Structural characterisation of proteins at the cohesion/replication interface

Zuzana Hodáková

University College London

and

The Francis Crick Institute

PhD Supervisors: Martin R. Singleton & Peter Cherepanov

A thesis submitted for the degree of

Doctor of Philosophy

University College London

February 2021

Declaration

I, Zuzana Hodáková, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Cohesion establishment and DNA synthesis are tightly regulated processes occurring at the replication fork. DNA synthesis is under the control of the replisome; a complex machinery of numerous proteins which mediate DNA unwinding and DNA synthesis. Factors interacting with the replisome facilitate replication-related events ahead and behind the fork. Arising impediments ahead of the replisome are resolved by specialised helicases and DNA repair complexes. Conversely, events occurring behind the fork focus not only on DNA synthesis but also on the joining of the newly synthesised sister chromatids. The protein complex cohesin is responsible for ensuring that the sisters are joined immediately after replication and remain held together until anaphase. The precise spatiotemporal relations of proteins at the replication fork have not been fully elucidated.

This thesis addresses two important questions of cohesion establishment using structural biology tools. The first focuses on a long outstanding question on the mechanisms of cohesin loading onto DNA. It exposes insights into the folding mechanism of cohesin upon loading, governed by its accessory complex, the cohesin loader. The thesis further describes the variety of modifications which can be applied to study cohesin, and characterises the overall architectures of the loader complexes. The second question studied in this thesis describes the link between cohesion and DNA replication via the small helicase Chl1. There is currently no structure of Chl1 available and therefore the presented Chl1 envelope is the first structural characterisation of this helicase, pointing to a conserved architecture amongst the XPD subfamily of helicases. Additional work focuses on studying a potential auto-inhibition mechanism contributing to the function of this helicase in response to replication stress.

Impact Statement

The ability to propagate is one of the essential characteristics of a living organism. Duplicating and passing on genetic material occurs in every cell cycle of a cell. Getting it right is crucial: aberrations in DNA replication and sister chromatid cohesion can lead to the development of hereditary mutations or chromosomal instability, a common hallmark of cancer, or even cell death. Structural characterisation of events at the replication fork are a matter of study for decades.

The evolutionary conservation and the related structural appearance of the proteins studied in this thesis provides clues to the understanding of the function of these as well as homologous proteins. This is highly important for medicinal purposes, as drug discovery is often led by insights gained by structural characterisation of proteins. Aberrations in chromosome segregation which lead to cancer development could, with the use of structure-guided drug design, be treated and corrected for. Furthermore, the protein targets studied in this thesis, cohesin and Chl1, are mutated in genetic disorders. Understanding their function is therefore crucial to be able to develop therapies and treatments for these disorders.

This thesis provides insights into complex, not fully understood processes essential for life in the hopes that they will inform and help guide future research. The information gained with this research is crucial for understanding how hereditary material is passed from generation to generation.

Acknowledgement

I would first like to thank my supervisor, Martin Singleton, for giving me the opportunity to do a PhD in his lab in the first place. I am very grateful for his mentoring and help throughout the years. I could not have imagined a better environment for a starting scientist. I would also like to thank my other supervisor, Peter Cherepanov, for adopting me into his lab, for the enlightening discussions about image processing and his insight in general.

I would also like to thank my thesis committee, Nic Tapon, Rob de Bruin and Alessandro Costa, for insightful discussions regarding my projects, and all the feedback received.

I would then like to thank both the Singleton lab and the Cherepanov lab for their endless support when things were stressful, and their incredible patience with my hyperactive moods. A very big thanks goes to Allison Ballandras-Colas: for her help in the lab, her career guidance but mainly the incredible wisdom that she has brought into my scientific and personal life. And secondly to Matthew Singer and Joshua Hope: my knights in lab coats that always brighten my day. I further wish to thank Frank Uhlmann and the whole CSL lab who have been kind and welcoming, taught me how cute yeast are under the microscope, and have been a wonderful lab to work with. I also want to thank all the STPs I have worked with for over the years, with a big thanks to Simone Kunzelmann from the SB STP, and Peter Faull, Fairouz Ibrahim and Shahid Mehmood from the Protseomics STP. A special thanks goes to Peter Rosenthal for his oversight over my data collections, and Andrea Nans, who has patiently answered every question I ever had (and there were many) and got me so interested in EM that I decided to pursue a career in academia beyond this PhD. In this respect, my thanks extends to Donald Benton and Antoni Wrobel who, aside from image processing, taught me that everything can be solved with a keg of beer.

I would also like to thank my amazing friend Flavia Roşianu who started this PhD with me and has been the biggest, most kindest support I could wish for. Not just during the PhD, but also on those stressful nights watching the Bake-Off. My other

thanks goes to my best friend Nela Halamová, whose rational opinions and terrible jokes got me through this PhD and will most likely also get me through the rest of my career.

Most importantly, I would like to thank my mother Alena, my role model and my favourite person in the world, who I thank for absolutely everything. For her guidance to make the right life choices, for her love and support. For being able to do this PhD, and to finish it.

Table of Contents

Table of Figures

List of Tables

Abbreviations

A. gossypii/Ag	*Ashbya gossypii*
A. thaliana/At	*Arabidopsis thaliana*
ABC	ATP-binding cassette
ATP	Adenosine triphosphate
bp	base pair
Bs	*Bacillus subtilis*
BS3	Bis(sulfosuccinimidyl)suberate
C. thermophilum/Ct	*Chaetomium thermophilum*
CAK	CDK-activating kinase
CCD	Charge-coupled device
CDK	Cyclin-dependent kinase
CdLS	Cornelia de Lange Syndrome
Chl1	Chromosome loss 1 protein
CIP box	Ctf4 interacting peptide
Cryo-EM	Electron cryo-microscopy
Cryo-ET	Electron cryo-tomography
CTCF	CCCTC-binding factor
CTD	C-terminal domain
CTF	Contrast transfer function
CV	Column volume
D-loop	Displacement loop
D. melanogaster	*Drosophila melanogaster*
DDD	Direct detection device
DDK	Dbf4-dependent kinase
DNA	Deoxyribonucleic acid
DQE	Detective quantum efficiency
DSB	Double-stranded break
dsDNA	Double-stranded DNA
E.coli	*Escherichia coli*
Eco1	Establishment of cohesion protein 1
EM	Electron microscopy
FA	Fanconi Anaemia
FACS	Fluorescence-activated cell sorting
FACT	Facilitates chromatin transactions
FAM	Fluorescein
FancJ	Fanconi Anaemia complement group J
Fe-S	Iron-Sulphur
FEG	Field emission gun
FEN1	Flap endonuclease 1
FPC	Fork protection complex
FRET	Fluorescence resonance energy transfer
FSC	Fourier Shell Correlation
G	Guanine
G1	Gap-1 phase
G4	Guanine quadruplex
GD	Globular domain
GINS	Go Ichi Ni San

GO	Graphene Oxide
GraFix	Gradient fixation
GST	Glutathione-S-transferase
H. sapiens/Hs	*Homo sapiens*
HD	Helicase domain
HDX-MS	Hydrogen Deuterium Exchange-Mass Spectrometry
HEAT	Huntingtin, elongation factor 3, A subunit and TOR
Hi-C	High-throughput chromosome capture
His tag	6xHistidine tag
HR	Homologous recombination
HU	Hydroxyurea
ICL	Interstrand crosslink
IEX	Ion exchange chromatography
IP	Immunoprecipitation
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
kDa	kilodalton
LB	Luria broth
LiAc	Lithium Acetate
LMNG	Lauryl Maltose Neopentyl Glycol
MCM/	Minichromosome maintenance 2-7
MCS	Multiple cloning site
MFT	Modulation transfer function
Mg^{2+}	Magnesium
MOI	Multiplicity of infection
MRX	Mre11-Rad50-Xrs2
MS	Mass Spectrometry
NER	Nucleotide excision repair
Nipbl	Nipped-B like protein
NMR	Nuclear magnetic resonance
NS	Negative staining
NTD	N-terminal domain
OG	Octyl glucoside
ONC	Overnight culture
ORC	Origin recognition complex
PBS	Phosphate buffer saline
PBS-T	PBS-Tween-20
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
Pf	*Pyrococcus furiosus*
pI	Isoelectric point
PK	Anti-V5 epitope tag
Pol α	DNA polymerase alpha
Pol δ	DNA polymerase delta
Pol ε	DNA polymerase epsilon
PP2A	Protein phosphatase 2A
PTA	Phosphotungstic acid
Py	*Pyrococcus yayanosii*
RBS	Roberts Syndrome
RFC	Replication factor-C

RNA	Ribonucleic acid
RPA	Replication protein A
RPC	Replisome progression complex
RSC	Remodels the structure of chromatin
Rtel1	Regulator of telomere elongation helicase 1
S-phase	Synthesis phase
S. cerevisiae/Sc	*Saccharomyces cerevisiae*
S. pombe/Sp	*Schizosaccharomyces pombe*
Sa	*Sulfolobus acidocaldarius*
Scc	Sister chromatid cohesion protein
SEC	Size exclusion chromatography
SEC-MALS	SEC with Multi-angle light scattering
SF	Superfamily
Sf9	*Spodoptera frugiperda*
Single-band ID	Single-band identification
Smc	Structural maintenance of chromosomes
SNR	Signal-to-noise
ssDNA	Single-stranded DNA
SST	Sodium silicotungstate
St	*Sulfurisphaera tokodaii*
Strep-tag	Double streptavidin tag
T-loop	Telomere loop
Ta	*Thermoplasma acidophilum*
TAD	Topologically-associated domain
TE	Tris/EDTA buffer
TEV	Tobacco etch virus
TFIIH	Transcription factor II H
TTD	Trichothiodystrophy
UA	Uranyl acetate
UV	Ultraviolet
WABS	Warsaw Breakage Syndrome
Wapl	Wings-apart like
WB	Western blot
WHD	Winged helix domain
XL-MS	Crosslinking-mass spectrometry
XP	Xeroderma Pigmentosum
XP/CS	Xeroderma Pigmentosum/Cockayne Syndrome
YP	Yeast peptone

Chapter 1. Introduction

1.1 The cell cycle

The ability to create offspring is one of the essential characteristics of a living organism. Each cell undergoes a cell cycle in which it duplicates all of its genetic material in order to pass it onto its progeny, the daughter cell. DNA gets duplicated in the synthesis phase (S-phase) of the cycle by the replication machinery. This machinery unwinds double-stranded deoxyribonucleic acid (dsDNA) into single strands which serve as templates for synthesis of new DNA. The total length of DNA spans over several meters and must be compressed into the nucleus, an organelle of only a few micrometres large in size. Proteins that shape the chromosomes belong to the structural maintenance of chromosomes (Smc) family. Proteins in this family are responsible for compacting the newly synthesised DNA into sister chromatids, each sister comprised of one template strand and one daughter strand (Burgers and Kunkel, 2017). The Smc complex cohesin is responsible for compaction as well as for sister chromatid cohesion. These sisters must remain together from replication until cell division in mitosis in order to be correctly segregated into the daughter cells (Uhlmann and Nasmyth, 1998). Aberrations in DNA replication, DNA repair and chromosome segregation are a common hallmark seen in cancers. DNA synthesis, cohesion establishment and chromosome compaction must therefore be tightly regulated to ensure successful cell division (Malumbres and Barbacid, 2009). The following sections will first describe the mechanism of sister chromatid cohesion by the cohesin complex, followed by the description of the proteins of the replication fork and how these two processes are linked.

1.2 Smc complexes

The findings that large macromolecular assemblies are responsible for chromosome segregation and condensation came in the late 20th century. A number of genes and their protein products, now known as the Smc proteins, were found to be important for accurate mitotic chromosome segregation across a range of species. Three complexes, named cohesin, condensin and Smc5/6, were identified. Deletion and rescue studies of various cohesin subunits and their mutants have specified its role

in sister chromatid cohesion, whereas the role of condensin was connected to DNA compaction (Strunnikov, Larionov and Koshland, 1993; Lehmann *et al.*, 1995; Guacci, Koshland and Strunnikov, 1997; Michaelis, Ciosk and Nasmyth, 1997; Losada, Hirano and Hirano, 1998; Toth *et al.*, 1999; Sumara *et al.*, 2000; Losada and Hirano, 2001). Cohesin, along with condensin and a third Smc complex, the Smc5/6 complex, belong to the Smc complex family. In addition to their role in DNA metabolism, they are all characterised by their 50nm ring shape created by the long anti-parallel coiled coils of their Smc protein components. The Smc family also contain non-Smc subunits, which interact with the Smc subunits and are essential for complex function, described in more detail in section 1.3 (Haering *et al.*, 2002).

1.2.1 Functions of Smc complexes

Functions of eukaryotic Smc complexes show more diversity than their bacterial counterparts but some of their functions overlap. For example, cohesin and condensin are both involved in chromosome compaction but contribute to it differently. In some cases two Smc complexes can work together as in the case of Smc5/6 and cohesin in damage-induced cohesion.

By isolating mutants which lead to defects in sister chromatid cohesion, four proteins were established as the core protein components of cohesin, namely Smc1, Smc3, Scc1 and Scc3. (Michaelis, Ciosk and Nasmyth, 1997; Toth *et al.*, 1999). Aside from sister chromatid cohesion, cohesin is involved in multiple other processes including DNA repair, chromosome organisation and transcription. Cohesin is loaded at chromosomal arms and at centromeres where it is particularly enriched to ensure sister chromatid cohesion until anaphase and to promote bipolar attachment of sister chromatids to the microtubules of the mitotic spindle (Tanaka *et al.*, 2000).

Localisation to chromosomal arms is most likely important for genome organisation and transcription. In budding yeast, cohesin on chromosome arms was found to localise between genes of converging transcription and subsequently relocated to more permanent locations (Lengronne *et al.*, 2004). In mammalian cells, the distribution of cohesin was found to depend on CCCTC-binding factor (CTCF). High-throughput chromosome capture (Hi-C) experiments revealed that CTCF organises

the mammalian genome into topologically-associated domains (TADs), chromosomal domains characteristic by intradomain contacts prevailing over interdomain interactions, where CTCF creates the “borders” of these domains (Eagen, 2018). CTCF, serving as a restriction for cohesin movement, is believed to be the basis behind cohesin-dependent loop extrusion, important for bringing promoters and enhancers together (Davidson *et al.*, 2016). Deletion of cohesin or the protein complex responsible for its loading, the cohesin loader, eliminates TADs but preserves higher compartmentalisation (Busslinger *et al.*, 2017; Haarhuis *et al.*, 2017; Rao *et al.*, 2017; Schwarzer *et al.*, 2017). Cohesin was also found to be involved in DNA repair through homologous recombination (HR) between sister chromatids, which relies on searching for alike sequences to use as templates for repairing double stranded breaks (DSBs). Cohesin’s localisation to DSBs was observed with both yeast and human complexes and depended on the recruitment to the sites of damage by the cohesin loader (Ström *et al.*, 2004; Bot *et al.*, 2017). This topic is further discussed in section 1.9.1.

The precise role of cohesin in mediating transcription has not been identified yet there is enough evidence to claim that cohesin plays an important role. Genome-wide profiling of mutants of cohesin and its loader complex often found in the cohesinopathy Cornelia de Lange Syndrome (CdLS) were found to cause a dysregulation of gene expression (Liu *et al.*, 2009). The so-called mediator, a coactivator that recruits RNA polymerase II to the core promoter in response to binding of transcription factors and enhancers, was found to physically associate with cohesin at active gene boundaries in embryonic stem cells, independent of CTCF. This was also found to be dependent on the cohesin loader (Kagey *et al.*, 2010). As mentioned above, loop extrusion mediated by cohesin can also contribute to transcription by bridging enhancers and promoters together (Davidson *et al.*, 2016). The transcriptional role can also be linked to cohesin’s DNA repair role. For example, cohesin was recently found to localise to DSBs with the chromatin remodeller PBAF to repress transcription at sites flanking the break, a common phenomenon seen at DSBs (Meisenberg *et al.*, 2019).

Condensin was identified in xenopus egg extracts as a five subunit protein complex responsible for condensation of chromatin into mitotic chromosomes, with two Smc

subunits and three non-Smc subunits (Hirano and Mitchison, 1994; Hirano, Kobayashi and Hirano, 1997). Condensin is a mechanochemical motor that translocates on DNA. It has been proposed that condensin plays an important role in transcription regulation as its localisation depends on transcription machineries and actively transcribed genes (Iwasaki *et al.*, 2015; Sutani *et al.*, 2015; Terakawa *et al.*, 2017). Condensin's translocation was further shown to be associated with the rate of transcription, more specifically by encountering the RNA polymerase II which slows down condensin (Brandão *et al.*, 2019). In higher eukaryotes, condensin exists as two isoforms, condensin I and condensin II. Whilst both are absolutely essential for proper chromosome condensation, their contribution to condensation varies (Ono *et al.*, 2003). Using Hi-C, condensin II was found to contribute to the formation of the helical scaffold of mitotic chromosomes with formed loops winding around the scaffold. Condensin I mediates formation of small "nested" loops within condensin II-mediated loops, which further compact chromatin (Gibcus *et al.*, 2018).

Concurrently with the discovery of cohesin and condensin, the six subunit Smc complex, the Smc5/6 complex, was identified. Smc5/6 complex is comprised of two Smc subunits, Smc5 and Smc6, and four non-Smc subunits, one of which is the DNA repair protein Smc5 (Lehmann *et al.*, 1995; Fousteri and Lehmann, 2000). Like cohesin, Smc5/6 complex can embrace two sister chromatids, important for its function in DNA repair where it promotes sister chromatid recombination (De Piccoli *et al.*, 2006; Kanno, Berta and Sjögren, 2015). It is recruited to replication forks by the actions of its loader, Nse5-Nse6, a process which is inhibited by both replication stress and natural pausing sites (Menolfi *et al.*, 2015). Mutants in the Smc5/6 complex and the Nse5-Nse6 complex both show errors in DNA repair (Bustard *et al.*, 2012). An atypical subunit for Smc complexes, the E3-SUMO ligase Mms21, also known as Nse2 (Andrews *et al.*, 2005; Potts and Yu, 2005; Zhao and Blobel, 2005), found in Smc5/6, is important for DNA repair functions where it is utilised to remove linkages between sister chromatids (Bermúdez-López *et al.*, 2010; Varejão *et al.*, 2018). Mms21 is further responsible for SUMOylation of cohesin, which is an important modification for cohesion upon DNA damage. Cohesin and Smc5/6 were found to co-localise at the replication fork suggesting a collaborative role in DNA maintenance (Almedawar *et al.*, 2012; McAleenan *et al.*, 2012). It is possible that another function of the Smc5/6 complex and its loader lies in linking SUMOylation

events to DNA repair (Bustard, Ball and Cobb, 2016). Nse5 of the Smc5/6 loader itself associates with multiple SUMOylating enzymes which, if deleted, render cells sensitive to DNA damaging agents, and SUMOylation of the Smc5/6 complex decreases in Nse5 mutants. Recently, Smc5/6 has been proposed to function with the Fanconi Anaemia pathway proteins in DNA repair (Rossi *et al.*, 2020).

1.3 The cohesin complex

1.3.1 Cohesin architecture

The architecture of cohesin will be described in the following sections using the yeast nomenclature unless otherwise stated. Proteins of *H. sapiens* will be denoted by the prefix *Hs*, with the exception of the Nipped-B like protein (Nipbl). All protein names are listed in Table 1, excluding meiotic isoforms.

Table 1 Proteins of the cohesin complex and associated establishment factors

Protein Name	*S. Cerevisiae*	*S. pombe*	*D. melanogaster*	*H. sapiens*	*B. subtilis*	*E. coli*
Smc	Smc1	Psm1	Smc1	SMC1A	Smc	MukB
Smc	Smc3	Psm3	Smc3	SMC3	Smc	MukB
Kleisin	Scc1	Rad21	Rad21	SCC1	ScpA	MukF
HEAT (or *WHD)	Scc3	Psc3	SA	SA1, SA2	ScpB*	MukE*
HEAT	Pds5	Pds5	Pds5	PDS5A, PDS5B		
HEAT	Wapl	Wpl1	Wapl	WAPL		
Loader	Scc2	Mis4	Nipped-B	NIPBL		
	Scc4	Ssl4	Mau-2	MAU2		
Acetyltransferase	Eco1	Cut1	Eco	ESCO1, ESCO2		

1.3.1.1 Smc proteins

Currently, three eukaryotic Smc complexes made of six distinct Smc subunits have been identified: the Smc5 and Smc6 proteins of the Smc5/6 complex, Smc2 and

Smc4 of the condensin complex and Smc1 and Smc3 of the cohesin complex (Figure 1.1). The bacterial counterparts for genome organisation is Smc/ScpAB and MukBEF with MukB being architecturally similar to Smc proteins (Cobbe and Heck, 2004). The similarity of prokaryotic and eukaryotic Smc complexes is striking, with Smc subunits being highly conserved, but there are differences found. For instance, whereas eukaryotic Smc complexes are formed of Smc protein heterodimers, bacterial MukBEF and Smc/ScpAB only contains one Smc protein, MukB and Smc, respectively, which function as homodimers (Nasmyth and Haering, 2005).

Smc proteins are characteristic by their long intramolecular anti-parallel coiled coils forming a ring structure. By folding upon itself, the Smc protein creates a stable dimerisation interface called the hinge on one end, and a second dimerisation interface on the distal end. The latter interface is formed by the N and C termini of the Smc protein to form the ATP-binding cassette (ABC) type ATPase. One Smc head dimerises with the Smc head of the second Smc protein in the complex, where both heads possess a Walker A and Walker B motif for the binding and hydrolysis of two adenosine triphosphate (ATP) molecules sandwiched between the heads. Similarly, the hinge domain is created by dimerisation of two Smc proteins (Saitoh *et al.*, 1994; Losada, Hirano and Hirano, 1998; Haering *et al.*, 2002). Both cohesin and condensin showed dimerisation of the Smc heterodimer at the hinge in rotary shadowing and electron microscopy (EM) experiments. Both complexes were also found to form closed rod-shaped and open V-shaped conformations while always remaining bound at the hinge. In the rod-shaped conformation the coiled coils are juxtaposed and ATPase heads engaged, unlike in the open conformation where the heads do not dimerise and the coils open up (Melby *et al.*, 1998; Anderson *et al.*, 2002).

1.3.1.2 The hinge domain

The hinge domain is created by a dimerisation between two intramolecular Smc subunits with a pseudo-2-fold symmetry both in bacteria and higher eukaryotes (Haering *et al.*, 2002; Kurze *et al.*, 2011). The dimerisation creates a lysine-rich basic channel. In its narrowest diameter of only 5Å no DNA can however be

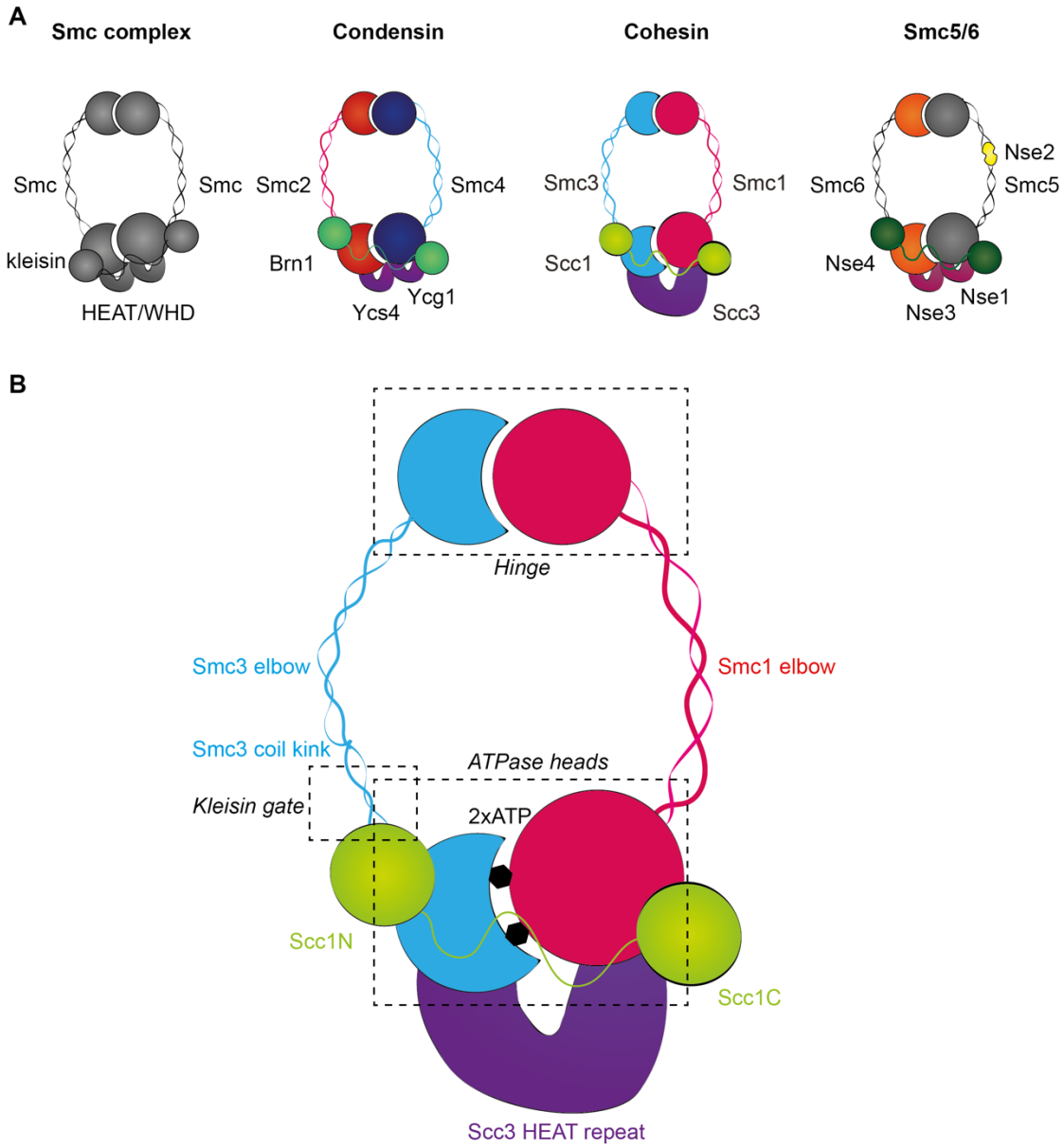

Figure 1.1 The architecture of eukaryotic Smc complexes

A All three complexes share an overall ring shaped architecture formed by their Smc subunits. Compared to condensin and cohesin, the Smc5/6 complex contains WHD subunits which envelop its kleisin subunit and an E3 ligase subunit. **B** Interfaces and regions of cohesin important for its function. ATP is represented as black hexagons.

accommodated. This channel is also present in condensin's Smc2/Smc4 hinge and the hinge of Smc5/6 complex but not in MukB hinge of the bacterial MukBEF Smc complex (Griese, Witte and Hopfner, 2010; Ku *et al.*, 2010; Li *et al.*, 2010; Alt *et al.*, 2017). The positively charged residues found at the three eukaryotic hinges are important for DNA binding as the hinge domain of cohesin is able to bind both dsDNA and single-stranded DNA (ssDNA), the latter with higher affinity (Hirano and Hirano, 2006). Similar is seen with the condensin hinge and Smc5/6 hinge (Griese, Witte and Hopfner, 2010; Alt *et al.*, 2017). In contrast, MukB hinge which does not contain this positive patch could not bind DNA (Li *et al.*, 2010). This suggests diversity between bacterial and eukaryotic Smc complexes. Cohesin can adopt open arm and closed arm conformations. In the closed arm conformation, the coiled coils of eukaryotic Smc proteins are juxtaposed closely together. This most likely obstructs the DNA binding patch and prevents DNA binding as demonstrated with the *Bacillus subtilis* (*Bs*) hinge construct with long coils which only weakly bind DNA, shortening the coils increasing the affinity to sub-micromolar range. Importantly, extensively shortening the coils led to loss of DNA binding suggesting that the coils play an important role in bacterial Smc complexes, whereas in eukaryotic Smc complexes even long coils do not interfere with DNA binding (Soh *et al.*, 2015).

1.3.1.3 Full-length Smc protein visualisation

Solving the structure of the MukB hinge with a 100Å of adjacent coiled coil has revealed that the coil contains a kink at a highly conserved proline (Figure 1.1-B), which introduces a shift in the coil axis by 15° (Li *et al.*, 2010). A similar phenomenon was seen with the *Pyrococcus furiosus* (*Pf*) Smc hinge and coils(Soh *et al.*, 2015). Additionally, the crystal structure of larger segments of the coil revealed that a region proximal to the ATPase heads contains a joint in its C-terminal coiled coils where three helix-loop repeats wrap around a long N-terminal helix (Diebold-Durand *et al.*, 2017). This region is strictly conserved among bacterial Smc and eukaryotic Smc3 of cohesin, suggesting that the joint is crucial for function. Reconstructions of the full Smc structure using sequences of the *Pf* hinge and coils, *Pyrococcus yayanosii* (*Py*) and *Bs* coiled coils and *Py* Smc heads with adjacent coils and joint showed that the Smc proteins form a rod where the N- and C-terminal helices interact 7 times from hinge to head. The 7th contact is the joint region which introduces an 11° tilt and

results in alignment of the Smc heads. This revealed a second dimerisation state of the heads, one in a rod conformation where no ATP is present and second where the heads are slightly tilted and ATP is bound (Diebold-Durand *et al.*, 2017). Indeed, a Cryo-EM structure of the ATPase heads bound to the sister chromatid cohesion protein (Scc) Scc1 were shown to have an asymmetric organisation and varying surface areas available for ATP hydrolysis (Muir *et al.*, 2020). Furthermore, the length of the coils follows a pattern. The coils can be truncated or extended if this pattern is preserved but any disruptions to it disrupt the ATP-dependent localisation (Bürmann *et al.*, 2017)

An important observation came from direct visualisation of MukBEF and budding yeast cohesin using negative staining (NS) EM. Both complexes were able to fold onto themselves, bringing the hinge domains close to the ATPase heads. This bend is facilitated by the “elbow”, a segment close to the centre of the Smc protein. Structure predictions showed that the elbow as well as the joint are present in all Smc proteins of Smc complexes and agree with the crystal structure of the MukB elbow (Bürmann *et al.*, 2019). Other kinks and structural features of the coil emergence are proposed to collaborate on conformational changes of the Smc complexes. For instance, angles at which the coiled coils emerge from MukB hinge are uneven, creating an asymmetric shape of the coiled coil ring. Along with kinks found in the MukB hinge crystal structure, this could aid in opening or folding the ring (Li *et al.*, 2010; Soh *et al.*, 2015).

The DNA repair complex Mre11-Rad50-Xrs2 (MRX) has a similar molecular composition to the Smc complexes, where the Rad50 subunit forms long intramolecular coiled coils similar to Smc proteins. Rad50 forms a Zinc hook on one end and two ATPase heads on the other. Recent publications have shown that the coiled coils of Rad50 form a rod near its hinge domain and upon DNA binding the coiled coils proximal to the ATPase heads engage and clamp around the DNA further forcing the coils into a rod shape (Park *et al.*, 2017; Käshammer *et al.*, 2019). A rod shape functional state was also observed in cohesin by single molecule experiments where cohesin cannot bypass obstacles which it could overcome if it formed an open ring (Stigler *et al.*, 2016). Recently a rod conformation was also observed by *in vivo* crosslinking upon ATP hydrolysis (Chapard *et al.*, 2019). Taken together, the DNA

binding complexes with such long coiled coils could share a common mechanism of substrate engagement and translocation.

1.3.1.4 Scc1

A heterodimer is however not enough for DNA entrapment. In order to capture DNA, cohesin and all known Smc complexes must form a tripartite ring comprised of two Smc subunits and a kleisin subunit, which acts as an interaction platform for accessory proteins which bind to Smc complexes. Kleisins are proteins which associate with Smc complexes and bridge the two Smc proteins. Cohesin contains an α kleisin subunit Scc1, whereas condensin for example contains a γ kleisin Brn1, or in the case of higher eukaryotes either a γ kleisin or a β kleisin in the two isoforms of condensin, condensin I and condensin II, respectively (Nasmyth and Haering, 2005).

The C-terminal domain of Scc1 (Scc1C) forms a winged helix domain that interacts with two C-terminal β strands of the Smc1 head (Gruber, Haering and Nasmyth, 2003; Haering *et al.*, 2004). The N terminus of Scc1 (Scc1N) folds into three helices where two of them form a four helix bundle with the coiled coil adjacent to the Smc3 head, but not the head itself. This rearrangement is interesting as the Smc3 head structurally closely resembles the Smc1 head (Gligoris *et al.*, 2014). Information on this arrangement came from crystal structures of a single Smc head bound to a segment of Scc1, but no crystal structures contained both Smc heads and Scc1, which would show whether their interaction is indeed supported by Scc1. It was not until experiments using rotary shadowing EM which have confirmed that Scc1 physically holds the two heads together in an asymmetric manner to ensure a closed cohesin ring, which is upon cleavage of Scc1 for sister chromatid release opened with the heads no longer associated together (Huis In 'T Veld *et al.*, 2014).

1.3.1.5 HEAT repeat proteins

Scc1 is further bound to a Huntingtin, elongation factor 3, PP2A subunit and TOR1 (HEAT) repeat protein Scc3 implicated in loading and unloading of cohesin from DNA. Scc3 is a hook-shaped protein composed exclusively of antiparallel α helices that makes extensive and often highly conserved contacts with the central region of

Scc1 (Hara *et al.*, 2014; Roig *et al.*, 2014). Scc3 binds to many proteins associated with cohesin necessary for its function. In vertebrates Scc3 exists as two isoforms, SA1 and SA2. They are structurally highly similar but were found to have distinct roles in chromosome compaction. SA1 was found to co-localise more with CTCF and its loss led to increases in long-range contacts whereas SA2 appeared to be more important for intra-TAD contacts. Interestingly, SA2 can take over SA1's function upon SA1 depletion but not the other way around (Kojic *et al.*, 2018).

Similar to Scc3, another HEAT-repeat protein Pds5 was found to bind to Scc1 in an alike fashion. The binding sites for Scc3 and Pds5 on Scc1 do not overlap yet both interact with the central region of Scc1 (Sumara *et al.*, 2000; Lee *et al.*, 2016; Muir *et al.*, 2016). Pds5 interaction with Scc1 was found to be essential for sister chromatid cohesion and Smc3 acetylation, a modification of the centromeric population of cohesin which only gets removed by proteolytic cleavage(Chan *et al.*, 2013). Interestingly, the site of interaction between Pds5 and Scc1 lies only a few amino acids away from Scc1's interaction site with Smc3. Because previous experiments have shown that sealing this interaction prevents the removal of cohesin by wings-apart like protein (Wapl), an accessory protein that is responsible for unloading (described below), it is possible that Pds5 binding helps remove cohesin by interactions with Wapl, as well as promoting cohesion establishment with establishment of cohesion protein 1. (Eco1), an acetyltransferase that acetylates the Smc3 head, described in section 1.4.2 (Chan *et al.*, 2012; Vaur *et al.*, 2012).

Wapl was found to remove cohesin from chromosomes in both yeast and higher eukaryotes, its depletion leading to inability of cells to resolve sister chromatids and increased levels of cohesin in prophase. Wapl is a HEAT repeat protein which forms a stable subcomplex with Pds5, and with cohesin through a shared interaction surface of Scc1 and Scc3. The interaction of Wapl with Pds5 is thought to be cohesin-dependent as depleting Scc1 results in lower levels of Pds5 associating with Wapl (Gandhi, Gillespie and Hirano, 2006; Kueng *et al.*, 2006). Wapl can be structurally divided into an elongated unstructured N terminus and a HEAT repeat C terminus. The C terminus was found to be essential for binding to cohesin, but insufficient on its own. The unfolded N terminus contacts both cohesin, specifically the Scc1/Scc3 subunits, and Pds5. Whereas binding of Wapl toScc1N requires Pds5 but not Scc3,

binding to Scc1C requires Scc3 but not Pds5. (Chatterjee *et al.*, 2013; Ouyang *et al.*, 2013). Importantly, binding of Wapl to Pds5 occurs through a conserved motif on the N terminus of Wapl, YSR, which is also found on sororin, a protein which protects cohesin from Wapl-dependent removal. Wapl and sororin compete for binding to Pds5, and both need cohesin to associate with Pds5 *in vivo* (Shintomi and Hirano, 2009; Ouyang *et al.*, 2016).

1.3.1.6 The cohesin loader

The cohesin loader is the most essential loading factor that interacts directly with cohesin to bridge it with DNA via other factors found on chromatin (Lopez-Serra *et al.*, 2014). The loader is composed of two proteins, Scc2 and Scc4. These proteins are not constitutive components of cohesin but are essential for cohesin loading on DNA (Toth *et al.*, 1999; Ciosk *et al.*, 2000). Cohesin gets loaded to sites unoccupied by nucleosomes as a consequence of chromatin remodelling by remodels the structure of chromatin (RSC) complex. The Scc2-Scc4 loader itself cannot bind DNA, but does so via other factors. In budding yeast, RSC not only remodels chromatin but also acts as a receptor for the cohesin loader (Lopez-Serra *et al.*, 2014; Muñoz *et al.*, 2019).

Structural analysis revealed that Scc4 is a tetratricopeptide repeat (TRP) superhelix with a central hydrophobic cavity that binds to the N terminus of Scc2 (Scc2N). Without Scc4, the Scc2N is likely disordered but adopts a folded conformation upon interaction (Chao *et al.*, 2015; Hinshaw *et al.*, 2015). Scc4 subunit contains a conserved patch that is involved in loading of cohesin to centromeric regions. This patch interacts with the phosphorylated Ctf19 protein of the yeast Ctf19 complex. Ctf19 gets phosphorylated by the Dbf4-dependent kinase (DDK) which is recruited to the kinetochores by the Ctf19 complex to recruit Scc2. Kinetochores, large protein complexes important for spindle microtubule attachment, are located on centromeres and in this way facilitate centromeric but not chromosome arm cohesion (Hinshaw *et al.*, 2017).

Association with chromosomes *in vivo* requires the full-length cohesin loader but *in vitro* the C terminus of Scc2 (Scc2C) is sufficient for loading activity (Murayama and

Uhlmann, 2014). Scc2 forms multiple contacts around the circumference of the cohesin ring. Interestingly, the region of Scc1N that binds Pds5 was found to also interact with Scc2 with the two interacting regions on Scc1 overlapping (Kikuchi *et al.*, 2016). Scc2 is a HEAT repeat protein with a similar hook-shaped architecture to Scc3 and Pds5. Its extreme C terminus has a globular domain (GD) 2 which is followed by 14 HEAT repeats which give the protein its hook shape. Two more GDs follow the hook-shaped body, namely GD1 and GD0. GD0 is proximal to the Scc2N module closely associated with Scc4. The hook-shape of the protein brings the GD2 into closer proximity with the GD1 (Kikuchi *et al.*, 2016; Chao *et al.*, 2017).

1.4 The cohesion cycle

1.4.1 Cohesin loading and turnover

Cohesin must be topologically loaded onto DNA in order for sister chromatid entrapment to occur. Synthesis of Scc1 and subsequent loading of cohesin by the cohesin loader occurs in the gap-1 phase (G1), potentially mediated by DDK activity for the attachment to the kinetochores, and by the RSC protein complex for attachment to nucleosome-free regions (Hinshaw *et al.*, 2017; Muñoz *et al.*, 2019). Pds5 and Wapl are responsible for the dynamic turnover of cohesin on DNA in G1 until a pool of cohesin gets acetylated and remains locked on chromatin until the anaphase step of mitosis (Gerlich *et al.*, 2006; Lopez-serra *et al.*, 2013). Pds5 and Scc2 bind to the same region of Scc1N, by which they could compete for binding to cohesin and therefore a shift towards unloading and loading, respectively (Kikuchi *et al.*, 2016). The cycle is summarised in Figure 1.2.

1.4.2 Establishment

Cohesin gets dynamically loaded and removed by Scc2-Scc4 or Pds5/Wapl subcomplex before DNA replication. In S phase, cohesin gets locked on DNA by acetylation and this cohesin remains bound to DNA until mitosis. The acetyltransferase Eco1 has been characterised as a crucial cohesion establishment factors (Toth *et al.*, 1999; Ivanov *et al.*, 2002). Eco1 acetylates the Smc3 head during replication on two conserved lysine residues, first at K112 followed by acetylation of K113 in yeast (K105 and K106 in vertebrates). Acetylation is required for long-term cohesion establishment but its absence does not perturb cohesin's association with

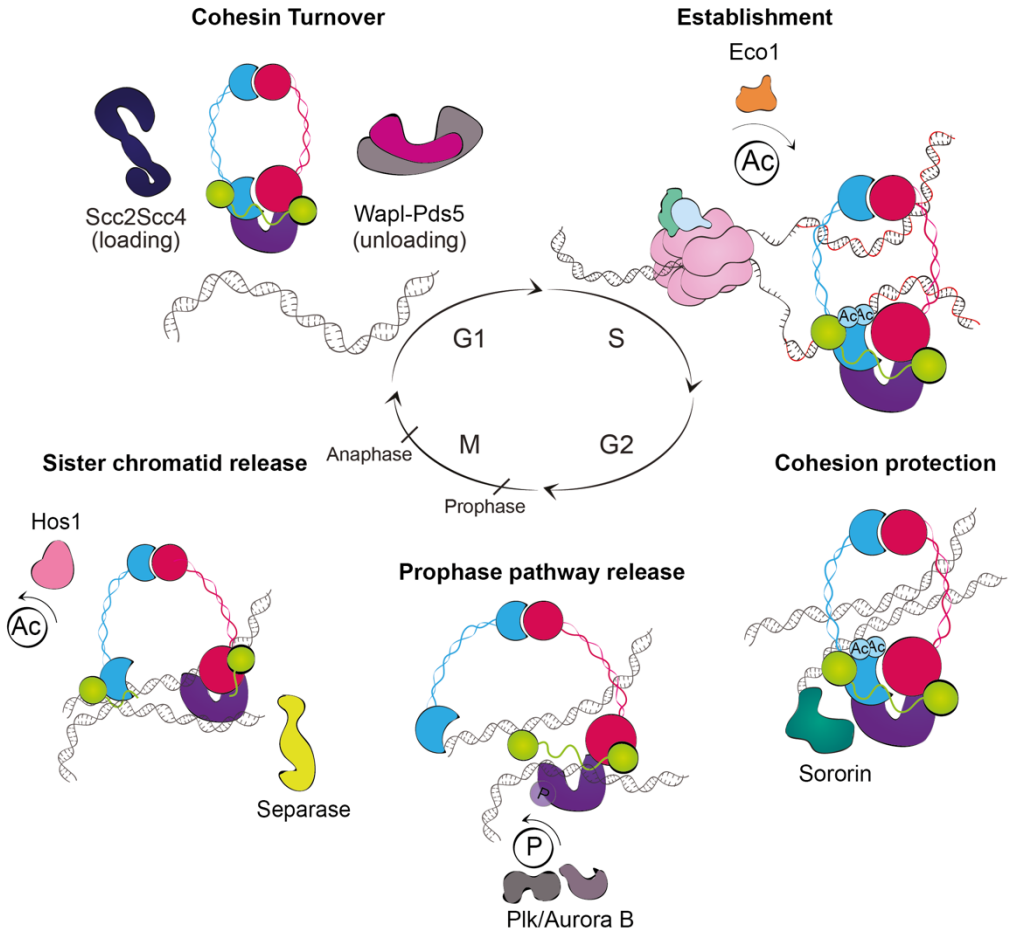

Figure 1.2 The cohesion cycle.

Dynamic turnover in G1 is governed by the loading activity of the Scc2-Scc4 cohesin loader and the unloading activity of Wapl-Pds5 subcomplex. In S phase, Eco1 acetylates the Smc3 head to establish sister chromatid cohesion, which is protected by sororin until mitosis. Cohesin is removed in anaphase either by the prophase pathway involving mitotic kinases, or proteolytically cleaved by separase. This is followed by cell division.

DNA (Ben-Shahar *et al.*, 2008; Unal *et al.*, 2008; Rowland *et al.*, 2009; Chao, Wade, *et al.*, 2017). Acetylation was found to be negligible for the ATPase activity of human cohesin, but conversely, ATP hydrolysis is essential to obtain acetylation of the Smc3's lysine residues. ATPase mutants are able to associate with DNA but do not remain bound potentially as a consequence of acetylation defects in such mutants (Ladurner *et al.*, 2014).

Sororin has been identified as a vertebral-specific protein that when overexpressed increases the association of cohesin with metaphase chromosomes, leading to failure of chromosome segregation in anaphase. Sororin associates with Pds5 to protect acetylated cohesin from removal for enduring cohesion of sister chromatids (Nishiyama *et al.*, 2010). ESCO2, one of the two vertebral orthologs of yeast Eco1, must also be present in order for sororin to get loaded (Lafont, Song and Rankin, 2010). Other factors contributing to cohesion protection is shugoshin which was found to protect cohesin by associating with protein phosphatase 2A (PP2A) and keeping cohesin and sororin in a hypophosphorylated state. This association is mediated by competing with Wapl for binding to Scc3 (Kitajima *et al.*, 2006; Liu, Rankin and Yu, 2012; Hara *et al.*, 2014).

1.4.3 Cohesin removal in anaphase

Two general mechanisms are responsible for removing mitotic cohesin from DNA in vertebrates: proteolytic cleavage in anaphase and the so-called prophase pathway. The larger cohesin pool is dynamically removed by the actions of Wapl in the prophase pathway, whereas a smaller, more stably bound cohesin population cannot be removed by the prophase pathway and is instead removed by proteolysis. In yeast, most of cohesin is removed by proteolytic cleavage, which forms only a minor proportion in vertebrates (Sumara *et al.*, 2000).

In metaphase-to-anaphase transition when sister chromatids are aligned on the metaphase plate, mitotic regulators Cdc20 and Cdh1 signal the activation of the E3 ubiquitin ligase anaphase-promoting complex/cyclosome (APC/C) which in turn targets the proteins sororin and securin for degradation (Rankin, Ayad and Kirschner, 2005; Qiao *et al.*, 2016). Securin is responsible for inhibition of the protease

separase, which is upon degradation of securin released from inhibition and cleaves the Scc1 subunit of acetylated hypophosphorylated cohesin (Uhlmann, Lottspeich and Nasmyth, 1999; Uhlmann *et al.*, 2000; Boland *et al.*, 2017). This results in the release of sister chromatids and their separation to opposite poles of the cell by the mitotic spindle microtubules. After separase cleaves Scc1 and cohesin dissociates from chromosomes, the protein Hos1 of the class I histone deacetylase family is responsible for deacetylating Smc3. This process is dependent on the proteolytic cleavage and does not occur whilst cohesin is still bound to chromosomes, but is essential for *de novo* cohesion establishment in the next S phase (Beckouët *et al.*, 2010; Borges *et al.*, 2010).

1.4.4 The prophase pathway

In vertebrates, the majority of cohesin is removed prior to anaphase by the prophase pathway (Waizenegger *et al.*, 2000). This pathway is independent of proteolytic cleavage of Scc1. Instead, it is initiated by Aurora-B and polo-like kinase (Plk) mitotic kinases phosphorylating the Scc3 subunit of cohesin and sororin (Sumara *et al.*, 2002; Hauf *et al.*, 2005). The pathway was found to be mediated by Wapl, as its depletion leads to prolonged residence time of cohesin on DNA (Gandhi, Gillespie and Hirano, 2006).

In cells depleted of shugoshin or Eco1 the premature separation observed can be rescued by depleting Wapl, suggesting that these two pathways act in opposite manners (Shintomi and Hirano, 2009). Similarly, sororin was found to be important in the presence of Wapl, indicating that an important function of sororin is to compete with Wapl (Nishiyama *et al.*, 2010). This was later confirmed by the structural characterisation of Pds5 showing both sororin and Wapl compete for binding to Pds5 (Ouyang *et al.*, 2016). Interestingly, acetylated cohesin is also reported to be removed by the prophase pathway, suggesting that acetylation as well as protection by sororin and other proteins is essential to maintain cohesion until anaphase (Uhlmann, 2016).

1.5 DNA exit gate

Release of DNA from the cohesin ring has been well characterised. Exit occurs through the kleisin gate, or exit gate, located between the Smc3 coiled coil adjacent to the head and Scc1N. In nonacetylated cohesin, or in conditions of compromised acetylation, Wapl is responsible for opening this gate and releasing cohesin from DNA. Identifying the exit gate came from experiments where potential exit points were sealed by crosslinking. Fusing the Smc3/Scc1N interface, but not Scc1/Smc1, extended cohesin's residence on chromosomes even in the absence of Eco1 (Chan *et al.*, 2012; Buheitel and Stemmann, 2013). Rotary shadowing and EM of human cohesin have shown that this gate indeed opens the cohesin ring and causes DNA release. Mutating residues at the Smc3/Scc1N interface showed open cohesin rings unable to close, leading to the loss of sister chromatid cohesion as the rings cannot close their exit gates (Huis In 'T Veld *et al.*, 2014). Biochemical experiments focusing on topologically loaded cohesin have further shown that Wapl opened the kleisin gate, a mechanism dependent on its association with Pds5. This process requires binding of ATP but not its hydrolysis, as unloading also occurred in the presence of ATPγS (Murayama and Uhlmann, 2015).

In vivo mutations which impair ATP hydrolysis but not ATP binding on the Smc1 head but not the Smc3 head contribute to cohesin unloading. These mutants can bypass the need of Eco1 for being locked on DNA (Elbatsh *et al.*, 2016). This functional asymmetry of the ATPase heads was found to stabilise the Scc1N which interacts with the Smc3 head's adjacent coil, suggesting that Smc1 head also has effects on the Smc3/Scc1 interface which gets disengaged during Wapl-dependent cohesin unloading (Beckouët *et al.*, 2016). Moreover, ATP hydrolysis was found to be a two-step operation. In the presence of Scc1, Smc1 hydrolysis is believed to trigger the hydrolysis at the Smc3 site, followed by opening of the ATPase heads (Marcos-Alcalde *et al.*, 2017).

Solving the structure of the ATPase heads bound by Scc1 has shown that a conformational change accompanying ATP binding remodels the interaction between the Scc1N and the Smc3 coiled coil near the head; the kleisin gate. Introduction of a nucleotide causes a rotational change of the coiled coils of both

Smc proteins towards each other. Because of a rotation in the coiled coil, the four helix bundle interaction found at the kleisin gate is abolished (Muir *et al.*, 2020). Akin to cohesin, studying the interactions of condensin's ATPase heads has shown that head engagement is a stepwise process, where first the Smc4 head must first bind ATP which allows dimerisation with the Smc2 head. Compromising the hydrolysis function of the Smc2 head has no effect on dimerisation with Smc4 head, but if Smc4 head is mutated, this completely abolishes dimerisation (Hassler *et al.*, 2019). In the Smc1/3 head interface, the surface of the Smc1 head contributing to the ATPase site is substantially larger, and has a similar effect; its destabilisation leads to dissociation of the kleisin gate, whereas the smaller ATPase site on Smc3 not having a drastic effect. This was further supported by experiments in which cells with mutated Smc1 ATPase site can cope with Eco1 deletion (Beckouët *et al.*, 2016; Muir *et al.*, 2020).

1.6 DNA entry and binding

Unlike DNA exit, DNA entry into the ring was a highly debated topic without a clear conclusion. Prior to the submission of this thesis but after concluding the cohesin project, two papers have characterised the DNA entry mechanisms, see Discussion. Findings prior to the publishing of these two papers are summarised in this section (section 1.6)

Two distinct DNA entry sites have been proposed. One describes the entry through the hinge domains and is supported by crosslinking experiments where sealing this interface led to a decrease of cohesin loading (Gruber *et al.*, 2006; Buheitel and Stemmann, 2013). The second theory proposes that DNA enters through the same route as DNA exit, with the stimulation of ATPase heads playing a role in structural rearrangements (Figure 1.3). Cohesin is known to topologically entrap DNA in an ATP-dependent manner. First, Scc2 stimulates ATP hydrolysis by the Smc heads which in turn enhances binding of cohesin to DNA. Amino acid substitutions in the Walker A motif, a mutation which results in an inability to hydrolyse ATP, without affecting its binding, abolish binding to DNA. Binding and stimulation of ATP hydrolysis would suggest that head opening also accompanies DNA entry (Ladurner *et al.*, 2014; Murayama and Uhlmann, 2014). Based on these findings a hypothesis where ATP hydrolysis at the heads causes a conformational change which opens

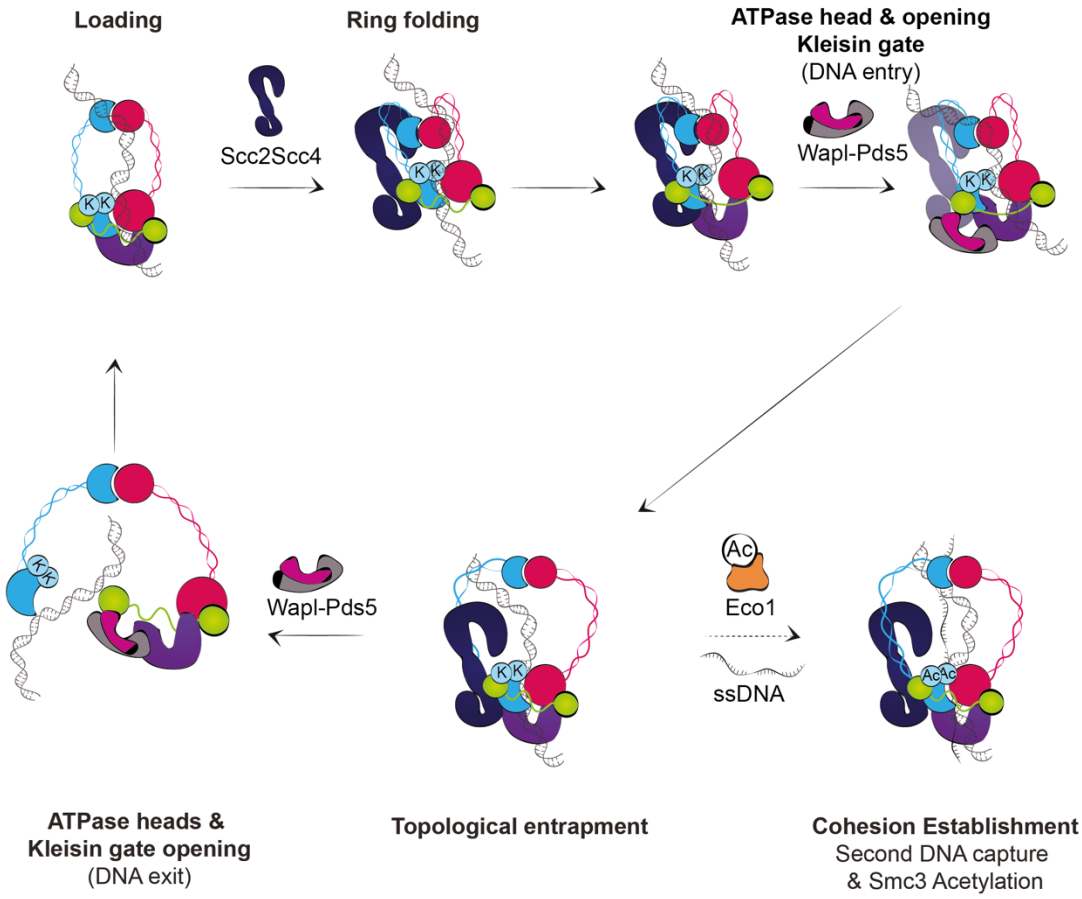

Figure 1.3 DNA entry and exit.

The cohesin ring folds to expose its DNA-sensing lysines upon DNA entry. ATPase head and kleisin gate opening most likely accompany topological DNA entrapment, followed by acetylation for permanent cohesion establishment. The remaining cohesin is removed by Wapl-Pds5.

the hinge as a potential DNA entry gate could not be excluded. The loader's stimulation of ATP hydrolysis is mediated by direct contacts with the Smc heads. Substitutions of these residues does not abolish loading, only the hydrolysis activity, suggesting different contacts are needed for binding and stimulation. Furthermore, the loader is known to bind to Scc1N. Another important factor for loading is the Scc3 subunit, as it was found that cohesin trimers shows significantly less loading onto DNA when compared to a loading reaction with a cohesin tetramer including Scc3 (Murayama and Uhlmann, 2014). The loader very clearly forms extensive contacts with the cohesin complex.

Cohesin is known to bind DNA, but where exactly it resides and at what time is not fully understood. Binding of DNA to the subunits enveloping the ATPase heads were found to occur in all Smc complexes. A recent crystal structure of condensin's Brn1 and Ycg1 subunits with DNA have revealed a "buckle and latch" mechanism where DNA binds to a conserved positively charged patch on Ycs1 and further enveloped by the Brn1 subunit. Importantly, this DNA binding patch was found essential for condensin's binding to chromosomes (Kschonsak *et al.*, 2017). A related observation was seen in the Smc5/6 complex where the winged helix domain (WHD) subunit Nse3 was shown to bind DNA (Zabrady *et al.*, 2016). Similarly in cohesin, the Scc1 subunit and the compartment formed between the ATPase heads and the kleisin subunit was shown to entrap DNA, and was suggested to be responsible for concatenating two sister chromatids rather than the lumen of the ring (Chapard *et al.*, 2019). *In vitro* DNA binding studies have revealed that cohesin binds dsDNA and subsequently shifts its preference to ssDNA as a second DNA capture (Murayama *et al.*, 2018). Such binding preference could have relevance in the context of DNA replication of the leading and lagging strand. It is possible that the overall cohesin conformation and DNA position depends on the function cohesin is carrying out. For example, for DNA replication and sister chromatid cohesion DNA must be topologically embraced in the ring lumen but upon translocation cohesin changes into a more juxtaposed conformation. Rod-shape conformation was observed in translocating cohesin where DNA resided in the ATPase head/kleisin compartment, but an open ring with coils further apart allows DNA to reside in the lumen too (Vazquez Nunez, Ruiz Avila and Gruber, 2019). It is possible that the compartment between Scc1/Scc3 and the ATPase heads is a transition step of DNA entry towards

entering the lumen of the ring. Conversely, DNA could also change between the two compartments.

