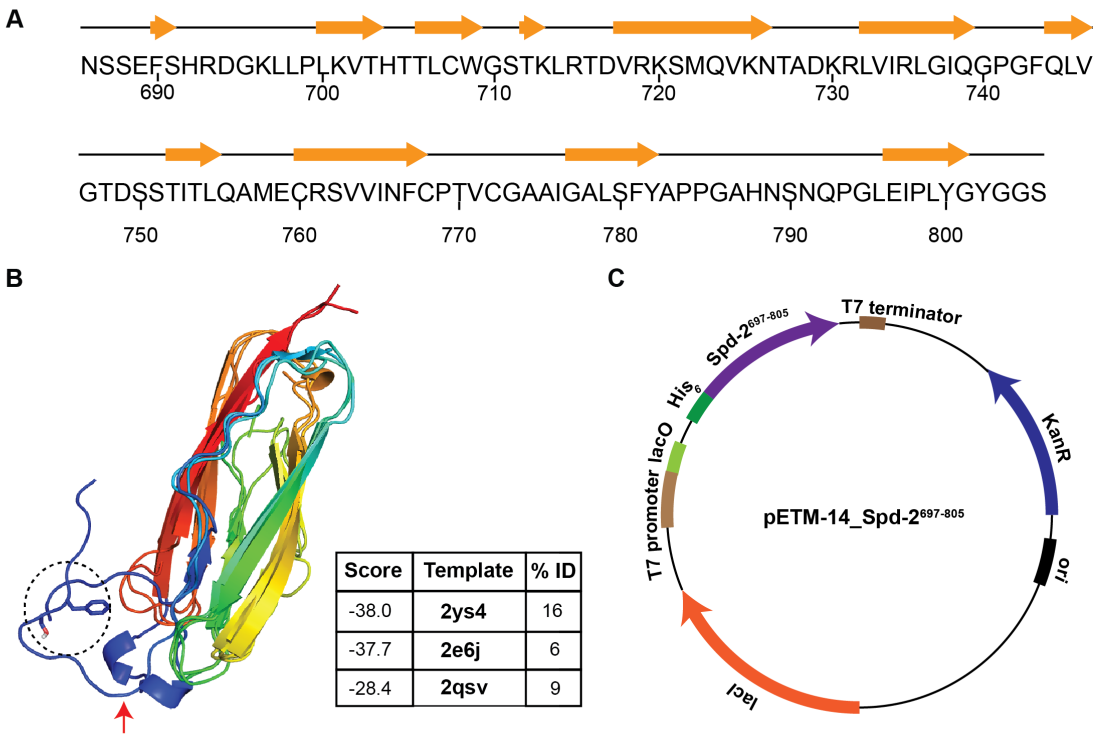
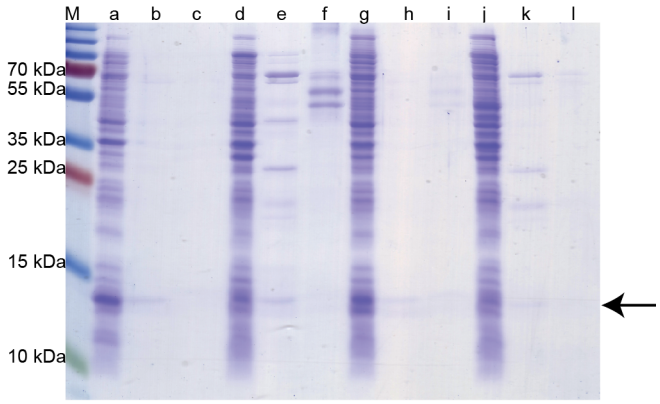


Figure 4.9 (**A**) Schematics showing the topology of Spd-2$^{686-805}$. β strands predicted by PSIPRED are indicated by *orange* arrows, and the protein sequence is shown underneath. Note a short strand is predicted for residues 690 to 691 by PSIPRED, but is not seen in the homology model (**B**). The construct cloned covers Leu697 to Ser805. (**B**) Left, overlay of the homology models predicted by FFAS. Right, quality of the FFAS hits (FFAS score lower than -9.5 indicates less than 3% of false positives) and % identity the models are based on (see **Figure 4.1B** for structures included for homology modelling). Phe690 and Ser691 (side chains highlighted) located in a loop region that was not part of the predicted core structure fold (dashed circle) and was hence excluded. The N-terminus of cloned construct is highlighted with the *red* arrow. (**C**) Plasmid map of the pETM-14 vector into which the Spd-2 ASH1 ORF was cloned. Diagram modified from [Cottee, 2014].

Construct	MW (kDa)	pI	Abs 0.1%	No. of Cys
HisSpd-2$^{697-805}$	13.976	9.27	0.713	4
Spd-2$^{697-805}$	11.949	9.30	0.834	4

Table 4.1 Details of Spd-2 ASH1 protein calculated using Protparam [Wilkins et al., 1999]. Due to the presence of cysteine residues, 5 mM β-mercaptoethanol (BME) was included in working buffer (WB) to prevent disulfide bonds.

A high-throughput expression screen was first carried out, following the standard protocol (**Appendix 8.1.6**), to determine the optimal growth conditions for HisSpd-2$^{697-805}$. The plasmid was transformed into 2 different *E. coli* strains, B834 (DE3) and Tuner (DE3). Cells were grown at 37 °C in LB or TB medium until the OD_{600} reached 0.6. Protein expression was then induced with 1 mM IPTG overnight at 21 °C. Magnetic Ni-NTA beads were used to purify the expressed protein, and gel analysis was followed to confirm the protein overexpression and solubility. As shown in **Figure 4.10**, HisSpd-2$^{697-805}$ was not largely overexpressed although a small amount of protein with the predicted molecular weight was purified.



Induced cell sample and Ni beads elution were loaded in a pair, next to each other. The growth conditions are as following:
a-c= B834 cells/LB medium. d-f= B834 cells/TB medium.
g-i= Tuner cells/LB medium. j-l= Tuner cells/TB medium

Figure 4.10 15% reducing SDS-PAGE gel analysis for the high-throughput trial to identify optimal growth conditions for HisSpd-2$^{697-805}$. The fusion protein was not largely overexpressed in all conditions (lane a, d, g, j). Protein band at expected molecular weight (highlighted by arrow) was detected for purification from B834 (DE3) cells grown in LB medium. Purified protein was eluted off the Ni-NTA beads using WB plus 300 mM imidazole, and a gel sample was taken for analysis (lane b, e, h, k). The beads were then resuspended in 2x SDS loading buffer for gel analysis to confirm a complete protein elution (lane c, f, i, l). M: PageRuler Plus Prestained Protein Ladder (Thermo Scientific).

Soluble protein was only obtained using B834 (DE3) cell strain and LB medium for growth. As a result, these conditions were used for large-scale expression and purification. Due to the low protein expression level, I conducted an 8 l production. Protein purification was carried out following the standard protocol (**Appendix 8.1.7**; **Figure 4.11A**). However, fine precipitate was observed when the protein was eluted off Ni-NTA column (Qiagen). I therefore diluted the volume of eluted sample by five-fold. The N-terminal His6 tag was then cleaved off using 3C protease, which was added at a protease to protein target ratio of 1 mg: 1.48 µmol. The reaction was dialysed into WB overnight at 4 °C. The cleavage reaction was concentrated to 5 ml, using a 3

kDa MWCO centrifugal concentrator (Amicon), before loading onto an S75 16/60 column (GE Healthcare) that was equilibrated into WB. Unfortunately, precipitation was observed upon concentration, so the sample was centrifuged at 15,000x g for 5 mins before the supernatant was collected for SEC. Gel analysis showed the elution of full length protein within the major SEC peak around 0.69 CV (**Figure 4.11B**).

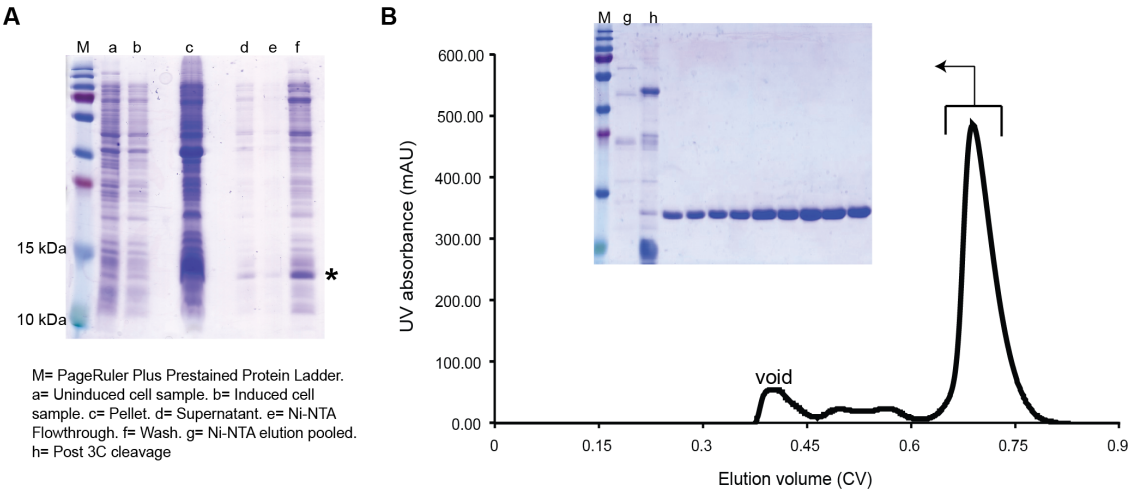
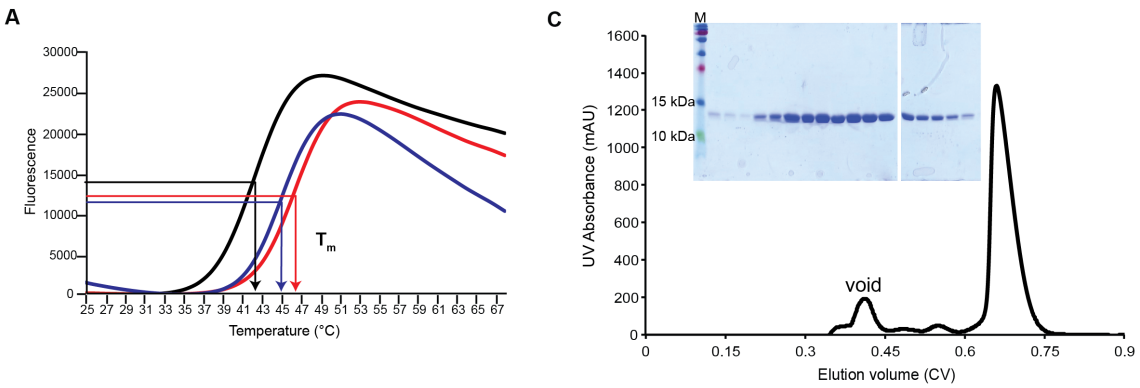


Figure 4.11 Purification of Spd-2$^{697-805}$. (**A**) 15% reducing SDS-PAGE gel analysis of Ni-NTA purification fractions. The asterisk indicates the band corresponding to HisSpd-2$^{697-805}$. (**B**) S75 16/60 SEC profile and the subsequent SDS-PAGE gel analysis showing the elution of full length protein around 0.69 CV. Elution with 50 mM Tris (pH 7.5), 150 mM NaCl and 5 mM BME. Following SEC, the protein sample was already at a high purity.

In order to optimise the buffer conditions and hence improve the protein solubility, I carried out a fluorescence-based thermal shift assay, ThermoFluor, for Spd-2$^{697-805}$. SYPRO Orange dye (Sigma) used in this assay will bind to hydrophobic patches upon protein denaturation. As a result, the change in dye fluorescence can be monitored as the temperature increases and the protein starts to unfold. The midpoint of this unfolding transition, T_m, is measured and compared for protein in different buffer conditions including pH, salt concentration and the addition of glycerol (**Figure 4.12**). The SEC fractions were pooled and the protein concentration was measured by a nanodrop (Thermo Scientific). 5 µl of the protein stock (1 mg/ml) was pipetted into a 96-well plate that contained 12.5 µl of premixed buffers. SYPRO Orange dye was then added to each well at a 1:500 *v/v* final dilution in a 25 µl reaction. The melting curves are shown in **Figure 4.12B**, which showed an unfolding transition suggesting Spd-2$^{697-805}$ had tertiary structure. Spd-2$^{697-805}$ was more stable at pH 6.5 and pH 7.0, and the salt concentration did not significantly affect its stability. The addition of 5/15% glycerol was also preferred. Based on the assay read-out, I then optimised the working buffer composition for purification of Spd-2$^{697-805}$. Because Nickel

ions have a pI around 6.5, Ni-NTA purification was performed at 50 mM Tris (pH7.3), 150 mM NaCl and 5mM BME. Protein was, however, eluted with 50 mM Tris (pH7.3) 150 mM NaCl, 300 mM imidazole, 5 mM BME and 5% Glycerol. The 3C cleavage reaction was dialysed into a buffer with lowered pH (20 mM MES (pH6.5), 150 mM NaCl, 5 mM BME and 5% glycerol). The untagged protein was further purified and exchanged into its storage buffer (20 mM Tris (pH7.0), 150 mM NaCl, 2 mM TCEP) via S75 16/60 SEC (**Figure 4.12C**). Note the final storage buffer was at a more basic pH and had no glycerol added which could impede nucleation events during crystallisation. The typical yield of SEC purified Spd-2$^{697-805}$ was ~3 mg protein per 1 l culture, a 3-fold increase on the yield was obtained when purification was conducted in buffer at pH 7.5.



B

pH/buffer

[NaCl]/ %glycerol	NaCitrate pH 4.0	Lactic Acid pH 4.5	PiCitrate pH 5.0	Succinate pH 5.5	MES pH 6.0	MOPS pH 6.5	PIPES pH 7.0	NaPi pH 7.5	HEPES pH 8.0	Bicine pH 8.5	BisTris pH 9.0	CHES pH 9.5
0/0	39.3	44.5	41.9	41.9	43.2	44.5	43.2	41.9	41.9	39.3	40.6	35.4
100 mM/0	38	41.9	41.9	41.9	41.9	43.2	43.2	41.9	35.4	39.3	39.3	35.4
200 mM/0	38	40.6	40.6	41.9	41.9	41.9	41.9	40.6	41.9	39.3	39.3	36.7
300 mM/0	39.3	39.3	40.6	41.9	41.9	43.2	41.9	41.9	40.6	40.6	39.3	36.7
400 mM/0	38	39.3	40.6	41.9	41.9	41.9	41.9	41.9	41.9	40.6	39.3	36.7
500 mM/0	38	39.3	40.6	41.9	40.6	41.9	41.9	41.9	41.9	40.6	39.3	36.7
0/10%	39.3	45.8	43.2	44.5	45.8	45.8	44.5	43.2	43.2	41.9	41.9	38
0/30%	41.9	45.8	44.5	45.8	47.1	47.1	45.8	43.2	44.5	43.2	43.2	39.3

Figure 4.12 Optimisation of the purification buffer for Spd-2$^{697-805}$. (**A**) The unfolding transition of Spd-2$^{697-805}$ obtained via ThermoFluor. Upon protein denaturation, SYPRO Orange Dye binds to the exposed hydrophobic core leading to an increase in fluorescence. T_m (the mid-point of unfolding transition) is compared in (**B**) for different buffer conditions. The protein sample was originally stored in 50 mM Tris (pH 7.5), 150 mM NaCl, 5 mM BME (*black*). Note the transition was shifted towards higher T_m at pH 6.5 (*red*) and pH 7.0 (*blue*) in the presence of 5% glycerol. (**B**) The calculated T_m values are summarised in a 96-well format. Each reaction contained 5 µl protein, 7.5 µl dye (1:500 final *v/v* ratio) and 12.5 µl buffer screen as shown. Results are colored according to magnitude, from *blue* (low) to *red* (high). (**C**) S75 16/60 SEC using the optimised buffer (20 mM Tris (pH 7.0), 150 mM NaCl, 2 mM TCEP). 15% reducing SDS-PAGE gel analysis showing the protein elution across the main peak. M: PageRuler Plus Prestained Protein Ladder (Thermo Scientific).

4.3.2 Biophysical characterisation of Spd-2 ASH1 domain

In order to characterise the oligomerisation state of the Spd-2 ASH1 domain in solution, I prepared a sample at 84 µM for SEC-MALS experiment. The sample was injected onto an S75

10/30 column (GE Healthcare) that was equilibrated into the optimised protein storage buffer. Spd-2$^{697-805}$ was eluted predominantly as a monomer (**Figure 4.13**).

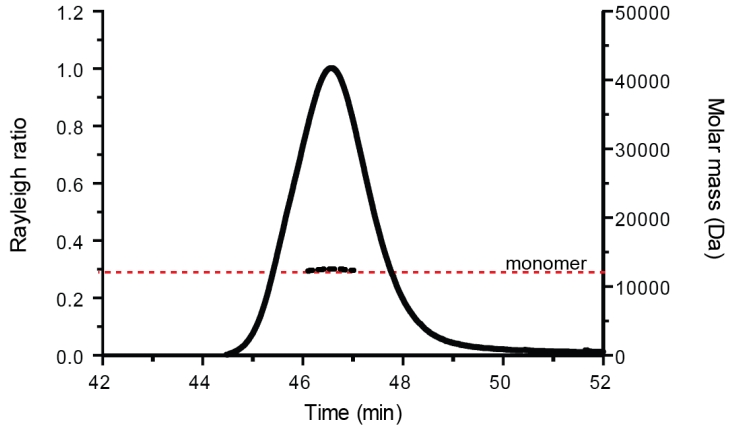


Figure 4.13 SEC-MALS analysis of Spd-2$^{697-805}$ showing the elution of monomeric species off an S75 10/30 column. Elution with 20 mM Tris (pH 7.0), 150 mM NaCl, 2 mM TCEP. A solid line represents the protein elution with its observed molecular weight shown in a dashed line across the peak. The expected mass of a monomer is indicated with a *red* dotted line across the graph.

A CD experiment was then carried out to check the secondary structure content of the purified ASH1 domain. Protein was diluted to 0.1 mg/ml before overnight dialysis into 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.0). Spectra were collected between 195 and 250 nm at 20 °C, and were averaged over four accumulations using the dialysis buffer as a blank measurement. The spectrum shows a trough at ~218 nm and a peak at ~197 nm that are both characteristic of a β-type profile (**Figure 4.14**). This indicates Spd-2$^{697-805}$ is rich in β strands, which is in agreement with the PSIPRED prediction and the FFAS predicted models.

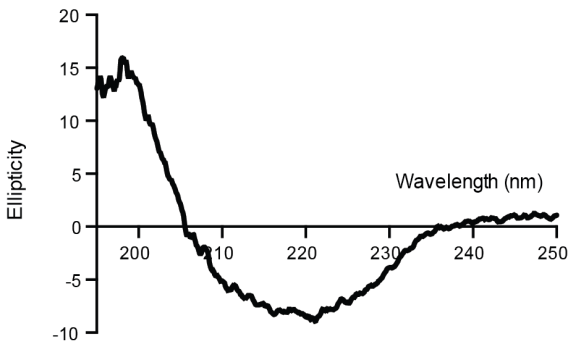


Figure 4.14 CD spectrum for Spd-2$^{697-805}$ carried out in 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.0) at 20 °C. Four scans were acquired and normalised with the dialysis buffer (10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.0)) as a blank measurement. A negative trough was detected at around 218 nm and a positive peak at 197 nm, both indicative of β strand content.

4.4 Structural characterisation of the Spd-2 ASH1 domain

4.4.1 Crystallisation trials with Spd-2$^{697-805}$

The SEC purified Spd-2$^{697-805}$ provided a higher yield in the optimised buffer conditions, and a homogenous sample was obtained that appeared to be correctly folded as suggested by CD and ThermoFluor analysis. The sample was therefore used for structural characterisation, and crystallisation trials were set up using the vapour-diffusion method.

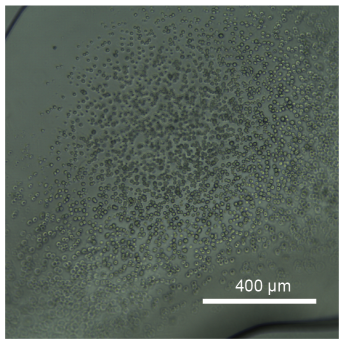
The protein sample was concentrated using 3 kDa MWCO centrifugal concentrator (Amicon), and was centrifuged at 15,000x g for 5 mins to remove any precipitate before putting down into the trays. I noticed that the sample showed precipitation over time, despite its concentration was not significantly changing as measured by a nanodrop (Thermo Scientific). A range of protein concentrations were used in the crystallisation trials summarised in **Figure 4.15A**.

The broad screens (Molecular Dimensions) were set up with a drop size of 0.2 µl at 50/75% protein: mother liquor *v/v* ratio. After 3-week incubation at 20 °C, no protein crystals were grown. The majority of the conditions showed clear drops or brownish precipitation at 40 and 50 mg/ml, and phase separations at 10 mg/ml. The formation of dark coloured heavy precipitate indicated the protein has denatured in these conditions, which were unfavourable and should be avoided. Interestingly, crystalline and spherulite formations were frequently observed in conditions where salt was used as a precipitant. In particular, two crystallisation conditions (150 mM sodium citrate (pH 5.5), 0.75-2.0 M ammonium sulphate; 10 mM sodium borate (pH 8.5), 0.75-1.5 M sodium citrate) consistently showed crystalline formation at higher protein concentration and phase separation at lower concentration (**Figure 4.15B**). I thus decided to optimise these particular mother liquor conditions including buffer pH and salt concentrations (**Figure 4.15C**). However, the optimisation screens produced no protein crystals and the majority of drops showed gel-like precipitates.

A

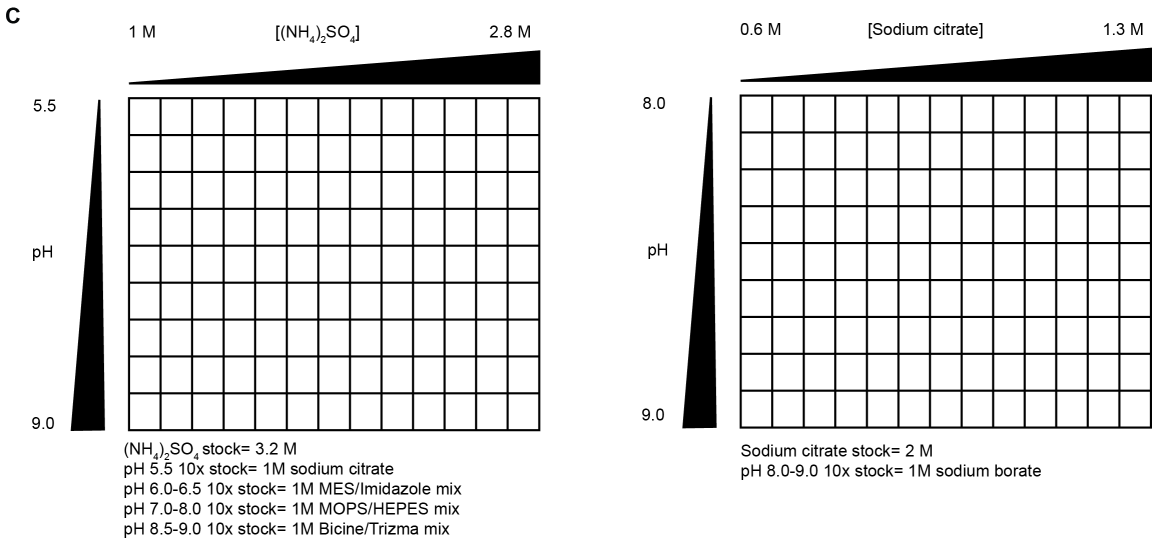
Protein concentration	Broad screens
10 mg/ml	Stura, Structure, JCSG+, Morpheus, PACT, Proplex, Midas, Clear Strategy I+II
20 mg/ml	Stura, Structure, JCSG+, Morpheus, PACT, Proplex, Midas, Clear Strategy I+II
40 mg/ml	Stura, Structure, JCSG+, Morpheus, PACT, Proplex, Midas, Clear Strategy I+II
50 mg/ml	Stura, Structure, JCSG+, Morpheus, PACT, Proplex

B



0.2 µl drop= 0.1 µl Spd-2$^{697-805}$ at 50 mg/ml and 0.1 µl mother liquor (10 mM sodium borate (pH 8.5), 1.0 M sodium citrate)

C



$(NH_4)_2SO_4$ stock= 3.2 M
pH 5.5 10x stock= 1M sodium citrate
pH 6.0-6.5 10x stock= 1M MES/Imidazole mix
pH 7.0-8.0 10x stock= 1M MOPS/HEPES mix
pH 8.5-9.0 10x stock= 1M Bicine/Trizma mix

Sodium citrate stock= 2 M
pH 8.0-9.0 10x stock= 1M sodium borate

Figure 4.15 Crystallisation trials acquired with Spd-2$^{697-805}$. (**A**) Diagram showing the protein concentrations attempted and the broad screens carried out. (**B**) One example of drops demonstrating spherulite-like precipitate. Clear drop or brownish precipitation was observed for the majority of conditions, expect the ones included salt as a precipitant. (**C**) Optimisation of the initial conditions that had formation of crystalline precipitate or spherulites. Optimisation screens were set up with protein at 50 mg/ml.

Since the purified Spd-2$^{697-805}$ did not crystalise, I wondered if the protein had flexible regions that might affect the ordered crystal lattice packing. In order to define the core structure region, I carried out limited proteolysis experiments using trypsin (Sigma) and subtilisin (Sigma, **Figure 4.16A**). Trypsin specifically cleaves at the C-terminal side of accessible Lys or Arg residues, whereas subtilisin is more non-specific.

40 µl reactions were prepared containing SEC purified Spd-2$^{675-805}$ protein at 1 mg/ml in the optimised WB supplemented with varying amounts of proteases (**Figure 4.16B**). Reactions were incubated at room temperature for 2 hrs, and 15% SDS-PAGE gel analysis was then followed to check whether a more stable core structure could be obtained (**Figure 4.16B**). Spd-2$^{697-805}$ showed no further truncation in trypsin digestion. When a higher ratio of subtilisin: target protein (1:50 *v/v* protease: protein ratio) was included in the reaction, two bands of lower molecular

weight were observed in gel analysis, although the full-length band was still retained. Even in the presence of significant amount of non-specific subtilisin protease, ~50% of the Spd-2$^{697-805}$ sample was still intact suggesting the protein was fairly stable.

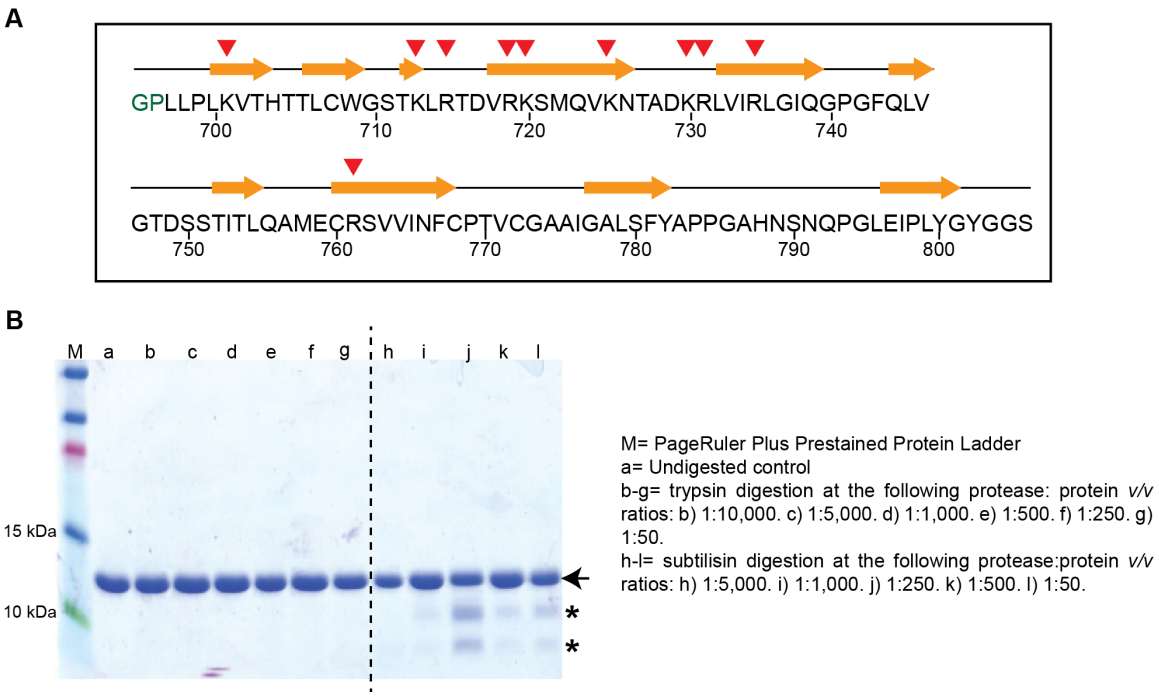


Figure 4.16 Limited proteolysis of Spd-2$^{697-805}$ in order to identify the core domain. (**A**) Schematics of Spd-2$^{697-805}$ with the highlighted Lys and Arg residues (*red* arrowheads), which could be accessible to trypsin digestion if located in an unfolded region. Note trypsin cleavage sites are not found in the C-terminal 40 residues hence subtilisin digestion was also carried out. (**B**) 15% reducing SDS-PAGE gel analysis following the reaction showed no truncated lower bands in trypsin digestion. Two minor bands (highlighted by asterisks) were observed in subtilisin digestion at higher protease: target protein ratio (1:50). However, the full-length band (highlighted by arrow) was still retained with intensity much stronger than the lower bands, indicating Spd-2$^{697-805}$ represents a highly stable domain.

Reductive methylation of Lys side chain has been shown to encourage the crystallisation of proteins that are homogenously purified but could not be packed into ordered crystal lattices due to the presence of flexible side chains [Walter et al., 2006]. I therefore conducted methylation of the Spd-2$^{697-805}$ protein. Because the reaction had to be performed in a buffer of pH 7.5, SEC was carried out in 50 mM HEPES (pH7.5), 150 mM NaCl and 5 mM BME. The pooled fractions were divided into 1 ml aliquots at 1 mg/ml. Each aliquot was mixed with 20 µl freshly prepared 1 M dimethylamineborane complex (DMAB; Fluka product 15584) and 40 µl 1 M formaldehyde (Fluka product 33220) for 2 hrs in the dark at 4 °C. This mixing was carried out in three rounds. 10 µl of 1 M DMAB was further added, following which the reaction was incubated overnight. The reaction led to a small amount of precipitated protein, which was then removed by centrifugation at 15,000x g for 5 mins before loading onto the S75 16/60 column (GE Healthcare). The reaction

was quenched as the methylated protein was purified on the SEC column, using the optimised storage buffer as the running buffer (the same protocol followed as the unmethylated protein, see **Section 4.3.1**). Mass spectrometry analysis of the final purified sample indicated that all four Lys side chains were methylated giving a mass of 12.12 kDa. Crystallisation trials were carried out with methylated Spd-2$^{697-805}$ at 23 mg/ml, a starting concentration suggested by the Stura solubility test. After 3 weeks of incubation at 20 °C, no protein crystals had grown and more trials were conducted with a concentration of 32 mg/ml. Unfortunately, this has still not yielded any crystal hits.

In summary, the crystallisation trials I have performed in this work have not produced protein crystals, despite the sample seemed pure and homogenous with a stably folded core structure. Further modification of the crystallisation parameters could be conducted including lowering the incubation temperature, optimising the mother liquor: protein ratio or testing other protein concentrations. Crystallisation trials have been set up at 50 mg/ml, which was likely the solubility limit of Spd-2$^{697-805}$ as further concentration led to immediate protein precipitation. It is possible that the protein does not have the intrinsic property to crystallise. Further analysis of the structure modelled on the FFAS hits indicated that a large number of hydrophobic residues were solvent exposed (**Figure 4.17**). This is unusual as such residues tend to be buried in the core. If this were the case, these residues would likely be involved in self-association or interaction with other proteins in order to be buried away from the solvent. As demonstrated in the RNA injection assay, the ASH1 domain is essential for Spd-2 scaffold formation either via multimerisation or binding to Cnn scaffolds. Therefore, its structure could help to understand the underlying mechanism. I then decided to attempt structure determination using NMR spectroscopy in collaboration with Dr Devon Sheppard, a post doc in the lab.

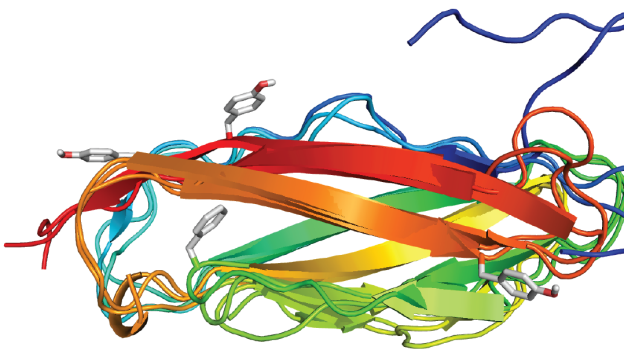


Figure 4.17 Side-on view of Spd-2$^{697-805}$ homology models showing the exposed hydrophobic Phe/Tyr side chains. The structure was modelled on the FFAS hits. All structures are coloured in rainbow format. For clarification, Phe/Tyr side chains are displayed for one model, and the other models show the same sidechain orientations.

4.4.2 Sample preparation for NMR spectroscopy

In order to confirm Spd-2$^{697-805}$ is well folded in solution and is suitable for structure determination with NMR spectroscopy, I expressed and purified ^{15}N-^{13}C double-labelled Spd-2$^{697-805}$ for ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) experiment. Due to the insufficient natural abundance levels of ^{15}N and ^{13}C, the isotropic enrichments of protein sample was achieved by using ^{13}C glucose and ^{15}N ammonium chloride (Sigma) as the only carbon and nitrogen sources during protein expression. pETM-14-Spd-2$^{697-805}$ plasmid was transformed into a non-methionine auxotrophic *E. coli* strain (BL21 (DE3) cells). A single colony was picked for overnight inoculation in LB medium (supplemented with 50 µg/ml Kanamycin antibiotic) at 37 °C, which was subsequently used as a starting culture for large-scale growth in LB media (supplemented with 50 µg/ml Kanamycin antibiotic). Cells were grown at 37 °C until OD_{600} reached 0.9, and were then harvested by centrifugation according to the usual protocol (**Appendix 8.1.7**). Each cell pellet, from 1 l culture, was used to inoculate 500 ml minimal medium (prepared as **Table 4.2**). Cells were shaking at 21 °C for 30 mins before the addition of 1 mM IPTG for overnight induction.

Composition	Final concentration
1x M9 salt (pH 7.2):	
Na_2HPO_4	26 mM
KH_2PO_4	22 mM
NaCl	8 mM
$MgSO_4$	2 mM
$CaCl_2$	0.1 mM
Thiamine	2.96 mM
(^{13}C-)glucose	0.4% *w/v*
$^{15}NH_4Cl$	1 mg/ml
Kanamycin	50 µg/ml

Table 4.2 Composition of minimal media for labelled protein expression.

Purification of ^{15}N-^{13}C Spd-2$^{697-805}$ followed the same protocol as the unlabelled protein. However, the SEC buffer was different because Tris is protonated. Therefore the protein was purified on an S75 16/60 column (GE Healthcare) that was equilibrated into 50 mM NaH_2PO_4/Na_2HPO_4 (pH 7.0), 150 mM NaCl and 2 mM DTT. Mass spectrometry analysis of the final purified sample indicated a mass of 12,606 Da. The expected mass of fully labelled ^{15}N-^{13}C Spd-2$^{697-805}$ is 12,623.8 Da,

hence a nearly 100% labelling efficiency was achieved. D_2O was added in the protein sample (350 µl at 1 mM) to a final concentration of 5% *v/v*. The heavy isotope deuterium of D_2O will exchange with the hydrogen atoms on the solvent, which would otherwise give rise to chemical shift peaks in the NMR spectrum. Deuterium has a different magnetic moment and spin because of its integer spin number, 1 instead of 1/2. As a result, this heavy isotope will not be detected in a spectrometer tuned to protons (^1H). This explains why Tris buffer is not suitable as an NMR solvent because it has non-exchangeable protons (i.e. it is protonated).

4.4.3 NMR data collection for assignment of Spd-2$^{697-805}$

A ^1H-^{15}N Heteronuclear Single Quantum Coherence (HSQC) spectrum was collected at 20 °C, as shown in **Figure 4.18**. The appearance of good peak dispersion in both dimensions indicated the protein was not aggregating and had well-defined secondary and tertiary structural contents amenable for NMR structure determination.

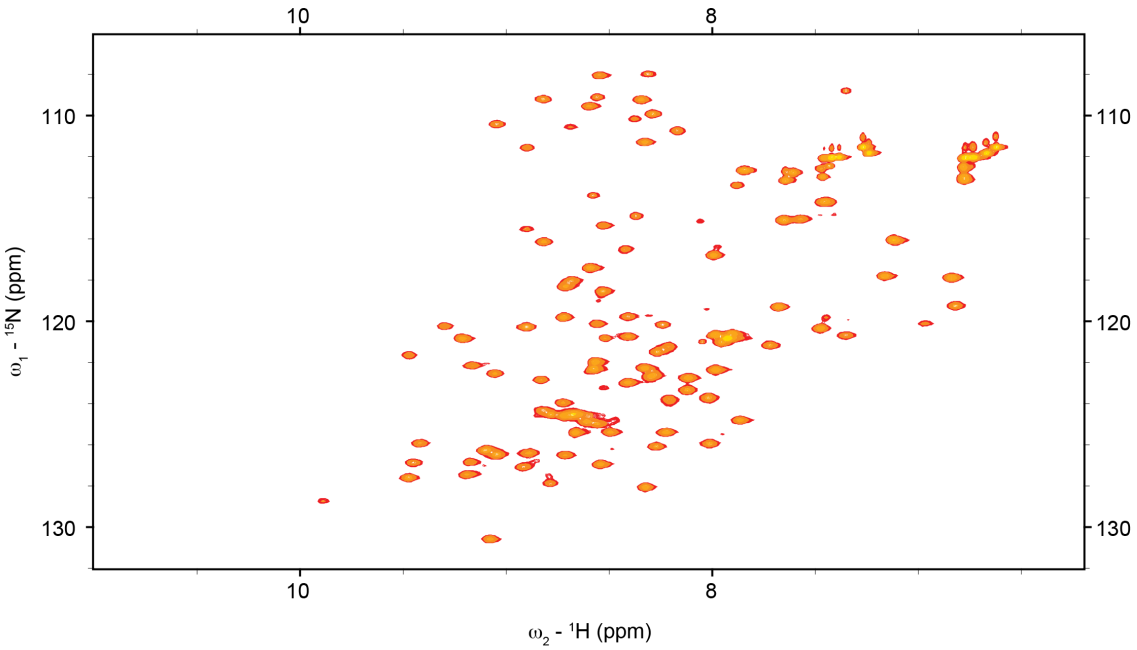


Figure 4.18 ^1H, ^{15}N-HSQC spectrum for $^{13}C/^{15}N$-Spd2$^{697-805}$ prepared at 1 mM in 50 mM NaH_2PO_4/Na_2HPO_4 (pH 7.0), 150 mM NaCl, 2 mM DTT and 5% *v/v* D_2O.

A series of NMR spectra were then collected on Bruker Avance II 500MHz triple resonance, pulse field gradient system equipped with cryoprobe, and were processed by Dr Devon Sheppard who then solved the NMR structure of the Spd-2 ASH1 domain.

4.5 Purification and characterisation of the Spd-2 ASH2 domain

4.5.1 High-throughput screen for s soluble fragment of the Spd-2 ASH2 domain

Given the functional importance of the Spd-2 ASH2 domain, I decided to attempt to purify and characterise this fragment *in vitro* in order to better understand its role *in vivo*. Based on the structure models built using FFAS hits, I designed 2 constructs one of which included a predicted α helix at the C-terminus (Spd-$2^{820\text{-}921/942}$; **Figure 4.19A**). In addition, a construct covering both the ASH1 and ASH2 domains were designed, since the soluble and well-folded ASH1 region might facilitate the expression and solubility of ASH2. All fragments were cloned into pETM-14 vector by T4 ligation (NEB). The plasmids were transformed into 2 different *E. coli* strains, B834 (DE3) and Tuner (DE3) cells, for protein expression in LB or TB medium (**Appendix 8.1.6**). The expressed protein was purified using Ni-NTA beads (Qiagen) and were subsequently analysed on a 15% SDS-PAGE gel. No protein overexpression was observed for HisSpd-$2^{823\text{-}921/942}$, but B834 (DE3) cells grown in LB medium showed overexpression of HisSpd-$2^{697\text{-}942}$ (**Figure 4.19B**). No soluble protein was detected for any of the tested constructs. Due to the poor expression level in the tested strains, I transformed the plasmids into Rosetta™ (DE3)pLysS competent cells that are designed to enhance the expression of eukaryotic proteins in bacterial cells. Although protein was overexpressed using the new strain, no soluble fraction could be obtained (**Figure 4.19C**).

Construct	MW (kDa)	pI	Abs 0.1%	No. of Cys
HisSpd-$2^{820\text{-}921}$	13.585	8.96	0	3
HisSpd-$2^{820\text{-}942}$	16.224	10.21	0	3
HisSpd-$2^{697\text{-}942}$	29.205	9.91	0.341	7

Table 4.3 Details of the Spd-2 constructs expressed and purified in high-throughput screen, as calculated using ProtParam [Wilkins et al., 1999]. Note Spd-$2^{820\text{-}921/942}$ is predicted to have no absorbance at 280 nm. Due to the presence of cysteine residues, 5 mM β-mercaptoethanol (BME) was included in working buffer (WB) to prevent disulfide bonds.

Further constructs were designed as summarised in **Table 4.4** and **Figure 4.20A**. All the fragments were cloned into pETM-14 vector, and the plasmids were transformed into different *E. coli* strains including BL21 (DE3), B834 (DE3) and Tuner (DE3) cells. Unfortunately, no soluble protein could be purified (**Figure 4.20B**).

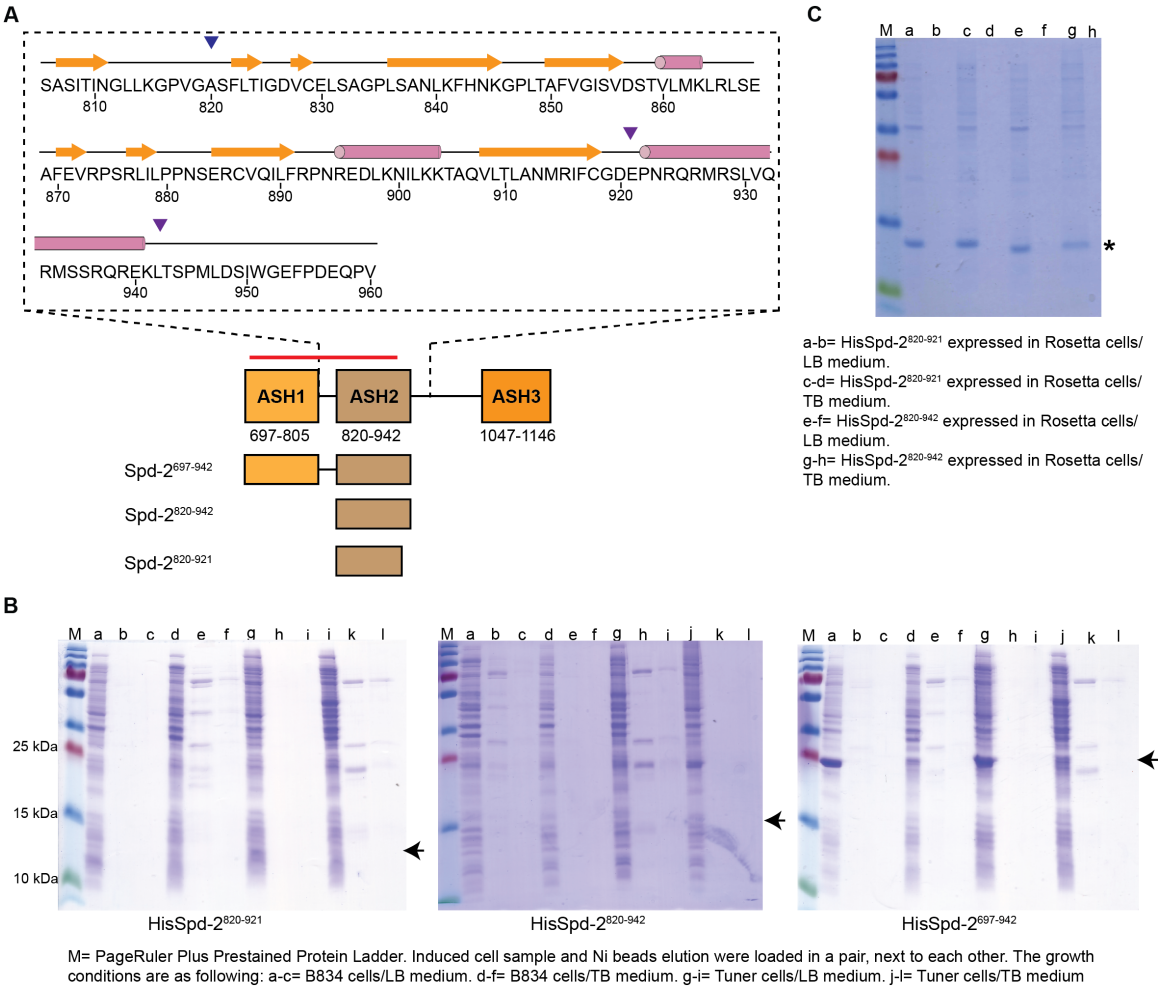
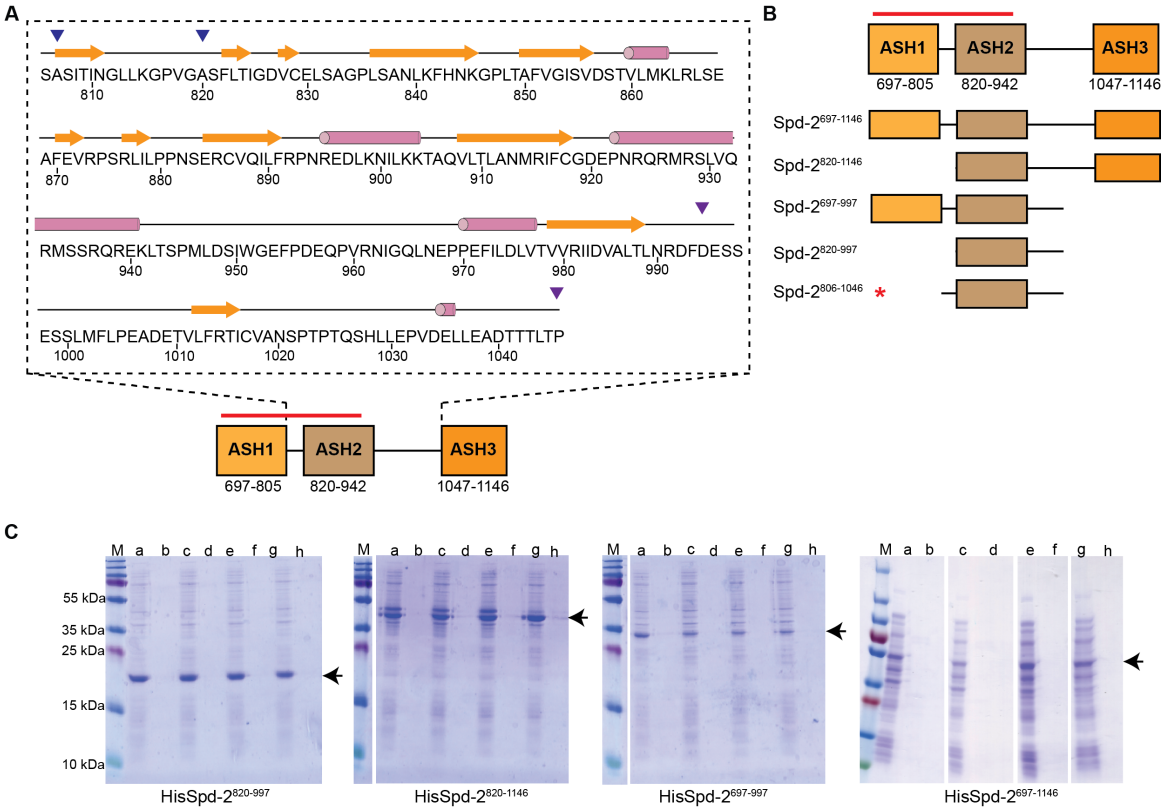


Figure 4.19 (**A**) Schematics showing the overall topology of Spd-2$^{805-960}$. Protein sequence is shown, and the β strands predicted by PSIPRED are indicated by *orange* arrows. FFAS hits-based structure model of the Spd-2 ASH2 domain indicated that the predicted strand of residues 806 to 811 provides a linker between ASH1 and ASH2 regions, and this linker was not included in the ASH2 construct. Two fragments were designed, Spd-2$^{820-921/942}$. Another construct was cloned including both ASH domains. (**B,C**) High-throughput expression and purification screen was performed for the constructs designed in (**A**). Different *E. coli* strains and growth media were tested. 15% reducing SDS-PAGE gel analysis was carried out to confirm the protein induction and solubility. All proteins were overexpressed, but not detected in Ni-NTA elution fractions. Arrows indicate the bands corresponding to HisSpd-2 constructs. Note in (**B**) gel samples were taken for protein eluted from Ni-NTA beads (lane b, e, h & k) and the beads resuspended in 2x SDS loading dye (lane c, f, i & l).

Construct	MW (kDa)	pI	Abs 0.1%	No. of Cys
HisSpd-2$^{820-997}$	22.52	7.82	0.244	3
HisSpd-2$^{697-997}$	35.43	9.12	0.437	7
HisSpd-2$^{820-1146}$	38.99	5.28	0.320	9
HisSpd-2$^{697-1146}$	51.90	6.04	0.433	13

Table 4.4 Details of the Spd-2 constructs expressed and purified in further high-throughput screen, as calculated using ProtParam [Wilkins et al., 1999]. Due to the presence of cysteine residues, 5 mM β-mercaptoethanol (BME) was included in working buffer (WB) to prevent disulfide bonds.



M= PageRuler Plus Prestained Protein Ladder. Induced cell sample and Ni beads elution were loaded in a pair, next to each other. The growth conditions are as following: a-b= B834 cells/LB medium. c-d= B834 cells/TB medium. e-f= Tuner cells/LB medium. g-h= Tuner cells/TB medium

Figure 4.21 (**A,B**) Further constructs were designed for the high-throughput screen. The N- and C-termini are highlighted with arrowheads (*blue* showing the N-terminus and *purple* showing the C-terminus) in (**A**). The different constructs trialled in the high-throughput test are listed in (**B**). Construct labelled with red asterisk was not trialed in the high-throughput test, but rather by large-scale purification discussed in **Section 4.5.3**. (**C**) 15% reducing SDS-PAGE gel analysis of induced cell samples and various Spd-2 constructs purified by magnetic Ni beads. Arrows indicate the bands corresponding to HisSpd-2 constructs. No soluble protein could be detected in the elution.

I attempted two further approaches trying to generate soluble ASH2 domain protein. First, an N-terminal MBP tag was fused with Spd-2$^{820-942}$, which can be cleaved off using 3C protease. MBP is highly soluble and hence would increase the solubility of the fusion protein. Second, I mutated the conserved Ser/Thr sites within the PR3 region (Spd-2$^{994-1044}$) to Asp/Glu residues (phosphomimetic mutations, hereafter referr to as 10D/E; **Figure 4.7D**). In the case of the Cnn PReM domain, the wild-type construct was not soluble unless phosphomimetic mutations were introduced. I wondered if the same situation might occur with the Spd-2 ASH2 domain. Although Spd-2 10A-GFP showed normal localisation at centrosomes in RNA injection assay, the presence of endogenous Spd-2 WT protein could affect the observation. To ensure the entire domain was covered, I have cloned a large construct containing residues 806 to 1046 with the 10D/E mutations (**Figure 4.20A**). Both MBPSpd-2$^{820-942}$ and HisSpd-2$^{806-1046}$ 10D/E constructs were subjected to large-scale expression and purification in Tuner (DE3) cells/TB media (**Table 4.5**).

Construct	MW (kDa)	pI	Abs 0.1%	No. of Cys
MBPSpd-2$^{820-942}$	56.740	6.82	1.169	3
Spd-2$^{820-942}$	14.268	10.35	0	3
HisSpd-2$^{803-1046}$	29.342	5.77	0.187	3
HisSpd-2$^{803-1046}$ 10D/E	29.622	4.9	0.186	3

Table 4.5 Details of the Spd-2 constructs expressed and purified on a large-scale. Due to the presence of cysteine residues, 5 mM β-mercaptoethanol (BME) was included in working buffer (WB) to prevent disulfide bonds.

4.5.2 Large-scale purification of MBPSpd-2$^{820-942}$

A 2 l production was conducted for MBPSpd-2$^{820-942}$, and the fusion protein was purified following the standard protocol (**Appendix 8.1.7**). MBP has an N-terminally fused His6 tag that enabled its purification by Ni-NTA affinity chromatography (**Figure 4.22A**). Protein eluted off Ni-NTA column (Qiagen) was mixed with 3C protease at a ratio of 1 mg 3C: 1.48 µmol target protein. The reaction was dialysed overnight at 4 °C against WB. A gel sample was taken to check the tag cleavage before the sample was concentrated to 5 ml, using a 3 kDa MWCO centrifugal concentrator (Amicon), for loading onto an S75 26/60 SEC column (GE Healthcare, **Figure 4.22C**). The elution profile showed two major peaks at 0.34 CV and 0.47 CV respectively. Gel analysis of SEC fractions across each peak indicated that free MBP tag was partially removed in the second peak, and the peak close to void (void volume of an S75 26/60 column= ~0.33 CV) contained untagged and tagged Spd-2$^{820-942}$ as well as MBP (**Figure 4.22B,C**). 3C tag cleavage was not complete after 20-hr digestion at 4 °C as shown via gel analysis. I therefore pooled fractions within the first SEC peak and resumed 3C cleavage for another 17 hrs at 4 °C after the supplement of fresh protease. Further digestion still did not yield complete tag cleavage, which suggested the protein might be aggregating and some protease recognition sites were therefore not accessible for cleavage.

In order to isolate the untagged Spd-2 protein, I carried out another purification step via cation exchange chromatography. The predicted pI for Spd-2$^{820-942}$ with or without the MBP tag is very different (6.82 and 10.35) that should allow purification based on the charge difference. At pH 7.5, the untagged protein would carry a net positive charge whereas the MBP tagged fusion protein would have an overall negative charge. The sample was dialysed overnight into 50 mM Tris (pH 7.5) at 4 °C. A mono S 5/50 GL (GE Healthcare) ion exchange column was used for purification,

which is prepacked with negatively charged beads. Ideally the column would bind to Spd-2$^{820-942}$ that would be positively charged at pH 7.5, and subsequently the fusion protein would be found in flow through because of the negative charge repulsion. However, the cleaved protein seemed to self-associate with MBPSpd-2$^{820-942}$, and both were collected in the flow through (**Figure 4.22D,E**). The failure to separate out untagged Spd-2 protein has hindered further biophysical experiments.

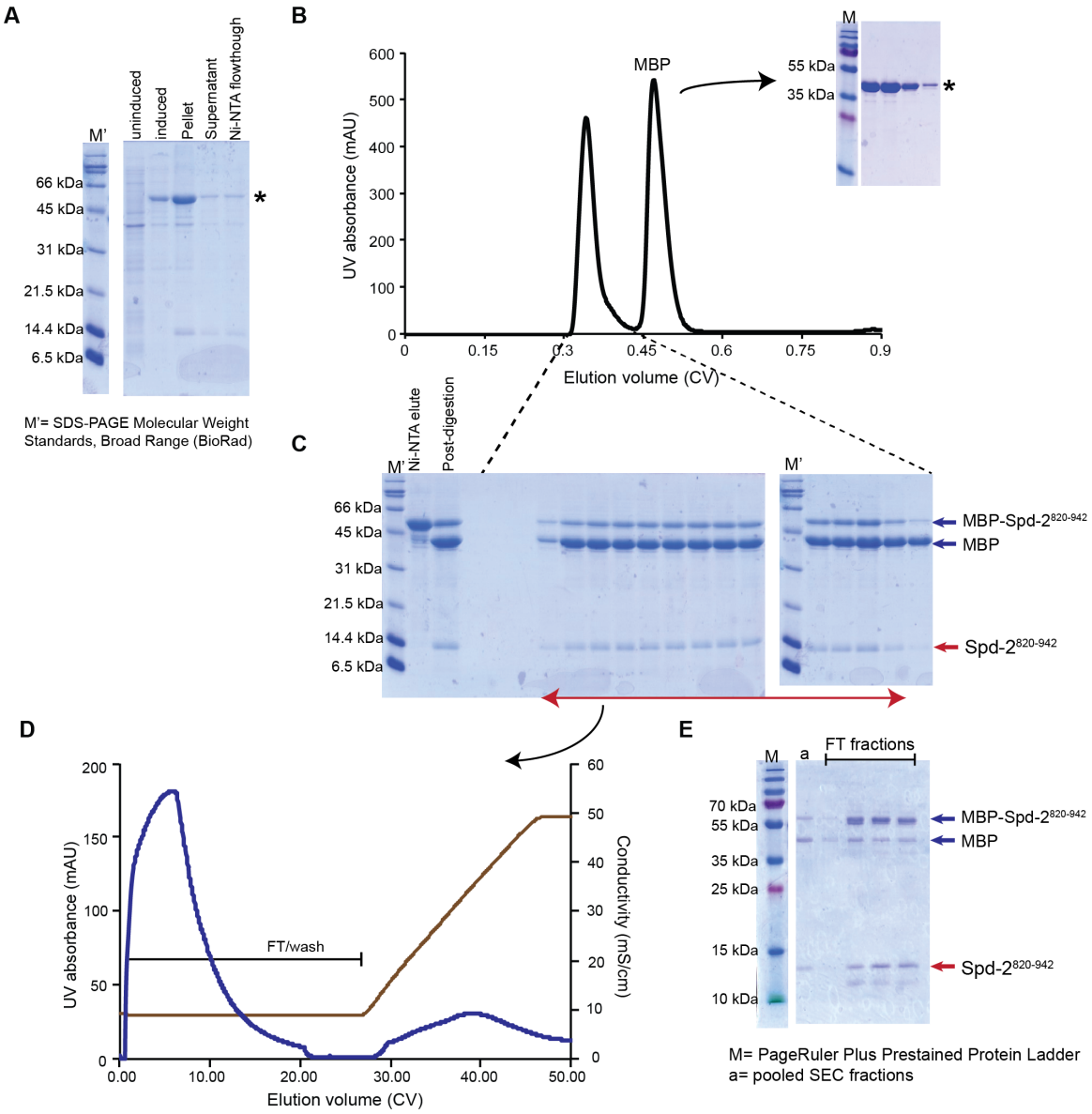


Figure 4.22 Purification of Spd-2$^{820-942}$. (**A**) 15% reducing SDS-PAGE gel analysis of Ni-NTA purification fractions. MBPSpd-2$^{820-942}$ was largely insoluble as majority of the fusion protein was present in the lysate pellet. Asterisk indicates the band corresponding to MBPSpd-2$^{820-942}$. (**B,C**) S75 26/60 SEC (**B**) and 15% reducing SDS-PAGE gel analysis (**C**) showing the incomplete tag cleavage and the failure to separate out untagged Spd-2$^{820-942}$, since mixture of proteins eluted in the peak close to void volume (i.e. ~0.33 CV). Elution with 50 mM Tris (pH 7.5), 150 mM NaCl and 5 mM BME. SEC fractions within the first elution peak (as highlighted on the gel) were pooled for monoS cation exchange chromatography. (**D,E**) Cation exchange (**D**), and the subsequent gel analysis (**E**) demonstrating the appearance of MBP, tagged and untagged forms of Spd-2$^{820-942}$ all in the flow through. Sample was loaded at 50 mM Tris (pH 7.5), 50 mM NaCl and 5 mM BME, and an elution gradient of 50 to 500 mM NaCl over 20 CV was applied.

