Molecular architecture of SAS-5 enables construction of a daughter centriole



Kacper Budzisław Rogala

Wolfson College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2015

Statement of Originality

I confirm that the work presented in this thesis is my own, and I have included the contribution statement detailing the collaborative work which formed a part of Chapter 4.

No part of this thesis has been accepted or is currently being submitted for any degree, diploma or certificate or other qualification in this University or elsewhere.

Kacper Budzisław Rogala

This thesis is dedicated to my Parents,
who have always believed in me

Acknowledgements

Although this thesis has only one name on its cover, there were a great many people who contributed not only directly to experiments presented on the following pages but also to shaping of my scientific interests and future goals. I am forever thankful to them.

First and foremost, I would like to express my sincerest gratitude to my supervisor, Dr John Vakonakis, who offered me to carry out the most exciting science project in the stimulating lab environment which had become my sandbox for three and a half years. John was not only a brilliant mind and a great researcher, from whom I could learn how good structural biology is done, but most of all, he was a mentor who guided me to completion of this thesis. I have been very fortunate to have a supervisor who cared so much about my work, and who gave me freedom to explore and pursue my own ideas. At the same time, he was able to weed out millions of potentially futile ideas, keeping me on the right track. Involvement and support of my supervisor helped me overcome an unexpected situation during the process of publishing work described in this thesis. He kept a sense of humour when I had lost mine.

I am deeply grateful to my co-supervisor, Professor Charlotte Deane, for involving me in various projects which taught me a lot about collaboration in science and made me a better researcher. Her career advice has also been particularly valuable to me, and I am thankful for that. I would also like to acknowledge members of my thesis committee, Professor Colin Kleanthous and Professor Christina Redfield, for their time and insights.

The members of the Vakonakis and Deane group have contributed a lot to both my professional and personal time at Oxford. There is not a single experiment in this thesis which I did not discuss with Dr George Hatzopoulos. His insightful comments and ideas were an inspiration to me. George's skill in the lab coupled with

his knowledge made him an invaluable living encyclopedia of research, available exactly when you need it, which in critical situations made a difference between a successful experiment and a failed one. I could not ask for a better colleague and friend. I would like to acknowledge all the past and present Vakonakis group members that I had the pleasure to work with: Leanne Slater, Dr Michéle Erat, Dr Dirk Reiter, Erin Cutts, Julia Busch, Jemma Day, Madeleine Butler, Jodie Ford and the numerous summer and Part II students who have come through the lab. I appreciate all the insightful discussions with members of the Deane group, and the regular group meetings which broadened my scientific horizons by introducing new ideas for my own research. I thank all past and present OPIG group members with whom I shared unforgettable Le Tour d'OPIG, punting and Christmas parties: Samuel Demharter, Malte Lücken, Dr Eoin Malins, Dr Jean-Paul Ebejer, Anthony Bradley, Saulo de Oliveira, Dr Bernhard Knapp, Jinwoo Leem, Jarosław Nowak, Dr Jamie Hill, Dr Konrad Krawczyk, Dr James Dunbar, Dr Hannah Edwards, Dr Reyhaneh Esmaielbeiki, Luis Ospina, Alistair Martin, Dr Markus Gerstel, Eleanor Law, Nicholas Pearce, Claire Marks, Cristian Regep, Florian Klimm, Hannah Patel, Dr Waqar Ali, Dr Henry Wilman, Dr Leila Alexander and Dr Sebastian Kelm.

I have appreciated the camaraderie and local expertise of the Schnell, Kleanthous, Newstead, Higgins and Garman groups on the ground floor of New Biochemistry, and lately those of Glycobiology Institute. I have drawn on expertise of many people from both Biochemistry, Chemistry and Structural Genomics Consortium departments, and I am thankful for their time, patience and advice: Dr David Staunton, Dr Edward Lowe, Sophie Gilbert, Dr Peter Appleford, Dr Yin Dong, Dr Thomas McCorvie, Dr Wyatt Yue and Professor Ian Campbell.

I am also eternally indebted to my programme director, Professor Elspeth Garman, who believed in me, and has been the spiritual help ever since I stepped into Oxford. She helped me make some of my most difficult decisions during my DPhil, and has always been there to listen and give advice. Furthermore, I am grateful to my college advisor, Dr Lucy Davison, who was an inexhaustible source of wisdom, and helped me make many crucial choices related to my career.

My attempt of understanding the role of SAS-5 in centriole assembly has led me to collaborate with a number of fantastic people who contributed an immense amount

of energy and time into this work: Dr Manuel Hilbert, Sarah Bianchi, Dr Nicola Dynes, Victor Pong, Dr Jun Yan, Dr Shabaz Mohammed, Professor Carol Robinson, Professor Pierre Gönczy and Professor Michel Steinmetz. You have enriched this research with ideas and results a single person could not possibly do.

I gratefully acknowledge funding agencies that made my DPhil work possible. I was funded by the Engineering and Physical Sciences Research Council studentship for four years, and was honoured to receive travel fellowships from Boehringer Ingelheim Fonds, Wolfson College and Biochemical Society to work abroad and to attend conferences.

My experience at Oxford would not have been the same if it was not for my class at the Doctoral Training Centre and for my fellow students at Wolfson College and my friends. Thank you for being there for me, Radosław Nowak, Mateusz Ujma, Yvonne Su, Jia Tsing Ng, Marta Wojnowska and Tamas Yelland.

Lastly, I would like to thank my family for all their love and encouragement. For my parents and my grandmother who have always unconditionally supported me in all my pursuits and took great interest in my research. And most of all, for Dipali, who not only was the most loving, kind, supportive and encouraging partner but also an inspirational scientist whose help and advice throughout my DPhil were absolutely invaluable. Thank you.

Abstract

In dividing cells, centrioles are duplicated once per cell cycle in a semi-conservative manner. A daughter centriole forms perpendicularly to the mother in a process templated by cartwheel-like structures. Cartwheels are found in centrioles of most eukaryotes, and are regarded as the key factor in establishing the nine-fold symmetry of centrioles. Cartwheels comprise the self-oligomerising protein SAS-6, recruitment of which to the mother centriole is mediated by direct binding to protein SAS-5 (also known as Ana2 or STIL). Although SAS-5 is an essential protein for centriole duplication, depletion of which completely terminates centrosome-dependent cell division, its exact role in this process has remained obscure. Using X-ray crystallography and a range of biophysical techniques, we have determined the molecular architecture of SAS-5. We show that SAS-5 forms a complex oligomeric structure, mediated by two self-associating domains: a trimeric coiled coil and a novel globular dimeric *Implico* domain. Disruption of either domain leads to centriole duplication failure in worm embryos, indicating that large SAS-5 assemblies are necessary for function. We propose that SAS-5 provides multivalent attachment sites that are critical for promoting assembly of SAS-6 into a cartwheel, and thus centriole formation.

Contents

List of Figures **xiii**

List of Tables **xvi**

List of Abbreviations **xvii**

List of Publications **xx**

1 Introduction **1**

1.1 Function of centrioles 1

1.1.1 Cell division 5

1.1.2 Basal bodies, cilia and flagella 7

1.2 Centriole biogenesis 11

1.2.1 Centriole multiplication 12

1.2.2 Assembly from unique components 12

1.2.3 The sequential order of events 14

1.2.4 Assembly of the pericentriolar material 16

1.2.5 Procentriole growth 17

1.3 Centriole architecture 20

1.3.1 The origin of centrioles 22

1.3.2 Molecular basis for centriolar symmetry 24

1.3.3 The unique cartwheels of *C. elegans* 30

1.3.4 Z-dimension and pinhead asymmetry 33

1.3.5 Doublets and triplets of microtubules and auxiliary proteins . 36

1.3.6 The A-C linker 39

1.3.7 Attachment of microtubules to cartwheels 40

1.4 Beyond the obvious . . . 44
1.4.1 SAS-5 defines a new protein family . . . 44
1.4.2 SAS-5 in cartwheel formation: interaction with SAS-6 . . . 47
1.4.3 Microcephaly and the SAS-5 / SAS-4 interaction . . . 54
1.4.4 Regulation of SAS-5 and SAS-6 proteins . . . 59
1.5 The aim of this work . . . 62

2 Solubilising SAS-5 **63**
2.1 Construction of pFloat expression system . . . 63
2.1.1 The host system . . . 64
2.1.2 Expression machinery . . . 65
2.1.3 Selection markers for cell growth . . . 68
2.1.4 Fusion tags . . . 69
2.1.5 Tag removal . . . 71
2.1.6 Selection marker for cloning . . . 73
2.1.7 Sequence fossils . . . 73
2.1.8 pFloat outlook . . . 74
2.2 Expression and purification of SAS-5_{FL} . . . 75
2.2.1 Initial cloning and expression trials . . . 75
2.2.2 Initial attempts of large scale expression and purification . . . 76
2.2.3 SAS-5 is a soluble aggregate . . . 78
2.2.4 The *sticky strand* . . . 79

3 Two folded domains of SAS-5 **84**
3.1 Sequence modules in SAS-5 family . . . 84
3.2 Structural architecture of SAS-5 . . . 85
3.2.1 Computational evaluation . . . 86
3.2.2 SAS-5 features two independently folded domains . . . 87
3.3 Coiled coil of SAS-5 is a moderately stable trimer . . . 92
3.3.1 Crystal structure of SAS-5_{CC} . . . 92
3.3.2 Trimer conservation . . . 97
3.3.3 Crystal structure of the extended coiled coil . . . 98
3.3.4 SAS-5_{CC} crystallography - summary . . . 99

3.4 SAS-5_{Imp} is a structurally novel dimeric domain 100

4 Oligomerisation of SAS-5 facilitates daughter centriole formation 113

4.1 Contribution statement 113
4.2 SAS-5 assembles into tetramers and hexamers in solution 114
4.3 Disrupted oligomerisation of SAS-5 117
4.3.1 Disruption of individual domains 118
4.3.2 Disruption of higher-order assemblies 118
4.4 Higher-order assemblies of SAS-5 are essential for centriole duplication *in vivo* 121
4.4.1 Phenotype of *C. elegans* SAS-5 124
4.4.2 Live cell imaging of mutant SAS-5 embryos 125
4.4.3 Centriolar localisation of SAS-5 mutants 125

5 Discussion 131

5.1 Eliminating insoluble protein bottleneck 132
5.1.1 Full-length SAS-5 for biochemical work 133
5.1.2 Challenges of working with modular proteins 135
5.2 *Implico* is a novel dimeric fold 137
5.3 Large oligomers of SAS-5 are important for centriole formation 139
5.3.1 Potential benefit from the avidity effect 139
5.3.2 Oligomeric state of SAS-5 140
5.3.3 Hypothetical mechanism of centriole formation in *C. elegans* . 141
5.3.4 Same but different: mechanism of centriole assembly in other species 144
5.4 Recent structural studies of SAS-5 orthologues 146
5.4.1 Interaction between STIL and PLK4 147
5.4.2 Dynein light chain and short intrinsically disordered segments of Ana2 148
5.4.3 Different modes of Ana2/STIL oligomerisation, similar downstream effects 150
5.4.4 New insights into centriole assembly pathway in flies and humans 152
5.5 The future of centriole research in *C. elegans* 154

6 Materials and methods **156**

6.1 Construction of pFloat vectors 156

6.2 Protein cloning, expression and purification 157

6.2.1 SAS-5 157

6.2.2 HRV-3C protease 158

6.3 Biophysical studies 158

6.3.1 Circular dichroism spectroscopy 158

6.3.2 Size-exclusion chromatography multi-angle light scattering . . 159

6.3.3 Native mass spectrometry 159

6.3.4 Negative-stain electron microscopy 160

6.4 X-ray crystallography 160

6.4.1 SAS-5_{CC} 160

6.4.2 SAS-$5_{CC\text{-}L2}$ 162

6.4.3 SAS-5_{Imp} 163

6.5 Molecular dynamics simulations 164

6.6 Nematode strains and RNA interference 164

6.7 Antibodies and immunofluorescence 165

6.8 Microscopy 165

6.9 Western blotting 166

6.10 Sequence notes 166

Appendix A **168**

Appendix B **170**

B.1 pFloat family of vectors 170

B.1.1 pFloat - full sequence 170

B.1.2 pFloat and pFloat-TAG - short sequences 173

B.1.2.1 pFloat 174

B.1.2.2 pFloat-STAG 174

B.1.2.3 pFloat-GST 174

B.1.2.4 pFloat-TRX 175

B.1.2.5 pFloat-Im9 175

B.1.2.6 pFloat-MBP 176

B.1.2.7 pFloat-SUMO 177
B.1.2.8 pFloat-HALO 177
B.1.2.9 pFloat-TF 178
B.1.2.10 pFloat-NusA 179
B.1.2.11 pFloat-MsyB 181
B.1.3 pFloat fluorescence fusions - short sequences 181
B.1.3.1 pFloat-YFP (monoVenus) 181
B.1.3.2 pFloat-eGFP 182
B.1.4 pFloat.fuse - short sequences 183
B.1.4.1 pFloat-MBP.fuse 183
B.1.4.2 pFloat-SUMO.fuse 184
B.1.4.3 pFloat-TRX.fuse 184
B.2 *sas-5* recoded for RNAi resistance 185
B.3 Expression constructs 186
B.3.1 SAS-5 proteins 186
B.3.1.1 MsyB-SAS-5_{FL} 186
B.3.1.2 MsyB-SAS-$5_{\Delta(282\text{-}295)}$ 186
B.3.1.3 MsyB-SAS-5_{FLEX} 187
B.3.1.4 SUMO-SAS-$5_{CC\text{-}L}$ 187
B.3.1.5 *C. briggsae* SUMO-SAS-$5_{(96\text{-}199)}$ 187
B.3.1.6 *C. brenneri* SUMO-SAS-$5_{(1\text{-}82)}$ 188
B.3.1.7 *C. remanei* SUMO-SAS-$5_{(98\text{-}206)}$ 188
B.3.1.8 *C. sinica* SUMO-SAS-$5_{(94\text{-}204)}$ 188
B.3.1.9 *C. tropicalis* SUMO-SAS-$5_{(94\text{-}199)}$ 188
B.3.2 HRV-$3C_{(1\text{-}181)}$ 189

Bibliography **190**

List of Figures

1.1 The centriole cycle 3
1.2 Centrioles as mitotic spindle organisers 4
1.3 *Xenopus* egg experiment 7
1.4 Multipolar spindle formation 8
1.5 Are centrioles indispensable? 9
1.6 Beating patterns of cilia and flagella 11
1.7 Multiplication of centrioles 13
1.8 Centriole assembly in *C. elegans* 15
1.9 Centriole assembly in other organisms 18
1.10 Dot-like localisation of PLK4 19
1.11 Model of procentriole assembly and growth 21
1.12 Architecture of centrioles and cartwheels 23
1.13 Hypothetical evolution of centrioles 25
1.14 Abnormal cartwheels in SAS-6 and Cep135 mutants 27
1.15 Self-oligomerisation of SAS-6 29
1.16 Cartwheels of *C. elegans* 31
1.17 Procentrioles in three dimensions 35
1.18 Doublet of microtubules 38
1.19 Microtubule triplets and the A-C linker 41
1.20 Potential involvement of CPAP in centriole architecture 45
1.21 SAS-5 family of proteins 48
1.22 Ana2 and SAS-6 work together to build large structures in flies 50
1.23 SAS-5 as a cytoplasm-centriole shuttle 52
1.24 SAS-5 / SAS-6 interaction 55
1.25 Potential cross-spoke connections in *Phlyctochytrium* and *Trichonympha* 56

1.26 SAS-5 / SAS-4 interaction 58

2.1 Vector map of pFloat 66
2.2 Gene expression circuit based on the *lac* operon 67
2.3 Plasmid replication control 68
2.4 Outline of the pFloat vectors 72
2.5 Purity assessment of recombinant proteins 77
2.6 *In vitro* protein aggregates of SAS-5_{FL} 81
2.7 *Sticky strand* drives aggregation of SAS-5 82
2.8 Removal of the *sticky strand* yields well-behaved SAS-5 83

3.1 Secondary structure and disorder predictions for SAS-5 86
3.2 Circular dichroism of isolated N- and C-terminal segments of SAS-5 . 89
3.3 Circular dichroism of isolated Middle segment of SAS-5 90
3.4 SAS-5 features two independently folded domains 91
3.5 Crystal structure of SAS-5_{CC} 94
3.6 Packing of molecules in the SAS-5_{CC} crystal 96
3.7 Computational relaxation of crystal packing-distorted SAS-5_{CC} 97
3.8 Hydrophobic interactions drive trimerisation of the coiled coil of SAS-5 102
3.9 Sequence conservation of SAS-5 coiled coil within the *Caenorhabditis* genus 103
3.10 Evolutionary conservation of the trimeric coiled coil of SAS-5 104
3.11 SAS-$5_{CC\text{-}L2}$ trimer intercalates in the crystal via both termini 106
3.12 Packing of molecules in the SAS-$5_{CC\text{-}L2}$ crystal 107
3.13 Zn^{2+} co-ordination and register shift in the crystal of SAS-$5_{CC\text{-}L2}$ . . . 108
3.14 Sequence conservation of SAS-5 *Implico* domain within the *Caenorhabditis* genus 109
3.15 SAS-5_{Imp} forms a stable dimer in solution 109
3.16 Crystal structure of SAS-5_{Imp} reveals a novel type of a protein dimer 110
3.17 Proline residues allow for tight turns in the compact SAS-5_{Imp} dimer 111
3.18 Packing of molecules in the SAS-5_{Imp} crystal 112

4.1 Hypothetical model of higher-order oligomerisation of SAS-5 115
4.2 Higher-order assemblies of SAS-5 in solution 116

4.3 Native mass spectrometry reveals presence of SAS-5 hexamers in solution 117
4.4 Point mutation disrupts the coiled-coil domain 119
4.5 Point mutation disrupts the *Implico* domain 120
4.6 Selective disruption of SAS-5 oligomerisation interfaces - CD 122
4.7 Selective disruption of SAS-5 oligomerisation interfaces - SEC-MALS 123
4.8 Higher-order oligomers of SAS-5 are essential for centriole duplication in worm embryos 126
4.9 Dual DIC and fluorescence time-lapse microscopy of oligomerisation mutant worm embryos 127
4.10 Expression levels of SAS-5 in mutant worm embryos 128
4.11 Defective centriolar localisation of SAS-5 oligomerisation mutants . . 130

5.1 Sequential assembly of SAS-5 oligomers 142
5.2 Hypothetical model of cartwheel nucleation 145

List of Tables

3.1 Crystallographic data collection and refinement statistics 93

A.1 Anisotropy correction statistics for native X-ray diffraction dataset of SAS-5_{CC}. 168

A.2 Protein concentrations used in CD experiments 169

A.3 Details of SEC-MALS experiments 169

List of Abbreviations

βTrCP	β-transducin repeats-containing protein
γTuRC	γ-tubulin ring complex
Imp	*Implico*
ccd	control of cell death
3D	three-dimensional
aa	amino acid
Ana2	anastral spindle 2
APC/C	anaphase-promoting complex/cyclosome
Asl	asterless
ASU	asymmetric unit
AUC	analytical ultra-centrifugation
Bld10p	bald 10 protein
Bld12p	bald 12 protein
bp	base pair
C-terminus	carboxyl-terminus
CC	coiled coil
CD	circular dichroism
Cep135	centrosomal protein of 135 kDa
Cep152	centrosomal protein of 152 kDa
Cep192	centrosomal protein of 192 kDa
CID	cartwheel inner densities
CP110	centriolar coiled-coil protein of 110 kDa
CPAP	centrosomal protein 4.1-associated protein
cryo-EM	cryo-electron microscopy
CV	column volume
DIC	differential interference contrast
DLS	Diamond Light Source
DM	density modification
DNA	deoxyribonucleic acid
dRI	differential refractive index
dsRNA	double-stranded RNA
DTT	dithiothreitol

EM	electron microscopy
ESRF	European Synchrotron Radiation Facility
FBXW5	F-Box/WD repeat-containing protein 5
FOM	figure of merit
FRAP	fluorescence recovery after photobleaching
GFP	green fluorescent protein
GST	glutathione S-transferase
HALO	haloalkane dehalogenase
HEPES-Na	sodium 4-(2-hydroxyethyl)piperazine-1-ethanesulfonate
HIV-1	human immunodeficiency virus type 1
HRV-3C	human rhinovirus 3C protease
IMAC	immobilised metal-affinity chromatography
IPTG	isopropyl-β-D-thiogalactoside
LacI	*lac* repressor
LacO	*lac* operator
LC8	dynein light chain of 8 kDa
MBP	maltose-binding protein
MCS	multiple cloning site
MD	molecular dynamics
MES	2-(N-morpholino)ethanesulfonic acid
MIP	microtubule inner protein
MME	monomethyl ether
MOPS	3-(N-morpholino)propanesulfonic acid
MPD	(RS)-2-methyl-2,4-pentanediol
MRSAD	molecular replacement with single-wavelength anomalous dispersion
MS	mass spectrometry
Mud	mushroom body defect
N-N	interaction between amino-terminal 'head' domains of SAS-6
N-terminus	amino-terminus
NCS	non-crystallographic symmetry
Ni-NTA	Ni-nitrilotriacetic acid
OMV	outer membrane vesicle
PB3	polo-box domain 3

PBS	phosphate-buffered saline
PCM	pericentriolar material
PDB ID	Protein Data Bank identification code
PEG	polyethylene glycol
PLK4	polo-like kinase 4
polyE	polyglutamic acid
PP2A	protein phosphatase 2
RBS	ribosome binding site
RMSD	root-mean-square deviation
RNA	ribonucleic acid
RNAi	RNA interference
RNAP	RNA polymerase
Rop	repressor of primer
SAD	single-wavelength anomalous dispersion
Sak	serum-inducible kinase/polo-like kinase-akin kinase
SAP	Sas6/Ana2 particle
SAS-4	spindle assembly abnormal 4
SAS-5	spindle assembly abnormal 5
SAS-6	spindle assembly abnormal 6
SAStubule	SAS-6, Ana2/SAS-5 tubule
SCF	seventeen kilodalton protein, cullin, F-box protein complex
SDS-PAGE	sodium dodecyl sulfate - polyacrylamide gel electrophoresis
SEC-MALS	size-exclusion chromatography multi-angle light scattering
SG	structural genomics
SPD-2	spindle-defective protein 2
STAN	STIL-Ana2 homology
STIL	stem cell leukaemia/T-cell acute lymphocytic leukaemia protein 1 interrupting locus
SUMO	small ubiquitin-like modifier
TEV	tobacco etch virus
TIM	truncated in microcephaly
TLS	translation-libration-screw-rotation
Tris	tris(hydroxymethyl)aminomethane
TRX	thioredoxin
UV	ultraviolet
ZYG-1	zygote defective protein 1

List of Publications

Over the course of my DPhil, from 2011 to 2015, I have contributed to the publication of the papers listed below. In particular, results presented in Rogala *et al.* (2015) form a significant part of this thesis, which was further expanded to also include data from auxiliary and unsuccessful experiments.

- Hatzopoulos, G. N., Erat, M. C., Cutts, E., **Rogala, K. B.**, Slater, L. M., Stansfeld, P. J., Vakonakis, I. Structural analysis of the G-Box domain of the microcephaly protein CPAP suggests a role in centriole architecture. *Structure 21*, (Nov. 2013), 2069-77.

- **Rogala, K. B.**, Dynes, N. J., Hatzopoulos, G. N., Yan, J., Pong, S. K., Robinson, C. V., Deane, C. M., Gönczy, P., Vakonakis, I. The *Caenorhabditis elegans* protein SAS-5 forms large oligomeric assemblies critical for centriole formation. *eLife 4* (May 2015), e07410.

- Palumbo, V., Pellacani, C., Heesom, K. J., **Rogala, K. B.**, Deane, C. M., Mottier-Pavie, V., Gatti, M., Bonaccorsi, S., Wakefield, J. G. Misato controls mitotic microtubule generation by stabilizing the tubulin chaperone protein-1 complex. *Current Biology 26*, (June 2015), 1777-83.

Chapter 1

Introduction

Microtubule networks, asymmetric inheritance, environmental sensing and locomotion are all but seemingly disconnected features of living cells that share one common element, a centriole. This barrel-shaped structure is an ancient macromolecular machine that is pivotal to formation of centrosomes, cilia and flagella, through which, it is engaged in three basic cellular activities: cell division, sensation and motion. In this chapter, I introduce the reader to the world of centrioles by explaining the fundamental roles centrioles play in a living organism. I review the recent progress in understanding the mechanisms which guide the assembly of centrioles, and pay particular attention to the first few steps in the assembly pathway. Deregulation of this pathway, arising from mutations in centriolar proteins, often leads to defects in the structure, function and numbers of centrioles, and is linked with a number of human pathologies, such as cancer and ciliopathies that affect many organ systems. I also discuss new advances that shed light on the fascinating molecular structure of centrioles, explaining how structure can mean function. Finally, I introduce SAS-5 as one of the key proteins involved in building new centrioles, which was the primary focus of this study.

1.1 Function of centrioles

Centrioles are present in most living organisms, from ancient protists to mammals. According to phylogenetic analysis, centrioles were present in the last common eukaryotic ancestor; however, they were lost several times during evolution, mostly in

lineages not contingent to their locomotive competency, such as fungi and higher plants [15, 51, 147].

A pair of orthogonally positioned centrioles, enveloped by a complex protein matrix, known as pericentriolar material (PCM), constitutes a centrosome which serves as the major microtubule-organising centre in animal cells (Figure 1.1A on page 3) [214].

Centrosomes sprout characteristic asters of microtubules that give rise to the mitotic spindle (Figure 1.1F on page 3 and Figure 1.2A on page 4), which in turn connects to chromosomes and segregates them into two daughter cells (Figure 1.2B on page 4). This mechanism has been of paramount importance in the evolution of the multi-cellular life on Earth by providing a framework where distribution of duplicated genetic material is both swift and reliable [36, 53]. Malfunctioning spindle apparatus leads to aberrant number of chromosomes in cells - a hallmark of cancer [121, 125, 241]. Another recognisable function of centrioles is to seed the growth of cilia and flagella - as basal bodies (Figure 1.1G on page 3). Flagella enable single cells, such as sperm, to swim, while cilia allow epithelial cells, such as those that line our air passages, to effectively clear out mucus towards throat, or serve as the chemo- and mechanosensory arsenal of the cell. While in dividing cells centrioles duplicate precisely once per cell cycle, multiciliated cell types are normally supplied with hundreds of basal bodies [44]. Centrioles also play integral part in establishment of cell polarities and in differential inheritance of various molecules during asymmetric cell division [34, 123, 279].

The striking nine-fold rotational symmetry of centrioles has captivated researchers for years (Figure 1.12 on page 23). Although symmetry is a common feature in nature, it is unclear why centrioles do not adopt eight-fold or ten-fold, but precisely nine-fold symmetry. Every duplication of centrioles meticulously reproduces the nine-fold symmetry, which appears to have been conserved throughout the evolution of eukaryotes [51]. Such conservation may reflect a complex evolutionary history of centrioles, where various constraints and trade-offs dictated their present shape. Centrioles seem to be required for mitosis in situations when fidelity and speed of chromosomal segregation are critical for development [218]. They constitute a 'special task force unit' which is activated to aid the cell in the event of severe pressure, but otherwise are

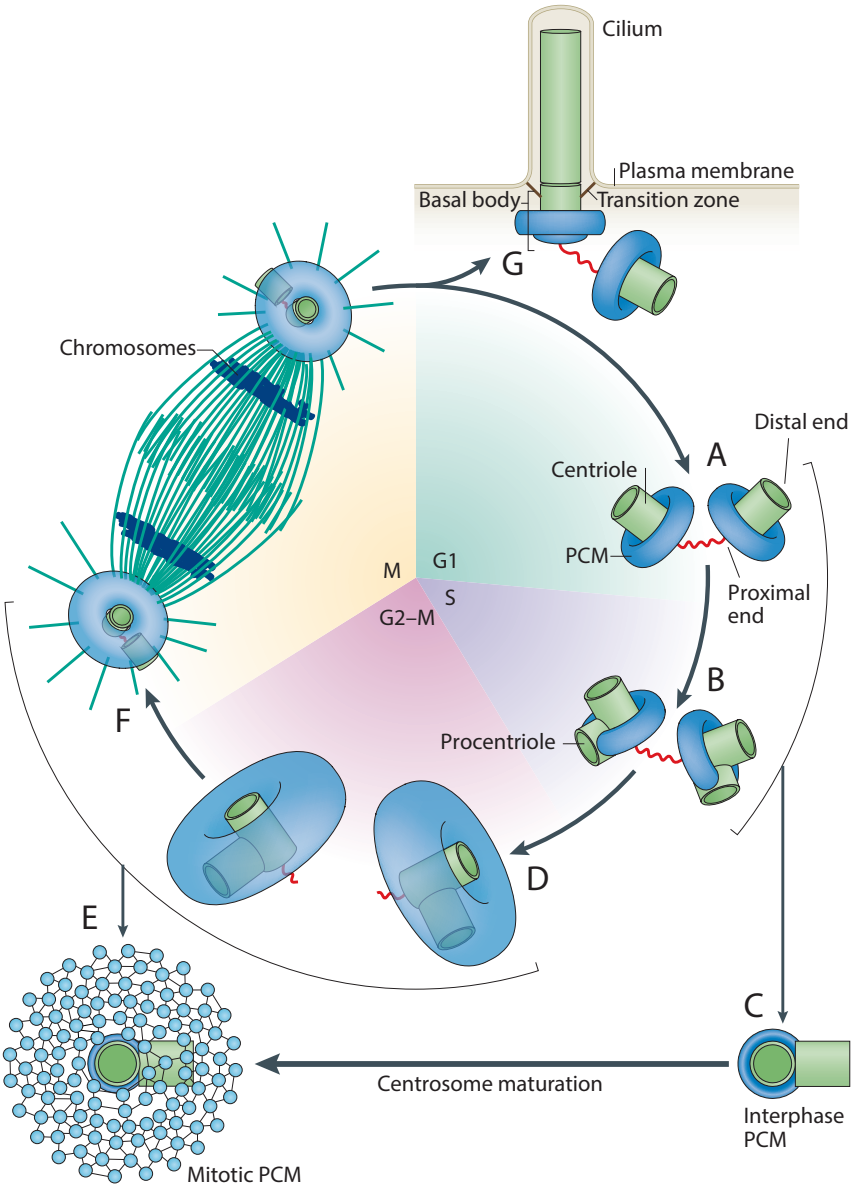


Figure 1.1: **The centriole cycle.** (**A**) A new cell that starts its life in the G1 phase of cell cycle inherits two centrioles (green cylinders), which are connected through their proximal ends via elastic fibres (red wave). By organising PCM (blue tori) around themselves, centrioles can form centrosomes. (**B**) At the transition from G1 to S phase, centrioles begin their duplication process - each assembling a procentriole (a centriole in an early stage of development) at right angles, adjacent to their proximal ends (the so-called 'engaged' conformation). (**C**) Interphasic centrioles organise very little PCM, and that PCM is largely recruited by the mature centrioles and not the procentrioles. (**D**) Once the cell has entered mitosis (G2-M), the fibres linking the pairs of centrioles are cut, allowing them to move apart. The PCM around the centrioles is rapidly increasing in size, and this expansion is thought to be organised in protein layers - as illustrated in **E**. (**F**) Such enlarged PCM enables nucleation and recruitment of an extensive network of microtubules, which assemble the mitotic spindle (green lines). With the exit from mitosis, the chromosomes (dark blue lines) are segregated to daughter cells, and the now-matured procentrioles 'disengage' from their orthogonal arrangement; however, they are still loosely linked via the elastic fibres. (**G**) In many non-dividing cells, the pair of centrioles migrates to the cell periphery, where the more mature centriole docks to the cell membrane, seeding growth of cilia as a basal body. Figure adapted from [63].

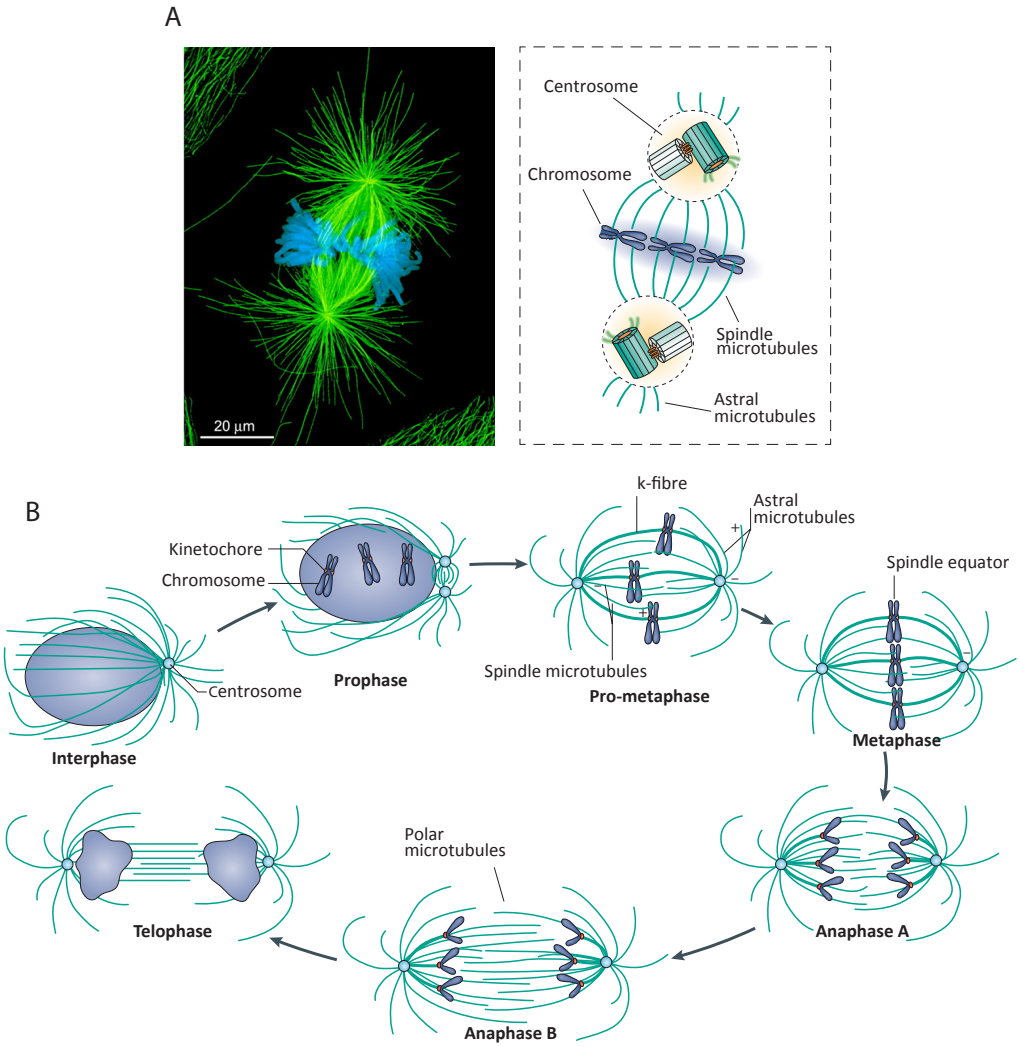


Figure 1.2: **Centrioles as mitotic spindle organisers.** (**A**) Left: newt lung epithelial cell fixed at metaphase - the precise moment when the chromosomes are aligned at the equator of the cell. Microtubules are shown in green and chromosomes in blue (image adapted from [248]). The asters of microtubules radiate from two minute foci - the centrosomes, which nucleate the mitotic spindle. Right: a schematic version of the image on the left, indicating the presence of two centriolar barrels at the foci of microtubule asters (adapted from [26]). (**B**) Mitosis undergoes distinct phases, defined largely by the organisation and position of chromosomes within the cell. During interphase, most of the nuclear chromatin is decondensed and a small number of microtubules radiates from the centrosome. Once the cell enters mitosis, chromosomes begin condensation inside the nucleus, and the centrosomes increase the microtubule nucleation levels. Prophase ends with breakdown of the nuclear envelope, which allows microtubules to penetrate the nuclear membrane and attach to chromosomes. During pro-methaphase, microtubule bundles known as kinetochore fibres (k-fibres) connect kinetochores with spindle poles, initiating the movement of chromosomes towards the spindle equator. The metaphase is defined by the complete alignment of all chromosomes at the equator, after which the cohesion between the sister chromatids is broken and the cell enters anaphase. Daughter chromosomes begin a poleward movement (anaphase A) due to shortening of k-fibres, and the spindle poles separate (anaphase B) as a result of a motor protein push through the overlapping polar microtubules. The nuclear envelope is then rebuilt around decondensing chromosomes (telophase), completing the cycle with cytokinesis which divides the cytoplasm and segregates two newly formed nuclei into two individual daughter cells (Figure adapted from [338]).

involved in performing other tasks. The analysis of eukaryotic phylogeny suggests that this other, basic task of centrioles, is to build cilia and flagella - because all species which grow cilia at any stage of their life cycle, also produce centrioles [51]. Organisms such as lower plants or ferns, which lack centrioles or cilia in adult stages, form both of these structures during spermatogenesis. In fact, centrioles as basal bodies are not only strictly required for nucleation of ciliary microtubules but also define the orientation and positioning of cilia within the cell membrane through distal appendages [220]. Therefore, any evolutionary benefits to cell division which came with the innovation of centrosomes must be linked with the capacity of the eukaryotic precursor organism to grow cilia. Furthermore, the ability of centrosomal centrioles to convert into basal bodies and nucleate cilia (and *vice versa*), suggests that the centrosome evolved by internalisation of basal bodies [15]. The diversity of centriolar morphologies could also possibly advocate the fact that such internalisation did not happen only once in evolution but many times. However, the exact moment or number of such events in different lineages is unknown.

1.1.1 Cell division

Within the centrosome, centrioles provide a structural scaffold for organising the PCM, and the PCM in turn anchors microtubules, either directly or through recruitment of microtubule-nucleating agents, such as γ-tubulin ring complexes (γTuRCs) [190]. During mitosis, in the process called centrosome maturation, the pericentriolar material increases in size through an intense recruitment of γTuRCs (Figure 1.1D on page 3). This leads to a dramatic rise in microtubule nucleation capacity of the centrosome, resulting in the radial expansion of both astral and spindle microtubules required for spindle formation. The majority of spindle microtubules in animal cells are indeed nucleated at centrosomes, which are indispensable for cell division during embryogenesis in many model systems. In numerous species, including the frog *Xenopus laevis*, centrioles are eliminated during oogenesis, rendering sperm a strict donor of centrioles, and ensuring that the fertilised zygotes carry the correct number of centrioles. The fertilisation of a *Xenopus* egg can be mimicked by pricking it with a needle, so that the egg follows a few rounds of cell cycle, but ultimately fails to divide at the end of each round (Figure 1.3 on page 7). In the classical experiment

by Klotz *et al.* [188], who injected centrioles into the egg while it was being pricked, the egg was able to divide and further develop into a morphologically regular tadpole that included only the maternal pool of chromosomes. Thus, it was confirmed that centrioles are essential for cell division during embryonic development in *Xenopus*. Analogous dependence was later found in fertilised embryos of the nematode worm *Caenorhabditis elegans* [179, 201, 250] and the fruit fly *Drosophila melanogaster* [314]. Furthermore, at the turn of the 19th century, Theodor Boveri entertained the thought that abnormal centrosome duplication can cause cancer [40]. In order to prove his theory, he increased the number of centrosomes in the zygote by producing dispermic sea urchin eggs. Presence of extra centrosomes led to assembly of multipolar spindles, which in turn caused unequal chromosome segregation and division into three or more severely aneuploid progeny (Figure 1.4 on page 8). Observing distinctively different developmental characteristics in each progeny, he concluded that chromosomes contribute to those differences, and that the adverse cellular features are a direct reflection of the imbalance in chromosome number [38, 39]. This observation eventually led him to propose that aneuploidy was the driver of tumourigenesis - a rather bold move for that time [40]. Today, it is widely accepted that aneuploidy is a common feature of cancer cells; however, it is still a matter of debate whether aneuploidy is a cause or a consequence of malignant transformation [128, 148]. Nevertheless, the above-mentioned experiments provided strong evidence for the major role of centrosomes in cell division, and particularly in assembling the bipolar mitotic spindle of animal cells.

In contrast to embryogenesis and Boveri's description of the centrosome as 'the organ for cell division', there is some evidence that in a number of somatic cells, centrosomes are dispensable for mitosis. Studies in flies and mammalian cells showed that in the absence of centrosomes, which occurs either naturally (as in higher plants or in early embryonic stages in mice) or is due to experimental manipulation (microsurgery or laser ablation), bipolar spindles can be assembled in the vicinity of chromosomes (Figure 1.5 on page 9; reviewed in [86, 213]). The alternative pathway responsible for such centrosome-independent spindle assembly is organised around protein Ran - a small guanosine triphosphatase that recruits a number of molecular motors and microtubule-bundling factors [168]. Although the centrosome-less mu-

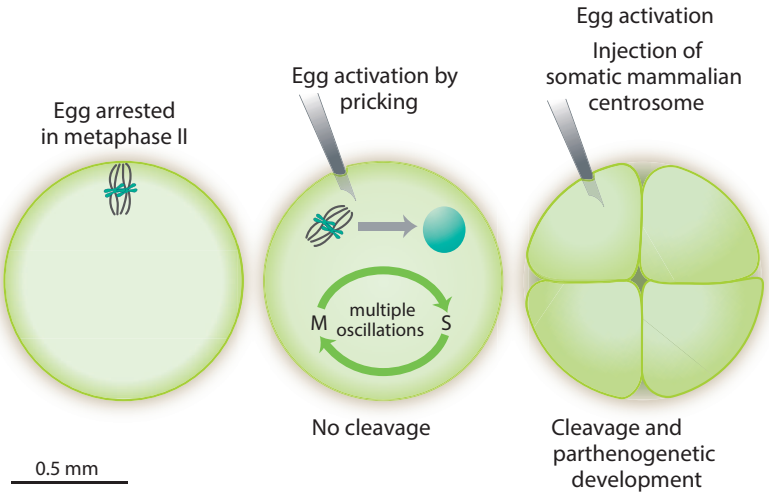


Figure 1.3: ***Xenopus* egg experiment.** Metaphase II-arrested oocytes can be activated by pricking, which triggers the cell to undergo multiple cell cycles. Ultimately, however, the oocyte fails to divide in the absence of a sperm cell. Only after the egg is injected with an isolated mammalian centrosome, it can divide and develop into a regular tadpole. Figure adapted from [35].

tant flies can develop into morphologically normal adults (Figure 1.5A on page 9), they do require centrosomes during the early embryonic stages that are characterised by both a very short cell cycle and a large proportion of asymmetric cell divisions, leading to two daughter cells of two different fates. After centrosome depletion, the spindle assembly process is severely slowed down, and despite the fact that these centrosome-less mutants take much longer to reach adulthood, they die earlier than the wild-type flies [21]. These findings strengthen the importance of centrosomes in embryogenesis, which not only requires swift and accurate chromosome segregation but also the ability to split cells asymmetrically with high precision. Furthermore, the shorter lifespan of centriole-less *Drosophila* mutants likely reflects the inability of these mutants to assemble cilia, which are essential for sensing by many neurons.

1.1.2 Basal bodies, cilia and flagella

Nearly every cell type in the human body contains thin projections stemming from its surface. These projections are known as primary cilia and can provide the cell with a several micrometre-long extension which reads its extracellular surroundings. Specialised versions of such projections - motile cilia and flagella, endow the cell with

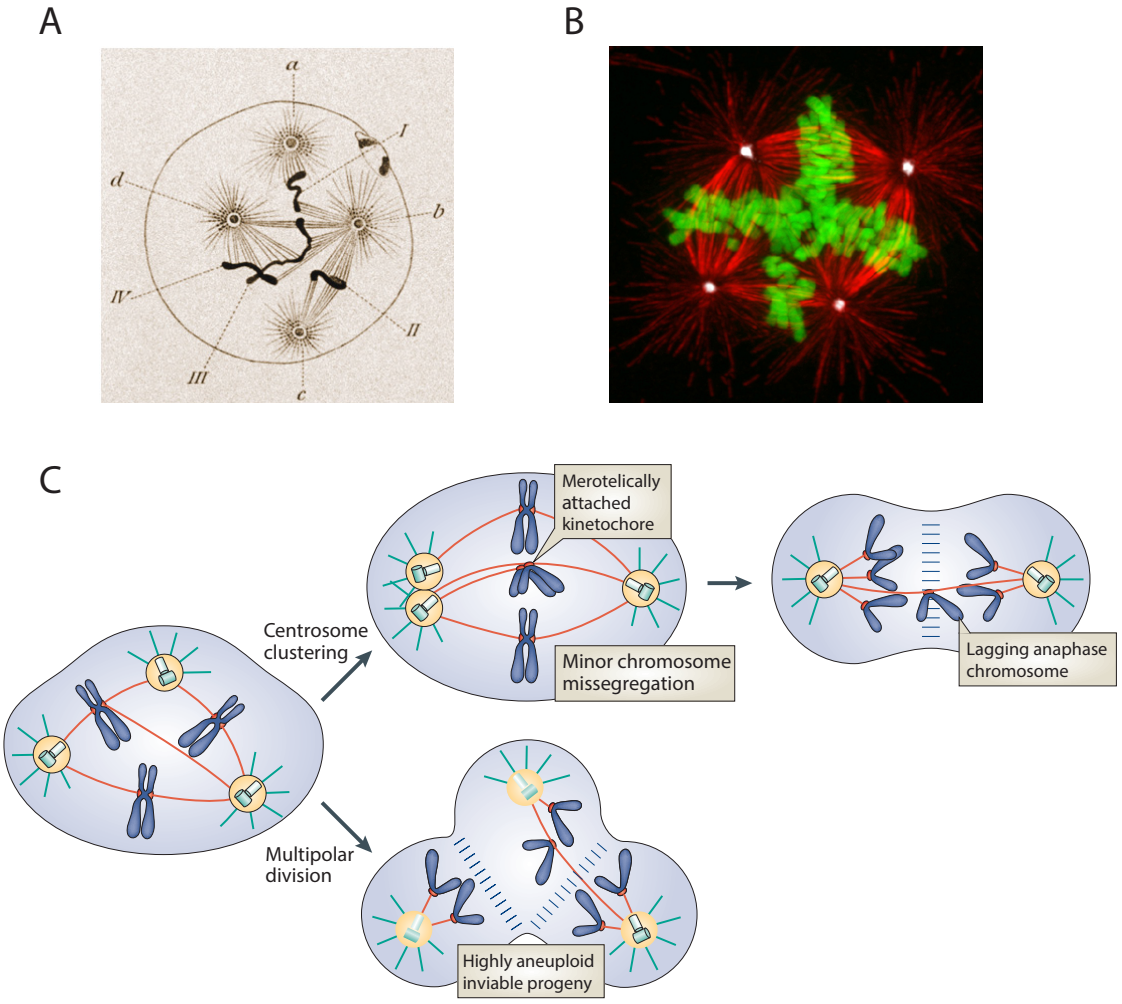


Figure 1.4: **Multipolar spindle formation.** (**A**) Theodor Boveri identified two fundamental events of cell division: centrosome duplication and chromosome duplication [39]. He also noticed that errors in cell division in sea urchin embryos lead to aberrant development and might cause cancer. In his drawing, an abnormal number of centrosomes (*a*, *b*, *c* and *d*) in sea urchin embryos causes unequal distribution of chromosomes into daughter cells, *i.e.* aneuploidy. (**B**) Fluorescence micrograph showing four centrosomes (white) which nucleate a tetrapolar spindle (microtubules in red) and pull the chromosomes (green) in four different directions. Reproduced with permission from Dr Neil Ganem, Boston University. (**C**) Multipolar mitotic spindles can lead to multipolar divisions, which produce highly aneuploid and, most of the time, inviable daughter cells. Frequently, however, multiple centrosomes cluster together to form two isolated groups, consequently organising a pseudo-bipolar spindle instead. Although such clustering allows the cells with supernumerary centrosomes to survive, it also increases chances of incorrect microtubule attachments to kinetochores, and therefore promotes chromosome missegregation. The so-called merotelic attachment occurs when two microtubules from the opposite poles attach to a single kinetochore. Pulling forces applied from both spindle poles will cause such chromatid to lag and likely missegregate or be lost altogether if trapped within the cleavage furrow. Figure adapted from [148].

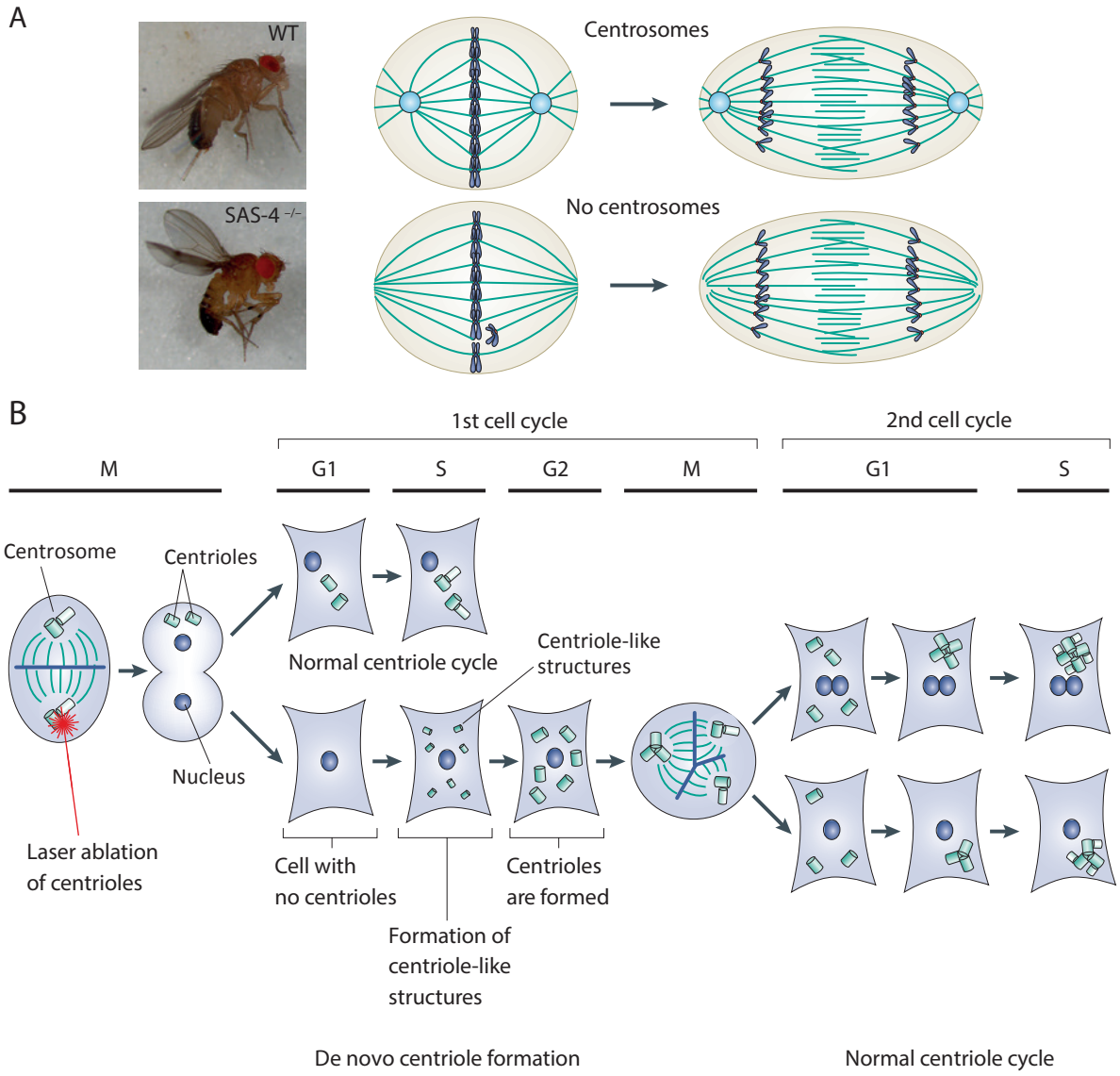


Figure 1.5: **Are centrioles indispensable?** (**A**) Although *Drosophila melanogaster* flies deficient in one of the core centriolar proteins, SAS-4 (see Section 1.2.3 *The sequential order of events* on page 14) lack centrosomes and cilia, they develop into a morphologically normal adult organism. The lack of cilia, however, completely eliminates their mechano- and chemosensory mechanisms, which results in the held-up wings and unnatural posture. Even though the assembly of the mitotic spindle is largely slowed down (indicated by the lagging chromosome) due to absence of centrosomes, most of the mutant cells are able to segregate their chromosomes normally. (**B**) In cases where centrioles in one of the mitotic spindle poles were removed experimentally with a laser, such dividing cell often produces one daughter cell with a normal number of centrioles and the other daughter cell with no centrioles. While the cell with two centrioles continues to divide normally, the cell with no centrioles begins to assemble multiple centriole-like structures (small green barrels) upon entering the S-phase. By the time the cell is about to enter mitosis, the centriole-like structures have developed into morphologically complete centrioles (larger green barrels). Such *de novo*-formed centrioles often assemble multipolar spindles, but do not mature centrosomes until the G1 phase of the following cell cycle, after which they resume cycles with normal centriolar duplication. Figures adapted from [26, 63, 195].

motile properties. Cilia are most common in specialised, non-dividing cell types, that do not assemble centrosomes. Centrioles of such cells can thus travel to the cell surface, and once attached to the cell membrane, they undergo structural rearrangements to form a basal body which seeds the growth of a cilium or a flagellum (Figure 1.1G on page 3; reviewed in [84, 219]). The function of cilia and flagella in different systems is to propel unicellular and small colonial organisms across fluids, and also to pass fluids over epithelial tissues, such as pulsative circulation of cerebrospinal fluid in brain and spinal chord, or debris clearance in the respiratory tract (Figure 1.6 on page 11). Moreover, the motile cilia of the embryonic node can generate a fluid flow necessary for the establishment of embryonic left-right asymmetry [245]. The difference between motile cilium and flagellum is more subtle than their number per cell. While it is true that there can be hundreds of cilia per cell and only up to a handful of flagella, as well as that they are structurally similar, they differ significantly in their length and beating patterns. Flagella are significantly longer than cilia, and exhibit an independent, whip-like, undulating, low-frequency, asymmetric stroking motion (Figure 1.6D on page 11). Cilia, on the other hand, perform a complex, coordinated sequence of high-frequency symmetrical beats, either synchronously or in a metachronal rhythm [197] (Figure 1.6E, F on page 11). During the photophobic response, a unicellular algae *Chlamydomonas reinhardtii*, can switch between the beating patterns of its two flagella: to either a ciliary (forward swimming) or flagellar (backward swimming) motion, effectively repositioning itself in order to escape an abrupt increase in light intensity [207, 281].

Primary cilia, which are not directly involved in motility, act as signal-receiving antennae. Environmental cues obtained through such surveillance can be processed by various cilia-specific receptors and channels sitting in the ciliary membrane. Such information is then converted into signalling cascades, which can reach the cell body and create a response. One of the prime examples of mechanosensation is the ability of passive cilia to perceive bending induced by the extracellular fluid flow [269]. Moreover, cilia mediate a number of signalling pathways, including Sonic Hedgehog, platelet-derived growth factor receptor or non-canonical Wnt, as well as the mating pathway in *C. reinhardtii* [108, 235].

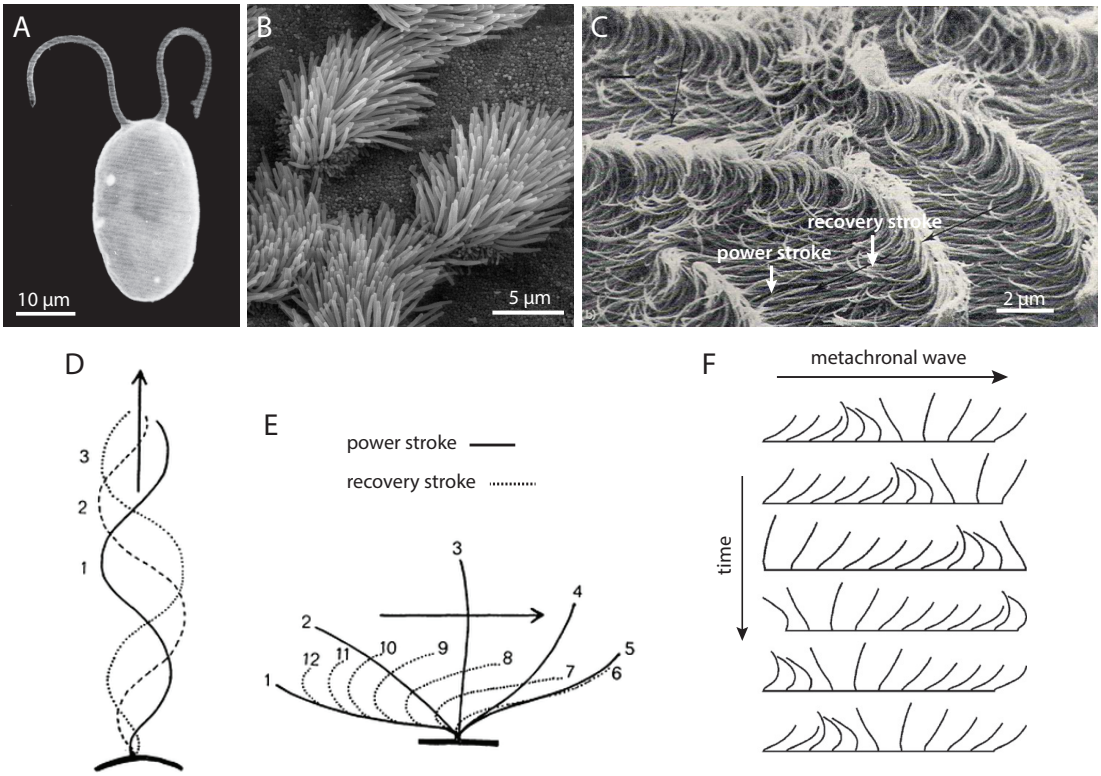


Figure 1.6: **Beating patterns of cilia and flagella**. Scanning electron micrographs of two flagella from green algae *Chlamydomonas* (**A**), cilia on the surface of mammalian lung trachea epithelium (**B** - note that lung epithelium has both ciliated and non-ciliated cells, image credit: Charles Daghlian, Dartmouth College), and cilia on the surface of a protozoan *Opalina*, beating in the highly coordinated metachronal rhythm in which cilia in a given row are at the same stage of the beat cycle (**C**). Schematic drawing of the waveforms created by a beating flagellum (**D**) and cilium (**E**). Successive stages of the beat cycle are numbered consecutively. The power and recovery strokes of ciliary beating are drawn with solid and dotted lines, respectively. Arrows indicate the effective direction of displacement of the medium. Progression of the metachronal wave is shown in (**F**), with successive stages of the cycle drawn in order, from top to bottom. Images reproduced from [81, 289, 321, 345].

1.2 Centriole biogenesis

The continuity of successive cell divisions relies on synchronously creating an extra copy of a centrosome, thus preparing for chromosome segregation in the upcoming cell cycle. Therefore, depending on the cell cycle stage, cells contain either one centrosome (two centrioles) in G1 or two centrosomes in G2 (two pairs of centrioles). A newly born cell inherits two centrioles of different ages: an immature 'daughter' centriole that was assembled in the previous cell cycle, and a mature 'mother' centriole that

was assembled at least two cell cycles ago. Starting in S-phase, each centriole (both mother and daughter) assembles a structure called procentriole that matures into a daughter centriole throughout the remainder of the cell cycle (Figure 1.1 on page 3; reviewed in [63, 242]). Even though the majority of proliferating cells tightly control their centriole numbers, there are interesting exceptions.

1.2.1 Centriole multiplication

In order to supply epithelia with hundreds of basal bodies required for cilia formation, centrioles emerge nearly simultaneously from large fibrous aggregates known as deuterosomes. Until recently, it has been assumed that the mass production of centrioles in multiciliated cells was mediated by two independent pathways: the canonical centriolar pathway, where several procentrioles form in association with the mother centriole, and the *de novo* acentriolar, deuterosome-based pathway (Figure 1.7A on page 13). Al Jord *et al.* [5] have revised this concept by showing that in fact there is only a single pathway, orchestrated by the centrosome. They revealed that the centrosomal daughter centriole in multiciliated cells supports multiple rounds of procentriole seeding by generating 'halos', *i.e.* procentrioles organised around spherical deuterosomes (Figure 1.7B on page 13). Following their release into the cytoplasm and elongation of their procentrioles, halos develop into flower-like shapes. Synchronised procentriole maturation during the flower stage proceeds with a simultaneous release of those newly matured centrioles from deuterosomes and their immediate docking at the cell membrane, thus seeding growth of hundreds of cilia. Effectively, almost all of the basal bodies in multiciliated cells originate from the daughter centriole, while the mother centriole contributes only a small fraction.

Another interesting exception to the canonical centriolar pathway, where procentrioles are generated adjacent to pre-existing mature centrioles, is the ability of cycling cells to assemble centrioles *de novo* after the resident centrioles have been eliminated experimentally (Figure 1.5B on page 9; reviewed in [26, 213]).

1.2.2 Assembly from unique components

Even though proteomic studies revealed hundreds of proteins involved in human centrosomes [7, 157], it was the combined efforts of genetics and RNAi-based methods

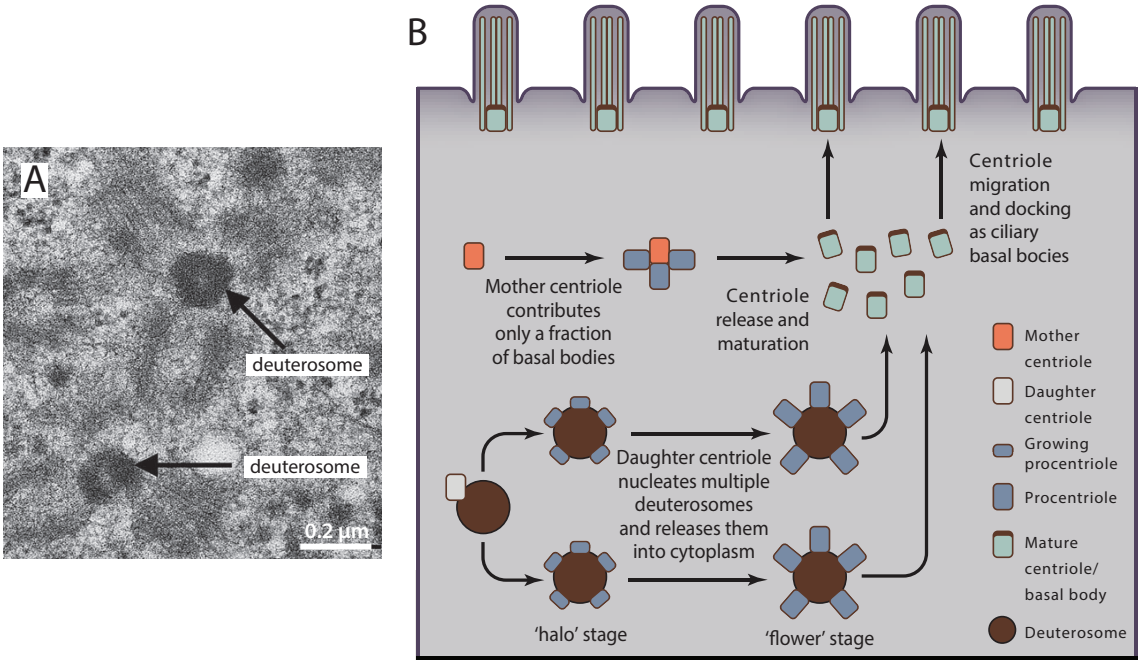


Figure 1.7: **Multiplication of centrioles.** (**A**) Electron micrograph of a deuterosome nucleating basal bodies in mouse tracheal epithelium cells. Although a few components of deuterosomes have been indentified [187, 353], our understanding of these mysterious structures is still rather vague. (**B**) Deuterosomes are central to biogenesis of a large number of basal bodies in multiciliated cells. Current model postulates that the centrosome itself (both mother and daughter centrioles) contribute only a small fraction of basal bodies via direct assembly, while the bulk of basal bodies is nucleated indirectly via deuterosomes. Recent evidence shows that the daughter centriole nucleates multiple deuterosomes, and then releases them into cytoplasm. Deuterosomes in turn, organise assembly of a large number of procentrioles that mature from the 'halo' stage to the 'flower' stage, and eventually migrate to cell membrane to become basal bodies and seed growth of multiple cilia. Figures adapted from [5, 344, 355].

that identified five unique gene products as essential for centriole assembly in early embryos of *C. elegans* [78, 88, 172, 179, 200, 201, 249, 263, 296]. These five key proteins turned out to be evolutionarily conserved, and even though more proteins from other species have since been implicated in centriole formation, this basic set of five proteins constitutes the core machinery which drives centriole biogenesis. Such a small number of proteins building enormous, on the molecular level, structures is characteristic of symmetrical assemblies. While the largest known asymmetrical macromolecule, the mammalian ribosome, can reach up to 30 nm in diameter [174], symmetrical structures, such as viral capsids or clathrin cages can extend to as much as 200 nm [75]. The common feature between those molecular giants is a small number of unique

components. The centriole is no different in this respect.

Two pioneering studies by Pelletier *et al.* [262] and Delattre *et al.* [87] shed light on the detailed course of events at the early stages of procentriole formation in *C. elegans*. Using antibodies and green fluorescent protein (GFP) fusions as markers for cellular localisation of the 'core five' proteins, they were able to determine the spatial and temporal distribution of these proteins throughout the cell cycle. Furthermore, by taking advantage of the strict donation of centrioles to the zygote by the sperm in embryos of *C. elegans*, they were able to learn the order of recruitment of these proteins to the emerging procentiole. Müller-Reichert and colleagues used a time-resolved fixation method to capture *C. elegans* embryos at various developmental stages [232]. Together with Leica, they adapted a tissue preservation technique of high-pressure freezing, followed by a dehydration step called freeze-substitution for ultrastructural imaging of whole worms. These procedures provided them with the necessary details to observe the effect of RNAi knock-down of individual genes on the morphology of small *C. elegans* centrioles (Figure 1.8A on page 15).

1.2.3 The sequential order of events

It was revealed that proteins ZYG-1 (zygote defective protein 1), SPD-2 (spindle-defective protein 2), SAS-4, SAS-5 and SAS-6 (spindle assembly abnormal 4, 5 and 6) act in a precisely orchestrated sequence of events to build a new centriole (Figure 1.8 on page 15) [87, 262]. As soon as the egg has been fertilised by the sperm, SPD-2 is recruited to paternal centrioles. In turn, this action triggers centriolar recruitment of ZYG-1, a protein kinase, required for bringing a complex between SAS-6 and SAS-5 to the procentriole. Finally, the SAS-6/SAS-5 complex guarantees recruitment of SAS-4, and subsequent addition of microtubules. Moreover, orthologues of the 'core five' proteins have also been identified in other systems, indicating that this cascade of events is not unique to nematodes. However, as usual in biology, there is a notable diversity in protein sequence between these orthologues, and there are many and varied requirements for the involvement of additional proteins in the process of centriole assembly, which reflects the extent of potential evolutionary adaptations across various species.

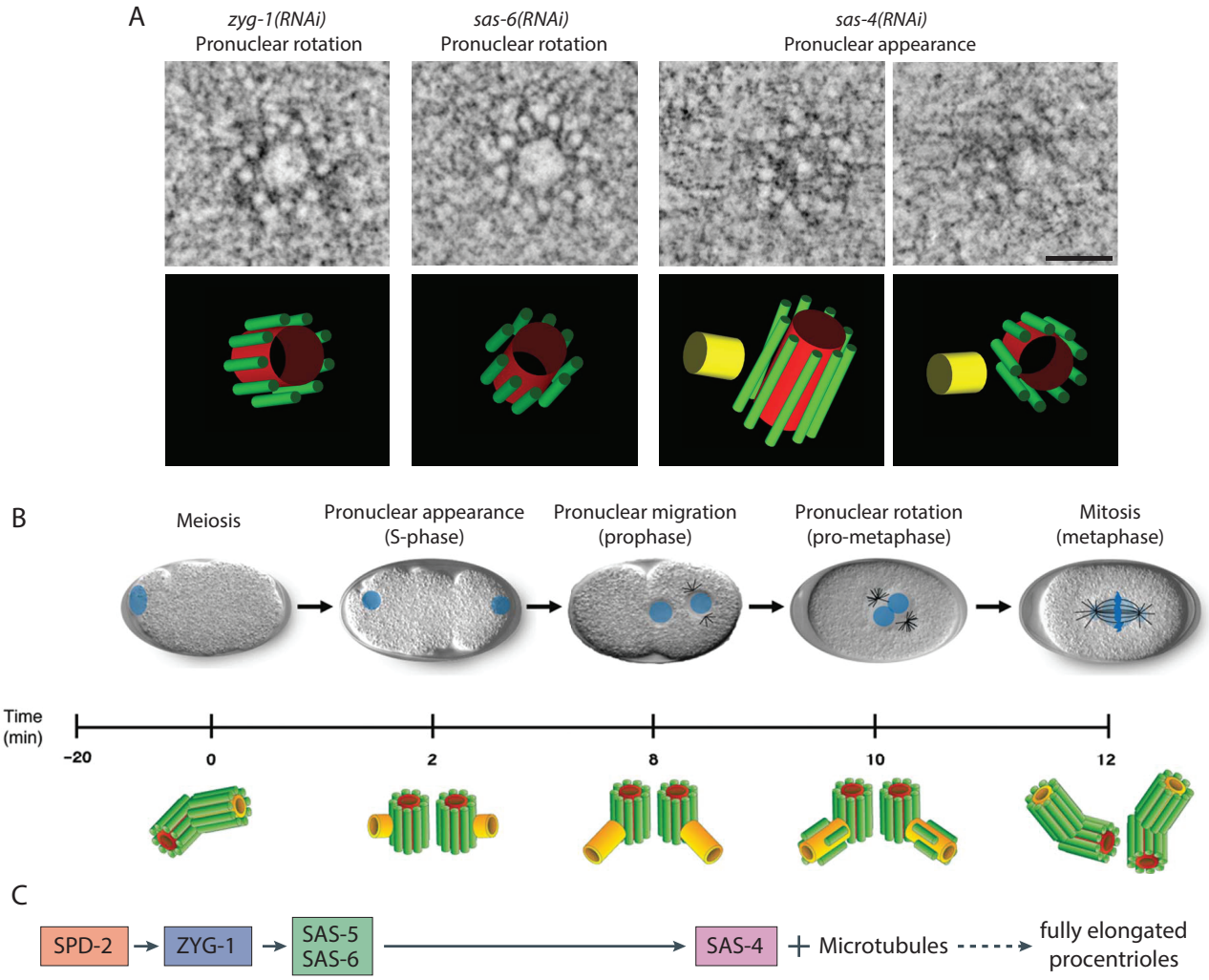


Figure 1.8: **Centriole assembly in *C. elegans*.** Pelletier *et al.* [262], using a time-resolved fixation method coupled with high-pressure freezing, were able to preserve the structure of worm early embryos at different stages of development. Working with wild-type embryos, they imaged centriole duplication process throughout mitosis. Then, using RNAi targeted against genes known to be important in centriole assembly pathway (**A**), they monitored defects in duplication of centrioles, ultimately revealing the order of protein recruitment early in the pathway (see Section 1.2.3 *The sequential order of events* on page 14). While *zyg-1(RNAi)* and *sas-6(RNAi)* embryos failed to assemble daughter centrioles altogether, *sas-4(RNAi)* successfully assembled a structure called the central tube, but eventually failed to attach microtubules. (**A**) Electron micrographs of RNAi-treated worm embryos, and 3D models of centrioles observed. Mother centriole - red, central tube of a growing procentriole - yellow, and microtubules - green. (**B**) Timeline of mitotic events. Top panel: centrioles of worm embryos at different mitotic stages are indicated in blue. Bottom panels show 3D models based on the images of *C. elegans* centrioles acquired with cryo-electron tomography. (**C**) The recruitment order of proteins involved in progression through the centriole duplication cycle. Following centriole disengagement, SPD-2, a PCM component of mother and daughter centrioles, recruits ZYG-1, which in turn recruits the SAS-5/SAS-6 complex. These two proteins (SAS-5 and SAS-6) build the central tube, and later recruit SAS-4 which ensures attachment of microtubules. Figures adapted from [63, 262].

While the vertebrate orthologue of SPD-2, known as Cep192 (centrosomal protein of 192 kDa; Figure 1.9 on page 18) [7], had been shown to be involved in many steps of centriole biogenesis, including duplication, growth, PCM recruitment and microtubule nucleation [122,356], in flies, SPD-2 appears to be completely dispensable for centriole assembly [93,117]. Nevertheless, *Drosophila* SPD-2 plays an essential role in pericentriolar material recruitment and centriole maturation [62]. It turns out that the seeding component for procentriole formation in *Drosophila* is a protein called Asterless (Asl; vertebrate orthologue - Cep152 [7]; Figure 1.9A on page 18) [31, 99, 335], which recruits downstream proteins in the assembly pathway. Interestingly, SPD-2 appears to be absent from the genome of planarians, which strictly assemble centrioles *de novo*, suggesting that SPD-2 may only be required in the instance when a procentriole is to be seeded next to a pre-existing mature centriole [18]. The work on human cells have shown that Cep192 and Cep152 are indeed integral components of the PCM, where they localise to a ring-like structure, and are required to seed growth of human centrioles [60, 99, 139, 177, 311].

1.2.4 Assembly of the pericentriolar material

The control mechanisms ensure that seeding of procentrioles happens only once per cell cycle and at exactly one site per mature centriole [329, 346], but what defines the exact location for seeding growth of a new centriole? Although a definite answer is not yet available, four independent studies used subdiffraction light microscopy techniques [114, 199, 228, 312] to reveal the layered structure of the pericentriolar material, which paved the way to understanding how this seemingly amorphous protein cloud can seed the growth of a new centriole. Soon after a procentriole disengages from its mother and turns into a daughter centriole (M to G1 transition), it becomes encapsulated by the first structured layer of PCM, $\sim$100 nm thick, that includes a toroid of Cep192 (Figure 1.9A, B on page 18). With the advancement of the cell cycle towards mitosis, Polo-like kinase 4 (PLK4) [27, 134, 186], a relative of ZYG-1 in *D. melanogaster* and vertebrates, specifically phosphorylates many cytosolic PCM components. Such phosphorylation acts as an on-switch to drive deposition of these cytosolic PCM proteins over the first PCM layer, thus increasing the radius and complexity of the PCM. A number of proteins, including Cep152 (Figure 1.9 on page 18),

cross-interact to build a lattice-like organised network that acts as a structural foundation for the recruitment of microtubule-nucleating complexes at the onset of mitosis [347]. In G1 and G2, these PCM layers on mature centrioles undergo a noticeable structural rearrangement - they form an exclusion zone that will house an emerging procentriole (Figure 1.10 on page 19 and Figure 1.1B on page 3).

PLK4 is a binding partner of both Cep192 and Cep152, and similarly to those proteins, PLK4 forms a toroid structure around the proximal part of mother and daughter centrioles in the early G1 phase (Figure 1.10, first two rows, page 19) [177]. Such PLK4 toroid (∼450 nm in diameter) is formed immediately after Cep192 is recruited to establish the first PCM layer of a maturing daughter centriole [256]. Subsequent addition of the second structured layer of PCM components (late G1), that include Cep152, triggers the expansion of the PLK4 ring to ∼600 nm in diameter. Upon G1-S transition, PLK4 distribution shrinks to a discreet dot-like structure at the surface of mother and daughter centrioles, setting the location of the forthcoming centriole duplication site (Figure 1.9C on page 18 and Figure 1.10C, D on page 19).

These observations suggest massive rearrangements in PCM of the mother and daughter centrioles, where after the site for procentriole assembly is selected, the exclusion zone forms by localised disassembly of PCM. The striking change of PLK4 localisation from the ring-like structure to a discreet dot suggests that a localised accumulation of all of the PLK4 is needed to precisely trigger the centriole assembly pathway. Given that PLK4 is a kinase, a simple model would suggest that phosphorylation of dot-co-localised centriolar proteins serves as an on-switch for centriole formation at the right time and place. However, the identities of both the timekeeper and the guide factor for PLK4 remain elusive [115]. Recently, a number of reports have shed light on the molecular details behind the activity of PLK4 in variaous systems, and included *i.a.* crystal structures of PLK4 bound to acidic peptides of SPD-2/Cep192 and Asl/Cep152 [256, 301, 304].

1.2.5 Procentriole growth

Centriole duplication begins at the transition between G1 and S of the cell cycle. Once the site for growth of a procentriole is chosen, displayed as the dot-like localisation of PLK4 (Figure 1.9C on page 18 and Figure 1.10 on page 19), the building of a

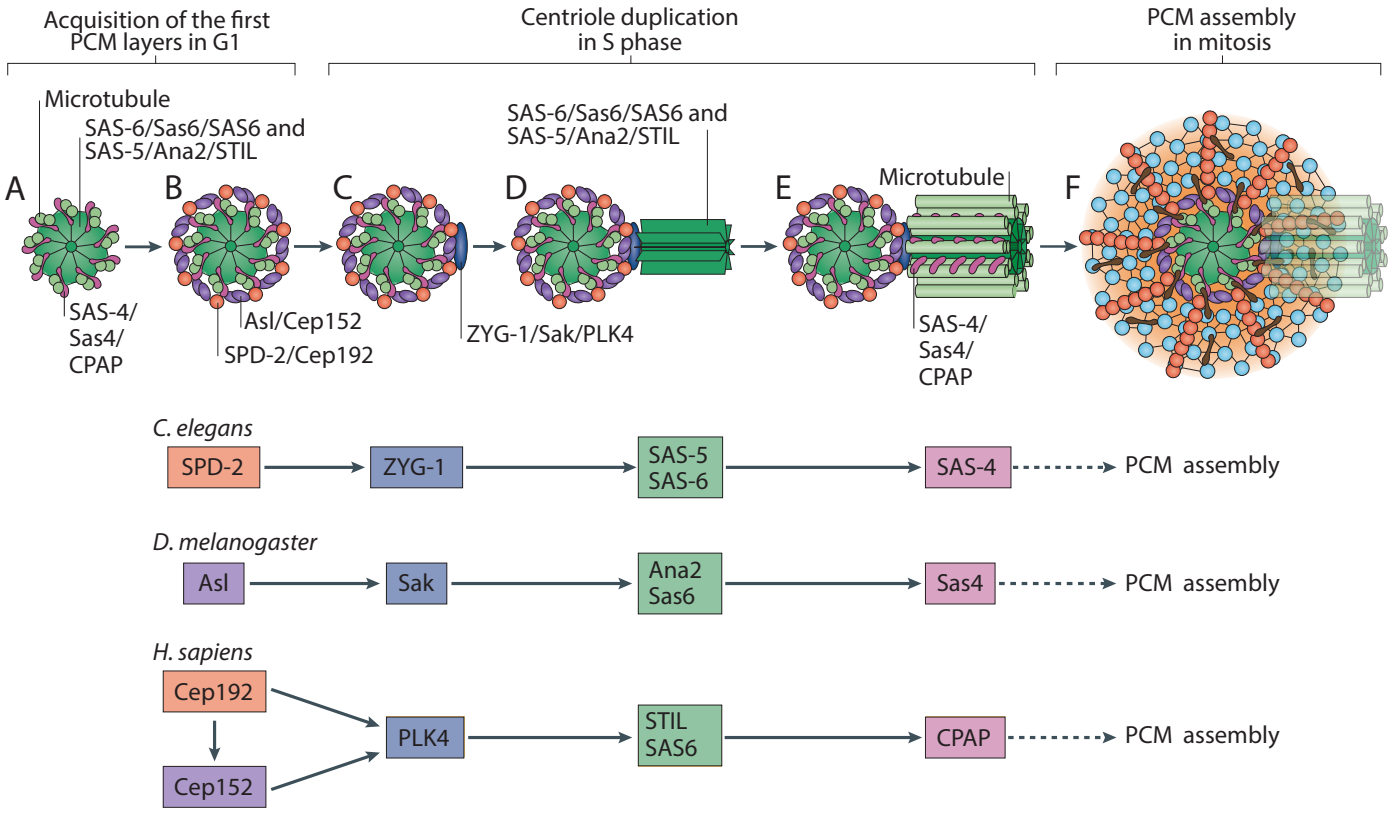


Figure 1.9: **Centriole assembly in other organisms.** Top row: schematic drawings of the maturation process of a daughter centriole which had just disengaged from its mother, and is now able to assemble a procentriole of its own. Diagrams below show our current understanding of the molecular pathway behind the process of centriole assembly in various model organisms: worms, flies and humans. Protein homologues and putative homologues share the same colouring. Names of some protein homologues, although similar, have not yet been unified to a single notation common for all species, and hence the erratic SAS-6, Sas6, SAS6, *etc.* Arrows indicate the sequence of protein recruitment to centrioles. (**A**, **B**) A newly nucleated daughter centriole after the exit from mitosis develops into a mother itself. First, it becomes surrounded by a few structured layers of PCM, components of which include *C. elegans* SPD-2, *D. melanogaster* Asl and a combination of their homologues, Cep192 and Cep152, in *H. sapiens*. (**C**) Centriole duplication in *C. elegans* begins with recruitment of protein kinase ZYG-1 (equivalent of Sak in *D. melanogaster* and PLK4 in *H. sapiens*) to a single site on the mother centriole by SPD-2/Asl/Cep192/Cep152 proteins. Note that the process of ZYG-1/Sak/PLK4 localisation to a single site may be more complex (see Section 1.2.4 *Assembly of the pericentriolar material* on page 16 and Figure 1.10 on page 19). (**D**, **E**) The protein kinase (ZYG-1, Sak or PLK4) then recruits SAS-5 (in *C. elegans*; Ana2 in *D. melanogaster* and STIL in *H. sapiens*) and SAS-6, either simultaneously as a complex or in a sequence. This results in formation of a cartwheel - either as a stack (shown here) or as a central tube/spiral in worms (see Section 1.3.3 *The unique cartwheels of C. elegans* on page 30). At this point, SAS-4 (in *C. elegans*; Sas4 in *D. melanogaster* and CPAP in *H. sapiens*) is recruited to centrioles and begins attachment of microtubules. (**F**) SAS-4 proteins are also involved in recruitment of mitotic PCM, which itself is made of a number of protein layers - reviewed in [347]. Figure adapted from [63].

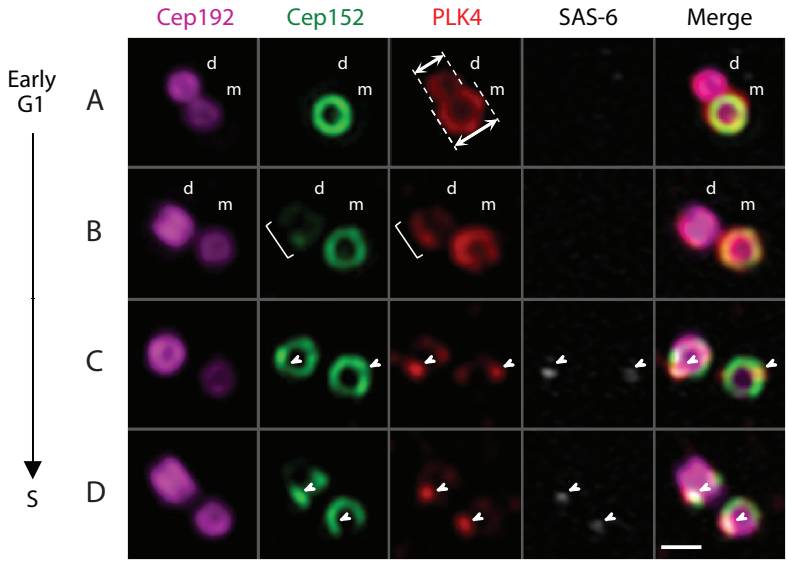


Figure 1.10: **Dot-like localisation of PLK4.** PLK4 switches its localisation pattern from a toroid to a focused dot. Super-resolution (3D structured-illumination microscopy) fluorescent micrographs of human cells, between early G1 and S phases, illustrate centriolar localisation of Cep192 (magenta), Cep152 (green), PLK4 (red) and SAS-6 (grey). (**A**) Cep192 forms the first PCM layer around the newly nucleated daughter centriole (d), which after disengagement is now also developing into a mother itself. Following Cep192 localisation, PLK4 (a binding partner of both Cep192 and Cep152) is recruited to centrioles. Arrows in the first row mark diameters of two PLK4 rings localised around the mother centriole (m) and the daughter centriole (d). (**B**) At this point, the second layer of PCM is deposited, which includes Cep152, and triggers expansion of the PLK4 toroid. Brackets in the second row indicate the position of the PLK4 toroid, co-localised with the toroid of Cep152. (**C**, **D**) Upon transition from G1 to S, PLK4 focuses its localisation to a dot-like structure. Arrowheads in rows C and D point to the dot-like signals of PLK4 at procentrioles, co-localised with SAS-6, and now overlapping only with a small part of the Cep152 toroid. Scale bar, 0.5 μm. Figure adapted from [256].

new centriole begins. In human cells, centriole biogenesis starts by localisation of the first structural proteins to the PLK4-dot, including SAS-6 [200] (Figure 1.10C, D on page 19), STIL (SCL/TAL1 interrupting locus; relative of *C. elegans* SAS-5 and *D. melanogaster* Ana2 - anastral spindle 2) [14, 182, 313, 323, 337], CPAP (centrosomal P4.1-associated protein; orthologue of *C. elegans* SAS-4) [154], Cep135 (relative of Bld10p in *C. reinhardtii* [143, 221]; no known relative in *C. elegans*) and γ-tubulin [186, 246] (Figure 1.9D, E on page 18). Cross-interaction of these proteins leads to assembly of a procentriolar seed in the early S-phase (Figure 1.11A on page 21), which then recruits a capping protein CP110 (centriolar coiled-coil protein of 110 kDa) to the distal tip of the procentriole (Figure 1.11B on page 21) [186] (Note: there are

three distinct regions along the longitudinal axis of a centriole: distal, central and proximal - see Section 1.3 *Centriole architecture* on page 20 for details). Such seed then elongates from its distal end to form a procentriole by inserting α-/β-tubulin dimers underneath the cap formed by CP110 (S and G2 phases). What factors define the final length of the centriole is still not completely understood. The centriole length can vary significantly between species, and even between cell types within the same species. The emerging evidence points to the antagonistic relationship between CPAP and CP110 as guardians of this process (Figure 1.11E on page 21). While CP110 limits the extension of centriolar microtubules by capping them at the plus end [110], overexpressed CPAP produces overly elongated centrioles [189, 295, 322].