1.6.1 Cohesin folding during DNA loading and entry

The functionality of a folded cohesin ring is currently debated, but it is widely accepted that cohesin undergoes a major conformational change upon loading. Several lines of evidence support this theory. Firstly, the cohesin loader was shown to bind to multiple subunits of the cohesin ring. The GD2 of Scc2C forms contacts with various cohesin domains including the hinge, ATPase heads and Scc1. The GD0 domain of Scc2C, which lies in the proximity of the N terminus, interacted with the DNA exit gate in crosslinking-mass spectrometry (XL-MS). Only upon a large conformational change, as is proposed for cohesin to undergo upon DNA loading (Figure 1.3), would these two regions interact. This was supported by glycerol gradients which have proved the interaction of cohesin with the isolated GD0 domain (Chao *et al.*, 2017). The conformational change theory is further supported by pulldown experiments where the isolated yeast hinge domain was able to interact with Scc3 as well as with the cohesin loader (Murayama and Uhlmann, 2015). The loader measures just under 30nm and is therefore not able to envelop the extended conformation of cohesin which measures about 50nm (Chao *et al.*, 2017).

Conformational changes could also accompany other steps of the cohesion cycle, as similar contacts were observed for Pds5 which was shown to bind to the hinge domain in fluorescence resonance energy transfer (FRET) experiments (McIntyre *et al.*, 2007). Binding of Pds5 was also visualised using EM where Pds5 was shown to form extensive contacts with Scc1 and Scc3, and also both the ATPase heads and hinge domain of the Smc subunits. The construct used for this study lacked the coiled coils of Smc proteins and it therefore remains to be elucidated how Pds5 changes the overall conformation of cohesin (Hons *et al.*, 2016). Pds5 and Scc2 furthermore have overlapping interacting regions on Scc1 and could potentially envelop cohesin in a similar fashion. This is further supported by *in vitro* experiments with purified proteins which show that addition of Pds5 to a loading reaction decreases the ATPase activity of cohesin and reduces the loading of cohesin onto DNA (Murayama and Uhlmann, 2015; Petela *et al.*, 2018). Another such observation came from

studying the related Smc5/6 complex and its loader Nse5-Nse6, which was found to bind at the Smc5/6 hinge(Duan *et al.*, 2009). Although there is no structural information on this loader, Nse6 subunit was predicted to be a HEAT repeat protein like Scc2 (Pebernard *et al.*, 2006). Folding of the ring could therefore be a common mechanism of loading with all Smc complexes.

The notion that cohesin loading, as well as unloading, is greatly reduced *in vitro* upon mutating Smc3 lysines which get acetylated raised the question how do these lysines contribute to DNA entry? The lysines were found to act as DNA sensors which upon contact with DNA enhance ATP hydrolysis. The importance of these lysines for DNA binding is highlighted by observations from the MRX complex where DNA was found to bind to a similar lysine-rich loop in the Rad50 head module (Rojowska *et al.*, 2014; Murayama and Uhlmann, 2015). The Rad50 head module and overall MRX architecture is greatly similar to that of the Smc complexes (Käshammer *et al.*, 2019). The DNA sensing lysines of cohesin are pointed inwards but could potentially become twisted outwards and exposed upon a large conformational change where the hinge and ATPase head domains would be brought close together by the cohesin loader. Later in the cell cycle their acetylation could prevent activating the DNA sensor which would lead to DNA exit, as the lysines point inwards in an unfolded cohesin ring where the DNA resides in the lumen.

The observation that addition of Pds5 and Wapl, but not Pds5 alone, stimulated loading raised the question whether this subcomplex is involved in a distinct loading step to the step stimulated by the cohesin loader, as well as unloading, potentially by similar mechanisms of weaking the kleisin gate (Murayama and Uhlmann, 2015). Pds5/Wapl requires only the presence but not hydrolysis of the nucleotide, unlike the loader complex. Findings that ATP hydrolysis is a two-step procedure that weakens the interaction of Scc1N/Smc3 coil (the kleisin gate) could be connected to the loading reaction. First, the cohesin loader binds cohesin and brings it to close proximity of DNA. Binding of the loader causes ATP hydrolysis and weakens the kleisin gate. Opening of the gate could be facilitated by Pds5-Wapl subcomplex as this complex is known to bind in the vicinity. Because the loader and Pds5-Wapl were found to bind the same region of Scc1N, it is probable that loading is a two-step process where each step requires a different complex. Alternatively, the loader

complex could be responsible for opening the gate itself. Speculatively, the hinge domain's DNA binding capacity could serve a stabilising role where the entering DNA binds the hinge domain to prevent escape during the second DNA capture (McIntyre *et al.*, 2007; Murayama and Uhlmann, 2015; Murayama *et al.*, 2018; Muir *et al.*, 2020).

1.7 Project 1 aims

In order to get a better understanding of the loading mechanism of cohesin, it is important to uncover the precise interactions between this complex and the Scc2-Scc4 cohesin loader. Such information could elucidate contacts at the ATPase heads or the kleisin gate, and disclose structural rearrangements that occur at and around the head module upon Scc2-Scc4 binding. Because of the large size of the cohesin complex and the loader, the aim of this project was to structurally characterise the contacts between these two complexes using Cryo-EM with the ambition to help elucidate the path for DNA entry in the topological binding of cohesin in sister chromatid cohesion.

1.8 DNA replication

All organisms must undergo DNA replication in order to survive. Both prokaryotes and eukaryotes contain a highly conserved enzymatic machinery for this task, although the level of complexity varies depending on species. For example, most bacteria have a circular genome and their replication is initiated from a solitary point, a single origin of replication. Eukaryotic organisms on the other hand contain multiple origins and have more complex replicative machineries (Costa, Hood and Berger, 2013). Preparation for DNA replication occurs in G1 and is followed by DNA synthesis in S phase. Any errors in the DNA structure are then corrected in G2 and the properly synthesised duplicated genome copies are segregated into daughter cells in mitosis. Synthesis of DNA occurs by first unwinding the dsDNA into two strands that serve as templates for new daughter strand synthesis. Synthesis is carried out by replicative polymerases whose responsibility is to add nucleotides to the growing DNA chain and proofread the accuracy of nucleotide matching (Leman and Noguchi, 2013). Because polymerases only synthesise new DNA in the 5'-3' direction, the replication fork is divided into the "leading" and "lagging" strand. The leading strand gets synthesised continuously, and faster contrasting with the lagging strand which is synthesised in discontinuous 100-200 base pair (bp) segments called Okazaki fragments. The Okazaki framents are subsequently ligated to create a continuous strand (Okazaki *et al.*, 1968).

The central player in eukaryotic replication is the minichromosome maintenance 2-7 (MCM2-7, or shortly MCM), a hexameric ATPases associated with diverse cellular activities (AAA+) ATPase, which is loaded onto DNA as an inactive head-to-head double hexamer (Figure 1.4) (Evrin *et al.*, 2009; Remus *et al.*, 2009). Loading of MCM occurs before replication and is dependent on Cdt1, which recruits one MCM hexamer and holds it in an open conformation for DNA loading (Tanaka and Diffley, 2002; Frigola *et al.*, 2017). MCM-Cdt1 is recruited to replication origins marked by the origin recognition complex (ORC) bound to Cell division control protein 6 (Cdc6) (Speck *et al.*, 2005; Fernández-Cid *et al.*, 2013). Subsequently, a second MCM-Cdt1 hexamer is recruited to DNA with Cdt1 dissociating after each hexamer is loaded (Yuan *et al.*, 2017; Miller *et al.*, 2019). The inactive double hexamer is then activated

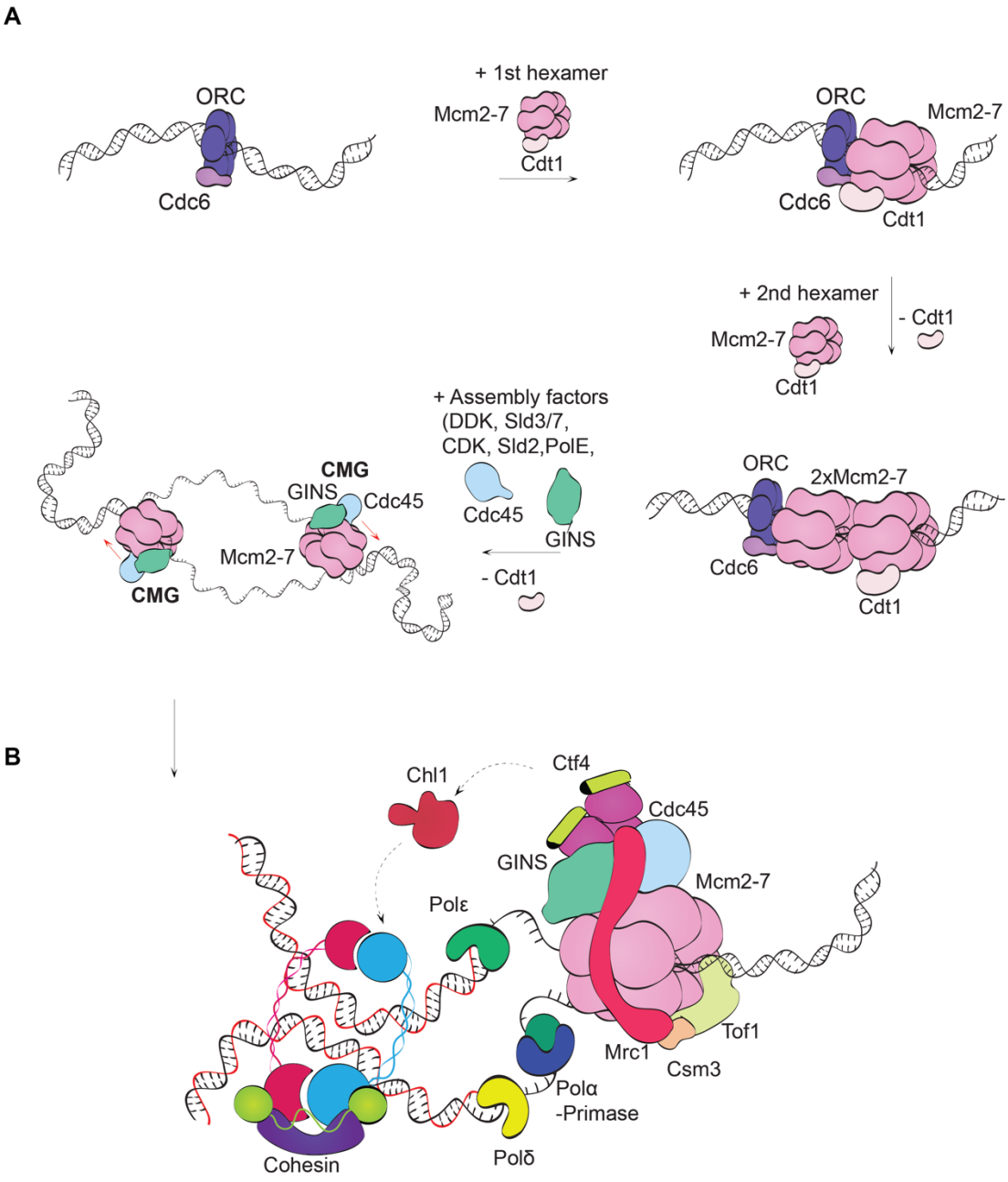

Figure 1.4 A simplified schematic of the replication fork.

A Steps of origin licensing, CMG activation and unwinding of DNA. **B** An assembled replisome and cohesion of newly synthesised sister chromatids. The contacts between cohesin and the replisome are mediated by Chl1 and are denoted in dashed arrows.

by the activities of DDK and the cyclin-dependent kinase (CDK) and other firing factors. The downstream result is the recruitment of Cdc45 and the Go Ichi Ni San (GINS) complex, composed of Sld5, Psf1, Psf2 and Psf3, to create the active CMG helicase (Tercero, 2000; Takayama *et al.*, 2003; Costa *et al.*, 2011; Yeeles *et al.*, 2015). The transition from the inactive MCM-bound DNA to the active CMG-bound DNA is accompanied by structural rearrangements within the MCM subunit that results in DNA untwisting and melting, further mediated by the Mcm10 firing factor (Van Deursen *et al.*, 2012; Abid Ali *et al.*, 2017; Douglas *et al.*, 2018).

DNA unwinding takes place in the 3'-5' direction mediated by CMG sliding along the leading strand to create the ssDNA template strands. The leading and lagging strand polymerases, DNA polymerase epsilon (pol ε) and DNA polymerase delta (pol δ), respectively, synthesise new DNA. Pol ε is incorporated stably into the replisome by forming interactions between its Dpb2 subunit and the GINS subunit Psf1 (Sengupta *et al.*, 2013). DNA polymerase α -primase (pol α-primase), specifically its primase domain, is responsible for synthesis of the ribonucleic acid (RNA) primer of the newly growing strands (Klinge *et al.*, 2009). A conformational switch in the complex is thought to be responsible for handing over the RNA primer to the polymerase for synthesis of the few initial nucleotides (Baranovskiy *et al.*, 2016). On the lagging strand, pol α synthesises only about 30 nucleotides after which synthesis is taken over by pol δ (Núñez-Ramírez *et al.*, 2011). The polymerase eventually reaches the previous Okazaki fragment where it displaces a few nucleotides from the 5' RNA/DNA primer and creates a 5' flap. Two nucleases remove the flaps, namely flap endonuclease 1 (FEN1) and Dna2. Discontinuity in lagging strand synthesis creates stretches of ssDNA, predominantly found at Okazaki fragment boundaries. The fragment boundaries are coated by Replication protein A (RPA) which protect against digestion by nucleases (Wold and Kelly, 1988; Yu *et al.*, 2014). FEN1 only cleaves short flaps uncoated by RPA, whereas Dna2 is responsible for removal of RPA-coated longer flaps (Bae *et al.*, 2001; Rossi and Bambara, 2006). Loading of the polymerases is a stepwise process. Both pol ε and pol δ interact with and are loaded by the trimeric clamp protein proliferating cell nuclear antigen (PCNA). PCNA is loaded by Replication Factor-C (RFC) family clamp loaders where Rfc1-RFC and Ctf18-RFC are needed to load PCNA for pol δ and ε, respectively (Liu *et al.*, 2020). Whilst pol ε is a highly active enzyme, pol δ's activity is greatly enhanced by PCNA

(Ganai, Osterman and Johansson, 2015; Stodola and Burgers, 2016). Interestingly, newly emerging evidence suggests that pol δ is not simply a lagging strand polymerase but is also involved in initiating the leading strand synthesis (Johnson *et al.*, 2015; Yeeles *et al.*, 2017).

1.8.1 The replisome progression complex

Activation of CMG allows DNA to be unwound, but additional factors are necessary to perform DNA synthesis, whether acting as recruiting platforms or actively participating in unwinding. Such factors include the fork protection complex (FPC): a heterotrimeric complex comprised of Mrc1, Tof1 and Csm3; the Ctf4 trimer, serving as an interaction hub of the replisome; facilitates chromatin transactions (FACT) histone chaperone, responsible for histone deposition onto newly synthesised strands; and Topoisomerase I responsible for removing supercoiling after DNA unwinding (Gambus *et al.*, 2006). The addition of these factors allow CMG to reach cellular rates of DNA synthesis *in* vitro (Yeeles *et al.*, 2017). These factors together form the replisome progression complex (RPC), depicted in Figure 1.5.

1.8.1.1 Ctf4

Ctf4 is another component of the RPC, serving as an interaction hub at the replisome for eukaryotic proteins. By assembling into a homotrimer, Ctf4 can simultaneously bind up to three interacting proteins, which are distinctive by sharing a Ctf4-binding peptide (CIP box) for binding to Ctf4's C-terminal domain (CTD). Two types of CIP-boxes exist. Type I-containing CIP-box proteins include Sld5 of GINS, Pol1 of pol α-primase, Dna2 involved in Okazaki fragment processing and HR, and the Chromosome loss 1 protein (Chl1) helicase implicated in sister chromatid cohesion (Gambus *et al.*, 2009; Simon *et al.*, 2014; Samora *et al.*, 2016). Screening for interactors of Ctf4 has further revealed the existence of type II CIP-box proteins which bind to a site on Ctf4 distinct from type I, although the overall architecture of a double-turn helix is present in both. The sequence conservation of the type II box is not as high as that of type I but structure of Ctf4 bound to this type II CIP-box were solved. Proteins identified to contain this motif included Tof2 and Dbp2 of pol ε (Villa *et al.*, 2016). The type I CIP-box amino acid sequence, DDIL, was found to be absolutely essential for the interaction with Ctf4. Amino acids flanking the DDIL motif

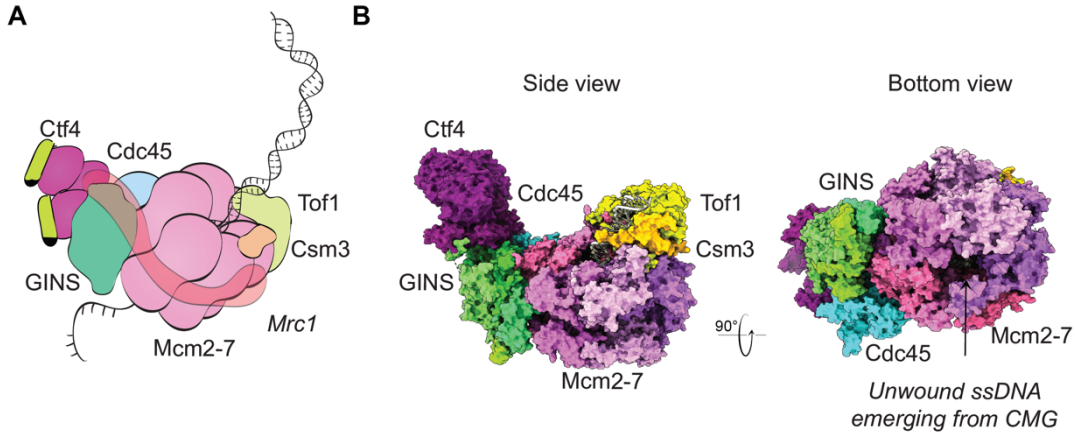

Figure 1.5 A partial replisome progression complex.

A Schematic representation of the CMG helicase bound by the FPC and Ctf4. **B** Cryo-EM structure of CMG unwinding dsDNA. Tof1 and Csm3 of the FPC travel in front of the fork. Ctf4 is bound to the GINS complex away from DNA. The density of Mrc1 is not clearly visible in EM maps. PDB:6SKL.

are variable. Whereas the immediate C-terminal amino acid to the motif can be readily mutated, the two amino acid preceding the motif were found to be an arginine followed by a hydrophobic residue in all proteins but Chl1, which instead has an aspartic acid and a glycine residue. The Chl1 CIP-box is the only type I CIP-box that could not be co-crystallised with Ctf4 (Simon *et al.*, 2014). Interestingly, GINS and Pol1 CIP-boxes have the highest affinity for Ctf4 (5μM and 25μM, respectively), followed by Dna2 (230μM) and Tof2 whose affinity was not measurable by fluorescence anisotropy. Albeit highly similar in sequence, additional contacts between Sld5 and Ctf4, including Sld5's isoleucine 3, could explain the higher affinity of the CIP-box for Ctf4 compared to Pol1. This could explain why Ctf4 is a constitutive component of the RPC through GINS of the CMG helicase. Structural characterisation of the Ctf4 trimer with EM showed the ability of three copies of GINS to occupy the three binding sites on Ctf4, whilst other interacting proteins cannot (Samora *et al.*, 2016; Baretic *et al.*, 2020).

1.8.1.2 Fork protection complex

The Mrc1/Tof1/Csm3 complex is positioned in front of the CMG and therefore forms the first contact with DNA, strengthening the otherwise loose association of MCM with dsDNA (Eickhoff *et al.*, 2019). A recent Cryo-EM structure of the CMG bound to FPC and Ctf4 revealed that gripping of DNA for tighter association is mediated by Tof1 and Csm3, leading to enhanced processivity by CMG (Figure 1.5-B). Tof1 forms extensive contacts with Mcm6, 4 and 7 of MCM, Csm3 in turn mainly contacting Tof1. In this structure the Mrc1 density could not be faithfully assigned as this subunit showed a high degree of flexibility. Supporting XL-MS experiments revealed that Mrc1 contacts multiple subunits around the whole replisome including Tof1 at the front of the fork, Mcm6/Mcm2 in the middle, and Pol ε, Ctf4 and Cdc45 behind the replication fork (Baretic *et al.*, 2020).

1.8.2 The RPC in replication stress

Progression of DNA replication is commonly challenged by the presence of lesions caused by endogenous and exogenous sources, resulting in replication pausing. Additionally, various DNA structures that are not readily resolved, presence of transcription machineries or tightly bound protein:DNA complexes impede the

progression of the fork and cause replication fork stalling. Replisome stalling is accompanied by the presence of ssDNA as a result of uncoupling between the replicative helicase and polymerase. The ssDNA is coated by RPA which not only serves as a mean of protection for the DNA but is also one of the three signals (along with the 9-1-1 and the Rad24-RFC complexes) of DNA damage to the downstream kinase Mec1 and its partner Ddc2 (Zou and Elledge, 2003; Zou, 2013). Mec1 is recruited to the sites of stalled forks and DNA damage and is considered the main S-phase checkpoint kinase. Mec1 acts indirectly on the downstream effector kinase Rad53 via Mrc1 of the FPC (Alcasabas *et al.*, 2001; Osborn and Elledge, 2003). Phosphorylation of Rad53 occurs as a downstream response to the Mec1 kinase S-phase checkpoint activation and results in and serves as a mean of observing replication stalling (Alcasabas *et al.*, 2001). Rad53 in turn phosphorylates Dbf4 of the DDK kinase, which inhibits origin firing and DNA synthesis, and Sld3, a firing factor, which normally regulates MCM loading but in the situation of stress, prevents loading of MCM onto origins that have already fired (Zegerman and Diffley, 2010). The DNA-sensing signals and signalling pathways are present from yeast to humans. A commonly used experimental approach for studying replication stress also activates this response, which involves the treatment of cells with hydroxyurea (HU), inhibition of ribonucleotide reductase and subsequent depletion of dNTP pools and replication stalling (Poli *et al.*, 2012).

The positioning of the FPC in front of the unwinding CMG helicase allows the sensing of stressors as well as mediating the speed of replication. Aside from mediating replication progression Mrc1 is also a part of the S-phase checkpoint which responds to replication stress via the Mec1-Mrc1-Rad53 signalling cascade. Under stress conditions Mrc1 could mediate the slowing down of the fork until the cause of fork pausing is resolved. At stalled replication forks either paused naturally or by stress, Mrc1 is able to uncouple DNA synthesis from the translocation of CMG by interacting both with Cdc45, Mcm6 and pol ε (Lou *et al.*, 2008; Komata *et al.*, 2009; Yeeles *et al.*, 2017). Furthermore, Tof1, which is also needed to mediate the speed of the fork, forms extensive contacts around the double-stranded helix ahead of the CMG, giving the helicase an enhanced grip on the DNA (Yeeles *et al.*, 2017; Baretic *et al.*, 2020). Very importantly, Tof1 is responsible for tightly gripping the DNA before its contacts with CMG, and also senses aberrant DNA structures ahead of the fork. In human

cells the Tof1 ortholog Timeless was found to sense guanine quadruplex (G4) DNAs; complex DNA structures that inhibit DNA synthesis and impede fork progression. G4 DNAs are guanine (G)-rich regions of the genome forming complex secondary structures as a result of guanine forming Hoogsten base pairing with itself to form the ring-like G4. The CMG helicase is unable to resolve complex DNA structures. The RPC must therefore recruit additional factors for the resolution of this DNA. Such factors include Timeless which in turn recruits the DDX11 helicase (human ortholog to Chl1) to resolve DNAs such as G4s (Lerner and Sale, 2019; Lerner *et al.*, 2020). The functions and modulations of the replisome progression complex are vast and remain enigmatic. It appears that the positioning of the FPC ahead of the CMG helicase and its ability to both signal and recruit various factors to the fork is crucial for responses to any potential inhibitions the fork may encounter (Baretic *et al.*, 2020).

1.8.3 The RPC and cohesion establishment

Cohesin gets dynamically loaded and removed from DNA in G1. In S-phase, a distinct mechanism involving cohesin acetylation locks it on DNA to assure that the newly replicated DNA in the form of sister chromatids remain together until cell division (see section 1.2). Some outstanding unresolved questions include; how does cohesion get established with respect to the progressing replisome? Does cohesin allow the replisome to pass through? Does it dissociate and re-associate after the replisome progresses? The cohesin ring is 50nm in diameter and should be able to pass the replisome by letting it in its lumen. Single-molecule experiments have revealed that the lumen is, indeed, larger, but cohesion adopts a more juxtaposed rod-shaped conformation when loaded on DNA. In light of these findings, the replisome would not be able to pass through the lumen (Stigler *et al.*, 2016). Tracking the fate of cohesin with photobleaching reveals that even if cohesin gets loaded during G1, the cohesin molecule remains bound to DNA after replication, suggesting that cohesin does not get loaded *de novo* but persists on DNA (Rhodes *et al.*, 2017). Upon binding a second DNA, cohesin shifts its preference from dsDNA to ssDNA *in vitro*. This could mean that cohesin is present in a stage of replication where full synthesis of two dsDNAs is not yet complete (Murayama *et al.*, 2018). The coordination of cohesin and the replisome is not fully understood. Furthermore,

persistent cohesion establishment by acetylation of the Smc3 subunit occurs after DNA replication but the precise spatiotemporal interactions are also not understood.

Two distinct pathways were identified that contribute to cohesion establishment at the replication fork. In yeast, Ctf18-RFC and the FPC are thought to promote Eco-1 dependent cohesion establishment, whereas the Ctf4-Chl1 axis is believed to work independently of cohesin acetylation (Xu, Boone and Brown, 2007; Borges *et al.*, 2013).

Eco1 is essential for cohesion establishment by acetylating Smc3 during replication, functioning downstream of the clamp loader Ctf18-RFC (Ben-Shahar *et al.*, 2008). Deletion of Ctf18-RFC leads to replication reduction and cohesion defects, conversely to the deletion of the leading strand clamp loader, Rfc1-RFC, which only leads to defects in replication. Ctf18-RFC loads PCNA, which enhances the processivity of the lagging strand polymerase. PCNA also recruits Eco1 and, indeed, deleting Ctf18 shows a decrease in Smc3 acetylation (Moldovan, Pfander and Jentsch, 2006; Liu *et al.*, 2020) The precise order of the recruitment of these factors to the fork as well as their exact interactions are not fully understood.

Studying the Chl1-Ctf4 axis in yeast has shown that deleting either protein results in cohesion defects. The GINS subunit of CMG interacts with Ctf4, which in turn binds Chl1. Chl1 then contacts yet unidentified cohesin subunits (Samora *et al.*, 2016). This pathway is described in more detail below (section 1.9). Intriguingly, Chl1 and Eco1, which are believed to operate in two distinct pathways, were also found to interact (Skibbens, 2004). The cohesion defects observed with Chl1 deletions however seem independent of Eco1 acetylation (Borges *et al.*, 2013). Human Chl1 was further found to interact with PCNA, Fen1 and Ctf18-RFC, lagging strand factors that promote DNA synthesis and cohesion (Farina *et al.*, 2008a).

1.9 The dual role of Chl1

The interactions identified in yeast may not be conserved in humans. For example, unlike in yeast, the interaction between human Tof1 and Chl1 orthologs, Timeless and DDX11, also promote sister chromatid cohesion in addition to mediating a

response to replication fork stalling. Conversely, Ctf4 and Chl1 in yeast were reported to interact in numerous papers(Petronczki *et al.*, 2004; Borges *et al.*, 2013; Samora *et al.*, 2016), but the human counterpart to Ctf4, AND-1, shows binding to the replisome and pol α, but its interactions with DDX11 and human cohesin are not well characterised (Im *et al.*, 2009). The following paragraphs summarise the current knowledge of the Chl1 interactions.

1.9.1 Observations in budding yeast

The interaction between Chl1 and Ctf4 is crucial for cohesion establishment. Ctf4 binds to the Sld5 subunit of GINS of CMG with the highest affinity out of all of its binding partners (Gambus *et al.*, 2009; Simon *et al.*, 2014). One site of the trimeric Ctf4 is constantly occupied by GINS, whereas the other two are available presumably for other interacting partners (see section 1.8.1.1). Two CMG complexes have been reported to bind to one Ctf4 trimer via GINS (Yuan *et al.*, 2019). Chl1 can occupy two sites of Ctf4 but not three (Samora *et al.*, 2016). To achieve correct cohesion establishment the Ctf4-Chl1 axis cannot be disrupted as this leads to lower levels of cohesin associated with the replisome as demonstrated in immunoprecipitation (IP) experiments. Deletion of Chl1 in budding yeast and mutations to the DDIL motif that mediate interaction with Ctf4 result in cohesion defects. Mutations to Chl1's helicase site however do not display any cohesion defects, suggesting that Chl1 does not perform any DNA-unwinding role in sister chromatid cohesion (Samora *et al.*, 2016). Instead, Chl1 together with Ctf4 could promote stable cohesion establishment in a parallel pathway with Eco1 (Ben-Shahar *et al.*, 2008). Deletion of either Chl1 or Ctf4 in S-phase synchronised *S. cerevisiae* results in decreased levels of Smc3 acetylation, comparable to cohesion defects seen with Eco1 deletions. This is suggestive of two pathways for Smc3 acetylation and further corroborated by the observation that a double mutant of Chl1 and Eco1 severely compromises vitality of the cells, suggesting that these proteins work in parallel rather than together to promote cohesion establishment (Borges *et al.*, 2013).

Conversely, mutations causing disruptions in Chl1's helicase activity show impairments under conditions of replication stress. Both deletion of Chl1 or helicase-impairing mutations reduce cohesin levels under HU conditions in S-phase, where

cohesin cannot localise to stalled replication forks. This localisation is not affected by ATPase dead mutants of Chl1, pointing to two distinct mechanisms of Chl1-dependent cohesin loading at the forks under normal and stress conditions (Delamarre *et al.*, 2019). Cohesin localises to DSBs during G2 for subsequent repair by HR, as well as to stalled forks in a distinct mechanism where it gets recruited by MRX (Tittel-Elmer *et al.*, 2012). MRX is responsible for generating short ssDNA gaps at DSBs, which are subsequently elongated by the Exo1 nuclease or the Sgs1-Dna2 helicase-nuclease in the process of fork resection, followed by coating by the protective RPA. MRX further collaborates with chromatin remodelling factors that promote cohesin loading under both normal and stress conditions. In stress conditions, cohesin loading is severely impacted if long stretches of ssDNA for loading are unavailable. It was found that this process also involves Chl1 and by these means Chl1 contributes to fork resection (Delamarre *et al.*, 2019). Interestingly, Dna2, which is also needed for resection is also a Ctf4-binding protein in yeast interacting through the DDIL motif like Chl1, GINS or pol α (Villa *et al.*, 2016). It has been observed that Ctf4 colocalises with Chl1 in conditions of replication stress induced by HU (Samora *et al.*, 2016). Ctf4 could also be colocalising to the sites of damage where both MRX and cohesin and associated proteins localise.

1.9.2 Observations in higher eukaryotes

Like its yeast counterpart, the human Chl1 ortholog DDX11 is also implicated in sister chromatid cohesion and DNA repair linked to DNA replication, albeit with distinctions in its interactions. As opposed to yeast, the roles of DDX11 in cohesion appear to be linked with Timeless, the human ortholog of Tof1 of the FPC, which co-IPs with DDX11 to mediate sister chromatid cohesion. The interaction between Timeless and DDX11 is mediated through a conserved EYE motif present in HD1 just after the Walker A motif of DDX11. A substitution of EYE amino acids to KAK leads to abolishment of binding. This interaction is important for sister chromatid cohesion as DDX11 mutants defective in binding to Timeless are unable to rescue a cohesion-defective phenotype (Cortone *et al.*, 2018). In yeast, this role is mediated by the Ctf4 homotrimer and is helicase activity-independent. This does not appear to be the case which appears to be false for human cells as DNA binding mutants of DDX11 cannot

rescue a sister chromatid cohesion defect in cells with a DDX11 deletion (Samora *et al.*, 2016; Faramarz *et al.*, 2020).

Timeless enhances DDX11's ATPase and helicase activities *in vitro,* where the increase in helicase activity can be attributed to the stimulation of DDX11 by Timeless to boost the interactions with DNA. *In vivo* depletion of DDX11 results in the downregulation of Timeless and vice versa. Interestingly, the interaction between Timeless and DDX11 is strongly enhanced in a condition of replication stress where the addition of HU results in a stronger interaction between the two when co-immunoprecipitated. Deleting either protein impairs fork progression *in vivo* as measured by DNA fiber track assays where the progression retardation is not worsened with a double deletion, suggesting a synergistic effect of the two proteins on fork progression under replication stress (Cali *et al.*, 2016). Deletion of ESCO2, one of the two variants of yeast Eco1 proteins in humans, shows synthetic lethality in cells from Warsaw breakage syndrome (WABS) patients (described in section 1.10.3) that carry a DDX11 mutation (Faramarz *et al.*, 2020). Deletion of DDX11 in HeLa cells results in a compromised genome integrity upon treatment with ultraviolet (UV) light or chemical agents such as cisplatin and methyl methanesulfonate (Shah *et al.*, 2013; Pisani *et al.*, 2019). Its helicase activities are very broad, being able to resolve forked structures, G4 DNA or displacement loops (D-loops) (Cali *et al.*, 2016) . A recent publication has revealed that DDX11's helicase activity is crucial for resolving G4 roadblocks that occur during DNA replication ahead of the replication fork. Timeless was found to be responsible for the sensing of DNA roadblocks and subsequent recruitment of DDX11, but excluded to be involved in the resolution of the complex DNA, which was attributed solely to DDX11 (Lerner *et al.*, 2020).

It is important to note that there might be yet undefined differences between yeast and metazoan pathways. Although both human and yeast Chl1 bind cohesin, a direct association with Tof1 has only been observed for the metazoan protein through the EYE motif. Sequence alignments show the presence of a similar "PYE" motif in *S. cerevisiae* but its functional aspects have not been identified. Likewise, the role of Chl1 in DSB-mediated cohesin recruitment has been observed only in yeast, but it is known that Chl1 plays a role in DNA damage response in both organisms. Structural characterisation of Chl1's closest homolog, XPD, revealed it also exhibits a dual role

where it only utilises its helicase activity in one (described in section 1.10.2). Structural characterisation of this protein has led to a closer understanding of its cellular function. Up to date, no such information exists about Chl1.

1.10 The XPD subfamily of helicases

Chl1 belongs to the XPD subfamily of proteins under the Superfamily 2 (SF2) helicases, one of the two largest families of helicases along with the Superfamily 1 (SF1). SF1 and SF2 helicases are characteristic by containing two RecA-like folds that form the helicase core of one monomer, and share multiple signature motifs which form the nucleotide or DNA binding pockets. The mode of operation for these helicases is the conversion of energy from ATP hydrolysis into conformational changes in the protein. The SF2 helicases process both RNA and DNA duplexes with the majority of subfamilies in SF2 unwinding DNA in a 3'-5' fashion. The XPD subfamily helicases are in this regard an exception to the SF2 family as they unwind DNA in the 5'-3' manner (Singleton, Dillingham and Wigley, 2007). Four proteins belong to the XPD subfamily; XPD, Chl1, Fanconi Anaemia complementation group J (FancJ) and Regulator of telomere elongation helicase 1 (Rtel1) (XPD, DDX11, FANCJ and RTEL1 in human nomenclature).

1.10.1 Structure of the XPD subfamily proteins

Structural characteristics of the SF2 are preserved in the XPD subfamily, including the RecA-like core domains and signature helicase motifs, despite the varying polarity of these proteins to the other members of its superfamily. In addition to its core domains, two other domains are predicted to be present in all XPD proteins but have been structurally characterised in only one member of the subfamily, the XPD protein (Figure 1.6). Adjacent to the first RecA-like fold, or, the helicase domain 1 (HD1) of the protein, XPD proteins are distinctive by their highly conserved iron-sulfur (Fe-S) cluster which is essential for their helicase activity, located in the Fe-S domain (Rudolf *et al.*, 2006). HD1 together with the second RecA-like fold, the helicase domain 2 (HD2), form the body of the protein. The fourth domain, when identified, folded into a novel fold resembling an arch-shaped conformation extended above the body of the protein and was thus named the Arch domain (Fan *et al.*, 2008; Liu *et al.*, 2008). These domains were identified in XPD and predicted for all members.

The helicase domains of these proteins are RecA-like ATPase motors characteristic by a central β sheet surrounded by α helices. The majority of the canonical motifs of the SF2 family helicases can also be found in the XPD protein, located in the two helicase domains. Motifs I, II, III, V and VI as well as the Q motif are responsible for ATP hydrolysis. Motif I, which corresponds to the helicase's Walker A motif, also known as the P-loop, sits in between a β strand and an α helix and has a highly conserved sequence of A/GXXXXGKT/S, with X denoting any amino acid. Conversely, motif II or the Walker B motif, which together with Walker A control the nucleotide binding and hydrolysis, is less conserved. The Walker B motif is incorporated into the "DEAH" box, a canonical motif of many helicases involved in RNA and DNA processing. Walker A motif interacts with the γ-phosphate of the nucleotide whilst Walker B motif binds the magnesium (Mg^{2+}) ion necessary for interaction with the phosphate (Ye *et al.*, 2004). Walker A and B motifs are found across all helicases (Walker *et al.*, 1982). The Q motif was identified as a unique motif for DEAD/DEAH box helicases located upstream of helicase motif I with an invariable glutamate residue contributing to ATP hydrolysis (Tanner *et al.*, 2003). Motif VI contains a highly conserved arginine residue and is thus also named the "arginine finger". It is located in HD2 but extends into the catalytic core of the opposite HD1, serving as a point of communication between the two domains (Scheffzek *et al.*, 1997). The movement of the two helicase domains with respect to one another is a result of the nucleotide hydrolysis, where the nucleotide is positioned between HD1 and HD2. Nucleotide hydrolysis leads to a conformational rearrangement of the HD1 and HD2; their movement and contacts formed between DNA and canonical motifs Ia, IV and the P motif driving the translocation on DNA (Cheng and Wigley, 2018). The first crystal structures of archaeal XPD revealed that the Arch domain forms a novel fold comprised of four α helices and four-stranded antiparallel β sheet. The Fe-S domain is located closer to the N terminus of the protein between motifs I and II. It is a four helix assembly binding to and stabilised by interactions with the four iron ions of the Fe-S cluster via the domain's four cysteines, three of which are absolutely conserved from archaea to eukaryotes. For clarity, only the cysteine residues are highlighted in Figure 1.6 as the Fe-S domain is located within HD1. The direct contact of the cysteines to the iron moieties of the cluster is crucial as mutations of three of these cysteines to serine results in the loss of the cluster. Only

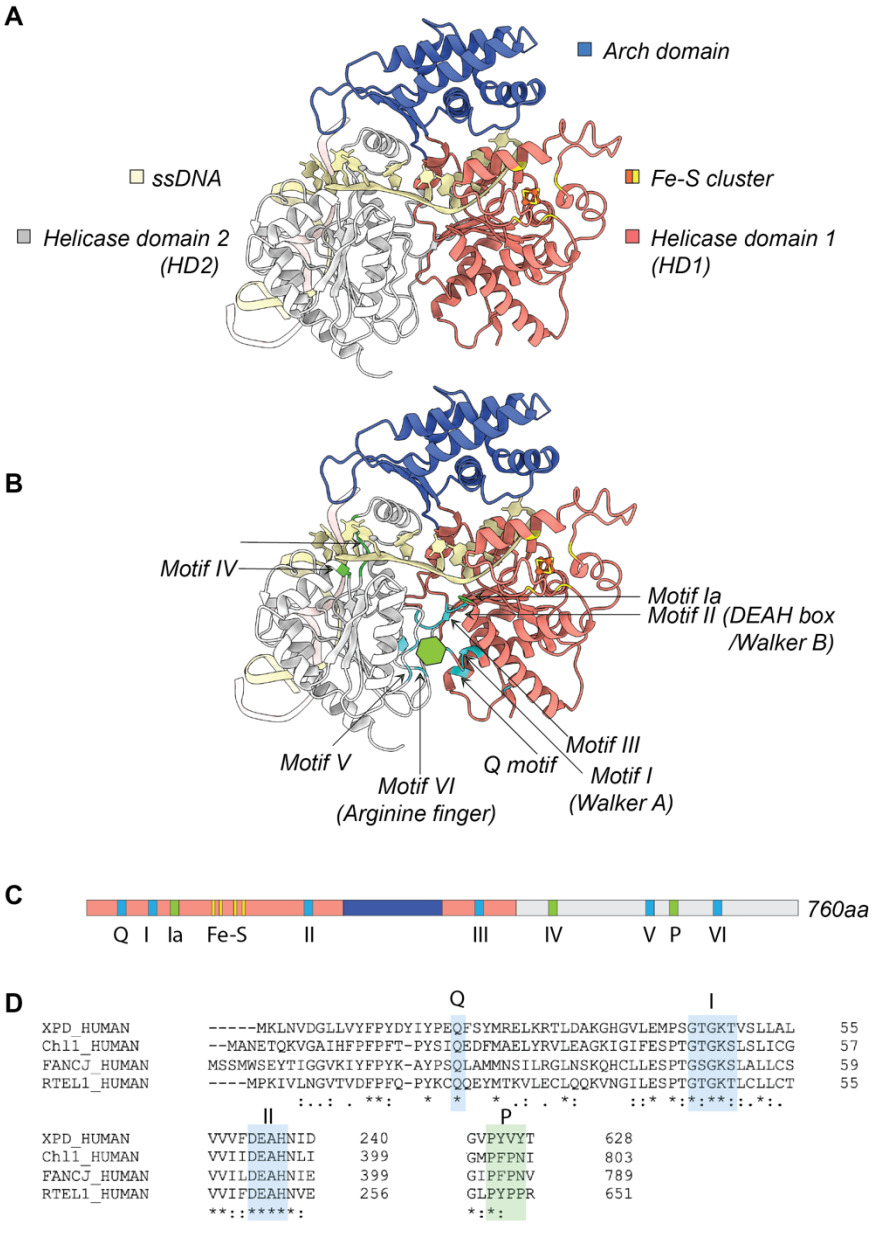

Figure 1.6 Human XPD from the Cryo-EM structure of TFIIH. (PDB 6RO4).
A Domain architecture of XPD. **B** Conserved signature motifs of SF1 and SF2 helicases. Motifs highlighted in cyan are responsible for nucleotide (green hexagon) binding. Motifs responsible for DNA binding are shown in lime. Alternative names of the motifs are in brackets. The schematics in A and B originate from the same structure, the ATP and DNA-bound XPD. **C** Schematic representation of the sequence of XPD and the location of the canonical motifs. **D** An example of sequence conservation of the canonical motifs. Motifs I and II (Walker A and B) are found in all helicases, and together with the adjacent Q motif and other motifs regulate ATP hydrolysis. The P motif is essential for translocating DNA.

a single cysteine residue (C102 in archaeal *Sa* structures) can tolerate this substitution. The loss of the cluster results in the loss of helicase activity, but not ATPase activity, which is stimulated by ssDNA binding. In its unoxidized state the cluster is present as a 4Fe-4S cluster but reduced in the cell to form a 3Fe-4S cluster. Its loss further results in a destabilisation of the Fe-S domain and an overall reduced stability of the protein as shown by the Apo-*Sa*XPD structure obtained by soaking XPD crystals in ferricyanide to remove the cluster (Liu *et al.*, 2008). Destabilisation of the cluster also leads to a somewhat disordered Arch domain, showing a less folded conformation towards the helicase domains and a partially unresolved sequence in the crystal structure (Rudolf *et al.*, 2006; Fan *et al.*, 2008; Liu *et al.*, 2008). The CTD of the XPD subfamily proteins, located in HD2, is one of the most variable regions. XPD and Chl1 proteins are structurally most related and are found in all eukaryotes. Conversely, prokaryotic life forms only contain XPD orthologs. FancJ and Rtel1 have evolved only in higher eukaryotes with their CTDs significantly larger and more specialised in comparison to XPD and Chl1 (Figure 1.7) (Wu, Suhasini and Brosh, 2009). Differences in the CTD can already be found with the same protein across different species. XPD is a part of the transcription factor II H (TFIIH) and interacts with p44 at its CTD. Since archaea do not possess TFIIH, the CTD is shorter and is missing the p44 interacting region (Kokic *et al.*, 2019).

1.10.1.1 Mechanism of DNA unwinding

Although some archaeal XPD crystal structures contained partially resolved DNA, solving the crystal structure of the bacterial ortholog of XPD, DinG, has revealed the operational mode of the helicase as this structure contained visible density of 10 bases of ssDNA (Figure 1.8). DNA binds several residues across the HD1 and HD2

domains. The interaction is mostly being formed by polar residues hydrogen bonding with the backbone of the DNA or by contacts between aromatic residues of HD2 with the bases of DNA forming π–π stacking. DNA is clamped from above by the Arch domain, which forms contacts with the Fe-S cluster to form a positively charged tunnel for DNA positioning and passage. DNA is translocated across the channel in a one ATP per base manner with three bases contacting the HD1 and seven contacting the HD2. The DNA bases are stacked onto one another but are flipped

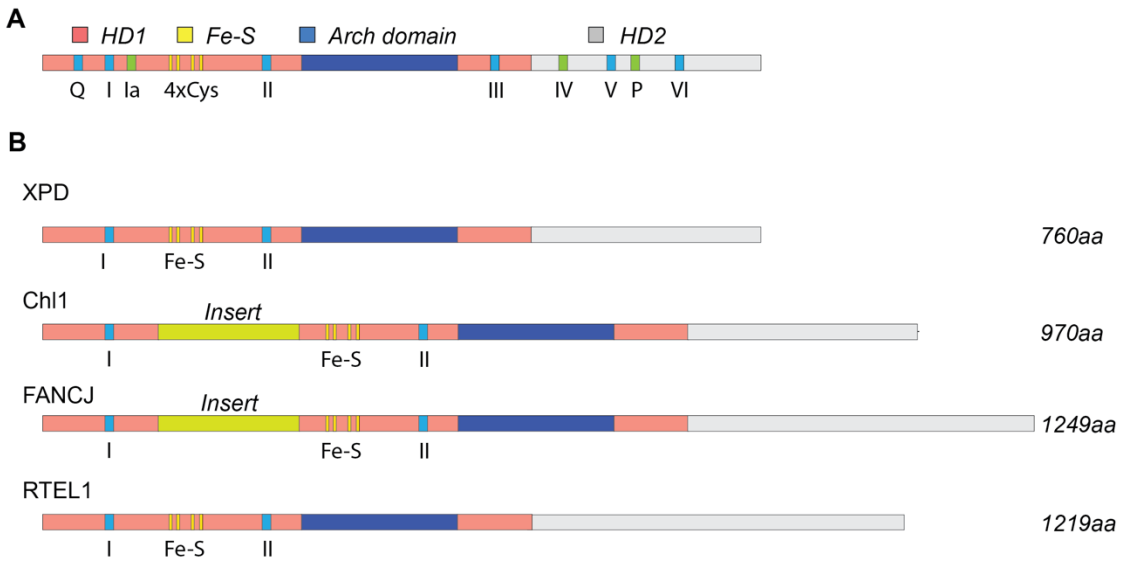

Figure 1.7 Sequence alignment of 4 human XPD subfamily proteins.

A Canonical motifs mapped to XPD structure. **B** Comparison of the lengths of human XPD subfamily proteins. Chl1 and FancJ both contain inserts between their Walker A and B motifs and these motifs are highlighted for comparison between sequences.

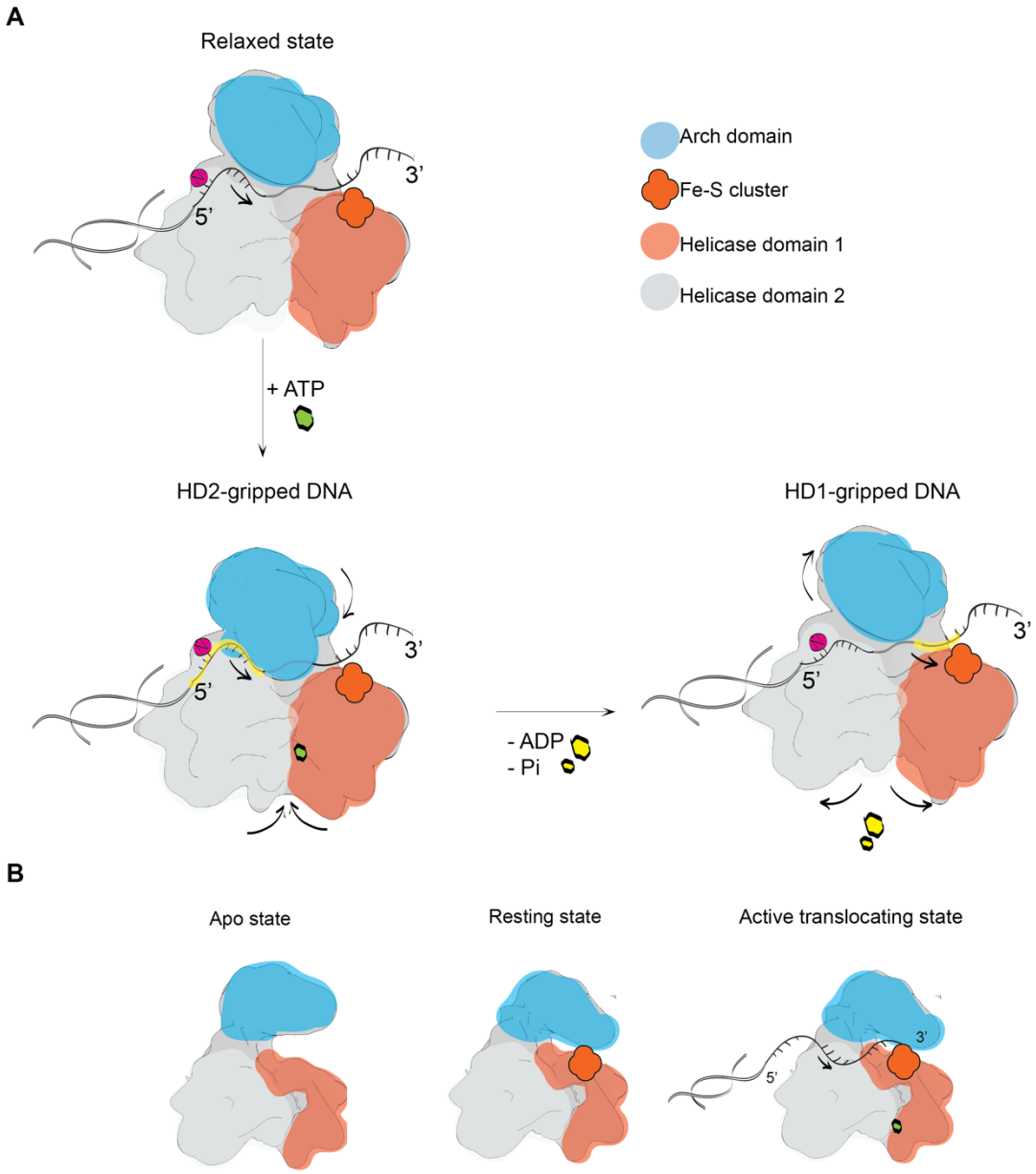

Figure 1.8 Translocation mechanisms of an XPD helicase.

A Conformational rearrangements leading to DNA translocation. Movement of the DNA is depicted by the location of the pink nucleotide N. The gripping of DNA by individual helicase domains during the translocation steps are highlighted in yellow. **B** Conformation of the Arch domain during various states of the protein.

out of this conformation upon encountering the P motif, a rigid body of two prolines in HD2. Upon ATP binding, the conformational change in HD2 results in the domain's sliding along the DNA with the HD1 holding its position, tightly gripping the DNA. ATP hydrolysis leads to relaxation of the protein into its free state. In this step the roles are reversed with the HD1 sliding along the DNA and HD2 gripping DNA more tightly (Cheng and Wigley, 2018).

1.10.2 Functions and DNA unwinding capabilities of the XPD subfamily

The gene encoding Chl1 was most likely the first chromosome loss mutant identified (Haber, 1974). It is a helicase with a strong preference of unwinding in the 5'-3' direction, although it is also able to unwind in the opposite direction if a long 3' ssDNA overhang is present (Gerring, Spencer and Hieter, 1990; Hirota and Lahti, 2000).The human ortholog DDX11 efficiently unwinds multiple types of DNA; forked duplex DNA with a minimum of 15 nucleotide 5' overhang, reminiscent of replication forks; D-loop structures, intermediates in HR; and G4 DNAs, replication fork-stalling G-rich sequences of the genome (Wu *et al.*, 2012). Interestingly, addition of RPA to helicase unwinding assays stimulates the helicase activity of DDX11 and leads to increased length of unwound DNA. The helicase activity is dependent on its ATPase activity as ATPase dead mutants do not separate DNA (Farina *et al.*, 2008b; Y. Wu *et al.*, 2012). Addition of a short 3' ssDNA overhang to a 5' 15 nucleotide overhang is required of Chl1 and leads to a marked increase in processivity. Such DNA structures resemble the replication fork (Wu *et al.*, 2012).

FANCJ was identified as a BRCA1 binding protein under the name of BACH-1 (Levran *et al.*, 2005; Litman *et al.*, 2005). Together with BRCA1, one of the initiators of HR, these proteins mediate DSB repair, where mutations in either or disruption of their binding often underlie breast cancer. The helicase activity of FANCJ for this role is essential (Cantor *et al.*, 2004). FANCJ further functions in the Fanconi Anaemia (FA) repair pathway. FANCJ's helicase activity is very versatile, being able to process even complex DNA structures like G4 DNA that negatively influence chromosomal stability (Wu, Shin-ya and Brosh, 2008). RTEL1, or regulator of telomere length 1, is the most recently evolved helicase of this family, also able to resolve various DNA structures. It is implicated in telomeric DNA maintenance where

it resolves telomere loops (T-loops), DNA structures that protect telomeric ends from degradation but must be temporarily resolved for replication (Vannier *et al.*, 2012). FANCJ further promotes disassembly of HR intermediates and D-loop structures, and resolves trinucleotide repeat hairpins, DNA structures often underlying many neurological disorders (Barber *et al.*, 2008; Frizzell *et al.*, 2014).

Identification of XPD came from studying patients with the Xeroderma pigmentosum (XP) disease, characterised by a defect in nucleotide excision repair (NER) that yields patients highly sensitive to UV light as they cannot process UV-induced DNA damage (Sung *et al.*, 1993). XPD is the only structurally characterised protein of the XPD subfamily helicases with the most characterised mechanism of action on a molecular level, and the only member present in bacterial and archaeal organisms (Wu, Suhasini and Brosh, 2009). All proteins are capable of unwinding DNA but unlike in simpler organisms, XPD of higher eukaryotes unwinds DNA as a part of the TFIIH. NER is the major pathway for processing lesions resulting from mutagenic substances such as chemotherapeutic agents or UV radiation. Such lesions are bulky and destabilising to the DNA duplex but do not share any other structural features. Enzymes involved in NER are thus good at processing a wide range of large lesions (Schärer, 2013). In TFIIH, XPD performs both NER and transcription-related roles but its DNA unwinding properties are dispensable for transcription and, conversely, essential for NER (Kuper *et al.*, 2014). XPD together with the helicase XPB open up the DNA around the lesion where only the ATPase activity of XPB is necessary (Coin, Oksenych and Egly, 2007). The helicase activity is provided by XPD. The CDK-activating kinase (CAK) complex, specifically its MAT1 subunit, is responsible for inhibiting XPD within TFIIH. Upon NER, TFIIH releases CAK in a process mediated and activated by a component of TFIIH, XPA, thus relieving the inhibition of XPD by MAT1 (Coin *et al.*, 2008).

The Cryo-EM structures of the TFIIH complex has revealed the molecular mechanism of the transition from the inactive to an active helicase (Figure 1.9-A,B). Eukaryotic XPD contains additional three helices in its Arch domain which contact the inhibitory MAT1 subunit of the CAK complex. This region in the Arch domain, described as the “plug”, likely evolved only in higher eukaryotes due to lack of TFIIH in simpler organisms. The plug occupies the same cleft as DNA (Figure 1.9-C) and

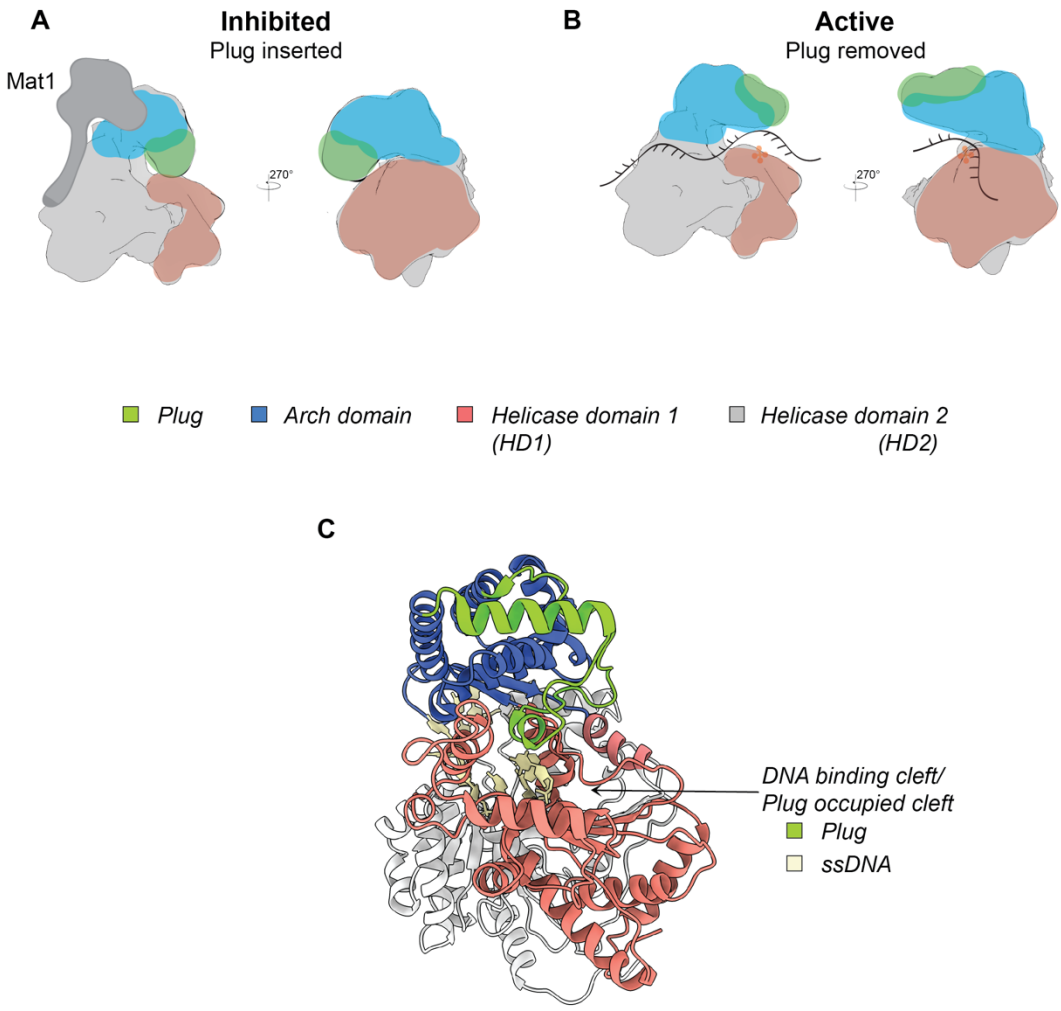

Figure 1.9 Inhibition mechanism of human XPD

A Auto-inhibited XPD has the plug domain occupying the DNA binding cleft. The inhibitory subunit MAT1 is shown in dark grey. **B** Rearrangements upon relieving Mat1 inhibition allow DNA to bind. **C** Auto-inhibited XPD with ssDNA from the actively unwinding XPD structure fitted in (PDBs:5OF4,6RO4). The plug and DNA binding sites overlap.

thus prevents DNA binding until a conformational change is induced by the activating partner XPA, which directly contacts the plug segment, resulting in the plug's removal from the DNA binding site (Greber *et al.*, 2017; Schilbach *et al.*, 2017; Kokic *et al.*, 2019). Therefore, the inactive conformation of XPD was elucidated to be due to auto-inhibition.