4.5.3 Large-scale purification and characterisation of the phosphomimetic Spd-2 ASH2 domain

In order to produce enough protein for subsequent biophysical and structural characterisation, I have conducted a 4 l large-scale expression for HisSpd-2$^{806-1046}$ 10D/E using Tuner (DE3) cells in TB medium. Purification then proceeded according to the standard protocol, using Ni-NTA chromatography followed by size exclusion chromatography (**Appendix 8.1.7**; **Figure 4.23A**). SEC was carried out using an S200 16/60 column (GE Heathcare), and using WB as a running buffer. The SEC elution profile and the subsequent gel analysis indicated that full-length protein co-migrated with bacterial contaminants in multiple peaks (**Figure 4.23B,C**). In particular, the first peak was close to the void volume of the column (void volume of an S200 16/60 column= ~0.33 CV) and could represent protein aggregates. Note a 3 l large-scale purification was also carried out for HisSpd-2$^{803-1046}$ WT construct, and both constructs showed similar SEC profiles (**Figure 4.23B**). However, the yield of HisSpd-2$^{803-1046}$ 10D/E was higher than WT after SEC purification, and was thus carried on for subsequent experiments. To further purify the protein by ion exchange chromatography, SEC fractions within the third elution peak were pooled and diluted until the sample was in a low ionic strength buffer (50 mM Tris (pH 7.5), 50 mM NaCl). Anion exchange was carried out using two buffers, A (50 mM Tris (pH 7.5), 5 mM BME) and B (50 mM Tris (pH 7.5), 1 M NaCl, 5 mM BME). The column was equilibrated to 50 mM NaCl before loading the sample. A linear gradient of NaCl from 50 to 500 mM over 100 ml was then applied to elute the column. Subsequent gel analysis showed His6-Spd-2$^{803-1046}$ 10D/E eluted as a peak with a mid-point of ~270 mM NaCl (**Figure 4.23D,E**). However, a lower contaminant band was also observed in the same elution peak. A minor amount of protein further eluted at a higher ionic strength ~330 mM NaCl showing less contamination (**Figure 4.23D,E**).

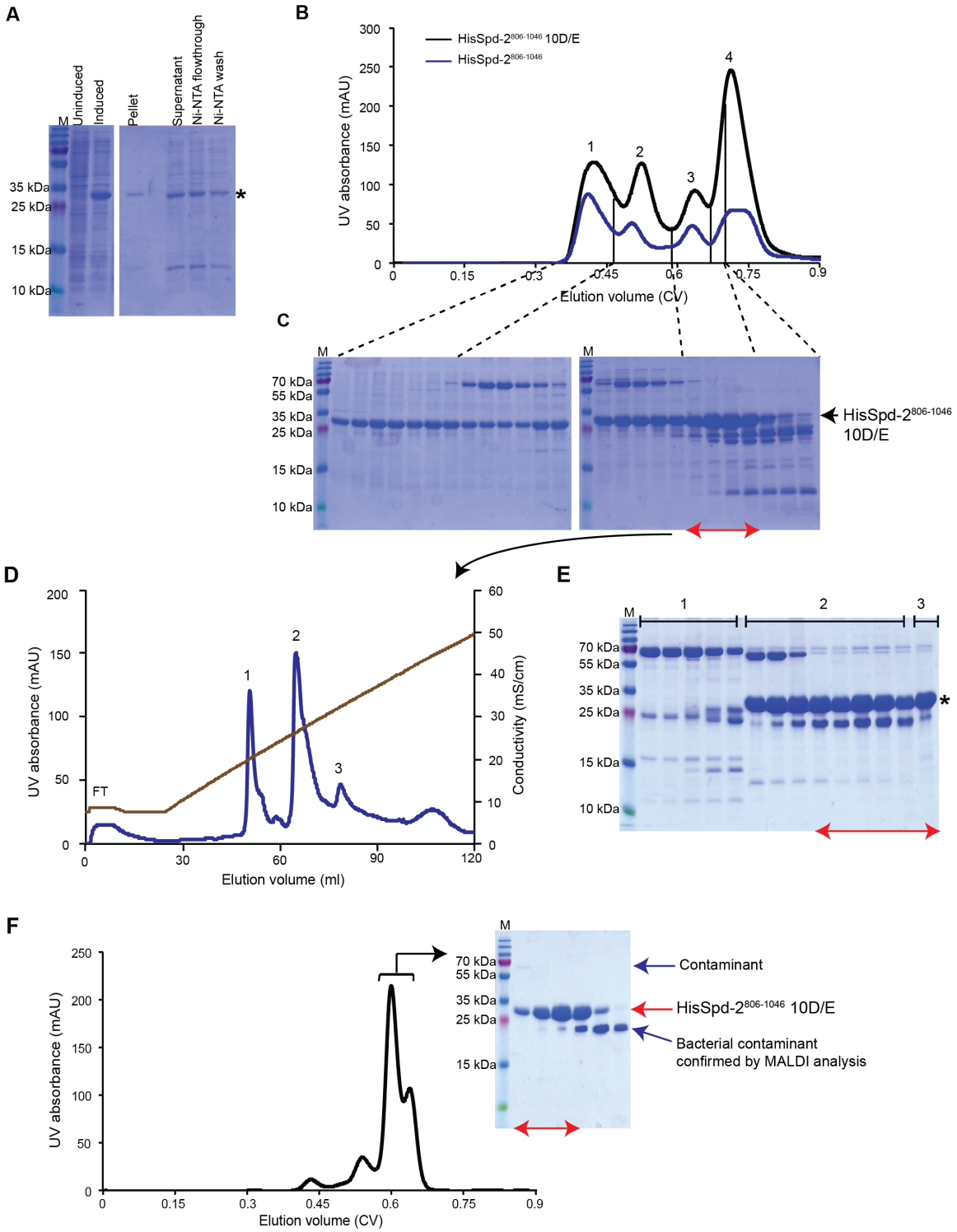


Figure 4.23 Purification of HisSpd-2$^{806-1046}$ 10D/E. (**A**) 15% reducing SDS-PAGE gel analysis of Ni-NTA purification fractions. Asterisk indicates the band corresponding to HisSpd-2$^{806-1046}$ 10D/E. (**B**) SEC elution profile for HisSpd-2$^{806-1046}$ WT and 10D/E from an S200 16/60 column. Elution with 50 mM Tris (pH 7.5), 150 mM NaCl and 5 mM BME. Note the two elution profiles (WT and 10D/E) look identical, but a higher yield of the 10D/E protein was produced; therefore only the mutant form was proceeded on for further experiments. Multiple elution peaks were observed, and gel samples were taken for each SEC peak. (**C**) 15% reducing SDS-PAGE gel analysis showing the elution of full-length protein across Peak1 with bacterial contaminants. Peak3 also contained HisSpd-2$^{806-1046}$ 10D/E, which co-migrated with lower molecular weight contaminants. SEC fractions within Peak3 were pooled for further purification by IEX. (**D,E**) Cation exchange helped to remove contaminants that were eluted at different ionic strengths, shown by SDS-PAGE analysis. Sample was loaded at 50 mM Tris (pH 7.5), 50 mM NaCl and 5 mM BME, and a linear gradient of NaCl from 50 to 500 mM over 100 ml was applied. Asterisk indicates the band corresponding to Spd-2$^{806-1046}$ 10D/E. (**F**) The pooled IEX fractions were loaded onto an S200 10/30 analytical column for buffer exchange into WB (50 mM Tris (pH 7.5), 150 mM NaCl and 5 mM BME). The sample appeared much cleaner than shown in **B**. The first three fractions were taken for subsequent experiments.

To exchange the IEX purified protein into WB, I concentrated the pooled fractions to 1 ml for loading onto an S200 10/30 GL column (GE Healthcare, **Figure 4.23F**). Gel analysis indicated that HisSpd-2$^{806-1046}$ 10D/E was much purer after this extra SEC step, and the lower molecular weight contaminant was eluted in a following peak that could then be removed. In order to check if the observed lower band was a degradation product from full-length protein, I have then cut out the gel band for MALDI mass spectrometry analysis at PNAC Facility University of Cambridge, which confirmed it was a bacterial contaminant.

The oligomerisation state of HisSpd-2$^{806-1046}$ 10D/E was investigated by SEC-MALS experiments. The final purified sample was prepared at 100 µM. A sample was also taken from the first elution peak of S200 16/60 SEC (**Figure 4.23B**). Both samples were injected onto an S200 10/30 column (GE Healthcare) that was equilibrated into WB. Protein fraction that was eluted close to the void volume of S200 16/60 column (GE Healthcare) showed the formation of multimeric species with a mass of 800 kDa (**Figure 4.24**). The other sample, further purified via IEX and analytical SEC, eluted much further into the column with mass estimation consistent of a monomeric species (**Figure 4.24**). A small preceding peak was also observed showing an average mass of a dimer. However, the molecular weight was highly consistent across the major elution peak on SEC-MALS profile with no obvious tendency to self-associate. As a result, the species eluted in the preceding peak might not be a dimer but a bacterial contaminant of similar size as seen in **Figure 4.23F**.

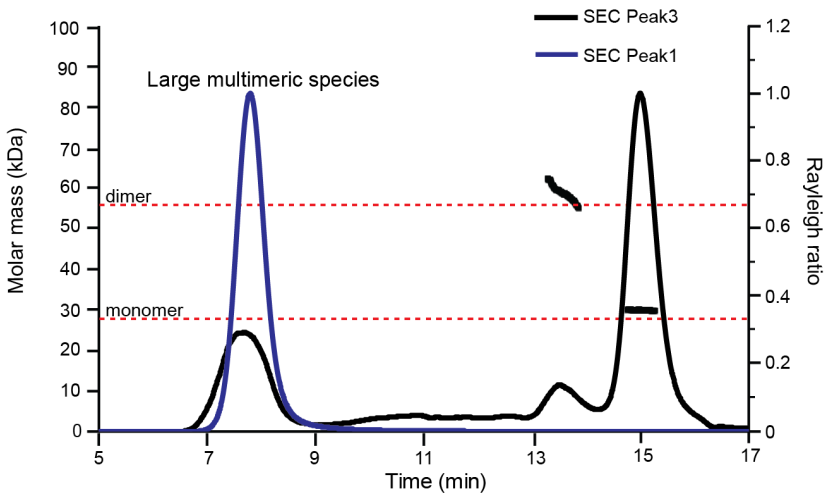


Figure 4.23 SEC-MALS analysis of HisSpd-2$^{806-1046}$ 10D/E at 100 µM. Elution with 50 mM Tris (pH 7.5), 150 mM NaCl and 5 mM BME. Solid lines indicate the relative Rayleigh ratio, and dash lines represent the observed molecular weight across the elution peak. Expected masses of a monomer and dimer are indicated with *red* dotted lines across the graph. The protein sample from Peak1 of S200 16/60 SEC (**Figure 4.23B**) eluted in the void peak. Protein taken from Peak3 and subsequently purified by IEX and analytical SEC (**Figure 4.23D-F**) eluted with an average mass of a monomer. Although a small preceding peak was observed containing species with mass of a dimer, a bacterial contaminant of similar size was observed on the gel as shown in **Figure 4.23F**. Additionally, the molecular weight was highly consistent across the major elution peak showing no tendency to further self-associate. Further experiments need to be followed in order to confirm whether the protein dimerises in solution (see text).

For further investigation, SEC-MALS experiment needs to be repeated at higher protein concentrations. If HisSpd-2$^{806-1046}$ 10D/E does form a dimer in solution, the amount of protein to be detected in the preceding peak would increase at a higher concentration; otherwise, the amount would stay the same. In addition, the upper contaminant band seen on the gel in **Figure 4.23F** should be identified by MALDI mass spectrometry analysis to show whether it is a bacterial contaminant or a disulfide-bonded dimer that was not fully reduced in the presence of 5 mM BME.

The secondary structure content of purified HisSpd-2$^{806-1046}$ 10D/E was then examined by CD experiments. The sample at 0.1 mg/ml was dialysed into 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.5) overnight at 4 °C. Four CD spectra were collected at 20 °C and normalised against a blank measurement taken with the dialysis buffer. The spectrum shows a negative trough around 214 nm indicating the protein is rich in β strands, which is consistent with the predicted secondary structure content (**Figure 4.24**).

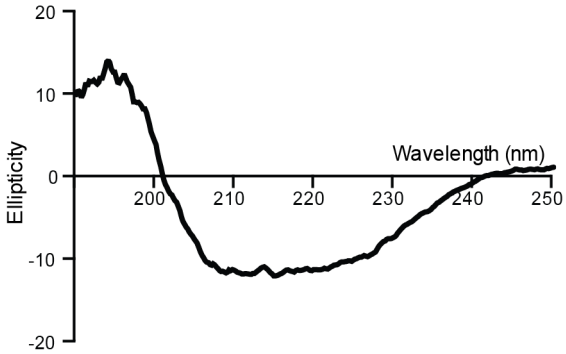


Figure 4.24 CD spectrum for HisSpd-2$^{806-1046}$ 10D/E carried out in at 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.5) 20 °C. Four scans were acquired and normalised with the dialysis buffer as a blank measurement. A positive peak is detected ~195 nm and a negative band is observed at 214 nm, both of which suggest the protein has secondary structure content and is rich in β strands.

Estimation of the final purified monomeric HisSpd-2$^{806-1046}$ 10D/E indicated a yield of 1 mg protein per 6 l culture. This low protein yield made crystallisation trials difficult, and hence no trials were performed.

4.6 Purification and characterisation of the Spd-2 ASH3 domain

4.6.1 Expression and purification of the Spd-2 ASH3 domain

As mentioned in **Chapter 4.2**, the Spd-2 ASH3 domain appeared to play a limiting role on the extent to which Spd-2 molecules can spread out in the PCM. The ASH1 and ASH3 domains were predicted to share a similar Ig-like fold, yet they seemed to have opposite functions in scaffold formation. It would be interesting to compare the structures of both domains and identify the potential difference between them. In addition, no structural insight was available from any Spd-2 homologs, and the fold was only predicted via bioinformatics tools (**Figure 4.1C**). As a result, I decided to characterise the ASH3 domain *in vitro*.

Based on the structure models built on the FFAS hits and the predicted secondary structural elements, a construct was designed covering residues 1047 to 1146 (**Figure 4.26A**). The PCR amplified construct was cloned into a pETM-14 vector by T4 ligation (NEB). The expressed fusion protein would have a 3C cleavable His6 tag at the N-terminus facilitating its purification by Ni-NTA chromatography. High-throughput purification trials were first performed to determine the optimal growth conditions and to assess the protein solubility (**Appendix 8.1.6**).

As shown in **Figure 4.26B**, protein was overexpressed, although not at a very high level, and a band at the expected molecular weight was detected in the Ni bead elution. A higher yield of soluble protein was obtained when B834 (DE3) cells were used for recombinant protein expression in LB medium. The construct details are summarised in **Table 4.6**.

Construct	MW (kDa)	pI	Abs 0.1%	No. of Cys
HisSpd-2$^{1047-1146}$	13.444	5.73	0.520	5
Spd-2$^{1047-1146}$	11.417	4.7	0.612	5

Table 4.6 Details of the ASH3 domain construct, calculated using ProtParam [Wilkins et al., 1999]. Due to the presence of cysteine residues, 5 mM β-mercaptoethanol (BME) was included in working buffer (WB) to prevent disulfide bonds.

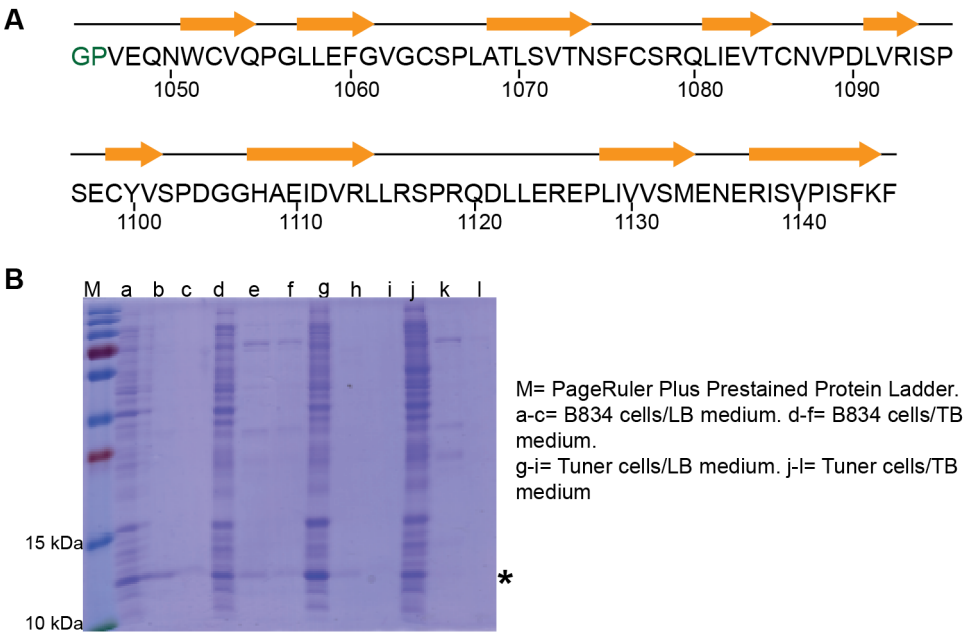


Figure 4.26 (**A**) Schematics of Spd-2$^{1047\text{-}1146}$. The predicted β strands (by PSIPRED) are indicated by *orange* arrows, and the protein sequence is shown underneath. The first two residues are from cleaved 3C recognition site and are coloured to *green*. (**B**) 15% reducing SDS-PAGE gel analysis of induced cell samples and elution from Ni-NTA beads. Gel sample was taken for protein eluted off the beads (lane b, e, h & k). Beads were also resuspended in 2x SDS loading dye for gel analysis to confirm protein was fully eluted with 300 mM imidazole (lane c, f, i & l). Asterisk indicates the band corresponding to HisSpd-2$^{1047\text{-}1146}$.

A 6 l large-scale expression was then carried out following the optimal growth conditions. Overexpressed fusion protein was purified using Ni-NTA chromatography, according to the standard procedure (**Appendix 8.1.7**; **Figure 4.27A**), and the eluted protein was subjected to tag cleavage with 3C protease. Digestion was carried out in which 1 mg of 3C was added per 1.48 µmol of target protein, and was dialysed overnight at 4 °C into WB. This reaction was concentrated to 5 ml (using a 3 kDa MWCO centrifugal concentrator (Amicon)) before loading onto an S75 16/60 column (GE Healthcare) for SEC, using WB as a running buffer (**Figure 4.27B**). Gel analysis showed the elution of pure full-length protein within the major peak (**Figure 4.27C**).

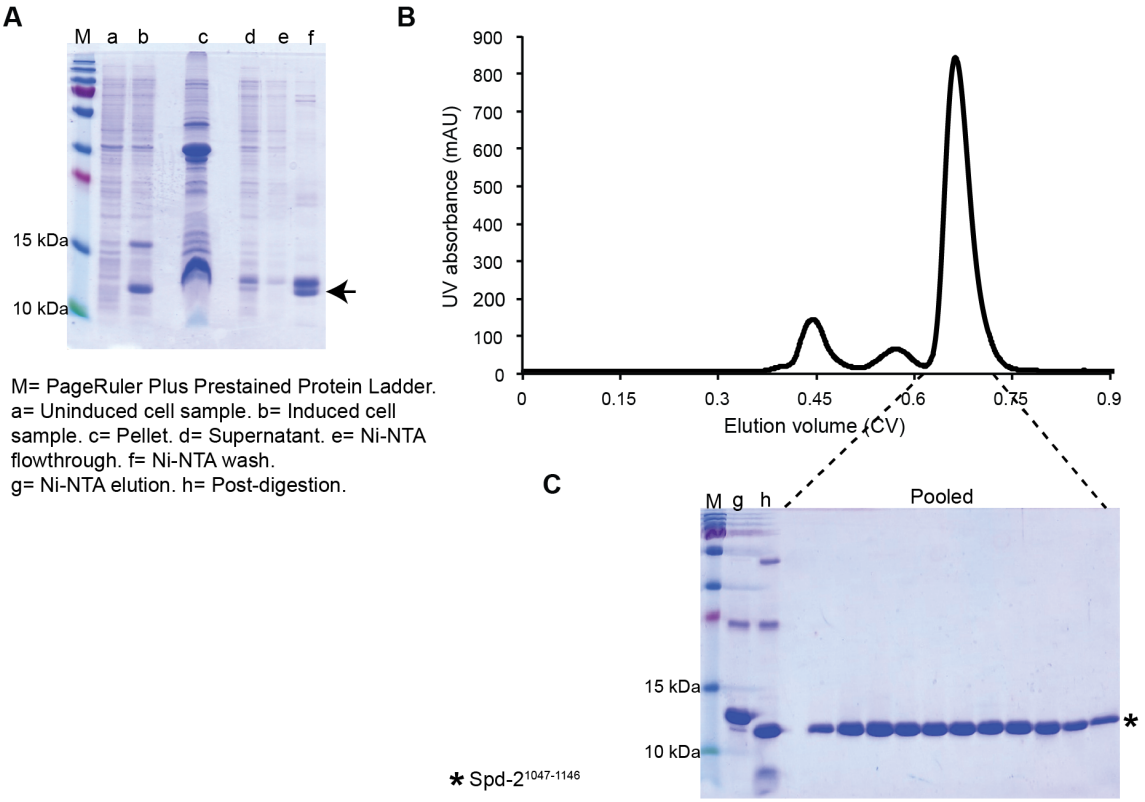


Figure 4.27 Purification of Spd-2$^{1047-1146}$. (**A**) 15% reducing SDS-PAGE gel analysis of Ni-NTA purification fractions. Arrow indicates the band corresponding to HisSpd-2$^{1047-1146}$. (**B**) Elution profile from an S75 16/60 column. Elution with 50 mM Tris (pH 7.5), 150 mM NaCl and 5 mM BME. (**C**) 15% reducing SDS-PAGE gel analysis of SEC fractions suggesting that Spd-2$^{1047-1146}$ eluted across the main peak was pure and homogenous, which were then pooled for further study.

4.6.2 Biophysical characterisation of the Spd-2 ASH3 domain

A SEC-MALS experiment was conducted to investigate the oligomerisation state of the Spd-2 ASH3 domain in solution. The sample was prepared at 88 µM and was injected onto an S75 10/30 column (GE Healthcare) that was equilibrated into WB. The protein eluted as a monomer (**Figure 4.28**).

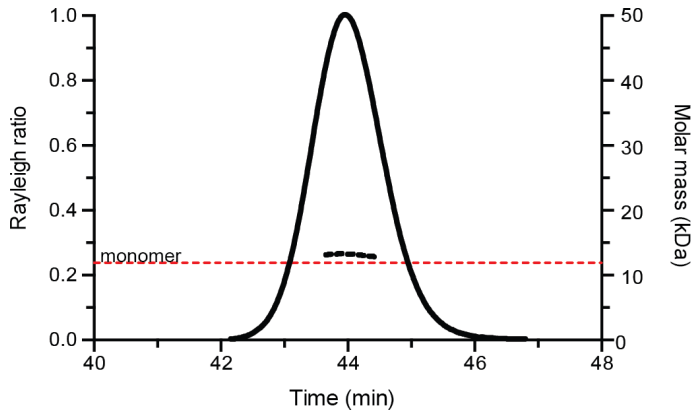


Figure 4.28 SEC-MALS analysis of Spd-2$^{1047-1146}$ showing the elution of a monomeric species. Elution with 50 mM Tris (pH 7.5), 150 mM NaCl and 5 mM BME. A solid line indicates the protein elution, and a dash line represents the observed molecular weight across the elution peak. The expected mass of a monomer is indicated with a *red* dotted line across the graph.

The secondary structural content of Spd-2$^{1047\text{-}1146}$ was analysed via CD experiments. Protein sample, at 0.1 mg/ml, was dialysed into 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.5) overnight at 4 °C. CD scans were then carried out at 20 °C with the measurement taken from dialysis buffer as a background subtraction. Four CD spectra were collected between 200 nm and 260 nm, showing a negative trough at ~214 nm characteristic of a β-type profile (**Figure 4.29A**). To further assess the stability of this structure fold, I performed a melting CD analysis by repeating spectrum collections at 2 °C temperature increments between 10 and 80 °C. Scans were performed between 200 and 240 nm. As the temperature increased, the negative trough characteristic of rich β strand content was shifted and the spectrum became one representing highly random coil content (**Figure 4.29B**). The ellipticity reading at 200 nm was then plotted against the temperature, and a cooperative transition was observed showing a T_m (mid-point of the unfolding transition) of ~52 °C (**Figure 4.29C**). This unfolding transition further indicated the protein had tertiary structural content.

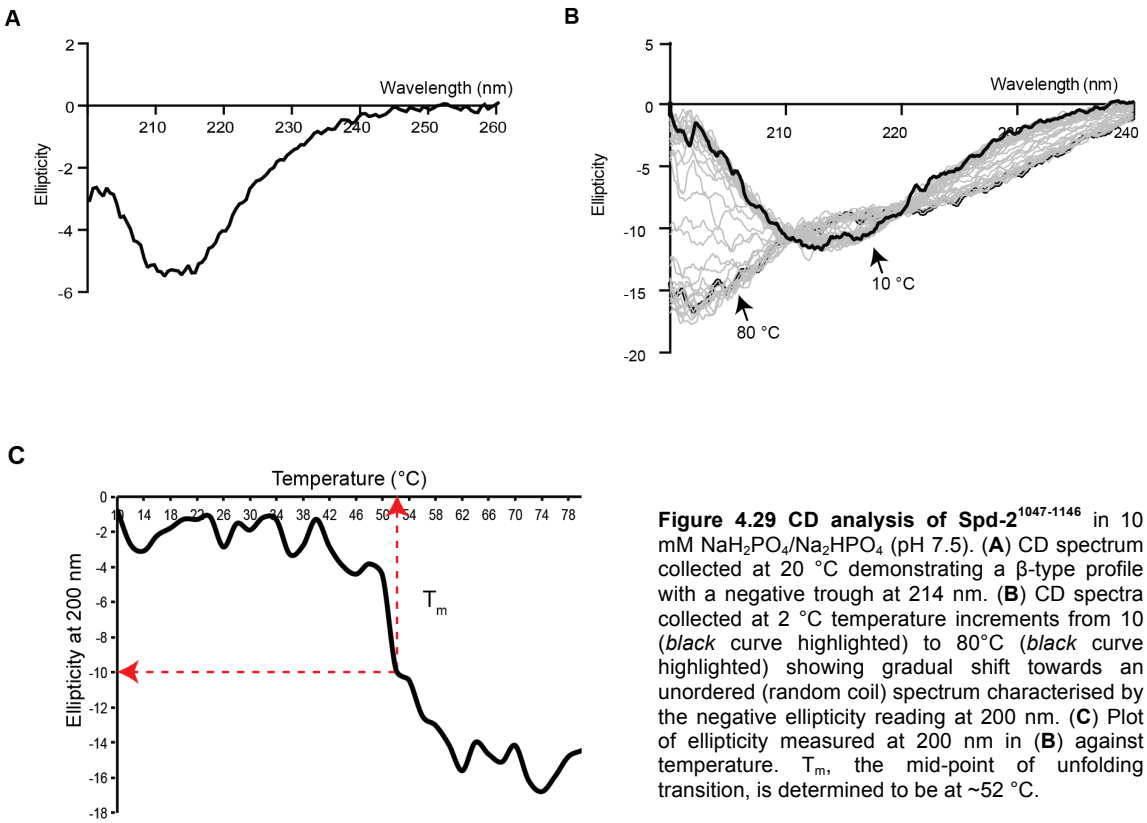


Figure 4.29 CD analysis of Spd-2$^{1047\text{-}1146}$ in 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.5). (**A**) CD spectrum collected at 20 °C demonstrating a β-type profile with a negative trough at 214 nm. (**B**) CD spectra collected at 2 °C temperature increments from 10 (*black* curve highlighted) to 80°C (*black* curve highlighted) showing gradual shift towards an unordered (random coil) spectrum characterised by the negative ellipticity reading at 200 nm. (**C**) Plot of ellipticity measured at 200 nm in (**B**) against temperature. T_m, the mid-point of unfolding transition, is determined to be at ~52 °C.

4.6.3 Analysis of the ASH1: ASH3 domain interaction

How might the ASH3 domain limit the extent of Spd-2 scaffold formation? Since ASH1 was found to be required for efficient scaffold assembly, I wondered if ASH3 might bind to ASH1 to inhibit its self-interaction or interaction with another PCM protein like Cnn, and so regulate the efficiency of scaffold assembly. To test this hypothesis, I analysed the potential ASH1: ASH3 domain interaction using NMR spectroscopy. NMR is highly sensitive in terms of its ability to study very weak protein-protein interactions. The ^{15}N-^{1}H HSQC spectrum of the labeled protein can be monitored when its unlabelled binding partner is titrated in, and any perturbation in chemical shifts is recorded. In the case of protein interaction, environment changes at the binding interface bring about perturbations in chemical shifts of the nuclei in this area. The HSQC spectrum of the ASH1 domain (Spd-2$^{697\text{-}805}$) was already assigned, and hence unlabelled Spd-2$^{1047\text{-}1146}$ was titrated in. This method would also allow the identification of any residues in ASH1 potentially involved in the interaction with ASH3.

Both protein samples were dialysed overnight into the same NMR buffer (50 mM NaH_2PO_4/Na_2HPO_4 (pH 7.0), 150 mM NaCl, 2 mM DTT) at 4 °C. A ^{15}N-^{1}H HSQC spectrum of Spd-2$^{697\text{-}805}$ was first collected to ensure any peak perturbation seen in the addition of Spd-2$^{1047\text{-}1146}$ was not due to changes in the buffer conditions. For NMR titration, a 350 µl reaction was set up containing ^{15}N-Spd-2$^{697\text{-}805}$ at 50 µM and Spd-2$^{1047\text{-}1146}$ at 1.5 mM. D_2O was added to a final concentration of 5% *v/v*. An HSQC spectrum was then acquired at 20 °C, and the addition of Spd-2$^{1047\text{-}1146}$ led to subtle changes in several peak positions (**Figure 4.30**). The chemical shift perturbation ($\Delta\omega$) was quantified for all peaks in order to identify the ones that were statistically significant. Calculation was carried out using the following formula,

$$\Delta\omega = \sqrt{\left(\left(\frac{\gamma_N}{\gamma_H}\right)\Delta\omega_N{}^2 + \Delta\omega_H{}^2\right)}$$

where $\Delta\omega_N$ is the change in chemical shift in ^{15}N dimension, $\Delta\omega_H$ is the change in chemical shift in ^{1}H dimension and $\frac{\gamma_N}{\gamma_H}$ is the ratio of gyromagnetic ratios.

The average of chemical shift perturbations was calculated to find the standard deviation that helps define the general shift alternation, and only those peaks with 1 standard deviation (s.t.d.)

from the mean were reported. However, no statistically significant shifts were observed in any residues implying no significant interaction between the ASH1 and ASH3 domains.

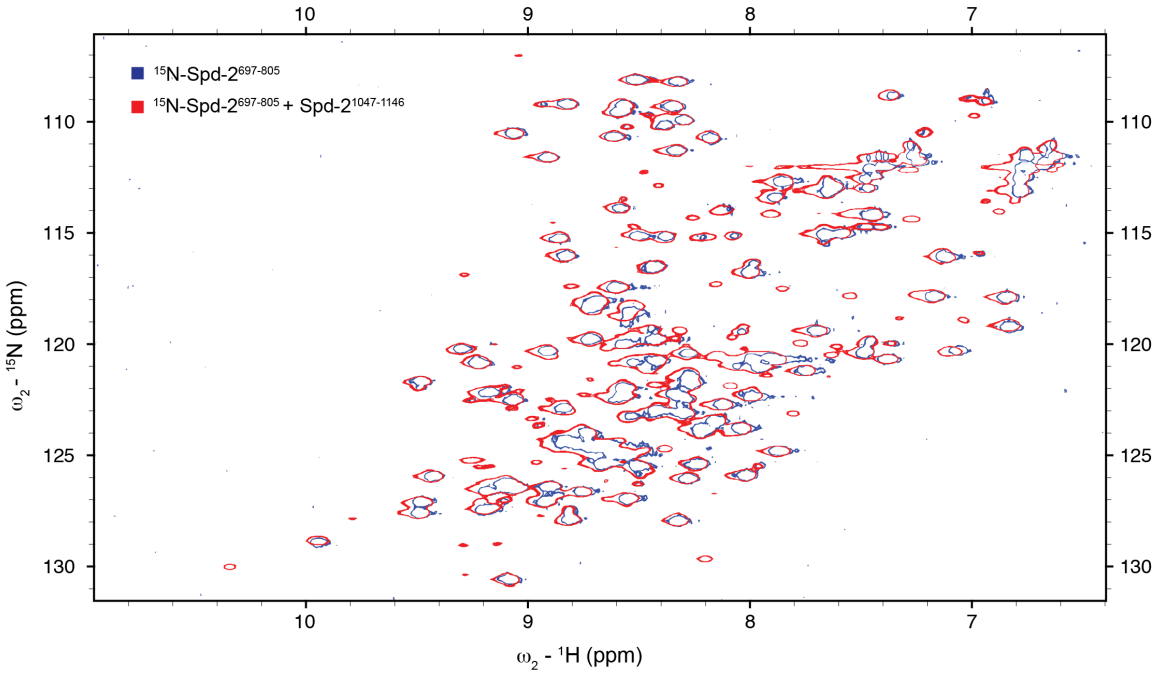


Figure 4.30 NMR study of ASH1: ASH3 domain interaction in solution. HSQC spectrum of ^{15}N-Spd-2$^{697-805}$ with or without the addition of 1.5 mM Spd-2$^{1047-1146}$ in 50 mM NaH_2PO_4/Na_2HPO_4 (pH 7.0), 150 mM NaCl, 2 mM DTT and 5% *v/v* D_2O.Note no significant shifts are observed in any residues.

4.7 Structure characterisation of the Spd-2 ASH3 domain

4.7.1 Crystallisation trials with the Spd-2 ASH3 domain

Spd-2$^{1047-1146}$ was purified to homogeneity after SEC, and it appeared to contain secondary and tertiary structural contents as suggested by CD experiments. I therefore set up crystallisation trials with this protein sample.

The protein was concentrated to 27 mg/ml using a 3 kDa MWCO centrifugal concentrator (Amicon), and was centrifuged at 15,000x g for 5 mins to remove any precipitate. I then conducted a 'Stura test' using the condition A7 from Stura Footprint Screen (20% PEG 4K, 0.2 M imidazole malate (pH 7.0); Molecular Dimensions) to determine whether 27 mg/ml would be a good starting concentration for crystallisation trials. 1 µl of protein stock was mixed with 1 µl of condition A7, and the appearance of precipitation was monitored under a microscope. An ideal concentration would show precipitation not immediately but not too slowly. As judged by the 'Stura test', this seemed a suitable concentration for initial crystallisation trials and therefore broad screens (Molecular Dimensions) were set up. Each drop was at 0.2 µl with a mother liquor: protein ratio of 50% or 75%. After 1-week incubation at 20 °C, most drops were clear and some showed phase separation, therefore the screening was repeated with a much higher protein concentration (52 mg/ml). This was close to the solubility limit of Spd-2$^{1047-1146}$, and a pellet of precipitate was observed when protein was concentrated to 58 mg/ml. Protein crystals grew after 1 week using the Stura/Macrosol screen (Molecular Dimensions), in 100 mM HEPES (pH 7.5), 2 M ammonium sulfate and 5% PEG 400 (**Figure 4.31A**). Multiple small crystals were observed in the drop, and were subsequently fished and snap frozen, with the mother liquor plus 20% *v/v* ethylene glycol for cryoprotection. In order to grow crystals of bigger sizes and in anticipation of the possible need to grow SeMet derived crystals for phasing purpose, optimisation screening was carried out with SeMet-Spd-2$^{1047-1146}$ (**Figure 4.31B**). Larger crystals were observed after 3 days (**Figure 4.31C**). Nevertheless, the native crystals diffracted to a high resolution, and the structure of Spd-2$^{1047-1146}$ was proven to be solvable by molecular replacement. Therefore, the SeMet crystals were not harvested for data collection.

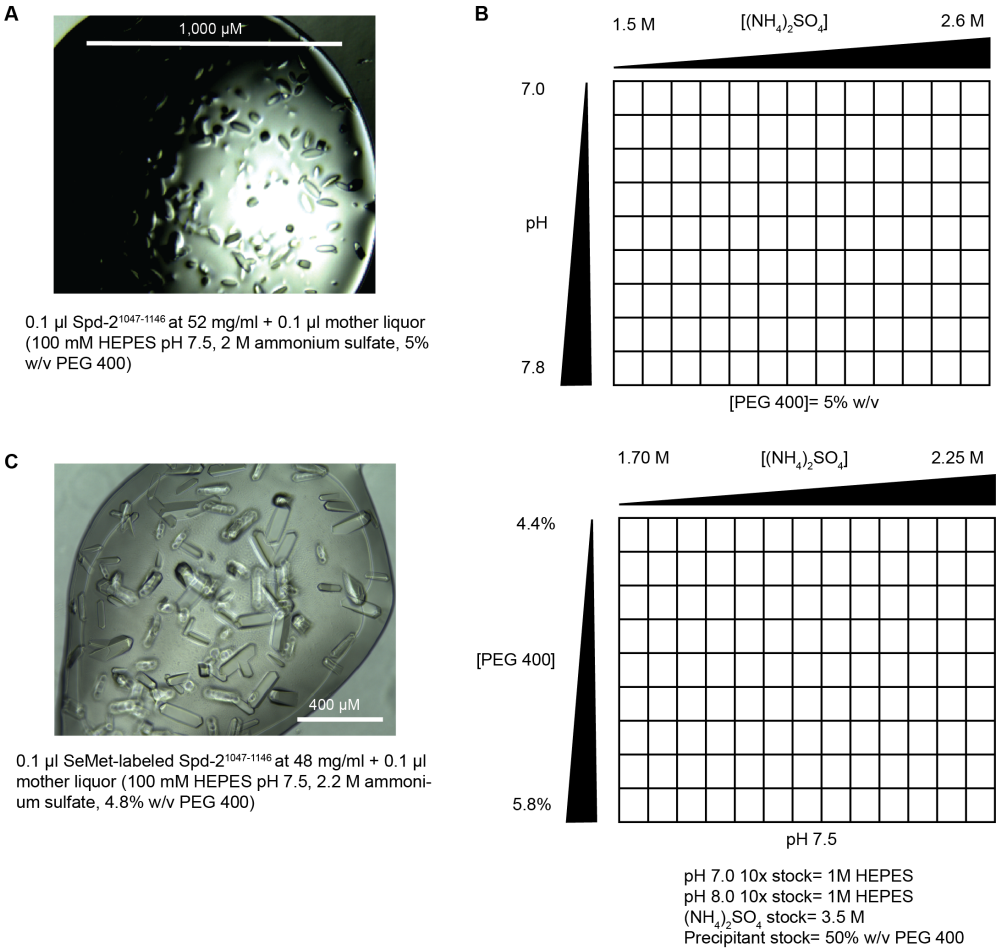


Figure 4.31 (**A**) Image of the initial crystal hit in Stura/Macrosol screen. (**B**) Optimisation screens were set up based on the condition in which the initial protein crystals were grown out. (**C**) Example of crystals obtained after the crystallisation conditions were optimised.

4.7.2 Data collection and processing

Data were collected by Paul Chappell at the I04-1 beamline at Diamond Light Source. Crystals were exposed to an X-ray beam of 0.92 Å wavelength. Three initial images were taken for the two harvested crystals, both of which showed diffraction patterns characteristic of a protein crystal. The initial images were indexed by Mosflm [Leslie and Powell, 2007] at the beamline, suggesting both crystals had I4 lattice symmetry. The data collection strategy was then automatically calculated by EDNA [Incardona et al., 2009] and Mosflm [Leslie and Powell, 2007]. Two full datasets were acquired by rotating the crystal through 180° with an exposure time of 0.2 sec and an oscillation angle of 0.20° (900 images).

Diffraction images were processed using Xia2 in the 3daii mode [Winter, 2010]. XDS [Kabsch, 2010] was used to index and integrate the reflections, from which the unit cell dimensions and possible space group were predicted and the intensity for each reflection was determined. The intensities of symmetry-related reflections were then scaled and averaged to give an I_{hkl} value for each unique reflection in process carried out by AIMLESS [Evans and Murshudov, 2013b]. This has produced two datasets both assigned to $I4_122$ space group with the same unit cell dimensions (**Table 4.7)**.

Dataset	**ASH3_1 (2.37 Å)**	**ASH3_2 (2.15 Å)**
Space group	$I4_122$	$I4_122$
Unit cell dimensions (Å)	a = 88.33, b = 88.33, c = 78.79	a = 88.68, b = 88.68, c = 78.99
Unit cell angles	α =β=γ= 90°	α =β=γ= 90°
Resolution (Å)	58.80-2.37	58.98-2.15
Number of unique observations (overall/inner/outer)	6529/96/423	8877/129/658
Completeness (overall/inner/outer)	98.8/99.4/89.7	99.9/99.6/99.0
Multiplicity (overall/inner/outer)	10.8/9.8/5.3	21.3/52.3/4.0
R_{merge} (overall/inner/outer)	0.102/0.049/0.642	0.101/0.032/0.801
R_{pim} (overall/inner/outer)	0.030/0.015/0.271	0.030/0.010/0.246
I/σI (overall/inner/outer)	16.8/34.9/2.4	21.3/52.3/4.0

Table 4.7 Statistics from processing of Spd-2$^{1047\text{-}1146}$ (ASH3) datasets.

According to the processing statistics, Dataset ASH3_2 provided a higher resolution (2.15 Å) with better data quality. The high-resolution shell data for Dataset ASH3_2 were still reliable judged by the high signal to noise ratio (I/σ= 4) and $CC_{1/2}$ value of 0.885. In addition, data within this shell showed nearly 100% completeness and a high multiplicity of 4. However, further analysis of the data quality revealed that some images showed more significant variations than the others, which was indicated by the higher R_{merge} values and were therefore excluded (**Figure 4.32**). Data processing was repeated using the first 300 and the last 150 images that showed lower R_{merge} values and hence higher internal consistency. Due to the high symmetry in the lattice, these 450 images (corresponding to 90° rotation of crystal) should still give sufficient multiplicity. Indeed, the new dataset (ASH3_3) demonstrated an overall multiplicity of nearly 6. Data were processed to a

high-resolution limit of 1.93 Å, and the statistics are summarised in **Table 4.8**. The values of I/σ and $CC_{1/2}$ were acceptable for the high-resolution shell data, and the data were still internally consistent as reflected by the R_{pim} value. Overall, the ASH3_3 dataset was suitable for further analysis.

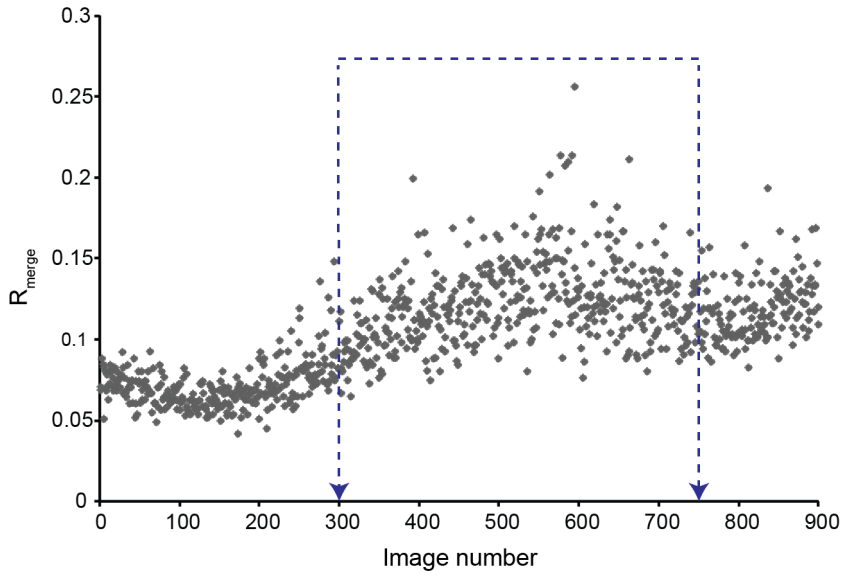


Figure 4.32 Analysis of data quality against all images. Reflections within images 301 to 749 showed more variation from the scaled intensity of a unique reflection. These images were then excluded in data processing. R_{merge} measures how consistent the redundant reflections are after scaling and merging. The last 150 images were also included in data processing for completeness (completeness overall/inner/outer= 96.7/89.1/99.1, if only the first 300 images were kept).

Dataset	**ASH3_3 (1.93 Å)**
Space group	$I4_122$
Unit cell dimensions (Å)	a = 88.58, b = 88.58, c = 79.00
Unit cell angles	α =β=γ= 90°
Resolution (Å)	62.64-1.93
Number of unique observations (overall/inner/outer)	12013/174/876
Completeness (overall/inner/outer)	99.1/99.9/99.4
Multiplicity (overall/inner/outer)	5.6/5.5/4.4
R_{merge} (overall/inner/outer)	0.081/0.039/0.566
R_{pim} (overall/inner/outer)	0.038/0.016/0.342
$CC_{1/2}$ (overall/inner/outer)	0.998/0.997/0.737
I/σI (overall/inner/outer)	15.1/39.2/2.4

Table 4.8 Statistics from processing of Spd-$2^{1047-1146}$ (ASH3) dataset with selected images.

In order to determine the number of copies of protein molecule in an asymmetric unit, the Matthews coefficient [Matthews, 1968] and solvent content were calculated in CCP4 program suite [Winn et al., 2011]. The solvent content of a crystal tends to range from 27 to 78% with the most common value being around 43% [Matthews, 1968]. Given the monomeric mass of Spd-2$^{1047\text{-}1146}$ and the unit cell dimensions, the Matthews coefficient was calculated to be 3.38 in the absence of non-crystallographic symmetry (NCS) and 1.69 if a dimer was present (**Table 4.9**). The latter gave a solvent content of ~27% around the edge of acceptance, whereas a solvent content of 64% was estimated for monomeric Spd-2$^{1047\text{-}1146}$ and hence was determined to be more likely.

Number of molecules per asymmetric unit	Matthews Coefficient	Solvent fraction	Probability
1	3.38	63.59%	0.96
2	1.69	27.19%	0.04

Table 4.9 Estimation of the number of copies of Spd-2$^{1047\text{-}1146}$ in the asymmetric unit. Note the probability shown above refers to all resolution ranges.

The NCS content was further confirmed by self-rotation function (SRF) analysis using MOLREP in CCP4 [Vagin and Teplyakov, 2010; Winn et al., 2011]. The chi (κ)= 180° section of SRF is shown in **Figure 4.33**, and no NCS was detected. The crystallographic 2-fold axes were observed along **a** and **b** directions, and a 4-fold axis along the **c** direction. There was no other peak with intensity above the noise that could be from non-crystallographic symmetry.

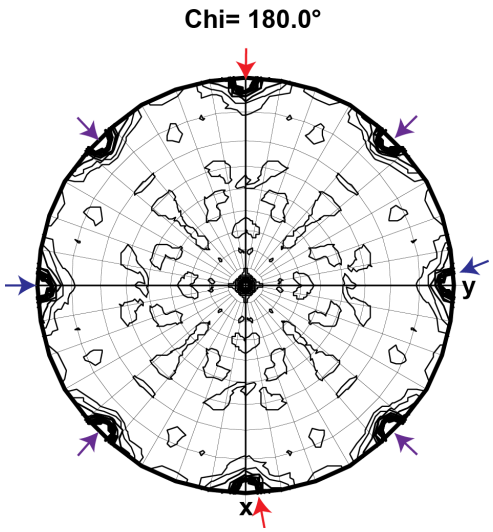


Figure 4.33 Self-rotation function analysis of Spd-2$^{1047\text{-}1146}$. Calculation was carried out in MOLREP using data between 2.14 and 62.64 Å. Note the high-resolution limit is usually adjusted to enhance the peak intensity. Peaks representing the crystallographic two-fold symmetry axes are highlighted by *red* and *blue* arrows respectively. The 4-fold axis along the **z**-direction interacts with the 2-folds generating additional 2-fold symmetries highlighted by *purple* arrows. All observed peaks could be explained by the crystallographic symmetry, therefore further supporting the presence of a monomer in the asymmetric unit.

4.7.3 Structure determination by molecular replacement

The dataset only contains the intensity for each reflection, and therefore the 'phase problem' needs to be solved in order to produce an electron density map and ultimately to build a structure. I obtained the structure solution for Spd-2$^{1047\text{-}1146}$ via molecular replacement (MR). Proteins that are homologous in amino acid sequence normally share a similar structural fold, as proposed by Max Perutz in the 1960s. In this way, a first model can be obtained if the structure of a homologous protein has already been determined. Although there is no structural information from Spd-2 homologs in any other species, I originally identified the ASH3 region as a domain by structure-based sequence analysis, which showed structural homology to MSP domain-containing proteins despite a low sequence identity (~15%). As a result, the top 3 FFAS hits, all with a score below -10, were used as search models, and only X-ray crystal structures were included. The sequence of Spd-2$^{1047\text{-}1146}$ was mapped onto the structures summarised in **Figure 4.34**, and the residues were trimmed to the last common atom using CHAINSAW [Stein, 2008]. In addition, the long flexible loops were manually removed in Pymol to further improve the search models. Note some proteins are dimeric, and individual chains were treated as an independent search model hence generating 5 search models overall.



PDB entry	Sequence identity
1M1S	16%
1ROW	15%
3IKK	12%

Figure 4.34 The aligned search models for Spd-2$^{1047\text{-}1146}$. The non-conserved residues were pruned to the last common atom, and the long flexible loops were trimmed. The following protein structures were modified to serve as the search models: WR4, a *C. elegans* major sperm protein (MSP) family member (PDB entry: 1M1S); chain A and chain B of MSP domain of the sperm-specific protein-19/SSP-19 (PDB entry: 1ROW); chain A and chain B of Human VAPB MSP domain (PDB entry: 3IKK). Sequence identity between the search model and Spd-2$^{1047\text{-}1146}$ is listed in the table, which also displays the colour of corresponding model.

MR was carried out with MOLREP in CCP4 [Vagin and Teplyakov, 2010; Winn et al., 2011]. The search model was rotated and then translated in the target unit cell, such that intra- and inter-molecular vectors respectively would align with the experimental Pattersons calculated from the measured reflection intensities (**Appendix 8.2.4**). All solutions are given a score based on the correlation coefficient between the observed intensities and those generated from the solution, and the packing function, which describes the steric overlap between molecules during packing. A 'contrast' value is also calculated, which compares a particular score to the average of other solutions. A contrast score greater than 2.5 is considered as a solution, whereas a value less than 1.5 is probably not a solution since it is not significantly different from the mean [Vagin and Teplyakov, 2010]. The top solution was obtained using chain A of PDB entry 3IKK (human VAPB MSP domain) with a contrast value of 1.71 and a score of 0.445 showing no molecular clashes. Solutions with similar packing were also found with other search models. As the solutions did not give very convincing scores, a polyAla model was generated where all residues were converted to Ala. This model was then used as a search model and produced a solution with a contrast score of 2.93, indicating a correct solution.

To further assess the solution, refinement was carried out using Refmac5 within CCP4 [Murshudov et al., 2011; Winn et al., 2011]. A rigid body refinement was first performed, in which the molecule as a whole was moved and rotated but the individual atoms were not. This was followed by restrained refinement, after which a big drop in R_{work} value to 35.52% was observed. In addition, positive residual densities in the Fo-Fc difference map became evident for the side chains (**Figure 4.35**), further suggesting the MR solution was correct.

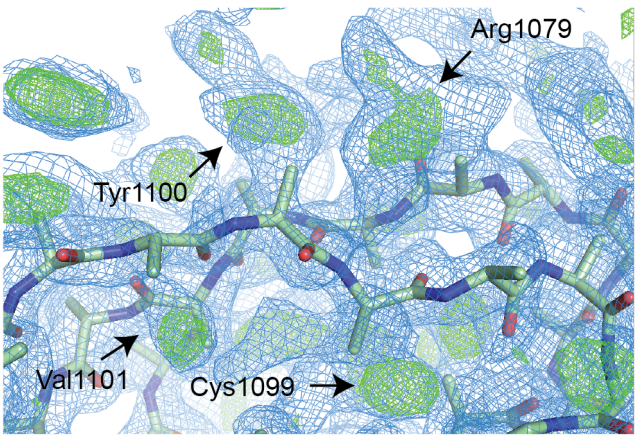


Figure 4.35 Initial refinement of Spd-2$^{1047-1146}$ structure after a solution was found by MOLREP. Map is viewed in Pymol, with the 2Fo-Fc map (*blue* density) contoured at 1.0σ and the Fo-Fc map (*green*= positive residual density; *red*= negative residual density) at 3.0σ. Pockets of residual density are evident into which the side chains can fit. Examples are highlighted with *black* arrows.

4.7.4 Refinement of the Spd-2$^{1047-1146}$ structure

After four rounds of restrained refinement in Refmac5, side chains were built into obvious densities (one example is shown in **Figure 4.36A**) and the polyAla model was partially mutated to the Spd-2$^{1047-1146}$ sequence. Buccaneer within CCP4 [Cowtan, 2006; Winn et al., 2011] was then used to autobuild model using the refined phases. Iterative rounds of refinement and rebuilding were carried out using Refmac5 [Murshudov et al., 2011] and Coot [Emsley et al., 2010]. In particular, the C-terminal strand, which was missing in the polyAla search model, was manually modelled based on the evident positive residual density (**Figure 4.36B**).

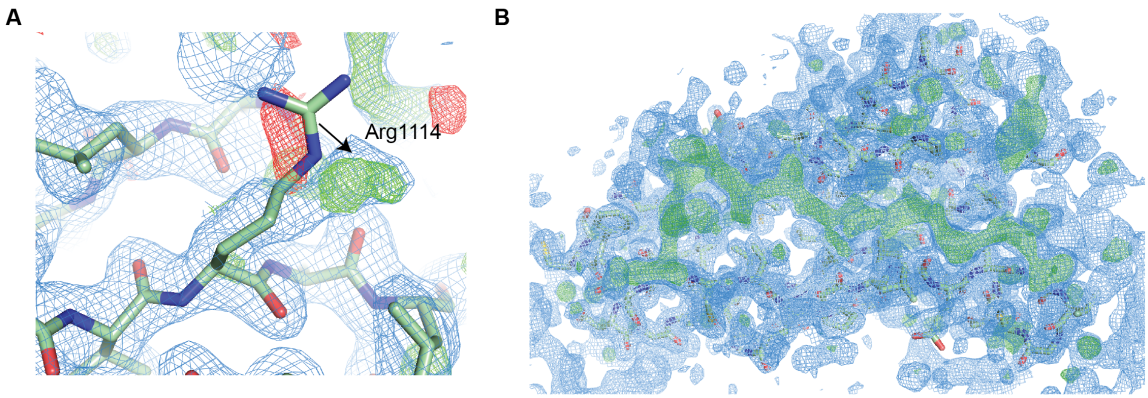


Figure 4.36 Refinement of the Spd-2$^{1047-1146}$ structure. View of maps resulting from four runs of refinement using Refmac5 in CCP4. The 2Fo-Fc map (blue density) is contoured at 1.0σ and the Fo-Fc map (*green*= positive residual density; *red*= negative residual density) at 3.0σ. (**A**) Both positive (*green*) and negative (*red*) residual densities are evident around the side chain of Arg1114 directing its correct orientation. (**B**) Positive residual densities appear for a C-terminal strand that is missing in the polyAla search model.