In summary, the 'core five' proteins, initially discovered in *C. elegans*, constitute the basic set required to seed growth of procentrioles, next to mother or daughter centrioles. These proteins are conserved to a large extent across the evolution; however, in *D. melanogaster* and human cells, some of the functions of the individual 'core five' proteins are shared by more than one protein.

1.3 Centriole architecture

A typical human centriole is a nine-fold symmetrical cylinder, 250 nm in diameter and 450 nm in length [124]. To put this into perspective, approximately 1000 ribosome particles could fit into one centriole [344]. Mature human centrioles can be dissected into three distinct parts along the longitudinal axis: distal - with microtubule doublets, central - with microtubule triplets, and proximal - with microtubule triplets and the cartwheel (Figure 1.12 on page 23). The cartwheel is a crucial structure that serves a symmetry-templating role in the centriole assembly process. In the transverse plane of the proximal region, cartwheels contain an inner ring, termed central hub (Figure 1.12D, F on page 23), from which a number of evenly distributed spokes emanates towards the periphery of the centriole. At the peripheral end of each spoke sits a structure called pinhead, which directly connects the spoke to yet another ring, known as the microtubule wall. At the proximal part of the centriole, this outer ring comprises an evenly distributed set of microtubule triplets (A-, B- and C-microtubules, also known as 'blades'). Effectively, a microtubule web is cre-

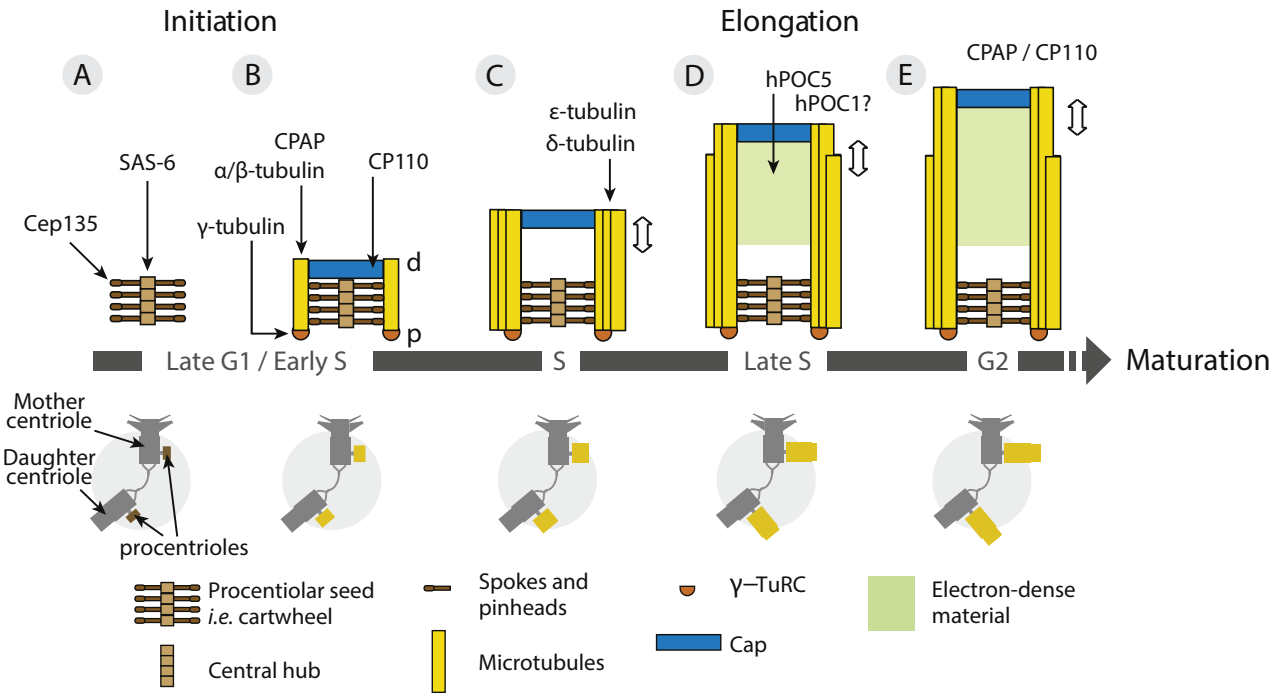


Figure 1.11: **Model of procentriole assembly and growth.** (**A**) Assembly of a procentriole begins at the transition between phases G1 and S, with the construction of a centriolar seed - the cartwheel (see Section 1.3 *Centriole architecture* on page 20), which itself is made of SAS-6 (central hub and spokes) and Cep135 (equivalent of Bld10p in *Chlamydomonas*; pinheads). (**B**) Recruitment of CPAP triggers nucleation of microtubules and their attachement to the pinhead. The exact mechanism of this is unknown, but it is possible that CPAP recruits γ-TuRC complexes to the proximal end of the cartwheel, which then nucleate unidirectional growth of microtubules in the proximal (p) to distal (d) direction. Protein CP110 localises to the distal end of the newly emerged procentriole, and forms a cap. The likely role of the cap is to stabilise the procentriole and control its growth. (**C**) Centrioles of worms only attach one tubule per pinhead, the A-tubule; however, flies and humans add an extra one tubule (B-tubule) or extra two tubules (B- and C-tubules), respectively (see Section 1.3.5 *Doublets and triplets of microtubules and auxiliary proteins* on page 36 for details). These extra tubules extend bidirectionally - until they reach the extent of the A-tubule, and are further stabilised by ϵ- and δ-tubulin. Over the duration of the S phase, procentrioles elongate up to $\sim$70 % of their final length. Centriole elongation is not very well studied, but some reports point to hPOC5 (human proteome of centriole 5) and hPOC1 proteins (**D**), as regulators of this process [16,171]. (**E**) The final extension of procentriole takes place in G2, where it is thought that the fine balance between CPAP (promotes growth by recruiting α- and β-tubulin to the distal end) and CP110 (inhibits growth by forming a cap) determines the ultimate procentriole length. Fully-elongated procentrioles mature throughout the M phase, and also until the early stages of the G1 phase of the next cycle, where they acquire distal appendages and organise PCM. Schematics below the cell cycle timeline illustrate the increasing size of procentrioles relative to the size of mother and daughter centrioles. Figure adapted from [17].

ated where triplets connect sideways through the structure called the A-C linker, and across through the pinhead and cartwheel spokes.

Small variations in individual components between species lead to noticeable differences in centriole morphology [15, 51]. Notable examples include the size and composition of microtubule blades, which vary between organisms and comprise either singlets in *C. elegans*, doublets in *D. melanogaster* or triplet microtubles in vertebrates. Although it is not exactly clear what factors facilitate assembly of microtubules into triplets or doublets, some evidence suggests that the tektin family of proteins could be involved in the process [6]. Moreover, the size of centrioles can range from 100 nm to 300 nm in diameter, and 100 nm to 5 μm in length. Furthermore, the architecture of the cartwheel itself has been shown to adopt two distinct conformations: a stack or a helix (see Section 1.3.3 *The unique cartwheels of C. elegans* on page 30).

1.3.1 The origin of centrioles

With a large number of proteins involved in their elegant symmetrical arrangement, centrioles appear extremely complex and irreducible. It is therefore difficult to imagine how something as intricate as centrioles could have possibly evolved, given that the fitness benefit to the cell would only apply once all of the components were in place and working together. If one component was removed, the structure would fall apart. Such thinking, however, in application to evolutionary history may be rather deceptive, as simple self-organising processes in chemistry and physics can indeed create structures of exceptional complexity. Regardless of the apparent complexity, the evolvability of centrioles depends solely on how complex the genetic instructions for assembling the structure are - exactly what computer scientists call algorithmic complexity [220].

To understand the algorithmic complexity of centrioles, we must find out what is the minimal set of genes that could build a useful centriole. 'Useful' in this case means that such structure will be of some fitness benefit to the cell. The main task of ancient centrioles was to template cilia, which requires a pre-assembled set of microtubules that can attach to cell membrane and then grow out as an antennae. It turns out that the canonical nine-fold symmetry of microtubule triplets in centrioles or

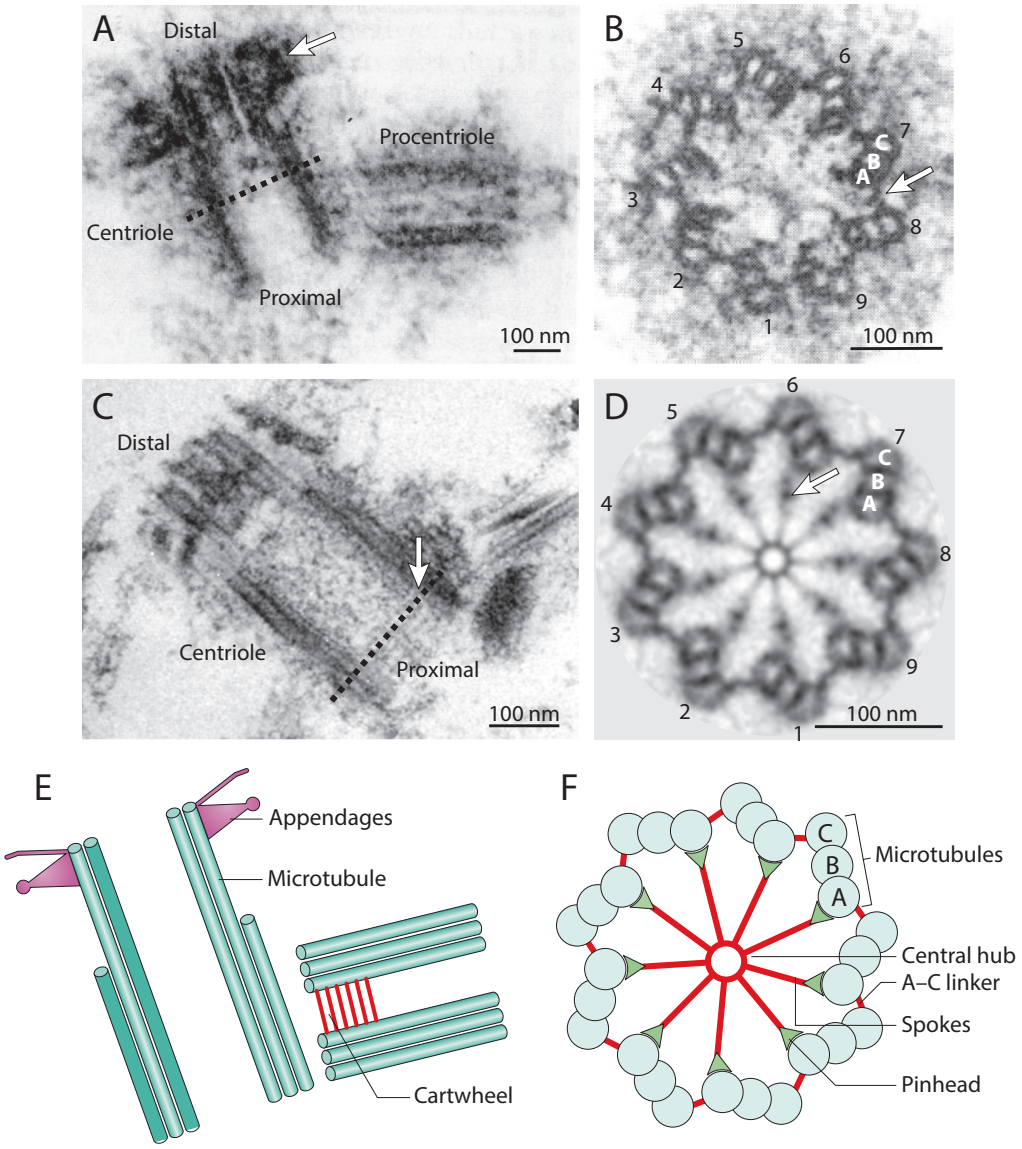


Figure 1.12: **Architecture of centrioles and cartwheels.** (**A**, **B**) Cryo-electron micrograph of resin-embedded human centrioles (mother centriole + procentriole). Part **A** shows centrioles in their longitudinal cross-section, indicating their proximal and distal ends, as well as distal and sub-distal appendages (arrow), which act as docking points for cytoplasmic microtubules, and can also anchor centrioles to cell membranes in the process of creating cilia. Dashed line indicates an approximate position of the transverse plane cross-section of the centriole shown in **B**. It reveals radial nine-fold symmetry of microtubules (triplets: A-, B- and C- tubules), which connect via the A-C linker (arrow in **B**) to form the outer wall ring structure. (**C**, **D**) Cryo-electron micrograph of resin-embedded *C. reinhardtii* centrioles. Part **C** shows the longitudinal cross-section of the centriole, indicating proximal and distal ends, as well as presence of a ∼100 nm tall cartweel (arrow). Part **D** is the transverse plane cross-section of the proximal region (dotted line in **C**), which includes the cartwheel. It depicts the central hub with nine spokes emanating from it towards the periphery of the centriole. The spokes connect with the microtubule wall via structure called pinhead (arrow). (**E**, **F**) Schematic representations of the human centriole and procentriole pair showed in the sections above. While the mother centriole has already lost its cartwheel, the procentriole still contains a stack of cartwheel rings (not very clear in **A**). Figure adapted from [124].

nine-fold symmetry of microtubule doublets in axonemes (Figure 1.13B on page 25; a microtubule-based protrusion which grows out of a basal body and provides structural scaffold for a growing cilium) is not absolutely essential for nucleating a functional cilium. The nine-fold symmetry of cilia has been lost in various species of insects, which can produce 3 to 70 doublet or triplet microtubules (Figure 1.13A on page 25), and in some extraordinary cases this number was found to reach 2500 doublet copies in a single cilium [282]. Although unusual, these giant centrioles are capable of nucleating perfectly motile cilia as basal bodies, or organising mitotic and meiotic spindles. Moreover, *C. reinhardtii* mutants in proteins Bld10p (Cep135) [143] and Bld12p (SAS-6) [237] produced centrioles with variable number of triplets instead of nine (Figure 1.13C, D on page 25). These mutant centrioles were still able to make cilia, suggesting that nine-fold symmetry is not an essential feature of centrioles, but perhaps the most optimal one. More importantly, it appears that microtubule triplets can form laterally connected arrays without the complete radial symmetry being present (Figure 1.13E on page 25) [143, 237, 286]. One can therefore speculate that if such a chunk of microtubules can perform any useful ciliary-related function, the selection process could have favoured these arrangements, which then over hundreds of millions of years could accumulate additional improvements, and eventually develop into the present-day cylinders of nine-fold symmetry (Figure 1.13F on page 25).

1.3.2 Molecular basis for centriolar symmetry

After revealing the identity of the 'core five' proteins which are essential in assembling a centriole, the next major step in understanding centriole biogenesis was to uncover the mechanisms by which centrioles attain the radial symmetry. Critical evidence came from the lab of Masafumi Hirono which investigated the mutants of SAS-6 and Cep135 homologues in *C. reinhardtii* [143, 237]. While *bld10*-null mutants were completely unable to assemble centrioles, Bld10p (Cep135) constructs which lacked ∼35 % of the C-terminal segment could produce centrioles of smaller diameter, with eight symmetrically arranged microtubule triplets instead of nine (Figure 1.14D on page 27). In these mutants, some of the cartwheel spokes did not connect to microtubule triplets, suggesting that the pinhead could have been defective (Figure 1.14C,

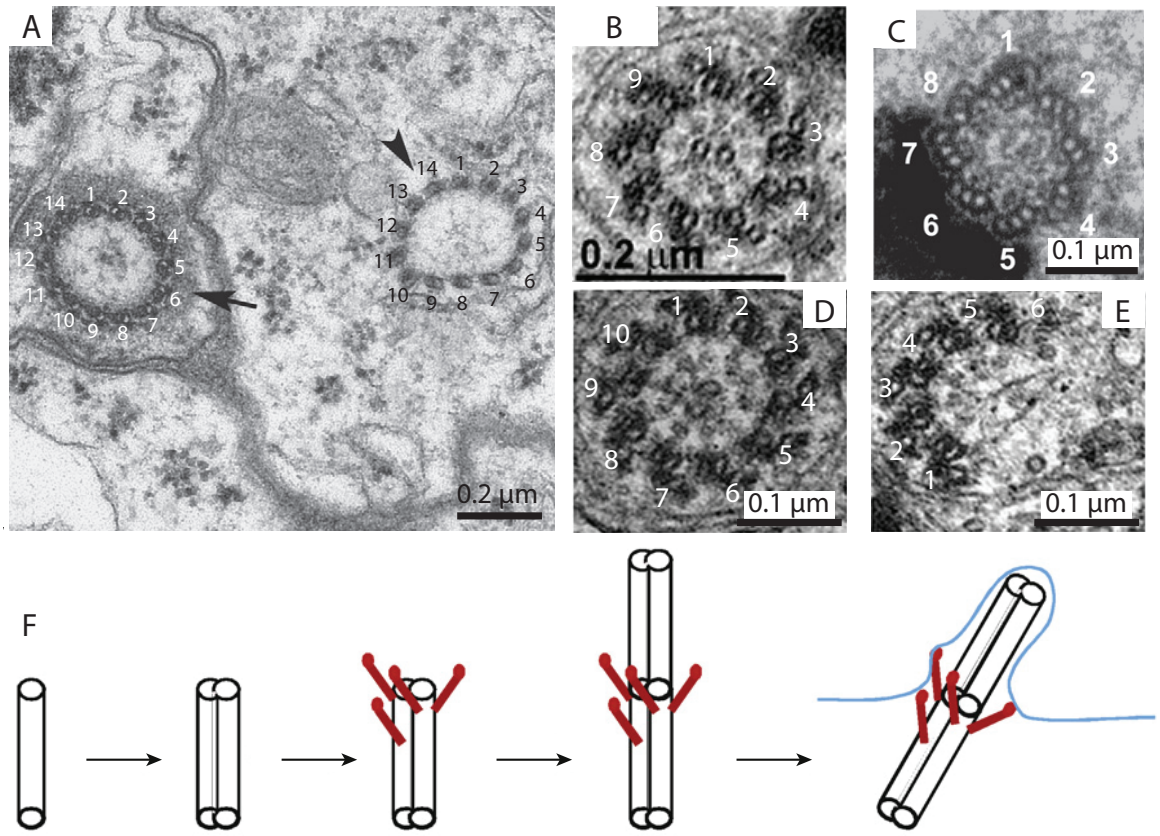


Figure 1.13: **Hypothetical evolution of centrioles.** (**A**) 14 doublet microtubules make a giant centriole (arrow) and axoneme (arrowhead) in the male germ cells of the proturan *Acerentomon*. (**B**) Electron micrographs of nine-fold symmetrical wild-type axoneme from male germ cells of *Drosophila*, made of nine doublet microtubules and a central pair. Mutations in genes for SAS-6 or Bld10p (Cep135 equivalent in *Chlamydomonas*) lead to centriole aberrations, such as reduced (**C** - *Chlamydomonas* centriole) or increased number of microtubule blades (**D** - *Drosophila* axoneme), and also incomplete ring structures (**E** - *Drosophila* axoneme). (**F**) A model of a proto-centriole evolution as proposed by Marshall [220]. With the emergence of microtubule inner proteins (MIP), stable microtubule doublets and triplets can form in the cell cytoplasm (see Section 1.3.5 *Doublets and triplets of microtubules and auxiliary proteins* on page 36). A class of microtubule-binding proteins evolves as proto-centriolar appendages, allowing the entire structure to dock to cell membranes. Further growth of microtubules forces the cell membrane to form a protrusion, which can act as a sensor of external environment or as a gliding machine - a proto-cilium. Figures adapted from [143, 220, 286].

G on page 27). Moreover, they showed that Bld12p (SAS-6) localises to the central part of the centriole - the cartwheel. *C. reinhardtii* mutants lacking *bld12* gene were also deficient in cartwheels (Figure 1.14B, F on page 27), and frequently produced centrioles of variable microtubule triplet symmetry: 7, 8, 9, 10 or 11. Furthermore, investigation of basal body formation from a ciliated protozoan - *Tetrahymena*

thermophila, revealed that its SAS-6 homologue localised to cartwheels of basal bodies [175], and that depletion of *sas-6* results in a defective assembly of new basal bodies and shortening of pre-existent cilia [74]. Conversely, over-expression of SAS-6 in various systems leads to formation of supernumerary centrioles or centriole-like structures, and in turn, results in increased frequency of multipolar spindles [200, 261, 286, 316]. These results suggested that while SAS-6 is both needed for seeding and maintaining the cartwheel, Cep135 appears to sit in the pinhead region, connecting the cartwheel spokes with the microtubule wall. The actual molecular architecture of cartwheels and the mechanism by which they dictate nine-fold symmetry was unknown at the time.

With the advances in modern experimental methods - electron microscopy and X-ray crystallography, the three-dimensional picture of centrioles' morphology had begun to take shape in 2010s. Two landmark studies by Kitagawa *et al.* [183] and van Breugel *et al.* [331] revealed crystal structures of *C. elegans* and *Danio rerio* SAS-6 proteins, as well as that of *C. renhardtii* SAS-6 homologue, Bld12p. These structures coupled with functional *in vivo* assays shed light on the molecular basis for the cartwheel structure, and thus the nine-fold symmetry of centrioles. SAS-6 is composed of three distinct domains: the N-terminal globular head, followed by a long stretch of α-helix in the middle, and a disordered C-terminus, which includes the KEN box - a cell-cycle-dependent degradation recognition motif (Figure 1.15A on page 29). The long α-helix is predicted to exhibit self-oligomerisation properties by forming an extended super-secondary structure, known as the coiled coil, where >2 α-helices (right-handed) twist around one another to form a left-handed super-coil [215].

Biophysical experiments showed that the coiled-coil domain of *C. elegans* SAS-6 forms a parallel dimer of apparent interaction strength $\sim$0.9 μM. Interestingly, the crystal structures uncovered an additional interaction site - between the pairs of N-terminal globular heads (the N-N interaction) (Figure 1.15B, D on page 29). A single residue of one head domain (I154 in *C. elegans*, F145 in *C. reinhardtii*, F131 in *D. rerio* and *H. sapiens*) forms a plug which inserts into a hydrophobic pocket of a neighbouring head domain (Figure 1.15D on page 29). This pocket is a part of the conserved PISA (present in SAS-6) motif [200]. The head to head (N-N) interaction appeared to be two orders of magnitude weaker ($\sim$110 μM for *C. elegans* SAS-6) than that of

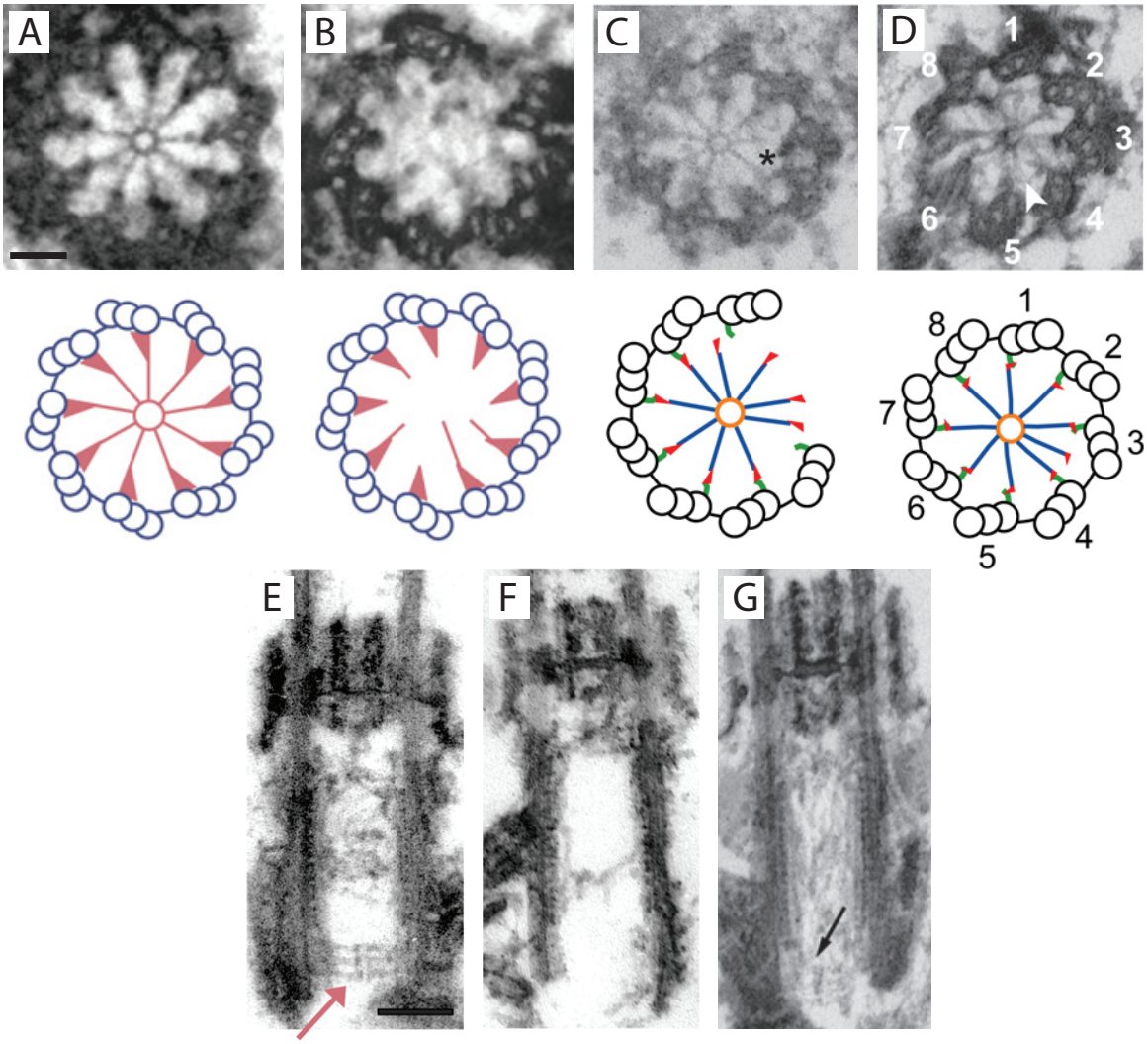


Figure 1.14: **Abnormal cartwheels in SAS-6 and Cep135 mutants.** Mutations in *C. reinhardtii* Bld12p (SAS-6) and Bld10p (Cep135) have marked effects on the centriolar cartwheel. The wild-type cartwheel is presented in **A** (transverse plane cross-section and its schematic drawing below) and **E** (longitudinal cross-section - notice several cartwheel layers indicated by the red arrow). While most of the centrioles in the *bld12* mutant were fragmented (similar to that of the axoneme in Figure 1.13E on page 25), the remaining ∼20 % of centrioles retained the circular arrangement of microtubules, but missed their carwheels (central hub and spokes). Such absence of cartwheels in their proximal parts (**B**, **F**), did not affect formation of pinheads. On the other hand, *bld10* mutants did not support formation of centrioles at all. Small deletions in Bld10p, however, produced shorter cartwheel spokes (C-terminal deletion), which were unable to efficiently connect to every single microtubule blade. It appeared that the connection between spokes and blades was weakened, and thus some blades were detached completely (**C**, **G**, star indicates a detached blade), or the circumference of the centriole wall decreased, accommodating eight triplets instead of nine (**D**), such that all spokes could reach to blades. However, this arrangement left one (the ninth) spoke unattached (indicated by the arrowhead in **D**). Scale bars in **A** - 50 nm and **E** - 100 nm. Figures adapted from [143, 237].

the coiled coil, and thus was detectable only at relatively high concentrations of SAS-6 in solution. Mutating the hydrophobic plug residue to glutamic or aspartic acids abolished oligomerisation. The rescue of centriole duplication in worm embryos under *sas-6(RNAi)* conditions with SAS-6 carrying the I154E mutation was unsuccessful, and the mutated version of *H. sapiens* SAS-6 N-N interface (F131D) or the coiled coil (90 % deletion) could not localise to centrioles in human cells, demonstrating the importance of both oligomerisation interfaces in centriole assembly.

Kitagawa *et al.* [183] analysed high concentrations of a *C. reinhardtii* SAS-6 construct, consisting of both oligomerisation interfaces, with an electron microscopy technique that is particularly suited for visualisation of elongated molecules (rotary metal shadowing), and noticed rings with characteristic spokes protruding outwards - a clear resemblance to a centriolar cartwheel (Figure 1.15C on page 29). Furthermore, purified *D. rerio* SAS-6 construct, also consisting of both oligomerisation interfaces, crystallised spontaneously in a test tube for van Breugel *et al.* [331]. Although the crystals were too small and did not diffract X-rays strongly enough to yield a structure, cryo-electron microscopy (cryo-EM) of these crystals revealed ring-shaped assemblies (Figure 1.15C on page 29). Thus, both studies arrived at the same conclusion: the existence of such two independent and self-interacting sites within SAS-6 results in formation of higher-order species in two steps. First, SAS-6 dimerises through the coiled-coil domain, forming a long tadpole-like structure. Then, these tadpole dimers associate via the N-N interface to create a curved chain with a potential to close on itself, and effectively form a ring structure - the cartwheel. Thus, such self-assembling property of SAS-6 leads to formation of a symmetrical cartwheel, which is essential for templating centrioles.

A few years later, van Breugel *et al.* [332] strengthened their earlier cryo-EM results of the *in crystallo D. rerio* SAS-6 rings. The group was investigating a potential of using SAS-6 as a drug target to combat *Leishmania major*, a parasite which is the main cause of leishmaniasis in the Middle East and North Africa, and relies on its flagella to move and sense the host environment. By screening a library of small molecules, they found a compound which interrupted the N-N homo-oligomerisation; however, it not only affected SAS-6 of the parasite but also to some extent *D. rerio* SAS-6. Because of the lack of specificity, such compound could not have been imme-

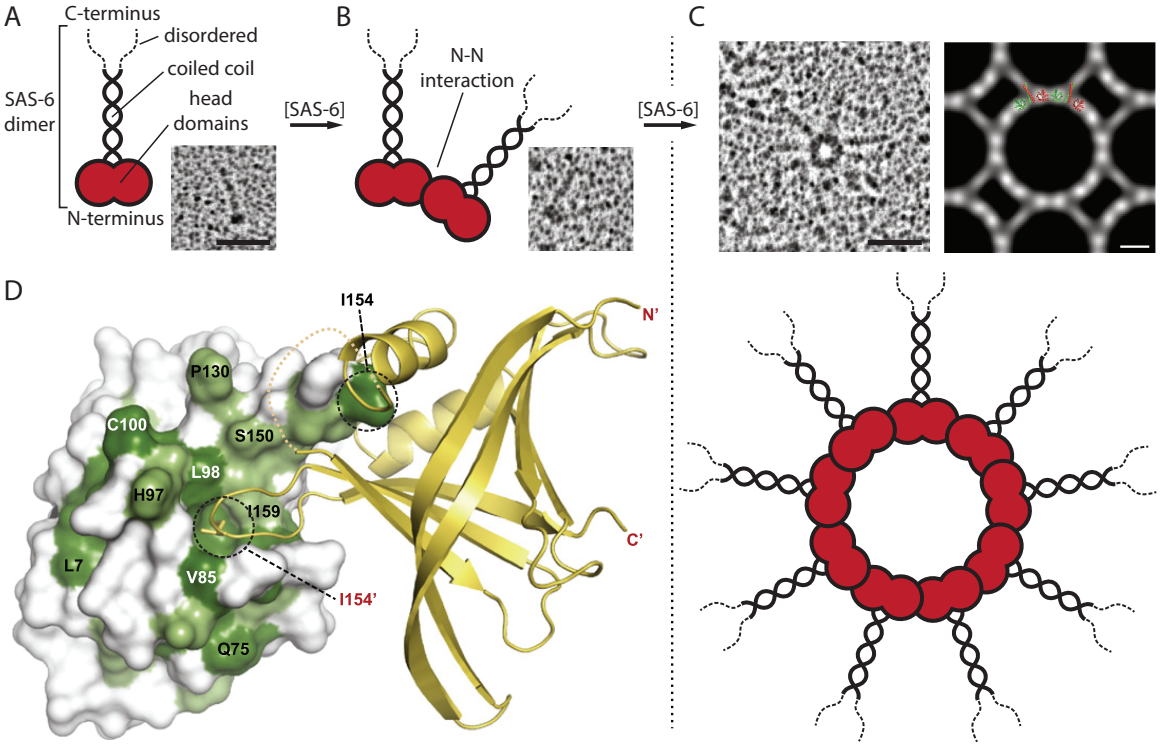


Figure 1.15: **Self-oligomerisation of SAS-6.** (**A**) SAS-6 exists as a dimer in the cytoplasm, self-interacting via the coiled-coil domain (CC, K_d ~1 μM). The CC domain is flanked by the head domain at the N-terminus, and a disordered segment at the C-terminus. Electron micrograph of *C. reinhardtii* SAS-6 reveals a 'tadpole' shape of SAS-6. Scale bar, 50 nm. (**B**) Although very weakly, the head-domains also self-interact (K_d ~110 μM), and form a V-shaped tetramer at higher protein concentrations. (**C**) Once the concentration of SAS-6 is significantly increased, the N-N interaction mediates closing of the large oligomer into a ring structure with characteristic spokes emanating radially - as seen by rotary-metal shadowing electron microscopy (top left; scale bar, 50 nm) or cryo-electron microscopy of SAS-6 crystals (top right; scale bar, 60 Å). (**D**) The N-N interaction relies on a hydrophobic plug residue (indicated by black dotted circles) of one head domain, which slots into a hydrophobic pocket of another head domain. The N-domains taking part in this interaction are shown in cartoon and surface representations. Evolutionarily conserved residues on the surface of *C. elegans* SAS-6 are coloured green. Figures adapted from [183, 331].

diately used in therapy, but nevertheless, it provided a proof of principle that SAS-6 oligomer could be disrupted by small molecule intervention. Following up on their quest to assess the ability of SAS-6 to impart the nine-fold symmetry of centrioles, van Breugel *et al.* [332] crystallised *L. major* SAS-6. This time, diffraction was much stronger than that of *D. rerio* microcrystals, allowing for a successful structure deter-

mination. It revealed an array of dazzling nine-fold symmetrical rings tightly packed into a crystal. Although this finding strengthened the previous observations that SAS-6 can form ring-like cartwheels [183, 331], it remained unanswered how, given the low affinity of the N-N interaction, a stable SAS-6 cartwheel can be assembled within a living cell.

1.3.3 The unique cartwheels of *C. elegans*

Although considered iconic in centriole research, *C. elegans* produce centrioles that are not quite the same as in other organisms. The centrioles in worm embryos are relatively small, 150 nm in length and 100 nm in diameter [262]. In the cross-section, instead of a ring and radial spokes that connect to microtubule triplets, the electron micrographs of *C. elegans* centrioles show a hollow tube, termed the central tube (Figure 1.16A on page 31), diameter of which is larger (∼60 nm) than that of the central hub of cartwheels (∼22 nm in diameter). The central tube of *C. elegans* centrioles is surrounded by singlet microtubules, arranged in nine-fold radial symmetry. Since the domain architecture of *C. elegans* SAS-6 is identical to SAS-6 of other species which promote cartwheel formation, it was unclear how the same protein could create a central tube instead. Hilbert *et al.* [142] solved the crystal structure of a dimer of *C. elegans* SAS-6. The crystallisation construct was specifically designed to include the CC dimer interface (mediated by a short stretch of truncated coiled-coil domain), and the N-N dimerisation interface was intentionally disrupted through mutagenesis of the plug residue in order to facilitate crystal packing. Interestingly, they concluded that they were unable to *in silico*-model a cartwheel ring from these *C. elegans* SAS-6 structures without introducing serious steric clashes between protein chains. Instead of a ring, SAS-6 seemed to be making a right-handed spiral with an open end (Figure 1.16B on page 31). It became apparent that the relative orientation of the coiled coil and the N-terminal head domain differed from those seen in *C. reinhardtii* and *D. rerio* by a ∼38 ° rotation (Figure 1.16C on page 31).

To follow up on this finding, they engineered a mutant SAS-6, which stabilised the otherwise very weak N-N interaction (I154W; Figure 1.16D, E on page 31) and allowed them to observe elongated SAS-6 spirals - both double and single, with electron microscopy (Figure 1.16F on page 31). Through computational modelling, Hilbert

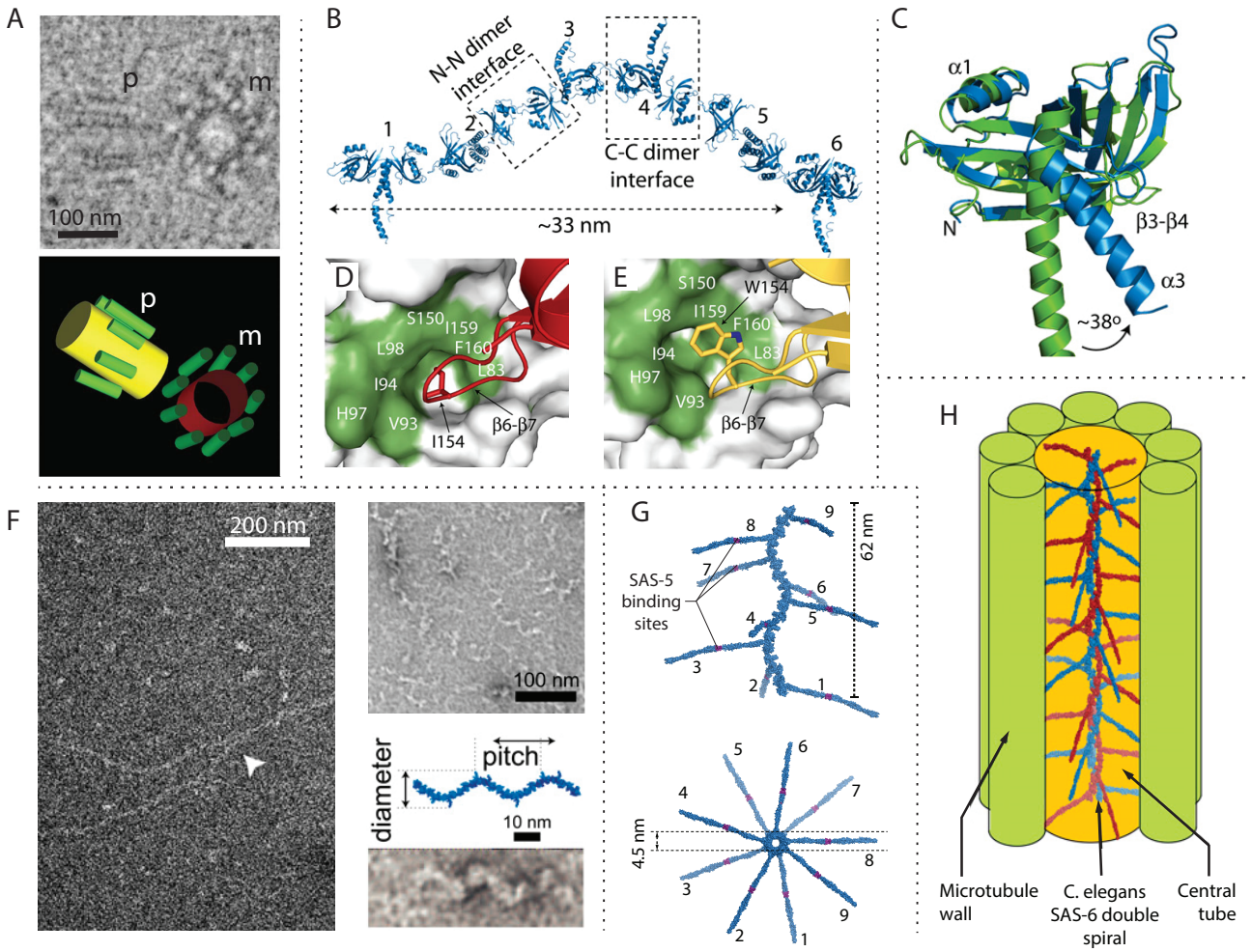


Figure 1.16: **Cartwheels of *C. elegans*.** (**A**) Centrioles of *C. elegans* are considerably smaller (∼100 nm) than those of humans. Electron tomographic slice (top) and corresponding 3D model (bottom) show a mother centriole (m, red) and a growing procentriole (p, yellow) during mitosis. Note the presence of the central tube surrounded by nine singlet microtubules. (**B**) Spatial modelling of a *C. elegans* cartwheel by superimposition yields a spiral structure instead of a flat ring, with pitch of ∼33 nm. (**C**) A swing of ∼38 ° of the coiled-coil (α3) of *C. elegans* SAS-6 (blue) relative to *C. renhardtii* (green) explains such spiral conformation. (**D**, **E**) Mutation in the hydrophobic plug, I154W, increases the affinity of the N-N interaction from ∼110 μM to ∼5 μM. Crystal structures of this mutant interface reveal that, compared to isoleucine (**D**), tryptophan (**E**) makes more contacts within the hydrophobic pocket, stabilising the interaction. Protein chain of the plug is drawn as cartoons, and the plug itself is drawn as sticks; protein chain of the pocket is drawn as surface, indicating conserved residues in green. (**F**) Electron micrographs of negatively stained N-N-stabilised I154W mutant SAS-6, which forms spirals *in vitro* (top right). These spirals also seem to intertwine into a double spiral (left) - the arrowhead marks the point where two strands come together. The model of the tight, 4.5-fold symmetric *C. elegans* SAS-6 spiral agrees with the spirals observed by electron microscopy (bottom right). (**G**) Model of the fully extended SAS-6 molecule forming a single spiral depicts how nine-fold symmetry can be achieved (top - side view, bottom - top view). The binding sites for SAS-5 are indicated (see Section 1.4.2 *SAS-5 in cartwheel formation: interaction with SAS-6* on page 47 for details). (**H**) Schematic of the *C. elegans* centriole showing how four turns of a SAS-6 double spiral could potentially build a central tube. Figures adapted from [142, 262].

et al. [142] were able to show that these spirals could potentially achieve nine-fold symmetry along the spiral axis without creating any steric clashes between residues (Figure 1.16G on page 31). The resulting spiral is very steep, with the pitch of ∼31 nm and only 4.5 SAS-6 dimers per turn. Such low number of SAS-6 molecules required to build the spiral cartwheel could be the reason for the seemingly hollow density of the central tube in *C. elegans* centrioles (Figure 1.16A, G on page 31). The diameter of the spiral can reach ∼72 nm (calculated based on the full stretch of *C. elegans* SAS-6 coiled coil), which is similar to the diameter of the central tube observed in *C. elegans* early embryos (∼60 nm).

Although the spirals could not be directly observed *in vivo*, the mechanistic structural biology approach allowed for demonstration that a single protein, after small evolutionary adjustments, could produce two distinct oligomerisation modes, a ring and a spiral. Fascinatingly, both of these modes could effectively arrive at the same output symmetry required for centriole assembly. However, since such an open-ended spiral is likely to be unstable (the N-N dimerisation is very weak if I154W mutation is not introduced), other proteins must be assisting SAS-6 in both nucleation and maintenance of the cartwheel/spiral structure. More experimental work is needed to confirm the identity of the SAS-6 spiral in *C. elegans* centrioles. Particularly promising could be applying the recent innovations for high-resolution cryo-electron tomography to study native centrioles of *C. elegans* as well as stabilised SAS-6 spirals *in vitro* [244].

Interestingly, the particular tadpole fold of SAS-6 resembles proteins involved in DNA repair (non-homologous end-joining), namely XRCC4 and XLF [10,165,206,303] (and also a new family member - PAXX [247]). While the coiled-coil tail of XRCC4 is uninterrupted and fully outstretched (as in SAS-6), the equivalent domain in XLF is broken up by random coil sequences. Thus, only a limited fragment of the XLF tail can form a coiled coil, while the remaining part folds back to reach up to head domains. Although XRCC4 and XLF are unlikely to form coiled-coil-mediated hetero-dimers, they readily form hetero-oligomers through head domains in a manner similar to that of SAS-6 [11,135,288,351]. Because XRCC4 and XLF also differ in the angle between head domain and coiled coil by ∼45 ° (as in *C. elegans* SAS-6 *vs. C. reinhardtii* SAS-6), they are able to form helical filaments that resemble the *C. elegans* SAS-6

spiral. These filaments are thought to be functionally relevant, potentially acting as a scaffold for recruitment of repair factors to damaged DNA.

1.3.4 Z-dimension and pinhead asymmetry

Revealing the elegant properties of SAS-6, which allow it to assemble into a cartwheel ring or a spiral, was a major step in understanding centrioles. It also opened the door for further molecular exploration of structure and assembly process of centrioles. Understanding how cartwheels stack in the dimension of centriole growth, the z-dimension, and how this stack connects to microtubule blades is the next molecular puzzle awaiting an answer. Moreover, while the role of cartwheel as a symmetry-templating structure is well-established, its centriole-maintaining role is not yet exactly clear. A typical human centriole is ∼500 nm long, and approximately 100 nm of this length is occupied by the stack of cartwheel rings. The unusual centrioles of *Trichonympha* can extend up to even a few micrometers in length, ∼90 % of which will be made of cartwheel layers. Some organisms keep cartwheels intact after a centriole has been fully developed (*Trichonympha*), while others, like *Chlamydomonas* and *Spermatozopsis*, reduce their stack from 7-10 layers in development to 2-4 layers once they have matured. Mammalian centrioles, on the other hand, completely dispose of their cartwheels as soon as mitosis begins [144].

Most recently, Paul Guichard and colleagues employed cryo-electron tomography in combination with sub-tomogram averaging to resolve the structure of the proximal region of the *Trichonympha* centriole [131, 132]. The resulting cryo-EM maps (from a mixture of three species of *Trichonympha*: *T. collaris*, *T. sphaerica*, and *T. campanula*, later referred to as *Trichonympha sp.*) featured the cartwheel stack at 38 Å resolution - after applying nine-fold symmetry averaging. They were also able to reconstruct the microtubule triplet, together with the pinhead and the A-C linker at 42 Å resolution (Figure 1.17A on page 35). A single cartwheel layer, which corresponds to SAS-6-contributed central hub and spokes, also appeared to contain extra densities inside of the central hub (Figure 1.17B on page 35). These extra densities, termed cartwheel inner densities (CID), also form a ring, diameter of which is smaller than that formed by SAS-6. Those two rings (SAS-6 and CID) make contacts exactly at the SAS-6 N-N interaction region, suggesting presence of yet unidentified proteins

capable of binding SAS-6 N-N dimers. Furthermore, in contrast to tight cartwheel stacks observed by van Breugel *et al.* [332] in the crystal of *L. major* SAS-6 (4.3 nm periodicity), the stacks observed in native *Trichonympha* centrioles had substantial spacings (8.5 nm periodicity; Figure 1.17C on page 35). Those spacings showed no notable density, suggesting that even though SAS-6 rings could potentially stack tightly on top of one another (as in the *L. major* SAS-6 crystal), there is preference for a slightly looser packing *in vivo*, allowing for some cushioning room between them. Such room could potentially be occupied by a disordered protein, which would act as a glue for the growing stack, but not necessarily show up as defined density due its inherent disorder.

Interestingly, in the z-dimension *in vivo*, two cartwheel spokes, each 50 nm long, merge approximately 20 nm away from the central hub. Thus, yet another periodicity (17 nm) is created, corresponding to the length of two tubulin dimers in the microtubule protofilament (Figure 1.17C on page 35). This spoke merger suggests that SAS-6 coiled-coil dimers must either supercoil further into a tetramer, or rather sit on top of each other, glued together by another protein. The thickness of the merged radial spoke increases further towards its periphery, where it fuses with the pinhead. While the pinhead had only been known as a simple merging platform between the spoke and the microtubules, Guichard *et al.* [132] have now demonstrated that in fact the pinhead has a complex structure, exhibiting a clear asymmetry in the z-dimension. It is made of a hooked structure called pinbody, which directly interacts with the cartwheel, and of two pinfeet that exhibit alternating spacing in the z-dimension (distal-to-proximal direction: pinfoot1-pinfoot2, 8 nm and pinfoot2-pinfoot1, 9 nm; Figure 1.17D on page 35). The asymmetry stems from the pinfeet, which are angled towards centriole's proximal end, where they make a direct contact with the A3 protofilament of the A-microtubule (protofilament numbering follows the Tilney-Linck convention [212]). This angle seems to be imparting directionality on microtubules, which align their plus ends towards the distal end of the centriole.

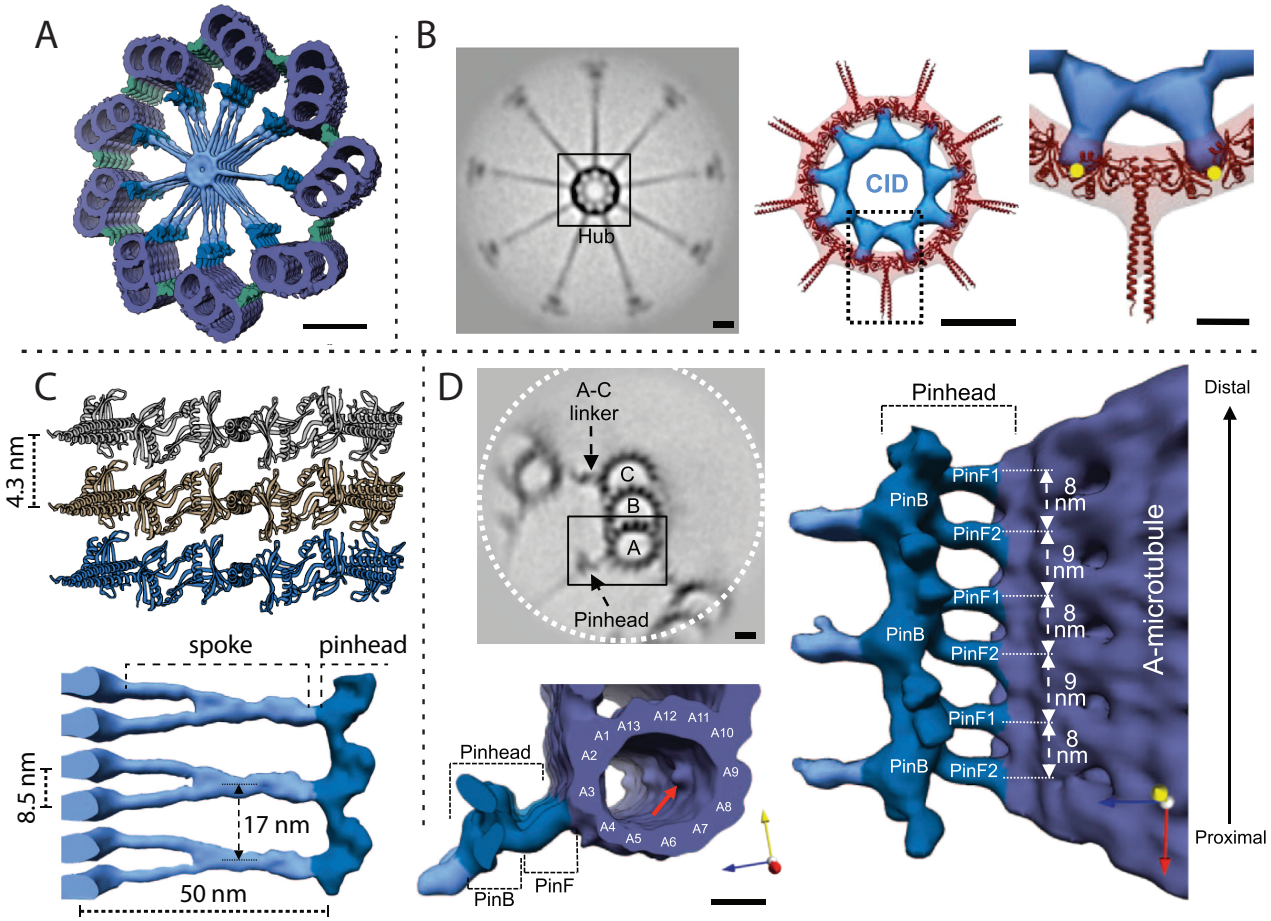


Figure 1.17: **Procentrioles in three dimensions.** (**A**) 3D map of a *Trichonympha* procentriole at 38 - 42 Å resolution. Features are coloured as follows: the cartwheel stack (light blue), pinhead (dark blue), microtubule triplets (purple) and A-C linker (green). Scale bar, 20 nm. (**B**) Cartwheel inner densities (CID) were found inside of the central hub of the cartwheel (left panel; note nine spokes that emanate from the central hub; scale bar, 10 nm). A model of SAS-6 with an incomplete coiled coil was fit into the central hub density; the CID volume is coloured blue (middle panel; scale bar, 10 nm). The boxed region (middle panel) is magnified in the panel on the right (scale bar, 2 nm). Note the interaction of the CID with SAS-6, close to the N-N interface. Yellow dots indicate positions of the hydrophobic residue serving as the plug in the N-N dimerisation (Y118 in *T. agilis*). (**C**) Stacking of SAS-6 cartwheels: in the crystal (top; side view), with periodicity of 4.3 nm, and in the *Trichonympha* procentriole (bottom; longitudinal cross section), with periodicity of 8.5 nm. Two adjacent spokes merge into a thicker arm of 17 nm periodicity. (**D**) The 3D map revealed further features of the pinhead (top left panel; scale bar, 10 nm): pinbody (PinB) and two pinfeet (PinF1 and PinF2). Boxed region of top left panel is magnified in the bottom-left panel (scale bar, 10 nm). The pinhead is ∼15 nm in length, and connects with the A-tubule via A3 protofilament. The two pinfeet are arranged asymmetrically along the z-axis (right panel), and show two distinct periodicities. Such asymmetry defines the polarisation of centrioles along their proximal-distal axis (the z-dimension). Figure adapted from [132].

1.3.5 Doublets and triplets of microtubules and auxiliary proteins

The mystery of how doublets or triplets of microtubules are formed has been puzzling researchers for years. Treating doublets of axonemal microtubules with increasing concentrations of ionic detergents results in a progressive dissolution of the doublets, eventually leaving only three to four protofilaments of the A-tubule intact. This last remaining fragment of microtubules is known as the ribbon (Figure 1.18A on page 38), and its resistance to detergents stems from the presence of proteins other than tubulin, which bind and stabilise the protofilaments [211]. These non-tubulin proteins include the family of tektins, as well as other coiled-coil proteins (Figure 1.18B-D on page 38) [6]. Cryo-electron tomography maps of axonemal doublet microtubules at ∼40 Å resolution, revealed A-tubule to contain 13 protofilaments, which took a slightly eliptical shape (8 % deformation) [239, 320]. The B-tubule was incomplete, with only a crescent of 10 protofilaments attached to the A-tubule in the region near the protofilaments A1 and A10 (Figure 1.18E on page 38; the numbering convention of protofilaments has evolved since the publication of these two papers - see [212]). The radius of the B-tubule was also somewhat larger (129 Å) than that of the A-tubule (112 Å), and corresponded to a microtubule made up of 15-protofilaments. The attachment sites between tubules A and B are termed inner and outer junctions (Figure 1.18E on page 38), and correspond to the density between protofilaments: A10, A11, B1 (outer junction) and A1, B10 (inner junction). Furthermore the partition region, which includes the ribbon structure, is made of protofilaments A1 and A10-A13, and covers the open part of the B-tubule, effectively separating the two tubules. Typical electron micrographs of single microtubules depict a hollow lumen surrounded by protofilament density. However, the doublets of microtubules have been shown to contain non-tubulin density around the partition region (Figure 1.18E on page 38), which is likely contributed by tektins and other ribbon-specific proteins such as Rib72 and Rib43a [210].

With the ever-increasing resolution of axonemal microtubule doublets, the picture of where exactly these extra proteins bind, and what type of tubulin isoform, α or β, is involved in such interaction, has started to take shape [45, 210, 216, 230]. Takashi Ishikawa and colleagues have recently produced a ∼19 Å resolution map of

the microtubule doublet from *Tetrahymena thermophila* using single-particle electron microscopy (Figure 1.18E on page 38) [216]. By decorating protofilaments with the head domain of monomeric kinesin, which specifically binds to β-tubulin, they were able to flag these isoforms and therefore understand the type of microtubule lattice present in each tubule. It turned out that both tubules are made of the B-type lattice (a type of lattice where α-tubulins on one protofilament contact other α-tubulins of the neighbouring protofilament - see Figure 1.18F on page 38), and effectively ended a few decades old controversy. The position of the seam (*i.e.* a single boundary between two protofilaments in the B-type lattice where instead of typical α-α lateral contacts, α subunits are adjacent laterally to β subunits [226] - see Figure 1.18F on page 38) of the A-tubule could not be unambiguously determined, but it was predicted to sit close to the outer junction, either between protofilaments A9 and A10 or between A10 and A11.

Besides the ribbon proteins, lumen of the microtubule doublets also contains other microtubule inner proteins (MIP), densities of which have now been catalogued with high precision; however, their protein identities are largely unknown [216]. The existence of a few abnormally wide distances between protofilaments revealed positions of bound MIPs. Those gaps between protofilaments together with the eliptical distortions in the A-tubule may suggest that MIPs play a major role in diversifying the structure of ciliary and centriolar microtubules from the cytoplasmic microtubules. It is not clear whether MIP proteins are added once a microtubule doublet has assembled, or perhaps are themselves needed to assemble a doublet. Given that centriolar microtubules likely contain the most highly modified tubulin of all cellular microtubules, assembling a doublet or a triplet of microtubules must require enzymatic activity and various post-translational modifications. Therefore, MIPs could potentially provide the recognition epitope for enzymes involved in catalysing these reactions [159].

The presence of microtubule triplets in centrioles is correlated with two additional tubulin isoforms, δ and ϵ [56]. Organisms which do not contain genes encoding for δ- and ϵ-tubulins, produce centrioles which lack the third tubule, and therefore are made of either doublet (*Drosophila*) or singlet (*C. elegans*) microtubules. Mutations in the gene for δ-tubulin in *Chlamydomonas* or complete silencing of this gene in

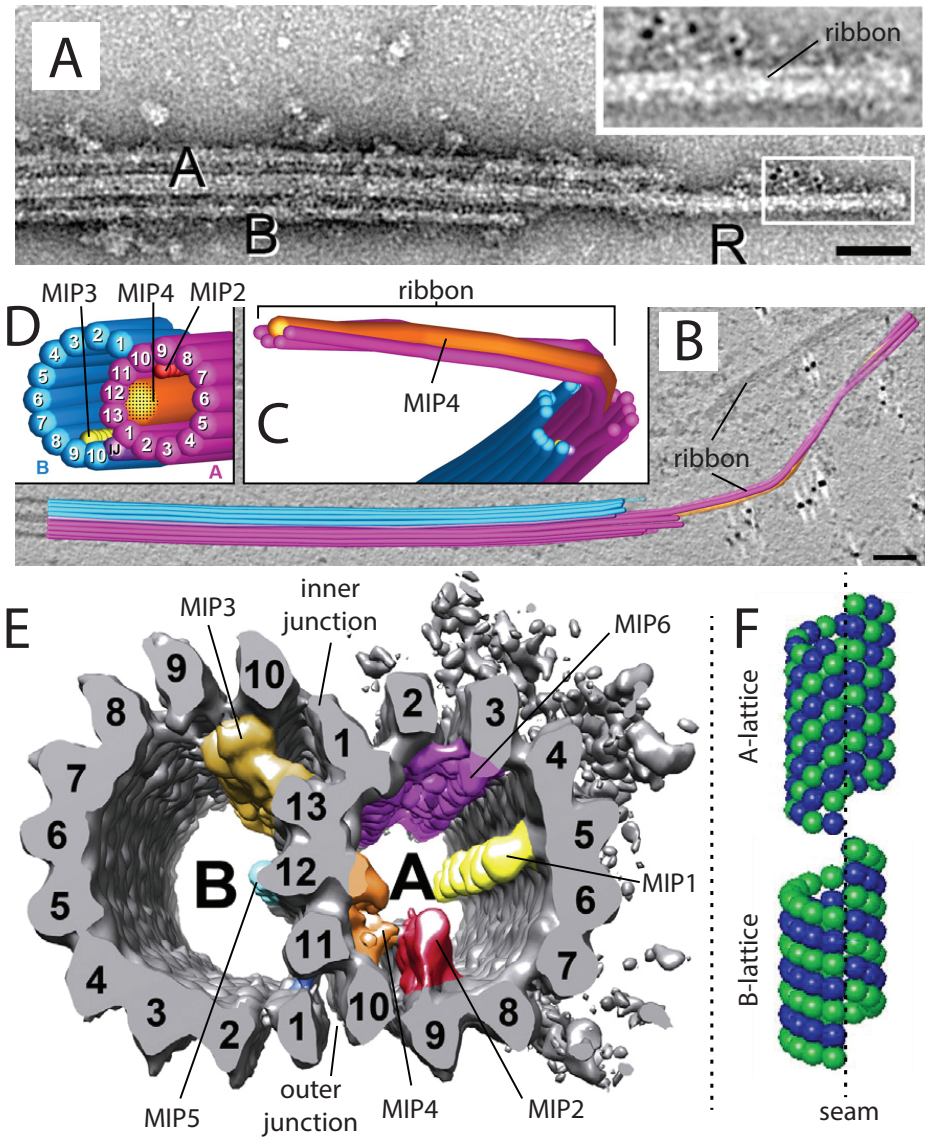


Figure 1.18: **Doublet of microtubules.** (**A**) Electron micrograph of a negatively stained doublet of microtubules ("A" marks the A-tubule and "B" marks the B-tubule), from sea urchin *Strongylocentrotus purpuratus* sperm flagella, transitioning into a ribbon ("R"). The ribbon is immuno-labelled with anti-*Sp*Rib74, which does not label intact microtubule doublets or A-tubules, but is specific to the side of the ribbon which faces the lumen of the A-tubule (MIP4 in C, D, E). Inset shows magnification of the boxed region. Scale bar, 50 nm. (**B**) Section from an immuno-cryo-electron tomogram showing transition from microtubule doublet to ribbon. Note that labelling only occurs at the side of the ribbon, which normally faces the lumen of the A-tubule. Reconstructed volumes are overlaid onto the doublet-ribbon structure, and shown in details in **C** (A-tubule in magenta, B-tubule in blue). Ribbon protofilaments correspond to A1, A11, A12, A13. (**D**) Microtubule doublet model as seen from the proximal end; various features are indicated (compare with the map in E). (**E**) ~19 Å resolution map of the microtubule doublet from *T. thermophila* cilia, as seen from its distal end. Various MIPs (MIP 1-6), tubule junctions (outer and inner), as well as positions of individual protofilaments are indicated. (**F**) Comparison of microtubule lattices: top panel - A-lattice, bottom panel - B-lattice. Two distinct colours (green and blue) indicate α- and β-tubulin. Details in Section 1.3.5 *Doublets and triplets of microtubules and auxiliary proteins* on page 36. Figures adapted from [210, 216, 226].

Paramecium, resulted in the loss of C-tubules with no effect on the functionality of basal bodies [97, 116]. ϵ-tubulin, on the other hand, appears to be essential for the assembly of centrioles and for the maintenance of microtubule triplets [344]. By using a cadmium-responsive gene promoter, Ross *et al.* [290] were able to observe depletion of *Tetrahymena* ϵ-tubulin in real time. Initially, triplets of microtubules started to disappear, and eventually entire basal bodies were lost.

Unlike axonemes, which uniformly contain doublets of microtubules along their entire length, centrioles vary the number of microtubules in different segments. The distal end of centrioles in most species is made of microtubule doublets, which in case of basal bodies, reach the transition plate and directly nucleate the axonemal doublets. In the proximal and central regions, however, an extra C-tubule is present. To shed light on the architecture of microtubule triplets, two independent studies by Guichard *et al.* [132] and Li *et al.* [205] obtained two cryo-electron tomography maps of basal body microtubules from the proximal region (*Trichonympha sp.*; 42 Å) and the central region (*C. reinhardtii*; 33 Å), respectively. The arrangement of protofilaments in tubules A and B was largely identical to that seen in axonemal doublets, with A-tubule also exhibiting eliptical distortion of 8-10 % (Figure 1.19A, C on page 41). Furthermore, similarly to the B-tubule, the additional C-tubule crescent was made of 10 protofilaments, and by attaching to the B-tubule (attachment sites: C1:B4 and C10:B8), it formed the second partition (B5-B8). Furthermore, a number of MIP densities was observed, locations of which partially overlapped with those seen in axonemes (Figure 1.19A, C on page 41).

1.3.6 The A-C linker

Although cartwheels initiate assembly of microtubule blades into a radially symmetrical structure, they do no seem to be responsible for maintenance of centrioles, as some organisms completely disassemble cartwheels once their centrioles have matured. Instead, the stability of microtubules is most likely achieved by the linker which connects the blades laterally into an uninterrupted, cylindrical wall. The density of the linker, as revealed by cryo-electron tomography of *C. reinhardtii* and *Trichonympha sp.* [132,205], runs between the A-tubule of one microtubule blade and the C-tubule of a neighbouring blade, and thus is known as the A-C linker (Figure 1.19B on page 41).

The linker completes the microtubule wall, which in turn provides compartmentalisation, *i.e.* inside or outside of the centriole. *Tetrahymena* basal bodies, integrity of which has been compromised through various mutations, were found to be flooded by cytoplasmic material, including ribosomes, and eventually led to the loss of microtubule triplets [260].

The attachment sites for the A-C linker seem to vary between *Trichonympha sp.* and *C. reinhardtii*. *Trichonympha* basal body at its proximal end appears to contain a clear linker running directly between protofilaments A8 and C9, which also exhibits asymmetry along the proximal-distal axis, and the periodicity of 8.5 nm (Figure 1.19A, B on page 41) [132]. On the other hand, the A-C linker in *Chlamydomonas* is not as explicit. It connects protofilament A6 with a broad region on C-tubule (Figure 1.19C on page 41), and its shape varies along the centriolar z-dimension [205]. This variation directly reflects the left-handed twist of the microtubule blades along the longitudinal axis, from 10 ° at the proximal end (measured as deviation of the triplet vector from the sides of an equilateral nonagon) to 1 ° at the distal end. Interestingly, such twisted centriolar barrels have been observed before, and while the twist angles reached up to 25 °, the diameter of the barrel remained constant [8, 255]. Finally, the triplets in *Chlamydomonas* basal bodies contained density similar to the pinhead observed in *Trichonympha*, but much more extensive, with the pinbody reaching towards the luminal part of the C-tubule. The identity of proteins residing in the A-C linker density is not known.

1.3.7 Attachment of microtubules to cartwheels

Length is an important variable in centriole biology, and one of those which are rather difficult to grasp. Microtubules are known to grow and shrink in the process called dynamic instability, which limits their length to certain range beyond which they become too unstable. However, microtubules themselves are much longer than molecular machines which organise and nucleate them - the centrioles. How do centrioles recognise when it is time to stop growing, or what determines the diameter of centrioles? While the answer to the second question has begun to take shape, the first question remains completely open.

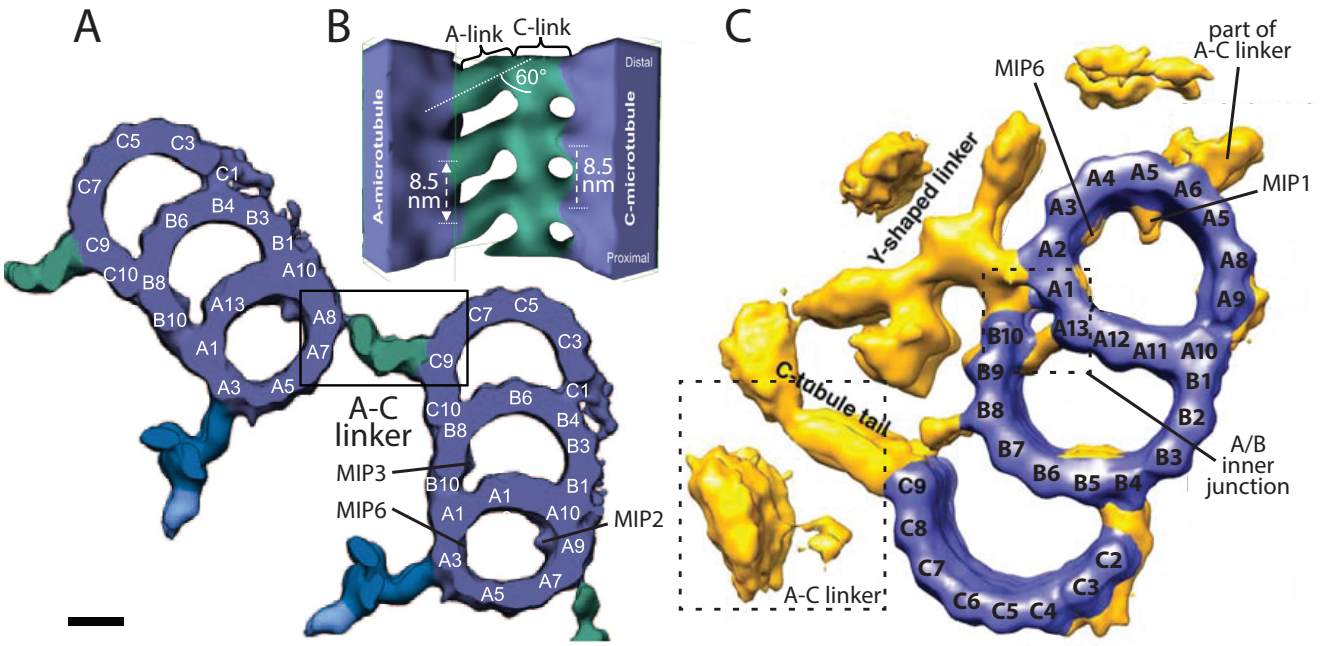


Figure 1.19: **Microtubule triplets and the A-C linker.** (**A**) Two connected *Trichonympha sp.* microtubule triplets as viewed from the distal end. The volume is a cryo-electron-tomographic reconstruction at 42 Å resolution. Positions of individual protofilaments are indicated. The A-C linker (boxed; green) connects two microtubule triplets via protofilaments A8 and C9. Scale bar, 10 nm. (**B**) Side view of the A-C linker region, revealing periodicity of 8.5 nm in the z-dimension. The A-C linker is composed of two parts, A- and C-links, related by a 60° angle, thus exhibiting polarity along the proximal-distal centriole axis. (**C**) Microtubule triplets in the central region of *Chlamydomonas* basal bodies, as viewed from the proximal end (Note: *Trichonympha* triplets in **A** are drawn from the opposite, distal end). Tubulin-contributed densities are in purple and non-tubulin-contributed in yellow. The non-tubulin densities were revealed after subtracting a pseudo-atomic model of a microtubule triplet from the complete basal body triplet density map. Some of the non-tubulin densities correspond to densities known from axonemal doublet microtubules, such as MIP1, MIP6 and the inner junction between A- and B-tubules. The other densities might be specific to centriolar triplets, the most prominent of which is the Y-shaped linker that stems from the A/B-tubule inner junction. The A-C linker of *Chlamydomonas* basal bodies is not as well defined as in *Trichonympha*. The A-link connects to protofilament A6 (A8 in *Trichonympha*), whereas the C-link shape seems to vary along the centriolar z-axis: from protofilament C9 at the distal end, to protofilaments C8 and C7 at the proximal end (C9 in *Trichonympha*). Figures adapted from [132, 205].

The diameter of centrioles appears to be largely controlled by the available length of SAS-6 coiled coil, with ∼0.15 nm axial raise per residue. Therefore, slight differences in centriole diameter between species can be directly explained by looking at the sequence of SAS-6 and comparing the lengths of predicted coiled coils. However, values calculated with this method are likely to be somewhat shorter than the actual

radius. This difference stems from the fact that cartwheel spokes do not touch microtubule blades directly, but are connected through the pinhead (Figure 1.12F on page 23). Studies of mutant *Chlamydomonas bld10* have revealed involvement of its gene product, protein Bld10p, in producing spokes of correct length [143]. Bld10p (vertebrate orthologue - Cep135) has been shown to localise to the distal part of cartwheel spokes, with its N-terminus located close to microtubule blades [143, 162]. Indeed, mutants with the N-terminal part of Bld10p removed produce spokes which are disconnected from the microtubule wall (Figure 1.14C on page 27). Moreover, the N-terminal-most fragments of Bld10p and Cep135 alone (residues 1-163 and 1-190, respectively) have been shown to directly bind and stabilise microtubules [52, 209]. Because the spokes are connected to microtubules through the pinhead, it is likely that Bld10p/Cep135 is a part of this structure - likely contributing to one of the pinfeet (Figure 1.17D on page 35). Furthermore, removal of the C-terminal portion of Bld10p resulted in shortened spokes and a smaller pinhead (Figure 1.14D on page 27), suggesting that this protein segment might not only be a part of the pinhead but also of the spoke itself [143]. The interaction of the C-terminal part of Cep135 with the coiled coil of SAS-6 further supports this hypothesis [209]. Given that Bld10p/Cep135 is predicted to be made almost entirely out of coiled coil (over 100 nm in length), it is indeed unlikely that a fully-stretched coiled coil of Bld10p/Cep135 could possibly fit within the pinhead alone. Instead, this coiled coil might lay adjacent to the merged pair of SAS-6 spokes in the distal spoke region. It is intriguing to speculate whether Cep135 is the glue which binds the spokes together. Nonetheless, the spoke merger is not long enough to explain the extent of the predicted coiled coil of Cep135, and therefore some part of this protein is also likely to reside (through bending perhaps) within the pinbody, and then continue towards the A-tubule as a part of one pinfoot.

It has been shown that the N-terminal region of *Chlamydomonas* Bld10p is not required for centriole assembly [143]. Thus, it is likely that another microtubule-associated protein could compensate for such a deletion. One line of evidence points to SAS-6 itself, which has been shown to bind to microtubules via its disordered C-terminal segment (residues 470-657 of *H. sapiens* SAS-6; $\sim$0.2 μM affinity to α/β-tubulin dimers) [133]. This region of SAS-6 in *C. elegans* (residues 406-492) is known to be a prime target of ZYG-1 phosphorylaton *in vitro*; however, mutating the phos-

phorylation sites on SAS-6 to alanine in worm embryos did not affect centriole assembly [203]. Because of the unaffected phenotype in those mutants, Lettman *et al.* [203] dismissed the possibility that SAS-6 is a critical target of ZYG-1 in centriole assembly *in vivo*. Nevertheless, it will be interesting to see whether the disordered tail of *C. elegans* SAS-6 could also bind to microtubules, and whether phosphorylation can have an effect on affinity. A positive outcome might suggest that there is another protein which compensates for the loss of binding between SAS-6 and microtubules.

Another potential candidate for bridging cartwheel spokes and microtubules is CPAP (vertebrate orthologue of *C. elegans* and *D. melanogaster* SAS-4), an essential protein for centriole assembly [179, 201], which has been shown to interact with microtubules (through its N-terminus) [67, 151, 153] and with Cep135 (through its C-terminus) [209]. In fact, CPAP has also been implicated in control of centriole length, with its over-expression leading to abnormally long centrioles [189, 295, 322]. Recently, Hatzopoulos *et al.* [140] have solved a high-resolution crystal structure (1.4 Å) of the C-terminal domain of *D. rerio* CPAP (known as G-BOX or TCP), revealing an elongated β-sheet structure with no apparent hydrophobic core. Those β-sheets were able to connect back-to-back and form a long fibril of ∼8 nm periodicity (Figure 1.20B on page 45). Similar periodicity was observed between the pinfeet and also between α/β-tubulin dimers in microtubules. Therefore, it was proposed that the G-BOX domain of CPAP could be a part of the pinbody and act as a molecular ruler in scaffolding centrioles in the z-dimension, prior to attachment of microtubules (Figure 1.20E on page 45). Because high concentrations of CPAP lead to longer G-BOX fibrils, it would be expected that cells which over-express CPAP, also produce longer centrioles [189, 295, 322]. Furthermore, as a modular protein, CPAP also contains a ∼24 nm stretch of coiled coil, N-terminal of which sit tubulin-binding (PN2-3 fragment) and microtubule-binding (A5N fragment) domains (Figure 1.20A on page 45). Hatzopoulos *et al.* [140] identified that the coiled coil of CPAP adopts a parallel arrangement, and therefore could potentially provide support to microtubule triplets (Figure 1.20C, D on page 45). They proposed that the coiled coil of CPAP might be a part of the Y-shaped densities observed by Li *et al.* [205], which run parallel to the microtubule blade on the luminal side (Figure 1.19C on page 41), effectively bridging the pinhead and the B/C-tubules. In fact, CPAP with both tubulin- and

microtubule-binding domains, might play a role in the process of nucleating B- and C-tubules. These speculations, however, do not explain why *C. elegans* SAS-4, modular architecture of which is almost identical to that of CPAP (Figure 1.20A on page 45), is required to build centrioles with singlet microtubules instead of triplets. Given that no Cep135-like protein has been identified in *C. elegans*, it would be interesting to find out how *C. elegans* SAS-4 and potentially also SAS-6, could recruit microtubules and attach them to the cartwheel in this system. Nevertheless, more work is needed to explain the mechanistic role of SAS-4/CPAP in the process of microtubule attachment to cartwheels, and also in tethering pericentriolar material to centrosomes.

1.4 Beyond the obvious

Cryo-electron tomograms of *Chlamydomonas* and *Trichonympha* basal bodies have given us an extremely powerful tool for making inferences into the roles played by centriole-assembling proteins. With the aid of these maps, we are now able to build models of how individual proteins can fit into the actual density of centrioles, thus reconstructing their interaction network in three dimensions. We must remember, however, that these density maps are averages between many "pictures" of almost identical content. Therefore, items which happen to have defined locations in those pictures are reinforced to give strong density. Other features, however, location of which differs from picture to picture, are completely lost in the averaging process because of their intrinsically dynamic or transient nature. Several other proteins besides SAS-6, Cep135 and CPAP have been shown to both localise to procentrioles and to be indispensable for their assembly. Yet, their locations within the centriole density cannot be unambiguously determined. One such protein is SAS-5 - the main focus of my doctoral research, described at length on the following pages of this thesis.

1.4.1 SAS-5 defines a new protein family

Following the discovery of the 'core five' essential factors for centriole duplication in *C. elegans*, the quest to locate orthologues of these proteins in other organisms moved relatively quickly; however, only four out of five were found with high certainty. One protein had eluded this pursuit, namely SAS-5. Finding orthologues of SAS-5 had

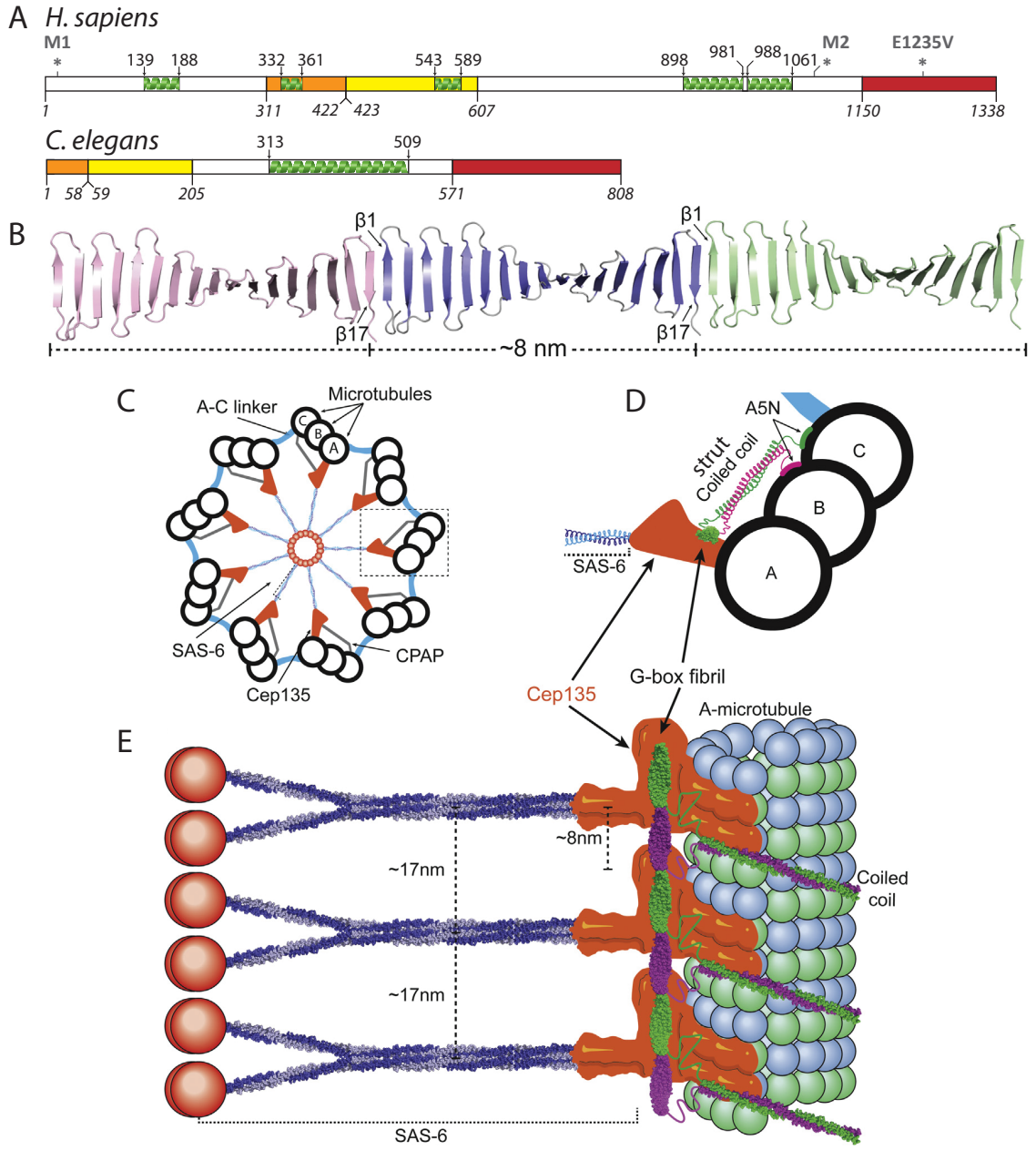


Figure 1.20: **Potential involvement of CPAP in centriole architecture.** (**A**) Comparison of domain architecture between *H. sapiens* CPAP and *C. elegans* SAS-4. Various features are indicated: tubulin binding (PN2-3, orange), microtubule binding (A5N, yellow), G-BOX domain (red), predicted coiled coil (green helix), microcephaly mutations (stars; M1, M2 - nonsense mutations). (**B**) G-BOX domain of *Danio rerio* CPAP forms a back-to-back β-sheet fibril in the crystal (consecutive protein chains shown in pink, purple and green). (**C**) Schematic drawing of centriole proximal region, showing cartwheel, pinheads (orange) and microtubule triplets connected laterally via A-C linkers. Hatzopoulos *et al.* [140] proposed that CPAP (grey), via its coiled-coil domain, may act as a hotizontal strut to link the pinhead with the triplet of microtubules. (**D**) Cep135 is likely the major component of the pinhead, receiving and stabilising the merged SAS-6 spoke. Cep135 also interacts with the G-BOX domain of CPAP (green), which may provide an attachment site for the coiled-coil strut. At the N-terminal end of the coiled-coil strut, the microtubule-binding A5N domains could connect with B and C tubules. (**E**) Given that CPAP fibril shows periodicity (~8 nm) similar to those of the α/β-tubulin dimer and the pinfeet, it could potentially act as a molecular ruler to evenly spread the cartwheels in the z-dimension. Figure adapted from [140].

proven a daunting task, mainly due to the unique primary sequence of this protein. The only direct homologues of SAS-5 are found within the nematode phylum, and the remote homology search in species other than nematodes is riddled with false positives. Thus, the link between SAS-5 and its orthologues in *Drosophila* and vertebrates was established much later, with the identification of proteins that displayed similar functions and cellular behaviours to SAS-5.