1.10.3 Structure-function relationship of XPD helicases in disease

The similarities in architecture as well as the presence of the conserved Fe-S cluster in the XPD subfamily domains point to a similar mechanism of action for DNA translocation. However, the possession of a helicase activity does not exclude the possibility for multiple functions not connected to DNA unwinding, as seen for XPD (Kuper *et al.*, 2014; Kokic *et al.*, 2019). The mutations arising in diseases with a disrupted XPD subfamily helicase have shed more light on the various functions of individual proteins. All diseases arising from the mutations in these helicases are inherited in an autosomal recessive manner.

Mutations in the XPD protein cause XP, Trichothiodystrophy (TTD) and XP with Cockayne's syndrome (XP/CS). All three disorders show a significantly enhanced sensitivity to UV light, but they differ greatly in cancer predisposition and aging acceleration as a result of mutations influencing the transcriptional role of TFIIH versus its repair role. Mutations in TTD were found to dominate in regions that disrupt the stability of TFIIH, between XPD and its binding partner p44 (Coin *et al.*, 1998). Mutations that disrupt the Fe-S cluster were also identified, disrupting the overall stability of the XPD protein and thus the TFIIH. This would explain the transcription-related developmental symptoms of TTD. Conversely, XP and XP/CS mutations show inability to remove UV-caused damage as a result of mutations hindering the helicase function. These mutations are predominantly found in HD2 targeting DNA or ATP-binding residues and their biochemical analysis directly shows defects in DNA unwinding (Fan *et al.*, 2008; Liu *et al.*, 2008; Kokic *et al.*, 2019).

FA arises from mutations in the FA pathway. 19 gene products are involved in this pathway with mutations in any resulting in FA symptoms: increased sensitivity to interstrand crosslinks (ICLs), bone marrow failure, predisposition to the development

of cancer One of the genes responsible for FA when mutated is FANCJ, a helicase of the XPD subfamily. The FA pathway is responsible for ICL removal, a type of DNA damage that may impede replication and transcription. The FA pathway also has a tight connection with HR pathways as FANCD1, alias BRCA2, is an important factor in HR. (Levitus *et al.*, 2005; Levran *et al.*, 2005; Ceccaldi, Sarangi and D'Andrea, 2016). Interestingly, FANCJ has roles in both FA and in HR, and was also found to mediate fork restart and replication origin firing suppression potentially by resolving DNA intermediates ahead of the fork (Cantor *et al.*, 2001; Raghunandan *et al.*, 2015). In HR, it works together with BRCA1 that interacts with the CTD of FANCJ to repair DSBs. Disruption of BRCA1 activity or the binding to FANCJ leads to breast cancer development, with regions forming the interactions being necessary for its helicase activity (Cantor *et al.*, 2004). FANCJ-mediated ICL repair important both for FA and fork restart is BRCA-1 independent, and has been found to be mediated by interactions with MLH-1 of the MutLα mismatch repair complex (Peng *et al.*, 2007). The differential interactions of FANCJ and mutations that disrupt the binding to its interactors thus dictate the loss of function and consequences of the mutation. Furthemore, studying the FA-related mutation in FANCJ has shown that this mutation causes an alanine to proline change, which disrupts the adjacent cysteine residue forming contacts with the Fe-S cluster. This results in the loss of helicase activity as an underlying cause for dysfunction of ICL repair (Wu *et al.*, 2010).

Mutations in DDX11 were found to cause WABS. Initial identification came from a male Polish patient in Warsaw suffering from developmental retardation symptoms including microcephaly and growth defects. The patient's cellular phenotype resembled that of FA, but additionally contained marks of premature sister chromatid separation which further increased when exposed to DNA crosslinking agents. Such behaviour has not been observed in FA but more closely resembled Roberts Syndrome (RBS), a cohesinopathy caused by the disruption of ESCO2. RBS however, does not show DNA damage defects like FA (Van Der Lelij *et al.*, 2010). These findings would agree with the to-date reported functions of DDX11, which include mediating cohesion establishment and a response to fork stalling. Indeed, premature sister chromatid separation as well as replication speed decline have been observed in cells of WABS patients, and mutations identified show a disruption in helicase activity of the protein. Comparison of cells from RBS and WABS patients

shows that RBS patients with an ESCO2 mutation rely primarily on DDX11 to mediate cohesion, and vice versa for WABS. In both cases, cohesion is achieved by Smc3 acetylation (Faramarz *et al.*, 2020). DDX11 mediates both cohesion and response to replication stress utilising its helicase activity where both are primarily mediated by interactions with the fork protection complex, specifically the Timeless subunit. The yeast counterpart, Chl1, does not require its helicase activity for cohesion, which was shown to be mediated by Ctf4. Such interaction exists in human cells, but interactions between Chl1 and Tof1 in yeast, which could mediate cohesion like in human cells, have not been identified or have changed during the course of evolution. The presence of both symptoms in WABS patients suggests that helicase-mediated cohesion is the primary way DDX11 contributes to cohesin, though this awaits further exploration in yeast. Both yeast and human cells however show a similar response to replication stress, and therefore at least this function is conserved across species (Wu *et al.*, 2012; van Schie *et al.*, 2020).

1.10.4 Structure-function relationship of Chl1 at the replication fork

Structural and biochemical experiments combined with observations from disease-related mutations in Chl1 point to a dual role of this protein at the replication fork. Its first role is to promote sister chromatid cohesion and this has been confirmed in simple and higher eukaryotes. Likewise, Chl1's role in responses to replication stress has also been observed in these organisms (Samora *et al.*, 2016; Faramarz *et al.*, 2020; Lerner *et al.*, 2020). Although the function is conserved, the interactions and biochemical properties of these proteins appear to vary across organisms (Figure 1.10). The requirements of the helicase activity of Chl1 remain particularly controversial as it was believed that this function is only necessary for the stress response. While this remains true, Tof1-mediated sister chromatid cohesion in human cells seems to require helicase activity, while the main cohesion pathway in yeast (also seen in humans) via the Ctf4 trimer does not. The details of Tof1-Chl1 interaction have not been well characterised in yeast and it is currently unknown whether this interaction could, too, mediate sister chromatid cohesion in a helicase-dependent way. Tof1 enhances the ATPase and helicase activities of Chl1 in human cells (Cali *et al.*, 2016). The motif responsible for Tof1 binding to Chl1 has been found to be located in Chl1's insert which is not found in all members of the XPD subfamily,

only in Chl1 and FANCJ (Figure 1.7). Sequence alignments show that this EYE motif is conserved across species and its mutations in human cells abolish the binding. Experiments where the entire Chl1 insert was removed, which should include the Tof1-binding motif, showed no effect on binding to Ctf4 in yeast but the interaction with Tof1 has not been studied (Samora *et al.*, 2016).

It is therefore possible that this region of Chl1 mediates its interactions with Tof1. Studies from human cells show that Tof1-mediated processes that include Chl1 are dependent on Chl1's helicase activity. It is therefore possible that this region not only mediates the connection to Tof1 but also its helicase function. Removal of the entire insert has not been tested on the helicase activity with human proteins and the interaction between Chl1 and Tof1 has not been well characterised in yeast.

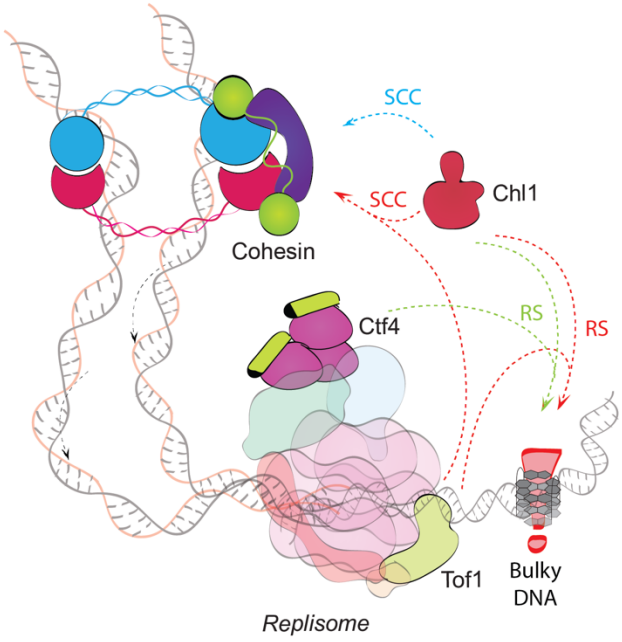

Figure 1.10 The dual role of Chl1.
Red arrows denote pathways confirmed in humans, green arrows denote pathways confirmed in yeast. Blue arrow shows a pathway observed in both organisms. The interaction between Tof1 and Chl1 in yeast is unclear. SCC: sister chromatid cohesion; RS: replication stress.

1.11 Project 2 aims

Given the lack of information on the structure of the XPD helicases and the varying function of Chl1 in yeast and humans, the primary aim of this project was to elucidate the structure of Chl1, focusing on identifying key features of this protein's architecture and observing the structure of the yet uncharacterised insert of Chl1. For this part both Cryo-EM and X-ray crystallography approaches were sought after. The secondary aim was to utilise any structural information obtained to explain the principles of the dual function of this helicase, focusing on *in vivo* studies in *S. cerevisiae* to investigate the function of Chl1 and its potential interaction with Tof1 at the replication fork.

Chapter 2. Theory of Cryo-EM

2.1 Techniques for structure determination

The three major techniques for protein structure determination are X-ray crystallography, Nuclear Magnetic Resonance (NMR) and electron cryo-microscopy (Cryo-EM). NMR is a suitable method for studying small disordered proteins and protein dynamics. Proteins of only a few dozen kilodaltons (kDa) can be studied with this technique. X-ray crystallography is the most widely used method for structure determination. The molecular weight range for this technique is broader than that for NMR. However, a major bottleneck of crystallography is the requirement for proteins to be able to pack into ordered crystals, making flexible proteins and flexible protein domains difficult to study. Over the past few years, increasing numbers of Cryo-EM structures have been deposited into the protein databank, showing how quick the gain of popularity of this method is. Cryo-EM single particle analysis (SPA) has several advantages over other techniques. These include the ability to image a wide range of molecular weights, visualisation of a larger spectrum of dynamic conformations from within one dataset and the ability to observe molecular assemblies at near-native conditions. Furthermore, recent developments in the field now allow Cryo-EM to achieve resolutions comparable to X-ray crystallography. The requirements for the amount of biological material make it more applicable than X-ray crystallography, which requires milligrams of protein compered to micrograms required for Cryo-EM (Jonić, Sorzano and Boisset, 2008; Egelman, 2016).

Cryo-EM is routinely used to study dynamics of viruses and large macromolecules such as ribosomes or proteasomes, but given the technical advancements is also applied to studying the structure of small proteins (Beckmann *et al.*, 2001; Herzik, Wu and Lander, 2019). By imaging protein samples in near-native conditions one can observe minimal restrictions to particle conformation, as opposed to crystal structures which only give a snapshot of the one possible protein conformation. Recent advancements have led to establishing time-resolved Cryo-EM methods which are further able to identify functional states of proteins in the range of milliseconds. Such range is usually unachievable by standard Cryo-EM, but the standard method is nonetheless used to solve major conformational states of

proteins as it is a reproducible method. Whilst still undergoing development, time-resolved Cryo-EM presents an exciting opportunity for the future and will allow uncovering a wide range of processing and conformational states of molecules which were not captured by standard Cryo-EM methods (Razinkov *et al.*, 2016; Feng *et al.*, 2017; Kontziampasis *et al.*, 2019; Rubinstein *et al.*, 2019). Another major breakthrough in the field has been the development of *in situ* Cryo-EM which allows studying protein structure and dynamics directly in a cell or an organism by combining plunge freezing, correlative microscopy, focused ion beam milling and electron cryo-tomography (Cryo-ET) to select a cell region for imaging by TEM. With this method structure of proteasomes, nuclear pores or ribosomes were solved within the context of the cell (Albert *et al.*, 2017; Guo *et al.*, 2018; Schaffer *et al.*, 2019; Toro-Nahuelpan *et al.*, 2020).

2.2 Sample preparation

2.2.1 Negative staining

Biological samples cannot withstand the high vacuum of the microscope and must therefore be inserted as solid samples. NS is a quick way to visualise the sample and observe its quality before preparing a more labour-extensive sample in amorphous ice. With NS one images the background rather than the molecules themselves. The protein sample is fixed on carbon-coated grids using heavy metal stains. After blotting away excess stain and air-drying the grid, particles remain embedded within a layer of stain which creates a strong contrast between the stain and the biological specimen due to high signal-to-noise ratio. This gives one an idea of the size and shape of the molecule, and homogeneity of the sample. The large grain size of heavy metal salts used for this technique limits the resolution to about 20Å (Ohi *et al.*, 2004). Multiple heavy stains with various properties exist, the most widely used being the acidic uranyl acetate (UA) or uranyl formate (UF) stain. Some proteins are particularly sensitive to acidic pH which causes them to degrade. In such cases, stains with a more neutral pH, including molybdenum or tungsten salts, can be used (Scarff *et al.*, 2018). Although providing strong contrast of the molecule, heavy stains flatten the sample which can lead to an artificial characterisation of the structure of the sample, and may cause sample dehydration and breakdown (De Carlo and Harris, 2011).

2.2.2 Vitrification

NS cannot be used for obtaining high resolution information. Sample needs to be placed in an environment which does not scatter electrons strongly, and which can reduce the radiation damage caused by the high voltage of electrons that reach the sample. Vitrification of samples for Cryo-EM allows the specimen to be kept in a native hydrated environment where the surrounding buffer composition does not scatter electrons strongly (Taylor and Glaeser, 1974). Vitrification relies on rapidly freezing the sample to form a solid layer of amorphous ice without dehydrating the sample or creating crystalline ice. In order to produce vitreous conditions, samples in aqueous solutions are applied to a grid and blotted to remove excess liquid leaving only a thin layer. Subsequently, the grid is plunged into liquid ethane cooled down to liquid nitrogen temperatures (Dubochet *et al.*, 1982; Passmore and Russo, 2016). Individual particles ideally position themselves in the holes of the grid, although some proteins show a preference for the carbon support of the grid and must therefore be frozen on a thin layer of support covering the holes (Discussed in 1.1.1.3). The thickness of the vitreous ice must be minimised to ideally be just above the largest diameter of the particle. Layers too thin will expose the particles to the air-water interface and denature. In contrast, too thick a layer will cause poor visibility of the particles, but even in thicker ice, the particles may reach the air-water interface (Glaeser and Han, 2017).

2.2.3 Choice of grid

Choice of grid support is also highly important. A grid consists of a support structure and a film spread over it. Because the supports are most often hydrophobic, aqueous solutions cannot spread across the grid unless the grid's properties are changed to hydrophilic. For that reason grids treated with low energy plasmas to remove any organic contaminants present on the surface. Argon, oxygen and hydrogen mixtures are often used for plasma cleaning, but plasma generated from residual air is also often used. This is often termed glow discharging (Passmore and Russo, 2016). Most commonly, copper supports are used for NS, with a layer of carbon deposited over them. For Cryo-EM however, thick films interfere with the quality and obtainable resolution as electrons must pass through these additional layers. Instead, so-called "open hole" grids with holey film or grids with a thin layer of either carbon or graphene

are used. Graphene consists of a monolayer of carbon atoms which do not create interfering signal and therefore does not hinder the attainable resolution. For Cryo-EM, the choice of support material and film influences the specimen's properties. Samples with the tendency to adhere to the support material are usually deposited on a thin layer of carbon or graphene to which they too adhere (Naydenova, Peet and Russo, 2019; Barski *et al.*, 2020). Using such supports can also change the orientation of particles, thus overcoming problems with preferred orientation. To eliminate specimen movement during irradiation fully gold grids are often used (Russo and Passmore, 2014). Additionally, grids can be treated with chemicals such as amylamine to reverse the charge of the surface. Amylamine creates a positively charged surface which too can help with particle orientation (Grassucci, Taylor and Frank, 2007).

2.2.4 Optimal sample quality for SPA

As a first step in structure determination, purification of a highly pure homogeneous sample is essential. Heterogeneous samples show conformational variations either due to internal flexibility, inconsistent stoichiometry or dissociation of subunits with low affinity. Complexes can be stabilised by optimising buffer composition, such as salt concentration, pH or presence of detergents, but even after extensive optimisation of these parameters the sample may remain heterogeneous. Another option can include the use of ligands to lock the protein in a given conformation, although the protein may not exhibit a significant change in conformation(Cheng *et al.*, 2015). In order to further stabilise the sample, proteins and protein complexes are often crosslinked with glutaraldehyde or bis(sulfosuccinimidyl)suberate (BS3) in solution, or can be stabilised with a more reproducible approach of gradient fixation (GraFix). In GraFix, a gradient of glycerol or sucrose and glutaraldehyde is prepared and the sample is centrifuged into the gradient, getting crosslinked as it passes through (Kastner *et al.*, 2008).

If sample heterogeneity arises from flexibility, another approach is to use monoclonal fragments antigen binding (Fab) ,an antibody fragment consisting of a variable region and a constant region of heavy and light chains, which bind to target protein with high specificity to form a rigid construct. Alternatively, nanobodies, which represent the

variable region of the heavy chain-only antibody from camelids, are also used for this purpose. Fabs are larger and can aid not only in stabilisation but also in particle alignment (discussed later), whereas the smaller-sized nanobodies have access to binding surfaces where Fabs cannot reach. By high affinity binding of a fab or a nanobody to the target protein, the flexibility can be reduced to particular conformations and increase molecular mass of the protein, giving a more defined particle for alignment. In certain cases they also improve particle orientation. Such approach is very useful for studying not only flexible but also small proteins (S. Wu *et al.*, 2012; Uchański, Pardon and Steyaert, 2020). Other options for dealing with a structurally heterogeneous sample include computational approaches, discussed later.

On several occasions, proteins are of great quality when purified or analysed by NS but upon freezing degrade rapidly. This is believed to be a consequence of particles touching the air-water interface where proteins get denatured. In a 1000Å thick ice film, particles are expected to collide with the air-water interface up to 1000x per second during the time it takes from applying the sample to plunge freezing using standard freezing protocols and equipment (Glaeser and Han, 2017). Several measures can be taken to avoid this. First, the blot times for freezing can be adjusted to ensure that particles are fully submerged in the ice layer. This does not prevent particles from touching the interface but if the ice is too thin all particles will be degraded. In order to avoid the air-water interface completely one can use detergent which will block access of the particles to the interface by forming a layer between the protein solution and the interface, or use grids with carbon or graphene film to which the particles will adhere, thus evading the air-water interface (Glaeser, 2018; D'Imprima *et al.*, 2019).

2.3 TEM

A typical electron microscope consists of an electron source, lenses to focus the beam and a detection system. Further components include apertures, vacuum pumps and valves. The vacuum system is essential for preventing electrons from interacting with air molecules and scattering, and therefore the microscope is kept under high vacuum. Electron sources like tungsten filaments or lanthanum

hexaboride crystals are now commonly used in lower end microscopes, with the most powerful machines using the field emission gun (FEG). The FEG beam emerges from a crystal sharpened tip and is more coherent with a smaller diameter than the beam from the other sources. The electron source is heated up to release electrons which are then accelerated by high voltage from 100-300kV.

The microscope has three types of lenses which focus the beam along its length: condenser, objective and projector lens. The condenser lens is placed after the electron source and is responsible for converging the beam into a parallel beam. The sample is placed below the condenser in the objective lens. This lens is responsible for the main magnification (up to 50x magnification). In the back focal plane of the objective lens is the objective aperture which is important for good contrast in single particle analysis. Further magnification of the image is provided by the projector lens, followed by detection of electrons by the detector The resolution is determined by the pixel spacing of a detector which sets the so-called Nyquist frequency. This determines the maximum resolution obtainable, where the maximum resolution is twice the pixel size (Orlova and Saibil, 2011).

2.3.1 Electrons and doses

Electrons exhibit a very short wavelength which is dependent on their energy state (0.02Å at 300kV). Because the highest theoretical resolution should depend on the wavelength of the radiation used, EM should allow for obtaining structures at atomic resolution. Indeed, sub-angstrom resolution has been achieved with radiation insensitive materials (O'Keefe *et al.*, 2001). Biological samples however are highly prone to radiation damage by the energy deposited from electrons to the specimen. Electrons which interact with the sample can either retain their energy (elastic scattering) or deposit their energy into the specimen (inelastic scattering). Alternatively, electrons can pass through the sample without an interaction (unscattered). Image formation depends on both scattered and unscattered electrons. Interference of the unscattered beam with low angle elastic scattering gives rise to phase contrast. Conversely, inelastic scattering contributes to higher background noise. Furthermore, high-angle elastic scattering is removed by the objective aperture and also contributes to amplitude contrast. This type of contrast

contributes to the image by only about 5-10% and does not give high resolution information but is especially important for obtaining image features. Despite not absorbing electrons, the biological specimen will change the phase of the electrons that pass through the sample which contributes to phase contrast and to high resolution. Ideally, a balance of electrons will be allowed to reach the sample to obtain both amplitude and phase contrast without damaging the sample and obtaining enough high resolution information. Furthermore, because inelastically scattered electrons deposit energy into the specimen, this type of electrons is responsible for inducing radiation damage. Usually a total dose up to 100 electrons/Å^2 is used to obtain high quality images of vitrified specimen, but even around 10electrons/Å^2 can damage the high resolution (Baker *et al.*, 2010; Orlova and Saibil, 2011; Cheng *et al.*, 2015; Glaeser, 2016).

Electrons which do not contribute to high resolution features are removed by the objective aperture and by an energy filter. The objective aperture is responsible for removing high angle elastically scattered electrons. The energy filter, which sits either in-column or is located post column before the detector, removes the inelastic scattering, particularly important in tomography, and improves the signal-to-noise (SNR) ratio. Both of these approaches enhance the amplitude contrast which is necessary for accurate reconstruction (Elmlund, Le and Elmlund, 2017).

2.3.2 Detectors

The total dose given the sample must be carefully chosen as too little electrons will not give sufficient contrast of the particles, whereas large doses lead to radiation damage. Insufficient electron dose leads to a very poor SNR because biological samples do not scatter electrons strongly. An improved SNR could be achieved by enhancing the beam intensity at the cost of radiation damage to the sample. Additionally, the detector itself adds noise to the image. This is described by the detective quantum efficiency (DQE). Quality of a detector is further characterised by the modulation transfer function (MTF) which describes how much contrast is transferred from a sample to the image at a given resolution (McMullan, Faruqi and Henderson, 2016).

Until 1990's, images recorded by an electron microscopes were deposited on film. Film was later replaced by charge-coupled devices (CCDs), both having a poor DQE. The revolution in Cryo-EM came by introducing detectors called Direct Detection Devices (DDDs). These sensors detect electrons directly, as opposed to indirect detection by CCDs where an electron is first converted into a photon. The new detectors are radiation-hard, have an improved DQE and a very fast readout (Kühlbrandt, 2014; McMullan *et al.*, 2014). This fast readout allows for compensation of beam-induced movements by recording movie stacks which are multiple frame recorded per exposure. The movement of the particles can be tracked and corrected for, reducing blurring of particles which hinders resolution. Furthemore, splitting into frames allows selection of frames based on amount of radiation damage inflicted, allowing for selection of frames based on the accumulated dose where high spatial frequencies are preserved (Brilot *et al.*, 2012; Campbell *et al.*, 2012; Li *et al.*, 2013).

2.4 Image collection

Images are taken in brightfield mode with an applied defocus, which is critical for enhancing contrast of the molecule from its low spatial frequencies, unfortunately at the cost of loosing high resolution information. Defocus is essential as it introduces a phase shift between scattered and unscattered electrons to enhance contrast, the larger the defocus the larger the phase shift and hence the contrast (Orlova and Saibil, 2011). Alternatively, one can use a phase plate, such as the Volta phase plate (VPP) which introduces a phase shift without the need of defocus, allowing visualisation of the specimen with greater ease without losing high resolution information (Danev *et al.*, 2014; Danev and Baumeister, 2016; Danev, Tegunov and Baumeister, 2017). For high-resolution structures, a 300kV instrument is most commonly used as higher acceleration reduces the amount of inelastic scattering which damages the sample. The disadvantage is the reduced contrast of the specimen compared to lower kV instruments(Egerton, 2014).

Collection of images is nowadays automated with several software, such as EPU, Leginon (Carragher *et al.*, 2000) or SerialEM (Mastronarde, 2005), allowing for fast automated data acquisition. The electron dose, range of defocus values used, and pixel size must be carefully chosen. The pixel size for SPA is usually kept to a

minimum, typically around 1Å, to allow for a high resolution to be obtained based on the Nyquist theorem. Most commonly images are collected without any tilt, although tilting the sample has been applied to overcome preferred particle orientation issues (Lyumkis, 2019). Introduction of DDDs allowed for collecting movies split into frames. Averaging these frames allows for correction of beam-induced motion and specimen drift. Furthermore, splitting movies into about a dozen frames allow for dose-weighing of individual frames, where one chooses those with sufficient SNR, least radiation damage and least motion. First few frames contain the lowest accumulated dose and hence have the highest resolution information preserved, but also highest amount of beam-induced motion. With progressing frames the dose used increases leading to accumulation of radiation damage, highest in the last few frames. Generally all frames are used for alignment and summing for initial processing whereas for a final three-dimensional (3D) model reconstruction the first and last few frames are removed (Campbell *et al.*, 2012; Li *et al.*, 2013; Grant and Grigorieff, 2015; Ripstein and Rubinstein, 2016; Zheng *et al.*, 2017).

Defocus values are usually kept between 1-4 um. Lens defocus and aberrations influence the contrast transfer function (CTF) of the microscope which causes resolution-dependent amplitude modulations and phase reversals of the image. CTF is a sine function whose periodicity is increased with increasing defocus. CTF crosses the zero multiple times at which points all information is lost. Increasing defocus increases CTF oscillations and therefore crossings at zero. Oscillations become more frequent at high frequencies (high resolution) and it is essential to correct for this effect by estimating defocus values and aberrations with great precision. By collecting data at different defocus values one accounts for information lost at the zero crossing of a particular defocus value, preserves both high resolution information of the sample whilst also obtaining enough amplitude contrast for alignments. Several softwares are now available which perform CTF estimation, including CTFFIND and GCTF (Zhou *et al.*, 1996; Rohou and Grigorieff, 2015; Zhang, 2016).

2.5 Image processing

Images acquired from a microscope are two-dimensional (2D) projections of the protein sample. The first step of the processing involves movie alignment to select high-quality unblurred images with satisfactory signal and low radiation damage accumulation from which the 2D projections of the sample, the particles, will be selected. Particle selection, or picking, is a process of marking particles with boxes with sufficient dimensions to select the whole particle, followed by particle extraction from the motion-corrected micrographs. Particles are either selected manually or through automatic approaches based on template picking or in a reference-free manner depending on pattern recognition(Cheng *et al.*, 2015). Recently many particle picking packages introduced function on neural networks and deep learning, including crYOLO, TOPAZ or Warp (Bepler *et al.*, 2019; Tegunov and Cramer, 2019; Wagner *et al.*, 2019). Picked particles are subsequently aligned based on their relative 2D orientations in a process called 2D classification, which is a reference-free manner of grouping particles into homogeneous subsets, or classes. Particle alignment into classes depends on the quality of the picking as well as available features of the sample. A heterogeneous flexible sample of low molecular weight will not align into classes as efficiently as a larger, more rigid protein target. If particles do not align well, the sample needs to be improved as it is essential to obtain well aligned 2D classes for an accurate 3D volume reconstruction (Joyeux and Penczek, 2002).

Because images are 2D projections of the 3D object, one can reconstruct the 3D volume according to the “projection-slice theorem” which states that the Fourier transform of a 2D projection of the sample is a 2D slice of the 3D Fourier transform of the object. By converting many 2D projections with known orientations into Fourier space, one can calculate the corresponding 2D slices within the 3D object in Fourier space 2D, and then use the inverse Fourier transform to obtain a 3D object in real space (Cheng *et al.*, 2015). Approaches to calculate a 3D volume include the “common-line” method which uses the fact that two 2D projections share a common line in the Fourier transform of a 3D object, and the relative orientations of 2D projections can be determined based on the shared lines and angles (Van Heel, 1987). Secondly, the projection-matching approach can be used where each 2D

projection is compared to a 3D projection generated computationally with an iterative improvement to the alignment process. Although this projection is initially not very accurate, each iteration leads to improvement and essentially to converging into a real structure (Penczek, Grassucci and Frank, 1994). The reference map used can either be created *ab initio* or a homolog structure can be used. Nowadays, new projection-matching based approaches are used to improve the convergence properties for an accurate 3D map generation. Such approaches include the maximum likelihood methods, such as those implemented in the software package Relion (Scheres *et al.*, 2007; Scheres, 2012), or the stochastic hill climbing first included in the package SIMPLE (Elmlund and Elmlund, 2012). Most software packages use similar projection-matching approaches. For all of them, critical parameters include sample homogeneity and conformational inflexibility. *Ab initio* reconstruction of a 3D volume depends on only a small subset of particles (with full datasets used to refine these initial volumes) and therefore it is critical that only particles which accurately represent the protein sample are chosen from well-aligned 2D classification. Following alignment, all software packages follow a similar pipeline of structure refinement, extended 3D subclassification or particle “polishing”, such as EMAN, PRIME, CryoSPARC or Relion (Tang *et al.*, 2007; Elmlund, Elmlund and Bengio, 2013; Punjani *et al.*, 2017; Zivanov *et al.*, 2018). Obtained structures must be validated in order to avoid incurred structures or false structural features to be calculated from noise or model bias. This includes for example collecting tilt pairs where one images the same spot at two different angles or by using the gold standard Fourier shell correlation (FSC), which is the correlation between two halves of the dataset (half maps) which are processed completely independently of each other. Introducing high resolution noise and observing the FSC prevents overfitting and inaccurate resolution determination, which is determined by the FSC at the cut-off value of 0.143 (Rosenthal and Henderson, 2003; Van Heel and Schatz, 2005; Scheres and Chen, 2012; Chen *et al.*, 2013).

Chapter 3. Materials & Methods

3.1 Cloning

The general cloning method used was Gibson assembly cloning, where a gene of interest with overhangs complementary to the linearised vector of choice is inserted within a pre-existing restriction site in the vector (for a list of plasmids and genes see Table 3). The gene of interest with overhangs was synthesised as a primer (GeneArt Sigma Aldrich) or amplified in a polymerase chain reaction (PCR) from a pre-existing plasmid or cDNA using primers with overhangs complementary to 5' and 3' insertion site sequences of the plasmid. In the latter approach, the amplified product was gel purified using QIAquick gel extraction kit (Qiagen), whereas the synthesised primer was used directly. The synthesised primer or product was assembled into a linearised plasmid using NEBuilder HiFi assembly Cloning kit (NEB) in a 2:1 insert:vector ratio. The reaction was incubated at 50°C for 15 minutes. 4µl of the reaction was subsequently transformed into XL-1 blue cells with positive clones selected based on antibiotic resistance selection.

Gene amplification PCR mix

5xPhusion buffer	10µl
Forward (Fw) primer 10µM	2µl
Reverse (Rv) primer 10µM	2µl
dNTPs 10mM	5µl
Template 50ng/ µl	1µl
Phusion polymerase	1µl
mQ	29µl
Total volume	50µl

Gene amplification PCR

98°C	30sec	
30 cycles of:		
	98°C	30sec
	55°C	30sec
	72°C	30sec/kb
72°C	1min/kb	
4°C	infinite	

HiFi assembly mix

Insert 100ng/ul	1µl
Vector 50ng/ul	1µl
2xHiFi Assembly mix	10µl
mQ	8µl
Total	20µl

3.1.1 Restriction enzyme digest

Prior to assembly into the plasmid, the plasmid of choice was linearised at desired restriction sites and gel purified. The reactions were performed with restrictions enzymes and CutSmart buffer (NEB) where all single enzymes exhibit 100% activity. For MultiBac cloning, cloning into MCS were carried out subsequently; first inserting the gene of interest into the multiple cloning site (MCS) 1, followed by transformation, plasmid purification using the QIAprep Spin Miniprep kit (Qiagen), restriction digest and insertion of gene of interest into MCS2 (see Table 3).

Restriction digest reaction mix

Enzyme A	1µl
Enzyme B	1µl
Plasmid	5µg
Buffer	5µl
H_2O	Xµl**
Total	50µl

**appropriate to make up to total volume

3.1.2 Site-directed mutagenesis

To introduce a mutations, a set of primers was designed using the webtool PrimerX. Following the reaction, non-mutated parental DNA was digested with Dpn1 enzyme (NEB) for 1 hour at 37°C.

Mutagenesis PCR reaction mix

10x Pfu buffer	5µl
dNTPs (10mM)	5µl
Fw primer 10µM	2µl
Rv primer 10µM	2µl
Template 50ng/µl	1µl
Pfu polymerase	1µl
H_2O	34µl
Total	50µl

Mutagenesis PCR

95°C	30sec	
16 cycles of:		
	95°C	30sec
	55°C	1min 16x
	68°C	1min/kb – 8min
4°C	infinite	

Table 2 List of mutagenic primers used in this study.

Protein	Mutation	Primer 1	Primer 2
*Ct*Chl1	E725D & Q728A	CGGTTCCTCTGACGA CATCCTGGCGCAGTA CTCCGACGC	GCGTCGGAGTACTGCG CCAGGATGTCGTCAGA GGAACCG
*Ct*Chl1	S723I	CAAGGGCGGTTCCAT TGACGACATCCTGGC GCAG	CTGCGCCAGGATGTCG TCAATGGAACCGCCCTT G
*Hs*Core^{ShortScc3_WB}	Smc1^{E1157Q}	CATTCTTCGTCCTGG ACCAGATCGACGCTG CCCTG	CAGGGCAGCGTCGATC TGGTCCAGGACGAAGA ATG
*Hs*Core^{ShortScc3_WB}	Smc3^{E1144Q}	CTACCTGTTCGACCA GATCGACCAGG	CCTGGTCGATCTGGTC GAACAGGTAG

3.1.3 Colony PCR

To rapidly test for positive clones, colony PCR was carried out using primers for the bacterial T7 promoter. Each colony was separately picked, briefly inserted and twisted in the colony PCR mix and subsequently inserted into 5ml of Luria broth (LB) media with the appropriate antibiotic to inoculate the overnight culture for subsequent plasmid purification.

Colony PCR mix	
10x MangoTaq buffer	5µl
50mM $MgCl_2$	1µl
dNTPs (10mM)	2µl
Fw primer (10µM)	1µl
Rv primer (10µM)	1µl
MangoTaq polymerase	0.75µl
H_2O	14.25µl
Total	25µl

Colony PCR		
94°C	5min	
22 cycles of:		
	94°C	30sec
	53°C	30sec
	72°C	30sec/kbp
72°C	1min/kbp	
4°C	infinite	

3.1.4 Yeast integration PCR

To assess correct integration with no tandem repeats, yeast-specific colony PCR was performed with primers annealing to genomic sites flanking the selective marker integration site. For exclusion of clones with tandem integration, a third set of primers is used which only result in a PCR product when tandem integration has occurred. For the PCR reaction, a yeast colony is first resuspended in 20µl of 15mM NaOH and boiled for 5-10min followed by brief centrifugation to remove cell debris.

Intergration PCR mix	
10x MangoTaq buffer	5µl
50mM $MgCl_2$	1µl
dNTPs (10mM)	1.5µl
Fw primer (10µM)	1µl
Rv primer (10µM)	1µl
Yeast template in NaOH	1µl
MangoTaq polymerase	0.5µl
H_2O	14.25µl
Total	25µl

Colony PCR		
94°C	5min	
35 cycles of:		
	94°C	30sec
	53°C	30sec
	72°C	1.5min
72°C	5min	
4°C	infinite	

3.1.5 Bacmid integration PCR

The transposition of the plasmid into the bacmid backbone is assessed with both the blue/white screen as well as with integration PCR using primers for sites flanking the Tn7 site used for transposition for bacmid production (see Section 3.3.3).

Bacmid integration PCR mix	
10x Taq buffer	5µl
50mM $MgCl_2$	1µl
dNTPs (10mM)	1µl
Fw primer (10µM)	1.5µl
Rv primer (10µM)	1.5µl
Bacmid 100ng/µl	1µl
MangoTaq polymerase	0.5µl
H_2O	38.5µl
Total	50µl

Colony PCR		
94°C	5min	
35 cycles of:		
	94°C	30sec
	53°C	30sec
	72°C	1.5min
72°C	5min	
4°C	infinite	

Table 3 Enzymes and overhang sequences used in these studies

Vector	Cloning site + *Restriction enzyme*	End	Overhang (in a 5'-3' direction)	Gene inserted
pUCDM	MCS1 *StuI*	5' 3'	AAGCGCGCGGAATTCAAA GG CTACGTCGACGAGCTCAC TA	ScSmc1, ScSmc1 ^{head}, HsSmc1^{head}, HsSmc1
pUCDM	MCS2 *SmaI*	5' 3'	GGGTGATCAAGTCTTCGT CG GCACCATGGCTCGAGATC CC	ScSmc3 ^{head}, HsSmc3^{head},ScSmc 3, HsSmc3
pFL*	MCS1 *StuI*	5' 3'	TGTACTTCCAGTCCGGCA GG CCTAGTCGCGGCCGCTTT CG	ScScc1, HsScc1, CtMiniChl1, CtMiniChl1^{v2,} Nipbl^{C}
pFL*	MCS2 *SmaI*	5' 3'	ACTCGACGAAGACTTGATC ACCC GGGTGATCAAGTCTTCGT CG	ScScc3, HsScc3
pET22b	T7 *NdeI*	5' 3'	ATAATTTTGTTTAACTTTAA GAAGGAGATATACAT GGTGGTGGTGGTGGTGCT CGAG	ScSmc3^{hinge}
pET28a – 6xHIS*	T7 *BamHI*	5' 3'	AAACCTGTATTTTCAGAGC GGATCC GTCGACGGAGCTCGAATT CG	Ag/At/Ct/Sc/Sp/Hs Chl1 Insert, CtCtf4^{CTD}, ScHinge^{1}, ScSmc1^{hinge}, HsHinge
pET28a – GST*	T7 *BamHI*	5' 3'	aaacctgtattttcagagcggatcc caaggcctgtacagaattcg	ScInsert^{v1}, ScInsert^{v2}, ScInsert^{v3}
pRS vectors	Gal 1-10 *SmaI*			

** Backbones of original plasmids with modifications*

Enzymes used for cloning:

MangoTaq™ DNA polymerase (Bioline), Taq polymerase (Thermofisher Scientific), Pfu polymerase (homemade, Singleton lab), Phusion™ polymerase (Thermofisher Scientific), Restriction enzymes: Bam-HI, NdeI, SmaI, StuI (NEB).

3.2 Protein Expression

3.2.1 Bacterial Expression

Cloned constructs containing either an N-terminal glutathione-S-transferase (GST) or 6xHistidine tag (His-tag) were transformed or co-transformed into BL21(DE3) competent cells using a standard heat shock protocol and plated on LB agar plates with the appropriate antibiotics. Colonies were selected after 18 hours at 37°C cells. A single colony overnight culture (ONC) was used to inoculate 6L or more of LB. Cells were grown at 37°C with constant shaking at 200rpm until the OD of 0.6 was reached. Cultures were then either directly induced with 1mM IPTG or cooled down to 18°C and induced and harvested after 4 and 18 hours, respectively, by centrifuging at 4000rpm for 20minutes. Pellets were then washed with lysis buffer and either processed directly or flash frozen in LN_2 to be kept at -80 °C.

3.2.2 Yeast growth conditions

All experiments with yeast culture, including *in vivo* experiments described in section 3.7, were performed at a 25°C temperature unless otherwise stated. Cultures were grown in Yeast Peptone (YP) media supplemented with a final concentration of 2% glucose (YPD). For selection of transformants, yeast nitrogen base (YNB) plates supplemented with a range of amino acids were used, where the amino acid used as the marker for positive selection was eluded. Amino acids at a concentration of 6mg/ml included Adenine, histidine, Tryptophan, Leucine, Uracil.

3.2.3 Yeast Expression

Saccharomyces cerevisiae (*Sc*) was selected as the expression strain. All genes of interest were cloned into integrating vectors kindly provided by John Diffley. The expression vectors contained a selective amino acid marker to complement the amino acid auxotrophy of the expression strain. To enable integration, the vector was

linearised in the selective marker and introduced into the cells to be integrated by HR. Depending on the vector used, positive clones were selected for by removing the vector's amino acid marker from the media. Yeast transformation was carried out using the lithium acetate (LiAc) method. Competent cells for transformation were first grown on YPD plates for 3 days at 30°C and subsequently used for a 5ml ONC. The ONC was used to inoculate 50ml of YPD media. After 4 hours at 30°C, the cells were centrifuged at 2500 rpm at 4°C, pellet washed with 10ml of H_2O and once with 4ml of LiAc/Tris-EDTA (LiAc/TE) buffer. Cells were then resuspended in 50ul of LiAc/TE buffer. 2µg of linearised plasmid was mixed with carrier DNA, the single-stranded salmon sperm DNA, which has been boiled for 5minutes at 95°C and cooled down on ice beforehand. 50µl of competent cells were added to the plasmid:carrier DNA mix and vortexed. For successful DNA uptake, 300µl of LiAc/TE buffer supplemented with polyethylene glycol (PEG) buffer was used to permeabilise the membrane, followed by a 5 second vortex. Cells were then incubated for 30 minutes at 30°C and subsequently heat-shocked for 15 minutes at 42°C. Cells were briefly placed on ice before centrifuging at 2500 rpm for 2 minutes. The supernatant was discarded, cells were resuspended in 300µL of 1M sorbitol and plated on plates containing the appropriate antibiotic. As a control, competent cells were transformed with carrier DNA only and plated on plates with and without the used antibiotic. Cells were grown for 3 days at 30°C before several colonies were re-streaked and grown for another day before being tested for successful integration.

A positive clone was re-streaked across a whole plate containing the appropriate marker and incubated overnight before taken from the plate and resuspended in a small volume of YPD media used for inoculation. For a 5L culture, 2.5L YP media supplemented with 2% raffinose was inoculated with cells and grown overnight at 220rpm at 30°C. The expected OD600 following this incubation is 2. Cells were back-diluted to OD_{600} 1 and expression was induced with addition of galactose to a final concentration of 2%. Cultures are incubated for 4 hours at 30°C with 220rpm shaking and harvested by centrifuging cultures at 2000rpm at 4°C for 10 minutes. Pellets were washed with lysis buffer and subsequently resuspended in fresh lysis buffer. The cell suspension was frozen as "popcorn" where suspension was added to LN_2 in a drop-wise manner. Yeast popcorn was mechanically lysed with a SPEX freezer mill 6875D (AXT) precooled to LN_2 temperature. Cell lysis was carried out in six

cycles of shaking and pausing to convert popcorn to cell powder which was stored in -80.

10x TE buffer
pH 7.5
100mM Tris-HCl
10mM EDTA
sterile filter, store at RT

50% PEG4000
50g PEG4000
dissolved in diH_2O to 100ml
sterile filter, store at RT

1M Lithium-Acetate
20.4g LiAc-dihydrate dissolved in diH_2O to 200ml end-volume
sterile filter, store at RT

LiAc/TE buffer
1ml 1M LiAc
1ml 10x TE buffer
8ml diH_2O
made fresh before use

LiAc/TE buffer + PEG
100µl 1M LiAc
100µl 10x TE buffer
800µl 50% PEG
make fresh before use

Lysis buffer
50mM HEPES pH8
250mM NaCl
0.5mM TCEP
10% glycerol
0.5mM EDTA pH8
10mM NaF
10mM Beta-glycerophosphate
EDTA-Free Protease Inhibitor cocktail 1 in 50 (Millipore)
Basemuncher Benzonase (Expedeon)

3.2.4 Insect cell

Plasmids containing two MCS were used for expression of proteins in insect cells. pFL-based plasmid or pFastBac were used for single protein expression and the protein was expressed under the polyhedrin promoter. For protein complex expression, pairs of protein were cloned into donor and acceptor plasmids into both MCS and recombined into a single plasmid using Cre recombinase (MultiBac) or cloned into pFL-based plasmids and used for co-infection with two viruses.

pUCDM cloning was carried out in PiR1 cells, pFL and MultiBac plasmid cloning was done in XL1 cells. Final products were transposed into DH10Bac or EmBacY cells using standard heat shock protocol, but left to regenerate for 4 hours shaking at 1000rpm at 37°C. Cells were plated on plates containing tetracyclin, gentamycin and kanamycin, supplemented with X-gal and Isopropyl ß-D-1-thiogalactopyranoside (IPTG) and incubated for 3 days at 37°C. Transposition of the plasmid into the cells occurs into the Tn7 transposition sites, resulting in the disruption of the *LacZ* gene and production of white colonies instead of blue. This allowed for selection of positive (white) clones, which were used to inoculate a 5ml LB ONC for bacmid production. The bacmid was purified using the isopropanol precipitation method as per standard protocol.

For transfection, a total of $1x10^6$ *Spodoptera frugiperda* 9 (Sf9) cells were added in a drop-wise fashion to a 6-well plate and left to adhere for 1 hour. 1µg of bacmid was mixed with 100µl of serum-free media (SFM), and either 5µl of GeneJuice Transfection Reagent (Merck) or 3µl of FuGene Transfection Reagent (Promega) was mixed with 100µl serum free media (Gibco). The two solutions were mixed and left at room temperature to up to 45 minutes. After incubation, further 800µl of SFM was added to the mixture to create the final transfection mixture. Media from the well with cells was removed and the final transfection mixture was added to the well. Plate was incubated in a humidified box at 27°C . If GeneJuice was used, media was exchanged after 6 hours post transfection. FuGene is not toxic to the cells and therefore not removed from the media. The plate was incubated for 3 days to create the P1 virus, followed by two viral amplifications to obtain the P2 and P3 viruses, respectively. Throughout the whole virus amplification, cell diameter and viability was monitored. Increased cell diameter suggested successful infection, and decrease in viability suggested effective viral replication and release into the surrounding media, therefore P2 and P3 viruses were harvested when their cell diameter increased by 30% and viability was reduced by 10-20%.

For protein expression, 300ml cultures in 2L roller flasks were prepared the day before infection to reach a cell count of $1.5x10^6$cells/ml on the day of infection. Each culture was infected with appropriate volume of P3 virus to obtain an multiplicity of

infection (MOI) of 1. Cells were harvested when the cell diameter expanded by 30% without a drop in cell viability. Cells were harvested at 2000rpm for 15 minutes. Pellets were either used directly or if frozen, EDTA-free Protease Inhibitor cocktail (Millipore) was added to each centrifuge tube and cells were snap frozen in LN_2.

3.3 Protein Purification

Table 4 Protein sequences used for construct generation in these studies

Protein abbreviation	Full protein name	Gene species	Sequence modification
*Hs*Hinge	Cohesin Smc1/Smc3 hinge domain	*H. sapiens*	Smc1a$^{481-687}$, Smc3$^{476-688}$
*Hs*Smc3^{hinge}	Cohesin Smc1/Smc3 hinge domain	*H. sapiens*	Smc3$^{476-688}$
*Sc*Hinge^{1}	Cohesin Smc1/Smc3 hinge domain	*S. cerevisiae*	Smc1$^{488-695}$, Smc3$^{488-704}$
*Sc*Hinge^{2}	Cohesin Smc1/Smc3 hinge domain	*S. cerevisiae*	Smc1$^{488-695}$, Smc3$^{488-690}$
*Sc*Scc2-Scc4	Cohesin loader	*S. cerevisiae*	Scc2^{FL}, Scc4^{FL}
Nipbl^{C}	Cohesin loader	*H. sapiens*	Nipbl$^{1164-2805}$
*Hs*Cohesin	Cohesin	*H. sapiens*	Smc1a^{FL}, Smc3^{FL}, Scc1^{FL},SA2^{FL}
*Sc*Cohesin	Cohesin	*S. cerevisiae*	Smc1^{FL}, Smc3^{FL}, Scc1^{FL},Scc3^{FL}
*Hs*Core	Cohesin Smc1 Smc3 head domain, Scc1, Scc3	*H. sapiens*	Smc1$^{1-225/1004-1233}$, Smc3$^{1-231/963-1217}$, Scc1^{FL},SA2^{FL}
*Hs*Core^{ShortScc3}	Cohesin Smc1 Smc3 head domain, Scc1, Scc3	*H. sapiens*	Smc1$^{1-225/1004-1233}$, Smc3$^{1-231/963-1217}$, Scc1^{FL},SA2$^{60-1080}$
*Hs*Core^{ShortScc3_WB}	Cohesin Smc1 Smc3 head domain, Scc1, Scc3	*H. sapiens*	Smc1$^{1-225/1004-1233}$, Smc1-E1157Q Smc3$^{1-231/963-1217}$, Smc3-E1144Q Scc1^{FL},SA2$^{60-1080}$
*Ct*Core	Cohesin Smc1 Smc3 head domain, Scc1, Scc3	*C. thermophilum*	Smc1$^{1-265/1140-1264}$, Smc3$^{1-229/951-1207}$, Scc1^{FL},Scc3$^{139-1119}$
*Ct*Chl1	Chl1	*C. thermophilum*	Chl1^{FL}

*Ct*MiniChl1	Chl1	*C. thermophilum*	$Chl1^{1\text{-}74\ +216\text{-}918}$
*Ct*MiniChl1^{v2}	Chl1	*C. thermophilum*	$CtChl1^{1\text{-}143\ +\ 208\text{-}918}$
*Ct*Chl1LongInsert	Chl1	*C. thermophilum*	
*Sc*Chl1Insert$^{GST\text{-}v1}$	Chl1	*S. cerevisiae*	$Chl1^{14\text{-}227}$
*Sc*Chl1Insert$^{GST\text{-}v2}$	Chl1	*S. cerevisiae*	$Chl1^{14\text{-}129}$
*Sc*Chl1Insert$^{GST\text{-}v3}$	Chl1	*S. cerevisiae*	$Chl1^{54\text{-}227}$
*Ag*Chl1Insert	Chl1	*A. gossyppii*	$Chl1^{50\text{-}179}$
*At*Chl1Insert	Chl1	*A. thaliana*	$Chl1^{51\text{-}188}$
*Hs*Chl1Insert	Chl1	*H. sapiens*	$Chl1^{56\text{-}226}$
*Sp*Chl1Insert	Chl1	*S. pombe*	$Chl1^{56\text{-}217}$
*Sc*Chl1Insert	Chl1	*S. cerevisiae*	$Chl1^{54\text{-}227}$
*Ct*Tof1^{CTD}	Tof1	*C. thermophilum*	$Tof1^{1\text{-}463}$

3.3.1 Bacterial

*Ct*Ctf4^{CTD} construct was purified as follows: the pellet was resuspended in 5:1 lysis buffer:pellet ratio and sonicated on ice for 5-7minutes at 40% amplitude using a sonicator (Branson). Lysed cells were centrifuged at 4°C for 45 minutes at 23,000rpm. The supernatant was filtered with a 5µm filter and incubated with NiNTA agarose beads (Qiagen) pre-equilibrated with lysis buffer. Beads were washed with 5 volumes of lysis buffer and 5 volumes of wash buffer and bound protein eluted with elution buffer. Protein was diluted in dilution buffer to decrease the salt concentration to 100mM, and loaded on Ion Exchange Chromatography (IEX) PorosQ column pre-equilibrated with buffer A. Protein was eluted with a gradient between buffer A and buffer B, concentrated with a 20kDa cut-off concentrator (Generon) and loaded on a pre-equilibrated Superose 6 10/300GL size exclusion chromatography (SEC) column (GE healthcare). Protein was either used directly for EM grids or snap frozen in LN_2 supplemented with 10% glycerol and stored at -80°C.

Hinge constructs were purified with pre-packed NiNTA columns (GE healthcare), eluted using a shallow gradient of wash buffer and elution buffer. Protein was diluted to 100mM NaCl and tag cleaved overnight using tobacco etch virus (TEV) protease. Untagged protein was loaded on a HiTrap Heparin HP column (Merck) pre-equilibrated with buffer A and eluted with a gradient of buffer A and B. Protein was concentrated with a 10kDa cut-off concentrator (Generon) and loaded on HiLoad

16/60 Superdex 75 column (GE healthcare) pre-equilibrated with GF buffer of 50mM, 150mM, or 300mM NaCl strength.

Chl1 inserts in His-tagged pET28a-based plasmids (see Table 4) were purified with NiNTA beads. For *C. thermophilum* and *H. sapiens* inserts, the tag removed overnight with TEV protease. Protein was again incubated with NiNTA beads for 30 minutes to capture the cleaved tag. The protein-containing flow-through was collected and concentrated using a 3kDa cut-off concentrator (Generon) and loaded on a Superdex 75 10/300 GL column (GE healthcare) equilibrated with GF buffer.

*Sc*Inserts in GST-tagged pET28a vectors (see Table 4) constructs were purified using GST beads using lysis and wash buffers with the same composition except the absence of imidazole. Elution was performed by adding 25mM reduced L-glutathione to the wash buffer.

Lysis buffer
50mM Hepes pH 8
500mM NaCl
20mM Imidazole
0.5mM TCEP
EDTA-Free Protease Inhibitor cocktail
1 in 50 (Millipore)
Basemuncher Benzonase (Expedeon)

Wash buffer
50mM Hepes pH 8
300mM NaCl
20mM Imidazole
0.5mM TCEP

Dilution buffer
50mM HEPES pH8
0.5mM TCEP

Buffer A
50mM Hepes pH 8
100mM NaCl
0.5mM TCEP

Buffer B
50mM Hepes pH 8
1000mM NaCl
0.5mM TCEP

GF buffer
50mM Hepes pH 8
150mM NaCl (unless otherwise stated)
0.5mM TCEP

3.3.2 Yeast

Cell powder containing yeast cohesin constructs was dissolved in 100ml of lysis buffer and mildly sonicated to shear DNA. Lysate was centrifuged for 50min at 4°C at 23,000 rpm and supernatant loaded on a pre-packet Streptactin column (Qiagen) equilibrated with lysis buffer. The column was washed with 10 column volumes (CV) of lysis buffer and eluted with elution buffer.

Lysis buffer	Elution buffer
50mM HEPES pH8	50mM HEPES pH8
250mM NaCl	250mM NaCl
0.5mM TCEP	0.5mM TCEP
10% glycerol	10% glycerol
0.5mM EDTA pH8	0.5mM EDTA pH8
10mM NaF	2.5mM Desthiobiotin
10mM Beta-glycerophosphate	
EDTA-Free Protease Inhibitor cocktail 1 in 50 (Millipore)	
Basemuncher Benzonase (Expedeon)	

3.3.3 Insect cell

All constructs expressed in insect cells contained a 2xStreptavidin tag (Strep-tag) on the N terminus, in case of protein complexes on one of the subunits. Harvested cells were resuspended in lysis buffer in a 2:1 buffer:pellet ratio and sonicated for 5-7minutes using a 5 second 20% pulse with 10 second breaks. Cells were subsequently centrifuged at 23,000rpm for 45 minutes at 4°C. The supernatant was applied to a pre-equilibrated Streptactin column (Qiagen), washed with at least 10CV of lysis buffer and eluted with 10CV of elution volume. All proteins were purified in the presence of HEPES at pH 8. For *Ct*Chl1, HEPES was substituted with TRIS pH 8.5.

IEX was carried out with PorosQ (GE healthcare) for cohesin constructs, HiTrap HP Q columns for *Ct*Chl1 and HiTrap HP S columns for *Ct*MiniChl1 and *Ct*MiniChl1^{v2}. All proteins were diluted with dilution buffer prior to IEX. The column was washed thoroughly and the protein sample eluted with a shallow gradient of buffer A and buffer B. Protein was then concentrated using an appropriate cut-off concentrator (Generon) and loaded on a pre-equilibrated SEC column. For EM, protein was taken from the fraction corresponding to the highest point of the elution peak and used directly or diluted and used. For crystallography, protein was concentrated to a desired concentration and used directly for crystal trays. Snap frozen protein was supplemented with 10% glycerol prior to freezing. For expression tests, Strep-Tactin Sepharose beads (IBA Lifesciences) were used instead of a prepacked column. Supernatant was incubated with pre-equilibrated beads for 2 hours. Beads were then centrifuged for 5 minutes at 1000rpm, the supernatant discarded and 10-20ml of lysis buffer added to beads. The beads in lysis buffer were applied to a gravity column, washed thoroughly and protein eluted with elution buffer.

Lysis buffer
50mM HEPES pH8
250mM NaCl
0.5mM TCEP
0.5mM EDTA pH8
EDTA-Free Protease Inhibitor cocktail 1 in 50 (Millipore)
Benzonase Benzonase (Expedeon)

Dilution buffer
50mM HEPES pH8
0.5mM TCEP

Buffer A
50mM Hepes pH 8
100mM NaCl
0.5mM TCEP

Elution buffer
50mM HEPES pH8
250mM NaCl
0.5mM TCEP
0.5mM EDTA pH8
2.5mM Desthiobiotin for Strep-Tactin beads

Buffer B
50mM Hepes pH 8
1000mM NaCl
0.5mM TCEP

GF buffer
50mM Hepes pH 8
150mM NaCl (unless otherwise stated)
0.5mM TCEP

3.4 Protein characterisation

3.4.1 Mass spectrometry

All mass spectrometry (MS) experiments were carried out by the Crick Proteomics STP.