While both R_{work} and R_{free} values dropped during further refinements, a great gap was observed between the R factors (~6%). The much lower R_{work} value suggested the model might be overfitted. The refinement strategy was therefore optimised, in which the weighting of experimental data relative to the chemistry restraints was changed to 0.1 and subsequently 0.01. This has tightened the stereochemistry of the model, and the R factors started to converge. In some cases, split conformations of side chains were also observed as shown in **Figure 4.37B**. Refinement was continued until the model adequately describes the dataset, which was judged by the R_{work} and R_{free} values reaching 18.96% and 22.50%. The protein chemistry of the final model was validated using Molprobity, and the statistics are summarised in **Table 4.11**.

	Spd-2$^{1047-1146}$
Residue seen in structure	Gln1049-Phe1147
Resolution range	62.64-1.93
R_{work} (%)	18.96
R_{free} (%)	22.50
Clashscore	4.21 (98th percentile)
Side chain with poor rotamers (%)	0.00
Ramachandran outliers (%)	0.00
Ramachandran favoured (%)	99.01
MolProbity score	1.20 (99th percentile)

Table 4.11 Refinement statistics for the final model. R factors were calculated with a test size of 5%. Average B-factor was 27.284.

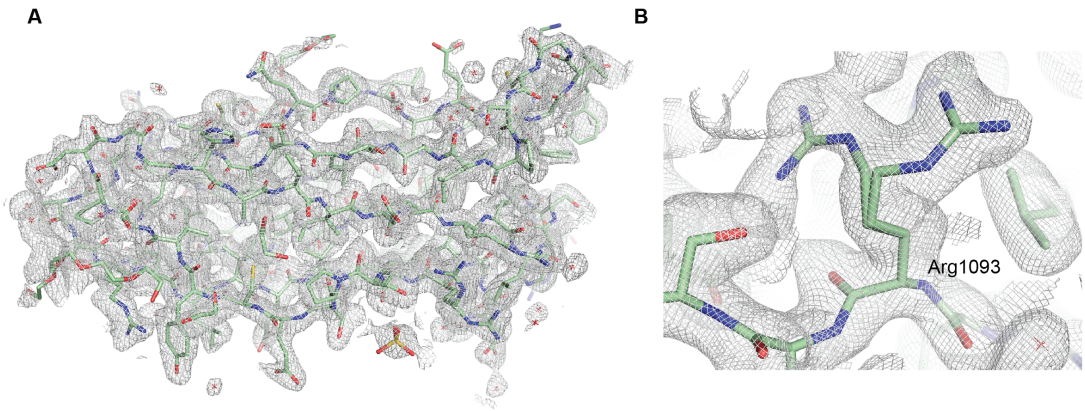


Figure 4.37 (**A**) Representative electron density for part of the refined model, contoured at 1.5σ. Water molecules were also modelled in the density map. (**B**) A close-up view of Arg1093 showing the static disorder that was modelled in with dual-conformation of the side chain. The 2Fo-Fc map was contoured at 1.0σ.

The final model comprises residues 1049 to 1146 with the N-terminal five residues missing, because no clear density is present in the region (**Figure 4.37A**). Note two of those missing residues are from the cleaved 3C recognition site and a Met residue is not part of the Spd-2 sequence. Apart from the polypeptide chain, 48 water molecules are fit into the density, so are three SO_4^{2-} groups and one PEG400 molecule that were present in the crystallisation condition. I have also manually modelled in two ethylene glycol molecules, which were added in the mother liquor as cryoprotectant for crystals during snap freezing.

4.8 Analysis of the Spd-2 ASH3 structure

4.8.1 The Spd-2 ASH3 domain adopts an immunoglobulin (Ig)-like fold

The Spd-2 ASH3 domain adopts an Ig-like β-sandwich fold that comprises opposing three-stranded and four-stranded β sheets (**Figure 4.38A**). The Ig superfamily represents one of the most abundant structure modules involved in a variety of cellular activities. In the Spd-2 ASH3 structure, seven antiparallel β strands are arranged into a Greek Key topology, in which the last strand is positioned adjacent to the first strand.

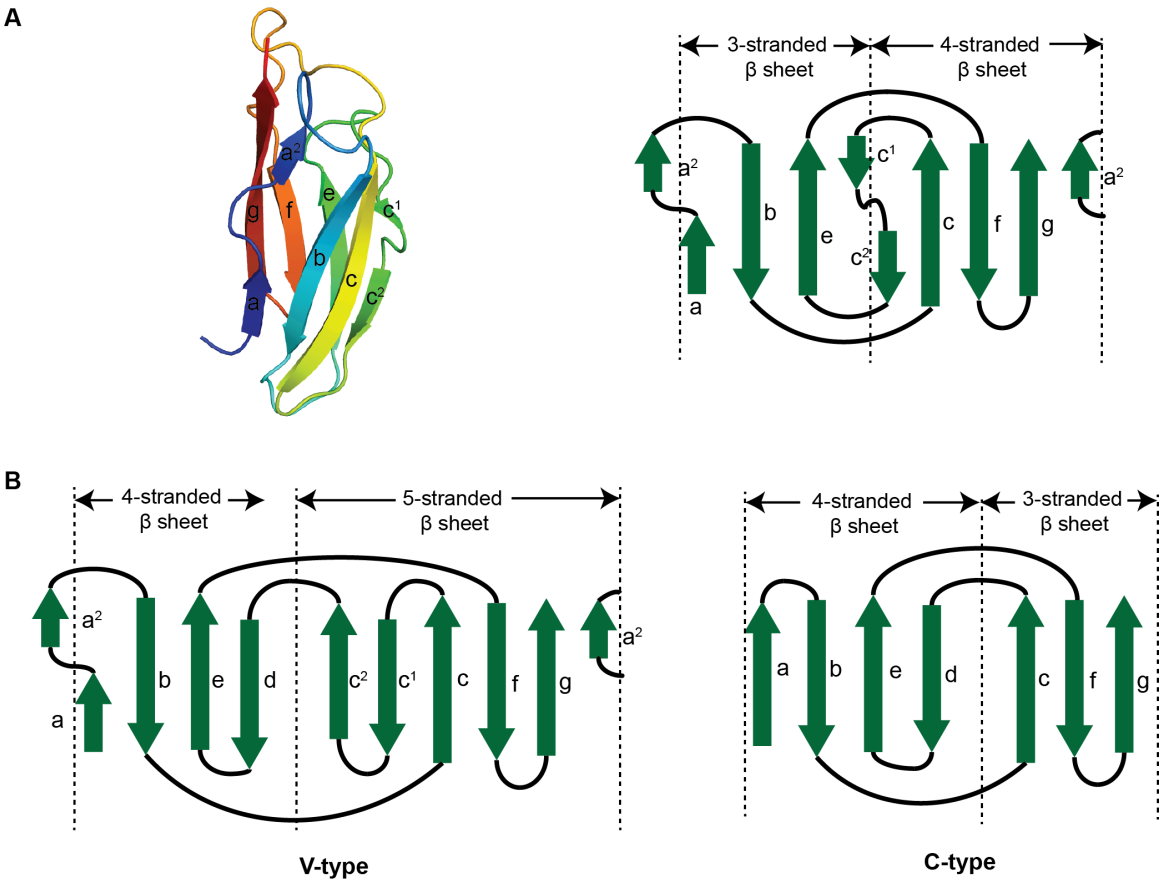


Figure 4.38 (**A**) The Spd-2 ASH3 domain shown in cartoon format with the displayed strand numbering. On the right is the 2D topology diagram according to the observed hydrogen bond pattern in the ASH3 structure. This is consistent with an S-type Ig fold, since the 4th strand is partially included in the 3-stranded sheet i.e. switching between the 2 β sheets. (**B**) 2D topology diagrams for 2 other types of Ig fold (V-set and C-set). Note the V-type domain has an extra β hairpin between c and d strands, which is missing in the C-type domain. In all 3 types, strand a shows antiparallel pairing with strand b; in addition, its parallel pairing with strand g is also detected in V- and S-type domains.

The Ig fold was first found in antibodies from which its name was defined [Hoffman et al., 2013]. Since an antibody has variable and constant domains, Ig domain has been initially classified into V-type and C-type [Bork et al., 1994; Clarke et al., 1999; Mitchell et al., 2015]. Later on, two more classes have been identified known as I-type and S-type [Bork et al., 1994; Clarke et al., 1999;

Mitchell et al., 2015]. V-type Ig folds contain a, b, e and d strands within one sheet, and a^2, g, f, c, c^1 and c^2 strands within the other sheet (**Figure 4.38B**). However, a^2 and c^2 strands are missing in the C-type Ig domain, and the I-type does not have a c^2 strand (**Figure 4.38B**). The S-type Ig fold, as seen in the case of the Spd-2 ASH3 structure, has no d strand, and its fourth strand ($c^1$and c^2) is partially included in the four-stranded β sheet rather than entirely in the three-stranded sheet as seen in the C-type.

Extensive hydrogen bonds are established between backbone amides and the carbonyl groups on adjacent chains to facillitate β strand-β strand pairing hence forming a stable β sheet (**Figure 4.39A**). In addition, the C-terminal half of the first strand is always directed slightly away from one sheet but towards the last strand in the other sheet to make parallel hydrogen bonding, subsequently linking the 2 sheets together (**Figure 4.38**; **Figure 4.39B**).

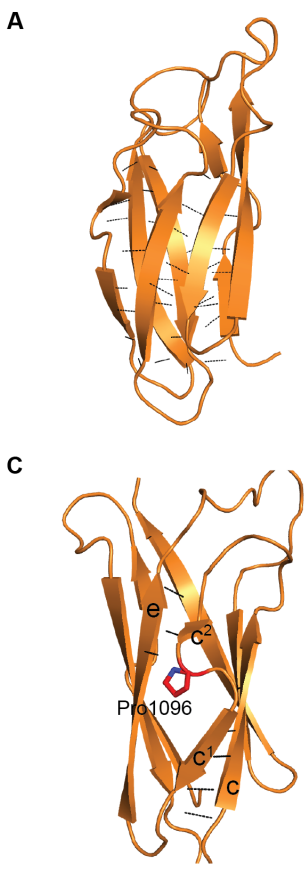


Figure 4.39 Intra-mainchain hydrogen bonding pattern observed in the Spd-2 ASH3 structure. (**A**) Conventional antiparallel β strand pairing via hydrogen bond interaction between the mainchain carbonyl group and the amide nitrogen, indicated by black dashed lines. (**B,C**) Side-on view of the two Pro residues that cause sharp kinks in the loop region between a and a^2 strands (**B**), and c^1 and c^2 strands (**C**). The presence of a Gly residue next to Pro introduces further destabilisation force within the a-a^2 loop. Such kinks have subsequently positioned the a^2 and c^2 strands for hydrogen bonding with strands g and e respectively. *Black* dashed lines highlight the antiparallel H-bonding. *Green* dashed lines represent the parallel H-bonding (**B**).

In this case, the c^1 strand (N-terminal half of the fourth strand that is part of the four-stranded sheet) is also involved in hydrogen bond interactions with the e strand that is part of the thee-stranded sheet, and therefore the fourth strand is considered to switch between the 2 sheets (**Figure 4.39C**). These sharp kinks are due to the presence of cis-Pro residue as highlighted in **Figure 4.39B,C**. Pro1055 and Pro1096 are likely important to stabilise the overall structural fold, since cis-Pro (at a^1-a^2/c^1-c^2

loop) to Ser mutation in the human VASP MSP domain (the homology model used for ASH3 structure solution) made the domain highly insoluble and unfolded [Shi et al., 2010].

4.8.2 Sequence and structure conservation of the Spd-2 ASH3 domain

The ASH3 domain is present in most Spd-2 family members, with the notable exception of worm SPD-2. I mapped the MSA for ASH3 onto its structure using Consurf [Glaser et al., 2003], (**Figure 4.40A,B**). Due to the alternating side chain orientation (successively pointing up and down) in antiparallel β sheet, it is clearly seen that the more conserved residues are buried in the core of the protein fold and the exposed residues are more variable in amino acid sequence (see, for example, magnified region in **Figure 4.40A**). There is no surface exposed region of particularly high sequence conservation, and the majority of the conserved residues appear to be structurally important. Chris Ponting carried out MSA using representative ASH domains among such domains identifiable at the time of publication, and he defined several consensus amino acids in each β strand [Ponting, 2006]. In the Spd-2 ASH3 domain structure, these amino acids are consistently involved in the hydrophobic core packing, and are presumably essential for the overall structural fold. Intriguingly, the invariant Asn residue at the C-terminal end of strand b, found in all ASH domains, is involved in extensive hydrogen bond interactions with the main chain carbonyl group and amide nitrogen at the tip of the domain (**Figure 4.40C**). This stabilises the conformation of b-c loop and c^2-e loop. Such interaction allows the divergence of side chains on residues in loop regions, which do not show high sequence conservation (**Figure 4.40C**).

However, Pro residues within loop region are highly conserved across evolution such as Pro1118 at e-f loop (**Figure 4.40B**; **Figure 4.41A**). Pro is frequently found in loops between β strands to accommodate turns in the sheet. Interestingly, the two cis-Pro residues, Pro1055 and Pro1096 mentioned above in **Section 4.8.1**, are consistently found in all fly species but not in the vertebrate homologs (**Figure 4.40B**). This suggests the kink and hence the S-type Ig fold might only be seen with ASH3 domains in flies. Further investigation into homologous system is required in order to understand the conservation of this particular type of Ig fold, which might highlight any potential divergence in evolution.

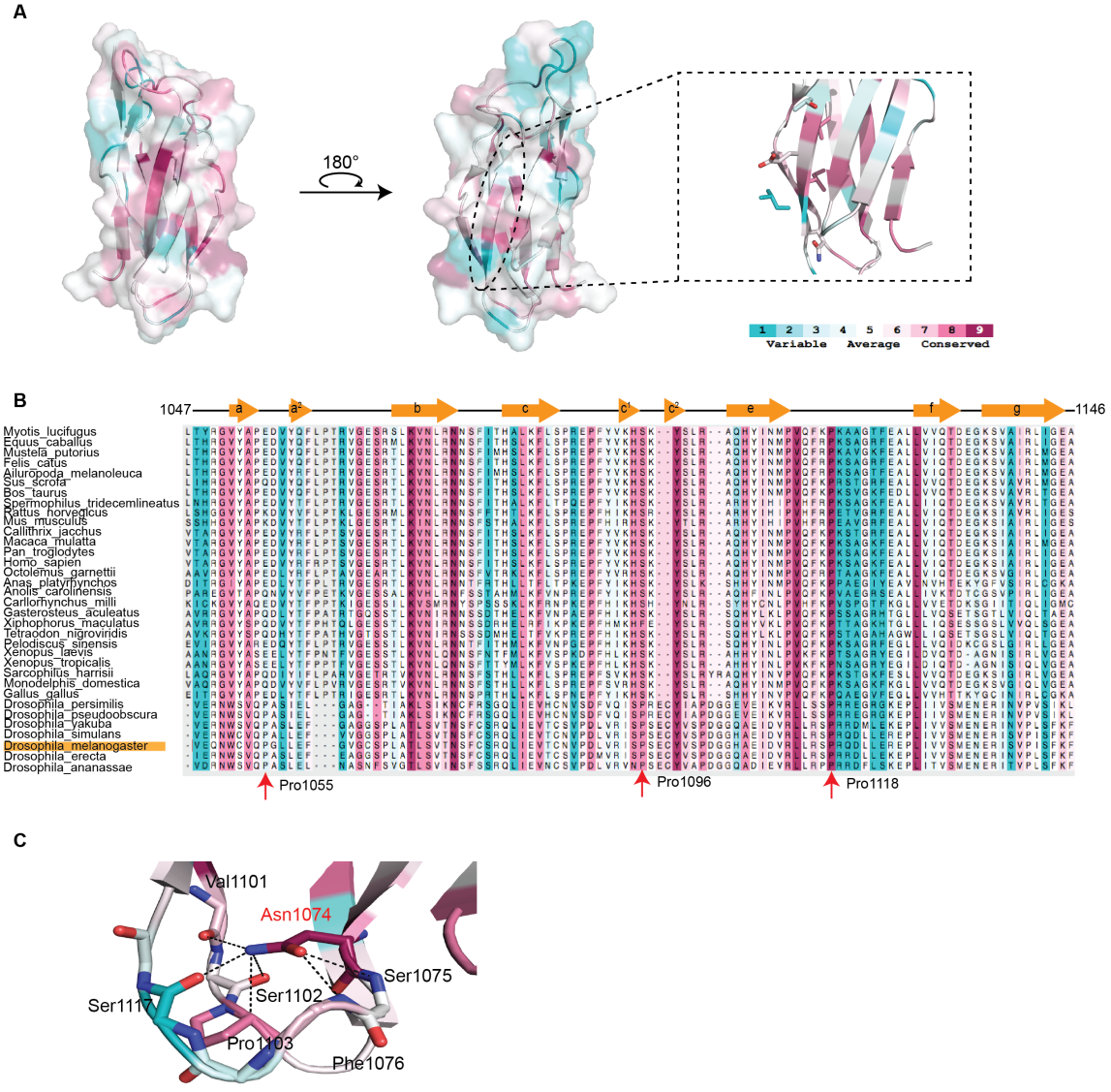


Figure 4.40 Analysis of the Spd-2 ASH3 domain conservation. (**A**) Spd-2 ASH3 domain structure in cartoon representation inside a translucent surface, both are coloured according to the Consurf conservation scale. The alternating side chain orientation is consistent with the sequence similarity, where the conserved residues tend to bury in the structure core and the surface exposed residues are more variable. One example is shown on the right with a close-up view of the strand c. (**B**) MSA output from Consurf, above which the assigned β strands in the Spd-2 ASH3 structure are displayed. Worms do not contain an ASH3 domain and hence are not included in the MSA. Note the a^2 and g strands are highly variable in amino acid sequence through evolution. The two cis-Pro residues at the kink of S-type Ig fold are highlighted with a *red* arrow, and both residues are only conserved in *Drosophila* homologs. (**C**) The well-conserved Asn1074 adopts a specific orientation at the tip of the ASH3 domain. Its side chain forms hydrogen bond interactions with the main chain of residues Ser1075, Phe1076, Val1101, Ser1102, Pro1103 and Ser1117 to stabilise the loop regions. Such interaction can tolerate divergence of the residue side chains. Indeed, these positions are not highly conserved in evolution.

Most of the solvent exposed residues are not evolutionarily conserved, but some do show conservation (**Figure 4.41A**). For instance, Thr1069 on strand b and Asp1111 on strand e are conserved, and are involved in an inter-strand hydrogen bond presumably providing stability for the β sheet (**Figure 4.41B**). A highly conserved hydrophobic residue, Leu1128, at the N-terminus of strand f is also solvent exposed (**Figure 4.41C**). In the Spd-2 ASH3 structure, this residue seems to pack against Pro1141 on the adjacent strand. Interestingly, another highly conserved

residue Tyr1100 is involved in cation-π interaction with the positively charged side chain of Arg1079 (**Figure 4.41C**). This has left the OH group at the tip of Tyr aromatic ring accessible to other proteins. Two Glu residues are also found in close vicinity, and might together be part of an interaction interface.

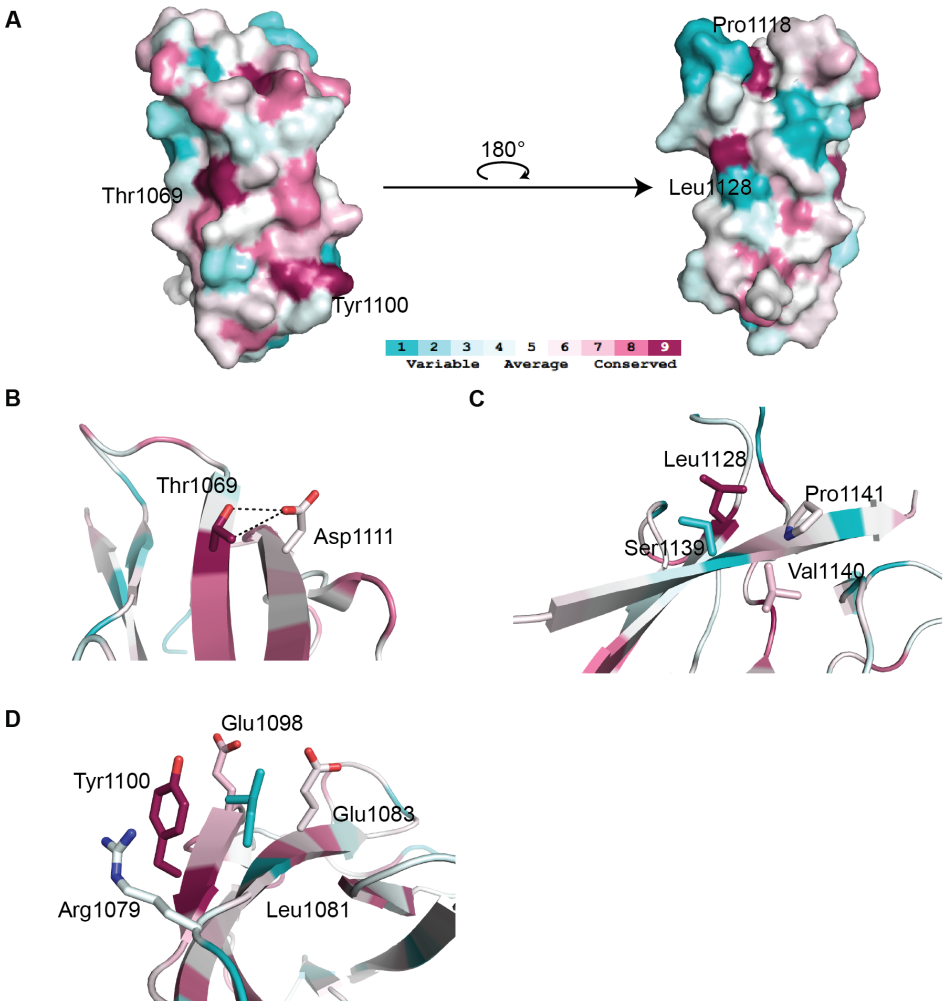


Figure 4.41 Sequence conservation analysis of surface exposed residues within the Spd-2 ASH3 structure. (**A**) Spd-2 ASH3 molecular surface coloured according to the Consurf conservation scale. The solvent exposed and evolutionarily conserved residues are highlighted, majority of which are structurally important. The conserved Pro residue is likely to be important for accommodating the turn between 2 strands. (**B**) Thr1069 forms an intra-chain hydrogen bond with Asp1111 on the adjacent strand. Asp1111 is substituted by an Asn residue in majority of the vertebrate sequences, which can also establish H-bonding with Thr hence linking the two strands. (**C**) The well-conserved Leu1128 packs against Pro1141. (**D**) Close-up view of cation-π interaction between Tyr and Arg. Tyr1100 also forms hydrophobic interaction with a cross chain Leu residue that is, however, highly variable in sequence. Note the aromatic OH group is not involved in any side chain interaction, which might be involved in interaction with other protein. There are two Glu residues that showed intermediate level of conservation and could be part of the binding interface.

4.8.3 Comparison to proteins with similar folds

The Spd-2 ASH3 domain was initially identified via FFAS search, and the majority of top hits contained MSP/MSP-like folds. Note that the MSP-like domain is shorter in length as it encompasses 107 to 109 amino acids, whereas the MSP domain normally contains 127 amino

acids. The ASH3 domain structure is solved by molecular replacement, using a MSP domain from Humans VAPB protein (PDB entry: 3IKK). This suggests that the two domains share great structural homology, even though the sequence identity is fairly low. The ASH3 domain is also structurally related to a PapD/PapD-like domain, as suggested by FFAS. Indeed, Bullock et al., recognised striking structural homology of *Ascaris suum* MSP domain, from its first crystal structure, with the N-terminal domain of bacterial chaperone protein PapD [Bullock et al., 1996].

Comparing the structure of the Spd-2 ASH3 domain with the three other domains mentioned above, great similarity was found in terms of the overall topology (**Figure 4.42**). In addition, I have also included the structure of OCRL (Lowe syndrome protein) ASH-RhoGAP domain (PDB entry: 3QIS) for fold comparison. Surprisingly this structure was not highlighted in FFAS search, despite it also contains an ASH domain. Superimposition of the Spd-2 ASH3 monomer onto the structure of SSP-19 (MSP-like domain), MSP, FimC (PapD-like domain) and ASH-RhoGAP yielded C^{α} RMSD value of 1.5033 Å over 85 C^{α} atoms, 1.4143 Å over 85 C^{α} atoms, 1.5008 Å over 83 C^{α} atoms and 1.9496 Å over 84 C^{α} atoms, respectively (**Figure 4.42**).

This high degree of structural similarity is referred primarily to residues within the strand region, whereas the loops do differ in length (highlighted in **Figure 4.42**). For instance, the 20-residue loop between f and g strands in MSP and SSP-19 is replaced by a 3-residue loop in Spd-2 ASH3 that does not contain a 3_{10} helix (highlighted with a *red* arrow in **Figure 4.42**). In all structure alignments, the a^2-b loop in Spd-2 ASH3 shows an outward curvature leading to an open ended conformation at one end of the domain fold, and this is not seen in other structures (highlighted with a *black* arrow in **Figure 4.42**). Further analysis of the amino acid sequence within this region shows the presence of two Gly residues surrounding a hydrophobic Val (Val1062), which is solvent-exposed (**Figure 4.43**). These Gly residues might confer high flexibility within the a^2-b loop, and Val1062 forms hydrophobic interaction with the long side chain arm of Lys1145 on the adjacent strand hence stabilising this splayed open conformation (**Figure 4.43**). Importantly, the loop regions within S-type Ig folds are often involved in protein binding and self-association. For instance, in OCRL protein the ASH and RhoGAP domains are joined by a hydrophobic interface involving a^2-b and e-f loops, which subsequently contribute to the stabilisation effect of one domain on the fold of the other [Pirruccello et al., 2010]. Interestingly, both loops show structural

divergence from the ones in the Spd-2 ASH3 domain. Therefore, differences in loop regions might provide functional specificity of S-type Ig folds in a diversity of proteins.

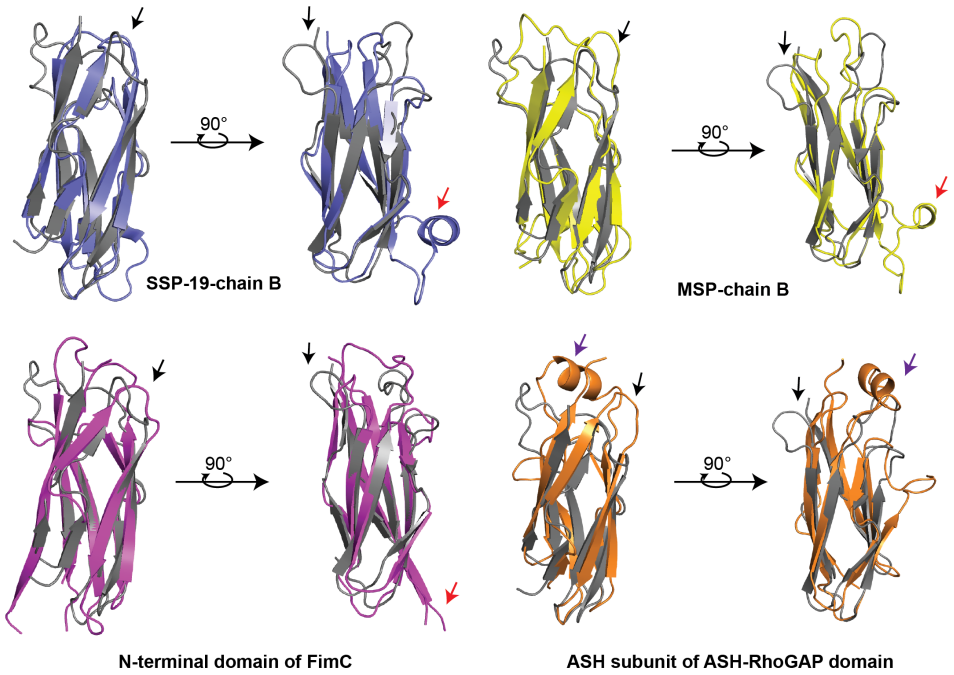


Figure 4.42 Comparison of the Spd-2 ASH3 domain structure to other protein domains (MSP, MPS-like, PapD-like and ASH-RhoGAP) with similar folds. The superimposition was carried out using Secondary Structure Match (SSM) in Superpose, CCP4 [Krissinel and Henrick, 2004; Winn et al., 2011]. The structures used here are from: SSP-19 (PDB entry: 1ROW), MSP (PDB entry: 1MSP), FimC (PDB entry: 3BWU) and ASH-RhoGAP (PDB entry: 3QIS). The highlighted loop regions are variable between structures. The f-g loop, highlighted with a *red* arrow, tends to be much longer in MSP/MSP-like and PapD domains; in some case, a 3_{10} helix is formed that can be involved in self-interaction or binding to other proteins. In the ASH-RhoGAP domain, the e-f loop, highlighted with a *purple* arrow, has a 3_{10} helix. The a^2-b loop in the Spd-2 ASH3 structure adopts an outward curvature, which is not seen in all other domain structures analysed here (highlighted with a *black* arrow).

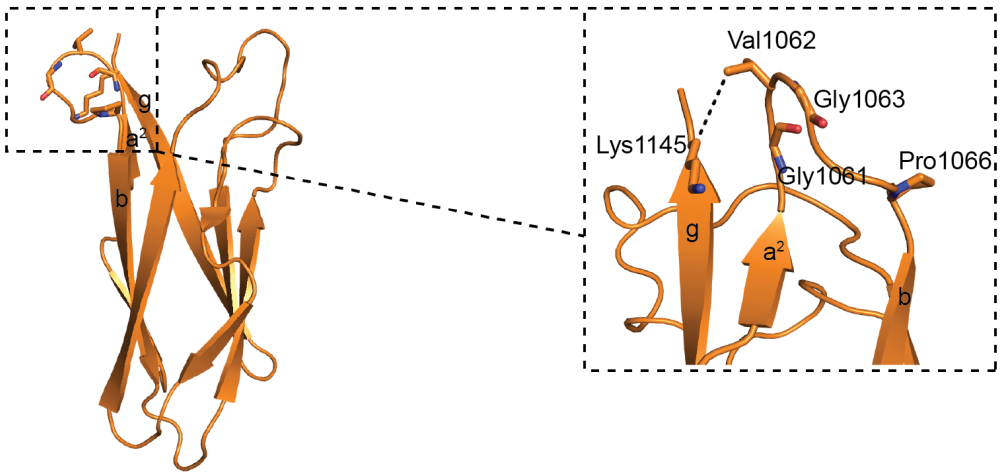


Figure 4.43 Spd-2 ASH3 structure showing outward curvature at one end of the fold. Left, the overall structural topology of the Spd-2 ASH3 domain. Right, focused view of the a^2-b loop showing an exposed Val residue surrounded by two Gly that might contribute to a high flexibility within the loop. Side chain of Val1062 forms hydrophobic interaction with the long side chain arm of Lys1145 on strand g. Note there is a Pro residue in the same loop region, which could also contribute to the high flexibility of the loop and hence the open conformation.

MSP proteins are known to form filament structures *in vivo* and the mechanism underlying this polymerisation process has been inferred from the crystal contacts [Baker et al., 2002], (**Figure 4.44A**). The g strand of MSP protein is involved in main chain hydrogen bond pairings to facilitate the self-association of dimers into subfilaments, and the lateral interactions between loop regions subsequently intertwine the 2 subfilaments into a filament [Baker et al., 2002]. Importantly, however, in the crystal lattice of Spd-2 ASH3, no such interaction is observed (**Figure 4.44B**). In fact, the monomers have rare contacts, and the top hit calculated by PISA server [Krissinel and Henrick, 2007] has a statistically insignificant CSS score of 0.005 (on the scale from 0 to 1 as interface relevance to complex formation increases; **Figure 4.44C**), suggesting no biologically relevant interface is present between the monomers. These results fit well with my analysis of the role of ASH3 in Spd-2 function in flies; although it is an intriguing possibility that Spd-2, like Cnn, has the ability to form a scaffold structure, ASH3 does not appear to be required for the formation of the Spd-2/Cnn scaffold (**Figure 4.6**).

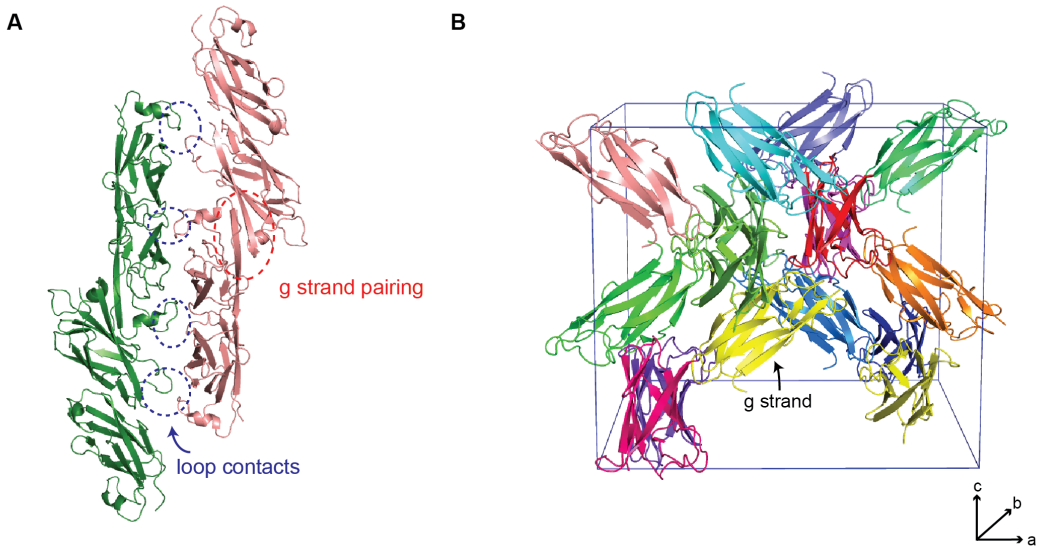


Figure 4.44. (**A**) Crystal structure of *C. elegans* MSP (PDB entry: 1GRW) showing the interfaces within and between the two adjacent helices (seen in the crystal packing). Within a helix, dimers associate via g strand pairing; between the two helices, loop regions show interactions. These crystal contacts are consistent with interaction sites present between subfilaments in MSP filaments (identified via Y2H screen) [Baker et al., 2002]. (**B**) Crystal packing in the Spd-2 ASH3 structure showing minimal contact between the monomers. The g strand (an example is highlighted) is not involved in any crystal contact.

4.9 Perspectives

At the time I started my studies, the importance of Spd-2 in PCM assembly in flies had been well recognised [Dix and Raff, 2007; Giansanti et al., 2008]. 3D super-resolution analysis (3D-SIM) of Spd-2-GFP at the centrosome had demonstrated its scaffold-like structure assembled around the mother centriole [Conduit et al., 2014b]. In contrast to Cnn, Spd-2 has two centrosomal fractions: centriole-associated and PCM-associated [Conduit et al., 2014b; Mennella et al., 2012]. Western blot analysis has shown that Spd-2 is specifically phosphorylated at the centrosome like Cnn. However, unlike Cnn, only a certain proportion of centrosomal Spd-2 molecules are phosphorylated. I started this project aiming to better understand how Spd-2 molecules contribute to the assembly of a PCM scaffold.

Bioinformatic analyses were first carried out to help identify evolutionarily conserved domains. Based on the MSA of full length Spd-2 and the FFAS search tool, I identified three ASH domains in *D. melanogaster* Spd-2 (a domain found in **A**SPM, **S**pd-2, **H**ydin, proteins; [Ponting, 2006]). The first two ASH domains (denoted as ASH1/2, Spd-2$^{697\text{-}805/820\text{-}942}$) are part of the highly conserved SPD-2 domain (Spd-2$^{707\text{-}928}$), while ASH3 (Spd-2$^{1047\text{-}1146}$) locates at the extreme C-terminus. A structurally important Asn residue was found at the end of a predicted β strand in all three ASH domains. ASH2 seems to differ from the structure of the other two ASH domains, because a short helix is predicted to form and it encompasses more residues (~120 amino acids instead of ~100 residues in ASH1/3). In the original paper identifying the ASH domain [Ponting, 2006], these differences meant that the ASH2 region was not identified as a potential ASH domain, although it is observed in all other Spd-2 homologs and it was predicted to resemble a MSP-domain like fold based on my FFAS search.

What is the function of these highly conserved ASH domain in Spd2/Cep192? The ASH domain has been identified in a diverse set of proteins that localise within three intracellular compartments, cilia, centrosome and Golgi apparatus [Ponting, 2006]. However, no common role has been determined for this domain. Yeast TRAPPII complex subunits are vesicle-tethering entities involved in Golgi transport, and all subunits contain a C-terminal ASH domain that is

required for centrosome targeting [Schou et al., 2014]. In another example, the ASH module is required to stabilise a folded domain as seen in the structure of OCRL, a Lowe syndrome protein [Pirruccello et al., 2010]. Therefore, it would be interesting to investigate the biological significance of the ASH domain in Spd-2. Using RNA injection approach with various domain-truncated mutants, I have shown a novel function of the ASH domain in centrosomal scaffold assembly. ASH1 and ASH2 are required for efficient scaffold formation, whereas ASH3 has little effect on the scaffold assembly and, if anything, it seems to play an inhibitory role. This result is consistent with the literature that the Spd-2 N-terminal half is involved in centriolar targeting via Asl interaction [Conduit et al., 2014b].

I attempted to identify the potential phosphorylation sites that might be involved in regulation of PCM scaffold assembly, since this normally only occurs at the centrosome. I identified three potential phosphorylated regions (PRs) that contained multiple Ser/Thr resiudes, at least some of which resembled CDK1 or PLK1 phosphorylation sites. Deletion of the first two PR regions, which contain centrosome-specific phosphorylation sites identified by MS/MS, had no influence on the Spd-2 scaffold formation. Removal of the third PR region, which is located between the ASH2 and ASH3 domains, showed a similar phenotype as deleting the ASH1/2 domain. However, a phosphomutant (10A) showed no defect. Because this region is close to the ASH2 domain, its deletion might have disrupted the protein domain fold. But due to the presence of endogenous protein, it still cannot be ruled out that these identified sites would be phosphorylated *in vivo*. At the time of writing, we have just obtained transgenic flies expressing a form of Spd-2 in which these 10 Ser/Thr residues in PR3 have been mutated to Ala. Preliminary experiments suggest that this protein was unable to rescue the female sterility associated with *spd-2* mutants. Clearly more work is required to assess the potential significance of this finding. In particular, I believe it remains an exciting possibility that, as is the case for Cnn, the ability of Spd-2 to help assemble a PCM scaffold will be regulated by phosphorylation.

How do Spd-2 molecules assemble into scaffold-like structures at the centrosome? Very recently, the lab has investigated the centrosomal localisation pattern of Spd-2-GFP in the absence of Cnn after colchicine treatment. Because Cnn is required for the maintenance of Spd-2 at the centrosome [Conduit et al., 2014b], which would otherwise be pulled away by MTs, embryos were

injected with colchicine to depolymerise the MTs and so allow Spd-2 to accumulate to a limited extent around mother centrioles even in the absence of Cnn. As MTs were depolymerised, small aggregates of Spd-2-GFP could be visualised around the mother centrioles, although the amount of scaffold formation was dramatically reduced compared to WT (unpublished data). This finding is difficult to interpret: clearly the ability of Spd-2 to form an expanded scaffold structure around the mother centriole relies heavily on Cnn, but it may be that Spd-2 can form a more limited scaffold structure without Cnn.

My data reveal that the ASH domains of Spd-2 play an important role in scaffold formation. This could be via promoting the ability of Spd-2 itself to form a scaffold, or by promoting the interaction between Spd-2 and Cnn, thus allowing Spd-2 to efficiently form a scaffold with Cnn. I wanted to characterise these domains *in vitro*, which would hopefully provide more insights into the *in vivo* mechanism. Both ASH1 and ASH3 could be purified from bacteria to near-homogeneity. Unfortunately, ASH1 did not crystallise, and in collaboration with Devon Sheppard I therefore used NMR spectroscopy for structure determination. Using X-ray crystallography, I solved structure for the ASH3 domain.

I attempted to purify the Spd-2 ASH2 domain. Unfortunately, the majority of constructs tested were insoluble, and despite repeated attempts, I was unsuccessful. It is possible that the other two ASH domains in Spd-2 are required to stabilise the fold of the ASH2 domain. However, purification of constructs covering ASH1-2, ASH2-3 and ASH1-3 has not yielded any soluble protein. I then cloned a longer fragment that covers the entire region between the C-terminus of ASH1 and the N-terminus of ASH3 (Spd-$2^{806-1046}$). Soluble protein was yielded and additionally, the phosphomimetic mutations further helped the expression of soluble HisSpd-$2^{806-1046}$ 10D/E. This longer fragment was further purified and characterised on a large-scale. The protein was largely aggregated, as indicated by SEC-MALS result and the incomplete His tag cleavage. A small amount of purified sample was monomeric and appeared to have secondary structural content, but its yield was insufficient for crystallisation trials. Further experiments, such as ThermoFluor assay, could be carried out to optimise the buffer conditions and to assess if the multimerisation state could be shifted towards monomeric. A SCWRL homology model for Spd-$2^{800-1046}$ constructed using its sequence alignment with MSP domain suggests that the C-terminal

120 residues are not included in the structure fold. However, any shorter constructs designed on the basis of secondary and tertiary structural predictions (PSIPRED and FFAS) could not be expressed as soluble protein. Limited proteolysis should be carried out on HisSpd-2$^{806-1046}$ 10D/E to help determine if a core structure fold of shorter length is present.

The observation of monomeric Spd-2 ASH domains in solution is very interesting, despite Spd-2 could form scaffold-like structures at the centrosome and Ig-like folds have been implicated in protein multimerisation. Therefore, any intrinsic self-assembly seems to be very low affinity. It is plausible that multimerisation can occur when phosphorylation takes place and subsequently triggers the formation of Spd-2 scaffolds at the PCM, a mechanism similar to Cnn scaffold assembly (**Chapter 2**).

Nonetheless, *in vivo* Spd-2 does rely on Cnn for the efficient formation of an expanded scaffold structure around the mother centriole, suggesting it might not be a simple multimerisation process underlying the assembly of Spd-2 scaffolds. It is plausible that the ASH domain is not only involved in oligomerisation but also binds to other PCM components like Cnn, in the process of building a scaffold-like structure. As outlined in **Chapter 3**, the CM2 domain of Cnn is required for centrosomal localisation. Spd-2 has been referenced as a potential binding site for Cnn at the centrosome [Conduit et al., 2010; Conduit et al., 2014b]. I therefore decided to probe for any potential interaction between the Spd-2 ASH domain and the Cnn-CM2 domain *in vitro*, which is described in the next chapter.

Chapter 5: Interaction study between the 2 conserved domains in Cnn and Spd-2

5.0 Introduction

5.0.1 Evidence for Cnn: Spd-2 interaction

Evidence from *in vivo* experiments suggested possible interactions between the conserved domains in Cnn and Spd-2 (**Chapter 2, Chapter 3 and Chapter 4**). In particular, the Spd-2 scaffold did not efficiently assemble if the Cnn scaffold was not properly assembled due to mutations in the PReM domain (**Chapter 2.5**). Additionally, in the absence of Cnn, the amount of Spd-2 scaffold formation was dramatically reduced, even after colchicine treatment (**Chapter 3.7 and Chapter 4.9**). These results suggest that Spd-2 at least partially relies on Cnn in order to efficiently expand into the PCM region. The ASH1 and ASH2 domains in Spd-2 appeared to be necessary for its scaffold assembly (**Chapter 4.2**). As a result, I suspected that ASH1 and ASH2 might be involved in a Cnn interaction. In addition, the evolutionarily conserved CM2 domain within Cnn was shown essential for its centrosomal localisation (**Chapter 3.7.2**). Previous studies in the lab have demonstrated that in the absence of Spd-2 the amount of Cnn localisation at centrosomes was greatly diminished, and anti-Spd-2 antibodies strongly suppress Cnn recruitment to centrosomes in embryos [Conduit et al., 2010; Dix and Raff, 2007]. Therefore, the Cnn-CM2 domain might be a potential binding partner for the Spd-2 ASH1 and ASH2 domains.

5.0.2 Aims

- Investigate whether an interaction is present between the conserved domains in Cnn and Spd-2
- Characterise the importance of this interaction, if any, *in vivo*

5.0.3 Key findings

Preliminary Microscale Thermophoresis (MST) and NMR studies identify a low affinity protein interaction between the Spd-2 ASH1 and Cnn-CM2 domains. The Cnn T1133E mutation, which is at the CM2 tetramer interface *in crystallo* (**Chapter 3**), does not influence CM2 binding to the ASH1 domain *in vitro*. Residues at the interaction interface are mapped onto the ASH1 homology model (based on the FFAS hits) by NMR, all of which appear to gather on one side of the Ig-like fold. Mutation clusters were designed based on the crystal structure of Cnn-CM2, in order to probe the biological significance of this putative interaction.

5.1 Probing a putative interaction between the ASH1 and CM2 domains in solution

5.1.1 Interaction characterisation via MST measurements

Microscale Thermophoresis (MST) measurement is based on the effect of Thermophoresis when a temperature gradient is applied. A temperature gradient will not only induce a flow of heat but also a flow of molecules in aqueous solution [Moran et al., 2011]. Measurement is carried out using a NanoTemper Monolith NT.115. In general, an experiment is set up by heating the sample with an Infrared (IR)-laser that is focussed on one area of the solution, from which the fluorescence is also monitored [Moran et al., 2011]. A serial dilution that contains 16 different samples are then analysed and compared through the same local area. A typical set of MST signal curves are shown in **Figure 5.1**.

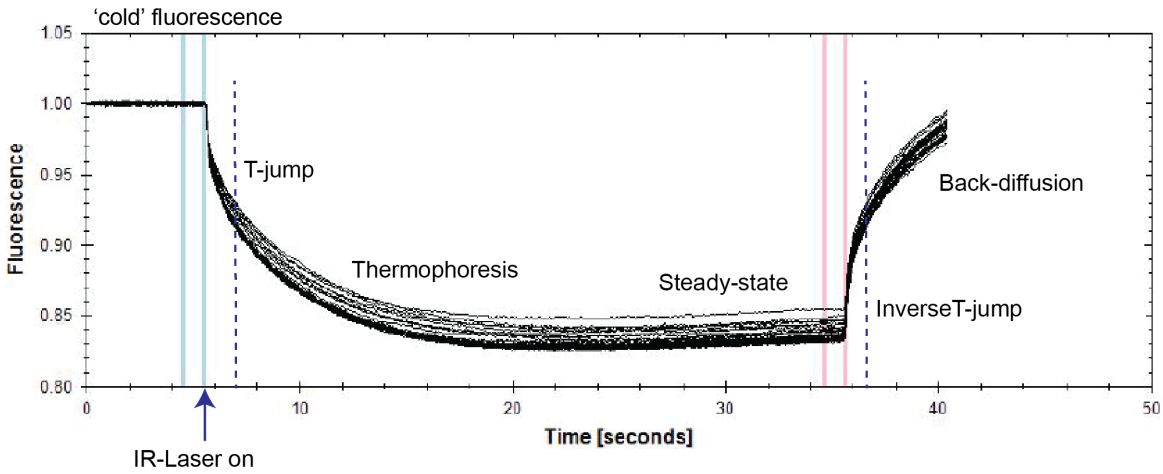


Figure 5.1 Example of typical MST curves seen with labelled Spd-2$^{697-805}$ upon the titration of Cnn$^{1082-1148_noTEV}$. Temperature-jump (T-jump) occurs immediately after the heating is turned on, and there is no mass flow during T-jump. Molecules then start to move away from the heat spot, in a process known as Thermophoresis, until a steady state is reached where no further molecular diffusion will take place. The laser is turned on for 30 sec at 20% or 40% laser power. Once the heating is off, molecules diffuse back due to the gradient established during Thermophoresis. The normalised fluorescence of unbound labelled molecules ('cold' fluorescence) and complexes are compared, and the difference is measured in T-jump and Themorphoresis. These differences are then plotted against the titrant concentration, from which the dissociation constant can be calculated.

Initially, the sample molecules are homogenously distributed in solution and show a constant fluorescence (also known as 'cold' fluorescence). The IR-Laser is then turned on, and this induces a rapid temperature-jump (T-jump) in the local solution volume. Whereas T-jump occurs immediately upon heating, thermophoretic motion of molecules is only seen after 10 sec [Moran et al., 2011]. Molecules in the heated spot tend to migrate away into the surrounding area until a

steady state is reached, where there is no net movement of molecules into or out of the monitored region [Christoph et al., 2010; Moran et al., 2011]. This creates a mass flow gradient. As a result, when the laser is turned off, molecules diffuse back in a process known as 'back diffusion'. Because this process is purely driven by mass diffusion, the molecular size can be deduced as bigger protein complexes tend to move more slowly. Additionally, when a fluorophore is located close to the interaction interface, its property and subsequently the fluorescence will change upon binding [Moran et al., 2011]. Therefore, the fluorescence change upon T-jump may also provide useful information about any binding event.

I first carried out an MST experiment using the primary amine labelled $Cnn^{1082-1148}$. Dye labelling was performed according to the standard protocol (**Appendix 8.1.11**), and the concentration of final labelled protein was adjusted to give a suitable fluorescence signal for MST, of ~1000 AU. The unlabelled and labelled protein samples were both buffer exchanged into 50 mM Tris (pH7.0), 200 mM NaCl, 1 mM DTT, 0.05% Tween-20, 10 µM $ZnCl_2$, using a spin column (NanoTemper). Note a small amount of $ZnCl_2$ was also added to the buffer to ensure all the Zn^{2+} binding sites in Cnn-CM2 were saturated. A 2-fold dilution series was prepared for Spd-$2^{697-805}$ from, 2800 µM, 1400 µM, 700 µM etc. down to 85 nM. 15 µl of the dilution series was mixed with 15 µl of labelled $Cnn^{1082-1148}$, followed by a 20 min-incubation at room temperature. These samples were centrifuged for 1 min before loading into the premium-coated MST capillaries (NanoTemper) via capillary action. MST scans were carried out with 30 sec-laser pulses at 20% and 40% laser power, and with 20% LED power. Three repeats of the MST measurement were performed. The time traces showed smooth molecular diffusion upon Thermophoresis, and no aggregation was evident. Analysis of the normalised Thermophoresis and T-jump indicated that there might be an interaction at very high protein concentrations, but the change in fluorescence was not significant (**Figure 5.2**). In addition, the shift in normalised fluorescence was only detected above a protein concentration of 200 µM suggesting the binding strength might be on the scale of micromolar to millimolar range.

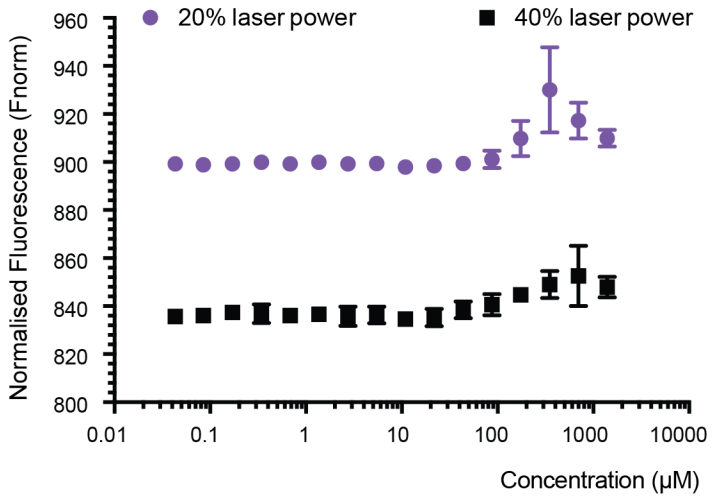


Figure 5.2 MST measurements to investigate whether the Spd-2 ASH1 domain interacts with the Cnn-CM2 domain that was labelled with primary amine-reactive dye in this assay. Three repeated measurements were taken, and the Thermophoresis + T-jump fluorescence signals were plotted against the $Cnn^{1082-1148}$ concentration, at two different MST powers. There is no significant change in Thermophoresis + T-jump with titrant concentration, although an increasing trend is observed. The complex formation needs to be further confirmed using other assays.

In order to confirm whether such weak binding is a true event, I repeated MST measurements with the titration of $Cnn^{1082-1148_noTEV}$ into primary amine-labelled Spd-$2^{697-805}$. Due to the requirement to include a high titrant concentration to detect a low affinity interaction, I used the non-TEV version of Cnn-CM2 that could reach a maximal protein concentration of ~5 mM. Both samples were buffer exchanged into 50 mM Tris (pH7.0), 200 mM NaCl, 1 mM DTT, 0.05% Tween-20, 1 μM $ZnCl_2$, using a spin column (NanoTemper). A dilution series was set up with the highest concentration of $Cnn^{1082-1148_noTEV}$ at 5300 μM that was diluted 2-fold to 2650 μM, 1325 μM etc. down to 162 nM. 15 μl of the dilution series was then mixed with 5 μl of labelled Spd-$2^{697-805}$ whose final concentration would yield a sufficient fluorescence reading at ~1000 AU. The same MST experiment parameters were followed as described above. More drastic Thermophoretic and T-jump behaviours were observed upon the titration of $Cnn^{1082-1148_noTEV}$, consistent with binding at high concentrations (**Figure 5.3**). Unfortunately, however, the unsaturated binding curve made estimation of the K_d value impossible.

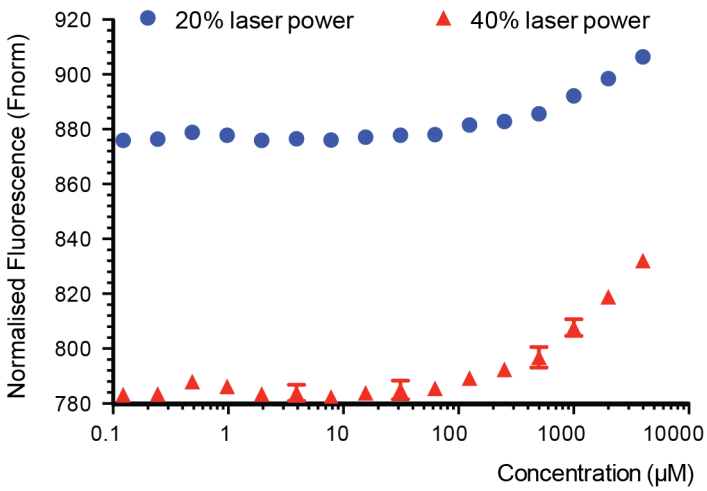


Figure 5.3 MST measurements to investigate whether Spd-2 ASH1 domain interacts with Cnn-CM2 domain that was labelled with a primary amine-reactive dye in this assay. Three repeated measurements were taken, and the Thermophoresis + T-jump fluorescence signals were plotted against the ASH1 concentration, at two different MST powers. Note an obvious shift in fluoresce is detected with the increasing titrant concentration, although the binding curve is not saturated.

5.1.2 Interaction characterisation via ^{1}H-^{15}N HSQC NMR measurements

With the assigned ^{1}H-^{15}N HSQC NMR spectrum for Spd-2$^{697-805}$, I decided to perform an NMR titration with Cnn$^{1082-1148}$ to map the position of any potential interaction interface. Both samples were dialysed into 50 mM NaH_2PO_4/Na_2HPO_4 (pH7.0), 220 mM NaCl, 1 mM DTT, 10 μM $ZnCl_2$, overnight at 4 °C. 5% *v/v* D_2O was present in a 350 μl reaction mixture. NMR spectra of 150 μM ^{15}N-Spd-2$^{697-805}$ and 50 μM ^{15}N-Spd-2$^{697-805}$ with the titration of 380 μM Cnn$^{1082-1148}$ were acquired at 20 °C, on Bruker Avance II 500 MHz spectrometer. Overlay of the two spectra showed several chemical shift perturbations, indicating a protein interaction albeit of low strength as indicated by the magnitude of change in peak positions (**Figure 5.4A**). This is consistent with the weak binding observed in MST assays described above. The peak shifts were then quantified following the formula $\Delta\omega = \sqrt{\left(\left(\frac{\gamma_N}{\gamma_H}\right)\Delta\omega_N{}^2 + \Delta\omega_H{}^2\right)}$, where $\Delta\omega_N$ is the change in chemical shift in ^{15}N dimension, $\Delta\omega_H$ is the change in chemcial shift in ^{1}H dimension and $\frac{\gamma_N}{\gamma_H}$ is the ratio of gyromagnetic ratios (**Figure 5.4B**). Interestingly, four regions of residues showed statistically significant chemical shift perturbations (1 to 2 standard deviations above the average). Mapping these residues onto the Spd-2 ASH1 model, predicted using structurally homologous domains, indicated the presence of two potential interaction interfaces, one of which donated the majority of the affected residues (**Figure 5.4D**). Surprisingly, the most significant changes in peak positions were observed mainly

for Gly residue that does not possess a large side chain. I therefore suspected backbone hydrogen bonding might be involved in the ASH1: CM2 domain interaction.

At this stage of the project, I solved the structure of $Cnn^{1082-1148}$, from which it was judged the TEV cleavage site might be flexible due to its missing electron density, and the residual TEV recognition sequence could be cleaved off to increase the protein solubility. Therefore I repeated the NMR binding experiment with the non-TEV form of CM2 ($Cnn^{1082-1148_noTEV}$). The same procedure was followed as outlined above. The new spectrum showed shifts in peak positions for the three regions that all gathered to one interface but not the region on the other end of the ASH1 modelled structure (**Figure 5.4C,D**). This suggested an unspecific binding of the TEV cleavage site onto the Spd-2 ASH1 domain, but also further confirmed the presence of a specific binding interface between the CM2 and ASH1 domains in solution. Note that conclusions still remain tentative at the time of writing, since the interaction residues are mapped on a predicted model, rather than the real structure, although the predictions contained high confidence as judged by FFAS scores (**Figure 4.2**, **Chapter 4**).