The early studies of the protein STIL in humans, mice and zebrafish were consistent with the function of SAS-5 previously described in worms; however, they were not met with general appreciation at first. While zebrafish *stil* mutants exhibited unusual centrosome deficiency in dividing cells [265], similar mutants in mice showed defects characteristic of erroneous cilia function (*e.g.* deviancy in neural tube formation, and randomised left-right assymetry) [156]. Later, it was revealed that mutations in human *stil* cause microcephaly, a neurodevelopmental disorder characterised by reduced head circumference [193]. Microcephaly was also found to be caused by mutations in *MCPH1*, *CDK5RAP2*, *ASPM*, and *CPAP/CENPJ* [348]. Because all of these genes produce centrosomal proteins, it was implied that STIL must be somewhat involved in the regulation of centrosome function. Later, three independent studies by Tang *et al.* [323], Arquint *et al.* [14] and Vulprecht *et al.* [337] showed that STIL is required for centriole duplication in human cells, and that its depletion leads to progressive loss of centrosomes and primary cilia. Conversely, over-expression of STIL results in over-duplication of daughter centrioles next to pre-existing mothers. Moreover, STIL was found to localise with growing procentrioles during G1-to-S transition, and remained associated with daughter centrioles until its degradation during metaphase-to-anaphase transition. Furthermore, over-expressed STIL was found in lung cancer cells and also, to some extent, in other tumours [103, 276]. Knock-down of STIL in those cancer cells *in vitro* delayed mitotic entry and induced apoptosis in p53-dependent manner [102]. Such tumourigenesis might be indirectly induced by the presence of supernumerary centrosomes which are formed after over-expression of STIL [121].

Working with flies, Stevens *et al.* [313] discovered that *D. melanogaster* protein Ana2, when over-expressed, has the ability to induce centriole over-duplication in *Drosophila* spermatocytes. Ana2 also proved to be a key centriole duplication factor

which localised to proximal end of the growing procentriole. They were able to locate a short segment, ∼90 aa in length, in the C-terminal region of Ana2 which showed significant sequence similarity to STIL (31 % identity and 48 % similarity), and therefore named it STAN, for STil-ANa2 homology (Figure 1.21 on page 48). Moreover, Stevens *et al.* [313] attempted to link this region as homologous to the C-terminus of SAS-5; however, sequence similarity between both proteins was drastically low (12 % identity, 26 % similarity), and therefore required further indirect experimental evidence to either refute or accept such hypothesis.

Arquint *et al.* [14] noticed that Ana2 and STIL share yet another region of high similarity, referred to as TIM (truncated in microcephaly). TIM is located at the very C-terminus of STIL, which is truncated in some patients suffering from microcephaly (Figure 1.21 on page 48). However, similarly to STAN, any resemblance of TIM to *C. elegans* SAS-5 based purely on primary amino acid sequence is highly disputable, and should be verified experimentally.

1.4.2 SAS-5 in cartwheel formation: interaction with SAS-6

SAS-6 is capable of assembling cartwheels and spirals as seeds for emerging procentrioles; however, it is obscure how a protein which requires either extra mutations or a boost in concentration to improve its own self-oligomerisation ability, could possibly form a cartwheel in the cell all by itself.

SAS-5 has been shown to physically interact with SAS-6 and drive progression of centriole assembly [142, 200, 275]. It is an essential protein, depletion of which completely terminates centrosome-dependent cell division. Over-expression of SAS-6 in unfertilised *Drosophila* eggs, that normally are devoid of centrosomes, drove the formation of hundreds of large irregular structures (Figure 1.22A, B on page 50) [261]. Interestingly, after co-over-expression of SAS-6 with Ana2 in *Drosophila* spermatocytes, these irregular structures turned into somewhat distinct particles called SAPs (Sas6/Ana2 particles), which stained for centriolar markers, recruited PCM and nucleated microtubule asters (Figure 1.22C, D on page 50) [313, 315]. SAPs often inserted themselves between two paired centrioles (Figure 1.22E-H on page 50), and closer inspection with electron tomography revealed that SAPs were made of highly organised structures called SAStubules (SAS-6, Ana2/SAS-5 tubules), which to some

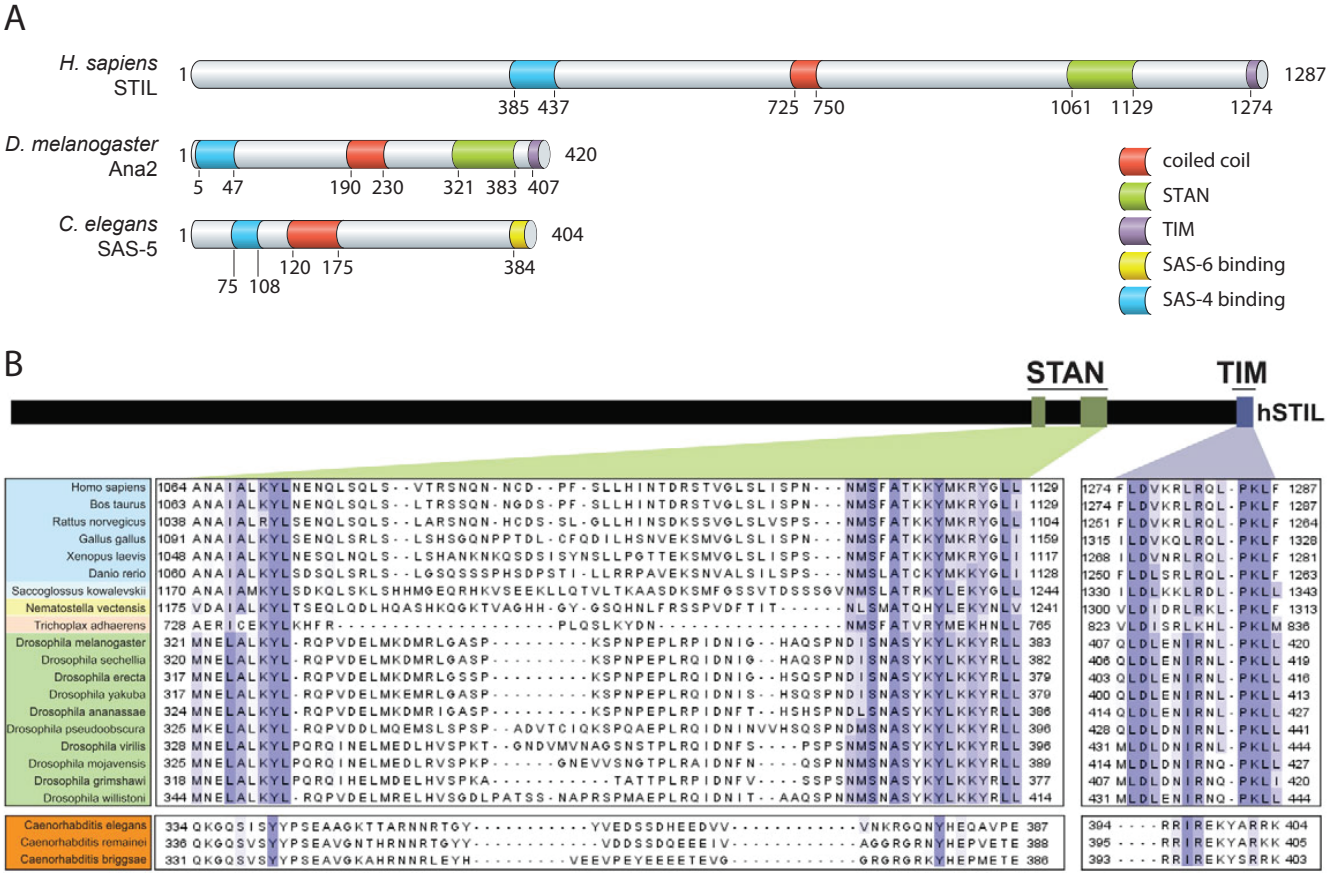


Figure 1.21: **SAS-5 family of proteins.** (**A**) Schematic illustration of various domains in SAS-5, Ana2 and STIL: predicted coiled coil (red), STAN motif (green), TIM motif (purple), SAS-6 binding region (yellow), SAS-4 binding region (blue). Note that the SAS-4 binding region in *C. elegans* SAS-5 is a prediction based on the proline-rich motif (see Section 1.4.3 *Microcephaly and the SAS-5 / SAS-4 interaction* on page 54 for details). (**B**) Sequence alignment of STAN and TIM motifs across SAS-5 family of proteins by Arquint *et al.* [14]. Columns scoring highly in similarity are coloured purple. While STIL and Ana2 show high degree of similarity within the primary sequence of STAN and TIM motifs (top box), nematode SAS-5 (bottom box) appears extremely divergent, prompting experimental verification of these hypothetical evolutionary connections.

extent resembled centriolar cartwheels in shape and size (Figure 1.22I-K on page 50). Furthermore, once Sak kinase (*Drosophila* orthologue of *C. elegans* ZYG-1 and vertebrate PLK4) was also over-expressed alongside SAS-6 and Ana2, then both SAS-6 and Ana2 began loading onto centrioles (Figure 1.22L-O on page 50), which in turn drove centriole over-duplication [315]. Within such over-duplicated centrioles, an extensive network of proteins was observed, which upon centriole disengagement (*i.e.* the process of removing the physical connection between the mother and daughter centrioles)

at the end of meiosis I, became stretched into long fibres (Figure 1.22P on page 50). Intriguingly, these Ana2 and SAS-6-containing fibers kept connecting the centrioles even after disengagement, which eventually led to severe centriole mis-segregation (Figure 1.22R on page 50). Overall, these results showed that Ana2 and SAS-6 are involved in promoting centriole duplication and cohesion (*i.e.* the act of holding duplicated centrioles together) between engaged mother and daughter centrioles, while Sak/PLK4 plays a regulatory role in both processes.

Interestingly, *C. elegans* SAS-5 was also implicated to be co-dependent on SAS-6 for centriole localisation. Embryos expressing mutant SAS-5 which was unable to bind SAS-6, were also failing to duplicate centrioles [88]. Moreover, depletion of either SAS-6 or STIL resulted in an analogous phenotype, where the other protein was no longer targeted to growing procentrioles [14, 323, 337]. Furthermore, FRAP (fluorescence recovery after photobleaching) experiments with labelled SAS-5 and STIL showed that unlike SAS-4 or SAS-6, which are recruited to centrioles only once per cell cycle, SAS-5 and STIL are highly dynamic proteins. Despite a stable population present at centrioles between S and mitosis, SAS-5 and STIL shuttle between cytoplasm and centrioles throughout the cell cycle (Figure 1.23 on page 52) [88, 337]. Such shuttling behaviour of SAS-5 may suggest that this protein takes a role of a chaperone transporter, delivering proteins to the site of the growing procentriole.

It is likely that SAS-6 is one of the proteins transported by SAS-5, where as a SAS-5/SAS-6 complex, they could co-assemble the cartwheel. Recently, in pursuit of understanding the nature of this interaction, two independent groups have arrived at the same conclusion: not more than twenty C-terminal amino acids of *C. elegans* SAS-5 are needed to bind to the middle region of SAS-6 coiled coil (Figure 1.24A, B, C on page 55), with affinity in the low micromolar range (2.7 μM) [142, 275] (interaction regions as mapped by Qiao *et al.* [275] SAS-5: aa 390-404, SAS-6: aa 261-288; interaction regions as mapped by Hilbert *et al.* [142]: SAS-5: aa 384-404, SAS-6: aa 261-290). Additionally, Qiao *et al.* [275] reported the stoichiometry of this interaction as 2:1 SAS-6 : SAS-5, as measured by isothermal titration calorimetry.

In the effort to understand binding between SAS-5 and SAS-6 in more detail, Qiao *et al.* [275] solved the crystal structure of the coiled coiled of *C. elegans* SAS-6 (aa 248-410). It was revealed that the coiled coil of SAS-6 could be divided into

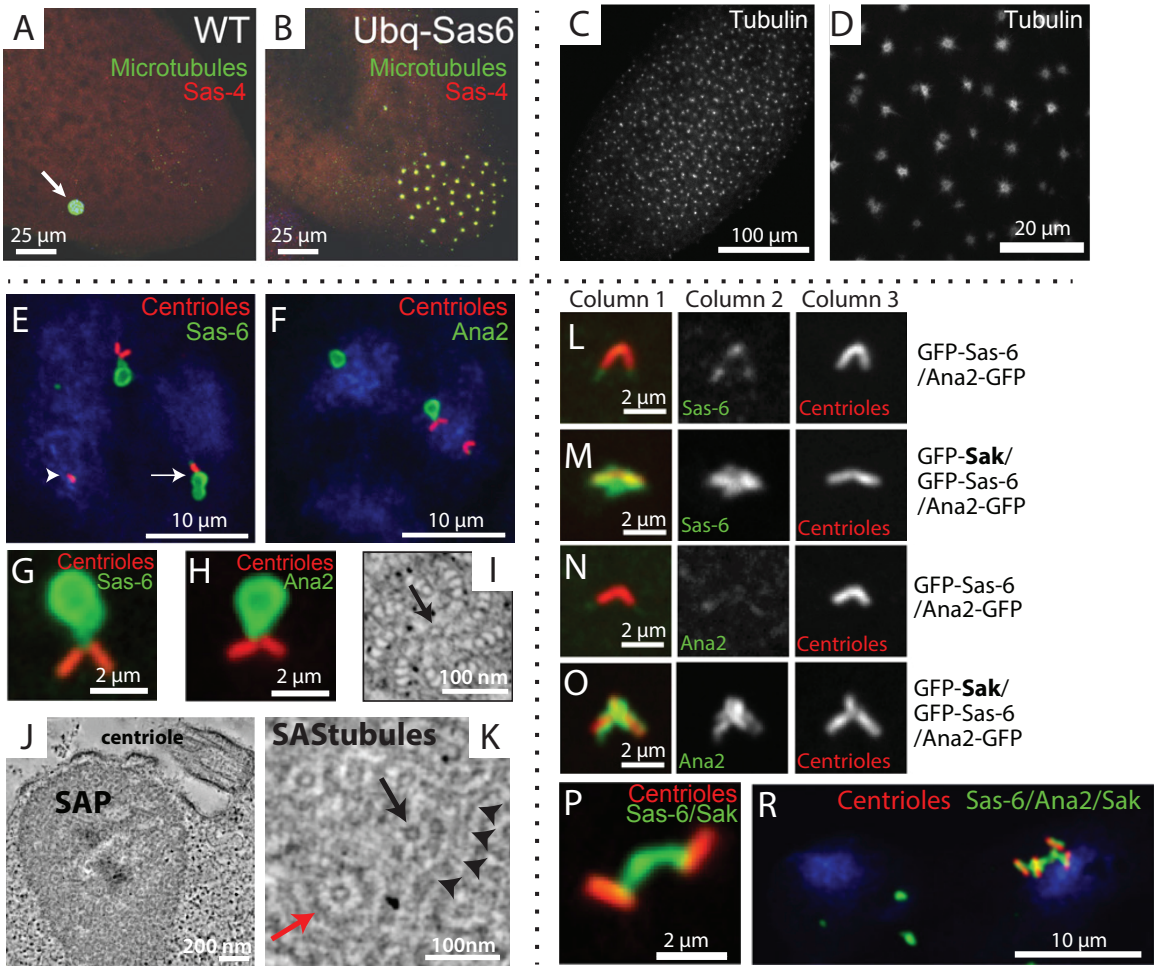


Figure 1.22: **Ana2 and SAS-6 work together to build large structures in flies.** (**A**, **B**) Over-expression of GFP-fused Sas-6 under control of the ubiquitin promoter drives *de novo* formation of large structures in unfertilised *Drosophila* eggs which normally lack centrioles (**B**). Wild type eggs are shown in **A**; the arrow points at the polar body (*i.e.* a small cell that normally buds off the egg cells during oogenesis). Microtubules (green) and Sas-4 (red) were immuno-stained with specific antibodies. (**B**, **C**) The effect of producing large structures in unfertilised *Drosophila* eggs is augmented by over-expression of Ana2 alongside Sas-6, both under control of the ubiquitin promoter (**C**). Magnified view of these structures in panel **D** reveals their aster-like character. The eggs were stained for α-tubulin with specific antibodies. (**E**, **F**, **G**, **H**) *Drosophila* primary spermatocytes expressing Ana2-GFP and GFP-Sas-6. Staining with antibodies specific to Sas-6 (E, G, green) and Ana2 (F, H, green), DNA (blue), centriolar marker GTU88* (red) reveals that both proteins form the body of these large particles, hence the name "SAP" for Sas6/Ana2 particles. SAPs are often inserted between two paired centrioles (magnified views in **G** and **H**). (**J**, **K**) Sections from an electron tomogram of SAPs (**K** - magnified view) reveal structured organisation of those particles, with circular tubules, called "SAStubules" (black arrow) connecting to an outer ring (red arrow) via spokes. SAStubules were imaged in two cross-sections: transverse plane (black arrow) and longitudinal (arrowheads). For comparison, a cross section through a spermatocyte centriole is shown in **I**; black arrow points at the cartwheel. Note that SAStubules and spokes show striking similarity to centriolar cartwheels in both size and morphology. Continued on the next page.

Figure 1.22: **Ana2 and SAS-6 work together to build large structures in flies**. (**L**, **M**, **N**, **O**) Over-expression of Sak/PLK4 kinase together with Ana2 and Sas-6 triggers loading of both Ana2 and Sas-6 onto centrioles, and centriole over-duplication. Compare panels **L** and **M** for Sas-6, and **N** and **O** for Ana2, before and after co-over-expression of Sak. Over-expressed proteins are indicated at the right-hand side of the panels. Column 2 shows staining for either Sas-6 or Ana2; column 3 shows staining for centriolar marker GTU88*; column 1 shows merged columns 2 and 3 (Ana2/Sas-6 - green, centrioles - red). (**P**) Centrioles (red) of a *Drosophila* spermatocyte in telophase of meiosis II. Sas-6 is co-over-expressed with Sak (both green), but not with Ana2, thus the centrioles are not over-duplicated. The centrioles have begun disengagement, but are still linked by a GFP-containing fibre. This fibre is similar to the GFP cross-links in Sas6/Ana2/Sak-over-expressed spermatocytes (**R**), which eventually lead to centriole mis-segregation into one daughter cell. Figures adapted from [261, 313, 315].

four segments which are predominantly either negatively or positively charged in the successive alternating fashion (Figure 1.24D on page 55). Such arrangement allows for an anti-parallel association of SAS-6 dimers into tetramers. Interestingly, this could explain why many electron micrographs of purified SAS-6 protein, on top of the typical tadpole-like structure, also contained a population of dumbbell-like shapes (Figure 1.24E on page 55). These anti-parallel dumbbells of SAS-6 are much weaker ($\sim$56 μM) than tadpoles ($\sim$0.9 μM), and therefore do not show up very ubiquitously on microscopy grids. Moreover, Qiao *et al.* [275] showed that the act of binding SAS-5 to SAS-6, being the stronger interaction (see below), removes any dumbbell self-associations of SAS-6.

In the absence of high-resolution structure of SAS-6 in complex with SAS-5, Qiao *et al.* [275] performed mutagenesis of surface accessible residues of SAS-6, around the mapped region of SAS-5 binding. Coupled with pull-down assays and isothermal titration calorimetry measurements, they were able to determine the identity of the most critical amino acids for this interaction: SAS-6 side - I279, M283, E275, E276, E278, E287, SAS-5 side - I396, R397, Y400, R403. Furthermore, they investigated the effect of disruptive mutations at these key positions on worm viability (on the SAS-6 side only), and found that the mutations abolished recruitment of SAS-6 to growing centrioles, which led to a lethal phenotype similar to that seen in worms where SAS-5 was depleted. Qiao *et al.* [275] attempted computational docking of a theoretical model of SAS-5 C-terminal helix onto the coiled coil of SAS-6, and proposed a potential binding mode between those proteins. Because such *in silico*

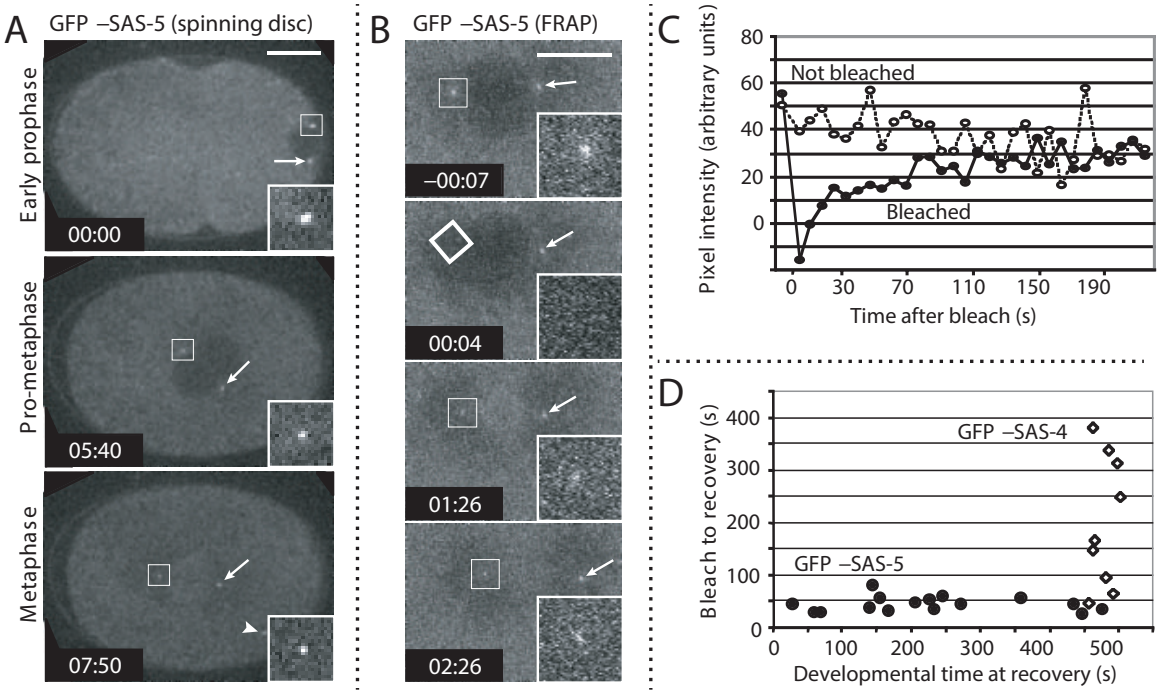


Figure 1.23: **SAS-5 as a cytoplasm-centriole shuttle.** (**A**) Spinning-disc confocal time-lapse microscopy images of *C. elegans* one-cell stage embryo expressing GFP-SAS-5. Cell-cycle stages are indicated on the left-hand side. Elapsed time is shown in the min:sec format. Insets show magnified images of boxed centrioles. Arrows points at the 2$^{\text{nd}}$ centriole. The arrowhead in the bottom panel points a GFP-positive focus in the cytoplasm. Such foci become more frequent as the worm develops. Scale bar, 10 μm. (**B**) FRAP experiment of GFP-SAS-5-positive centrosomes. Confocal sections of one-cell stage *C. elegans* embryos show a nucleus and surrounding centrosomes (GFP foci). Panels follow the photobleaching experiment, with the elapsed time indicated in min:sec (00:00 represents the end of photobleaching). The thick square in the 2$^{\text{nd}}$ panel from the top indicates the targeted area. Insets show magnified images of boxed centrosomes. Arrows point to the 2$^{\text{nd}}$ centrosome. Scale bar, 10 μm. (**C**) Plot of pixel intensity of centrosomes as a function of time after photobleaching (filled circles). Control experiment of unbleached centrosomes is plotted as open circles. The intensities have been background subtracted (average cytoplasm fluorescence). Note the drop in intensity of centrosomes right after photobleaching, and the quick recovery to normal within ∼80 s. (**D**) Plot of recovery time after photobleaching as a function of developmental time at recovery (pronuclear meeting - 0 s, cytokinesis - 430 s). In contrast to GFP-SAS-4 which is recovered only in the next cell cycle, GFP-SAS-5 show continuous protein exchange, allowing it to rapidly recover after photobleaching at any stage of cell cycle. Figure adapted from [88].

approaches are still very much in their infancy, the only reliable route for explaining the true nature of the interaction between SAS-5 and SAS-6 is through experiment. Thus, it still remains a challenge to obtain a high-resolution crystal structure of SAS-6 in complex with SAS-5.

Two structural homologues of SAS-6, XRCC4 and XLF, form helical filaments to potentially aid recruitment and stabilisation of DNA double-strand break repair enzymes. One of these enzymes is likely to be DNA ligase IV (LigIV), which forms a defined complex with the coiled-coil tail of XRCC4 [94, 278, 303, 350]. Because the stoichiometry (1:2 LigIV:XRCC4) and the structural features involved in this interaction (random coil for LigIV and coiled coil for XRCC4) resemble those of SAS-5 and SAS-6, it is interesting to speculate whether the SAS-5/SAS-6 complex can form in a similar fashion.

Finally, because *C. elegans* SAS-6 has never been observed to form ring-like cartwheels (Hilbert *et al.* [142] study of SAS-6 spirals came out later), regardless of concentration it was studied at, Qiao *et al.* [275] set out to examine whether SAS-5 could be the tadpole-to-ring turning factor for SAS-6. They reported that recombinantly produced full-length SAS-5 formed aggregates (Figure 1.24F on page 55); however, once added to purified SAS-6, it was released from those aggregates and formed arch-like structures instead. The authors concluded that since the diameter of these arch-like structures was ∼60 nm, they could have possibly captured a half-assembled central tube of *C. elegans* centrioles. Given the rather poor SAS-5 quality used in these studies, one must be careful not to over-interpret the resulting electron micrographs, as highly contaminated and aggregated samples can yield many unexpected results. In order to avoid confirmation bias [240], better protein samples are needed.

The most exciting fact about these two studies by Hilbert *et al.* [142] and Qiao *et al.* [275] is that SAS-5 binds to a precise location along the long cartwheel spoke created by the coiled coil of SAS-6, approximately 18 nm away from the central hub. Given that SAS-5 is also a coiled-coil protein, and therefore minimum a dimer, it could potentially bridge these spokes, creating yet another ring of nine-fold symmetry. Intriguingly, a structure that fits this description was found in the lumen of the cartwheel of the chytridiomycete fungus *Phlyctochytrium irregulare*, approximately 15 nm away from the central hub (Figure 1.25A, C on page 56) [227]. A hint of such spoke cross-link was also seen in basal bodies of *Trichonympha* (Figure 1.25B on page 56) [118], which is in contrast to much higher resolution tomography reconstructions by Guichard *et al.* [131, 132]. Perhaps by utilising the new CRISPR-Cas technology

for editing genomes [163], it will be possible to investigate this potential cross-link further, but first an orthologue of SAS-5 would need to be identified in *P. irregulare* and *T. agilis*.

1.4.3 Microcephaly and the SAS-5 / SAS-4 interaction

Another protein which could board the "SAS-5 train" en-route to the growing centriole is SAS-4. In fact, CPAP (vertebrate orthologue of SAS-4, also known as CENPJ), has been shown to bind STIL directly [323], and mutations which negatively affect this interaction can cause microcephaly [182]. In two idependent studies, Hatzopoulos *et al.* [140] and Cottee *et al.* [71] solved crystal structures of *D. rerio* CPAP G-BOX domain and *D. melanogaster* SAS-4 G-BOX domain in complex with short fragments of STIL (Hatzopoulos *et al.*: 1.7 Å and 2.4 Å resolution; Figure 1.26A on page 58; Cottee *et al.*: 2.7 Å resolution) and Ana2 (Cottee *et al.*: 2.6 Å resolution). It was revealed that the strength of this interaction was in low micromolar range (0.55 μM for *H. sapiens* and 1.7 μM for *D. rerio*) and that the common microcephaly mutation E1235V weakened this affinity ten-fold. They identified a conserved proline-rich motif within STIL (PRxxPxP) which, if deleted, reduced binding affinity to CPAP 25-fold (Figure 1.26B on page 58). Hatzopoulos *et al.* [140] suggested that such motif can also be found in *C. elegans* SAS-5, which despite the lack of overall sequence similarity between SAS-5 and STIL, might be the only convincing primary sequence fossil advocating common ancestry of those proteins (Figure 1.26C on page 58).

Furthermore, Cottee *et al.* [71] determined that the G-BOX domain of *C. elegans* SAS-4 is required for the centriole localisation of this protein, and for successful incorporation into the outer centriolar wall. To further investigate whether the failure of SAS-4$^{\Delta\text{G-BOX}}$ localisation into the vicinity of the microtubule wall was due to broken interaction with SAS-5, they attempted *in vitro* pull-downs of SAS-4 with two purified fragments of SAS-5 (Figure 1.26D on page 58) immobilised on beads: the N-terminal half (aa 1-202) and the C-terminal half (aa 203-404). While the N-terminal half of SAS-5 successfully pulled full-length SAS-4, it was unable to pull shorter fragments of SAS-4 (neither SAS-4N: aa 1-523 nor SAS-4C: aa 506-808, which is the G-BOX-containing fragment). Based on this experiment, Cottee *et al.* [71] concluded that G-BOX domain of *C. elegans* SAS-4 appears both essential and insufficient

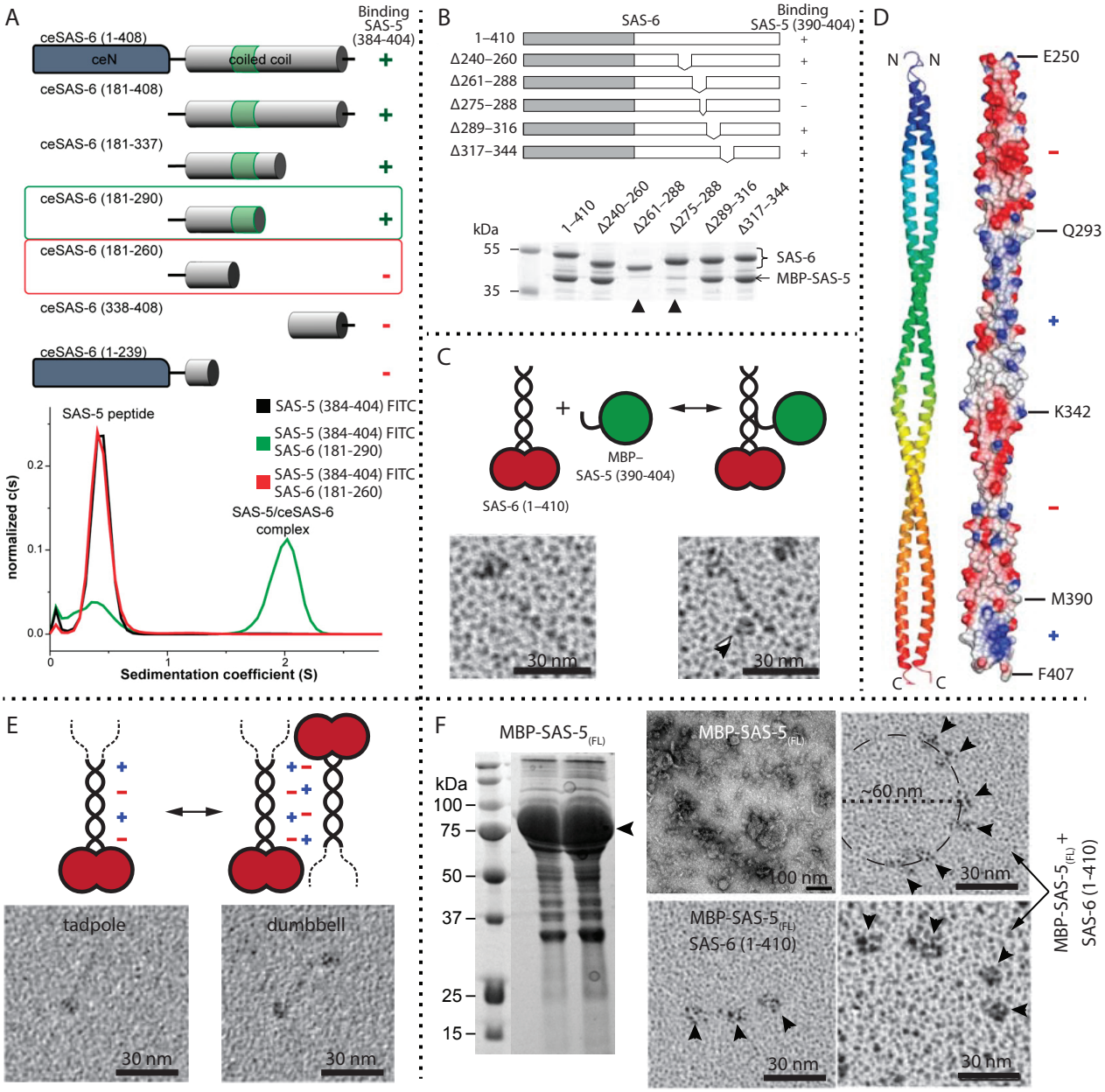


Figure 1.24: **SAS-5 / SAS-6 interaction.** (**A**) After expressing a range of *C. elegans* SAS-6 coiled-coil constructs, Hilbert *et al.* [142] tested them for binding to SAS-$5_{(384\text{-}404)}$ with analytical ultra-centrifugation (AUC) velocity experiments. Fluorescently-labelled (fluorescein isothiocyanate) SAS-$5_{(384\text{-}404)}$ was detected during sedimentation, and revealed the critical site for SAS-5 binding on SAS-6 as residues 261-290 (coloured green). Bottom panel shows a plot of overlaid sedimentation distributions from AUC experiments. Compare sedimentation distributions of SAS-$5_{(384\text{-}404)}$ (black trace), SAS-$5_{(384\text{-}404)}$ + SAS-$6_{(181\text{-}260)}$ (red trace), and SAS-$5_{(384\text{-}404)}$ + SAS-$6_{(181\text{-}290)}$ (green trace). Note that SAS-$5_{(384\text{-}404)}$ forms a complex with SAS-$6_{(181\text{-}290)}$ (green trace), thus its sedimentation co-efficient increases from ∼0.4 S to ∼2 S. (**B**) Qiao *et al.* [275] performed *in vitro* pull-down experiments of MBP-tagged SAS-$5_{(390\text{-}404)}$ by using SAS-$6_{(1\text{-}410)}$ as bait. SDS-PAGE of purified proteins (bottom panel) show two deletion constructs of SAS-6 (Δ261-288 and Δ275-288) that fail in SAS-5 pull-down (arrowheads). (**C**) Rotary shadowing electron microscopy provided visualisation of the binding of MBP-SAS-$5_{(390\text{-}404)}$ to the coiled-coil region of SAS-$6_{(1\text{-}410)}$. apo-SAS-6 normally forms a tadpole shape, but once MBP-SAS-$5_{(390\text{-}404)}$ is added, an extra density adjacent to the coiled coil appears (arrowhead). Continued on the next page.

Figure 1.24: **SAS-5/SAS-6 interaction**. (**D**) Crystal structure of SAS-$6_{(248\text{-}410)}$ represented as a rainbow-coloured cartoon (left), and as electrostatic-potential-coloured surface (right). Note the alternating character of electrostatic potential across the length of the coiled coil. Boundary residues between sections are indicated. (**E**) A likely explanation for the the dumbbell-shaped SAS-6 particles in rotary shadowing EM preparations. Two SAS-6 dimers (tadpoles) bind via alternating electrostatic interactions in an anti-parallel fashion. (**F**) Qiao *et al.* [275] attempted to purify MBP-tagged full-length SAS-5 (MBP-SAS-5_{FL}). SDS-PAGE (left panel) shows partially purified MBP-SAS-5_{FL} (arrowhead). This protein was negatively stained and visualised with electron microscopy - note the aggregated material (top left in the right panel of four micrographs). Qiao *et al.* [275] report that after mixing this preparation of MBP-SAS-5_{FL} with pure SAS-$6_{(1\text{-}410)}$, the aggregates of MBP-SAS-5_{FL} disappeared, leaving an arch-like structure of ∼60 nm in diameter (as seen in rotary shadowing preparations; top right, bottom right and bottom left electron micrographs of the right panel). Figures adapted from [142, 275].

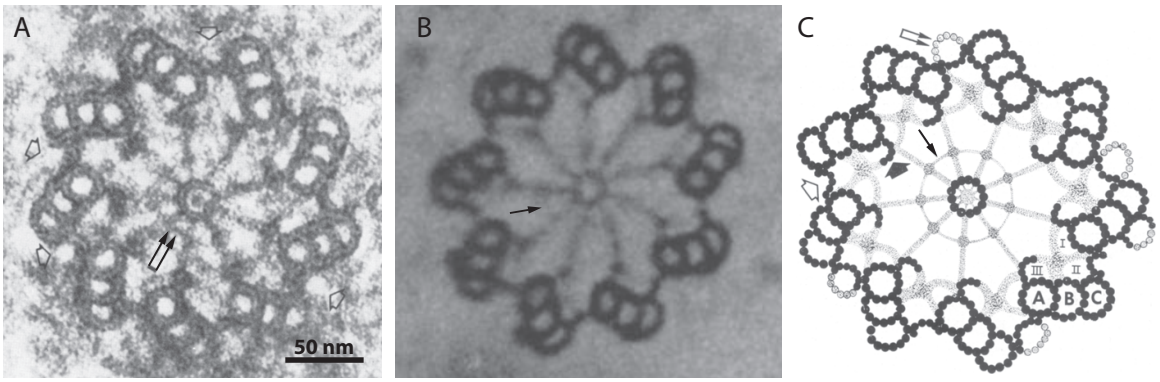


Figure 1.25: **Potential cross-spoke connections in *Phlyctochytrium* and *Trichonympha*.** (**A**) Electron tomogram showing a transverse plane cross-section through a centriole of fungus *P. irregulare*. Double arrow points at the fibre which connects adjacent spokes, approximately 15 nm from the central hub. (**B**) Transverse section of a *Trichonympha* basal body in the proximal region. Arrow points at the faint density linking adjacent cartwheel spokes. (**C**) Schematic drawing based on the transverse section of *P. irregulare* centriole from **A**. Note that in 1970s, it was believed that the central hub was made of a 13-protofilament microtubule, and so did the pinhead and A-C linker densities, but only partially. A single arrow points at the cross-spoke fibre. However intriguing such cross-spoke densities might be, in order to avoid confirmation bias [240], more experimental evidence is needed to verify their existence. Figures adapted from [118, 227].

for SAS-5 binding, indirectly suggesting that there might be another domain within SAS-4 that could potentially bind SAS-5 in tandem with G-BOX. However, these conclusions might have been negatively affected by the state of proteins used for the binding assays. It was assumed that the shorter fragments of SAS-4, expressed *in vitro* in coupled transcription/translation reactions, are perfectly folded proteins, and therefore absence of their bands from the gel suggests lack of binding between SAS-5

and SAS-4. It is not unlikely, however, that the shorter fragments of SAS-4 were aggregated, and therefore did not bind to immobilised SAS-5, explaining the false negative result.

Nevertheless, these studies have shown that the interaction between STIL and CPAP appears to be very extensive, involving the entire length of the G-BOX domain (as seen in solution nuclear magnetic resonance experiments performed by Hatzopoulos *et al.* [140]). Mutations in residues involved in this interaction result in noticeable decrease in binding affinity, with the microcephaly mutations being sufficiently mild to support some degree of binding at cellular protein levels. The molecular details of the SAS-5 / SAS-5 complex in *C. elegans* remain to be elucidated.

Finally, Zheng *et al.* [354] also solved the structure of the SAS-4 G-BOX domain from *D. melanogaster* (2.7 Å resolution), but not in complex with Ana2. In order to compensate for the lack of high-resolution information on the interaction between SAS-4 and Ana2, and also to follow up on their previous work which implied that SAS-4 could act as a scaffold for tethering pericentriolar material [127], they created various mutants of SAS-4/CPAP for use in functional assays in *Drosophila* and human cells as well as in yeast two-hybrid assays. They reported that these mutants show reduced levels of recruitment for various pericentriolar proteins including γ-tubulin, Cep152 and PCNT, and concluded that G-BOX domain of SAS-4/CPAP must therefore be involved in tethering of those complexes to centrioles. However, it appears that in some of the mutant cells in the study by Zheng *et al.* [354], centriole structures themselves were also affected by those G-BOX mutations, suggesting that the centriole assembly process did not complete successfully in the first place. Therefore, it is rather unlikely that the G-BOX of SAS-4/CPAP is directly involved in tethering of those pericentriolar proteins, but instead it is the interaction between SAS-4/CPAP and Ana2/STIL that was affected by those mutations, thus leading to incomplete loading of SAS-4/CPAP onto procentrioles by Ana2/STIL. Indeed, recruitment of PCM is defective in G-BOX domain mutants; however, this is likely an indirect effect caused by the failure in centriole assembly. Yet, since full-length SAS-4 has been shown to be involved in recruiting PCM in various systems, it is likely that a domain other than G-BOX is directly responsible for PCM recruitment. Although the mutagenesis studies did not allow Zheng *et al.* [354] to unambiguously determine which side of

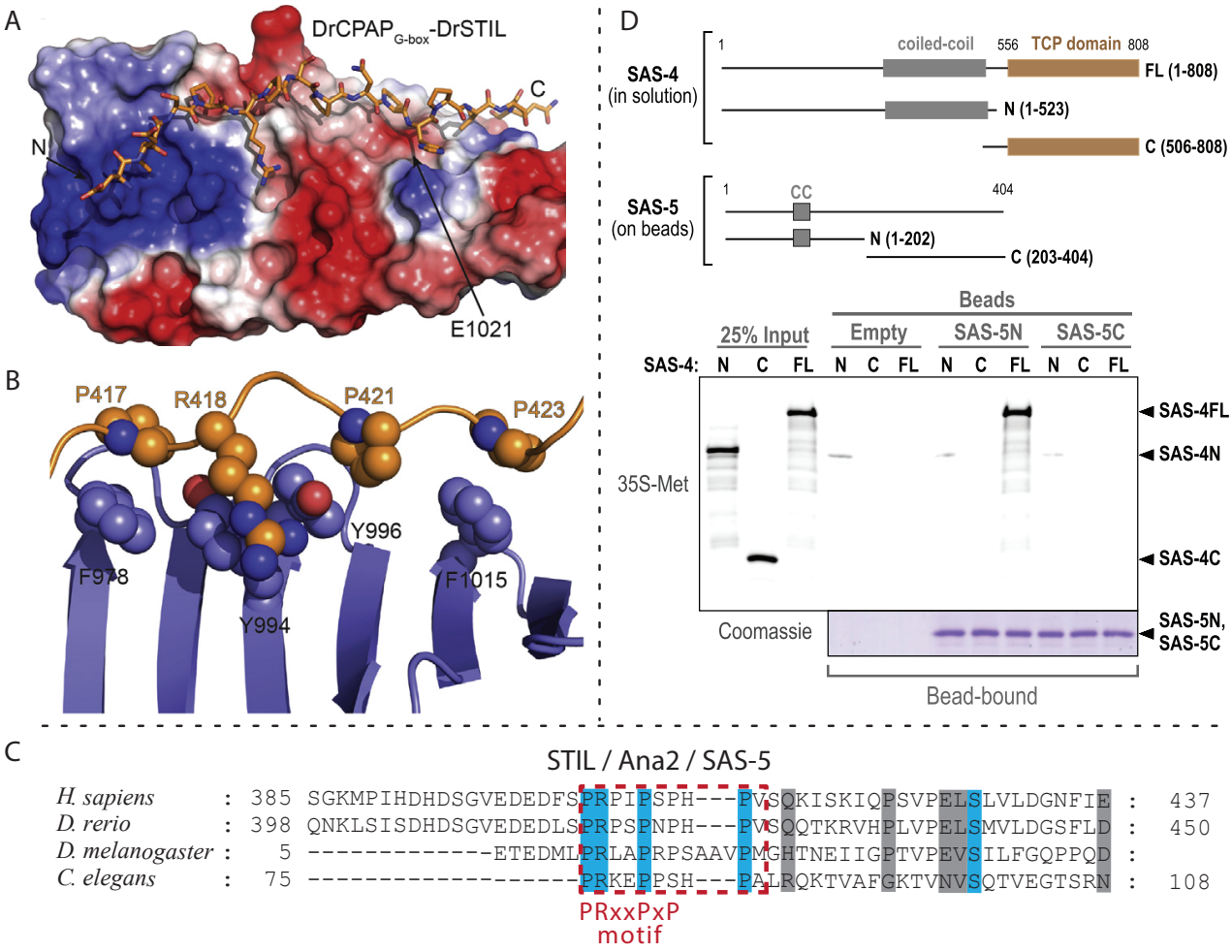


Figure 1.26: **SAS-5 / SAS-4 interaction**. (**A**, **B**) Crystal structure of *D. rerio* CPAP$_{(953\text{-}1106)}$ G-BOX domain in complex with a fragment of *D. rerio* STIL$_{(413\text{-}428)}$. Panel **A** shows CPAP in surface representation (eletrostatic potential colouring) and STIL in sticks (orange, hetero-atom colouring). Position of a glutamate residue (E1021, equivalent of *H. sapiens* E1235) mutated in microcephaly is indicated. Panel **B** depicts the hydrophobic interactions involving the proline-rich motif of STIL, PRxxPxP. Both proteins in cartoon representation (CPAP - purple, STIL - orange); residues forming contacts are shown in space-filling representation, and are coloured by hetero-atom. (**C**) Multiple sequence alignment of the CPAP G-BOX-interacting region of STIL and its relatives: *C. elegans* SAS-5 and *D. melanogaster* Ana2. The proline-rich region is boxed. Columns coloured blue and grey denote universally conserved and highly conserved residues, respectively. (**D**) Interaction studies of *C. elegans* SAS-5 and SAS-4 by Cottee *et al.* [71]. Top panel shows the schematic of construct boundaries used in this experiment. Bottom panel shows SDS-PAGE of the *in vitro* pull-down experiment of two His$_6$-tagged constructs of SAS-5 immobilised on Ni-NTA beads: SAS-5N and SAS-5C (bait). Three constructs of SAS-4 (prey, ^{35}S-Met labelled) were expressed *in vitro* in coupled transcription/translation reactions: SAS-4FL, SAS-4N, SAS-4C. Autoradiograph revealed that out of all SAS-5 / SAS-4 construct combinations, only SAS-5N / SAS-4FL showed binding. Coomassie-stained gel shows bead-immobilised SAS-5. Figures adapted from [71, 140].

the G-BOX domain serves as the binding surface for Ana2, the resulting mutations in the G-BOX domain interfered with binding not only to Ana2, but also to Bld10p (Cep135) - as judged by pull-down assays from *Drosophila* cell extracts (immobilised SAS-4 was used as bait). Mapping these mutations onto the CPAP/STIL (SAS-4/Ana2) structures by Hatzopoulos *et al.* [140] and Cottee *et al.* [71] suggests that Cep135 shares with STIL the same side of the CPAP G-BOX β-sheet for binding, or perhaps that binding of STIL to CPAP must precede binding of Cep135 for the ternary complex between STIL, CPAP and Cep135 to form. Further structural work may provide novel insights into this matter, and perhaps also shed light on the process of tethering microtubules to centrioles and subsequent recruitment of PCM.

1.4.4 Regulation of SAS-5 and SAS-6 proteins

In proliferating cells, new centrioles are made only once per cell cycle. They are also made exactly at the right place - next to a mother. To precisely orchestrate such a unique event, cells employ tight control of protein levels, as over-expression of the essential proteins leads to abnormal centriole numbers, while their depletion completely arrests centriole assembly. SAS-6 was found to be ubiquitinated by SCF-FBXW5, a protein complex composed of an E3-ubiquitin ligase SCF (Skp, Cullin, F-box proteins) and an F-box protein FBXW5 (F-Box/WD repeat-containing protein 5), which directs SCF to SAS-6 [274]. In turn, SCF-FBXW5 is negatively regulated by PLK4 which phosphorylates FBXW5, and effectively suppresses its ability to target SAS-6 for degradation. When the centriole duplication cycle starts at G1-to-S transition, the levels of PLK4 are increased, which lifts inhibition from SAS-6 (by supressing FBXW5), and therefore promotes cartwheel formation. At the end of mitosis, when mother and daughter centrioles have disengaged, SAS-6 is being targeted for degradation by yet another E3-ubiquitin ligase - APC/C. In fact, both STIL and SAS-6 contain the KEN box - a short destruction motif [13,316], recognised by Cdh1 and Cdc20 - the coactivators of APC/C [55]. Targeted degradation of both SAS-6 and STIL, between mitosis and G1, restricts the procentriole formation to only one event per cell cycle.

The cartwheel is assembled for the purpose of scaffolding the radial symmetry of centrioles. This is particularly apparent in mammalian cells, where once the symmetry

is in place and the microtubule triplets are fully grown, the cartwheel is no longer needed to hold the centriole together. The major cartwheel components, SAS-6 and STIL, are progressively dissociated from mitotic centrosomes [13]. It is likely that STIL is a direct substrate of Cdk1 which becomes activated following the nuclear membrane breakdown, and may hyperphosphorylate STIL [13, 48]. Once out of the centrosome, SAS-6 and STIL, are monoubiquitinated by APC/C before the mitotic exit [13, 14]. Interestingly, Arquint and Nigg [13] noticed that microcephaly nonsense mutations, which truncate STIL by either 49 or 69 amino acids at the C-terminus (TIM motif) [193], also remove the KEN box located in that region. These versions of STIL were resistant to APC/C-targeted degradation, leading to accumulation of the protein and thus centriole amplification.

Besides the control of protein levels throughout the cell cycle, an important aspect of regulating centriole biogenesis is targeted localisation of newly translated proteins to the site of a growing procentriole. *C. elegans* kinase ZYG-1 (Sak in *Drosophila*, PLK4 in vertebrates), acts upstream from SAS-5 and SAS-6 in centriole biogenesis, and promotes cartwheel assembly [27, 134, 186, 249]. The exact mechanistic role of ZYG-1 in this pathway, however, has seen some controversy. Kitagawa *et al.* [180] reported that ZYG-1 phosphorylates SAS-6 at S123 *in vitro*, which they found was crucial for centriole formation *in vivo* and for SAS-6 maintenance at the *C. elegans* central tube. Later, Lettman *et al.* [203] divided this process into two separate events. They showed that in the first step, ZYG-1 physically binds to SAS-6, which in turn leads to targeting of SAS-6/SAS-5 complex to centrioles. In the second step, through enzymatic activity of ZYG-1, the cartwheel is formed. In contrast to study by Kitagawa *et al.* [180], Lettman *et al.* [203] found that phosphorylation of SAS-6 at any potential site is dispensable for centriole assembly. Lettman *et al.* [203] speculated that instead of targeting SAS-6 for phosphorylation, ZYG-1 could possibly target SAS-5, as binding of SAS-5 at the coiled coil of SAS-6 is close to where ZYG-1 also binds (SAS-5 binds to SAS-6 residues 261-288, while ZYG-1 binds to SAS-6 residues 210-255). They were not able to explain, however, why binding of SAS-6 to SAS-5 is needed for SAS-6 recruitment to the mother centriole. It was recognised that testing SAS-5 / SAS-6 relationship in biochemical assays has been limited by the lack of available recombinant full-length SAS-5. This puzzle remains as an important future

goal in centriole biology.

Another line of evidence for SAS-5 control points to the catalytic subunit (LET-92) of the protein phosphatase 2 (PP2A), which had been shown to interact with SAS-5 (via SAS-5's C-terminal half), while the other subunits of PP2A: PAA-1 (structural) and SUR-6 (regulatory) interact with SAS-6 instead [181, 310]. A ternary complex involving SAS-5/SAS-6 and all subunits of PP2A was also observed. Kitagawa *et al.* [181] showed that PP2A dephosphorylates SAS-5 (but not SAS-6), and that this event promotes targeting of the SAS-5/SAS-6 complex to centrioles. Furthermore, Song *et al.* [310] proposed that such dephosphorylation of SAS-5 by PP2A protects SAS-5 and ZYG-1 from degradation, and therefore by controlling levels of those proteins, PP2A also promotes centriole duplication.

What is the centriole targeting mechanism of SAS-5 and SAS-6 in other systems? Two studies reported that at the G1-to-S transition in human cells, SAS-6 localises transiently to the mother centriole, before it can be released towards the outer wall of the mother, where it will co-localise with STIL and initiate procentriole growth through cartwheel assembly. The authors of those studies, however, report two different mechanisms for such transient localisation of SAS-6. Keller *et al.* [170] observed that SAS-6 accumulates into a torus structure around the base of the mother centriole, while Fong *et al.* [109] detected transient localisation of SAS-6 inside the lumen of the mother centriole instead. Interestingly, SAS-6 was shown to remain in the cytoplasm primarily as a homodimer, while its oligomerisation into a ring occurred only at centrioles [170]. The estimated concentration of SAS-6 in the cytoplasm was $\sim$0.1 μM, and up to 10 μM at centrioles. Furthermore, Fong *et al.* [109] have found that PLK4 and STIL are required for release of SAS-6 that localised to the mother centriole lumen, which poses a question whether these proteins could mimic the interactions seen earlier in *C. elegans*.

With a number of recent studies, great progress has been made in understanding the link between PLK4, STIL and SAS-6 [98, 191, 231, 252]. STIL/Ana2 forms a complex with PLK4 through its coiled-coil domain. This interaction leads to activation of PLK4 by promoting self-phosphorylation of the auto-inhibitory loop near the kinase domain of PLK4 [185, 231]. Once the kinase is activated, PLK4 phosphorylates STIL/Ana2 at multiple sites, many of which are located within the conserved STAN

motif. Phosphorylation of residues S1108 and S1116 (*H. sapiens* STIL; S370 is the *D. melanogaster* Ana2 equivalent of S1116) leads to increased centriole targeting of STIL. Furthermore, phosphorylated STAN motif allows for binding of STIL to SAS-6, which in turn triggers cartwheel formation.

Many interesting questions remain. What is the mechanisms of SAS-6 release from the torus/lumen of mother centriole? How is PLK4 ring condensed into a dot? How does the interaction between STIL and SAS-6 lead to cartwheel assembly?

1.5 The aim of this work

Centriole assembly is a tightly regulated process, achieving duplication only once per cell cycle at a precise location in the cell. Abnormal centriole numbers lead to developmental defects and cancer. Several hundred proteins are involved in the regulation of this process, but only a handful of them are absolutely essential for making a new centriole. Protein called SAS-6 is involved in seeding the growth of new centrioles by constructing a cartwheel-like structure. However, timely nucleation of the cartwheel in nematode worms *Caenorhabditis elegans* also requires another protein, called SAS-5. It forms a complex with SAS-6, and this interaction is needed to target both proteins to the site from which a new centriole is to emerge. Phosphorylation events have been shown to critically regulate involvement of SAS-5 in centriole assembly; however, the exact mechanistic role of SAS-5 in this process is not clearly understood. Depletion of SAS-5 in worm embryos leads to a failure in centriole assembly and therefore a lethal phenotype. This limits our ability to explain the function of SAS-5 *in vivo*.

Structural biology has been of vital importance in understanding the molecular mechanisms underlying the processes that lead to centriole duplication. In my doctoral work, I set out to dissect SAS-5 with biochemical and biophysical methods in order to build an understanding of its structural architecture. Knowing the structure of SAS-5 would allow me to alter its chemistry through mutagenesis, and therefore open new avenues for probing its function in *C. elegans* embryos.

Chapter 2

Solubilising SAS-5

Large quantities of purified protein, which is well-behaved and non-aggregated, are normally required for biophysical and structural work. While a method for recombinant expression and purification of *C. elegans* SAS-6 was well-established in the lab [142, 183], its binding partner - SAS-5, had been notoriously difficult to obtain. A number of labs (including ours) have attempted expression of full-length SAS-5 (residues 1-404) in various systems, including bacteria, insect and mammalian cells [203, 275]. However, they arrived at a common conclusion: SAS-5 is expressed in insoluble form, and even if marginally soluble, the protein product is aggregated, and therefore unsuitable for biochemical and structural work. This difficulty has hindered the progress of understanding the function of SAS-5, the exact role of which in centriole assembly is still unknown since its discovery in 2004 [78, 88, 296]. In this chapter, I will describe the groundwork which allowed me to address this problem.

2.1 Construction of pFloat expression system

At the start of my DPhil in 2012, there were no reports of any attempts to study SAS-5 with biochemistry. While in both our and collaborating labs, very short fragments of SAS-5 (up to 50 amino acids) were shown to express in soluble form, any attempts at acquiring full-length SAS-5 were plagued with insoluble protein. In order to effectively address this problem, we decided to go back to the drawing board.

In theory, obtaining pure recombinant protein is fairly straightforward. A gene is cloned into an expression vector, which is then transformed into a host. Expression of the gene is then induced, and the protein is ready for purification. In practice,

however, obstacles can occur at any stage, and the later they do, the more time is wasted. To avoid this dissatisfying scenario, we focused on improving the early stages of this process, so that a number of options can be tried simultaneously. By exploring broader set of conditions, we hoped to escape the bottleneck of no expression or no solubility. To aid this endeavour, we constructed a customised line of entero-bacterial expression vectors that we called *pFloat* (Figure 2.1 on page 66).

2.1.1 The host system

Beyond phosphorylation, *C. elegans* SAS-5 has never been reported to undergo more complex post-translational modifications, such as glycosylation. Therefore, we believed that even though SAS-5 is an eukaryotic protein, expressing it in a prokaryotic host was not unreasonable.

Recombinant protein expression in *Escherichia coli* is by far the most cost-effective route of acquiring large quantities of protein material for biophysical work. *E. coli* has extremely fast growth kinetics, with doubling time of 20 to 30 min in rich media. Moreover, high cell densities per litre of liquid culture can be readily achieved ($<$1 x 10^{10} cell / mL) [299] with rich complex media, composition of which is relatively inexpensive. There are numerous bacterial strains available for recombinant protein expression. The most widely used strain - BL21(DE3) [77], similarly to other paternal B strains, does not carry genes for Lon and OmpT proteases, which are known for active degradation of foreign proteins and can wreak havoc once the cells are lysed. Moreover, in BL21(DE3) strain, the host specificity gene *hsdS* is removed, which normally regulates the function of the endonuclease HsdR and the methylase HsdM enzymes. The system of these three proteins methylates DNA at specific sequences and cleaves DNA which is not methylated. Therefore, disabling *hsdS* prevents plasmid loss. Finally, BL21(DE3) carries a copy of the bacteriophage T7 RNA polymerase (RNAP), which is installed under transcriptional control of *lac*UV5 promoter. *lac*UV5 is a stronger derivative of the wild-type *lac* promoter, and is insensitive to repression by high levels of glucose, therefore suitable for growth in rich media [49]. We also favoured BL21(DE3) over other *E. coli* hosts, mainly due to the capability of the T7 expression to give target protein yields close to 50% of the total protein mass in the cell.

Finally, the codon bias of eukaryotic *sas-5* was expected to be completely different than that of *E. coli*, and thus in order to avoid any ribosome stalling, we used the re-coded version of *sas-5* (devoid of any rare bacterial codons; see Appendix Section B.2 on page 185 for the exact sequence) in our expression trials.

2.1.2 Expression machinery

Choosing BL21(DE3) as a host, with T7 RNAP-led expression [95, 318], requires certain regulatory sequences to be encoded by the plasmid carrying target genes. These include T7 promoter and terminator, as well as *lac* operon (Figure 2.1 on page 66). *lac*UV5 promoter, which controls expression of the T7 RNAP in bacterial genome, can give rise to a significant level of basal expression of T7 RNAP, even in the absence of the inducer. This in turn leads to basal expression of target genes, and in case when the target genes are toxic, to stalled growth of bacteria and lower proteins yields. Also, if the target protein is easily degradable, such basal expression can result in higher amounts of truncation products. Therefore, to minimise such basal expression, two sequences are needed in the expression plasmid: *lac* repressor (LacI) and *lac* operator (*lacO*). *lacO* sequence is normally inserted between the T7 promoter and the target gene, so that LacI protein, which specifically binds to the *lacO* DNA element, can suppress the so-called "leaky" expression of target genes by physically blocking the access of T7 RNAP to target gene sequences (Figure 2.2 on page 67).

Sequences of T7 promoter, T7 terminator, *lac* operator and *lac* repressor have already been put together in a commercial vector system called pET (from Novagen), and therefore served as a good starting backbone on top of which a new pFloat vector could be built. Besides the above-mentioned T7 regulatory sequences, pET contains the pBR322 and f1 origins of replication (*ori*), RNA I / RNA II and repressor of primer (*rop*) (Figure 2.1). These extra elements constitute regulatory machinery for copying the plasmid and for maintaining its low copy number within bacterial cells. These elements were originally copied into pET from the classic pBR322 plasmid, which belongs to the ColE1 family of plasmids, and was constructed by Francisco Bolivar and Raymond Rodriguez in 1977 [33]. ColE1 plasmids use RNA species for initiation of replication, and RNA-RNA interactions for a steady copy number

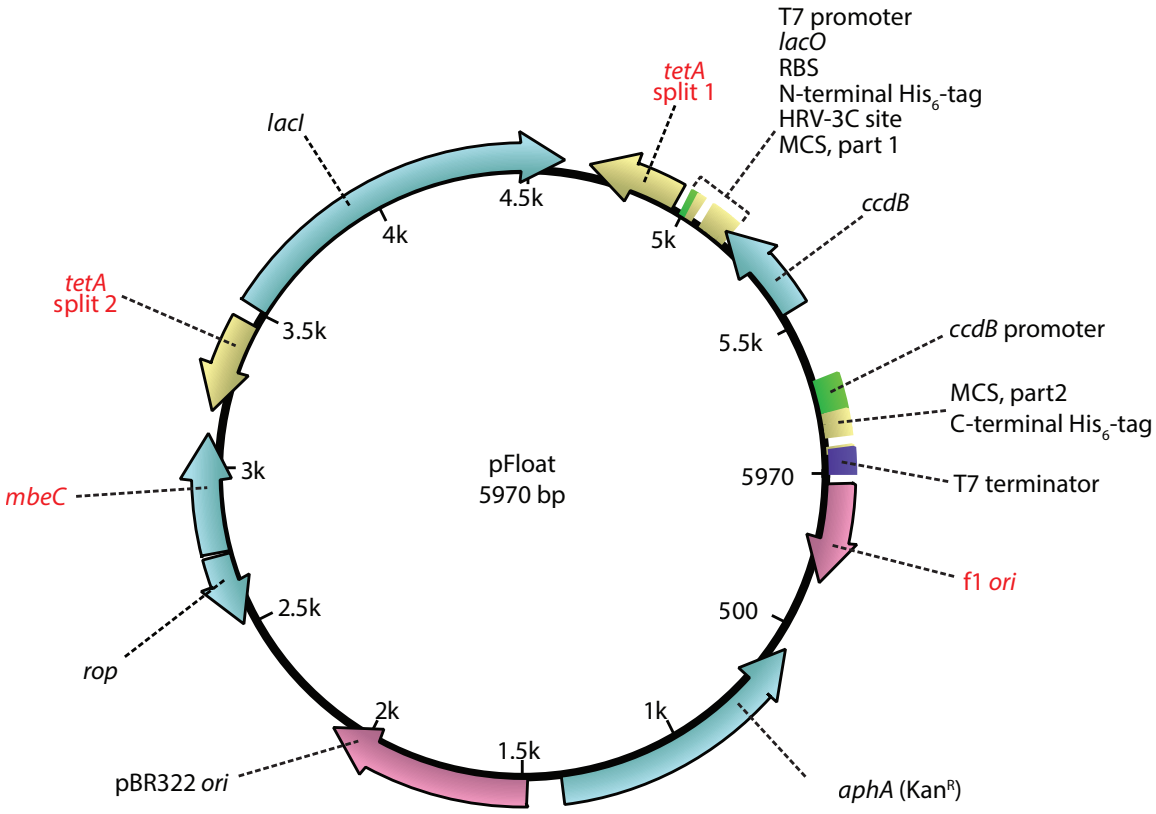


Figure 2.1: **Vector map of pFloat.** The basic pFloat vector is 5970 bp long and includes a number of regulatory sequences which drive and regulate plasmid replication (pBR322 *ori*, repressor of plasmid - *rop*), transcription (T7 promoter, T7 terminator, lac operator - *lacO*), lac repressor - *lacI*) and protein expression (ribosome binding site - RBS, TATA-box). pFloat also carries a kanamycin-resistance cassette with the gene for aminoglycoside 3'-phosphotransferase (*aphA* or *kanR*). In order to facilitate quicker positive colony screening, pFloat includes the control of cell death (*ccd*) cassette which is inserted inside of the multiple cloning site (MCS) and includes a gene promoter and a toxin/anti-toxin gene system: *ccdA* and *ccdB*. The *ccdA* anti-toxin gene carries a nonsense mutation, which truncates the CcdA gene product prematurely, thus allowing the CcdB toxin to act as a selection marker for cloning. pFloat also contains some sequence fossils which have been carried over from the original vectors used for its assembly (red labels). These include f1 *ori*, *mbeC*, and split tetracycline efflux pump *tetA*. The vector map was drawn using UGENE [253].

maintenance, of 15-20 plasmid copies per cell. Once a long RNA primer (555 nt), RNA II, is transcribed from RNA II promoter by host RNA polymerase, it strongly hybridises with the plasmid *ori* (Figure 2.3 on page 68). This hybrid is recognised by RNase H, which cleaves off a part of RNA II, leaving a free 3' hydroxyl group. This creates a primer for the replication of the entire plasmid by host DNA polymerase I [272].

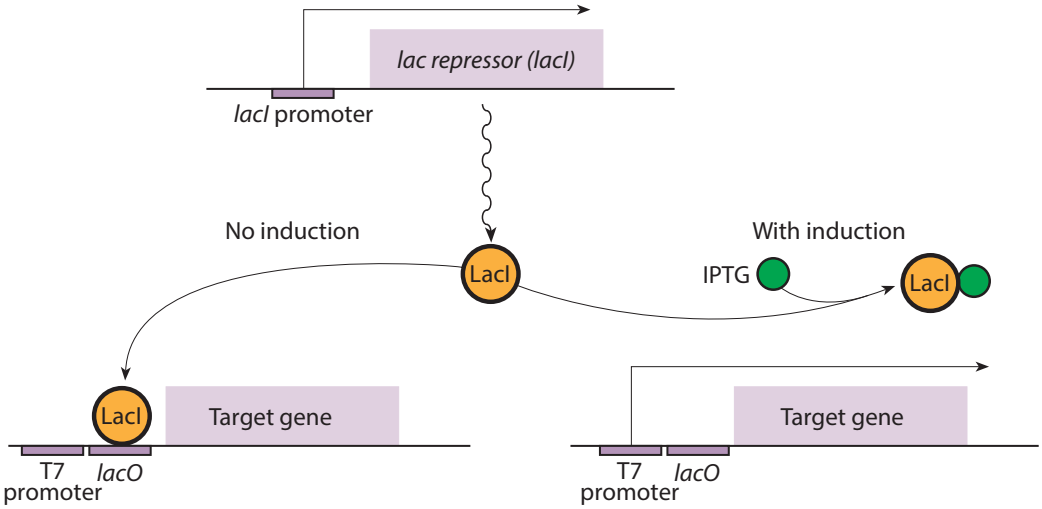


Figure 2.2: **Gene expression circuit based on the *lac* operon.** pFloat vectors use inducible gene expression system based on the *lac* operon and T7 transcription machinery. Basal expression of the target gene from the *lac* operon is switched off by a repressor protein LacI (encoded by the *lacI* gene). LacI specifically binds to the operator sequence which lies immediately downstream of the T7 promoter, thus effectively inhibiting transcription of the target gene by T7 RNAP. Supplementing growth medium with lactose or its non-metabolisable analogue isopropyl-β-D-thiogalactoside (IPTG) promotes dissociation of LacI from *lacO* (IPTG physically interacts with LacI), thereby allowing transcription to commence. Figure adapted from [272].

The regulation of ColE1 plasmid copy number is facilitated by RNA I, a short RNA countertranscript (108 nt) of RNAII. Because RNA I is fully complementary to a part of RNA II sequence, the interaction between RNA I and RNA II leads to formation of a double stranded RNA helix. This duplex prevents RNase H from processing RNA II, and therefore plasmid replication no longer takes place. Because RNA I is encoded by the plasmid, replication of which is negatively regulated by RNA I, more plasmid copies result in more RNA I being transcribed. When bacteria grow and divide, the concentration of RNA I falls, thus lifting repression from RNA II and allowing for plasmid replication to continue. Such negative feedback loop defines the average copy number of plasmid per cell. In addition to RNA I as a negative regulator of replication, ColE1 encodes a small protein called Rop (63 amino acids). Rop forms a dimeric α-helical bundle which enhances affinity between RNA I and RNA II, so that processing of RNA II is blocked even at low concentrations of RNA I (Figure 2.3 on page 68). Deletions or mutations in *rop* result in increased plasmid copy numbers [272]. We have incorporated all these elements into pFloat.

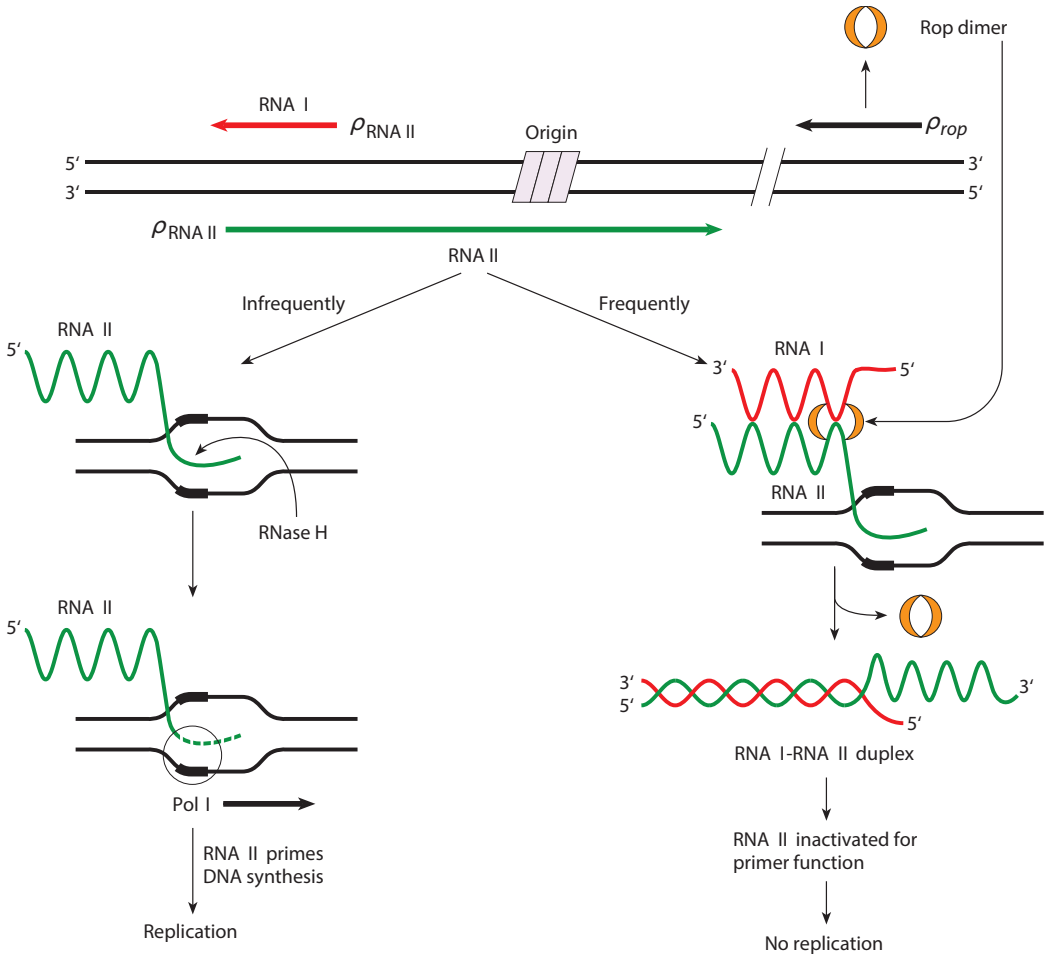


Figure 2.3: **Plasmid replication control.** ColE1-derived plasmids, such as pBR322, pET or pFloat, rely on RNA species for initiation of their replication, and RNA-RNA interactions for maintenance of their low copy number (∼15 plasmid copies per cell). To initiate plasmid replication, RNA II transcript (green) is processed by RNase H to serve as a primer. However, replication priming by RNA II is controlled by RNA I (red, countertranscript of RNA II) which forms an RNA duplex with RNA II, thus inhibiting the process and regulating plasmid copy number. Binding affinity between RNA II and RNA I is enhanced by binding of Rop (repressor of primer) protein dimer (orange crescents), which further decreases plasmid copy number. $\rho_{\mathrm{RNA\ II}}$, $\rho_{\mathrm{RNA\ I}}$ and ρ_{rop} are promoters for RNA II, RNA I and *rop*, respectively. Figure adapted from [272].

2.1.3 Selection markers for cell growth

Stopping growth of plasmid-free cells is usually achieved by inclusion of a selection marker in the plasmid backbone. The most popular form of selection in the *E. coli* system are antibiotic resistance genes.

Ampicillin is probably the most widely used antibiotic for this purpose, as it prevents cross-linking of peptidoglycan chains in bacterial cell wall. Bacteria growing

in presence of ampicillin produce weakened cell walls and eventually burst due to high internal osmotic pressure. Resistance to ampicillin is conferred by β-lactamase, an enzyme which hydrolyses β-lactam ring of any penicillin-type antibiotics, including ampicillin. However, *E. coli* also secrete translated β-lactamase into the growth medium, and therefore continuous expression of β-lactamase leads to degradation of the antibiotic in a couple of hours, allowing cells which do not carry the plasmid to increase in number during cultivation.

Thus, instead of ampicillin, we decided to choose kanamycin as the primary antibiotic. Resistance to kanamycin is conferred by cytoplasmic enzyme aminoglycoside phosphotransferase (Figure 2.1 on page 66) which covalently phosphorylates kanamycin, and unlike β-lactamase, it is not secreted to growth medium. Moreover, kanamycin is a strong bactericidal agent which binds to the 30S subunit of the ribosome, and actively inhibits its translocation step during protein synthesis, which leads to stalled protein production and cell death.

2.1.4 Fusion tags

The use of fusion tags to aid protein purification is relatively common in any modern structural biology laboratory. Specialised purification resins are available which can bind to target tags. However, besides serving as a purification aid, a good fusion tag also provides means for detection, and increases yields of target proteins. In some cases, it may be difficult to distinguish the target protein from other *E. coli* proteins with SDS-PAGE, if over-expression of the target gene is not overwhelmingly large. In such cases, a fusion tag can be detected by Western blotting, as commercial antibodies are available for most of the tags. Moreover, some fusion tags are extremely soluble in aqueous solutions themselves, and therefore can assist in solubilising target proteins. A number of such solubility enhancers has been described in the literature, with characteristic negative charge (theoretical pI between 3 and 6) as the common theme between them. Such high negative charge allows them to reach concentrations in solution beyond 100 mg/mL (for review see [178]). Large size of these enhancers, however, can lead to misjudgement of the effective yields of the target protein, as larger proteins stain better in SDS-PAGE.

With pFloat, we aimed to create a system of multiple vectors with compatible cloning properties, carrying a wide range of fusion tags, so that testing for soluble protein yields can be achieved in a high-throughput manner. The basic pFloat (Figure 2.4 on page 72) contains a hexahistidine (His_6) tag which serves as a purification aid for immobilised metal-affinity chromatography (IMAC) [145, 160, 208, 268]. His-tagged proteins are normally purified through immobilisation on a highly cross-linked nitrilotriacetic acid (NTA) agarose matrix loaded with transition metal ions: Ni^{2+}, Co^{2+}, Cu^{2+} or Zn^{2+} [146, 267]. These ions bind to NTA with four coordination sites, and leave the remaining two sites in the coordination sphere free to associate with the imidazole ring of histidine residues, effectively forming a specific and tight interaction at physiological conditions (for review see [32]).

To streamline solubility assessment and the purification process, each of the pFloat vectors contains a His_6-tag, followed by another tag which either increases solubility or allows for fluorescence detection (Figure 2.4 on page 72 and Appendix Section B.1.2 on page 173). We have incorporated the following tags into pFloat:

1. His_6-tag (pFloat; both N- and C-terminal fusions) [37]
2. S-tag (pFloat-STAG) [176]
3. Glutathione S-transferase (pFloat-GST) [308]
4. Thioredoxin (pFloat-TRX) [198]
5. Colicin E9 immunity protein (pFloat-Im9) ([184], Kleanthous Laboratory - personal communication.)
6. Maltose-binding protein (pFloat-MBP) [89]
7. Small ubiquitin-like modifier (pFloat-SUMO) [217]
8. Haloalkane dehalogenase (pFloat-HALO) [251]
9. Trigger factor (pFloat-TF) [30, 104, 243]
10. N utilisation substance protein A (pFloat-NusA) [82]
11. Acidic protein MsyB (pFloat-MsyB; both N- and C-terminal fusions) [319, 358]

12. Yellow fluorescent protein (pFloat-YFP; mVenus) [236]

13. Enhanced green fluorescent protein (pFloat-eGFP) [66, 326]

2.1.5 Tag removal

Once the solubility tag has improved protein yields and the His_6-tag helped with purification, these tags are usually no longer needed in downstream biophysical applications. Removal of tags is particularly important for structural studies where a tag can adversely affect crystallisation of proteins, although some reports suggest that if suitably engineered, large tags can aid the crystallisation process [69, 309]. In order to allow for flexibility of choice: to produce a protein with a tag or without it, we have included a short recognition sequence (LEVLFQGP) between the fusion tag and the target gene (Figure 2.4 on page 72). This sequence is specifically recognised by human rhinovirus 3C protease (HRV-3C) [64, 65, 222]. The high specificity and processivity of HRV-3C, coupled with a rather simple protocol for in-house production of recombinant His_6-tagged HRV-3C, were the decisive factors in choosing HRV-3C for this role. HRV-3C cleaves between Q and G residues of the recognition sequence, leaving only two amino acids (GP) as an artefact of digestion (in case of the N-terminal fusion).

There are other proteases available for enzymatic digestion of proteins, such as tobacco etch virus (TEV) protease [169, 257], thrombin [161], enterokinase [59] and factor Xa [161], as well as methods for chemical digestion [307]; however, in our experience, neither of them is as processive and cost-effective as HRV-3C.

Besides the standard pFloat, we have also created the pFloat.fuse vector line as an aid in crystallisation of difficult proteins which cannot otherwise be crystallised alone. Three tags, MBP, SUMO and TRX, were employed as permanent fusion partners (see Appendix Section B.1.4 on page 183). In order to avoid any repulsive forces between MBP and the target protein, highly charged C-terminal residues were mutated to alanine. In these vectors, the HRV-3C recognition site has been moved between the His_6-tag and the MBP/SUMO/TRX tag, so that after purification, only the His_6-tag and not the crystallisation tag is removed by enzymatic digestion.

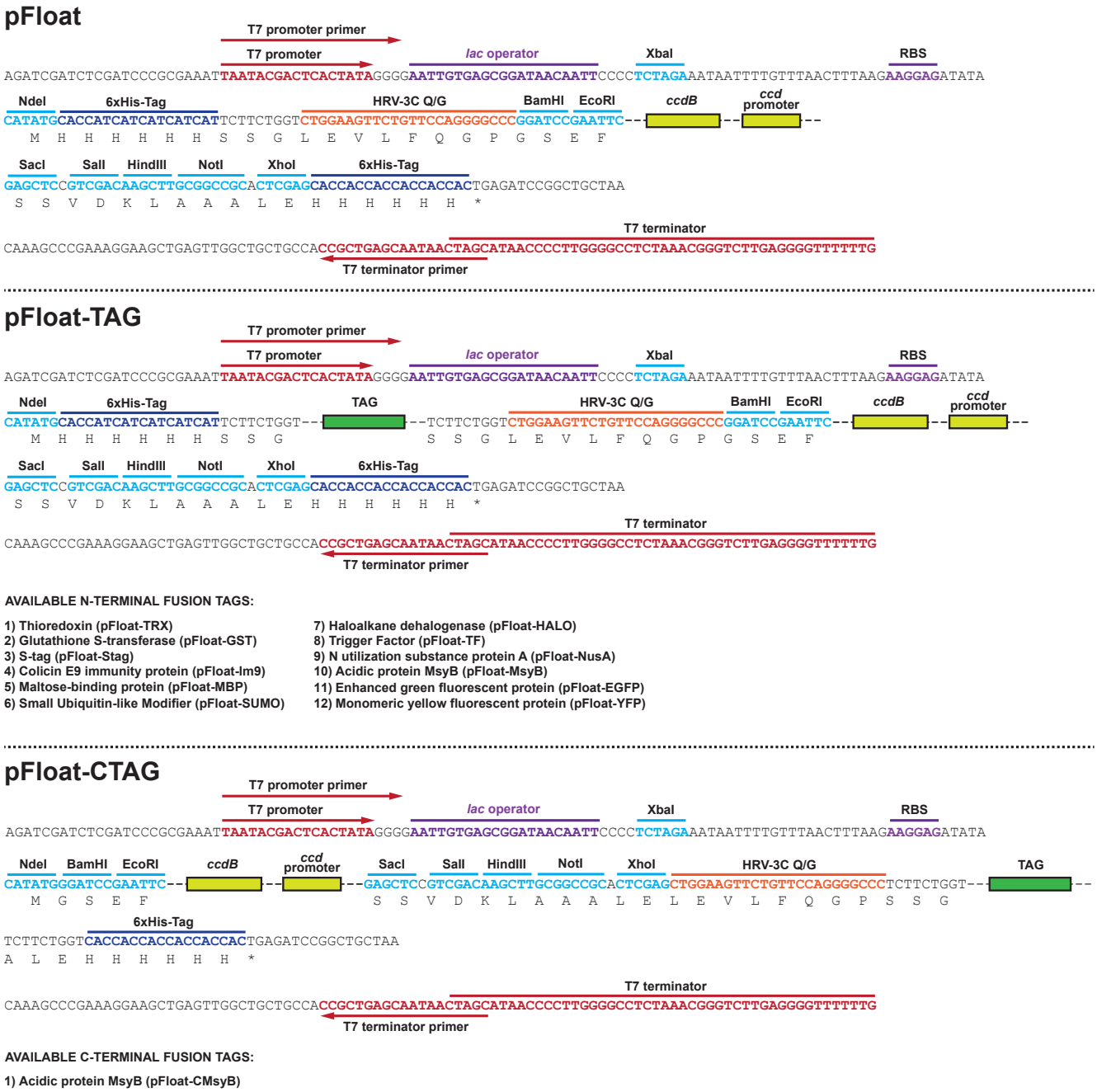


Figure 2.4: **Outline of the pFloat vectors.** pFloat, pFloat-TAG and pFloat-CTAG vectors use T7 promoter/terminator to drive transcription of target genes; positions of hybridisation for T7 primers used in sequencing are indicated. *lac* operator is inserted between the T7 promoter and the ribosome binding site (RBS). The open reading frame starts at the NdeI site (first methionine after the TATA box), and follows with the N-terminal His_6-tag and a HRV-3C cleavage site (LEVLFQGP - cleavage between Q and G); both sequences are separated by a flexible linker (SSG). Multiple cloning site includes a number of sites recognised by rare restriction enzymes, and also a *ccd* cassette with *ccdB* toxic gene and *ccd* constitutive promoter. An optional non-cleavable C-terminal His_6-tag is also available. Typically, a target gene of interest is inserted adjacent to the HRV-3C site, with a STOP codon at the end, so that the output gene product looks as follows: (His_6)-SSG-(3C site)-(target gene)-STOP. TAG and CTAG versions of pFloat offer expression of an extra fusion tag to either provide detection or enhance protein yield and solubility. Available tags are listed, and their positions in the expression construct are indicated.

2.1.6 Selection marker for cloning

Parallel cloning of target genes into multiple expression vectors can be a tedious exercise. However, once the cloning is completed, screening for *E. coli* colonies carrying the correct clone takes up even more time. To address this issue, we have employed an extra selection marker to streamline this process. *ccd* (control of cell death) is a toxin/antitoxin module which codes for CcdB poison and the CcdA antidote proteins [24, 229]. CcdB targets DNA gyrase, an essential bacterial topoisomerase responsible for disentangling DNA before replication and transcription. CcdB binds gyrase with nanomolar affinity, either in its free form, or while it is associated with DNA. The former leads to inactivation, while the latter leads to poisoning by forming a roadblock for polymerases [79]. Both poisoning and inactivation can be rescued by CcdA which actively dissociates CcdB-gyrase complex by binding to CcdB with picomolar affinity, in the process called rejuvenation [85]. Alternatively, the R462C mutation in gyrase GyrA subunit renders gyrase CcdB-resistant.

To speed up our cloning efforts, we have placed a *ccd* cassette within the multiple cloning site (MCS) of pFloat (Figure 2.1 on page 66 and Figure 2.4 on page 72). The cassette carries both *ccdA* and *ccdB* genes under control of the wild-type constitutive promoter. However, the *ccdA* coding sequence is interrupted by a a number of mutations rendering the gene product inactive. During cloning, a target gene is inserted into the MCS, and therefore replaces the *ccd* cassette, effectively eliminating any possible expression of the *ccdB* gene. Therefore, cells transformed with plasmids that incorporated the target gene are viable (insertional inactivation of *ccdB*), while the other cells, transformed with empty plasmids, express poisonous CcdB and perish. This method of cloning leaves no background and is 100% positive (minus potential rare nonsense mutants that can inactivate the *ccdB* gene). To propagate the pFloat plasmids, CcdB-resistant cells are used, *e.g.* DB3.1 strain which carries the gyrA R462C mutation.

2.1.7 Sequence fossils

On top of useful elements, pFloat contains opportunist hitch-hikers which we have not eliminated yet (Figure 2.1 on page 66). Although harmless, they take up unnecessary space on the plasmid, and may contribute to increased chances of replication errors.