3.4.1.1 Single-band ID and Intact Molecular weight determination

Single-band identification (Single-band ID), also known as gel band identification, and intact molecular weight determination MS were used to confirm the identity of purified proteins. For single-band ID, protein was run on a BisTris SDS-PAGE gel. Bands for analysis were excised and digested with trypsin which created peptides between lysine and arginine residues. Individual peptides were then mapped onto provided protein sequence and abundance of the protein and other potential proteins from the expression system was determined. For intact molecular weight determination-MS the sample was kept in the GF buffer.

3.4.1.2 Hydrogen-Deuterium Exchange

Samples for Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS) were prepared by 10-fold dilutions from 5μM CtChl1 protein in deuterated or nondeuterated buffers. In-line pepsin-immobilized column was used for protein digestion. For labelling experiments, protein was incubated for 10 s, 100 s, and 1000 s at room temperature. All HDX-MS experiments were performed in triplicate. Sequence coverage and deuterium uptake were analysed by using ProteinLynx Global Server (Waters) and DynamX (Waters) programs, respectively.

3.4.2 Western blot

Purified protein identity or contents of crosslinked species were confirmed with Western blot (WB). All primary and secondary antibodies are listed below. Briefly, denatured proteins were separated by SDS-PAGE, transferred onto a nitrocellulose membrane using a semi-dry TurboBlot transfer system (BioRad) and quality of the transfer checked with Ponceau S stain. Subsequently, the membrane was blocked for 1 hour using 5% milk in phosphate buffer saline (PBS) supplemented with 0.1% Tween-20 (PBS-T) followed by an overnight incubation with the primary antibody at

4°C. Following washing with PBS-T, the membrane was incubated secondary antibody for 1 hour at room temperature followed by several washes with PBS-T. The signal was detected using Amersham ECL™ Western blotting Detection reagents (GE Healthcare) with ImageQuant LAS 4000 imager (GE Healthcare). For membrane re-probing, the second round of primary antibody incubation was supplemented with 0.03% sodium azide for secondary antibody HRP deactivation.

List of primary antibodies:

Mouse anti-Rad21 (ab154769, Abcam), Mouse anti-SA2 (ab155081, Abcam), Mouse anti-Smc1 (ab21583, Abcam), Mouse anti-Smc3 (ab9263, Abcam), Mouse anti-strep (MB2017, Bioworld), Mouse anti-His (ab18184, Abcam), Mouse Anti-HA (sc-9372, SantaCruz Biotechnology), Rabbit Anti-V5 (ab15828, Abcam), Mouse anti-tubulin (T9026, Sigma), Anti-GST HRP conjugate (GERPN1236, Merck).

List of secondary antibodies:

Polyclonal Goat Anti-mouse Immunoglobulins/HRP (P044801-2, Agilent), Polyclonal Goat Anti-rabbit Immunoglobulins/HRP (P044701-2, Agilent).

3.4.3 SEC-MALS

To confirm the molecular weight and determine the oligomeric states of individual proteins and protein complexes, Size exclusion chromatography with Multi-angle light scattering (SEC-MALS) was kindly performed by Ian Taylor. Proteins were loaded on an Superdex200 10/300 column (GE Healthcare) pre-equilibrated with GF buffer supplemented with 0.05% sodium azide, followed by sample injection into the Dawn 8+ MALS system (Wyatt).

3.4.4 Thermal stability

Thermal stability measurements were carried out using the Prometheus system (Nanotemper) which measures the effects of a range of pHs on thermal unfolding of proteins and their aggregation. Bis-Tris buffers ranging from 6 to 9 in 0.5 increments with a 150mM NaCl and 0.5mM TCEP. Melting temperatures were plotted in Prism.

3.5 Protein-protein and protein-DNA interactions

3.5.1 Glycerol gradients

Gradients were prepared from two starting solutions, low and high density solutions. Low density solution was dispensed into a polypropylene tube (Beckmann) followed by dispensing the high density solution below. Solutions were made to resemble the purification buffer of the proteins studied, supplemented with the necessary glycerol concentration. The gradient was prepared using a gradient master (Biocomp) and left to set at 4°C for a minimum of 1 hour. A concentrated sample of 100-200µl was loaded onto the gradients. Gradients were subsequently centrifuged depending on the molecular weight of the proteins analysed. Gradients were separated into 300µl fractions and analysed with SDS-PAGE.

Table 5 Parameters of glycerol gradients used in these studies

Sample in gradients	Glycerol gradient	Centrifugation time	Centrifugation speed
*Hs*Core	10-25%	18 hours	33,000 rpm
Full-length *Hs/S* cohesin	10-50%	16 hours	33,000 rpm
*Ct*Chl1	10-30%	16 hours	50,000 rpm

3.5.2 Crosslinking

To stabilise macromolecular complexes, GraFix is often the method of choice (Stark, 2010). This includes introducing a crosslinking reagent into the high density solution when preparing a gradient. Glutaraldehyde (Grade I, 25% in H_2O, Sigma Aldrich) was introduced into the high density solution of a glycerol gradient. GraFix gradients were prepared with final concentrations ranging from 0.05%-0.2%. Alternatively, in solution crosslinking was performed using a final concentration of 0.06-0.8% glutaraldehyde.

3.5.3 Pulldowns

Purified proteins with either a his-tag or strep-tag were used to study protein-protein interactions *in vitro.* His-tagged proteins were incubated with pre-equilibrated NiNTA Agarose beads (Qiagen) for 1 hour, strep beads were incubated with Strep-Tactin Sepharose beads (IBA Lifesciences) for 1 hour. Afterwards beads were centrifuged at 1000rpm for 3 min and washed 3x with wash buffer supplemented with 5% BSA, 0.025% NP-40, glycerol or unsupplemented. Beads were again centrifuged, buffer removed and 2x SDS sample buffer was added to the beads. Beads were boiled, centrifuged at 13,000rpm for 1min and the top of the liquid was loaded on SDS-PAGE gel for analysis. Untagged protein, protein with a different tag or cell lysates were used as negative controls to determine non-specific binding.

Wash buffer	Elution buffer
50mM HEPES pH8	50mM HEPES pH8
150mM NaCl	150mM NaCl
0.5mM TCEP	0.5mM TCEP
	2.5mM Desthiobiotin

3.5.4 Electromobility shift assays

All DNA used is listed in table 6. To assess protein binding to DNA, either 1% agarose or 6% polyacrylamide native gels were used for electromobility shift assays (EMSAs). A constant DNA concentration was incubated with protein of increasing concentrations to visualise the shift of DNA suggestive of binding. 10µl mixture of either protein or protein:DNA was loaded into the wells of a prechilled gel and ran at 4°C in 1xTBE buffer (Novex). Agarose gels were ran at 80V for 20 min, native gels at 5mA for 75min. Fluorescein (FAM)-labelled DNA was used in all experiments and the gels were visualised using Typhoon Fla 9500 (GE Healthcare). After imaging, gels were also stained with InstantBlue™ coomassie stain to confirm the presence of the proteins. 1µM of *Ct*Chl1 in GF buffer was incubated with increasing DNA concentrations in the presence of ATPγS nucleotide and $MgCl_2$ unless otherwise stated. 1µM of *Hs*Hinge, *Hs*Smc3^{hinge} or ScHinge[2] was incubated in the presence of either ssDNA or dsDNA.

3.5.5 Fluorescence anisotropy

Fluorescence anisotropy measurements were kindly performed by the Biophysics/Structural Biology STP. Anisotropy was measured in a 3x3 mm quartz cuvette using a JASCO FP-8500 fluorescence spectrometer equipped with polarizers. To determine the affinity of Chl1 for ssDNA, 10 nM FAM-labelled ssDNA^4 was titrated with *Ct*Chl1 or *Ct*MiniChl1^{v2} solutions also containing 10nM of FAM ssDNA^4. The fluorescence anisotropy was measured after each addition at 484nm/ 520nm excitation and emission wavelength, with 10nm band width. Experiments were performed in buffer containing 50 mM HEPES pH 8.0, 50 mM NaCl and 0.5 mM TCEP at 25°C.

Table 6 DNA used in this study.

For dsDNA only the top strand is shown. The use of individual DNA is described in Results chapters where necessary.

Name	Description	Sequence 5'-3'
dsDNA^1	10 bp dsDNA	CAGCTCCATG
dsDNA^2	12 bp dsDNA	CAGCTCCATGAG
dsDNA^3	21 bp dsDNA	CAGCTCCATGAGCAGCTCCAT
dsDNA^4	22 bp dsDNA	CCCAGTACGACGGCCAGTGCGC
ssDNA^2	12 base ssDNA	CAGCTCCATGAG
ssDNA^3	21 base ssDNA	CAGCTCCATGAGCAGCTCCAT
ssDNA^4	22 base ssDNA	CCCAGTACGACGGCCAGTGCGC

3.6 Structural Analysis

3.6.1 Negative stain EM

For negative stain grids, all proteins were taken straight from the elution peak of the SEC column and used at a concentration of 0.1µM. Copper 400 mesh grids coated with carbon (EM resolutions) were glow discharged with the carbon side exposed for 30seconds at 45mA. 4µl of protein sample was applied to the glow-discharged carbon for 60 seconds. Majority of the protein solution was blotted away and thegrid stained by either 2% UA, sodium silicotungstate (SST) or sodium phosphotungstate (PTA) by applying rotational movements of the grid on top of 4 drops of the negative stain in a sequential manner. Excess negative stain was blotted away and the grid was left to air dry.

3.6.2 Cryo-EM

For the highest resolution dataset, *Ct*Chl1 was prepared as follows: the sample was taken from the elution peak after SEC and concentration was adjusted to 2µM. Lauryl Maltose Neopentyl Glycol (LMNG) detergent of a final concentration of 0.003% was added to the sample prior to vitrification. Non-glow discharged fresh C-flats 1.2/1.3 Au (EM resolutions) were used for vitrification with Vitrobot mark IV (FEI) at room temperature at 95% humidity. Sample was applied for 60 seconds and blotted with a blot force of -1 for 2.5 seconds. Grids were transferred and stored in LN_2. For grid types used during screening and for individual datasets, refer to Table 7 below.

Table 7 Grid types and conditions used for screening

Sample	**Grid type**	**Glow-discharge parameters**	**Grid treatment/Detergent addition to sample**
*Ct*Chl1 (for 8Å reconstruction)	C-flat 1.2/1.3 300 Mesh Au	No glow discharging	+0.003% LMNG final concentration
*Ct*Chl1/*Ct*Chl1+ssDNA/*Ct*Chl1+E06 nanobody/ *Ct*MiniChl1, *Ct*MiniChl1^{v2}	C-flat 1.2/1.3 300 Mesh Au	No glow discharging	+0.003% LMNG or +0.1% OG (final concentrations)
*Ct*Chl1	UltrAuFoil 1.2/1.3 300 Mesh Au	45mA 4min	
*Ct*Chl1, *Ct*Chl1 + *Ct*Ctf4^{CTD} Crosslinked	UltrAuFoil 1.2/1.3 300 Mesh Au	45mA 4min	+ Graphene Oxide
*Ct*Chl1 + *Ct*Ctf4^{CTD} Crosslinked	Ultrathin carbon 400 Mesh Cu (“Lacey” grids)	45mA 60sec	

*Ct*Chl1 + *Ct*Ctf4^{CTD} Crosslinked	Quantifoil 1.2/1.3 400 Cu Mesh	45mA 60sec	+Amylamine
*Hs*Core	Quantifoil 1.2/1.3 400 Cu Mesh	45mA 60sec	

For graphene oxide (GO) grid preparation, GO (Sigma Aldrich) was mixed with mQ in a 1:8 ratio and centrifuged at 500xg to remove debris. UltrAuFoil R1.2/1.3 grids were glow discharged for 4 minutes at 45 mA followed by GO application on the glow-discharged side for 4 minutes before blotting off the excess and washing the grid three times with mQ. Grids were made fresh before freezing.

For each grid type, blot times of 2-5 seconds were tested, keeping the blot force constant at -1. For amylamine treatments, grids were glow discharged in the presence of 20µl of amylamine deposited onto a filter paper placed within the glow discharger using the same parameters as with non-amylamine grids. For grids with the octyl glucoside (OG) detergent, a final concentration of 0.1% was added to the protein solution prior to vitrification.

3.6.3 Grid screening and data collection

Negative stain grid and initial cryo grid screening was performed on a CCD camera using 120kV G2 Spirit Twin TEM (Thermofisher Scientific) with a single-tilt side entry holder or a 626 side-entry cryo-holder (Gatan), respectively. For further cryo grid screening, grids were clipped and loaded in a 12-slot cassette placed in the NanoCab onto the 200kV Talos Arctica with a Falcon III camera (Thermofisher Scientific). Data collection was performed in linear mode, specific parameters for data sets are provided in Table 8. Defocus parameters ranged from -3 to -1.5µm changing in half µm increments. High-resolution data collection was performed on a 300kV Titan Krios equipped with a K2 camera and an energy filter operating in counting mode.

Table 8 Image collection parameters

Sample	Microscope	Pixel size	Total dose
*Ct*Chl1, CtChl1 + ssDNA/E06 nanobody, CtMiniChl1, *Ct*MiniChl1^{V2}	Talos Arctica	1.26Å	Up to 85e^-/Å^2
*Ct*Chl1 + CtCtf4^{CTD}	Talos Arctica	1.61Å	Up to 85e^-/Å^2
*Ct*Chl1	Titan Krios	0.839Å	74e^-/Å^2

3.6.4 Image processing

Initial datasets were motion-corrected using MotionCorr2 (Zheng *et al.*, 2017) and CTF estimation performed using GCTF(Zhang, 2016). Particles were picked semi-automatically using Eman2.2 boxer and extracted in Relion3. Particles were imported into CryoSparc2 for initial 2D classification. After several rounds of classification to eliminate unwanted particles and artifacts such as crystalline ice, 2D classes were used for template-based picking with Gautomatch. Particles were extracted and classified as before. First initial model was built in CryoSparc2 (Punjani *et al.*, 2017) and refined in Relion3(Scheres, 2012; Zivanov *et al.*, 2018), final model built with Sidesplitter(Ramlaul *et al.*, 2020). For full pipeline see Figure 3.1.

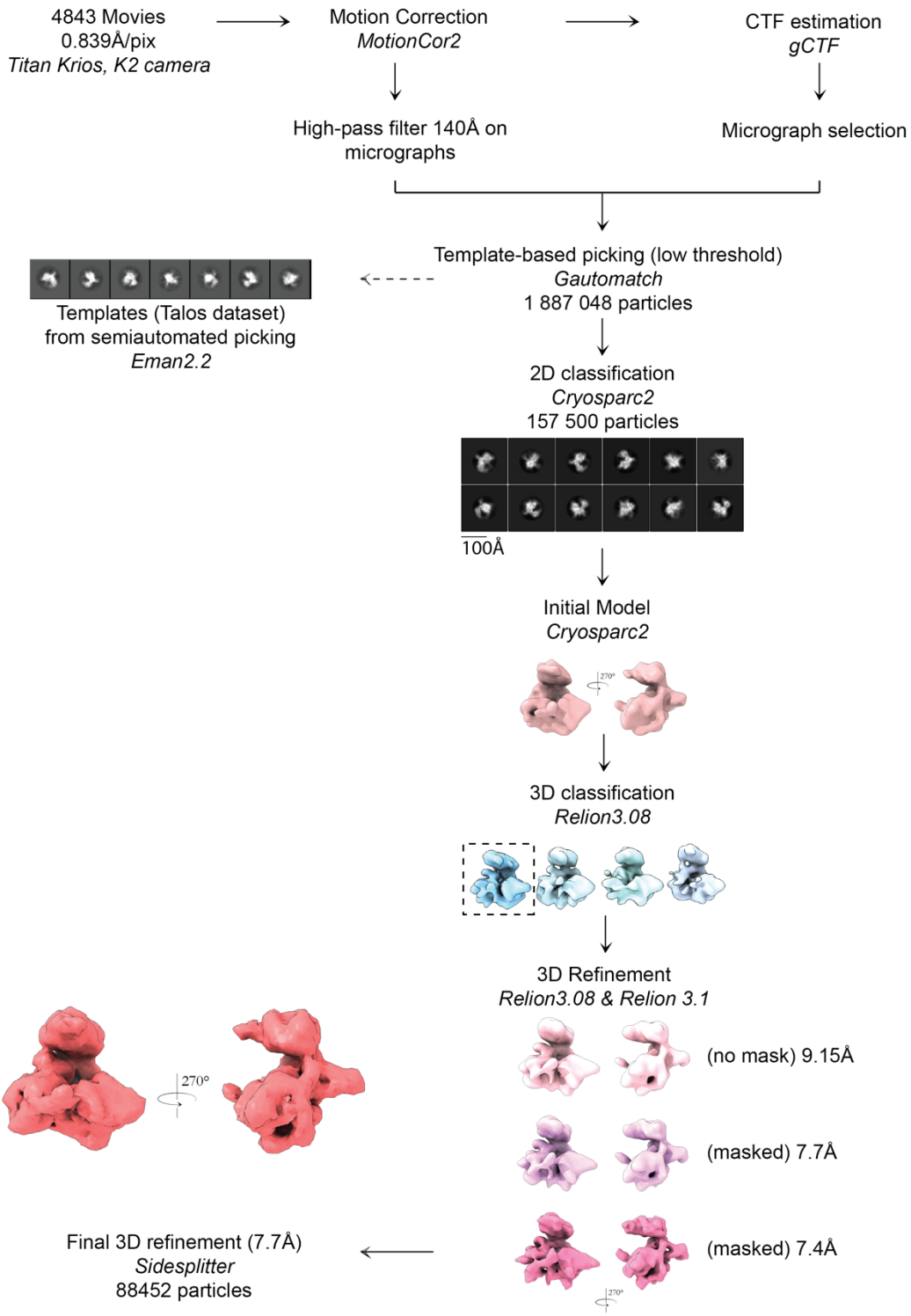

Figure 3.1 Image processing pipeline for the 7.7Å *Ct*Chl1 structure.
Softwares used for individual processing steps are shown in italics. All 3D volumes were visualised in ChimeraX.

3.6.5 Crystallisation trials

All conditions screened are included in Table 9. Screens were purchased by the Structural Biology STP.

Screens used:
CrystalScreen HT, Index, Natrix HT, PEG-Ion (Hamptons), Wizard I, II, III, IV (Jena Bioscience), JCSG + HTS, Morpheus, PACT Premier, PGA, PEG I, PEG II, Structure I&II (Molecular Dimensions), Classics I&II, JCSG Core I, II, III, IV, Mb Class Suite II, Nucleix (Qiagen), PEG/AS (Crick Structural Biology STP).

Cryoprotectants used: 10-30% glycerol, MPD, Cryo-oil

Table 9 Crystallisation screens used in these studies

ScHinge

Protein concentration	**Screen**	**Temperature**	**Drop size (protein+reservoir)**	**DNA:protein ratio**
5mg/ml	JCSG I	RT	100nl+100nl	2:1
5mg/ml	JCSG II	RT	100nl+100nl	2:1
5mg/ml	Wizard I&II	RT	100nl+100nl	2:1
5mg/ml	Wizard III&IV	RT	100nl+100nl	2:1

*Hs*Hinge

Protein concentration	**Screen**	**Temperature**	**Drop size (protein+reservoir)**	**DNA:protein ratio**
5mg/ml	JCSG I	RT	100nl+100nl	2:1
5mg/ml	JCSG II	RT	100nl+100nl	2:1
5mg/ml	Wizard I&II	RT	100nl+100nl	2:1
5mg/ml	Wizard III&IV	RT	100nl+100nl	2:1

5mg/ml	JCSG I	RT	100nl+100nl	2:1
8.5mg/ml	JCSG I	RT	150nl+150nl	0.8:1
8.5mg/ml	JCSG II	RT	150nl+150nl	0.8:1
8.5mg/ml	JCSG III	RT	150nl+150nl	0.8:1
8.5mg/ml	JCSG IV	RT	150nl+150nl	0.8:1
8.5mg/ml	Natrix	RT	150nl+150nl	0.8:1
10mg/ml	JCSG I	RT	150nl+150nl	0.8:1
10mg/ml	JCSG II	RT	150nl+150nl	0.8:1
10mg/ml	JCSG III	RT	150nl+150nl	0.8:1
10mg/ml	JCSG IV	RT	150nl+150nl	0.8:1
10mg/ml	Morpheus	RT	150nl+150nl	0.8:1
13mg/ml	JCSG+	RT	100nl+100nl	-
13mg/ml	Classics I	RT	100nl+100nl	-
13mg/ml	JCSG I	RT	100nl+100nl	-
13mg/ml	JCSG II	RT	100nl+100nl	-
13mg/ml	JCSG III	RT	100nl+100nl	-
13mg/ml	JCSG IV	RT	100nl+100nl	-
13mg/ml	Classics I	RT	100nl+100nl	2:1
13mg/ml	Classics II	RT	100nl+100nl	2:1
13mg/ml	JCSG I	RT	100nl+100nl	2:1
13mg/ml	JCSG II	RT	100nl+100nl	2:1
13mg/ml	JCSG III	RT	100nl+100nl	2:1
13mg/ml	JCSG IV	RT	100nl+100nl	2:1
13mg/ml	Classics I	4°C	100nl+100nl	-
13mg/ml	Classics II	4°C	100nl+100nl	-
13mg/ml	JCSG I	4°C	100nl+100nl	-
13mg/ml	JCSG II	4°C	100nl+100nl	-
13mg/ml	JCSG III	4°C	100nl+100nl	-
13mg/ml	JCSG IV	4°C	100nl+100nl	-
13mg/ml	Classics I	4°C	100nl+100nl	2:1
13mg/ml	Classics II	4°C	100nl+100nl	2:1
13mg/ml	JCSG I	4°C	100nl+100nl	2:1
13mg/ml	JCSG II	4°C	100nl+100nl	2:1

13mg/ml	JCSG III	4°C	100nl+100nl	2:1
13mg/ml	JCSG IV	4°C	100nl+100nl	2:1

*Hs*Hinge crystallisation conditions used for shooting

Protein concentration	**Screen and well**	**Temperature**	**DNA:protein ratio**	**Condition**
13mg/ml	JCSG IV – D9	RT	-	1.26M Tri-sodium citrate, 10% Glycerol, 0.09M HEPES pH 7.5
13mg/ml	JCSG IV – D9	RT	2:1	1.26M Tri-sodium citrate, 10% Glycerol, 0.09M HEPES pH 7.5
13mg/ml	JCSG IV – D9	RT	-	2.4M Ammonium Sulfate, 0.1M Bicine pH 9
13mg/ml	JCSG IV – D9	RT	2:1	2.4M Ammonium Sulfate, 0.1M Bicine pH 9
10mg/ml	JCSG III – D6	RT	0.8:1	2.4M Ammonium Sulfate, 0.1M HEPES pH 6.5
10mg/ml	JCSG IV – A2	RT	0.8:1	0.2M Lithium Sulfate, 1.2M Sodium di-hydrogen phosphate, 0.8M Potassium di-hydrogen phosphate, 0.1M Glycine pH 10.5

*Hs*Smc3^{hinge}

Protein concentration	**Screen**	**Temperature**	**Drop size (protein+reservoir)**	**DNA:protein ratio**
10mg/ml	JCSG I	RT	150nl+300nl	0.8:1
10mg/ml	JCSG II	RT	150nl+300nl	0.8:1
10mg/ml	JCSG III	RT	150nl+300nl	0.8:1
10mg/ml	JCSG IV	RT	150nl+300nl	0.8:1
10mg/ml	Natrix HT	RT	150nl+300nl	0.8:1

12mg/ml	JCSG I	RT	150nl+300nl	0.8:1
12mg/ml	JCSG II	RT	150nl+300nl	0.8:1
12mg/ml	JCSG III	RT	150nl+300nl	0.8:1
12mg/ml	JCSG IV	RT	150nl+300nl	0.8:1
12mg/ml	Morpheus	RT	150nl+300nl	0.8:1

Nipbl^{C}

Protein concentration	**Screen**	**Temperature**	**Drop size (protein+reservoir)**
9mg/ml	JCSG I	RT	150nl+150nl
9mg/ml	JCSG I	RT	150nl+150nl
9mg/ml	JCSG III	RT	150nl+150nl
9mg/ml	JCSG IV	RT	150nl+150nl
9mg/ml	Morpheus	RT	150nl+150nl
9mg/ml	PACT Premier	RT	150nl+150nl
9mg/ml	PEG/AS	RT	150nl+150nl

Nipbl^{C} condition used for shooting

Protein concentration	**Screen and well**	**Temperature**	**Condition**
9mg/ml	PEG/AS – E9	RT	1.8M Ammonium Sulfate, 0.1M PIPES pH 7

*Ct*Chl1

Protein concentration	**Screen**	**Temperature**	**Drop size (protein+ reservoir)**	**Protein:DNA ratio (if applicable)**
15mg/ml	JCSG+ HTS	RT	150nl+150 nl	3:1

10mg/ml	JCSG+ HTS	RT	150nl+150 nl	3:1
9mg/ml	JCSG I	RT	150nl+150 nl	3:1
9mg/ml	JCSG II	RT	150nl+150 nl	3:1
9mg/ml	JCSG III	RT	150nl+150 nl	3:1
9mg/ml	JCSG IV	RT	150nl+150 nl	3:1
9mg/ml	CrystalScreen HT	RT	150nl+150 nl	3:1
9mg/ml	Index	RT	150nl+150 nl	3:1
9mg/ml	Mb Classics Suite II	RT	150nl+150 nl	3:1
9mg/ml	Morpheus	RT	150nl+150 nl	3:1
9mg/ml	Natrix HT	RT	150nl+150 nl	3:1
9mg/ml	Nucleix	RT	150nl+150 nl	3:1
9mg/ml	PACT premier	RT	150nl+150 nl	3:1
9mg/ml	PEG I	RT	150nl+150 nl	3:1
9mg/ml	PEG II	RT	150nl+150 nl	3:1
9mg/ml	PEG Ion	RT	150nl+150 nl	3:1
9mg/ml	PGA	RT	150nl+150 nl	3:1

9mg/ml	Structure I&II	RT	150nl+150 nl	3:1
12mg/ml	JCSG I	RT	150nl+150 nl	3:1
12mg/ml	JCSG II	RT	150nl+150 nl	3:1
12mg/ml	JCSG III	RT	150nl+150 nl	3:1
12mg/ml	JCSG IV	RT	150nl+150 nl	3:1
12mg/ml	Natrix HT	RT	150nl+150 nl	3:1

*Ct*MiniChl1

Protein concentration	Screen	Temperature	Drop size (protein+reservoir)
11mg/ml	JCSG I	RT	150nl+150nl
11mg/ml	JCSG II	RT	150nl+150nl
11mg/ml	JCSG III	RT	150nl+150nl
11mg/ml	JCSG IV	RT	150nl+150nl

*Ct*MiniChl1^{V2}

Protein concentration	Screen	Temperature	Drop size (protein+reservoir)
13mg/ml	JCSG I	RT	150nl+150nl
13mg/ml	JCSG II	RT	150nl+150nl
13mg/ml	JCSG III	RT	150nl+150nl
13mg/ml	JCSG IV	RT	150nl+150nl
13mg/ml	Natrix HT	RT	150nl+150nl

3.7 *In vivo* yeast experiments

3.7.1 Transformation

Transformation was performed to introduce a tag to the endogenous locus of Tof1. Overall, the transformation was performed as described in section 3.2.3. After confirmation of positive clones using integration PCR, clones were expanded on YPD plates and either used for further experiments or flash frozen with 10% glycerol.

3.7.2 Cell cycle arrest

Mating type a budding yeast were used in all yeast experiments. For cell synchronisation in G1, cells are treated with α factor (synthesised by the Peptide Chemistry STP), a mating pheromone of the opposite "sex" of budding yeast, mating type α. This treatment is repeated three times for 55 minutes starting with log phase cultures (OD_{600} = 0.2-0.3). Cells arrest themselves in G1 and the arrest is determined by the presence of schmoos, and by fluorescence-activated cell sorting (FACS). For synchronised release into S-phase, G1-arrested cells were filtered with media without alpha factor and released into media either supplemented with 200mM HU or unsupplemented.

3.7.3 FACS

In order to identify the correct timepoint for co-IP sample collection, as well as confirm a successful G1 arrest, 1ml of yeast culture was taken at specific timepoints for FACS analysis where the content of DNA in the cells is analysed. The 1ml culture was centrifuged at 13,000rpm, supernatant was aspirated and cells were resuspended in pre-chilled 70% ethanol. Cells were centrifuged again and resuspended in 500µl of 50mM TRIS pH 7.5 supplemented with RNase. Samples were incubated for at least 4 hours at 27°C before centrifuged for 1 min at 13,000rpm and resuspended in FACS buffer supplemented with 0.5ug/ml propionium iodide (Sigma Aldrich). Samples were sonicated for 10 seconds and measured with FACSCalibur (Becton Dickinson) according to standard protocol for haploid yeast cells. Data was analysed using FlowJo.

FACS buffer

200mM TRIS pH 7.5

210 mM NACl

78mM $MgCl_2$

3.7.4 Co-immunoprecipitation

Cells of OD_{600} = 0.3 were pelleted and resuspended in IP lysis buffer. Glass beads were added and sample was lysed using a cell breaker prechilled to 4°C with 14 rounds of 7 second breaking followed by 7 second rest to achieve sufficient cell breakage as judged by an optical microscope. The sample was subsequently centrifuged at 13,000rpm for 10minutes at 4°C and supernatant applied to pre-equilibrated IgG Dynabeads Talon (Invitrogen) as a preclear step and incubated for 1 hour spinning. The unbound fraction was then applied onto Dynabeads with Protein-A (Invitrogen) bound to an antibody appropriate for the tag of the protein (Anti-HA probe SantaCruz; Anti-V5 antibody Abcam) incubated at the same conditions as previously, followed by extensive washing and elution into a 2xSDS loading buffer. Samples were boiled for 5minutes at 65°C and analysed using WB.

IP lysis buffer

50mM HEPES pH 7.5

150mM NaCl

0.5mM TCEP

10% glycerol

1x cOmplete EDTA free protease inhibitor cocktail tablet

Basemuncher Benzonase (Expedeon)

RNase (Sigma Aldrich)

0.1% NP-40

Chapter 4. Results 1 – Structural characterisation of cohesion establishment

4.1 Establishment of successful expression systems

4.1.1 Summary of structure and function of cohesin and its loader

In 2020 two structures showing the mechanism of DNA entry into the cohesin ring were published. This chapter describes experiments carried out to understand this process without the knowledge of these recently published mechanisms.

4.1.2 Design of rigid constructs for structural analysis

The large molecular weight of the cohesin complex makes it an appropriate candidate for structural analysis by Cryo-EM. However, as this technique relies on precise particle alignment for high resolution structure determination, cohesin's flexibility poses a problem for the alignments. Constructs used in this study were designed to represent the conformation which cohesin is believed to adopt upon binding by the cohesin loader complex. Cohesin constructs created therefore lack the flexible regions, namely the Smc coiled coils, but still include all regions previously found and predicted to form contacts with the loader (Figure 4.1) (Chao, Murayama, *et al.*, 2017). All constructs used in this study are listen in Table 4 (see Methods). The construct designed to resemble the body of cohesin, termed the core, lacks the coiled coils which were instead replaced by two intramolecular linkers, leaving only the most proximal coiled coil segments to the heads in the constructs. Each linker connects two helical segments of the same Smc head, which then dimerise to form cohesin's ATPase. As Scc1 and Scc3 are both important for cohesin's function and interactions, these proteins were co-expressed with the Smc1 and Smc3 ATPase heads to create the core construct. The hinge domain was designed based on available crystal structure of *T.maritima* comprised of the donut-shaped dimer comprised of helices and sheets and a short stretch of coiled coil. The hinge domain was not linked to the core by a linker but expressed as a separate protein complex. Nipbl C-terminal segment, termed Nipbl^{C}, the human ortholog of Scc2, was designed based on the crystal structure of *Ashbya gossypii* (*Ag*Scc2, PDB: 5ME3). This structure represents the C terminus of *Ag*Scc2 (*Ag*Scc2^{C}), a

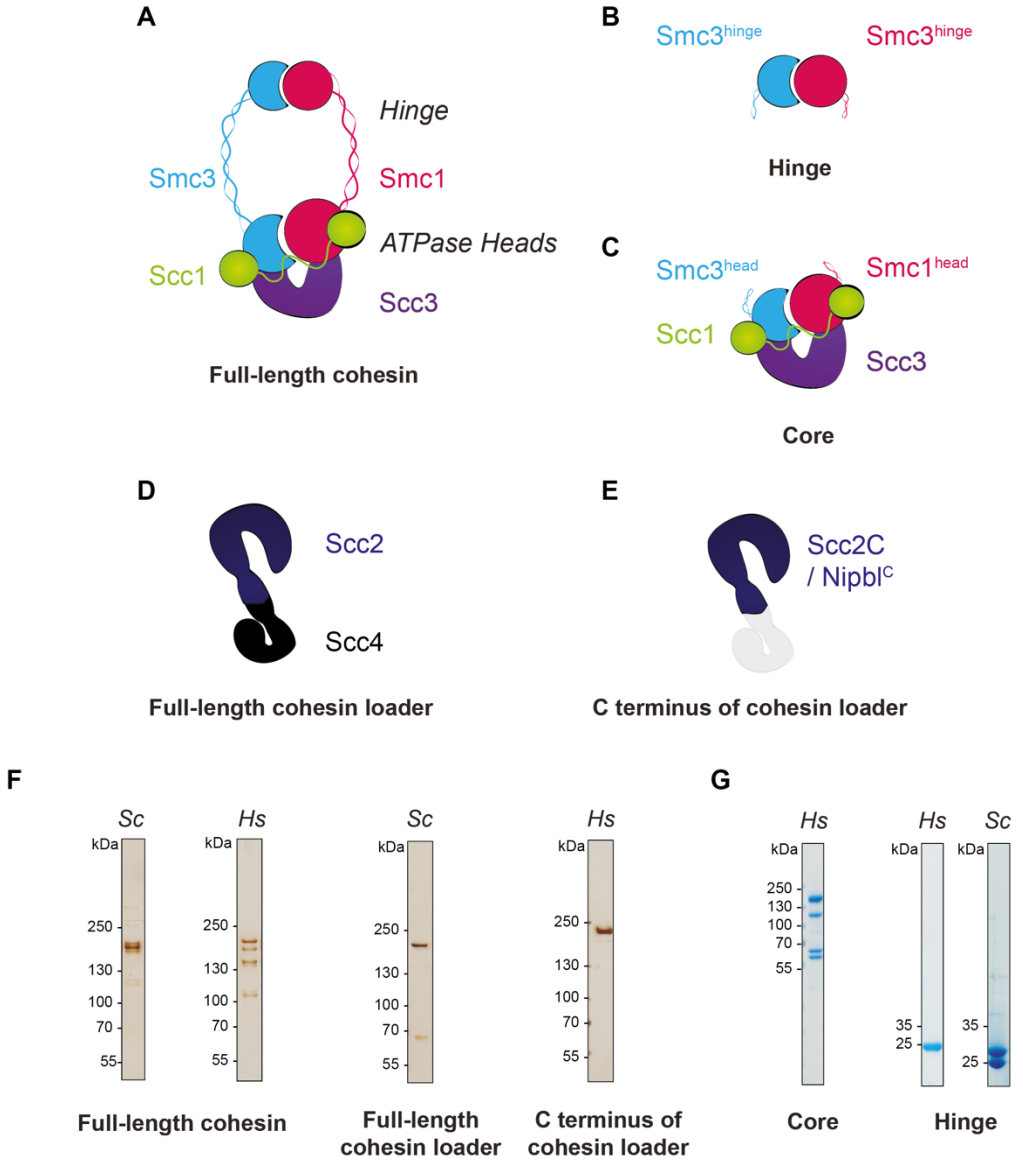

Figure 4.1 The design of cohesin constructs.

A Protein subunits of the full-length cohesin complex. **B** The isolated hinge domain construct. **C** Truncated core construct, where the coiled coils and the hinge are absent. **D** Protein subunits of the full-length cohesin loader. **E** Nipbl^{C} , the C terminus of the human homolog of Scc2. **F** Silver-stained gels of purified full-length cohesins and cohesin loaders. **G** Coomasie-stained isolated cohesin domains. All constructs in F and G are described in more detail below.

structured segment of the cohesin loader that was shown to be sufficient for *in vitro* loading activity (Chao *et al.*, 2015).

4.1.3 Choice of expression system

Three eukaryotic systems, baculovirus, yeast and mammalian, were screened for expression of the human core construct. Out of those, all cohesin constructs expressed with highest yields using the baculovirus expression system. Yeast expression was low and due to lack of antibodies for individual subunits often difficult to confirm, whereas expression in mammalian cells resulted in more signal from endogenous full-length Smc1 and Smc3 proteins than the recombinant Smc1 and Smc3 heads (not shown). The insect cell system was therefore chosen for expression. Full-length cohesin was subsequently also tested in this expression system and because of good yields was also later purified from insect cells. When choosing the expression system for Nipbl^{C} and the hinge domains, all proteins were tested and expressed in the same expression systems as in reports of crystal structures. Nipbl^{C} was expressed in insect cells whereas both *Sc*Hinge and *Hs*Hinge were expressed in *Eschericha coli (E. coli).*

4.1.4 Purification of the hinge domains

*Hs*Hinge purification has shown that under conditions with higher salt the protein can adopt a monomeric state where the Smc3 hinge purifies as a separate peak (Figure 4.2-A). This has been previously reported with bacterial Smc proteins where Smc3 could be purified without Smc1 (Haering *et al.*, 2002). By increasing the salt concentration to 300mM NaCl two well-defined peaks were obtained where the presence of *Hs*Hinge dimer and *Hs*Smc3^{hinge} monomer were confirmed by single-band ID and SEC-MALS. Individual peaks were then purified in 50mM salt for crystallisation. Even under low salt, the *Hs*Smc3^{hinge} monomer never formed a homodimer as the elution volume remained the same as in the combined sample purification under higher salt.

The equivalent of *Hs*Hinge construct was made with budding yeast proteins. The first construct, termed *Sc*Hinge^{1}, showed a C-terminal truncation in its Smc3 subunit, resulting in two species of Smc3 hinge binding to the Smc1 hinge (Figure 4.2-B).

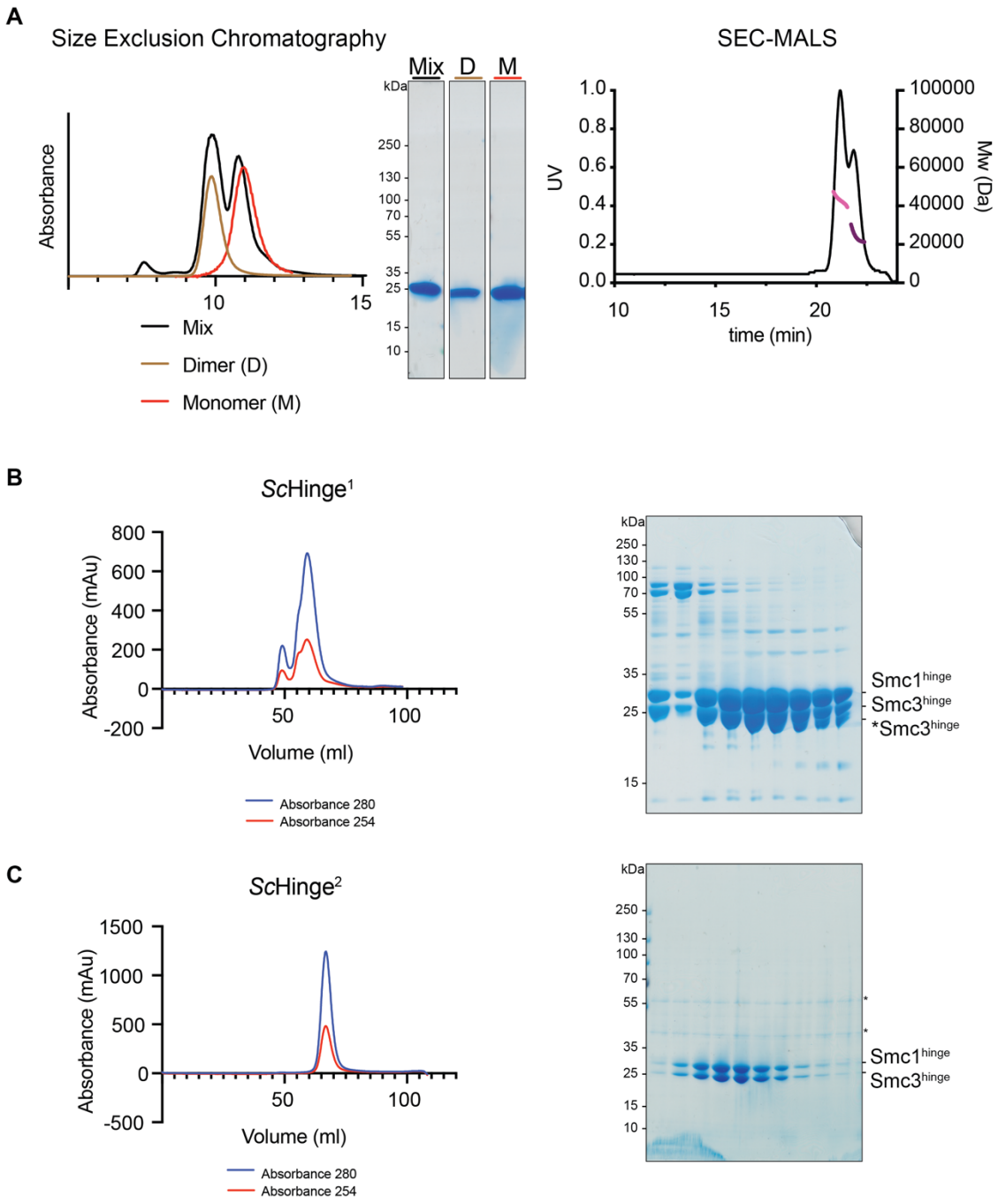

Figure 4.2 Purification of human and yeast hinge domains.

A *Hs*Hinge forms a dimer in low salt. **B** The Smc3 subunit of the *Sc*Hinge purifies as two subspecies. Both interact with the Smc1^{hinge} subunit. **C** Purification of the 2nd generation of *Sc*Hinge shows a stable Smc3^{hinge} subunit. Asterisks denote contaminants from the loading dye used.

Molecular weight determination by MS revealed the boundaries of the C-terminal truncation and therefore the construct design was revisited and the sequence of Smc3 hinge amended to remove the amino acids missing from the truncated Smc3 hinge. The second construct, *Sc*Hinge^2, was shown to purify in a single peak with no truncations (Figure 4.2-C).

4.1.5 Purification of cohesin and the loader constructs

All components were expressed in insect cells and tagged with a 2xStrep tag. For cohesin constructs only the Scc1 subunit was tagged. Both Nipbl^C and *Sc*Scc2-Scc4 were purified by affinity chromatography and SEC. Nipbl^C did not tolerate low salt and was purified into 250mM salt. Full-length human and yeast cohesin required an additional step to other strep-tagged constructs because of higher amounts of contaminants co-purifying with these protein complexes. An additional IEX step was introduced before SEC to remove contaminants.

4.2 Cohesin's affinity for the loader decreases with cohesin arm removal

4.2.1 Full-length cohesin interacts with the loader

To confirm interaction of cohesin with the cohesin loader, purified full-length human and yeast cohesin and two cohesin loader constructs were subjected to analytical SEC and glycerol gradient analysis (see Table 5 in Methods for details on all glycerol gradient parameters). *Hs*Cohesin and Nipbl^C were mixed in 1:1 molar ratio and loaded on a glycerol gradient. Gradients showed that in the gradient with both of *Hs*Cohesin and Nipbl^C samples, *Hs*Cohesin's sedimentation in the gradient is mildly shifted (Figure 4.3-A). *Sc*Cohesin and *Sc*Scc2-Scc4 showed a similar gradient elution pattern (not shown) as previously published in Chao *et al*. To gain stronger evidence of the *in vitro* interaction, *Hs*Cohesin and Nipbl^C were mixed in 3:1 ratio for a complete saturation of *Hs*Cohesin and analysed by SEC. The elution profile shows that the strength of the interaction between *Hs*Cohesin and Nipbl^C is weak but a complex does form (Figure 4.3-B). Experiments were carried out under no nucleotide conditions and could potentially be improved by addition of ATP or DNA. Upon binding of the cohesin

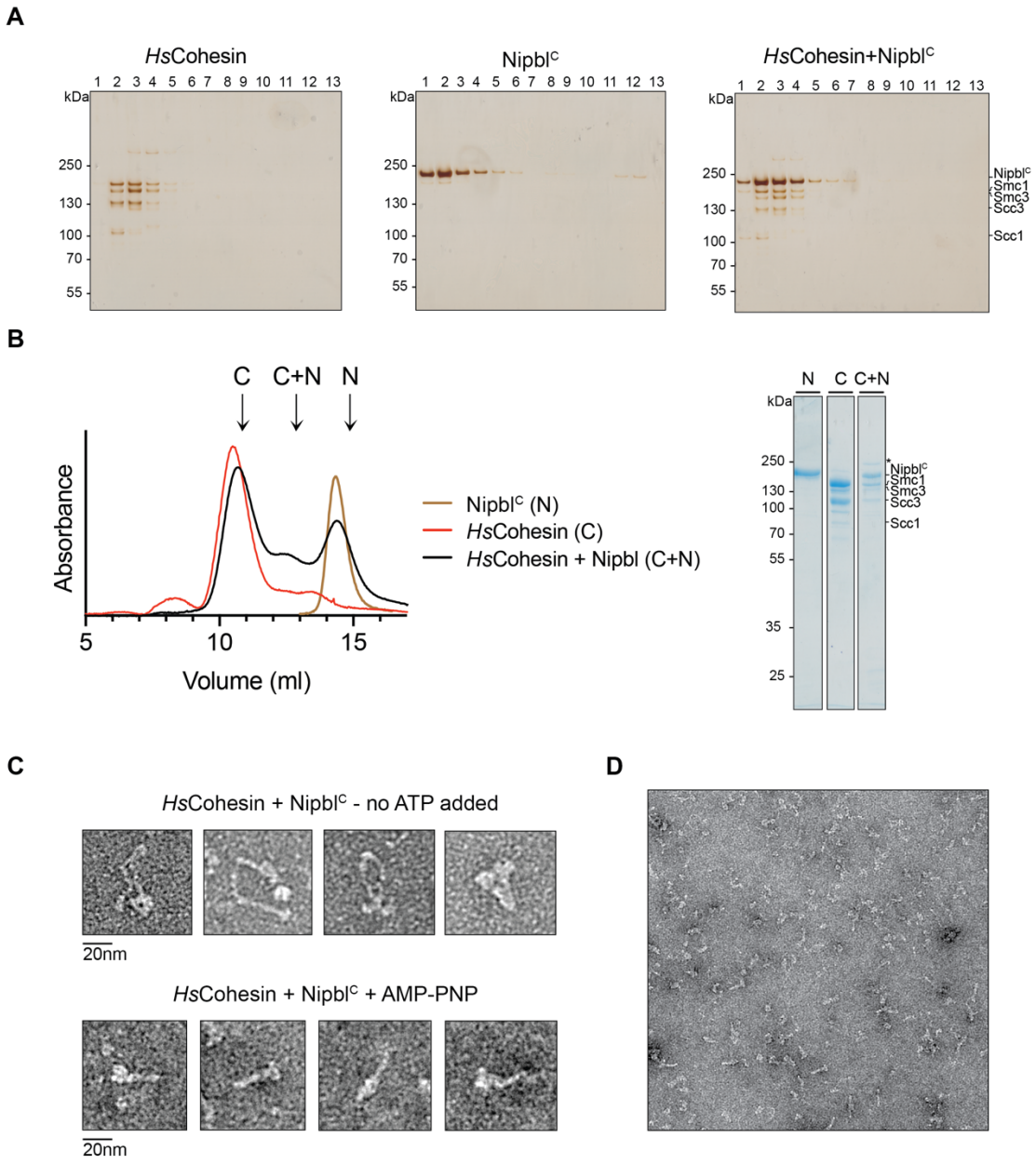

Figure 4.3 Interactions between full-length *Hs*Cohesin and its loader.

A The interaction of *Hs*Cohesin and Nipbl^{C} on glycerol gradients. **B** The complex between *Hs*Cohesin and Nipbl^{C} forms on SEC. Arrows for each peak represent a fraction shown with SDS-PAGE from the elution of the combined sample. Asterisk denotes a contaminant from purification. **C** Negatively stained particles of nucleotide-free *Hs*Cohesin adopting open conformations. Addition of a nucleotide results in coiled coil folding. **D** Negative stain micrograph of the AMP-PNP containing sample from C shows the coiled coils emerge from cohesin's rigid body in various directions. The body is bound by Nipbl^{C}.

loader and ATP, cohesin has been reported to adopt a more closed, rigid conformation where the coiled coils are expected to fold to bring the hinge and head domains in close proximity. To confirm this observation for human cohesin, the proteins were visualised by negative staining either in the absence or presence of nucleotide and the loader. *Hs*Cohesin showed a high degree of internal flexibility, its coiled coils adapting various conformations. The closed conformation was achieved by the addition of AMP-PNP and Nipbl^{C}, agreeing with previously published results of yeast cohesin (Figure 4.3-C). The closed conformation can be observed in the absence of nucleotide but with much less occurrence suggesting that a nucleotide enhances the folding of the coils and binding by the loader. Despite a more locked conformational change, *Hs*Cohesin coils were not rigid; they protruded outwards from the more rigid body of the complex but showed no distinct preference on their trajectory and remained relatively flexible when folded (Figure 4.3-D).

4.2.2 Truncated cohesin does not interact with the loader

The flexibility of the Smc coils of cohesin have shown to be problematic for structural analysis as both X-ray crystallography and Cryo-EM need predominant conformations for high resolution structure determination. As described in section 4.1.2, the constructs *Hs*Core and *Hs*Hinge which represent the structured rigid regions of cohesin were used to study the interaction with Nipbl^{C}.

Binding of Nipbl^{C} to *Hs*Core and *Hs*Hinge was investigated using glycerol gradients (Figure 4.4-A-C). Under conditions of no nucleotide or addition of both AMP-PNP and DNA[1] there was no evident binding of *Hs*Hinge to the rest of the proteins. *Hs*Core and Nipbl^{C} eluted in the same fractions either when loaded separately or together on a gradient. To further evaluate whether *Hs*Core binds to Nipbl^{C}, analytical SEC (Figure 4.4-D) and pulldowns (not shown) were performed. Strep-tagged *Hs*Core was used for pulldown experiments, however non-specific binding of Nipbl^{C} to the beads was identified and attempts to reduce non-specific binding were unsuccessful. Analytical SEC showed two separate peaks with *Hs*Core eluting in a clearly distinct

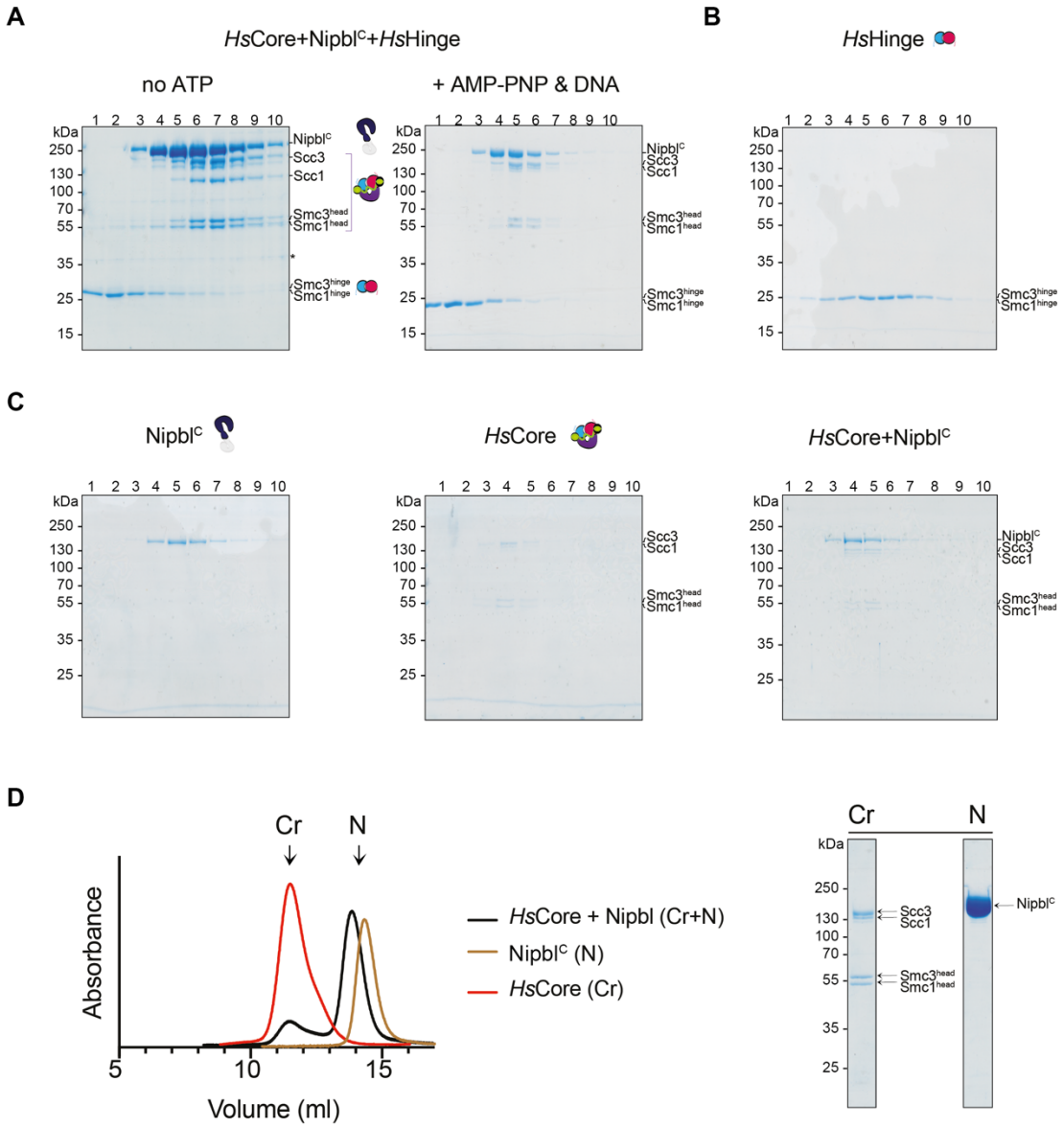

Figure 4.4 The interaction between *Hs*Core, *Hs*Hinge and Nipbl^{C}.

A *Hs*Hinge does not interact with *Hs*Core or Nibpl^{C} on glycerol gradients. This is independent of the presence of a nucleotide or dsDNA[2]. Asterisks mark a contamination from the loading dye. **B** *Hs*Hinge-only gradient in the presence of AMP-PNP and DNA as in A. **C** Glycerol gradients of Nipbl^{C}, *Hs*Core, and a combined sample, suggesting an interaction. For all gradients, the fractions are shown as numbers above the gels. **D** Nipbl^{C} and *Hs*Core do not interact on SEC. Arrows for each peak represent the fraction from the combined sample run, shown with SDS-PAGE. Only a proportion of the *Hs*Core sample was used in the combined sample run.

peak from Nipbl^{C}. Taken together, these results clearly show that affinity between *Hs*Cohesin and Nipbl^{C} are further reduced with the removal of the coiled coils, which therefore likely play a role in stabilising the formation of this complex or contain highly important binding sites. Without the coils, DNA (dsDNA^{2} or dsDNA^{3}, see Table 6) was not found to stabilise this complex either.

4.3 Structural studies show high heterogeneity and flexibility of the loader and cohesin domains

4.3.1 Negative stain EM and Cryo-EM of cohesin constructs and complexes

Despite the lack of a visible interaction on SEC, the sedimentation profile of *Hs*Core and Nipbl^{C} in glycerol gradients suggested an interaction *in vitro.* Since complex formation between the full-length *Hs*Cohesin and Nipbl^{C} also shows an excess of free proteins compared to the complex, it is plausible that Nipbl^{C} has low affinity for cohesin or requires additional components to strengthen the interaction. Similarly, binding of Nipbl^{C} to the truncated *Hs*Core might suffer from a low affinity or transient binding resulting in no complex formation on SEC, and only a partial complex formation in glycerol gradients.

To analyse the potential complex of *Hs*Core with Nipbl^{C} , the fractions of glycerol gradients showing an interaction were negatively stained and imaged (Figure 4.5-A). After 2D classification of 45,000 particles, *Hs*Core particles did not successfully classify. However, 2D classes of Nipbl^{C} 2D classes alone were visible, unbound to *Hs*Core. Nipbl^{C} has a distinctive hook-shaped architecture highly similar to Scc2 proteins with an available crystal structure and is therefore easily identifiable (see Section 4.3.2). To exclude the possibility of the acidic pH of UA stain breaking the complex, PTA and SST stains were also used, but the particle quality was not improved (not shown). Similarly, the complex was vitrified and cryo grids analysed, but no intact particles were found (Figure 4.5-B).

In an attempt to stabilise the *Hs*Core and Nipbl^{C} complex, proteins were crosslinked together using GraFix (Kastner *et al.*, 2008; Stark, 2010). Imaging separate fractions of the gradient has not yielded particles of better homogeneity (Figure 4.5-C-E).