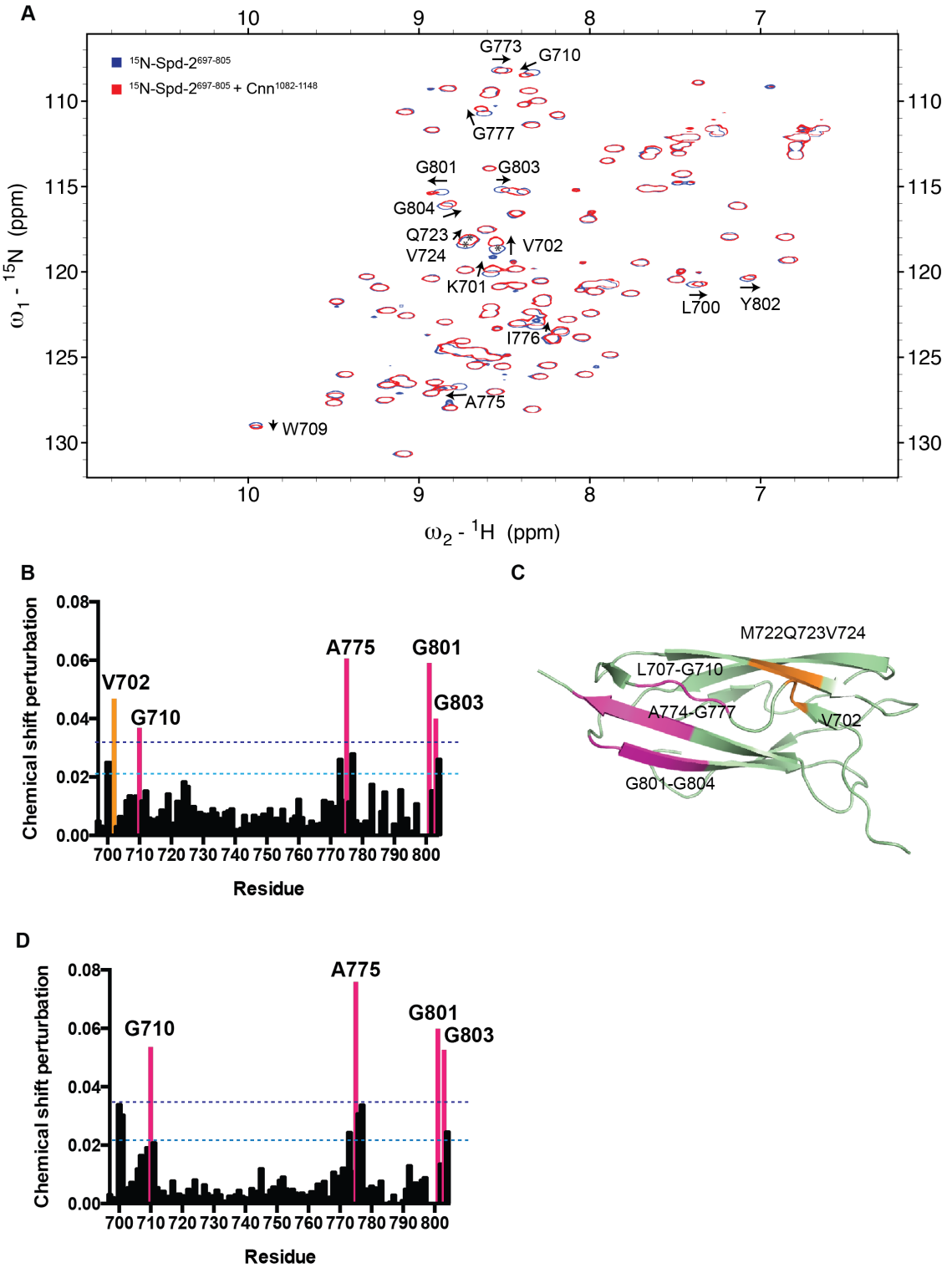


Figure 5.4 NMR measurement of interaction between the Spd-2 ASH1 and Cnn-CM2 domains. (**A**) Overlay of HSQC spectrum of ^{15}N-Spd-2$^{697-805}$ with or without the addition of 380 μM Cnn$^{1082-1148}$, in 50 mM NaH_2PO_4/Na_2HPO_4 (pH 7.0), 220 mM NaCl, 1 mM DTT, 10 μM $ZnCl_2$. Those peaks showing a shift in position are highlighted, with asterisks denoting the ones that were not chemically perturbed after titration of the non-TEV version of Cnn-CM2. (**B,D**) Quantification of the chemical shift perturbation in the presence of Cnn$^{1082-1148}$ (**B**) or $^{Cnn1082-1148_noTEV}$ (**D**). Four residues show significant peak position shift with a magnitude of two standard deviations (SD) above the average (highlighted in *pink*). Note the surrounding residues also demonstrate chemical shift perturbations with a magnitude of one SD from the average. Chemical shift perturbation of V702 (highlighted in *orange*) is only detected in the presence Cnn$^{1082-1148}$, indicating an unspecific binding of ASH1 at the TEV cleavage site. (**C**) Homology model of Spd-2$^{697-805}$ based on the FFAS hits showing the three highlighted regions in *pink*, centered around G710, A775 and G801 respectively, that might be involved in Cnn-CM2 binding. V702 is proven to be an unspecific binding site for the TEV cleavage sequence (shown in *orange*).

5.1.3 T1133E mutation does not affect the CM2: ASH1 interaction

As outlined in **Chapter 3.7.2**, a single point mutation, Thr1133 to Glu, within the Cnn-CM2 domain has disrupted the assembly of Cnn scaffold-like structure at centrosomes. Based on insights from the tetrameric crystal structure of Cnn-CM2, this mutation could disrupt the assembly of dimers into a stable tetramer structure (**Section 3.5**). In order to further investigate whether this mutation influenced the interaction with Spd-2 ASH1, I performed an NMR binding experiment, using ^{15}N-Spd-2$^{697-805}$ in the presence of Cnn$^{1082-1148}$ T1133E. The exact same procedures were followed as for the wild-type construct. Spectra were recorded at 20 °C, and were overlaid to identify any changes in peak positions (**Figure 5.5A**). In addition, the chemical shift perturbations were quantified and compared to WT. The same residues were observed at the interaction interface, suggesting this single point mutation has not influenced interaction of the Cnn-CM2 domain with the Spd-2 ASH1 domain. However, the NMR measurement was only conducted at one protein concentration (380 μM). In order to confirm the binding affinity between the CM2 mutant and ASH1, I performed MST measurements with the T1133E mutant protein. As mentioned above, Cnn$^{1082-1148_noTEV}$ could reach much higher protein solubility, and a more drastic change in Thermophoretic + T-jump behaviour was observed when labelled Spd-2$^{697-805}$ was titrated with varying amount of Cnn$^{1082-1148_noTEV}$. As a result, I decided to monitor the fluorescence change in primary amine-labelled Spd-2$^{697-805}$ upon the addition of Cnn$^{1082-1148_noTEV}$ T1133E.

MST measurements were repeated three times, and the normalised fluorescence in Thermophoresis and T-jump was plotted against the titrant concentration, as shown in **Figure 5.5B**. However, I did notice the “cold fluorescence” signal for the T1133E mutant construct before MST measurement increased with the increasing protein concentration. In addition, the plotted F_{norm} values showed higher background noise, especially at 20% MST power. This might be due to the aggregation-prone property of the T1133E mutant, which is indicated by its increasing diffusion coefficient at a higher protein concentration in the NMR DOSY measurement (**Chapter 3.6.3**). As a result, I remain cautious in comparing the MST binding curves between WT and the T1133E mutant, and the NMR binding assay seems a more reliable alternative to characterise the protein interaction in this case.

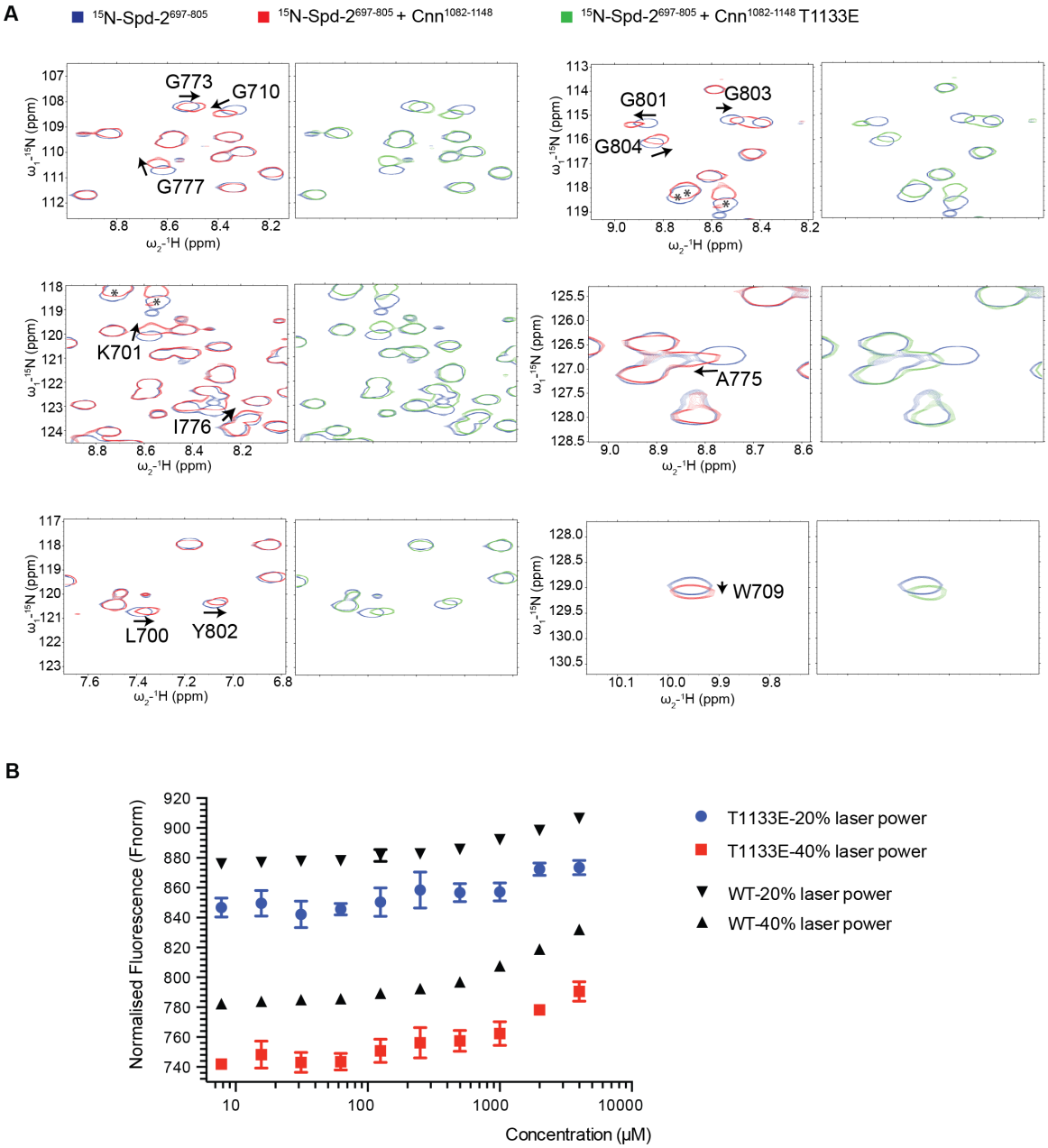


Figure 5.5 Analysis of protein interaction between Spd-2 ASH1 and the Cnn-CM2 T1133E mutant. (**A**) ^{15}N HSQC sections of Spd-2$^{697-805}$ in the presence of 380 μM Cnn$^{1082-1148}$ WT or T1133E mutant, in 50 mM NaH_2PO_4/Na_2HPO_4 (pH 7.0), 220 mM NaCl, 1 mM DTT and 10 μM $ZnCl_2$. Asterisks indicate residues unspecifically bound to the TEV cleavage sequence. There is no change in chemical shift perturbations comparing the WT and T1133E mutant, indicating the point mutation has not affected the ASH1 binding at tested concentration. This was further confirmed with the quantification of peak shifts (not shown). (**B**) Thermophoresis assay was performed to measure the strength of interaction between ASH1 and the CM2 T1133E mutant, at 2 different MST powers. The normalised fluorescence, in Thermophoresis and T-jump, was plotted against the titrant concentration. Binding curves are displayed for both WT and the point mutant. Note a higher background noise is observed for ASH1: CM2 T1133E binding titration.

5.2 Design of mutations to dissect the interaction between the CM2 and ASH1 domains

In order to determine the biological importance of a CM2: ASH1 domain interaction, it is necessary to design mutations of the binding interface and observe the consequence of such mutations on complex formation *in vitro* and *in vivo*.

By mapping the CM2 sequence conservation onto its structure, I observed that there were several conserved and charged residues on the top face of the CC that were not involved in tetramer or dimer formation, and so could potentially be available for binding to Spd-2 in the context of a dimer or tetramer (**Figure 5.6**).

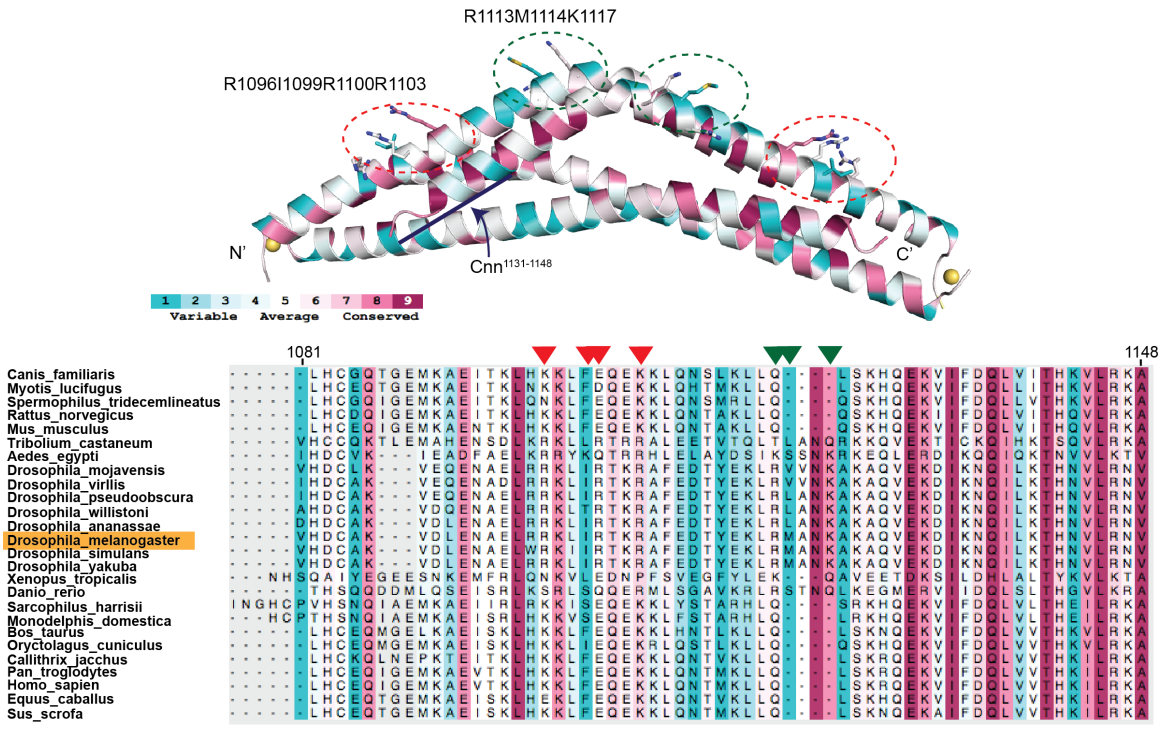


Figure 5.6 Design of mutation clusters that might weaken or abolish the binding of Cnn-CM2 to Spd-2 ASH1. Top, structure of Cnn-CM2 coloured according to the Consurf conservation score. Clusters of residues mutated to Ala are highlighted. Bottom, MSA of CM2 across evolution showing the position of two mutational clusters (highlighted by arrowheads of different colours matching with the loop colours in top figure). Cnn$^{1131-1148}$, highlighted with a *blue* bar, is not involved in ASH1 interaction (see main text). Note that majority of these residues are conserved evolutionarily in terms of their amino acid properties, and some residues show conservation only across fly species. Residue numbers for *D. melanogaster* are indicated, above the sequence alignment.

Preliminary data from NMR binding assay suggested the deletion of C-terminal 18 residues of the Cnn-CM2 domain did not weaken/abolish its binding to Spd-2$^{697-805}$ (not shown). I therefore

suspected this putative CM2: ASH1 domain interaction would mainly involve residues 1082 to 1130, and I designed two clusters of mutations to target localised sets of residues that are relatively conserved in amino acid property in this region. Residues were mutated to Ala by site-directed mutagenesis (Agilent). These constructs are currently awaiting testing.

5.3 Future experiments

In this chapter, I have presented the preliminary data for characterisation of the CM2: ASH1 domain interaction *in vitro*. Due to time restriction, I was not able to fully characterise this putative binding event, and many of the experiments I present here require further repetition and optimisation to test if the weak interaction I am observing is physiologically relevant. In future experiments, I would like to analyse the influence of mutational clusters within the Cnn-CM2 region on its binding strength to the Spd-2 ASH1 domain using NMR binding assay as described above. If I can confirm which amino acids in each domain are responsible for this protein interaction, I plan to mutate these residues and test their *in vivo* significance using the RNA injection approach.

Chapter 6: Purification and characterisation of Asterless (Asl)

6.0 Introduction

6.0.1 Asterless (Asl)

Drosophila Asl is predicted to be a highly coiled coil protein. Homologs of the protein can be found in species from flies to vertebrates (where it is called Cep152) [Andersen et al., 2003; Blachon et al., 2008; Bonaccorsi et al., 1998; Varmark et al., 2007]. However, there is no homolog of Asl in worms. Asl/Cep152 family members all appear to be involved in centriole duplication [Blachon et al., 2008; Varmark et al., 2007]. In flies, Asl provides a molecular link between centriole duplication and PCM assembly due to its role in Spd-2 recruitment [Conduit et al., 2010]. Indeed, cells lacking centrioles can still form PCM if Asl is overexpressed, suggesting it is involved in the initiation of PCM assembly [Baumbach et al., 2015].

Several labs have identified interactions between Asl/Cep152 and either Sak/PLK4 or Sas4/CPAP [Cizmecioglu et al., 2010; Dzhindzhev et al., 2010; Hatch et al., 2010; Kim et al., 2013; Park et al., 2014; Sonnen et al., 2013] in flies and vertebrates. As described in the **Introduction**, Sak/PLK4 is at the top of the core centriole assembly pathway [Kleylein-Sohn et al., 2007; Peel et al., 2007; Pelletier et al., 2006]. In flies, Asl recruits Sak to the side of the mother centriole to initiate centriole duplication [Dzhindzhev et al., 2010; Novak et al., 2014]. In vertebrates, Cep152 appears to cooperate with Cep192 to recruit PLK4 to mother centrioles [Kim et al., 2013; Park et al., 2014; Sonnen et al., 2013], while in worms ZYG-1, is recruited to mother centrioles exclusively by SPD-2 [Kemp et al., 2004; Pelletier et al., 2004]. In flies, Asl is first recruited to new-born centrioles towards the end of mitosis at about the time the new centrioles disengage from their mothers [Novak et al., 2014]. This recruitment is Sas-4 dependent, and it appears to be a crucial step in allowing new centrioles to mature into mother centrioles that can recruit PCM and that are capable of supporting new rounds of centriole duplication ([Novak et al., 2014]; personal communication with Dr Zsofia A. Novak). The interaction between Cep152/Asl

and CPAP/Sas-4 is also observed in vertebrates, but the significance of this interaction is less clear [Cizmecioglu et al., 2010].

At the beginning of my project these protein interactions were identified, but were not structurally characterised. Recently, several structures have been published of the PLK4 cryptic polo box (CPB) that is involved in Asl/Cep152 binding [Shimanovskaya et al., 2014; Slevin et al., 2012]. In addition, a crystal structure is also solved for protein complex formed between vertebrate PLK4 and Cep152 [Park et al., 2014]. In the **Perspective** section of this chapter, I discussed some of my work in regard to these published structures.

6.0.2 Aims

- Identify putative functional domains of Asl.
- Purify and characterise domains of Asl that might provide insight into its biological role.

6.0.3 Key findings

The N-terminal fragment of Asl (Asl^{1-86}) known to be involved in Sak binding does not form a well-folded protein when expressed and purified from bacteria. Interestingly, Asl^{1-356} behaved predominantly as a dimer in solution, while its first 180 residues were not required for this self-association. The C-terminal fragment of Asl that includes the Sas4 interaction region showed a tendency to form trimers in solution. Overall, Asl contains at least 2 self-interaction domains that showed different oligomerisation states in solution. Such homo-oligomerisation might be necessary for the ring-like localisation of Asl around the centrioles [Fu and Glover, 2012; Mennella et al., 2012].

6.1 Bioinformatic analysis of Asl

6.1.1 Asl proteins are highly coiled-coil, showing little sequence conservation

MSA was first performed using Asl/Cep152 sequences from flies and vertebrates. The vertebrate Cep152 proteins are much longer than the fly homologs, making it difficult to judge the accuracy of full-length sequence alignment. At the time I carried out this project, the canonical isoform of HsCep152 was Isoform 3 comprising 1650 residues. More recently, a longer isoform (Isoform 4) was identified showing Isoform 3 had residues 1156 to 1121 missing. Therefore, both Isoforms were included in my bioinformatic analysis. I was not able to detect any region with enriched sequence conservation, apart from the first 50 residues (denoted as Conserved Region 1 (CR1), **Figure 6.1A-C**). In particular, an enrichment of negatively charged residues is observed within the CR1 region (**Figure 6.1C**). Dzhindzhev et al., carried out GST-pull down assays using truncated Asl fragments to investigate the Asl: Sak interaction *in vitro* [Dzhindzhev et al., 2010]. A minimal domain of residues 24 to 47 was found essential for this interaction, suggesting that CR1 comprises the region of Asl/Cep152 that interacts with Sak/PLK4. This is very recently shown to be the case when the structure of this interaction interface is solved for the human proteins [Park et al., 2014], although it is important to note that this work was published after I had performed the studies I describe in this chapter.

In addition, Hatch et al., carried out MSA with vertebrate Cep152 sequences and identified a region termed as CR2 (residues 1207 to 1295 in Isoform 3; residues 1263 to 1351 in Isoform 4) towards the C-terminus of the full-length protein [Hatch et al., 2010]. This region could only be found in vertebrates but not flies, which explains why I did not identify this region in my MSA analysis.

I carried out further bioinformatic analysis using secondary structural prediction tools. Secondary structural prediction by PSIPRED [Jones, 1999] was conducted on several Asl homologous proteins, which all indicated a large alpha-helical content in Asl/Cep152 proteins (not shown). This was further confirmed by COILs analysis [Lupas et al., 1991] from which the potential coiled coil (CC) domains were assigned (**Figure 6.1D**). Surprisingly, DmAsl is predicted to contain more CC domains than HsCep152 (**Figure 6.1A,B,D**). For instance, CR1 is predicted to form a CC in

DmAsl, but not in HsCep152. The 2 CC domains predicted for HsCep152, by COILs, show structural homology to a structural maintenance of chromosome (SMC)-like domain (SMC-A, SMC-B) [Guernsey et al., 2010]. Isoforms 3 and 4 differ by the presence of an extra short CC corresponding to the missing residues in the former protein sequence. Interestingly, the last 347 residues in DmAsl are found to bind Sas-4 [Dzhindzhev et al., 2010], and a much longer C-terminal region is mapped on Cep152 (residues 784 to 1654 of Isoform3) to bind CPAP (Sas-4 in *Drosophila*) [Cizmecioglu et al., 2010]. This interaction seems to be conserved in evolution, but no obvious sequence similarity can be identified in these regions.

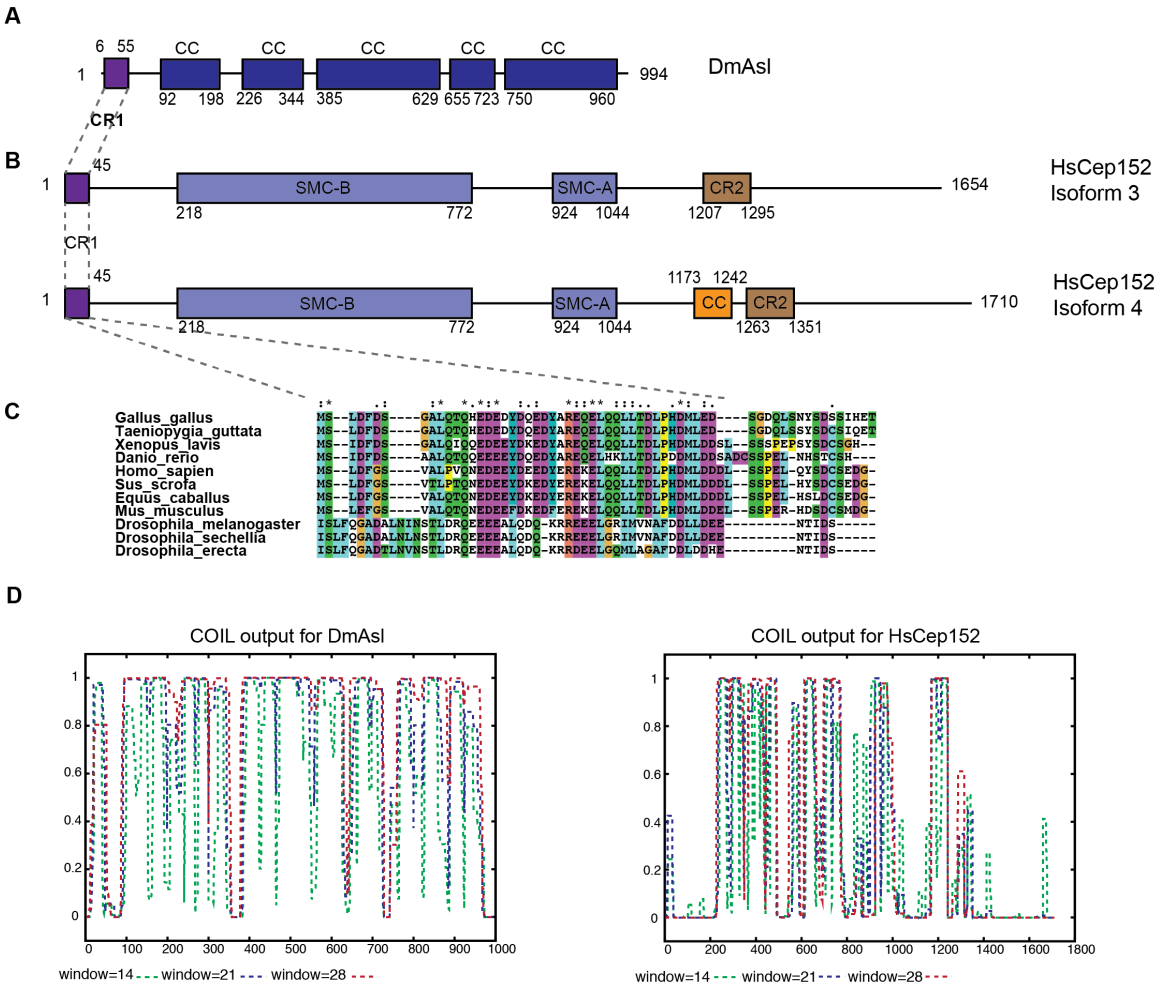


Figure 6.1 Bionformatic analysis of DmAsl and HsCep152. (**A,B**) Domain delineation for DmAsl and HsCep152 (Isoform3 and Isoform4). CR1 (conserved region 1) is found sufficient and essential for Sak binding. CR2 (conserved region 2) shows sequence conservation across vertebrate species. Its role in centriolar targeting suggests this domain might be functionally equivalent to Asl$^{648\text{-}994}$, which is required for centriolar targeting in flies. The predicted coiled-coil (CC) domains in central region of Cep152 are classified as structural maintenance of chromosome (SMC)-like domains. The longer isoform of Cep152 has 56 extra residues inserted between residues 1156 and 1211, which are predicted to form a CC domain. (**C**) Selected sequences from MSA analysis of Asl and Cep152 showing conservation within the CR1 domain. In particular, the negatively charged residues (Asp, D; Glu, E) are highly conserved in evolution. Invariant residues are highlighted with asterisks, conserved residues with colons and semi-conserved residues with full stops. Residues are coloured according to the ClustalX colour scheme. (**D**) COILs server output showing the predicted CC domains for Asl and Cep152-Isoform 4. Note a CC is predicted for the CR1 domain in DmAsl, which is not the case for HsCep152.

6.1.2 Rationalisation of the Asl constructs designed for *in vitro* characterisation

Due to the significant sequence divergence for Asl/Cep152 homologs, I have performed a MSA analysis for Asl proteins from fly species only. This alignment allows regions with more enriched conservation to be determined. The N-terminal 356 residues showed sequence similarity, especially towards the first 68 residues. For example, *D. melanogaster* Asl$^{1-356}$ had 37% sequence similarity to *Apis melifera* Asl$^{1-356}$, and *D. melanogaster* Asl$^{1-68}$ had 59.4% sequence similarity to *A. melifera* Asl$^{1-68}$. Therefore, I designed a construct covering residues 1 to 356. Truncated fragments were designed by successively trimming back from the C-terminus with the shortest construct covering the 68 highly conserved residues (**Figure 6.2A**). Asl$^{1-356}$ was cloned into a pETM-14 vector that contains an N-terminal His6 tag via T4 ligation (NEB). The shorter fragments were cloned by site-directed mutagenesis (Agilent) on pETM-14-Asl$^{1-356}$, after which a double stop codon "TAATAA" was inserted into the ORF.

Dzhindzhev et al., have also mapped the interaction region for Sas4 on Asl to the C-terminal 347 residues (Asl$^{648-994}$) [Dzhindzhev et al., 2010], and I therefore cloned Asl$^{648-994}$ into a pETM-14 vector. From the MSA of *Drosophila* Asl proteins, I noticed a gap in sequence conservation around residues 720 to 740, which is consistent with a break between 2 long continuous CCs (**Figure 6.2B**). I hence designed another shorter fragment covering the last CC domain at the extreme C-terminus (**Figure 6.2A**). Both constructs were PCR amplified from the cDNA encoding full-length Asl protein, and were subsequently cloned into the NcoI/NotI double digested pETM-14 vector backbone by In-Fusion cloning (Clontech).

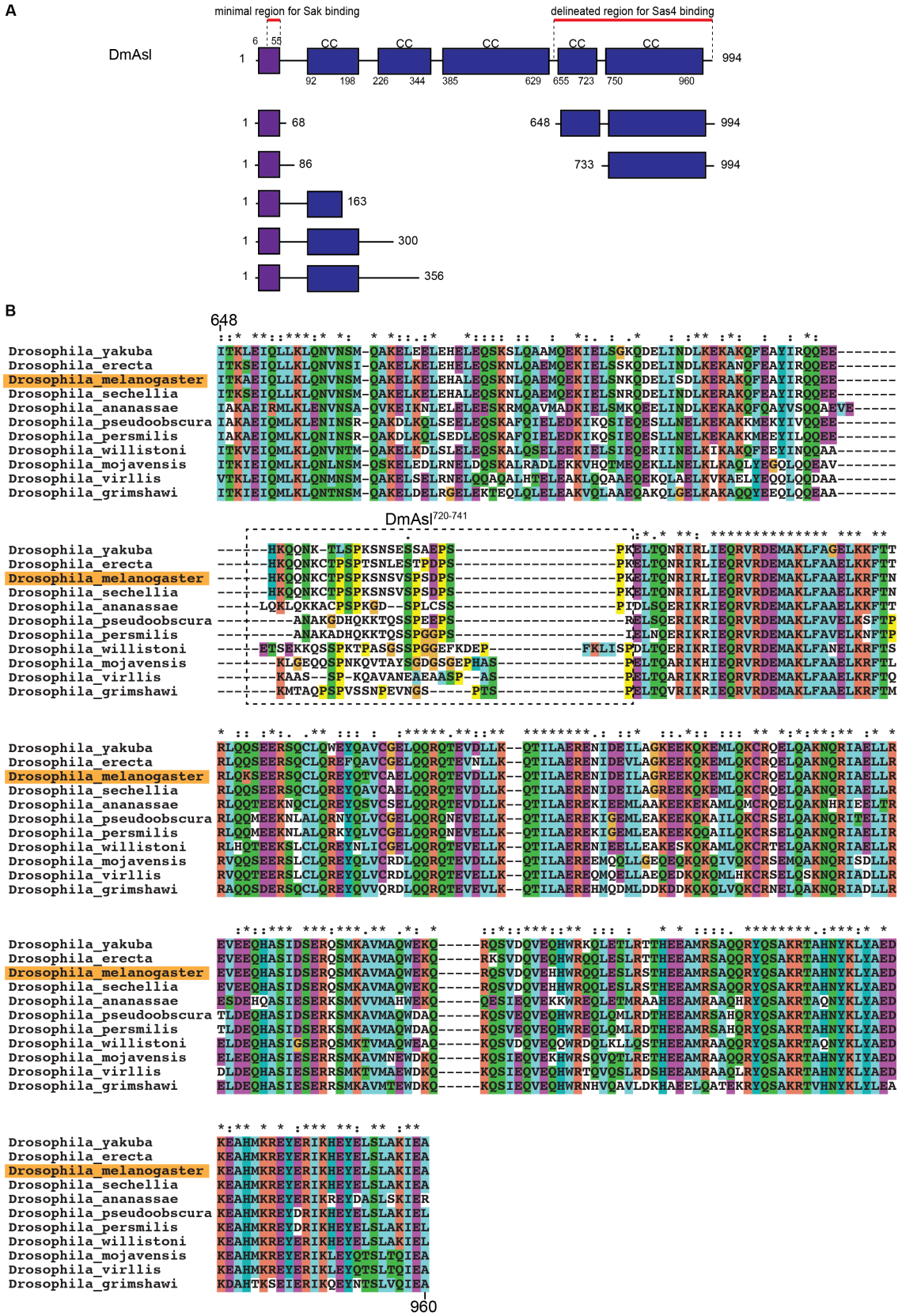


Figure 6.2 Asl constructs designed for *in vitro* purification and characterisation. (**A**) Schematics showing the overall topology of DmAsl, and the different constructs that are designed. All the N-terminal constructs cover the minimal region required for Sak binding. Two C-terminal constructs are designed, one of which contains only the most C-terminal CC domain. (**B**) MSA analysis for *Drosophila* Asl homologs showing the break in sequence conservation between residues 720 and 741 where several Pro residues are observed. Pro residues could break the CC continuity, and hence its insertion might explain the correlation of this region with a linker between two CC domains as shown in (**A**). Residue numbers for *D. melanogaster* are indicated above the sequence alignment. Invariant residues are highlighted with asterisks, conserved residues with colons and semi-conserved residues with full stops. Residues are coloured according to the ClustalX colour scheme.

6.2 Purification and characterisation of Asl N-terminal fragments

6.2.1 Screening for soluble N-terminal fragments

For biophysical and structural characterisations *in vitro*, a large amount of purified protein is required. I therefore decided to use the bacterial system for the expression of various Asl constructs. As mentioned in **Section 6.1.2**, all constructs were cloned into a pETM-14 vector with a 3C protease-cleavable His6 tag at the N-terminus. In a high-throughput expression and purification screen, all plasmids were transformed into both B834 (DE3) and Tuner (DE3) cells, which were then grown in LB or TB medium. Details of the screened constructs are summarised in **Table 6.1**. Cells were grown according to the standard protocol (**Appendix 8.1.6**). Protein expression was induced by the addition of 1 mM IPTG, overnight at 21 °C. The overexpressed protein was then purified using Ni-NTA magnetic beads (Qiagen, **Appendix 8.1.6**). Gel samples were taken for the induced cell pellet and protein eluted off Ni-NTA beads to check the expression of soluble protein. As shown in **Figure 6.3**, all proteins were overexpressed at the expected sizes, apart from HisAsl$^{1\text{-}68}$ for which no induced band was observed. Gel analysis of Ni bead elution showed that a small amount of soluble protein could be purified for each construct. I then decided to express and purify the shortest (HisAsl$^{1\text{-}86}$) and the longest N-terminal fragments (HisAsl$^{1\text{-}356}$) on a large scale.

Construct	MW (kDa)	pI	Abs 0.1%	No. of Cys
HisAsl$^{1\text{-}68}$	10	4.59	0.15	0
HisAsl$^{1\text{-}86}$	12	4.94	0.249	0
HisAsl$^{1\text{-}163}$	21	5.13	0.143	1
HisAsl$^{1\text{-}300}$	37	5.23	0.311	3
HisAsl$^{1\text{-}356}$	43	5.09	0.299	4

Table 6.1 Details of the His-tagged Asl N-terminal constructs, calculated using Protparam [Wilkins et al., 1999]. Due to the presence of cysteine residues, 5 mM β-mercaptoethanol (BME) was included in working buffer (WB) to prevent disulfide bonds. Note for HisAsl$^{1\text{-}68/1\text{-}86}$ constructs, no cysteine residue is present in the protein sequence and hence no reducing agent was included in WB.

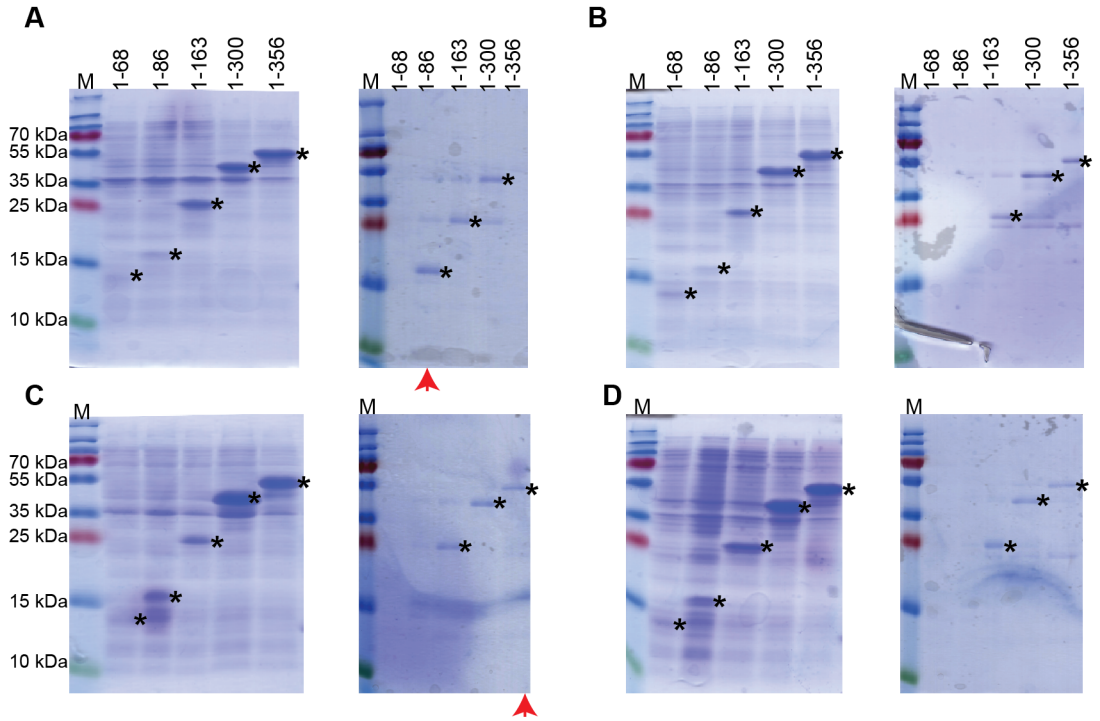


Figure 6.3 High-throughput expression and purification screen for soluble N-terminal fragments, analysed by 15% reducing SDS-PAGE gels**.** Different cell types and growth media were assessed for each construct. All cell cultures were grown at 37 °C until OD_{600} reached 0.6. 1mM IPTG was then added to induce protein expression overnight at 21 °C. A detailed protocol can be found in **Appendix 8.1.6**. (**A,B**) Plasmids were transformed into B834 (DE3) cells that were grown in LB (**A**) or TB (**B**) media. (**C,D**) Protein overexpression using Tuner (DE3) cells in LB (**C**) or TB (**D**) media. Gels on the left showing protein induction, and gels on the right showing soluble protein eluted off Ni-NTA magnetic beads. HisAsl$^{1-68}$ is not largely overexpressed, and no soluble fraction could be obtained. Two constructs are brought forward for large-scale purification, and the optimal growth conditions for both constructs are highlighted with *red* arrows. Asterisks indicate the bands corresponding to HisAsl constructs. M: PageRuler Plus Prestained Protein Ladder (Thermo Scientific).

6.2.2 Purification of HisAsl$^{1-86}$

For large-scale expression, 8 l LB media were inoculated with an overnight culture of B834 (DE3) cells that were transformed with pETM-14-Asl$^{1-86}$ plasmid. Cells were grown at 37 °C until OD_{600} reached 0.6. 1 mM IPTG was added to induce the protein overexpression overnight at 21 °C. Protein purification proceeded following the standard protocol (**Appendix 8.1.7**), using a Ni-NTA column (Qiagen) (**Figure 6.4A**). This was followed by size exclusion chromatography to further purify the protein (**Figure 6.4B**). The pooled Ni-NTA elution was concentrated to 5 ml using a 3 kDa MWCO centrifugal concentrator (Amicon) before loading onto an S75 16/60 column (GE Healthcare) that was pre-equilibrated in WB. Subsequent SDS-PAGE gel analysis revealed that the full-length protein was eluted around 0.54 CV, showing a high purity despite a minor amount of lower molecular weight contaminants (**Figure 6.4C**). HisAsl$^{1-86}$ is ~12 kDa, a globular protein of similar size is expected to elute from the same SEC column around 0.7 CV [EMBL, 2015]. The

much earlier elution observed with HisAsl$^{1-86}$ might be due to its extended α-helical conformation or its oligomerisation in solution.

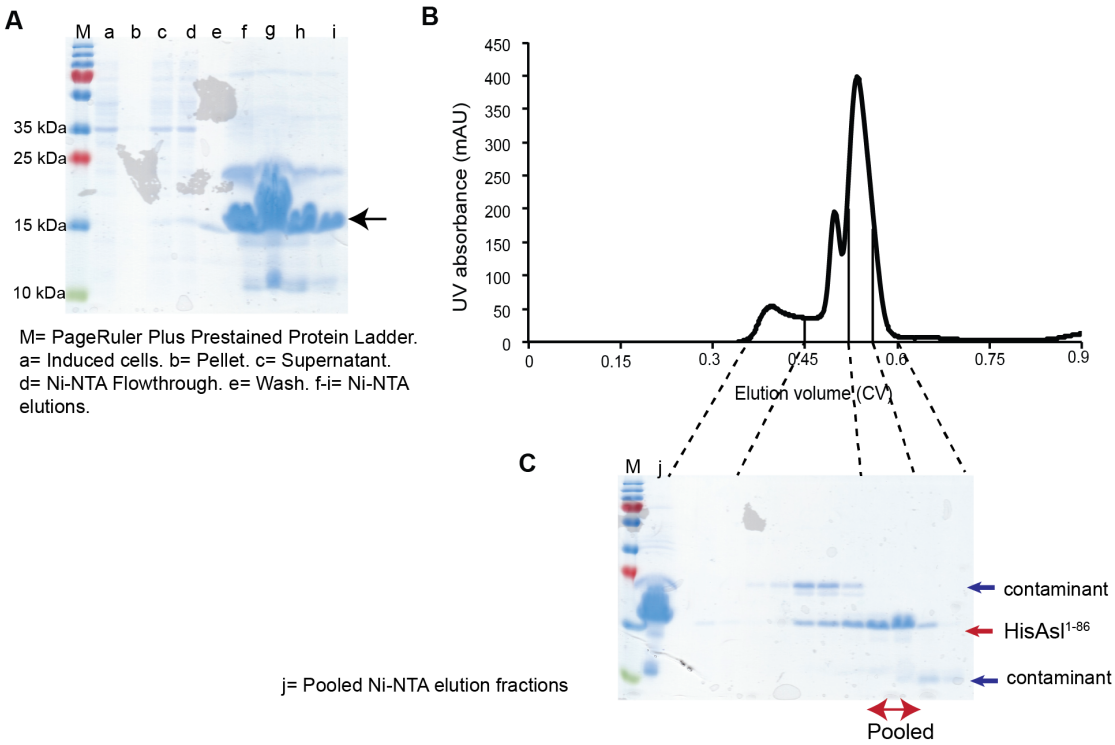


Figure 6.4 Purification of HisAsl$^{1-86}$. (**A**) 15% reducing SDS-PAGE gel analysis of Ni-NTA purification fractions. Arrow indicates the band corresponding to HisAsl$^{1-86}$. (**B**) SEC trace from an S75 16/60 column. Elution with 50 mM Tris (pH 7.5), 150 mM NaCl. The subsequent 15% reducing SDS-PAGE gel analysis of SEC peak fractions (**C**) indicates the elution of HisAsl$^{1-86}$ in the major peak around 0.54 CV. The preceding peak contains higher molecular weight contaminants. Only fractions (highlighted) from the major elution peak are pooled for further experiments.

6.2.3 Biophysical characterisation of HisAsl$^{1-86}$

SEC-MALS analysis was conducted on the purified protein to assess its oligomerisation state in solution. Sample was prepared at 417 μM for injection onto an S75 10/30 column (GE Healthcare) equilibrated with WB. Protein was eluted with an observed molecular weight consistent with a monomer (**Figure 6.5**).

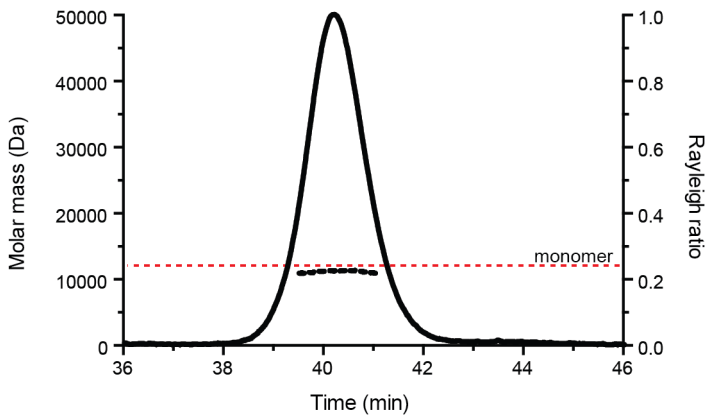


Figure 6.5 SEC-MALS analysis of HisAsl$^{1-86}$ at 470 μM in 50 mM Tris (pH 7.5), 150 mM NaCl. Protein eluted off an S75 10/30 column as a monomeric species. The solid line indicates the relative Rayleigh ratio representing protein elution, and the dashed line shows the observed molecular weight across the peak. The expected mass of a monomer is indicated with a *red* dotted line across the graph.

The secondary structural content of HisAsl$^{1-86}$ was confirmed via CD experiment. Protein sample at 0.1 mg/ml was dialysed overnight at 4 °C into 10 mM NaH_2PO_4/Na_2HPO_4 (pH7.5). Four CD scans were performed at 20 °C, and the ellipticity was measured at wavelengths between 195 nm and 250 nm. Dialysis buffer was used as a blank measurement for background subtraction. The averaged CD spectrum is displayed in **Figure 6.6**. Two negative bands are observed at ~201 nm and ~222 nm respectively. In α-helical type profile, 2 troughs are expected at 208 nm and 222 nm, showing similar ellipticity values. However, in this case, the ellipticity reading is much more negative at 201 nm, and the profile seems to be more shifted towards a random coil type in which a negative band is expected at 195 nm. Overall, it is difficult to judge whether the protein is properly folded only based on the results from CD experiment. I therefore decided to investigate this by NMR spectroscopy.

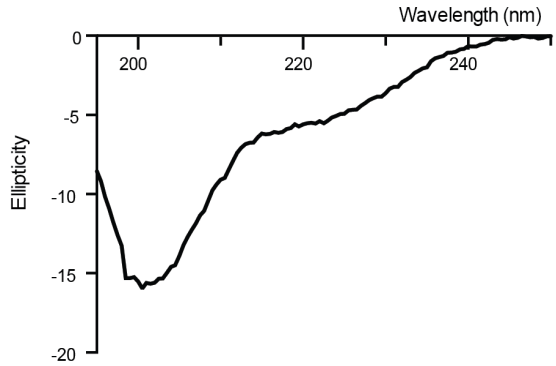


Figure 6.6 Secondary structural content of HisAsl$^{1-86}$ was investigated by CD experiment in 10 mM NaH_2PO_4/Na_2HPO_4 (pH7.5) at 20 °C. Two negative troughs are observed at ~201 nm and ~222 nm. If the protein is highly α helical, two negative peaks are expected at 208 and 222 nm. The first peak, in this case, is more shifted towards a random coil type profile where a largely negative band is expected at 195 nm. In addition, the two troughs in α type spectrum normally have similar ellipticity readings, which is not the case for HisAsl$^{1-86}$. It is difficult to conclude the folding state only based on the results from CD analysis.

Following the same protocol described for production of ^{15}N-labelled Spd2$^{697-805}$ (see **Section 4.4.2**), ^{15}N-labelled HisAsl$^{1-86}$ was also expressed using BL21 (DE3) cells that were grown in minimal media (**Table 6.2**) at 21 °C overnight with 1 mM IPTG induction. Protein purification was carried out following the same protocol as for the unlabelled native protein. The SEC fractions were pooled and concentrated to 18 mg/ml before dialysing into the NMR buffer (50 mM NaH_2PO_4/Na_2HPO_4 (pH 7.5), 150 mM NaCl). A 350 µl sample was prepared with a final concentration of 5% *v/v* D_2O. A ^{1}H-^{15}N HSQC spectrum was collected at 20 °C on Bruker Avance II 500MHz spectrometer. The spectrum showed severe peak overlaps in the central region, hence making peak picking very difficult (**Figure 6.7**). In addition, peaks were observed with chemical shifts ranging from 6.5 ppm to 8.5 ppm, strongly suggesting the residues are largely in a similar

chemical environment and that the protein is largely unstructured in solution. As a result, I did not pursue this construct for further structural characterisation.

Composition	Final concentration
1x salt solution:	
Na_2HPO_4	42 mM
KH_2PO_4	22 mM
NaCl	8 mM
Trace elements (pH 7.0):	
EDTA	270 µM
$FeCl_3$	48 µM
$ZnCl_2$	3 µM
$CuCl_2$	1 µM
$CoCl_2$	1 µM
H_3BO_3	2.5 µM
Yeast Nitrogen base	1.5 g/l
CaCl	0.1 mM
Thiamine	1 mM
Glucose	0.5% w/v
$^{15}NH_4Cl$	1 mg/ml
Kanamycin	50 µg/ml

Table 6.2 Composition of minimal medium for the expression of ^{15}N-labelled HisAsl$^{1-86}$. Note the yeast nitrogen base does not contain any amino acids or ammonium sulfate, which is supplemented externally.

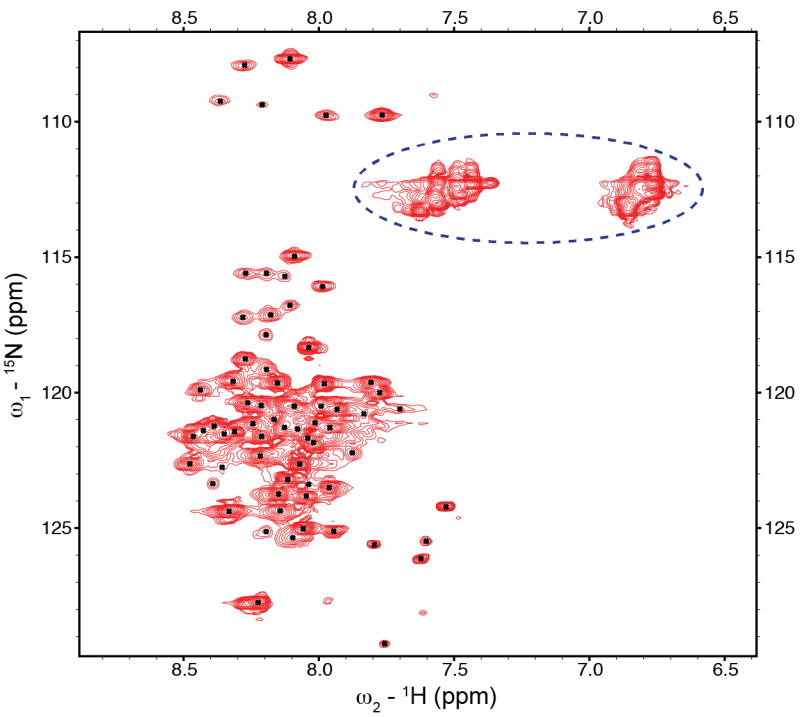


Figure 6.7 ^{1}H-^{15}N NMR spectrum of HisAsl$^{1-86}$ in 50 mM NaH_2PO_4/Na_2HPO_4 (pH 7.5), 150 mM NaCl. Attempts were made to pick putative peaks for backbone amides (*black* dots) and Asn and Gln side chain amides (dashed circle). Significant peak overlap is observed, especially in the central region of the spectrum. The narrow range of chemical shifts experienced along both dimensions (^{1}H and ^{15}N) also suggests the protein is not properly folded.

6.2.4 Purification of HisAsl$^{1-356}$

HisAsl$^{1-356}$ was expressed and purified from Tuner (DE3) cells that were grown in 2 l of LB media. The N-terminal His6 tag of the recombinant protein allowed purification via Ni-NTA chromatography following the standard protocol (**Appendix 8.1.7**). Gel analysis of protein sample eluted off the Ni-NTA column showed several contaminant bands being co-eluted (**Figure 6.8A**). Protein was further purified by size exclusion chromatography, using WB as the running buffer. Full-length protein eluted off an S200 16/60 column (GE Healthcare) around 0.47 CV, and the major contaminant band was separated as it eluted ~0.15 CV later (**Figure 6.8B**). However, a ladder of lower molecular weight bands was observed with the elution of full-length protein, and the Intensity of these lower bands increased over time when the sample was stored at 4 °C (not shown). This suggested that these shorter species were likely degradation products. I decided to conduct limited proteolysis on HisAsl$^{1-356}$ to try to identify a more stable core structure.

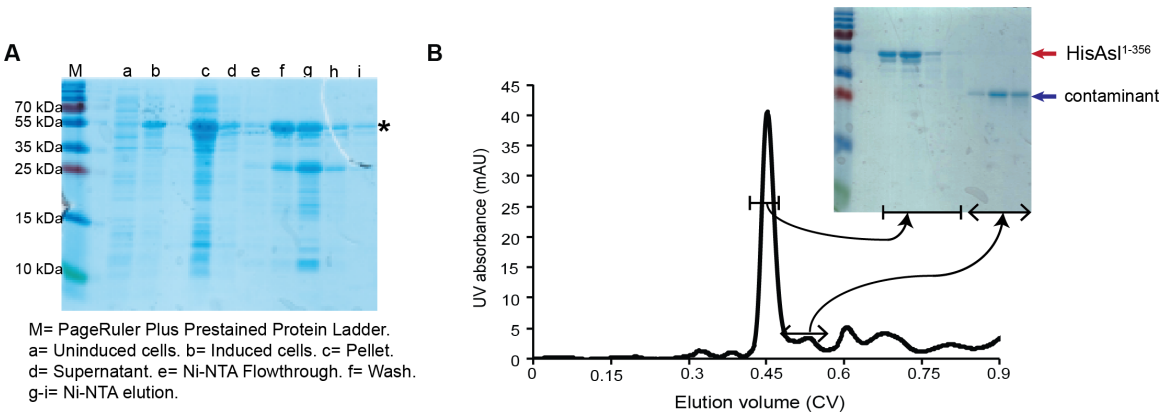


Figure 6.8 Large-scale purification of HisAsl$^{1-356}$. (**A**) 15% reducing SDS-PAGE gel analysis of Ni-NTA purification fractions. Full-length protein (indicated by an asterisk) is overexpressed, although majority is in the pellet (insoluble fraction). Several contaminant bands are observed in the Ni-NTA elution. (**B**) Left, SEC elution profile of an S200 16/60 column (elution with 50 mM Tris (pH 7.5), 150 mM NaCl and 5 mM BME); Right, the subsequent 15% reducing SDS-PAGE gel analysis of SEC fractions. HisAsl$^{1-356}$ eluted in the major peak around 0.47 CV, and a ladder of shorter species is also detected within the same fractions, as shown by SDS-PAGE gel analysis.

6.2.5 Limited proteolysis of HisAsl$^{1-356}$

A 40 µl reaction was set up with target protein at 1 mg/ml, and an increasing amount of trypsin protease was added (**Figure 6.9A**). All reactions were left at room temperature for 2 hrs before a 15% SDS-PAGE gel analysis was performed to check if a more stable fragment was obtained. An initial trial indicated that limited proteolysis has occurred forming several lower molecular weight species (**Figure 6.9A**). Nonetheless, a major band (secondary species) was detected with a mass

just above 35 kDa, as shown via gel analysis. Further addition of trypsin led to a more severe truncation from that secondary species, and ladder of protein bands was obtained. The target protein: protease *w/w* ratio was subsequently fine tuned between 5,000 to 1 and 1,000 to 1 so that the amount of secondary species could be maximised (**Figure 6.9B**).