These include f1 *ori*, which is more of a fossil and an artefact of cloning the whole chunk of pET backbone rather than a useful feature. Prior to the development of cycle sequencing, f1 origin was used for production of single-stranded DNA template for use in sequencing. Today, this feature is not of much benefit any more.

Yet another artefact of cloning is the split tetracycline resistance gene (tetracycline efflux pump), flanking the *lac* repressor (*lacI*) cassette. It is likely that the authors of the pET vector inserted the *lacI* cassette into the pBR322's tet^R gene to accelerate their cloning efforts, as rendering cells tetracyclin-sensitive allows for more efficient colony screening.

Finally, an even older fossil dating all the way back to the ColE1 plasmid, is the *mbeC* gene. This gene is a part of the ColE1 *mob* region, responsible for coding a full complement of mobilisation genes which are involved in the preparation of a plasmid to be transferred from one bacterial cell to another, in the process called bacterial conjugation [41, 336]. Conjugation is normally disabled in cloning vectors through removal of parts of the mobilization machinery. pBR322 has only *mbeC* left, and therefore avoids accidental recombinant DNA transfer to wild bacterial strains.

The prototype pFloat vector has all these fossil genes present - an artefact of plasmid ancestry. The future versions of pFloat, however, will have these regions removed, in an effort to create a minimal expression vector, with only necessary elements in place.

2.1.8 pFloat outlook

Although hundreds of various vectors were available for recombinant protein expression in *E. coli*, there had been no vector set which would support efficient cloning and high-throughput screening for protein solubility. The speed of screening for different tags had been hindered by slow cloning protocols which would require designing custom primers for each vector. In the last decade, with the advent of modern cloning methods, such as Gateway, Gibson Assembly, Ligation Independent Cloning, In-Fusion and Restriction Free cloning, a number of commercial and non-commercial groups have tried to address this question. They created suits of vectors for high-throughput cloning and screening, where target insert could be introduced into multiple vectors simultaneously [25, 30, 47, 68, 76, 96, 285, 298]. The layout of these vectors,

however, did not always satisfy the needs of our lab, and would require multiple changes to be adapted to our needs. These included plasmid copy number, exclusive expression in bacterial hosts to minimise the plasmid size, selection markers, proteolytic cleavage sites, fusion tags and flexible spacers between protein features. Instead, we decided to take a step back and create a vector set from scratch, with all the elements we deemed crucial for successful solubility screening and further structural work. This is how pFloat was born. Simple design of the basic pFloat vector allowed us to build on it by either adding extra tags for difficult purifications, or customising other features specifically required for the task in hand.

In summary, the current version of pFloat (Figure 2.1 on page 66) has a pBR322 replicon (low-copy number), is kanamycin resistant, and uses T7 expression system. The expression cassette is flanked by T7 promoter and terminator, with *lac* operator serving as a leaky expression controller. The polylinker site includes an N-terminal His_6-tag and a HRV-3C cleavage sequence (Figure 2.4 on page 72). Target proteins can be inserted with seamless cloning methods, using HRV-3C site to hybridise the forward primer, and any sequence before the T7 terminator to hybridise the reverse primer. MCS is also present for restriction enzyme cloning. Colony screening is no longer required due to positive-selection provided by the *ccd* cassette. Such basic pFloat vector (6.0 kbp) was expanded to also include a wide range of purification, fluorescence and solubility-enhancer tags, and also remodelled to incorporate C-terminal fusions, as well as permanent fusions - for crystallisation purposes. See Appendix Section B.1 on page 170 for all pFloat sequences.

2.2 Expression and purification of SAS-5_{FL}

Once the construction of the new pFloat set of vectors for parallel cloning had been completed, we began testing its potential to express soluble material from a gene known to be recalcitrant in any expression system tested so far, *sas-5*.

2.2.1 Initial cloning and expression trials

The common design of all pFloat vectors allowed us to clone one PCR-amplified insert of full-length SAS-5 (residues 2-404; recoded to minimise the number of rare bacterial

codons) into 12 different pFloat vectors in parallel. We followed this up by performing small-scale expression trials in BL21(DE3) cells, at two expression temperatures: 37 °C (for 4 h) and 18 °C (for 16 h). Following cell harvest and lysis, we carried out test purification, and test cleavage with HRV-3C. The results we intriguing: while most of the vectors, yielded insoluble or marginally soluble expression of SAS-5_{FL}, a few of them, including pFloat-SUMO, pFloat-Im9, and pFloat-MsyB, gave soluble protein. The highest soluble expression of SAS-5_{FL} was achieved with pFloat-MsyB, and the MsyB-SAS-5_{FL} fusion product tested positive for cleavage with HRV-3C. In general, expression of SAS-5_{FL} at 18 °C for 16 h gave better results than expression at 37 °C for 4 h.

It appears that in comparison to well-characterised solubility enhancers, such as MBP or SUMO, the relatively unknown MsyB performed exceedingly well. MsyB is a highly acidic protein (theoretical pI: 3.58), which has been implicated in enhancement of protein export in *E. coli* [330]. It has proved a great solubility booster (in *E. coli*) of various protein targets known to be prone to aggregation, such as bovine enterokinase, TEV protease and rbcL (large subunit of RuBisCO) [319, 358]. Such excellent performance of MsyB in solubilising other proteins likely stems from the inherent negative charge of the protein at physiological conditions. Through repulsive forces, the high net charge may prevent aggregation of target proteins through shielding, and for proteins that fold rather slowly, such shielding will provide them with enough time to fold into correct conformations.

2.2.2 Initial attempts of large scale expression and purification

Once we had identified pFloat-MsyB as a vector supporting soluble expression of SAS-5_{FL} in *E. coli*, we carried out large-scale expression in BL21(DE3), and attempted to establish a purification protocol for SAS-5_{FL}. Early in our efforts, however, we found that the protein did not bid efficiently to any of our IMAC columns tested, Ni-NTA and Co-TALON (∼10 % bound). We reasoned that perhaps SAS-5 formed a large oligomer, which caused occlusion of the His_6-tag, and hence affected its binding ability to any type of IMAC resin. From other experiments we learned that insoluble SAS-5_{FL} can in fact be solubilised in 8 M urea - through unfolding of the polypeptide chain. However, removal of urea through dialysis or rapid dilution did not allow for

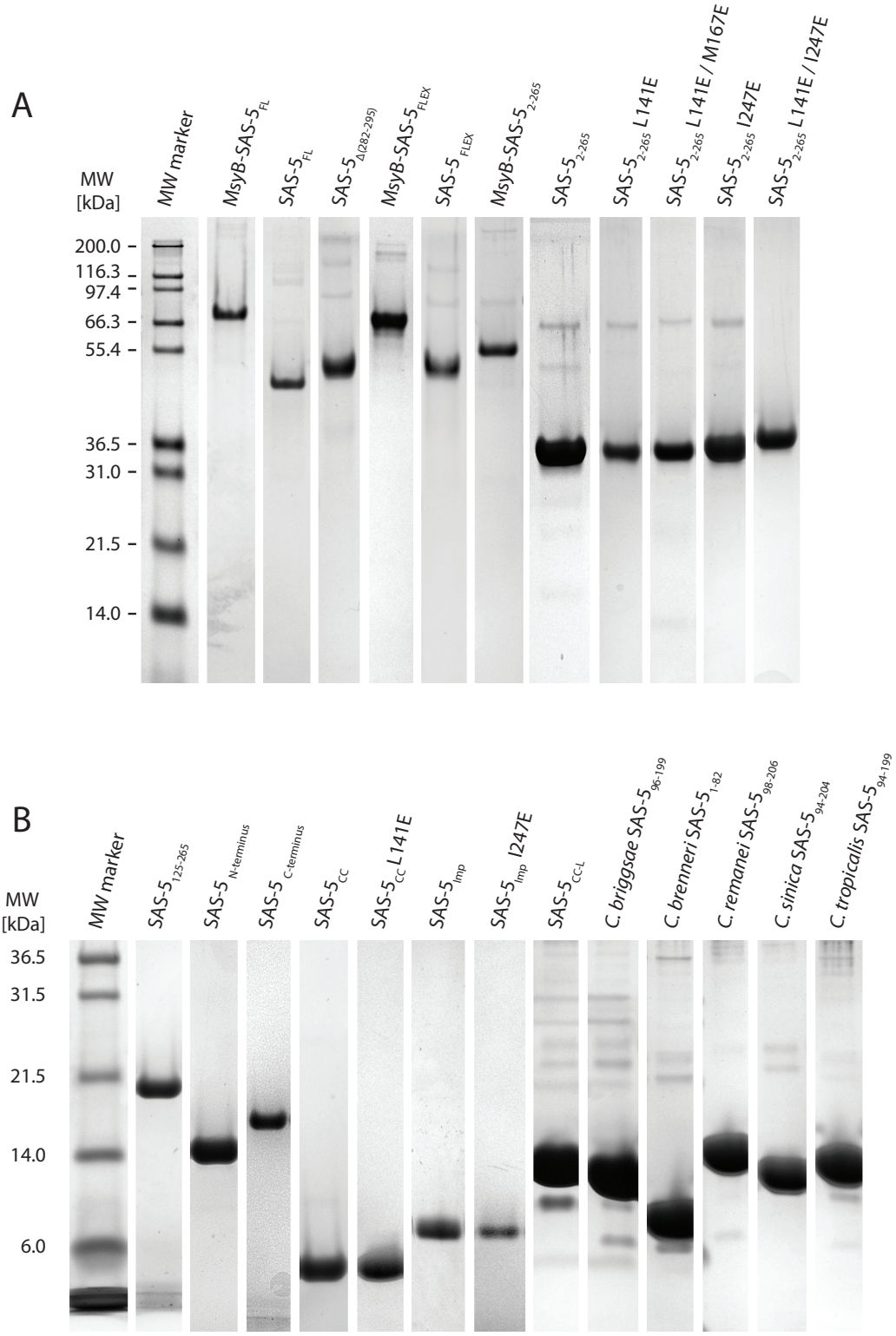


Figure 2.5: **Purity assessment of recombinant proteins.** (**A**, **B**) Coomassie-stained SDS-PAGE of recombinant protein samples used in this study. Panel **A** shows proteins with molecular weight over 30 kDa, and panel **B** proteins smaller than that. Individual gel lanes were cut and matched in size to a common molecular-weight marker along the vertical axis. The level of contrast between the lanes was equalized.

correct re-folding of SAS-5_{FL}, which precipitated out of solution instead. Therefore, in order to relax the potential physical constraints on the His_6-tag, but not completely unfold the protein, we used 3 M urea in IMAC purifications. To our satisfaction, >50 % of SAS-5_{FL} in 3 M urea was bound to Ni-NTA resin. Furthermore, subsequent dialysis against buffer containing no urea, did not cause precipitation.

Having a large amount of IMAC-purified SAS-5_{FL}, we managed to enrich it to 99 % purity using a combination of anion-exchange and size-exclusion chromatography (Figure 2.5 on page 77, which shows all recombinant proteins used in this study). Although the purity of SAS-5_{FL} was increasing in the process, we noticed that concentrating the protein (via centrifugal filtration devices, also known as centrifugal concentrators) for size-exclusion chromatography injections, caused large losses of the protein. After screening various buffer conditions, as well as different types of membranes for protein filtration, it emerged that SAS-5_{FL} behaved significantly better in high-salt buffers (>300 mM NaCl) and did not adhere to polyethersulphone membranes. However, over time of days at 4 °C, the protein slowly started to aggregate and eventually sedimented to the bottom of a tube. Long-term storage of SAS-5_{FL} at low temperatures was also not possible, because freeze-thawing seemed to cause precipitation of the protein.

2.2.3 SAS-5 is a soluble aggregate

Despite various obstacles in precipitation of SAS-5_{FL}, we managed to produce enough pure protein for its biophysical assessment. Characterisation of purified SAS-5_{FL} using circular dichroism (CD) revealed presence of protein aggregates, a conclusion also supported by size-exclusion chromatography multi-angle light scattering (SEC-MALS) and negative-stain electron microscopy (EM).

The circular dichroism spectrum of SAS-5_{FL} did not assume the characteristic shape of either α or β proteins (Figure 2.6A, B on page 81). Instead, it resembled spectra of polyacid and polybase peptides which have a tendency to aggregate over time (Figure 2.6C on page 81) [91, 136]. Similarly to the aggregated polyglutamic acid (polyE), SAS-5_{FL} exhibited very low CD signal, with minima at 210 nm and 225 nm, suggesting presence of a β structure. In the polyE study, Digambaranath *et al.* [91], was able to show that the initial polyE CD spectra contained mainly helical

and random coil structures; however, over time these were converted into β structures upon aggregation.

SEC-MALS showed that SAS-5_{FL} scatters light very strongly; however, together with a low amount of protein passing the column (indicated by a very low reading of differential refractive index - dRI), these results suggested that very large species were present in the sample (Figure 2.6D on page 81). Indeed, the protein eluted in the void volume of Superose 6 column, and reached several gigadaltons in size.

Moreover, electron microscopy inspection of negatively stained SAS-5_{FL} molecules revealed that the protein forms extremely large spherical species $\sim$50 nm in diameter (Figure 2.6E on page 81), which closely resemble protein-lipid structures of outer membrane vesicles (OMV) from bacteria (Figure 2.6F on page 81) [158]. Interestingly, it seems that bacteria use OMVs to release the stress, induced by bulk-mass of misfolded and aggregated proteins which accumulated over time inside of the periplasm [192, 223]. It would be interesting to speculate whether some of the exposed hydrophobic parts of SAS-5_{FL} could form non-specific interactions with lipids within the cell of *E. coli*. Over-production of MsyB-SAS-5_{FL}, which due to the strong solubilising effect of MsyB, instead of driving formation of inclusion bodies, could potentially allow SAS-5_{FL} to aggregate into vesicles of various sizes. Most of those vesicles must have been too large to interact with divalent ions during various IMAC purifications; however, they were flexible or permeable enough to allow HRV-3C to cleave the MsyB-tag off.

2.2.4 The *sticky strand*

Even though SAS-5_{FL} has proved to be a soluble aggregate, we found that shorter fragments of SAS-5 can exist in solution as non-aggregated and well-behaved species (in particular the SAS-$5_{(2\text{-}265)}$ construct - see Section 4.2 *SAS-5 assembles into tetramers and hexamers in solution* on page 114). Therefore, we reasoned that the aggregation of SAS-5 must stem from a short motif within the C-terminal part of the protein (between residues 266 and 404). We sought to locate the aggregation motif by expressing a series of MsyB-tagged C-terminal truncation constructs, and testing them for their ability to bind IMAC resin, without 3 M urea. Interestingly, we observed a significant increase in solubility when comparing a construct encompassing residues

2-279 with one containing residues 2-296 (Figure 2.7A on page 82, compare lanes 3 and 4). Secondary structure prediction by PSIPRED v3.3 [225] suggested a β-strand between residues 282 and 295 (Figure 3.1A on page 86). Excising this region from SAS-5_{FL} (*i.e.* construct SAS-$5_{\Delta(282\text{-}295)}$) or replacing it with a Gly-Ser-Ala-rich flexible linker of equal length (SAS-5_{FLEX}) resulted in proteins with circular dichroism spectra characteristic of mainly α-helical proteins (Figure 2.8 on page 83), and with improved IMAC binding properties (Figure 2.7B on page 82). These results suggest that the predicted β-strand (henceforth the '*sticky strand*') between residues 282 and 295 of SAS-5 drives formation of large protein aggregates.

In summary, we cloned SAS-5_{FL} into a range of pFloat vectors (Figure 2.4 on page 72), and found that pFloat-MsyB gives strong soluble expression of SAS-5_{FL}. However, purification of MsyB-SAS-5_{FL} was not effective without 3 M urea which rendered the His_6-tag more accessible for IMAC binding. Biophysical analysis of pure SAS-5_{FL} showed that the protein forms soluble aggregates (Figure 2.6 on page 81), which explained why SAS-5_{FL} was unable to bind IMAC columns efficiently. Further dissection of the C-terminal region of SAS-5_{FL} uncovered a small motif (residues 282-295) responsible for driving the protein into the aggregated state (Figure 2.7 on page 82). Excision of this motif yielded perfectly soluble protein, which behaved well in biophysical assays (Figure 2.8 on page 83).

Finally, those results suggest that MsyB is capable of solubilising even aggregated proteins. Therefore, using such an effective tag should be routinely followed up by biophysical analysis of the protein state in solution to avoid misinterpretation of results.

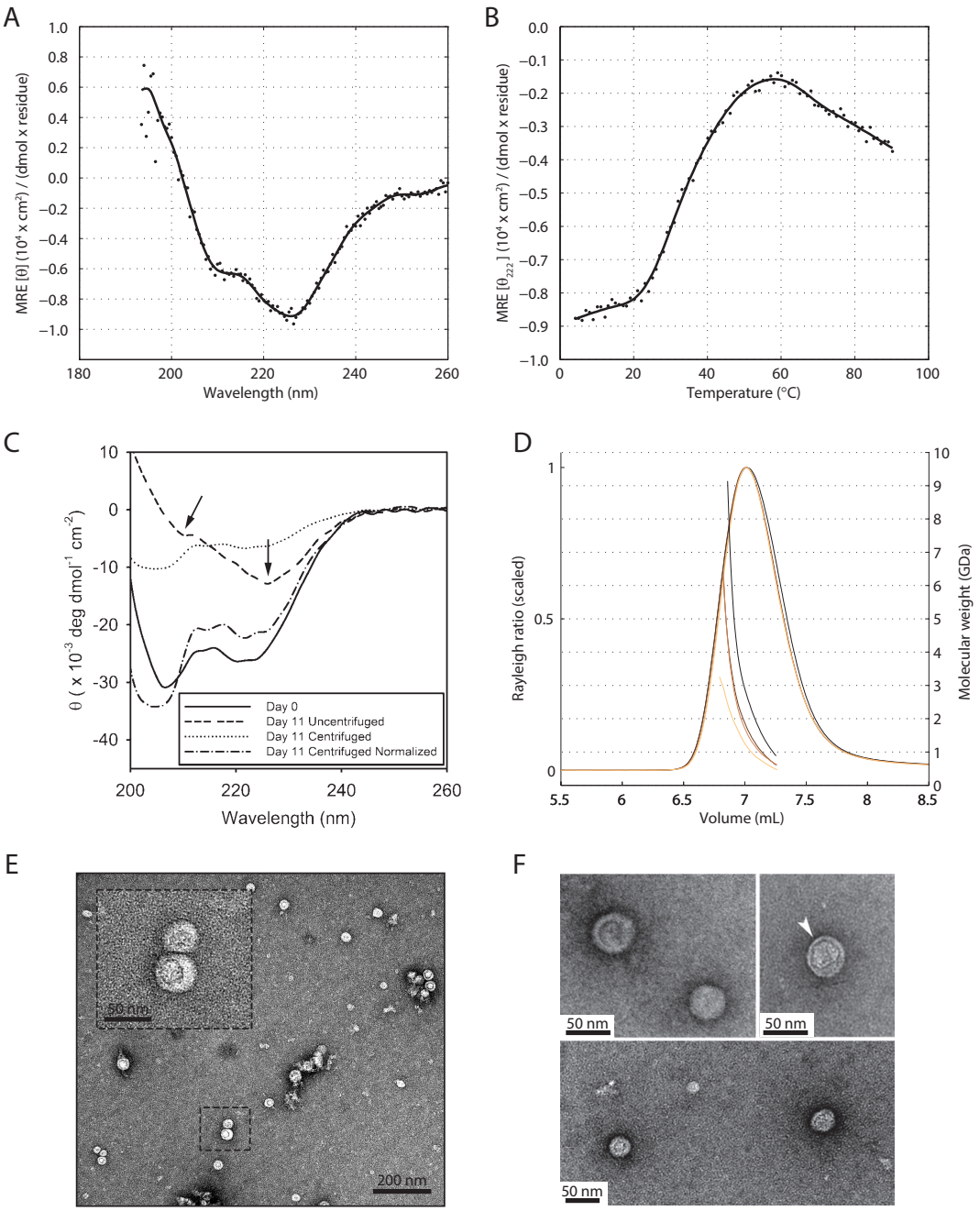


Figure 2.6: ***In vitro* protein aggregates of SAS-5_{FL}.** (**A**) CD spectrum of SAS-5_{FL} collected at 10 °C and plotted as per residue molar elipticity against wavelength, and (**B**) thermal unfolding profile of SAS-5_{FL} monitored by CD recorded at 222 nm, and plotted as per residue molar elipticity against temperature. Note the non-co-operative melting profile. The CD spectrum in panel **A** is of very low signal and resembles spectra of aggregated polyglutamic acid (polyE) [91, 136], reproduced in panel **C**. Arrows point to the relevant spectrum, indicating two common minima between SAS-5_{FL} and polyE, at 210 nm and 225 nm. (**D**) Overlay of multiple chromatograms from SEC-MALS of SAS-5_{FL}, coloured according to injection concentration (black - the highest, light orange - the lowest). Scaled light scattering intensity is plotted against elution volume. Calculated molecular weights are shown as horizontal dotted lines over chromatogram peaks and correspond to the right-hand axis. SAS-5_{FL} elutes in the void volume of Superose 6 10/300 column (∼7 ml), suggesting protein aggregation. (**E**) Representative electron micrograph of negatively stained SAS-5_{FL}; the inset shows a high-magnification view of the boxed section. SAS-5_{FL} appears to form large particles of variable size, which resemble bilayered outer membrane vesicles (OMV) from bacteria [158], reproduced in **F**.

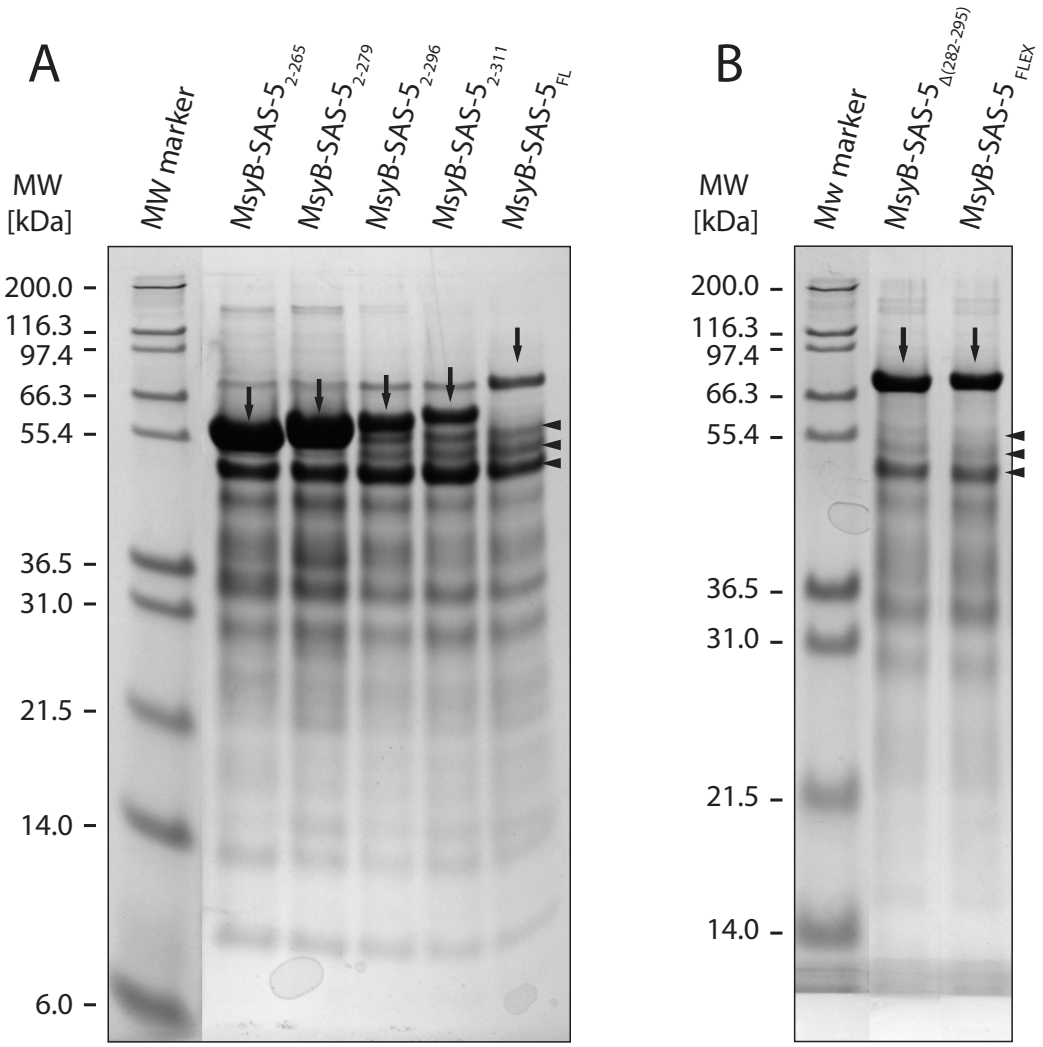


Figure 2.7: ***Sticky strand* drives aggregation of SAS-5.** (**A**, **B**) Coomassie-stained SDS-PAGE of metal-affinity purified SAS-5 constructs. MsyB-tagged SAS-5_{FL} does not bind very well to the IMAC column (lane 6; arrows indicate non-degraded protein construct). Instead, products of proteolytic degradation (arrowheads) show better resin binding than the full-length, non-degraded protein. Expressing multiple constructs of MsyB-SAS-5 truncated at the C-terminus, reveals a dramatic change in binding after truncation of residues 280-296 (compare lanes 3 and 4). Once this protein segment is deleted (MsyB-SAS-$5_{\Delta(282\text{-}295)}$) or substituted with a flexible linker (MsyB-SAS-5_{FLEX}), binding of the non-degraded protein product to metal-affinity resin is much higher than of its degradation products (**B**, compare to lane 6 of panel **A**).

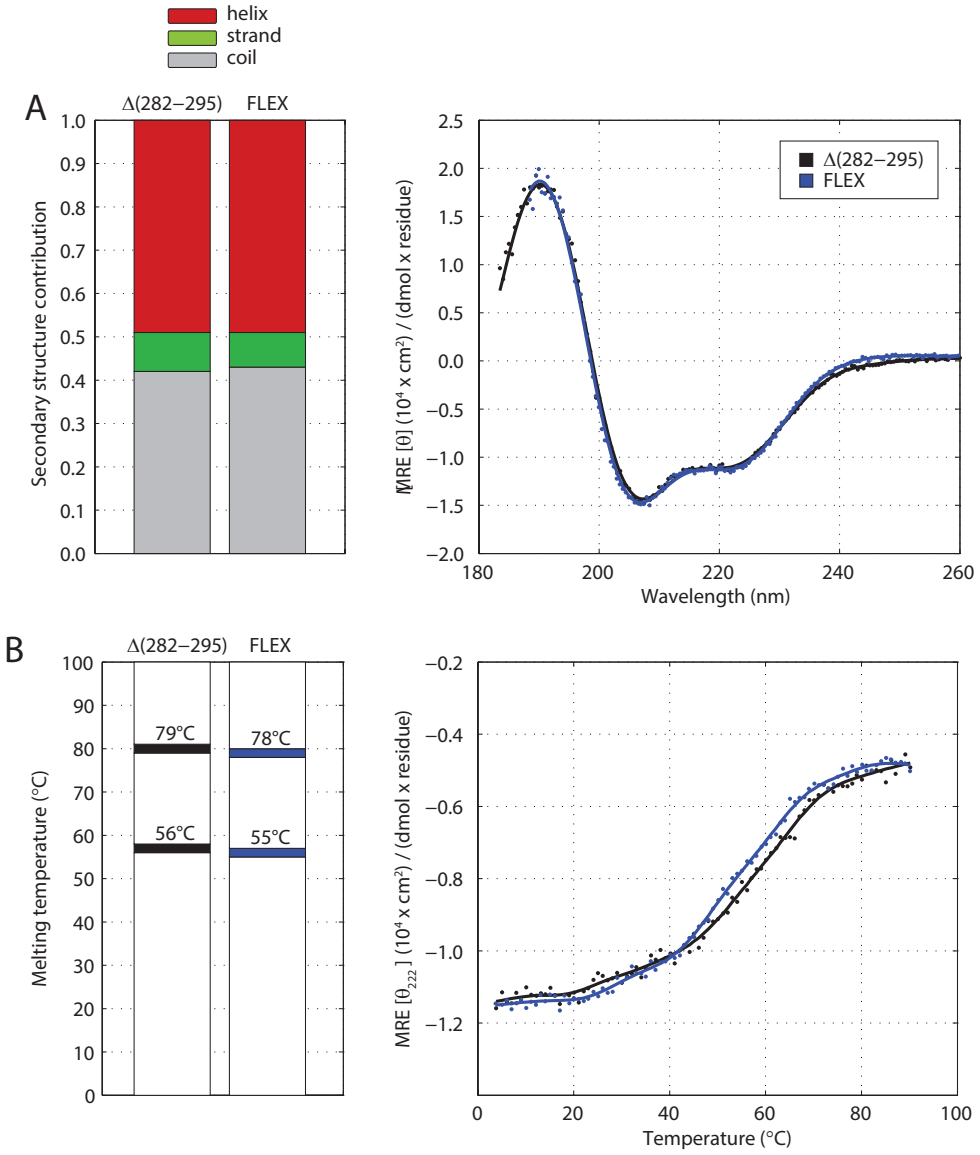


Figure 2.8: **Removal of the *sticky strand* yields well-behaved SAS-5.** (**A**) Overlaid CD spectra of SAS-$5_{\Delta(282\text{-}295)}$ and SAS-5_{FLEX} collected at 10 °C and plotted as per residue molar elipticity against wavelength. Bar charts on the left depict contribution of each secondary-structure elements (random coil in grey, β-strand in green, α-helix in red) to the CD spectrum, deconvoluted with CDSSTR [164]. Note that this is a semi-quantitative analysis. (**B**) Thermal unfolding profiles of SAS-$5_{\Delta(282\text{-}295)}$ and SAS-5_{FLEX} recorded at 222 nm, and plotted as per residue molar elipticity against temperature. Bars on the left depict melting transition temperatures observed for each sample.

Chapter 3

Two folded domains of SAS-5

In the previous chapter, I explained the steps we took to produce soluble SAS-5 for biophysical and structural work. Large quantities of purified protein material are often needed to characterise individual proteins, and even more so, when the function of the protein in question is completely unknown. Admittedly, elucidation of the SAS-5 function in centriole assembly has been hindered by its own essential requirement for the process itself [78, 88, 296]. Worm embryos, where SAS-5 has been depleted with RNAi, show a defect in centriole duplication, which in turn leads to a lethal phenotype. Structural biology has been instrumental in shedding light on the molecular mechanisms underlying cellular processes, and in the last five years, it helped to establish the basis for the nine-fold symmetry of centrioles [142, 183, 331, 332], and to explain a number of relationships between essential proteins in centriole assembly [71,140,256,301,304]. In this chapter, I will describe how we used structural biology as a tool for understanding the molecular structure of SAS-5. The premise of those experiments was to find out what SAS-5 looked like at the atomic level, and to use that knowledge to make specific changes in the chemistry of SAS-5, effectively altering the original properties in a controlled fashion. Introduction and assessment of those small chemical changes *in vivo*, as opposed to depleting the entire protein altogether, may shed light on the role of SAS-5 in centriole assembly.

3.1 Sequence modules in SAS-5 family

SAS-5 has no direct primary sequence homologues outside of the *Nematoda* phylum. The orthologues of SAS-5 in *Drosophila* and in vertebrates (Ana2 and STIL) share

similar mutant phenotype upon depletion and over-expression. Likewise, these orthologues share homology only within their own group. Therefore, not only proteins in the SAS-5 family are distinct from one another but also the entire SAS-5/Ana2/STIL family of proteins appears to be rather unique - a likely reason why besides the coiled coil (see Section 3.2.1 *Computational evaluation* on page 86), no other known structural folds have been found within the primary sequence of SAS-5. However, there appears to be a common ground between the members of the family, characterised by either short fragments of primary sequence similarity or by common binding partners (Figure 1.21A on page 48).

Two regions of sequence similarity between Ana2 and STIL are found within STAN (STil/ANa2) and TIM (truncated in microcephaly) motifs (Figure 1.21A on page 48) [14,313]. Both motifs were also suggested to be present in *C. elegans* SAS-5; however, due to incredibly weak sequence similarity, such claims need further experimental validation. The predicted location of those motifs in SAS-5 partially overlaps with the last 20 amino acids of SAS-5, which had been previously found to interact with SAS-6 (Figure 1.24 on page 55) [142,275]. However, it remains to be established in molecular detail whether the interaction between SAS-6 and Ana2/STIL in flies and vertebrates is similar to that of SAS-6 and SAS-5 in worms [98, 191, 231, 252]. Furthermore, a highly conserved proline-rich motif of STIL and Ana2 (PRxxPxP), takes the central role in the interaction between Ana2/STIL and SAS-4/CPAP (Figure 1.26 on page 58) [71, 140]. Also, such proline-rich region, although not exactly following the motif of Ana2/STIL, could potentially exist in *C. elegans* SAS-5, and be considered as the putative site for SAS-4 binding [140]. However, whether the mode of *C. elegans* SAS-5 binding to SAS-4 resembles that of Ana2/STIL and SAS-4/CPAP remains an open question.

3.2 Structural architecture of SAS-5

Structural biology is a reductionist tool. A protein is normally evaluated for its ability to form stable and folded domains - the tertiary structure. If these domains fold independently, it usually is warranted to study them in isolation. Such reductionist approach is particularly helpful in situations where the protein under investigation

has a modular architecture (*i.e.* is made of many different modules or domains), so that the roles of individual modules can be determined.

3.2.1 Computational evaluation

Typically, before commencing any experimental structural work, exhaustive computational predictions are made to guide design of expression constructs. In case of SAS-5, disorder (DISOPRED [341]) and secondary-structure prediction (PSIPRED [225]) algorithms suggested that SAS-5 could be divided into three segments: N-terminal (residues 1-120), Middle (residues 125-264) and C-terminal (residues 269-404) (Figure 3.1 on page 86).

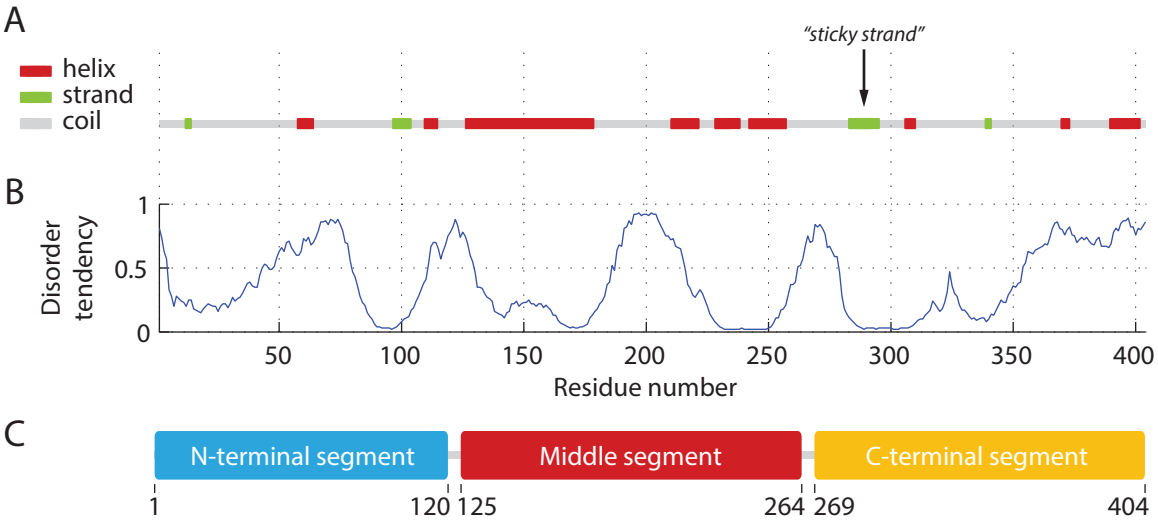


Figure 3.1: **Secondary structure and disorder predictions for SAS-5.** (**A**) PSIPRED-predicted [225] secondary-structure elements of SAS-5 and their relative positions within the primary sequence. (**B**) DISOPRED3-predicted [341] intrinsic protein disorder of SAS-5, plotted as probability versus position in the primary sequence. (**C**) Schematic illustrating division of SAS-5 into three segments: N-terminal, Middle and C-terminal, based on the results from **A** and **B**.

Both the N-terminal and the C-terminal segments, are predicted to be largely unstructured, with only sporadic secondary-structure elements present. While the N-terminal segment of SAS-5 is expected to make contacts with SAS-4 (through the proline-rich region) [140], the C-terminal segment has been shown to interact with the coiled coil of SAS-6 (through 20 C-terminal residues) [142, 275].

Unlike the N- and C-terminal segments, a dense packing of α-helices is expected within the Middle segment. Moreover, according to disorder (DISOPRED [341])

and coiled-coil (Multicoil2 [328]) predictions, these helices are expected to have a defined tertiary structure, with the longest helix within this segment (residues 125-178) potentially super-coiling into a dimeric coiled coil.

Overall, SAS-5 is predicted to be made of a number of short and ordered regions, interspersed by highly flexible linkers, *i.e.* a dynamic structure with a few potentially stable, folded domains.

Based on analogous predictions, Dr Manuel Hilbert (Paul Sherrer Institut, Switzerland) previously produced and crystallised a short fragment of SAS-5 (residues 210-265), which diffracted to ∼20 Å resolution (Dr Vakonakis, personal communication). These preliminary studies proved that parts of SAS-5 can be crystallised, and served as basis for our further investigation.

3.2.2 SAS-5 features two independently folded domains

In Chapter 2, I have described how parallel cloning into multiple pFloat vectors allowed us to obtain soluble SAS-5_{FL} which upon biophysical examination was found to be aggregated (Figure 2.6 on page 81). This aggregation, however, was eliminated after excision of the *sticky strand* (Figure 2.7 on page 82 and Figure 2.8 on page 83). Effectively, two well-behaved protein constructs were obtained: SAS-$5_{\Delta(282\text{-}295)}$ (*sticky strand* excised) and SAS-5_{FLEX} (*sticky strand* replaced with a Gly-Ser-Ala-rich flexible linker).

Interestingly, thermal unfolding of the soluble SAS-$5_{\Delta(282\text{-}295)}$ and SAS-5_{FLEX} constructs revealed a two-step cooperative transition, as monitored by circular dichroism (Figure 2.8B on page 83). Such two-step transition indicates that during the unfolding process, SAS-5 takes three distinct states: folded, intermediate (or partially unfolded) and unfolded. Although the second transition was relatively subtle, these data are suggestive of the presence of two independently folded domains within SAS-5.

In order to locate these domains, we divided SAS-5 into three segments - based on our initial bioinformatics analyses: N-terminal (residues 2-122), Middle (residues 125-264) and C-terminal (residues 269-404) (Figure 3.1C on page 86). After cloning, expression in *E. coli*, and purification, we analysed the three segments with biophysics. Experiments with the C-terminal segment were particularly difficult because of the inclusion of the *sticky strand* in the sequence. This aggregating sequence caused the

C-terminal construct to precipitate out of solution upon concentrating. Yet, at low concentrations, it stayed in solution, allowing us to collect biophysical data.

Circular dichroism measurements verified our earlier computational predictions regarding the disorder level in the N- and C- segments: CD spectra of isolated SAS-5 N-terminal (residues 2-122) and C-terminal (residues 269-404) fragments showed no persistent secondary structure or cooperative thermal unfolding (Figure 3.2 on page 89). Interestingly, the Middle segment showed a highly α-helical character and a very well pronounced two-step cooperative transition profile (Figure 3.3 on page 90), a hint of which was visible earlier - in the CD spectra of SAS-$5_{\Delta(282\text{-}295)}$ and SAS-5_{FLEX} (Figure 2.8B on page 83).

Our previous sequence-based computational predictions suggested that the Middle segment contains a coiled coil (residues 125-180), as well as three tightly-spaced α-helices (residues 210-265) (Figure 3.1A on page 86). These two distinct elements lie adjacent to one another, and are separated by a presumed disordered linker of $\sim$30 residues. We expressed and purified both the coiled coil (SAS-5_{CC}) and the predicted helical region (SAS-5_{Imp} for *Implico*, see below; Figure 3.4A on page 91) and assessed their structural properties with CD. SAS-5_{CC} displayed a double-minimum spectrum, a feature typical for α-helical coiled coils (Figure 3.4B on page 91). Furthermore, SAS-5_{CC} exhibited moderate thermal stability in isolation, with the apparent melting transition temperature (Tm) of $\sim$37 °C (Figure 3.4C on page 91). In contrast, CD of SAS-5_{Imp} revealed a very stable (Tm $\sim$72 °C) α-helical domain (Figure 3.4B, C on page 91). These two distinct melting temperatures undoubtedly are a reflection of the two transitions seen in the unfolding profile of the Middle segment of SAS-5 (SAS-$5_{(125\text{-}265)}$; Figure 3.3B on page 90). We attribute the first melting transition of SAS-$5_{(125\text{-}265)}$ (Tm of $\sim$50 °C) to unfolding of the coiled-coil domain, and the second (Tm of $\sim$70 °C) to unfolding of the *Implico* domain.

Overall, these findings led us to conclude that the segment between residues 125-265 contains two independently folded domains, which are also the only structured elements of SAS-5.

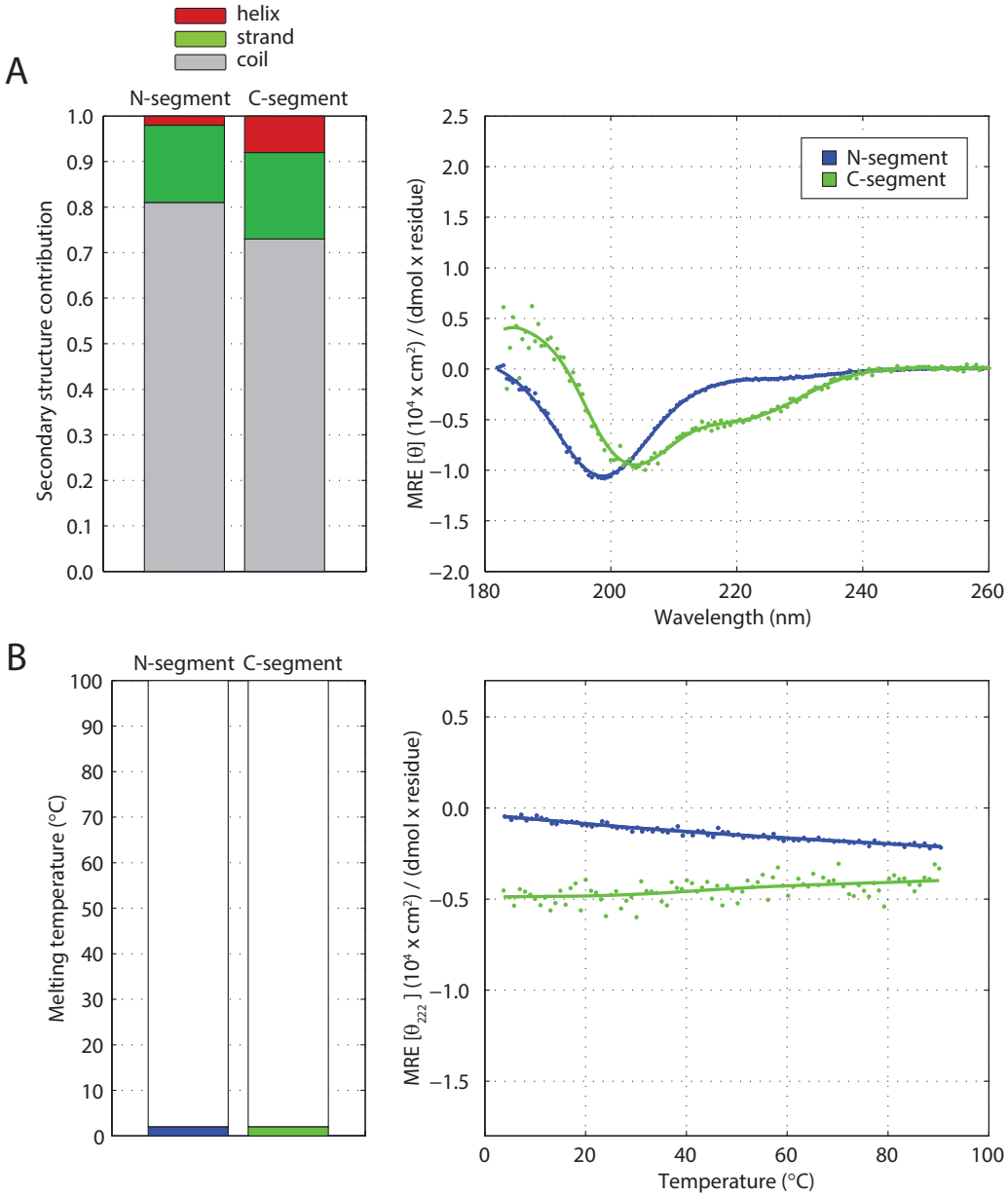


Figure 3.2: **Circular dichroism of isolated N- and C-terminal segments of SAS-5.** (**A**) Overlaid CD spectra of SAS-$5_{(2\text{-}122)}$ (N-segment; blue) and SAS-$5_{(269\text{-}404)}$ (C-segment; green) collected at 10 °C and plotted as per-residue molar elipticity against wavelength. Bar charts on the left depict contribution of each secondary-structure elements (random coil in grey, β-strand in green, α-helix in red) to the CD spectrum, deconvoluted with CDSSTR [164]. Note that this is a semi-quantitative analysis. (**B**) Thermal unfolding profiles of SAS-$5_{(2\text{-}122)}$ (N-segment) and SAS-$5_{(269\text{-}404)}$ (C-segment) recorded at 222 nm, and plotted as per-residue molar elipticity against temperature. Bars on the left depict melting transition temperatures observed for each sample. Note that the unfolding profiles are relatively flat for both constructs, confirming that N- and C-terminal segments of SAS-5 have no persistent tertiary structure *in vitro*.

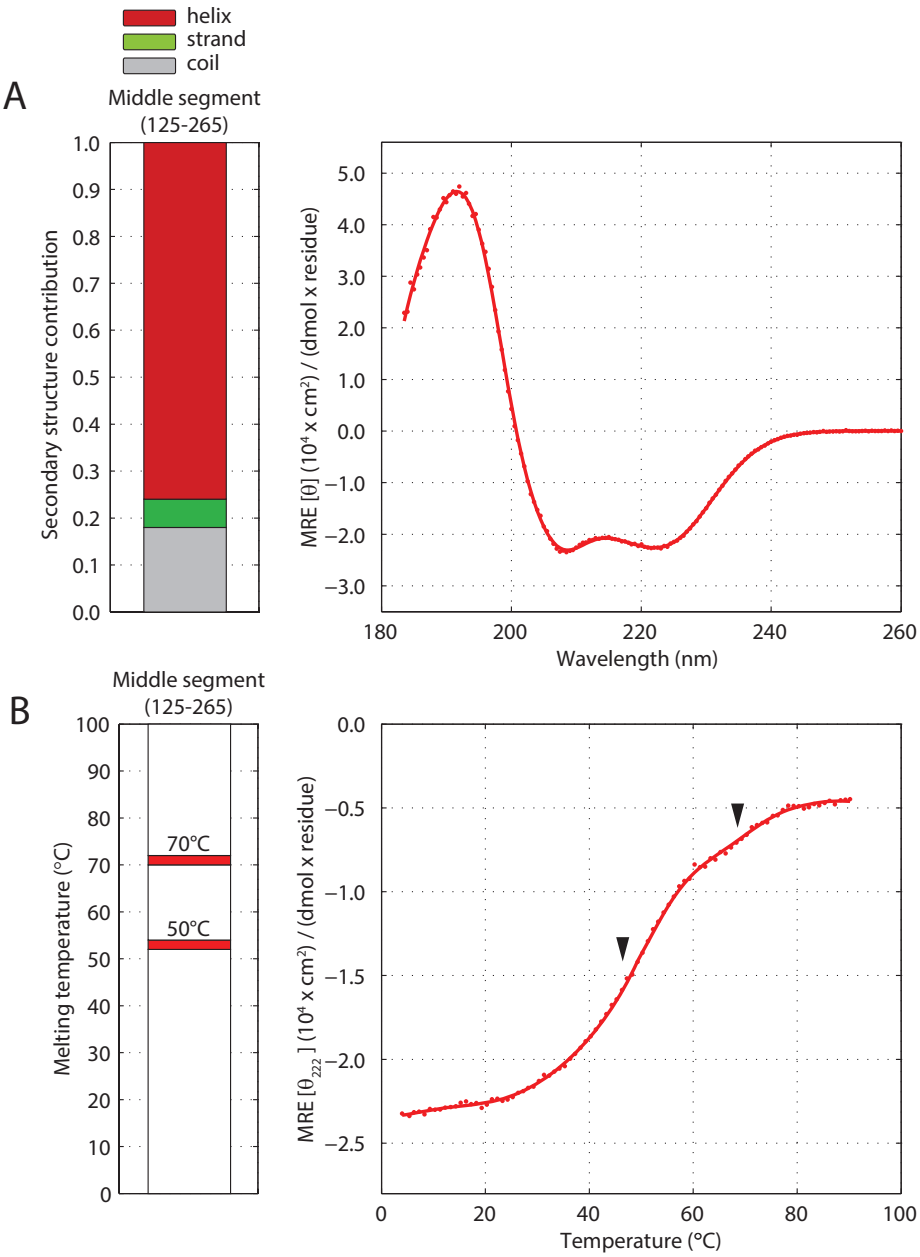


Figure 3.3: **Circular dichroism of isolated Middle segment of SAS-5.** (**A**) Overlaid CD spectra of SAS-$5_{(125\text{-}265)}$ (Middle segment) collected at 10 °C and plotted as per-residue molar elipticity against wavelength. Bar charts on the left depict contribution of each secondary-structure elements (random coil in grey, β-strand in green, α-helix in red) to the CD spectrum, deconvoluted with CDSSTR [164]. Note that this is a semi-quantitative analysis. (**B**) Thermal unfolding profiles of SAS-$5_{(125\text{-}265)}$ (Middle segment) recorded at 222 nm, and plotted as per-residue molar elipticity against temperature. Bars on the left depict melting transition temperatures observed for each sample. Note the two unfolding transitions (arrowheads), suggesting presence of intermediate species.

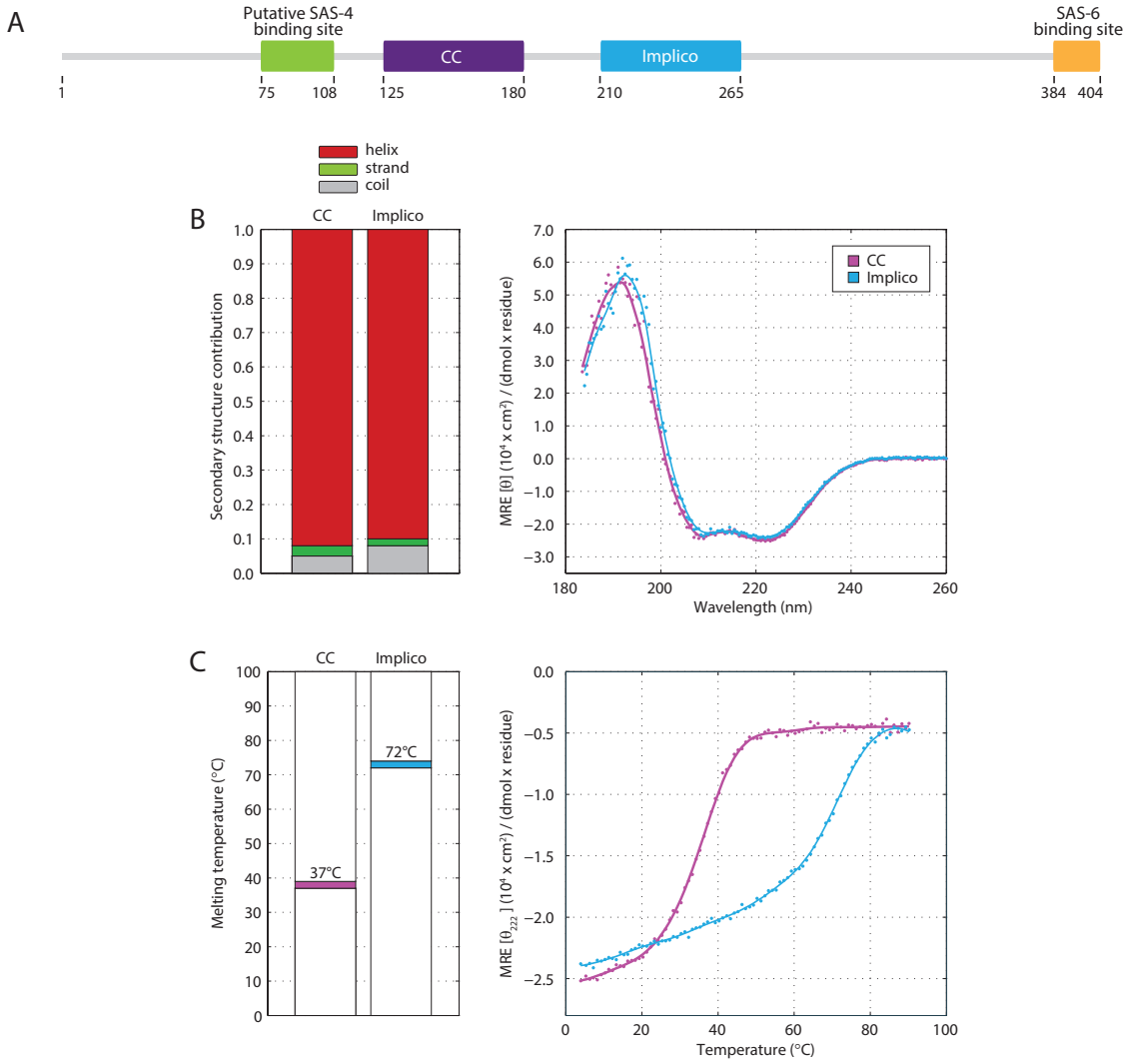


Figure 3.4: **SAS-5 features two independently folded domains.** (**A**) Schematic drawing of positions of two domains, *Implico* and coiled coil, within the primary sequence of SAS-5. Putative SAS-4 binding site [140] and the SAS-6 binding site [142,275] are also indicated for reference. (**B**) Overlaid CD spectra of SAS-5_{CC} (residues 125-180) and SAS-5_{Imp} (residues 210-265) collected at 10 °C and plotted as per-residue molar elipticity against wavelength. Bar charts on the left depict contribution of each secondary-structure elements (random coil in grey, β-strand in green, α-helix in red) to the CD spectrum, deconvoluted with CDSSTR [164]. Note that this is a semi-quantitative analysis. (**C**) Thermal unfolding profiles of SAS-5_{CC} and SAS-5_{Imp} recorded at 222 nm, and plotted as per-residue molar elipticity against temperature. Bars on the left depict melting transition temperatures observed for each sample. Note that while both domains are expected to be almost completely identical in terms of secondary structure composition (**A**), they differ significantly in their folding kinetics (**B**).

3.3 Coiled coil of SAS-5 is a moderately stable trimer

After discovering that SAS-5 is made of two independently folded domains, whilst the rest of the protein is disordered, we set out to learn more about these two domains with X-ray crystallography.

3.3.1 Crystal structure of SAS-5_{CC}

We have obtained orthorhombic $P2_12_12_1$ native crystals of SAS-5_{CC} (residues 125-180), which diffracted X-rays to 1.8 Å resolution. We approach the crystallographic phase problem with multiple methods; however, only one of them had proved to be successful in the end. We expected the coiled-coil construct of SAS-5 to be made exclusively of α-helices, and therefore performed extensive molecular replacement calculations with a library of α-helical peptide fragments, and also with a large library of idealised polyalanine coiled-coil structures, varying in oligomeric state, length, supercoil pitch and supercoil radius. Given that our data extended beyond 2.0 Å resolution, we also attempted *ab initio* protein solution with tools such as AMPLE [29, 327] and ARCIMBOLDO [287, 291, 292]; however, due to anisotropic nature of the dataset (*i.e.* directional dependence in diffraction quality, where a crystal diffracts to higher resolution in one direction, but not in the other, see [294]), and the presence of pseudo-translational symmetry (*i.e.* a situation when non-crystallographic symmetry operators are close to true crystallographic symmetry, and therefore lead to uncertainties in space group assignments, see [359]), none of these approaches were successful. Eventually, after screening a large number of heavy-metal compounds for binding to SAS-5_{CC} crystals, we were able to experimentally determine the phases for structure solution by single anomalous dispersion (SAD) using trimethyl lead acetate derivatised crystals (Table 3.1 on page 93, which includes all crystallographic data and refinement statistics). The structure of SAS-5_{CC} suggested a parallel trimeric coiled coil, although this arrangement was distorted by crystal packing at both termini (Figure 3.5 on page 94).

It is not unusual to observe such distortions for a short coiled coil, such as SAS-5_{CC}, which is only 55 amino acids in length. The coiled coil may partially unwrap to facilitate its own packing into a three-dimensional lattice. The distortion observed at the N-terminus of the coiled coil involves exactly such unwrapping, where the ends of

Protein	$SAS\text{-}5_{CC}$	$SAS\text{-}5_{CC}$ (Pb-derivative)	$SAS\text{-}5_{CC\text{-}L2}$	$SAS\text{-}5_{CC\text{-}L2}$ (Hg-derivative)	$SAS\text{-}5_{Imp}$	$SAS\text{-}5_{Imp}$ (Hg-derivative)
PDB ID	4YV4	-	-	-	4YNH	-
Space group	$P2_12_12_1$	$P2_12_12_1$	$C222_1$	$C222_1$	$P2_1$	$P2_1$
Unit cell (Å, °)	a = 46.52 b = 55.10 c = 191.50	a = 46.32 b = 52.99 c = 191.55	a = 56.21 b = 62.86 c = 134.33	a = 55.10 b = 63.29 c = 134.51	a = 27.18 b = 36.29 c = 42.60 β = 97.50	a = 27.85 b = 36.65 c = 42.13 β = 97.74
Beamline	DLS/I04	DLS/I03	DLS/I04	DLS/I04-1	ESRF ID14-4	DLS/I02
Wavelength (Å)	0.9795	0.9470	0.8855	0.9174	0.900	1.007
Resolution range (Å)	41.90-1.80 (1.84-1.80)*†	41.70-1.90 (1.95-1.90)‡	44.79-1.72 (1.78-1.72)	44.84-2.60 (2.69-2.60)	18.49-1.00 (1.05-1.00)	41.75-1.70 (1.74-1.70)
R_{merge}	0.047 (1.092)	0.089 (1.088)	0.195 (4.416)	0.414 (4.005)	0.108 (0.763)	0.065 (0.541)
R_{pim}	0.034 (0.988)	0.014 (0.680)	0.057 (1.623)	0.077 (0.783)	0.044 (0.314)	0.017 (0.151)
Completeness (%)	98.1 (97.8)	80.1 (33.6)§	100 (100)	98.8 (94.0)	99.4 (98.3)	85.6 (39.4)¶
Multiplicity	2.9 (2.9)	9.3 (3.6)	12.5 (8.0)	29.4 (25.9)	6.9 (6.8)	18.4 (13.6)
Mean I/σ(I)	7.4 (0.7)	13 (0.9)	8.1 (0.3)	9.0 (0.8)	12.0 (4.2)	26.5 (4.5)
Phasing						
No. of heavy atom sites	-	2	-	2	-	2
Resolution (Å)	-	41.70-1.90	-	44.84-2.60	-	41.75-1.70
FOM DM	-	0.66	-	0.57	-	0.75
Refinement statistics						
R_{work} **(reflections)**	20.9 % (30118)	-	25.1 % (25703)	-	12.8 % (42016)	-
R_{free}**(reflections)**	23.8 % (1600)	-	27.0 % (1242)	-	14.9 % (2232)	-
Number of atoms						
Protein atoms	3647	-	1527	-	2040 (including H)	-
Ligands	25	-	2	-	-	-
Water	261	-	70	-	124	-
Average B factors (Å²)						
Protein atoms	50.9	-	55.58	-	10.9	-
Water	46.5	-	55.33	-	23.5	-
RMSD from ideal values		-		-		-
Bonds / angles (Å / °)	0.01 / 0.95	-	0.014 / 1.71	-	0.007 / 1.152	-
MolProbity statistics						
Ramachandran favored (%)	99.5	-	100	-	99.2	-
Ramachandran disallowed (%)	0	-	0	-	0	-
Clashscore (percentile)	2.49 (99th)	-	0.99 (99th)	-	0.98 (98th)	-
MolProbity score (percentile)	1.11 (100th)	-	1.26 (98th)	-	0.79 (99th)	-

Table 3.1: **Crystallographic data collection and refinement statistics**.
* Values in parentheses are for highest-resolution shell
† Anisotropic diffraction of 2.17 Å, 2.59 Å and 1.73 Å (highest resolution) along the a^*, b^* and c^* axes, respectively, based on a mean I/σ(I) > 2.0 criterion.
‡ Anisotropic diffraction of 2.18 Å, 3.03 Å and 2.17 Å (highest resolution) along the a^*, b^* and c^* axes, respectively, based on a mean I/σ(I) > 2.0 criterion.
§ Anisotropic diffraction; 98.0% completeness to 2.26 Å.
¶ Anisotropic diffraction; 98.8% completeness to 2.03 Å.

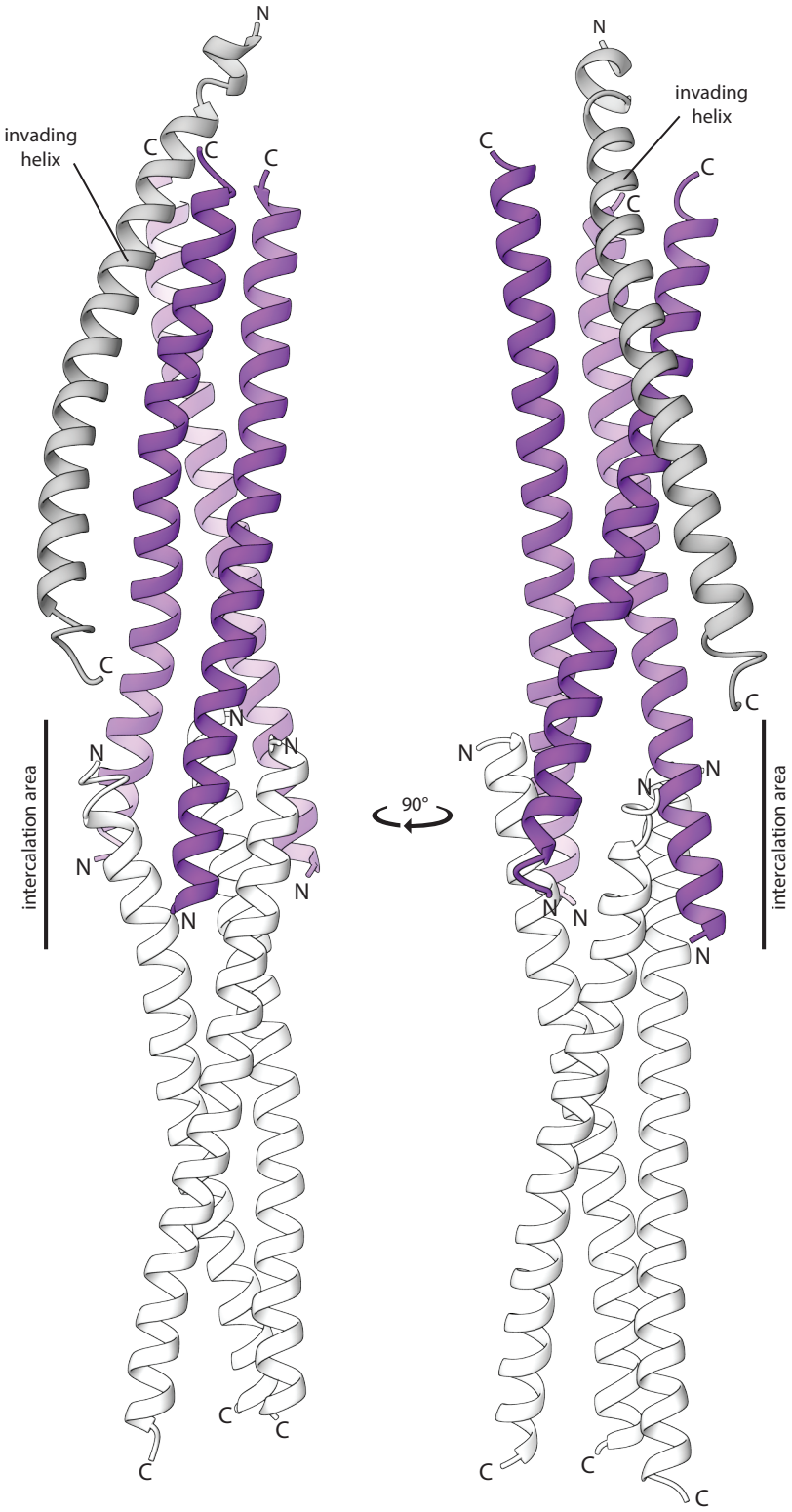


Figure 3.5: **Crystal structure of SAS-5$_{CC}$.** SAS-5$_{CC}$ is drawn in cartoon representation. In order to form crystal contacts, the trimer of SAS-5$_{CC}$ unwrapped at its termini. The N-terminus of SAS-5$_{CC}$ (purple) intercalates with the N-terminus of its symmetry-related trimer mate (white). The intercalation area is indicated. The C-terminus of SAS-5$_{CC}$ (purple) is invaded by an α-helix (grey) positioned anti-parallel to SAS-5$_{CC}$, distorting the coiled coil interaction at that end of the molecule.

successive triple helical bundles intercalate in the crystal. This distortion enables the coiled coils to pack along the longitudinal axis (crystallographic c-axis) (Figure 3.6A on page 96). On the other hand, the distortion present at the C-terminus of the coiled coil allows for packing of the molecules along the lateral axis (crystallographic a-axis) (Figure 3.6B, C on page 96). Besides trimeric bundles, an extra SAS-5_{CC} polypeptide chain is involved in lateral packing. This extra chain also has an α-helical configuration, but it is not involved in coiled-coil formation. It invades the C-termini of two parallel trimers (positioned anti-parallel to one another), effectively glueing them together into a stack. Finally, vertical packing (crystallographic b-axis) is achieved without any unwrapping of the trimers, and involves side-to-side contacts between the stacks created after longitudinal and lateral packing (Figure 3.6D on page 96). Effectively, eight molecules of SAS-5_{CC} are present in one asymmetric unit: two parallel trimers, and two extra α-helices which bridge the trimers in the lateral axis (Figure 3.6A on page 96, dashed box).

To correct for these distortions, we performed 45 ns-long atomistic molecular dynamics (MD) simulations in explicit solvent of the crystal structure of SAS-5_{CC}. We monitored the average distance between the centres of the three helices in the coiled coil (at both N- and C-termini), which rapidly decreased from the initial 17 Å (N-terminus) and 16 Å (C-terminus) to ∼11 Å (both N- and C-termini) within 3 ns from the start of the simulation (Figure 3.7A, B on page 97), leading to convergence into a stable and compact trimer (Figure 3.7C on page 97). However, even though the overall fold of the final simulated model resembled a trimeric super-coil, the canonical knob-into-holes interactions were only present between two chains within the trimer - as evaluated by SOCKET [339]. This could either suggest that the simulation did not proceed for long enough to observe such convergence into a canonical trimeric coiled-coil, or perhaps that protein regions beyond those captured in the crystal are needed to stabilise the overall canonical fold. In any case, at the current state of computational technology, the results from such simulations need to be analysed with caution.

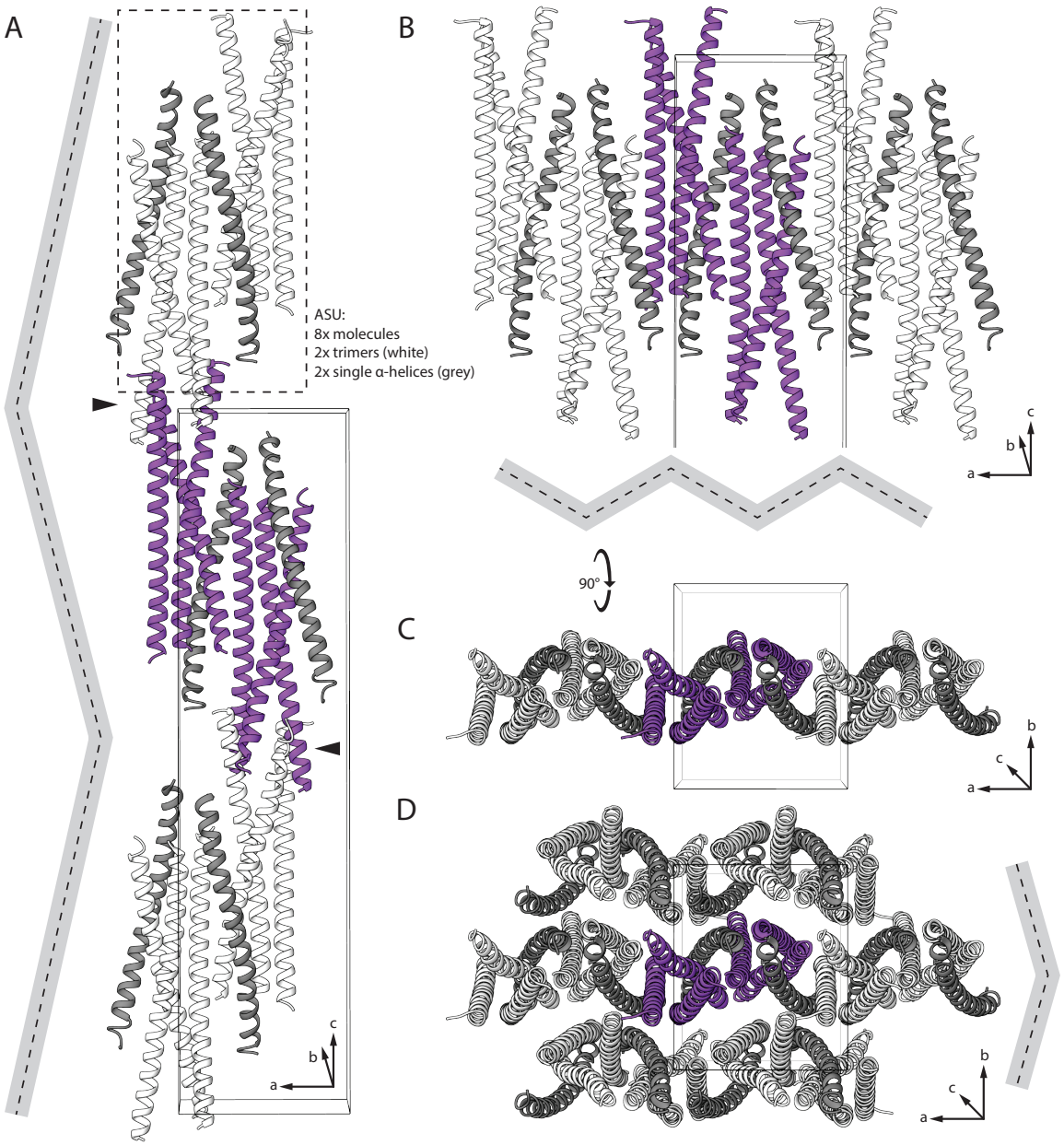


Figure 3.6: **Packing of molecules in the SAS-5$_{CC}$ crystal.** (**A**) The asymmetric unit (ASU; dashed box) is composed of 8× SAS-5$_{CC}$ chains (2× trimers - in white and 2× invading α-helices in grey). Molecules of one ASU pack longitudinally (crystallographic c-axis) via N-terminal intercalation (arrowheads). Unit cell is drawn around the molecules and the axes are indicated. Schematic drawing on the left (grey with a dashed line in the centre) indicates molecule displacement along the packing axis. (**B**, **C**) The invading α-helices connect SAS-5$_{CC}$ trimers laterally (crystallographic a-axis), forming a stack. (**D**) Stack-to-stack contacts achieve vertical packing (crystallographic b-axis) of the ASU with no distortions to SAS-5$_{CC}$ trimer.

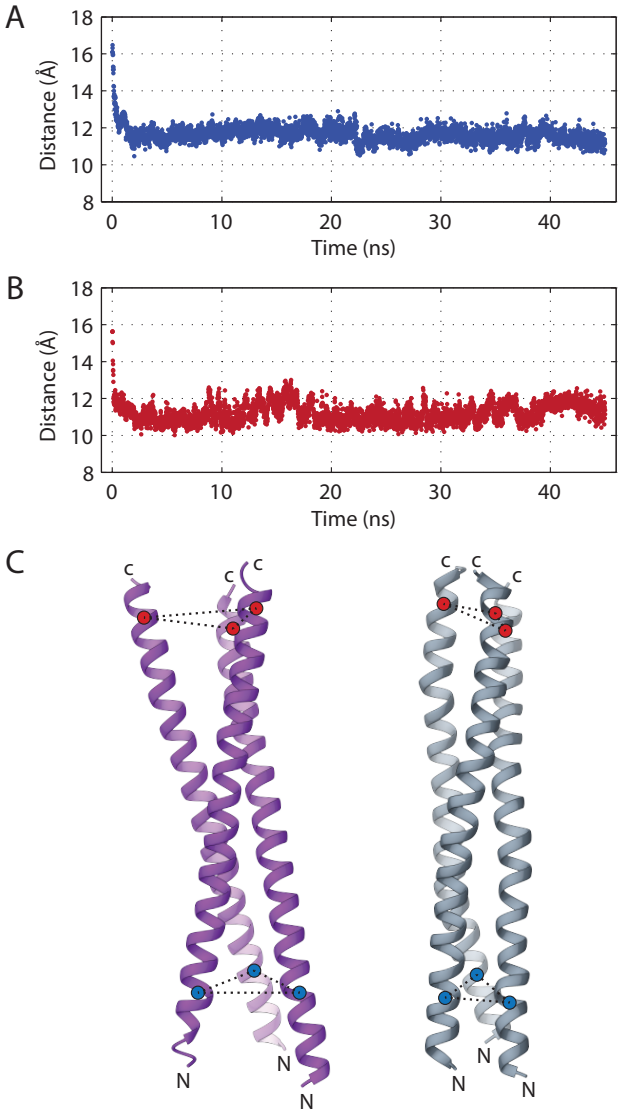


Figure 3.7: **Computational relaxation of crystal packing-distorted SAS-5$_{CC}$.** (**A**, **B**) Average distances between the centres of the terminal α-helices in the SAS-5$_{CC}$ trimer, plotted as a function of time in molecular dynamics simulations. Distances were measured at the N-terminus W133 (blue) and at the C-terminus F174 (red). Initially opened ends of the SAS-5$_{CC}$ trimer decreased rapidly from the average distance of ~16 nm to ~11 nm as the structure converged into a more compact trimeric arrangement. (**C**) Cartoon-represented SAS-5$_{CC}$ trimer structure at the beginning of the MD simulation (as in the crystal structure; left, purple) and at the end (right, grey). Centres of α-helices, average distances between which were plotted in panels **B** and **C**, are shown as blue (at W133) and red (at F174) spheres.

3.3.2 Trimer conservation

In order to verify the oligomeric state of SAS-5$_{CC}$ in solution, whilst avoiding any potential non-specific effects of studying this overly short construct in isolation (*e.g.* unwrapping - as seen in the crystal), we performed SEC-MALS experiments on a slightly longer coiled-coil construct (SAS-5$_{CC\text{-}L1}$, residues 101-206). The results

demonstrated formation of a trimer in solution, in a concentration-dependent manner (Figure 3.8A on page 102), with apparent K_d of ~3 μM (Figure 3.8B on page 102). Both the crystallographic model and MD simulations show a series of hydrophobic residues (W133, M137, L141, I144, I148, M167, M171 and F174) forming the core of the coiled coil (Figure 3.8C, D on page 102). Interestingly, we found that all these residues are either conserved or conservatively substituted in SAS-5 homologues in other *Caenorhabditis* species, suggesting functional relevance (Figure 3.9 on page 103). To investigate the conservation of the trimeric coiled-coil domain across nematode evolution, we expressed constructs corresponding to SAS-$5_{CC\text{-}L1}$ from *Caenorhabditis brenneri*, *Caenorhabditis briggsae*, *Caenorhabditis remanei*, *Caenorhabditis sinica* and *Caenorhabditis tropicalis*. In all cases, SEC-MALS revealed similar oligomerisation modes and apparent affinities to *C. elegans* SAS-$5_{CC\text{-}L1}$ (Figure 3.10 on page 104).

3.3.3 Crystal structure of the extended coiled coil

Finally, in pursuit of a different perspective on the ability of SAS-5_{CC} to pack within a crystal, we crystallised yet another construct featuring the coiled coil of SAS-5: N-terminally extended variant SAS-$5_{CC\text{-}L2}$ (residues 106-180). Native $C222_1$ crystals of SAS-$5_{CC\text{-}L2}$ diffracted to ~2.0 Å resolution. Similarly to the shorter SAS-5_{CC} construct, we were unable to solve the structure with molecular replacement or *ab initio* methods, but eventually determined the phases with SAD from mercury acetate-derivatised crystals after extensive heavy-metal screening (Table 3.1 on page 93).

Although the crystallisation construct spanned amino acids 106-180, we did not observe electron density for the first 11 amino acids (residues 106-116). Nevertheless, extending the N-terminus of the crystallisation construct by additional 19 residues (SAS-5_{CC} spans residues 125-180) introduced a number of changes to crystal packing. First, instead of eight molecules in the asymmetric unit (as in the $P2_12_12_1$ crystals of SAS-5_{CC}), SAS-$5_{CC\text{-}L2}$ crystallised with three molecules per asymmetric unit, strongly indicating the inherent trimeric state of the coiled coil (Figure 3.11 on page 106). Moreover, the N-terminal extension of the construct retained intercalation as means for longitudinal crystal packing (crystallographic c-axis); however, in the SAS-$5_{CC\text{-}L2}$ $C222_1$ crystals, the intercalation occurs between both the N- and

C-termini of successive trimer molecules, and not between pairs of N-termini alone (as in the SAS-5_{CC} $P2_12_12_1$ crystals).

This longitudinal packing was evident in the first crystals of SAS-$5_{CC\text{-}L2}$ which took shape of very thin needles, and only developed into thicker rods after cross-seeding into a different condition. Interestingly, we found that these thicker crystals grew only in the presence of a divalent cation. The best results, size-wise, were achieved with Zn^{2+}. Indeed, packing along both vertical and lateral axes (Figure 3.12 on page 107) was revealed as the result of a presumed Zn^{2+} coordination by carboxylate anions from two glutamates (E131) and two aspartates (D135): one glutamate/aspartate pair from each coiled-coil trimer at the longitudinal intercalation area (Figure 3.13A, B on page 108). These contacts were further supported by two R134 which formed salt bridges with the glutamates and the aspartates. It appears that these Zn-facilitated contacts between the side-chains of the coiled-coil trimers substituted the chain invasion-based vertical contacts of SAS-5_{CC} in $P2_12_12_1$ crystals. Effectively, such crystal packing allowed for SAS-$5_{CC\text{-}L2}$ molecules to form distinct trimers, middle part of which did not form any outside contacts and thus was completely undisturbed (Figure 3.12 on page 107). Unfortunately, this packing also enforced a register shift on the coiled coil: 0, +3, +4 (for each consecutive chain in the trimer; Figure 3.13C on page 108), likely rendering any contacts between the chains of the coiled coil found in this crystal structure physiologically irrelevant, and also highlighting the risks and pitfalls of coiled-coil X-ray crystallography.

3.3.4 SAS-5_{CC} crystallography - summary

The process of crystallisation and structure solution of both SAS-5 coiled-coil constructs (SAS-5_{CC} and SAS-$5_{CC\text{-}L2}$) had proved to be extremely challenging, and yet, despite overcoming these obstacles, the resulting structures were partially distorted by crystal packing forces (Figure 3.5 on page 94 and Figure 3.11 on page 106). Satisfyingly, however, the crystal structure of SAS-5_{CC} trimer was not register-shifted, and even though both of its termini intercalated with other molecules in the crystal, the middle region of the trimer contained useful structural data, which could serve as basis for mutagenesis. Moreover, the MD-relaxed structure of SAS-5_{CC} provided further relevant contact information for almost all the residues within the coiled coil

of SAS-5. Taken together, both the solution data and the crystal structures suggest that the trimeric coiled coil is an evolutionary conserved feature of nematode SAS-5 proteins.

3.4 SAS-5_{Imp} is a structurally novel dimeric domain

Following structural characterisation of the coiled-coil domain of SAS-5, we next turned our attention to the second independently folded domain, SAS-5_{Imp}. We found that this newly recognised domain is highly conserved amongst *Caenorhabditis* species (Figure 3.14 on page 109); however, its primary sequence carries no evident similarity to any known protein family, and appears unique to nematodes. Importantly, through SEC-MALS experiments, we found that that SAS-5_{Imp} forms a dimer at a wide range of concentrations investigated (Figure 3.15 on page 109), suggesting that the dimer is very stable.

Furthermore, we obtained monoclinic $P2_1$ crystals, which diffracted X-rays to 1.6 Å initially, and 1.0 Å following optimisation. Given the ultra-high resolution of X-ray diffraction, we attempted solving SAS-5_{Imp} structure with direct methods, but with no success. Only after screening numerous heavy-atom derivatives, we were able to determine phases by SAD from mercury acetate-derivatised crystals (Table 3.1 on page 93). Crystal structure of SAS-5_{Imp} revealed a dimer composed of interlocked chains arranged in an anti-parallel fashion (Figure 3.16A on page 110). Each chain features three α-helices (A1 to A3 and B1 to B3) that predominantly interact in a pairwise manner (A1-B3, A2-B2 and A3-B1). Proline residues at the linkers between α-helices allow for tight 90 ° turns, resulting in an overall very compact dimer of just $\sim$4 nm in diameter (Figure 3.17 on page 111). The dimer interface involves a number of hydrophobic residues (I219, I225, A229, L230, I232, I233, L237, F243, I247, V250, L251) buried inside of the protein core, effectively organising a perfectly globular domain, with its polar and charged residues positioned at the surface (Figure 3.16B on page 110). Thus, breaking the dimer is expected to unfold the SAS-5_{Imp} domain. Also, given the high thermal stability of SAS-5_{Imp}, it is expected that stable dimers will form even at very low protein concentrations.

This interlocked tight dimer is a novel structural feature that we named *Implico*, from Latin 'to entangle'. To some extent, the two faces of *Implico* differ in charge

distribution. One face is abundant in positively charged residues, while the other is dominated by the negative charge (Figure 3.18A on page 112). Such property allowed the protein to form tight crystal contacts, packing two dimer molecules in a relatively small unit cell (Figure 3.18B, C on page 112). Moreover, unlike a typical protein macromolecule that forms crystals of ∼50 % solvent content, *Implico* crystallised with only ∼23 % solvent content, thus explaining the unusually high resolution of its X-ray diffraction data.

In summary, our structural work revealed that SAS-5 is an oligomeric protein with disordered N- and C-terminal segments. It harbours two independently folded domains, coiled coil and *Implico*, which are flanked and connected by flexible sections. This arrangement creates a modular protein with many functional units assembled into a single polypeptide chain. While the folded domains self-oligomerise into a trimer (coiled coil) and a dimer (*Implico*), flexible regions may undergo coupled folding-and-binding transitions upon interaction with their targets (such as SAS-4 and SAS-6). Understanding how the coiled coil and the *Implico* influence centriole assembly was our next goal - see Chapter 4.