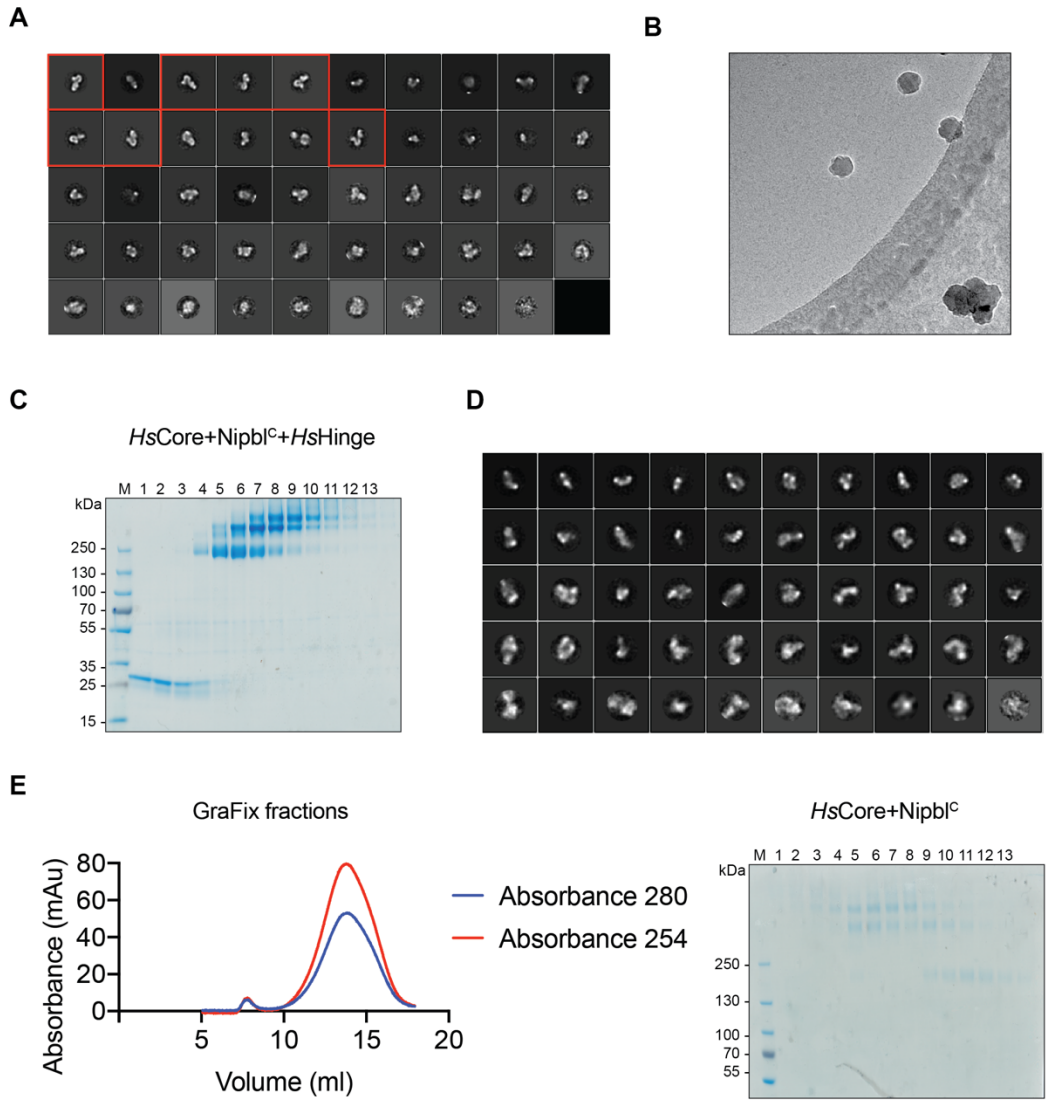

Figure 4.5 Interactions between the cohesin and loader constructs with EM.

A 2D classification reveals a separate class of Nipbl^{C} (most prominent classes highlighted in red boxes). The remaining classes could represent *Hs*Core but show a high degree of heterogeneity. **B** An example micrograph of vitrified *Hs*Core. **C** GraFix of the cohesin components and Nipbl^{C}. **D** 2D classification of negatively stained particles from fraction 10 of the GraFix gradient in C. The analysis shows flexible heterogeneous complexes. **E** Analysis of GraFix fractions, which did not contain *Hs*Hinge, with SEC show that despite the differences in molecular weight, the crosslinked complexes cannot be separated.

Nipbl^{C} classes were no longer present, suggesting that the protein has been crosslinked to *Hs*Core. Because the classes were not uniform and of insufficient quality to build a 3D volume, this led to the assumption that this complex still has high internal flexibility and is unsuitable for further structural analysis. Alternatively, the complex may exist in various stoichiometries as suggested by the presence of multiple bands on the GraFix gel that would hinder accurate alignments. An attempt to separate the crosslinked complexes seen on the GraFix gradient resulted in a single elution peak on SEC with all classes eluting more or less together, and therefore this approach was not further pursued.

4.3.2 Negative stain EM shows similarity of the human and yeast cohesin loaders

To confirm that Nipbl^{C} particles are indeed not bound to *Hs*Core in uncrosslinked samples, Nipbl^{C}-only negatively stained grids were prepared and particles classified to compare the resulting classes. Averaging the particles yielded classes where the overall protein architecture was visible, confirming that previous results indeed showed Nipbl^{C} only. Nipbl^{C} resembles the crystal structure of *Ag*Scc2^{C} (Chao *et al.*, 2017, PDB:5ME3) and *Chaetomium thermophilum (Ct)* Scc2^{C} *(Ct*Scc2^{C}*)* (Kikuchi *et al.*, 2016, PDB:5T8V). The Nipbl^{C} construct, which was based on these crystal structures, is about 50kDa larger than the *Ag*Scc2^{C}, because it contains the GD0 domain that is not present in this crystal structure (Figure 4.6-A). In addition, Nipbl^{C} contains an additional 200 amino acids on its extreme C terminus. This sequence can also be found in other species, but is not present in the *Ct*Scc2^{C} crystal structure.

To compare the human Nipbl^{C} and yeast loader *Sc*Scc2-Scc4 was negatively stained, imaged and 2D classified (Figure 4.6-B,C). The overall shape of the Scc2 subunit of the yeast loader and Nipbl^{C} are highly similar. Although Nipbl^{C} is larger, the 200 amino acids at the C terminus of Nipbl^{C} are not distinguishable. *Sc*Scc2-Scc4 classes show the flexible Scc4 module which can rotate around a central point in the vicinity of the GD0 domain. The N terminus of Scc2 in yeast is relatively short; only around 400 amino acids including the GD0 domain. Full-length Nipbl however

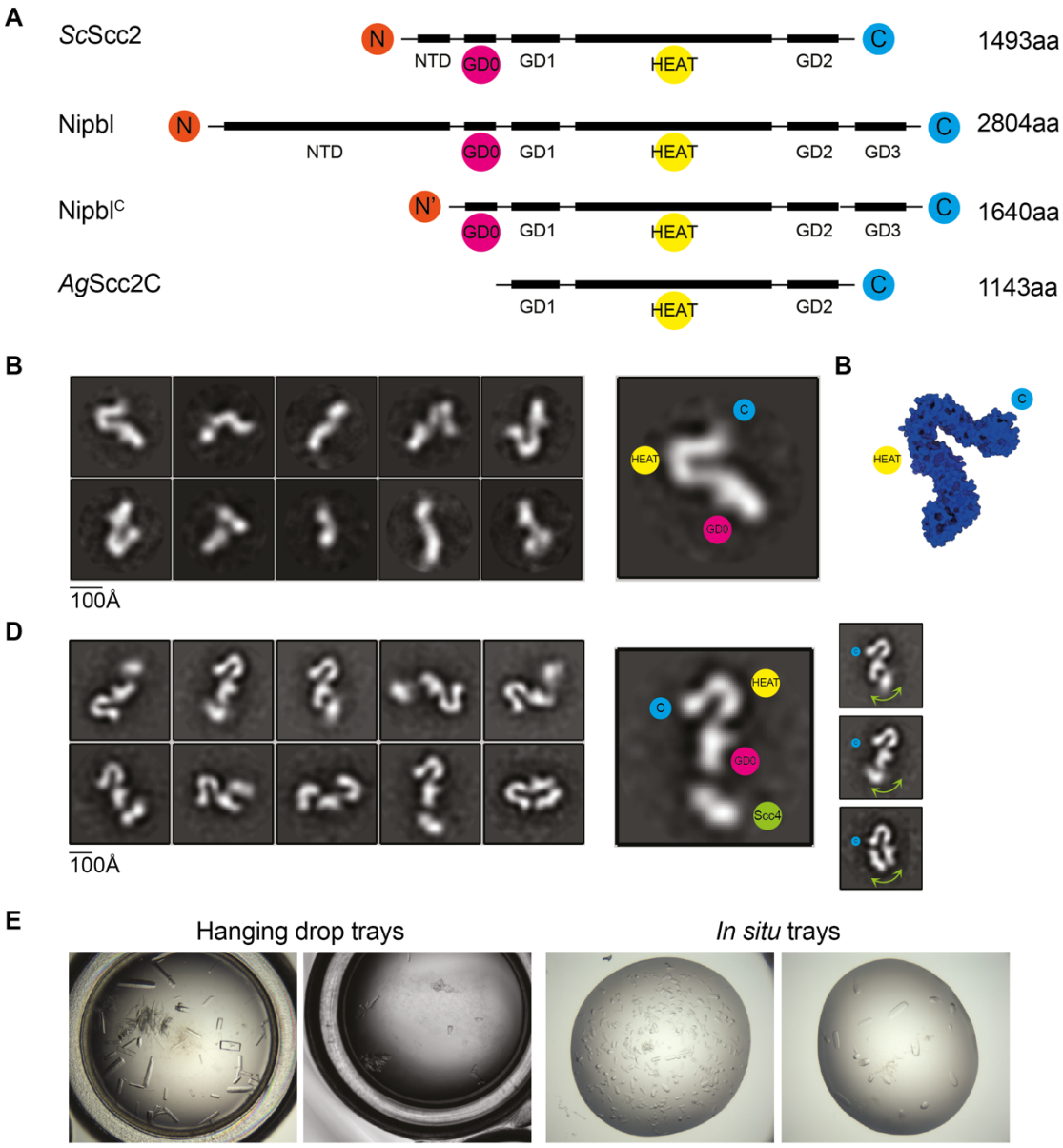

Figure 4.6 Structural overview of cohesin loaders.

A Domain architecture of Scc2 and Nipbl. Nipbl^{C} was constructed by removing the NTD of Nipbl. N' marks the new N terminus of Nipbl^{C}. **B** Crystal structure of the *Ag*Scc2C. PDB:5me3. A schematic representation of the crystal structure's sequence in shown in A. **C** 2D classification of Nipbl^{C} from UA-stained sample. The HEAT repeat domain responsible for the loader's hook shape is visible. NipblC closely resembles its yeast homolog Scc2^{C}. **D** 2D classification of *Sc*Scc2-Scc4 from UA-stained samples. The Scc4 subunit is flexible, its movement is depicted with green arrows. **E** Crystals of Nibpl^{C} from hanging drop trays and *in situ* trays in the same crystallisation condition.

has an N terminus larger by over two-fold, despite Scc4 length being similar in yeast and human. Due to this the overall architecture of the full-length human cohesin loader might be different to that of yeast to accommodate for other functions of Nipbl (Luna-Peláez *et al.*, 2019).

4.3.3 Crystallisation trials of Nipbl^{C}

The Nipbl^{C} construct was successfully classified into 2D classes. Although its hook shape resembles that of other species, Nipbl^{C} is larger with an additional domain of which there is no structural information. Crystallisation trials were performed with the aim to study the structure of Nipbl^{C}. Trials identified a single optimal condition for crystal growth, which was successfully reproduced in hanging drops to obtain larger crystals (Figure 4.6-D). Cryoprotection with various cryoprotectants has led to only low resolution diffraction, in several cases crystals were dissolved by the cryoprotectant. To test if the crystals diffract at all, they were grown on plates suitable for *in situ* shooting. No high resolution diffraction was observed suggesting that the lack of high resolution diffraction is most likely an intrinsic property of the crystals possibly due to the flexible nature of the protein.

4.3.4 DNA binding to the loaders varies across species

The cohesin loader was previously reported to bind DNA, the Scc2 subunit being mainly responsible for the binding, with a strong preference for dsDNA over ssDNA (Murayama and Uhlmann, 2014; Chao *et al.*, 2015). Because binding of DNA to the loader could be crucial for understanding recruitment and entry of DNA to the cohesin ring, the binding capacity of both *Sc*Scc2-Scc4 and Nipbl^{C} were tested by EMSAs (Figure 4.7-A). As *Sc*Scc2-Scc4 was not able to penetrate native PAGE gels, DNA gel shifts were performed in agarose gels, Nipbl^{C} in native gels. Equal amounts of FAM-dsDNA^{3} were incubated in the presence of increasing concentrations of individual proteins. Because *Sc*Scc2-Scc4 was able to bind DNA more efficiently than Nipbl^{C}, the yeast loader was used for structural characterisation of DNA binding. To draw conclusions about low affinity of Nipbl^{C} for DNA full-length Nipbl protein bound to human Scc4 would be required for these experiments. Furthermore, EMSAs with lower DNA concentrations would need to be performed, as loss of DNA

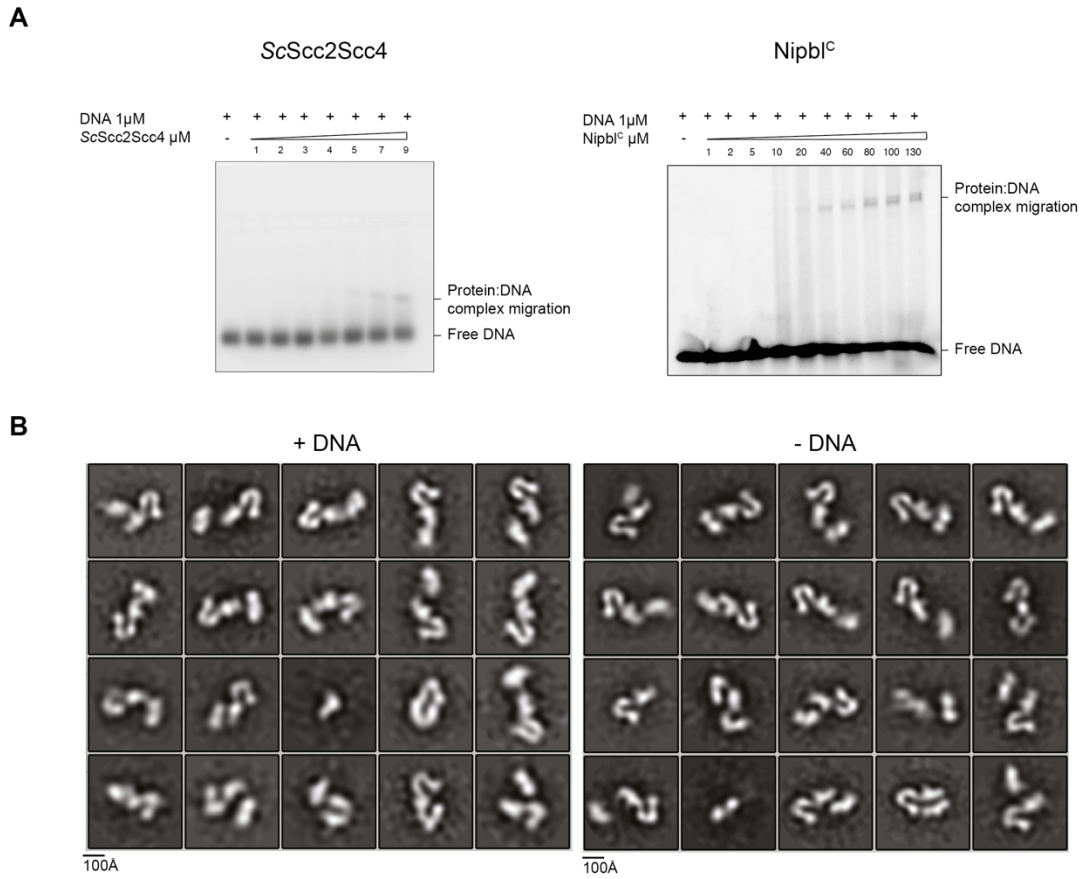

Figure 4.7 DNA binding to cohesin loaders.

A *Sc*Scc2-Scc4 and Nipbl^{C} EMSAs in agarose and native gels, respectively. The yeast loader appears to bind DNA more efficiently. **B** Comparison of 2D classes with and without DNA shows no significant change in the conformation of *Sc*Scc2-Scc4.

in the unbound lane could also suggest DNA binding but is not apparent in the below presented gels due to high amounts of DNA used.

To compare the loader structures with and without DNA, *Sc*Scc2-Scc4 was either directly negatively stained or mixed with DNA prior to staining. After eliminating contamination identified by an initial round of 2D classification, particles were then broadly 2D-classified to capture any variations in their 2D classes but no differences were observed (Figure 4.7-B). Furthermore, no DNA was present in any of the 2D classes, possibly because of low affinity and fast dissociation rate. DNA is not expected to be visible by negative staining and it therefore cannot be concluded whether DNA is or is not bound to the protein. For such conclusion, a higher resolution structure would be required but given the high flexibility of the complex, structure determination by Cryo-EM was not pursued.

4.3.5 Core construct optimisation

Topological entrapment of sister chromatids is an ATP-dependent process. The loader loads cohesin onto DNA in an ATP-dependent fashion where it increases cohesin's ATP hydrolysis rates. This is further supported by observations that ATPase dead mutants are unable to load cohesin on DNA (Murayama and Uhlmann, 2014). ATP hydrolysis is generally accompanied by structural rearrangements in the protein. To identify these changes, *Hs*Core was not only studied in the presence of Nipbl^{C} but also in its absence, to observe the structural configuration of the ATPases prior to nucleotide hydrolysis.

The human cohesin core construct was therefore studied in parallel to the interaction and crosslinking studies of cohesin and its loader (Figure 4.8). The SEC elution profile of *Hs*Core showed a single peak, suggestive of an intact complex but the analysis of this elution with negative staining showed a heterogeneous sample. Attempts to 2D classify particles identified a single protein as the most populated 2D class. This protein structurally resembled the HEAT-repeat protein Scc3. Its presence as a single class points to its existence as a single subunit unbound to the

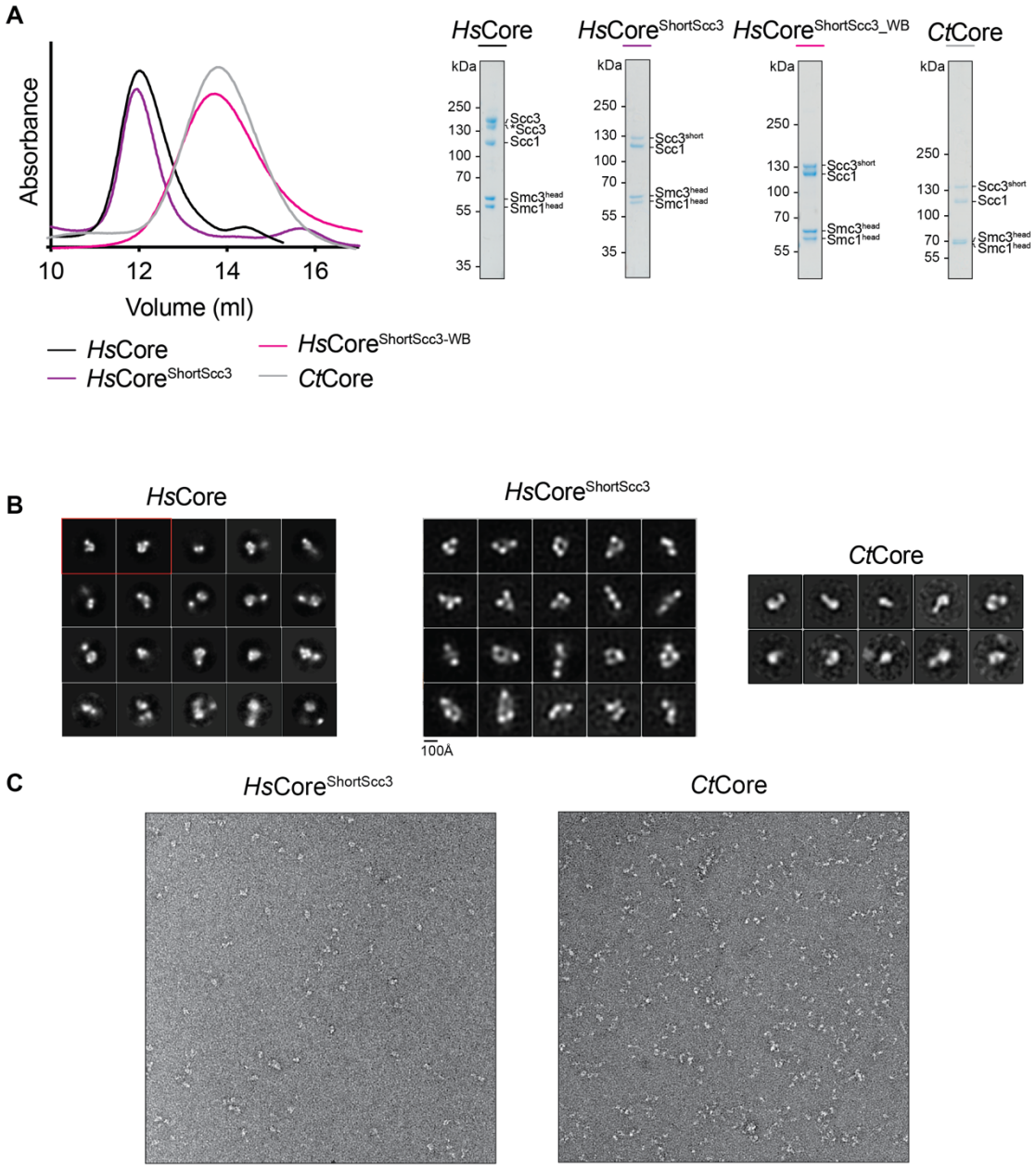

Figure 4.8 Optimisation of the core constructs for EM analysis.

A SEC analysis of four core constructs suggests a more compact conformation of the *Ct* and Walker B mutant construct. **B** 2D classes oof the core constructs. With *Hs*Core, the Scc3 subunit is clearly visible as a separate class (highlighted in a red box), but is not present after the modification of the Scc3 construct for *Hs*Core^{ShortScc3}. **C** Micrographs of two core constructs with the most alike sequence design.

complex, possibly due to the complex falling apart. Further experiments have revealed an instability of Scc3 resulting in a degradation visible on SDS-PAGE. It is important to note that the instability of Scc3 could have been present in previous samples too, but because Nipbl^{C} and Scc3 share a highly similar overall shape, the presence of Scc3 as single particles could have gone undetected. These observations led to a re-design of the construct, where Scc3 was shortened based on the available crystal structure (Roig *et al.*, 2014). The new *Hs*Core with shortened Scc3, termed *Hs*Core^{ShortScc3}, was not prone to degradation as judged by SDS-PAGE. This led to improved 2D classes, where classification of 22,000 particles did not reveal a separate class for Scc3. Overall, the construct did not appear more homogeneous despite the addition of AMP-PNP and $MgCl_2$ which should induce head engagement. To confirm that the addition of nucleotide and subsequent engagement of the ATPases indeed does not stabilise this complex, *Hs*Core^{ShortScc3} with two WalkerB motif mutations Smc1^{E1157Q} and Smc3^{E1144Q} were analysed, termed *Hs*Core^{ShortScc3_WB}. These mutations allow binding but not hydrolysis of a nucleotide. Introducing these mutations had no effect on the complex, which remained flexible with inadequate 2D classes to progress further (not shown).

Studying the human core constructs have not resulted in the formation of a stable complex. *S. cerevisiae* Core (*Sc*Core) and *C. thermopilum* Core (*Ct*Core) were therefore constructed for testing. *Sc*Core construct contained full-length Scc3 protein, whereas *Ct*Core contained a shortened Scc3 based on the Scc3 crystal structure as previously done with the *Hs*Core^{ShortScc3} construct. Whereas *Sc*Core failed to express, *Ct*Core showed good yields and high purity with a single elution peak on SEC. Moreover, this protein could be purified both at 4°C and at room temperature, suggesting higher stability than the previous *Hs*Core construct. Negative stain micrographs suggested a slight improvement in particle quality, but after classification of several thousand particles it was concluded that this construct was not optimal for structural analysis either, as the classes obtained were highly heterogeneous (Figure 4.8-B,C).

Interestingly, the four constructs showed distinct elution times on SEC despite being very similar in their molecular weight. The largest construct *Hs*Core is about 330kDa in weight; by shortening Scc3 to obtain a more stable construct, *Hs*Core^{ShortScc3} is

lower in size by 30kDa and yet elutes in the same volume as *Hs*Core. The Walker B mutant has the same molecular weight as *Hs*Core^{ShortScc3} but its elution volume is significantly shifted (Figure 4.8-A). This could suggest a more compact closed conformation of the ATPase heads in mutants which cannot hydrolyse ATP. Although no ATP was added to the purified protein, some ATP could have been bound to the ATPases during protein expression in the cells and because of the introduced mutations was never hydrolysed. Intriguingly, the *Ct*Core construct is almost the exact same molecular weight as *Hs*Core^{ShortScc3} but also elutes later. This could be due to changes between organisms and stability as *Ct*Core has shown to be more stable than *Hs*Core^{ShortScc3} when analysed by NS. Importantly, these changes in elution cannot be attributed to differential DNA binding, as all complexes are treated with benzonase and the absence of DNA confirmed by UV spectroscopy.

4.3.6 Crystallisation of cohesin hinge domains

The basic inner patch of the hinge domain can interact with both ssDNA and dsDNA, an event not influenced by the presence of ATP (Hirano and Hirano, 2006). DNA binds the region between the hinge and the adjacent coiled coil, a region that is obstructed by the coil if cohesin is in a closed, rod-shaped conformation. It has been proposed that this region becomes uncovered after a conformational change which results in pulling the coils apart and exposing the DNA binding site on the hinge (Soh *et al.*, 2015). Despite multiple structures of the hinge domain being solved (Haering *et al.*, 2002; Li *et al.*, 2010; Kurze *et al.*, 2011; Alt *et al.*, 2017), there is no DNA-bound hinge structure.

In an attempt to get a better understanding of the nature of the interaction between DNA and the hinge domain of cohesin, *Sc*Hinge^2, *Hs*Hinge and the monomeric *Hs*Smc3^{hinge} were examined for their DNA binding properties using EMSAs (Figure 4.9-A). Comparing ScHinge^2 and *Hs*Hinge, both construct were shown to both bind DNA efficiently and with similar affinities. Although *Hs*Hinge only contained a short stretch of the coils emerging from each monomer, *Hs*Smc3^{hinge} could potentially have an even more exposed DNA binding surface. When the dimer was separated, *Hs*Hinge and *Hs*Smc3^{hinge} showed no preference for ssDNA^2 or dsDNA^2 (Figure 4.9-B) but to estimate the affinities more precisely, EMSAs with less DNA concentrations

would need to be performed, as there is no loss of DNA from the unbound fraction visible.

The proteins were then subjected to crystallisation trials with dsDNA[1] (Figure 4.9-C,D). For *Hs*Hinge, four concentrations were tested both in the presence and absence of DNA. The highest concentration was tested first, which resulted in several conditions producing distinct crystals. Crystals were cryoprotected with different cryoprotectants but no diffraction was observed. Screening a range of concentrations revealed that lower concentrations appear more favourable and several new crystallising conditions were identified where the protein did not crystallise in higher concentrations. Crystals from these conditions were tested for diffraction *in situ* but no diffraction was observed from any of these crystals, suggesting that the *Hs*Hinge construct was not stable enough, and could not be further stabilised by DNA. No hits were found for *Hs*Smc3^{hinge}.

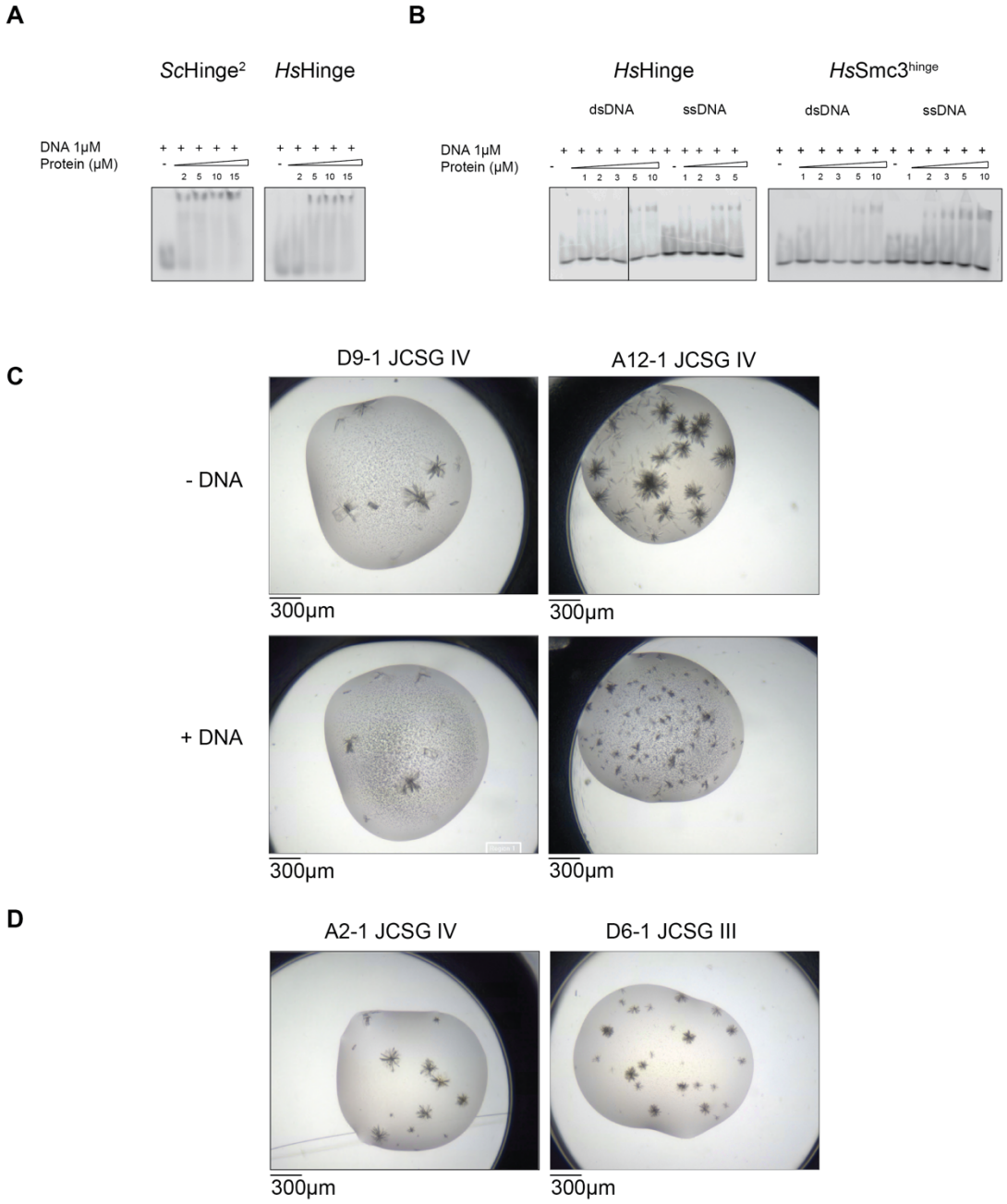

Figure 4.9 DNA binding properties of the hinge domains.

A DNA binding to yeast and human hinge heterodimers. **B** DNA binding to the monomeric and heterodimeric human hinge. **C** Two most successful conditions for *Hs*Hinge at 13mg/ml concentration, either with or without 10bp DNA. **D** *Hs*Hinge at 8mg/ml concentration with dsDNA. Crystals were shot in situ but no diffraction was observed.

Chapter 5. Results 2 – Characterisation of the Chl1 helicase

Chl1 is a 5'-3' helicase belonging to the XPD subfamily along with three other proteins: Rtel1, FancJ and XPD. Structure predictions suggest a common architecture for all proteins yet only the structure of the XPD protein has been solved (Rudolf *et al.*, 2006; Fan *et al.*, 2008; Liu *et al.*, 2008). This protein is known to bind multiple factors found at the replication fork, including cohesin, Ctf4 and Tof1, where it is believed to help establish sister chromatid cohesion and aid in the restart of stalled replication forks and in HR after DNA damage (Cali *et al.*, 2016; Samora *et al.*, 2016; Delamarre *et al.*, 2019). Despite having numerous identified interactions and functions, its precise role in these processes is unknown.

5.1 Purification and characterisation of CtChl1

Previous expression of the Chl1 protein in the lab has shown that the *S. cerevisiae* Chl1 expressed in insect cells gives low yields. The *C. thermophilum* Chl1 (*Ct*Chl1) was therefore chosen as the next test subject as proteins of this thermophilic fungus are often used in crystallography for their thermostability and high yields. *Ct*Chl1 was expressed in insect cells and purified using affinity chromatography and size exclusion chromatography. The protein yields were significantly higher than for the yeast ortholog and could be purified to high purity (Figure 5.1-A,B). *Ct*Chl1showed no decreased stability in low salts and was therefore purified into 150mM NaCl.

To test the optimal buffer for the highest protein stability, thermal shift assays were performed where Bis-Tris was used for screening due to its large buffering capacity (Figure 5.1-C). These assays showed that pH 8.5 is the most optimal pH for the protein. Conversely, buffers with pH far below the isoelectric point (pI) of *Ct*Chl1 (pI = 7.6) showed the fastest protein unfolding with increasing temperature. SEC-MALS was performed to determine the oligomeric state of the protein across a range of concentrations. At all three concentrations tested *Ct*Chl1 eluted within the same elution volume in a monomeric form (Figure 5.1-D).

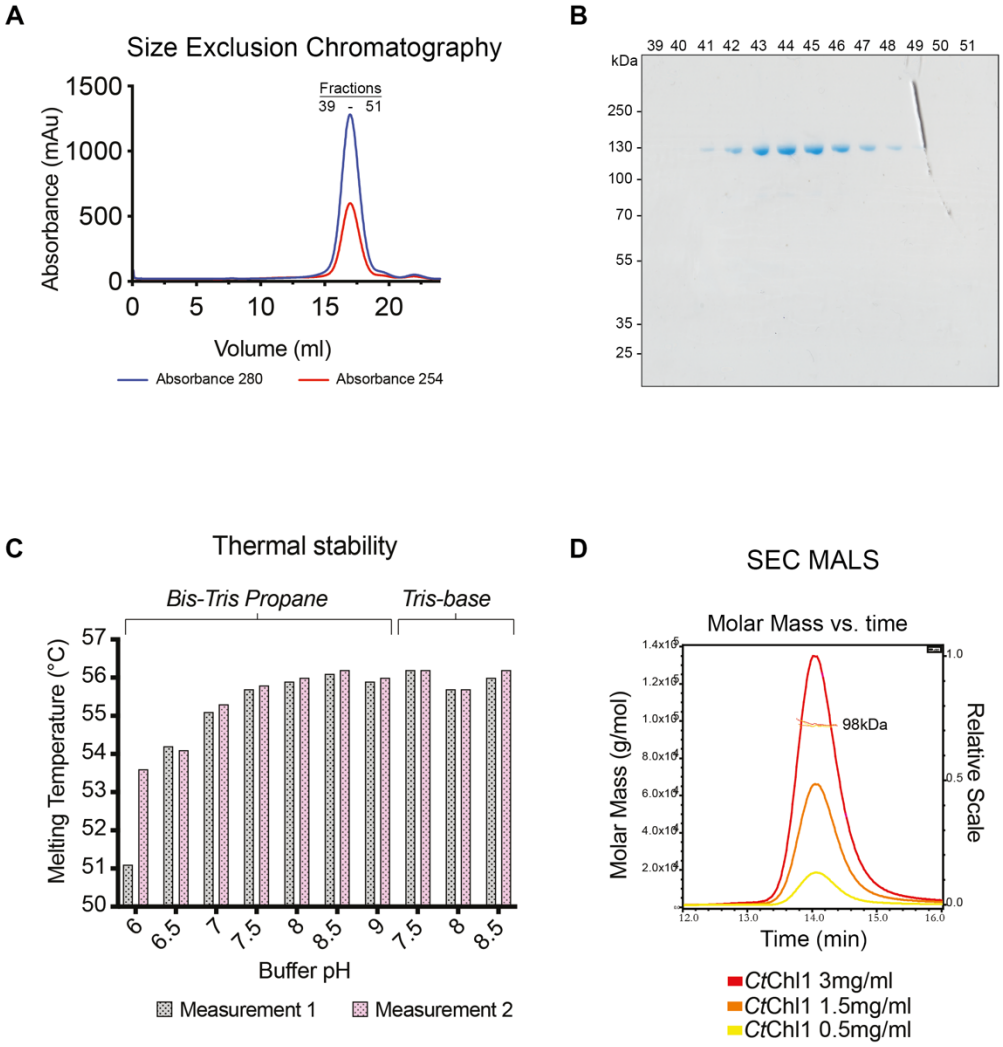

Figure 5.1 Characterisation of *Ct*Chl1.

A SEC elution profile. **B** SDS-PAGE analysis of the elution peak from A. **C** Thermal stability assays. The pH range around the isoelectric point of the protein (7.6) is the most favourable. **D** *Ct*Chl1 does not form oligomers.

5.2 Structural insights into Chl1

5.2.1 Predicted domain architecture

The published structure prediction for Chl1, as described in Samora *et al,* is based on the known structure of XPD and the estimation that all XPD subfamily proteins share a common architecture: two helicase domains separated by an Arch domain which orchestrates DNA unwinding together with the highly essential Fe-S cluster. This is supported by sequence alignments and structure predictions. There is no homology or structural information on the 20kDa “insert” domain of Chl1 and FancJ, a domain which lies adjacent to their WalkerA motifs. At least in humans, inserts of both proteins form interactions which contribute to replication stress responses. Sequence alignments suggest that these inserts may have some difference in properties, for example FancJ’s insert is positively charged as opposed to the negatively charged insert of Chl1.

5.2.2 Negative staining

Two structural approaches were carried out in parallel to obtain the structure of Chl1: X-ray crystallography and Cryo-EM. Whereas crystallising the protein has not been successful (discussed in section 5.2.6 and Discussion), staining *Ct*Chl1 with the heavy stain UA has shown a well-behaved monomeric protein sample with no aggregates. Collection of around 100 micrographs of negatively stained *Ct*Chl1 and subsequent classification resulted in the first 3D volume of the protein (Figure 5.2-A). Because of negative staining resolution limitation, the position of individual domains of Chl1 were not determinable. Despite the protein’s small size of only 100kDa, this promising first glimpse has led to focusing on Cryo-EM as the method of choice to obtain the structure of *Ct*Chl1.

5.2.3 Cryo grid optimisation

The initial optimisation of cryo grids involved screening grid types, concentrations and blot times (Figure 5.2-B-D). Initial screening was performed on a 120kV screening microscope but in many instances the particles were not visible . Even in thinner ice it was difficult to distinguish between particles and noise. The screening of grids was therefore performed on the 200kV instrument with a DDD where the

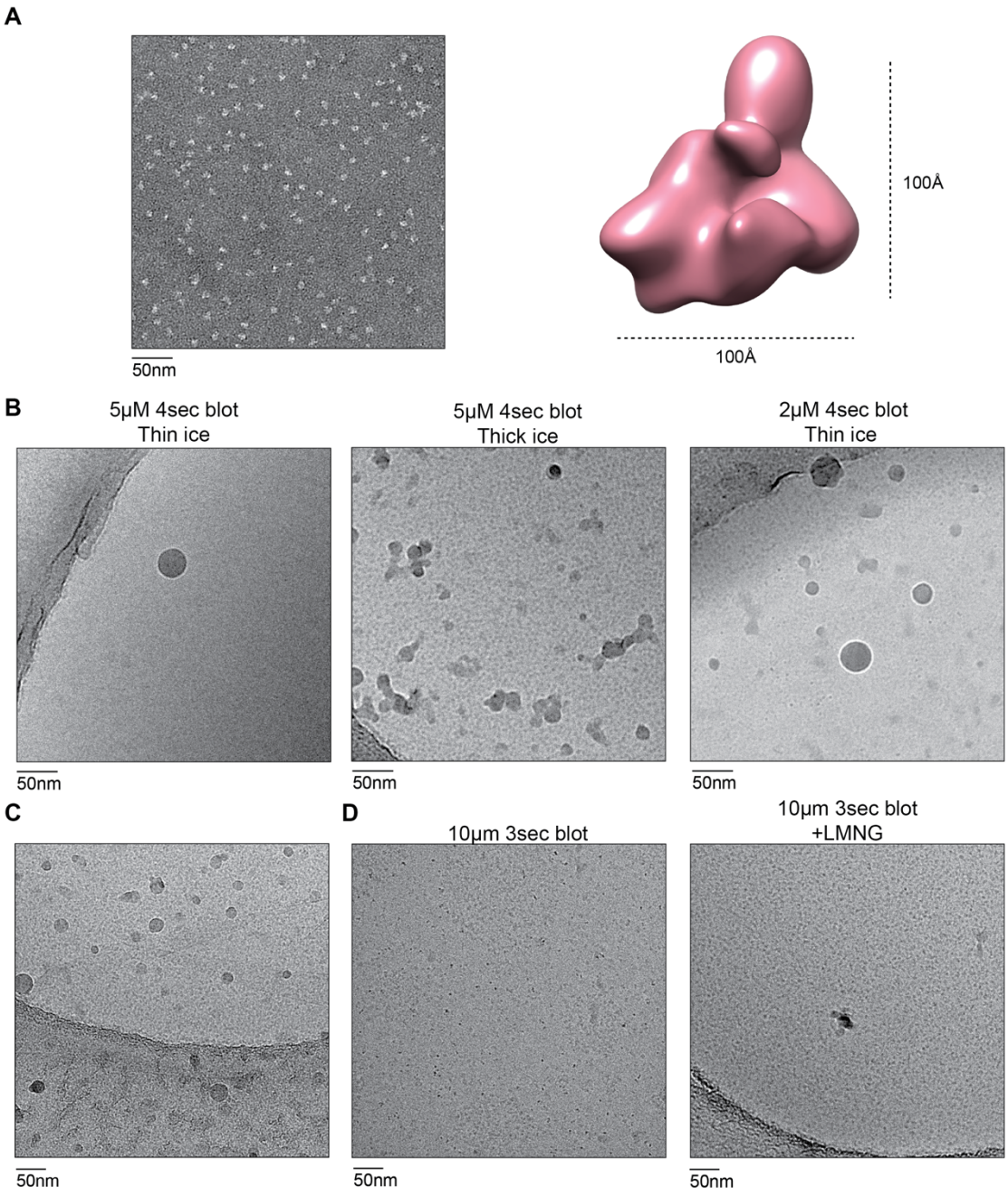

Figure 5.2 Negative staining and freezing conditions optimisation.

A UA stained grids and the first 3D envelope of *Ct*Chl1. **B** Optimisation of open hole Quantifoil grid freezing conditions. **C** Quantifoil grids coated with GO. The folds of the coating are visible near the carbon support. **D** C-flat grids with and without detergent.

visibility of particles in ice was greatly enhanced, as opposed to the previously used 120kV screening microscope with a CCD. Out of the screening conditions tested (see Table 6 in Methods), fresh open-hole C-flat grids not subjected to glow discharging showed the most optimal particle behaviour. Initial vitrification showed that particles localised in the open holes but were not uniformly distributed in ice, forming small clusters that would interfere with data collection. This issue was overcome by the addition of the LMNG detergent in low concentrations, which resulted in the separation of particles and their uniform distribution across the hole (Figures 5.2-D & 5.3-A).

5.2.4 Structure of *Ct*Chl1

The first map of *Ct*Chl1 was obtained from a data collection from C-flat grids with LMNG imaged with a 200kV instrument. A total of 133 665 particles contributed to the reconstruction of a 12Å Cryo-EM envelope of *Ct*Chl1 with several weakly visible features of its overall architecture (Figure 5.3-B-D). The protein was shown to be under 100Å in diameter. The small molecular weight and lack of strong features prevented obtaining secondary structure information. Despite this, 3D volumes of the protein showed the overall shape of *Ct*Chl1: a larger body of the protein presumably formed by the two helicase domains, and a domain separate from the body extending above it, reminiscent of the XPD Arch domain. Given the low resolution, no definite statements about the architecture of *Ct*Chl1 could be made.

To complement the lack of visible secondary structure features the frozen sample was imaged on the 300kV Titan Krios equipped with a K2 camera and an energy filter(Figures 5.4 and 5.6). A total of approximately 1.8 million particles was collected and upon classifying into 2D classes further features of the protein became apparent when compared to the Talos dataset. The final refinement converged at 7.7Å resolution as determined by the FSC curve. The overall architecture of *Ct*Chl1 agrees with the observed domains of XPD and predicted domains for Chl1 (Figure 5.5). *Ct*Chl1 can be separated into three domains: the two helicase domains, HD1 (which includes the Fe-S domain) and HD2, and the Arch domain. The Arch domain is clearly visible as it extends above the two helicase domains. Fitting XPD maps into the *Ct*Chl1 density reveals that the Arch domain of *Ct*Chl1 is larger. It is important to

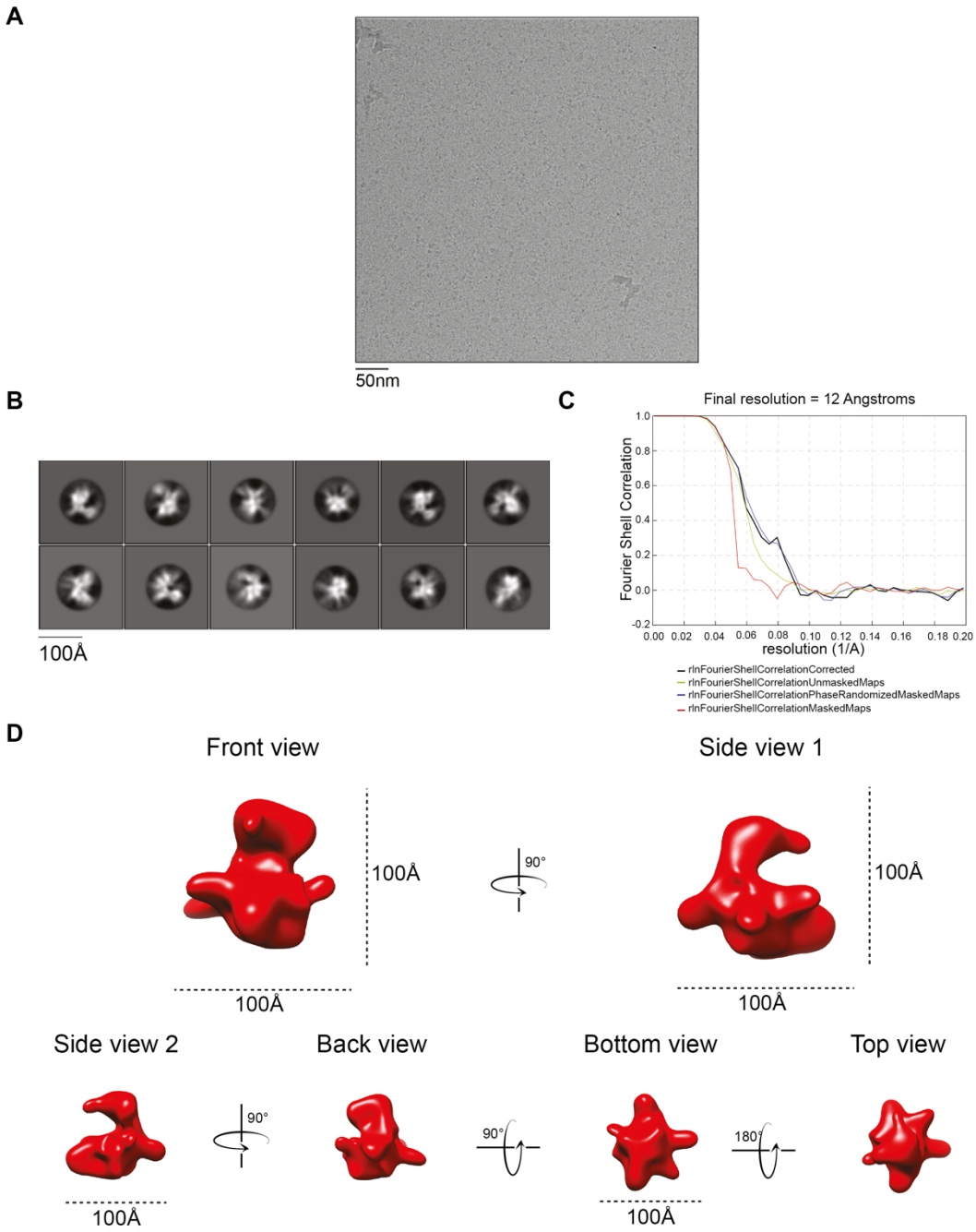

Figure 5.3 Low-resolution information on the architecture of *Ct*Chl1.

A C-flat grids with 2µM *Ct*Chl1 and LMNG detergent used for data collections. **B** 2D classification of *Ct*Chl1 form the Talos Arctica dataset. **C** The Fourier shell correlation (FSC) curve from Relion3.08 of the final 3D model. **D** Views of the final 3D model with the corresponding angle distributions.

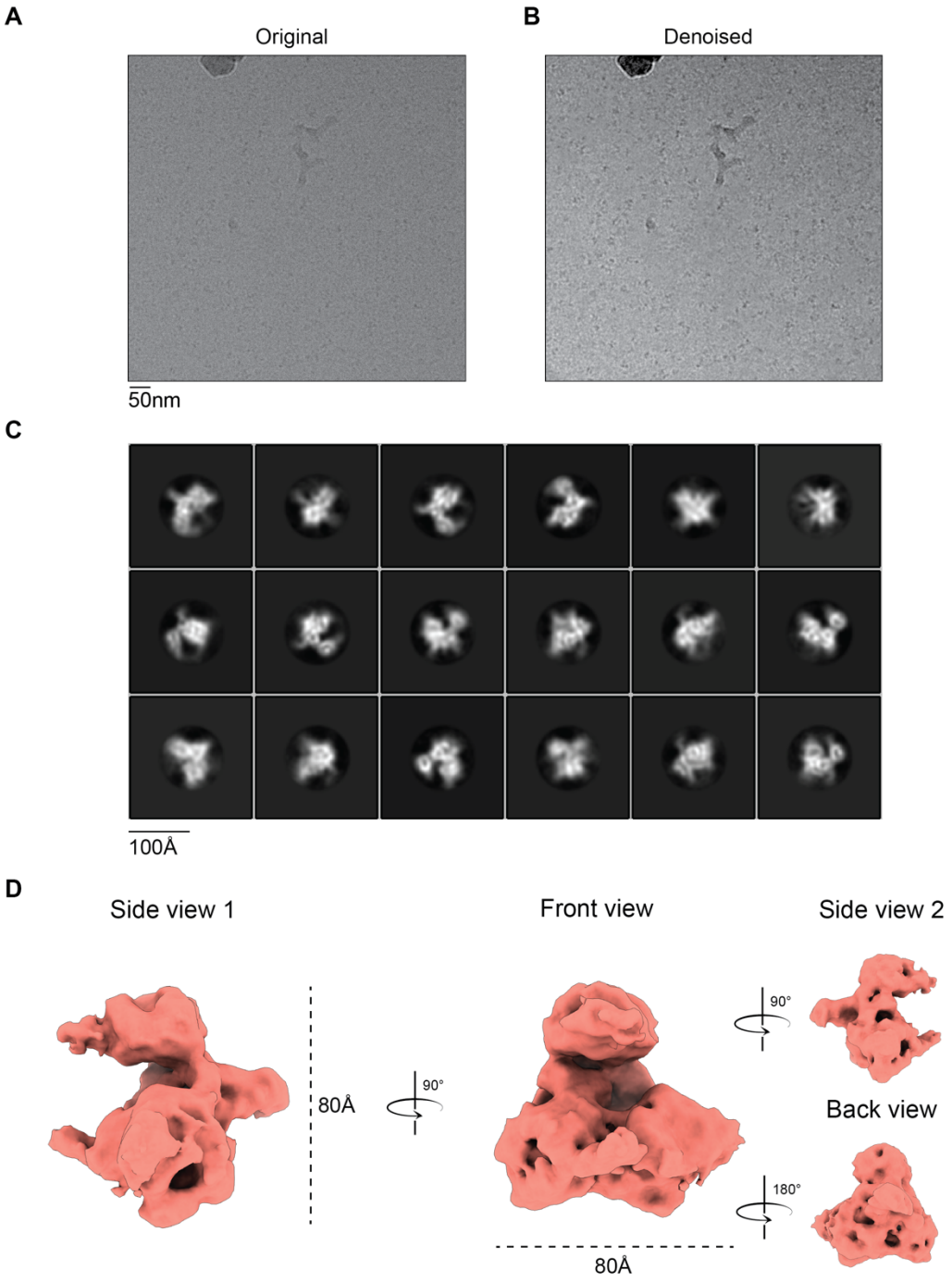

Figure 5.4 The 7.7Å reconstruction of *Ct*Chl1

A A representative micrograph from the data collection. **B** the same micrograph as in A denoised in CrYOLO to visualise particles. **C** Final 2D classes. **D** Final reconstruction of *Ct*Chl1 using Sidesplitter.

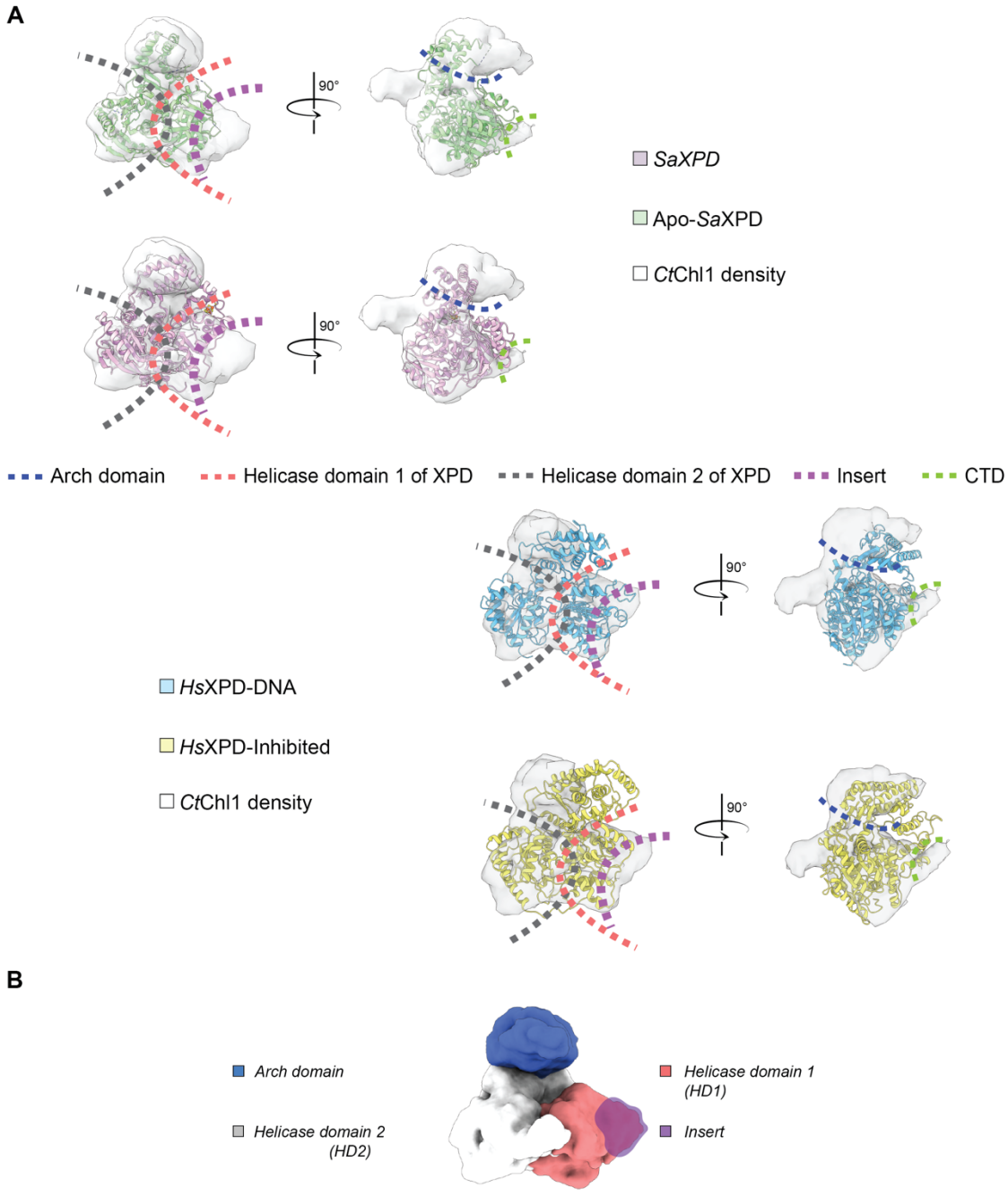

Figure 5.5 The architecture of *Ct*Chl1.

A Fitting of nucleotide-free Apo-*Sa*XPD and nucleotide-bound *Sa*XPD, human XPD protein from TFIIH in its DNA-bound conformation (DNA not shown) and inhibited state into the *Ct*Chl1 density. **B** Overall description of *Ct*Chl1 architecture.

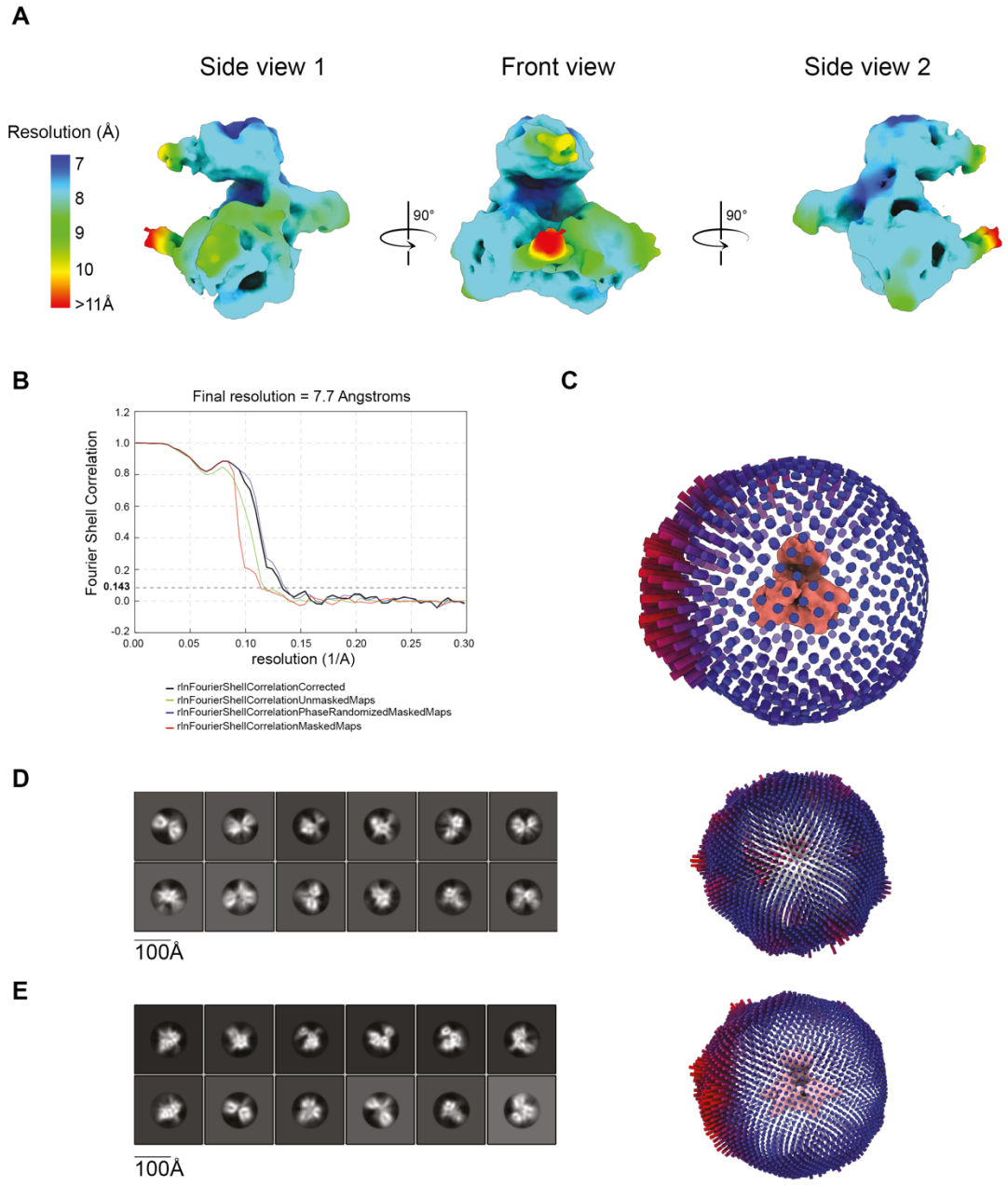

Figure 5.6 Local resolution and angular distributions of *Ct*Chl1.