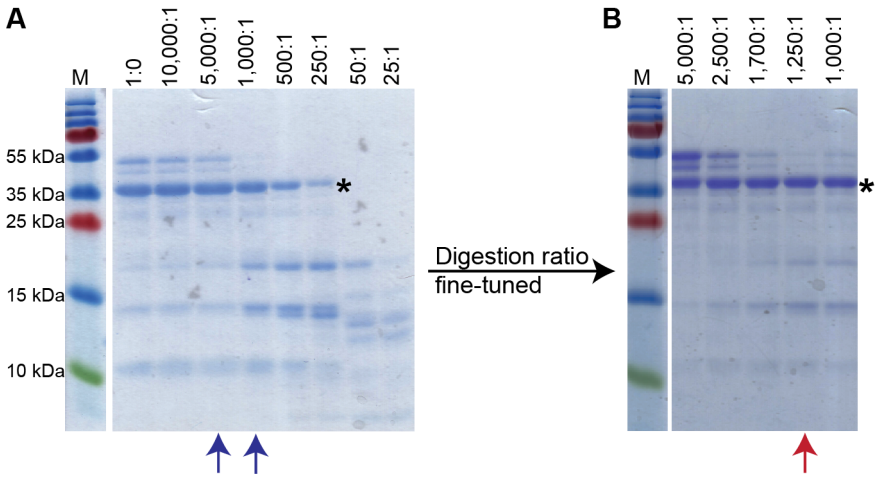


Figure 6.9 Small-scale trypsin digestion to find the optimal protein: trypsin *w/w* ratio. 15% reducing SDS-PAGE gel analysis was performed. A dominant digestion band (highlighted with an asterisk) is detected, and the addition of trypsin to a ratio of higher than 1,000:1 leads to a ladder of lower molecular weight bands. I further fine-tuned the protein: trypsin *w/w* ratio between 5,000:1 and 1,000:1 (highlighted with *blue* arrows in **A**), so that the intensity of dominate band (secondary species, highlighted with an asterisk) would be maximised and the amount of other shorter species is minimised. The optimal protein: trypsin *w/w* ratio is 1,250:1 (highlighted with a *red* arrow in **B**). M: PageRuler Plus Prestained Protein Ladder (Thermo Scientific).

As the optimal trypsin digestion ratio was determined, I carried out a large scale limited proteolysis with the Ni-NTA purified HisAsl$^{1\text{-}356}$. The Ni elution was dialysed into WB and concentrated to 5 ml before being supplemented with trypsin at protein: protease *w/w* ratio of 1,250 to 1. The reaction was left at room temperature for 2 hrs, and was then loaded onto an S200 16/60 column (GE Healthcare) for SEC, using WB as the running buffer. This was followed by gel analysis to check the separation of digestion products (**Figure 6.10A,B**). The secondary species eluted around 0.48 CV of the SEC column, although other shorter species from trypsin digestion could not be fully removed.

To further purify this more stable fragment and hence maximise the accuracy in mass spectrometry analysis, I carried out an ion exchange purification step, which utilises the difference in strength of protein binding to cation beads packed within a monoQ 5/50 GL column (GE Healthcare). The pooled SEC fractions were loaded onto the ion exchange column that was equilibrated to 150 mM NaCl (**Figure 6.10C,D**). Two buffers were used in anion exchange

purification, A (50 mM Tris (pH7.5), 5 mM BME) and B (50 mM Tris (pH 7.5), 1 M NaCl, 5 mM BME). A linear gradient of NaCl from 150 to 650 mM over 50 ml was applied to elute the column. The lower molecular weight contaminants were removed with their elution at 230-250 mM NaCl. The dominant digestion product eluted as a sharp peak with a mid-point of ~260 mM NaCl, despite two higher bands were still visible but in a much lower amount.

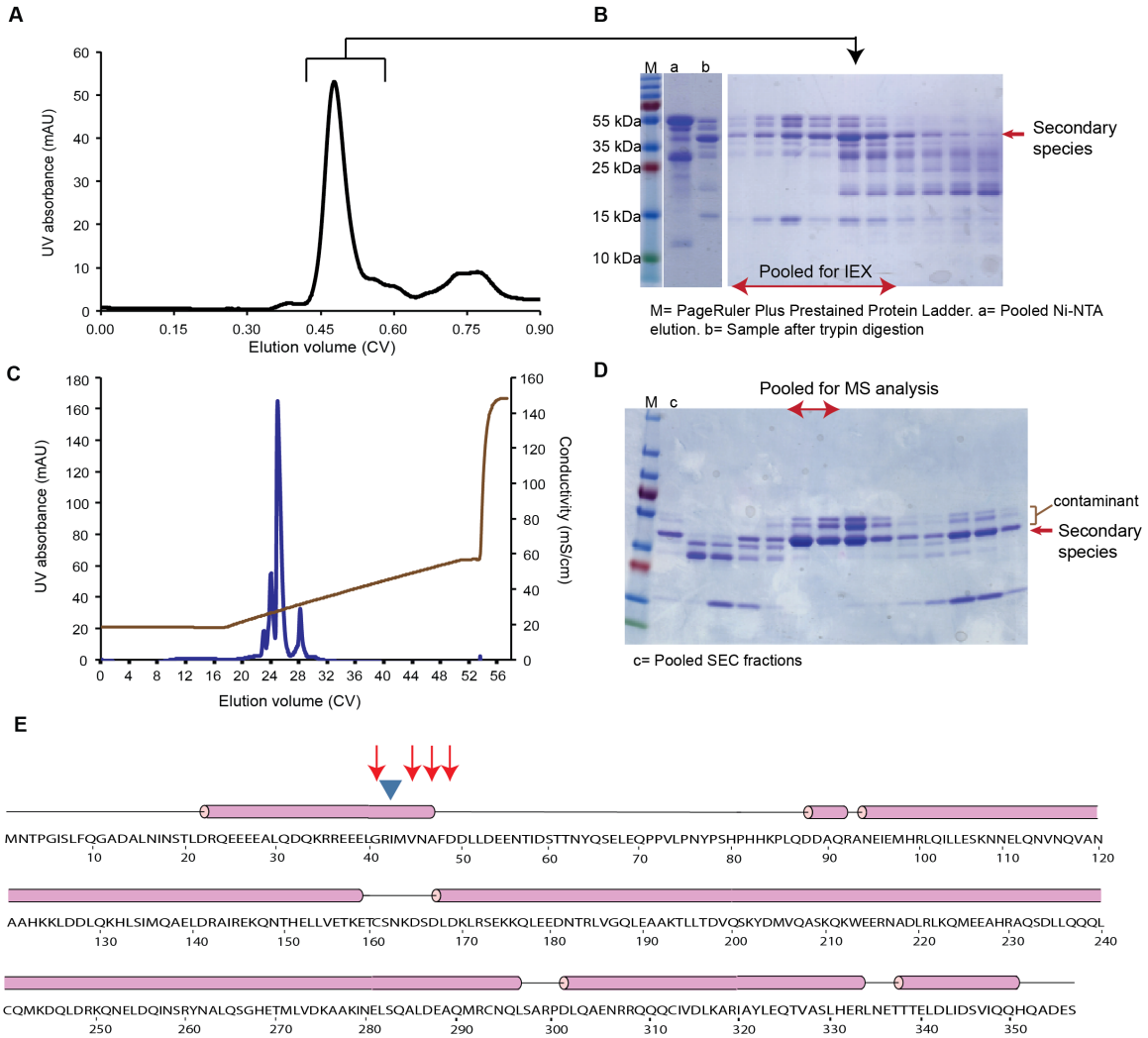


Figure 6.10 Purification of HisAsl$^{1\text{-}356}$ after trypsin digestion. (**A,B**) Elution profile from an S200 16/60 column. Elution with 50 mM Tris (pH 7.5), 150 mM NaCl and 5 mM BME. The 15% reducing SDS-PAGE gel analysis of peak fractions indicates that the more stable fragment is eluted in the major peak together with several lower molecular weight contaminants. (**C,D**) Anion exchange purification for the pooled SEC fractions, using a Mono Q 5/50 GL column. 15% reducing SDS-PAGE gel analysis showing the elution of dominant species in one sharp peak with a mid-point of ~260 mM NaCl. Higher molecular weight contaminants are still not fully removed in the peak fractions. (**E**) Sequence of Asl$^{1\text{-}356}$ showing the trypsin cleavage site used (arrowhead) during limited proteolysis. Four other constructs are designed around this site, with various N-termini marked by the *red* arrows (Asl$^{441/443/447/449\text{-}356}$).

The two peak fractions from IEX were pooled (highlighted in **Figure 6.10D**) and sent for mass spectrometry analysis. This showed that a species of 36.34 kDa was formed corresponding to

residues 43 to 356 when compared to the sequence of HisAsl$^{1-356}$. This indicated removal of N-terminal His6 tag and a small proportion of the N-terminus (*blue* arrowhead, **Figure 6.10E**).

6.2.6 Purification of HisAsl$^{47-356}$

Based on the results from limited proteolysis experiment, I designed new constructs with successively trimmed N-termini (*red* arrows, **Figure 6.10E**). These fragments were cloned into a pETM-14 vector via the same protocol for pETM-14-Asl$^{1-356}$. A high-throughput screen was first conducted to help choose the optimal growth conditions and the construct with a better yield hence suitable for large-scale expression and purification. All five constructs showed protein overexpression, and soluble protein could be obtained in all cases (**Figure 6.11**). However, a larger amount of soluble HisAsl$^{47-356}$ was expressed using B834 (DE3) cells grown in TB media. As a result, this construct was pursued for large-scale production.

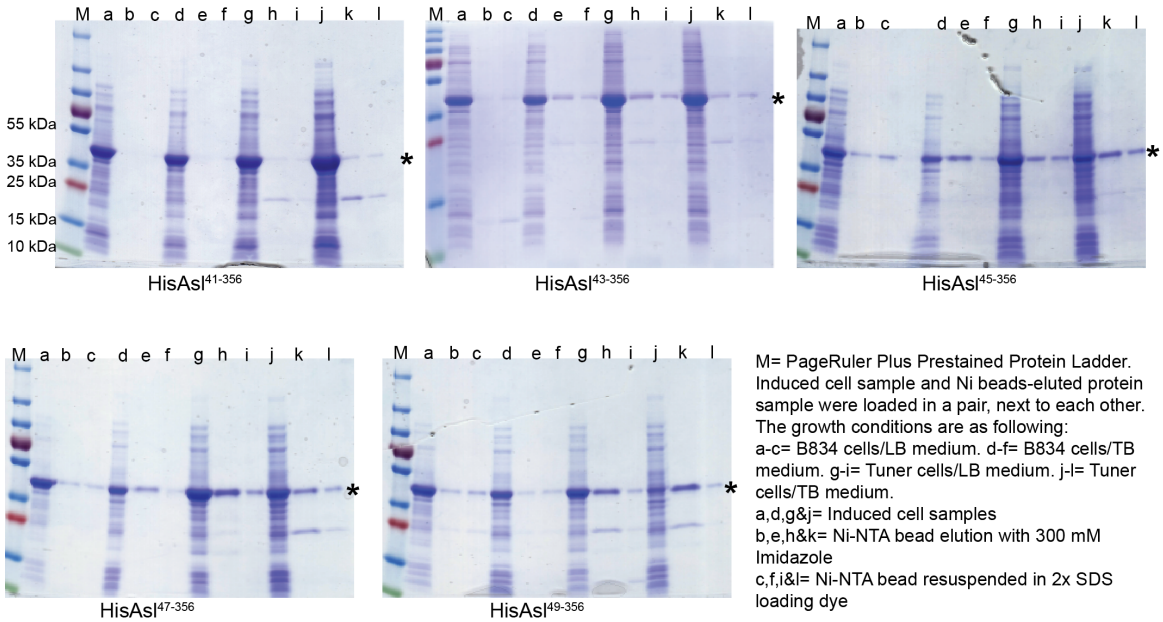


Figure 6.11 High-throughput screening for a soluble Asl N-terminal construct. 15% reducing SDS-PAGE gel analysis was performed, and asterisks indicate the bands corresponding to HisAsl constructs. All constructs showed overexpression, and some soluble protein could be purified. HisAsl$^{47-356}$ showed the largest yield of soluble protein from Ni-NTA purification, and was therefore pursued for large-scale production.

2 l B834 (DE3)/TB cell cultures transformed with pETM14-Asl$^{47-356}$ were grown (construct details in **Table 6.3**), and the cells were harvested according to the standard protocol (**Appendix 8.1.7**). Protein was purified via Ni-NTA chromatography followed by size exclusion chromatography.

Surprisingly, two major species were observed at the stage of Ni-NTA purification, even though only the full-length protein was overexpressed and detected in the supernatant of bacterial cell lysate (**Figure 6.12A**). The two constructs could be separated via SEC, since the full-length fragment eluted around 0.48 CV and the shorter species eluted around 0.6 CV (**Figure 6.12B,C**).

Construct	MW (kDa)	pI	Abs 0.1%	No. of Cys
HisAsl$^{47-356}$	38.3	5.27	0.339	4
Asl$^{47-356}$	36.2	4.98	0.358	4

Table 6.3 Construct details calculated by Protparam. Due to the presence of cysteine residues, 5 mM β-mercaptoethanol (BME) was included in working buffer (WB) to prevent disulfide bonds.

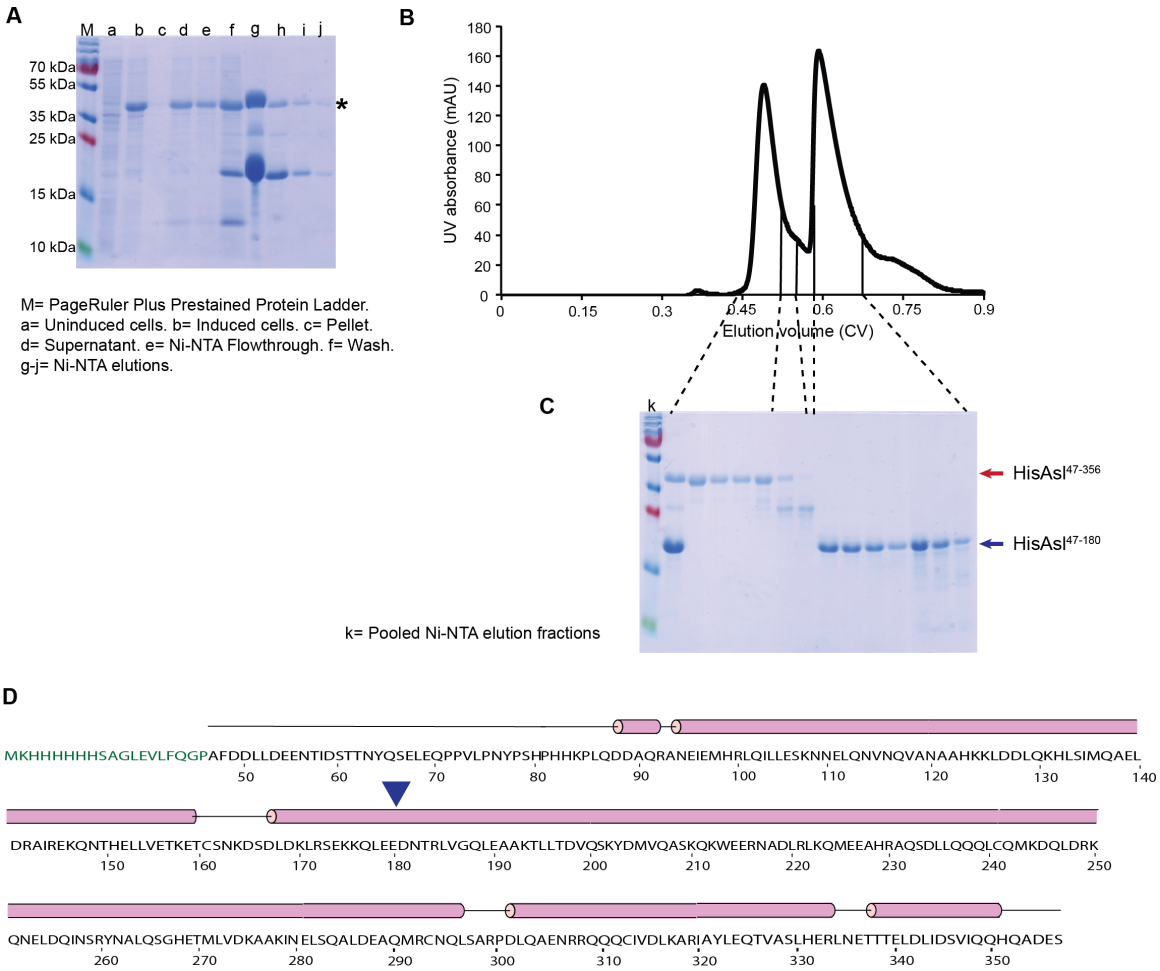


Figure 6.12 Large-scale purification of HisAsl$^{47-356}$. (**A**) 15% reducing SDS-PAGE gel analysis of Ni-NTA purification steps. Asterisk indicates the band corresponding to HisAsl$^{47-356}$. Protein was overexpressed, and a large proportion was soluble. However, a lower molecular weight band was observed at all stages of Ni purification. This band was later confirmed to be a degradation product from the full-length protein, by in-gel mass spectrometry and N-terminal sequencing. (**B**) S200 16/60 SEC elution profile for HisAsl$^{47-356}$. Elution with 50 mM TrIs (pH 7.5), 150 mM NaCl and 5 mM BME. Two major peaks are detected, and the peak fractions were taken for gel analysis. (**C**) 15% reducing SDS-PAGE gel analysis of SEC fractions in (**B**). Full-length protein eluted across the first peak (~0.48 CV), and the shorter species eluted further into the separating volume of the column. (**D**) Sequence of HisAsl$^{47-356}$ showing the C-terminal truncation (arrowhead). The His6 tag and 3C cleavage site are coloured to *green*. Predicted secondary structural elements are displayed above the primary sequence.

Mass spectrometry analysis and N-terminal sequencing were conducted in order to identify this shorter construct, and it covers residues 47 to 180 with the removal of C-terminal residues (**Figure 6.12D**). Both constructs were stable in solution, and hence were subjected to further studies.

6.2.7 Biophysical characterisation of HisAsl$^{47\text{-}180/356}$

The secondary structural contents of both constructs were checked by CD experiments at 20 °C. Sample at 0.1 mg/ml was dialysed overnight into 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.5). CD spectra were collected between 190 nm and 260 nm, and the measurement from dialysis buffer was used as background subtraction. High α-helical content profiles were observed for both Asl protein fragments, which was consistent with the PSIPRED prediction (**Figure 6.13**).

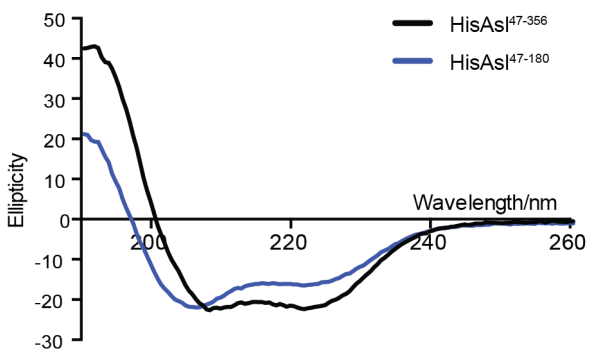


Figure 6.13 CD analysis of His6 tagged Asl N-terminal fragments in 10 mM NaH_2PO_4/Na_2HPO_4 (pH 7.5) at 20 °C. Double negative troughs were observed at 208 nm and 222 nm, respectively. Both samples showed α type profiles, which were consistent with the PSIPRED prediction..

Both samples were prepared for SEC-MALS analysis to investigate the oligomerisation of Asl N-terminal fragments in solution (see **Figure 6.14** for molar concentrations tested for each construct). An S200 10/30 column (GE Healthcare) was used for protein elution, using WB as the running buffer. Interestingly, HisAsl$^{47\text{-}356}$ existed predominantly as a dimer in solution whereas the shorter fragment was predominantly monomeric (**Figure 6.14A,B**). This result suggests that residues 181 to 356 are essential for Asl dimerisation.

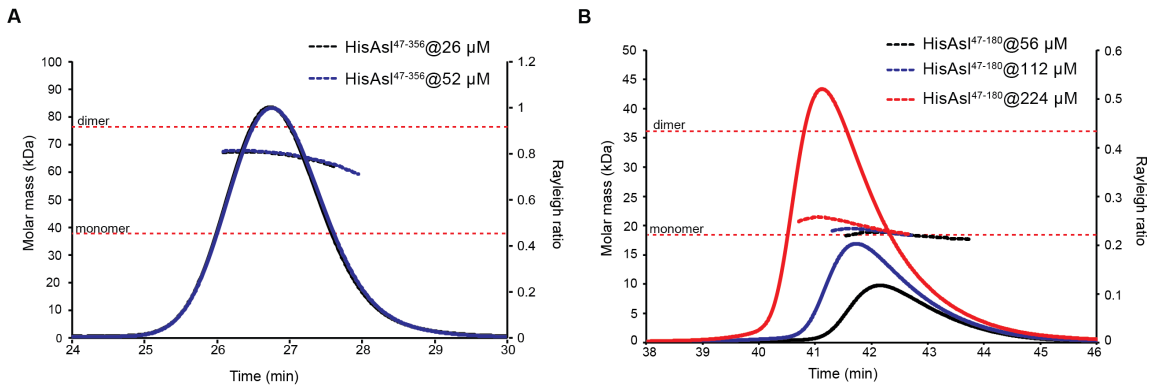


Figure 6.14 SEC-MALS analysis for HisAsl$^{47-356}$ (A) and HisAsl$^{47-180}$ (B) using an S200 10/30 column. Elution with 50 mM Tris (pH 7.5), 150 mM NaCl and 5 mM BME. Solid lines represent the Rayleigh ratio, and dashed lines indicate the observed molecular weight eluted across the peak. The expected masses of a monomer and a dimer are indicated with *red* dotted lines across the graph. The longer fragment eluted largely as a dimer in a single peak. Dimerisation is disrupted for HisAsl$^{47-180}$, suggesting the C-terminal 176 residues are essential for the efficient self-association in solution. HisAsl$^{47-180}$ does show some tendency towards self-association, but the peak appears to be predominantly monomeric.

HisAsl$^{47-356}$ had low protein solubility in WB (~5 mg/ml), and I attempted to find an optimal buffer using a fluorescence-based thermal shift assay (ThermoFluor). However, this assay is not ideal for highly coiled coil protein due to the lack of a hydrophobic core that would be seen in folded globular protein. Indeed, the increase in fluorescence upon heat-activated protein unfolding was not significant and no typical unfolding transition curve could be detected in the tested conditions (**Figure 6.15A**).

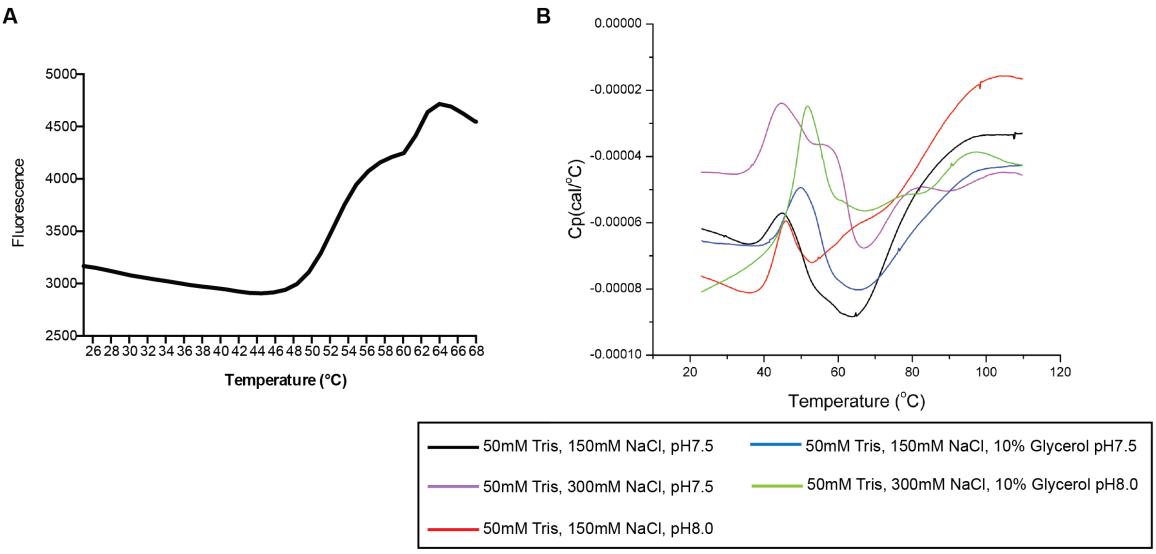


Figure 6.15 Finding the optimal storage buffer for HisAsl$^{47-356}$. (**A**) The unfolding transition of HisAsl$^{47-356}$, in 50 mM Tris (pH 7.5), 150 mM NaCl and 5 mM BME, obtained via ThermoFluor. No significant increase in fluorescence is detected upon heat-activated melting. There might not be an exposure of the protein core that can bind the SYPRO Orange dye when the helical protein unfolds. (**B**) Differential Scanning Calorimetry (DSC) studies were conducted to measure the T_m of HisAsl$^{47-356}$ in different buffer conditions. A cooperative unfolding transition curve was observed when the pH of WB was increased to pH8.0 (*red* curve). The addition of glycerol largely stabilised the protein with an increase in T_m from 44.8 °C (*black* curve) to 49.3 °C (*blue* curve), and the T_m was increased by another 3 °C (to 52 °C) when the protein was in 50 mM Tris (pH 8.0), 300 mM NaCl, 10% glycerol (*green* curve). This buffer was in turn used as the storage buffer for HisAsl$^{47-356}$.

I therefore used Differential Scanning Calorimetry (DSC) to directly measure the melting temperature (T_m) of the protein in different buffer conditions. In this type of experiment, the protein sample is heated at a constant rate and the amount of heat absorbed by the sample is plotted against temperature (known as a DSC curve); as the temperature increases the sample eventually reaches its T_m, and this melting process results in an endothermic peak in the DSC curve. Due to the requirement for large amounts of protein and the longer period of time taken for each scan, I only tested the influence of salt concentration (150 mM and 300 mM), pH value (pH 7.5 and pH 8.0) and the addition of 10% glycerol. The measured T_m was then compared with the one obtained with original WB (*black* curve, **Figure 6.15B**). Increasing the buffer pH largely stabilised the protein sample as judged by the shape of its melting curve representative of a cooperative unfolding transition (*red* curve, **Figure 6.15B**). The addition of glycerol and the higher salt concentration further stabilised the sample with an increase in T_m from 44.8 °C to 52.3°C (*green* curve, **Figure 6.15B**). Exchanging HisAsl$^{47-356}$ into the optimised storage buffer (50 mM Tris (pH 8.0), 300 mM NaCl, 10% glycerol, 1 mM TCEP) led to a 2-fold increase in protein solubility.

6.2.8 Attempts in structural characterisation of Asl$^{47-180/356}$

Both Asl$^{47-180}$ and Asl$^{47-356}$ proteins were taken forward for crystallisation trials. The N-terminal His6 tag was cleaved off using 3C protease at 1 mg 3C: 1.48 µmol Asl protein, dialysed overnight at 4 °C into WB. The cleavage reaction was exchanged into the optimised storage buffer via S200 16/60 SEC. Crystallisation trays were set up for Asl$^{47-180}$ at 20 mg/ml and Asl$^{47-356}$ at 10 mg/ml. The longer fragment was much less soluble, even in the optimal storage buffer, as fine precipitate started to appear when protein concentration reached beyond 10 mg/ml. The sample was centrifuged at 15,000x g for 5 mins before the supernatant was carefully pipetted out for putting down into trays. Initially, broad screens of crystallisation conditions (Molecular Dimensions) were conducted using the high-throughput screens from Molecular Dimensions (**Figure 6.16A**; **Appendix 8.1.10**). After 3-week incubation at 21 °C, trays for Asl$^{47-356}$ showed drops mainly with brownish precipitation, suggesting the protein was denatured in these crystallisation conditions. In contrast, Asl$^{47-180}$ crystallisation trials showed high proportion of clear drops, which were still

observed even after 1 month. I therefore performed crystallisation trials at much higher protein concentrations up to 50 mg/ml (**Figure 6.16A**). While the majority of conditions showed formation of heavy precipitation or occurrence of phase separation, multiple needle-like microcrystal bundles were grown in Condition B12 of PACT screen (100 mM MES (pH 6.0), 10 mM $ZnCl_2$, 20% PEG 6K; **Figure 6.16B**). The drop contained 0.15 µl of Asl^{47-180} at 50 mg/ml and 0.05 µl of mother liquor. These microcrystals were then harvested for data collection: this confirmed that they were protein crystals but unfortunately they diffracted poorly to only 30 Å. Optimisation of the initial crystallisation condition was then followed, in order to grow single crystals of bigger sizes and better morphologies that could diffract to a higher resolution for structure determination (**Figure 6.16C-E**).

The presence of multiple small crystals in a drop suggested the crystal growth was likely over-nucleated. Several optimisation strategies were followed. First, the protein or precipitant concentration was reduced, as the original conditions might have led to an extended nucleation phase, which in fact requires a higher degree of supersaturation than crystal growth. An optimisation screen was designed to fine-tune the precipitant concentration around the original condition that has produced microcrystals of poor diffraction quality (**Figure 6.16C**), and a tray was set up using protein stock at 40 mg/ml. Although the needle-like crystalline bundles were reproducible, no crystal of better morphologies could be produced.

Second, the protein to mother liquor ratio was optimised. Initially each drop contained 50% or 75% protein stock. In order to slow down the precipitation process, an increased proportion of mother liquor was trialled (40% protein stock and 60% mother liqour). This will not alter the final precipitant concentration in a drop, but allows a more gentle rise in protein and precipitant concentrations until equilibrium between mother liquor conditions in the well and in the drop are reached. The optimisation screen was conducted to test the new protein: mother liquor ratio at the same drop size (0.2 µl), but no individual crystals grew.

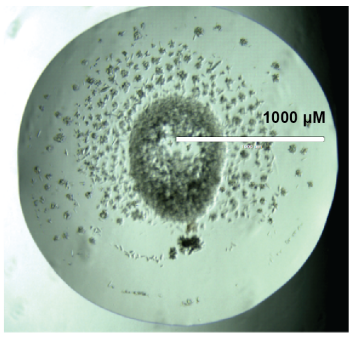
Third, streak seeding was applied in a fresh drop of protein solution to introduce crystalline nuclei that can initiate the crystal growth. The microcrystalline seeds were produced from small needle bundles that had grown in the initial condition from the broad screen. The seed stock was

prepared by vortexing the original crystals in the presence of a PTFE bead (Hampton). The stock was then serially diluted by 10-fold, as shown in **Figure 6.16D**, using the mother liquor. A small amount of diluted seed stock was included in each drop of the optimisation screen. Heavy precipitations were observed in screens supplemented with 10-fold or 100-fold diluted seed solution. The addition of 1000-fold diluted seed stock, however, still produced tiny needle-like crystalline bundles.

A

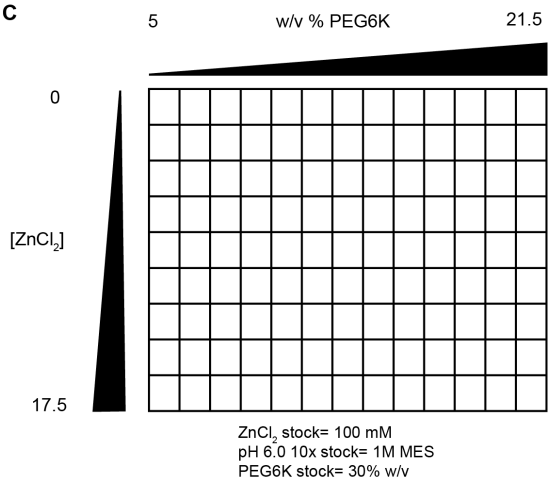
Construct	Protein concentration	Broad screens
AslA$^{47-356}$	10 mg/ml	Stura, Structure, JCSG+, Morpheus, PACT
AslA$^{47-180}$	20, 40, 45 & 50 mg/ml	Stura, Structure, JCSG+, Morpheus, PACT

B

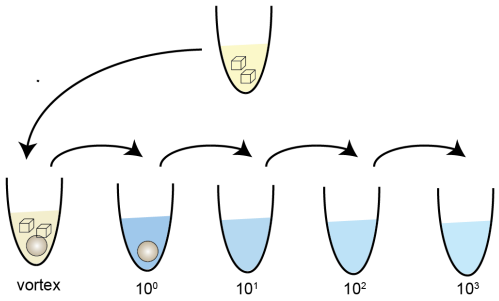


0.2 µl drop= 0.15 µl AslA$^{47-180}$ at 50 mg/ml and 0.05 µl mother liquor (100 mM MES, pH6.0, 10 mM $ZnCl_2$, 20% w/v PEG6K)

C



D



E

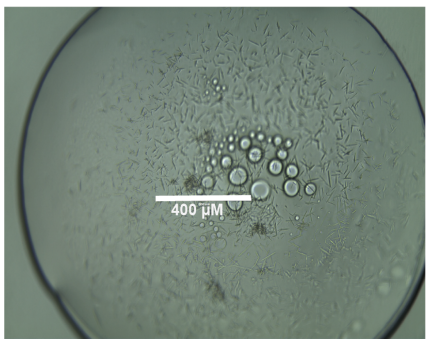


Figure 6.16 Crystallisation trials for AslA$^{47-180/356}$. Drop size was kept constant at 0.2 µl. (**A**) Broad screens were set up for both constructs with a drop size of 0.2 µl at 50% or 75% protein: mother liquor *v/v* ratio. A range of concentrations was trialled for AslA$^{47-180}$ due to its higher solubility, and heavy phase separations were observed in trays set up at lower concentrations. (**B**) Microcrystals of AslA$^{47-180}$ grew in the broad screen. **(C**) Optimisation screening was conducted to try to obtain crystals that could diffract to higher atomic resolution. Trays were set up at 75% and 40% protein: mother liquor *v/v* ratio using 40 mg/ml protein stock. (**D**) Production of microcrystalline seeds using a seed bead kit. A drop containing the microcrystals was resuspended in 50 µl mother liquor. Crystals were mechanically crushed by vortexing in the presence of a PTFE bead. This highly concentrated seed stock was then serially diluted by 10-fold with fresh mother liquor to create various concentrations of seed solution. 50 nl of such seed solution was added to each drop in the subsequent crystallization experiment. Diagram modified from [Cottee, 2014]. (**E**) The best crystals grew in optimisation screen. No single crystal could be produced.

Unfortunately, optimisation screening did not produce any crystal that could diffract to a better resolution. I therefore wanted to analyse this sample by NMR spectroscopy to determine whether it might be suitable for NMR structure determination. A 1D NMR spectrum was first collected for native Asl$^{47-180}$ at 20 °C to confirm the protein folding in solution. A 350 µl sample was prepared at 100 µM in 10 mM NaH_2PO_4/Na_2HPO_4 (pH 8.0), 100 mM NaCl, with the supplement of 5% *v/v* D_2O. The 1D NMR spectrum of Asl$^{47-180}$ is shown in **Figure 6.17**.

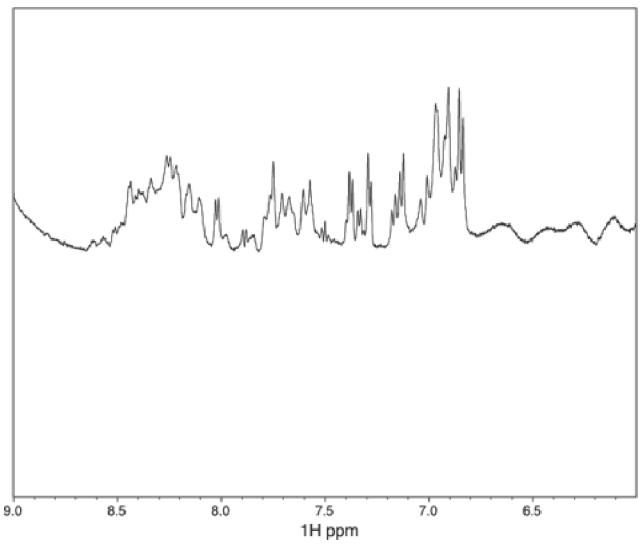


Figure 6.17 1D NMR spectrum for Asl$^{47-180}$ in 10 mM NaH_2PO_4/Na_2HPO_4 (pH 8.0), 100 mM NaCl, 5% *v/v* D_2O**.** Chemical shifts corresponding to backbone and side chain amides were observed in a narrow range with peaks located mainly between 6.8 ppm and 8.5 ppm. This suggested that the residues were not experiencing distinct environment as expected in a fully folded protein. Sharp distinctive peaks were observed between 6.8 ppm and 8.0 ppm, but the peaks at 8.0-8.5 ppm were severely overlapping further indicating Asl$^{47-180}$ might be partially folded in solution.

The peaks corresponding to protons within the amide region were fairly broad, although some peaks were sharper. This indicated some protons were not in distinctive chemical environment and hence the peak was an average of all the different environments that proton was found in [Cable, 2012]. Nevertheless, other protons were in a folded region of the protein and thus a sharp peak was obtained corresponding to a particular chemical environment that proton was found in. In addition, the peaks did not impart a broad range of chemical shifts, which would be seen in a well-folded protein due to the protons occupying different positions. Overall, the 1D NMR spectrum of Asl$^{47-180}$ suggested that the protein might be partially folded in solution. A core structure might be defined by carrying out limited proteolysis on this construct. As other projects started to work at this stage, I didn't pursue this further. Recently, the crystal structures of HsPLK4-CPB in complex with Cep152 and Cep192 were published [Park et al., 2014], and will be discussed in the context of my work in the **Perspective** section.

6.3 Purification and characterisation of the AsI C-terminal fragments

6.3.1 Purification of HisAsI$^{648/733-994}$

Initially a small-scale expression screen was carried out for the AsI C-terminal fragments (HisAsI$^{648/733-994}$), in an attempt to check whether soluble protein could be obtained and to determine the optimal growth condition. Gel analysis indicated that a small amount of soluble HisAsI$^{648-994}$ was expressed in Tuner (DE3) cells (**Figure 6.18**).

Construct	MW (kDa)	pI	Abs 0.1%	No. of Cys
HisAsI$^{648-994}$	44	8.47	0.493	4
HisAsI$^{733-994}$	34	8.84	0.595	3

Table 6.5 Details of the AsI C-terminal constructs included in the high-throughput screen. Due to the presence of cysteine residues, 5 mM β-mercaptoethanol (BME) was included in working buffer (WB) to prevent disulfide bonds.

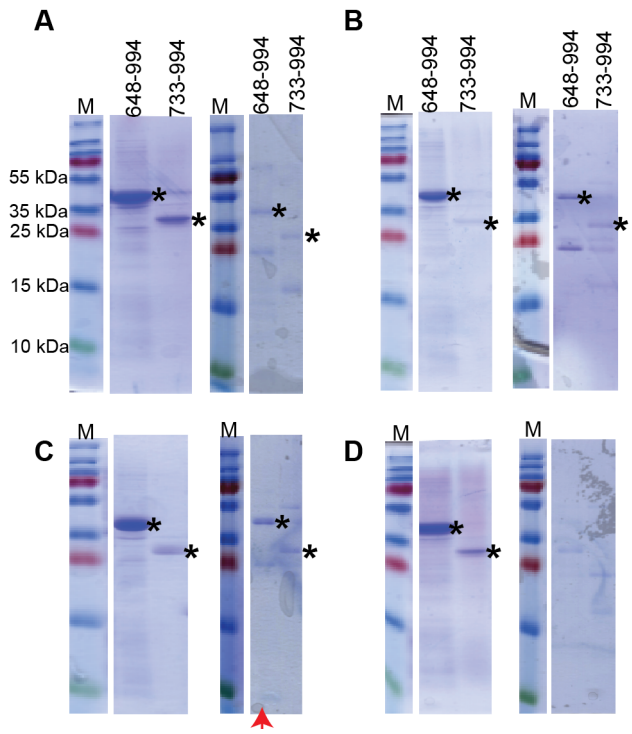


Figure 6.18 High-throughput expression and purification screen for a soluble C-terminal fragment of AsI. Different cell types and growth media were assessed for each construct. 15% reducing SDS-PAGE gel analysis was performed to confirm the protein overexpression and solubility. All cell cultures were grown at 37 °C until OD_{600} reached 0.6. 1 mM IPTG was then added to induce protein expression overnight at 21 °C. (**A,B**) Plasmids were transformed into B834 (DE3) cells that were grown in LB (**A**) or TB (**B**) media. (**C,D**) Protein overexpression using Tuner (DE3) cells in LB (**C**) or TB (**D**) media. Asterisks indicate the bands corresponding to HisAsI$^{648/733-994}$. Left, gels showing protein induction; Right, gels showing soluble protein eluted off Ni-NTA magnetic beads. The expressed HisAsI$^{733-994}$ was largely insoluble, and was therefore not pursued for large-scale purification. The optimal growth condition for HisAsI$^{648-994}$ is highlighted with a *red* arrow.

1 l of LB medium was then used for a large-scale protein production of HisAsI$^{648-994}$. Successful protein expression was confirmed via gel analysis that showed an obvious overexpression band for the construct. Protein purification was then performed using the standard protocol (**Appendix 8.1.7**). Cell lysate was purified by Ni-NTA chromatography. Unfortunately, the majority of protein was insoluble and little protein was eluted off the Ni column (Qiagen). Due to the poor solubility of

HisAsl$^{648/733-994}$, I decided to clone both constructs into a pETM-44 vector that contains an N-terminal MBP tag.

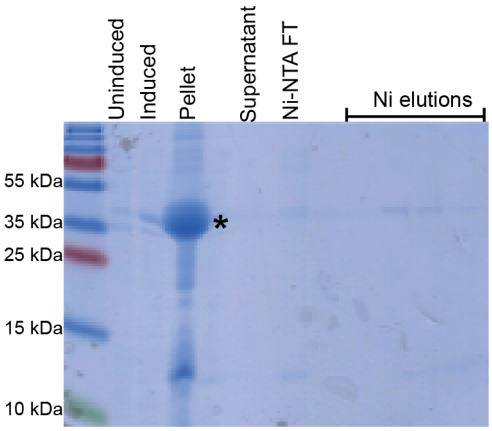


Figure 6.19 Ni-NTA purification of HisAsl$^{648-994}$. 15% reducing SDS-PAGE analysis of Ni-NTA purification fractions indicated that the majority of expressed fusion protein (indicated by an asterisk) was in the insoluble fraction of cell lysate. This construct was subsequently cloned into a pETM-44 vector that encodes for an N-terminal MBP tag.

6.3.2 Purification of MBPAsl$^{648/733-994}$

The MBP fusion proteins were expressed using B834 (DE3) cells in 1 l of TB medium, and purified using Ni-NTA column (**Figure 6.20A**; protocol see **Appendix 8.1.7**). The construct details are listed in **Table 6.6**. The Ni-NTA elution was concentrated to 5 ml using a 10 kDa MWCO centrifugal concentrator (Amicon), before loading onto an S200 16/60 column (GE Healthcare) for size exclusion chromatography, using WB as the running buffer. Gel analysis of the SEC eluted fractions indicated that MBPAsl$^{648-994}$ was eluted in multiple peaks over much of the separating volume of the column (**Figure 6.20B**). In addition, one of those peaks was close to the void volume (void volume of an S200 16/60 column= ~0.33 CV). This result suggested that the sample was not homogenous, and formed a range of species with differing stoichiometries. A lower molecular weight contaminant was also co-eluted with the full-length protein, which was likely to be free MBP tag. A similar elution profile was obtained for MBPAsl$^{733-994}$, but most protein eluted in a major peak around 0.47 CV (**Figure 6.21A,B**). The estimated yield of SEC purified MBPAsl$^{733-994}$ in the non-void peak was 1 mg per 1 l culture, which was much higher than MBPAsl$^{648-994}$ at 0.1 mg per 1 l culture. As a result, the shorter fragment was taken forward for further characterisation.

Construct	MW (kDa)	pI	Abs 0.1%	No. of Cys
MBPAsl$^{648-994}$	83.95	6.37	1.046	4
MBPAsl$^{733-994}$	73.99	6.39	1.166	3

Table 6.6 Details of the MBP-tagged Asl C-terminal constructs pursued for large-scale expression and purification. Due to the presence of cysteine residues, 5 mM β-mercaptoethanol (BME) was included in working buffer (WB) to prevent disulfide bonds.

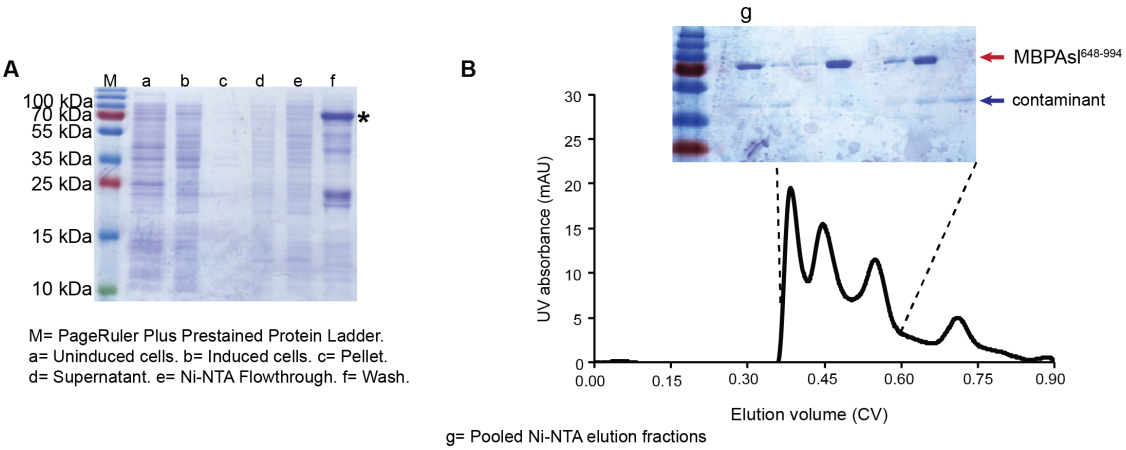


Figure 6.20 Purification of MBPAsl$^{648-994}$. (**A**) 15% reducing SDS-PAGE gel analysis of Ni-NTA purification steps. Asterisk indicates the band corresponding to MBPAsl$^{648-994}$. (**B**) Elution profile of an S200 16/60 column (elution with 50 mM Tris (pH 7.5), 150 mM NaCl and 5 mM BME) and the subsequent 15% reducing SDS-PAGE gel analysis showing the elution of full-length protein over multiple peaks. A lower molecular weight contaminant co-eluted in all peak fractions.

SEC fractions within the major elution peak of MBPAsl$^{733-994}$ were pooled for tag cleavage with 3C protease at a ratio of 1 mg 3C: 1.84 µmol target protein. The reaction was dialysed overnight at 4 °C into WB at pH 7.5. The MBP tag cleavage was complete as indicated by gel analysis (**Figure 6.21C**). However, the sample was visibly forming white precipitate, and no untagged protein was obtained after SEC purification (**Figure 6.21D**). Although the predicted isoelectric point of untagged protein was 8.85, I wondered if the observation of protein precipitates after tag cleavage was due to the buffer pH (pH 7.5) being too close to the experimental isoelectric point. As a result, the SEC buffer was adjusted to pH 7.0 with a higher salt concentration (300 mM) and the addition of 5% glycerol. Due to the low yield of MBP fusion protein, I decided to carry out a 6 l production. The harvested cell pellet was resuspended in lysis buffer that contained 20 mM Tris (pH 7.5), 200 mM NaCl, 5 mM BME (optimised WB) plus protease inhibitor cocktail tablet (Roche), DNase (Sigma) and Lysozyme (Sigma). Cells were lysed using a homogeniser (Avestin), and the lysate was purified via MBPTrap HP affinity purification (**Appendix 8.1.7**). Note that a Ni-NTA column (Qiagen) was used previously for the purification of MBPAsl$^{733-994}$; however, several contaminants were co-eluted with the full-length protein, and a large amount of fusion protein was lost in the 50 mM imidazole wash step (**Figure 6.21A**). MBPTrap HP column (GE

Healthcare) provides a higher specificity for MBP-tagged protein and might therefore reduce the amount of unspecific binding of bacterial contaminants. The column was also washed with a smaller volume of buffer before protein elution and would therefore reduce the amount of protein lost in the wash step. Protein was eluted off the MBPTrap column using the optimised WB plus 10 mM maltose. The 3C tag cleavage reaction was dialysed overnight at 4 °C into the optimised WB at pH 7.0.

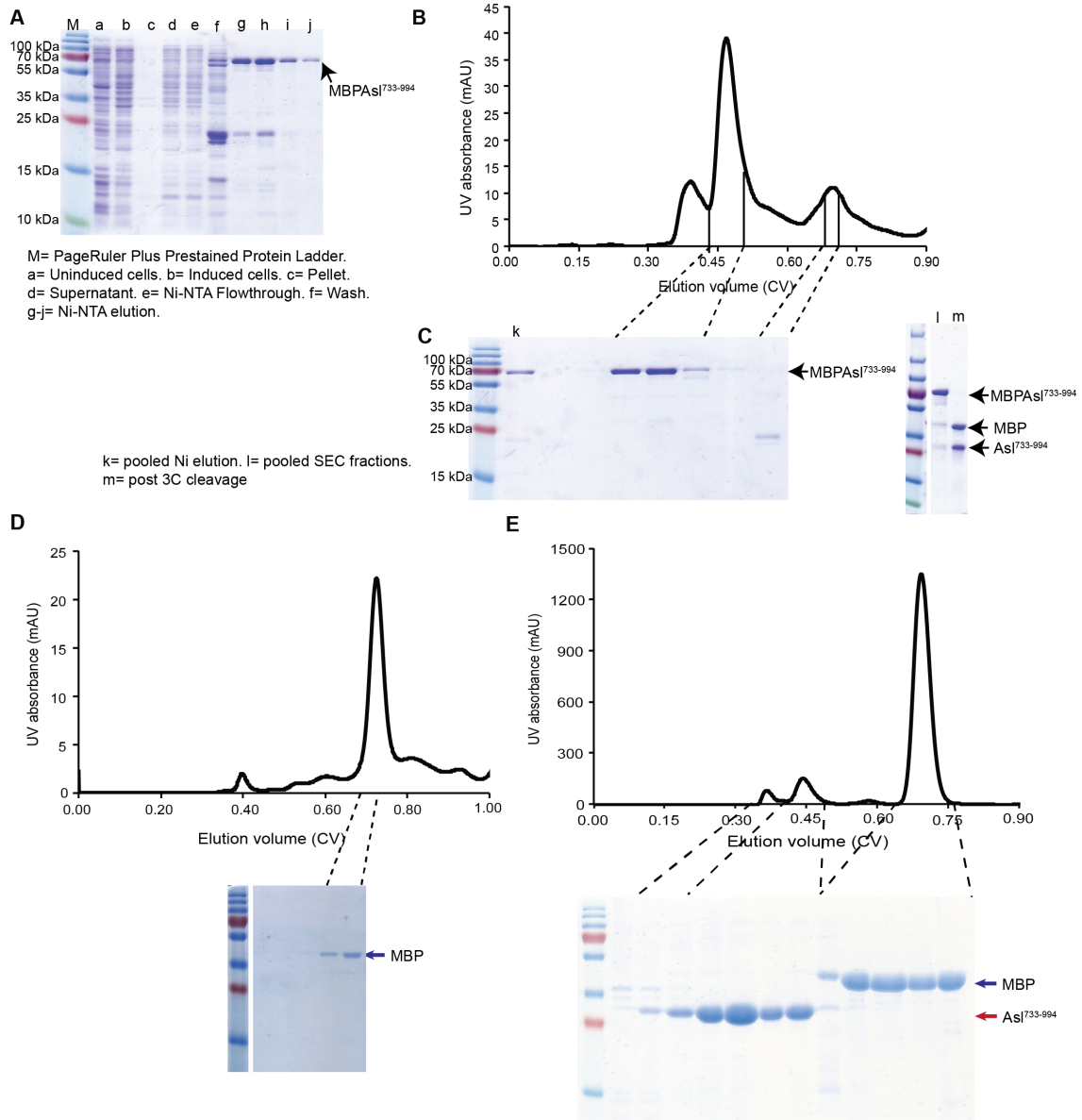


Figure 6.21 Large-scale purification of Asl$^{733-994}$. (**A**) 15% reducing SDS-PAGE gel analysis of Ni-NTA purification steps. (**B**) Elution profile for MBPAsl$^{733-994}$ off an S200 16/60 SEC column. Elution with 50 mM Tris (pH 7.5), 150 mM NaCl, 5 mM BME. (**C**) Subsequent 15% reducing SDS-PAGE gel analysis of SEC fractions showing the elution of full-length protein within the major peak around 0.47 CV (left). The pooled SEC fractions were subjected to tag cleavage by 3C protease (right). (**D**) Top, SEC elution of untagged Asl$^{733-994}$ in a buffer containing 50 mM Tris (pH 7.5), 150 mM NaCl and 5 mM BME. Bottom, 15% reducing SDS-PAGE gel analysis of fractions across the major peak showing the elution of free MBP tag. (**E**) Top, SEC elution of untagged Asl$^{733-994}$, from a 6 l production, in a buffer containing 50 mM Tris (pH 7.0), 300 mM NaCl, 5% glycerol and 5 mM BME. Bottom, 15% reducing SDS-PAGE gel analysis of SEC fractions showing the elution of untagged Asl$^{733-994}$ protein within a peak around 0.44 CV. The optimised buffer conditions successfully increased the yield of Asl$^{733-994}$.

Sample in a volume of 10 ml was loaded onto an S200 26/60 column (GE Healthcare), which was equilibrated into 50 mM Tris (pH 7.0), 300 mM NaCl, 5% glycerol and 5 mM BME (optimised SEC buffer), for SEC purification. The subsequent gel analysis showed elution of Asl$^{733-994}$ within a peak around 0.44 CV (**Figure 6.21E**).

6.3.3 Biophysical characterisation of Asl$^{733-994}$

In order to determine the oligomerisation state of untagged Asl$^{733-994}$ in solution, a SEC-MALS experiment was conducted. The SEC purified protein was concentrated using a 10 kDa MWCO centrifugal concentrator (Amicon) to 32 µM, 64 µM and 128 µM. All 3 samples were injected onto an S200 10/30 column (GE Healthcare) equilibrated in the optimised SEC buffer. Interestingly, Asl$^{733-994}$ eluted predominantly as a dimer at the lowest tested concentration with the tendency to further self-associate into a trimer as the protein concentration increased (**Figure 6.22**).

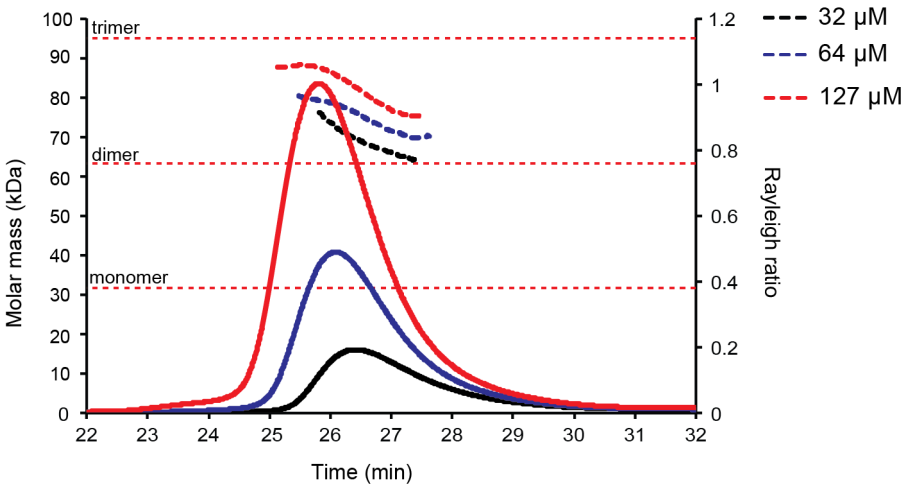


Figure 6.22 SEC-MALS analysis of untagged Asl$^{733-994}$ to investigate its oligomerisation state in solution. Elution with 50 mM Tris (pH 7.0), 300 mM NaCl, 5% glycerol and 5 mM BME. Solid lines indicate the relative Rayleigh ratio, and dashed lines show the observed molecular weight across the elution peak. Expected masses of a monomer and successive oligomers are indicated with *red* dotted lines across the graph. Asl$^{733-994}$ shows a tendency to self-associate into a trimer with the increasing protein concentration.

I then investigated the secondary structural content of purified Asl$^{733-994}$ via CD experiments. Protein sample at 0.1 mg/ml was dialysed overnight at 4 °C into 10 mM NaH_2PO_4/Na_2HPO_4 (pH7.0). CD spectra were measured at room temperature between 190 nm and 260 nm. Spectra were averaged over four accumulations, and normalised with a blank measurement taken with dialysis buffer. The spectrum does not show double negative troughs at 218 nm and 222 nm

typical of an alpha-helical type profile (**Figure 6.23**). Nonetheless, the absence of a negative peak at 200 nm expected for an unfolded protein suggests that the protein might be partially folded.

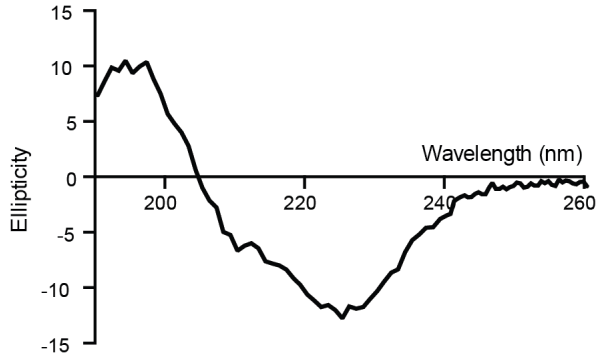


Figure 6.23 CD experiment to assess the secondary structural content of purified AsI$^{733-994}$ in 10 mM NaH_2PO_4/Na_2HPO_4 (pH7.0) at 20 °C**.** The profile does not show double negative troughs, at 208 nm and 222 nm respectively, typical of protein enriched with α-helices. However, there is no negative peak at 200 nm either, which is expected for an unfolded protein. It is likely the protein might be partially folded. This needs to be further confirmed with, for example, NMR spectroscopy.

It is difficult to draw conclusion purely based on the results from CD analysis, and further experiment such as 1D NMR spectrum collection should be followed in order to confirm whether the protein is folded properly in solution.