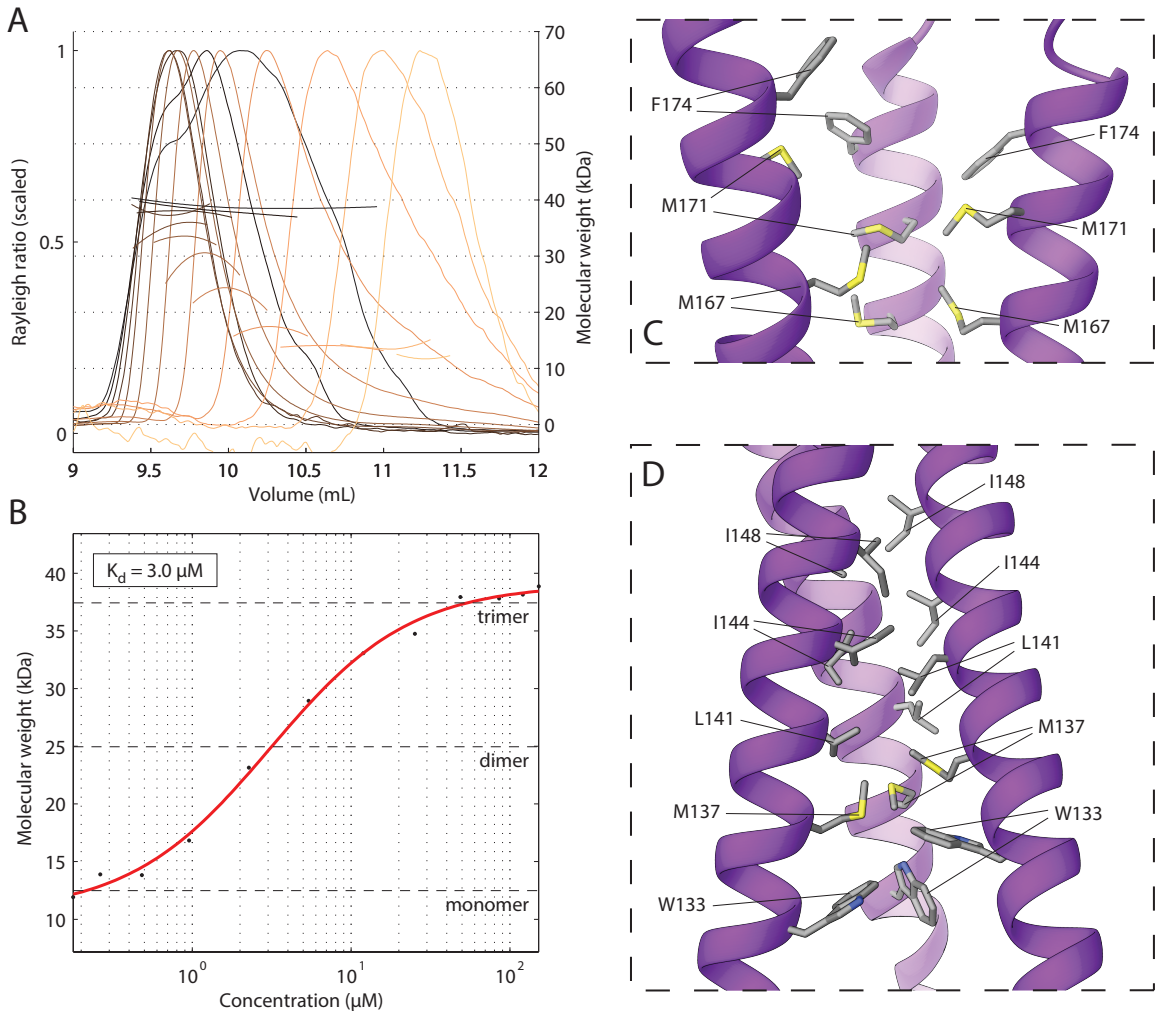


Figure 3.8: **Hydrophobic interactions drive trimerisation of the coiled coil of SAS-5.** (**A**) Overlaid SEC-MALS chromatograms of SAS-$5_{\text{CC-L1}}$ 1:1 dilution series, showing scaled light scattering intensity versus elution volume from a Superdex 75 10/300 column. Calculated molecular weights are shown as horizontal lines over chromatogram peaks, and correspond to the right-hand axis. Datasets are coloured according to injection concentration of the protein (black - the highest, light orange - the lowest). Chromatograms of samples at the highest two injection concentrations show evidence of size-exclusion column overload with protein material. These two samples reach the mass of a trimer, and their peaks instead of shifting towards earlier elution, are broadening. Note the long horizontal lines running across broad elution peaks, indicating stable mass across the entire peak length. Expected monomer mass: 12.5 kDa (**B**) Average molecular weight from SEC-MALS experiments is plotted as a function of on-column protein concentration. Apparent K_d and molecular sizes of expected protein species are indicated. (**C**, **D**) Molecular interactions within the hydrophobic core of the SAS-5_{CC} crystal structure. Protein chains are drawn as cartoon (purple), with relevant C-terminal (**C**) and N-terminal (**D**) hydrophobic residues drawn as sticks (grey, hetero-atom colouring). Mutational substitutions to charged residues at positions 141 and 167 disrupted oligomerisation of the coiled coil.

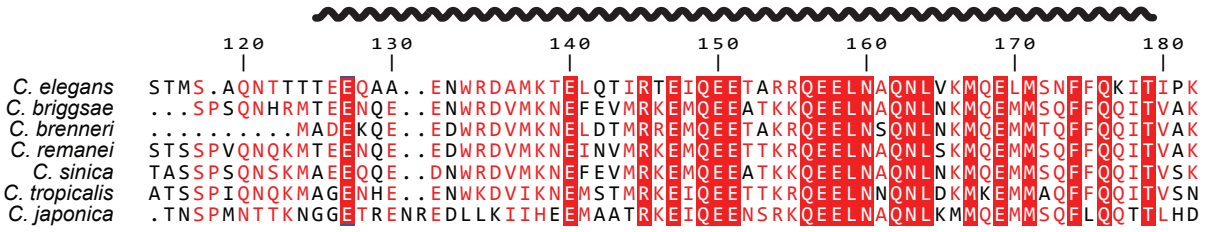


Figure 3.9: **Sequence conservation of SAS-5 coiled coil within the *Caenorhabditis* genus.** SAS-5 coiled-coil sequences from *Caenorhabditis* genus were aligned with MergeAlign [61] and drawn with ESPript 3 [283]. A schematic drawing of protein secondary structure, as observed in the crystal structure of SAS-5_{CC} is shown on top. Universally conserved residues are highlighted in red; conservatively substituted residues are coloured red. Note that the sequence of *C. brenneri* SAS-5 is N-terminally shorter than those of the other six *Caenorhabditis* SAS-5 proteins. Neither Wormbase nor UniProt databases contained information for the missing N-terminal sequence in *C. brenneri* SAS-5, which was further confirmed by the search in the *C. brenneri* genome. It is not unlikely that the missing portion of the *C. brenneri* SAS-5 sequence will eventually emerge, as the *C. brenneri* shotgun sequencing project has not yet been completed.

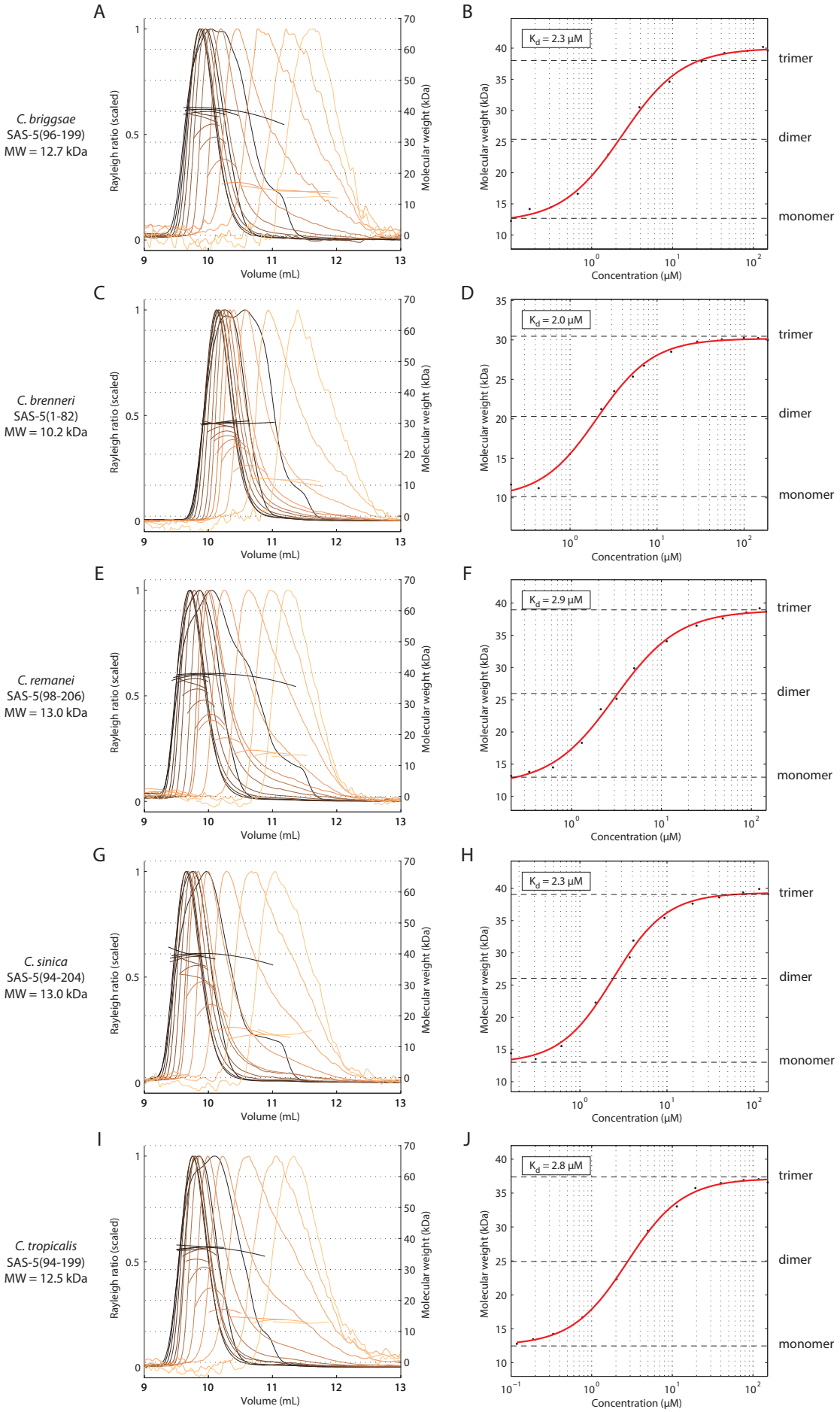


Figure 3.10: **Evolutionary conservation of the trimeric coiled coil of SAS-5.**

Figure 3.10: **Evolutionary conservation of the trimeric coiled coil of SAS-5**. (**A**, **B**) Overlaid SEC-MALS chromatograms of *C. briggsae* SAS-$5_{(96\text{-}199)}$ (equivalent of *C. elegans* SAS-$5_{CC\text{-}L1}$) 1:1 dilution series (**A**), showing scaled light scattering intensity versus elution volume from a Superdex 75 10/300 column. Calculated molecular weights are shown as horizontal lines over chromatogram peaks, and correspond to the right-hand axis. Datasets are coloured according to injection concentration of the protein (black - the highest, light orange - the lowest). Chromatograms of the most concentrated samples show evidence of size-exclusion column overload with protein material. These samples reach the mass of a trimer and their peaks instead of shifting towards earlier elution, are broadening. Note the long horizontal lines running across broad elution peaks, indicating stable mass across the entire peak length. Average molecular weight from SEC-MALS experiments is plotted as a function of on-column protein concentration (**B**). Apparent K_d and molecular sizes of expected protein species are indicated. Similar data from *C. brenneri* SAS-$5_{(1\text{-}82)}$ (**C**, **D**), *C. remanei* SAS-$5_{(98\text{-}206)}$ (**E**, **F**), *C. sinica* SAS-$5_{(94\text{-}204)}$ (**G**, **H**) and *C. tropicalis* SAS-$5_{(94\text{-}199)}$ (**I**, **J**).

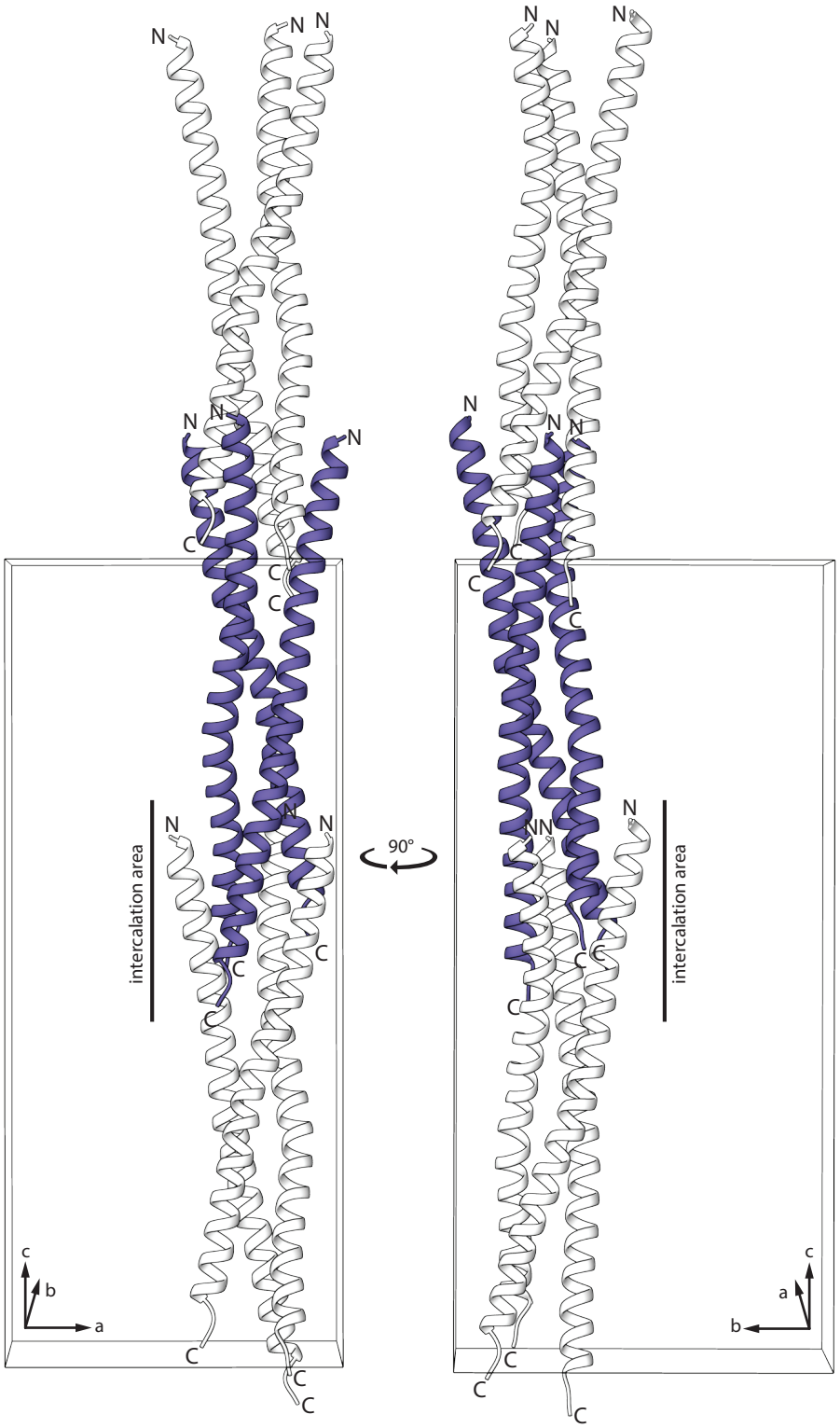


Figure 3.11: **SAS-$5_{CC\text{-}L2}$ trimer intercalates in the crystal via both termini.** Crystal structure of SAS-$5_{CC\text{-}L2}$ in cartoon representation. In order to facilitate crystal contacts, the trimer of SAS-$5_{CC\text{-}L2}$ unwrapped at its termini. The N-terminus of SAS-$5_{CC\text{-}L2}$ (dark blue) intercalates with the C-terminus of its symmetry-related trimer mate (white), effectively forming a back-to-back (N-to-C) continuous chain of molecules along the longitudinal axis (crystallographic c-axis). The intercalation areas are indicated, and the unit cell is drawn as a box around the molecules. The asymmetric unit contains only 3 molecules (dark blue).

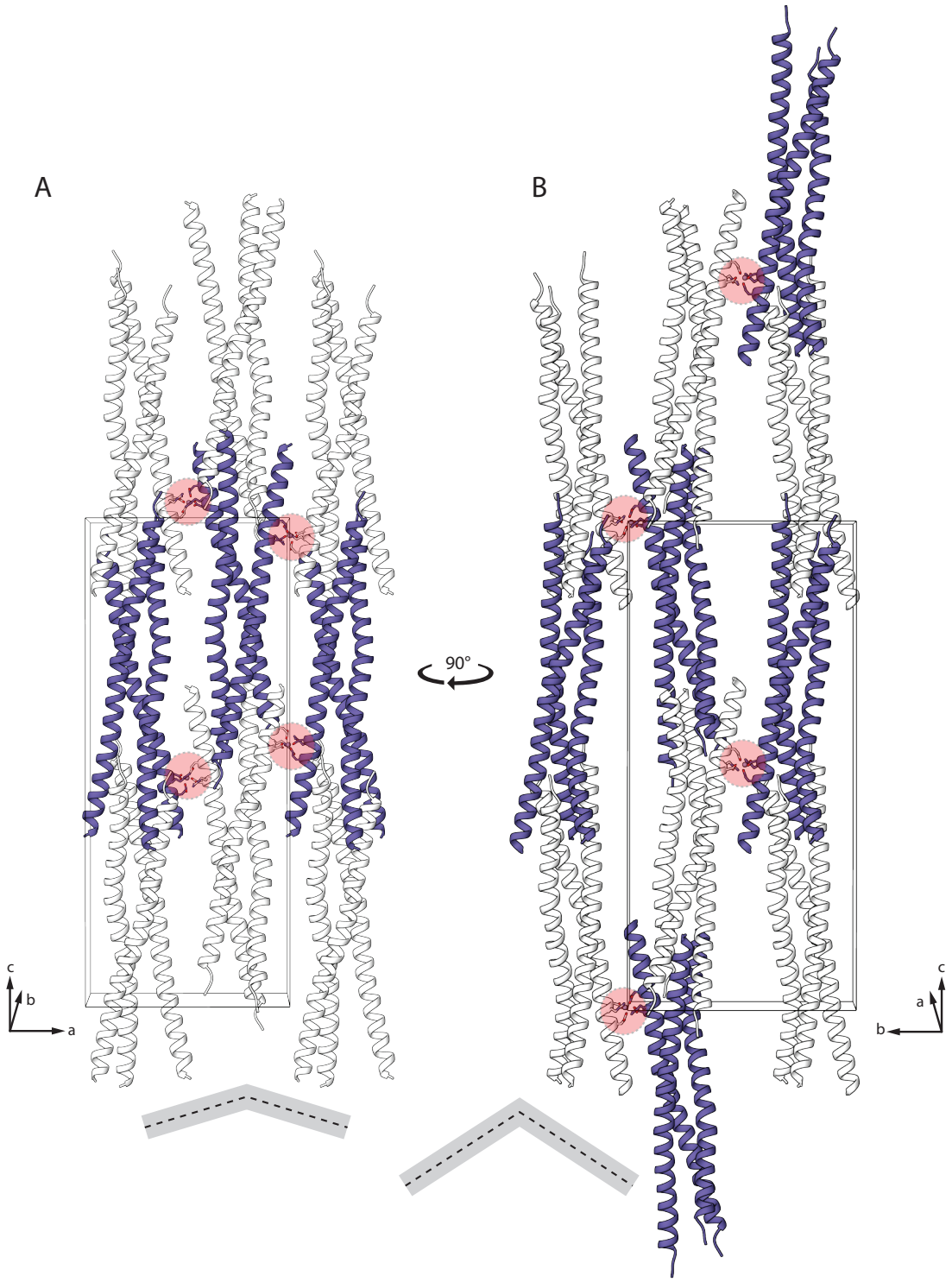


Figure 3.12: **Packing of molecules in the SAS-5$_{\text{CC-L2}}$ crystal.** (**A**, **B**) The asymmetric unit is composed of 3× SAS-5$_{\text{CC-L2}}$ chains (alternating dark blue / white trimers). Unit cell is drawn around the molecules and the axes are indicated. Schematic drawing at the bottom (grey with a dashed line in the centre) indicates molecule displacement along the packing axis. Lateral (**A**; crystallographic a-axis) and vertical (**B**; crystallographic b-axis) crystal contacts are facilitated via local interactions between neighbouring α-helices (red circles). Even though the solvent channels in a-dimension are much narrower than those in b-dimension, there is enough space for the middle region of the coiled coil to remain unrestrained. The SAS-5$_{\text{CC-L2}}$ trimer makes twice as many contacts in a-dimension than in b-dimension.

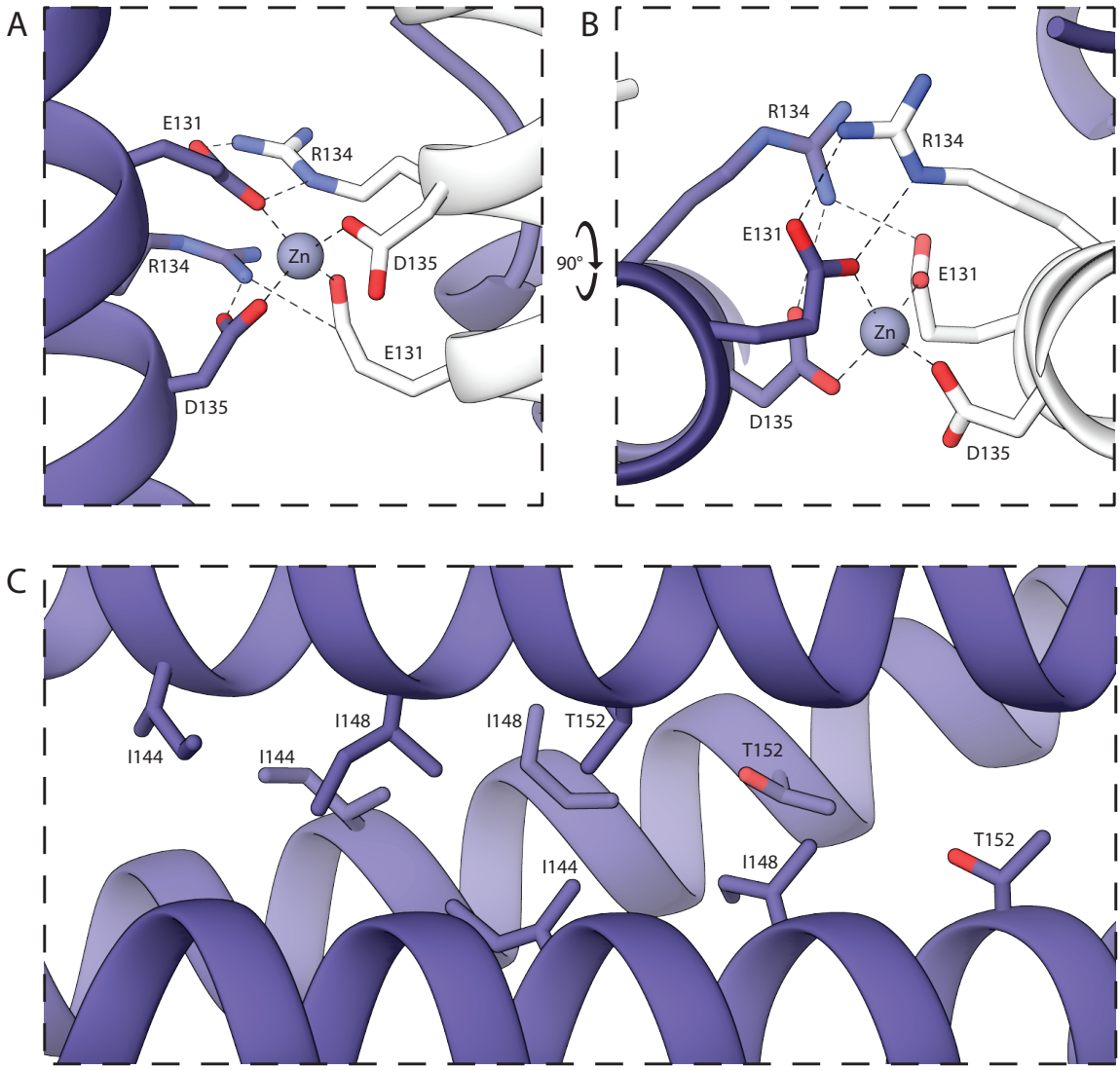


Figure 3.13: **Zn^{2+} co-ordination and register shift in the crystal of SAS-$5_{CC\text{-}L2}$.** (**A**, **B**) At the intercalation site between two laterally-packed SAS-$5_{CC\text{-}L2}$ trimers (see Figure 3.12A on page 107), four negatively charged residues (2× E131 and 2× D135) co-ordinate a Zn^{2+} cation. This site is the only lateral packing contact bridging SAS-$5_{CC\text{-}L2}$ trimers in the crystal. This interaction is further strengthened by salt bridges between residues E131/D135 and R143. Protein chains are drawn as cartoons, with relevant residues shown as sticks (coloured by hetero-atom). (**C**) The middle region of SAS-$5_{CC\text{-}L2}$ showing residues that make internal contacts within the core of the trimer. Note that while in the canonical parallel coiled coil corresponding residues in all chains lay in one plane perpendicular to the coiled-coil axis, the crystal structure of SAS-$5_{CC\text{-}L2}$ deviates from this rule, as *e.g.* residues I144, I148 and T152 are a part of such a plane. It appears that the structure of SAS-$5_{CC\text{-}L2}$ has been distorted by crystal packing forces, imposing a residue register shift of 0, +3 and +4.

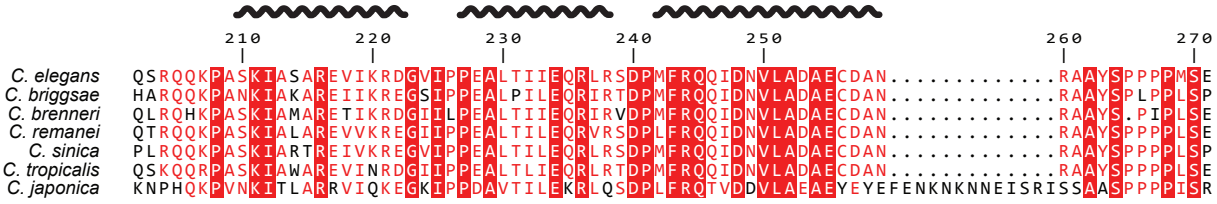


Figure 3.14: **Sequence conservation of SAS-5 *Implico* domain within the *Caenorhabditis* genus.** SAS-5 *Implico* sequences from *Caenorhabditis* genus were aligned with MergeAlign [61] and drawn with ESPript 3 [283]. A schematic drawing of protein secondary structure, as observed in the crystal structure of SAS-5_{Imp} is shown on top. Universally conserved residues are highlighted in red; conservatively substituted residues are coloured red.

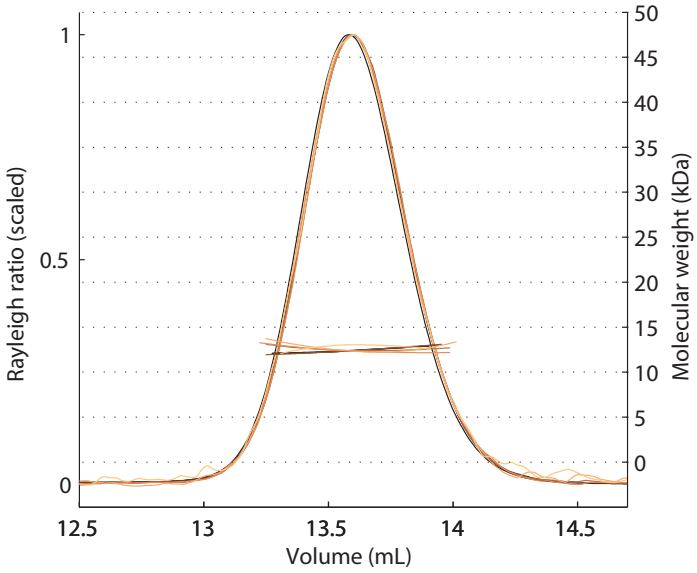


Figure 3.15: **SAS-5_{Imp} forms a stable dimer in solution.** Overlaid SEC-MALS chromatograms of SAS-5_{Imp} 1:1 dilution series, showing scaled light scattering intensity versus elution volume from a Superdex 75 10/300 column. Calculated molecular weights are shown as horizontal lines over chromatogram peaks, and correspond to the right-hand axis. Datasets are coloured according to injection concentration of the protein (black - the highest, light orange - the lowest). Note that SAS-5_{Imp} forms a stable dimer at all concentrations tested. Expected monomer mass: 6.5 kDa.

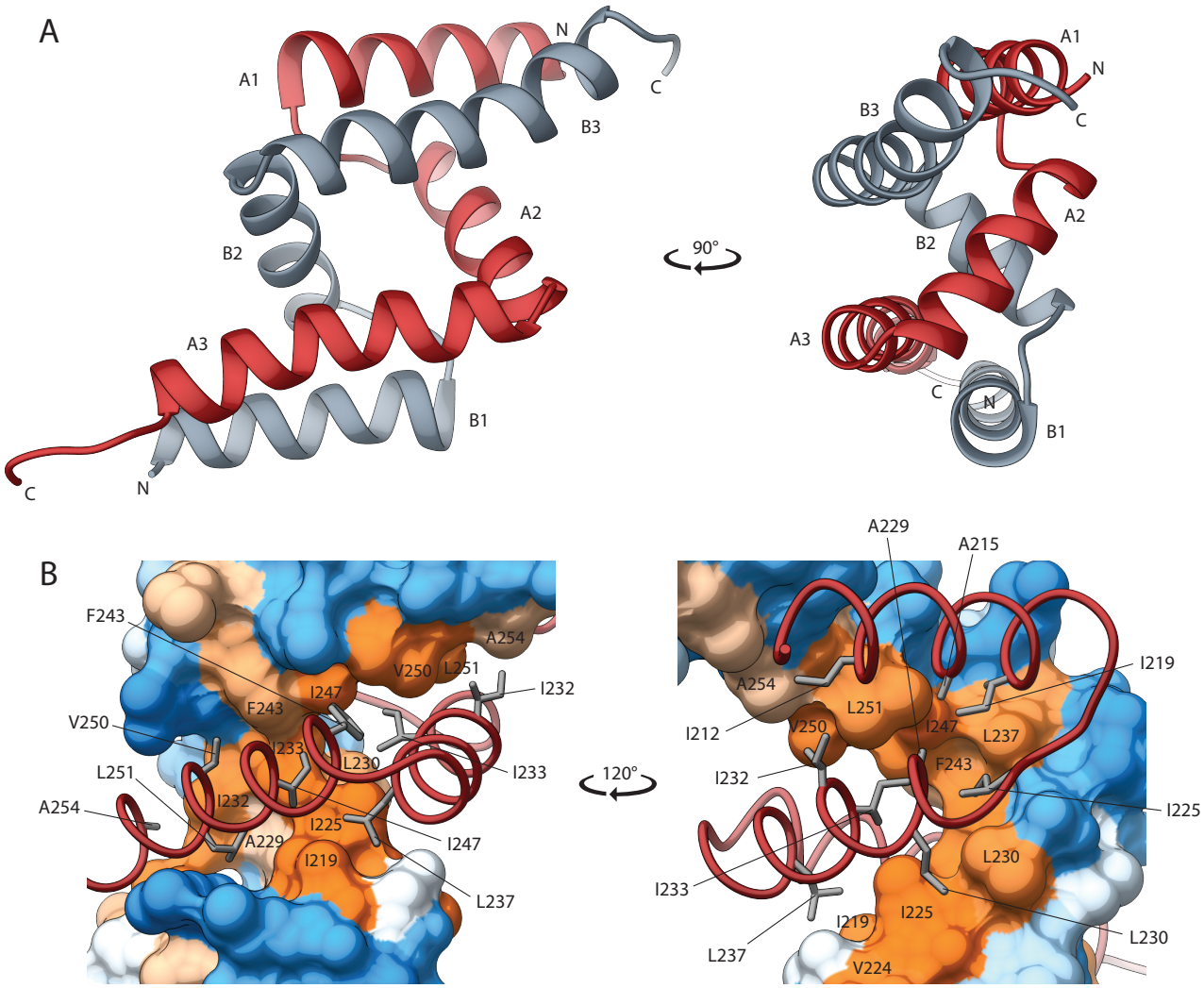


Figure 3.16: **Crystal structure of SAS-5$_{\mathrm{Imp}}$ reveals a novel type of a protein dimer.** (**A**) Two orthogonal views of the crystal structure of SAS-5$_{\mathrm{Imp}}$, drawn as cartoons. The α-helices for both protein chains (red and grey) are indicated as A1-A3 and B1-B3. Note that SAS-5$_{\mathrm{Imp}}$ forms an interlocked anti-parallel dimer. (**B**) Dimer interface of the SAS-5$_{\mathrm{Imp}}$ molecule is shown in two views related by a 120 ° rotation. One protein chain is represented as surface and the other as ribbon. Residues which make contacts at the dimer interface are shown as sticks on the ribbon chain, and their position at the other chain is indicated on the protein surface. Surface colouring follows the Kyte-Doolittle hydrophobicity scale [194], and is represented as a gradient from blue to orange, with blue corresponding to -4.5 in this scale (most hydrophilic), white to 0 and orange to 4.5 (most hydrophobic). Note that the core of SAS-5$_{\mathrm{Imp}}$ is exclusively made of hydrophobic residues.

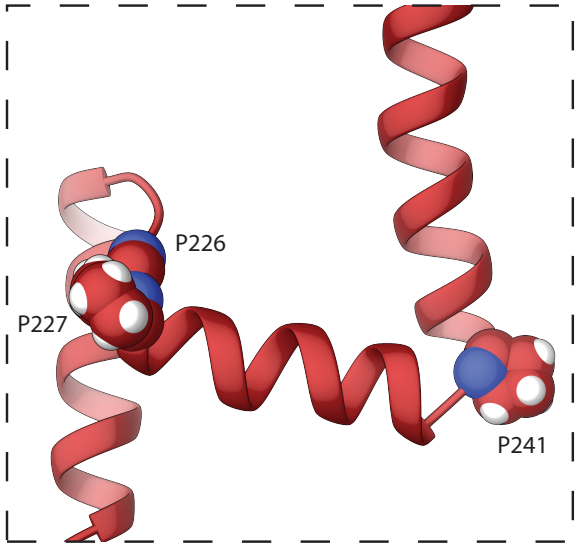


Figure 3.17: **Proline residues allow for tight turns in the compact SAS-5$_{\mathbf{Imp}}$ dimer.** A zig-zag structure of a single SAS-5$_{\mathrm{Imp}}$ chain is drawn as a cartoon. Location of proline residues at tight turns between α-helices is indicated (space-filling representation, hetero-atom colouring).

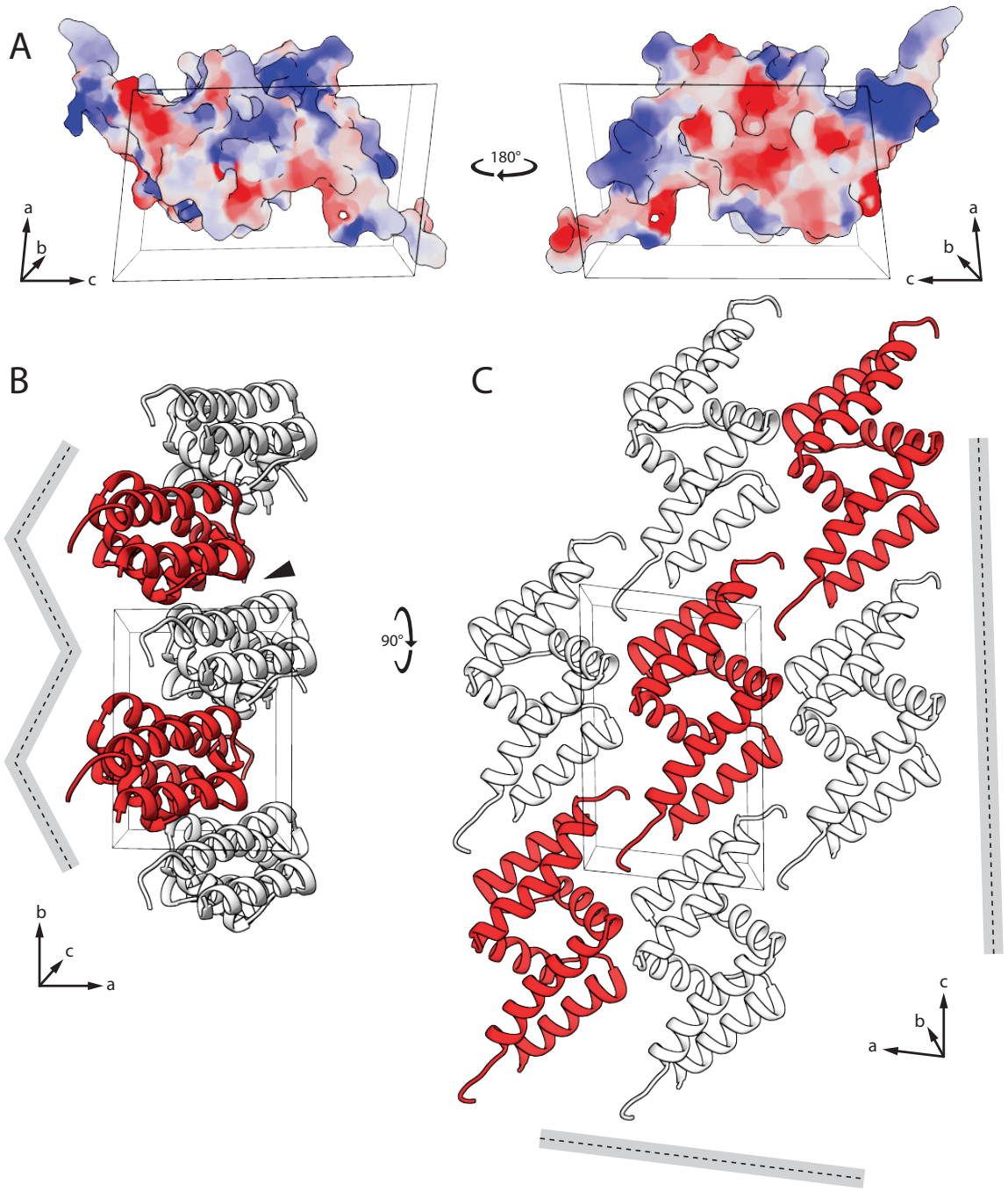


Figure 3.18: **Packing of molecules in the SAS-5$_{\mathbf{Imp}}$ crystal.** (**A**) Crystal packing interface between dimers of SAS-5$_{\mathrm{Imp}}$ (location of this interface is indicated by the arrowhead in **B**). Protein molecules are represented as surface and coloured by electrostatic potential (blue - positive charge, red - negative charge, at neutral pH). Note that one surface is dominated by negative charge (right), and the other by positive charge (left), but to smaller extent. (**B**) Charge difference between the two faces of SAS-5$_{\mathrm{Imp}}$ dimer enabled tight packing in the vertical axis (crystallographic b-axis). Schematic drawing on the left (grey with a dashed line in the centre) indicates molecule displacement along the packing axis. (**C**) SAS-5$_{\mathrm{Imp}}$ packs extremely tightly in the a-c plane, using every available surface to make contacts, thus completely filling the available space, and leaving only minor fraction for solvent. Note that this crystal diffracted X-rays to 1.0 Å resolution.

Chapter 4

Oligomerisation of SAS-5 facilitates daughter centriole formation

In the previous chapter, I described the route which led us to understanding of the molecular architecture of SAS-5. Those results revealed that on top of two disordered termini, which serve as binding platforms for other centriolar proteins, SAS-5 is composed of two independently folded domains, the coiled coil and the *Implico*. Both of these domains were found to self-oligomerise, but in two distinct modes. The parallel coiled coil is a moderately stable, concentration-dependent trimer, *i.e.* depending on its concentration, it can either take monomeric, dimeric or trimeric states. On the other hand, the anti-parallel *Implico* is a strong and stable dimer at all concentrations tested. Together, the two oligomerisation domains of SAS-5 have the potential to drive formation of protein assemblies of order higher than merely dimers or trimers. In this chapter, I will describe the efforts we took to understand the higher-order oligomerisation of SAS-5, and also explain how we used this knowledge to shed some light on the role of these SAS-5 oligomers in centriole assembly.

4.1 Contribution statement

Part of the work presented in this chapter was done in collaboration with laboratories of Professor Dame Carol Robinson at the Department of Chemistry, University of Oxford and Professor Pierre Gönczy at the Swiss Institute for Experimental Cancer Research, Swiss Federal Institute of Technology in Lausanne, Switzerland. I expressed, purified and buffer-exchanged SAS-$5_{(2\text{-}265)}$ into a suitable buffer for mass

spectrometry (MS) work. Then, working closely with Dr Jun Yan of the Robinson Laboratory, we acquired and analysed the native MS spectra of SAS-$5_{(2\text{-}265)}$. I recoded *sas-5* for RNAi resistance, introduced oligomerisation mutations, and inserted the recoded gene into a suitable vector, which was then used to coat gold particles for biolistic bombardment of worms. Worm bombardment, isolation of lines expressing SAS-5 variants, and cell biology assays testing centriole duplication and protein localisation were performed by Dr Nicola Dynes and Mr Sheng Kai Pong of the Gönczy Laboratory. All other biochemical and biophysical work was carried out by me.

4.2 SAS-5 assembles into tetramers and hexamers in solution

Given that SAS-5 comprises two independently self-oligomerising domains, we speculated what oligomeric state would SAS-5 adopt if it was expressed as one continuous polypeptide chain. We envisioned that SAS-5 would readily dimerise even at low concentrations via the *Implico* domain (Figure 4.1A on page 115), as oligomerisation of this domain is inherent to its own folding into a globular domain, and therefore its self-oligomerisation will be much stronger compared to that of the coiled coil. We envisioned that the *Implico*-mediated SAS-5 dimer would constitute the basic unit of oligomerisation, and hypothesised that once the concentration of SAS-5 is increased, the *Implico*-mediated dimers could then come together to form higher-order configurations: tetramers and hexamers, through concentration-dependent oligomerisation of the coiled coil (Figure 4.1B, C on page 115).

To test such hypothesis, we performed SEC-MALS experiments on MsyB-tagged SAS-5_{FLEX}, which revealed concentration-dependent oligomerisation of SAS-5 from dimers to an approximately equimolar population of tetramers and hexamers (Figure 4.2A, B on page 116). Difficulty of working with SAS-5_{FLEX}, which due to its disordered N- and C-terminal segments was particularly prone to proteolytic degradation during the purification process, prompted us to use an MsyB-tagged SAS-5 construct that lacked the disordered C-terminus (SAS-$5_{(2\text{-}265)}$), and therefore was less susceptible to proteolysis. We were able to obtain large quantities (tens of milligrams) of pure

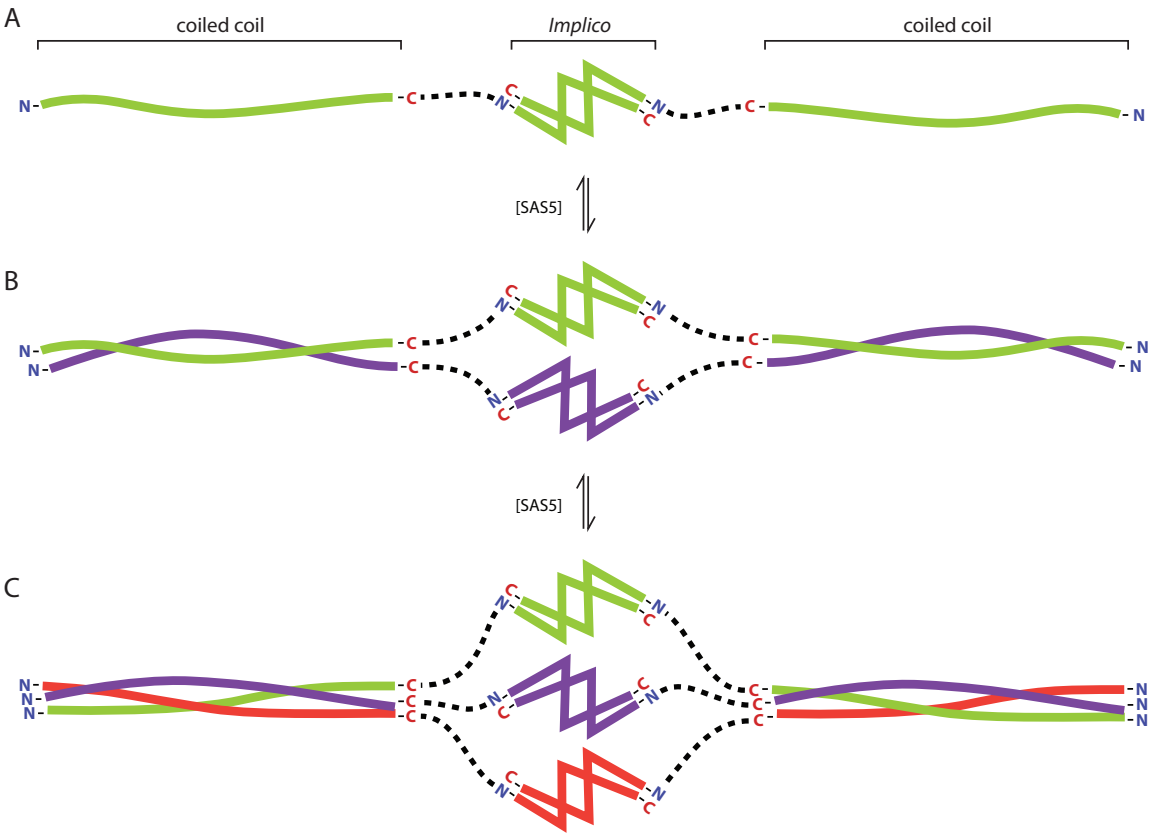


Figure 4.1: **Hypothetical model of higher-order oligomerisation of SAS-5.** (**A**) The coiled coil and *Implico* domains are shown as one continuous polypeptide chain of SAS-5 in schematic representation. Note that the disordered N- and C-termini were omitted in this drawing. At low protein concentrations, *Implico* mediates strong dimer formation, which exists as the basic unit of higher-order oligomerisation of SAS-5. (**B**, **C**) Once the concentration of SAS-5 is increased, potentially after localisation to centrioles, *Implico*-mediated dimers could come together through the coiled coil to form, first, a tetramer (**B**; additional purple dimer) and ultimately a hexamer (**C**; additional red dimer) in a concentration-dependent manner, thus saturating all the binding sites in a compact oligomer. This model of oligomerisation does not take into account protein aggregation mediated by the *sticky strand* at the C-terminus of SAS-5.

and non-degraded MsyB-SAS-$5_{(2\text{-}265)}$, enabling us to also use higher protein concentrations for SEC-MALS injections. SEC-MALS experiments with MsyB-SAS-$5_{(2\text{-}265)}$ (Figure 4.2C, D on page 116) produced analogous results to MsyB-SAS-5_{FLEX}. However, despite using MsyB as a solubility enhancer, the consistency of protein changed into a gel at very high concentrations (∼2 mM), which could only be solubilised with high salt (∼2 M NaCl). Although the protein returned into solution after the salt treatment, it contained a large amount of aggregated material, which skewed SEC-MALS results. Thus, it was technically not feasible to use SEC-MALS for detection of pure hexamers of SAS-5 (or perhaps even higher-order species) due to protein aggregation.

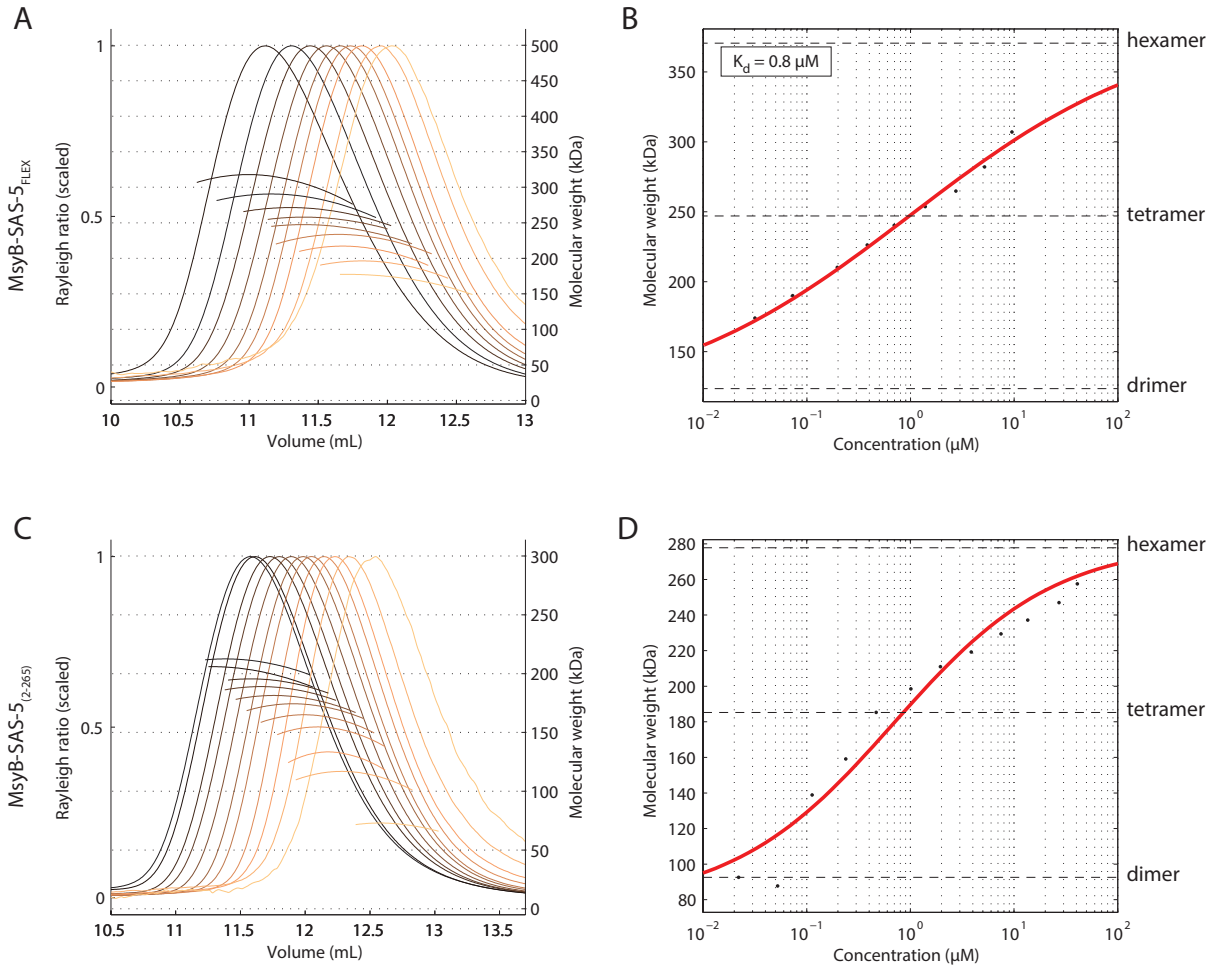


Figure 4.2: **Higher-order assemblies of SAS-5 in solution.** (**A**) Overlaid SEC-MALS chromatograms of MsyB-tagged SAS-5_{FLEX} 1:1 dilution series, showing scaled light scattering intensity versus elution volume from a Superose 6 10/300 column. Calculated molecular weights are shown as horizontal lines over chromatogram peaks, and correspond to the right-hand axis. Datasets are coloured according to injection concentration of the protein (black - the highest, light orange - the lowest). Expected MsyB-SAS-5_{FLEX} monomer mass: 61.7 kDa (**B**) Average molecular weight from SEC-MALS experiments is plotted as a function of on-column protein concentration. Apparent K_d and molecular sizes of expected protein species are indicated. At the highest on-column concentration in this dilution series (9.5 μM), MsyB-SAS-5_{FLEX} reached an equimolar hexamer-to-tetramer ratio. Further SAS-5 up-concentration for column injections resulted in aggregated protein (see Section 4.2 on page 114 for details). The experimental data were fitted under the assumption of ultimate hexamer formation and yielded K_d values comparable to those of the SAS-5 coiled coil in isolation (compare panel **B** with Figure 3.8B). (**C**, **D**) Similar data from MsyB-tagged SAS-$5_{(2\text{-}265)}$. At the highest on-column concentration point (41 μM), MsyB-SAS-$5_{(2\text{-}265)}$ reaches a 3:1 hexamer to tetramer ratio. Expected MsyB-SAS-$5_{(2\text{-}265)}$ monomer mass: 46.3 kDa.

In order to bypass such limitation, we used native mass spectrometry, which in contrary to SEC-MALS, examines individual protein species rather than their bulk average, and the protein material is not diluted in the process. The results from native MS of SAS-$5_{(2\text{-}265)}$ also supported the presence of higher-order protein assemblies up to hexamers in solution (Figure 4.3 on page 117). However, since the experiment was conducted in the gas phase, energy of collisions between the protein oligomer and neutral residual gas caused SAS-5 oligomers to fall apart (*i.e.* both *Implico* and coiled-coil domains break at some rate), and hence detectable presence of intermediate species: monomers, trimers and pentamers.

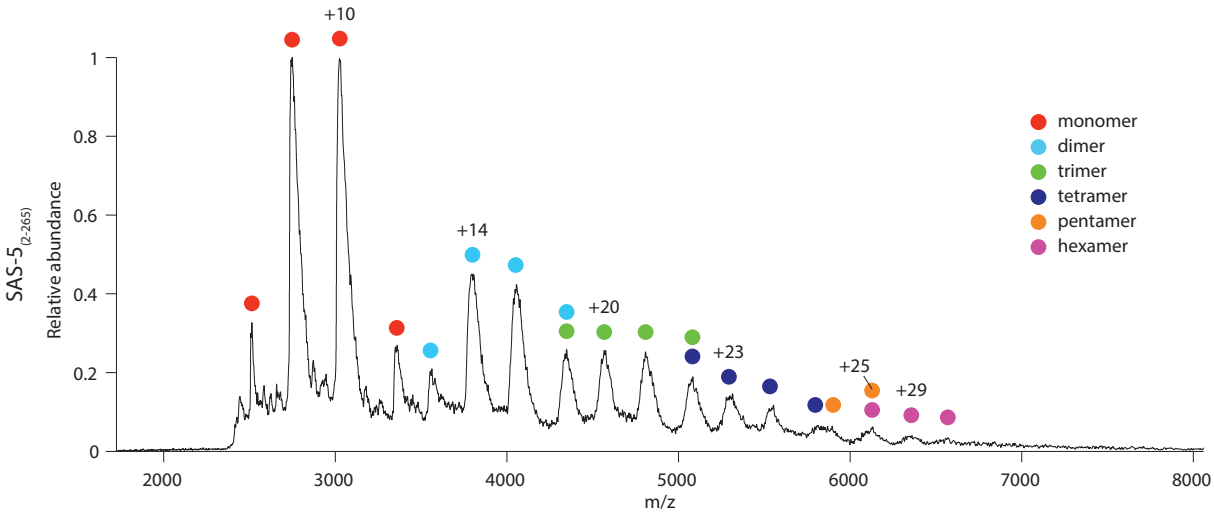


Figure 4.3: **Native mass spectrometry reveals presence of SAS-5 hexamers in solution.** Native mass spectrometry electrospray ionization spectrum of 20 μM SAS-$5_{(2\text{-}265)}$ is plotted as relative abundance of protein species versus mass to charge ratio (m/z). Protein oligomeric forms and their charged states are indicated next to the corresponding peaks. Presence of odd-numbered oligomeric forms (monomers, trimers and pentamers) in the spectrum is likely due to in-flight breakdown of higher-order oligomeric forms of SAS-5.

Taken together, these data support the notion that the coiled coil of SAS-5 promotes further association of *Implico*-mediated dimers towards higher-order assemblies.

4.3 Disrupted oligomerisation of SAS-5

Through X-ray crystallography and a range of biophysical techniques, we have established the molecular basis for the oligomerisation of SAS-5. In order to use this knowledge in probing the function of SAS-5 *in vivo*, mutants defective in their ability

to form higher-order species were required. Such mutants would also be useful in the thorough examination of the architectural model of higher-order assemblies of SAS-5, which we proposed earlier (see Section 4.2 on page 114).

4.3.1 Disruption of individual domains

We reasoned that disruptive mutagenesis of SAS-5 oligomerisation would be most effective if we targeted the core hydrophobic interactions which drive the oligomerisation process. Based on the crystal structures of SAS-5_{CC} and SAS-5_{Imp} which we solved earlier (see Chapter 3), we were able to locate a number of residues, that upon dimerisation of the *Implico* or trimerisation of the coiled coil, are in physical contact with their corresponding partners of the neighbouring chains (*e.g.* I247 of chain A interacts with I247 of chain B, see Figure 3.8D on page 102 and Figure 3.16B on page 110). Mutating these amino acids to Glu or Arg completely altered their original chemical properties, and effectively acted as a disrupting device that severely impaired oligomerisation attributes of the coiled coil and *Implico* domains.

We found that mutations in positions *a* and *d* of one of the strongest heptad repeats of SAS-5_{CC}, L141E and L144E, showed significant reduction of α-helical content in the coiled coil, and a striking reduction in stability (Tm $<$10 °C), as measured by CD (Figure 4.4 on page 119). Moreover, mutants of SAS-5_{Imp}, I247E and I247R, showed absence of α-helical structure and a complete loss of co-operativity upon thermal unfolding, as measured by CD (Figure 4.5 on page 120).

4.3.2 Disruption of higher-order assemblies

We showed that surgical introduction of mutations in the hydrophobic core of individual folded domains of SAS-5 severely disrupts the quaternary structure of these domains. To probe the effects of these disruptions further, we engineered the same substitutions in the SAS-$5_{(2\text{-}265)}$ construct to selectively disrupt each oligomerisation interface whilst leaving the other one intact. We found that the thermal unfolding profile of the SAS-$5_{(2\text{-}265)}$ L141E mutant, as monitored by CD, showed reduction in stability of the coiled coil by approximately 25 °C, whereas stability of the *Implico* domain was not affected (Figure 4.6 on page 122, blue trace). Unlike the globular *Implico* domain, the coiled coil has a large and elongated self-interaction surface, thus

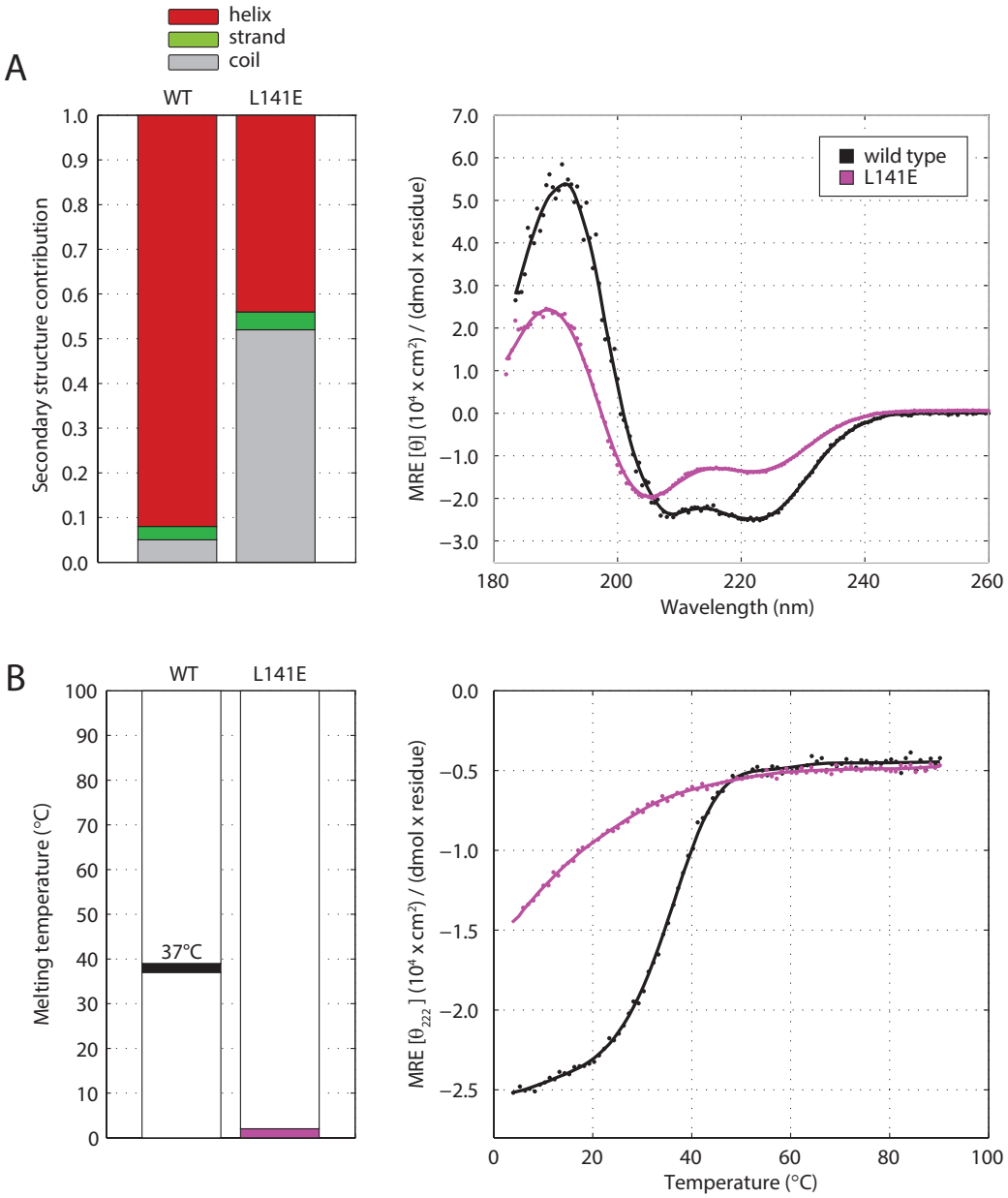


Figure 4.4: **Point mutation disrupts the coiled-coil domain.** (**A**) Overlaid CD spectra of SAS-5_{CC} wild type (black) and L141E mutant (magenta) collected at 10 °C and plotted as per-residue molar elipticity against wavelength. Bar charts on the left depict contribution of each secondary-structure elements (random coil in grey, β-strand in green, α-helix in red) to the CD spectrum, deconvoluted with CDSSTR [164]. Note that this is a semi-quantitative analysis. (**B**) Thermal unfolding profiles of SAS-5_{CC} wild type and L141E mutant were recorded at 222 nm, and plotted as per-residue molar elipticity against temperature. Bars on the left depict melting transition temperatures observed for each sample. Note the α-helical signal loss of the second minimum (222 nm; **A**) and a shift of the 193 nm and 208 nm peaks to shorter wavelengths. Also, the melting transition has shifted (**B**) from 37 °C to a value close to 0 °C. CD spectra and unfolding profiles of the I144E mutant were similar to those of L141E.

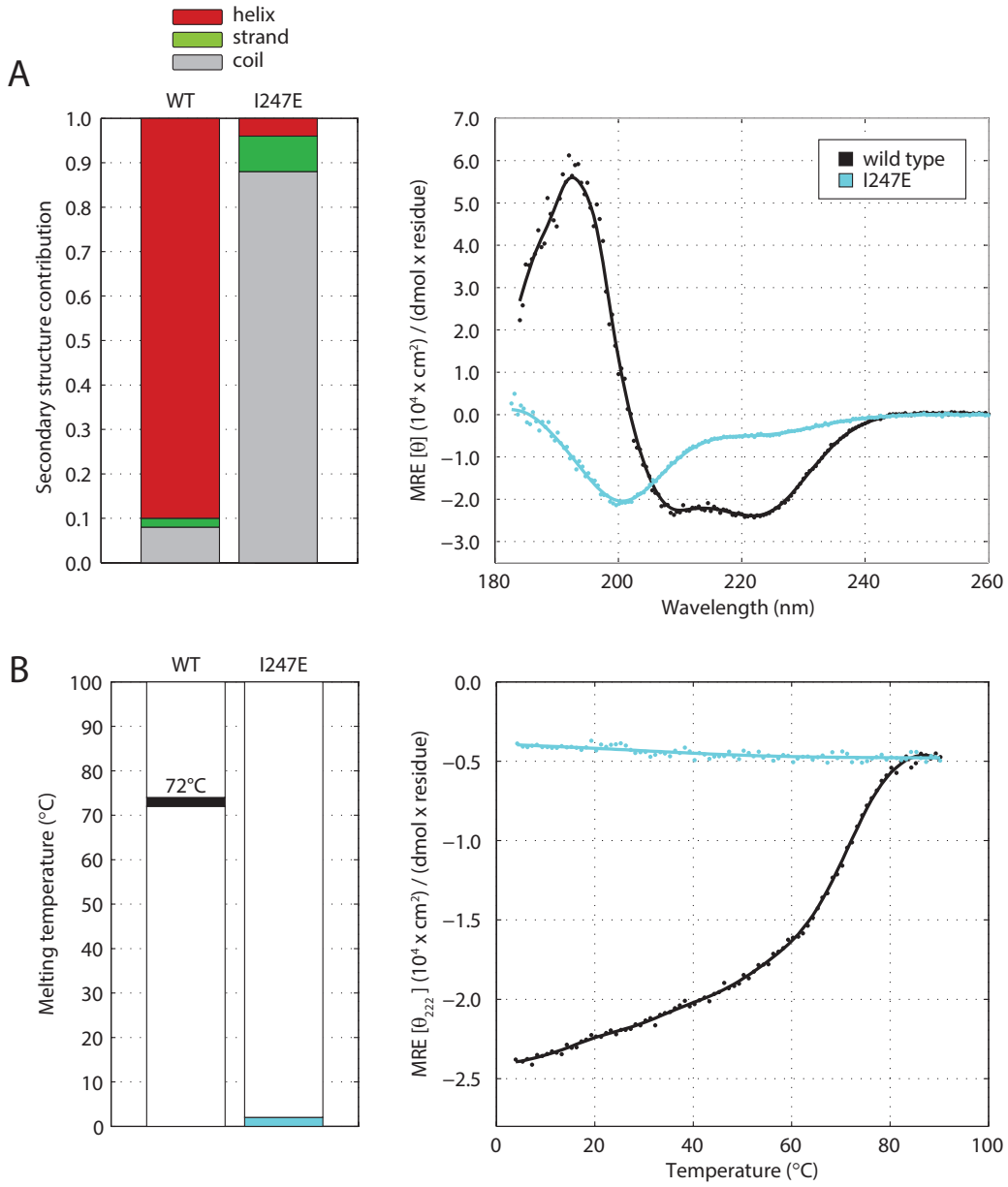


Figure 4.5: **Point mutation disrupts the *Implico* domain.** (**A**) Overlaid CD spectra of SAS-5_{Imp} wild type (black) and I247E mutant (cyan) collected at 10 °C and plotted as per-residue molar elipticity against wavelength. Bar charts on the left depict contribution of each secondary-structure elements (random coil in grey, β-strand in green, α-helix in red) to the CD spectrum, deconvoluted with CDSSTR [164]. Note that this is a semi-quantitative analysis. (**B**) Thermal unfolding profiles of SAS-5_{Imp} wild type and I247E mutant were recorded at 222 nm, and plotted as per-residue molar elipticity against temperature. Bars on the left depict melting transition temperatures observed for each sample. Note the dramatic loss of α-helical signal at 193 nm and 222 nm (**A**), and a shift of the 193 nm and 208 nm peaks to shorter wavelengths. Also, the co-operative melting transition has completely disappeared (**B**), indicating that the I247E mutation led to a loss of the tertiary structure. CD spectra and unfolding profiles of the I247R mutant were similar to those of I247E.

disrupting it at only one point along the super-coil is bound to be incomplete. Therefore, based on the SAS-5_{CC} crystal structure, we introduced additional substitutions to core hydrophobic residues in the C-terminal part of the coiled coil (SAS-$5_{(2\text{-}265)}$ L141E/L164E, L141E/M167E and L141E/M171E, see Figure 3.8C on page 102). All three mutation pairs effectively abrogated coiled-coil formation, again with no effect on the *Implico* domain (Figure 4.6 on page 122, cyan trace).

As anticipated, disruptions of the coiled coil in either SAS-$5_{(2\text{-}265)}$ L141E or L141E/ M167E mutants yielded a stable dimer in SEC-MALS experiments (Figure 4.7A, B on page 123), mediated by the *Implico* domain. The single coiled-coil mutant (SAS-$5_{(2\text{-}265)}$ L141E) shows a minimal transition in mass and peak elution volume with increasing injection concentrations - an indication of partially functional coiled-coil domain. This behaviour is almost completely eliminated in the double coiled-coil mutant (SAS-$5_{(2\text{-}265)}$ L141E/ M167E), indicative of a much stronger disruption.

Furthermore, we found that a SAS-$5_{(2\text{-}265)}$ I247E mutant displayed a single-step thermal unfolding profile as measured by CD, with melting temperature similar to that of the SAS-5 coiled coil in isolation (compare Figure 4.6 on page 122, red trace with Figure 3.4 on page 91, magenta trace). SEC-MALS analysis of the same construct shows a concentration-dependent trimerisation, again consistent with formation of the coiled coil of SAS-5 alone (compare Figure 4.7C, D on page 123 with Figure 3.8A, B on page 102). Finally, a SAS-5 mutant where both oligomerisation interfaces are disrupted, SAS-$5_{(2\text{-}265)}$ L141E/I247E, showed no co-operativity upon thermal unfolding, as measured by CD (Figure 4.6 on page 122, green trace), and remained monomeric in SEC-MALS experiments (Figure 4.7E on page 123).

Together, our results support the notion that the two domains of SAS-5 self-associate independently from one another, and that together they allow formation of higher-order SAS-5 assemblies.

4.4 Higher-order assemblies of SAS-5 are essential for centriole duplication *in vivo*

Selective disruption of the oligomerisation domains of SAS-5 by single point mutations *in vitro* constitutes a set of precise tools to probe the role of higher-order assemblies

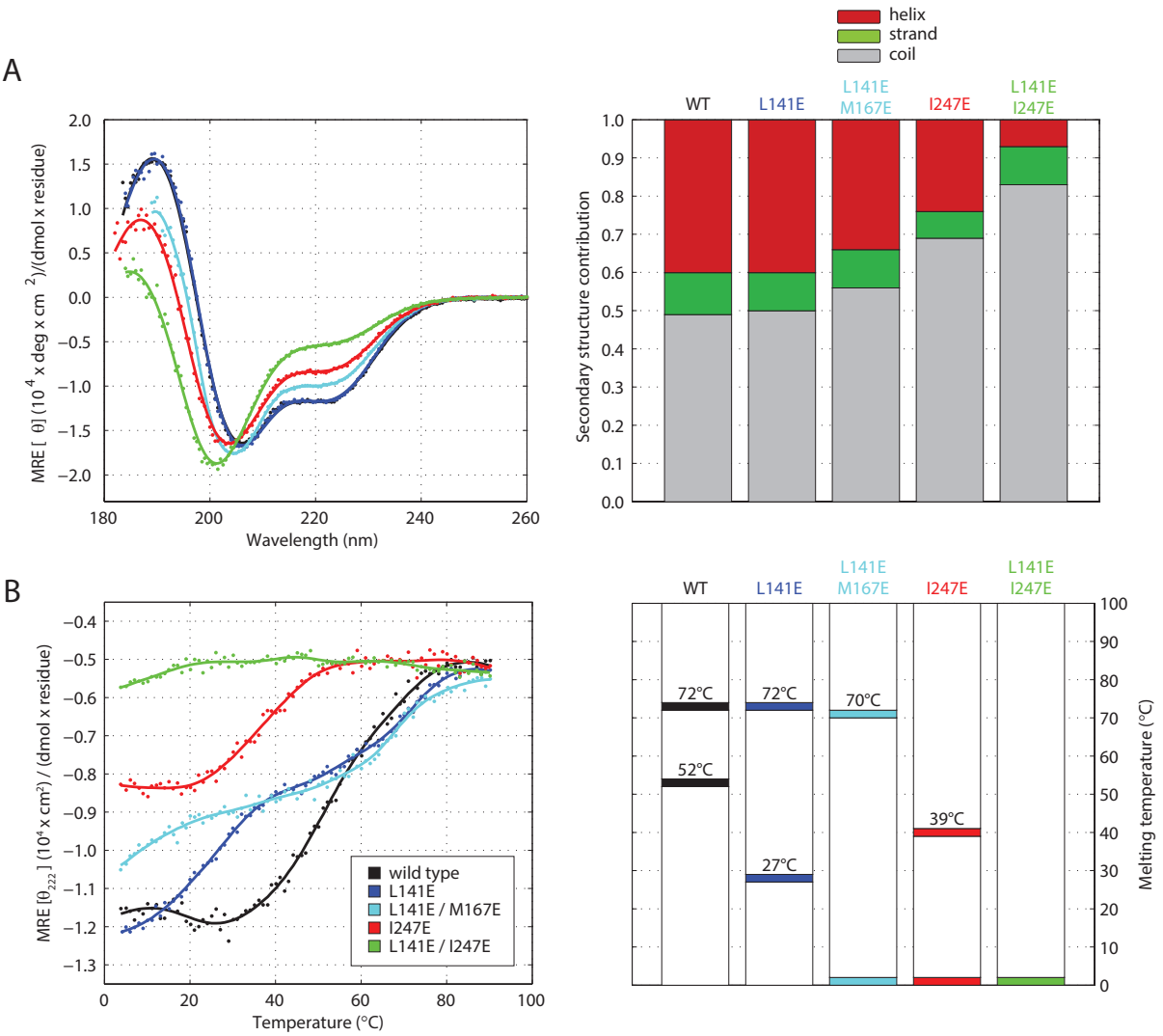


Figure 4.6: **Selective disruption of SAS-5 oligomerisation interfaces.** (**A**) Overlaid CD spectra of SAS-$5_{(2\text{-}265)}$ wild type (black) and mutants that disrupt the coiled coil (L141E - blue, L141E/M167E - cyan), the *Implico* (I247E - red), or both (L141E/I247E - green) oligomerisation interfaces. Data (collected at 10 °C) is plotted as per-residue molar elipticity against wavelength. Bar charts on the right depict contribution of each secondary-structure elements (random coil in grey, β-strand in green, α-helix in red) to the CD spectrum, deconvoluted with CDSSTR [164]. This is a semi-quantitative analysis. Note that the wild-type SAS-$5_{(2\text{-}265)}$ has a considerable contribution from a random coil (first minimum at 205 nm instead of 208 nm, as in all-α-helical proteins), which reflects the inclusion of the N-terminal segment (residues 2-122) in this construct. Also, introduction of mutations progressively reduces the α-helical signal at 193 nm and 222 nm, and the first two peaks (193 nm and 205 nm) shift to shorter wavelengths. (**B**) Thermal unfolding profiles of the same proteins, recorded at 222 nm and plotted as per-residue molar elipticity against temperature. Bars on the right depict melting transition temperatures observed for each sample. Note the progressive loss of protein tertiary structure after introduction of mutations. Unlike in case of the isolated coiled coil, the L141E mutation in the coiled coil does not completely disrupt this domain (T_m ~27 °C; compare with Figure 4.4B on page 119). This is likely due to the neighbouring and strongly oligomerising *Implico*. Addition of an extra M167E mutation, however, shifts the melting transition to a value below 0 °C. Mutating both oligomerisation domains (L141/I247E) leads to a complete loss of co-operative unfolding. CD spectra and unfolding profiles of L141E/L164E and L141E/M171E were similar to those of L141E/M167E.

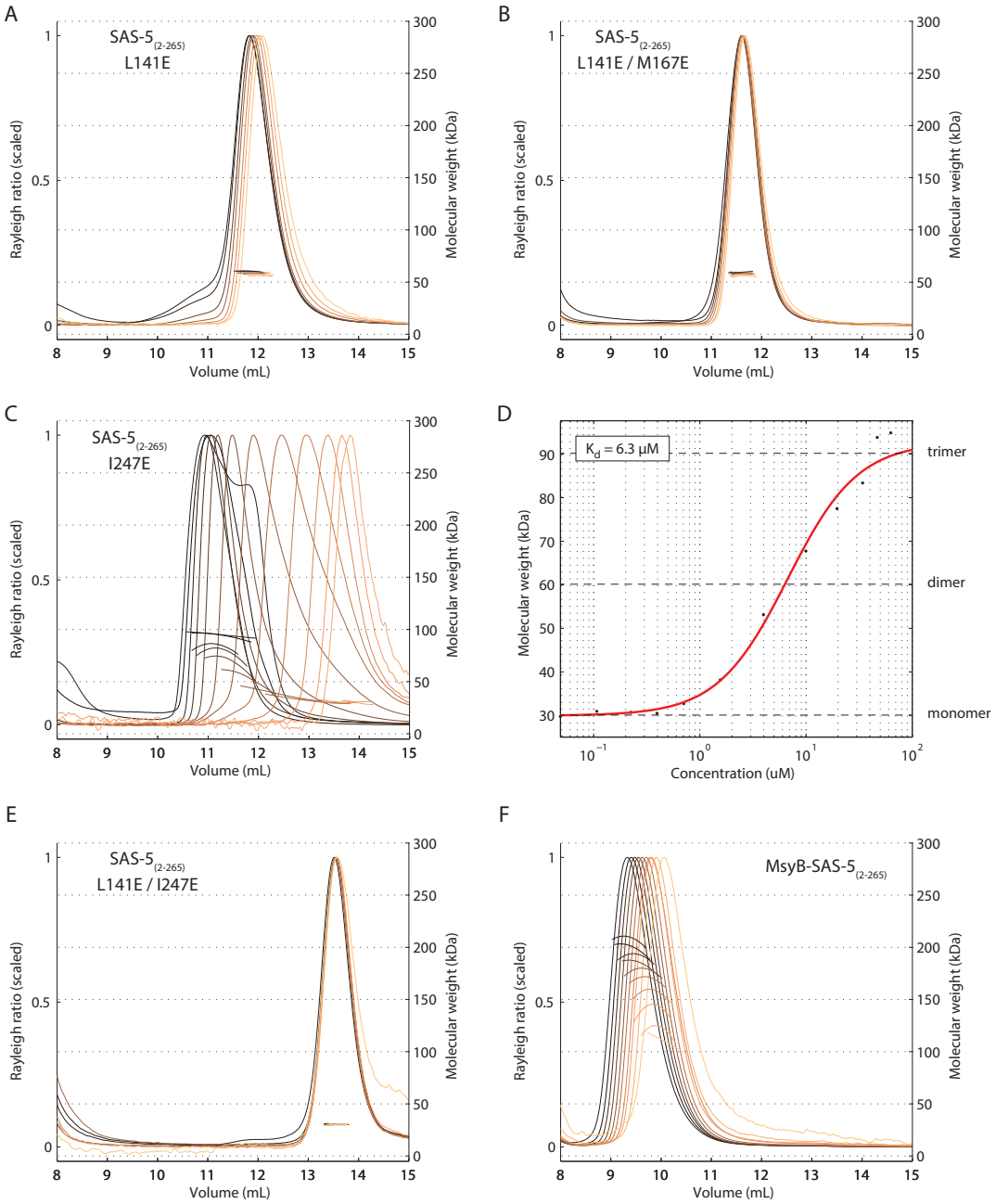


Figure 4.7: **Selective disruption of SAS-5 oligomerisation interfaces - SEC-MALS.** (**A**) Overlaid SEC-MALS chromatograms of SAS-$5_{(2\text{-}265)}$ wild type (MsyB-tagged; **F**) and mutants that disrupt the coiled coil (L141E - **A**, L141E/M167E - **B**), the *Implico* (I247E - **C**), or both (L141E/I247E - **E**) oligomerisation interfaces. Scaled light scattering intensity is plotted versus elution volume from a Superdex 200 10/300 column. Calculated molecular weights are shown as horizontal lines over chromatogram peaks, and correspond to the right-hand axis. Datasets are coloured according to injection concentration of the protein (black - the highest, light orange - the lowest; 1:1 dilution series). Expected monomer masses: SAS-$5_{(2\text{-}265)}$ - 30.1 kDa, MsyB-SAS-$5_{(2\text{-}265)}$ - 46.3 kDa. Chromatograms of SAS-$5_{(2\text{-}265)}$ I247E mutant at the highest two injection concentrations show evidence of size-exclusion column overload with protein material. These two samples reach the mass of a trimer and their peaks instead of shifting towards earlier elution, are broadening. Note the long horizontal lines running across broad elution peaks, indicating stable mass across the entire peak length. (**D**) Average molecular weight from SEC-MALS experiments of the SAS-$5_{(2\text{-}265)}$ I247E *Implico* mutant is plotted as a function of on-column protein concentration. Apparent K_d and molecular sizes of expected protein species are indicated. Note the similarity of the K_d to the coiled coil in isolation (Figure 3.8B on page 102).

of SAS-5 in centriole duplication *in vivo*. For those *in vivo* studies, we used the iconic model organism, the nematode worm *Caenorhabditis elegans*, which is in fact the organism where the involvement of SAS-5 in centriole assembly was originally established [78, 88, 296].

4.4.1 Phenotype of *C. elegans* SAS-5

Similarly to many other animal species, *C. elegans* centrioles are eliminated during oogenesis, leaving sperm as the sole contributor of a centriole pair to the newly fertilised embryo. The first round of centriole duplication in the zygote yields two centrosomes, each with a pair of centrioles. During the first mitotic division, the centrosomes are positioned at the opposite sides of the nucleus, from where they direct assembly of a bipolar spindle. Once the genetic material has been segregated, the bipolar spindle disassembles and the cell is split into two daughter cells following cytokinesis. Each daughter cell inherits one centrosome, *i.e.* a pair of centrioles. At this point, a new cell cycle begins, and the centrioles undergo separation and duplication. This is followed by another bipolar spindle assembly in each blastomere, and results in a signature 4-cell stage configuration after mitotic exit (Figure 4.8A on page 126).

This situation is remarkably different in worms treated with *sas-5* RNAi. RNAi treatment of worms relies on the introduction of double-stranded RNA (dsRNA), which is complementary to the gene of interest, to an adult hermaphrodite worm. Without any alterations to worm DNA, the treatment with *sas-5* dsRNA leads to formation of oocytes that are completely devoid of any of translated *sas-5*. The effect of such depletion is then monitored during embryonic development after fertilisation. RNAi does not work efficiently in sperm, and therefore upon fertilisation the sperm contributes two normal centrioles to the zygote. Each of these paternally-contributed centrioles recruits pericentriolar material, so that together, they can drive bipolar spindle assembly during the first mitotic division [88]. However, due to depletion of the maternal pool of SAS-5, centriole duplication does not occur. What follows is truly harmful to worm embryos. Because blastomeres inherited only one paternal centriole each, they can no longer organise a bipolar spindle. Instead, a monopolar spindle is assembled, which unlike the bipolar spindle, is not able to efficiently segregate genetic

material. As a result, the development is stalled, and a four-cell configuration is never observed in *sas-5 (RNAi)* embryos (Figure 4.8B on page 126).

4.4.2 Live cell imaging of mutant SAS-5 embryos

In order to assess the role of SAS-5 oligomerisation *in vivo*, we generated *C. elegans* transgenic animals expressing either GFP-SAS-5 (wild-type), or GFP-SAS-5 L141E or GFP-SAS-5 I247E mutant versions. All exons in the *sas-5* sequence were recoded in the transgenic constructs to confer resistance to RNAi directed against endogenous *sas-5*.

To monitor the effect of transgene introduction on centriole duplication, and hence worm viability, we used time-lapse differential interference contrast (DIC) microscopy coupled with RNAi-mediated depletion of endogenous *sas-5*.

We confirmed the expression of the GFP-fused SAS-5 by Western blot analysis and live imaging after *sas-5* RNAi depletion (Figure 4.9 on page 127 and Figure 4.10 on page 128). Importantly, using time-lapse DIC microscopy to analyse the three transgenic lines generated in this study, we found that expression of GFP-SAS-5 fully rescued centriole duplication failure seen under *sas-5 (RNAi)* (Figure 4.8C on page 126). All RNAi-treated embryos which expressed the RNAi-resistant wild-type copy of GFP-SAS-5 progressed to the 4-cell stage, analogously to wild-type embryos which were not treated with RNAi.

In stark contrast, GFP-SAS-5 mutants that disrupt either the coiled coil (L141E) or the *Implico* oligomerisation domains (I247E), were unable to rescue the *sas-5 (RNAi)* phenotype, such that all embryos underwent monopolar spindle assembly in each blastomere at the second cell cycle, and thus failed to result in a 4-cell configuration (Figure 4.8D, E on page 126).

Those observations suggest that both SAS-5 oligomerisation interfaces, and hence formation of SAS-5 assemblies larger then merely dimers or trimers, are essential for SAS-5 function *in vivo*.