A Local resolution, reconstructed in Cryosparc2, estimates the least resolved density to be in HD1. **B** Final resolution of *Ct*Chl1 as judged by the FSC curve. **C** Angle distribution of the particles contributing to the final model. Red bars represent the most represented orientations, blue bars represent less represented orientations. **D** The angle distribution changes with introducing the OG detergent. **E** A merged dataset of LMNG and OG detergent datasets shows that the majority of particles contributing to the highest resolution map comes from the LMNG detergent dataset. *All angle distributions show the same (front) view.

note that bacterial and archaeal proteins are overall smaller in size than eukaryotic, and a difference in Arch domain sizes across the XPD subfamily is expected. Indeed, this has been observed for XPD in TFIIH where the plug region of the Arch domain enlarges this domain and such region is not found in the prokaryotic structures. In this structure the DNA binding cleft of XPD is occupied by the plug of the Arch domain as a result of the inhibitory effect of MAT1 on XPD (refer to Section 1.10.2 and Figure 1.9) . XPD is therefore in a more closed conformation even in the absence of DNA.

Conversely, *Ct*Chl1, which is in its nucleotide-free state, shows that the Arch domain is not folded towards the helicase domains. The DNA-free conformation for XPD and Chl1 might therefore differ. Fitting in the structures of XPD further reveals that the handedness of the *Ct*Chl1 map is correct, as the HD2 appears to correspond to the left side of the CtChl1 density. Structure prediction suggested similar folding of HD2 of Chl1 to XPD despite HD2 not being strongly conserved (Samora *et al.*, 2016). Within the HD2, the extreme C-terminal of proteins in the XPD subfamily has diverse lengths, with XPD having the shortest C-tail. Compared to XPD, the tail of Chl1 is about 50 amino acids longer, which could explain empty densities when either XPD structure is fitted into the *Ct*Chl1 map. As there is no structural information available on FancJ or Rtel1, the increasing size of HD2 cannot be further compared.

Comparing the right-sided density of the map to the HD1 of any XPD structure is more difficult, as there are expected variations between Chl1 and XPD. Chl1 proteins contain an additional insertion of around 150 amino acids in HD1, which is located just after the helicase motif II. The *Ct*Chl1 map shows an enlarged HD1 compared to XPD, where the additional densities have been attributed to the location of the insert (Figure 5.5). The obtained resolution does not allow to build any segment of *Ct*Chl1 *de novo*. The resolution could not be improved by any modifications in the processing pipeline. A decrease in the number of particles did not cause a decrease in resolution, suggesting that collecting a larger volume of data would not solve this problem. The insert and the surrounding HD1 is the least resolved segment of the map (Figure 5.6-A). One of the factors contributing to this is the particle orientation distribution. Although all orientations are represented there is a clear preference for the HD2 orientation where the lack of particles oriented in the HD1 direction could lead to poor resolution of this domain (Figure 5.6-C). The main factor that likely

contributes to the lowest resolution of this domain is the flexibility of the insert, which is predicted to be partially disordered by secondary structure prediction servers such as PsiPred. A change in particle orientation could improve the map of the surrounding HD1, as the dataset is anisotropic, but would not lead to obtaining high resolution information on the disordered insert. To confirm the insert's flexibility is the cause of low resolution in the insert area, a change in particle orientation was achieved by changing the detergent used for vitrification (Figure 5.6-D). Low amounts of OG were added to the sample prior to vitrification and a dataset with the same parameters was collected on the 300kV instrument. This dataset has introduced more particles with the HD1 orientation but merging the two datasets shows that there is no improvement in the final resolution (Figure 5.6-E). Thus, the low resolution of the insert can be attributed to its flexibility.

5.2.5 Studying the conformational states of Chl1

The above presented structure is the *Ct*Chl1 protein in its nucleotide and DNA free form. Because ATP binding, ATP hydrolysis and DNA binding are essential for the function of helicases, *Ct*Chl1 was supplemented with a nonhydrolyzable ATP analog ATPγS or in the presence of both ATPγS and ssDNA^4 (supplemented with $MgCl_2$). This was followed by negative stained to determine whether *Ct*Chl1 changes its conformation in response to nucleotide and DNA binding in a similar fashion to XPD. Conformational changes could not be determined from the very low resolution negative staining data. The sample was therefore vitrified in the presence of ATPγS and ssDNA^4 to gain higher resolution information, which would show whether the nucleotide and DNA induce a conformational change. The same grid preparation , including the addition of LMNG, was used for the nucleotide and DNA bound sample as for the *Ct*Chl1. Classification of a final number of 93 013 particles has not revealed any changes to the structure as judged by the good agreement of the *Ct*Chl1 map and the nucleotide and ssDNA-bound *Ct*Chl1 map (ssDNA-*Ct*Chl1) (Figure 5.7). Despite the low visibility of secondary structure features, the Arch domain, which is visible in both maps in an extended open conformation above the body of the helicase.

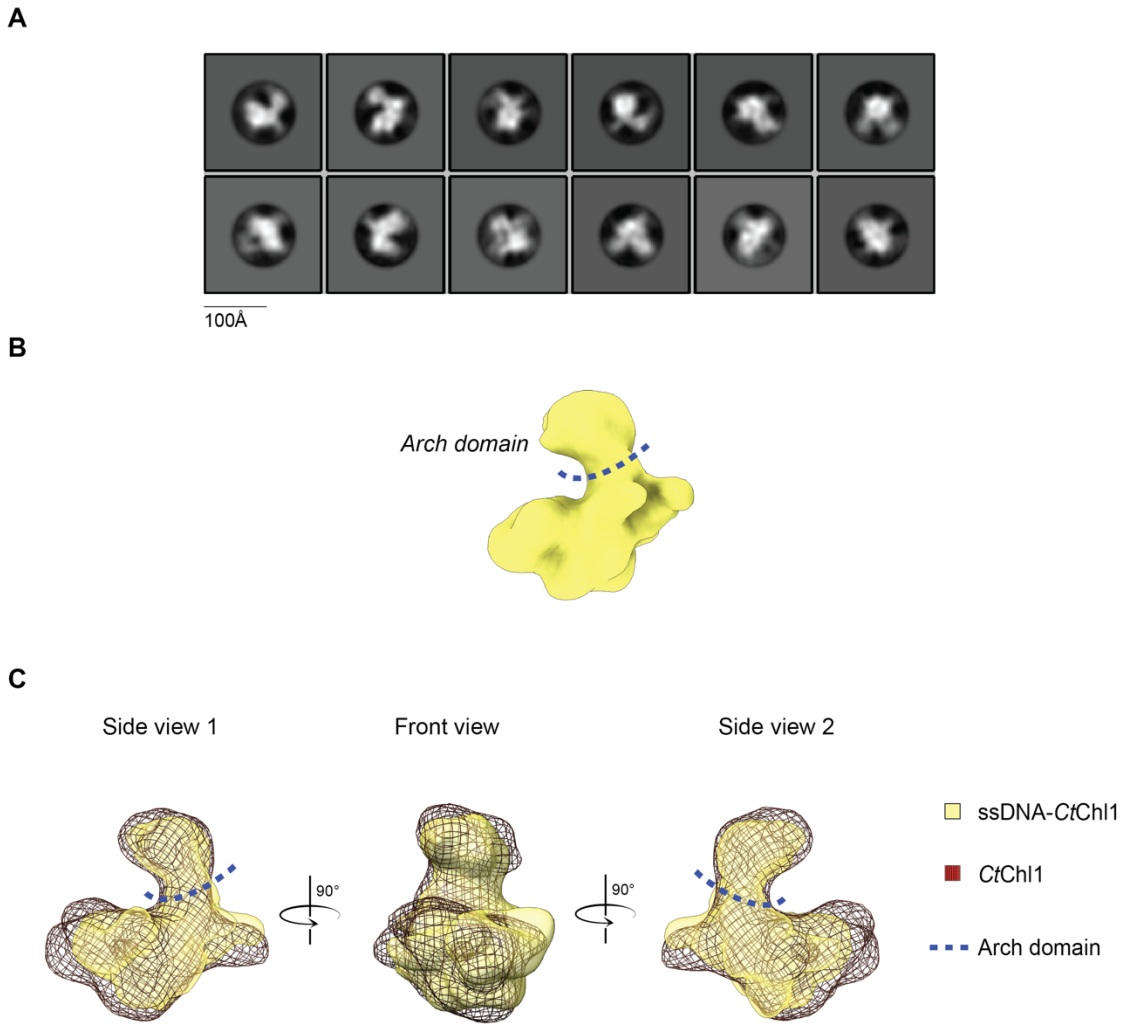

Figure 5.7 Architecture of *Ct*Chl1 supplemented with DNA and nucleotide.
A Final 2D classification of 93 013 particles. **B** The resulting 3D model (side view) shows an extended (open) Arch domain. **C** Comparison of nucleotide-free and nucleotide and DNA-bound proteins' 3D volumes shows no significant changes in domain organisation.

5.2.6 Crystallisation trials of CtChl1

In parallel with Cryo-EM studies, *Ct*Chl1 was subjected to crystallisation trials. The protein was purified as previously described, concentrated to various concentrations ranging from 9 to 15 mg/ml and used in a wide variety of crystallisation screens (see Table 9). Three formulations of the protein were used: protein with no additives, protein with AMP-PNP, or protein with both AMP-PNP and ssDNA[4]. Only one condition was found to form microcrystals, which was similar with all three formulation (Figure 5.8-A,B). Further expansion of the crystallisation drops showed no microcrystals. No further manipulation of the conditions showed any crystallising events. In order to identify whether a contamination to the screen occurred, a new batch of the screen was used, which did not show the same microcrystals as previously (Figure 5.8-C). Seeding the microcrystals into a freshly made plate with this condition also did not result in crystallisation events (Figure 5.8-D). It was therefore concluded that the microcrystals formed were due to a contamination of the well of the screen, which could not be identified. Conversely to *Ct*Chl1, XPD structures published show that this protein can be crystallised in the absence of DNA, as well as in a state where its Fe-S cluster is disrupted. Despite this disruption leading to a partially disorganised region, the protein still does crystallise (Fan *et al.*, 2008; Liu *et al.*, 2008). Compared to XPD, *Ct*Chl1 contains the additional insert in its HD1. It is therefore possible that this insert hinders the formation of a crystal lattice because it does not have a strong secondary structure.

5.2.7 Structural studies of Mini-Chl1

In order to investigate whether the insert truly is hindering crystallisation and to map its precise location, a *C. thermophilum* construct of the so-called MiniChl1 (*Ct*MiniChl1) was constructed based on the previously published MiniChl1 in *S. cerevisiae* (Samora *et al.*, 2016). The isoelectric point of *Ct*MiniChl1 compared to *Ct*Chl1 was significantly changed due to removal of numerous negatively charged residues found in the insert region and was therefore polished with a cation exchange chromatography column (Figure 5.9-A). The protein eluted within the same volume as full-length *Ct*Chl1 despite being by 20kDa smaller compared to the 100kDa *Ct*Chl1, suggesting that its conformation is not more compact than its full-length version.

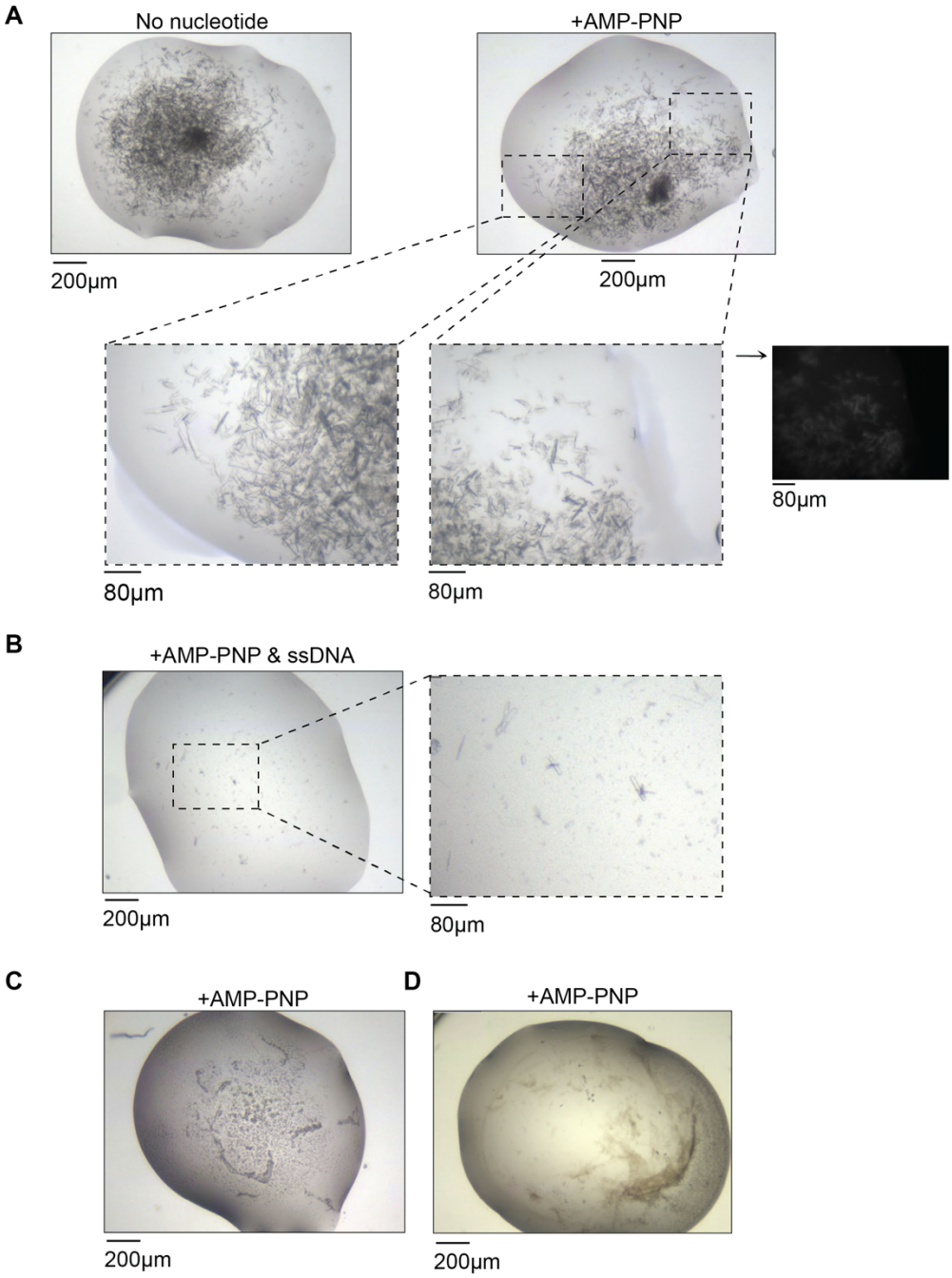

Figure 5.8 Crystallisation trials with *Ct*Chl1.

A *Ct*Chl1 with and without nucleotide. The crystallisation drop with AMP-PNP has been visualised with both visible and UV light to confirm the microcrystals are protein. **B** *Ct*Chl1 with nucleotide and ssDNA. **C** Repeating the crystallisation trial with a fresh screen. **D** Seeding of *Ct*Chl1 from AMP-PNP drop into a self-prepared condition.

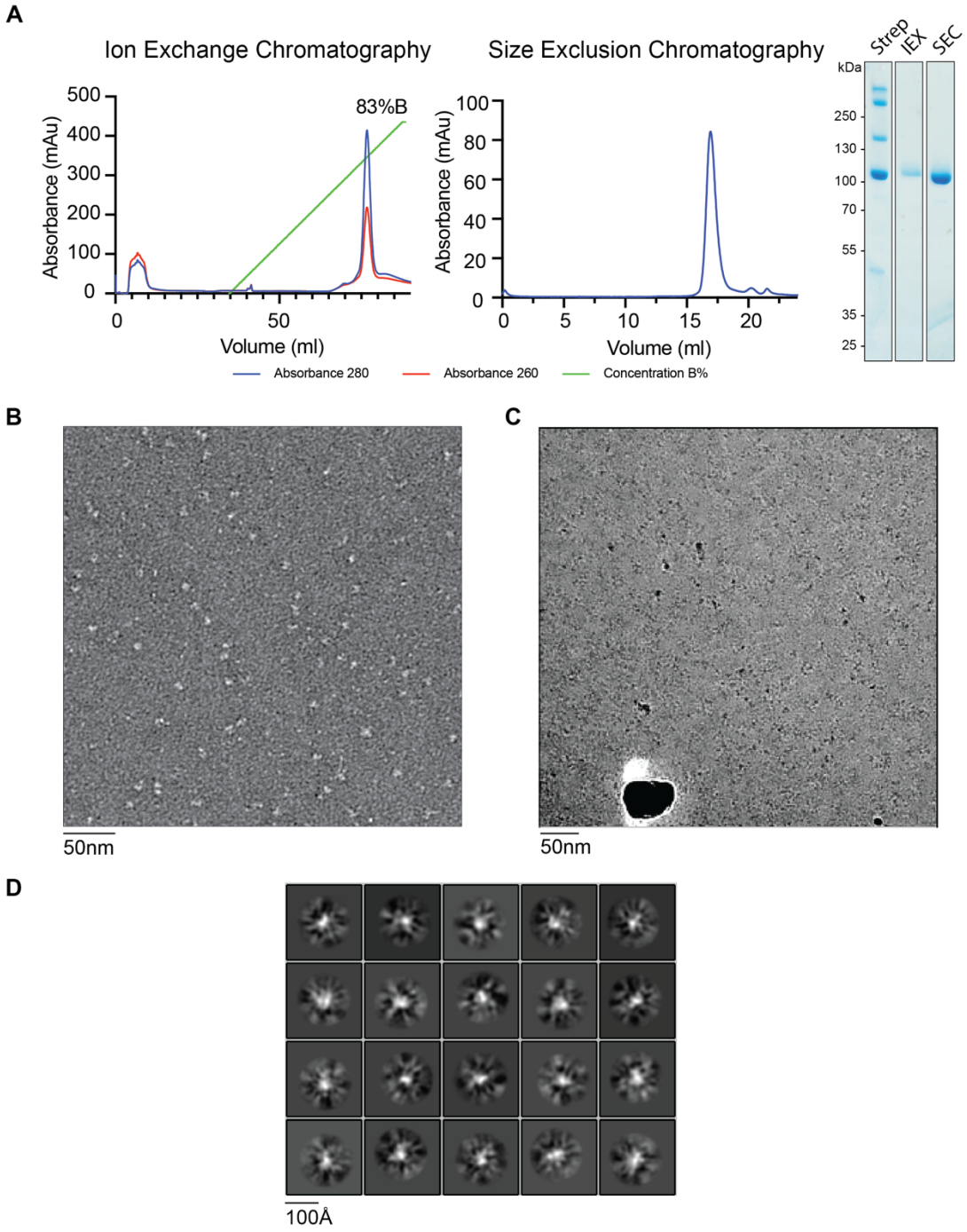

Figure 5.9 *Ct*MiniChl1 purification and structural analysis.

A *Ct*MiniChl1 purifies with cellular contaminants which can be removed with IEX and SEC polishing. **B** UA staining reveals mild heterogeneity in the sample. **C** Denoised micrograph of the frozen sample. Particles are difficult to see. **D** Repeated 2D classification cannot converge to clear 2D classes.

Negative staining of *Ct*MiniChl1 with UA has shown that this protein is an intact monomer but indeed more heterogeneous (Figure 5.9-B). Despite the heterogeneity, the protein was frozen in the same conditions as full-length *Ct*Chl1 and imaged on the Talos Arctica tin the attempt to obtain an envelope that could be used for signal subtraction from the full-length *Ct*Chl1 to locate the insert. The 80kDa protein was difficult to distinguish from ice due to its low SNR, therefore particle picking was performed with the use of the CrYOLO denoising filter but even then the extensive 2D classification of particles could not "purify" the sample from noise (Figure 5.9-C,D). Furthemore, the reduction of size to 80kDa reduced the number of features available for alignment where classification could not converge to any reliable 2D classes with a defined architecture. Alternatively, removal of the insert from this construct has introduced further instability to the protein. Two likely explanations to these observations are that the insert is need for overall stability of the protein, or simply that the boundaries of the insert or the linker substitute length and thus the construct design were not predicted correctly.

5.2.8 Isolation of the Chl1 Insert

To gain more information on the structural features of the insert, the corresponding sequence of *Ct* as well as of 5 other species have been expressed in bacteria (Figure 5.10-A). Around 25% identity was observed between inserts of any two species with the exception of *Arabidopsis thaliana* (*At*) and *Schizosaccharomyces pombe* (*Sp*) inserts where the sequence identity was only 19%, as judged by multiple sequence alignments. The insert of *Ag* Chl1 was the only construct which did not express. The expressed inserts were between 15-20kDa. The *Ct* and the *Hs* inserts (*Ct*Chl1Insert and *Hs*Chl1Insert) showed the highest level of expression and were therefore studied further. Scaling up the expression of both has resulted in an increased number of contaminants. The *Ct*Chl1Insert was optimised to purify as a single elution peak on SEC but the final yields of the sample never reached sufficient levels for crystallisation trials (Figure 5.10-B). Similar was observed for the *Hs*Chl1Insert.

The predicted insert boundaries arose from the XPD crystal structure and sequence (Samora *et al.*, 2016). Purification of the *Ct*Chl1Insert has occasionally resulted in the protein eluting in the void volume, suggesting it is not particularly stable. To

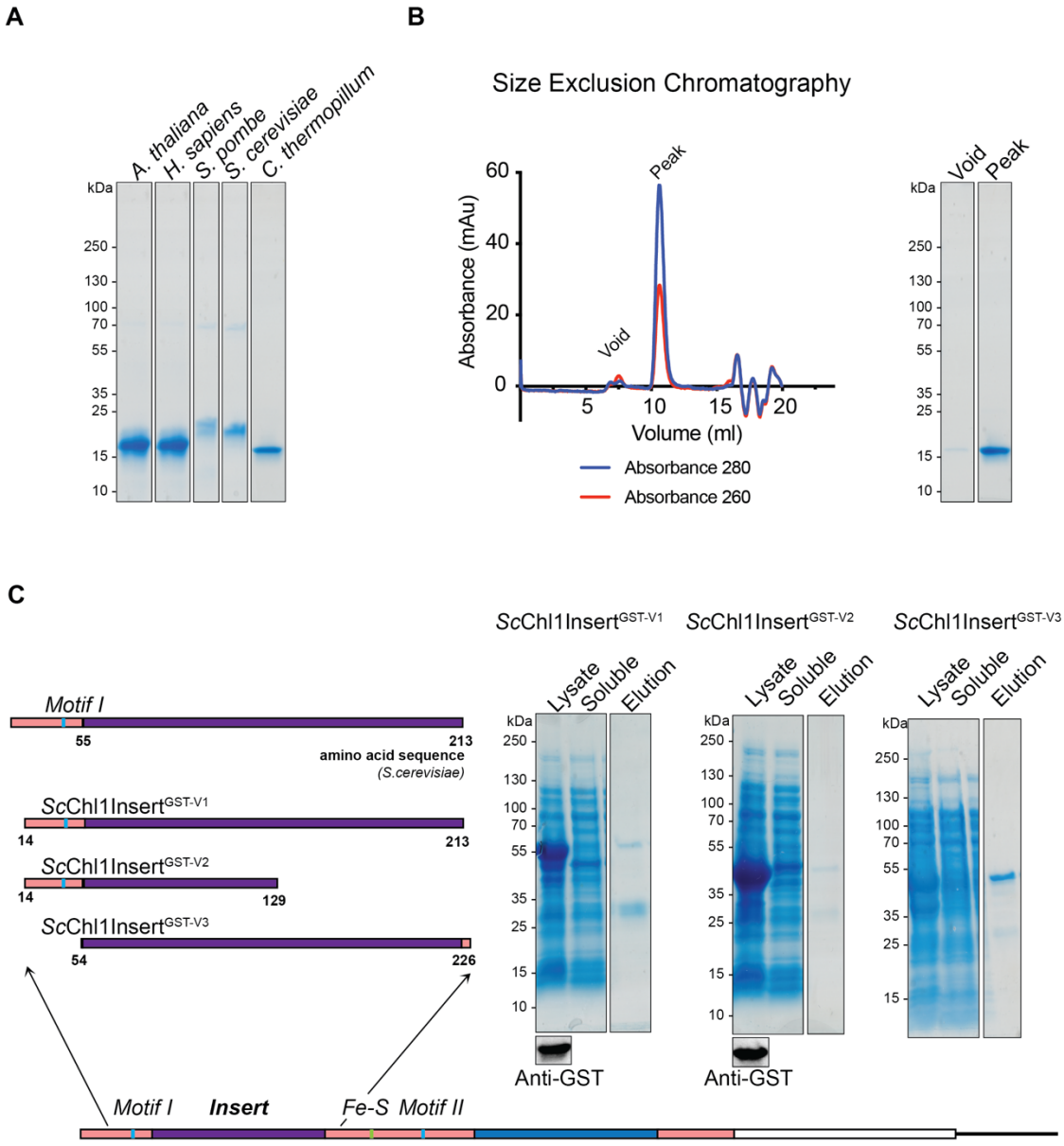

Figure 5.10 Isolated Chl1 inserts.

A Expression of Chl1 inserts from 5 different species. Only one insert did not express. **B** Final step of the purification of *Ct*Chl1Insert. The elution profile and SDS-PAGE show a stable protein, but of low yields. **C** A schematic of the sequence of the three constructs of the yeast Chl1 insert and their expression tests. The identity of the GST-tagged insert in the insoluble fraction was confirmed with WB using an anti-GST antibody.

create a more stable construct for both structural and functional analysis, three different constructs of GST-tagged *S. cerevisiae* Insert (*Sc*Chl1Insert^{GST-v1}, *Sc*Chl1Insert^{GST-v2}, *Sc*Chl1Insert^{GST-v3}) were cloned based on structure predictions using PsiPred and the XPD structures (Figure 5.10-C). The choice of organism and tag allowed for investigating the insert's functional role in *S. cerevisiae* (described in section 5.4.2.2). Initially, two inserts were tested for expression with a variation in the C terminus border only.Expressing the two constructs showed that the CTD of the insert is most likely disordered as its shortening resulted in an insoluble peptide even when tagged with GST, a tag used for solubilising proteins. This assumption was made based on the fact that a His-tagged *Sc*Chl1Insert was previously expressed as a soluble protein. Resting on these observations a third insert construct *Sc*Chl1Insert^{GST-v3} was cloned, which yielded a soluble protein. Taken together, this would suggest the insert contains a disordered region which, when unprotected, yields the protein insoluble. This unfolded region in its CTD is most likely followed by a sequence with a stronger secondary structure. Removal of these few amino acids at the CTD exposes the disordered region and yields the protein insoluble as judged by the difference between ScChl1Insert^{GST-v1} and ScChl1Insert^{GST-v3}.

5.2.9 Construct optimisation

Internal deletion of the insert from *Ct*Chl1 to create the *Ct*MiniChl1 construct has resulted in an unstable protein. The insert could therefore be required for the overall stability of the Chl1 protein. Given that the construct design was not structure-guided, the precise boundaries of the insert domain were not known.

5.2.9.1 Hydrogen deuterium exchange

In order to investigate the nature of the protein and gain more understanding of the insert domain boundaries, HDX-MS was performed to obtain information about the folding of the protein. This technique focuses on measuring the exchange between hydrogens in the protein and deuterium uptake from the solvent over time, followed by trypsin digest and MS analysis. The structured regions do not uptake deuterium but flexible regions exchange deuterium much more readily. The insert domain spans approximately 150 amino acids just after the Walker A motif of the helicase. HDX-MS results show that the sequence of the insert proximal to the Walker A motif has

a secondary structure, consistent with predictions. This folded region then transits into a more-or-less disordered state, with the C-terminal region of the insert appearing completely disordered (Figure 5.11-A,B). These results agree with the observations from the *Sc* insert construct optimisation. Interestingly, the Tof1-binding motif is expected to lie in this disordered region based on observations with human proteins and sequence alignments (Lerner *et al.*, 2020). These observations could be responsible for the inability of the protein to crystallise as well as the lower resolution of HD1 compared to the rest of the protein. The XPD structure with the disrupted Fe-S cluster shows a partially unresolved HD1. Despite the protein sample being orange in colour which is common for Fe-S cluster-containing proteins (Rudolf *et al.*, 2006), the regions surrounding the cluster were also studied with HDX-MS to exclude the possibility of the HD1 being not resolved due to the loss of the cluster. HDX-MS data agrees with the binding mode of XPD's cysteines to the Fe-S ions, showing structured regions surrounding the cluster (Figure 5.11-C,D).

5.2.9.2 Optimisation of the Mini Chl1 construct

Based on the expression tests and HDX-MS data obtained the design of the Mini Chl1 construct was revisited. Compared to the first design, the new *Ct*MiniChl1^{v2} only lacks the disordered region of the insert, keeping the folded regions just after the Walker A motif intact. The disordered region was replaced with a flexible linker and resulted in a 99kDa protein. Purification of this construct followed similar steps to the first design but purified in higher yields with significantly less contaminants (Figure 5.12-A). Analysis of *Ct*MiniChl1^{v2} using SEC and negative staining has shown that this protein is monomeric and more intact than the first design (Figure 5.12-B). Structural analysis of the frozen sample has shown that *Ct*MiniChl1^{v2} aligns into 2D classes with higher precision than the first design but is on the border of what is obtainable with regards to alignments due to its size (Figure 5.12-D). The resolution does not allow to observe the precise location of the disordered region of the insert when comparing the map of *Ct*MiniChl1^{v2} with the full-length protein map. Given its more desirable behaviour than the first design, *Ct*MiniChl1^{v2} was subjected to crystallisation trials, where identical screens were used as for the full-length protein. Trials have yielded several promising conditions (Figure 5.12-C). Comparing these

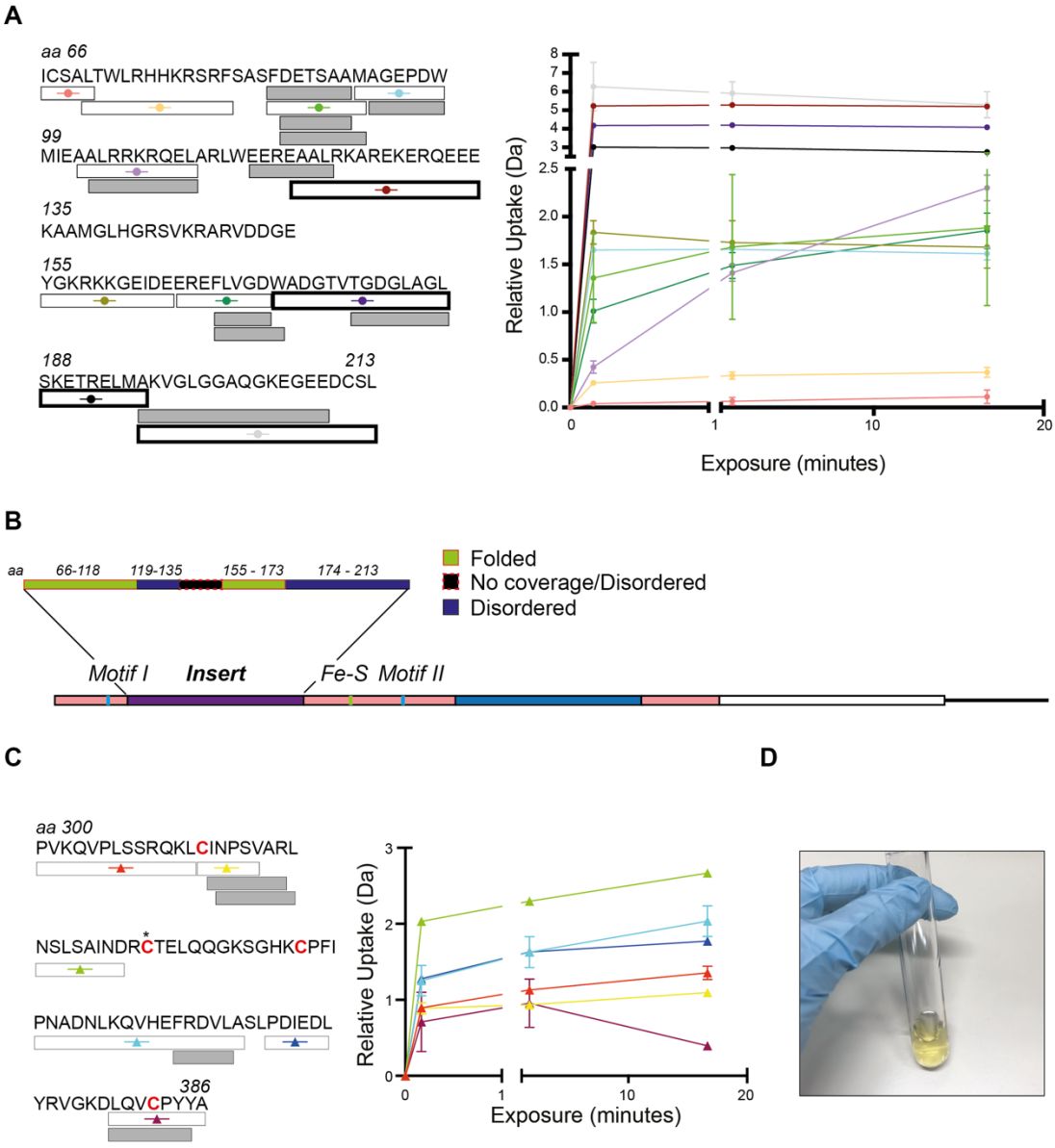

Figure 5.11 HDX-MS characterisation of *Ct*Chl1.

A Deuterium uptake curves for the *Ct*Chl1 insert. Peptide coverage is represented in boxes mapped on the sequence of the insert. Coloured boxes are matched to the colours of the uptake curves. Grey boxes illustrate peptides whose uptake curves are not shown. **B** Overview of the folding of the insert based on the expression tests and the HDX-MS data. **C** Same as in A for the Fe-S cluster, showing its intactness. The asterisk denotes the variable cysteine of the four cysteines binding the cluster. **D** The typical yellow colour of purified Fe-S-containing protein sample.

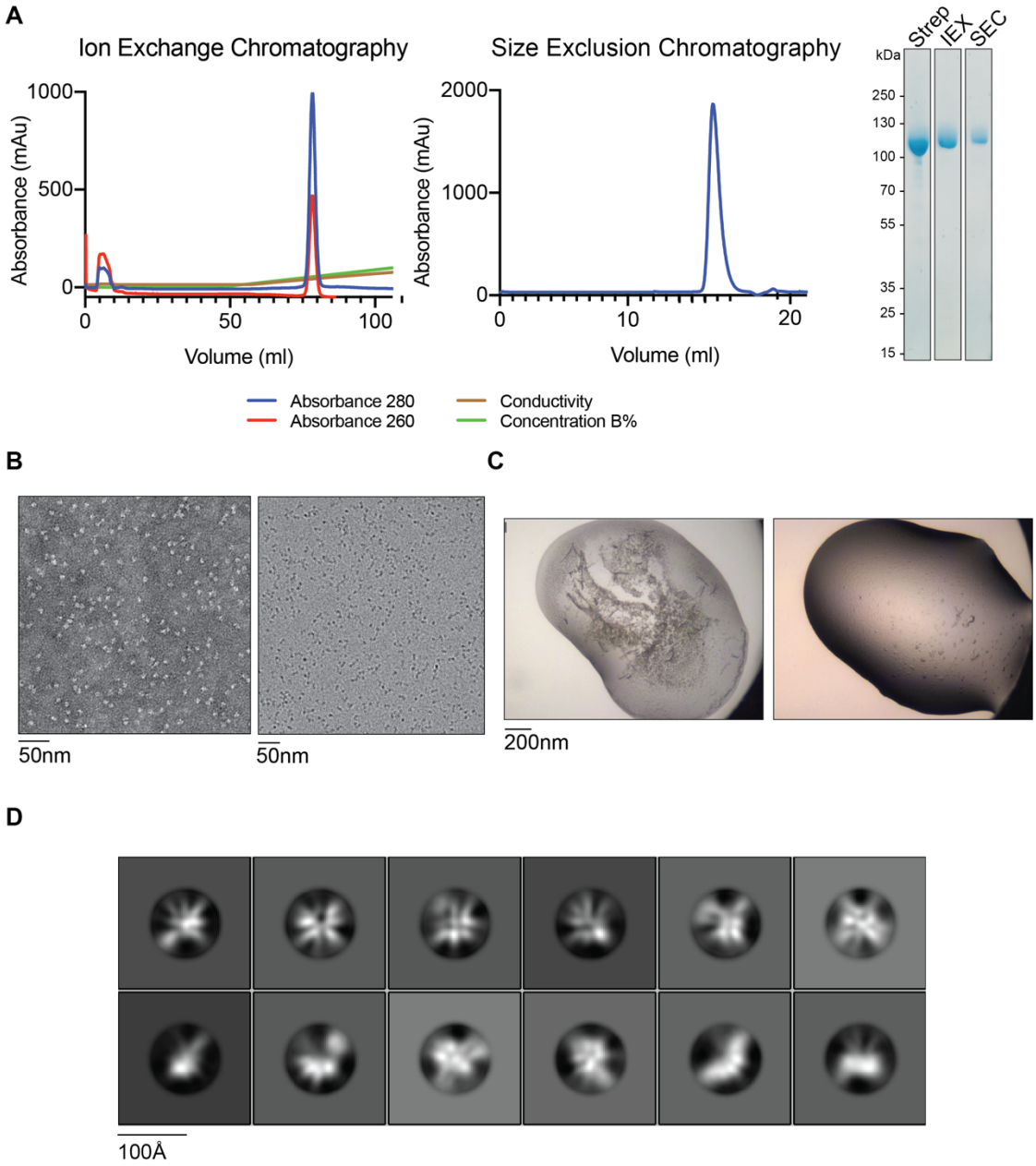

Figure 5.12 Purification and structural studies of the optimised Mini Chl1 construct. **A** The modified Mini construct purifies with significantly less contaminants than the previous construct. **B** Negative staining and specimen vitrification result in well-behaved visible particles. **C** Promising drops from crystallisation trials. **D** 2D classes of the 99kDa protein.

promising conditions with those of full-length *Ct*Chl1, *Ct*MiniChl1^{v2} , although requiring more optimisation, shows a higher likelihood to crystallise.

5.3 Chl1 interactions

Given the size and the internal flexible domain of Chl1, the structural characterisation of this protein has proven difficult. To overcome the limitation of size, increasing the mass by binding Chl1 to an interaction partner would allow for more accurate alignments of particles and potentially a higher resolution.

5.3.1 Interactions of Chl1 with proteins of the replication fork

The best characterised binding partner of Chl1 is Ctf4, a homotrimeric protein stably associated with the replisome. Its binding partners all share a common DDIL motif through which they bind to the CTD of Ctf4. Unlike the beta-propeller and helical folds of the CTD, the N-terminal domain (NTD) of Ctf4 is more disordered and is not involved in binding of the DDIL motif-containing proteins (Simon *et al.*, 2014; Samora *et al.*, 2016). Because the NTD of Ctf4 is not required for interactions with Chl1, the CTD construct of the *C. thermopilum* Ctf4 (*Ct*Ctf4^{CTD}) was expressed. *Ct*Ctf4^{CTD} was expressed in bacteria and purified with a three step purification. The resulting protein was visualised using negative staining which showed that this protein forms a homotrimer as previously reported for the yeast homolog(Figure 5.13) (Simon *et al.*, 2014). In addition, *Ct*Core, which was previously used for EM studies (refer to Chapter 4), contains subunits reported to interact with Chl1 in *S. cerevisiae* (Samora *et al.*, 2016), and was therefore also tested for interaction *in vitro*.

To test the interaction between *Ct*Chl1 and *Ct*Ctf4^{CTD} the proteins were first subjected to SEC analysis and pulldown experiments. Using streptavidin beads the two proteins did not pull down together (Figure 5.14-D). The pulldowns are also unsuccessful if performed against the His-tag on *Ct*Ctf4^{CTD}. Similar was observed with the modified cohesin complex *Ct*Core. Analysing the SEC elution profile of the combined *Ct*Chl1 and *Ct*Ctf4^{CTD} sample showed a single peak that would be suggestive of an interaction between the two proteins but subsequent SDS-PAGE analysis of the fractions show that the protein samples' elution profiles most likely overlap rather than show an interaction of the two (Figure 5.14-A,B). Overlays of

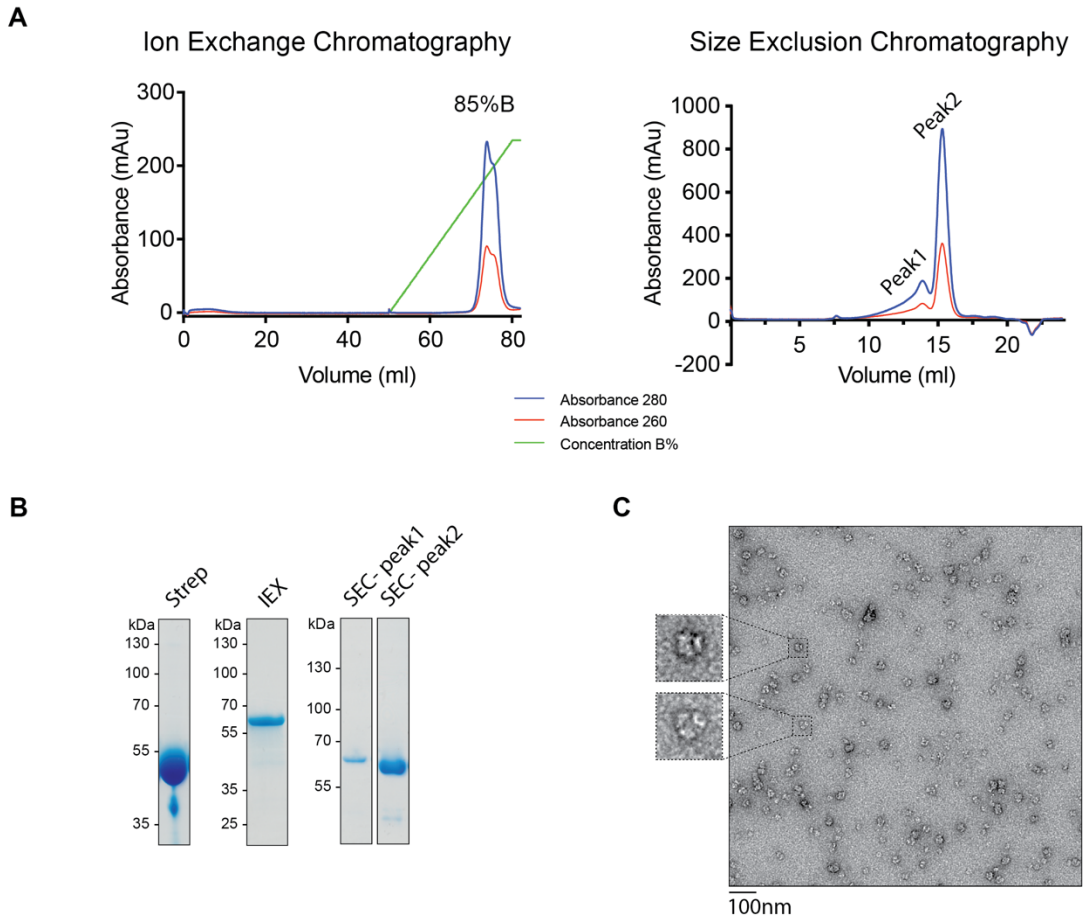

Figure 5.13 *Ct*Ctf4^{CTD} purification and structural analysis.

A & B The protein expression and purification results in a clean protein of high yields. **C** Negative staining shows the protein oligomerises.

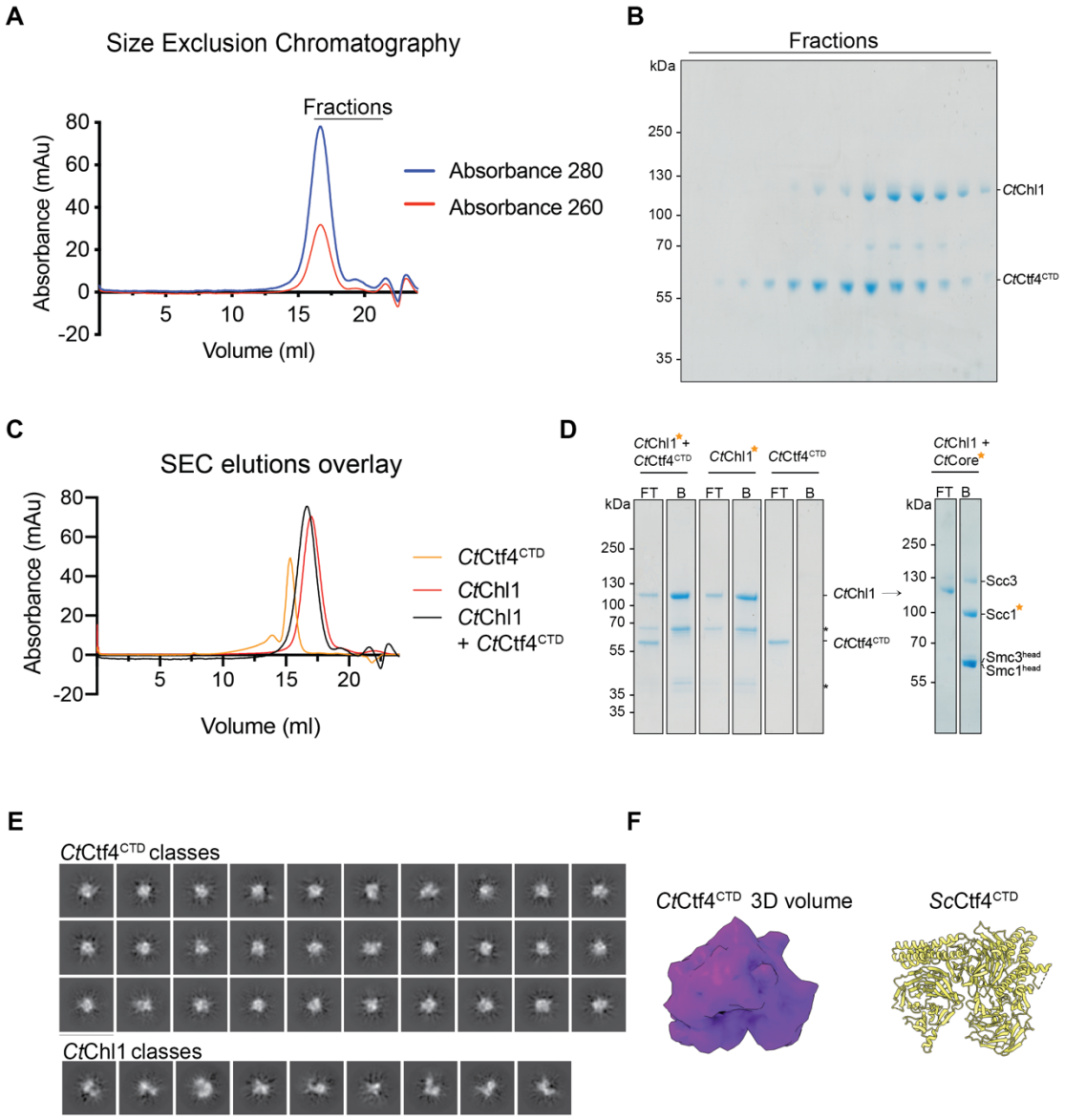

Figure 5.14 Interactions between *Ct*Chl1 and *Ct*Ctf4CTD.

A SEC analysis shows both proteins elute in the same peak. **B** SDS-PAGE suggests elution overlap. **C** Overlay of elution profiles for the combined sample and the individual proteins shows a shift in *Ct*Ctf4^{CTD}'s elution profile. **D** Pulldowns using streptavidin beads do not show an interaction. Proteins with the strep-tag are marked with a yellow star. FT-flow-through, B-beads. **E** Non-crosslinked *Ct*Chl1 and *Ct*Ctf4^{CTD} samples from Cryo-EM sample preparation classify separately. **F** Reconstructed *Ct*Ctf4^{CTD} 3D volume agrees with the yeast homolog structure.

single protein elution profiles however show that both peaks are slightly shifted when the sample is joined (Figure 5.14-C). This would suggest they interact very weakly *in vitro*. Indeed, the affinity of Ctf4 for Chl1 is much lower than for its two other binding partners, GINS and Pol α, and unlike these two proteins whose DDIL-containing peptides could be co-crystallised with Ctf4, no such structure was obtained for Chl1 (Simon *et al.*, 2014).

The interaction observed *in vivo* between Chl1 and Ctf4 could not be fully reproduced with purified proteins in pulldown experiments, but the observed shift in SEC elution profiles raised a question whether this complex, although very transiently, exists *in vitro*. In the attempt to capture this complex, a large dataset of the combined vitrified *Ct*Chl1 and *Ct*Ctf4^{CTD} sample was collected. As observed previously, *Ct*Ctf4^{CTD} localises mainly on the carbon of the grids and therefore a carbon support grid was used for this collection. 2D classification of the dataset revealed no complex formation between the two proteins, and the two proteins classified into individual 2D classes (Figure 5.14-E). Given the decreased signal-to-noise ratio with the carbon support, *Ct*Chl1 2D classes were distinguishable but no 3D volume could be built to confirm the proteins identity. Conversely, *Ct*Ctf4^{CTD} was successfully reconstructed into a 3D volume (Figure 5.14-F). The yeast structure of Ctf4^{CTD} matches the Cryo-EM map of *Ct*Ctf4^{CTD} when fitted in, confirming its identity.

In parallel to obtaining data for the non-crosslinked samples, *Ct*Chl1 was subjected to both in solution and in gradient crosslinking with glutaraldehyde (Figure 5.15-A). Crosslinking *Ct*Chl1 to *Ct*Core resulted in a single band when analysed on a silver-stained gel. Classifying particles from negatively stained grids revealed a variety of 2D classes, of which a proportion represented *Ct*Chl1 only (Figure 5.15-B,C). The remaining classes could represent Chl1 bound to cohesin, but the crosslinked *Ct*Core alone structurally resembled the 2D classes obtained for the combined sample. All of these classes remained heterogeneous and given the complex *Ct*Core sample behaviour, optimisation of this complex was not further pursued.

Since cohesin needs ATP to engage its ATPase heads, reactions with *Ct*Core were prepared in the presence of ATPγS. The interaction between Ctf4 and Chl1 *in vivo* was reported to be ATP-independent as the ATPase dead mutant does not affect

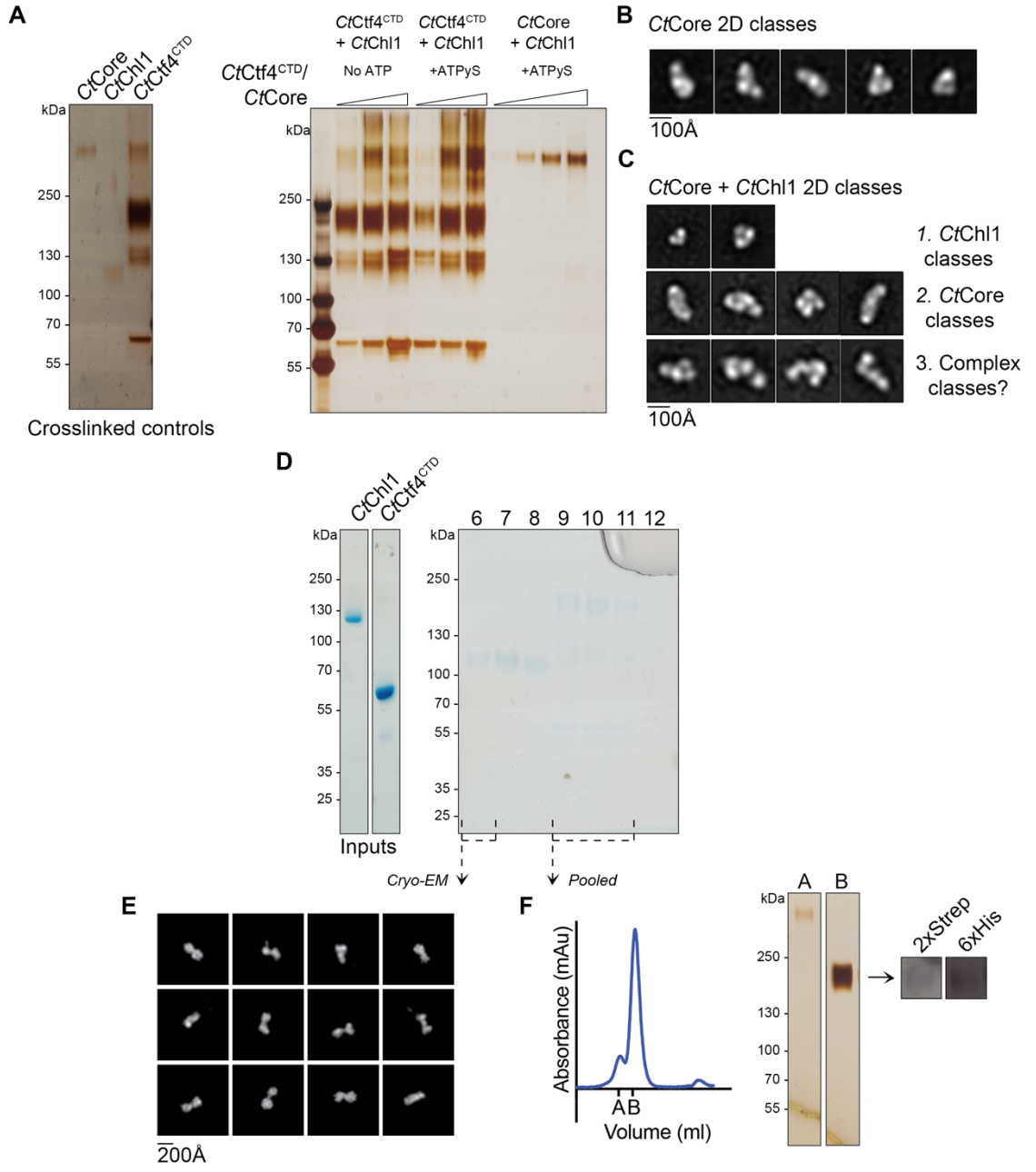

Figure 5.15 Crosslinking of Chl1 to its potential interaction partners.

A In-solution crosslinking with glutaraldehyde. **B** Crosslinked *Ct*Core and **C** the combined sample of *Ct*Core and *Ct*Chl1. Three types of 2D classes are visible. The third could represent a crosslinked complex of cohesin and Chl1. **D** Crosslinking with GraFix and the **E** the fraction used for Cryo-EM and **F** the elution profile of three pooled fractions highlighted. The Cryo-EM sample in E likely represents two Chl1 molecules crosslinked together.

binding in *S. cerevisiae* (Samora *et al.*, 2016). The in solution crosslinking was however performed both in the presence and absence of the non-hydrolysable analog, whereas GraFix preparation was in the absence of a nucleotide. The resulting crosslinking with both GraFix and an in solution crosslinking approach has showed that *Ct*Ctf4^{CTD} alone crosslinks into multiple species as visible on the gel (Figure 5.15-D). The early fractions of the gradient show a single band with a different molecular weight to the uncrosslinked controls, as well as fractions where crosslinking resulted in multiple bands. The fraction of the gradient showing a single band on the gel was therefore chosen for Cryo-EM analysis. Vitrification and subsequent processing of the sample has resulted in a low resolution model in which individual proteins could not be assigned. This is because crosslinked samples were found to preferentially bind to carbon supports and thus open hole grids were omitted. Lacey grids with a thin carbon support were used instead, which resulted in a reduction in the SNR. A follow-up analysis of the sample used has shown that this vitrified sample only contained the *Ct*Chl1 protein. The classes obtained most likely show the *Ct*Chl1 protein crosslinked onto a second *Ct*Chl1 molecule, resulting in a complex with a diameter significantly larger than a monomeric *Ct*Chl1 particle (Figure 5.15-E). The crosslinked *Ct*Chl1 yielded a 3D volume of very poor quality and therefore the approach of crosslinking *Ct*Chl1 to improve sample stability for structure determination was not sought after again. WB analysis of GraFix fractions showing multiple bands revealed that the highest molecular weight band, found in later fractions, contains both strep-tagged *Ct*Chl1 and His-tagged *Ct*Ctf4^{CTD}. Pooling these fractions together and their subsequent separation on SEC shows that this complex can be separated from the remaining proteins (Figure 5.15-F). It is most likely that a minor proportion of this sample could have contained both proteins and thus gave a signal on a WB gel, but it did not appear that the samples had crosslinked into a stable complex.

The interaction of Chl1 and Ctf4 was shown to exist *in vivo* (Samora *et al.*, 2016). Because this interaction is weak and cannot be reproduced with purified proteins, a mutant version of Chl1 was purified which should have enhanced binding affinities towards Ctf4. In this mutant version, the CIP-box of Chl1 was modified to resemble the binding motif of the Sld5 subunit of GINS, a constitutive binding partner of Ctf4. In addition to the D D/E IL motif present in all Ctf4 binding proteins, GINS forms

additional contacts with Ctf4, including the I and A residues flanking the CIP-box of Sld5. Such residues are not present in Chl1 or Pol2. The wild-type sequence of *Ct*Chl1, SDEILQ, was therefore mutated to resemble the IDDILA sequence of Sld5 and this construct was termed *Ct*Chl1^{IDDILA}. Pulldown experiments using *Ct*Chl1^{IDDILA} and *Ct*Ctf4^{CTD} have however not led to both proteins pulling down together (Figure 5.16-A). Two reasons could explain the observed behaviour: either the interaction does not exist *in vitro* or the affinity of the two proteins is very little and potentially requires another component for efficient binding.