6.4 Perspectives

In this chapter I have attempted to purify fragments of the Asl protein containing the regions thought to interact with either Sak or Sas-4, as first identified by Dzhindzhev et al., 2010. A minimal region of residues 24 to 55 is essential for Sak binding both *in vitro* and *in vivo* [Dzhindzhev et al., 2010], and I therefore designed various N-terminal constructs with the longest length of 356 residues and the shortest length of 68 residues. Although the shortest N-terminal fragment is largely insoluble, sufficient HisAsl$^{1-86}$ protein could be purified for biophysical characterisation. SEC-MALS showed that the protein behaves as a monomer in solution. However, this protein is not folding properly, as demonstrated by the poor peak dispersions in ^{1}H-^{15}N NMR spectrum. Very recently, the structure of the CPB from HsPLK4 was published in complex with a 60-residue N-terminal fragment from HsCep152 [Park et al., 2014]. The HsPLK4 CPB consists of two PB domains, each comprising seven β strands and a α-helix, arranged into an X-shaped dimer (**Figure 6.24A**). Based on the MSA of full-length Asl sequences from various species, the 60-residue fragment in the solved structure complex is mapped onto residues 2 to 66 within DmAsl. Consistent with the oligomerisation state that I have observed on SEC-MALS profile of HisAsl$^{1-86}$, HsCep152$^{1-60}$ exists as a monomer binding to the concave surface of CPB, forming multiple hydrophobic and hydrogen bond interactions with the α1 helix of CPB1 (**Figure 6.24B**). The complex of 2:1 stoichiometry formed by a dimer of Sak CPB and a monomeric N-terminal fragment of DmAsl was further demonstrated by Shimanovskaya et al.,, although no crystal structure of such complex could be obtained [Shimanovskaya et al., 2014]. The short helix of Cep152$^{1-60}$ is aligned antiparallel to the α1 helix of CPB1 thus forming tightly packed coiled-coil interface [Park et al., 2014], (**Figure 6.24B**). This might provide an explanation for the unfolded state of Asl$^{1-86}$ in solution. It is plausible that the fragment would only fold properly once it binds to the CPB, as the formation of side chain contacts might constrain the conformation of residues into a α-helix in order to accommodate the coiled-coil interactions.

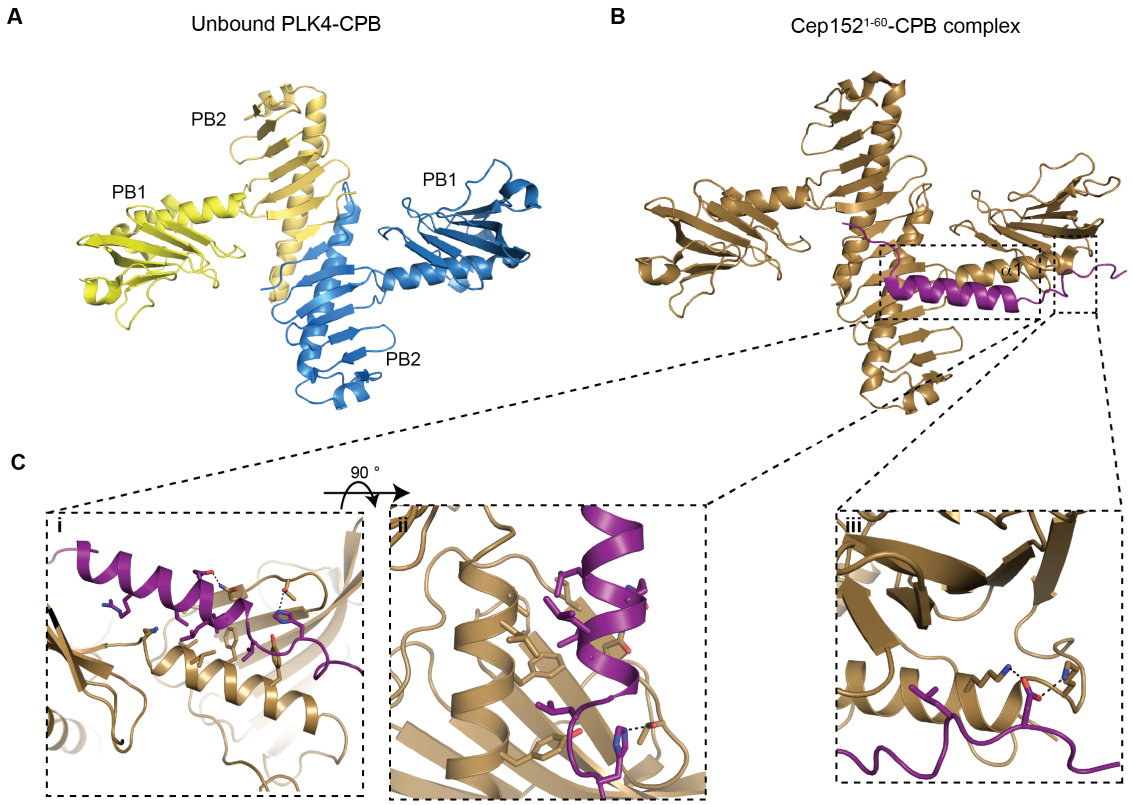


Figure 6.24 Atomic structure of human PLK4-cryptic polo box (CPB) in complex with a 60-mer fragment of Cep152 (Cep152$^{1-60}$). (**A**) Structure of the unbound CPB (PDB entry: 4N9J) revealing the dimeric interface formed between PB2 in different monomers. Each monomer comprises PB1 and PB2. (**B**) Structure of CPBs bound with Cep152$^{1-60}$ (PDB entry: 4N7V). Multiple hydrophobic and hydrogen bond (highlighted in *black* dashed line) interactions occur between the two α-helices, as highlighted in (**C**). Note the helix of Cep152 is aligned antiparallel to the α_1 helix of CPB allowing tightly packed CC interactions (**i**, **ii**). A hydrogen bond network is observed at C-terminus of Cep152$^{1-60}$ involving a D-rich motif that is conserved within the CPB-Cep152/Cep192 binding interface (**iii**).

It is also interesting that, although the N-terminal 180 residues are observed as monomers in solution, Asl$^{47-356}$ appears predominantly dimeric. This observation is consistent with a very recent publication that found the same self-association state for Asl$^{1-357}$ [Klebba et al., 2015]. What is the importance of Asl dimerisation via its N-terminus? In the GST-pull down assay carried out by Dzhindzhev et al., Asl$^{1-200}$ was not able to bind Sak whereas Asl$^{1-300}$ could (even though the internal deletion of Asl$^{24-58}$ can abolish the binding of full-length Asl to Sak CPB), [Dzhindzhev et al., 2010]. This suggests that Asl$^{200-300}$ is important for Sak interaction, although these residues are not directly involved in the binding interface. One possibility is that these residues might be required for stabilisation of the assembled Asl: Sak complex via dimerisation of Asl NT. Without the atomic structure of Asl$^{47-356}$, it is difficult to map the dimerisation interface. It would be interesting to delete residues 181 to 356, a dimerisation region narrowed down by SEC-MALS analysis, and assess the influence of this deletion in centriole duplication. It might have important

implications for the molecular mechanisms of centriole duplication if Asl needs to function as a dimer.

Apart from its N-terminal dimerisation region, Asl contains another self-ssociation domain at the C-terminus. Klebba et al., detected homotrimer formation via Asl$^{626-994}$ at 50 μM in solution [Klebba et al., 2015]. My *in vitro* characterisation of Asl$^{733-994}$ also identified a self-interaction to form predominantly dimeric and trimeric species towards higher protein concentrations, although the sample is not entirely trimeric even at 128 μM. This may indicate that the shorter fragment cannot oligomerise as efficiently as the longer fragment. Intriguingly, these authors identified a putative interaction between the recombinantly expressed Asl$^{626-994}$ and the Sak CPB [Klebba et al., 2015]. This interaction was found to stabilise Sak and was sufficient to trigger centriole amplification in Asl-depleted S2 cells. This observation contradicts the earlier publication by Dzhindzhev et al., who could not detect any interaction between Asl$^{501/701-994}$ and Sak using GST-pull down assays [Dzhindzhev et al., 2010].

The difference in interaction results mentioned above might be explained by the different protocols followed for protein binding assays. Dzhindzhev et al., produced Asl fragments by *in vitro* transcription-translation, and the reaction mixture containing the expressed protein was then incubated with PLK4-GST beads [Dzhindzhev et al., 2010]. In the later study performed by Klebba et al., the recombinant proteins were mixed at a molar ratio of 1:4 Asl$^{622-994}$: GST-CPB [Klebba et al., 2015]. Because it is unclear what molar ratio of the protein components was incubated at in the early study, it might be possible that the binding strength is low in solution, and only when an excess amount of CPB is present Asl$^{622-994}$ is then able to interact. Further *in vitro* binding assays such as MST and Surface Plasmon Resonance (SPR) should be carried out to confirm this interaction. It would be important to solve the atomic structure of Asl$^{622-994}$: CPB complex in order to identify the residues involved in this interaction. Mutational study could then be followed to investigate the specific role of Asl$^{622-994}$: CPB interaction, if any, *in vivo*.

The C-terminal 330 residues of Asl (Asl$^{664-994}$) were also found to interact with Spd-2 by yeast-two-hybrid assay [Conduit et al., 2014b]. In this project, I have established a protocol for purifying large amount of Asl$^{733-994}$. This fragment could then be used for interaction study with Spd-2 and

for crystallisation trials. The complex oligomerisation pattern observed within full-length Asl protein, with its N-terminal region involved in dimerisation and C-terminal region in trimerisation, might be essential for the concentration-dependent PLK4 kinase stability around the centriole wall. Currently, this idea remains highly speculative. Therefore, obtaining an atomic structure for the Asl C-terminal region could help to reveal the oligomerisation interface and to explore the potential biological significance of trimer formation using *in vivo* techniques.

There are many interesting questions remaining to be investigated, and solving the atomic structures of Asl functional domains would be essential for understanding the biological mechanism *in vivo*. Unfortunately the individual homo-oligomerisation domain might not fold properly in solution, as demonstrated in this project, making the structural study very difficult. Co-crystallisation of Asl with its binding partners seems a better alternative for future studies.

Chapter 7: Discussion

7.1 Summary

My thesis work has focussed on trying to understand the molecular mechanisms underlying PCM assembly in *Drosophila*. During centrosome maturation, the PCM expands greatly in size to recruit many components including γ-tubulin and consequently increases its MT nucleation ability [Conduit et al., 2015]. For decades, the PCM was regarded as an amorphous matrix of unstructured and/or partially folded proteins and is therefore inherently difficult to study. The underlying molecular scaffold, which recruits all the other PCM components, is composed of highly coiled coil proteins. Such intrinsic flexibility allows the proteins to interact with many partners and in turn allows the accumulation of hundreds of proteins at the PCM. The PCM is crucial to the function of the centrosome, and therefore a detailed understanding how the PCM is assembled and organised is essential.

Recent studies carried out in the lab have highlighted the importance of two key PCM proteins, Cnn and Spd-2, in flies [Conduit et al., 2014b; Conduit et al., 2014a; Conduit et al., 2015]. Intriguingly, PCM assembly is completely abolished when both proteins are absent, and partially disrupted if either protein is missing. Asl provides a molecular platform linking centriole duplication and centrosome maturation. In this thesis, I have identified functional domains within these essential PCM proteins. Using X-ray crystallography and NMR spectroscopy, I have obtained high-resolution structures that highlight the potential interfaces present in these proteins in *Drosophila*. The biological importance of these interfaces was then investigated *in vivo*. In addition, biophysical characterisation of the functional domains generated further molecular insights. These results together have provided important information towards understanding how Cnn and Spd-2 proteins assemble into a molecular scaffold specifically at the centrosome.

Cnn is recruited, possibly as parallel dimers, to its binding sites around the centriolar wall, and a highly conserved CM2 domain is necessary for its localisation at the centrosome (**Chapter 3**). Once localised at the centrosome, Cnn proteins are phosphorylated by active Polo kinases that are concentrated there (**Chapter 2**). Multisite phosphorylation within the Cnn PReM domain promotes the self-interaction of Cnn molecules into higher order oligomers, which then spread outwards into the peripheral region of the PCM. Interestingly, the CM2 domain is also essential for the maintenance of assembled Cnn scaffolds, which is a novel function not being identified in any other homologs (**Chapter 3**). Based on the crystal structure of Cnn-CM2 and the biological data obtained so far, I suspect that this highly conserved motif stabilises the phosphorylation-driven Cnn scaffolds by forming lock-like tetramers via the extreme C-termini of different Cnn molecules.

When Cnn scaffold-like structures did not assemble efficiently at the centrosome, Spd-2 did not fully expand into the PCM region (**Chapter 2 and Chapter 4**). This result suggests that, unlike Cnn, Spd-2 may not be able to self-assemble into a scaffold, and so might rely on Cnn assembly for its outward expansion into the PCM. Although I have preliminary result suggesting that the Spd-2 ASH1 domain may bind to CM2 (**Chapter 5**), this has yet to be proven and will be a focus of future experiments. The biological importance of this interaction will also be investigated *in vivo*. One possibility is that Spd-2 might bind to Cnn scaffolds via its ASH domains for expansion into the PCM region. ASH domains are highly conserved in evolution, it would therefore be interesting to investigate whether such domains are also involved in the PCM scaffold formation in other species. The assembled molecular scaffolds of Cnn and Spd-2 then recruit many other PCM components such as Grip71WD, Aurora-A and γ-tubulin, which subsequently nucleate and anchor MTs at the centrosome [Conduit et al., 2014b; Dobbelaere et al., 2008; Megraw et al., 1999; Megraw et al., 2001; Megraw et al., 2002; Vaizel-Ohayon and Schejter, 1999; Zhang and Megraw, 2007].

7.2 PCM expansion during mitosis

7.2.1 Temporal and spatial control

The centrosome cycle is a highly regulated cellular process, during which each centriole only duplicates once every cycle and the newly formed daughter centriole only recruits PCM once it is disengaged from the mother. Recent studies suggest that Asl is only incorporated at the disengaged daughter centriole and hence converts it into a new mother for centriole duplication in the next cell cycle [Novak et al., 2014]. In fly embryos, Asl also plays an important role at the mother centriole to recruit mitotic PCM by helping to recruit Spd-2 and Cnn [Conduit et al., 2010].

At the onset of mitosis, the PCM expands drastically in size and increases its ability to nucleate MTs, preparing the centrosome for chromosome segregation. In this thesis, I have shown that Cnn phosphorylation by Polo kinase is essential for its PCM scaffold assembly in flies. During the course of my project, similar observations were made in *C. elegans* system where SPD-5, a potential functional homolog of Cnn, forms scaffold-like structures upon PLK-1 phosphorylation both *in vivo* and *in vitro* [Woodruff et al., 2015]. Although Polo kinase is also present during interphase, it is likely largely inactive and only activated at later stages of the G2 phase [Bruinsma et al., 2012; Daub et al., 2008]. One possible explanation is that Polo binding and subsequent phosphorylation of a particular protein is normally primed either by self-phosphorylation or phosphorylation by another kinase such as a Cyclin-dependent kinase (CDK) [Archambault and Glover, 2009; Zitouni et al., 2014]. This would ensure the centrosome maturation is properly coupled to the progression of the cell cycle. Although I failed to identify phosphorylation sites within *D. melanogaster* Spd-2, Polo binding sites were identified within *C. elegans* SPD-2, which matches with a CDK1/2 consensus site [Decker et al., 2011]. CDK1/2 kinase is largely active during mitosis hence correlating with the activation of Polo/PLK1 kinase.

Centrosome maturation is not only temporally regulated, it also specifically occurs around centrioles reflecting a spatial control. Phosphorylation seems to mediate the conversion of freely diffusing molecules into the PCM matrix. Centrioles may present binding sites for PCM components. For example, Spd-2 appears to be recruited, at least in part, via binding to Asl (a centriole protein) [Conduit et al., 2010; Conduit et al., 2014b; Giansanti et al., 2008]. In addition,

the centrosomal localisations of Spd-2 and Polo kinase largely overlap ([Fu and Glover, 2012]; unpublished data); therefore the Spd-2-dependent recruitment of Cnn also allows efficient Cnn phosphorylation by Polo. Intriguingly, Cnn proteins in the phosphomimetic form can form cytoplasmic foci, suggesting the presence of Polo specifically around centrioles may limit the region of centrosome assembly. Moreover, an important feature emerging from my work is that the protein interactions are of low affinity *in vitro*. This, however, might play an essential role *in vivo* to restrict the timing and localisation of these interactions and hence ensure scaffold assembly does not occur in the cytoplasm. In *C. elegans*, SPD-5 and SPD-2 proteins exist as monomers with no detected protein interaction in the cytoplasm [Wueseke et al., 2014]. One key question would therefore be how are these interactions regulated to ensure robust centrosome assembly.

7.2.2 Feedback mechanism

In *Drosophila*, Cnn and Spd-2, the two essential PCM proteins, are co-dependent on each other to create a positive feedback that can drive the expansion of PCM scaffolds at the mitotic centrosome [Conduit et al., 2014b; Conduit et al., 2015]. In fly embryos, Spd-2 is required for Cnn incorporation to the centrosome, and it is largely co-localised with Polo at the centrosome [Conduit et al., 2010; Conduit et al., 2014b; Fu and Glover, 2012]. Cnn, however, is also essential for the maintenance of Spd-2 scaffold at the centrosome [Conduit et al., 2014b; Fu and Glover, 2012]. Therefore, Cnn scaffolds allow more Spd-2 and Polo proteins to accumulate at the centrosome, which in turn allow more Cnn to be incorporated into the PCM region.

7.2.3 A unified model of centrosome maturation?

Although the mechanism underlying mitotic PCM assembly in flies and *C. elegans* start to unveil in recent years, it remains unclear whether CDK5RAP2 and Cep192, the vertebrate homologs of Cnn and Spd-2, also constitute a molecular scaffold that recruits many other PCM components in vertebrate cells. Both proteins play important role during mitotic PCM assembly in vertebrates. In

addition, Pericentrin is also essential for centrosome maturation during mitosis. PLP however is dispensable for mitotic PCM formation, but is required for efficient recruitment of PCM proteins during interphase in flies. Such differences might reflect the essence of multiple components in higher order species due to its evolutionary complexity. The vertebrate homologs of Cnn, Spd-2 and PLP are much larger in size and predicted to contain many regions of coiled coils, which make them strong candidates for scaffold proteins. Intriguingly, the number of ASH domains within Spd-2/Cep192 increases significantly from *Drosophila* to Human and its worm homolog has the lowest number (from my FFAS search). In my thesis, ASH domains are shown to be essential for PCM expansion of Spd-2 scaffolds. Thus, I think the difference in protein domain organisation is also closely related to the organism complexity.

7.3 Future directions

How is the CM2 domain involved in Cnn/CDK5RAP2 localisation at the centrosome? Previous studies have suggested a Pericentrin: CM2 interaction is required to target CDK5RAP2 to the centrosome [Kim and Rhee, 2014; Wang et al., 2010], however PLP is not essential for Cnn recruitment during mitosis in flies [Conduit et al., 2010; Lerit et al., 2015; Richens et al., 2015]. Since the primary sequence and the biological function of CM2 are highly conserved through evolution, it is likely that a common binding partner might be present at the centrosome for Cnn and CDK5RAP2. As mentioned in the **Introduction**, the recruitment of Cnn appears to depend on Spd-2 and Asl from antibody injection experiments in fly embryos [Conduit et al., 2010]. However, previous results from Y2H assay carried out in the lab did not show any potential interaction between the C-terminal 383 residues of Cnn including the CM2 domain and the fragments of Spd-2 or Asl [Conduit et al., 2014b]. This cannot rule out the possibility that a particular protein construct was not expressed and folded properly hence inhibiting the binding. In addition, due to the lack of domain boundary information in previous studies, fragments used in the Y2H assay could be less than optimal. In my thesis, I have established protocols for purification of several functional domains within Cnn, Spd-2 and Asl proteins, which can then be used for interaction analysis with the CM2 domain in future experiments.

How is Spd-2 targeted to centrioles? A Y2H assay has demonstrated an interaction between the Asl C-terminal region ($Asl^{663-994}$) and the Spd-2 N-terminal 380 residues [Conduit et al., 2014b]. In addition, the observation of fast fluorescence recovery from Spd-2 protein lacking the ASH domains, after photobleaching, further supported the idea that the Spd-2 N-terminus is involved in centriolar targeting. I have established a protocol to produce decent yields of purified $Asl^{733-994}$ that can then be used to map its interaction region within Spd-2. Based on bioinformatic analysis, the Spd-2 N-terminal region does not contain a high secondary structure content. Nevertheless, a short CC domain is predicted at the N-terminus and might be a potential target for binding analysis.

How is Spd-2 scaffold assembly regulated at the centrosome? Spd-2 at the centrosome presents in two forms: phosphorylated and non-phosphorylated. Previous work in *C. elegans* indicated that the phosphorylated Spd-2 limits the amount of PCM expansion at the centrosome [Decker et al., 2011; Yang and Feldman, 2015]. But which sites are specifically phosphorylated at the centrosome and consequently regulate the Spd-2 scaffold assembly in flies? This needs to be investigated in the future. Post-translation modification can subsequently strengthen the binding of Spd-2 molecules onto Cnn scaffold structures or might promote the self-association of Spd-2 molecules at the centrosome.

Due to the large protein size and difficulty to express the full-length construct, in my thesis work I have recombinantly expressed protein fragments of shorter lengths using a bacterial system. However, several functional domains still cannot be expressed in soluble form or purified to a decent yield. A possible alternative in the future is to use *Drosophila* S2 cells for large-scale protein expression and purification. In addition, some protein fragments, even though they can be purified *in vitro*, were not fully folded. This might be due to poor construct design, but can also be an intrinsic property of the protein that possess flexibility for binding many partners. As a result, protein fragments may be co-expressed with their putative interaction partners to allow the construct to adopt a folded state upon binding. Purification and structural characterisation of such complexes would also help to identify potential interface residues for *in vivo* study.

7.4 Conclusion

In summary, my results have shown that Cnn localisation at the centrosome involves an evolutionarily conserved CM2 domain, and that Cnn is phosphorylated by Polo kinase within a phosphorylation-driven multimerisation (PReM) domain to drive the mitotic PCM expansion specifically at the centrosome. The CM2 domain then stabilises the assembled Cnn scaffolds against MT pulling forces. Spd-2, another essential PCM protein, is recruited to the centriole via its N-terminal region. The C-terminal ASH domains are involved in expansion of Spd-2 proteins into the PCM region. My results provide molecular understanding of how PCM scaffolds might be assembled and organised at the mitotic centrosome, which, until now, has remained largely uncharacterised. The future investigation would aim to help understand how these PCM proteins are incorporated at structural level and the molecular regulation to ensure centrosome maturation specifically occur around centrioles.

8.0: Appendix

8.1 Standard laboratory methods

8.1.1 Primer design, PCR amplification and ligation

For the non-gateway cloned vectors, primers were designed carrying ~15 bp homology to the vector backbone and ~15 bp to the target sequence. The target sequence was amplified using the Phusion PCR Master Mix (NEB) (**Table 8.1,8.2**). DpnI (Roche) treatment was followed to digest the double-stranded template DNA (1 µl DpnI to 50 µl reaction mix for 1 hr at 37 °C). The resulting fragment carried flanking arms that were homologous to the insertion sites of the vector. For a fragment that did not contain the same restriction sites as for ligation cloning, double restriction digestion was then performed, and the fragment was cloned into the end-destination vector via T4 ligation (NEB) (**Table 8.3**). However, in some cases, the restriction sites for vector insertion were also found within the inserts, and InFusion® cloning (Clontech) was used as an alternative (**Table 8.4**). This involves a homologous recombination between the insert and vector. For the pETM vectors cloned in this thesis, the inserts were always cloned in using NcoI (Roche) and NotI (Roche) restriction sites. This created an insertion of Gly residue after the initiator Met in order to maintain the ORF. For constructs that did not contain a Met in sequence, one was inserted.

Reaction component	Volume (µl)
2x Phusion Master Mix	25
DNA template (10 ng/µl)	1
Forward primer (10 µM)	2.5
Reverse primer (10 µM)	2.5
aH_2O	19

Table 8.1 Set up of a PCR reaction.

Segment	Cycles	Temperature	Time
1	1	98 °C	2 minutes
2	35	98 °C	10 seconds
		T_m+3 °C	30 seconds
		72 °C	1 minute/kb of fragment length
3	1	72 °C	10 minutes

Table 8.2 PCR cycling parameters. Note the annealing temperature is 3 °C above the T_m for both forward and reverse primers.

Reaction component	Volume
Insert	3-fold molar ratio of vector volume
Vector	2 µl – volume of insert
T4 ligase (NEB)	1 µl
aH_2O	8 µl

Table 8.3 T4 ligation reaction is set up as above, and is incubated at room temperature for at least 2 hours. A molar ratio of 1:3 insert: vector is used in all reactions.

Reaction component	Amount
Insert	50-200 ng
Vector	100-300 ng
aH_2O	To final of 10 µl

Table 8.4 In-Fusion reaction set up. The insert and vector are mixed at a 2:1 molar ratio. The 10 µl of vector + insert DNA + H_2O are added to In-Fusion Dry-down pellet, and mix well by pipetting up and down. The reaction is incubated at 37 °C for 15 mins, followed by another 15 mins at 50 °C. The mixture is allowed to cool down to 4 °C before diluting with 40 µl TE buffer (10 mM Tris-HCl, 1 mM EDTA) pH 8.0.

8.1.2 Gateway cloning

For Gateway®-cloned vectors, primers were designed to incorporate att*B* sites into the PCR product. The forward and reverse primers were designed as following:

Forward primer: 5'–GGGGACAAGTTTGTACAAAAAAGCAGGCTTC (~18 gene-specific nucleotides)–3'

Reverse primer: 5'–GGGGACCACTTTGTACAAGAAAGCTGGGTC (~18 gene-specific nucleotides)–3'

The PCR amplified fragment was first cloned into a Donor vector via attBP reaction (Life Technologies) and subsequently into a Destination vector via attLR reaction (Life Technologies). In this thesis, the Destination vectors used include pRNA-GFPNT, pRNA-GFPCT, pUbq-GFPNT

and pUbq-GFPCT. BP and LR reactions were set up according to manufacturer's protocol (Life Technologies).

8.1.3 Site-directed mutagenesis

Point mutations and internal deletions were introduced within the full-length cDNA using QuikChange II XL Site-Directed Mutagenesis (Agilent). The desired mutation (or deletion) was in the middle of the primer with 15 bp of correct sequence on both sides. Primers had a melting temperature of ≥ 78 °C, with a minimum GC content of 40%. The reactions were prepared as in **Table 8.5**, and the cycling parameters are listed in **Table 8.6**. Each reaction was then supplemented with 1 µl DpnI and incubated for 1 hr at 37 °C to digest the template DNA.

Reaction component	Volume (µl)
10x reaction buffer	5
DNA template (10 ng/µl)	1
Forward primer (10 µM)	125 ng
Reverse primer (10 µM)	125 ng
dNTP mix	1
QuikSolution	3
aH2O	To final of 50 µl
PfuUltra HF DNA polymerase	1

Table 8.5 Set up for a site-directed mutagenesis reaction.

Segment	Cycles	Temperature	Time
1	1	95 °C	1 minute
2	18	95 °C	50 seconds
		60 °C	50 seconds
		68 °C	1 minute/kb of plasmid length
3	1	68 °C	7 minutes

Table 8.6 Cycling parameters for the QuikChange II XL method.

8.1.4 Bacterial transformation

All bacterial transformations were carried out following the heat shock protocol. 1 µl of ~100 ng/µl DNA was added to 30 µl chemically competent *E. coli* cells in a 15 ml falcon tube, and were

subsequently incubated on ice for 30 mins. Cells were heat shocked at 42 °C for 1.5 mins, before a 5 min-incubation on ice. 270 µl 2x YT medium was added, and the cells were allowed to recover at 37 °C with shaking for 1 hr. Cells were then plated onto LB agar plates (LB medium, 16.7 g/l agar) containing the appropriate antibiotic for overnight growth. For Zeocin^{TM}-resistant plasmids, low salt LB medium was used for making the agar plates.

8.1.5 Small-scale expression and purification trials

In order to assess the optimal growth condition for each construct and to highlight those constructs that would give a higher level of expression and solubility, a high throughput screen was devised. In such trials, each fragment was tested in the following conditions:

1) Cell type- B834 (DE3) and Tuner (DE3) (otherwise stated in the main chapter)

2) Media- LB and TB

Each construct was transformed into *E. Coli* cells using the standard protocol mentioned above. 5 ml of LB with appropriate antibiotic was added to each well of a 24 deep-well block (Thermo Fisher Scientific). Each was then inoculated with a single colony from the plated transformation. The sealed block was shaken at 37 ^{o}C overnight. On the next day, 100 µl of the overnight inoculation was added to 5 ml media in a 24 deep-well block (Thermo Scientific). Cells were grown at 37 ^{o}C until OD_{600} reached 0.6-0.8. 1 mM IPTG was then added to induce the overnight protein expression at 21 ^{o}C. 400/OD_{600} µl of culture was transferred into a PCR plate, which was then centrifuged at 4,000 rpm at 4 ^{o}C for 15 mins to pellet down the cells. To each well, 100 µl of 1x SDS loading dye was added to resuspend the pellets. These gel samples were frozen at -21 ^{o}C, which would later be analysed on SDS-PAGE for the level of protein overexpression. The remaining cells were harvested and lysed using the Lysis buffer (**Table 8.7**). The lysate supernatant was transferred to each well of a flat-bottomed 96-well plate with magnetic Ni-NTA beads (Qiagen). The plate was left on shaker for 30 mins at room temperature at 600 rpm, in order to well mix the supernatant and magnetic beads. The unbound supernatant was carefully removed while placing the plate on a magnet for at least 1 min. To wash away any unspecific binding, 200 µl of Wash buffer (**Table 8.7**) was pipetted into each well. After mixing on a shaker for 5 mins, the plate was placed on magnet for 1 min and the buffer was removed. This wash step

has been repeated before elution. Beads were either eluted using Elution buffer or were boiled in 35 µl of 3x SDS-PAGE loading dye for direct analysis by SDS-PAGE.

Buffer	Components
Lysis buffer	Working buffer plus 10 mM imidazole and 2% v/v Tween20
Wash buffer	Working buffer plus 20 mM imidazole and 0.05% v/v Tween20
Elution buffer	Working buffer plus 500 mM imidazole and 0.05% v/v Tween 20

Table 8.7 Buffers used in high throughput expression trials. The lysis buffer was further supplemented with protease inhibitor cocktail (Roche; 1 tablet per 40 ml buffer), DNase I (50 µg/ml) and Lysozyme (1 mg/ml).

8.1.6 Large-scale protein expression and purification

Unless otherwise stated, the procedures described below were followed for the purification of proteins documented in this thesis. The protein fragments were expressed in preferred cell types (as chosen in high throughput screen). The 1 l culture was inoculated with 15 ml of overnight culture, and cells were then grown at 37 °C (shaken at 180 rpm) until the OD_{600} reached 0.6-0.8. 1 mM IPTG was added to induce the recombinant protein expression at 21 °C overnight. Cells were harvested by centrifugation at 4,000x g for 20 mins at 10 °C. The cell pellet was resuspended in PBS (Dulbecco's, Sigma), and repelleted by centrifugation at 4,000x g for 20 mins. The supernatant was discarded, and the pellet was weighed and stored at -80 °C.

The frozen bacterial cell pellet was thawed at room temperature, and resuspended in Lysis buffer (Working Buffer (WB) supplemented with protease inhibitor cocktail (Roche, 1 tablet per 40 ml buffer), 50 µg/ml Bovine pancreas DNase I (Sigma) and 1 mg/ml hen egg-white lysozyme (Sigma); see **Table 8.8**). Typically, 4 ml of Lysis buffer was used per gram of cell pellet, and the cell suspension was stirred at 4 °C for at least 30 mins to ensure a homogenous resuspension. Cells were lysed using an Emulsiflex C5 homogeniser (Avestin) at 4 °C. The insoluble fraction was separated by centrifugation at 19,500 rpm for 30 mins. The His6-tagged protein was purified using Ni-NTA superflow cartridge (Qiagen) that was pre-equilibrated with 15 CV of WB. The clarified supernatant was supplemented with imidazole before loaded onto the Ni-NTA column (**Table 8.8**). This is to reduce the amount of non-specific binding. The column was then washed with 50CV of WB supplemented to 25-50 mM imidazole (**Table 8.8**). Protein was eluted in WB plus 300 mM imidazole, and the concentration of each eluted fraction was measured by a

nanodrop (Thermo Scientific). For the purification of MBP-tagged proteins, a 5 ml MBPTrap HP column (GE Healthcare) was used. The column was pre-equilibrated in WB before sample loading, and was subsequently washed with the same buffer for 10 CV. Elution was carried out in WB supplemented to 10 mM maltose.

Construct	Working buffer	Loading [Imidazole]	Wash [Imidazole]
HisCnn$^{455/462/490-608/622}$	50 mM NaH_2PO_4/Na_2HPO_4 (pH7.5), 300 mM NaCl, 5 mM BME	(HTP procedure) 10 mM imidazole, 2% Tween-20	20 mM imidazole, 0.05% Tween-20
HisCnn$^{455-622}$	50 mM Tris (pH7.5), 150 mM NaCl, 5 mM BME	20 mM	50 mM, 50 CV
MBPCnn$^{403/462-608}$	50 mM Tris (pH7.5), 150 mM NaCl, 5 mM BME	(MBPTrap purification) Same as WB	Same as WB, 10 CV
MBPCnn$^{462-608}$ 10A	50 mM Tris (pH7.5), 150 mM NaCl, 5 mM BME	(MBPTrap purification) Same as WB	Same as WB, 10 CV
MBPCnn$^{403-608}$ L7A,10D/E, 10D/E+L7A	50 mM Tris (pH7.5), 150 mM NaCl, 5 mM BME	(MBPTrap purification) Same as WB	Same as WB, 10 CV
MBP	50 mM Tris (pH7.5), 150 mM NaCl	(MBPTrap purification) Same as WB	Same as WB, 5 CV
MBPCnn$^{443/490-608}$ 10D/E	50 mM Tris (pH7.5), 150 mM NaCl, 5 mM BME	(MBPTrap purification) Same as WB	Same as WB, 10 CV
MBPCnn$^{490-608}$ 10D/E	50 mM Tris (pH7.5), 150 mM NaCl, 5 mM BME	(Reverse GSTrap/Ni-NTA purification) Same as WB	Same as WB, 10 CV
diLipoyl-Cnn$^{1082-1148}$ WT/DCtoAA/T1133E	50 mM Tris (pH7.5), 150 mM NaCl, 5 mM BME	10 mM	25 mM, 50 CV
Cnn$^{1082-1148}$ WT/DCtoAA/T1133E	Buffer for reverse Ni purification: 50 mM Tris (pH 7.0), 300 mM NaCl, 5 mM BME	(Reverse Ni-NTA purification; 3 passes) 20 mM	WB, 5 CV
Cnn$^{1082-1148}$ WT/DCtoAA/T1133E	WB: 50 mM Tris (pH 7.0), 150 mM NaCl, 5 mM BME (or 1 mM TCEP for crystallisation)		
Lipoyl-Cnn$^{1082-1148_noTEV}$ WT/T1133E	50 mM Tris (pH7.5), 150 mM NaCl, 5 mM BME	10 mM	25 mM, 30 CV
Cnn$^{1082-1148_noTEV}$ WT/T1133E	Buffer for reverse Ni purification: 50 mM Tris (pH7.0), 300 mM NaCl, 5 mM BME	(Reverse Ni-NTA purification; 3 passes) 20 mM	WB, 5 CV
Cnn$^{1082-1148_noTEV}$ WT/T1133E	WB: 50 mM Tris (pH 7.0), 200 mM NaCl, 5 mM BME		
HisSpd2$^{697-805}$	50 mM Tris (pH 7.5), 150 mM NaCl, 5 mM BME	10 mM	25 mM, 50 CV
HisSpd2$^{697/820-921/942/1146}$	50 mM NaH_2PO_4/Na_2HPO_4 (pH7.5), 150 mM NaCl, 5 mM BME	(HTP procedure) 10 mM imidazole, 2% Tween-20	20 mM imidazole, 0.05% Tween-20

Construct	Working buffer	Load [Imidazole]	Wash [Imidazole]
HisSpd2$^{697/820-997}$	50 mM NaH_2PO_4/Na_2HPO_4 (pH7.0), 150 mM NaCl, 5 mM BME	(HTP procedure) 10 mM imidazole, 2% Tween-20	20 mM imidazole, 0.05% Tween-20
MBPSpd2$^{820-942}$	50 mM Tris (pH7.5), 150 mM NaCl, 5 mM BME	(MBPTrap purification) Same as WB	Same as WB, 10 CV
HisSpd2$^{803-1046}$ WT/10D/E	50 mM Tris (pH 7.5), 150 mM NaCl, 5 mM BME	10 mM	25 mM, 50 CV
HisSpd2$^{1047-1146}$	50 mM Tris (pH 7.5), 150 mM NaCl, 5 mM BME	10 mM	25 mM, 50 CV
HisAsl$^{1-68/86/163/300/356}$	50 mM NaH_2PO_4/Na_2HPO_4 (pH7.5), 300 mM NaCl, 5 mM BME	(HTP procedure) 10 mM imidazole, 2% Tween-20	20 mM imidazole, 0.05% Tween-20
HisAsl$^{1-86/356}$	50 mM Tris (pH 7.5), 150 mM NaCl, 5 mM BME	20 mM	50 mM, 50 CV
HisAsl$^{47-180/356}$	50 mM Tris (pH 7.5), 150 mM NaCl, 5 mM BME	20 mM	50 mM, 50 CV
HisAsl$^{648/733-994}$	50 mM NaH_2PO_4/Na_2HPO_4 (pH7.5), 300 mM NaCl, 5 mM BME	(HTP procedure) 10 mM imidazole, 2% Tween-20	20 mM imidazole, 0.05% Tween-20
HisAsl$^{648-994}$	50 mM Tris (pH 7.5), 150 mM NaCl, 5 mM BME	20 mM	50 mM, 5 CV
MBPAsl$^{648/733-994}$	50 mM Tris (pH7.5), 150 mM NaCl, 5 mM BME	(Ni-NTA purification) 20 mM	50 mM, 50 CV
MBPAsl$^{733-994}$ (optimised protocol)	20 mM Tris (pH7.5), 200 mM NaCl, 5 mM BME	(MBPTrap purification) same as WB	Same as WB, 5 CV

Table 8.8 Purification details for the constructs purified in this thesis. Working buffer (WB) is listed for each construct, together with the buffer details for Ni-NTA chromatography. WB plus 300 mM imidazole was always used for the Ni-NTA elution. To elute protein from the MBPTrap column, WB plus 10 mM maltose was used.

The recombinant protein, after fusion tag cleavage (as specified in main chapters), was further purified via HiLoad-Superdex (GE Healthcare) chromatography. Columns of appropriate bead pore size (either Superdex 75 or Superdex 200) were chosen depending on the molecular weight of the target protein. Gel filtration was performed using an AKTA purifier (GE healthcare).

Proteins were concentrated using centrifugal concentrators (Amicon) with appropriate MWCO. Identities were confirmed by electrospray mass spectrometry (David Staunton, Department of Biochemistry, University of Oxford; Lingzhi Gong, Department of Chemistry, University of Oxford).

8.1.7 SDS-PAGE gel electrophoresis

For the analysis of protein overexpression in cell culture, 400/OD_{600} µl aliquots were taken for pre- and post-induction, pelleted down and resuspended in 100 µl 1x SDS loading dye (**Table 8.9**). To analyse the purity of SEC eluted fractions, 10 µl of the sample was mixed with 2x SDS loading dye and heated to 95 °C for 5 min. In cases where the UV readout was low indicating a relatively diluted sample, 10 µl Stratagene beads (Agilent) were mixed with 100 µl sample, and the mixture was incubated at room temperature for 10 mins before pelleting down by centrifugation. Pellet containing the concentrated protein was resuspended in 10 µl 1x loading dye. Samples were loaded on to a 15% SDS-PAGE gel (**Table 8.9**), which was then run in 1x SDS-PAGE running buffer (**Table 8.9**) at 200 V until completion. The gels were stained by Coomassie blue staining (**Table 8.9**).

Solution	Composition	Concentration
1x loading dye	Tris-HCl, pH6.8	50 mM
	Glycerol	10%
	SDS	2%
	Bromophenol Blue	0.1%
	DTT	50 mM
15% SDS-PAGE gel	*Running gel:*	
	Acrylamide/methylene bisacrylamide solution (37.5:1 ratio)	15% v/v
	Tris-HCl, pH8.8	400 mM
	SDS	0.1%
	Ammonium Persulphate (APS)	0.1%
	Tetramethylethylenediamine (TEMED)	0.01%
	Stacking gel:	
	Acrylamide/methylene bisacrylamide solution (37.5:1 ratio)	5.1% v/v
	Tris-HCl, pH6.8	130 mM
	SDS	0.1%
	APS	0.1%
	TEMED	0.01%
1x SDS-PAGE running buffer	Tris	25 mM
	Glycine	190 mM
	SDS	1%

Solution	Composition	Concentration
Comassie Blue staining solution	Coomassie Brilliant Blue R-250	0.12% w/v
	Methanol	40 %
	Glacial Acetic Acid	10%
Destaining solution	Methanol	30%
	Glacial Acetic Acid	10%

Table 8.9 Solutions used in SDS-PAGE.

8.1.8 Multi Angle Light Scattering (MALS)

Samples were prepared at desired concentrations as detailed throughout the **Result** sections. Size exclusion chromatography was performed to separate discrete oligomeric species prior to mass determination. The scattering of light from each sample was detected by a Dawn Heleos-II light scattering detector and an Optilab-rex refractive index monitor. The analysis was carried out using ASTRA 5.3.4.14 software assuming a dn/dc value of 0.186 ml/g. (Acknowledgements to Dr Steven Johnson for his help with MALS experiments).

8.1.9 Differential Scanning Calorimetry (DSC)

Samples were exchanged into different buffers using an S200 10/30 analytical column (GE Healthcare), and were concentrated to 10 μM for DSC. The experiments were performed using the standard protocol on CSC 6300 NanoDSC III machine. The temperature was increased from 20 to 110 °C at 1 °C/min, and any difference in the heat capacity was recorded. The heating scan of buffer was first performed to produce baseline signals, followed by the scan of sample solution. Thermodynamic analysis was conducted using CpCalc program, which allows one to calculate the excess heat capacity of a sample. The melting temperature was then determined for protein under each buffer condition. (Acknowledgements to David Staunton for his help on DSC scans.)

8.1.10 Crystallisation trials

Crystallisation trials were carried out using the vapor-diffusion sitting drop method. The mother liquor and protein stock were mixed at 1:1 or 1:3 ratios in 200 nl drops. All trials were set up in the MRC Crystallisation Plates that were sealed with a StarSeal Advanced Polyolefin Film (Type D; Star Lab) for incubation at 21 °C. Broad screens (Molecular Dimensions) listed in **Table 8.10** were first set up to determine the initial crystallisation conditions. Unless stated, all screens were trialled for the protein.

Broad Screen	Properties
Stura	Screen the protein precipitant solubility curve by varying the pH values (from 4.5 to 8.5) and the concentrations of PEGs/salts.
Structure I&II	Sparse-matrix screen sampling different types of salts, polymers, organics and pH.
JCSG+	Optimised sparse-matrix screen designed based on previously successful crystallisation conditions; test a wide range of pH, salts and precipitates.
Morpheus	Incorporate 49 low molecular weight components that could promote initial crystal formation and lattice stability.
PACT	Screen the effect of pH, anions and cations, using PEG as the precipitate.
Proplex	Sparse matrix screen containing crystallisation conditions optimised for protein complexes.
ClearStrategy I&II	Screen a wide range of pHs, salts, cations and PEGs.
MIDAS	Screen alternative polymeric precipitants while focusing on a narrow range of pH and salt concentrations.

Table 8.10 96-well high throughput screens used in the crystallisation trials.

8.1.11 Protein labelling and purification for use with MicroScale Thermophoresis (MST)

Monolith™ protein labelling kit Red-NHS (Amine Reactive) was used to label the protein. The dye NT-647 carries a reactive NHS-ester group that modifies primary amines present in amino acids such as Lys and the N-terminal amino group. Prior to labelling, the protein was buffer exchanged into WB containing 1 mM TCEP (Labelling buffer) to remove BME that can otherwise be labelled with the dye.

For protein labelling, the protein concentration was adjusted to 20 µM using Labeling buffer in 100 µl. 30 µl of 100% DMSO (Sigma) was added to the solid fluorescent dye and mixed thoroughly by

vortexing. The dissolved dye was then adjusted to 60 µM (the stock was ~435 µM) using Labeling buffer. Protein and dye were mixed at a 1:1 ratio in a final volume of 200 µl, and then incubated at room temperature for 30 mins in the dark.

For purification, the buffer exchange column (GE Healthcare) was first equilibrated and washed with 3x 3 ml of the analysis buffer. The labelling reaction was added to the center of the column and let sample enter the bed completely. Note the volume of labelling reaction was adjusted to 500 µl after the sample has entered the column bed, using the analysis buffer. The flow through was discarded. Labelled sample was eluted using 600 µl of the analysis buffer, and collected in 100 µl aliquots.

8.1.12 mRNA preparation

mRNA was produced via *in vitro* transcription using mMESSAGE mMACHINE® kit. For preparation, the plasmid DNA, which contains a T3 RNA polymerase promoter site, was linearised using AscI restriction enzyme (10 U per 1 µg DNA; NEB). The digestion was incubated at 37 °C for 10 hrs followed by heat inactivation at 65 °C for 15 mins. To terminate the reaction, 1/20th volume of 0.5M EDTA, 1/10th volume of 3M Sodium acetate and 2 volumes of Ethanol were added and well mixed before chilling at -20 °C for 1 hr. The DNA was pelleted by centrifugation at top speed for 20 mins (at 4 °C), the supernatant was carefully pipetted away, and the pellet was washed with 1 ml of 70% ethanol by centrifugation for another 5 mins. The supernatant was discarded, and the pellet was dried at 55 °C with the lid open for 10 to 15 mins. Pelleted DNA was resuspended in 10 µl of RNase free water (Ambion™).

An *in vitro* transcription reaction was assembled as shown in **Table 8.11**, and was incubated for 2 hrs at 37 °C. The resulting mRNA was purified using RNeasy Minielute cleanup kit (Qiagen), according to manufacturer's protocol. In order to obtain sufficiently high concentrations, eluted RNA was re-applied to the spin column for re-elution. The concentration was measured via the A_{230} using a nanodrop (Thermo Scientific) in "ssRNA" mode. All samples were adjusted with RNase free water to a concentration of 2 mg/ml, and stored at -80 °C in 4 µl aliquots.

Reaction component	Volume (µl)
Nuclease free water	To a final of 20 µl
2x NTP/CAP	10
10x reaction buffer	2
Linearised DNA	1 µg
Enzyme mix	2

Table 8.11 *In vitro* transcription reaction set up.

8.1.13 Image presentation and analysis

All the images are displayed as 'smooth zoom' impressions, and the image contrast was adjusted to use the full range of pixel intensities. The levels and contrast in control and experimental images were adjusted in the same way; however, in some cases, the experimental images were adjusted independently to allow better visualisation, and this has been stated in the main text (**Chapter 3**).

For phosphomutants within the Cnn PReM domain (**Chapter 2**) and the Spd-2 ASH domain mutants (**Chapter 4**), radial profiling (i.e. the fluorescence intensity across the centrosome) was conducted in order to effectively compare the embryo injection results. Images were first scaled to maximise the resolution for radial profiling. The centre of mass of the centrosome was calculated using the 'analyse particles' (centre of mass) macro on the most central Z plane of the centrosome in Fiji. A concentric ring was then drawn using the calculated centre, around which the average fluorescence was measured. The background was subtracted using the average cytosolic signal, and the profile was normalised so that peak intensity was equal to 1. The profile was mirror imaged to produce a symmetric centrosomal profile. Measurements were carried out on 10 centrosomes, and the averaged distribution was plotted.

8.2 X-ray crystallography

8.2.1 Data collection

Once protein crystal is grown, the X-ray data is the only structural experimental data collected for that protein. Reflections of the incident X-ray beam, which is diffracted from lattice planes in the crystal according to Bragg's Law, are collected. Crystals are first fished and frozen in liquid nitrogen to protect them from radiation damage. Cryoprotectant is included to avoid the formation of ice crystals that would otherwise produce a very strong signal in the X-ray diffraction image. In this thesis, data collection was performed at a synchrotron radiation source using the rotation method, under a cryo-condition in the range of 100-120 K. The crystal was mounted onto a goniometer head and centred in the X-ray beam path. Diffraction images were collected by rotating the crystal about a single axis with a small oscillation angle that typically ranges from 0.1 to 1°. When the reciprocal lattice plane intersects with a sphere, known as the Eward Sphere, certain points on that plane would be in the diffracting condition and can then be detected[12]. As the crystal is rotated, different reciprocal lattice planes will be in the correct orientation for diffraction and thus different spots are observed. In addition, only a limited amount of the plane can be detected within the narrow oscillation range. Therefore, a continuous series of such images are collected for a complete dataset.

Indexing software automatically finds all spots on the diffraction images, and assigns Miller indices (h, k and l) to each spot, from which information about the lattice group can be deduced. Based on this, it is possible to calculate a strategy to collect a complete dataset that contains all the unique reflections within a given resolution range, while reducing the amount of time exposed to X-ray beam and hence the radiation damage to a sample. The minimum number of degrees needed to collect a complete dataset is defined by 180°/N, where N is the highest symmetry fold. Note even for a crystal in P1 space group, 360° rotation is not required because each reflection has another one from the opposite side of the same lattice plane, known as Friedel's pair, and the two are same in intensity but opposite in phase (Freidel's law; will be discussed in more details later). This is, however, not true for anomalous data collection. Since the data collected will be used in the calculation of electron density map, it is important to collect as complete a dataset as

[1] http://www-structmed.cimr.cam.ac.uk/Course/Adv_diff2/Diffraction2.html
[2] http://xray0.princeton.edu/~phil/Facility/Guides/XrayDataCollection.html

possible, so to include all the unique reflections. Although a complete dataset can be obtained via rotation through a minimum range, more than the minimal value is always taken in order to measure most reflections several times (known as multiplicity).

$$\text{Completeness} = \frac{\text{Number of unique reflections observed}}{\text{Total number of unique reflections}}$$

$$\text{Multiplicity} = \frac{\text{Total number of reflections observed}}{\text{Number of unique reflections}}$$

8.2.2 Data processing

Once the data is collected, each reflection spot on the diffraction images is found. Some crystal symmetry operations produce specific “footprints” in the diffraction pattern and therefore the Laue group can be predicted based on the observed reflections. The reflections are then mapped onto their reciprocal lattices, based on which the unit cell dimensions are determined. Auto-indexing software predicts what the image will look like from the cell parameters and crystal orientation, which is superimposed onto the real image and the parameters are optimised until the two images fit [Smyth and Martin, 2000]. The intensity of each reflection is then measured, which is a summation of pixel counts in that spot region with subtraction of the background. After data integration, the dataset contains a list of unique reflections (h, k, l) with their measured intensities. Often each unique reflection is observed multiple times due to the symmetry operation, and the intensities of such symmetry related reflections are expected to be identical. However, this is always not the case due to variations including beamline fluctuations at synchrotrons, orientation of the crystal in the beam path or radiation damage in the crystal [Evans, 2006a]. Therefore, these factors need to be taken into account during data processing. In addition, some reflections are recorded over several images and must be combined. However, in order to combine such partially distributed reflections, a scaling factor needs to be applied. Finally, the intensities for symmetry related and multiply measured reflections are averaged to calculate an I_{hkl} value for each unique reflection in the final dataset.

After data processing, the quality of diffraction data is assessed to determine the consistency of multiply measured reflections and the high-resolution limit. Data quality is quantified using several statistics indicators. These indicators are calculated against resolution range, or between diffraction images, or crystal rotation. Due to the large number of unique reflections in a dataset, reflections within similar resolution ranges are grouped into one shell/bin and the statistics is reported for an average of all the reflections within that shell/bin. In this thesis, I have used the following statistics indicators to assess the data quality.

R_{merge} measures how different the independent measurements of the intensity of a unique reflection is from the average intensity for that reflection. This therefore reflects the internal consistency of individual observations. However, this indicator does not take into account of multiplicity. R_{merge} value increases with the data multiplicity, even though the merged data are improved by averaging multiple observations [Evans and Murshudov, 2013a]. Alternatively, the redundancy-corrected R_{pim} value is calculated and more accurately describes the precision within the averaged measurements.

$$R_{merg} = \frac{\sum_{hkl} \sum_{i=1}^{N} |I_{hkl}(i) - \overline{I_{hkl}}|}{\sum_{hkl} \sum_{i=1}^{N} I_{hkl}(i)}$$

$$R_{pim} = \frac{\sum_{hkl} \sqrt{\frac{1}{N-1}} \sum_{i=1}^{N} |I_{hkl}(i) - \overline{I_{hkl}}|}{\sum_{hkl} \sum_{i=1}^{N} I_{hkl}(i)}$$

In addition, the data quality is also measured with a correlation coefficient, known as $CC_{1/2}$. The unmerged data is randomly split into two subsets, each of which contains half of all the measurements for a unique reflection [Karplus and Diederichs, 2012]. The average intensities in each subset are correlated, with the $CC_{1/2}$ value expressed as a fraction of perfect correlation [Karplus and Diederichs, 2012]. The signal-to-noise level is calculated, and a general rule of thumb is that the data is likely unreliable when the I/σI value falls below 2 [Evans, 2006a]. However, other statistics quantifications should also be taken into account for decision of the high-resolution cut-off. For high-resolution reflections, the signal is often much weaker and therefore results in a lower I/σI value. This also leads to a lower consistency within the measurements for the outer shell data. In addition, the data completeness in each resolution shell is also important, which could otherwise affect map quality in structure determination. The

statistics indicators mentioned above could also be used to determine any outliers of images that should not be included in the dataset because the intensities of reflections within these images are more variable than the average in other images.

At the end stage of data processing, the amplitude of each reflection ($|\mathrm{F}|$) is calculated. If the true intensity were known, a direct square root can be taken for the value of $|\mathrm{F}|$. However, the experimentally observed intensities do have errors that are then corrected for by the Lorentz and Polarisation (LP) factors in the equation below [Buerger, 1940; Drenth, 2010]. Both parameters are dependent on the experimental set up.

$$|\mathrm{F_{hkl}}| = \sqrt[2]{\frac{\mathrm{I}}{\mathrm{LP}}}$$

8.2.3 The Phase problem

In the final, processed dataset, each unique reflection is given a Miller index h,k,l, which corresponds to its position in the reciprocal lattice, and its associated amplitude. The structure factor, which is a Fourier transform of the electron density distribution in the crystal unit cell, is calculated in an equation:

$$\mathrm{F_{hkl}} = \mathrm{V} \iiint_{0}^{1} \rho(x\ y\ z)\ \exp[2\pi i(hx + ky + lz)]\ dx\ dy\ dz$$

The ultimate goal is to determine the electron density at every position x, y, z in the unit cell [Drenth, 2010]. The structure factor is defined by two components: the amplitude and the phase of each reflection.

$$\mathrm{F_{hkl}} = |\mathrm{F_{hkl}}|\exp\ (i\phi_{hkl})$$

Note the amplitude of the structure factor is measurable, however the phase angles cannot be determined directly from the diffraction data. This therefore raises the phase problem, which needs to be solved in order to produce an electron density map. An atomic structure can then be built into that density. In the next two sections, I will summarise the two methods I have used in this thesis to solve this problem.

8.2.4 Phase determination by Molecular Replacement (MR)

In molecular replacement method, an initial set of protein phase angles is provided from a search model. Search model is normally based on the structure of a protein that is similar in primary sequence and is therefore structurally homologous. In some cases, however, proteins do not share a high sequence similarity, but they do share an overall similar structural fold. This is seen in the example of Spd-2 ASH3 structure determination.

The search model is first placed in the target unit cell, a process that requires precise orientation and position. Each molecule is specified with 3 rotational angles and 3 translation angles, which overall creates a 6-dimensional problem. However, this can be separated into two 3D problems when a rotation function (RF) is first computed, the orientated molecule is then superimposed onto the target molecule following the 3D translation function (TF). These functions can be better understood by considering the Patterson map, which will be described below (**Figure 8.1**).

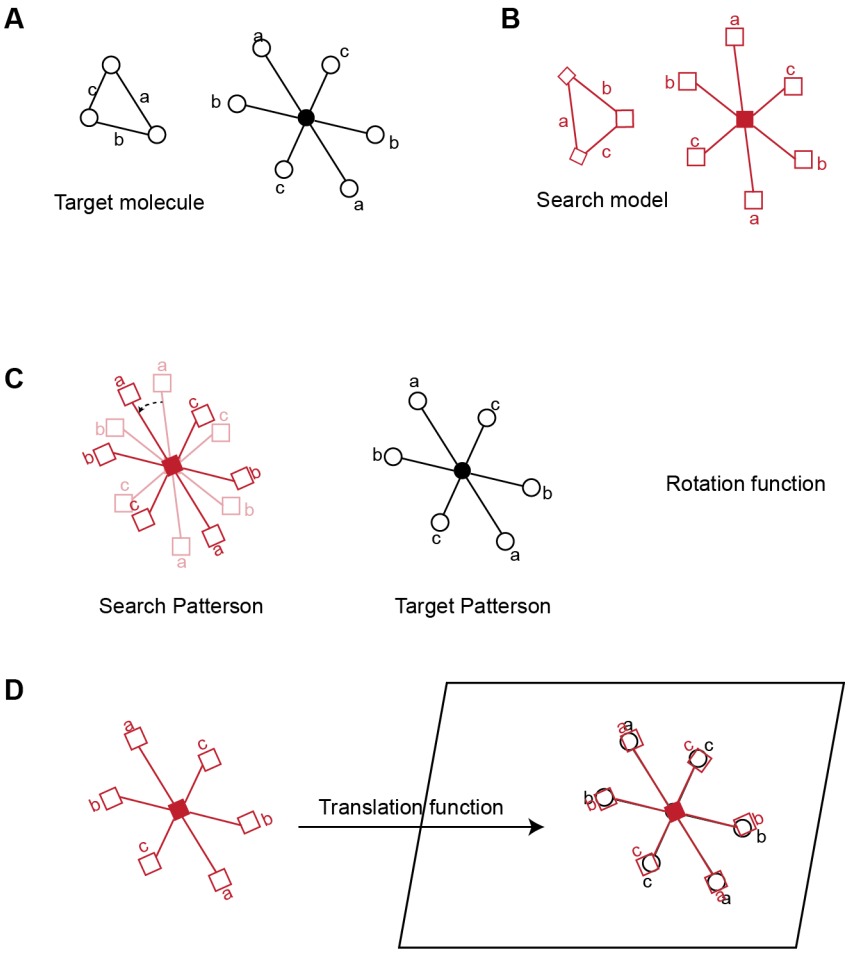


Figure 8.1 The Patterson function and molecular replacement. (**A**) Left, a target molecule with no crystal structure (target). Right, Patterson calculated from the measured intensities of reflections. (**B**) Left, a search model of the target molecule. Right, Patterson calculated from the model. (**C**) The search Patterson map is rotated such that its orientation aligns with the target map. (**D**) The translation function then positions the correctly orientated search Patterson in the unit cell matching the target Patterson. Phase information can now be calculated from this solution. Diagram modified from [Cottee, 2014] with permission.