4.4.3 Centriolar localisation of SAS-5 mutants

There are two sequential steps in the life of a SAS-5 molecule that lead to building of a new centriole. First, following translation in the cytoplasm, SAS-5 needs to travel

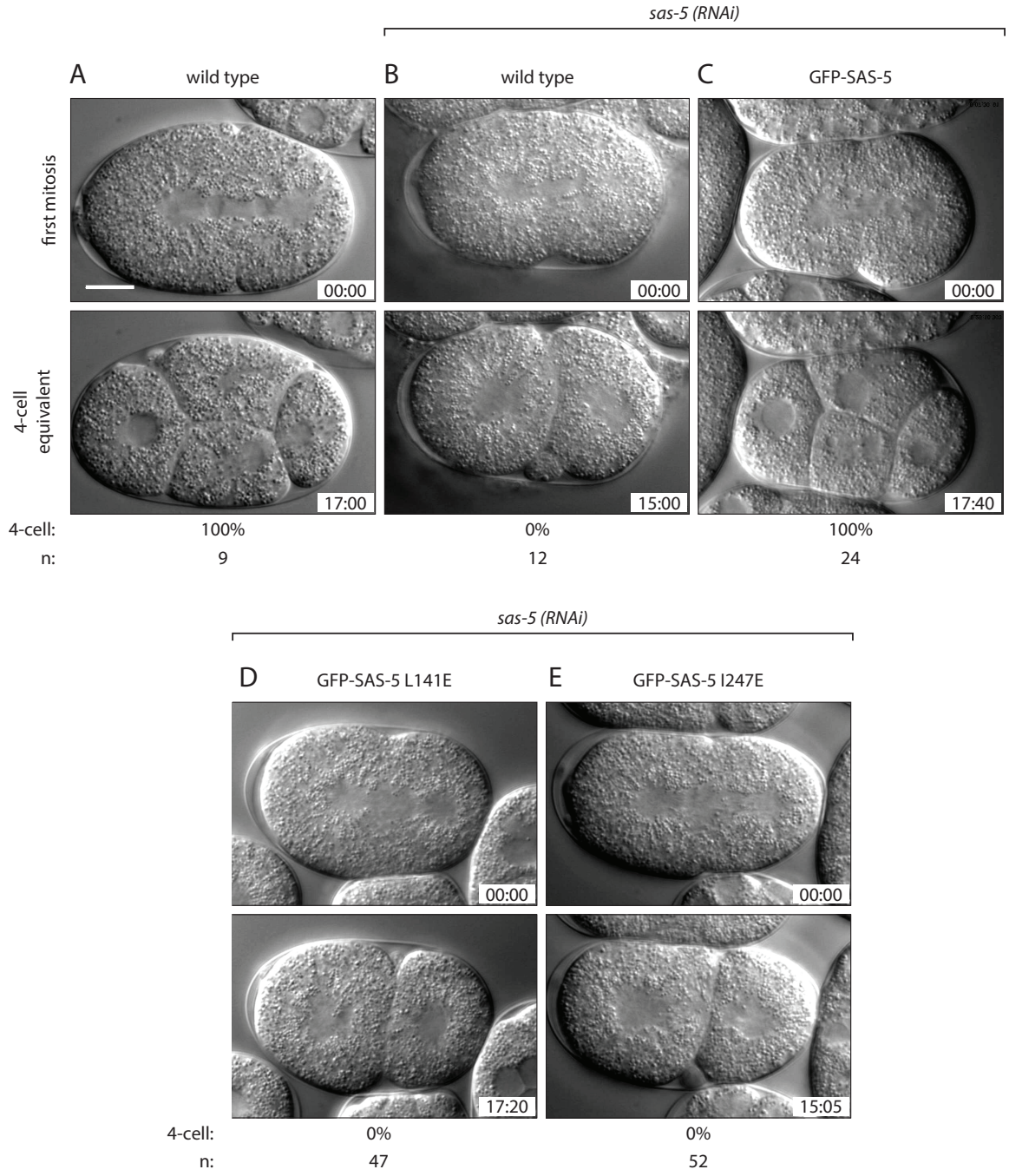


Figure 4.8: **Higher-order oligomers of SAS-5 are essential for centriole duplication in worm embryos.** Wild type (**A**, **B**), *gfp::sas-5* (**C**), *gfp::sas-5[L141E]* (**D**), and *gfp::sas-5[I247E]* (**E**) adult worms were subjected to *sas-5* RNAi or left untreated (**A**). The resulting embryos were imaged by time-lapse differential interference contrast (DIC) microscopy. Embryos were imaged at the onset of cleavage furrow ingression (top image), and at the end of the second cell cycle (bottom image). Elapsed time from the beginning of the cleavage furrow ingression at the first mitosis is shown in the min:sec format. Below every panel, proportion of embryos that reach the four-cell stage and the number of embryos (n) analysed per condition are indicated. Scale bar, 10 μm. Note that unlike the GFP-SAS-5 control (**C**), embryos which express GFP-SAS-5 L141E (**D**) and GFP-SAS-5 I247E (**E**) mutants fail to rescue the *sas-5 (RNAi)* phenotype.

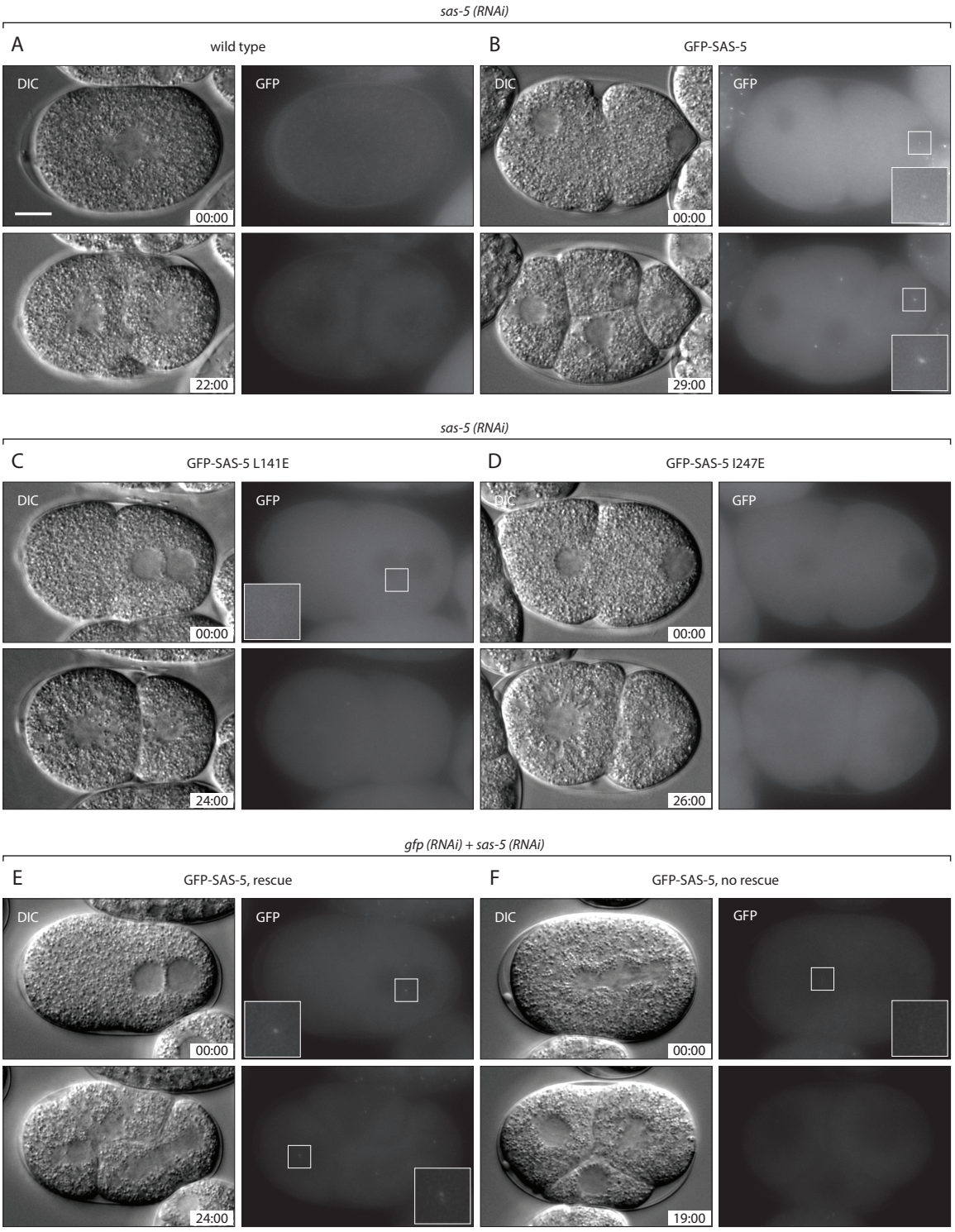


Figure 4.9: **Dual DIC and fluorescence time-lapse microscopy of oligomerisation mutant worm embryos.** Wild type (**A**), *gfp::sas-5* (**B**, **E**, **F**), *gfp::sas-5[L141E]* (**C**), and *gfp::sas-5[I247E]* (**D**) adult worms were treated with *sas-5* RNAi (**A**-**D**) or simultaneous *gfp* and *sas-5* RNAi (**E**, **F**). The resulting embryos were imaged by dual DIC and fluorescence time-lapse microscopy. Embryos were imaged at the the one-cell stage (top image) and ∼25 min later (bottom image). The exact elapsed time from taking the first image (top) is shown in the min:sec format. GFP images are maximum intensity projections of 8 z-sections taken at 0.5 μm intervals. Below every panel, proportion of embryos that reach the four-cell stage and the number of embryos (n) analysed per condition are indicated. 2.5× magnifications of boxed regions are shown in insets. Scale bar, 10 μm.

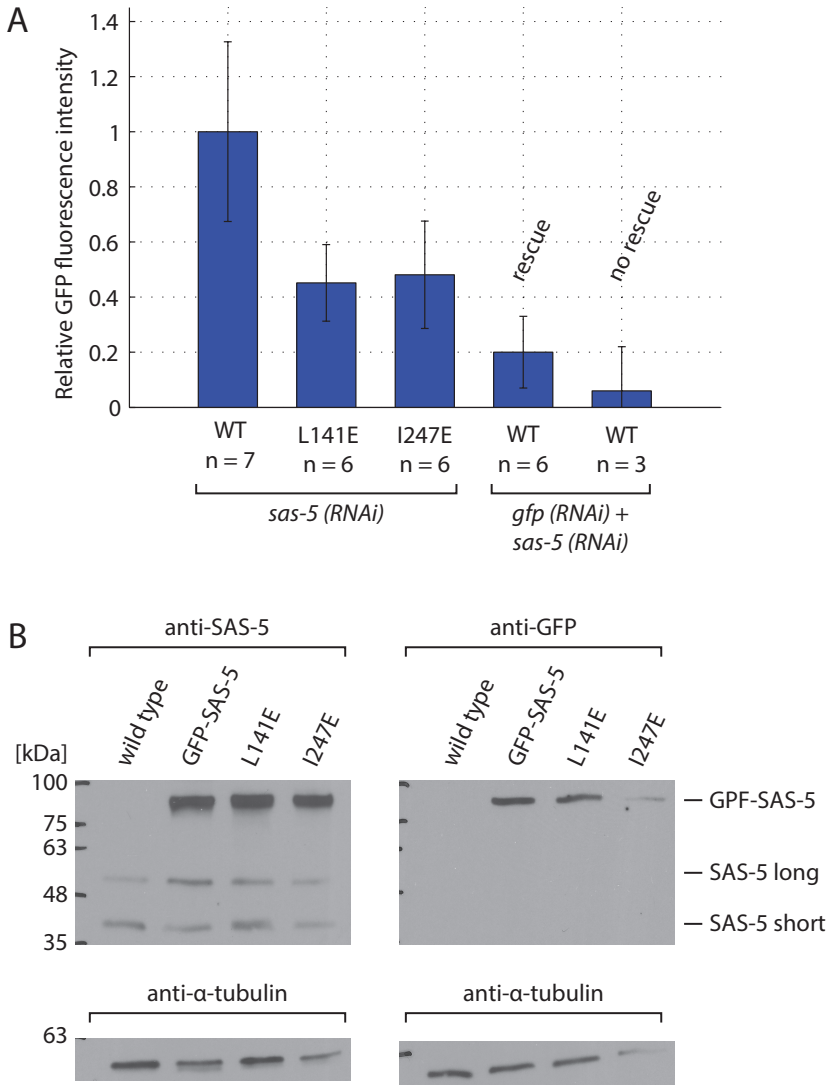


Figure 4.10: **Expression levels of SAS-5 in mutant worm embryos.** (**A**) Bar chart of averaged and wild-type-normalised fluorescence intensity of GFP in live one-cell stage embryos. Note that successful *sas-5 (RNAi)* phenotype rescue after dual RNAi treatment depended on the levels of exogenous GFP-SAS-5 expression (compare bars #1, #4 and #5). Also, *gfp* RNAi-reduced expression of GFP-SAS-5 to levels below those of either of the GFP-SAS-5 mutants (compare bars #2, #3 and #4), were able to rescue the *sas-5 (RNAi)* phenotype. This indicates that the lack of rescue in the two strains of GFP-SAS-5 mutants is not due to insufficient protein expression levels. To ensure that the *sas-5 (RNAi)* rescue was not due to dilution of the *sas-5* RNAi with *gfp* RNAi, all *gfp::sas-5[L141E]* embryos were scored under identical conditions, which confirmed that all those embryos exhibited the *sas-5 (RNAi)* phenotype (data not shown). Finally, GFP-SAS-5 expression at ∼5 % (also treated with *sas-5* RNAi; bar #5), were not enough to provide rescue, further indicating that *sas-5 (RNAi)* is functional in those assays. (**B**) All *C. elegans* strains used in this study were analysed by Western blotting (50 adult worms each) with antibodies against either SAS-5, GFP, or α-tubulin (loading control). Note that endogenous SAS-5 is expressed in two isoforms: full-length isoform 'a' (404 aa; Wormbase, F35B12.5a) and a shorter isoform 'c' (288 aa; Wormbase, F35B12.5c), which is likely the result of trans-splicing. Exogenous expression of *gfp::sas-5* produces the longer isoform only, possibly due to disturbed splice sites after recoding for RNAi resistance.

to the site of a procentriole formation. Second, once it has localised to the centriole, SAS-5 builds the cartwheel (together with SAS-6), thus seeding assembly of a new centriole.

Having established that higher-order oligomers of SAS-5 are important for centriole duplication, we wondered which of the two above-mentioned steps had been affected by the oligomerisation mutations. Using live imaging, we monitored distribution of GFP-fused SAS-5 in one-cell stage embryos treated with *sas-5* RNAi. While GFP-SAS-5 was observed in foci at centrosomes (Figure 4.9B on page 127), such foci were not seen for either of the oligomerisation mutants (Figure 4.9C, D on page 127). Thus, we wondered whether the absence of mutant foci was a real result or rather a false negative due to limited intensity of GFP fluorescence. In order to investigate the localisation of SAS-5 using a more sensitive approach, we analysed fixed one-cell stage embryos with immuno-fluorescence assays using antibodies which specifically recognise GFP or worm centrioles (IFA marker; Figure 4.11A-C on page 130). Similarly to the results obtained with live imaging, GFP-SAS-5 was shown to be enriched at centrioles (Figure 4.11A, D on page 130 and [88]). Intriguely, distributions of SAS-5 molecules observed for the two oligomerisation mutants were distinct from those of GFP-SAS-5. In the case of GFP-SAS-5 L141E, up to ∼30 % of centrosomes exhibited a clear GFP focus, with the remaining ones exhibiting either no focus, or only a weak/diffuse focus (Figure 4.11B, D on page 130). In the case of GFP-SAS-5 I247E, the vast majority of embryos exhibited no detectable GFP focus at centrioles (Figure 4.11C, D on page 130).

These results revealed that the self-oligomerisation is important in the very first step that SAS-5 takes on its path to centriole assembly - in localisation to the procentriole. It is evident that the *Implico* domain plays a particularly critical role in this respect. Moreover, the presence of GFP-SAS-5 L141E at centrioles in one-third of embryos is intriguing. Given that this mutant completely failed to rescue the *sas-5 (RNAi)* phenotype, it raises the possibility that formation of higher-order assemblies of SAS-5 is essential for function (step 2: seeding centriole growth), even if centriolar localisation has been achieved (step 1).

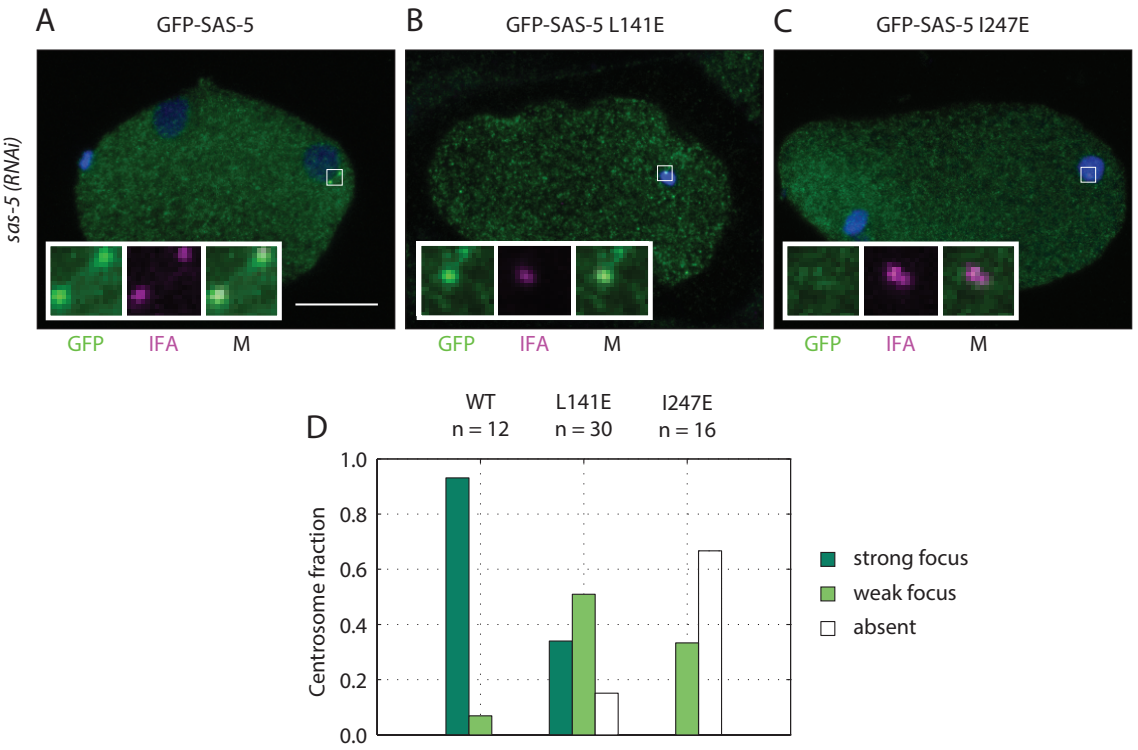


Figure 4.11: **Defective centriolar localisation of SAS-5 oligomerisation mutants.** (**A**-**C**) *sas-5 (RNAi)*-treated adult worm embryos carrying a gene-expression cassette for GFP-SAS-5 (**A**), GFP-SAS-5 L141E (**B**), GFP-SAS-5 I247E (**C**) were fixed and analysed by immuno-fluorescence. The embryos were stained with antibodies against GFP (green) and the centriolar marker IFA (magenta); DNA is visualized in blue. Scale bar, 10 μm. 5× magnifications of boxed regions are shown in insets. (**D**) Bar chart showing scored localisation of GFP to centrosomes (absent, weak focus or strong focus) in one-cell stage embryos, as shown in panels **A**-**C**. Note that while GFP-SAS-5 strongly localised to centrosomes in ∼95 % of cases, the GFP-SAS-5 L141E mutant did so only in ∼30 % of cases, and the SAS-5 I247E mutant showed no detectable GFP focus in most cases.

Chapter 5

Discussion

Centriole duplication is the fundamental cellular process which leads to formation of a new centriole (procentriole) next to a pre-existing centriole (mother or daughter). Further maturation allows centrioles to recruit PCM and form specialised cellular organelles called centrosomes. Cells carrying two centrosomes are able to organise a bipolar mitotic spindle and thus accurately segregate their chromosomes into two daughter cells in a timely fashion. At the onset of mitosis, cells carrying a number of centrosomes different than two face adverse consequences of forming either a monopolar spindle (one centrosome) or a multipolar spindle (more than two centrosomes). While the monopolar spindle leads to a complete failure of chromosome segregation, the multipolar spindle segregates chromosomes unequally, and thus leads to aneuploid progeny - a hallmark of cancer cells. Therefore, dividing cells impose a strict control on centriole numbers by tightly regulating centriole duplication to only one event per cell cycle. The process of centriole duplication, however, is integral not only to cell division. In fact, the ancestral role of centrioles was to act as basal bodies and to nucleate growth of cilia and flagella - the sensory and motility apparatus of cells. Therefore, errors in centriole assembly contribute to a variety of complex disorders called ciliopathies, which stem from defective or absent cilia.

Although the process of duplicating centrioles involves a large number of complex events, coordinated by hundreds of different proteins acting throughout many stages of cell cycle, it appears that only a small subset of proteins (the 'core five') is absolutely essential to drive centriole assembly [78, 88, 172, 179, 200, 201, 249, 263, 296]. One such protein, SAS-6, plays a critical role in establishment of the nine-fold symmetry - the

most recognisable feature of centrioles [237]. Through self-oligomerisation, SAS-6 can form a structure called cartwheel, which looks like a ring with nine equally-spaced spokes emanating from it [183, 331, 332]. The cartwheel is involved in the earliest stages of centriole assembly, where it takes the scaffolding role, providing support and templating the construction of the outer wall of microtubules. Formation of the cartwheel itself requires very high concentrations of SAS-6, which is normally seen only at centrioles [170, 183]. Given that SAS-6 is the seed for centriole growth, cells must therefore tightly control concentration and localisation of SAS-6 in order to avoid random centriole nucleation events in their cytoplasm. Such random centriole nucleations could potentially lead to multipolar spindle formation and cancer. We speculate that cells might employ a mechanism promoting localised up-concentration of SAS-6, thereby limiting centriole formation to the right place and the right time. We hypothesise that the factor involved in such mechanism is protein SAS-5, which belongs to the 'core five' essential components of the centriole duplication pathway [78,88,296]. In this chapter, I will discuss the results of my doctoral research, primary aim of which was to discover the structure of SAS-5, and to unravel its function in the mechanism of centriolar cartwheel formation.

5.1 Eliminating insoluble protein bottleneck

The primary requirement for biochemical and biophysical experiments *in vitro* is availability of soluble target protein. Depending on the complexity of the protein of interest, the task of obtaining well-behaved samples can take months or even years. Such task is particularly difficult if a new family of proteins is to be examined, and no previous knowledge is available to narrow down the search for optimal conditions, which would guarantee stability of the target protein in biochemical assays. Heterologous expression of eukaryotic target genes often yields insoluble protein, or soluble but aggregated protein. A multitude of factors can influence such scenario, with the major effect seen from the lack of appropriate folding machinery in the heterologous host (*e.g.* protein chaperones and post-translational modification enzymes), leading to protein misfolding and subsequent aggregation or inclusion body formation. Proteins also aggregate in the absence of their interaction partner, and can often be solubilised by co-expression with such partner proteins.

5.1.1 Full-length SAS-5 for biochemical work

Obtaining full-length SAS-5 for biochemical work has been a challenge for years. The attempts were made both by our laboratory and by others [203, 275]; however, the results were unsatisfactory: SAS-5 was insoluble or if marginally soluble, it was severely degraded and aggregated. The unavailability of soluble and well-behaved full-length SAS-5 limited the progress of understanding its role in centriole duplication. One approach of bypassing this bottleneck was to work with short peptides of SAS-5, which could be either synthesised or purified from bacterial hosts [142, 275]. Biophysical work performed by Hilbert *et al.* [142] and Qiao *et al.* [275] used such peptides to probe the ability of relatively short C-terminal fragments of SAS-5 to interact with SAS-6. Although informative about specific interaction interfaces, short peptides cannot provide information on the properties of full-length proteins.

Our approach to counter the insolubility problem involved design of a customised line of *E. coli* expression vectors which we called pFloat (Figure 2.4 on page 72). By fusing SAS-5_{FL} to a range of solubility tags, we have obtained large quantities of soluble protein. Once we had established the purification protocol for SAS-5_{FL}, we were able to subject the protein to a series of biophysical experiments, which revealed that SAS-5_{FL} formed aggregates *in vitro* (Figure 2.6 on page 81). Further experiments showed that such aggregation was due to a $\sim$20 amino acid segment predicted to form a β-strand - the *sticky strand* (Figure 2.7 on page 82). Position of the *sticky strand* is outside of the SAS-5 region implicated in SAS-6 binding. Although we cannot completely exclude the possibility that the *sticky strand* is a physiologically-relevant region through which SAS-5 multimerises, we speculate that this is not the case. First, even at low protein concentrations, the aggregates formed by SAS-5_{FL}, as observed by EM, are proportional in size to centrioles of *C. elegans* themselves (Figure 2.6E on page 81). Thus, given their large size, if SAS-5 aggregates were indeed a relevant *in vivo* feature of worms, the remaining space to accommodate other centriolar components, including SAS-6, SAS-4 and microtubules, would be significantly limited. It is likely that *C. elegans* SAS-5 has co-evolved with other cellular machinery, and therefore worms are able to control the state of SAS-5 in their cytoplasm, as such SAS-5 aggregates have not been observed in worm embryos. Interestingly, however, prolonged over-expression of GFP-fused SAS-5 in *C. elegans*

leads to formation of scattered GFP foci within the cytoplasm of worm cells [88]. It is intriguing to speculate whether *C. elegans* employ protection mechanisms which could prevent SAS-5 from aggregation. At basal expression levels, such potential internal mechanisms could steadily repress aggregation of SAS-5; however, once the expression levels are elevated (due to constitutive expression via a strong promoter in the GFP-SAS-5 cassette) and SAS-5 slowly accumulates in the cytoplasm, the protection mechanism may no longer cope with the overabundance of SAS-5, and as a result some SAS-5 molecules will start aggregating via the *sticky strand*. It remains to be seen whether such mechanisms exist, and if these GFP foci are indeed the *sticky strand*-triggered aggregates of SAS-5.

Given that heterologous expression of SAS-5 produces soluble aggregates (Figure 2.6 on page 81), and that potentially even over-expression of SAS-5 in *C. elegans* may lead to such aggregation [88], we postulate that *in vivo* SAS-5 forms a complex with a yet unidentified partner using the *sticky strand* described in this study. Such complex may be a part of the protection mechanism aimed at retaining non-aggregated SAS-5 in the cell. It would be interesting to speculate whether protein phosphatase 2 (PP2A), which had previously been shown to interact with SAS-5 *in vivo* and to control SAS-5 levels [181, 310], may potentially bind to the *sticky strand* of SAS-5. Alternatively, a protection mechanism might employ phosphorylation, a durable covalent modification, that if applied to the *sticky strand*, it will shift the chemical charge of this region to negative, and thus effectively render SAS-5 less likely to self-aggregate. Since protein phosphorylation is a common post-translational modification which regulates the spatio-temporal interactions within the centriole duplication pathway, such mechanism is not unlikely. In fact, SAS-5 was reported as a target of a kinase [181, 203]; however, neither the identity of the kinase nor the exact phosphorylation site have been determined yet.

Taken together, we believe that the *sticky strand* has a physiological role, either as a binding platform for another protein - such as PP2A, or as a target site for phosphorylation, but not as an oligomerisation interface. Hence, while deletion of the *sticky strand* could potentially disturb protein function *in vivo*, our *in vitro* experiments aimed at oligomerisation and structure of SAS-5 are unlikely to be affected. More work is required to shed light on the exact physiological role of the *sticky strand*

and the mechanisms that control it.

Excising or replacing the *sticky strand* with a flexible linker allowed us to produce a non-aggregated full-length SAS-5_{FLEX} (Figure 2.8 on page 83), which can now be used in biochemical and biophysical experiments to advance our understanding of the role of SAS-5 in centriole assembly.

This feat, however, would not have been possible without the use of the pFloat expression vector system (Figure 2.4 on page 72), which allowed for streamlined testing of a range of solubilising protein fusion tags. This eventually led to identification of the *sticky strand* through incremental biochemical work (Figure 2.7 on page 82). pFloat system has been since successfully used in a variety of projects in our lab and beyond.

5.1.2 Challenges of working with modular proteins

Even though elimination of the *sticky strand* yielded soluble full-length SAS-5, purification and working with this protein construct has still remained a challenge. The challenge stems from the molecular architecture of SAS-5 which is composed of multiple modules that effectively constitute a versatile molecular machine within one polypeptide chain (Figure 3.4A on page 91). Modular architecture is a common feature among specialised eukaryotic proteins, which attained multiple domains through domain-shuffling events between various genomic locations [259]. Such domains are usually folding independently of one another, and are connected by flexible linkers, allowing for unconstrained spatial movement of individual domains. Proteins with disordered linkers, however, have a significantly shorter half-life *in vivo* than proteins that do not [333]. Both terminal and internal disordered linkers can be cleaved by the proteasome, and are thought to act as the initiation sites for degradation, allowing for efficient fractionation of multidomain proteins to smaller pieces and individual modules [271]. Interestingly, changes in number and length of disordered regions is linked with altered half-life, which may indicate that disordered regions can contribute to evolutionary tuning of protein half-life in cells [333]. Modularity gives proteins an edge in performance and processivity due to inclusion of many functional modules at the precise location of protein involvement, thus mediating highly specific interactions

with a number of binding partners simultaneously [19, 137]. Degradation limits such versatility by disconnecting these modules at a given rate.

The degradation rate can also be thought of as the factor which determines the life span of modular proteins within a cell. This effect is particularly prominent during heterologous expression of modular proteins, because cell lysis disturbs the carefully controlled environment of the cell, and eliminates any regulation from endogenous proteases and phosphatases. This leads to inevitable series of degradation events targeting parts of the over-expressed proteins, which lack defined tertiary structure, such as disordered and flexible linkers between modules. Effectively, a ladder made of non-degraded, partially degraded and completely degraded target proteins is created (Figure 2.7A on page 82). In some cases chromatography techniques can be used to separate such degradation products; however, if some modules of the target protein self-oligomerise, separating these truncation products becomes quite arduous, if not completely unfeasible.

C. elegans SAS-5 is one such modular protein. It is composed of the two independently folded and self-oligomerising domains (Figure 3.4 on page 91), which are flanked by long stretches of disordered N- and C-terminal segments (Figure 3.2 on page 89), able to interact with two other essential centriole duplication proteins, SAS-4 [71, 140] and SAS-6 [142, 275]. Such architectural composition renders purification of SAS-5_{FLEX} an extremely challenging task. In our pursuit to obtain pure full-length protein, we employed a number of extreme measures. These included addition of protease inhibitors to all buffer solutions used for purification, as well as swift processing of the protein at low temperatures. In our experience, keeping both protein and buffer solutions chilled, as well as working uninterrupted from the moment of lysing the cells to freezing the pure protein for storage, were key to preparation of non-degraded SAS-5_{FLEX} for biochemical and biophysical assays. However, such protocol is extremely gruelling, and perhaps an alternative method should be devised to allow for production of modular proteins at large quantities. One way of handling this problem might be to induce insoluble expression of SAS-5_{FLEX}, and thus protect the protein from damaging effects of proteases, which could then be removed before refolding of SAS-5_{FLEX}. However, the process of protein refolding might be equally challenging because SAS-5 could require protein chaperones to avoid misfolding. As

an alternative to host-cell based expression systems, it might be interesting to attempt cell-free synthesis of SAS-5_{FLEX}. Such cell-free expression systems, however, are normally prepared from cell extracts, and therefore are abundant in non-specific proteases [50]. To avoid proteolytic degradation altogether, it might be necessary to use a cell-free system in which all the components of transcription, translation as well as protein folding machinery are recombinantly purified and reconstituted [302].

5.2 *Implico* is a novel dimeric fold

In our quest to understanding the role of SAS-5 in the mechanism of centriole assembly, we analysed the molecular architecture of SAS-5 with structural and biophysical methods. Our results revealed that SAS-5 can form strong dimers through its *Implico* domain (Figure 3.15 on page 109). Crystal structure of the *Implico* domain revealed that it constitutes a novel dimerisation interface (Figure 3.16 on page 110), which has no known homologues outside of the SAS-5 family of proteins.

This finding is particularly interesting in the light of advancement of structural biology, where the percentage of novel fold discoveries relative to total protein structures being solved is now <1 % [73]. With the establishment of multiple centres for structural genomics (SG) at the turn of the 21^{st} century, the number of protein structures solved by experimental methods has increased almost exponentially. Broad range of these initiatives allowed to expand the structural coverage of the protein universe, leading to better understanding of the biochemical repertoire encoded within the genomes. Although every new experimental protein structure contributes to the global repository of structural data, one of the most exciting and rewarding discoveries a structural biologist can make is to find a novel structure. The novelty arising from such structures is invaluable in broadening our understanding of the protein fold space, and may be used to infer the structure and function of other family members (*protein family* constitutes a set of structurally similar protein domains, which are also likely to perform similar biochemical functions).

Over the last decade, with technical advancements in whole-genome sequencing, the number of annotated protein sequences has considerably increased. Solving structures of all of these proteins would not be feasible as such experiments are inherently slow due to their unpredictable nature. The cost of protein structure determination

is also rather expensive, *e.g.* at Protein Structure Initiative centres, the average cost per structure in 2004 was $138,000, with a structure of the first Pfam [107] family member costing $1.0 million, and $2.2 million for the first SCOP [234] superfamily/fold protein [54]. Experimentally-determined structures of the first members of a protein family serve as templates for calculation of tens of thousands of comparative models for related proteins, and contribute to the database of unique protein structures [28, 266] which is used to improve protein structure prediction methods [92], as well as protein design and engineering [106, 349].

Since their establishment, SG centres have become the main driving force behind discovery of new protein families; however most recently, the pace of such discoveries has rapidly slowed down [73, 173]. Because the conservation of the protein structure is three to ten times higher than the protein sequence [155], the total of over 110,000 entries in the Protein Data Bank (as of 2015) represents only 148 protein folds in the SCOP2 database [9]. *Implico* domain due to its rare primary sequence defines a new family of small dimeric proteins. Also, due to its inherent topology, where two polypeptide chains intertwine in a unique way to form an incredibly stable globular domain, *Implico*, to our best knowledge, constitutes a novel protein fold, as searches for structural similarity in the Protein Data Bank with DALI server [150] were not successful.

A relatively small size (∼12 kDa) and the unique anti-parallel dimeric fold of *Implico* has a great potential application in protein design and engineering. One such application could be engineering of chimeric antibodies against HIV-1 (human immunodeficiency virus type 1). Unlike most bacterial and viral pathogens, which display a dense array of antigens at their surface, a single HIV-1 virion presents only ∼14 envelope spikes [357]. Such sparse spike display limits the ability of antibodies to achieve bivalent binding through interaction with two adjacent spikes, as the distance between HIV-1 spikes is larger than the potential span of two Fab arms [342]. By fusing C_H3 domains of antibodies to *Implico* domain, it would be possible to create a tetrameric (single N- or C-terminal fusion) or an octameric structure (double N- and C-terminal fusion), with *Implico* essentially serving as both bridging and oligomerising factor. Such chimera will not only increase the span between Fabs but also increase binding strength towards antigens by providing multivalent attachment

sites. Similar fusions can be used for linking other proteins. Given that *Implico* dimer is of anti-parallel topology, a particularly useful application would be to fuse it with short dimeric anti-parallel coiled coils, which otherwise are not stable enough for crystallisation. *Implico* fusion could potentially stabilise their dimeric state, and also provide extra surface for crystal contacts, which in case of coiled-coil crystallisation is often relatively limited.

5.3 Large oligomers of SAS-5 are important for centriole formation

In this study, we show that *Implico* domain is inherently folded into a strong dimeric structure, which cannot be disrupted even in the instance of protein dilution to concentrations as low as 20 nM (Figure 4.2 on page 116). By comparison, the dimer mediated by the coiled coil of both *C. elegans* and *C. reinhardtii* SAS-6 is approximately 50-times weaker ($\sim$1 μM) [183], and yet its human homologue readily forms dimers in cell cytoplasm [170]. Since detection of a 12 kDa protein (size of the *Implico* domain) at concentrations below 20 nM by SEC-MALS is quite challenging, it may not be feasible to determine the exact dissociation constant for *Implico* with this technique. However, it is not unreasonable to assume that the affinity of the *Implico* self-interaction is much stronger than 20 nM, and therefore similarly to *H. sapiens* SAS-6 coiled-coil dimers, SAS-5 will also form dimers (*Implico*-mediated) in the cytoplasm of *C. elegans* embryos. Given the moderate strength of SAS-5/SAS-6 and SAS-5/SAS-4 interactions, we envision that *Implico*-mediated SAS-5 dimers would stochastically associate with SAS-6 [88, 142, 203, 275] and SAS-4 [71, 140] in worm cytoplasm.

5.3.1 Potential benefit from the avidity effect

Our work indicates that formation of these strong *Implico*-mediated SAS-5 dimers are crucial for SAS-5 localisation to centrioles (Figure 4.11C, D on page 130). This observation suggests that the mechanism responsible for centriolar localisation of SAS-5 either recognizes surface of the globular *Implico* domain itself, or alternatively,

depends on a protein association that requires two copies of a distinct SAS-5 epitope, and hence the SAS-5 dimer.

While affinity refers to the binding strength of a single interaction between a molecule and its ligand, avidity describes the synergy effect, where binding strength is enhanced due to presence of multiple binding sites for the ligand [233]. We speculate that the interaction of SAS-5 with SAS-6, known to be necessary for localisation of both proteins to centrioles [88, 203, 275], benefits from such an avidity effect. The avidity effect depends on the strength of affinity and the binding valence (number of sites). The valence of SAS-5 upon dimerisation increases to two, and further to four and six at higher protein concentrations. Thus, once the highly valent oligomer of SAS-5 interacts with two or more SAS-6 dimer molecules, it will only completely dissociate after all the SAS-6 binding sites have also dissociated. The rate of dissociation of the whole SAS-5 oligomer from SAS-6 is therefore much slower than the rate of individual dissociation events between individual binding sites. Effectively, total binding strength (*i.e.* avidity) between SAS-5 and SAS-6 is increased.

It will be interesting to determine whether L141E and I247E mutations of SAS-5, which disrupt its oligomerisation properties and hence reduce its valency for SAS-6, prevent recruitment of SAS-6 to the procentriole, or rather interfere with assembly of SAS-6 into a cartwheel once localisation has taken place. In either case, failure to localise SAS-5 (and SAS-6) to centrioles prevents downstream recruitment of SAS-4, and therefore centriole formation does not proceed [78, 88, 203, 275].

5.3.2 Oligomeric state of SAS-5

A well-defined foci of GFP-SAS-5 at centrioles suggests that following centriolar localisation, the concentration of SAS-5 at centrioles is higher than in the cytoplasm (Figure 4.11A, D on page 130). We propose that once the cytoplasmic *Implico*-mediated SAS-5 dimers have localised to centrioles and their concentration increased, SAS-5 dimers then associate further through the coiled-coil domain to form higher-order assemblies. Our SEC-MALS results indicate that SAS-5 is able to form tetramers already at 1 μM concentration, while a mixture of tetramers and hexamers is already available at 10 μM (Figure 4.2B on page 116). Two independent oligomerisation interfaces have a potential to create protein oligomers much larger than tetramers or

hexamers. This includes open-ended protein chains or networks [20]. However, our native MS analysis of SAS-$5_{(2\text{-}265)}$ construct at 20 μM did not reveal species larger than hexamers (Figure 4.3 on page 117). Therefore, we believe that it is likely that higher-order assemblies of SAS-5 preferentially cap at a hexameric size and only a small proportion of them forms larger open-ended species. Further experimentation at higher concentrations of SAS-5 is needed to explore this possibility. Given that the hexamer of SAS-5 would form through sequential association of three SAS-5 dimers (Figure 4.1 on page 115), effectively saturating all possible self-binding sites, we speculate that such hexameric arrangement should be both kinetically and entropically favoured over open-ended SAS-5 chains. While dissociation of the compact SAS-5 hexamer would require simultaneous disruption of two SAS-5 coiled coils, an open-ended SAS-5 oligomer will already dissociate after a single coiled-coil disruption. Moreover, in the sequential assembly model, once a monomeric coiled-coil chain of a SAS-5 dimer associates with another monomeric coiled-coil chain of another SAS-5 dimer, folding into the first coiled-coil structure, it effectively forms a tetramer, although the second coiled coil has not yet been assembled (Figure 5.1 on page 142). At this point, the number of degrees of freedom for the two remaining, non-associated, monomeric coiled-coil chains is reduced, and thus their association into the second folded coiled-coil structure would become entropically favoured.

5.3.3 Hypothetical mechanism of centriole formation in *C. elegans*

Localisation of SAS-5 to centrioles is weakened by disruption of SAS-5 coiled-coil assemblies with mutations targeting the hydrophobic core of the coiled coil (L141E; Figure 4.11B, D on page 130). Such weakened localisation may be due to mutations impacting the avidity mechanism towards SAS-6, that we postulated earlier. Interestingly, our data indicates that formation of higher-order SAS-5 assemblies is necessary for function beyond merely localisation. It appears that even though the GFP-SAS-5 L141E mutant does localise to procentrioles in ∼30 % of embryos, it is unable to rescue centriole duplication upon depletion of endogenous SAS-5 with RNAi (Figure 4.8D on page 126). It is certain that some amount of wild-type SAS-5, present

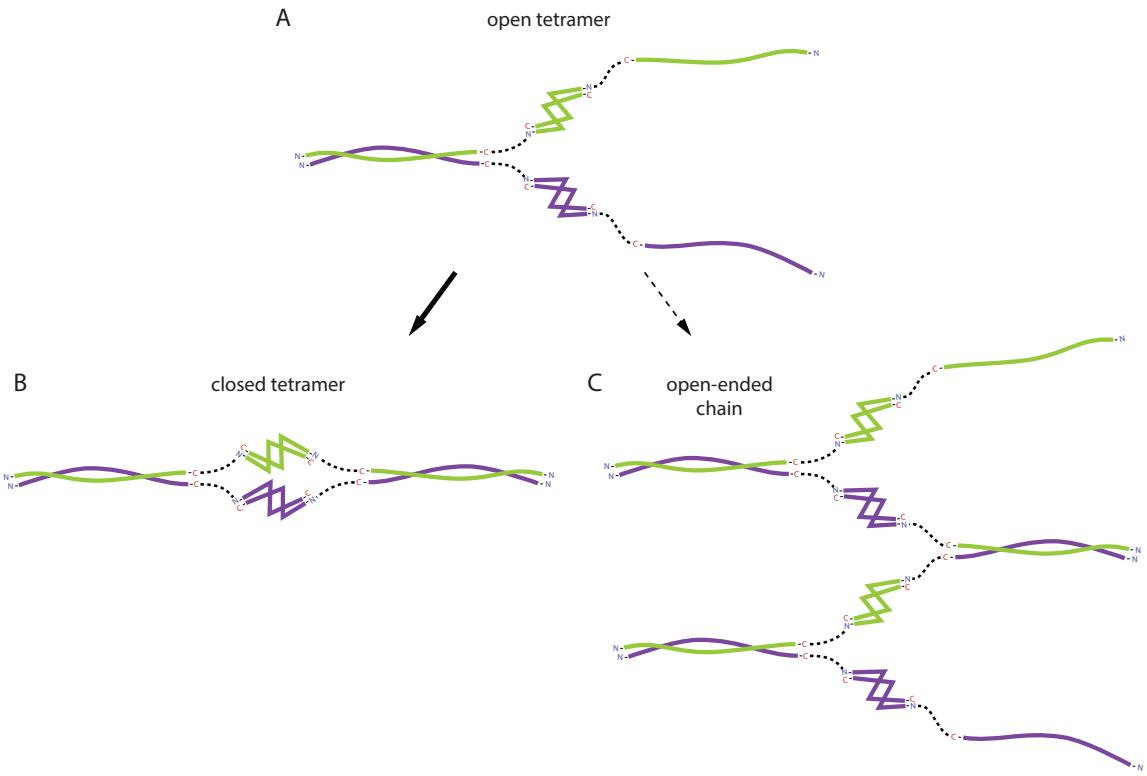


Figure 5.1: **Sequential assembly of SAS-5 oligomers.** With increasing concentration of SAS-5, the coiled-coil domain of individual *Implico*-mediated dimers will come together to form an open tetramer (**A**). We argue that it would be both kinetically and entropically favoured if the tetramer closed into a compact structure through dimerisation of the remaining coiled-coil domains (**B**) rather than by associating with yet another open tetramer (or dimer) to form open-ended chains of SAS-5 (**C**).

in sperm centrioles, is passed on to the early *C. elegans* embryo of the GFP-SAS-5 L141E mutant. However, because SAS-5 is an incredibly dynamic protein which shuttles between the cytoplasm and centrioles [88], we expect that any paternal wild-type SAS-5, that was contributed by the sperm centrioles, will rapidly exchange with the much larger maternal protein pool. Thus, we expect that GFP-SAS-5 L141E at centrioles, as observed by our fluorescence imaging studies (Figure 4.11B, D on page 130), is indeed the maternal mutant protein which localised to centrioles.

It is interesting to speculate whether SAS-5 is simply a transporter of SAS-6 (and potentially also of SAS-4) to growing procentrioles, or that perhaps SAS-5 also acts in an additional, more structural capacity, as has been suggested elsewhere [70,203,275]. Our studies provide evidence that higher-order assemblies of SAS-5 are indeed crucial for localisation of SAS-5 to centrioles, although with distinct outcomes for disruption of each oligomerisation domain. Given that the coiled-coil mutant of SAS-5 (GFP-SAS-5 L141E) was able to localise to centrioles in ~30 % of embryos, but failed

at centriole duplication, we believe that the role of large oligomeric assemblies of SAS-5 is not exclusive to centriolar localisation. This suggestion is consistent with the requirement for the mutual centriole localisation of the SAS-5/SAS-6 complex to initiate formation of the *C. elegans* central tube [200]. However, because FRAP experiments revealed rapid exchange of SAS-5 molecules between centrioles and the cytoplasm [88], suggestion that SAS-5 performs a structural role in centriole assembly is rather controversial. We wondered whether these two observations could be reconciled.

It appears that prior to successful assembly of *C. elegans* central tube, two sequential (but to some extent independent) events must take place. First, after translation in the cytoplasm, SAS-5 and SAS-6 localise to the site of a nascent centriole. Second, this centriole-localised pool of SAS-5 and SAS-6 acts together to build a new centriole. We speculate that large assemblies of SAS-5 could play a role in both of these events, effectively facilitating centriole assembly by working together with SAS-6, although in two different, non-exclusive modes. Firstly, large assemblies of SAS-5 could expedite co-localisation of the SAS-5/SAS-6 complex to centrioles by providing multivalent attachment sites for SAS-6. Secondly, SAS-5 could act as a nucleating agent in the process of cartwheel assembly (see below). Finally, such large oligomers of SAS-5 may be transiently located at the site of the procentriole, or alternatively, could correspond to only a small fraction of total SAS-5 in the cell, and thus escape detection in FRAP experiments [88].

Once the centriolar localisation of SAS-5 has been achieved, and both SAS-5 and SAS-6 are present at centrioles, a rapid increase in local concentration of SAS-6 is required for it to begin assembling a cartwheel spiral (or a ring in species other than *C. elegans*). We speculate that large oligomeric assemblies of SAS-5 could bind multiple SAS-6 dimers and bring them together in a confined space at the site of the procentriole formation (Figure 5.2 on page 145). ZYG-1, which also localises to centriole duplication site and also binds SAS-6 (although independently of SAS-5 [203]) may provide further enhancement to local concentration of SAS-6. Given the weak affinity (K_d $\sim$110 μM [183]) of the N-N interaction between globular head-domains of SAS-6 (Note: N-N interaction facilitates further oligomerisation of SAS-6 towards rings and spirals), such increase in local concentration of SAS-6 may provide

a nucleation point for SAS-6 oligomerisation.

Moreover, if SAS-5 forms defined hexamers as argued above, such hexamers have capacity to recruit up to six SAS-6 dimers [142, 275] per hexamer. Thus, one SAS-5 hexamer could potentially provide enough SAS-6 material to build just over one turn of the proposed SAS-6 spiral [142], but only ⅔ of a flat nine-fold symmetrical ring [183, 331, 332]. It is possible that through its hexameric architecture, SAS-5 provides the initial structural guidance for SAS-6 to oligomerise into a certain configuration (spiral vs. ring). Such guidance would otherwise be lost in SAS-5 mutants that disrupt oligomeric architecture of SAS-5 or the ability of SAS-5 to interact with SAS-6 [88, 203, 275], therefore explaining the failure of such mutants in assembling centrioles.

5.3.4 Same but different: mechanism of centriole assembly in other species

Over the last couple of years a number of studies have investigated orthologues of SAS-5 and SAS-6 in other organisms, offering insights into centriole assembly mechanisms outside *C. elegans*. At the beginning of the centriole duplication cycle in G1 phase, PLK4 encircles the pre-existing centrioles (mother and daughter), and upon transition to S phase, the localisation of PLK4 dramatically changes into a defined dot-like structure, which then later becomes the site of procentriole growth [177]. Ohta *et al.* [252] and Kratz *et al.* [191] revealed that PLK4 physically interacts with STIL (vertebrate equivalent of *C. elegans* SAS-5), and offered suggestion that such binding may protect PLK4 from degradation. They also proposed that protection of PLK4 by STIL binding may be critical in achieving the dot-like localisation of PLK4 [252], and hence formation of only one procentriole at a specific location next to a pre-existing centriole. However, it is not clear how transition from a dispersed ring to a focused dot is achieved mechanistically after STIL and PLK4 binding, *i.e.* what is the factor which differentiates protection from degradation at one specific site, but not anywhere else. This important question remains to be answered, as our understanding of the canonical centriole duplication process, where a pre-existing centriole assembles only a single procentriole, directly depends on the circle-to-dot localisation shift.

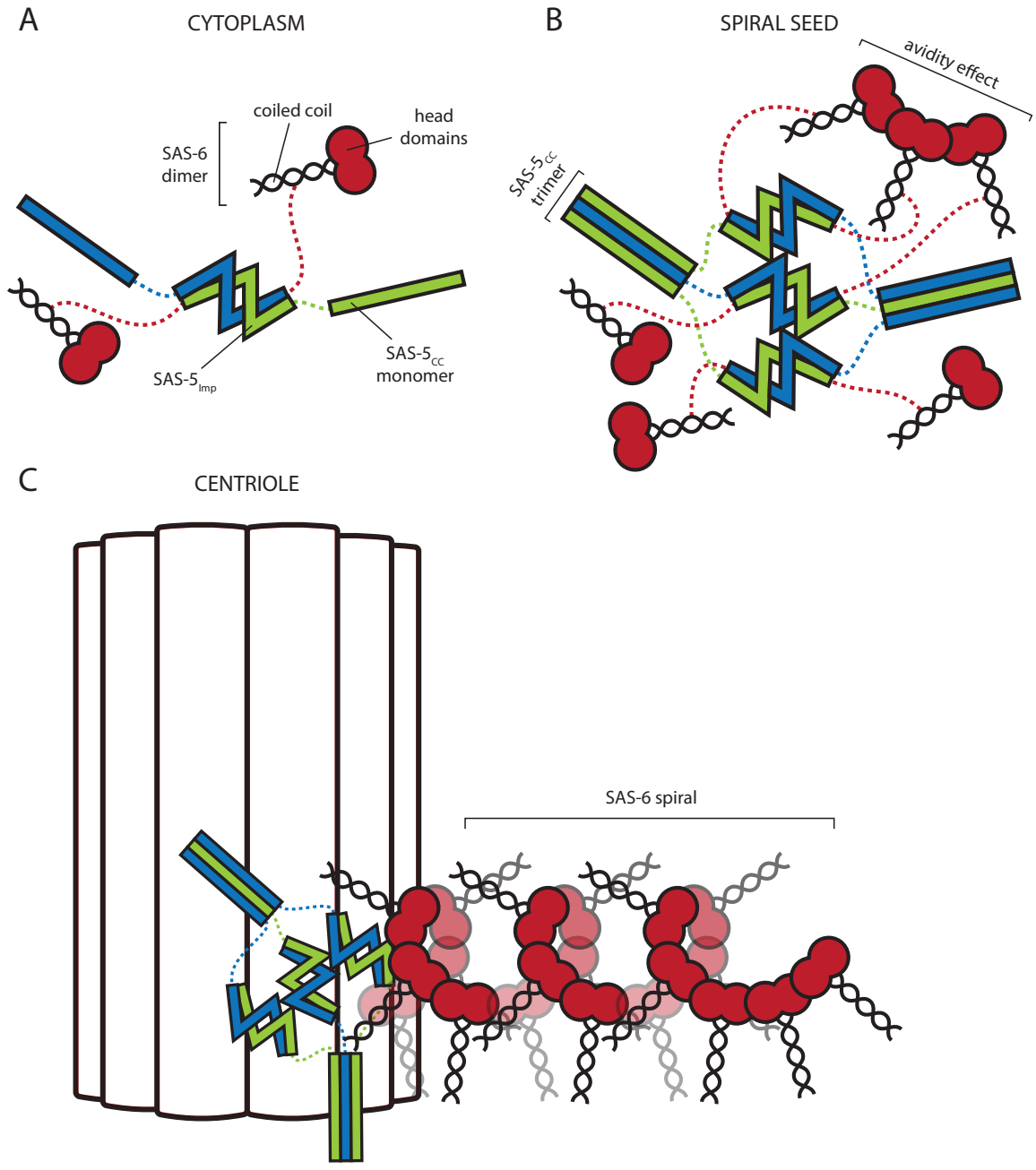


Figure 5.2: **Hypothetical model of cartwheel nucleation.** (**A**) *Implico*-mediated dimers of SAS-5 may associate with SAS-6 dimers in cell cytoplasm, depending on the relative levels of each protein. (**B**) After SAS-5 is localised to the centriole, its own local concentration increases and the protein forms higher-order assemblies. Such assemblies recruit multiple SAS-6 dimers in a confined space, and thus also increase the local concentration of SAS-6 through avidity effects. (**C**) Once the critical concentration of SAS-6 is reached, nucleation of the centriolar cartwheel/central tube begins.

Ohta *et al.* [252] showed that PLK4 phosphorylates STIL at numerous sites (including STAN motif, which harbours the epitope for STIL/SAS-6 binding), and that such phosphorylation acts as a trigger for the interaction between STIL and SAS-6. It appears that the change of the chemical environment around the STAN motif of STIL to more acidic increases the affinity of SAS-6 binding, which otherwise might go undetected in experiments with an unmodified version of STIL [14,323,337]. Interestingly, Ohta *et al.* [252] show that binding between STIL and SAS-6 facilitates loading of SAS-6 onto centrioles, and thus allows for further centriole assembly. In an independent study by Dzhindzhev *et al.* [98], it was revealed that similarly to vertebrate STIL, Ana2 (*Drosophila* equivalent of SAS-5) also interacts with PLK4 and that this interaction leads to phosphorylation of Ana2, and to later complex formation between Ana2 and SAS-6.

These two models of centriolar protein loading (vertebrate and *Drosophila*) are in contrast to the current *C. elegans* model, where no phosphorylation is required for the complex between SAS-5 and SAS-6 to form [142, 275]. Also, in *C. elegans*, SAS-5 and SAS-6 appear to be recruited simultaneously to centrioles as a complex [87, 262], rather than in sequence. While in vertebrates and flies, PLK4 recruits STIL through direct interaction, *C. elegans* ZYG-1 (nematode equivalent of PLK4) appears to physically interact with SAS-6 instead [203]. If the elementary role of ZYG-1/PLK4 binding to either SAS-5/STIL/Ana2 or SAS-6 is to solely recruit these proteins to the mother centriole, then perhaps these two seemingly divergent modes of PLK-4 binding have a common ground. Both the sequential recruitment of STIL and then SAS-6 (through phosphorylation of STIL) by PLK4 in vertebrates/flies or the simultaneous recruitment of SAS-5/SAS-6 complex by ZYG-1 will effectively arrive at precisely the same outcome, where all the proteins are present at procentrioles.

5.4 Recent structural studies of SAS-5 orthologues

In a number of recent studies in systems other than *C. elegans*, a line of evidence has emerged which strongly agrees with our data that self-oligomerisation of SAS-5 is important for its localisation to centrioles. While Ohta *et al.* [252] and Kratz *et al.* [191] were the first to report the physical interaction between STIL and PLK4, they were unable to accurately narrow down the binding site for STIL on PLK4 using

co-immunoprecipitation (Co-IP) assays (see below). The reverse of this experiment, however, *i.e.* locating the binding site for PLK4 on STIL, proved to be rather precise for Ohta *et al.* [252] who narrowed it down to the predicted coiled-coil region of STIL. The variable accuracy of Co-IP experiments reflects the limitations of such approaches for rigorous protein work. In an independent study, Moyer *et al.* [231] using GST pull-downs of partially purified proteins from insect cells, strengthened the above-mentioned data by showing that indeed PLK4 binds to the central region of STIL that includes the predicted coiled-coil domain.

5.4.1 Interaction between STIL and PLK4

Eventually, Arquint *et al.* [12] presented a crystal structure of the complex between human PLK4 and STIL, confirming that STIL indeed binds to PLK4 through its coiled-coil domain (residues 720-751), with nanomolar affinity (∼280 nM) and 1:1 stoichiometry. Interestingly, PLK4 appears to interact with the coiled coil of STIL via two distinct regions: (1) the C-terminal most polo-box domain 3 (PB3, residues 880-970) (Note: PLK4 has three distinct polo-boxes, the role of which is to regulate localisation, substrate binding and kinase activity), and (2) the N-terminal linker region (L1, residues 265-570) (Co-IP experiments), that lies between the kinase domain and the polo-boxes.

Work by Arquint *et al.* [12] has clarified the earlier conflicting evidence about the STIL-binding region on PLK4 [191, 252]. The crystal structure revealed the PB3 domain of PLK4 adopting the canonical polo-box fold, *i.e.* a six-stranded anti-parallel β-sheet and an α-helix, which interacts with the predicted coiled-coil domain of STIL. The interaction between STIL α-helix and PB3 α-helix is predominantly driven by hydrophobic contacts, further supported by backbone-backbone hydrogen bonds.

Intriguingly, the predicted coiled-coil domain of STIL in the context of STIL/ PLK4(PB3) binding is monomeric, and does not form a coiled coil with another STIL molecule in the crystal (or any other type of a higher-order oligomer for that matter). Arquint *et al.* [12] suggest, however, that the interaction between the α-helix of STIL and the α-helix of PLK4(PB3) mimics a coiled-coiled interaction, which may be a slight over-interpretation. Quick inspection of the crystal structure reveals that although it is true that STIL α-helix faces PLK4(PB3) α-helix with the side which

is predicted to form crucial hydrophobic contacts in a coiled coil (through positions a and d in the heptad repeat), there is only one instance of a typical knob-into-holes interaction between these two α-helices (as detected by SOCKET [339]). Also, there is only a single ionic interaction (between Q737 of STIL and S956 of PLK4) which stabilises the binding interface and resembles a typical interaction between positions e and g of the heptad repeat in a canonical coiled coil. These two instances of coiled-coil-like features may not be enough to justify the overall coiled-coil character of this protein-protein interaction.

5.4.2 Dynein light chain and short intrinsically disordered segments of Ana2

The central region of Ana2 (*Drosophila* orthologue of STIL), which also includes the predicted coiled-coil domain, has been shown to form higher-order assemblies by Slevin *et al.* [305]. Moreover, the authors revealed exciting new details of the interaction between Ana2 and dynein light chain (LC8), which appears to be important in proper spindle positioning during *Drosophila* larval brain development [305,340]. LC8 is a promiscuous protein which binds a wide range of substrates by recognising short linear motifs that are normally intrinsically disordered, but fold upon binding to LC8. The main role of homodimeric LC8 is to promote dimerisation and structural stabilisation of substrates, and to also link these substrates to transport via microtubule or actin routes [277]. Slevin *et al.* [305] solved two crystal structures of LC8 bound to two distinct short regions of Ana2: residues 159-168 (K_d $\sim$1 μM) and 237-246 (K_d $\sim$13 μM), which flank the central coiled-coil domain. Individually, these two short fragments of Ana2 fold into β-strands upon contact with LC8 (2:2 stoichiometry), and constitute an extension to the homo-dimerisation interface of LC8, which itself is a ten-stranded (two times five) anti-parallel β-sandwich flanked by two α-helices.

Although initially the authors were unable to produce a recombinant central region of Ana2 (residues 156-251), they eventually succeeded after co-expressing it with LC8. Through SEC-MALS experiments, they revealed that $Ana2_{(156\text{-}251)}$ forms a complex with LC8 in solution; however, the exact stoichiometry of binding was unclear. In the effort to determine relative contributions of each protein species to the

Ana2$_{(156\text{-}251)}$/LC8 complex, they fused LC8 with a SNAP-tag, effectively differentiating the mass of the SNAP-LC8 fusion (30 kDa) from the mass of Ana2$_{(156\text{-}251)}$ (11 kDa). Interestingly, the complex of Ana2$_{(156\text{-}251)}$ and SNAP-LC8 eluted as a broad peak, with the highest mass corresponding to species composed of four Ana2$_{(156\text{-}251)}$ and eight LC8 molecules. Because each Ana2$_{(156\text{-}251)}$ molecule has two LC8 binding sites, this result suggested a 4:8 stoichiometry. The authors concluded that LC8 fulfils its primary role by stabilising the tetramer conformation of Ana2$_{(156\text{-}251)}$. They also suggested that the SNAP-LC8 variant is not as effective in providing such stabilisation - possibly because of the fused SNAP-tag which may be causing steric hindrance to Ana2 tetramerisation. We would like to add a few important points to this interpretation, as the above-mentioned experimental evidence leaves some space for a debate.

While it is possible that LC8 enhances and stabilises tetramerisation of Ana2, it is difficult to unambigously confirm this in the absence of apo-Ana2 data - indeed, the authors found it challenging to obtain apo-Ana2$_{(156\text{-}251)}$. Interestingly, the SEC-MALS experiment with SNAP-LC8 may have provided a glimpse into the solution behaviour of Ana2, where, depending on the concentration, a broad range of oligomerisation species are present (similarly to SAS-5 from our study). Such a broad elution peak is a hallmark of a weak oligomer at a transition concentration. Because the authors did not specify whether the injection concentrations of Ana2/LC8 complexes (Ana2$_{(156\text{-}251)}$/LC8 and Ana2$_{(156\text{-}251)}$/SNAP-LC8) were identical, it remains unclear if the SNAP-tag fusion could actually influence stability of the Ana2/LC8 complex. Instead, the broad elution peak of the Ana2$_{(156\text{-}251)}$/SNAP-LC8 complex may reflect a moderate strength of Ana2 self-tetramerisation (unknown affinity) or the low-micromolar affinity of LC8 towards Ana2, or a combination of both. Neither of those scenarios can be unambiguously confirmed or refuted unless more data on apo-Ana2 becomes available. In fact, some answers to these puzzles have recently arrived with the study by Cottee *et al.* [70] - see below.

5.4.3 Different modes of Ana2/STIL oligomerisation, similar downstream effects

Cottee *et al.* [70] presented evidence for the importance of the self-oligomerising domain of Ana2 (*Drosophila* equivalent of SAS-5) in centriole assembly. The authors showed that a 37 amino acids-long central region of Ana2 (residues 193-229; predicted to form a coiled coil) oligomerised into a stable tetramer across a wide range of concentrations. They also solved a crystal structure of $Ana2_{(193\text{-}229)}$, showing that instead of a coiled coil, this domain of Ana2 forms a parallel tetrameric α-helical bundle. The authors reported that they were unable to extract the exact affinity strength of Ana2 homo-interaction due to incredibly stable nature of this oligomer. This finding cleared the earlier uncertainty regarding the stabilising role of LC8 after binding to two short epitopes of Ana2 [305]. It appears that instead of simply stabilising Ana2 tetramer, which is already a strong oligomer by itself, the LC8/Ana2 interaction has a more complex role. LC8 was shown to be part of a ternary complex between Ana2 and Mushroom body defect (Mud) protein by Wang *et al.* [340] (Co-IP assays). Mud is a known regulator of spindle positioning, and therefore the LC8/Ana2 interaction is also likely a part of this process. Interestingly, formation of the Mud/LC8/Ana2 ternary complex appears not to be co-dependent on prior interaction between Ana2 and LC8, as yeast two-hybrid assays and Co-IP experiments revealed that Mud can interact independently with both Ana2 and LC8. The relationship between these three proteins appears to be somewhat complex because if the role of Ana2 in spindle positioning was simply to anchor the centrosome to cell cortex during asymmetric cell division (through Mud which localises both to cell cortex and to PCM [340]), then the requirement for LC8 in this context would be conceptually redundant. Instead, interactions between these three proteins are likely changing in time and space, and one can envision that Ana2-bound LC8 could provide attachment sites for dynein, a microtubule-associated minus-end directed motor (walking towards centrosomes), to actively facilitate positioning of the centrosome, and hence the entire spindle, towards the cell cortex. Although studies of mechanistic involvement of centrosomes in spindle positioning are still in their infancy, the work by Wang *et al.* [340] and Slevin *et al.* [305] pave the way towards exciting new discoveries.

Unlike SAS-5, which uses two independent domains to form large oligomers, Ana2 and STIL appear to be relatively simpler in this respect, with only the central coiled-coil region having so far been reported to form oligomeric species [70, 305]. An interesting question arises as to whether Ana2/STIL protein could possibly harbour a second oligomerisation domain besides the coiled coil, that would be equivalent to the *Implico* domain of *C. elegans* SAS-5. It is difficult to identify such oligomerisation region in the primary amino acid sequence or via secondary structure alignments, as both the sequences and the secondary structural elements of SAS-5 and Ana2/STIL have significantly diverged. Moreover, besides the N-terminal domain of STIL (not present in Ana2) and the coiled-coil region, the remaining segments of STIL are predicted to be largely disordered, suggesting that the *Implico* domain of *C. elegans* SAS-5 may not have a direct equivalent in Ana2 or STIL. This prediction was indeed confirmed by Arquint *et al.* [12] who co-expressed full-length human STIL with truncated STIL versions. The two over-expressed STIL constructs were fused to a distinct tag (FLAG or Myc), allowing for their pull-down and differential identification after Co-IP. Although such assays might be prone to error, these results strongly suggested that besides the central coiled-coil region, STIL does not contain any other self-oligomerising domain. Such divergence in organisation of large oligomers between SAS-5 and Ana2/STIL may reflect differences in cartwheel assembly and hence centriole architecture between those species [142, 262]. It is likely that the divergence between members of the SAS-5/Ana2/STIL family of proteins is even more extensive, as no homologue of SAS-5 has yet been identified in the classical model organism for centriolar sudies - *C. reinhardtii*, or in the nematode species *Trichinella spiralis*, which is expected to feature centrioles with flat SAS-6 cartwheels rather than spirals (Note: *T. spiralis* SAS-6 features small alanine residue instead of a nematode-conserved bulky hydrophobic residue at the key position regarded as the switch between a ring- and a spiral-type oligomerisation) [142]. Thus, it is possible that *C. elegans* SAS-5 has co-evolved with SAS-6 in order to accommodate the structural shift of cartwheel assembly, from a ring to a spiral.

5.4.4 New insights into centriole assembly pathway in flies and humans

The crystal structure of the predicted coiled-coil region of Ana2 (residues 193-229) revealed a parallel, symmetrical tetrameric oligomer [70]. Interestingly, although the homo-oligomerisation of $Ana2_{(193\text{-}229)}$ is driven in part by contacts typically found in canonical coiled coils - through residues corresponding to *a* and *d* positions in the heptad repeat, there are also important contacts formed by the *e* and *g* residues, which are completely buried within the tetramer interface. While the central helical bundle of Ana2 formed a stable tetramer, its equivalent in human STIL (residues 717-758) appeared to be much weaker, displaying concentration-dependent oligomerisation. Such self-oligomerisation of $STIL_{717\text{-}758}$, however, is not observed in context of the STIL/PLK4(PB3) complex [12], highlighting a possible switch mechanism between apo-STIL tetramer and PLK4-bound STIL monomer. Interestingly, mutations which disrupt binding between the coiled-coil region of STIL and PLK4(PB3) do not affect the interaction between STIL and the N-terminal segment of PLK4 (residues 1-570; contains L1 - the second STIL-binding region), indicating that the two distinct STIL-binding domains of PLK4: L1 (residues 265-570) and PB3 (residues 880-970), interact with the coiled coil of STIL via two distinct modes.

For disruptive mutagenesis, Arquint *et al.* [12] targeted mostly residues of the STIL α-helix which formed hydrophobic contacts with PLK4(PB3), with the exception of Q737 (salt bridge with $PLK4_{S956}$) and Q739 (hydrogen bond to $PLK4_{G922}$). Cottee *et al.* [70] attempted a rather brute-force method to disrupt Ana2 tetramer. After mutating almost 30% of the short α-helical tetrameric bundle of Ana2 (all ten *d* and *g* positions of the heptad-repeat were targeted), they found that Ana2 is no longer efficiently targeted to centrioles, and causes centriole assembly failure. It turns out that mutations specifically designed by these two groups to either disrupt the STIL/PLK4(PB3) complex [12] or the homo-tetramer [70] are actually clustered on the same face of the STIL central α-helix, and thus affect both events simultaneously: PLK4 binding and self-oligomerisation. Therefore, using the above-mentioned mutants of STIL central α-helix, it is not possible to distinguish whether the centriole assembly process failed due to disrupted PLK4 binding or self-assembly or both. More work is required to design better mutants that can unambiguously differentiate

between these two events. Moreover, because these two events share the same α-helix interface of STIL, it is likely that PLK4 binding and self-assembly of STIL are mutually exclusive, posing a question of how STIL can shift between the two states during the centriole assembly process. While the affinity of the interaction between STIL α-helix and PLK4(PB3) is in the nanomolar range [12], Cottee *et al.* [70] reported that the self-tetramerisation of STIL is much weaker (micromolar range). The latter observation suggests that after centriolar localisation of tetrameric STIL, and subsequent binding to PLK4(PB3), STIL is likely to be trapped in the monomeric state, and only occasionally get released to self-assemble into tetramers. On the other hand, if STIL was as strong as Ana2 (or SAS-5 for that matter) in self-oligomerisation, then it will be able to escape the PLK4-bound state much more frequently, and through self-oligomerisation-propelled avidity effect, trigger nucleation of SAS-6 cartwheels. The discrepancy between Ana2 and STIL affinity for self-interaction should be addressed with more experimental work to clarify this matter.

In summary, despite minor differences, recent insights into the early steps of centriole assembly pathway of both *Drosophila* and vertebrates, support the model where tetrameric STIL is recruited to centrioles by Cep192- [60, 98, 139] and Cep152-held cloud of PLK4 [177, 256, 311]. Via its central α-helix, STIL binds to two distinct sites on PLK4: L1 and PB3 [12]. While binding to PB3 likely disrupts self-oligomerisation of STIL, binding to L1 is predicted to shield PLK4 from degradation (see below). Either of those binding events appears to also trigger activation of PLK4 kinase [12, 185, 231], and subsequent auto-phosphorylation of PLK4. Phosphorylated PLK4 is predicted to activate ubiquitin-dependent proteosomal pathway (via βTrCP-SCF) [130, 149], which acts to degrade activated PLK4, therefore eliminating majority of the centrosomal PLK4 cloud, and leaving only a dot-like PLK4 population that is protected from degradation by prior binding to STIL. PLK4(L1) contains the DSGHAT motif (residues 284-289), which upon phosphorylation is recognised by βTrCP-SCF, leading to PLK4 degradation. It is not unlikely that the event of STIL binding to PLK4(L1) shields this motif [12, 252] and protects it from recognition by degradation machinery. Activated PLK4 also phosphorylates STAN motif of STIL [98, 191, 252], which then leads to recruitment of SAS-6 to centrioles (via interaction with STIL), and subsequent cartwheel formation.

An immense progress in understanding of the early events of centriole assembly pathway has been made in recent years, providing novel insights into interactions between proteins that drive centriole assembly in flies and humans. These studies have given a fresh new perspective and a solid foundation to answering a number of exciting new questions. How is What is the mechanism of PLK4 activation by STIL binding? Why does STIL only protect a sub-population of PLK4 from degradation, and how does this translate to dot-localisation of PLK4? What does STIL/PLK4(L1) complex look like, and what is the strength of this interaction? What happens to STIL after it binds PLK4(PB3) and switches from a tetramer to a monomer? How stable is the complex between PLK4(PB3) and STIL - is dissociation a frequent event? What does the binding interface between STIL and SAS-6 look like? Does STIL/Ana2/SAS-5 family of proteins play a structural role in centrioles, or is it purely required for nucleation of SAS-6 cartwheels/spirals?

5.5 The future of centriole research in *C. elegans*

While the molecular details of centriole assembly pathway in vertebrates and flies appear to accumulate at a great pace, similar studies are still missing for worms. Even though the pioneering *C. elegans* work, which first established the order of centriolar protein recruitment [87, 262], gave worms a head start over studies in other model organisms, it was quickly outperformed by work in human cells and in *Drosophila*, and has since remained far behind. These two model systems (human cells and flies) have had two critical advantages over *C. elegans*, which allowed them to make quick progress in revealing the molecular details of the centriole assembly pathway: fast genetics and availability of single-molecule-based super-resolution imaging methods. To establish a stable mutant line in *C. elegans*, it used to take up to several months, as opposed to a week or two in other systems. Moreover, recent advances in fluorescence imaging technology [83] were not available for work on whole organisms, due to limited imaging depth. Small size of *C. elegans* centrioles, compared to those of flies or human, was also a significant obstacle. *C. elegans* studies have been strongly relying on classical fluorescence microscopy, optical resolution of which is limited by diffraction of visible light, to ∼250 nm in lateral direction and ∼600 nm in axial direction [152]. A number of recent technological advances, however, have made it

possible for the *C. elegans* field to finally catch up with their peers. First, MosSCI (Mos1-mediated single-copy insertion) system was established in worms [112, 113], which unlike the biolistic bombardment method [270], it allows for efficient creation of single-insertion mutant lines within a couple of weeks only. Moreover, CRISPR-Cas9 system was adapted for the work with worms, and holds a great potential for fast and accurate genome editing, including not only loss-of-function mutations but also precise knock-ins of fusion proteins, such as GFP [90, 111]. Finally, a major advancement in super-resolution microscopy has lifted the limit of depth for imaging, and made it possible to detect molecules up to $\sim$10 μm deep inside an animal, with accuracy of up to 20 nm resolution [284, 334].

With the advent of new tools for genome manipulation and microscopy, the future of research on *C. elegans* centrioles is starting to look bright again. Given that the centriole assembly pathway of *C. elegans* appears to be the least complex of all centriole-bearing organisms, it may once again prove to be the system of choice for centriolar research. It will be exciting to see if more can be learned about the mystery events surrounding the very first steps of centriole assembly pathway in *C. elegans*, following recruitment of ZYG-1 to centrioles. Does ZYG-1 localisation to centrioles resemble that of PLK4? Is there are dot-like localisation of ZYG-1, and if yes, then what is the mechanism behind it? Is SAS-5 and SAS-6 indeed recruited to centrioles as a complex, or rather the recruitment proceeds in sequence? Does *C. elegans* SAS-6 really form a spiral cartwheel, and if yes, then how does SAS-5 fit into this process? These and other exciting questions await new discoveries!

Chapter 6

Materials and methods

6.1 Construction of pFloat vectors

pFloat was derived from pET30a (Novagen, Madison, WI) and further engineered to meet needs of a modern structural biology lab (see Section 2.1 *Construction of pFloat expression system* on page 63). The resulting polylinker site of pFloat includes N-terminal His_6- and solubility tags, and a human rhinovirus 3C protease (HRV-3C) cleavage sequence (LEVLFQGP). In order to accelerate construct screening, we also inserted a *ccd* gene cassette, amplified from pDONR211 (Invitrogen, Grand Island, NY), inside the multi-cloning site (Figure 2.4 on page 72). A number of various solubility or fluorescence tags included: thioredoxin (pFloat-TRX), glutathione S-transferase (pFloat-GST), S-tag (pFloat-STAG), colicin E9 immunity protein (pFloat-Im9), maltose-binding protein (pFloat-MBP), small ubiquiting-like modifier (pFloat-SUMO), haloalkane dehalogenase (pFloat-HALO), trigger factor (pFloat-TF), N utilization substance protein A (pFloat-NusA), acidic protein MsyB (pFloat-MsyB), yellow fluorescent protein (pFloat-YFP) and enhanced green fluorescent protein (pFloat-EGFP).

Tag sequences were amplified either from the pOPIN [30], pGEX (GE Healthcare, Little Chalfont, UK) vector systems, or directly from genomic DNA of *Escherichia coli* strain BL21. All amplifications were performed using high-fidelity Phusion polymerase (New England Biolabs, Ipswich, MA), and the resulting set of pFloat vectors was constructed using seamless cloning methods (Gibson Assembly [119, 120], New England Biolabs). See Appendix Section B.1 on page 170 for all pFloat sequences.

6.2 Protein cloning, expression and purification

6.2.1 SAS-5

Fragments of *sas-5* from *C. brenneri*, *C. remanei*, *C. sinica* and *C. tropicalis* were synthesised by IDT (Coralville, IA). *C. briggsae sas-5* was amplified from genomic DNA (gifted by Marie-Anne Félix, Ecole Normale Supérieure, France). Full-length or gene fragments encoding SAS-5 constructs were cloned into pFloat vectors (See Appendix Section B.3.1 on page 186 for expression construct sequences) and transformed into *E. coli* strain BL21(DE3) Rosetta2 (Novagen). The cells were grown at 37 °C until mid-log phase (∼0.6 optical density measured at 600 nm), at which point the temperature was lowered to 18 °C, and protein production induced by isopropyl-β-D-1-thiogalactopyranoside (IPTG, 0.5 mM final concentration).

After ∼16 h incubation, cells were harvested by centrifugation (15 min at 6,000 RPM, JLA-8.1000, Beckman Coulter, Brea, CA), and resuspended in PBS (phosphate-buffered saline) solution (10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4) supplemented with 350 mM NaCl, 0.1 mM phenylmethanesulfonyl fluoride, a protease inhibitor cocktail (cOmplete, Roche, Basel, Switzerland), 0.25 mg/ml lysozyme and 10 μg/ml DNAse I. Resuspended cells were disrupted by sonication while chilled on ice, and cell lysate was cleared by centrifugation (2 h at 20,000 RPM, JA-25.50, Beckman Coulter) and filtration (0.22 μm syringe filter, EMD Millipore, Billerica, MA). Imidazole at 20 mM final concentration was added to clarified lysates prior to 2 h incubation with Ni-NTA (QIAGEN, Venlo, Netherlands) metal-affinity resin pre-equilibrated with PBS supplemented with 350 mM NaCl and 20 mM imidazole. Resin beads were washed with 50 column volumes (CV) of PBS supplemented with 350 mM NaCl, 1 mM β-mercaptoethanol and 20 mM imidazole, and protein was eluted with 6 CV of PBS supplemented with 350 mM NaCl, 1 mM β-mercaptoethanol and 500 mM imidazole. Eluted proteins were dialysed against 10 mM Tris or Na_2HPO_4, 50 mM NaCl, 3 mM β-mercaptoethanol, pH 8.0 or pH 6.5.

Further purification of proteins was performed by anion- (HiTrap Q HP, GE Healthcare) or cation-exchange (HiTrap SP HP, GE Healthcare) chromatography in 10 mM Tris or Na_2HPO_4, 50 mM NaCl, 3 mM β-mercaptoethanol, pH 8.0 (for anion-exchange) or pH 6.5 (for cation-exchange) running buffer, and proteins were eluted

using a NaCl gradient to 2 M. His_6- and solubility-tags were removed, whenever necessary, by 16 h incubation with HRV-3C protease at 4 °C. The final purification step involved size-exclusion chromatography using HiLoad Superdex 75 16/60, Superdex 200 16/60 or Superose 6 16/70 columns (GE Healthcare) equilibrated in 10 mM Tris, 300 mM NaCl, 2 mM β-mercaptoethanol, pH 7.5 buffer. All purification steps were performed at 4 °C.

Pure protein fractions were buffer-exchanged by dialysis, and concentrated by centrifugal ultra-filtration. Protein concentration was estimated by UV absorption at 280 nm, and protein identity confirmed by electrospray ionisation mass-spectrometry.

6.2.2 HRV-3C protease

Human rhinovirus 3C protease (residues 1-181) was cloned into pFloat immediately after the His_6 sequence. The resulting construct did not include a HRV-3C cleavage site, and thus the His_6-tag was uncleavable by HRV-3C (see Appendix Section B.3.2 on page 189 for the complete sequence of the expression construct). Following transformation of the plasmid into *E. coli* BL21(DE3) Rosetta2 strain (Novagen), the cells were grown and harvested in the same way as SAS-5 (see Section 6.2.1 *SAS-5* on page 157). HRV-3C purification was performed using one-step IMAC (Ni-NTA resin), and the protocol was identical to that of SAS-5 (see Section 6.2.1 *SAS-5* on page 157). Following one-step IMAC, 98 %-pure HRV-3C was aliquoted (200 μl), flash-frozen in liquid nitrogen and stored at -20 °C.

6.3 Biophysical studies

6.3.1 Circular dichroism spectroscopy

Far-UV CD spectra (260-180 nm) were measured using a Jasco (Easton, MD) J-815 Spectropolarimeter at 10 °C, in a QS glass cuvette (1 mm light path, Hellma Analytics, Müllheim, Germany). Protein samples were buffer-exchanged to a 10 mM NaH_2PO_4, 100 mM NaF, 1 mM β-mercaptoethanol, pH 7.5 buffer prior to measurements. CD sample concentrations are shown in Table A.2 on page 169. Data were truncated at high-tension voltage exceeding 800 V. After buffer/cuvette background subtraction, data were smoothed using a Savitzky-Golay filter [293]. CD spectra were

deconvoluted using CDSSTR [164], accessed through the DICHROWEB server [343]. Measurements of thermal stability monitored CD signal at 222 nm while temperature was increased by 1 °C/min between 4 °C and 90 °C.

6.3.2 Size-exclusion chromatography multi-angle light scattering

Chromatographic separations were performed at 20 °C using analytical Superdex 75 10/300, Superdex 200 10/300 or Superose 6 10/300 columns (GE Healthcare) and a Shimadzu (Kyoto, Japan) chromatography system. MALS analysis was carried out in-line after SEC using a Heleos8+ multi-angle light scattering detector and an Optilab T-rEX refractive index (RI) detector (Wyatt Technology, Goleta, CA). 1:1 dilution series of protein samples in 10 mM Tris, 300 mM NaCl, 2 mM β-mercaptoethanol, pH 7.5 buffer were sequentially injected into the SEC-MALS system, and the resulting MALS, RI and UV traces were processed in ASTRA 6 (Wyatt Technology). The traces were baseline-corrected, and the peak shapes were corrected for band broadening. The molecular mass (M_w) was calculated via light scattering. Zimm multipole expansion was used to retrieve the particle size information (radius of gyration, r_g) and second virial coefficient (A_2). On-column protein concentration was calculated from the differential RI, assuming dn/dc of 0.1850 ml/g. SEC-MALS data were fitted with a two-state model to derive association parameters, and the resulting K_d was estimated from the inflection point of the fit. Details of protein concentrations used in SEC-MALS experiments, monomeric molecular weights and columns used are shown in Table A.3 on page 169.

6.3.3 Native mass spectrometry

Protein samples in 250 mM ammonium acetate, pH 7.5 buffer, were analysed by native mass spectrometry using a modified Synapt G1 High Definition Mass Spectrometry Quadrupole Time of Flight instrument (Waters, Milford, MA) [46], calibrated using 10 mg/ml CsI. 2 μl aliquots of sample were delivered by nano-electrospray ionisation via gold-coated capillaries prepared in-house [141]. Instrument parameters during data collection were as follows: source pressure 6.0 mbar, capillary voltage 1.20 kV, cone voltage 50 V, trap energy 10 V, bias voltage 5 V and trap pressure 0.0163 mbar.

6.3.4 Negative-stain electron microscopy

SAS-5_{FL} protein samples (5 μl, 0.1 mg/ml) in 10 mM Tris, 300 mM NaCl, 2 mM β-mercaptoethanol, pH 7.5 buffer were applied to glow-discharged home-made carbon-coated copper mesh grids and allowed to adsorb for 1 min. After blotting off any excess liquid, negative stain was applied by adding 3 μl of 1 % wt/vol uranyl acetate to the grid. Blotting and staining was repeated 2 times, and the grids were then allowed to air-dry for 10 min. The grids were imaged in a JEOL (Tokyo, Japan) JEM-2010 electron microscope equipped with a high brightness LaB_6 filament and an ORIUS SC1000 (Model 832) camera (Gatan, Pleasanton, CA). Images were acquired using an accelerating voltage of 200 kV, at a nominal magnification of 16,000× to 52,000×.

6.4 X-ray crystallography

Crystals were obtained using the sitting-drop vapour diffusion technique. A Mosquito robot (TTP LabTech, Melbourn, UK) was used to setup 200 nl drops with 1:1 and 1.3:0.7 ratios of protein to mother liquor.

Model quality was assessed by MolProbity [58]. For graphical representation, we used UCSF Chimera [264]. Initial crystal inspection and data collection took place at the DLS beamline I04-1.

6.4.1 SAS-5_{CC}

For SAS-5_{CC}, protein in 10 mM Tris pH 7.5, 200 mM NaCl buffer and at concentration of 20 mg/ml was mixed with 0.1 M MOPS/HEPES-Na pH 7.5, 12.5 % wt/vol PEG 1000, 12.5 % wt/vol PEG 3350, 12.5 % wt/vol MPD, 0.03 M $NaNO_3$, 0.03 M Na_2HPO_4, 0.03 M $(NH_4)_2SO_4$. Crystals developed at 4 °C in 30 days, were flash-cooled in liquid nitrogen and diffracted to a maximum resolution of 1.73 Å at the Diamond Light Source (DLS, Harwell, UK) beamline I04. SAS-5_{CC} crystallised in a high-symmetry space group $P2_12_12_1$ with eight molecules per asymmetric unit. Crystallographic data (statistics in Table 3.1 on page 93) were processed in XDS [167] and scaled in XSCALE [166]. Scaling revealed substantial diffraction anisotropy, which

was corrected using the Diffraction Anisotropy Server (UCLA Molecular Biology Institute [294, 317]) whilst applying a high resolution cut-off of 1.8 Å. Statistics of the anisotropically-corrected data are shown in Table A.1 on page 168.