5.3.2 Nanobodies

The search for a stable strong interacting partner for *Ct*Chl1 has been unsuccessful. Increasing the mass of the particles and potential stabilisation of Chl1 was therefore not achieved. An alternative to native interacting partner for increasing mass are nanobodies, single chain llama antibodies, which were developed to specifically bind *Ct*Chl1. Four nanobodies produced by Hybrigenics were tested for expression. All four expressed with a high amount of *E. coli* contaminants that could not be reduced by changing expression temperatures or purifying the cells from a periplasmic component of the cell. These contaminants most likely represent molecular chaperones, such as the Hsp70 at the 70kDa mark, but their identity was not confirmed. A three-step purification of affinity chromatography, IEX and SEC was applied to all four nanobodies (Figure 5.16-B-D). The nanobodies were subsequently used for Cryo-EM grid preparation as for previous collections, using C-flats with either LMNG or OG detergent. No observable additional density for nanobodies was observed.

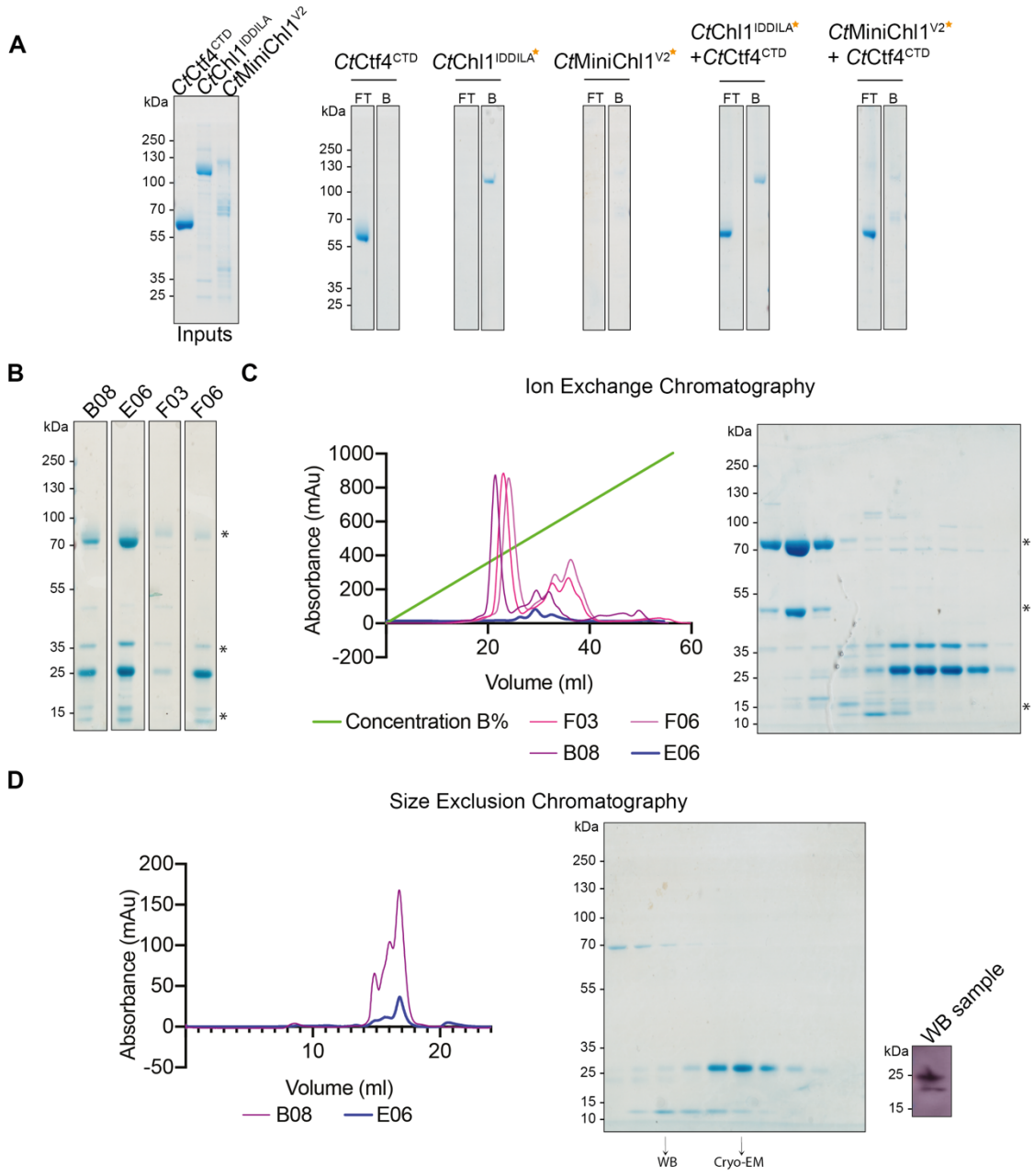

Figure 5.16 Forming interactions with *Ct*Chl1 for Cryo-EM.

A Mutating the residues around *Ct*Chl1's CIP-box to mimic Sld5 do not result in a complex formation *in vitro.* **B** All four nanobodies express with a high amount of cellular contaminants, denoted by an asterisk. **C** IEX shows the nanobodies can be partially polished. **D** Further polishing with SEC yields a cleaner sample. The identity of the nanobody was confirmed by WB against the His-tag on the N terminus of the nanobody. SDS-PAGE analysis of purification steps is shown for the E06 nanobody only.

5.4 Functional studies of Chl1

The Chl1 interactions important for sister chromatid cohesion in budding yeast have been linked with Ctf4. The replication stress response is a less well characterised process but is known to involve Chl1 (Samora *et al.*, 2016; Cortone *et al.*, 2018; Lerner *et al.*, 2020). To investigate the functions of Chl1 at the replication fork, both *in vitro* approaches and *in vivo* characterisation in *S. cerevisiae* were undertaken.

5.4.1 DNA binding activity of Chl1

The yeast as well as human Chl1 protein are reported to bind and unwind DNA *in vitro*. To confirm *Ct*Chl1 is able to bind DNA, EMSAs were performed. The binding of *Ct*Chl1 to DNA was therefore tested. *Ct*Chl1 efficiently bound to both FAM-ssDNA^{4} and FAM-dsDNA^{4} showing no preference for the DNA type (Figure 5.17-A). It is possible that a yet undefined segment of *Ct*Chl1 binds dsDNA. In order to gain a better understanding at DNA binding to *Ct*Chl1 and to optimise DNA binding conditions for other experiments, DNA binding affinity was determined using fluorescence anisotropy measurements, kindly performed and analysed by Dr Simone Kunzelmann (Figure 5.17-B). *Ct*Chl1 was titrated against a fixed concentration of FAM-ssDNA^{4} at 10nm. Addition of *Ct*Chl1 in excessive amounts did not reach a plateau in the affinity curve, suggesting that the protein has very low affinity for DNA. Conversely, *Ct*MiniChl1^{v2} bound the same DNA concentration with much higher affinity with a calculated Kd = 0.038μM ± 0.005 μM.

Previous reports from yeast have shown that ATPase and helicase activity are not necessary for sister chromatid cohesion, an event co-dependent on the Chl1-Ctf4 interaction. This could be due to an inhibitory effect of Ctf4 on Chl1. *Ct*Ctf4^{CTD} was titrated into the EMSA reactions at increasing concentrations with a constant concentration of *Ct*Chl1. *Ct*Ctf4^{CTD} did not bind any DNA and did not influence the DNA binding to Chl1 either in the presence or absence of a nucleotide (Figure 5.17-C,D).

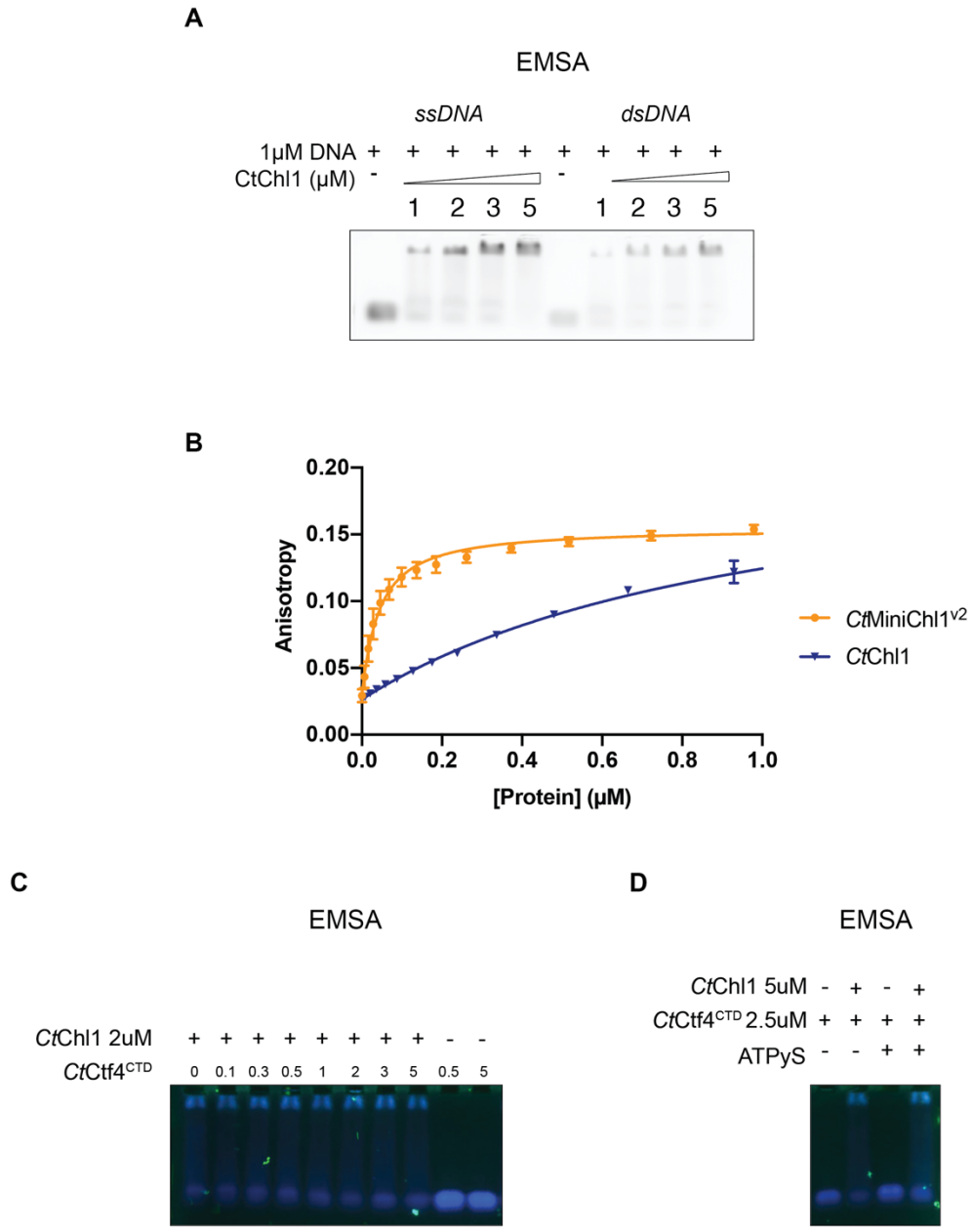

Figure 5.17 DNA binding properties of *Ct*Chl1.

A *Ct*Chl1 can bind both ssDNA and dsDNA. **B** Anisotropy measurements showing weak affinity of *Ct*Chl1 for ssDNA. Removal of the insert results in an over 25-fold increase in affinity. **C** *Ct*Ctf4^{CTD} has no effect on ssDNA binding to *Ct*Chl1. **D** The same reaction as in C in the presence of a non-hydrolysable nucleotide analog.

5.4.2 In vivo characterisation of Chl1

To gain a better understanding of the function of Chl1 in DNA synthesis, yeast were used as the model organism. Information on the role of Chl1 in sister chromatid cohesion has been partially explored in this organism (refer to section 1.9.1) , but its role in repairing stalled replication forks and potentially DNA repair has not been thoroughly investigated.

5.4.2.1 Strain selection

In order to study the role of Chl1 in replication stress, Chl1 was endogenously tagged with a HA tag in *S. cerevisiae* (*Sc*Chl1-HA). Furthermore, the MiniChl1 construct described previously (Samora *et al.*, 2016) was created and tagged with HA (*Sc*MiniChl1-HA). Both strains were a kind gift from the Uhlmann lab. The presence of the tag was first analysed using TCA extraction (Figure 5.18-A,B) , and to exclude any possibilities of the tag causing perturbations in the cell cycle, the progression through S phase was confirmed using FACS.

5.4.2.2 Interactions of Chl1

Cohesion establishment and replication stress are events both occurring in S-phase of the cell cycle. To identify the timepoint at which Chl1 interactions would be analysed, a time-course FACS experiment was carried out. Replication was halted at a timepoint where early origins have started firing but replication was still ongoing. This was achieved by a G1 arrest using alpha factor with mat a untagged yeast strain, which arrests itself in G1 upon the pheromone addition. This was followed by the release into S-phase by washing off the pheromone replication progression. The harvesting timepoint was decided upon using a wild-type untagged strain (Figure 5.18-C) and subsequently confirmed in *Sc*Chl1-HA strain for both normal S-phase progression and cell cycle arrest (Figure 5.18-D)

Interactions of Chl1 at the replication fork were studied by performing an IP against the HA tag on the proteins and subsequent analysis of the IP by MS with a focus to identify protein-protein interactions with the HA-tagged *Sc*Chl1. The design of the

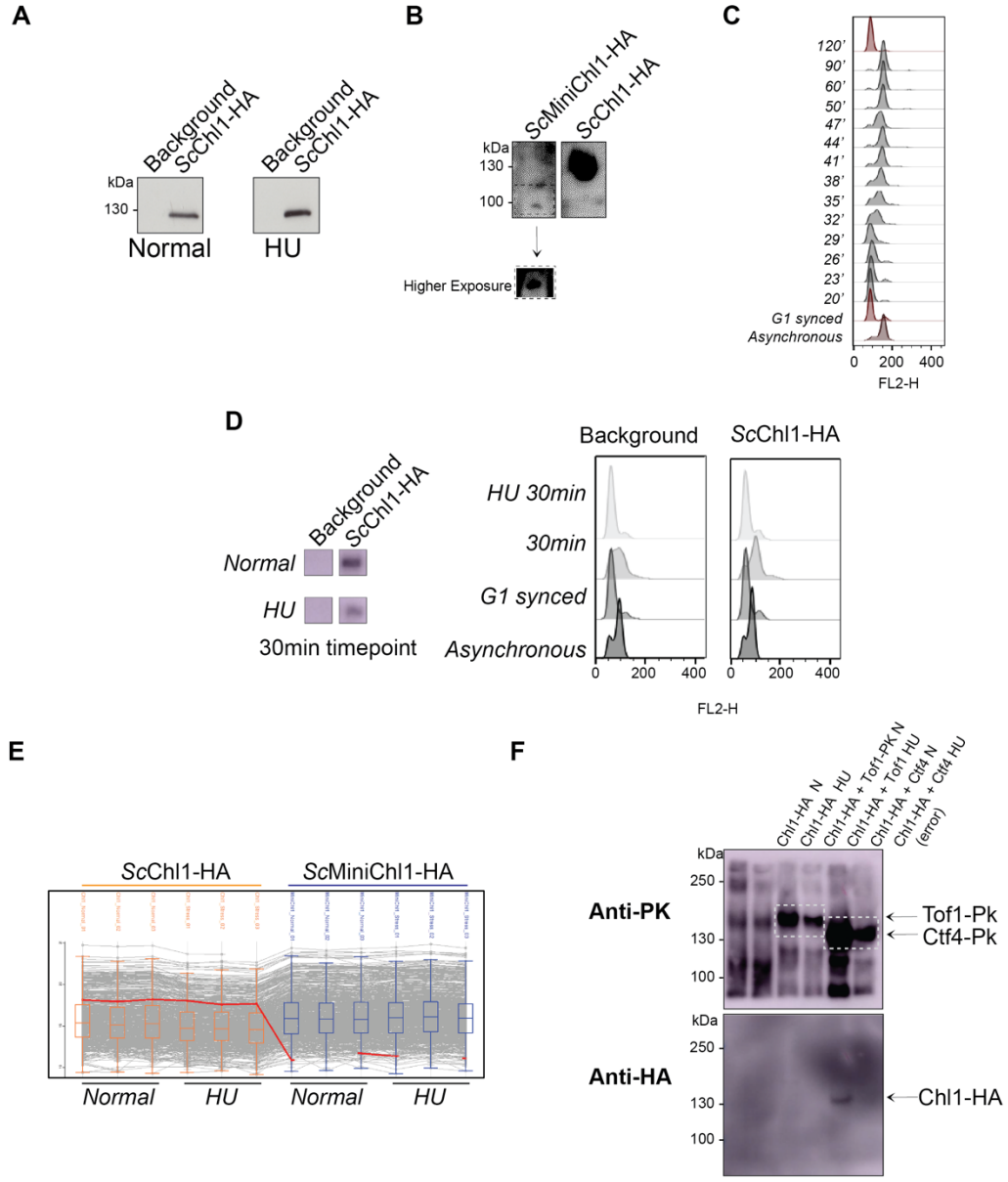

Figure 5.18 *In vivo* *Sc*Chl1 characterisation.

A TCA extraction confirming the presence and level of expression of the HA-tag in the endogenous Chl1 locus. **B** Same as in A for *Sc*MiniChl1 under conditions of no replication stress. **C** FACS of cell cycle progression showing visible DNA duplication at the 30 minute timepoint. **D** Successful co-IP of *Sc*Chl1-HA at a 30 minute timepoint in normal and HU conditions. FACS profile confirmed the replication arrest. **E** Detection of *Sc*Chl1-HA and *Sc*MiniChl1-HA constructs by MS. **F** Ctf4 interacts with Chl1 *in vivo.* No interaction with Tof1 was observed.

experiment aimed at identifying change in these interactions in unperturbed versus stressed conditions at the fork, as well as the difference between Chl1 with and without its insert. Whereas *Sc*MiniChl1-HA was not detectable by MS, *Sc*Chl1-HA was visible in the IP fractions and detectable with MS (Figure 5.18-E). The resulting list of proteins detected by MS has shown no significant results with regards to events at the replication fork (Table 10). A new approach was therefore sought after. For this reason, GST-tagged constructs of the yeast Chl1 insert (refer to Section 5.2.8) were created with the idea to capture any interacting proteins from the yeast lysate. Although the construct was expressed, this approach was not tested but instead, targeted IP approach was sought after.

5.4.2.3 Strain tagging and Targeted IPs

Tof1 and Chl1 interact in mammalian cells, and their interaction is stronger upon stressing the cells, an observation which yet awaits confirmation in yeast. In order to gain a perspective on this interaction, *Sc*Chl1-HA strain was tagged at the endogenous Tof1 locus with anti-V5 epitope tag (SV5-P-K, or shortly PK), creating the *Sc*Chl1-HA-Tof1-PK strain. The interaction between Chl1 and Ctf4 was also evaluated using the strain *Sc*Chl1-HA-Ctf-PK kindly gifted by the Uhlmann lab. The presence of the tag was confirmed using TCA extraction (Figure 5.18-F).

Targeted IP was performed against the PK tag on the Tof1 or Ctf4 proteins, with *Sc*Chl1-HA serving as a control. Cells were synchronised and released into S-phase either into normal media or into media supplemented with hydroxyurea to induce replication stress, and harvested at a 30 minute timepoint. Following the IP, the blots were first probed against the PK tag following re-probing for the HA tag. The results show that Chl1 and Ctf4 interact under normal conditions. The interaction under HU conditions could not be assessed due to human error. No interaction between Tof1 and Chl1 could be seen.

Table 10 Expression changes in replication stress compared to normal condition

Greater expression in stress condition	Reduced expression in stress condition
Ubiquitin Fusion Degradation protein 1	DRS1 helicase
Probable family 17 glucosidase SCW10	Rpn1 - 26S proteasome
Pyruvate carboxylase 1/2	Phenylalanine-tRNA ligase alpha subunit
ATP-dependent RNA helicase DHH1	Periodic tryptophan protein 2
Aminopeptidase Y	Alpha-1,2 mannosyltransferase KTR1
U3 small nucleolar RNA-associated protein 13	Phosphoinositide phosphatase SAC1
Nicotinamidase	26S proteasome regulatory subunit RPN12
	Probable secreted beta-glucosidase UTH1
	Protein TOS1
	U3 small nucleolar RNA-associated protein 7
	AcetylCoA acetyltransferase
	Glycerol-3-phosphate dehydrogenase
	U3 small nucleolar RNA-associated protein 10
	Eukaryotic translation initiation factor 2A
	Mitogen activated protein kinase SLT2
	NADP-dependent alcohol dehydrogenase 6
	Nucleolar complex-associated protein 3
	Ribosome biogenesis protein MAK21
	Ribosomal RNA-processnig protein 12
	GTP-binding protein YPT6

Chapter 6. Discussion

6.1 The interactions of cohesin and the cohesin loader in sister chromatid cohesion

Ever since the discovery of the cohesin complex much of research has focused on identifying the mechanism by which it concatenates the two sisters. Initial findings included the identification of the cohesin loader, a protein complex essential for loading cohesin onto DNA. Previous work in the Singleton lab has led to the solving of two important structures: the C-terminal section of the *Ag*Scc2^{C} and the *Ag*Scc4-Scc2^{N} structure (Chao *et al.*, 2015; Chao, Murayama, *et al.*, 2017). These structures revealed that the loader contacts the ring around the whole circumference. Secondly, they showed that the loading reaction is catalysed by the C terminus of Scc2. The first finding supports the hypothesis where cohesin undergoes a significant conformational change during loading. Whereas the release of DNA has been long understood, it has not been until this year when two structures of the cohesin complex bound to its loader have been solved using proteins of *S.pombe* (Higashi *et al.*, 2020) and *H. sapiens* (Shi *et al.*, 2020) which have unravelled the mechanism by which DNA enters the ring. The experimental chapter of this thesis agrees with these published results. The results described in this thesis are discussed first, followed by a summary of the key findings from the two published papers.

The function of cohesin's coiled coils

Full-length human cohesin can be expressed in insect cells with final yields in the milligram range. Pulldown of cohesin using the strep-tag on the flexible Scc1 subunit leads to obtaining the entire complex with no contaminants. The complex can be purified to high purity without any significant losses of yields or disruption to the stoichiometry. In this study, *Hs*Cohesin was shown to be a flexible complex, agreeing with previously published studies. In addition to this, the folding of the human cohesin ring has been confirmed in our structural studies. Negatively stained *Hs*Cohesin showed a drastic conformational change induced by the cohesin loader, Nipbl^{C}. The two globular domains, the hinge and the ATPase heads, are normally spatially separated by the long coiled coils. We observed that the coil folding brings the ATPases to the hinge domains as a result of ATP-dependent Nipbl^{C} binding. As a

consequence of the observed flexibility and conformational changes to cohesin we have generated constructs which retain the loader's observed and proposed binding sites but reduce the flexibility, creating a construct more suitable for Cryo-EM. We observe that *Hs*Cohesin tolerates the removal of its coiled coils, and even increases obtainable yields from insect cells when compared to the full-length protein. The removal of the coils however introduced heterogeneity to the core complex as seen with negative staining of *Hs*Core. The changes I observed to the ATPase heads upon coil removal suggest that the coils may partially restrict the movement of the head module. This would agree with the observations that an ATPase dead mutant of the core complex, the Walker B mutant E1154Q, is also heterogeneous. This construct can bind but not hydrolyse ATP and thus heterogeneity is most likely the result of the coil removal rather than the failure to engage the ATPase heads. The binding of the nucleotide and subsequent dimerisation of the heads may be rigidifying the globular head components of the Smc subunits, but may not be exerting this effect over the proximal coils.

In vitro interactions between cohesin and its loader

Interactions observed in this study suggest that, with the exception of the coils emerging from the ATPase heads, the remaining coils are not crucial for the interactions with the loader. Furthermore, the removal of the coils led to the loss of recruitment of the hinge domains to the core module. The hinges may therefore be brought closer to the body as a response to the structural rearrangements in the coils. Binding of the hinge to the core module occurs during loading (Murayama and Uhlmann, 2015), but the affinity might be significantly reduced without the mechanical pulling of the hinge closer to the core by the folded coils. Interaction studies of Nipbl^{C} with cohesin show a low affinity for cohesin *in vitro*. By removing the coiled coils, I have observed a reduction in the affinity as judged by SEC experiments and glycerol gradients. Despite the presence of both samples in the same fractions of the gradient, the resolution of these gradients is not high enough to conclude an interaction. The interaction may require additional components which were not included in this study, such as a longer DNA sequences The DNA used in glycerol gradient experiments could have been of insufficient length to mediate this interaction.

The conserved architecture of the cohesin loader

The additional structural analyses carried out in this project involved studying the human cohesin loader and the hinge domains in isolation. I have characterised the Nipbl^{C} loader using negative staining which revealed a near identical structure to *Ag*Scc2^{C} and *Ct*Scc2^{C} as well as my characterisation of the full-length yeast loader. The hook-shaped architecture of the Scc2 subunit is conserved across multiple species, suggesting an importance of this feature. EM analysis further confirmed the high flexibility of the yeast Scc4 subunit which is bound to Scc2 by wrapping around its N terminus. Compared to the published Scc2^{C} structures, Nibpl^{C} formed crystals in several conditions with the resulting crystals not showing any diffraction. The construct was not optimised further as the C-terminus is highly reminiscent of the published structures and would most likely not show any drastic differences. Nipbl performs other roles outside of sister chromatid cohesion, including roles in transcription(Luna-Peláez *et al.*, 2019). Its central and N-terminal segments, only found in higher eukaryotes, perform the transcription-related function but this region is not predicted to have a strong secondary structure and is not included in the constructs used in this study. Optimisation of the constructs to include regions beyond the CTD would most likely not be beneficial for crystallography purposes. Similarly, no crystals of DNA-bound hinge domains were obtained. The choice of DNA sequence and its length can influence crystal packing and therefore more sequences would need to be tested for successful crystallisation.

DNA binding to the cohesin-loader subunits

DNA binding is one of the key features of cohesin and although multiple subunits were found to contact DNA, the precise regions and residues within the full ring are not known. The obtained results in this project demonstrate that both human and yeast proteins, specifically the C-terminal loader protein and the hinge domain, can bind DNA *in vitro*. The hinge domains show a similar affinity for ds and ssDNA. Furthermore, DNA binding to monomeric Smc3 hinge or heterodimeric hinge domain shows no change in affinity for the DNA. This is consistent with observations that the DNA binding patch is formed at the interface of the two Smc proteins where both most likely contribute to the DNA binding in the same fashion. Subjecting the cohesin loaders to DNA binding assays has showed that both yeast and human proteins bind DNA. The Scc2N-Scc4 module which does not engage with cohesin was previously

reported to not bind DNA (Chao *et al.*, 2015). The affinity of the loaders, or individual subunits of the loaders, for DNA cannot be accurately judged from the studies shown as the DNA concentrations were higher than ideal concentrations for DNA gel shift assays. Likewise, assays where binding affinities could be accurately determined were not performed.

Structural insights into topological cohesin loading

The Cryo-EM structures of fission yeast and human cohesin complexes bound to their loaders, obtained by the Uhlmann and Yu labs, respectively (Higashi *et al.*, 2020; Shi *et al.*, 2020), have revealed the precise contacts necessary for their interaction. The maps revealed a three-layered structure of the cohesin complex bound to Nipbl^{C}/Scc2, with the loader contacting each layer through a distinct region (Figure 6.1). The three layers are further bridged by DNA with at least 15 base length of the DNA directly contacting the loader. The structures revealed that although head engagement depends on the presence of ATP, the dimerisation is promoted and strengthened by Nipbl^{C}/Scc2, which packs against the ATPases with its hook-shaped central segment, and by DNA. Together, they promote conformational changes to the ATPases where the rearrangement of the Smc3 head and the proximal coiled coil by the NTD of the loader (the 'N' handle) and DNA bring the head closer to the opposite head.

Furthermore, both structures revealed that the kleisin gate opens during DNA loading. Rearrangements in cohesin weaken this gate which allows DNA entry and subsequent DNA "gripping" where DNA gets tightly bound by the loader and cohesin subunits. This is predicted to be followed by DNA passage through the ATPase gates, supported by findings that in this state, ATP is bound but not yet hydrolysed, suggesting the gripping state is the first state of DNA loading. The N handle of Nipbl^{C} contacts regions proximal to the kleisin gate, weakening this interactions to allow for DNA passage. Once through the kleisin gate, a channel formed between the ATPase heads and the N handle of the loader, which tightly grip DNA. DNA in this state is positioned to contact the DNA-sensing lysines on the Smc3 heads. Subsequent ATP hydrolysis mediated by the hook of the loader causes head disengagement and passage of the DNA through the ATPase head gate, leading to topological entrapment.

The hinge domain was shown to adopt a half open conformation in the human cohesin Cryo-EM structure which together with the crystal structure solved shows that this conformation can accommodate ssDNA. Interestingly, the Scc3 subunit contacts the hinge at residues which overlap with DNA binding. This could mean that the two contact DNA at different steps of the loading reaction. Scc3 was found to contact DNA in the gripping state. It is possible that in the topological embracement the hinge contacts DNA instead of Scc3.

The regions of cohesin shown to bind to the loader have all been included in the constructs used in the study presented in this thesis. Taken together, the conservation of the kink in the Smc3 coil, the structural appearance of Scc2 of the loader, and the observation of coil folding across many species, including bacterial, all suggest a conserved mechanism for cohesin loading (Anderson *et al.*, 2002; Li *et al.*, 2010; Soh *et al.*, 2015; Chao *et al.*, 2017; Diebold-Durand *et al.*, 2017; Higashi *et al.*, 2020; Shi *et al.*, 2020). Multiple of these observations have been seen and confirmed by this study.

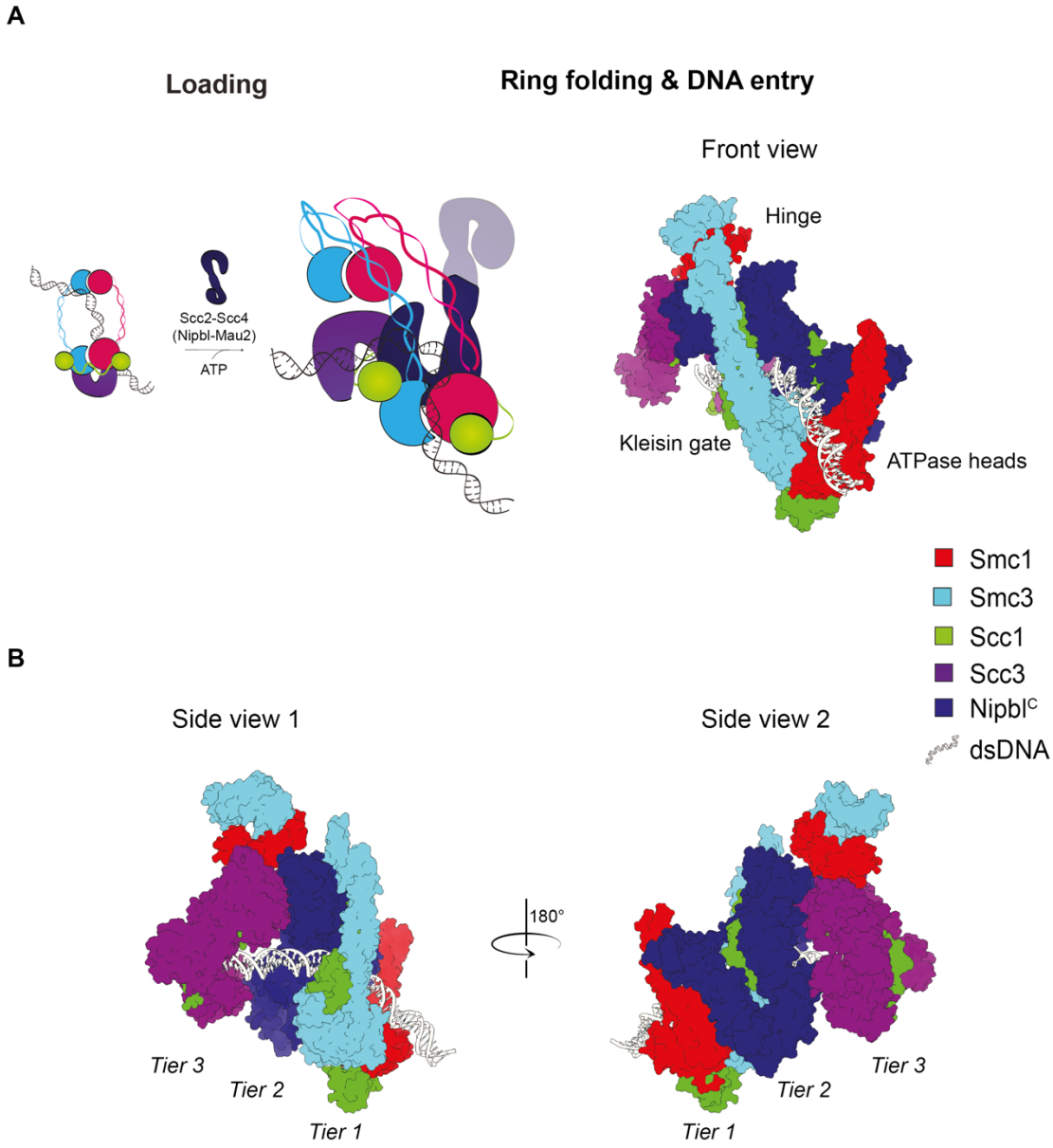

Figure 6.1 DNA entry into the cohesin ring.

A Cohesin ring folds during loading. A schematic representation (left) of the Cryo-EM map (right) of the human cohesin bound to its loader. **B** The three tiers of cohesin, all contacted by the cohesin loader. (PDB:6WG3).

6.2 The structure of the Chl1 helicase and implications for its function at the replication fork

Being the first chromosome loss mutant identified, the existence of the Chl1 protein has been known for over 40 years (Haber, 1974). Despite its numerous appearances in the literature, the structure of Chl1 has not been determined. Belonging to the XPD subfamily of proteins, its architecture was predicted but not experimentally confirmed. Up to the date of this thesis submission there is no available structure for this protein. Using Cryo-EM I have characterised the Chl1 protein of *Chaetomium thermophilum* at a 7.7Å resolution. The presented *Ct*Chl1 is therefore a novel structure of this helicase.

The architecture of Chl1 and implications for its function

Despite the lack of high resolution features, several statements can be made from this study. The major finding confirms the architecture conservation in the XPD subfamily, revealing three predicted domains characteristic of the XPD subfamily proteins. Chl1, like the XPD protein, is comprised of two RecA-like helicase domains separated by a third domain, the Arch domain. HD1 further contains the Fe-S cluster essential for helicase activity. The Arch domain extends above the two helicase domains (Fan *et al.*, 2008; Liu *et al.*, 2008). In the presented *Ct*Chl1 structure, the Arch domain adopts an open conformation where it is not folded towards the helicase domains. A similar observation has only been found with XPD structures with a disrupted Fe-S cluster. MS analysis together with the biochemical properties of *Ct*Chl1 strongly suggest an intact Fe-S cluster. The regions surrounding the cluster were shown to adopt a secondary structure, which would be impossible with a disrupted cluster and these regions would instead be unfolded. These structural changes were observed in structures of Apo-XPD where even the conserved cysteines are unresolved as a result of high degree of flexibility arising from the missing cluster. *Ct*Chl1 cysteines were found to lie in folded regions of the protein. Despite the lack of peptide coverage for all cysteines, those with poor coverage are still located in close proximity to folded regions which are not found in the Apo-XPD structures.

The translocation mechanism of the bacterial XPD homolog, DinG, showed the importance of the Arch domain for DNA unwinding (Cheng and Wigley, 2018). Its folding towards the helicase domains creates a tunnel for DNA passage. Together with the Fe-S cluster, the Arch domain facilitates DNA translocation in concert with ATP hydrolysis-dependent movement of the helicase domains. This agrees with other XPD structures which show a folded Arch domain upon nucleotide or DNA binding. Neither the nucleotide-free nor the nucleotide and DNA-bound *Ct*Chl1 proteins show the Arch domain in a closed conformation. Although it cannot be ruled out that DNA has not remained bound to the protein as it was not visible on negatively stain grids, the non-hydrolysable analog of ATP should alone induce a significant conformational change. The likelihood of the disrupted cluster being a cause is very low, yet several other possibilities could explain this observation. First, Chl1 does not translocate DNA in a similar fashion than XPD. This is highly unlikely as a high conservation of both nucleotide and DNA binding motifs are present in *Ct*Chl1 as well as other XPD subfamily proteins. This points to their importance in DNA unwinding shared by all XPD helicases. Secondly, Chl1 may have a different resting state than XPD proteins. The presence of ATP in these helicases should prime the protein for DNA engagement, accompanied with structural rearrangements. No such rearrangements have been visualised for *Ct*Chl1. This possibility cannot be ruled out as such rearrangements may not be visible at the resolution obtained for the nucleotide-free and bound sample.

The third alternative points to an auto-inhibited state of Chl1 (Figure 6.2). Auto-inhibition has been observed for the eukaryotic XPD, which in eukaryotes is a part of the TFIIH. The transition from an inactive to an active helicase occurs by relieving the inhibitory effect of MAT1 on XPD and subsequent activation by XPA of the TFIIH. The structural rearrangements which occur during this transition lead to freeing up the DNA binding channel, which is otherwise occupied by the plug, a short segment of the Arch domain of XPD. Since TFIIH is not present in prokaryotes, self-inhibition is not expected in archaeal and bacterial proteins such as DinG, further reinforced by the lack of the plug segment in these organisms. In higher eukaryotes, sequence alignments of human XPD, FANCJ and RTEL1 all show the 60 amino acid large plug carrying a net negative charge with on average 20% glutamines present in this segment. The exception to this is the human Chl1, where this region is positively

charged. The plug, like DNA, occupies the same cleft therefore the negative charge most likely serves an important function in binding. The different charge of the Arch domain region in Chl1 corresponding to the plug could explain why this domain is not folded towards the helicase body. However, if the protein is auto-inhibited, a certain segment must be responsible for the function the plug segment performs in XPD.

The insert of Chl1 and its speculated role in auto-inhibition

Compared to XPD, Chl1 contains an additional 200 amino acids after its canonical motif I. These residues form a negatively charged domain present in all Chl1 proteins. The insert has no homology to any deposited structures and its fold is unknown. Comparison of the Cryo-EM maps obtained in this study to human XPD structures shows an enlarged HD1 in comparison to XPD. The overall domain architecture of the left side of the EM map agrees with the structure of HD2 of XPD and thus the insert domain is believed to be correctly placed in the protein structure. Using a combination of Cryo-EM, MS analysis and biochemical characterisation, I have addressed the structure of this insert. Agreeing with structure predictions, the insert sits in the HD1 of the protein. Preliminary data on the mutant version of Chl1, *Ct*MiniChl1^{v2}, show a missing density in HD1 further confirming that the HD1 is larger due to the presence of the insert, as well as confirming the handedness of the *Ct*Chl1 EM maps. The secondary structure of the insert transits from a folded state into an unstructured C-terminus. I have isolated the insert of five different species, all expressed as a soluble domain, showing that the boundaries of the insert are conserved throughout the animal kingdom. Further validation of the inserts' architecture was observed in expression tests of GST-fused insert constructs. These have shown that imprecise manipulations of the construct length, which terminate the sequence in the disordered region, render the protein insoluble. As the predicted insert's boundaries reach the more conserved HD1 it becomes folded again. Keeping these amino acids in the construct supports solubility. Disordered sequences are flexible owing to the lack of a rigid secondary structure. This flexibility is projected onto the local resolution estimation of the *Ct*Chl1 map, which shows the lowest resolution of the protein being in this region. Aside from the insert, the remaining part of HD1 reaches the highest resolution despite the particle orientation bias towards HD2, suggesting the low local resolution is due to flexibility. Similarly, the Arch domain reaches the highest resolution in my Cryo-EM map. This region was shown

as a flexible domain in many XPD structures, but appears relatively rigid compared to the remaining domains of the protein. This observation further supports the hypothesis that the insert rather than the Arch domain mediate the inhibition.

The insert found in Chl1 is also found in FancJ. Interestingly, it is not present in Rtel1. Sequence alignments of the insert from Chl1 and FancJ show that although their positions are similar, their overall charge is drastically different. FancJ, which contains its MLH1 binding motif in this insert, has a positively charged insert compared to the negative charge of the Chl1 insert. Conversely, FancJ contains the negatively charged plug which Chl1 does not. The proximity of this insert to the region corresponding to the plug-binding segment of XPD could potentially mean that the insert segment takes over the role of auto-inhibition. Chl1's insert is predicted to contain a protein-protein interaction motif responsible for binding to Tof1. Although better conserved in vertebrates, it is also present in simpler eukaryotes. Binding of Chl1 to Tof1 in human cells is necessary for sister chromatid cohesion and resolution of complex DNA that impede DNA replication (Cali *et al.*, 2016; Cortone *et al.*, 2018; Lerner *et al.*, 2020). The helicase function of Chl1 is essential for these processes. These observations would suggest that Tof1 causes structural rearrangements to this domain which alleviate the potential inhibition and allow for DNA binding (Figure 7.2). This would further be supported by observations that ATPase and helicase activities of Chl1 are significantly enhanced by Tof1 in human cells. I conducted an expression test of Tof1's CTD for the purpose of observing whether this hypothesis is true, but the expression was not successful.

If the insert is truly blocking the DNA binding site, it would also prevent the Arch domain from coming closer towards the Fe-S cluster, as this region would be occupied by the insert density. Interestingly, preliminary data on the mutant version of *Ct*Chl1 with the insert removed (CtMiniChl1^{v2}) show a variation in obtained 2D classes. In the classification, the Arch domain, which is normally a predominant feature in numerous classes, is not visible despite being the highest resolution segment of the Cryo-EM map of the full-length protein. It is possible that removal of the insert allows the Arch domain to adopt a more closed conformation. The insert could prevent DNA binding together and restrict the Arch domain movement, a function necessary for DNA unwinding.

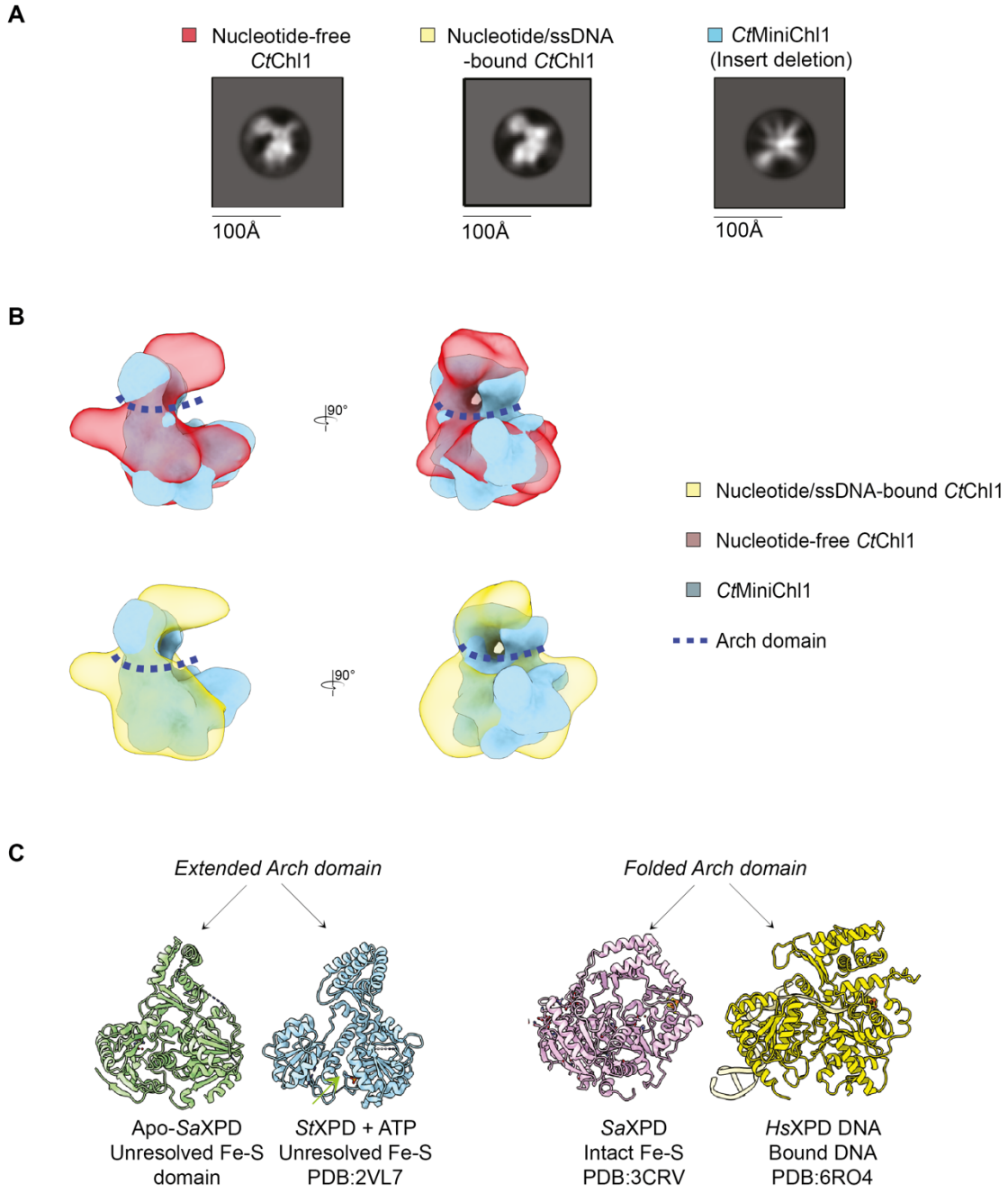

Figure 6.2 The potential mechanism of Chl1 auto-inhibition.

A Most represented 2D classes of three samples with **B** an overlay of their initial 3D models. The absence of the Arch domain in *Ct*MiniChl1^{v2} is notable. **C** Examples of XPD structures with opened and closed Arch domains. The Arch domain movement is found in all species with a solved crystal structure.

Interactions of Chl1 and Ctf4

I confirmed the interaction of Chl1 and Ctf4 in *S. cerevisiae* using targeted IPs as previously published (Borges *et al.*, 2013; Samora *et al.*, 2016). Due to human error this interaction could not be compared between normal and stress conditions. Two other approaches to compare this interaction were unsuccessful, namely interaction proteomics and pulldowns with recombinant proteins. It has previously been suggested that the affinity of Ctf4 for Chl1 is very low despite the two binding via a conserved motif that is shared with even the high affinity binding interaction partners of Ctf4, such as Sld5 of GINS. Attempts to crosslink these two proteins or enhance their binding using site-directed mutagenesis to change the DDIL motif to mimic Sld5 has not shown promise. The low affinity of the two proteins towards each other therefore hinders their structural characterisation as a complex.

DNA binding properties of Chl1

The low affinity of CtChl1 for ssDNA raises a question whether Chl1 requires a binding partner for efficient DNA binding. Characterisation of the interaction between Chl1 and Ctf4 in sister chromatid cohesion in yeast shows no involvement of Chl1's helicase activity in cohesion establishment (Samora *et al.*, 2016). The CIP-box of Chl1 located in HD2 of the protein would not sterically clash with the DNA binding patch predicted from XPD crystal structures. Preliminary DNA binding experiments of the effect of *Ct*Ctf4^{CTD} on *Ct*Chl1 DNA binding has shown no inhibition. The lack of high affinity of these two proteins however hinders experiments which could further study Ctf4's effect over Chl1's helicase. Conversely, protein-protein interactions which influence DNA binding have been observed with the human Chl1 protein, DDX11, and the human ortholog of Tof1, Timeless. Timeless enhances both ATP hydrolysis and DNA binding activities of DDX11, recruiting it to the replication fork for unwinding DNA roadblocks without itself unwinding any DNA (Cali *et al.*, 2016; Lerner *et al.*, 2020). Structural and biophysical characterisation of *Ct*Chl1 raised the question whether the insert prevents DNA binding. Fluorescence anisotropy measurements with the full-length protein shown in this study do not reach a saturated state, but conversely, results with *Ct*MiniChl1^{V2} show that this construct can bind DNA with a 25-fold higher affinity with a Kd of 38nM. Similar binding is observed with prokaryotic XPDs. Eukaryotic XPD was shown to be in an auto-inhibited state before its release from inhibition by XPA. The structural rearrangements in eukaryotic XPD needed for DNA binding point to a difference

between species. It is possible that, like XPD, eukaryotic Chl1 is also in an auto-inhibited state. Together with the absence of a visible Arch domain in this construct, these results point to a mechanism where structural rearrangements in the insert domain are required for DNA binding to the protein (Figure 6.3).

Interactions of Chl1 at the replication fork and future work

The obtained structural information informs on the behaviour of the protein. Chl1 is not an ideal candidate for X-ray crystallography due to its flexible region. Its size however limits the achievable resolution with Cryo-EM both due to difficulties in particle picking as a result of decreased visibility and contrast, as well as the lack of features for successful alignments. Since the primary technique applied in this thesis is Cryo-EM, suggestions for future work mainly focus on this approach. In order to obtain a higher resolution structure of Chl1, a complex between Chl1 and another interacting partner should be formed. Studying the cohesin complex has been shown to be challenging and since the precise binding location of Chl1 is not known, structural characterisation of this interaction should first be narrowed down to individual regions of the cohesin complex with techniques such as IPs, pulldowns, peptide arrays or proximity-based approaches like FRET. Alternatively, the interaction between Chl1 and Tof1 could further be exploited. Tof1 is a large protein which interacts with Csm3, the MCM ATPase, DNA and Parp1 (Lerner *et al.*, 2020). The residues which mediate the interaction with Chl1 has not been identified in Tof1 but could potentially lie in the regions which are not occupied by Csm3. This is because Csm3 and Tof1 form a tight interaction where the expression of full-length Tof1 must be accompanied by expression of Csm3. The region unoccupied by Csm3 is the CTD of Tof1 but as mentioned above, expressing this sequence in isolation has not been successful and therefore may need to be co-expressed with Csm3.

To identify the mechanisms of Chl1's helicase activity, it is important to obtain data on DNA unwinding. For such experiments, the difference between insert-containing and insert-free proteins could be of use. Furthermore, binding of Chl1 to Tof1 under replication stress conditions in *S. cerevisiae* should be pursued. An internal deletion of the insert and its subsequent comparison to the full-length protein may elucidate new mechanisms by which Chl1 mediates the stress response in yeast.

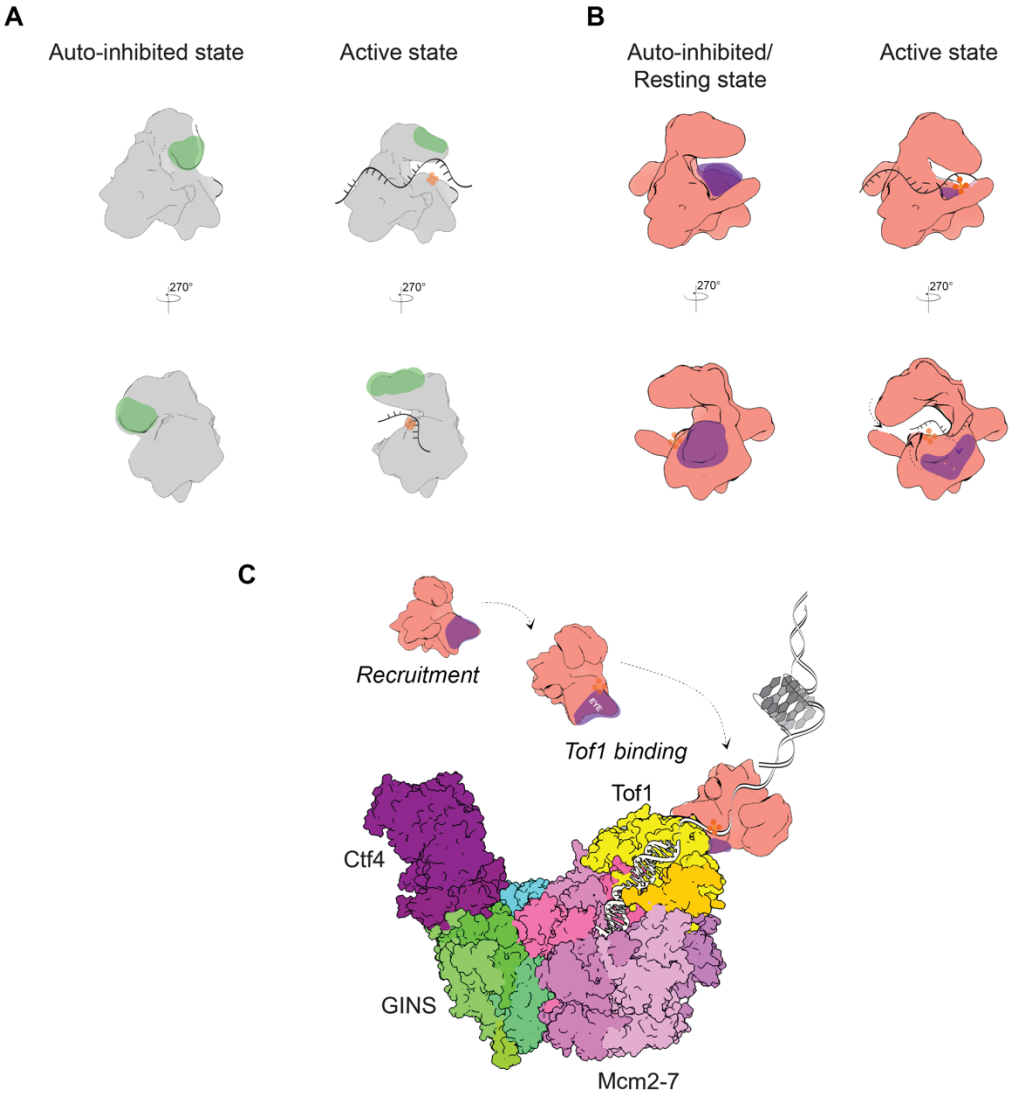

Figure 6.3 A hypothesised mechanism of Chl1 action at the fork.

A The observed auto-inhibited state of the XPD protein. **B** The hypothesised auto-inhibited (resting) state of Chl1. Chl1's insert (purple) occupies the DNA binding site on HD1, similarly to the plug segment (green in A) of XPD. Conformational rearrangements mediated by the reorganisation of the insert segment, potentially induced by protein-protein interactions, can lead to relieving this state, freeing the DNA binding site and switching the protein into an active DNA-unwinding (active) helicase. In this arrangement, the Arch domain would be in close contact with the Fe-S cluster in HD1. **B** Auto-inhibition can potentially be relieved by binding to Tof1 via the EYE motif in Chl1's insert, allowing the helicase to bind DNA and resolve obstacles ahead of the fork.

Reference List

Abid Ali, F., Douglas, M. E., Locke, J., Pye, V. E., Nans, A., Diffley, J. F. X. and Costa, A. (2017) Cryo-EM structure of a licensed DNA replication origin, *Nature Communications*, 8, 1–10.

Albert, S. et al. (2017) Proteasomes tether to two distinct sites at the nuclear pore complex, *Proceedings of the National Academy of Sciences of the United States of America*, 114, 13726–13731.

Alcasabas, A. A. et al. (2001) Mrc1 transduces signals of DNA replication stress to activate Rad53, *Nature Cell Biology*, 3, 958–965.

Almedawar, S., Colomina, N., Bermúdez-López, M., Pociño-Merino, I. and Torres-Rosell, J. (2012) A SUMO-dependent step during establishment of sister chromatid cohesion, *Current Biology*, 22, 1576–1581.

Alt, A. et al. (2017) Specialized interfaces of Smc5/6 control hinge stability and DNA association, *Nature Communications*, 8, 1–14.

Anderson, D. E., Losada, A., Erickson, H. P. and Hirano, T. (2002) Condensin and cohesin display different arm conformations with characteristic hinge angles, *The Journal of Cell Biology*, 156, 419–424.

Andrews, E. A., Palecek, J., Sergeant, J., Taylor, E., Lehmann, A. R. and Watts, F. Z. (2005) Nse2, a Component of the Smc5-6 Complex, Is a SUMO Ligase Required for the Response to DNA Damage, *Molecular and Cellular Biology*, 25, 185–196.

Bae, S.-H., Bae, K.-H., Kim, J.-A. and Seo, Y.-S. (2001) RPA governs endonuclease switching during processing of Okazaki fragments in Eukaryotes, *Nature*, 412, 326–333.

Baker, L. A., Smith, E. A., Bueler, S. A. and Rubinstein, J. L. (2010) The resolution dependence of optimal exposures in liquid nitrogen temperature electron cryomicroscopy of catalase crystals, *Journal of Structural Biology*, 169, 431–437.

Baranovskiy, A. G., Babayeva, N. D., Zhang, Y., Gu, J., Suwa, Y., Pavlov, Y. I. and Tahirov, T. H. (2016) Mechanism of Concerted RNA-DNA Primer Synthesis by the Human Primosome, *Journal of Biological Chemistry*, 291, 10006–10020.

Barber, L. J. et al. (2008) RTEL1 Maintains Genomic Stability by Suppressing Homologous Recombination, *Cell*, 135, 261–271.

Baretic, D., Jenkyn-Bedford, M., Aria, V., Cannone, G., Skehel, M. and Yeeles, J. T. P. (2020) Cryo-EM Structure of the Fork Protection Complex Bound to CMG at a Replication Fork, *Molecular Cell*, 78, 1–15.

Barski, M. S., Minnell, J. J., Hodakova, Z., Pye, V. E., Nans, A., Cherepanov, P. and Maertens, G. N. (2020) Cryo-EM structure of the deltaretroviral intasome in complex with the PP2A regulatory subunit B56γ, *Nature Communications*, 11, 1–10.

Beckmann, R., Spahn, C. M. T., Eswar, N., Helmers, J., Penczek, P. A., Sali, A., Frank, J. and Blobel, G. (2001) Architecture of the protein-conducting channel

associated with the translating 80S ribosome, *Cell*, 107, 361–372.