An inverse Fourier transformation of the structure factors carrying amplitudes and phases produces an electron density map. Fourier transformation of the measured intensities produces a Patterson map that describes the interatomic vectors[1] (**Figure 8.1A,B**). Each vector starts at the origin of the cell, and points in both directions due to the centrosymmetry (**Figure 8.1A,B**). The magnitude of the peak is proportional to the product of the atomic masses of the atoms involved, and the distance between the origin and the peak is equal to the interatomic vector. In the presence of a search model, a predicted Patterson map can be calculated, which is then rotated and translated to align with the experimental Patterson map (**Figure 8.1C,D**).

In this thesis, I have used MOLREP program in CCP4 wizard [Vagin and Teplyakov, 2010; Winn et al., 2011]. The search model undergoes RF and TF modifications as well as a packing function (PF). PF describes any steric clashes between different copies of the molecule in the unit cell, and a score is given that estimates the likelihood of the molecular packing on a scale of 0 to 1 (no clash)[3]. The solutions are scored based on the correlation coefficient, between the observed intensities and the calculated values from the solution, times PF. A contrast value is also calculated that describes the difference of the highest score (associated with the top solution) from the average, and hence how unique the top solution is. In most cases, a contrast value of higher than 2.5 means the solution is definite [Vagin and Teplyakov, 2010]. Nonetheless, further refinement needs to be carried out in order to confirm the solution, as a correct solution should improve during refinement.

Self-rotation function can provide information about the number of monomers in an asymmetric unit, which can help guide the subsequent MR job. The Patterson map is rotated upon itself as explained in **Figure 8.2**. Peaks, higher than background noises, are given when the maps overlap in the presence of crystallographic and/or non-crystallographic symmetry. However, the non-crystallographic symmetry could align with the crystallographic symmetry and hence might not be detected. Matthew's coefficient should also be calculated to help determine the number of NCS copies in an asymmetric unit.

[3] http://ccp4wiki.org/~ccp4wiki/wiki/index.php?title=Molecular_replacement_with_MolRep

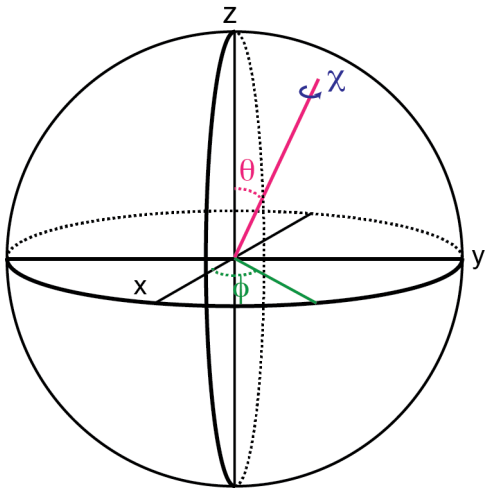


Figure 8.2 Construction of self-rotation function in spherical angles. The Patterson is first rotated from the *z* axis by θ degrees followed by ϕ degrees from the *x* axis, and the last rotation takes place around the final axis by χ degrees. Diagram modified from[4].

8.2.4 Anomalous diffraction in Phase solution

For normal protein atoms that contain free electrons with absorption wavelength far from the X-ray wavelength, the diffracted beam has a 180° phase difference with respect to the incident beam. Due to the centrosymmetry, each reflection, (h, k, l), in a dataset is related to another reflection, (-h, -k, -l), that scatters from the opposite face of the same lattice plane [Drenth, 2010]. These related reflections are equal in amplitude, but opposite in phases/direction, according to Friedel's law. The related reflections are known as a Friedel pair.

In the presence of protein bound heavy atoms (anomalously scattering atom) such as Zn^{2+} in the Cnn-CM2 protein crystals, the incident beam is absorbed, because the X-ray wavelength approaches the absorption wavelength of Zn^{2+}, and is later emitted at another wavelength. This is because when a bound electron interacts with the X-ray, it is excited to a different shell and then falls back to the original shell but with a small difference in energy. In the presence of anomalous scattering, the diffracted beam from the heavy atoms does not differ by 180° in phase from the incident beam (**Figure 8.3A**). Alternatively, the scattering is contributed by:

$$f_{anom} = f + \Delta f + if'' = f' + if''$$

where Δ*f* is the change in scattering factor along the horizontal axis (a real component) in Argand diagram and *if"* is the change along the vertical axis (an imaginary component) [Drenth, 2010].

[4] http://www.stampcollectingproject.org/crystallography/symmetry-and-sg/self-rotation-function

Note *i* indicates a counterclockwise rotation of 90°. As a consequence of anomalous scattering, the atomic scattering vector is rotated counterclockwise, and the structure factors F(h, k, l) and F(-h, -k, -l) are no longer equivalent in amplitudes and have different phase angles [Drenth, 2010], (**Figure 8.3B**). This difference, however, is useful to determine the position of heavy atom substructure. Anomalous difference is defined by:

$$\Delta|\mathrm{F}|_{\mathrm{anom}} = \{|\mathrm{F}_{\mathrm{PH}}(+)| - |\mathrm{F}_{\mathrm{PH}}(-)|\}\frac{f'}{2\,f''}$$

where $|F_{PH}(+)|$ represents the amplitude of the structure factor for reflection (h, k, l) and $|F_{PH}(-)|$ for reflection (-h, -k, -l), scaled by f'/2f" [Drenth, 2010]. The squared values of $\Delta|F|_{anom}$ are then used to calculate an anomalous Patterson map.

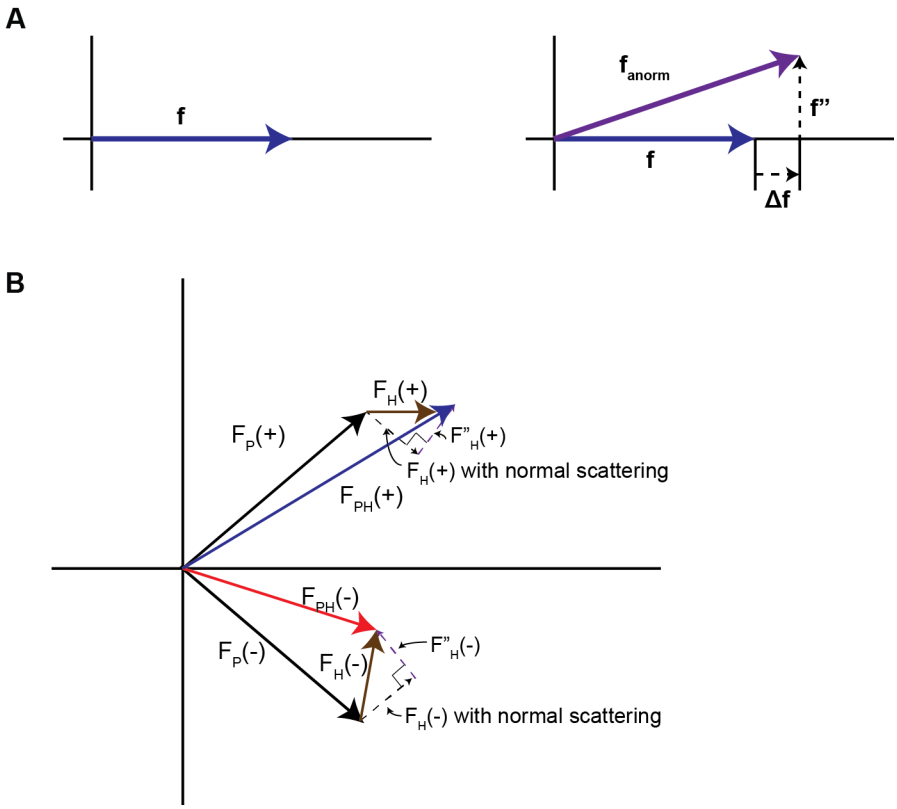


Figure 8.3 Phase shift in anomalous scattering breaks down Friedel's law. (**A**) The atomic scattering factors for a normal scatter (Left) and an anomalous scatter (Right). The anomalous contribution contains two components: a real part Δf and an imaginary part f". The f" component leads to a 90° phase shift. Under normal scattering conditions, f has a 180° phase difference to the incoming X-ray beam. Diagram modified from [Drenth, 2010]. (**B**) Argand diagram showing the structure factors for a Friedel pair from a heavy atom derivative. Each reflection, F_{PH}, can be decomposed into 2 scattering components, F_P from all protein atoms and F_H from the bound heavy atoms. In the absence of anomalous scattering, F_P (+) and F_P (-) are symmetrical, and so are F_H (+) and F_H (-); therefore, F_{PH} (+) and F_{PH} (-) are equal in amplitude but opposite in phase. Under anomalous scattering conditions, the anomalous component, f", contributes a 90° phase shift and the Friedel mates are no longer equivalent in terms of phase or amplitude.

Using the Patterson method, locations of heavy atoms can be derived. Once the heavy atom positions are known, their structure factors can be calculated and are then used to determine the

phase angles for other reflections. This can be illustrated using the Argand diagram to draw a Harker construction, as shown in **Figure 8.4**. In a Singlewavelength Anomalous Dispersion (SAD) experiment, 2 phase solutions are possible for each reflection. This ambiguity can be removed by collecting the data at an alternative wavelength, which will result in different values of f' and f". In this thesis, I did not perform a Multiplewavelength Anomalous Dispersion (MAD) experiment. Nonetheless, the incorrect phase solution will not improve during the subsequent phase improvement and is therefore discarded. Once the phases are determined, the structure factors are calculated and Fourier transformed to produce an electron density map into which the structure can be built.

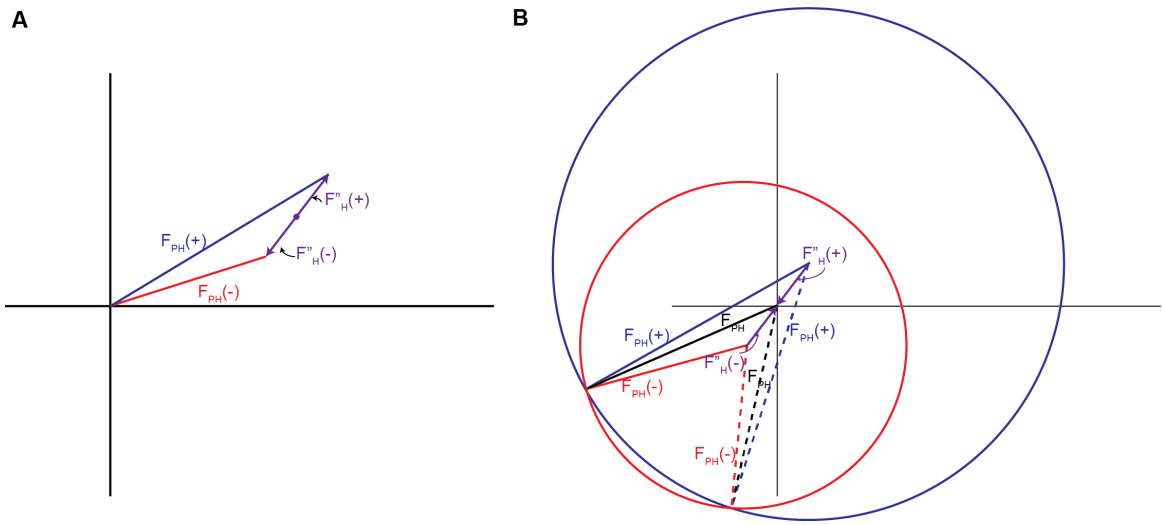


Figure 8.4 Harker construction to show phase solutions in an SAD experiment. (**A**) Once the localisation of heavy atom substructure is defined, the anomalous component, F", is phased for both Friedel's mates. The amplitudes of F_{PH} (+), F_{PH} (-) and F" are all measurable. (**B**) Circles are plotted with the origins at the starting points of vectors $F"_H$ (+) and $F"_H$ (-), taking the amplitudes of F_{PH} (+) and F_{PH} (-) respectively. This gives rises to 2 intersection points, both of which can explain phase solution for the non-anomalous structure factor (F_{PH}). The subsequent phase improvement narrows down the answer to a single phase solution, since only the correct set of phase angles can be improved.

8.2.5 Structure model refinement

When the preliminary structure model is available after initial phasing and building or positioning of the search model, structure factors calculated from such model are in poor agreement with the observed structure factors. Iterative rounds of model refinement are then followed. The agreement between calculated and experimental amplitudes is represented by an R-factor as defined by:

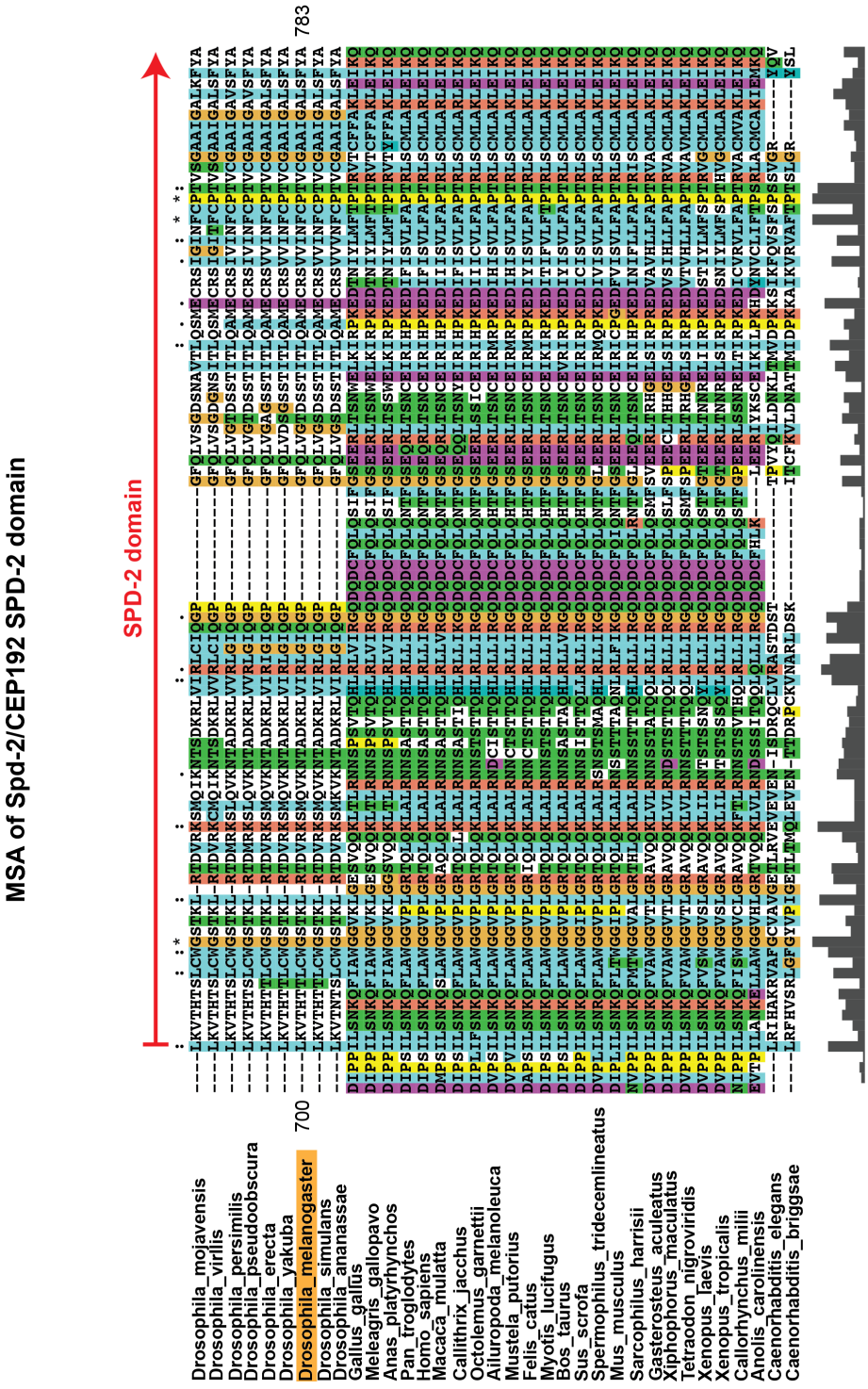
$$R = \frac{\sum_{hkl} \left| |F_{obs}| - |F_{calc}| \right|}{\sum_{hkl} |F_{obs}|}$$

In each refinement, a new electron density map is calculated using phases from the refined model and the originally measured amplitudes for each reflection. The phase improvement would lead to a more accurate map, into which the structure can be more accurately modelled. Several refinement methods are normally applied in order to improve the structure model such that the observed diffractions can be adequately explained. The adjustment of the model includes changing the positions and the temperature factors (B factors) for all atoms in the structure. In some cases, especially for high-resolution structure, occupancies are split as to accommodate side chains with alternative rotamer conformations. A *priori* knowledge about the stereochemistry of the protein, such as the bond angle and length, is also used. In the presence of non-crystallographic symmetry, NCS restrains are set to average phases in different molecules since they should be similar.

After refinement additional rebuilding is normally needed, and this was carried out manually in Coot [Emsley et al., 2010] in this thesis. Two types of maps from refinement are useful: $2F_{obs}$-F_{calc} map and F_{obs}-F_{cal} difference map. F_{obs} is a structure factor calculated using the model-derived phases and the experimentally measured amplitudes; whereas F_{cal} is calculated using both the model-derived phases and amplitudes. In the difference map, peaks above 4σ intensity show positive residual density for unexplained features and negative density for mis-interpretated features in the current model. Both maps are used to guide manual modification of the structure in Coot.

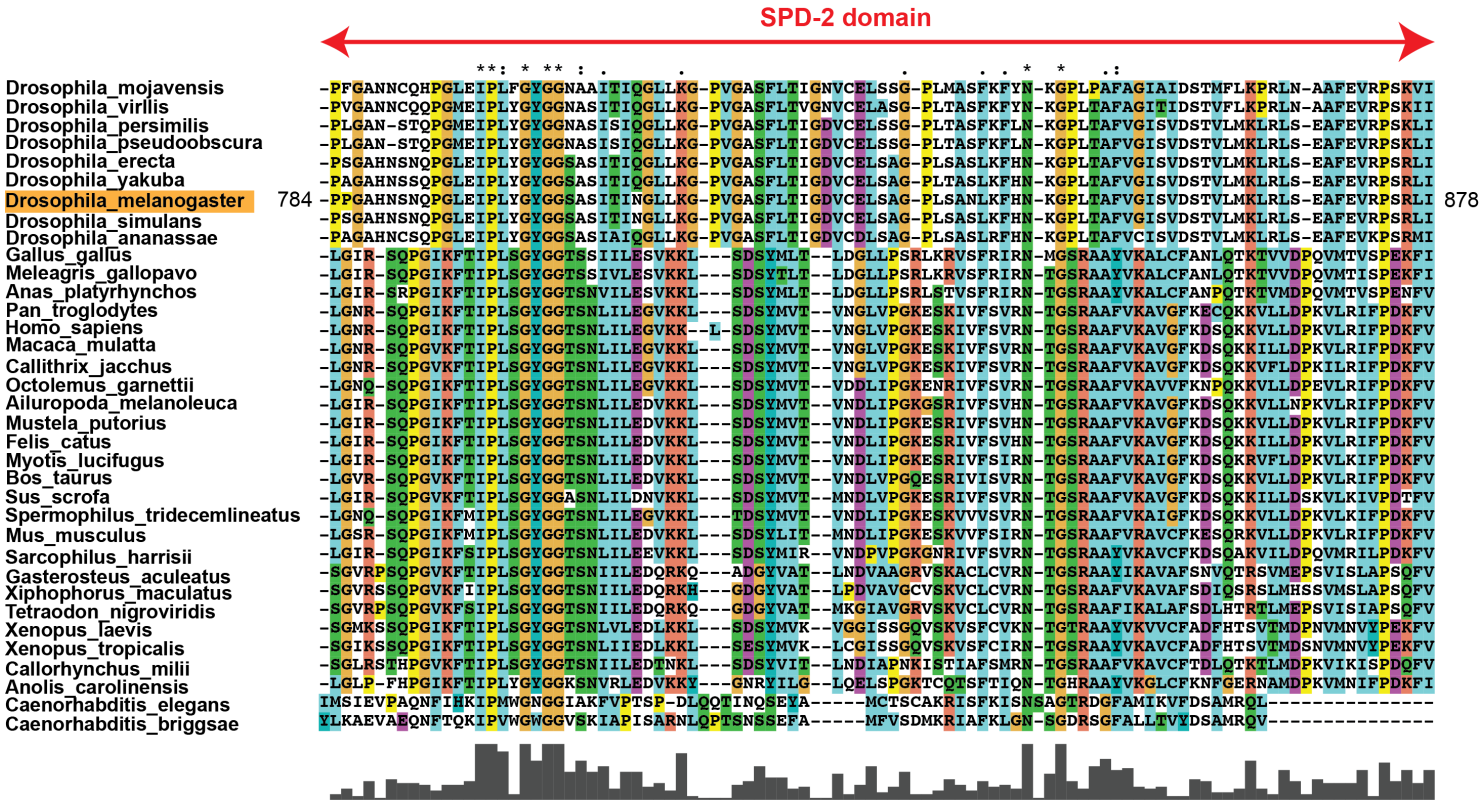
However, there could be a model bias because the new map is calculated using phases from the model, and the R value can give false impression of a good model. For this reason, 5% of all reflections in a dataset are excluded from the refinement and used to calculate the R_{free} (test set) giving an independent measure of the model accuracy. The remaining reflections are included in the refinement process, and are referred to as the working set (R_{work}). Both values are therefore considered in all the refinements in this thesis.

8.3 MSA of Spd-2 homologs reveal conservation within the SPD-2 domain

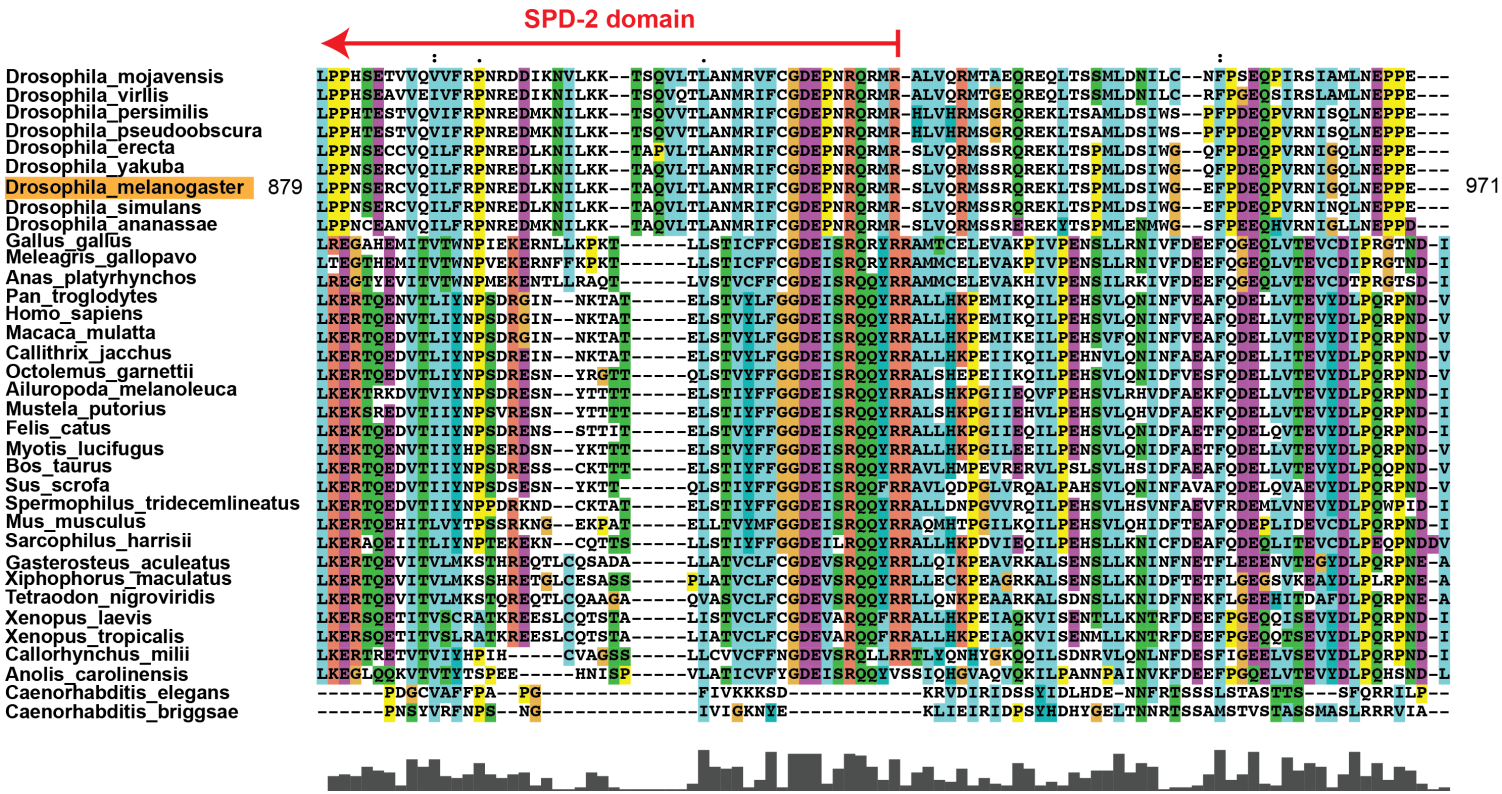


Identical, conserved and semi-conserved residues are indicated by asterisks, colons and full-stops respectively.
Sequence numbers for *D. melanogaster* are shown.
Quality curve is displayed below the alignment.

MSA of Spd-2/CEP192 SPD-2 domain



MSA of Spd-2/CEP192 SPD-2 domain



8.4 Purification and characterisation of HisCnn$^{93-185}$

Cnn$^{93-185}$ includes the conserved CM1 domain (Cnn$^{97-172}$), and was PCR amplified from a cDNA encoding the full length *D. melanogaster* Cnn. The PCR fragment was then cloned into a pETM-14 vector, which encodes for an N-terminal His6 tag, according to the standard procedure (**Appendix 8.1.1**). A high-throughput expression and purification screen was carried out (protocol see **Appendix 8.1.6**; **Table 8.12**), in order to find the optimal growth condition for this construct. The subsequent gel analysis showed that a greater yield of soluble protein was obtained using Tuner (DE3) cells in TB medium.

Construct	MW (kDa)	pI	Abs 0.1%	No. of Cys
HisCnn$^{93-185}$	13.00	6.71	0.115	0

Table 8.12 Details of the His6-tagged Cnn N-terminal construct, calculated using Protparam [Wilkins et al., 1999].

A 1 l large-scale production was performed, using the optimised growth conditions, following the standard procedure (**Appendix 8.1.7**). The working buffer contains 50 mM Tris (pH 8.0) and 150 mM NaCl. Cells were lysed, and the cell lysate was clarified by centrifugation (**Appendix 8.1.7**). The supernatant was supplemented to 20 mM imidazole, and was then loaded onto a 1 ml Ni-NTA column (Qiagen, **Figure 8.5A**). In order to remove any non-specific binders, the column was washed with WB plus 50 mM imidazole for 50 CV. Protein was eluted from the Ni-NTA cartridge using WB plus 300 mM imidazole. HisCnn$^{93-185}$ was further purified by size exclusion chromatography, using WB as the running buffer (**Figure 8.5B**). Full-length protein eluted around the void volume from an S200 16/60 column (GE Healthcare), suggesting the formation of large molecular weight complexes. A second elution peak was obtained, containing protein of the expected length and a minor amount of higher molecular weight contaminant.

To investigate the oligomerisation state of the purified HisCnn$^{93-185}$, I took samples from both SEC peaks for SEC-MALS experiment. The measurement was carried out at 90 µM, using a Superose 6 10/30 column (GE healthcare), using WB for the running buffer. Fraction taken from the major SEC peak (close to the void volume) showed the formation of multimeric aggregates, whereas the protein eluted within the second SEC peak stayed dimeric in solution (**Figure 8.6**). However, the observed molecular weight for HisCnn$^{93-185}$ showed large variations across each elution peak.

This indicated poor sample homogeneity, and the construct was therefore not pursued for further study.

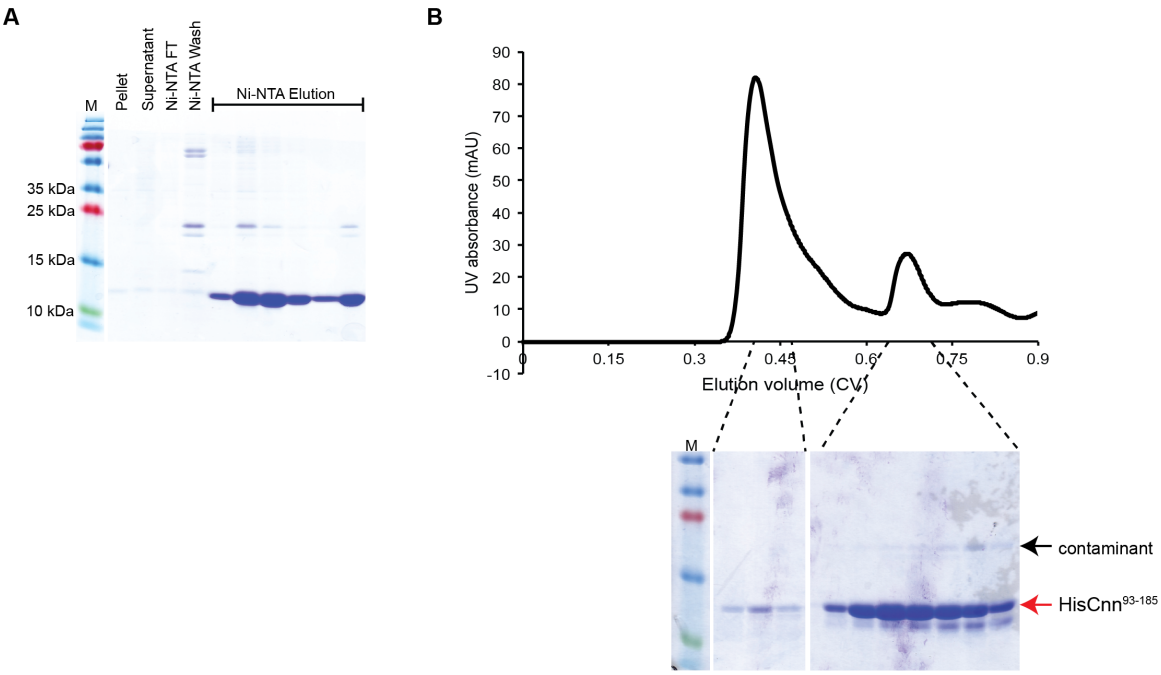


Figure 8.5 Purification of HisCnn$^{93-185}$. (**A**) Ni-NTA purification for HisCnn$^{93-185}$. Elution fractions contained largely the full-length protein at expected molecular weight, but minor amount of contaminants were also present in the sample. (**B**) S200 16/60 SEC purification and the subsequent SDS-PAGE analysis for HisCnn$^{93-185}$. Both elution peaks contained the fusion protein, and the majority eluted close to the void. The second peak contained contaminants. Fractions within each peak were pooled for SEC-MALS analysis.

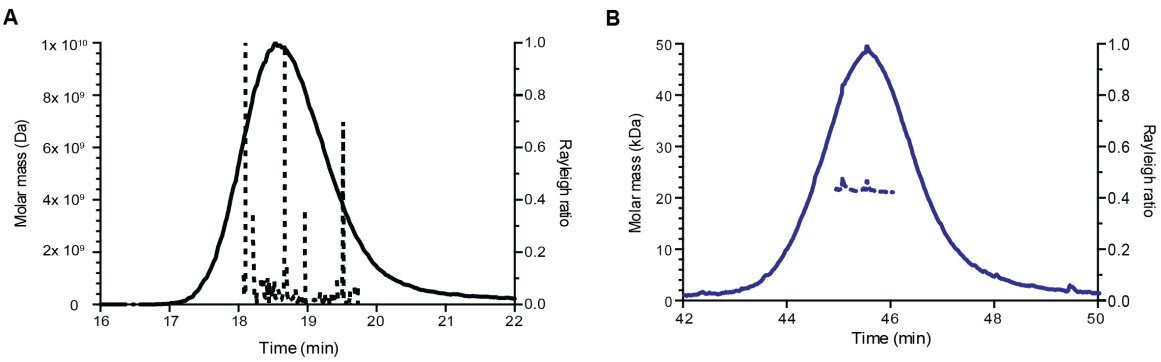


Figure 8.6 Oligomerisation state of the purified HisCnn$^{93-185}$ protein investigated by SEC-MALS, at 90 µM. The solid line shows the relative Rayleigh ratio, and the dashed line shows the observed mass across each elution peak. (**A**) Fraction taken from the major SEC peak on S200 16/60 elution (**Figure 8.6B**) was shown to aggregate into multimeric species. (**B**) Fraction taken from the second SEC peak on S200 16/60 elution was shown to form dimers in solution.

Bibliography

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr,* 66**,** 213-221.

Alva, V., Syamala Devi, D. P. & Sowdhamini, R. 2008. COILCHECK: an interactive server for the analysis of interface regions in coiled coils. *Protein Pept Lett,* 15**,** 33-38.

Andersen, J. S., Wilkinson, C. J., Mayor, T., Mortensen, P., Nigg, E. A. & Mann, M. 2003. Proteomic characterization of the human centrosome by protein correlation profiling. *Nature,* 426**,** 570-574.

Archambault, V. & Glover, D. M. 2009. Polo-like kinases: conservation and divergence in their functions and regulation. *Nature Reviews,* 10**,** 265-275.

Azimzadeh, J. & Marshall, W. F. 2010. Building the centriole. *Curr Biol,* 20**,** R816-825.

Baker, A. M. E., Roberts, T. M. & Stewart, M. 2002. 2.6 Å Resolution Crystal Structure of Helices of the Motile Major Sperm Protein (MSP) of Caenorhabditis elegans. *J Mol Biol,* 319**,** 491-499.

Balestra, F. R., Strnad, P., Flucklger, I. & Gonczy, P. 2013. Discovering regulators of centriole biogenesis through siRNA-based functional genomics in human cells. *Dev Cell,* 25**,** 555-571.

Barbelanne, M. & Tsang, W. Y. 2014. Molecular and cellular basis of autosomal recessive primary microcephaly. *Biomed Res Int,* 2014**,** 547986.

Barr, A. R. & Gergely, F. 2007. Aurora-A: the maker and breaker of spindle poles. *J Cell Sci,* 120**,** 2987-2996.

Barr, A. R., Kilmartin, J. V. & Gergely, F. 2010. CDK5RAP2 functions in centrosome to spindle pole attachment and DNA damage response. *J Cell Biol,* 189**,** 23-39.

Barrera, J. A., Kao, L. R., Hammer, R. E., Seemann, J., Fuchs, J. L. & Megraw, T. L. 2010. CDK5RAP2 regulates centriole engagement and cohesion in mice. *Dev Cell,* 18**,** 913-926.

Basten, S. G. & Giles, R. H. 2013. Functional aspects of primary cilia in signaling, cell cycle and tumorigenesis. *Cilia,* 2**,** 6.

Basto, R., Brunk, K., Vinogradova, T., Peel, N., Franz, A., Khodjakov, A. & Raff, J. W. 2008. Centrosome Amplification Can Initiate Tumorigenesis in Flies. *Cell,* 133**,** 1032-1042.

Basto, R., Lau, J., Vinogradova, T., Gardiol, A., Woods, C. G., Khodjakov, A. & Raff, J. W. 2006. Flies without centrioles. *Cell,* 125**,** 1375-1386.

Baumbach, J. 2014. *Studying centrosome formation and the consequences of centrosome loss in Drosophila melanogaster.* Doctor of Philosophy, University of Oxford.

Baumbach, J., Levesque, M. P. & Raff, J. W. 2012. Centrosome loss or amplification does not dramatically perturb global gene expression in Drosophila. *Biol Open,* 1**,** 983-993.

Baumbach, J., Novak, Z. A., Raff, J. W. & Wainman, A. 2015. Dissecting the function and assembly of acentriolar microtubule organizing centers in Drosophila cells in vivo. *PLoS Genet,* 11**,** e1005261.

Bazzi, H. & Anderson, K. V. 2014. Acentriolar mitosis activates a p53-dependent apoptosis pathway in the mouse embryo. *Proc Natl Acad Sci U S A,* 111**,** E1491-1500.

Berdnik, D. & Knoblich, J. A. 2002. Drosophila Aurora-A is required for centrosome maturation and actin-dependent asymmetric protein localization during mitosis. *Curr Biol,* 12**,** 640-647.

Berns, M. W. & Richardson, S. M. 1997. Continuation of mitosis after selective laser microbeam destruction of the centriolar region. *J Cell Biol,* 75**,** 977-982.

Bettencourt-Dias, M. & Glover, D. M. 2007. Centrosome biogenesis and function: centrosomics brings new understanding. *Nat Rev Mol Cell Biol,* 8**,** 451-463.

Bettencourt-Dias, M., Hildebrandt, F., Pellman, D., Woods, G. & Godinho, S. A. 2011. Centrosomes and cilia in human disease. *Trends Genet,* 27**,** 307-315.

Bettencourt-Dias, M., Rodrigues-Martins, A., Carpenter, L., Riparbelli, M., Lehmann, L., Gatt, M. K., Carmo, N., Balloux, F., Callaini, G. & Glover, D. M. 2005. SAK/PLK4 is required for centriole duplication and flagella development. *Curr Biol,* 15**,** 2199-2207.

Blachon, S., Gopalakrishnan, J., Omori, Y., Polyanovsky, A., Church, A., Nicastro, D., Malicki, J. & Avidor-Reiss, T. 2008. Drosophila asterless and vertebrate Cep152 Are orthologs essential for centriole duplication. *Genetics,* 180**,** 2081-2094.

Bonaccorsi, S., Giansanti, M. G. & Gatti, M. 1998. Spindle self-organization and cytokinesis during male meiosis in asterless mutants of Drosophila melanogaster. *JOURNAL OF CELL BIOLOGY,* 142**,** 751-761.

Bonaccorsi, S., Giansanti, M. G. & Gatti, M. 2000. Spindle assembly in Drosophila neuroblasts and ganglion mother cells. *Nat Cell Biol,* 2**,** 54-56.

Bond, J., Roberts, E., Mochinda, G. H., Hampshire, D. J., Scott, S., Askham, J. M., Springell, K., Mahadevan, M., Crow, Y. J., Markham, A. F., Walsh, C. A. & Woods, C. G. 2002. ASPM is a major determinant of cerebral cortical size. *Nat Genet,* 32**,** 316-320.

Bond, J., Roberts, E., Springell, K., Lizarraga, S. B., Scott, S., Higgins, J., Hampshire, D. J., Morrison, E. E., Leal, G. F., Silva, E. O., Costa, S. M., Baralle, D., Raponi, M., Karbani, G., Rashid, Y., Jafri, H., Bennett, C., Corry, P., Walsh, C. A. & Woods, C. G. 2005. A centrosomal mechanism involving CDK5RAP2 and CENPJ controls brain size. *Nat Genet,* 37**,** 353-355.

Bork, P., Holm, L. & Sander, C. 1994. The immunoglobulin fold. Structural classification, sequence patterns and common core. *J Mol Biol,* 242**,** 309-320.

Bornens, M. 2002. Centrosome composition and microtubule anchoring mechanisms. *Curr Opin Cell Biol,* 14**,** 25-34.

Bornens, M. 2012. The centrosome in cells and organisms. *Science,* 335**,** 422-426.

Bornens, M., Paintrand, M., Berges, J., Marty, M. C. & Karsenti, E. 1987. Structural and chemical characterization of isolated centrosomes. *Cell Motil Cytoskeleton,* 8**,** 238-249.

Boveri, T. 1914. *The origin of malignant tumours*.

Breslow, D. K., Koslover, E. F., Seydel, F., Spakowitz, A. J. & Nachury, M. V. 2013. An in vitro assay for entry into cilia reveals unique properties of the soluble diffusion barrier. *J Cell Biol,* 203**,** 129-147.

Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O. S., Wonrhein, C. & Womack, T. O. 2011. BUSTER version 2.10.1. Cambridge, United Kingdom: Global Phasing Ltd.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. 2003. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr D Biol Crystallogr,* 59**,** 2023-2030.

Broekhuis, J. R., Leong, W. Y. & Jansen, G. 2013. Regulation of cilium length and intraflagellar transport. *Int Rev Cell Mol Biol,* 303**,** 101-138.

Bruinsma, W., Macurek, L., Freire, R., Lindqvist, A. & Medema, R. H. 2014. Bora and Aurora-A continue to activate Plk1 in mitosis. *J Cell Sci,* 127**,** 801-811.

Bruinsma, W., Raaijmakers, J. A. & Medema, R. H. 2012. Switching Polo-like kinase-1 on and off in time and space. *Trends Biochem Sci,* 37**,** 534-542.

Buchman, J. J., Tseng, H. C., Zhou, Y., Frank, C. L., Xie, Z. & Tsai, L. H. 2010. Cdk5rap2 interacts with pericentrin to maintain the neural progenitor pool in the developing neocortex. *Neuron,* 66**,** 386-402.

Buerger, M. J. 1940. The Correction of X-Ray Diffraction Intensities for Lorentz and Polarization Factors. *Proc Natl Acad Sci U S A,* 26**,** 637-642.

Bullock, T. L., Roberts, T. M. & Stewart, M. 1996. 2.5 Å Resolution Crystal Structure of the Motile Major Sperm Protein (MSP) ofAscaris suum. *J Mol Biol,* 263**,** 284-296.

Cable, J. 2012. What can NMR do for you? - Part One.

Callaini, G., Whitfield, W. G. & Riparbelli, M. G. 1997. Centriole and centrosome dynamics during the embryonic cell cycles that follow the formation of the cellular blastoderm in Drosophila. *Exp Cell Res,* 234**,** 183-190.

Carvalho-Santos, Z., Machado, P., Branco, P., Tavares-Cadete, F., Rodrigues-Martins, A., Pereira-Leal, J. B. & Bettencourt-Dias, M. 2010. Stepwise evolution of the centriole-assembly pathway. *J Cell Sci,* 123**,** 1414-1426.

Castellanos, E., Dominguez, P. & Gonzalez, C. 2008. Centrosome dysfunction in Drosophila neural stem cells causes tumors that are not due to genome instability. *Curr Biol,* 18**,** 1209-1214.

Chavali, P. L., Putz, M. & Gergely, F. 2014. Small organelle, big responsibility: the role of centrosomes in development and disease. *Philos Trans R Soc Lond B Biol Sci,* 369.

Chen, C. T., Hehnly, H., Yu, Q., Farkas, D., Zheng, G., Redick, S. D., Hung, H. F., Samtani, R., Jurczyk, A., Akbarian, S., Wise, C., Jackson, A., Bober, M., Guo, Y., Lo, C. & Doxsey, S. 2014. A unique set of centrosome proteins requires pericentrin for spindle-pole localization and spindle orientation. *Curr Biol,* 24**,** 2327-2334.

Choi, Y. K., Liu, P., Sze, S. K., Dai, C. & Qi, R. Z. 2010. CDK5RAP2 stimulates microtubule nucleation by the gamma-tubulin ring complex. *J Cell Biol,* 191**,** 1089-1095.

Christoph, J., Philipp, B., Ulrich, R., Dieter, B. & Stefan, D. 2010. Protein-binding assays in biological liquids using microscale thermophoresis. *Nature Communications,* 1**,** 1-7.

Cizmecioglu, O., Arnold, M., Bahtz, R., Settele, F., Ehret, L., Haselmann-Weiss, U., Antony, C. & Hoffmann, I. 2010. Cep152 acts as a scaffold for recruitment of Plk4 and CPAP to the centrosome. *J Cell Biol,* 191**,** 731-739.

Clarke, J., Cota, E., Fowler, S. B. & Hamill, S. J. 1999. Folding studies of immunoglobulin-like β-sandwich proteins suggest that they share a common folding pathway. *Structure,* 7**,** 1145-1153.

Clarke, P. R. & Zhang, C. 2008. Spatial and temporal coordination of mitosis by Ran GTPase. *Nat Rev Mol Cell Biol,* 9**,** 464-477.

Conduit, P. T. 2010. *Regulation of centrosome size in Drosophila.* PhD, Univeristy of Cambrdige.

Conduit, P. T., Brunk, K., Dobbelaere, J., Dix, C. I., Lucas, E. P. & Raff, J. W. 2010. Centrioles regulate centrosome size by controlling the rate of Cnn incorporation into the PCM. *Curr Biol,* 20**,** 2178-2186.

Conduit, P. T., Feng, Z., Richens, J. H., Baumbach, J., Wainman, A., Bakshi, S. D., Dobbelaere, J., Johnson, S., Lea, S. M. & Raff, J. W. 2014a. The centrosome-specific phosphorylation of Cnn by Polo/Plk1 drives Cnn scaffold assembly and centrosome maturation. *Dev Cell,* 28**,** 659-669.

Conduit, P. T. & Raff, J. W. 2015. Different Drosophila cell types exhibit differences in mitotic centrosome assembly dynamics. *Curr Biol,* 25**,** R650-651.

Conduit, P. T., Richens, J. H., Wainman, A., Holder, J., Vicente, C. C., Pratt, M. B., Dix, C. I., Novak, Z. A., Dobbie, I. M., Schermelleh, L. & Raff, J. W. 2014b. A molecular mechanism of mitotic centrosome assembly in Drosophila. *Elife,* 3**,** e03399.

Conduit, P. T., Wainman, A. & Raff, J. W. 2015. Centrosome function and assembly in animal cells. *Nat Rev Mol Cell Biol,* 16**,** 611-624.

Cottee, M. A. 2014. *Building the Drosophila centriole: a structural investigation of the centriolar proteins SAS-6, SAS-4 and Ana2.* Ph.D, University of Oxford.

Cottee, M. A., Muschalik, N., Johnson, S., Leveson, J., Raff, J. W. & Lea, S. M. 2015. The homo-oligomerisation of both Sas-6 and Ana2 is required for efficient centriole assembly in flies. *eLife,* 4**,** e07236.

Cottee, M. A., Muschalik, N., Wong, Y. L., Johnson, C. M., Johnson, S., Andreeva, A., Oegema, K., Lea, S. M., Raff, J. W. & van Breugel, M. 2013. Crystal structures of the CPAP/STIL complex reveal its role in centriole assembly and human microcephaly. *Elife,* 2**,** e01071.

Cowtan, K. 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr,* 62**,** 1002-1011.

Crick, F. H. C. 1953. The packing of α-helices: simple coiled-coils. *Acta Crystallogr D Biol Crystallogr,* 6**,** 689-697.

Cromer, D. T. 1983. Calculation of Anomalous Scattering Factors at Arbitrary Wavelengths. *J Applied Cryst,* 16**,** 437-438.

Culver, B. P., Meehl, J. B., Giddings, T. H., Jr. & Winey, M. 2009. The two SAS-6 homologs in Tetrahymena thermophila have distinct functions in basal body assembly. *Mol Biol Cell,* 20**,** 1865-1877.

da Costa, K. S., Leal, E., dos Santos, A. M., Lima e Lima, A. H., Alves, C. N. & Lameira, J. 2014. Structural analysis of viral infectivity factor of HIV type 1 and its interaction with A3G, EloC and EloB. *PLoS One,* 9**,** e89116.

Daub, H., Olsen, J. V., Bairlein, M., Gnad, F., Oppermann, F. S., Korner, R., Greff, Z., Keri, G., Stemmann, O. & Mann, M. 2008. Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle. *Mol Cell,* 31**,** 438-448.

David, A., Liu, F., Tibellus, A., Vulprecht, J., Wald, D., Rothermei, U., Ohana, R., Seitel, A., Metzger, J., Ashery-Padan, R., Meinzer, H., Grone, H., Izraeli, S. & Kramer, A. 2014. Lack of centrioles and primary cilia in STIL−/− mouse embryos. *Cell Cycle,* 13**,** 2859-2868.

Dawe, H. R., Farr, H. & Gull, K. 2007. Centriole/basal body morphogenesis and migration during ciliogenesis in animal cells. *J Cell Sci,* 120**,** 7-15.

Debec, A., Sullivan, W. & Bettencourt-Dias, M. 2010. Centrioles: active players or passengers during mitosis? *Cell Mol Life Sci,* 67**,** 2173-2194.

Decker, M., Jaensch, S., Pozniakovsky, A., Zinke, A., O'Connell, K. F., Zachariae, W., Myers, E. & Hyman, A. A. 2011. Limiting amounts of centrosome material set centrosome size in C. elegans embryos. *Curr Biol,* 21**,** 1259-1267.

Delattre, M., Canard, C. & Gonczy, P. 2006. Sequential protein recruitment in C. elegans centriole formation. *Curr Biol,* 16**,** 1844-1849.

Delaval, B. & Doxsey, S. J. 2010. Pericentrin in cellular function and disease. *J Cell Biol,* 188**,** 181-190.

Delgehyr, N., Sillibourne, J. & Bornens, M. 2005. Microtubule nucleation and anchoring at the centrosome are independent processes linked by ninein function. *J Cell Sci,* 118**,** 1565-1575.

Dix, C. I. & Raff, J. W. 2007. Drosophila Spd-2 recruits PCM to the sperm centriole, but is dispensable for centriole duplication. *Curr Biol,* 17**,** 1759-1764.

Dobbelaere, J., Josue, F., Suijkerbuijk, S., Baum, B., Tapon, N. & Raff, J. 2008. A genome-wide RNAi screen to dissect centriole duplication and centrosome maturation in Drosophila. *PLoS Biol,* 6**,** e224.

Dos Santos, H. G., Abia, D., Janowski, R., Mortuza, G., Bertero, M. G., Boutin, M., Guarin, N., Mendez-Giraldez, R., Nunez, A., Pedrero, J. G., Redondo, P., Sanz, M., Speroni, S., Teichert, F., Bruix, M., Carazo, J. M., Gonzalez, C., Reina, J., Valpuesta, J. M., Vernos, I., Zabala, J. C., Montoya, G., Coll, M., Bastolla, U. & Serrano, L. 2013. Structure and non-structure of centrosomal proteins. *PLoS One,* 8**,** e62633.

Drenth, J. 2010. *Principles of protein X-ray crystallography*, Springer.

Dzhindzhev, N. S., Tzolovsky, G., Lipinszki, Z., Schneider, S., Lattao, R., Fu, J., Debski, J., Dadlez, M. & Glover, D. M. 2014. Plk4 phosphorylates Ana2 to trigger Sas6 recruitment and procentriole formation. *Curr Biol,* 24**,** 2526-2532.

Dzhindzhev, N. S., Yu, Q. D., Weiskopf, K., Tzolovsky, G., Cunha-Ferreira, I., Riparbelli, M., Rodrigues-Martins, A., Bettencourt-Dias, M., Callaini, G. & Glover, D. M. 2010. Asterless is a scaffold for the onset of centriole assembly. *Nature,* 467**,** 714-718.

EMBL. 2015. https://www.embl.de/pepcore/pepcore_services/protein_purification/chromatography/hiload16-60_superdex75/ [Online].

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

Evans, L., Mitchison, T. & Kirschner, M. 1985. Influence of the centrosome on the structure of nucleated microtubules. *J Cell Biol,* 100**,** 1185-1191.

Evans, P. 2006a. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr,* 62**,** 72-82.

Evans, P. 2006b. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr,* 62**,** 72-82.

Evans, P. R. & Murshudov, G. N. 2013a. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr,* D69**,** 1204-1214.

Evans, P. R. & Murshudov, G. N. 2013b. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr,* 69**,** 1204-1214.

Faheem, M., Naseer, M. I., Rasool, M., Chaudhary, A. G., Kumosani, T. A., Ilyas, A. M., Pushparaj, P., Ahmed, F., Algahtani, H. A., Al-Qahtani, M. H. & Saleh Jamal, H. 2015. Molecular genetics of human primary microcephaly: an overview. *BMC Med Genomics,* 8 Suppl 1**,** S4.

Fan, J., Griffiths, A. D., Lockhart, A., Cross, R. A. & Amos, L. A. 1996. Microtubule minus ends can be labelled with a phage display antibody specific to alpha-tubulin. *Journal of molecular biology,* 259**,** 325-330.

Firat-Karalar, E. N. & Stearns, T. 2014. The centriole duplication cycle. *Phil Trans R Soc B,* 369**,** 20130460.

Fish, J. L., Dehay, C., Kennedy, H. & Huttner, W. B. 2008. Making bigger brains-the evolution of neural-progenitor-cell division. *J Cell Sci,* 121**,** 2783-2793.

Foe, V. E. & Alberts, B. M. 1983. Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis. *J Cell Sci,* 61**,** 31-70.

Fong, K. W., Choi, Y. K., Rattner, J. B. & Qi, R. Z. 2008. CDK5RAP2 is a pericentriolar protein that functions in centrosomal attachment of the gamma-tubulin ring complex. *Mol Biol Cell,* 19**,** 115-125.

Fry, A. M. 2015. Solving the centriole disengagement puzzle. *Nat Cell Biol,* 17**,** 3-5.

Fu, J. & Glover, D. M. 2012. Structured illumination of the interface between centriole and peri-centriolar material. *Open Biol,* 2**,** 120104.

Gadde, S. & Heald, R. 2004. Mechanisms and molecules of the mitotic spindle. *Curr Biol,* 14**,** R797-805.

Ganem, N. J., Godinho, S. A. & Pellman, D. 2009. A mechanism linking extra centrosomes to chromosomal instability. *Nature,* 460**,** 278-282.

Ganss, B. & Jheon, A. 2004. Zinc finger transcription factors in skeletal development. *Crit Rev Oral Biol Med,* 15**,** 282-297.

Geimer, S. & Melkonian, M. 2004. The ultrastructure of the Chlamydomonas reinhardtii basal apparatus: identification of an early marker of radial asymmetry inherent in the basal body. *J Cell Sci,* 117**,** 2663-2674.

Genin, A., Desir, J., Lambert, N., Biervliet, M., Van Der Aa, N., Pierquin, G., Killian, A., Tosi, M., Urbina, M., Lefort, A., Libert, F., Pirson, I. & Abramowicz, M. 2012. Kinetochore KMN network gene CASC5 mutated in primary microcephaly. *Hum Mol Genet,* 21**,** 5306-5317.

Gergely, F. & Basto, R. 2008. Multiple centrosomes: together they stand, divided they fall. *Genes Dev.,* 22**,** 2291-2296.

Giansanti, M. G., Bucciarelli, E., Bonaccorsi, S. & Gatti, M. 2008. Drosophila SPD-2 is an essential centriole component required for PCM recruitment and astral-microtubule nucleation. *Curr Biol,* 18**,** 303-309.

Giansanti, M. G., Gatti, R. & Bonaccorsi, S. 2001. The role of centrosomes and astral microtubules during asymmetric division of Drosophila neuroblasts. *Development,* 128**,** 1137-1145.

Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. & Ben-Tai, N. 2003. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics,* 19**,** 163-164.

Godinho, S. A. & Pellman, D. 2014. Causes and consequences of centrosome abnormalities in cancer. *Philos Trans R Soc Lond B Biol Sci,* 369.

Godinho, S. A., Picone, R., Burute, M., Dagher, R., Su, Y., Leung, C. T., Polyak, K., Brugge, J. S., Thery, M. & Pellman, D. 2014. Oncogene-like induction of cellular invasion from centrosome amplification. *Nature,* 510**,** 167-171.

Gomez-Ferreria, M., Bashkurov, M., Helbig, A. O., Larsen, B., Pawson, T., Gingras, A. C. & Pelletier, L. 2012. Novel NEDD1 phosphorylation sites regulate c-tubulin binding and mitotic spindle assembly. *J Cell Sci,* 125**,** 3745-3751.

Gomez-Ferreria, M. A., Rath, U., Buster, D. W., Chanda, S. K., Caldwell, J. S., Rines, D. R. & Sharp, D. J. 2007. Human Cep192 is required for mitotic centrosome and spindle assembly. *Curr Biol,* 17**,** 1960-1966.

Gonczy, P. 2012. Towards a molecular architecture of centriole assembly. *Nat Rev Mol Cell Biol,* 13**,** 425-435.

Gopalakrishnan, J., Guichard, P., Smith, A. H., Schwarz, H., Agard, D. A., Marco, S. & Avidor-Reiss, T. 2010. Self-assembling SAS-6 multimer is a core centriole building block. *J Biol Chem,* 285**,** 8759-8770.

Gordon, D. J., Resio, B. & Pellman, D. 2012. Causes and consequences of aneuploidy in cancer. *Nat Rev Genet,* 13**,** 189-203.

Goshima, G., Mayer, M., Zhang, N., Stuurman, N. & Vale, R. D. 2008. Augmin: a protein complex required for cenrosome-independent microtubule generation within the spindle. *J Cell Biol,* 181**,** 421-429.

Gotz, M. & Huttner, W. B. 2005. The cell biology of neurogenesis. *Nat Rev Mol Cell Biol,* 6**,** 777-788.

Gould, R. R. & Borisy, G. G. 1977. The pericentriolar material in Chinese hamster ovary cells nucleates microtubule formation. *J Cell Biol,* 73**,** 601-615.