Phase information for SAS-5_{CC} was obtained from a highly redundant dataset with maximum resolution of 1.9 Å collected on Pb-derivatised crystals at a wavelength of 0.947 Å at the DLS beamline I03, anomalous signal of which extended to 2.55 Å (CC_{ano} of 14 %). For Pb-derivatisation, the native SAS-5_{CC} crystals were incubated with 1 mM trimethyl lead (IV) acetate for 16 h prior to cooling. After applying a resolution cut-off of 3.35 Å (at this resolution $<d''/sig>$ dropped below 1.3), we performed a sub-structure search with SHELXD [300], which yielded two disconnected populations of solutions. Solutions displaying higher CC_{all} and CC_{weak} values (∼35 % and ∼33 %, respectively) included two heavy-atom sites of high occupancy (1.00 and 0.99). Occupancy for other potential sites dropped rapidly below 0.12, suggesting that a possible solution was found. We extended the resolution cut-off to 1.9 Å in order to carry out refinement of the heavy-atom sites, as well as to determine the correct sub-structure enantiomorph via cycles of density modification (DM) and protein backbone tracing in SHELXE [300]. A clear difference in contrast and connectivity between the two hands allowed us to resolve the enantiomorph ambiguity, and returned an interpretable electron density map (mapCC of 0.33). An initial model was build into this map using BUCCANEER [72] in cycles with density modifications by PARROT [352] (365 residues built, 215 identified out of 427 residues in the asymmetric unit). Iterative model building with COOT [101] and refinement against the anisotropically-corrected native data in Buster 2.10 using automatic NCS (non-crystallographic symmetry) and TLS (translation-libration-screw-rotation) restraints [306] yielded the final SAS-5_{CC} model.

Due to pseudo-translational symmetry, arising from similarities between non-crystallographic symmetry operators to true crystallographic symmetry, the correct space group was ambiguous until the structure was solved, built and refined in a number of potential space groups. Most solutions in orthorhombic space groups: $P2_12_12$, $P2_122_1$, $P22_12_1$ and P222, could not support building of all protein chains without overlaps with symmetry-related molecules. Owing to screw symmetry in all three crystallographic axes, $P2_12_12_1$ space group provided enough space in the unit cell to

accommodate SAS-5_{CC} molecules without any steric clashes.

6.4.2 SAS-$5_{CC\text{-}L2}$

For SAS-$5_{CC\text{-}L2}$, protein in 10 mM HEPES pH 7.5, 150 mM NaCl, 1 mM DTT buffer and at concentration of 15 mg/ml was mixed with 0.01 M $ZnCl_2$, 0.1M sodium acetate pH 5.0, 20 % wt/vol PEG 6,000 [238]. Because only small (<5 μm), over-nucleated crystal needles developed after incubation at 18 °C for 3 days, we set up a number of optimisation screens to determine which divalent cation was most suited for crystal growth, and which pH/[precipitant] combination could yield the largest crystals. It turned out that Zn^{2+} promoted growth of the longest needle crystals (∼20 μm); however, these were not thick enough (only ∼1 μm) for synchrotron data collection.

In order to improve the thickness of crystals, we performed micro-seeding of crushed crystals, generated via vortexing with a Seed Bead (Hampton Research, Aliso Viejo, CA) for 1 min, into multiple sparse-matrix crystallisation screens by mixing protein (15 mg/ml, 150 nl) with mother liquor (135 nl) and crystal micro-seeds (15 nl, 1:1000 seed stock) in 300 nl drops using a Mosquito robot (TTP LabTech). Seeding into 0.2 M NH_4NO_3 pH 6.3, 20 % wt/vol PEG 3,350 [238] yielded thicker crystals (∼4 μm) which were cryo-protected by brief immersion in mother liquor supplemented with 20 % vol/vol glycerol, flash-cooled in liquid nitrogen, and diffracted to 1.72 Å at the DLS beamline I04. SAS-$5_{CC\text{-}L2}$ crystallised in a high-symmetry space group $C222_1$ with three molecules per asymmetric unit. Diffraction data were processed in XDS [167] and scaled in XSCALE [167] (Table 3.1 on page 93).

Phase information was obtained from a highly redundant dataset with maximum resolution of 2.6 Å collected on Hg-derivatised crystals at 0.917 Å wavelength, at the DLS beamline I04-1. For Hg-derivatisation, native SAS-$5_{CC\text{-}L2}$ crystals were incubated with 1 mM mercury (II) acetate for 16 h prior to cooling. Initial attempts at structure solution with SAD were unsuccessful due to low anomalous signal extending to ∼3.3 Å. We attempted optimisation of the anomalous signal by integrating various ranges of data, and found that the first 825 ° of data collected at one crystal centring provided the strongest anomalous signal, extending up to 2.8 Å (CC_{ano} of 12 %). Such truncation of data allowed us to find a SAD solution with SHARP [43], which located

7 heavy-atom sites of relatively low occupancy (between 0.49 and 0.12), and the overall phasing power of 0.592 and figure-of-merit (FOM) of 0.20 for acentric and 0.13 for centric reflections. Although density modifications with SOLOMON [2] successfully resolved the enantiomorph ambiguity, the resulting map was uninterpretable to any auto-building software. However, several areas of the electron density displayed distinctly connected features into which we manually built idealised poly-alanine α-helices, and used the resulting model to perform MRSAD (molecular replacement with single-wavelength anomalous dispersion) with PHASER [224], which refined the heavy-atom sites to only two high-occupancy Hg atoms, effectively improving the SAD solution (overall figure-of-merit of 0.57). The fine features of the resulting map also improved, and cycles of manual model re-building in COOT [101] and automated re-building in PHENIX.autobuild [325] eventually traced the entire protein molecule. The initial auto-building after MRSAD resulted in 114 residues built and 32 identified, out of 225 residues in the asymmetric unit. Iterative model building with COOT [101] and refinement against the native data in Buster 2.10 using automatic NCS and TLS restraints [306] yielded the final SAS-$5_{\text{CC-L2}}$ model.

6.4.3 SAS-5_{Imp}

For SAS-5_{Imp}, protein in 10 mM HEPES pH 7.5, 150 mM NaCl, 1 mM DTT buffer and at concentration of 10 mg/ml was mixed with 0.1 M MES/imidazole pH 6.4, 10 % wt/vol PEG 20,000, 20 % vol/vol PEG MME 550, 0.1 M carboxylic acids mixture (0.02 M sodium formate, 0.02 M ammonium acetate, 0.02 M trisodium citrate, 0.02 M sodium potassium *L*-tartrate, 0.02 M sodium oxamate) [129]. Crystals developed at 18 °C in 3 days, were cryo-protected by brief immersion in mother liquor supplemented with 20 % vol/vol glycerol, flash-cooled in liquid nitrogen and diffracted to 1.0 Å at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) beamline ID14-4. SAS-5_{Imp} crystallised in a low-symmetry space group $P2_1$ with two molecules per asymmetric unit. Diffraction data were processed in MOSFLM [202] and scaled in Aimless [105] (Table 3.1 on page 93).

Phase information was obtained from a highly redundant dataset with maximum resolution of 1.7 Å collected on Hg-derivatised crystals at a wavelength of 1.007 Å at the DLS beamline I02. For Hg-derivatisation, native SAS-5_{Imp} crystals were incubated

with 2 mM mercury (II) acetate for 16 h prior to cooling. Phasing by SAD was performed using PHENIX.autosol [3, 324] which located and refined 2 Hg sites to produce a density map with initial figure-of-merit of 0.39. Initial model building was done with PHENIX.autobuild (86 residues built and 51 identified out of 116 residues in the asymmetric unit) [325]. Iterative model building was carried out with COOT [101] and refinement against the native data in PHENIX.refine [4] using explicit hydrogens and atom-level anisotropic restraints.

6.5 Molecular dynamics simulations

All MD simulations were performed using GROMACS 4.5 [273] and the GROMOS96 53a6 force field [254]. The crystal structure of SAS-5_{CC} was placed in a cubic simulation box of 1452 nm^3 volume with periodic boundary conditions. The protein was solvated by adding 10,637 explicit SPC water molecules. Na^+ and Cl^- ions were added to the final concentration of 0.1 M. The system was first minimised with the steepest descend algorithm, equilibrated in constant temperature (NVT, 300 K) for 100 ps, and followed by equilibration in constant temperature and pressure (NPT, 1 bar) for 500 ps. The pressure was controlled with an isotropic Parrinello-Rahman barostat [258] applied to the entire system with a time constant of 2.0 ps and compressibility of 4.5×10^{-5} bar^{-1}. The temperature was controlled with two velocity rescaling thermostats [23] applied to the protein and solvent, with a time constant of 0.1 ps. Electrostatic interactions were calculated using the particle-mesh Ewald summation method [1, 80]. Ten replicate simulations of 45 ns duration were performed in the Oxford Advanced Research Computing facility [280].

6.6 Nematode strains and RNA interference

The genomic sequence of *sas-5* (Wormbase, F35B12.5a) was recoded for resistance to RNAi directed against the endogenous gene (complete sequence in Appendix Section B.2 on page 185), and synthesised by GenScript (Piscataway, NJ). Intronic sequences were kept intact, as well as the bases neighbouring the splice sites: 3 bases at 3' ends and 6 bases at 5' ends of exons. The codon adaptation index was preserved at the level of the wild type (WT: 0.777, recoded: 0.712). The recoded

gene was amplified and cloned into pIC26 using the SpeI restriction site. pIC26 contains a *pie-1* promoter and 3'UTR cassette, as well as a GFP coding sequence, which includes introns (specifically designed for efficient expression of GFP in *C. elegans*), and is located upstream of and in frame with *sas-5*, effectively yielding GFP-SAS-5 fusion protein upon expression [57]. *C. elegans* culture was performed according to standard procedures [42]. GFP-SAS-5 transgenic animals were generated by biolistic bombardment method [270]. Integrated lines were recovered after bombardment with *gfp::sas-5 L141E* (strain GZ1301; isIs51*{pie-1::gfp::sas-5[recL141E]}*) and *I247E* (strain GZ1302; isIs52*{pie-1::gfp::sas-5[recI247E]}*), but GFP-SAS-5 bombardment led to the recovery of non-integrated lines. For these experiments a line with <50 % transmission of the extra chromosomal array was used (strain GZ1300; isEx5*{pie-1::gfp::sas-5[rec]}*). RNAi was carried out by selecting L3-L4 wild-type hermaphrodites and feeding them for 26 h at 24 °C using the *sas-5(RNAi)* feeding strain, which targets nucleotides 301-1170 of the genomic sequence [88].

6.7 Antibodies and immunofluorescence

Embryos were prepared for immunofluorescence as described previously [126], with a fixation time of 2 min in -20 °C methanol and an additional blocking step in PBS, 2 % wt/vol bovine serum albumin before antibody incubation. Primary and secondary antibodies were diluted in PBS as follows: rabbit anti-GFP (gift from Viesturs Simanis, EPFL, 1/500), mouse anti-IFA ([204], 1/50), goat anti-GFP (Abcam, Cambridge, UK, ab6673, 1/250), goat anti-rabbit Alexa Fluor 488, goat anti-mouse Alexa Fluor 568, donkey anti-goat Alexa Fluor 488, donkey anti-mouse Alexa Fluor 568 (Life Technologies, Grand Island, NY, all 1/500). Primary antibodies were incubated overnight at 4 °C and secondary incubations were at 20 °C for 1 h. 1 μg/ml Hoechst 33,258 (Sigma-Aldrich, St. Louis, MO) was used to visualise DNA.

6.8 Microscopy

Time-lapse differential interference contrast (DIC) microscopy of early embryos was carried out as described previously [126], recording one image every 5 s at 24 °C. A Zeiss (Oberkochen, Germany) LSM 700 confocal microscope was used for fluorescence

microscopy. Z-sections were imaged at intervals of ∼0.3 μm. Fluorescence images in Figure 4.11 on page 130 are maximum intensity projections. Dual time-lapse and fluorescence DIC imaging was performed on a Zeiss Axioplan 2 [22]. The motorised filter wheel, two external shutters, and a 1392 × 1040 pixel, 12-bit Photometrics CoolSNAP ES2 (Photometrics, Tucson, AZ) were controlled by μManager [100]. Images were acquired with an exposure time of 100 ms for the DIC and 200 ms for the GFP channel using the Zeiss Filter Set 10 (GFP). The GFP images presented in Figure 4.9 on page 127 are maximum intensity projections of 8 z-sections acquired with an interval of 0.5 μm. For quantification of fluorescence intensities, the integrated intensity was measured from three areas of fixed size within each embryo and then averaged. The average background fluorescence of wild-type (GFP-negative) embryos was subtracted, then values were normalised to the average value of the GFP-SAS-5 wild-type intensity. All images were processed and fluorescence intensities measured using ImageJ [297].

6.9 Western blotting

Embryos were synchronised by bleaching adult worms that hatched and matured at 24 °C. 50 adult worms were then collected manually, boiled in Laemmli buffer [196] for 5 min, sonicated, and loaded on a 4-12 % wt/vol SDS-PAGE (Biorad, Hercules, CA). Proteins were transferred to a 0.45 μm nitrocellulose membrane, blocked with 5 % wt/vol milk powder in PBS, 0.1 % vol/vol Tween-20, and blotted with rabbit anti-SAS-5 (full length, 1/1000, [88]) or rabbit anti-GFP (1/1000, gift from Viesturs Simanis, EPFL), then re-blotted with mouse anti-α-tubulin (1/10,000, DM1α, Sigma-Aldrich) antibodies. Immuno-complexes were visualized using anti-rabbit or anti-mouse HRP conjugated secondary antibodies at 1/5000 dilution (Promega, Madison, WI), chemiluminescence kit (Roche), and X-ray films (Fujifilm, Tokyo, Japan).

6.10 Sequence notes

SAS-5 protein sequences in this study correspond to UniProt database entries Q20010 (*C. elegans*), Q60M48 (*C. briggsae*), G0NH40 (*C. brenneri*) and E3LPN2 (*C. remanei*). *C. sinica* and *C. tropicalis* SAS-5 sequences were identified from the respec-

tive genomes by Blast searches using Wormbase interface [138]. For *C. sinica*, the University of Edinburgh *Caenorhabditis sp. 5* genome project (Csp5_scaffold_00353) was used, while for *C. tropicalis*, the Genome Institute at Washington University *Caenorhabditis sp. 11* genome project (Csp11.Scaffold629) served as reference. The flexible linker used in SAS-5_{FLEX} to replace the *sticky strand* segment has the sequence SGAAGSSGAAGSSG.

Appendix A

Resolution (Å)	Number of observed reflections		Redundancy		Completeness (%)		R_{merge} (%)		Mean $I/\sigma(I)$	
	Before	After	Before	After	Before	After	Before	After	Before	After
8.05	1428	1428	2.3	2.3	92.3	92.3	2.9	2.9	22.3	21.9
5.69	2617	2598	2.5	2.5	92.1	92.2	3.3	3.3	22.3	21.9
4.65	3322	3317	2.6	2.6	94.4	94.4	4.0	4.0	22.4	22
4.03	3949	3923	2.6	2.6	93.3	93.2	3.6	3.6	22.7	22.3
3.6	4581	4527	2.7	2.7	95.8	95.8	3.7	3.8	21.7	21.3
3.29	4925	4957	2.6	2.6	95.1	95.3	4.1	4.1	19.8	19.4
3.04	5760	5703	2.8	2.8	96.9	96.7	5.0	5.0	17.1	16.8
2.85	6026	6003	2.8	2.8	96.3	96.4	5.1	5.1	14.2	13.9
2.68	6696	6649	2.9	2.9	97.8	97.8	6.5	6.4	12.1	11.9
2.55	7098	7098	2.9	3	98.6	98.5	7.5	7.5	10.1	9.9
2.43	7178	7136	2.8	2.8	98.2	98.1	9.4	9.3	8.2	8.1
2.32	7824	7781	3	3	98.9	99.1	11.8	11.7	6.8	6.7
2.23	7844	7233	2.8	2.6	97.4	91.1	15.9	14.9	5.2	5.4
2.15	8551	6609	3	2.3	99.0	79.6	20.4	15.9	4.4	5.3
2.08	8757	5726	3	2	99.0	66.9	29.3	20.0	3.1	4.4
2.01	8707	4542	2.9	1.5	98.7	51.8	41.8	22.9	2.3	4.1
1.95	9329	3355	3	1.1	99.1	33.9	59.1	24.0	1.7	4.2
1.9	9150	2341	2.9	0.7	98.1	20.5	93.3	30.2	1.1	3.8
1.85	9824	1532	3	0.5	99.2	11.6	114.0	28.6	0.9	4.5
1.8	9748	502	2.8	0.1	97.9	3.9	143.1	31.8	0.8	3.9
total	133314	92,960	2.9	2	97.6	68.2	4.9	4.4	8.2	11.3

Table A.1: **Anisotropy correction statistics for native X-ray diffraction dataset of SAS-5_{CC}.** Derived from the UCLA Diffraction Anisotropy Server [317].

Protein name	Concentration (μM)
SAS-5$_{\mathrm{FL}}$	1.5
SAS-5$_{\Delta(282\text{-}295)}$	3
SAS-5$_{\mathrm{FLEX}}$	3
SAS-5$_{(2\text{-}265)}$	6
SAS-5$_{(2\text{-}265)\ \mathrm{L141E}}$	6
SAS-5$_{(2\text{-}265)\ \mathrm{L141E/M167E}}$	6
SAS-5$_{(2\text{-}265)\ \mathrm{I247E}}$	6
SAS-5$_{(2\text{-}265)\ \mathrm{L141E/I247E}}$	6
SAS-5$_{(125\text{-}265)}$	10
SAS-5$_{\mathrm{CC}}$	20
SAS-5$_{\mathrm{CC\ L141E}}$	20
SAS-5$_{\mathrm{Imp}}$	20
SAS-5$_{\mathrm{Imp\ I247E}}$	20
SAS-5$_{\mathrm{N\text{-}terminus}}$	20
SAS-5$_{\mathrm{C\text{-}terminus}}$	5

Table A.2: **Protein concentrations used in CD experiments.**

Protein name	Monomeric MW (kDa)	Size exclusion column	Minimum concentration (μM)*	Maximum concentration (μM)*
SAS-5$_{\mathrm{FL}}$	46.2	Superose 6	0.10	0.57
MsyB-SAS-5$_{\mathrm{FLEX}}$	61.7	Superose 6	0.03	9.5
MsyB-SAS-5$_{(2\text{-}265)}$ (Figure 4.2 C, D)	46.3	Superose 6	0.02	41
MsyB-SAS-5$_{(2\text{-}265)}$ (Figure 4.7 F)	46.3	Superdex 200	0.05	41
SAS-5$_{(2\text{-}265)\ \mathrm{L141E}}$	30.1	Superdex 200	1.32	63
SAS-5$_{(2\text{-}265)\ \mathrm{L141E/M167E}}$	30.1	Superdex 200	1.16	63
SAS-5$_{(2\text{-}265)\ \mathrm{I247E}}$	30.1	Superdex 200	0.05	63
SAS-5$_{(2\text{-}265)\ \mathrm{L141E/I247E}}$	30.1	Superdex 200	0.82	63
SAS-5$_{\mathrm{Imp}}$	6.5	Superdex 75	2.9	292
SAS-5$_{\mathrm{CC\text{-}L}}$	12.5	Superdex 75	0.18	152
C. briggsae SAS-5$_{(96\text{-}199)}$	12.7	Superdex 75	0.10	149
C. brenneri SAS-5$_{(1\text{-}82)}$	10.2	Superdex 75	0.21	186
C. remanei SAS-5$_{(98\text{-}206)}$	13.0	Superdex 75	0.21	146
C. sinica SAS-5$_{(94\text{-}204)}$	13.0	Superdex 75	0.16	145
C. tropicalis SAS-5$_{(94\text{-}199)}$	12.5	Superdex 75	0.12	152

Table A.3: **Details of SEC-MALS experiments.**
* Refers to on-column concentration calculated from the protein differential refractive index.

Appendix B

B.1 pFloat family of vectors

B.1.1 pFloat - full sequence

```
   1 CGAGATCGAT CTCGATCCCG CGAAATTAAT ACGACTCACT ATAGGGGAAT TGTGAGCGGA
  61 TAACAATTCC CCTCTAGAAA TAATTTTGTT TAACTTTAAG AAGGAGATAT ACATATGCAC
 121 CATCATCATC ATCATTCTTC TGGTCTGGAA GTTCTGTTCC AGGGGCCCGG ATCCGAATTC
 181 TCATTATATT CCCCAGAACA TCAGGTTAAT GGCGTTTTTG ATGTCATTTT CGCGGTGGCT
 241 GAGATCAGCC ACTTCTTCCC CGATAACGGA GACCGGCACA CTGGCCATAT CGGTGGTCAT
 301 CATGCGCCAG CTTTCATCCC CGATATGCAC CACCGGGTAA AGTTCACGGG AGACTTTATC
 361 TGACAGCAGA CGTGCACTGG CCAGGGGGAT CACCATCCGT CGCCCGGGCG TGTCAATAAT
 421 ATCACTCTGT ACATCCACAA ACAGACGATA ACGGCTCTCT CTTTTATAGG TGTAAACCTT
 481 AAACTGCATT TCACCAGCCC CTGTTCTCGT CAGCAAAAGA GCCGTTCATT TCAATAAACC
 541 GGGCGACCTC AGCCATCCCT TCCTGATTTT CCGCTTTCCA GCGTTCGGCA CGCAGACGAC
 601 GGGCTTCATT CTGCATGGTT GTGCTTACCA GACCGGAGAT ATTGACATCA TATATGCCTT
 661 GAGCAACTGA TAGCTGTCGC TGTCAACTGT CACTGTAATA CGCTGCTTCA TAGCATACCT
 721 CTTTTTGACA TACTTCGGGT ATACATATCA GTATATATTC TTATACCGCA AAAATCAGCG
 781 CGCAAATACG CATACTGTTA TCTGGCTTTT AGTAAGCCGA GCTCCGTCGA CAAGCTTGCG
 841 GCCGCACTCG AGCACCACCA CCACCACCAC TGAGATCCGG CTGCTAACAA AGCCCGAAAG
 901 GAAGCTGAGT TGGCTGCTGC CACCGCTGAG CAATAACTAG CATAACCCCT TGGGGCCTCT
 961 AAACGGGTCT TGAGGGGTTT TTTGCTGAAA GGAGGAACTA TATCCGGATT GGCGAATGGG
1021 ACGCGCCCTG TAGCGGCGCA TTAAGCGCGG CGGGTGTGGT GGTTACGCGC AGCGTGACCG
1081 CTACACTTGC CAGCGCCCTA GCGCCCGCTC CTTTCGCTTT CTTCCCTTCC TTTCTCGCCA
1141 CGTTCGCCGG CTTTCCCCGT CAAGCTCTAA ATCGGGGGCT CCCTTTAGGG TTCCGATTTA
1201 GTGCTTTACG GCACCTCGAC CCCAAAAAAC TTGATTAGGG TGATGGTTCA CGTAGTGGGC
```

1261 CATCGCCCTG ATAGACGGTT TTTCGCCCTT TGACGTTGGA GTCCACGTTC TTTAATAGTG
1321 GACTCTTGTT CCAAACTGGA ACAACACTCA ACCCTATCTC GGTCTATTCT TTTGATTTAT
1381 AAGGGATTTT GCCGATTTCG GCCTATTGGT TAAAAAATGA GCTGATTTAA CAAAAATTTA
1441 ACGCGAATTT TAACAAAATA TTAACGTTTA CAATTTCAGG TGGCACTTTT CGGGGAAATG
1501 TGCGCGGAAC CCCTATTTGT TTATTTTTCT AAATACATTC AAATATGTAT CCGCTCATGA
1561 ATTAATTCTT AGAAAAACTC ATCGAGCATC AAATGAAACT GCAATTTATT CATATCAGGA
1621 TTATCAATAC CATATTTTTG AAAAAGCCGT TTCTGTAATG AAGGAGAAAA CTCACCGAGG
1681 CAGTTCCATA GGATGGCAAG ATCCTGGTAT CGGTCTGCGA TTCCGACTCG TCCAACATCA
1741 ATACAACCTA TTAATTTCCC CTCGTCAAAA ATAAGGTTAT CAAGTGAGAA ATCACCATGA
1801 GTGACGACTG AATCCGGTGA GAATGGCAAA AGTTTATGCA TTTCTTTCCA GACTTGTTCA
1861 ACAGGCCAGC CATTACGCTC GTCATCAAAA TCACTCGCAT CAACCAAACC GTTATTCATT
1921 CGTGATTGCG CCTGAGCGAG ACGAAATACG CGATCGCTGT TAAAAGGACA ATTACAAACA
1981 GGAATCGAAT GCAACCGGCG CAGGAACACT GCCAGCGCAT CAACAATATT TTCACCTGAA
2041 TCAGGATATT CTTCTAATAC CTGGAATGCT GTTTTCCCGG GGATCGCAGT GGTGAGTAAC
2101 CATGCATCAT CAGGAGTACG GATAAAATGC TTGATGGTCG GAAGAGGCAT AAATTCCGTC
2161 AGCCAGTTTA GTCTGACCAT CTCATCTGTA ACATCATTGG CAACGCTACC TTTGCCATGT
2221 TTCAGAAACA ACTCTGGCGC ATCGGGCTTC CCATACAATC GATAGATTGT CGCACCTGAT
2281 TGCCCGACAT TATCGCGAGC CCATTTATAC CCATATAAAT CAGCATCCAT GTTGGAATTT
2341 AATCGCGGCC TAGAGCAAGA CGTTTCCCGT TGAATATGGC TCATAACACC CCTTGTATTA
2401 CTGTTTATGT AAGCAGACAG TTTTATTGTT CATGACCAAA ATCCCTTAAC GTGAGTTTTC
2461 GTTCCACTGA GCGTCAGACC CCGTAGAAAA GATCAAAGGA TCTTCTTGAG ATCCTTTTTT
2521 TCTGCGCGTA ATCTGCTGCT TGCAAACAAA AAAACCACCG CTACCAGCGG TGGTTTGTTT
2581 GCCGGATCAA GAGCTACCAA CTCTTTTTCC GAAGGTAACT GGCTTCAGCA GAGCGCAGAT
2641 ACCAAATACT GTCCTTCTAG TGTAGCCGTA GTTAGGCCAC CACTTCAAGA ACTCTGTAGC
2701 ACCGCCTACA TACCTCGCTC TGCTAATCCT GTTACCAGTG GCTGCTGCCA GTGGCGATAA
2761 GTCGTGTCTT ACCGGGTTGG ACTCAAGACG ATAGTTACCG GATAAGGCGC AGCGGTCGGG
2821 CTGAACGGGG GGTTCGTGCA CACAGCCCAG CTTGGAGCGA ACGACCTACA CCGAACTGAG
2881 ATACCTACAG CGTGAGCTAT GAGAAAGCGC CACGCTTCCC GAAGGGAGAA AGGCGGACAG
2941 GTATCCGGTA AGCGGCAGGG TCGGAACAGG AGAGCGCACG AGGGAGCTTC CAGGGGGAAA
3001 CGCCTGGTAT CTTTATAGTC CTGTCGGGTT TCGCCACCTC TGACTTGAGC GTCGATTTTT
3061 GTGATGCTCG TCAGGGGGGC GGAGCCTATG GAAAAACGCC AGCAACGCGG CCTTTTTACG
3121 GTTCCTGGCC TTTTGCTGGC CTTTTGCTCA CATGTTCTTT CCTGCGTTAT CCCCTGATTC

3181 TGTGGATAAC CGTATTACCG CCTTTGAGTG AGCTGATACC GCTCGCCGCA GCCGAACGAC

3241 CGAGCGCAGC GAGTCAGTGA GCGAGGAAGC GGAAGAGCGC CTGATGCGGT ATTTTCTCCT

3301 TACGCATCTG TGCGGTATTT CACACCGCAT ATATGGTGCA CTCTCAGTAC AATCTGCTCT

3361 GATGCCGCAT AGTTAAGCCA GTATACACTC CGCTATCGCT ACGTGACTGG GTCATGGCTG

3421 CGCCCCGACA CCCGCCAACA CCCGCTGACG CGCCCTGACG GGCTTGTCTG CTCCCGGCAT

3481 CCGCTTACAG ACAAGCTGTG ACCGTCTCCG GGAGCTGCAT GTGTCAGAGG TTTTCACCGT

3541 CATCACCGAA ACGCGCGAGG CAGCTGCGGT AAAGCTCATC AGCGTGGTCG TGAAGCGATT

3601 CACAGATGTC TGCCTGTTCA TCCGCGTCCA GCTCGTTGAG TTTCTCCAGA AGCGTTAATG

3661 TCTGGCTTCT GATAAAGCGG GCCATGTTAA GGGCGGTTTT TTCCTGTTTG GTCACTGATG

3721 CCTCCGTGTA AGGGGGATTT CTGTTCATGG GGGTAATGAT ACCGATGAAA CGAGAGAGGA

3781 TGCTCACGAT ACGGGTTACT GATGATGAAC ATGCCCGGTT ACTGGAACGT TGTGAGGGTA

3841 AACAACTGGC GGTATGGATG CGGCGGGACC AGAGAAAAAT CACTCAGGGT CAATGCCAGC

3901 GCTTCGTTAA TACAGATGTA GGTGTTCCAC AGGGTAGCCA GCAGCATCCT GCGATGCAGA

3961 TCCGGAACAT AATGGTGCAG GGCGCTGACT TCCGCGTTTC CAGACTTTAC GAAACACGGA

4021 AACCGAAGAC CATTCATGTT GTTGCTCAGG TCGCAGACGT TTTGCAGCAG CAGTCGCTTC

4081 ACGTTCGCTC GCGTATCGGT GATTCATTCT GCTAACCAGT AAGGCAACCC CGCCAGCCTA

4141 GCCGGGTCCT CAACGACAGG AGCACGATCA TGCGCACCCG TGGGGCCGCC ATGCCGGCGA

4201 TAATGGCCTG CTTCTCGCCG AAACGTTTGG TGGCGGGACC AGTGACGAAG GCTTGAGCGA

4261 GGGCGTGCAA GATTCCGAAT ACCGCAAGCG ACAGGCCGAT CATCGTCGCG CTCCAGCGAA

4321 AGCGGTCCTC GCCGAAAATG ACCCAGAGCG CTGCCGGCAC CTGTCCTACG AGTTGCATGA

4381 TAAAGAAGAC AGTCATAAGT GCGGCGACGA TAGTCATGCC CCGCGCCCAC CGGAAGGAGC

4441 TGACTGGGTT GAAGGCTCTC AAGGGCATCG GTCGAGATCC CGGTGCCTAA TGAGTGAGCT

4501 AACTTACATT AATTGCGTTG CGCTCACTGC CCGCTTTCCA GTCGGGAAAC CTGTCGTGCC

4561 AGCTGCATTA ATGAATCGGC CAACGCGCGG GGAGAGGCGG TTTGCGTATT GGGCGCCAGG

4621 GTGGTTTTTC TTTTCACCAG TGAGACGGGC AACAGCTGAT TGCCCTTCAC CGCCTGGCCC

4681 TGAGAGAGTT GCAGCAAGCG GTCCACGCTG GTTTGCCCCA GCAGGCGAAA ATCCTGTTTG

4741 ATGGTGGTTA ACGGCGGGAT ATAACATGAG CTGTCTTCGG TATCGTCGTA TCCCACTACC

4801 GAGATGTCCG CACCAACGCG CAGCCCGGAC TCGGTAATGG CGCGCATTGC GCCCAGCGCC

4861 ATCTGATCGT TGGCAACCAG CATCGCAGTG GGAACGATGC CCTCATTCAG CATTTGCATG

4921 GTTTGTTGAA AACCGGACAT GGCACTCCAG TCGCCTTCCC GTTCCGCTAT CGGCTGAATT

4981 TGATTGCGAG TGAGATATTT ATGCCAGCCA GCCAGACGCA GACGCGCCGA GACAGAACTT

5041 AATGGGCCCG CTAACAGCGC GATTTGCTGG TGACCCAATG CGACCAGATG CTCCACGCCC

```
5101 AGTCGCGTAC CGTCTTCATG GGAGAAAATA ATACTGTTGA TGGGTGTCTG GTCAGAGACA
5161 TCAAGAAATA ACGCCGGAAC ATTAGTGCAG GCAGCTTCCA CAGCAATGGC ATCCTGGTCA
5221 TCCAGCGGAT AGTTAATGAT CAGCCCACTG ACGCGTTGCG CGAGAAGATT GTGCACCGCC
5281 GCTTTACAGG CTTCGACGCC GCTTCGTTCT ACCATCGACA CCACCACGCT GGCACCCAGT
5341 TGATCGGCGC GAGATTTAAT CGCCGCGACA ATTTGCGACG GCGCGTGCAG GGCCAGACTG
5401 GAGGTGGCAA CGCCAATCAG CAACGACTGT TTGCCCGCCA GTTGTTGTGC CACGCGGTTG
5461 GGAATGTAAT TCAGCTCCGC CATCGCCGCT TCCACTTTTT CCCGCGTTTT CGCAGAAACG
5521 TGGCTGGCCT GGTTCACCAC GCGGGAAACG GTCTGATAAG AGACACCGGC ATACTCTGCG
5581 ACATCGTATA ACGTTACTGG TTTCACATTC ACCACCCTGA ATTGACTCTC TTCCGGGCGC
5641 TATCATGCCA TACCGCGAAA GGTTTTGCGC CATTCGATGG TGTCCGGGAT CTCGACGCTC
5701 TCCCTTATGC GACTCCTGCA TTAGGAAGCA GCCCAGTAGT AGGTTGAGGC CGTTGAGCAC
5761 CGCCGCCGCA AGGAATGGTG CATGCAAGGA GATGGCGCCC AACAGTCCCC CGGCCACGGG
5821 GCCTGCCACC ATACCCACGC CGAAACAAGC GCTCATGAGC CCGAAGTGGC GAGCCCGATC
5881 TTCCCCATCG GTGATGTCGG CGATATAGGC GCCAGCAACC GCACCTGTGG CGCCGGTGAT
5941 GCCGGCCACG ATGCGTCCGG CGTAGAGGAT
```

Relevant sequence features are highlighted as follows:

(consecutively from 1 to 5970)

T7 promoter | *lacO* | Ribosome binding site | START codon | N-terminal His_6-tag | N-terminal HRV-3C cleavage site | Multiple cloning site, part 1 | *ccdB* | *ccd* promoter | Multiple cloning site, part 2 | C-terminal His_6-tag | T7 terminator | f1 *ori* | KanR | pBR332 *ori* | *rop* | *mbeC* | Split *tetA*, part 2 | *lacI* | Split *tetA*, part 1

B.1.2 pFloat and pFloat-TAG - short sequences

Shortened sequences of all pFloat-TAG vectors and their expected protein products are listed in this section. The toxic *ccd* cassette has been removed for clarity. See B.1.1 on page 170 as reference for the position of the *ccd* cassette within the MCS.

The following elements are common for all pFloat vectors:

His_6-tag | flexible linkers | solubility tag | HRV-3C cleavage site | TEV cleavage site | Multiple cloning site

B.1.2.1 pFloat

```
  1 ATGCACCATC ATCATCATCA TTCTTCTGGT CTGGAAGTTC TGTTCCAGGG GCCCGGATCC
 61 GAATTCGAGC TCCGTCGACA AGCTTGCGGC CGCACTCGAG CACCACCACC ACCACCACTG
121 AGATCCGGCT GCTAA

  1 MHHHHHHSSG LEVLFQGPGS EFELRRQACG RTRAPPPPPL RSGC
```

B.1.2.2 pFloat-STAG

```
  1 ATGCACCATC ATCATCATCA TTCTTCTGGT AAAGAAACCG CTGCTGCTAA ATTCGAACGC
 61 CAGCACATGG ACAGCTCTTC TGGTCTGGAA GTTCTGTTCC AGGGGCCCGG ATCCGAATTC
121 GAGCTCCGTC GACAAGCTTG CGGCCGCACT CGAGCACCAC CACCACCACC ACTGAGATCC
181 GGCTGCTAA

  1 MHHHHHHSSG KETAAAKFER QHMDSSSGLE VLFQGPGSEF ELRRQACGRT RAPPPPPLRS
 61 GC
```

B.1.2.3 pFloat-GST

```
  1 ATGCACCATC ATCATCATCA TTCTTCTGGT ATGTCCCCTA TACTAGGTTA TTGGAAAATT
 61 AAGGGCCTTG TGCAACCCAC TCGACTTCTT TTGGAATATC TTGAAGAAAA ATATGAAGAG
121 CATTTGTATG AGCGCGATGA AGGTGATAAA TGGCGAAACA AAAAGTTTGA ATTGGGTTTG
181 GAGTTTCCCA ATCTTCCTTA TTATATTGAT GGTGATGTTA AATTAACACA GTCTATGGCC
241 ATCATACGTT ATATAGCTGA CAAGCACAAC ATGTTGGGTG GTTGTCCAAA AGAGCGTGCA
301 GAGATTTCAA TGCTTGAAGG AGCGGTTTTG GATATTAGAT ACGGTGTTTC GAGAATTGCA
361 TATAGTAAAG ACTTTGAAAC TCTCAAAGTT GATTTTCTTA GCAAGCTACC TGAAATGCTG
421 AAAATGTTCG AAGATCGTTT ATGTCATAAA ACATATTTAA ATGGTGATCA TGTAACCCAT
481 CCTGACTTCA TGTTGTATGA CGCTCTTGAT GTTGTTTTAT ACATGGACCC AATGTGCCTG
541 GATGCGTTCC CAAAATTAGT TTGTTTTAAA AAACGTATTG AAGCTATCCC ACAAATTGAT
601 AAGTACTTGA AATCCAGCAA GTATATAGCA TGGCCTTTGC AGGGCTGGCA AGCCACGTTT
661 GGTGGTGGCG ACCATCCTCC AAAATCGGAT CTGTCTTCTG GTCTGGAAGT TCTGTTCCAG
721 GGGCCCGGAT CCGAATTCGA GCTCCGTCGA CAAGCTTGCG GCCGCACTCG AGCACCACCA
```

```
781 CCACCACCAC TGAGATCCGG CTGCTAA

  1 MHHHHHHSSG MSPILGYWKI KGLVQPTRLL LEYLEEKYEE HLYERDEGDK WRNKKFELGL
 61 EFPNLPYYID GDVKLTQSMA IIRYIADKHN MLGGCPKERA EISMLEGAVL DIRYGVSRIA
121 YSKDFETLKV DFLSKLPEML KMFEDRLCHK TYLNGDHVTH PDFMLYDALD VVLYMDPMCL
181 DAFPKLVCFK KRIEAIPQID KYLKSSKYIA WPLQGWQATF GGGDHPPKSD LSSGLEVLFQ
241 GPGSEFELRR QACGRTRAPP PPPLRSGC
```

B.1.2.4 pFloat-TRX

```
  1 ATGCACCATC ATCATCATCA TTCTTCTGGT AGCGATAAAA TTATTCACCT GACTGACGAC
 61 AGTTTTGACA CGGATGTACT CAAAGCGGAC GGGGCGATCC TCGTCGATTT CTGGGCAGAG
121 TGGTGCGGTC CGTGCAAAAT GATCGCCCCG ATTCTGGATG AAATCGCTGA CGAATATCAG
181 GGCAAACTGA CCGTTGCAAA ACTGAACATC GATCAAAACC CTGGCACTGC GCCGAAATAT
241 GGCATCCGTG GTATCCCGAC TCTGCTGCTG TTCAAAAACG GTGAAGTGGC GGCAACCAAA
301 GTGGGTGCAC TGTCTAAAGG TCAGTTGAAA GAGTTCCTCG ACGCTAACCT GGCCTCTTCT
361 GGTCTGGAAG TTCTGTTCCA GGGGCCCGGA TCCGAATTCG AGCTCCGTCG ACAAGCTTGC
421 GGCCGCACTC GAGCACCACC ACCACCACCA CTGAGATCCG GCTGCTAA

  1 MHHHHHHSSG SDKIIHLTDD SFDTDVLKAD GAILVDFWAE WCGPCKMIAP ILDEIADEYQ
 61 GKLTVAKLNI DQNPGTAPKY GIRGIPTLLL FKNGEVAATK VGALSKGQLK EFLDANLASS
121 GLEVLFQGPG SEFELRRQAC GRTRAPPPPP LRSGC
```

B.1.2.5 pFloat-Im9

```
  1 ATGCACCATC ATCATCATCA TTCTTCTGGT GAACTGAAGC ATAGCATTAG TGATTATACA
 61 GAAGCTGAAT TTTTACAGCT TGTAACAACA ATTTGTAATG CGGACACTTC CAGTGAAGAA
121 GAACTGGTTA AATTGGTTAC ACACTTTGAG GAAATGACTG AGCACCCTAG TGGTAGTGAT
181 TTAATATATT ACCCAAAAGA AGGTGATGAT GACTCACCTT CAGGTATTGT AAACACAGTA
241 AAACAATGGC GAGCCGCTAA CGGTAAGTCA GGATTTAAAC AGTCTTCTGG TCTGGAAGTT
301 CTGTTCCAGG GGCCCGGATC CGAATTCGAG CTCCGTCGAC AAGCTTGCGG CCGCACTCGA
```

```
361 GCACCACCAC CACCACCACT GAGATCCGGC TGCTAA

  1 MHHHHHHSSG ELKHSISDYT EAEFLQLVTT ICNADTSSEE ELVKLVTHFE EMTEHPSGSD
 61 LIYYPKEGDD DSPSGIVNTV KQWRAANGKS GFKQSSGLEV LFQGPGSEFE LRRQACGRTR
121 APPPPPLRSG C
```

B.1.2.6 pFloat-MBP

```
   1 ATGCACCATC ATCATCATCA TTCTTCTGGT AAAATCGAAG AAGGTAAACT GGTAATCTGG
  61 ATTAACGGCG ATAAAGGCTA TAACGGTCTC GCTGAAGTCG GTAAGAAATT CGAGAAAGAT
 121 ACCGGAATTA AAGTCACCGT TGAGCATCCG GATAAACTGG AAGAGAAATT CCCACAGGTT
 181 GCGGCAACTG GCGATGGCCC TGACATTATC TTCTGGGCAC ACGACCGCTT TGGTGGCTAC
 241 GCTCAATCTG GCCTGTTGGC TGAAATCACC CCGGACAAAG CGTTCCAGGA CAAGCTGTAT
 301 CCGTTTACCT GGGATGCCGT ACGTTACAAC GGCAAGCTGA TTGCTTACCC GATCGCTGTT
 361 GAAGCGTTAT CGCTGATTTA TAACAAAGAT CTGCTGCCGA ACCCGCCAAA AACCTGGGAA
 421 GAGATCCCGG CGCTGGATAA AGAACTGAAA GCGAAAGGTA AGAGCGCGCT GATGTTCAAC
 481 CTGCAAGAAC CGTACTTCAC CTGGCCGCTG ATTGCTGCTG ACGGGGGTTA TGCGTTCAAG
 541 TATGAAAACG GCAAGTACGA CATTAAAGAC GTGGGCGTGG ATAACGCTGG CGCGAAAGCG
 601 GGTCTGACCT TCCTGGTTGA CCTGATTAAA AACAAACACA TGAATGCAGA CACCGATTAC
 661 TCCATCGCAG AAGCTGCCTT TAATAAAGGC GAAACAGCGA TGACCATCAA CGGCCCGTGG
 721 GCATGGTCCA ACATCGACAC CAGCAAAGTG AATTATGGTG TAACGGTACT GCCGACCTTC
 781 AAGGGTCAAC CATCCAAACC GTTCGTTGGC GTGCTGAGCG CAGGTATTAA CGCCGCCAGT
 841 CCGAACAAAG AGCTGGCAAA AGAGTTCCTC GAAAACTATC TGCTGACTGA TGAAGGTCTG
 901 GAAGCGGTTA ATAAAGACAA ACCGCTGGGT GCCGTAGCGC TGAAGTCTTA CGAGGAAGAG
 961 TTGGCGAAAG ATCCACGTAT TGCCGCCACT ATGGAAAACG CCCAGAAAGG TGAAATCATG
1021 CCGAACATCC CGCAGATGTC CGCTTTCTGG TATGCCGTGC GTACTGCGGT GATCAACGCC
1081 GCCAGCGGTC GTCAGACTGT CGATGAAGCC CTGAAAGACG CGCAGACTTC TTCTGGTCTG
1141 GAAGTTCTGT TCCAGGGGCC CGGATCCGAA TTCGAGCTCC GTCGACAAGC TTGCGGCCGC
1201 ACTCGAGCAC CACCACCACC ACCACTGAGA TCCGGCTGCT AA

   1 MHHHHHHSSG KIEEGKLVIW INGDKGYNGL AEVGKKFEKD TGIKVTVEHP DKLEEKFPQV
  61 AATGDGPDII FWAHDRFGGY AQSGLLAEIT PDKAFQDKLY PFTWDAVRYN GKLIAYPIAV
```

```
121 EALSLIYNKD LLPNPPKTWE EIPALDKELK AKGKSALMFN LQEPYFTWPL IAADGGYAFK
181 YENGKYDIKD VGVDNAGAKA GLTFLVDLIK NKHMNADTDY SIAEAAFNKG ETAMTINGPW
241 AWSNIDTSKV NYGVTVLPTF KGQPSKPFVG VLSAGINAAS PNKELAKEFL ENYLLTDEGL
301 EAVNKDKPLG AVALKSYEEE LAKDPRIAAT MENAQKGEIM PNIPQMSAFW YAVRTAVINA
361 ASGRQTVDEA LKDAQTSSGL EVLFQGPGSE FELRRQACGR TRAPPPPPLR SGC
```

B.1.2.7 pFloat-SUMO

```
  1 ATGCACCATC ATCATCATCA TTCTTCTGGT AGCGATAGCG AAGTGAACCA GGAAGCGAAA
 61 CCGGAAGTTA AACCGGAAGT GAAACCGGAA ACCCATATTA ATCTGAAAGT TAGCGACGGC
121 AGCAGCGAAA TCTTTTTTAA AATTAAAAAA ACCACCCCGC TGCGTCGCCT GATGGAAGCC
181 TTTGCGAAAC GTCAGGGTAA AGAAATGGAT AGCCTGCGCT TTCTGTATGA CGGCATCCGT
241 ATTCAGGCCG ATCAGACCCC GGAAGACCTG GATATGGAAG ACAACGATAT TATTGAAGCG
301 CATCGCGAAC AGATCTCTTC TGGTCTGGAA GTTCTGTTCC AGGGGCCCGG ATCCGAATTC
361 GAGCTCCGTC GACAAGCTTG CGGCCGCACT CGAGCACCAC CACCACCACC ACTGAGATCC
421 GGCTGCTAA

  1 MHHHHHHSSG SDSEVNQEAK PEVKPEVKPE THINLKVSDG SSEIFFKIKK TTPLRRLMEA
 61 FAKRQGKEMD SLRFLYDGIR IQADQTPEDL DMEDNDIIEA HREQISSGLE VLFQGPGSEF
121 ELRRQACGRT RAPPPPPLRS GC
```

B.1.2.8 pFloat-HALO

```
  1 ATGCACCATC ATCATCATCA TTCTTCTGGT TCCGAAATCG GTACTGGCTT TCCATTCGAC
 61 CCCCATTATG TGGAAGTCCT GGGCGAGCGC ATGCACTACG TCGATGTTGG TCCGCGCGAT
121 GGCACCCCTG TGCTGTTCCT GCACGGTAAC CCGACCTCCT CCTACGTGTG GCGCAACATC
181 ATCCCGCATG TTGCACCGAC CCATCGCTGC ATTGCTCCAG ACCTGATCGG TATGGGCAAA
241 TCCGACAAAC CAGACCTGGG TTATTTCTTC GACGACCACG TCCGCTTCAT GGATGCCTTC
301 ATCGAAGCCC TGGGTCTGGA AGAGGTCGTC CTGGTCATTC ACGACTGGGG CTCCGCTCTG
361 GGTTTCCACT GGGCCAAGCG CAATCCAGAG CGCGTCAAAG GTATTGCATT TATGGAGTTC
421 ATCCGCCCTA TCCCGACCTG GGACGAATGG CCAGAATTTG CCCGCGAGAC CTTCCAGGCC
481 TTCCGCACCA CCGACGTCGG CCGCAAGCTG ATCATCGATC AGAACGTTTT TATCGAGGGT
```

```
 541 ACGCTGCCGA TGGGTGTCGT CCGCCCGCTG ACTGAAGTCG AGATGGACCA TTACCGCGAG
 601 CCGTTCCTGA ATCCTGTTGA CCGCGAGCCA CTGTGGCGCT TCCCAAACGA GCTGCCAATC
 661 GCCGGTGAGC CAGCGAACAT CGTCGCGCTG GTCGAAGAAT ACATGGACTG GCTGCACCAG
 721 TCCCCTGTCC CGAAGCTGCT GTTCTGGGGC ACCCCAGGCG TTCTGATCCC ACCGGCCGAA
 781 GCCGCTCGCC TGGCCAAAAG CCTGCCTAAC TGCAAGGCTG TGGACATCGG CCCGGGTCTG
 841 AATCTGCTGC AAGAAGACAA CCCGGACCTG ATCGGCAGCG AGATCGCGCG CTGGCTGTCT
 901 ACTCTGGGCT CTTCTGGTCT GGAAGTTCTG TTCCAGGGGC CCGGATCCGA ATTCGAGCTC
 961 CGTCGACAAG CTTGCGGCCG CACTCGAGCA CCACCACCAC CACCACTGAG ATCCGGCTGC
1021 TAA
```

```
   1 MHHHHHHSSG SEIGTGFPFD PHYVEVLGER MHYVDVGPRD GTPVLFLHGN PTSSYVWRNI
  61 IPHVAPTHRC IAPDLIGMGK SDKPDLGYFF DDHVRFMDAF IEALGLEEVV LVIHDWGSAL
 121 GFHWAKRNPE RVKGIAFMEF IRPIPTWDEW PEFARETFQA FRTTDVGRKL IIDQNVFIEG
 181 TLPMGVVRPL TEVEMDHYRE PFLNPVDREP LWRFPNELPI AGEPANIVAL VEEYMDWLHQ
 241 SPVPKLLFWG TPGVLIPPAE AARLAKSLPN CKAVDIGPGL NLLQEDNPDL IGSEIARWLS
 301 TLGSSGLEVL FQGPGSEFEL RRQACGRTRA PPPPPLRSGC
```

B.1.2.9 pFloat-TF

```
   1 ATGCACCATC ATCATCATCA TTCTTCTGGT CAAGTTTCAG TTGAAACCAC TCAAGGCCTT
  61 GGCCGCCGTG TAACGATTAC TATCGCTGCT GACAGCATCG AGACCGCTGT TAAAAGCGAG
 121 CTGGTCAACG TTGCGAAAAA AGTACGTATT GACGGCTTCC GCAAAGGCAA AGTGCCAATG
 181 AATATCGTTG CTCAGCGTTA TGGCGCGTCT GTACGCCAGG ACGTTCTGGG TGACCTGATG
 241 AGCCGTAACT TCATTGACGC CATCATTAAA GAAAAAATCA ATCCGGCTGG CGCACCGACT
 301 TATGTTCCGG GCGAATACAA GCTGGGTGAA GACTTCACTT ACTCTGTAGA GTTTGAAGTT
 361 TATCCGGAAG TTGAACTGCA GGGTCTGGAA GCGATCGAAG TTGAAAAACC GATCGTTGAA
 421 GTGACCGACG CTGACGTTGA CGGCATGCTG GATACTCTGC GTAAACAGCA GGCGACCTGG
 481 AAAGAAAAAG ACGGCGCTGT TGAAGCAGAA GACCGCGTAA CCATCGACTT CACCGGTTCT
 541 GTAGACGGCG AAGAGTTCGA AGGCGGTAAA GCGTCTGATT TCGTACTGGC GATGGGCCAG
 601 GGTCGTATGA TCCCGGGCTT TGAAGACGGT ATCAAAGGCC ACAAAGCTGG CGAAGAGTTC
 661 ACCATCGACG TGACCTTCCC GGAAGAATAC CACGCAGAAA ACCTGAAAGG TAAAGCAGCG
 721 AAATTCGCTA TCAACCTGAA GAAAGTTGAA GAGCGTGAAC TGCCGGAACT GACTGCAGAA
```

```
 781 TTCATCAAAC GTTTCGGCGT TGAAGATGGT TCCGTAGAAG GTCTGCGCGC TGAAGTGCGT
 841 AAAAACATGG AGCGCGAGCT GAAGAGCGCC ATCCGTAACC GCGTTAAGTC TCAGGCGATC
 901 GAAGGTCTGG TAAAAGCTAA CGACATCGAC GTACCGGCTG CGCTGATCGA CAGCGAAATC
 961 GACGTTCTGC GTCGCCAGGC TGCACAGCGT TTCGGTGGCA ACGAAAAACA AGCTCTGGAA
1021 CTGCCGCGCG AACTGTTCGA AGAACAGGCT AAACGCCGCG TAGTTGTTGG CCTGCTGCTG
1081 GGCGAAGTTA TCCGCACCAA CGAGCTGAAA GCTGACGAAG AGCGCGTGAA AGGCCTGATC
1141 GAAGAGATGG CTTCTGCGTA CGAAGATCCG AAAGAAGTTA TCGAGTTCTA CAGCAAAAAC
1201 AAAGAACTGA TGGACAACAT GCGCAATGTT GCTCTGGAAG AACAGGCTGT TGAAGCTGTA
1261 CTGGCGAAAG CGAAAGTGAC TGAAAAAGAA ACCACTTTCA ACGAGCTGAT GAACCAGCAG
1321 GCGTCTTCTG GTCTGGAAGT TCTGTTCCAG GGGCCCGGAT CCGAATTCGA GCTCCGTCGA
1381 CAAGCTTGCG GCCGCACTCG AGCACCACCA CCACCACCAC TGAGATCCGG CTGCTAA

   1 MHHHHHHSSG QVSVETTQGL GRRVTITIAA DSIETAVKSE LVNVAKKVRI DGFRKGKVPM
  61 NIVAQRYGAS VRQDVLGDLM SRNFIDAIIK EKINPAGAPT YVPGEYKLGE DFTYSVEFEV
 121 YPEVELQGLE AIEVEKPIVE VTDADVDGML DTLRKQQATW KEKDGAVEAE DRVTIDFTGS
 181 VDGEEFEGGK ASDFVLAMGQ GRMIPGFEDG IKGHKAGEEF TIDVTFPEEY HAENLKGKAA
 241 KFAINLKKVE ERELPELTAE FIKRFGVEDG SVEGLRAEVR KNMERELKSA IRNRVKSQAI
 301 EGLVKANDID VPAALIDSEI DVLRRQAAQR FGGNEKQALE LPRELFEEQA KRRVVVGLLL
 361 GEVIRTNELK ADEERVKGLI EEMASAYEDP KEVIEFYSKN KELMDNMRNV ALEEQAVEAV
 421 LAKAKVTEKE TTFNELMNQQ ASSGLEVLFQ GPGSEFELRR QACGRTRAPP PPPLRSGC
```

B.1.2.10 pFloat-NusA

```
   1 ATGCACCATC ATCATCATCA TTCTTCTGGT AACAAAGAAA TTTTGGCTGT AGTTGAAGCC
  61 GTATCCAATG AAAAGGCGCT ACCTCGCGAG AAGATTTTCG AAGCATTGGA AAGCGCGCTG
 121 GCGACAGCAA CAAAGAAAAA ATATGAACAA GAGATCGACG TCCGCGTACA GATCGATCGC
 181 AAAAGCGGTG ATTTTGACAC TTTCCGTCGC TGGTTAGTTG TTGATGAAGT CACCCAGCCG
 241 ACCAAGGAAA TCACCCTTGA AGCCGCACGT TATGAAGATG AAAGCCTGAA CCTGGGCGAT
 301 TACGTTGAAG ATCAGATTGA GTCTGTTACC TTTGACCGTA TCACTACCCA GACGGCAAAA
 361 CAGGTTATCG TGCAGAAAGT GCGTGAAGCC GAACGTGCGA TGGTGGTTGA TCAGTTCCGT
 421 GAACACGAAG GTGAAATCAT CACCGGCGTG GTGAAAAAAG TAAACCGCGA CAACATCTCT
 481 CTGGATCTGG GCAACAACGC TGAAGCCGTG ATCCTGCGCG AAGATATGCT GCCGCGTGAA
```

```
 541 AACTTCCGCC CTGGCGACCG CGTTCGTGGC GTGCTCTATT CCGTTCGCCC GGAAGCGCGT
 601 GGCGCGCAAC TGTTCGTCAC TCGTTCCAAG CCGGAAATGC TGATCGAACT GTTCCGTATT
 661 GAAGTGCCAG AAATCGGCGA AGAAGTGATT GAAATTAAAG CAGCGGCTCG CGATCCGGGT
 721 TCTCGTGCGA AAATCGCGGT GAAAACCAAC GATAAACGTA TCGATCCGGT AGGTGCTTGC
 781 GTAGGTATGC GTGGCGCGCG TGTTCAGGCG GTGTCTACTG AACTGGGTGG CGAGCGTATC
 841 GATATCGTCC TGTGGGATGA TAACCCGGCG CAGTTCGTGA TTAACGCAAT GGCACCGGCA
 901 GACGTTGCTT CTATCGTGGT GGATGAAGAT AAACACACCA TGGACATCGC CGTTGAAGCC
 961 GGTAATCTGG CGCAGGCGAT TGGCCGTAAC GGTCAGAACG TGCGTCTGGC TTCGCAACTG
1021 AGCGGTTGGG AACTCAACGT GATGACCGTT GACGACCTGC AAGCTAAGCA TCAGGCGGAA
1081 GCGCACGCAG CGATCGACAC CTTCACCAAA TATCTCGACA TCGACGAAGA CTTCGCGACT
1141 GTTCTGGTAG AAGAAGGCTT CTCGACGCTG GAAGAATTGG CCTATGTGCC GATGAAAGAG
1201 CTGTTGGAAA TCGAAGGCCT TGATGAGCCG ACCGTTGAAG CACTGCGCGA GCGTGCTAAA
1261 AATGCACTGG CCACCATTGC ACAGGCCCAG GAAGAAAGCC TCGGTGATAA CAAACCGGCT
1321 GACGATCTGC TGAACCTTGA AGGGGTAGAT CGTGATTTGG CATTCAAACT GGCCGCCCGT
1381 GGCGTTTGTA CGCTGGAAGA TCTCGCCGAA CAGGGCATTG ATGATCTGGC TGATATCGAA
1441 GGGTTGACCG ACGAAAAAGC CGGAGCACTG ATTATGGCTG CCCGTAATAT TTGCTGGTTC
1501 GGTGACGAAG CGTCTTCTGG TCTGGAAGTT CTGTTCCAGG GGCCCGGATC CGAATTCGAG
1561 CTCCGTCGAC AAGCTTGCGG CCGCACTCGA GCACCACCAC CACCACCACT GAGATCCGGC
1621 TGCTAA
```

```
   1 MHHHHHHSSG NKEILAVVEA VSNEKALPRE KIFEALESAL ATATKKKYEQ EIDVRVQIDR
  61 KSGDFDTFRR WLVVDEVTQP TKEITLEAAR YEDESLNLGD YVEDQIESVT FDRITTQTAK
 121 QVIVQKVREA ERAMVVDQFR EHEGEIITGV VKKVNRDNIS LDLGNNAEAV ILREDMLPRE
 181 NFRPGDRVRG VLYSVRPEAR GAQLFVTRSK PEMLIELFRI EVPEIGEEVI EIKAAARDPG
 241 SRAKIAVKTN DKRIDPVGAC VGMRGARVQA VSTELGGERI DIVLWDDNPA QFVINAMAPA
 301 DVASIVVDED KHTMDIAVEA GNLAQAIGRN GQNVRLASQL SGWELNVMTV DDLQAKHQAE
 361 AHAAIDTFTK YLDIDEDFAT VLVEEGFSTL EELAYVPMKE LLEIEGLDEP TVEALRERAK
 421 NALATIAQAQ EESLGDNKPA DDLLNLEGVD RDLAFKLAAR GVCTLEDLAE QGIDDLADIE
 481 GLTDEKAGAL IMAARNICWF GDEASSGLEV LFQGPGSEFE LRRQACGRTR APPPPPLRSG
 541 C
```

B.1.2.11 pFloat-MsyB

```
  1 ATGCACCATC ATCATCATCA TTCTTCTGGT ACCATGTACG CAACGCTTGA AGAAGCCATT
 61 GACGCTGCAC GCGAAGAATT TCTTGCAGAC AACCCCGGCA TCGACGCCGA AGATGCGAAT
121 GTGCAACAGT TCAATGCCCA AAAATACGTT TTGCAGGACG GCGACATCAT GTGGCAAGTT
181 GAGTTTTTTG CCGACGAAGG GGAAGAAGGT GAATGTTTAC CTATGCTTAG CGGTGAAGCC
241 GCGCAAAGTG TTTTTGATGG CGACTATGAT GAGATAGAGA TACGCCAGGA GTGGCAGGAA
301 GAGAATACAT TACATGAATG GGACGAGGGG GAATTTCAGC TTGAGCCACC GCTGGATACC
361 GAGGAAGGAC GCGCAGCAGC TGATGAGTGG GATGAACGTT CTTCTGGTCT GGAAGTTCTG
421 TTCCAGGGGC CCGGATCCGA ATTCGAGCTC CGTCGACAAG CTTGCGGCCG CACTCGAGCA
481 CCACCACCAC CACCACTGAG ATCCGGCTGC TAA

  1 MHHHHHHSSG TMYATLEEAI DAAREEFLAD NPGIDAEDAN VQQFNAQKYV LQDGDIMWQV
 61 EFFADEGEEG ECLPMLSGEA AQSVFDGDYD EIEIRQEWQE ENTLHEWDEG EFQLEPPLDT
121 EEGRAAADEW DERSSGLEVL FQGPGSEFEL RRQACGRTRA PPPPPLRSGC
```

B.1.3 pFloat fluorescence fusions - short sequences

B.1.3.1 pFloat-YFP (monoVenus)

```
  1 ATGCACCATC ATCATCATCA TGAAAACCTG TACTTTCAGG GCGGTTCTAT GGTGAGCAAG
 61 GGCGAGGAGC TGTTCACCGG GGTGGTGCCC ATCCTGGTCG AGCTGGACGG CGACGTAAAC
121 GGCCACAAGT TCAGCGTGTC CGGCGAGGGC GAGGGCGATG CCACCTACGG CAAGCTGACC
181 CTGAAGCTGA TCTGCACCAC CGGCAAGCTG CCCGTGCCCT GGCCCACCCT CGTGACCACC
241 CTGGGCTACG GCCTGCAGTG CTTCGCCCGC TACCCCGACC ACATGAAGCA GCACGACTTC
301 TTCAAGTCCG CCATGCCCGA AGGCTACGTC CAGGAGCGCA CCATCTTCTT CAAGGACGAC
361 GGCAACTACA AGACCCGCGC CGAGGTGAAG TTCGAGGGCG ACACCCTGGT GAACCGCATC
421 GAGCTGAAGG GCATCGACTT CAAGGAGGAC GGCAACATCC TGGGGCACAA GCTGGAGTAC
481 AACTACAACA GCCACAACGT CTATATCACC GCCGACAAGC AGAAGAACGG CATCAAGGCC
541 AACTTCAAGA TCCGCCACAA CATCGAGGAC GGCGGCGTGC AGCTCGCCGA CCACTACCAG
601 CAGAACACCC CCATCGGCGA CGGCCCCGTG CTGCTGCCCG ACAACCACTA CCTGAGCTAC
661 CAGTCCAAGC TGAGCAAAGA CCCCAACGAG AAGCGCGATC ACATGGTCCT GCTGGAGTTC
721 GTGACCGCCG CCGGGATCAC TCTCGGCATG GACGAGCTGT ACAAGTCTTC TGGTCTGGAA
```

```
781 GTTCTGTTCC AGGGGCCCGG ATCCGAATTC GAGCTCCGTC GACAAGCTTG CGGCCGCACT
841 CGAGCACCAC CACCACCACC ACTGAGATCC GGCTGCTAA

  1 MHHHHHHENL YFQGGSMVSK GEELFTGVVP ILVELDGDVN GHKFSVSGEG EGDATYGKLT
 61 LKLICTTGKL PVPWPTLVTT LGYGLQCFAR YPDHMKQHDF FKSAMPEGYV QERTIFFKDD
121 GNYKTRAEVK FEGDTLVNRI ELKGIDFKED GNILGHKLEY NYNSHNVYIT ADKQKNGIKA
181 NFKIRHNIED GGVQLADHYQ QNTPIGDGPV LLPDNHYLSY QSKLSKDPNE KRDHMVLLEF
241 VTAAGITLGM DELYKSSGLE VLFQGPGSEF ELRRQACGRT RAPPPPPLRS GC
```

B.1.3.2 pFloat-eGFP

```
  1 ATGCACCATC ATCATCATCA TGAAAACCTG TACTTTCAGG GCGGTTCTAT GGTGAGCAAG
 61 GGCGAGGAGC TGTTCACCGG GGTGGTGCCC ATCCTGGTCG AGCTGGACGG CGACGTAAAC
121 GGCCACAAGT TCAGCGTGTC CGGCGAGGGC GAGGGCGATG CCACCTACGG CAAGCTGACC
181 CTGAAGTTCA TCTGCACCAC CGGCAAGCTG CCCGTGCCCT GGCCCACCCT CGTGACCACC
241 CTGACCTACG GCGTGCAGTG CTTCAGCCGC TACCCCGACC ACATGAAGCA GCACGACTTC
301 TTCAAGTCCG CCATGCCCGA AGGCTACGTC CAGGAGCGCA CCATCTTCTT CAAGGACGAC
361 GGCAACTACA AGACCCGCGC CGAGGTGAAG TTCGAGGGCG ACACCCTGGT GAACCGCATC
421 GAGCTGAAGG GCATCGACTT CAAGGAGGAC GGCAACATCC TGGGGCACAA GCTGGAGTAC
481 AACTACAACA GCCACAACGT CTATATCATG GCCGACAAGC AGAAGAACGG CATCAAGGTG
541 AACTTCAAGA TCCGCCACAA CATCGAGGAC GGCAGCGTGC AGCTCGCCGA CCACTACCAG
601 CAGAACACCC CCATCGGCGA CGGCCCCGTG CTGCTGCCCG ACAACCACTA CCTGAGCACC
661 CAGTCCGCCC TGAGCAAAGA CCCCAACGAG AAGCGCGATC ACATGGTCCT GCTGGAGTTC
721 GTGACCGCCG CCGGGATCAC TCTCGGCATG GACGAGCTCT ACAAGTCTTC TGGTCTGGAA
781 GTTCTGTTCC AGGGGCCCGG ATCCGAATTC GAGCTCCGTC GACAAGCTTG CGGCCGCACT
841 CGAGCACCAC CACCACCACC ACTGAGATCC GGCTGCTAA

  1 MHHHHHHENL YFQGGSMVSK GEELFTGVVP ILVELDGDVN GHKFSVSGEG EGDATYGKLT
 61 LKFICTTGKL PVPWPTLVTT LTYGVQCFSR YPDHMKQHDF FKSAMPEGYV QERTIFFKDD
121 GNYKTRAEVK FEGDTLVNRI ELKGIDFKED GNILGHKLEY NYNSHNVYIM ADKQKNGIKV
181 NFKIRHNIED GSVQLADHYQ QNTPIGDGPV LLPDNHYLST QSALSKDPNE KRDHMVLLEF
```

```
241 VTAAGITLGM DELYKSSGLE VLFQGPGSEF ELRRQACGRT RAPPPPPLRS GC
```

B.1.4 pFloat.fuse - short sequences

B.1.4.1 pFloat-MBP.fuse

```
   1 ATGCACCATC ATCATCATCA TTCTTCTGGT CTGGAAGTTC TGTTCCAGGG GCCCAAAATC
  61 GAAGAAGGTA AACTGGTAAT CTGGATTAAC GGCGATAAAG GCTATAACGG TCTCGCTGAA
 121 GTCGGTAAGA AATTCGAGAA AGATACCGGA ATTAAAGTCA CCGTTGAGCA TCCGGATAAA
 181 CTGGAAGAGA AATTCCCACA GGTTGCGGCA ACTGGCGATG GCCCTGACAT TATCTTCTGG
 241 GCACACGACC GCTTTGGTGG CTACGCTCAA TCTGGCCTGT TGGCTGAAAT CACCCCGGAC
 301 AAAGCGTTCC AGGACAAGCT GTATCCGTTT ACCTGGGATG CCGTACGTTA CAACGGCAAG
 361 CTGATTGCTT ACCCGATCGC TGTTGAAGCG TTATCGCTGA TTTATAACAA AGATCTGCTG
 421 CCGAACCCGC CAAAAACCTG GGAAGAGATC CCGGCGCTGG ATAAAGAACT GAAAGCGAAA
 481 GGTAAGAGCG CGCTGATGTT CAACCTGCAA GAACCGTACT TCACCTGGCC GCTGATTGCT
 541 GCTGACGGGG GTTATGCGTT CAAGTATGAA AACGGCAAGT ACGACATTAA AGACGTGGGC
 601 GTGGATAACG CTGGCGCGAA AGCGGGTCTG ACCTTCCTGG TTGACCTGAT TAAAAACAAA
 661 CACATGAATG CAGACACCGA TTACTCCATC GCAGAAGCTG CCTTTAATAA AGGCGAAACA
 721 GCGATGACCA TCAACGGCCC GTGGGCATGG TCCAACATCG ACACCAGCAA AGTGAATTAT
 781 GGTGTAACGG TACTGCCGAC CTTCAAGGGT CAACCATCCA AACCGTTCGT TGGCGTGCTG
 841 AGCGCAGGTA TTAACGCCGC CAGTCCGAAC AAAGAGCTGG CAAAAGAGTT CCTCGAAAAC
 901 TATCTGCTGA CTGATGAAGG TCTGGAAGCG GTTAATAAAG ACAAACCGCT GGGTGCCGTA
 961 GCGCTGAAGT CTTACGAGGA AGAGTTGGCG AAAGATCCAC GTATTGCCGC CACTATGGAA
1021 AACGCCCAGA AAGGTGAAAT CATGCCGAAC ATCCCGCAGA TGTCCGCTTT CTGGTATGCC
1081 GTGCGTACTG CGGTGATCAA CGCCGCCAGC GGTCGTCAGA CTGTCGATGC AGCCCTGGCA
1141 GCCGCGCAGA CTGCAGCCGC AGGATCCGAA TTCGAGCTCC GTCGACAAGC TTGCGGCCGC
1201 ACTCGAGCAC CACCACCACC ACCACTGAGA TCCGGCTGCT AA

   1 MHHHHHHSSG LEVLFQGPKI EEGKLVIWIN GDKGYNGLAE VGKKFEKDTG IKVTVEHPDK
  61 LEEKFPQVAA TGDGPDIIFW AHDRFGGYAQ SGLLAEITPD KAFQDKLYPF TWDAVRYNGK
 121 LIAYPIAVEA LSLIYNKDLL PNPPKTWEEI PALDKELKAK GKSALMFNLQ EPYFTWPLIA
 181 ADGGYAFKYE NGKYDIKDVG VDNAGAKAGL TFLVDLIKNK HMNADTDYSI AEAAFNKGET
 241 AMTINGPWAW SNIDTSKVNY GVTVLPTFKG QPSKPFVGVL SAGINAASPN KELAKEFLEN
```

```
301 YLLTDEGLEA VNKDKPLGAV ALKSYEEELA KDPRIAATME NAQKGEIMPN IPQMSAFWYA
361 VRTAVINAAS GRQTVDAALA AAQTAAAGSE FELRRQACGR TRAPPPPPLR SGC
```

B.1.4.2 pFloat-SUMO.fuse

```
  1 ATGCACCATC ATCATCATCA TTCTTCTGGT CTGGAAGTTC TGTTCCAGGG GCCCAGCGAT
 61 AGCGAAGTGA ACCAGGAAGC GAAACCGGAA GTTAAACCGG AAGTGAAACC GGAAACCCAT
121 ATTAATCTGA AAGTTAGCGA CGGCAGCAGC GAAATCTTTT TTAAAATTAA AAAAACCACC
181 CCGCTGCGTC GCCTGATGGA AGCCTTTGCG AAACGTCAGG GTAAAGAAAT GGATAGCCTG
241 CGCTTTCTGT ATGACGGCAT CCGTATTCAG GCCGATCAGA CCCCGGAAGA CCTGGATATG
301 GAAGACAACG ATATTATTGA AGCGCATCGC GAACAGATCG GTGGTGGATC CGAATTCGAG
361 CTCCGTCGAC AAGCTTGCGG CCGCACTCGA GCACCACCAC CACCACCACT GAGATCCGGC
421 TGCTAA

  1 MHHHHHHSSG LEVLFQGPSD SEVNQEAKPE VKPEVKPETH INLKVSDGSS EIFFKIKKTT
 61 PLRRLMEAFA KRQGKEMDSL RFLYDGIRIQ ADQTPEDLDM EDNDIIEAHR EQIGGGSEFE
121 LRRQACGRTR APPPPPLRSG C
```

B.1.4.3 pFloat-TRX.fuse

```
  1 ATGCACCATC ATCATCATCA TTCTTCTGGT CTGGAAGTTC TGTTCCAGGG GCCCAGCGAT
 61 AAAATTATTC ACCTGACTGA CGACAGTTTT GACACGGATG TACTCAAAGC GGACGGGGCG
121 ATCCTCGTCG ATTTCTGGGC AGAGTGGTGC GGTCCGTGCA AAATGATCGC CCCGATTCTG
181 GATGAAATCG CTGACGAATA TCAGGGCAAA CTGACCGTTG CAAAACTGAA CATCGATCAA
241 AACCCTGGCA CTGCGCCGAA ATATGGCATC CGTGGTATCC CGACTCTGCT GCTGTTCAAA
301 AACGGTGAAG TGGCGGCAAC CAAAGTGGGT GCACTGTCTA AAGGTCAGTT GAAAGAGTTC
361 CTCGACGCTA ACCTGGCCGG TAGCGCGGGA TCCGAATTCG AGCTCCGTCG ACAAGCTTGC
421 GGCCGCACTC GAGCACCACC ACCACCACCA CTGAGATCCG GCTGCTAA

  1 MHHHHHHSSG LEVLFQGPSD KIIHLTDDSF DTDVLKADGA ILVDFWAEWC GPCKMIAPIL
 61 DEIADEYQGK LTVAKLNIDQ NPGTAPKYGI RGIPTLLLFK NGEVAATKVG ALSKGQLKEF
```

```
121 LDANLAGSAG SEFELRRQAC GRTRAPPPPP LRSGC
```

B.2 *sas-5* recoded for RNAi resistance

```
   1 ATGAACAACT ATGATGATCT GCCGTGTAGT ATCTATTTTA AGAAGCCAAC TGgtgagttg
  61 ttttcgatat aaaagtgtta aaattgccaa tttagTCCAA GAATTCGTTG ACCAGCCACG
 121 TGTTTTTGAG GACAGTGAGG TTCCGGCTTT CCAGGAGGTG TTGCAATTGC CGCAGGAGCG
 181 TACTAAGCCG CCGGTTCCGA GTACCCAACC GATCGTTGCT GCCGTTGAGG TTGCTAAAAA
 241 GAAGTCTTGC TTGTCTGCCC CAAAGCCGCG TAAGGAACCG CCGTCTCACC CGGCTTTGCG
 301 TCAGAAGACT GTGGCTTTCG GTAAGACTGT GAATGTGTCT CAGACCGTCG AGgtaattaa
 361 attaaaattg ttccatggca taattccact aaattatctt tcagGGAACT TCCCGTAACT
 421 CTAAGAAGGT TTTGGCCTCT ACTATGAGTG CTCAGAACAC CACTACCACC GAGGAGCAAG
 481 CTGCCGAGAA TTGGCGTGAT GCCATGAAGA CTGAGTTGCA GACCATTCGT ACTGAAATCC
 541 AGGAGGAGAC CGCTCGCCGT CAGGAGGAAT TGAACGCCCA GAACTTGGTG AAGATGCAGG
 601 AGCTGATGTC TAATTTTTTT CAGAgtatgt ttgaatatat attatcccat aaattcaata
 661 aataataatt tttcagAGAT TACCATCCCG AAGCAACAGG CTATTGAGCC GGTGGAGAAG
 721 GATAAAGAGA ACTTCCACGA ATCTCCGCGC CAGTCTCGTC AACAAAAACC GGCCAGTAAA
 781 gtgagaattc agataaaaaa taaatcaaaa agtactccgt tttaagATCG CCTCTGCTCG
 841 TGAGGTGATT AAGCGTGATG GAGTGATTCC GCCGGAGGCT CTGACTATCA TCGAACAGCG
 901 TTTGCGTTCT GATCCGATGT TCCGTCAGCA GATCGACAAC GTGCTGGCAG ACgtgagtag
 961 tgtatttaaa tccgatagaa atagtttgtt ttcagGCTGA GTGCGACGCC AATCGTGCCG
1021 CTTACAGTCC GCCGCCGCCG ATGAGTGAGG TGCGTTATGG TAGTGGAGTT AACCCGGCCC
1081 TGATGCGTGA GACTTTGACC GTTGAGCGTT CTATCCGTTA CGATAATGGT TTGGCCAGTA
1141 TCGACAGTCG TCAATGGACT AACGAGCGTC GCGACAATCG TGgtaattta atttttcaat
1201 attcaaacac agaattcgtc cgctagaatt gtataaatat aataattcac gcattccaaa
1261 atgtatcaaa taattcatta aaaatgtttg aaagttgata accattcgta gtgaaaacac
1321 tttatactac gaaatgattt taagatgaaa tcctcaaatc agttgaaaaa tttgaattaa
1381 aagtcaaaaa tgagaagtct cttcactatt caaaaaaacc ttttctagCT CCGGATTCTT
1441 ATCGCACTTA TGAGCCAGAC CAGCCGTGCC ATTCTTTGTA CCAGAAAGGA CAGTCTATCT
1501 CTTACTACCC ATCTGAGGCT GCCGGTAAGA CCACTGCCCG TAATAACCGT ACCGGTTATT
1561 ATGTTGAGGA CTCTTCTGAC CACGAGGAAG ACGTGGTTGT TAACAAGCGT GGTCAAAATT
1621 ACCATGAACA GGCTGTTCCG GAAACCCCGG CCGAGCGTGA GCGTCGCATC CGCGAGAAGT
```

```
1681 ATGCCCGTCG TAAGTAA
```

Exons (capital letters) - highlighted with alternating yellow and blue; **introns (small letters)** - no highlight.

B.3 Expression constructs

B.3.1 SAS-5 proteins

B.3.1.1 MsyB-SAS-5$_{FL}$

```
  1 MHHHHHHSSG TMYATLEEAI DAAREEFLAD NPGIDAEDAN VQQFNAQKYV LQDGDIMWQV
 61 EFFADEGEEG ECLPMLSGEA AQSVFDGDYD EIEIRQEWQE ENTLHEWDEG EFQLEPPLDT
121 EEGRAAADEW DERSSGLEVL FQGPNNYDDL PCSIYFKKPT VQEFVDQPRV FEDSEVPAFQ
181 EVLQLPQERT KPPVPSTQPI VAAVEVAKKK SCLSAPKPRK EPPSHPALRQ KTVAFGKTVN
241 VSQTVEGTSR NSKKVLASTM SAQNTTTTEE QAAENWRDAM KTELQTIRTE IQEETARRQE
301 ELNAQNLVKM QELMSNFFQK ITIPKQQAIE PVEKDKENFH ESPRQSRQQK PASKIASARE
361 VIKRDGVIPP EALTIIEQRL RSDPMFRQQI DNVLADAECD ANRAAYSPPP PMSEVRYGSG
421 VNPALMRETL TVERSIRYDN GLASIDSRQW TNERRDNRAP DSYRTYEPDQ PCHSLYQKGQ
481 SISYYPSEAA GKTTARNNRT GYYVEDSSDH EEDVVVNKRG QNYHEQAVPE TPAERERRIR
541 EKYARRK
```

N-terminal His$_6$-tag | flexible linkers | solubility tag | HRV-3C cleavage site
SAS-5 - no highlight, *sticky strand* or FLEX replacement

B.3.1.2 MsyB-SAS-5$_{\triangle(282\text{-}295)}$

```
  1 MHHHHHHSSG TMYATLEEAI DAAREEFLAD NPGIDAEDAN VQQFNAQKYV LQDGDIMWQV
 61 EFFADEGEEG ECLPMLSGEA AQSVFDGDYD EIEIRQEWQE ENTLHEWDEG EFQLEPPLDT
121 EEGRAAADEW DERSSGLEVL FQGPNNYDDL PCSIYFKKPT VQEFVDQPRV FEDSEVPAFQ
181 EVLQLPQERT KPPVPSTQPI VAAVEVAKKK SCLSAPKPRK EPPSHPALRQ KTVAFGKTVN
241 VSQTVEGTSR NSKKVLASTM SAQNTTTTEE QAAENWRDAM KTELQTIRTE IQEETARRQE
301 ELNAQNLVKM QELMSNFFQK ITIPKQQAIE PVEKDKENFH ESPRQSRQQK PASKIASARE
361 VIKRDGVIPP EALTIIEQRL RSDPMFRQQI DNVLADAECD ANRAAYSPPP PMSEVRYGSG
```

421 VNPADNGLAS IDSRQWTNER RDNRAPDSYR TYEPDQPCHS LYQKGQSISY YPSEAAGKTT
481 ARNNRTGYYV EDSSDHEEDV VVNKRGQNYH EQAVPETPAE RERRIREKYA RRK

B.3.1.3 MsyB-SAS-5$_{\mathbf{FLEX}}$

1 MHHHHHHSSG TMYATLEEAI DAAREEFLAD NPGIDAEDAN VQQFNAQKYV LQDGDIMWQV
61 EFFADEGEEG ECLPMLSGEA AQSVFDGDYD EIEIRQEWQE ENTLHEWDEG EFQLEPPLDT
121 EEGRAAADEW DERSSGLEVL FQGPNNYDDL PCSIYFKKPT VQEFVDQPRV FEDSEVPAFQ
181 EVLQLPQERT KPPVPSTQPI VAAVEVAKKK SCLSAPKPRK EPPSHPALRQ KTVAFGKTVN
241 VSQTVEGTSR NSKKVLASTM SAQNTTTTEE QAAENWRDAM KTELQTIRTE IQEETARRQE
301 ELNAQNLVKM QELMSNFFQK ITIPKQQAIE PVEKDKENFH ESPRQSRQQK PASKIASARE
361 VIKRDGVIPP EALTIIEQRL RSDPMFRQQI DNVLADAECD ANRAAYSPPP PMSEVRYGSG
421 VNPASGAAGS SGAAGSSGDN GLASIDSRQW TNERRDNRAP DSYRTYEPDQ PCHSLYQKGQ
481 SISYYPSEAA GKTTARNNRT GYYVEDSSDH EEDVVVNKRG QNYHEQAVPE TPAERERRIR
541 EKYARRK

B.3.1.4 SUMO-SAS-5$_{\mathbf{CC\text{-}L}}$

1 MHHHHHHSSG SDSEVNQEAK PEVKPEVKPE THINLKVSDG SSEIFFKIKK TTPLRRLMEA
61 FAKRQGKEMD SLRFLYDGIR IQADQTPEDL DMEDNDIIEA HREQISSGLE VLFQGPTVEG
121 TSRNSKKVLA STMSAQNTTT TEEQAAENWR DAMKTELQTI RTEIQEETAR RQEELNAQNL
181 VKMQELMSNF FQKITIPKQQ AIEPVEKDKE NFHESPRQSR QQ

B.3.1.5 *C. briggsae* SUMO-SAS-5$_{\mathbf{(96\text{-}199)}}$

1 MHHHHHHSSG SDSEVNQEAK PEVKPEVKPE THINLKVSDG SSEIFFKIKK TTPLRRLMEA
61 FAKRQGKEMD SLRFLYDGIR IQADQTPEDL DMEDNDIIEA HREQISSGLE VLFQGPTVEG
121 TSRHTKKAIA SPSQNHRMTE ENQEENWRDV MKNEFEVMRK EMQEEATKKQ EELNAQNLNK
181 MQEMMSQFFQ QITVAKPSAE ESQDREKENW YEQSRHARQQ

B.3.1.6 *C. brenneri* SUMO-SAS-5$_{(1\text{-}82)}$

```
  1 MHHHHHHSSG SDSEVNQEAK PEVKPEVKPE THINLKVSDG SSEIFFKIKK TTPLRRLMEA
 61 FAKRQGKEMD SLRFLYDGIR IQADQTPEDL DMEDNDIIEA HREQISSGLE VLFQGPMADE
121 KQEEDWRDVM KNELDTMRRE MQEETAKRQE ELNSQNLNKM QEMMTQFFQQ ITVAKKPSIE
181 SGEREKENHS EARQLRQH
```

B.3.1.7 *C. remanei* SUMO-SAS-5$_{(98\text{-}206)}$

```
  1 MHHHHHHSSG SDSEVNQEAK PEVKPEVKPE THINLKVSDG SSEIFFKIKK TTPLRRLMEA
 61 FAKRQGKEMD SLRFLYDGIR IQADQTPEDL DMEDNDIIEA HREQISSGLE VLFQGPTVEG
121 TSRHSKKALA SPSSTSSPVQ NQKMTEENQE EDWRDVMKNE INVMRKEMQE ETTKRQEELN
181 AQNLSKMQEM MSQFFQQITV AKPATDNTHR EKENLYEPTR QTRQQ
```

B.3.1.8 *C. sinica* SUMO-SAS-5$_{(94\text{-}204)}$

```
  1 MHHHHHHSSG SDSEVNQEAK PEVKPEVKPE THINLKVSDG SSEIFFKIKK TTPLRRLMEA
 61 FAKRQGKEMD SLRFLYDGIR IQADQTPEDL DMEDNDIIEA HREQISSGLE VLFQGPTVEG
121 TSRHSKKALA PSSTASSPSQ NSKMAEEQQE DNWRDVMKNE FEVMRKEMQE EATKKQEELN
181 AQNLNKMQEM MSQFFQQITV SKQPAGGESL DNNKENRYER SQPLRQQ
```

B.3.1.9 *C. tropicalis* SUMO-SAS-5$_{(94\text{-}199)}$

```
  1 MHHHHHHSSG SDSEVNQEAK PEVKPEVKPE THINLKVSDG SSEIFFKIKK TTPLRRLMEA
 61 FAKRQGKEMD SLRFLYDGIR IQADQTPEDL DMEDNDIIEA HREQISSGLE VLFQGPTVEG
121 TSRNSKKASS ATSSPIQNQK MAGENHEENW KDVIKNEMST MRKEIQEETT KRQEELNNQN
181 LDKMKEMMAQ FFQQITVSNK PAIVDSERGK ENLYEPRQSK QQ
```

B.3.2 HRV-3C$_{(1\text{-}181)}$

```
  1 MHHHHHHSSG GPNTEFALSL LRKNIMTITT SKGEFTGLGI HDRVCVIPTH AQPGDDVLVN
 61 GQKIRVKDKY KLVDPENINL ELTVLTLDRN EKFRDIRGFI SEDLEGVDAT LVVHSNNFTN
121 TILEVGPVTM AGLINLSSTP TNRMIRYDYA TKTGQCGGVL CATGKIFGIH VGGNGRQGFS
181 AQLKKQYFVE KQ
```

N-terminal His$_6$-tag flexible linkers (SSG) HRV-3C protease - no highlight

Bibliography

[1] Abraham, M. J., and Gready, J. E. Optimization of parameters for molecular dynamics simulation using smooth particle-mesh Ewald in GROMACS 4.5. *Journal of Computational Chemistry 32*, 9 (July 2011), 2031–2040.

[2] Abrahams, J. P., and Leslie, A. G. Methods used in the structure determination of bovine mitochondrial F_1 ATPase. *Acta Crystallographica Section D: Biological Crystallography 52*, 1 (Jan. 1996), 30–42.

[3] Adams, P. D., Afonine, P. V., Bunkóczi, 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., and Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography 66*, 2 (Feb. 2010), 213–221.

[4] Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D: Biological Crystallography 68*, 4 (Apr. 2012), 352–367.

[5] Al Jord, A., Lemaître, A.-I., Delgehyr, N., Faucourt, M., Spassky, N., and Meunier, A. Centriole amplification by mother and daughter centrioles differs in multiciliated cells. *Nature 516*, 7529 (Dec. 2014), 104–107.

[6] Amos, L. A. The tektin family of microtubule-stabilizing proteins. *Genome Biology 9*, 7 (Jan. 2008), 229.

[7] Andersen, J. S., Wilkinson, C. J., Mayor, T., Mortensen, P., Nigg, E. A., and Mann, M. Proteomic characterization of the human centrosome by protein correlation profiling. *Nature 426*, 6966 (Dec. 2003), 570–574.

[8] Anderson, R. G. The three-dimensional structure of the basal body from the rhesus monkey oviduct. *The Journal of Cell Biology 54*, 2 (Aug. 1972), 246–265.

[9] Andreeva, A., Howorth, D., Chothia, C., Kulesha, E., and Murzin, A. G. SCOP2 prototype: a new approach to protein structure mining. *Nucleic Acids Research 42*, D1 (Jan. 2014), D310–D314.

[10] Andres, S. N., Modesti, M., Tsai, C. J., Chu, G., and Junop, M. S. Crystal structure of human XLF: a twist in nonhomologous DNA end-joining. *Molecular Cell 28*, 6 (Dec. 2007), 1093–1101.

[11] Andres, S. N., Vergnes, A., Ristic, D., Wyman, C., Modesti, M., and Junop, M. A human XRCC4-XLF complex bridges DNA. *Nucleic Acids Research 40*, 4 (Feb. 2012), 1868–1878.

[12] Arquint, C., Gabryjonczyk, A.-M., Imseng, S., Böhm, R., Sauer, E., Hiller, S., Nigg, E. A., and Maier, T. STIL binding to Polo-box 3 of PLK4 regulates centriole duplication. *eLife 4* (July 2015), e07888.

[13] Arquint, C., and Nigg, E. A. STIL microcephaly mutations interfere with APC/C-mediated degradation and cause centriole amplification. *Current Biology 24*, 4 (Feb. 2014), 351–360.

[14] Arquint, C., Sonnen, K. F., Stierhof, Y.-D., and Nigg, E. A. Cell-cycle-regulated expression of STIL controls centriole number in human cells. *Journal of Cell Science 125*, 5 (Mar. 2012), 1342–1352.

[15] Azimzadeh, J. Exploring the evolutionary history of centrosomes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 369*, 1650 (Sept. 2014), 20130453.

[16] Azimzadeh, J., Hergert, P., Delouvée, A., Euteneuer, U., Formstecher, E., Khodjakov, A., and Bornens, M. hPOC5 is a centrin-binding protein required for assembly of full-length centrioles. *The Journal of Cell Biology 185*, 1 (Apr. 2009), 101–114.

[17] Azimzadeh, J., and Marshall, W. F. Building the centriole. *Current Biology 20*, 18 (Sept. 2010), R816–R825.

[18] Azimzadeh, J., Wong, M. L., Downhour, D. M., Sánchez Alvarado, A., and Marshall, W. F. Centrosome loss in the evolution of planarians. *Science 335*, 6067 (Jan. 2012), 461–463.

[19] Babu, M. M., Kriwacki, R. W., and Pappu, R. V. Versatility from protein disorder. *Science 337*, 6101 (Sept. 2012), 1460–1461.

[20] Baldwin, A. J., Lioe, H., Hilton, G. R., Baker, L. A., Rubinstein, J. L., Kay, L. E., and Benesch, J. L. P. The polydispersity of αB-crystallin is rationalized by an interconverting polyhedral architecture. *Structure 19*, 12 (Dec. 2011), 1855–1863.

[21] Basto, R., Lau, J., Vinogradova, T., Gardiol, A., Woods, C. G., Khodjakov, A., and Raff, J. W. Flies without centrioles. *Cell 125*, 7 (June 2006), 1375–1386.

[22] Bellanger, J. M., and Gönczy, P. TAC-1 and ZYG-9 form a complex that promotes microtubule assembly in *C. elegans* embryos. *Current Biology 13*, 17 (Sept. 2003), 1488–1498.

[23] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Di-Nola, A., and Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics 81*, 8 (1984), 3684.

[24] Bernard, P., and Couturier, M. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *Journal of Molecular Biology 226*, 3 (Aug. 1992), 735–745.

[25] Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D. I., and Owens, R. J. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Research 35*, 6 (Jan. 2007), e45.

[26] Bettencourt-Dias, M., and Glover, D. Centrosome biogenesis and function: centrosomics brings new understanding. *Nature Reviews Molecular Cell Biology 8*, 6 (June 2007), 451–463.

[27] Bettencourt-Dias, M., Rodrigues-Martins, A., Carpenter, L., Riparbelli, M., Lehmann, L., Gatt, M. K., Carmo, N., Balloux, F., Callaini, G., and Glover, D. M. SAK/PLK4 is required for centriole duplication and flagella development. *Current Biology 15*, 24 (Dec. 2005), 2199–2207.

[28] Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., and Schwede, T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research 42*, W1 (July 2014), W252–W258.

[29] Bibby, J., Keegan, R. M., Mayans, O., Winn, M. D., and Rigden, D. J. AMPLE: a cluster-and-truncate approach to solve the crystal structures of small proteins using rapidly computed *ab initio* models. *Acta Crystallographica Section D: Biological Crystallography 68*, 12 (Dec. 2012), 1622–1631.

[30] Bird, L. E. High throughput construction and small scale expression screening of multi-tag vectors in *Escherichia coli*. *Methods 55*, 1 (Sept. 2011), 29–37.

[31] Blachon, S., Gopalakrishnan, J., Omori, Y., Polyanovsky, A., Church, A., Nicastro, D., Malicki, J., and Avidor-Reiss, T. *Drosophila asterless* and vertebrate Cep152 Are orthologs essential for centriole duplication. *Genetics 180*, 4 (Dec. 2008), 2081–2094.

[32] Block, H., Maertens, B., Spriestersbach, A., Brinker, N., Kubicek, J., Fabis, R., Labahn, J., and Schäfer, F. Immobilized-metal affinity

chromatography (IMAC): a review. *Methods in Enzymology 463* (Jan. 2009), 439–473.

[33] Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., Boyer, H. W., Crosa, J. H., and Falkow, S. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene 2*, 2 (Jan. 1977), 95–113.

[34] Bornens, M. Organelle positioning and cell polarity. *Nature Reviews Molecular Cell Biology 9*, 11 (Nov. 2008), 874–886.

[35] Bornens, M. The centrosome in cells and organisms. *Science 335*, 6067 (Jan. 2012), 422–426.

[36] Bornens, M., and Azimzadeh, J. Origin and evolution of the centrosome. In *Eukaryotic Membranes and Cytoskeleton: Origins and Evolution*, G. Jékely, Ed., vol. 607 of *Advances in Experimental Medicine and Biology*. Springer, Jan. 2007, pp. 119–129.

[37] Bornhorst, J. A., and Falke, J. J. Purification of proteins using polyhistidine affinity tags. In *Applications of Chimeric Genes and Hybrid Proteins Part A: Gene Expression and Protein Purification*, J. Thorner, S. D. Emr, and J. N. Abelson, Eds., vol. 326 of *Methods in Enzymology*. Academic Press, Jan. 2000, pp. 245–254.

[38] Boveri, T. Über die Befruchtung der Eier von *Ascaris megalocephala*. *Sitzungsberichte der Gesellschaft für Morphologie und Physiologie in München 3* (1887), 71–80.

[39] Boveri, T. Zellen-Studien II. Die Befruchtung und Teilung des Eies von *Ascaris megalocephala*. *Jenaische Zeitschrift für Naturwissenschaft 22* (1888), 685–882.

[40] Boveri, T. Concerning the origin of malignant tumours by Theodor Boveri. Translated and annotated by Henry Harris. *Journal of Cell Science 121*, Supplement 1 (Jan. 2008), 1–84.

[41] Boyd, A. C., Archer, J. A., and Sherratt, D. J. Characterization of the ColE1 mobilization region and its protein products. *Molecular and General Genetics 217*, 2-3 (June 1989), 488–498.

[42] Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics 77*, 1 (May 1974), 71–94.

[43] Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallographica Section D: Biological Crystallography 59*, 11 (Nov. 2003), 2023–2030.

[44] Brooks, E. R., and Wallingford, J. B. Multiciliated cells. *Current Biology 24*, 19 (Oct. 2014), R973–R982.

[45] Bui, K. H., Yagi, T., Yamamoto, R., Kamiya, R., and Ishikawa, T. Polarity and asymmetry in the arrangement of dynein and related structures in the *Chlamydomonas* axoneme. *The Journal of Cell Biology 198*, 5 (Oct. 2012), 913–925.

[46] Bush, M. F., Hall, Z., Giles, K., Hoyes, J., Robinson, C. V., and Ruotolo, B. T. Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. *Analytical Chemistry 82*, 22 (Nov. 2010), 9557–9565.

[47] Cabrita, L. D., Dai, W., and Bottomley, S. P. A family of *E. coli* expression vectors for laboratory scale and high throughput soluble protein production. *BMC Biotechnology 6* (Jan. 2006), 12.

[48] Campaner, S., Kaldis, P., Izraeli, S., and Kirsch, I. R. Sil phosphorylation in a Pin1 binding domain affects the duration of the spindle checkpoint. *Molecular and Cellular Biology 25*, 15 (Aug. 2005), 6660–6672.

[49] Cantrell, S. A. Vectors for the expression of recombinant proteins in *E. coli*. In *E. coli plasmid vectors: methods and applications*, N. Casali and A. Preston,

Eds., vol. 235 of *Methods in Molecular Biology*. Humana Press, 2003, pp. 257–275.

[50] Carlson, E. D., Gan, R., Hodgman, C. E., and Jewett, M. C. Cell-free protein synthesis: Applications come of age. *Biotechnology Advances 30*, 5 (Sept. 2012), 1185–1194.

[51] Carvalho-Santos, Z., Azimzadeh, J., Pereira-Leal, J. B., and Bettencourt-Dias, M. Tracing the origins of centrioles, cilia, and flagella. *The Journal of Cell Biology 194*, 2 (July 2011), 165–175.

[52] Carvalho-Santos, Z., Machado, P., Alvarez-Martins, I., Gouveia, S. M., Jana, S. C., Duarte, P., Amado, T., Branco, P., Freitas, M. C., Silva, S. T., Antony, C., Bandeiras, T. M., and Bettencourt-Dias, M. BLD10/CEP135 is a microtubule-associated protein that controls the formation of the flagellum central microtubule pair. *Developmental Cell 23*, 2 (Aug. 2012), 412–424.

[53] Cavalier-Smith, T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *International Journal of Systematic and Evolutionary Microbiology 52*, Pt 2 (Mar. 2002), 297–354.

[54] Chandonia, J.-M., and Brenner, S. E. The impact of structural genomics: expectations and outcomes. *Science 311*, 5759 (Jan. 2006), 347–351.

[55] Chang, L., and Barford, D. Insights into the anaphase-promoting complex: a molecular machine that regulates mitosis. *Current Opinion in Structural Biology 29* (Dec. 2014), 1–9.

[56] Chang, P., and Stearns, T. δ-tubulin and ϵ-tubulin: two new human centrosomal tubulins reveal new aspects of centrosome structure and function. *Nature Cell Biology 2*, 1 (Jan. 2000), 30–35.

[57] Cheeseman, I. M., and Desai, A. A combined approach for the localization and tandem affinity purification of protein complexes from metazoans. *Science's STKE 2005*, 266 (Jan. 2005), pl1.

[58] Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography 66*, 1 (Jan. 2010), 12–21.

[59] Choi, S. I., Song, H. W., Moon, J. W., and Seong, B. L. Recombinant enterokinase light chain with affinity tag: expression from *Saccharomyces cerevisiae* and its utilities in fusion protein technology. *Biotechnology and Bioengineering 75*, 6 (Dec. 2001), 718–724.

[60] Cizmecioglu, O., Arnold, M., Bahtz, R., Settele, F., Ehret, L., Haselmann-Weiss, U., Antony, C., and Hoffmann, I. Cep152 acts as a scaffold for recruitment of Plk4 and CPAP to the centrosome. *The Journal of Cell Biology 191*, 4 (Nov. 2010), 731–739.

[61] Collingridge, P. W., and Kelly, S. MergeAlign: improving multiple sequence alignment performance by dynamic reconstruction of consensus multiple sequence alignments. *BMC Bioinformatics 13* (Jan. 2012), 117.

[62] 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., and Raff, J. W. A molecular mechanism of mitotic centrosome assembly in *Drosophila*. *eLife 3* (Jan. 2014), e03399.

[63] Conduit, P. T., Wainman, A., and Raff, J. W. Centrosome function and assembly in animal cells. *Nature Reviews Molecular Cell Biology 16*, 10 (Sept. 2015), 611–624.

[64] Cordingley, M. G., Callahan, P. L., Sardana, V. V., Garsky, V. M., and Colonno, R. J. Substrate requirements of human rhinovirus 3C protease for peptide cleavage *in vitro*. *The Journal of Biological Chemistry 265*, 16 (June 1990), 9062–9065.

[65] Cordingley, M. G., Register, R. B., Callahan, P. L., Garsky, V. M., and Colonno, R. J. Cleavage of small peptides *in vitro* by human rhinovirus

14 3C protease expressed in *Escherichia coli*. *Journal of Virology 63*, 12 (Dec. 1989), 5037–5045.

[66] Cormack, B. P., Valdivia, R. H., and Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene 173*, 1 (Jan. 1996), 33–38.

[67] Cormier, A., Clément, M.-J., Knossow, M., Lachkar, S., Savarin, P., Toma, F., Sobel, A., Gigant, B., and Curmi, P. A. The PN2-3 domain of centrosomal P4.1-associated protein implements a novel mechanism for tubulin sequestration. *The Journal of Biological Chemistry 284*, 11 (Mar. 2009), 6909–6917.

[68] Correa, A., Ortega, C., Obal, G., Alzari, P., Vincentelli, R., and Oppezzo, P. Generation of a vector suite for protein solubility screening. *Frontiers in Microbiology 5* (Jan. 2014), 67.

[69] Corsini, L., Hothorn, M., Scheffzek, K., Sattler, M., and Stier, G. Thioredoxin as a fusion tag for carrier-driven crystallization. *Protein Science 17*, 12 (Dec. 2008), 2070–2079.

[70] Cottee, M. A., Muschalik, N., Johnson, S., Leveson, J., Raff, J. W., and Lea, S. M. The homo-oligomerisation of both Sas-6 and Ana2 is required for efficient centriole assembly in flies. *eLife 4* (May 2015), e07236.

[71] Cottee, M. A., Muschalik, N., Wong, Y. L., Johnson, C. M., Johnson, S., Andreeva, A., Oegema, K., Lea, S. M., Raff, J. W., and van Breugel, M. Crystal structures of the CPAP/STIL complex reveal its role in centriole assembly and human microcephaly. *eLife 2* (Jan. 2013), e01071.

[72] Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallographica Section D: Biological Crystallography 62*, 9 (Sept. 2006), 1002–1011.

[73] Cuff, A., Redfern, O., and Orengo, C. Classification of protein structures. In *Computational Structural Biology: Methods and Applications*,

T. Schwede and M. C. Peitsch, Eds. World Scientific Publishing, May 2008, pp. 153–188.

[74] Culver, B. P., Meehl, J. B., Giddings, T. H., and Winey, M. The two SAS-6 homologs in *Tetrahymena thermophila* have distinct functions in basal body assembly. *Molecular Biology of the Cell 20*, 6 (Mar. 2009), 1865–1877.

[75] Cureton, D. K., Massol, R. H., Saffarian, S., Kirchhausen, T. L., and Whelan, S. P. J. Vesicular stomatitis virus enters cells through vesicles incompletely coated with clathrin that depend upon actin for internalization. *PLOS Pathogens 5*, 4 (Apr. 2009), e1000394.

[76] Curiel, J. A., de Las Rivas, B., Mancheño, J. M., and Muñoz, R. The pURI family of expression vectors: a versatile set of ligation independent cloning plasmids for producing recombinant His-fusion proteins. *Protein Expression and Purification 76*, 1 (Mar. 2011), 44–53.

[77] Daegelen, P., Studier, F. W., Lenski, R. E., Cure, S., and Kim, J. F. Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL606 and BL21(DE3). *Journal of Molecular Biology 394*, 4 (Dec. 2009), 634–643.

[78] Dammermann, A., Müller-Reichert, T., Pelletier, L., Habermann, B., Desai, A., and Oegema, K. Centriole assembly requires both centriolar and pericentriolar material proteins. *Developmental Cell 7*, 6 (Dec. 2004), 815–829.

[79] Dao-Thi, M.-H., Van Melderen, L., De Genst, E., Afif, H., Buts, L., Wyns, L., and Loris, R. Molecular basis of gyrase poisoning by the addiction toxin CcdB. *Journal of Molecular Biology 348*, 5 (May 2005), 1091–1102.

[80] Darden, T., York, D., and Pedersen, L. Particle mesh ewald: an n·log(n) method for ewald sums in large systems. *The Journal of Chemical Physics 98*, 12 (1993), 10089.

[81] Dauptain, A., Favier, J., and Bottaro, A. Hydrodynamics of ciliary propulsion. *Journal of Fluids and Structures 24*, 8 (Nov. 2008), 1156–1165.

[82] Davis, G. D., Elisee, C., Newham, D. M., and Harrison, R. G. New fusion protein systems designed to give soluble expression in *Escherichia coli*. *Biotechnology and Bioengineering 65*, 4 (Nov. 1999), 382–388.

[83] Davis, I. The 'super-resolution' revolution. *Biochemical Society Transactions 37*, 5 (Oct. 2009), 1042–1044.

[84] Dawe, H. R., Farr, H., and Gull, K. Centriole/basal body morphogenesis and migration during ciliogenesis in animal cells. *Journal of Cell Science 120*, Pt 1 (Jan. 2007), 7–15.

[85] De Jonge, N., Garcia-Pino, A., Buts, L., Haesaerts, S., Charlier, D., Zangger, K., Wyns, L., De Greve, H., and Loris, R. Rejuvenation of CcdB-poisoned gyrase by an intrinsically disordered protein domain. *Molecular Cell 35*, 2 (July 2009), 154–163.

[86] Debec, A., Sullivan, W., and Bettencourt-Dias, M. Centrioles: active players or passengers during mitosis? *Cellular and Molecular Life Sciences 67*, 13 (July 2010), 2173–2194.

[87] Delattre, M., Canard, C., and Gönczy, P. Sequential protein recruitment in *C. elegans* centriole formation. *Current Biology 16*, 18 (Sept. 2006), 1844–1849.

[88] Delattre, M., Leidel, S., Wani, K., Baumer, K., Bamat, J., Schnabel, H., Feichtinger, R., Schnabel, R., and Gönczy, P. Centriolar SAS-5 is required for centrosome duplication in *C. elegans*. *Nature Cell Biology 6*, 7 (July 2004), 656–664.

[89] di Guan, C., Li, P., Riggs, P. D., and Inouye, H. Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene 67*, 1 (July 1988), 21–30.

[90] Dickinson, D. J., Ward, J. D., Reiner, D. J., and Goldstein, B. Engineering the *Caenorhabditis elegans* genome using Cas9-triggered homologous recombination. *Nature Methods 10*, 10 (Oct. 2013), 1028–1034.

[91] Digambaranath, J. L., Dang, L., Dembinska, M., Vasyluk, A., and Finke, J. M. Conformations within soluble oligomers and insoluble aggregates revealed by resonance energy transfer. *Biopolymers 93*, 4 (Apr. 2010), 299–317.

[92] Dill, K. A., and MacCallum, J. L. The protein-folding problem, 50 years on. *Science 338*, 6110 (Nov. 2012), 1042–1046.

[93] Dix, C. I., and Raff, J. W. *Drosophila* Spd-2 recruits PCM to the sperm centriole, but is dispensable for centriole duplication. *Current Biology 17*, 20 (Oct. 2007), 1759–1764.

[94] Doré, A. S., Furnham, N., Davies, O. R., Sibanda, B. L., Chirgadze, D. Y., Jackson, S. P., Pellegrini, L., and Blundell, T. L. Structure of an Xrcc4-DNA ligase IV yeast ortholog complex reveals a novel BRCT interaction mode. *DNA Repair 5*, 3 (Mar. 2006), 362–368.

[95] Dubendorff, J. W., and Studier, F. W. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with *lac* repressor. *Journal of Molecular Biology 219*, 1 (May 1991), 45–59.

[96] Dümmler, A., Lawrence, A.-M., and de Marco, A. Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. *Microbial Cell Factories 4* (Jan. 2005), 34.

[97] Dutcher, S. K., and Trabuco, E. C. The *UNI3* gene is required for assembly of basal bodies of *Chlamydomonas* and encodes δ-tubulin, a new member of the tubulin superfamily. *Molecular Biology of the Cell 9*, 6 (June 1998), 1293–1308.

[98] Dzhindzhev, N. S., Tzolovsky, G., Lipinszki, Z., Schneider, S., Lattao, R., Fu, J., Debski, J., Dadlez, M., and Glover, D. M. Plk4 phosphorylates Ana2 to trigger Sas6 recruitment and procentriole formation. *Current Biology 24*, 21 (Nov. 2014), 2526–2532.

[99] Dzhindzhev, N. S., Yu, Q. D., Weiskopf, K., Tzolovsky, G., Cunha-Ferreira, I., Riparbelli, M., Rodrigues-Martins, A., Bettencourt-Dias, M., Callaini, G., and Glover, D. M. Asterless is a scaffold for the onset of centriole assembly. *Nature 467*, 7316 (Oct. 2010), 714–718.

[100] Edelstein, A., Amodaj, N., Hoover, K., Vale, R., and Stuurman, N. Computer control of microscopes using μManager. In *Current Protocols in Molecular Biology*, vol. 92. John Wiley & Sons, Inc., Oct. 2010, pp. 14.20.1–14.20.17.

[101] Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography 66*, 4 (Apr. 2010), 486–501.

[102] Erez, A., Castiel, A., Trakhtenbrot, L., Perelman, M., Rosenthal, E., Goldstein, I., Stettner, N., Harmelin, A., Eldar-Finkelman, H., Campaner, S., Kirsch, I., and Izraeli, S. The *SIL* gene is essential for mitotic entry and survival of cancer cells. *Cancer Research 67*, 9 (May 2007), 4022–4027.

[103] Erez, A., Perelman, M., Hewitt, S. M., Cojacaru, G., Goldberg, I., Shahar, I., Yaron, P., Muler, I., Campaner, S., Amariglio, N., Rechavi, G., Kirsch, I. R., Krupsky, M., Kaminski, N., and Izraeli, S. Sil overexpression in lung cancer characterizes tumors with increased mitotic activity. *Oncogene 23*, 31 (July 2004), 5371–5377.

[104] Esaki, K., Terashima, Y., Toda, E., Yoshinaga, S., Araki, N., Matsushima, K., and Terasawa, H. Expression and purification of human FROUNT, a common cytosolic regulator of CCR2 and CCR5. *Protein Expression and Purification 77*, 1 (May 2011), 86–91.

[105] Evans, P. R., and Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography 69*, 7 (July 2013), 1204–1214.

[106] Fersht, A. R. From the first protein structures to our current knowledge of protein folding: delights and scepticisms. *Nature Reviews Molecular Cell Biology 9*, 8 (Aug. 2008), 650–654.

[107] Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., and Bateman, A. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research 44*, D1 (Jan. 2016), D279–D285.

[108] Fliegauf, M., Benzing, T., and Omran, H. When cilia go bad: cilia defects and ciliopathies. *Nature Reviews Molecular Cell Biology 9*, 1 (Jan. 2007), 88–88.

[109] Fong, C. S., Kim, M., Yang, T. T., Liao, J.-C., and Tsou, M.-F. B. SAS-6 assembly templated by the lumen of cartwheel-less centrioles precedes centriole duplication. *Developmental Cell 30*, 2 (July 2014), 238–245.

[110] Franz, A., Roque, H., Saurya, S., Dobbelaere, J., and Raff, J. W. CP110 exhibits novel regulatory activities during centriole assembly in *Drosophila*. *The Journal of Cell Biology 203*, 5 (Dec. 2013), 785–799.

[111] Friedland, A. E., Tzur, Y. B., Esvelt, K. M., Colaiácovo, M. P., Church, G. M., and Calarco, J. A. Heritable genome editing in *C. elegans* via a CRISPR-Cas9 system. *Nature Methods 10*, 8 (Aug. 2013), 741–743.

[112] Frøkjær-Jensen, C., Davis, M. W., Ailion, M., and Jorgensen, E. M. Improved Mos1-mediated transgenesis in *C. elegans*. *Nature Methods 9*, 2 (Feb. 2012), 117–118.

[113] Frøkjær-Jensen, C., Davis, M. W., Hopkins, C. E., Newman, B. J., Thummel, J. M., Olesen, S.-P., Grunnet, M., and Jorgensen, E. M. Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nature Genetics 40*, 11 (Nov. 2008), 1375–1383.

[114] Fu, J., and Glover, D. M. Structured illumination of the interface between centriole and peri-centriolar material. *Open Biology 2*, 8 (Aug. 2012), 120104.

[115] FÄśrat-Karalar, E. N., and Stearns, T. The centriole duplication cycle. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 369*, 1650 (Sept. 2014), 20130460.

[116] Garreau de Loubresse, N., Ruiz, F., Beisson, J., and Klotz, C. Role of δ-tubulin and the C-tubule in assembly of *Paramecium* basal bodies. *BMC Cell Biology 2*, 1 (Jan. 2001), 1–7.

[117] Giansanti, M. G., Bucciarelli, E., Bonaccorsi, S., and Gatti, M. *Drosophila* SPD-2 is an essential centriole component required for PCM recruitment and astral-microtubule nucleation. *Current Biology 18*, 4 (Feb. 2008), 303–309.

[118] Gibbons, I. R., and Grimestone, A. V. On flagellar structure in certain flagellates. *The Journal of Biophysical and Biochemical Cytology 7* (July 1960), 697–716.

[119] Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R.-Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z.-Q., Segall-Shapiro, T. H., Calvey, C. H., Parmar, P. P., Hutchison, C. A., Smith, H. O., and Venter, J. C. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science 329*, 5987 (July 2010), 52–56.

[120] Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods 6*, 5 (May 2009), 343–345.

[121] Godinho, S. A., and Pellman, D. Causes and consequences of centrosome abnormalities in cancer. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 369*, 1650 (Sept. 2014), 20130467.

[122] Gomez-Ferreria, M. A., Rath, U., Buster, D. W., Chanda, S. K., Caldwell, J. S., Rines, D. R., and Sharp, D. J. Human Cep192 is

required for mitotic centrosome and spindle assembly. *Current Biology 17*, 22 (Nov. 2007), 1960–1966.

[123] Gönczy, P. Mechanisms of asymmetric cell division: flies and worms pave the way. *Nature Reviews Molecular Cell Biology 9*, 5 (May 2008), 355–366.

[124] Gönczy, P. Towards a molecular architecture of centriole assembly. *Nature Reviews Molecular Cell Biology 13*, 7 (July 2012), 425–435.

[125] Gönczy, P. Centrosomes and cancer: revisiting a long-standing relationship. *Nature Reviews Cancer 15*, 11 (Oct. 2015), 639–652.

[126] Gönczy, P., Schnabel, H., Kaletta, T., Amores, A. D., Hyman, T., and Schnabel, R. Dissection of cell division processes in the one cell stage *Caenorhabditis elegans* embryo by mutational analysis. *The Journal of Cell Biology 144*, 5 (Mar. 1999), 927–946.

[127] Gopalakrishnan, J., Mennella, V., Blachon, S., Zhai, B., Smith, A. H., Megraw, T. L., Nicastro, D., Gygi, S. P., Agard, D. A., and Avidor-Reiss, T. Sas-4 provides a scaffold for cytoplasmic complexes and tethers them in a centrosome. *Nature Communications 2* (June 2011), 359.

[128] Gordon, D. J., Resio, B., and Pellman, D. Causes and consequences of aneuploidy in cancer. *Nature Reviews Genetics 13*, 3 (Mar. 2012), 189–203.

[129] Gorrec, F. The MORPHEUS protein crystallization screen. *Journal of Applied Crystallography 42*, 6 (Nov. 2009), 1035–1042.

[130] Guderian, G., Westendorf, J., Uldschmid, A., and Nigg, E. A. Plk4 trans-autophosphorylation regulates centriole number by controlling βTrCP-mediated degradation. *Journal of Cell Science 123*, 13 (July 2010), 2163–2169.

[131] Guichard, P., Desfosses, A., Maheshwari, A., Hachet, V., Dietrich, C., Brune, A., Ishikawa, T., Sachse, C., and Gönczy, P. Cartwheel architecture of *Trichonympha* basal body. *Science 337*, 6094 (Aug. 2012), 553.

[132] Guichard, P., Hachet, V., Majubu, N., Neves, A., Demurtas, D., Olieric, N., Flückiger, I., Yamada, A., Kihara, K., Nishida, Y., Moriya, S., Steinmetz, M. O., Hongoh, Y., and Gönczy, P. Native architecture of the centriole proximal region reveals features underlying its 9-fold radial symmetry. *Current Biology 23*, 17 (Sept. 2013), 1620–1628.

[133] Gupta, H., Badarudeen, B., George, A., Thomas, G. E., Gireesh, K. K., and Manna, T. K. Human SAS-6 C-terminus nucleates and promotes microtubule assembly *in vitro* by binding to microtubules. *Biochemistry 54*, 41 (Oct. 2015), 6413–6422.

[134] Habedanck, R., Stierhof, Y.-D., Wilkinson, C. J., and Nigg, E. A. The Polo kinase Plk4 functions in centriole duplication. *Nature Cell Biology 7*, 11 (Nov. 2005), 1140–1146.

[135] Hammel, M., Rey, M., Yu, Y., Mani, R. S., Classen, S., Liu, M., Pique, M. E., Fang, S., Mahaney, B. L., Weinfeld, M., Schriemer, D. C., Lees-Miller, S. P., and Tainer, J. A. XRCC4 protein interactions with XRCC4-like factor (XLF) create an extended grooved scaffold for DNA ligation and double strand break repair. *The Journal of Biological Chemistry 286*, 37 (Sept. 2011), 32638–32650.

[136] Hammes, G. G., and Schullery, S. E. Structure of macromolecular aggregates. I. Aggregation-induced conformational changes in polypeptides. *Biochemistry 7*, 11 (Nov. 1968), 3882–3887.

[137] Han, J.-H., Batey, S., Nickson, A. A., Teichmann, S. A., and Clarke, J. The folding and evolution of multidomain proteins. *Nature Reviews Molecular Cell Biology 8*, 4 (Apr. 2007), 319–330.

[138] Harris, T. W., Baran, J., Bieri, T., Cabunoc, A., Chan, J., Chen, W. J., Davis, P., Done, J., Grove, C., Howe, K., Kishore, R., Lee, R., Li, Y., Muller, H.-M., Nakamura, C., Ozersky, P., Paulini, M., Raciti, D., Schindelman, G., Tuli, M. A., Van Auken, K., Wang, D., Wang, X., Williams, G., Wong, J. D., Yook, K., Schedl, T., Hodgkin, J., Berriman, M., Kersey, P., Spieth, J., Stein, L., and Sternberg,

P. W. WormBase 2014: new views of curated biology. *Nucleic Acids Research 42*, D1 (Jan. 2014), D789–D793.

[139] Hatch, E. M., Kulukian, A., Holland, A. J., Cleveland, D. W., and Stearns, T. Cep152 interacts with Plk4 and is required for centriole duplication. *The Journal of Cell Biology 191*, 4 (Nov. 2010), 721–729.

[140] Hatzopoulos, G. N., Erat, M. C., Cutts, E., Rogala, K. B., Slater, L. M., Stansfeld, P. J., and Vakonakis, I. Structural analysis of the G-box domain of the microcephaly protein CPAP suggests a role in centriole architecture. *Structure 21*, 11 (Nov. 2013), 2069–2077.

[141] Hernández, H., and Robinson, C. V. Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nature Protocols 2*, 3 (Jan. 2007), 715–726.

[142] Hilbert, M., Erat, M. C., Hachet, V., Guichard, P., Blank, I. D., Flückiger, I., Slater, L., Lowe, E. D., Hatzopoulos, G. N., Steinmetz, M. O., Gönczy, P., and Vakonakis, I. *Caenorhabditis elegans* centriolar protein SAS-6 forms a spiral that is consistent with imparting a nine-fold symmetry. *Proceedings of the National Academy of Sciences of the United States of America 110*, 28 (July 2013), 11373–11378.

[143] Hiraki, M., Nakazawa, Y., Kamiya, R., and Hirono, M. Bld10p constitutes the cartwheel-spoke tip and stabilizes the 9-fold symmetry of the centriole. *Current Biology 17*, 20 (Oct. 2007), 1778–1783.

[144] Hirono, M. Cartwheel assembly. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 369*, 1650 (Sept. 2014), 20130458.

[145] Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R., and Stüber, D. Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. *Nature Biotechnology 6*, 11 (Nov. 1988), 1321–1325.

[146] Hochuli, E., Döbeli, H., and Schacher, A. New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. *Journal of Chromatography 411* (Dec. 1987), 177–184.

[147] Hodges, M. E., Scheumann, N., Wickstead, B., Langdale, J. A., and Gull, K. Reconstructing the evolutionary history of the centriole from protein components. *Journal of Cell Science 123*, 9 (May 2010), 1407–1413.

[148] Holland, A. J., and Cleveland, D. W. Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nature Reviews Molecular Cell Biology 10*, 7 (July 2009), 478–487.

[149] Holland, A. J., Lan, W., Niessen, S., Hoover, H., and Cleveland, D. W. Polo-like kinase 4 kinase activity limits centrosome overduplication by autoregulating its own stability. *The Journal of Cell Biology 188*, 2 (Jan. 2010), 191–198.

[150] Holm, L., and Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Research 38*, suppl 2 (July 2010), W545–W549.

[151] Hsu, W.-B., Hung, L.-Y., Tang, C.-J. C., Su, C.-L., Chang, Y., and Tang, T. K. Functional characterization of the microtubule-binding and -destabilizing domains of CPAP and d-SAS-4. *Experimental Cell Research 314*, 14 (Aug. 2008), 2591–2602.

[152] Huang, B., Bates, M., and Zhuang, X. Super-resolution fluorescence microscopy. *Annual Review of Biochemistry 78* (July 2009), 993–1016.

[153] Hung, L.-Y., Chen, H.-L., Chang, C.-W., Li, B.-R., and Tang, T. K. Identification of a novel microtubule-destabilizing motif in CPAP that binds to tubulin heterodimers and inhibits microtubule assembly. *Molecular Biology of the Cell 15*, 6 (June 2004), 2697–2706.

[154] Hung, L. Y., Tang, C. J., and Tang, T. K. Protein 4.1 R-135 interacts with a novel centrosomal protein (CPAP) which is associated with the γ-tubulin complex. *Molecular and Cellular Biology 20*, 20 (Oct. 2000), 7813–7825.

[155] Illergård, K., Ardell, D. H., and Elofsson, A. Structure is three to ten times more conserved than sequence - a study of structural response in protein cores. *Proteins 77*, 3 (Nov. 2009), 499–508.

[156] Izraeli, S., Lowe, L. A., Bertness, V. L., Good, D. J., Dorward, D. W., Kirsch, I. R., and Kuehn, M. R. The *SIL* gene is required for mouse embryonic axial development and left-right specification. *Nature 399*, 6737 (June 1999), 691–694.

[157] Jakobsen, L., Vanselow, K., Skogs, M., Toyoda, Y., Lundberg, E., Poser, I., Falkenby, L. G., Bennetzen, M., Westendorf, J., Nigg, E. A., Uhlen, M., Hyman, A. A., and Andersen, J. S. Novel asymmetrically localizing components of human centrosomes identified by complementary proteomics methods. *The EMBO Journal 30*, 8 (Apr. 2011), 1520–1535.

[158] Jang, K.-S., Sweredoski, M. J., Graham, R. L. J., Hess, S., and Clemons, W. M. Comprehensive proteomic profiling of outer membrane vesicles from *Campylobacter jejuni*. *Journal of Proteomics 98* (Feb. 2014), 90–98.

[159] Janke, C., and Bulinski, J. C. Post-translational regulation of the microtubule cytoskeleton: mechanisms and functions. *Nature Reviews Molecular Cell Biology 12*, 12 (Dec. 2011), 773–786.

[160] Janknecht, R., de Martynoff, G., Lou, J., Hipskind, R. A., Nordheim, A., and Stunnenberg, H. G. Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus. *Proceedings of the National Academy of Sciences of the United States of America 88*, 20 (Oct. 1991), 8972–8976.

[161] Jenny, R. J., Mann, K. G., and Lundblad, R. L. A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. *Protein Expression and Purification 31*, 1 (Sept. 2003), 1–11.

[162] Jerka-Dziadosz, M., Gogendeau, D., Klotz, C., Cohen, J., Beisson, J., and Koll, F. Basal body duplication in Paramecium: the key role of Bld10 in assembly and stability of the cartwheel. *Cytoskeleton 67*, 3 (Mar. 2010), 161–171.

[163] Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science 337*, 6096 (Aug. 2012), 816–821.

[164] Johnson, W. C. Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins 35*, 3 (May 1999), 307–312.

[165] Junop, M. S., Modesti, M., Guarné, A., Ghirlando, R., Gellert, M., and Yang, W. Crystal structure of the Xrcc4 DNA repair protein and implications for end joining. *The EMBO Journal 19*, 22 (Nov. 2000), 5962–5970.

[166] Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D: Biological Crystallography 66*, 2 (Feb. 2010), 133–144.

[167] Kabsch, W. XDS. *Acta Crystallographica Section D: Biological Crystallography 66*, 2 (Feb. 2010), 125–132.

[168] Kalab, P., and Heald, R. The RanGTP gradient - a GPS for the mitotic spindle. *Journal of Cell Science 121*, 10 (May 2008), 1577–1586.

[169] Kapust, R. B., and Waugh, D. S. Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expression and Purification 19*, 2 (July 2000), 312–318.

[170] Keller, D., Orpinell, M., Olivier, N., Wachsmuth, M., Mahen, R., Wyss, R., Hachet, V., Ellenberg, J., Manley, S., and Gönczy, P. Mechanisms of HsSAS-6 assembly promoting centriole formation in human cells. *The Journal of Cell Biology 204*, 5 (Mar. 2014), 697–712.

[171] Keller, L. C., Geimer, S., Romijn, E., Yates, J., Zamora, I., and Marshall, W. F. Molecular architecture of the centriole proteome: the conserved WD40 domain protein POC1 is required for centriole duplication and length control. *Molecular Biology of the Cell 20*, 4 (Mar. 2009), 1150–1166.

[172] Kemp, C. A., Kopish, K. R., Zipperlen, P., Ahringer, J., and O'Connell, K. F. Centrosome maturation and duplication in *C. elegans* require the coiled-coil protein SPD-2. *Developmental Cell 6*, 4 (Apr. 2004), 511–523.

[173] Khafizov, K., Madrid-Aliste, C., Almo, S. C., and Fiser, A. Trends in structural coverage of the protein universe and the impact of the Protein Structure Initiative. *Proceedings of the National Academy of Sciences of the United States of America 111*, 10 (Mar. 2014), 3733–3738.

[174] Khatter, H., Myasnikov, A. G., Natchiar, S. K., and Klaholz, B. P. Structure of the human 80S ribosome. *Nature 520*, 7549 (Apr. 2015), 640–645.

[175] Kilburn, C. L., Pearson, C. G., Romijn, E. P., Meehl, J. B., Giddings, T. H., Culver, B. P., Yates, J. R., and Winey, M. New *Tetrahymena* basal body protein components identify basal body domain structure. *The Journal of Cell Biology 178*, 6 (Sept. 2007), 905–912.

[176] Kim, J.-S., and Raines, R. T. Ribonuclease S-peptide as a carrier in fusion proteins. *Protein Science 2*, 3 (Dec. 2008), 348–356.

[177] 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., and Lee, K. S. Hierarchical recruitment of Plk4 and regulation of centriole biogenesis by two centrosomal scaffolds, Cep192 and Cep152. *Proceedings of the National Academy of Sciences of the United States of America 110*, 50 (Dec. 2013), E4849–E4857.

[178] Kimple, M. E., Brill, A. L., and Pasker, R. L. Overview of affinity tags for protein purification. In *Current Protocols in Protein Science*, vol. 73. John Wiley & Sons, Inc., Jan. 2013, pp. 9.9.1–9.9.23.

[179] Kirkham, M., Müller-Reichert, T., Oegema, K., Grill, S., and Hyman, A. A. SAS-4 is a *C. elegans* centriolar protein that controls centrosome size. *Cell 112*, 4 (Feb. 2003), 575–587.

[180] Kitagawa, D., Busso, C., Flückiger, I., and Gönczy, P. Phosphorylation of SAS-6 by ZYG-1 is critical for centriole formation in *C. elegans* embryos. *Developmental Cell 17*, 6 (Dec. 2009), 900–907.

[181] Kitagawa, D., Flückiger, I., Polanowska, J., Keller, D., Reboul, J., and Gönczy, P. PP2A phosphatase acts upon SAS-5 to ensure centriole formation in *C. elegans* embryos. *Developmental Cell 20*, 4 (Apr. 2011), 550–562.

[182] Kitagawa, D., Kohlmaier, G., Keller, D., Strnad, P., Balestra, F. R., Flückiger, I., and Gönczy, P. Spindle positioning in human cells relies on proper centriole formation and on the microcephaly proteins CPAP and STIL. *Journal of Cell Science 124*, 22 (Nov. 2011), 3884–3893.

[183] Kitagawa, D., Vakonakis, I., Olieric, N., Hilbert, M., Keller, D., Olieric, V., Bortfeld, M., Erat, M. C., Flückiger, I., Gönczy, P., and Steinmetz, M. O. Structural basis of the 9-fold symmetry of centrioles. *Cell 144*, 3 (Feb. 2011), 364–375.

[184] Kleanthous, C., Kühlmann, U. C., Pommer, A. J., Ferguson, N., Radford, S. E., Moore, G. R., James, R., and Hemmings, A. M. Structural and mechanistic basis of immunity toward endonuclease colicins. *Nature Structural Biology 6*, 3 (Mar. 1999), 243–252.

[185] Klebba, J. E., Buster, D. W., McLamarrah, T. A., Rusan, N. M., and Rogers, G. C. Autoinhibition and relief mechanism for Polo-like kinase 4. *Proceedings of the National Academy of Sciences of the United States of America 112*, 7 (Feb. 2015), E657–E666.

[186] Kleylein-Sohn, J., Westendorf, J., Le Clech, M., Habedanck, R., Stierhof, Y.-D., and Nigg, E. A. Plk4-induced centriole biogenesis in human cells. *Developmental Cell 13*, 2 (Aug. 2007), 190–202.

[187] Klos Dehring, D. A., Vladar, E. K., Werner, M. E., Mitchell, J. W., Hwang, P., and Mitchell, B. J. Deuterosome-mediated centriole biogenesis. *Developmental Cell 27*, 1 (Oct. 2013), 103–112.

[188] Klotz, C., Dabauvalle, M. C., Paintrand, M., Weber, T., Bornens, M., and Karsenti, E. Parthenogenesis in *Xenopus* eggs requires centrosomal integrity. *The Journal of Cell Biology 110*, 2 (Feb. 1990), 405–415.

[189] Kohlmaier, G., Loncarek, J., Meng, X., McEwen, B. F., Mogensen, M. M., Spektor, A., Dynlacht, B. D., Khodjakov, A., and Gönczy, P. Overly long centrioles and defective cell division upon excess of the SAS-4-related protein CPAP. *Current Biology 19*, 12 (June 2009), 1012–1018.

[190] Kollman, J. M., Merdes, A., Mourey, L., and Agard, D. A. Microtubule nucleation by γ-tubulin complexes. *Nature Reviews Molecular Cell Biology 12*, 11 (Nov. 2011), 709–721.

[191] Kratz, A.-S., Bärenz, F., Richter, K. T., and Hoffmann, I. Plk4-dependent phosphorylation of STIL is required for centriole duplication. *Biology Open 4*, 3 (Jan. 2015), 370–377.

[192] Kulp, A., and Kuehn, M. J. Biological functions and biogenesis of secreted bacterial outer membrane vesicles. *Annual Review of Microbiology 64*, 1 (Oct. 2010), 163–184.

[193] Kumar, A., Girimaji, S. C., Duvvari, M. R., and Blanton, S. H. Mutations in STIL, encoding a pericentriolar and centrosomal protein, cause primary microcephaly. *The American Journal of Human Genetics 84*, 2 (Feb. 2009), 286–290.

[194] Kyte, J., and Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology 157*, 1 (May 1982), 105–132.

[195] La Terra, S., English, C. N., Hergert, P., McEwen, B. F., Sluder, G., and Khodjakov, A. The *de novo* centriole assembly pathway in HeLa cells: cell cycle progression and centriole assembly/maturation. *The Journal of Cell Biology 168*, 5 (Feb. 2005), 713–722.

[196] Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature 227*, 5259 (Aug. 1970), 680–685.

[197] Lauga, E., and Powers, T. R. The hydrodynamics of swimming microorganisms. *Reports on Progress in Physics 72*, 9 (Aug. 2009), 096601.

[198] LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F., and McCoy, J. M. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Nature Biotechnology 11*, 2 (Feb. 1993), 187–193.

[199] Lawo, S., Hasegan, M., Gupta, G. D., and Pelletier, L. Subdiffraction imaging of centrosomes reveals higher-order organizational features of pericentriolar material. *Nature Cell Biology 14*, 11 (Nov. 2012), 1148–1158.

[200] Leidel, S., Delattre, M., Cerutti, L., Baumer, K., and Gönczy, P. SAS-6 defines a protein family required for centrosome duplication in *C. elegans* and in human cells. *Nature Cell Biology 7*, 2 (Feb. 2005), 115–125.

[201] Leidel, S., and Gönczy, P. SAS-4 is essential for centrosome duplication in *C. elegans* and is recruited to daughter centrioles once per cell cycle. *Developmental Cell 4*, 3 (Mar. 2003), 431–439.

[202] Leslie, A. G. W., and Powell, H. R. Processing diffraction data with mosflm. In *Evolving Methods for Macromolecular Crystallography*, R. J. Read and J. L. Sussman, Eds., vol. 245 of *NATO Science Series*. Springer, 2007, pp. 41–51.

[203] Lettman, M. M., Wong, Y. L., Viscardi, V., Niessen, S., Chen, S.-H., Shiau, A. K., Zhou, H., Desai, A., and Oegema, K. Direct binding of SAS-6 to ZYG-1 recruits SAS-6 to the mother centriole for cartwheel assembly. *Developmental Cell 25*, 3 (May 2013), 284–298.

[204] Leung, B., Hermann, G. J., and Priess, J. R. Organogenesis of the *Caenorhabditis elegans* intestine. *Developmental Biology 216*, 1 (Dec. 1999), 114–134.

[205] Li, S., Fernandez, J.-J., Marshall, W. F., and Agard, D. A. Three-dimensional structure of basal body triplet revealed by electron cryotomography. *The EMBO Journal 31*, 3 (Feb. 2012), 552–562.

[206] Li, Y., Chirgadze, D. Y., Bolanos-Garcia, V. M., Sibanda, B. L., Davies, O. R., Ahnesorg, P., Jackson, S. P., and Blundell, T. L. Crystal structure of human XLF/Cernunnos reveals unexpected differences from XRCC4 with implications for NHEJ. *The EMBO Journal 27*, 1 (Jan. 2008), 290–300.

[207] Li, Z., Wakao, S., Fischer, B. B., and Niyogi, K. K. Sensing and responding to excess light. *Annual Review of Plant Biology 60*, 1 (June 2009), 239–260.

[208] Lilius, G., Persson, M., Bülow, L., and Mosbach, K. Metal affinity precipitation of proteins carrying genetically attached polyhistidine affinity tails. *European Journal of Biochemistry 198*, 2 (June 1991), 499–504.

[209] Lin, Y.-C., Chang, C.-W., Hsu, W.-B., Tang, C.-J. C., Lin, Y.-N., Chou, E.-J., Wu, C.-T., and Tang, T. K. Human microcephaly protein CEP135 binds to hSAS-6 and CPAP, and is required for centriole assembly. *The EMBO Journal 32*, 8 (Apr. 2013), 1141–1154.

[210] Linck, R., Fu, X., Lin, J., Ouch, C., Schefter, A., Steffen, W., Warren, P., and Nicastro, D. Insights into the structure and function of ciliary and flagellar doublet microtubules: tektins, Ca^{2+}-binding proteins, and stable protofilaments. *The Journal of Biological Chemistry 289*, 25 (July 2014), 17427–17444.

[211] Linck, R. W., and Norrander, J. M. Protofilament ribbon compartments of ciliary and flagellar microtubules. *Protist 154*, 3-4 (Oct. 2003), 299–311.

[212] Linck, R. W., and Stephens, R. E. Functional protofilament numbering of ciliary, flagellar, and centriolar microtubules. *Cell Motility and the Cytoskeleton 64*, 7 (July 2007), 489–495.

[213] Loncarek, J., and Khodjakov, A. *Ab ovo* or *de novo*? Mechanisms of centriole duplication. *Molecules and Cells 27*, 2 (Feb. 2009), 135–142.

[214] Lüders, J., and Stearns, T. Microtubule-organizing centres: a re-evaluation. *Nature Reviews Molecular Cell Biology 8*, 2 (Feb. 2007), 161–167.

[215] Lupas, A. N., and Gruber, M. The structure of α-helical coiled coils. *Advances in Protein Chemistry 70* (Jan. 2005), 37–78.

[216] Maheshwari, A., Obbineni, J. M., Bui, K. H., Shibata, K., Toyoshima, Y. Y., and Ishikawa, T. α- and β-tubulin lattice of the axonemal microtubule doublet and binding proteins revealed by single particle cryo-electron microscopy and tomography. *Structure 23*, 9 (July 2015), 1584–1595.

[217] Malakhov, M. P., Mattern, M. R., Malakhova, O. A., Drinker, M., Weeks, S. D., and Butt, T. R. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *Journal of Structural and Functional Genomics 5*, 1-2 (Jan. 2004), 75–86.

[218] Marshall, W. F. What is the function of centrioles? *Journal of Cellular Biochemistry 100*, 4 (Mar. 2007), 916–922.

[219] Marshall, W. F. Basal bodies: platforms for building cilia. In *Ciliary Function in Mammalian Development*, B. K. Yoder, Ed., vol. 85 of *Current Topics in Developmental Biology*. Elsevier BV, Jan. 2008, pp. 1–22.

[220] Marshall, W. F. Centriole evolution. *Current Opinion in Cell Biology 21*, 1 (Feb. 2009), 14–19.

[221] Matsuura, K., Lefebvre, P. A., Kamiya, R., and Hirono, M. Bld10p, a novel protein essential for basal body assembly in *Chlamydomonas*: localization to the cartwheel, the first ninefold symmetrical structure appearing during assembly. *The Journal of Cell Biology 165*, 5 (June 2004), 663–671.

[222] Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sisson, W., Villafranca, J. E., Janson, C. A., McElroy, H. E., and Gribskov, C. L. Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell 77*, 5 (June 1994), 761–771.

[223] McBroom, A. J., and Kuehn, M. J. Release of outer membrane vesicles by Gram-negative bacteria is a novel envelope stress response. *Molecular Microbiology 63*, 2 (Jan. 2007), 545–558.

[224] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. Phaser crystallographic software. *Journal of Applied Crystallography 40*, 4 (Aug. 2007), 658–674.

[225] McGuffin, L. J., Bryson, K., and Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics 16*, 4 (Apr. 2000), 404–405.

[226] McIntosh, J. R., Morphew, M. K., Grissom, P. M., Gilbert, S. P., and Hoenger, A. Lattice structure of cytoplasmic microtubules in a cultured mammalian cell. *Journal of Molecular Biology 394*, 2 (Dec. 2009), 177–182.

[227] McNitt, R. Centriole ultrastructure and its possible role in microtubule formation in an aquatic fungus. *Protoplasma 80*, 1 (Jan. 1974), 91–108.

[228] Mennella, V., Keszthelyi, B., McDonald, K. L., Chhun, B., Kan, F., Rogers, G. C., Huang, B., and Agard, D. A. Subdiffraction-resolution fluorescence microscopy reveals a domain of the centrosome critical for pericentriolar material organization. *Nature Cell Biology 14*, 11 (Nov. 2012), 1159–1168.

[229] Miki, T., Park, J. A., Nagao, K., Murayama, N., and Horiuchi, T. Control of segregation of chromosomal DNA by sex factor F in *Escherichia coli*. Mutants of DNA gyrase subunit A suppress *letD* (*ccdB*) product growth inhibition. *Journal of Molecular Biology 225*, 1 (May 1992), 39–52.

[230] Movassagh, T., Bui, K. H., Sakakibara, H., Oiwa, K., and Ishikawa, T. Nucleotide-induced global conformational changes of flagellar dynein arms revealed by *in situ* analysis. *Nature Structural & Molecular Biology 17*, 6 (June 2010), 761–767.

[231] Moyer, T. C., Clutario, K. M., Lambrus, B. G., Daggubati, V., and Holland, A. J. Binding of STIL to Plk4 activates kinase activity to promote centriole assembly. *The Journal of Cell Biology 209*, 6 (June 2015), 863–878.

[232] Müller-Reichert, T., Mäntler, J., Srayko, M., and O'Toole, E. Electron microscopy of the early *Caenorhabditis elegans* embryo. *Journal of Microscopy 230*, 2 (May 2008), 297–307.

[233] Murphy, K. M. *Janeway's immunobiology*, 8 ed. Garland Science, 2011.

[234] Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *Journal of Molecular Biology 247*, 4 (Apr. 1995), 536–540.

[235] Nachury, M. V. How do cilia organize signalling cascades? *Philosophical Transactions of the Royal Society of London B: Biological Sciences 369*, 1650 (Sept. 2014), 20130465.

[236] Nagai, T., Ibata, K., Park, E. S., Kubota, M., Mikoshiba, K., and Miyawaki, A. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnology 20*, 1 (Jan. 2002), 87–90.

[237] Nakazawa, Y., Hiraki, M., Kamiya, R., and Hirono, M. SAS-6 is a cartwheel protein that establishes the 9-fold symmetry of the centriole. *Current Biology 17*, 24 (Dec. 2007), 2169–2174.

[238] Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I., and Perrakis, A. Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallographica Section D: Biological Crystallography 61*, 10 (Oct. 2005), 1426–1431.

[239] Nicastro, D., Schwartz, C., Pierson, J., Gaudette, R., Porter, M. E., and McIntosh, J. R. The molecular architecture of axonemes revealed by cryoelectron tomography. *Science 313*, 5789 (Aug. 2006), 944–948.

[240] Nickerson, R. S. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology 2*, 2 (1998), 175–220.

[241] Nigg, E. A., and Raff, J. W. Centrioles, centrosomes, and cilia in health and disease. *Cell 139*, 4 (Nov. 2009), 663–678.

[242] Nigg, E. A., and Stearns, T. The centrosome cycle: Centriole biogenesis, duplication and inherent asymmetries. *Nature Cell Biology 13*, 10 (Oct. 2011), 1154–1160.

[243] Nishihara, K., Kanemori, M., Yanagi, H., and Yura, T. Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*. *Applied and Environmental Microbiology 66*, 3 (Mar. 2000), 884–889.

[244] Nogales, E., and Scheres, S. H. W. Cryo-EM: a unique tool for the visualization of macromolecular complexity. *Molecular Cell 58*, 4 (May 2015), 677–689.

[245] Nonaka, S., Tanaka, Y., Okada, Y., Takeda, S., Harada, A., Kanai, Y., Kido, M., and Hirokawa, N. Randomization of left-right asymmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein. *Cell 95*, 6 (Dec. 1998), 829–837.

[246] Oakley, C. E., and Oakley, B. R. Identification of γ-tubulin, a new member of the tubulin superfamily encoded by *mipA* gene of *Aspergillus nidulans*. *Nature 338*, 6217 (Apr. 1989), 662–4.

[247] Ochi, T., Blackford, A. N., Coates, J., Jhujh, S., Mehmood, S., Tamura, N., Travers, J., Wu, Q., Draviam, V. M., Robinson, C. V., Blundell, T. L., and Jackson, S. P. DNA repair. PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science 347*, 6218 (Jan. 2015), 185–188.

[248] O'Connell, C. B., and Khodjakov, A. L. Cooperative mechanisms of mitotic spindle formation. *Journal of Cell Science 120*, 10 (May 2007), 1717–1722.

[249] O'Connell, K. F., Caron, C., Kopish, K. R., Hurd, D. D., Kemphues, K. J., Li, Y., and White, J. G. The *C. elegans zyg-1* gene encodes a regulator of centrosome duplication with distinct maternal and paternal roles in the embryo. *Cell 105*, 4 (May 2001), 547–558.

[250] O'Connell, K. F., Maxwell, K. N., and White, J. G. The *spd-2* gene is required for polarization of the anteroposterior axis and formation of the sperm asters in the *Caenorhabditis elegans* zygote. *Developmental Biology 222*, 1 (June 2000), 55–70.

[251] Ohana, R. F., Encell, L. P., Zhao, K., Simpson, D., Slater, M. R., Urh, M., and Wood, K. V. HaloTag7: a genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification. *Protein Expression and Purification 68*, 1 (Nov. 2009), 110–120.

[252] Ohta, M., Ashikawa, T., Nozaki, Y., Kozuka-Hata, H., Goto, H., Inagaki, M., Oyama, M., and Kitagawa, D. Direct interaction of Plk4 with STIL ensures formation of a single procentriole per parental centriole. *Nature Communications 5* (Oct. 2014), 5267.

[253] Okonechnikov, K., Golosova, O., Fursov, M., and UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics 28*, 8 (Apr. 2012), 1166–1167.

[254] Oostenbrink, C., Villa, A., Mark, A. E., and van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry 25*, 13 (Oct. 2004), 1656–1676.

[255] Paintrand, M., Moudjou, M., Delacroix, H., and Bornens, M. Centrosome organization and centriole architecture: their sensitivity to divalent cations. *Journal of Structural Biology 108*, 2 (Jan. 1992), 107–128.

[256] 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., and Lee, K. S. Molecular basis for unidirectional scaffold switching of human Plk4 in centriole biogenesis. *Nature Structural & Molecular Biology 21*, 8 (Aug. 2014), 696–703.

[257] Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A., and Dougherty, W. G. Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. *Analytical Biochemistry 216*, 2 (Feb. 1994), 413–417.

[258] Parrinello, M. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics 52*, 12 (1981), 7182.

[259] Patthy, L. Modular assembly of genes and the evolution of new functions. In *Genetica*, M. Long, Ed., vol. 118 of *Contemporary Issues in Genetics and Evolution*. Kluwer Academic Publishers, July 2003, pp. 217–231.

[260] Pearson, C. G., Osborn, D. P. S., Giddings, T. H., Beales, P. L., and Winey, M. Basal body stability and ciliogenesis requires the conserved component Poc1. *The Journal of Cell Biology 187*, 6 (Dec. 2009), 905–920.

[261] Peel, N., Stevens, N. R., Basto, R., and Raff, J. W. Overexpressing centriole-replication proteins *in vivo* induces centriole overduplication and *de novo* formation. *Current Biology 17*, 10 (May 2007), 834–843.

[262] Pelletier, L., O'Toole, E., Schwager, A., Hyman, A. A., and Müller-Reichert, T. Centriole assembly in *Caenorhabditis elegans*. *Nature 444*, 7119 (Nov. 2006), 619–623.

[263] Pelletier, L., Ozlü, N., Hannak, E., Cowan, C., Habermann, B., Ruer, M., Müller-Reichert, T., and Hyman, A. A. The *Caenorhabditis elegans* centrosomal protein SPD-2 is required for both pericentriolar material recruitment and centriole duplication. *Current Biology 14*, 10 (May 2004), 863–873.

[264] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. UCSF Chimera - a visualization system for exploratory research and analysis. *Journal of Computational Chemistry 25*, 13 (Oct. 2004), 1605–1612.

[265] Pfaff, K. L., Straub, C. T., Chiang, K., Bear, D. M., Zhou, Y., and Zon, L. I. The zebra fish *cassiopeia* mutant reveals that SIL is required for mitotic spindle organization. *Molecular and Cellular Biology 27*, 16 (Aug. 2007), 5887–5897.

[266] Pieper, U., Webb, B. M., Dong, G. Q., Schneidman-Duhovny, D., Fan, H., Kim, S. J., Khuri, N., Spill, Y. G., Weinkam, P., Hammel, M., Tainer, J. A., Nilges, M., and Sali, A. ModBase, a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Research 42*, D1 (Jan. 2014), D336–D346.

[267] Porath, J. Immobilized metal ion affinity chromatography. *Protein Expression and Purification 3*, 4 (Aug. 1992), 263–281.

[268] Porath, J., Carlsson, J., Olsson, I., and Belfrage, G. Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature 258*, 5536 (Dec. 1975), 598–599.

[269] Praetorius, H. A., and Spring, K. R. Bending the MDCK cell primary cilium increases intracellular calcium. *Journal of Membrane Biology 184*, 1 (Nov. 2001), 71–79.

[270] Praitis, V., Casey, E., Collar, D., and Austin, J. Creation of low-copy integrated transgenic lines in *Caenorhabditis elegans*. *Genetics 157*, 3 (Mar. 2001), 1217–1226.

[271] Prakash, S., and Matouschek, A. Protein unfolding in the cell. *Trends in Biochemical Sciences 29*, 11 (Nov. 2004), 593–600.

[272] Primrose, S. B., and Twyman, R. M. *Principles of gene manipulation and genomics*, 7 ed. Blackwell Publishing, 2006.

[273] Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., and Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics 29*, 7 (Apr. 2013), 845–854.

[274] Puklowski, A., Homsi, Y., Keller, D., May, M., Chauhan, S., Kossatz, U., Grünwald, V., Kubicka, S., Pich, A., Manns, M. P., Hoffmann, I., Gönczy, P., and Malek, N. P. The SCF-FBXW5 E3-ubiquitin

ligase is regulated by PLK4 and targets HsSAS-6 to control centrosome duplication. *Nature Cell Biology 13*, 8 (Aug. 2011), 1004–1009.

[275] Qiao, R., Cabral, G., Lettman, M. M., Dammermann, A., and Dong, G. SAS-6 coiled-coil structure and interaction with SAS-5 suggest a regulatory mechanism in *C. elegans* centriole assembly. *The EMBO Journal 31*, 22 (Nov. 2012), 4334–4347.

[276] Ramaswamy, S., Ross, K. N., Lander, E. S., and Golub, T. R. A molecular signature of metastasis in primary solid tumors. *Nature Genetics 33*, 1 (Jan. 2003), 49–54.

[277] Rapali, P., Szenes, A., Radnai, L., Bakos, A., Pál, G., and Nyitray, L. DYNLL/LC8: a light chain subunit of the dynein motor complex and beyond. *FEBS Journal 278*, 17 (Sept. 2011), 2980–2996.

[278] Recuero-Checa, M. A., Doré, A. S., Arias-Palomo, E., Rivera-Calzada, A., Scheres, S. H. W., Maman, J. D., Pearl, L. H., and Llorca, O. Electron microscopy of Xrcc4 and the DNA ligase IV-Xrcc4 DNA repair complex. *DNA Repair 8*, 12 (Dec. 2009), 1380–1389.

[279] Reina, J., and Gonzalez, C. When fate follows age: unequal centrosomes in asymmetric cell division. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 369*, 1650 (Sept. 2014), 20130466.

[280] Richards, A. *University of Oxford Advanced Research Computing*, Aug. 2015.

[281] Ringo, D. L. Flagellar motion and fine structure of the flagellar apparatus in *Chlamydomonas*. *The Journal of Cell Biology 33*, 3 (June 1967), 543–571.

[282] Riparbelli, M. G., Dallai, R., and Callaini, G. The insect centriole: a land of discovery. *Tissue and Cell 42*, 2 (Apr. 2010), 69–80.

[283] Robert, X., and Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research 42*, W1 (July 2014), W320–W324.

[284] Robin, F. B., McFadden, W. M., Yao, B., and Munro, E. M. Single-molecule analysis of cell surface dynamics in *Caenorhabditis elegans* embryos. *Nature Methods 11*, 6 (June 2014), 677–682.

[285] Rocco, C. J., Dennison, K. L., Klenchin, V. A., Rayment, I., and Escalante-Semerena, J. C. Construction and use of new cloning vectors for the rapid isolation of recombinant proteins from *Escherichia coli*. *Plasmid 59*, 3 (May 2008), 231–237.

[286] Rodrigues-Martins, A., Bettencourt-Dias, M., Riparbelli, M., Ferreira, C., Ferreira, I., Callaini, G., and Glover, D. M. DSAS-6 organizes a tube-like centriole precursor, and its absence suggests modularity in centriole assembly. *Current Biology 17*, 17 (Sept. 2007), 1465–1472.

[287] Rodríguez, D. D., Grosse, C., Himmel, S., González, C., de Ilarduya, I. M., Becker, S., Sheldrick, G. M., and Usón, I. Crystallographic *ab initio* protein structure solution below atomic resolution. *Nature Methods 6*, 9 (Sept. 2009), 651–653.

[288] Ropars, V., Drevet, P., Legrand, P., Baconnais, S., Amram, J., Faure, G., Márquez, J. A., Piétrement, O., Guerois, R., Callebaut, I., Le Cam, E., Revy, P., de Villartay, J.-P., and Charbonnier, J.-B. Structural characterization of filaments formed by human Xrcc4-Cernunnos/XLF complex involved in nonhomologous DNA end-joining. *Proceedings of the National Academy of Sciences of the United States of America 108*, 31 (Aug. 2011), 12663–12668.

[289] Rosenbaum, J. L., and Witman, G. B. Intraflagellar transport. *Nature Reviews Molecular Cell Biology 3*, 11 (Nov. 2002), 813–825.

[290] Ross, I., Clarissa, C., Giddings, T. H., and Winey, M. ϵ-tubulin is essential in *Tetrahymena thermophila* for the assembly and stability of basal bodies. *Journal of Cell Science 126*, 15 (Aug. 2013), 3441–3451.

[291] Sammito, M., Millán, C., Frieske, D., Rodríguez-Freire, E., Borges, R. J., and Usón, I. ARCIMBOLDO_LITE : single-workstation

implementation and use. *Acta Crystallographica Section D: Biological Crystallography 71*, 9 (Aug. 2015), 1921–1930.

[292] Sammito, M., Millán, C., Rodríguez, D. D., de Ilarduya, I. M., Meindl, K., De Marino, I., Petrillo, G., Buey, R. M., de Pereda, J. M., Zeth, K., Sheldrick, G. M., and Usón, I. Exploiting tertiary structure through local folds for crystallographic phasing. *Nature Methods 10*, 11 (Nov. 2013), 1099–1101.

[293] Savitzky, A., and Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry 36*, 8 (July 1964), 1627–1639.

[294] Sawaya, M. R. Methods to refine macromolecular structures in cases of severe diffraction anisotropy. In *Structural Genomics*, Y. W. Chen, Ed., vol. 1091 of *Methods in Molecular Biology*. Humana Press, Jan. 2014, pp. 205–214.

[295] Schmidt, T. I., Kleylein-Sohn, J., Westendorf, J., Le Clech, M., Lavoie, S. B., Stierhof, Y.-D., and Nigg, E. A. Control of centriole length by CPAP and CP110. *Current Biology 19*, 12 (June 2009), 1005–1011.

[296] Schmutz, C., and Spang, A. Knockdown of the centrosomal component SAS-5 results in defects in nuclear morphology in *Caenorhabditis elegans*. *European Journal of Cell Biology 84*, 1 (Jan. 2005), 75–82.

[297] Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods 9*, 7 (July 2012), 671–675.

[298] Scholz, J., Besir, H., Strasser, C., and Suppmann, S. A new method to customize protein expression vectors for fast, efficient and background free parallel cloning. *BMC Biotechnology 13* (Jan. 2013), 12.

[299] Sezonov, G., Joseleau-Petit, D., and D'Ari, R. *Escherichia coli* physiology in Luria-Bertani broth. *Journal of Bacteriology 189*, 23 (Dec. 2007), 8746–8749.

[300] Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallographica Section D: Biological Crystallography 66*, 4 (Apr. 2010), 479–485.

[301] Shimanovskaya, E., Viscardi, V., Lesigang, J., Lettman, M. M., Qiao, R., Svergun, D. I., Round, A., Oegema, K., and Dong, G. Structure of the *C. elegans* ZYG-1 cryptic polo box suggests a conserved mechanism for centriolar docking of Plk4 kinases. *Structure 22*, 8 (Aug. 2014), 1090–1104.

[302] Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. Cell-free translation reconstituted with purified components. *Nature Biotechnology 19*, 8 (Aug. 2001), 751–755.

[303] Sibanda, B. L., Critchlow, S. E., Begun, J., Pei, X. Y., Jackson, S. P., Blundell, T. L., and Pellegrini, L. Crystal structure of an Xrcc4-DNA ligase IV complex. *Nature Structural Biology 8*, 12 (Dec. 2001), 1015–1019.

[304] Slevin, L. K., Nye, J., Pinkerton, D. C., Buster, D. W., Rogers, G. C., and Slep, K. C. The structure of the Plk4 cryptic polo box reveals two tandem polo boxes required for centriole duplication. *Structure 20*, 11 (Nov. 2012), 1905–1917.

[305] Slevin, L. K., Romes, E. M., Dandulakis, M. G., and Slep, K. C. The mechanism of dynein light chain LC8-mediated oligomerization of the Ana2 centriole duplication factor. *The Journal of Biological Chemistry 289*, 30 (July 2014), 20727–20739.

[306] Smart, O. S., Womack, T. O., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., and Bricogne, G. Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallographica Section D: Biological Crystallography 68*, 4 (Apr. 2012), 368–380.

[307] Smith, B. J. Chemical cleavage of proteins. In *Basic Protein and Peptide Protocols*, J. M. Walker, Ed., vol. 32 of *Methods in Molecular Biology*. Humana Press, Jan. 1994, pp. 297–309.

[308] Smith, D. B., and Johnson, K. S. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene 67*, 1 (July 1988), 31–40.

[309] Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P., and Kobe, B. Crystal structures of fusion proteins with large-affinity tags. *Protein Science 12*, 7 (July 2003), 1313–1322.

[310] Song, M. H., Liu, Y., Anderson, D. E., Jahng, W. J., and O'Connell, K. F. Protein phosphatase 2A-SUR-6/B55 regulates centriole duplication in *C. elegans* by controlling the levels of centriole assembly factors. *Developmental Cell 20*, 4 (Apr. 2011), 563–571.

[311] Sonnen, K. F., Gabryjonczyk, A.-M., Anselm, E., Stierhof, Y.-D., and Nigg, E. A. Human Cep192 and Cep152 cooperate in Plk4 recruitment and centriole duplication. *Journal of Cell Science 126*, 14 (July 2013), 3223–3233.

[312] Sonnen, K. F., Schermelleh, L., Leonhardt, H., and Nigg, E. A. 3D-structured illumination microscopy provides novel insight into architecture of human centrosomes. *Biology Open 1*, 10 (Oct. 2012), 965–976.

[313] Stevens, N. R., Dobbelaere, J., Brunk, K., Franz, A., and Raff, J. W. *Drosophila* Ana2 is a conserved centriole duplication factor. *The Journal of Cell Biology 188*, 3 (Feb. 2010), 313–323.

[314] Stevens, N. R., Raposo, A. A. S. F., Basto, R., St Johnston, D., and Raff, J. W. From stem cell to embryo without centrioles. *Current Biology 17*, 17 (Sept. 2007), 1498–1503.

[315] Stevens, N. R., Roque, H., and Raff, J. W. DSas-6 and Ana2 coassemble into tubules to promote centriole duplication and engagement. *Developmental Cell 19*, 6 (Dec. 2010), 913–919.

[316] Strnad, P., Leidel, S., Vinogradova, T., Euteneuer, U., Khodjakov, A., and Gönczy, P. Regulated HsSAS-6 levels ensure formation of a single

procentriole per centriole during the centrosome duplication cycle. *Developmental Cell 13*, 2 (Aug. 2007), 203–213.

[317] Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D., and Eisenberg, D. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America 103*, 21 (May 2006), 8060–8065.

[318] Studier, F. W., and Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology 189*, 1 (May 1986), 113–130.

[319] Su, Y., Zou, Z., Feng, S., Zhou, P., and Cao, L. The acidity of protein fusion partners predominantly determines the efficacy to improve the solubility of the target proteins expressed in *Escherichia coli*. *Journal of Biotechnology 129*, 3 (May 2007), 373–382.

[320] Sui, H., and Downing, K. H. Molecular architecture of axonemal microtubule doublets revealed by cryo-electron tomography. *Nature 442*, 7101 (July 2006), 475–478.

[321] Tamm, S. L., and Horridge, G. A. The relation between the orientation of the central fibrils and the direction of beat in cilia of *Opalina*. *Proceedings of the Royal Society of London B: Biological Sciences 175*, 1040 (Aug. 1970), 219–233.

[322] Tang, C.-J. C., Fu, R.-H., Wu, K.-S., Hsu, W.-B., and Tang, T. K. CPAP is a cell-cycle regulated protein that controls centriole length. *Nature Cell Biology 11*, 7 (July 2009), 825–831.

[323] Tang, C.-J. C., Lin, S.-Y., Hsu, W.-B., Lin, Y.-N., Wu, C.-T., Lin, Y.-C., Chang, C.-W., Wu, K.-S., and Tang, T. K. The human microcephaly protein STIL interacts with CPAP and is required for procentriole formation. *The EMBO Journal 30*, 23 (Nov. 2011), 4790–4804.

[324] Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., Afonine, P. V., Zwart, P. H., and Hung, L. W. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallographica Section D: Biological Crystallography 65*, 6 (June 2009), 582–601.

[325] Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., Read, R. J., and Adams, P. D. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallographica Section D: Biological Crystallography 64*, 1 (Jan. 2008), 61–69.

[326] Thastrup, O., Tullin, S., Poulsen, L. K., and Bjørn, S. P. Method of detecting biologically active substances., Aug. 1996.

[327] Thomas, J. M. H., Keegan, R. M., Bibby, J., Winn, M. D., Mayans, O., and Rigden, D. J. Routine phasing of coiled-coil protein crystal structures with AMPLE. *IUCrJ 2*, 2 (Mar. 2015), 198–206.

[328] Trigg, J., Gutwin, K., Keating, A. E., and Berger, B. Multicoil2: predicting coiled coils and their oligomerization states from sequence in the twilight zone. *PLOS ONE 6*, 8 (Jan. 2011), e23519.

[329] Tsou, M.-F. B., and Stearns, T. Mechanism limiting centrosome duplication to once per cell cycle. *Nature 442*, 7105 (Aug. 2006), 947–951.

[330] Ueguchi, C., and Ito, K. Multicopy suppression: an approach to understanding intracellular functioning of the protein export system. *Journal of Bacteriology 174*, 5 (Mar. 1992), 1454–1461.

[331] 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., and Zuber, B. Structures of SAS-6 suggest its organization in centrioles. *Science 331*, 6021 (Mar. 2011), 1196–1199.

[332] van Breugel, M., Wilcken, R., McLaughlin, S. H., Rutherford, T. J., and Johnson, C. M. Structure of the SAS-6 cartwheel hub from *Leishmania major*. *eLife 3* (Jan. 2014), e01812.

[333] van der Lee, R., Lang, B., Kruse, K., Gsponer, J., Sánchez de Groot, N., Huynen, M. A., Matouschek, A., Fuxreiter, M., and Babu, M. M. Intrinsically disordered segments affect protein half-life in the cell and during evolution. *Cell Reports 8*, 6 (Sept. 2014), 1832–1844.

[334] Vangindertael, J., Beets, I., Rocha, S., Dedecker, P., Schoofs, L., Vanhoorelbeeke, K., Hofkens, J., and Mizuno, H. Super-resolution mapping of glutamate receptors in *C. elegans* by confocal correlated PALM. *Scientific Reports 5* (Sept. 2015), 13532.

[335] Varmark, H., Llamazares, S., Rebollo, E., Lange, B., Reina, J., Schwarz, H., and Gonzalez, C. Asterless is a centriolar protein required for centrosome function and embryo development in *Drosophila*. *Current Biology 17*, 20 (Oct. 2007), 1735–1745.

[336] Varsaki, A., Moncalián, G., Garcillán-Barcia, M. d. P., Drainas, C., and de la Cruz, F. Analysis of ColE1 MbeC unveils an extended ribbon-helix-helix family of nicking accessory proteins. *Journal of Bacteriology 191*, 5 (Mar. 2009), 1446–1455.

[337] Vulprecht, J., David, A., Tibelius, A., Castiel, A., Konotop, G., Liu, F., Bestvater, F., Raab, M. S., Zentgraf, H., Izraeli, S., and Krämer, A. STIL is required for centriole duplication in human cells. *Journal of Cell Science 125*, 5 (Mar. 2012), 1353–1362.

[338] Walczak, C. E., Cai, S., and Khodjakov, A. Mechanisms of chromosome behaviour during mitosis. *Nature Reviews Molecular Cell Biology 11*, 2 (Feb. 2010), 91–102.

[339] Walshaw, J., and Woolfson, D. N. SOCKET: a program for identifying and analysing coiled-coil motifs within protein structures. *Journal of Molecular Biology 307*, 5 (Apr. 2001), 1427–1450.

[340] Wang, C., Li, S., Januschke, J., Rossi, F., Izumi, Y., Garcia-Alvarez, G., Gwee, S. S. L., Soon, S. B., Sidhu, H. K., Yu, F., Matsuzaki, F., Gonzalez, C., and Wang, H. An Ana2/Ctp/Mud complex regulates spindle orientation in *Drosophila* neuroblasts. *Developmental Cell 21*, 3 (Sept. 2011), 520–533.

[341] Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F., and Jones, D. T. The DISOPRED server for the prediction of protein disorder. *Bioinformatics 20*, 13 (Sept. 2004), 2138–2139.

[342] West, A. P., Scharf, L., Scheid, J. F., Klein, F., Bjorkman, P. J., and Nussenzweig, M. C. Structural insights on the role of antibodies in HIV-1 vaccine and therapy. *Cell 156*, 4 (Feb. 2014), 633–648.

[343] Whitmore, L., and Wallace, B. A. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers 89*, 5 (May 2008), 392–400.

[344] Winey, M., and O'Toole, E. Centriole structure. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 369*, 1650 (Sept. 2014), 20130457.

[345] Witman, G. B. Introduction to cilia and flagella. In *Ciliary and Flagellar Membranes*, R. A. Bloodgood, Ed. Plenum Press, 1990, pp. 1–30.

[346] Wong, C., and Stearns, T. Centrosome number is controlled by a centrosome-intrinsic block to reduplication. *Nature Cell Biology 5*, 6 (June 2003), 539–544.

[347] Woodruff, J. B., Wueseke, O., and Hyman, A. A. Pericentriolar material structure and dynamics. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 369*, 1650 (Sept. 2014), 20130459.

[348] Woods, C. G., Bond, J., and Enard, W. Autosomal recessive primary microcephaly (MCPH): a review of clinical, molecular, and evolutionary findings. *The American Journal of Human Genetics 76*, 5 (May 2005), 717–728.

[349] Woolfson, D. N., Bartlett, G. J., Burton, A. J., Heal, J. W., Niitsu, A., Thomson, A. R., and Wood, C. W. *De novo* protein design: how do we expand into the universe of possible protein structures? *Current Opinion in Structural Biology 33* (Aug. 2015), 16–26.

[350] Wu, P.-Y., Frit, P., Meesala, S., Dauvillier, S., Modesti, M., Andres, S. N., Huang, Y., Sekiguchi, J., Calsou, P., Salles, B., and Junop, M. S. Structural and functional interaction between the human DNA repair proteins DNA ligase IV and XRCC4. *Molecular and Cellular Biology 29*, 11 (June 2009), 3163–3172.

[351] Wu, Q., Ochi, T., Matak-Vinkovic, D., Robinson, C. V., Chirgadze, D. Y., and Blundell, T. L. Non-homologous end-joining partners in a helical dance: structural studies of XLF-XRCC4 interactions. *Biochemical Society Transactions 39*, 5 (Oct. 2011), 1387–1392.

[352] Zhang, K. Y., Cowtan, K., and Main, P. Combining constraints for electron-density modification. In *Macromolecular Crystallography Part B*, J. F. Riordan and B. L. Vallee, Eds., vol. 277 of *Methods in Enzymology*. Academic Press, Jan. 1997, pp. 53–64.

[353] Zhao, H., Zhu, L., Zhu, Y., Cao, J., Li, S., Huang, Q., Xu, T., Huang, X., Yan, X., and Zhu, X. The Cep63 paralogue Deup1 enables massive *de novo* centriole biogenesis for vertebrate multiciliogenesis. *Nature Cell Biology 15*, 12 (Dec. 2013), 1434–1444.

[354] Zheng, X., Gooi, L. M., Wason, A., Gabriel, E., Mehrjardi, N. Z., Yang, Q., Zhang, X., Debec, A., Basiri, M. L., Avidor-Reiss, T., Pozniakovsky, A., Poser, I., Saric, T., Hyman, A. A., Li, H., and Gopalakrishnan, J. Conserved TCP domain of Sas-4/CPAP is essential for pericentriolar material tethering during centrosome biogenesis. *Proceedings of the National Academy of Sciences of the United States of America 111*, 3 (Jan. 2014), E354–E363.

[355] Zhou, F., and Roy, S. SnapShot: Motile Cilia. *Cell 162*, 1 (July 2015), 224–224.e1.

[356] Zhu, F., Lawo, S., Bird, A., Pinchev, D., Ralph, A., Richter, C., Müller-Reichert, T., Kittler, R., Hyman, A. A., and Pelletier, L. The mammalian SPD-2 ortholog Cep192 regulates centrosome biogenesis. *Current Biology 18*, 2 (Jan. 2008), 136–141.

[357] Zhu, P., Liu, J., Bess, J., Chertova, E., Lifson, J. D., Grisé, H., Ofek, G. A., Taylor, K. A., and Roux, K. H. Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature 441*, 7095 (June 2006), 847–852.

[358] Zou, Z., Cao, L., Zhou, P., Su, Y., Sun, Y., and Li, W. Hyper-acidic protein fusion partners improve solubility and assist correct folding of recombinant proteins expressed in *Escherichia coli*. *Journal of Biotechnology 135*, 4 (July 2008), 333–339.

[359] Zwart, P. H., Grosse-Kunstleve, R. W., Lebedev, A. A., Murshudov, G. N., and Adams, P. D. Surprises and pitfalls arising from (pseudo)symmetry. *Acta Crystallographica Section D: Biological Crystallography 64*, 1 (Jan. 2008), 99–107.