Graser, S., Stierhof, Y. D. & Nigg, E. A. 2007. Cep68 and Cep215 (Cdk5rap2) are required for centrosome cohesion. *J Cell Sci,* 120**,** 4321-4331.

Greenan, G., Brangwynne, C. P., Jaensch, S., Gharakhani, J., Julicher, F. & Hyman, A. A. 2010. Centrosome size sets mitotic spindle length in Caenorhabditis elegans embryos. *Curr Biol,* 20**,** 353-358.

Grosse-Kunstleve, R. W. & Adams, P. D. 2003. Substructure search procedures for macromolecular structures. *Acta Crystallogr D Biol Crystallogr,* 59**,** 1966-1973.

Guernsey, D. L., Jiang, H., Hussin, J., Arnold, M., Bouyakdan, K., Perry, S., Babineau-Sturk, T., Beis, J., Dumas, N., Evans, S. C., Ferguson, M., Matsuoka, M., Macgillivray, C., Nightingale, M., Patry, L., Rideout, A. L., Thomas, A., Orr, A., Hoffmann, I., Michaud, J. L., Awadalla, P., Meek, D. C., Ludman, M. & Samuels, M. E. 2010. Mutations in centrosomal protein CEP152 in primary microcephaly families linked to MCPH4. *Am J Hum Genet,* 87**,** 40-51.

Guichard, P., Chretien, D., Marco, S. & Tassin, A. M. 2010. Procentriole assembly revealed by cryo-electron tomography. *EMBO J,* 29**,** 1565-1572.

Guichard, P., Desfosses, A., Maheshwari, A., Hachet, V., Dietrich, C., Brune, A., Ishikawa, T., Sachse, C. & Gonczy, P. 2012. Cartwheel architecture of Trichonympha basal body. *Science,* 337**,** 553.

Guichard, P., Hachet, V., Majubu, N., Neves, A., Demurtas, D., Olieric, N., Fluckiger, I., Yamada, A., Kihara, K., Nishida, Y., Moriya, S., Steinmetz, M. O., Hongoh, Y. & Gonczy, P. 2013. Native architecture of the centriole proximal region reveals features underlying its 9-fold radial symmetry. *Curr Biol,* 23**,** 1620-1628.

Gunawardane, R. N., Martin, O. C., Cao, K., Zhang, L., Dej, K., Iwamatsu, A. & Zheng, Y. 2000. Characterization and reconstitution of Drosophila gamma-tubulin ring complex subunits. *J Cell Biol,* 151**,** 1513-1524.

Habedanck, R., Stierhof, Y. D., Wilkinson, C. J. & Nigg, E. A. 2005. The Polo kinase Plk4 functions in centriole duplication. *Nat Cell Biol,* 7**,** 1140-1146.

Hakoshima, T. 2005. Leucine zippers. *eLS*.

Hamill, D. R., Severson, A. F., Carter, J. C. & Bowerman, B. 2002. Centrosome maturation and mitotic spindle assembly in C-elegans require SPD-5, a protein with multiple coiled-coil domains. *developmental cell,* 3**,** 673-684.

Hanafusa, H., Kedashiro, S., Tezuka, M., Funatsu, M., Usami, S., Toyoshima, F. & Matsumoto, K. 2015. PLK1-dependent activation of LRRK1 regulates spindle orientation by phosphorylating CDK5RAP2. *Nat Cell Biol,* 17**,** 1024-1035.

Haren, L., Stearns, T. & Luders, J. 2009. Plk1-dependent recruitment of gamma-tubulin complexes to mitotic centrosomes involves multiple PCM components. *PLoS One,* 4**,** e5976.

Hatch, E. M., Kulukian, A., Holland, A. J., Cleveland, D. W. & Stearns, T. 2010. Cep152 interacts with Plk4 and is required for centriole duplication. *J Cell Biol,* 191**,** 721-729.

Hatzopoulos, G. N., Erat, M. C., Cutts, E., Rogala, K. B., Slater, L. M., Stansfeld, P. J. & Vakonakis, I. 2013. Structural analysis of the G-box domain of the microcephaly protein CPAP suggests a role in centriole architecture. *Structure,* 21**,** 2069-2077.

Heald, R., Tournebize, R., Blank, T., Sandaltzopoulos, R., Becker, P., Hyman, A. A. & Karsenti, E. 1996. Self-organization of microtubules into bipolar spindles around artificial chromosomes in Xenopus egg extracts. *Nature,* 382.

Heald, R., Tournebize, R., Habermann, A., Karsenti, E. & Hyman, A. 1997. Spindle assembly in Xenopus egg extracts: respective roles of centrosomes and microtubule self-organization. *J Cell Biol,* 138**,** 615-628.

Herzog, W. & Weber, K. 1977. In vitro assembly of pure tubulin into microtubules in the absence of microtubule-associated proteins and glycerol. *Proc Natl Acad Sci U S A.,* 74**,** 1860-1864.

Heuer, J. G., Li, K. J. & Kaufman, T. C. 1995. 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.

Hilbert, M., Erat, M. C., Hachet, V., Guichard, P., Blank, I. D., Flucklger, I., Slater, L., Lowe, E. D., Hatzopoulos, G. N., Steinmetz, M. O., Gonczy, P. & Vakonakis, I. 2013. Caenorhabditis elegans centriolar protein SAS-6 forms a spiral that is consistent with imparting a ninefold symmetry. *Proc Natl Acad Sci U S A.,* 110**,** 11373-11378.

Hiraki, M., Nakazawa, Y., Kamiya, R. & Hirono, M. 2007. Bld10p constitutes the cartwheel-spoke tip and stabilizes the 9-fold symmetry of the centriole. *Curr Biol,* 17**,** 1778-1783.

Hirono, M. 2014. Cartwheel assembly. *Philos Trans R Soc Lond B Biol Sci,* 369.

Hoffman, R., Benz Jr, E. J., Silberstein, L. E., Heslop, H., Weitz, J. & Anastalsi, J. 2013. *Hematology: diagnosis and treatment*, Elsevier Health Sciences.

Howard, J. & Hyman, A. A. 2003. Dynamics and mechanics of the microtubule plus end. *Nature,* 422**,** 753-758.

Huangfu, D., Liu, A., Rakeman, A. S., Murcia, N. S., Niswander, L. & Anderson, K. V. 2003. Hedgehog signalling in the mouse requires intraflagellar transport proteins. *Nature,* 426**,** 83-87.

Hussain, M. S., Baig, S. M., Neumann, S., Nurnberg, G., Farooq, M., Ahmad, I., Alef, T., Hennies, H. C., Technau, M., Altmuller, J., Frommolt, P., Thiele, H., Noegel, A. A. & Nurnberg, P. 2012. A truncating mutation of CEP135 causes primary microcephaly and disturbed centrosomal function. *Am J Hum Genet,* 90**,** 871-878.

Hussain, M. S., Baig, S. M., Neumann, S., Peche, V. S., Szczepanski, S., Numberg, G., Tariq, M., Jameel, M., Khan, T. N., Fatima, A., Malik, N. A., Ahmad, I., Altmuller, J., Frommolt, P., Thiele, H., Hohne, W., Yigit, G., Wollnik, B., Neubauer, B. A., Numberg, P. & Noegel, A. A. 2013. CDK6 associates with the centrosome during mitosis and is mutated in a large Pakistani family with primary microcephaly. *Hum Mol Genet,* 22**,** 5199-5214.

Huttner, W. B. & Kosodo, Y. 2005. Symmetric versus asymmetric cell division during neurogenesis in the developing vertebrate central nervous system. *Curr Opin Cell Biol,* 17**,** 648-657.

Ibrahim, R., Messaoudi, C., Chichon, F. J., Celati, C. & Marco, S. 2009. Electron tomography study of isolated human centrioles. *Microsc Res Tech,* 72**,** 42-48.

Incardona, M. F., Bourenkov, G. P., Levik, K., Pieritz, R. A., Popov, A. N. & Svensson, O. 2009. EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *J Synchrotron Radiat,* 16**,** 872-879.

Insolera, R., Bazzi, H., Shao, W., Anderson, K. V. & Shi, S. H. 2014. Cortical neurogenesis in the absence of centrioles. *Nat Neurosci,* 17**,** 1528-1535.

Jackson, A. P., Eastwood, H., Bell, S. M., Adu, J., Toomes, C., Carr, I. M., Roberts, E., Hampshire, D. J., Crow, Y. J., Mighell, A. J., Karbani, G., Jafri, H., Rashid, Y., Mueller, R. F., Markham, A. F. & Woods, C. G. 2002. Identification of microcephalin, a protein implicated in determining the size of the human brain. *Am J Hum Genet,* 71**,** 136-142.

Jana, S. C., Marteil, G. & Bettencourt-Dias, M. 2014. Mapping molecules to structure: unveiling secrets of centriole and cilia assembly with near-atomic resolution. *Curr Opin Cell Biol,* 26**,** 96-106.

Jang, Y. J., Ma, S., Terada, Y. & Erikson, R. L. 2002. Phosphorylation of threonine 210 and the role of serine 137 in the regulation of mammalian polo-like kinase. *J Biol Chem,* 277**,** 44115-44120.

Jaroszewski, L., Li, Z., Cai, X. H., Weber, C. & Godzik, A. 2011. FFAS server: novel features and applications. *Nucleic Acids Res,* 39**,** W38-44.

Jaroszewski, L., Rychlewski, L., Li, Z., Li, W. & Godzik, A. 2005. FFAS03: a server for profile--profile sequence alignments. *Nucleic Acids Res,* 33**,** W284-288.

Jia, Y., Fong, K. W., Choi, Y. K., See, S. S. & Qi, R. Z. 2013. Dynamic recruitment of CDK5RAP2 to centrosomes requires its association with dynein. *PLoS One,* 8**,** e68523.

Jiang, K. & Aknmanova, A. 2010. Microtubule tip-interacting proteins: a view from both ends. *Current Opinion in Cell Biology,* 23**,** 94-101.

Jones, D. T. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol,* 292**,** 195-202.

Joukov, V., De Nicolo, A., Rodrigues, A., Walter, J. C. & Livingston, D. M. 2010. Centrosomal protein of 192 kDa (Cep192) promotes centrosome-driven spindle assembly by engaging in organelle-specific Aurora A activation. *Proc Natl Acad Sci U S A.,* 107**,** 21022-21027.

Joukov, V., Walter, J. C. & De Nicolo, A. 2014. The Cep192-organized aurora A-Plk1 cascade is essential for centrosome cycle and bipolar spindle assembly. *mOL cELL,* 55**,** 578-591.

Kabsch, W. 2010. XDS. *Acta. Cryst. ,* D66**,** 125-132.

Kalab, P., Pralle, A., Isacoff, E. Y., Heald, R. & Weis, K. 2006. Analysis of a RanGTP-regulated gradient in mitotic somatic cells. *Nature,* 440**,** 697-701.

Kao, L. R. & Megraw, T. L. 2009. Centrocortin cooperates with centrosomin to organize Drosophila embryonic cleavage furrows. *Curr Biol,* 19**,** 937-942.

Kapitein, L. C. & Hoogenraad, C. C. 2011. Which way to go? Cytoskeletal organization and polarized transport in neurons. *Mol Cell Neurosci.,* 46**,** 9-20.

Karplus, P. A. & Diederichs, K. 2012. Linking crystallographic model and data quality. *Science,* 336**,** 1030-1033.

Ke, Y. & Yang, W. 2014. Primary cilium: an elaborate structure that blocks cell division? *Gene,* 547**,** 175-185.

Keating, T. J. & Borisy, G. G. 2000. Immunostructural evidence for the template mechanism of microtubule nucleation. *Nat Cell Biol,* 2**,** 352-357.

Keller, L. C., Romijn, E. P., Zamora, I., Yates, J. R., 3rd & Marshall, W. F. 2005. Proteomic analysis of isolated chlamydomonas centrioles reveals orthologs of ciliary-disease genes. *Curr Biol,* 15**,** 1090-1098.

Kellogg, D. R., Moritz, M. & Alberts, B. M. 1994. The centrosome and ceullar organisation. *Annu. Rev. Biochem,* 63**,** 639-674.

Kemp, C. A., Kopish, K. R., Zipperlen, P., Ahringer, J. & O'Connell, K. F. 2004. Centrosome maturation and duplication in C. elegans require the coiled-coil protein SPD-2. *developmental cell,* 6**,** 511-523.

Khodjakov, A., Cole, R. W., Oakley, B. R. & Rieder, C. L. 2000. Centrosome-independent mitotic spindle formation in vertebrates. *Curr Biol,* 10**,** 59-67.

Khodjakov, A., Rieder, C. L., Sluder, G., Cassels, G., Sibon, O. & Wang, C. L. 2002. De novo formation of centrosomes in vertebrate cells arrested during S phase. *J Cell Biol,* 158**,** 1171-1181.

Kim, S. & Rhee, K. 2014. Importance of the CEP215-pericentrin interaction for centrosome maturation during mitosis. *PLoS One,* 9**,** e87016.

Kim, T. S., Park, J. E., Shukla, A., Choi, S., Murugan, R. N., Lee, J. H., Ahn, M., Rhee, K., Bang, J. K., Kim, B. Y., Loncarek, J., Erikson, R. L. & Lee, K. S. 2013. Hierarchical recruitment of Plk4 and regulation of centriole biogenesis by two centrosomal scaffolds, Cep192 and Cep152. *Proc Natl Acad Sci U S A,* 110**,** E4849-4857.

Kitagawa, D., Vakonakis, I., Olieric, N., Hilbert, M., Keller, D., Olieric, V., Bortfeld, M., Erat, M. C., Fluckiger, I., Gonczy, P. & Steinmetz, M. O. 2011. Structural basis of the 9-fold symmetry of centrioles. *Cell,* 144**,** 364-375.

Klebba, J. E., Galletta, B. J., Nye, J., Plevock, K. M., Buster, D. W., Hollingsworth, N. A., Slep, K. C., Rusan, N. M. & Rogers, G. C. 2015. Two Polo-like kinase 4 binding domains in Asterless perform distinct roles in regulating kinase stability. *J Cell Biol,* 208**,** 401-414.

Kleylein-Sohn, J., Westendorf, J., Le Clech, M., Habedanck, R., Stierhof, Y. D. & Nigg, E. A. 2007. Plk4-induced centriole biogenesis in human cells. *Dev Cell,* 13**,** 190-202.

Knoblich, J. A. 2010. Asymmetric cell division: recent developments and their implications for tumour biology. *Nature Reviews Molecular Cell Biology,* 11**,** 849-860.

Kollman, J. M., Merdes, A., Mourey, L. & Agard, D. A. 2011. Microtubule nucleation by gamma-tubulin complexes. *Nat Rev Mol Cell Biol,* 12**,** 709-721.

Kollman, J. M., Polka, J. K., Zelter, A., Davis, T. N. & Agard, D. A. 2010. Microtubule nucleating gamma-TuSC assembles structures with 13-fold microtubule-like symmetry. *Nature,* 466**,** 879-882.

Konchanczyk, T., Drozd, A. & Krezel, A. 2015. Relationship between the architecture of zinc coordination and zinc binding affinity in proteins – insights into zinc regulation. *Metallomics,* 7.

Kratz, A. S., Barenz, F., Richter, K. T. & Hoffmann, I. 2015. Plk4-dependent phosphorylation of STIL is required for centriole duplication. *Biol Open,* 4**,** 370-377.

Krishna, S. S., Majumdar, I. & Grishin, N. V. 2003. Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res,* 31**,** 532-550.

Krissinel, E. & Henrick, K. 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr,* 60**,** 2256-2268.

Krissinel, E. & Henrick, K. 2007. Inference of macromolecular assemblies from crystalline state. *J Mol Biol,* 372**,** 774-797.

Kumar, A., Girimaji, S. C., Duvvari, M. R. & Blanton, S. H. 2009. Mutations in STIL, encoding a pericentriolar and centrosomal protein, cause primary microcephaly. *Am J Hum Genet,* 84**,** 286-290.

Kwon, M., Godinho, S. A., Chandhok, N. S., Ganem, N. J., Azioune, A., Thery, M. & Pellman, D. 2008. Mechanisms to suppress multipolar divisions in cancer cells with extra centrosomes. *Genes Dev,* 22**,** 2189-2203.

Laitaoja, M., Valjakka, J. & Janis, J. 2013. Zinc coordination spheres in protein structures. *Inorg Chem,* 52**,** 10983-10991.

Lambrus, B. G., Uetake, Y., Clutario, K. M., Daggubati, V., Synder, M., Sluder, G. & Holland, A. J. 2015. p53 protects against genome instability following centriole duplication failure. *J Cell Biol,* 210**,** 63-77.

Lane, H. A. & Nigg, E. A. 1996. Antibody microinjection reveals an essential role for human polo-like kinase 1 (Plk1) in the functional maturation of mitotic centrosomes. *J Cell Biol,* 135**,** 1701-1713.

Laos, T., Cabral, G. & Dammermann, A. 2015. Isotropic incorporation of SPD-5 underlies centrosome assembly in C. elegans. *Curr Biol,* 25**,** R648-649.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. & Higgins, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23**,** 2947-2948.

Lawo, S., Bashkurov, M., Mullin, M., Ferreria, M. G., Kittler, R., Habermann, B., Tagliaferro, A., Poser, I., Hutchins, J. R., Hegemann, B., Pinchev, D., Buchholz, F., Peters, J. M., Hyman, A. A., Gingras, A. C. & Pelletier, L. 2009. HAUS, the 8-subunit human Augmin complex, regulates centrosome and spindle integrity. *Curr Biol,* 19**,** 816-826.

Lawo, S., Hasegan, M., Gupta, G. D. & Pelletier, L. 2012. Subdiffraction imaging of centrosomes reveals higher-order organizational features of pericentriolar material. *Nat Cell Biol,* 14**,** 1148-1158.

Lebedev, A. A. & Isupov, M. N. 2012. Space group validation with Zanuda. *CCP4 newletter,* 48.

Ledbetter, M. C. & Porter, K. R. 1964. Morphology of microutubules of plant cell. *Science,* 144**,** 872-874.

Lee, J. C. & Timasheff, S. N. 1975. Reconstitution of microtubules from purified calf brain tubulin. *Biochemistry,* 14**,** 5183-5187.

Lee, K. & Rhee, K. 2011. PLK1 phosphorylation of pericentrin initiates centrosome maturation at the onset of mitosis. *J Cell Biol,* 195**,** 1093-1101.

Lee, S. & Rhee, K. 2010. CEP215 is involved in the dynein-dependent accumulation of pericentriolar matrix proteins for spindle pole formation. *Cell Cycle,* 9**,** 774-783.

Leidel, S. & Gonczy, P. 2005. Centrosome duplication and nematodes: recent insights from an old relationship. *Dev Cell,* 9**,** 317-325.

Lerit, D. A., Jordan, H. A., Poulton, J. S., Fagerstrom, C. J., Galletta, B. J., Peifer, M. & Rusan, N. M. 2015. Interphase centrosome organization by the PLP-Cnn scaffold is required for centrosome function. *J Cell Biol,* 210**,** 79-97.

Leslie, A. G. W. & Powell, H. R. 2007. Processing Diffraction Data with Mosflm. *Evolving Methods for Macromolecular Crystallography,* 245**,** 41-51.

Lettman, M. M., Wong, Y. L., Viscardi, V., Niessen, S., Chen, S. H., Shiau, A. K., Zhou, H., Desai, A. & Oegema, K. 2013. Direct binding of SAS-6 to ZYG-1 recruits SAS-6 to the mother centriole for cartwheel assembly. *Dev Cell,* 25**,** 284-298.

Li, J. B., Gerdes, J. M., Haycraft, C. J., Fan, Y., Teslovich, T. M., May-Simera, H., Li, H., Blacque, O. E., Li, L., Leitch, C. C., Lewis, R. A., Green, J. S., Parfrey, P. S., Leroux, M. R., Davidson, W. S., Beales, P. L., Guay-Woodford, L. M., Yoder, B. K., Stormo, G. D., Katsanis, N. & Dutcher, S. K. 2004. Comparative Genomics Identifies a Flagellar and Basal Body Proteome that Includes the BBS5 Human Disease Gene. *Cell,* 117**,** 541-552.

Li, S., Fernandez, J. J., Marshall, W. F. & Agard, D. A. 2012. Three-dimensional structure of basal body triplet revealed by electron cryo-tomography. *EMBO J,* 31**,** 552-562.

Li, S., Wang, H. & Groth, C. 2014. Drosophila neuroblasts as a new model for the study of stem cell self-renewal and tumour formation. *Bioscience Reports,* 34**,** 401-414.

Lin, T. C., Neuner, A., Schlosser, Y. T., Scharf, A. N., Weber, L. & Schiebel, E. 2014. Cell-cycle dependent phosphorylation of yeast pericentrin regulates gamma-TuSC-mediated microtubule nucleation. *Elife,* 3**,** e02208.

Lin, Y. C., Chang, C. W., Hsu, W. B., Tang, C. J., Lin, Y. N., Chou, E. J., Wu, C. T. & Tang, T. K. 2013. Human microcephaly protein CEP135 binds to hSAS-6 and CPAP, and is required for centriole assembly. *EMBO J,* 32**,** 1141-1154.

Lizarraga, S. B., Margossian, S. P., Harris, M. H., Campagna, D. R., Han, A. P., Blevins, S., Mudbhary, R., Barker, J. E., Walsh, C. A. & Fleming, M. D. 2010. Cdk5rap2 regulates centrosome function and chromosome segregation in neuronal progenitors. *Development,* 137**,** 1907-1917.

Lodish, H., Berk, A., Zipursky, S. & al., e. 2000. Microtubule Dynamics and Motor Proteins during Mitosis. *Molecular Cell Biology. 4th edition.* New York: W. H. Freeman.

Loncarek, J., Hergert, P., Magidson, V. & Khodjakov, A. 2008. Control of daughter centriole formation by the pericentriolar material. *Nat Cell Biol,* 10**,** 322-328.

Lucas, E. P. & Raff, J. W. 2007. Maintaining the proper connection between the centrioles and the pericentriolar matrix requires Drosophila centrosomin. *J Cell Biol,* 178**,** 725-732.

Lupas, A., Van Dyke, M. & Stock, J. 1991. Predicting coiled coils from protein sequences. *Science,* 252**,** 1162-1164.

Mahen, R. & Venkitaraman, A. R. 2012. Pattern formation in centrosome assembly. *Current Opinion in Cell Biology,* 24**,** 14-23.

Mahjoub, M. R. 2013. The importance of a single primary cilium. *Organogenesis,* 9**,** 61-69.

Mahoney, N. M., Goshima, G., Douglass, A. D. & Vale, R. D. 2006. Making microtubules and mitotic spindles in cells without functional centrosomes. *Curr Biol,* 16**,** 564-569.

Malicki, J. & Avidor-Reiss, T. 2014. From the cytoplasm into the cilium: Bon voyage. *Organogenesis,* 10**,** 138-157.

Manning, J. A., Shalini, S., Risk, J. M., Day, C. L. & Kumar, S. 2010. A direct interaction with NEDD1 regulates gamma-tubulin recruitment to the centrosome. *PLoS One,* 5**,** e9618.

Marthiens, V., Rujano, M. A., Pennetier, C., Tessier, S., Paul-Gilloteaux, P. & Basto, R. 2013. Centrosome amplification causes microcephaly. *Nat Cell Biol,* 15**,** 731-740.

Martinez-Campos, M., Basto, R., Baker, J., Kernan, M. & Raff, J. W. 2004. The Drosophila pericentrin-like protein is essential for cilia/flagella function, but appears to be dispensable for mitosis. *J Cell Biol,* 165**,** 673-683.

Mason, J. M. & Arndt, K. M. 2004. Coiled coil domains: stability, specificity, and biological implications. *Chembiochem,* 5**,** 170-176.

Matsuura, K., Lefebvre, P. A., Kamiya, R. & Hirono, M. 2004. Bld10p, a novel protein essential for basal body assembly in Chlamydomonas: localization to the cartwheel, the first ninefold symmetrical structure appearing during assembly. *J Cell Biol,* 165**,** 663-671.

Matthews, B. W. 1968. Solvent content of protein crystals. *J Mol Biol,* 33**,** 491-497.

McCall, K. A., Huang, C. & Fierke, C. A. 200. Function and Mechanism of Zinc Metalloenzymes. *J. Nutr.,* 130**,** 14375-14465.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. 2007. Phaser crystallographic software. *J Appl Crystallogr,* 40**,** 658-674.

McCoy, A. J. & Read, R. J. 2010. Experimental phasing: best practice and pitfalls. *Acta Crystallogr D Biol Crystallogr,* 66**,** 458-469.

Megraw, T. L., Kao, L. R. & Kaufman, T. C. 2001. Zygotic development without functional mitotic centrosomes. *Current Biology,* 11**,** 116-120.

Megraw, T. L., Kilaru, S., Turner, F. R. & Kaufman, T. C. 2002. The centrosome is a dynamic structure that ejects PCM flares. *J Cell Sci,* 115**,** 4707-4718.

Megraw, T. L., Li, K. J., Kao, L. R. & Kaufman, T. C. 1999. The Centrosomin protein is required for centrosome assembly and function during cleavage in Drosophila. *Development,* 126**,** 2829-2839.

Megraw, T. L., Sharkey, J. T. & Nowakowski, R. S. 2011. Cdk5rap2 exposes the centrosomal root of microcephaly syndromes. *Trends Cell Biol,* 21**,** 470-480.

Meng, L., Park, J. E., Kim, T. S., Lee, E. H., Park, S. Y., Zhou, M., Bang, J. K. & Lee, K. S. 2015. Bimodal Interaction of Mammalian Polo-Like Kinase 1 and a Centrosomal Scaffold, Cep192, in the Regulation of Bipolar Spindle Formation. *Mol Cell Biol,* 35**,** 2626-2640.

Mennella, V., Keszthelyi, B., McDonald, K. L., Chhun, B., Kan, F., Rogers, G. C., Huang, B. & Agard, D. A. 2012. Subdiffraction-resolution fluorescence microscopy reveals a domain of the centrosome critical for pericentriolar material organization. *Nat Cell Biol,* 14**,** 1159-1168.

Mitchell, A., Chang, H. Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S., Sangrador-Vegas, A., Scheremetjew, M., Rato, C., Yong, S. Y., Bateman, A., Punta, M., Attwood, T. K., Sigrist, C. J., Redaschi, N., Rivoire, C., Xenarios, I., Kahn, D., Guyot, D., Bork, P., Letunic, I., Gough, J., Oates, M., Haft, D., Huang, H., Natale, D. A., Wu, C. H., Orengo, C., Sillitoe, I., Mi, H., Thomas, P. D. & Finn, R. D. 2015. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res,* 43**,** D213-221.

Mitchison, T. J. 1993. Localization of an exchangeable GTP binding site at the plus end of microtubules. *Science,* 261**,** 1044-1047.

Mitelman, F., Johansson, B., Mertens, F. & editors 2015. Mitelman Database of Chromosome Aberrations and Genes Fusions in Cancer. http://cgap.nci.nih.gov/Chromosomes/Mitelman.

Moran, J., Christoph, J. W., Dieter, B., Philipp, B. & Stefan, D. 2011. Molecular interaction studies using microscale thermophoresis. *Assay Drug Dev Technol,* 9**,** 342-353.

Moritz, M., Braunfeld, M. B., Fung, J. C., Sedat, J. W., Alberts, B. M. & Agard, D. A. 1995a. Three-dimensional structural characterization of centrosomes from early Drosophila embryos. *J Cell Biol,* 130**,** 1149-1159.

Moritz, M., Braunfeld, M. B., Guenebaut, V., Heuser, J. & Agard, D. A. 2000. Structure of the gamma-tubulin ring complex: a template for microtubule nucleation. *Nat Cell Biol,* 2**,** 365-370.

Moritz, M., Braunfeld, M. B., Sedat, J. W., Alberts, B. & Agard, D. A. 1995b. Microtubule nucleation by gamma-tubulin-containing rings in the centrosome. *Nature,* 378**,** 638-640.

Moritz, M., Zheng, Y., Alberts, B. M. & Oegema, K. 1998. Recruitment of the gamma-tubulin ring complex to Drosophila salt-stripped centrosome scaffolds. *J Cell Biol,* 142**,** 775-786.

Moudjou, M., Bordes, N., Paintrand, M. & Bornens, M. 1996. gamma-Tubulin in mammalian cells: the centrosomal and the cytosolic forms. *J Cell Sci,* 109 (Pt 4)**,** 875-887.

Moutinho-Pereira, S., Stuurman, N., Afonso, O., Hornsveld, M., Aguiar, P., Goshima, G., Vale, R. D. & Maiato, H. 2013. Genes involved in centrosome-independent mitotic spindle assembly in Drosophila S2 cells. *Proc Natl Acad Sci U S A.,* 110**,** 19808-19813.

Muller, H., Schmidt, D., Steinbrink, S., Mirgorodskaya, E., Lehmann, V., Habermann, K., Dreher, F., Gustavsson, N., Kessler, T., Lehrach, H., Herwig, R., Gobom, J., Ploubidou, A., Boutros, M. & Lange, B. M. 2010. Proteomic and functional analysis of the mitotic Drosophila centrosome. *EMBO J,* 29**,** 3344-3357.

Murphy, S. M., Urbani, L. & Stearns, T. 1998. The mammalian gamma-tubulin complex contains homologues of the yeast spindle pole body components spc97p and spc98p. *J Cell Biol,* 141**,** 663-674.

Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr,* 67**,** 355-367.

Nakazawa, Y., Hiraki, M., Kamiya, R. & Hirono, M. 2007. SAS-6 is a cartwheel protein that establishes the 9-fold symmetry of the centriole. *Curr Biol,* 17**,** 2169-2174.

Nicholas, A. K., Khurshid, M., Desir, J., Carvalho, O. P., Cox, J. J., Thornton, G., Kausar, R., Ansar, M., Ahmad, W., Verloes, A., Passemard, S., Misson, J. P., Lindsay, S., Gergely, F., Dobyns, W. B., Roberts, E., Abramowicz, M. & Woods, C. G. 2010. WDR62 is associated with the spindle pole and is mutated in human microcephaly. *Nat Genet,* 42**,** 1010-1014.

Nigg, E. A. 2006. Origins and consequences of centrosome aberrations in human cancers. *Int J Cancer,* 119**,** 2717-2723.

Nigg, E. A., Cajanek, L. & Arquint, C. 2014. The centrosome duplication cycle in health and disease. *FEBS Lett,* 588**,** 2366-2372.

Nigg, E. A. & Raff, J. W. 2009. Centrioles, centrosomes, and cilia in health and disease. *Cell,* 139**,** 663-678.

Nigg, E. A. & Stearns, T. 2011. The centrosome cycle: Centriole biogenesis, duplication and inherent asymmetries. *Nat Cell Biol,* 13**,** 1154-1160.

Nishi, H., Shaytan, A. & Panchenko, A. R. 2014. Physicochemical mechanisms of protein regulation by phosphorylation. *Front Genet,* 5**,** 270.

Nogales, E., Whittaker, M., Milligan, R. A. & Downing, K. H. 1999. High-resolution model of the microtubule. *Cell,* 96**,** 79-88.

Nogales, E., Wolf, S. G. & Downing, K. H. 1998. Structure of the alpha beta tubulin dimer by electron crystallography. *Nature,* 391**,** 199-203.

Novak, Z. A., Conduit, P. T., Wainman, A. & Raff, J. W. 2014. Asterless licenses daughter centrioles to duplicate for the first time in Drosophila embryos. *Curr Biol,* 24**,** 1276-1282.

O'Connell, C. B. & Khodjakov, A. L. 2007. Cooperative mechanisms of mitotic spindle formation. *J Cell Sci,* 120**,** 1717-1722.

O'Connell, K. F., Caron, C., Kopish, K. R., Hurd, D. D., Kemphues, K. J., Li, Y. & White, J. G. 2001. 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.

O'Connell, K. F., Maxwell, K. N. & White, J. G. 2000. The spd-2 gene is required for polarization of the anteroposterior axis and formation of the sperm asters in the Caenorhabditis elegans zygote. *Dev Biol,* 222**,** 55-70.

O'Driscoll, M., Jackson, A. P. & Jeggo, P. A. 2006. Microcephalin: a causal link between impaired damage response signalling and microcephaly. *Cell Cycle,* 5**,** 2339-2344.

O'Rourke, B. P., Gomez-Ferreria, M. A., Berk, R. H., Hackl, A. M., Nicholas, M. P., O'Rourke, S. C., Pelletier, L. & Sharp, D. J. 2014. Cep192 controls the balance of centrosome and non-centrosomal microtubules during interphase. *PLoS One,* 9**,** e101001.

Oakley, B. R., Oakley, C. E., Yoon, Y. & Jung, M. K. 1990. Gamma-tubulin is a component of the spindle pole body that is essential for microtubule function in Aspergillus nidulans. *Cell,* 61**,** 1289-1301.

Oegema, K., Wiese, C., Martin, O. C., Milligan, R. A., Iwamatsu, A., Mitchison, T. J. & Zheng, Y. 1999. Characterization of two related Drosophila gamma-tubulin complexes that diifer in their ability to nucleate microtubules. *J Cell Biol,* 144**,** 721-733.

Ohkura, H., Hagan, I. M. & Glover, D. M. 1995. The conserved Schizosaccharomyces pombe kinase plo1, required to form a bipolar spindle, the actin ring, and septum, can drive septum formation in G1 and G2 cells. *Genes Dev.,* 9**,** 1059-1073.

Ohta, K., Shiina, N., Okumura, E., Hisanaga, S., Kishimoto, T., Endo, S., Gotoh, Y., Nishida, E. & Sakai, H. 1993. Microtubule nucleating activity of centrosomes in cell-free extracts from Xenopus eggs: involvement of phosphorylation and accumulation of pericentriolar material. *J Cell Sci,* 104 (Pt 1)**,** 125-137.

Ohta, M., Ashikawa, T., Nozaki, Y., Kozuka-Hata, H., Goto, H., Inagaki, M., Oyama, M. & Kitagawa, D. 2014. Direct interaction of Plk4 with STIL ensures formation of a single procentriole per parental centriole. *Nat Commun,* 5**,** 5267.

Oliveira, R. A., Hamilton, R. S., Pauli, A., Davis, I. & Nasmyth, K. 2010. Cohesin cleavage and Cdk inhibition trigger formation of daughter nuclei. *Nat Cell Biol,* 12**,** 185-192.

Pagan, J. K., Marzio, A., Jones, M. J., Saraf, A., Jallepalli, P. V., Florens, L., Washburn, M. P. & Pagano, M. 2015. Degradation of Cep68 and PCNT cleavage mediate Cep215 removal from the PCM to allow centriole separation, disengagement and licensing. *Nat Cell Biol,* 17**,** 31-43.

Palazzo, R. E., Vogel, J. M., Schnackenberg, B. J., Hull, D. R. & Wu, X. 2000. Centrosome maturation. *Curr Top Dev Biol,* 49**,** 449-470.

Park, S. Y., Park, J. E., Kim, T. S., Kim, J. H., Kwak, M. J., Ku, B., Tian, L., Murugan, R. N., Ahn, M., Komiya, S., Hojo, H., Kim, N. H., Kim, B. Y., Bang, J. K., Erikson, R. L., Lee, K. W., Kim, S. J., Oh, B. H., Yang, W. & Lee, K. S. 2014. Molecular basis for unidirectional scaffold switching of human Plk4 in centriole biogenesis. *Nat Struct Mol Biol,* 21**,** 696-703.

Peel, N., Stevens, N. R., Basto, R. & Raff, J. W. 2007. Overexpressing centriole-replication proteins in vivo induces centriole overduplication and de novo formation. *Curr Biol,* 17**,** 834-843.

Pelletier, L., O'Toole, E., Schwager, A., Hyman, A. A. & Muller-Reichert, T. 2006. Centriole assembly in Caenorhabditis elegans. *Nature,* 444**,** 619-623.

Pelletier, L., Ozlu, N., Hannak, E., Cowan, C., Habermann, B., Ruer, M., Muller-Reichert, T. & Hyman, A. A. 2004. The Caenorhabditis elegans centrosomal protein SPD-2 is required for both pericentriolar material recruitment and centriole duplication. *Curr Biol,* 14**,** 863-873.

Petronczki, M., Lenart, P. & Peters, J. M. 2008. Polo on the Rise-from Mitotic Entry to Cytokinesis with Plk1. *Dev Cell,* 14**,** 646-659.

Petry, S., Groen, A. C., Ishihara, K., Mitchison, T. J. & Vale, R. D. 2013. Branching microtubule nucleation in Xenopus egg extracts mediated by augmin and TPX2. *Cell,* 152**,** 768-777.

Peyre, E. & Morin, X. 2012. An oblique view on the role of spindle orientation in vertebrate neurogenesis. *Dev Growth Differ,* 54**,** 287-305.

Piel, M., Meyer, P., Khodjakov, A., Rieder, C. L. & Bornens, M. 2000. The respective contributions of the mother and daughter centrioles to centrosome activity and behavior in vertebrate cells. *J Cell Biol,* 149**,** 317-330.

Piel, M., Nordberg, J., Euteneuer, U. & Bornens, M. 2001. Centrosome-dependent exit of cytokinesis in animal cells. *Science,* 291**,** 1550-1553.

Piel, M., Tulu, U. S., Wadsworth, P. & Cassimeris, L. 2004. Centrosome maturation: Measurement of microtubule nucleation throughout the cell cycle by using GFP-tagged EB1. *PNAS,* 101**,** 1584-1588.

Pirruccello, M., Swan, L. E., Folta-Stogniew, E. & Camilli, P. D. 2010. Recognition of the F&H motif by the Lowe syndrome protein OCRL. *Nature Structural & Molecular Biology,* 18**,** 789-795.

Ponting, C. P. 2006. A novel domain suggests a ciliary function for ASPM, a brain size determining gene. *Bioinformatics,* 22**,** 1031-1035.

Qiao, R., Cabral, G., Lettman, M. M., Dammermann, A. & Dong, G. 2012. SAS-6 coiled-coil structure and interaction with SAS-5 suggest a regulatory mechanism in C. elegans centriole assembly. *EMBO J,* 31**,** 4334-4347.

Quintyne, N. J., Reing, J. E., Hoffelder, D. R., Gollin, S. M. & Saunders, W. S. 2005. Spindle multipolarity is prevented by centrosomal clustering. *Science,* 307**,** 127-129.

Richens, J. H., Barros, T. P., Lucas, E. P., Peel, N., Pinto, D. M., Wainman, A. & Raff, J. W. 2015. The Drosophila Pericentrin-like-protein (PLP) cooperates with Cnn to maintain the integrity of the outer PCM. *Biol Open,* 4**,** 1052-1061.

Rieder, C. L. 2005. Kinetochore fiber formation in animal somatic cells: dueling mechanisms come to a draw. *Chromosoma,* 114**,** 310-318.

Ring, D., Hubble, R. & Kirschner, M. 1982. Mitosis in a cell with multiple centrioles. *J Cell Biol,* 94**,** 549-556.

Ringo, D. L. 1967. Flagellar motion and fine structure of the flagellar apparatus in Chlamydomonas. *J Cell Biol,* 33**,** 543-571.

RIparbelli, M. G. & Callaini, G. 2011. Male gametogenesis without centrioles. *Dev Biol,* 349**,** 427-439.

Roberts, D. B. 1986. *Drosophila: a practical approach,* Oxford, IRL Press.

Rogala, K. B., Dynes, N. J., Hatzopoulos, G. N., Yan, J., Pong, S. K., Robinson, C. V., Deane, C. M., Gonczy, P. & Vakonakis, I. 2015. The Caenorhabditis elegans protein SAS-5 forms large oligomeric assemblies critical for centriole formation. *Elife,* 4**,** e07410.

Roque, H., Wainman, A., Richens, J., Kozyrska, K., Franz, A. & Raff, J. W. 2012. Drosophila Cep135/Bld10 maintains proper centriole structure but is dispensable for cartwheel formation. *J Cell Sci,* 125**,** 5881-5886.

Roubin, R., Acquaviva, C., Chevrier, V., Sedjal, F., Zyss, D., Birnbaum, D. & Rosnet, O. 2013. Myomegalin is necessary for the formation of centrosomal and Golgi-derived microtubules. *Biol Open,* 2**,** 238-250.

Roy, S. 2009. The motile cilium in development and disease: emerging new insights. *Bioessays,* 31**,** 694-699.

Roy, S. 2011. Cilia and Hedgehog: When and how was their marriage solemnized? *Differentiation,* 83**,** S43-S48.

Santamaria, A., Wang, B., Elowe, S., Malik, R., Zhang, F., Bauer, M., Schmidt, A., Sillje, H. H., Korner, R. & Nigg, E. A. 2011. The Plk1-dependent phosphoproteome of the early mitotic spindle. *Mol Cell Proteomics,* 10**,** M110 004457.

Sawin, K. E., Lourenco, P. C. & Snaith, H. A. 2004. Microtubule nucleation at non-spindle pole body microtubule-organizing centers requires fission yeast centrosomin-related protein mod20p. *Curr Biol,* 14.

Schnackenberg, B. J., Khodjakov, A., Rieder, C. L. & Palazzo, R. E. 1998. The disassembly and reassembly of functional centrosomes in vitro. *Proc Natl Acad Sci U S A,* 95**,** 9295-9300.

Schou, K. B., Morthorst, S. K., Christensen, S. T. & Pedersen, L. B. 2014. Identification of conserved, centrosome-targeting ASH domains in TRAPPII complex subunits and TRAPPC8. *Cilia,* 3**,** 6.

Seki, A., Coppinger, J. A., Jang, C. Y., Yates, J. R. & Fang, G. 2008. Bora and the kinase Aurora a cooperatively activate the kinase Plk1 and control mitotic entry. *Science,* 320**,** 1655-1658.

Sherwood, D. & Cooper, J. 2011. *Crystals, X-rays and proteins,* United States, Oxford University Press.

Shi, J., Lua, S., Tong, J. S. & Song, J. 2010. Elimination of the native structure and solubility of the hVAPB MSP domain by the Pro56Ser mutation that causes amyotrophic lateral sclerosis. *Biochemistry,* 49.

Shimanovskaya, E., Viscardi, V., Lesigang, J., Lettman, M. M., Qiao, R., Svergun, D., Round, A., Oegema, K. & Dong, G. 2014. Structure of the C. elegans ZYG-1 cryptic polo box suggests a conserved mechanism for centriolar docking of Plk4 kinases. *Structure,* 22**,** 1090-1104.

Siegrist, S. E. & Doe, C. Q. 2007. Microtubule-induced cortical cell polarity. *Genes Dev,* 21**,** 483-496.

Silkworth, W. T., Nardi, I. K., Scholl, L. M. & Cimini, D. 2009. Multipolar spindle pole coalescence is a major source of kinetochore mis-attachment and chromosome mis-segregation in cancer cells. *PLoS ONE,* 4**,** E6564.

Sir, J. H., Barr, A. R., Nicholas, A. K., Carvalho, O. P., Khurshid, M., Sossick, A., Reichelt, S., D'Santos, C., Woods, C. G. & Gergely, F. 2011. A primary microcephaly protein complex forms a ring around parental centrioles. *Nat Genet,* 43**,** 1147-1153.

Sir, J. H., Putz, M., Daly, O., Morrison, C. G., Dunning, M., Kilmartin, J. V. & Gergely, F. 2013. Loss of centrioles causes chromosomal instability in vertebrate somatic cells. *J Cell Biol,* 203**,** 747-756.

Slevin, L. K., Nye, J., Pinkerton, D. C., Buster, D. W., Rogers, G. C. & Slep, K. C. 2012. The structure of the plk4 cryptic polo box reveals two tandem polo boxes required for centriole duplication. *Structure,* 20**,** 1905-1917.

Smyth, M. S. & Martin, J. H. J. 2000. X ray crystallography. *Mol Pathol,* 53**,** 8-14.

Snell, W. J., Dentler, W. L., Haimo, L. T., Binder, L. I. & Rosenbaum, J. L. 1974. Assembly of chick brain tubulin onto isolated basal bodies of Chlamydomonas reinhardi. *Science,* 185**,** 357-360.

Sonnen, K. F., Gabryjonczyk, A. M., Anselm, E., Stierhof, Y. D. & Nigg, E. A. 2013. Human Cep192 and Cep152 cooperate in Plk4 recruitment and centriole duplication. *J Cell Sci,* 126**,** 3223-3233.

Sonnen, K. F., Schermelleh, L., Leonhardt, H. & Nigg, E. A. 2012. 3D-structured illumination microscopy provides novel insight into architecture of human centrosomes. *Biology Open,* 0000**,** 1-12.

Spang, A., Geissier, S., Grein, K. & Schiebel, E. 1996. gamma-Tubulin-like Tub4p of Saccharomyces cerevisiae is associated with the spindle pole body substructures that organize microtubules and is required for mitotic spindle formation. *J Cell Biol,* 134**,** 429-441.

Stearns, T., Evans, L. & Kirschner, M. 1991. Gamma-tubulin is a highly conserved component of the centrosome. *Cell,* 65**,** 825-836.

Stearns, T. & Kirschner, M. 1994. In vitro reconstitution of centrosome assembly and function: the central role of gamma-tubulin. *Cell,* 76**,** 623-637.

Stein, N. 2008. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J Applied Cryst,* 41.

Stevens, N. R., Raposo, A. A. S. F., Basto, R., Johnston, D. S. & Raff, J. W. 2007. From Stem Cell to Embryo without Centrioles. *Curr Biol,* 17**,** 1498-1503.

Stevens, N. R., Roque, H. & Raff, J. W. 2010. DSas-6 and Ana2 coassemble into tubules to promote centriole duplication and engagement. *Dev Cell,* 19**,** 913-919.

Sun, T. & Hevner, R. F. 2014. Growth and folding of the mammalian cerebral cortex: from molecules to malformations. *Nat Rev Neurosci,* 15**,** 217-232.

Sung, C. H. & Leroux, M. R. 2013. The roles of evolutionarily conserved functional modules in cilia-related trafficking. *Nat Cell Biol,* 15**,** 1387-1397.

Sunkel, C. E. & Glover, D. M. 1988. polo, a mitotic mutant of Drosophila displaying abnormal spindle poles. *J Cell Sci,* 89 (Pt 1)**,** 25-38.

Takahashi, M., Yamagiwa, A., Nishimura, T., Mukai, H. & Ono, Y. 2002. Centrosomal proteins CG-NAP and kendrin provide microtubule nucleation sites by anchoring gamma-tubulin ring complex. *Mol Biol Cell,* 13**,** 3235-3245.

Tang, C. J., Fu, R. H., Wu, K. S., Hsu, W. B. & Tang, T. K. 2009. CPAP is a cell-cycle regulated protein that controls centriole length. . *Nat Cell Biol,* 11**,** 825-831.

Tang, C. J., Lin, S. Y., Hsu, W. B., Lin, Y. N., Wu, C. T., Lin, Y. C., Chang, C. W., Wu, K. S. & Tang, T. K. 2011. The human microcephaly protein STIL interacts with CPAP and is required for procentriole formation. *EMBO J,* 30**,** 4790-4804.

Terada, Y., Uetake, Y. & Kuriyama, R. 2003. Interaction of Aurora-A and centrosomin at the microtubule-nucleating site in Drosophila and mammalian cells. *J Cell Biol,* 162**,** 757-763.

Terwilliger, T. C. 2002. Automated structure solution, density modification and model building. *Acta Crystallogr D Biol Crystallogr,* 58**,** 1937-1940.

Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J. & Adams, P. D. 2008. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr,* 64**,** 61-69.

Thorton, G. K. & Woods, C. G. 2009. Primary microcephaly: do all roads lead to Rome? *Trends Genet,* 25**,** 501-510.

Tilney, L. G., Bryan, J., Bush, D. J., Fujiwara, K., Mooseker, M. S., Murphy, D. B. & Snyder, D. H. 1974. Microtubules: evidence for 13 protofilaments. *J Cell Biol,* 59**,** 267-275.

Tsou, M. F. & Stearns, T. 2006. Controlling centrosome number: licenses and blocks. *Curr Opin Cell Biol,* 18**,** 74-78.

Uehara, R. & Goshima, G. 2010. Functional central spindle assembly requires de novo microtubule generation in the interchrosomal region during anaphase. *J Cell Biol,* 191**,** 259-267.

Uetake, Y., Loncarek, J., Nordberg, J. J., English, C. N., La Terra, S., Khodjakov, A. & Sluder, G. 2007. Cell cycle progression and de novo centriole assembly after centrosomal removal in untransformed human cells. *J Cell Biol,* 176**,** 173-182.

Uetake, Y. & Sluder, G. 2010. Prolonged prometaphase blocks daughter cell proliferation despite normal completion of mitosis. *Curr Biol,* 20**,** 1666-1671.

UniProt, C. 2015. UniProt: a hub for protein information. *Nucleic Acids Res,* 43**,** D204-212.

Vagin, A. & Teplyakov, A. 2010. Molecular replacement with MOLREP. *ACta Crystallogr Sect D,* 66**,** 22-25.

Vaizel-Ohayon, D. & Schejter, E. D. 1999. Mutations in centrosomin reveal requirements for centrosomal function during early Drosophila embryogenesis. *Curr Biol,* 9**,** 889-898.

Valente, E. M., Rosti, R. O., Gibbs, E. & Gleeson, J. G. 2014. Primary cilia in neurodevelopmental disorders. *Nat Rev Neurol,* 10**,** 27-36.

van Breugel, M., Hirono, M., Andreeva, A., Yanagisawa, H. A., Yamaguchi, S., Nakazawa, Y., Morgner, N., Petrovich, M., Ebong, I. O., Robinson, C. V., Johnson, C. M., Veprintsev, D. & Zuber, B. 2011. Structures of SAS-6 suggest its organization in centrioles. *Science,* 331**,** 1196-1199.

van Breugel, M., Wilcken, R., McLaughlin, S. H., Rutherford, T. J. & Johnson, C. M. 2014. Structure of the SAS-6 cartwheel hub from Leishmania major. *Elife,* 3**,** e01812.

Vandre, D. D., Centonze, V. E., Peloquin, J., Tombes, R. M. & Borisy, G. G. 1991. Proteins of the mammalian mitotic spindle: phosphorylation/dephosphorylation of MAP-4 during mitosis. *J Cell Sci,* 98 (Pt 4)**,** 577-588.

Varmark, H., Llamazares, S., Rebollo, E., Lange, B., Reina, J., Schwarz, H. & Gonzalez, C. 2007. Asterless is a centriolar protein required for centrosome function and embryo development in Drosophila. *Curr Biol,* 17**,** 1735-1745.

Veland, I. R., Awan, A., Pedersen, L. B., Yoder, B. K. & Christensen, S. T. 2009. Primary cilia and signaling pathways in mammalian development, health and disease. *Nephron Physiol,* 111**,** 39-53.

Venkatesh, T. & Suresh, P. S. 2014. Emerging roles of MCPH1: expedition from primary microcephaly to cancer. *Eur J Cell Biol,* 93**,** 98-105.

Verde, F., Berrez, J.-M., Antony, C. & Karsenti, E. 1991. Taxol-induced Microtubule Asters in Mitotic Extracts of XenopusEggs: Requirement for Phosphorylated Factors and Cytoplasmic Dynem. *J Cell Biol,* 112**,** 1177-1187.

Verde, I., Pahlke, G., Salanova, M., Zhang, G., Wang, S., Coletti, D., Onuffer, J., Jin, S. L. C. & Conti, M. 2001. Myomegalin Is a Novel Protein of the Golgi/Centrosome That Interacts with a Cyclic Nucleotide Phosphodiesterase. *J Biol Chem,* 276**,** 11189-11198.

Vitre, B. D. & Cleveland, D. W. 2012. Centrosomes, chrosome instability (CIN) and aneuploidy. *Curr Opin Cell Biology,* 24**,** 809-815.

Wainman, A., Buster, D. W., Duncan, T., Metz, J., Ma, A., Sharp, D. & Wakefield, J. G. 2009. A new Augmin subunit, Msd1, demonstrates the importance of mitotic spindle-templated microtubule nucleation in the absence of functioning centrosomes. *Genes Dev,* 23**,** 1876-1881.

Walczak, C. E. & Heald, R. 2008. Mechanisms of mitotic spindle assembly and function. *Int. Rev. Cytol.,* 265**,** 111-158.

Walshaw, J. & Woolfson, D. N. 2001. SOCKET: a program for identifying and analysing coiled-coil motifs within protein structures *J Mol Biol,* 307**,** 1427-1450.

Walter, T. S., Meier, C., Assenberg, R., Au, K.-F., Ren, J., Verma, A., Nettleship, J. E., Owens, R. J., Stuart, D. I. & Grimes, J. M. 2006. Lysine methylation as a routine rescue strategy for protein crystallization. *Structure,* 14**,** 1617-1622.

Wang, X., Ching, Y. P., Lam, W. H., Qi, Z., Zhang, M. & Wang, J. H. 2000. Identification of a common protein association region in the neuronal Cdk5 activator. *J Biol Chem,* 275**,** 31763-31769.

Wang, Z., Wu, T., Shi, L., Zhang, L., Zheng, W., Qu, J. Y., Niu, R. & Qi, R. Z. 2010. Conserved motif of CDK5RAP2 mediates its localization to centrosomes and the Golgi complex. *J Biol Chem,* 285**,** 22658-22665.

Wiese, C. & Zheng, Y. 2000. A new function for the gamma-tubulin ring complex as a microtubule minus-end cap. *Nat Cell Biol,* 2**,** 358-364.

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D. & Hochstrasser, D. F. 1999. Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol,* 112**,** 531-552.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr,* 67**,** 235-242.

Winter, g. 2010. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Cryst.*

Wong, S. Y. & Reiter, J. F. 2008. The primary cilium at the crossroads of mammalian hedgehog signaling. *Curr Top Dev Biol,* 85**,** 225-260.

Wong, Y. L., Anzola, J. V., Davis, R. L., Yoon, M., Motamedi, A., Kroll, A., Seo, C. P., Hsia, J. E., Kim, S. K., Mitchell, J. W., Mitchell, B. J., Desai, A., Gahman, T. C., Shiau, A. K. & Oegema, K. 2015. Reversible centriole depletion with an inhibitor of Polo-like kinase 4. *Science,* 348**,** 1155-1160.

Woodruff, J. B., Wueseke, O. & Hyman, A. A. 2014. Pericentriolar material structure and dynamics. *Philos Trans R Soc Lond B Biol Sci,* 369**,** 1-10.

Woodruff, J. B., Wueseke, O., Viscardi, V., Mahamid, J., Ochoa, S. D., Bunkenborg, J., Widlund, P. O., Pozniakovsky, A., Zanin, E., Bahmanyar, S., Zinke, A., Hong, S. H., Decker, M., Baumeister, W., Andersen, J. S., Oegema, K. & Hyman, A. A. 2015. Centrosomes. Regulated assembly of a supramolecular centrosome scaffold in vitro. *Science,* 348**,** 808-812.

Wueseke, O., Bunkenborg, J., Hein, M. Y., Zinke, A., Viscardi, V., Woodruff, J. B., Oegema, K., Mann, M., Andersen, J. S. & Hyman, A. A. 2014. The C. elegans pericentriolar material components SPD-2 and SPD-5 are monomeric in the cytoplasm prior to incorporation into the PCM matrix. *Mol Biol Cell,* 25**,** 2984-2992.

Yamashita, Y. M., Mahowald, A. P., Perlin, J. R. & Fuller, M. T. 2007. Asymmetric inheritance of mother versus daughter centrosome in stem cell division. *Science,* 315**,** 518-521.

Yang, R. & Feldman, J. L. 2015. SPD-2/CEP192 and CDK are limiting for microtubule-organizing center function at the centrosome. *Curr Biol,* 25**,** 1924-1931.

Yang, Z., Loncarek, J., Khodjakov, A. & Rieder, C. L. 2008. Extra centrosomes and/or chromosomes prolong mitosis in human cells. *Nat Cell Biol,* 10**,** 748-751.

Yu, T. W., Mochida, G. H., Tischfield, D. J., Sgaier, S. K., Flores-Sarnat, L., Sergi, C. M., Topcu, M., McDonald, M. T., Barry, B. J., Felie, J. M., Sunu, C., Dobyns, W. B., Folkerth, R. D., Barkovich, A. J. & Walsh, C. A. 2010. Mutations in WDR62, encoding a centrosome-associated protein, cause microcephaly with simplified gyri and abnormal cortical architecture. *Nat Genet,* 42**,** 1015-1020.

Zhang, J. & Megraw, T. L. 2007. Proper recruitment of gamma-tubulin and D-TACC/Msps to embryonic Drosophila centrosomes requires Centrosomin Motif 1. *Mol Biol Cell,* 18**,** 4037-4049.

Zheng, Y., Wong, M. L., Alberts, B. & Mitchison, T. 1995. Nucleation of microtubule assembly by a gamma-tubulin-containing ring complex. *Nature,* 378**,** 578-583.

Zhu, F., Lawo, S., Bird, A., Pinchev, D., Ralph, A., Richter, C., Muller-Reichert, T., Kittler, R., Hyman, A. A. & Pelletier, L. 2008. The mammalian SPD-2 ortholog Cep192 regulates centrosome biogenesis. *Curr Biol,* 18**,** 136-141.

Zimmerman, W. C., Sillibourne, J., Rosa, J. & Doxsey, S. J. 2004. Mitosis-specific anchoring of gamma tubulin complexes by pericentrin controls spindle organization and mitotic entry. *Mol Biol Cell,* 15**,** 3642-3657.

Zitouni, S., Nabais, C., Jana, S. C., Guerrero, A. & Bettencourt-Dias, M. 2014. Polo-like kinases: structural variations lead to multiple functions. *Nat Rev Mol Cell Biol,* 15**,** 433-452.

Zwart, P. H., Grosse-Kunstleve, R. W. & Adams, P. D. 2005. Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation. *CCP4 newletter,* 43.

Zwicker, D., Decker, M., Jaensch, S., Hyman, A. A. & Julicher, F. 2014. Centrosomes are autocatalytic droplets of pericentriolar material organized by centrioles. *Proc Natl Acad Sci U S A,* 111**,** E2636-2645.