Beckouët, F., Hu, B., Roig, M. B., Sutani, T., Komata, M., Uluocak, P., Katis, V. L., Shirahige, K. and Nasmyth, K. (2010) An Smc3 Acetylation Cycle Is Essential for Establishment of Sister Chromatid Cohesion, *Molecular Cell*, 39, 689–699.

Beckouët, F. et al. (2016) Releasing Activity Disengages Cohesin's Smc3/Scc1 Interface in a Process Blocked by Acetylation, *Molecular Cell*, 61, 563–574.

Ben-Shahar, T. R., Heeger, S., Lehane, C., East, P., Flynn, H., Skehel, M. and Uhlmann, F. (2008) Eco1-Dependent Cohesin Sister Chromatid Cohesion, *Science*, 321, 563–566.

Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A. J. and Berger, B. (2019) Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs, *Nature Methods*, 16, 1153–1160.

Bermúdez-López, M., Ceschia, A., de Piccoli, G., Colomina, N., Pasero, P., Aragón, L. and Torres-Rosell, J. (2010) The Smc5/6 complex is required for dissolution of DNA-mediated sister chromatid linkages, *Nucleic Acids Research*, 38, 6502–6512.

Boland, A., Martin, T. G., Zhang, Z., Yang, J., Bai, X., Chang, L. and Scheres, S. H. W. (2017) Cryo-EM structure of a metazoan separase – securin complex at near-atomic resolution, *Nature Structural & Molecular Biology*, 24, 414–418.

Borges, V., Lehane, C., Lopez-Serra, L., Flynn, H., Skehel, M., Ben-Shahar, T. R. and Uhlmann, F. (2010) Hos1 Deacetylates Smc3 to Close the Cohesin Acetylation Cycle, *Molecular Cell*, 39, 677–688.

Borges, V., Smith, D. J., Whitehouse, I. and Uhlmann, F. (2013) An Eco1-independent sister chromatid cohesion establishment pathway in S. cerevisiae, *Chromosoma*, 122, 121–134.

Bot, C., Pfeiffer, A., Giordano, F., Edara, D. M., Dantuma, N. P. and Ström, L. (2017) Independent Mechanisms Recruit the Cohesin Loader Protein NIPBL to Sites of DNA Damage, *Journal of Cell Science*, 130, 1134–1146.

Brandão, H. B., Paul, P., van den Berg, A. A., Rudner, D. Z., Wang, X. and Mirny, L. A. (2019) RNA polymerases as moving barriers to condensin loop extrusion, *Proceedings of the National Academy of Sciences of the United States of America*, 116, 20489–20499.

Brilot, A. F., Chen, J. Z., Cheng, A., Pan, J., Harrison, S. C., Potter, C. S., Carragher, B., Henderson, R. and Grigorieff, N. (2012) Beam-induced motion of vitrified specimen on holey carbon film, *Journal of Structural Biology*, 177, 630–637.

Buheitel, J. and Stemmann, O. (2013) Prophase pathway-dependent removal of cohesin from human chromosomes requires opening of the Smc3-Scc1 gate., *The EMBO journal*, 32, 666–76.

Burgers, P. M. J. and Kunkel, T. A. (2017) Eukaryotic DNA Replication Fork, *Annual Review of Biochemistry*, 86, 417–438.

Bürmann, F., Basfeld, A., Vazquez Nunez, R., Diebold-Durand, M. L., Wilhelm, L. and Gruber, S. (2017) Tuned SMC Arms Drive Chromosomal Loading of

Prokaryotic Condensin, *Molecular Cell*, 65, 861-872.e9.

Bürmann, F. et al. (2019) A folded conformation of MukBEF and cohesin, *Nature Structural and Molecular Biology*, 26, 227–236.

Busslinger, G. A., Stocsits, R. R., Lelij, P. Van Der, Axelsson, E., Tedeschi, A. and Galjart, N. (2017) Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl, *Nature Letters*, 544, 503–507.

Bustard, D. E., Menolfi, D., Jeppsson, K., Ball, L. G., Dewey, S. C., Shirahige, K., Sjögren, C., Branzei, D. and Cobb, J. A. (2012) During replication stress, non-Smc element 5 (Nse5) is required for Smc5/6 protein complex functionality at stalled forks, *Journal of Biological Chemistry*, 287, 11374–11383.

Bustard, D. E., Ball, L. G. and Cobb, J. A. (2016) Non-Smc element 5 (Nse5) of the Smc5/6 complex interacts with SUMO pathway components, *Biology Open*, 5, 777–785.

Cali, F., Bharti, S. K., Perna, R. Di, Jr, R. M. B. and Pisani, F. M. (2016) Tim / Timeless , a member of the replication fork protection complex , operates with the Warsaw breakage syndrome DNA helicase DDX11 in the same fork recovery pathway, *Nucleic Acids Research*, 44, 705–717.

Campbell, M. G. et al. (2012) Movies of ice-embedded particles enhance resolution in electron cryo-microscopy, *Structure*, 20, 1823–1828.

Cantor, S., Drapkin, R., Zhang, F., Lin, Y., Han, J., Pamidi, S. and Livingston, D. M. (2004) The BRCA1-associated protein BACH1 is a DNA helicase targeted by clinically relevant inactivating mutations, *Proceedings of the National Academy of Sciences of the United States of America*, 101, 2357–2362.

Cantor, S. B. et al. (2001) BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function, *Cell*, 105, 149–160.

De Carlo, S. and Harris, J. R. (2011) Negative staining and cryo-negative staining of macromolecules and viruses for TEM, *Micron*, 42, 117–131.

Carragher, B., Kisseberth, N., Kriegman, D., Milligan, R. A., Potter, C. S., Pulokas, J. and Reilein, A. (2000) Leginon: An automated system for acquisition of images from vitreous ice specimens, *Journal of Structural Biology*, 132, 33–45.

Ceccaldi, R., Sarangi, P. and D'Andrea, A. D. (2016) The Fanconi anaemia pathway: New players and new functions, *Nature Reviews Molecular Cell Biology*, 17, 337–349.

Chan, K.-L., Maurici Brunet, R., Hu, B., Beckouët, F., Metson, J. and Nasmyth, K. (2012) Cohesin's DNA Exit Gate Is Distinct from Its Entrance Gate and Is Regulated by Acetylation, *Cell*, 150, 961–974.

Chan, K. L., Gligoris, T., Upcher, W., Kato, Y., Shirahige, K., Nasmyth, K. and Beckouët, F. (2013) Pds5 promotes and protects cohesin acetylation, *Proceedings of the National Academy of Sciences of the United States of America*, 110, 13020–13025.

Chao, W. C. H., Murayama, Y., Costa, A., Singleton, M. R., Costa, A., Uhlmann, F. and Singleton, M. R. (2015) Structural Studies Reveal the Functional Modularity of the Scc2-Scc4 Cohesin Loader, *Cell Reports*, 12, 719–725.

Chao, W. C. H., Wade, B. O., et al. (2017) Structural Basis of Eco1-Mediated Cohesin Acetylation, *Scientific Reports*, 7, 1–11.

Chao, W. C. H., Murayama, Y., et al. (2017) Structure of the cohesin loader Scc2, *Nature Communications*, 8, 13952.

Chapard, C., Jones, R., van Oepen, T., Scheinost, J. C. and Nasmyth, K. (2019) Sister DNA Entrapment between Juxtaposed Smc Heads and Kleisin of the Cohesin Complex, *Molecular Cell*, 75, 224-237.e5.

Chatterjee, A., Zakian, S., Hu, X. and Singleton, M. R. (2013) Structural insights into the regulation of cohesion establishment by Wpl1, *The EMBO Journal*, 32, 677–687.

Chen, S., McMullan, G., Faruqi, A. R., Murshudov, G. N., Short, J. M., Scheres, S. H. W. and Henderson, R. (2013) High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy, *Ultramicroscopy*, 135, 24–35.

Cheng, K. and Wigley, D. B. (2018) DNA translocation mechanism of an XPD family helicase, *eLife*, 7, 1–16.

Cheng, Y., Grigorieff, N., Penczek, P. A. and Walz, T. (2015) Primer A Primer to Single-Particle Cryo-Electron Microscopy, *Cell*, 161, 438–449.

Ciosk, R., Shirayama, M., Shevchenko, Anna, Tanaka, T., Toth, A., Shevchenko, Andrej and Nasmyth, K. (2000) Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins, *Molecular Cell*, 5, 243–254.

Cobbe, N. and Heck, M. M. S. (2004) The Evolution of SMC Proteins : Phylogenetic Analysis and Structural Implications, *Molecular Biology and Evolution*, 21, 332–347.

Coin, F., Marinoni, J. C., Rodolfo, C., Fribourg, S., Pedrini, A. M. and Egly, J. M. (1998) Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH, *Nature Genetics*, 20, 184–188.

Coin, F., Oksenych, V., Mocquet, V., Groh, S., Blattner, C. and Egly, J. M. (2008) Nucleotide Excision Repair Driven by the Dissociation of CAK from TFIIH, *Molecular Cell*, 31, 9–20.

Coin, F., Oksenych, V. and Egly, J. M. (2007) Distinct Roles for the XPB/p52 and XPD/p44 Subcomplexes of TFIIH in Damaged DNA Opening during Nucleotide Excision Repair, *Molecular Cell*, 26, 245–256.

Cortone, G., Zheng, G., Pensieri, P., Chiappetta, V., Tat, R., Malacaria, E., Pichierri, P., Id, H. Y. and Id, F. M. P. (2018) Interaction of the Warsaw breakage syndrome DNA helicase DDX11 with the replication fork-protection factor Timeless promotes sister chromatid cohesion, *PLoS Genetics*, 14, 1–22.

Costa, A., Ilves, I., Tamberg, N., Petojevic, T., Nogales, E., Botchan, M. R. and Berger, J. M. (2011) The structural basis for MCM2-7 helicase activation by GINS and Cdc45, *Nature Structural and Molecular Biology*, 18, 471–479.

Costa, A., Hood, I. V. and Berger, J. M. (2013) Mechanisms for Initiating Cellular

DNA Replication, *Annual Review of Biochemistry*, 82, 25–54.

D'Imprima, E., Floris, D., Joppe, M., Sánchez, R., Grininger, M. and Kühlbrandt, W. (2019) Protein denaturation at the air-water interface and how to prevent it, *eLife*, 8, 1–18.

Danev, R., Buijsse, B., Khoshouei, M., Plitzko, J. M. and Baumeister, W. (2014) Volta potential phase plate for in-focus phase contrast transmission electron microscopy, *Proceedings of the National Academy of Sciences*, 111, 15635–15640.

Danev, R. and Baumeister, W. (2016) Cryo-EM single particle analysis with the Volta phase plate, 1–14.

Danev, R., Tegunov, D. and Baumeister, W. (2017) Using the Volta phase plate with defocus for cryo-EM single particle analysis, 1–9.

Davidson, I. F. et al. (2016) Rapid movement and transcriptional re-localization of human cohesin on DNA, *The EMBO journal*, 35, 1–15.

Delamarre, A. et al. (2019) Molecular Cell MRX increases chromatin accessibility at stalled replication forks to promote nascent DNA resection and cohesin loading, *Molecular cell*, 77, 1–16.

Van Deursen, F., Sengupta, S., de Piccoli, G., Sanchez-Diaz, A. and Labib, K. (2012) emboj.2012.69.pdf, *The EMBO journal*, 31, 2195–2206.

Diebold-Durand, M. L. et al. (2017) Structure of Full-Length SMC and Rearrangements Required for Chromosome Organization, *Molecular Cell*, 67, 334-347.e5.

Douglas, M. E., Ali, F. A., Costa, A. and Diffley, J. F. X. (2018) The mechanism of eukaryotic CMG helicase activation, *Nature*, 555, 265–268.

Duan, X., Yang, Y., Chen, Y., Arenz, J., Rangi, G. K., Zhao, X. and Ye, H. (2009) Architecture of the Smc5 / 6 Complex of Saccharomyces cerevisiae Reveals a Unique Interaction between the Nse5-6 Subcomplex and the Hinge Regions of Smc5 and Smc6 * □, *The Journal of Biological Chemistry*, 284, 8507–8515.

Dubochet, J., Chang, J. J., Freeman, R., Lepault, J. and McDowall, A. W. (1982) Frozen aqueous suspensions, *Ultramicroscopy*, 10, 55–61.

Eagen, K. P. (2018) Principles of Chromosome Architecture Revealed by Hi-C, *Trends in Biochemical Sciences*, xx, 1–10.

Egelman, E. H. (2016) The Current Revolution in Cryo-EM, *Biophysical Journal*, 110, 1008–1012.

Egerton, R. F. (2014) Choice of operating voltage for a transmission electron microscope, *Ultramicroscopy*, 145, 85–93.

Eickhoff, P. et al. (2019) Molecular Basis for ATP-Hydrolysis-Driven DNA Translocation by the CMG Helicase of the Eukaryotic Replisome, *Cell Reports*, 28, 2673-2688.e8.

Elbatsh, A. M. O. et al. (2016) Cohesin Releases DNA through Asymmetric ATPase-Driven Ring Opening, *Molecular Cell*, 61, 575–588.

Elmlund, D. and Elmlund, H. (2012) SIMPLE: Software for ab initio reconstruction

of heterogeneous single-particles, *Journal of Structural Biology*, 180, 420–427.

Elmlund, D., Le, S. N. and Elmlund, H. (2017) High-resolution cryo-EM: the nuts and bolts, *Current Opinion in Structural Biology*, 46, 1–6.

Elmlund, H., Elmlund, D. and Bengio, S. (2013) PRIME: Probabilistic initial 3D model generation for single-particle cryo-electron microscopy, *Structure*, 21, 1299–1306.

Evrin, C., Clarke, P., Zech, J., Lurz, R., Sun, J., Uhle, S., Li, H., Stillman, B. and Speck, C. (2009) A double-hexameric MCM2-7 complex is loaded onto origin DNA during licensing of eukaryotic DNA replication, *Proceedings of the National Academy of Sciences of the United States of America*, 106, 20240–20245.

Fan, L., Fuss, J. O., Cheng, Q. J., Arvai, A. S., Hammel, M., Roberts, V. A., Cooper, P. K. and Tainer, J. A. (2008) XPD Helicase Structures and Activities: Insights into the Cancer and Aging Phenotypes from XPD Mutations, *Cell*, 133, 789–800.

Faramarz, A., Balk, J. A., van Schie, J. J. M., Oostra, A. B., Ghandour, C. A., Rooimans, M. A., Wolthuis, R. M. F. and de Lange, J. (2020) Non-redundant roles in sister chromatid cohesion of the DNA helicase DDX11 and the SMC3 acetyl transferases ESCO1 and ESCO2, *PloS one*, 15, e0220348.

Farina, A., Shin, J., Kim, D., Bermudez, V. P., Kelman, Z., Seo, Y. and Hurwitz, J. (2008a) Studies with the Human Cohesin Establishment Factor , ChlR1, *The Journal of Biological Chemistry*, 283, 20925–20936.

Farina, A., Shin, J., Kim, D., Bermudez, V. P., Kelman, Z., Seo, Y. and Hurwitz, J. (2008b) Studies with the Human Cohesin Establishment Factor , ChlR1, 283, 20925–20936.

Feng, X. et al. (2017) A Fast and Effective Microfluidic Spraying-Plunging Method for High-Resolution Single-Particle Cryo-EM, *Structure*, 25, 663-670.e3.

Fernández-Cid, A. et al. (2013) An ORC/Cdc6/MCM2-7 Complex Is Formed in a Multistep Reaction to Serve as a Platform for MCM Double-Hexamer Assembly, *Molecular Cell*, 50, 577–588.

Fousteri, M. I. and Lehmann, A. R. (2000) A novel SMC protein complex in Schizosaccharomyces pombe contains the Rad18 DNA repair protein, *The EMBO Journal*, 19, 1691–1702.

Frigola, J. et al. (2017) Cdt1 stabilizes an open MCM ring for helicase loading, *Nature Communications*, 8, 1–10.

Frizzell, A., Nguyen, J. H. G., Petalcorin, M. I. R., Turner, K. D., Boulton, S. J., Freudenreich, C. H. and Lahue, R. S. (2014) RTEL1 inhibits trinucleotide repeat expansions and fragility, *Cell Reports*, 6, 827–835.

Gambus, A., Jones, R. C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R. D. and Labib, K. (2006) GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks, *Nature Cell Biology*, 8, 358–366.

Gambus, A., Van, F., Polychronopoulos, D., Foltman, M., Jones, R. C., Edmondson, R. D., Calzada, A. and Labib, K. (2009) A key role for Ctf4 in

coupling the MCM2-7 helicase to DNA polymerase a within the eukaryotic replisome, *The EMBO Journal*, 28, 2992–3004.

Ganai, R. A., Osterman, P. and Johansson, E. (2015) Yeast DNA polymerase ε catalytic core and holoenzyme have comparable catalytic rates, *Journal of Biological Chemistry*, 290, 3825–3835.

Gandhi, R., Gillespie, P. J. and Hirano, T. (2006) Human Wapl Is a Cohesin-Binding Protein that Promotes Sister-Chromatid Resolution in Mitotic Prophase, *Current Biology*, 16, 2406–2417.

Gerlich, D., Koch, B., Dupeux, F., Peters, J. M. and Ellenberg, J. (2006) Live-Cell Imaging Reveals a Stable Cohesin-Chromatin Interaction after but Not before DNA Replication, *Current Biology*, 16, 1571–1578.

Gerring, S. L., Spencer, F. and Hieter, P. (1990) The CHL 1 (CTF 1) gene product of Saccharomyces cerevisiae is important for chromosome transmission and normal cell cycle progression in G2/M., *The EMBO Journal*, 9, 4347–4358.

Gibcus, J. H. et al. (2018) A pathway for mitotic chromosome formation, *Science*, 359.

Glaeser, R. M. (2016) *Specimen Behavior in the Electron Beam*. 1st edn, *Methods in Enzymology*. 1st edn. Elsevier Inc.

Glaeser, R. M. (2018) Proteins, interfaces, and cryo-EM grids, *Current Opinion in Colloid and Interface Science*, 34, 1–8.

Glaeser, R. M. and Han, B.-G. (2017) Opinion: hazards faced by macromolecules when confined to thin aqueous films, *Biophysics Reports*, 3, 1–7.

Gligoris, T. G. et al. (2014) Closing the cohesin ring: Structure and function of its Smc3-kleisin interface, *Science*, 346, 963–967.

Grant, T. and Grigorieff, N. (2015) Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6, *eLife*, 4, 1–19.

Grassucci, R. A., Taylor, D. J. and Frank, J. (2007) Preparation of macromolecular complexes for cryo-electron microscopy, *Nature Protocols*, 2, 3239–3246.

Greber, B. J., Nguyen, T. H. D., Fang, J., Afonine, P. V., Adams, P. D. and Nogales, E. (2017) The cryo-electron microscopy structure of human transcription factor IIH, *Nature*, 549, 414–417.

Griese, J. J., Witte, G. and Hopfner, K. (2010) Structure and DNA binding activity of the mouse condensin hinge domain highlight common and diverse features of SMC proteins, *Nucleic Acids Research*, 38, 3454–3465.

Gruber, S., Arumugam, P., Katou, Y., Kuglitsch, D., Helmhart, W., Shirahige, K. and Nasmyth, K. (2006) Evidence that Loading of Cohesin Onto Chromosomes Involves Opening of Its SMC Hinge, *Cell*, 127, 523–537.

Gruber, S., Haering, C. H. and Nasmyth, K. (2003) Chromosomal Cohesin Forms a Ring, *Cell*, 112, 765–777.

Guacci, V., Koshland, D. and Strunnikov, A. (1997) A direct link between sister chromatid cohesion and chromosome condensation revealed through the

analysis of MCD1 in S. cerevisiae, *Cell*, 91, 47–57.

Guo, Q. et al. (2018) In Situ Structure of Neuronal C9orf72 Poly-GA Aggregates Reveals Proteasome Recruitment, *Cell*, 172, 696-705.e12.

Haarhuis, J. H. I. et al. (2017) The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension, *Cell*, 169, 693-707.e14.

Haber, J. E. (1974) Bisexual mating behavior in a diploid of Saccharomyces cerevisiae: evidence for genetically controlled non random chromosome loss during vegetative growth, *Genetics*, 78, 843–858.

Haering, C. H., Lo, J., Hochwagen, A. and Nasmyth, K. (2002) Molecular Architecture of SMC Proteins and the Yeast Cohesin Complex, *Molecular Cell*, 9, 773–788.

Haering, C. H., Schoffnegger, D., Nishino, T., Helmhart, W., Nasmyth, K. and Löwe, J. (2004) Structure and stability of cohesin's Smc1-kleisin interaction, *Molecular Cell*, 15, 951–964.

Hara, K., Zheng, G., Qu, Q., Liu, H., Ouyang, Z., Chen, Z., Tomchick, D. R. and Yu, H. (2014) Structure of cohesin subcomplex pinpoints direct shugoshin-Wapl antagonism in centromeric cohesion, *Nature Publishing Group*, 21, 864–870.

Hassler, M. et al. (2019) Structural Basis of an Asymmetric Condensin ATPase Cycle, *Molecular Cell*, 74, 1175-1188.e9.

Hauf, S., Roitinger, E., Koch, B., Dittrich, C. M., Mechtler, K. and Peters, J. M. (2005) Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2, *PLoS Biology*, 3, 0419–0432.

Van Heel, M. (1987) Angular Reconstitution: A posteriori assignment of projection directions for 3D reconstruction, *Ultramicroscopy*, 21, 111–124.

Van Heel, M. and Schatz, M. (2005) Fourier shell correlation threshold criteria, *Journal of Structural Biology*, 151, 250–262.

Herzik, M. A., Wu, M. and Lander, G. C. (2019) High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM, *Nature Communications*, 10, 1–9.

Higashi, T. L. et al. (2020) A Structure-Based Mechanism for DNA Entry into the Cohesin Ring, *Molecular Cell*, 79, 917-933.e9.

Hinshaw, S. M., Makrantoni, V., Kerr, A., Marston, A. L. and Harrison, S. C. (2015) Structural evidence for Scc4-dependent localization of cohesin loading, *eLife*, 4, 1–15.

Hinshaw, S. M., Makrantoni, V., Harrison, S. C. and Marston, A. L. (2017) The Kinetochore Receptor for the Cohesin Loading Complex, *Cell*, 171, 72-84.e13.

Hirano, M. and Hirano, T. (2006) Opening closed arms: Long-distance activation of SMC ATPase by hinge-DNA interactions, *Molecular Cell*, 21, 175–186.

Hirano, T., Kobayashi, R. and Hirano, M. (1997) Condensins, Chromosome Condensation Protein Complexes Containing XCAP-C, XCAP-E and a Xenopus Homolog of the Drosophila Barren Protein, *Cell*, 89, 511–521.

Hirano, T. and Mitchison, T. J. (1994) A heterodimeric coiled-coil protein required for mitotic chromosome condensation in vitro, *Cell*, 79, 449–458.

Hirota, Y. and Lahti, J. M. (2000) Characterization of the enzymatic activity of hChlR1 , a novel human DNA helicase, *Nucleic Acids Research*, 28.

Hons, M. T., Huis Int Veld, P. J., Kaesler, J., Rombaut, P., Schleiffer, A., Herzog, F., Stark, H. and Peters, J. M. (2016) Topology and structure of an engineered human cohesin complex bound to Pds5B, *Nature Communications*, 7, 1–11.

Huis In 'T Veld, P. J., Herzog, F., Ladurner, R., Davidson, I. F., Piric, S., Kreidl, E., Bhaskara, V., Aebersold, R. and Peters, J. (2014) Characterization of a DNA exit gate in the human cohesin ring, *Science*, 346, 968–972.

Im, J. S., Ki, S. H., Farina, A., Jung, D. S., Hurwitz, J. and Lee, J. K. (2009) Assembly of the Cdc45-Mcm2-7-GINS complex in human cells requires the Ctf4/And-1, RecQL4, and Mcm10 proteins, *Proceedings of the National Academy of Sciences of the United States of America*, 106, 15628–15632.

Ivanov, D., Schleiffer, A., Eisenhaber, F., Mechtler, K., Haering, C. H. and Nasmyth, K. (2002) Eco1 is a novel acetyltransferase that can acetylate proteins involved in cohesion, *Current Biology*, 12, 323–328.

Iwasaki, O., Tanizawa, H., Kim, K. D., Yokoyama, Y., Corcoran, C. J., Tanaka, A., Skordalakes, E., Showe, L. C. and Noma, K. I. (2015) Interaction between TBP and Condensin Drives the Organization and Faithful Segregation of Mitotic Chromosomes, *Molecular Cell*, 59, 755–767.

Johnson, R. E., Klassen, R., Prakash, L. and Prakash, S. (2015) A Major Role of DNA Polymerase δ in Replication of Both the Leading and Lagging DNA Strands, *Molecular Cell*, 59, 163–175.

Jonić, S., Sorzano, C. O. S. and Boisset, N. (2008) Comparison of single-particle analysis and electron tomography approaches: An overview, *Journal of Microscopy*, 232, 562–579.

Joyeux, L. and Penczek, P. A. (2002) Efficiency of 2D alignment methods, *Ultramicroscopy*, 92, 33–46.

Kagey, M. H. et al. (2010) Mediator and cohesin connect gene expression and chromatin architecture, *Nature*, 467, 430–435.

Kanno, T., Berta, D. G. and Sjögren, C. (2015) The Smc5/6 Complex Is an ATP-Dependent Intermolecular DNA Linker, *Cell Reports*, 12, 1471–1482.

Käshammer, L., Saathoff, J.-H., Lammens, K., Gut, F., Bartho, J., Alt, A., Kessler, B. and Hopfner, K.-P. (2019) Mechanism of DNA End Sensing and Processing by the Mre11-Rad50 Complex., *Molecular Cell*, 76, 1–13.

Kastner, B. et al. (2008) GraFix: Sample preparation for single-particle electron cryomicroscopy, *Nature Methods*, 5, 53–55.

Kikuchi, S., Borek, D. M., Otwinowski, Z., Tomchick, D. R. and Yu, H. (2016) Crystal structure of the cohesin loader Scc2 and insight into cohesinopathy, *Proceedings of the National Academy of Sciences of the United States of America*, 113, 12444–12449.

Kitajima, T. S., Sakuno, T., Ishiguro, K. I., Iemura, S. I., Natsume, T., Kawashima,

S. A. and Watanabe, Y. (2006) Shugoshin collaborates with protein phosphatase 2A to protect cohesin, *Nature*, 441, 46–52.

Klinge, S., Ñez-Ramírez, R., Llorca, O. and Pellegrini, L. (2009) 3D architecture of DNA Pol α reveals the functional core of multi-subunit replicative polymerases, *EMBO Journal*, 28, 1978–1987.

Kojic, A., Cuadrado, A., De Koninck, M., Giménez-Llorente, D., Rodríguez-Corsino, M., Gómez-López, G., Le Dily, F., Marti-Renom, M. A. and Losada, A. (2018) Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization, *Nature Structural & Molecular Biology*, 25, 496–504.

Kokic, G., Chernev, A., Tegunov, D., Dienemann, C., Urlaub, H. and Cramer, P. (2019) Structural basis of TFIIH activation for nucleotide excision repair, *Nature Communications*, 10, 1–9.

Komata, M., Bando, M., Araki, H. and Shirahige, K. (2009) The Direct Binding of Mrc1, a Checkpoint Mediator, to Mcm6, a Replication Helicase, Is Essential for the Replication Checkpoint against Methyl Methanesulfonate-Induced Stress, *Molecular and Cellular Biology*, 29, 5008–5019.

Kontziampasis, D. et al. (2019) A Cryo--EM Grid Preparation Device for Time--Resolved Structural Studies, *IUCrJ*, 6, 1–8.

Kschonsak, M. et al. (2017) Structural Basis for a Safety-Belt Mechanism That Anchors Condensin to Chromosomes Article Structural Basis for a Safety-Belt Mechanism That Anchors Condensin to Chromosomes, *Cell*, 171, 588-600.e24.

Ku, B., Lim, J. H., Shin, H. C., Shin, S. Y. and Oh, B. H. (2010) Crystal structure of the MukB hinge domain with coiled-coil stretches and its functional implications, *Proteins: Structure, Function and Bioinformatics*, 78, 1483–1490.

Kueng, S., Hegemann, B., Peters, B. H., Lipp, J. J., Schleiffer, A., Mechtler, K. and Peters, J. (2006) Wapl Controls the Dynamic Association of Cohesin with Chromatin, *Cell*, 955–967.

Kühlbrandt, W. (2014) The resolution revolution, *Science*, 343, 1443–1444.

Kuper, J., Braun, C., Elias, A., Michels, G., Sauer, F., Schmitt, D. R., Poterszman, A., Egly, J. M. and Kisker, C. (2014) In TFIIH, XPD Helicase Is Exclusively Devoted to DNA Repair, *PLoS Biology*, 12.

Kurze, A. et al. (2011) A positively charged channel within the Smc1 / Smc3 hinge required for sister chromatid cohesion, *The EMBO journal*, 30, 364–378.

Ladurner, R., Bhaskara, V., Huis In 'T Veld, P. J., Davidson, I. F., Kreidl, E., Petzold, G. and Peters, J. M. (2014) Cohesin's ATPase activity couples cohesin loading onto DNA with Smc3 acetylation, *Current Biology*, 24, 2228–2237.

Lafont, A. L., Song, J. and Rankin, S. (2010) Sororin cooperates with the acetyltransferase Eco2 to ensure DNA replication-dependent sister chromatid cohesion, *Proceedings of the National Academy of Sciences of the United States of America*, 107, 20364–20369.

Lee, B. G., Roig, M. B., Jansma, M., Petela, N., Metson, J., Nasmyth, K. and Löwe, J. (2016) Crystal Structure of the Cohesin Gatekeeper Pds5 and in Complex with Kleisin Scc1, *Cell Reports*, 14, 2108–2115.

Lehmann, A. R., Walicka, M., Griffiths, D. J., Murray, J. M., Watts, F. Z., McCready, S. and Carr, A. M. (1995) The rad18 gene of Schizosaccharomyces pombe defines a new subgroup of the SMC superfamily involved in DNA repair., *Molecular and Cellular Biology*, 15, 7067–7080.

Van Der Lelij, P., Chrzanowska, K. H., Godthelp, B. C., Rooimans, M. A., Oostra, A. B., Stumm, M., Zdzienicka, M. Z., Joenje, H. and de Winter, J. P. (2010) Warsaw Breakage Syndrome , a Cohesinopathy Associated with Mutations in the XPD Helicase Family Member DDX11 / ChlR1, *The American Journal of Human Genetics*, 86, 262–266.

Leman, A. R. and Noguchi, E. (2013) The Replication Fork: Understanding the Eukaryotic Replication Machinery and the Challenges to Genome Duplication, *Genes*, 4, 1–32.

Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G. P., Itoh, T., Watanabe, Y., Shirahige, K. and Uhlmann, F. (2004) Cohesin relocation from sites of chromosomal loading to places of convergent transcription., *Nature*, 430, 573–8.

Lerner, L. K. et al. (2020) Timeless couples G-quadruplex detection with processing by DDX 11 helicase during DNA replication , *The EMBO Journal*, 1–18.

Lerner, L. K. and Sale, J. E. (2019) Replication of G quadruplex DNA, *Genes*, 10, 1–25.

Levitus, M. et al. (2005) The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J, *Nature Genetics*, 37, 934–935.

Levran, O. et al. (2005) The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia, *Nature Genetics*, 37, 931–933.

Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., Agard, D. A. and Cheng, Y. (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM, *Nature Methods*, 10, 584–590.

Li, Y., Schoeffler, A. J., Berger, J. M. and Oakley, M. G. (2010) The Crystal Structure of the Hinge Domain of the Escherichia coli Structural Maintenance of Chromosomes Protein MukB, *Journal of Molecular Biology*, 395, 11–19.

Litman, R., Peng, M., Jin, Z., Zhang, F., Zhang, J., Powell, S., Andreassen, P. R. and Cantor, S. B. (2005) BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCJ, *Cancer Cell*, 8, 255–265.

Liu, H. et al. (2008) Structure of the DNA Repair Helicase XPD, *Cell*, 133, 801–812.

Liu, H., Rankin, S. and Yu, H. (2012) Phosphorylation-enabled binding of SGO1 – PP2A to cohesin protects sororin and centromeric cohesion during mitosis, *Nature Cell Biology*, 15, 40–49.

Liu, H. W., Bouchoux, C., Panarotto, M., Kakui, Y., Patel, H. and Uhlmann, F. (2020) Division of Labor between PCNA Loaders in DNA Replication and Sister

Chromatid Cohesion Establishment, *Molecular Cell*, 1–14.

Liu, J. et al. (2009) Transcriptional Dysregulation in NIPBL and Cohesin Mutant Human Cells, *PLOS Biology*, 7.

Lopez-serra, L., Lengronne, A., Borges, V., Kelly, G. and Uhlmann, F. (2013) Budding Yeast Wapl Controls Sister Chromatid Cohesion Maintenance and Chromosome Condensation, *Current Biology*, 23, 64–69.

Lopez-Serra, L., Kelly, G., Patel, H., Stewart, A. and Uhlmann, F. (2014) The Scc2 – Scc4 complex acts in sister chromatid cohesion and transcriptional regulation by maintaining nucleosome-free regions, *Nature Genetics*, 46, 1147–1151.

Losada, A., Hirano, M. and Hirano, T. (1998) Identification of Xenopus SMC protein complexes required for sister chromatid cohesion, *Genes and Development*, 12, 1986–1997.

Losada, A. and Hirano, T. (2001) Intermolecular DNA interactions stimulated by the cohesin complex in vitro: Implications for sister chromatid cohesion, *Current Biology*, 11, 268–272.

Lou, H., Komata, M., Katou, Y., Guan, Z., Reis, C. C., Budd, M., Shirahige, K. and Campbell, J. L. (2008) Mrc1 and DNA Polymerase ε Function Together in Linking DNA Replication and the S Phase Checkpoint, *Molecular Cell*, 32, 106–117.

Luna-Peláez, N. et al. (2019) The Cornelia de Lange Syndrome-associated factor NIPBL interacts with BRD4 ET domain for transcription control of a common set of genes, *Cell Death and Disease*, 10.

Lyumkis, D. (2019) Challenges and opportunities in cryo-EM single-particle analysis, *Journal of Biological Chemistry*, 294, 5181–5197.

Malumbres, M. and Barbacid, M. (2009) Cell cycle, CDKs and cancer: A changing paradigm, *Nature Reviews Cancer*, 9, 153–166.

Marcos-Alcalde, I. et al. (2017) Two-step ATP-driven opening of cohesin head, *Scientific Reports*, 7, 3266.

Mastronarde, D. N. (2005) Automated electron microscope tomography using robust prediction of specimen movements, *Journal of Structural Biology*, 152, 36–51.

McAleenan, A., Cordon-Preciado, V., Clemente-Blanco, A., Liu, I. C., Sen, N., Leonard, J., Jarmuz, A. and Aragón, L. (2012) SUMOylation of the α-kleisin subunit of cohesin is required for DNA damage-induced cohesion, *Current Biology*, 22, 1564–1575.

McIntyre, J., Muller, E. G. D., Snydsman, B. E., Trisha, N. and Uhlmann, F. (2007) In vivo analysis of cohesin architecture using FRET in the budding yeast Saccharomyces cerevisiae, *The EMBO Journal*, 26, 3783–3793.

McMullan, G., Faruqi, A. R., Clare, D. and Henderson, R. (2014) Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy, *Ultramicroscopy*, 147, 156–163.

McMullan, G., Faruqi, A. R. and Henderson, R. (2016) *Direct Electron Detectors*.

1st edn, *Methods in Enzymology*. 1st edn. Elsevier Inc.

Meisenberg, C., Pinder, S. I., Hopkins, S. R., Wooller, S. K., Benstead-Hume, G., Pearl, F. M. G., Jeggo, P. A. and Downs, J. A. (2019) Repression of Transcription at DNA Breaks Requires Cohesin throughout Interphase and Prevents Genome Instability, *Molecular Cell*, 73, 212-223.e7.

Melby, T. E., Ciampaglio, C. N., Briscoe, G. and Erickson, H. P. (1998) The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: Long, antiparallel coiled coils, folded at a flexible hinge, *Journal of Cell Biology*, 142, 1595–1604.

Menolfi, D., Delamarre, A., Lengronne, A., Pasero, P. and Branzei, D. (2015) Essential Roles of the Smc5/6 Complex in Replication through Natural Pausing Sites and Endogenous DNA Damage Tolerance, *Molecular Cell*, 60, 835–846.

Michaelis, C., Ciosk, R. and Nasmyth, K. (1997) Cohesins : Chromosomal Proteins that Prevent Premature Separation of Sister Chromatids, *Cell*, 91, 35–45.

Miller, T. C. R., Locke, J., Greiwe, J. F., Diffley, J. F. X. and Costa, A. (2019) Mechanism of head-to-head MCM double-hexamer formation revealed by cryo-EM., *Nature*, 575.

Moldovan, G., Pfander, B. and Jentsch, S. (2006) PCNA Controls Establishment of Sister Chromatid Cohesion during S Phase, *Molecular Cell*, 23, 723–732.

Muir, K. W., Kschonsak, M., Li, Y., Metz, J., Haering, C. H. and Panne, D. (2016) Structure of the Pds5-Scc1 Complex and Implications for Cohesin Function, *Cell Reports*, 14, 2116–2126.

Muir, K. W., Li, Y., Weis, F. and Panne, D. (2020) The structure of the cohesin ATPase elucidates the mechanism of SMC–kleisin ring opening, *Nature Structural and Molecular Biology*, 27, 233–239.

Muñoz, S., Minamino, M., Casas-Delucchi, C. S., Patel, H. and Uhlmann, F. (2019) A Role for Chromatin Remodeling in Cohesin Loading onto Chromosomes, *Molecular Cell*, 74, 664-673.e5.

Murayama, Y., Samora, C. P., Kurokawa, Y., Iwasaki, H. and Uhlmann, F. (2018) Establishment of DNA-DNA Interactions by the Cohesin Ring, *Cell*, 172, 1–13.

Murayama, Y. and Uhlmann, F. (2014) Biochemical reconstitution of topological DNA binding by the cohesin ring, *Nature*, 505, 367–371.

Murayama, Y. and Uhlmann, F. (2015) DNA Entry into and Exit out of the Cohesin Ring by an Interlocking Gate Mechanism, *Cell*, 163, 1628–1640.

Nasmyth, K. and Haering, C. H. (2005) The Structure and Function of Smc and Kleisin Complexes, *Annual Review of Biochemistry*, 74, 595–648.

Naydenova, K., Peet, M. J. and Russo, C. J. (2019) Multifunctional graphene supports for electron cryomicroscopy, *Proceedings of the National Academy of Sciences of the United States of America*, 116, 11718–11724.

Nishiyama, T. et al. (2010) Sororin mediates sister chromatid cohesion by antagonizing Wapl, *Cell*, 143, 737–749.

Núñez-Ramírez, R. et al. (2011) Flexible tethering of primase and DNA Pol α in the eukaryotic primosome, *Nucleic Acids Research*, 39, 8187–8199.

O'Keefe, M. A. et al. (2001) Sub Angstrom high-resolution transmission electron microscopy at 300 keV, *Ultramicroscopy*, 89, 215–241.

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

Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K. and Sugino, A. (1968) Mechanism of DNA chain growth, I. Possible discontinuity and unusual secondary structure of newly synthesized chains, *Proceedings of the National Academy of Sciences of the United States of America*, 59, 598–605.

Ono, T., Losada, A., Hirano, M., Myers, M. P., Neuwald, A. F. and Hirano, T. (2003) Differential contributions of condensin I and condensin II to mitotic chromosome architecture in vertebrate cells, *Cell*, 115, 109–121.

Orlova, E. V. and Saibil, H. R. (2011) Structural analysis of macromolecular assemblies by electron microscopy, *Chemical Reviews*, 111, 7710–7748.

Osborn, A. J. and Elledge, S. J. (2003) Mrc1 is a replication fork component whose phosphorylation in response to DNA replication stress activates Rad53, *Genes and Development*, 17, 1755–1767.

Ouyang, Z., Zheng, G., Song, J., Borek, D. M., Otwinowski, Z., Brautigam, C. A., Tomchick, D. R., Rankin, S. and Yu, H. (2013) Structure of the human cohesin inhibitor Wapl, *Proceedings of the National Academy of Sciences*, 110, 11355–11360.

Ouyang, Z., Zheng, G., Tomchick, D. R., Luo, X. and Yu, H. (2016) Structural Basis and IP6 Requirement for Pds5-Dependent Cohesin Dynamics, *Molecular Cell*, 62, 248–259.

Park, Y. B., Hohl, M., Padjasek, M., Jeong, E., Jin, K. S., Krężel, A., Petrini, J. H. J. and Cho, Y. (2017) Eukaryotic Rad50 functions as a rod-shaped dimer, *Nature Structural and Molecular Biology*, 24, 248–257.

Passmore, L. A. and Russo, C. J. (2016) *Specimen Preparation for High-Resolution Cryo-EM*. 1st edn, *Methods in Enzymology*. 1st edn. Elsevier Inc.

Pebernard, S., Wohlschlegel, J., McDonald, W. H., Yates, J. R. and Boddy, M. N. (2006) The Nse5-Nse6 Dimer Mediates DNA Repair Roles of the Smc5-Smc6 Complex, *Molecular and Cellular Biology*, 26, 3336–3336.

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

Peng, M., Litman, R., Xie, J., Sharma, S., Brosh, R. M. and Cantor, S. B. (2007) The FANCJ/MutLα interaction is required for correction of the cross-link response in FA-J cells, *EMBO Journal*, 26, 3238–3249.

Petela, N. J. et al. (2018) Scc2 Is a Potent Activator of Cohesin's ATPase that Promotes Loading by Binding Scc1 without Pds5, *Molecular Cell*, 70, 1134-1148.e7.

Petronczki, M., Chwalla, B., Siomos, M. F., Yokobayashi, S., Helmhart, W., Deutschbauer, A. M., Davis, R. W., Watanabe, Y. and Nasmyth, K. (2004) Sister-chromatid cohesion mediated by the alternative polymerase- α -associated protein Ctf4 is essential for chromatid disjunction during meiosis II, *Journal of Cell Science*, 16, 3547–3559.

De Piccoli, G. et al. (2006) Smc5-Smc6 mediate DNA double-strand-break repair by promoting sister-chromatid recombination, *Nature Cell Biology*, 8, 1032–1034.

Pisani, F. M., Napolitano, E., Napolitano, L. M. R. and Onesti, S. (2019) Molecular and Cellular Functions of the Warsaw Breakage Synfrome DNA Helicase DDX11, *Genes*, 9.

Poli, J., Tsaponina, O., Crabbe, L., Keszthelyi, A., Pantesco, V., Chabes, A., Lengronne, A. and Pasero, P. (2012) dNTP pools determine fork progression and origin, *The EMBO journal*, 31, 883–894.

Potts, P. R. and Yu, H. (2005) Human MMS21/NSE2 Is a SUMO Ligase Required for DNA Repair, *Molecular and Cellular Biology*, 25, 7021–7032.

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

Qiao, R. et al. (2016) Mechanism of APC/CCDC20 activation by mitotic phosphorylation, *Proceedings of the National Academy of Sciences of the United States of America*, 113, E2570–E2578.

Raghunandan, M., Chaudhury, I., Kelich, S. L., Hanenberg, H. and Sobeck, A. (2015) FANCD2, FANCJ and BRCA2 cooperate to promote replication fork recovery independently of the Fanconi Anemia core complex, *Cell Cycle*, 14, 342–353.

Ramlaul, K., Palmer, C. M., Nakane, T. and Aylett, C. H. S. (2020) Mitigating local over-fitting during single particle reconstruction with SIDESPLITTER, *Journal of Structural Biology*, 211, 107545.

Rankin, S., Ayad, N. G. and Kirschner, M. W. (2005) Sororin, a substrate of the anaphase- promoting complex, is required for sister chromatid cohesion in vertebrates, *Molecular Cell*, 18, 185–200.

Rao, S. S. P. et al. (2017) Cohesin Loss Eliminates All Loop Domains, *Cell*, 171, 305-320.e24.

Razinkov, I., Dandey, V. P., Wei, H., Zhang, Z., Melnekoff, D., Rice, W. J., Wigge, C., Potter, C. S. and Carragher, B. (2016) A new method for vitrifying samples for cryoEM, *Journal of Structural Biology*, 195, 190–198.

Remus, D., Beuron, F., Tolun, G., Griffith, J. D., Morris, E. P. and Diffley, J. F. X. (2009) Concerted Loading of Mcm2-7 Double Hexamers around DNA during DNA Replication Origin Licensing, *Cell*, 139, 719–730.

Rhodes, J. D. P., Haarhuis, J. H. I., Grimm, J. B., Rowland, B. D., Lavis, L. D. and Nasmyth, K. A. (2017) Cohesin Can Remain Associated with Chromosomes during DNA Replication, *Cell Reports*, 20, 2749–2755.

Ripstein, Z. A. and Rubinstein, J. L. (2016) *Processing of Cryo-EM Movie Data*.

1st edn, *Methods in Enzymology*. 1st edn. Elsevier Inc.

Rohou, A. and Grigorieff, N. (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs, *Journal of Structural Biology*, 192, 216–221.

Roig, M. B., Löwe, J., Chan, K. L., Beckouët, F., Metson, J. and Nasmyth, K. (2014) Structure and function of cohesin's Scc3/SA regulatory subunit, *FEBS Letters*, 588, 3692–3702.

Rojowska, A., Lammens, K., Seifert, F. U., Direnberger, C. and Feldmann, H. (2014) Structure of the Rad 50 DNA double-strand break repair protein in complex with DNA, 33, 2847–2859.

Rosenthal, P. B. and Henderson, R. (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy, *Journal of Molecular Biology*, 333, 721–745.

Rossi, F. et al. (2020) SMC 5/6 acts jointly with Fanconi anemia factors to support DNA repair and genome stability , *EMBO reports*, 21, 1–19.

Rossi, M. L. and Bambara, R. A. (2006) Reconstituted Okazaki fragment processing indicates two pathways of primer removal, *Journal of Biological Chemistry*, 281, 26051–26061.

Rowland, B. D. et al. (2009) Building Sister Chromatid Cohesion : Smc3 Acetylation Counteracts an Antiestablishment Activity, *Molecular Cell*, 33, 763–774.

Rubinstein, J. L., Guo, H., Ripstein, Z. A., Haydaroglu, A., Au, A., Yip, C. M., Trani, J. M. D., Benlekbira, S. and Kwoka, T. (2019) Shake-it-off: A simple ultrasonic cryo-EM specimen-preparation device Rubinstein John L., *Acta Crystallographica Section D: Structural Biology*, 75, 1063–1070.

Rudolf, J., Makrantoni, V., Ingledew, W. J., Stark, M. J. R. and White, M. F. (2006) The DNA Repair Helicases XPD and FancJ Have Essential Iron-Sulfur Domains, *Molecular Cell*, 23, 801–808.

Russo, C. J. and Passmore, L. A. (2014) Ultrastable gold substrates for electron cryomicroscopy, *Science*, 346, 1377–1381.

Saitoh, N., Goldberg, I. G., Wood, E. R. and Earnshaw, W. C. (1994) ScII: An abundant chromosome scaffold protein is a member of a family of putative ATPases with an unusual predicted tertiary structure, *Journal of Cell Biology*, 127, 303–318.

Samora, C. P., Saksouk, J., Goswami, P., Wade, B. O., Singleton, M. R., Bates, P. A., Lengronne, A., Costa, A. and Uhlmann, F. (2016) Ctf4 Links DNA Replication with Sister Chromatid Cohesion Establishment by Recruiting the Chl1 Helicase to the Replisome, *Molecular Cell*, 63, 371–384.

Scarff, C. A., Fuller, M. J. G., Thompson, R. F. and Iadaza, M. G. (2018) Variations on negative stain electron microscopy methods: Tools for tackling challenging systems, *Journal of Visualized Experiments*, 2018, 1–8.

Schaffer, M. et al. (2019) A cryo-FIB lift-out technique enables molecular-resolution cryo-ET within native Caenorhabditis elegans tissue, *Nature Methods*,

16, 757–762.

Schärer, O. D. (2013) Nucleotide excision repair in Eukaryotes, *Cold Spring Harbor Perspectives in Biology*, 5, 1–19.

Scheffzek, K., Ahmadian, M. R., Kabsch, W., Wiesmuller, L., Lautwein, A., Schmitz, F. and Wittinghofer, A. (1997) Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP- hydrolysis reaction of Ras, *Nature*, 17, 498–502.

Scheres, S. H. W., Gao, H., Valle, M., Herman, G. T., Eggermont, P. P. B., Frank, J. and Carazo, J. M. (2007) Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization, *Nature Methods*, 4, 27–29.

Scheres, S. H. W. (2012) RELION: Implementation of a Bayesian approach to cryo-EM structure determination, *Journal of Structural Biology*, 180, 519–530.

Scheres, S. H. W. and Chen, S. (2012) Prevention of overfitting in cryo-EM structure determination, *Nature Methods*, 9, 853–854.

van Schie, J. J. M. et al. (2020) Warsaw Breakage Syndrome associated DDX11 helicase resolves G-quadruplex structures to support sister chromatid cohesion, *Nature Communications*, 11.

Schilbach, S., Hantsche, M., Tegunov, D., Dienemann, C., Wigge, C., Urlaub, H. and Cramer, P. (2017) Structures of transcription pre-initiation complex with TFIIH and Mediator, *Nature*, 551, 204–209.

Schwarzer, W. et al. (2017) Two independent modes of chromatin organization revealed by cohesin removal, *Nature*, 551, 51–56.

Sengupta, S., Van Deursen, F., De Piccoli, G. and Labib, K. (2013) Dpb2 Integrates the Leading-Strand DNA Polymerase into the Eukaryotic Replisome, *Current Biology*, 23, 543–552.

Shah, N., Inoue, A., Woo, S., Beishline, K., Lahti, J. M. and Noguchi, E. (2013) Roles of ChlR1 DNA helicase in replication recovery from DNA damage, *Experimental Cell Research*, 319, 2244–2253.

Shi, Z, Shi, Zhubing, Gao, H., Bai, X. and Yu, H. (2020) Cryo-EM structure of the human cohesin-NIPBL-DNA complex, *Science*, 0981, 1–13.

Shintomi, K. and Hirano, T. (2009) Releasing cohesin from chromosome arms in early mitosis: Opposing actions of Wapl-Pds5 and Sgo1, *Genes and Development*, 23, 2224–2236.

Simon, A. C. et al. (2014) A Ctf4 trimer couples the CMG helicase to DNA polymerase a in the eukaryotic replisome, *Nature*, 510, 293–297.

Singleton, M. R., Dillingham, M. S. and Wigley, D. B. (2007) Structure and Mechanism of Helicases and Nucleic Acid Translocases.

Skibbens, R. V (2004) Chl1p, a DNA Helicase-like Protein in Budding Yeast, Functions in Sister-Chromatid Cohesion, *Genetics*, 166, 33–42.

Soh, Y. et al. (2015) Molecular Basis for SMC Rod Formation and Its Dissolution upon DNA Binding, *Molecular Cell*, 57, 290–303.

Speck, C., Chen, Z., Li, H. and Stillman, B. (2005) ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA, *Nature Structural and Molecular Biology*, 12, 965–971.

Stark, H. (2010) *GraFix : Stabilization of Fragile Macromolecular Complexes for Single Particle Cryo-EM*, *Methods in Enzymology*. Elsevier Masson SAS.

Stigler, J., Gamze, O., Koshland, D. E. and Greene, E. C. (2016) Single-Molecule Imaging Reveals a Collapsed Article Single-Molecule Imaging Reveals a Collapsed Conformational State for DNA-Bound Cohesin, *Cell Reports*, 15, 988–998.

Stodola, J. L. and Burgers, P. M. (2016) Resolving individual steps of Okazaki-fragment maturation at a millisecond timescale, *Nature Structural and Molecular Biology*, 23, 402–408.

Ström, L., Lindroos, H. B., Shirahige, K. and Sjögren, C. (2004) Postreplicative recruitment of cohesin to double-strand breaks is required for DNA repair, *Molecular Cell*, 16, 1003–1015.

Strunnikov, A. V., Larionov, V. L. and Koshland, D. (1993) SMC1: An essential yeast gene encoding a putative head-rod-tail protein is required for nuclear division and defines a new ubiquitous protein family, *Journal of Cell Biology*, 123, 1635–1648.

Sumara, I., Vorlaufer, E., Gieffers, C., Peters, B. H. and Peters, J. (2000) Characterization of vertebrate cohesin complexes and their regulation in prophase, *Journal of Cell Biology*, 151, 749–761.

Sumara, I., Vorlaufer, E., Stukenberg, P. T., Kelm, O., Redemann, N., Nigg, E. A. and Peters, J. M. (2002) The dissociation of cohesin from chromosomes in prophase is regulated by polo-like kinase, *Molecular Cell*, 9, 515–525.

Sung, P., Bailly, V., Weber, C., Thompson, L. H., Prakash, L. and Prakash, S. (1993) Human xeroderma pigmentosum group D gene encodes a DNA helicase, *Nature*, 365, 852–855.

Sutani, T. et al. (2015) Condensin targets and reduces unwound DNA structures associated with transcription in mitotic chromosome condensation, *Nature Communications*, 6.

Takayama, Y., Kamimura, Y., Okawa, M., Muramatsu, S., Sugino, A. and Araki, H. (2003) GINS, a novel multiprotein complex required for chromosomal DNA replication in budding yeast, *Genes and Development*, 17, 1153–1165.

Tanaka, S. and Diffley, J. F. X. (2002) Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2-7 during G1 phase, *Nature Cell Biology*, 4, 198–207.

Tanaka, T., Fuchs, J., Loidl, J. and Nasmyth, K. (2000) Cohesin ensures bipolar attachment of microtubules to sister centromeres and resists their precocious separation, *Nature Cell Biology*, 2.

Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I. and Ludtke, S. J. (2007) EMAN2: An extensible image processing suite for electron microscopy, *Journal of Structural Biology*, 157, 38–46.

Tanner, N. K., Cordin, O., Banroques, J., Doère, M. and Linder, P. (2003) The Q motif: A newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis, *Molecular Cell*, 11, 127–138.

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

Tegunov, D. and Cramer, P. (2019) Real-time cryo-electron microscopy data preprocessing with Warp, *Nature Methods*, 16, 1146–1152.

Terakawa, T., Bisht, S., Eeftens, J. M., Dekker, C., Haering, C. H. and Greene, E. C. (2017) The condensin complex is a mechanochemical motor that translocates along DNA, *Science*, 358, 672–676.

Tercero, J. A. (2000) DNA synthesis at individual replication forks requires the essential initiation factor Cdc45p, *The EMBO Journal*, 19, 2082–2093.

Tittel-Elmer, M., Lengronne, A., Davidson, M. B., Bacal, J., François, P., Hohl, M., Petrini, J. H. J., Pasero, P. and Cobb, J. A. (2012) Cohesin Association to Replication Sites Depends on Rad50 and Promotes Fork Restart, *Molecular Cell*, 48, 98–108.

Toro-Nahuelpan, M., Zagoriy, I., Senger, F., Blanchoin, L., Théry, M. and Mahamid, J. (2020) Tailoring cryo-electron microscopy grids by photo-micropatterning for in-cell structural studies, *Nature Methods*, 17, 50–54.

Toth, A., Ciosk, R., Uhlmann, F., Galova, M., Schleiffer, A. and Nasmyth, K. (1999) Yeast Cohesin complex requires a conserved protein , Eco1p (Ctf7), to establish cohesion between sister chromatids during DNA replication Attila To, *Genes & Development*, 13, 320–333.

Uchański, T., Pardon, E. and Steyaert, J. (2020) Nanobodies to study protein conformational states, *Current Opinion in Structural Biology*, 60, 117–123.

Uhlmann, F., Wernic, D., Poupart, M. a, Koonin, E. V and Nasmyth, K. (2000) Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast., *Cell*, 103, 375–386.

Uhlmann, F. (2016) SMC complexes: from DNA to chromosomes, *Nature Reviews Molecular Cell Biology*, 17, 399–412.

Uhlmann, F., Lottspeich, F. and Nasmyth, K. (1999) Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1, *Nature*, 400, 37–42.

Uhlmann, F. and Nasmyth, K. (1998) Cohesion between sister chromatids must be established during DNA replication, *Current Biology*, 8, 1095–1102.

Unal, E., Heidinger-Pauli, J. M., Kim, W., Guacci, V., Onn, I., Gygi, S. P. and Koshland, D. E. (2008) A Molecular Determinant for the Establishment of Sister Chromatid Cohesion, *Science*, 321.

Vannier, J. B., Pavicic-Kaltenbrunner, V., Petalcorin, M. I. R., Ding, H. and Boulton, S. J. (2012) RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity, *Cell*, 149, 795–806.

Varejão, N., Ibars, E., Lascorz, J., Colomina, N., Torres-Rosell, J. and Reverter,

D. (2018) DNA activates the Nse2/Mms21 SUMO E3 ligase in the Smc5/6 complex, *The EMBO Journal*, 37, 1–16.

Vaur, S., Feytout, A., Vazquez, S. and Javerzat, J. P. (2012) Pds5 promotes cohesin acetylation and stable cohesin-chromosome interaction, *EMBO Reports*, 13, 645–652.

Vazquez Nunez, R., Ruiz Avila, L. B. and Gruber, S. (2019) Transient DNA Occupancy of the SMC Interarm Space in Prokaryotic Condensin, *Molecular Cell*, 75, 209-223.e6.

Villa, F. et al. (2016) Ctf4 Is a Hub in the Eukaryotic Replisome that Links Multiple CIP-Box Proteins to the CMG Helicase Article Ctf4 Is a Hub in the Eukaryotic Replisome that Links Multiple CIP-Box Proteins to the CMG Helicase, 385–396.

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

Waizenegger, I. C., Hauf, S., Meinke, A. and Peters, J. (2000) Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase., *Cell*, 103, 399–410.

Walker, J. E., Saraste, M., Runswick, M. J. and Gay, N. J. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold, *EMBO Journal*, 1, 945–951.

Wold, M. S. and Kelly, T. (1988) Purification and characterization of replication protein A, a cellular protein required for in vitro replication of simian virus 40 DNA., *Proceedings of the National Academy of Sciences of the United States of America*, 85, 2523–2527.

Wu, S. et al. (2012) Fabs enable single particle cryoEM studies of small proteins, *Structure*, 20, 582–592.

Wu, Y., Sommers, J. A., Suhasini, A. N., Leonard, T., Deakyne, J. S., Mazin, A. V., Shin-ya, K., Kitao, H. and Brosh, R. M. (2010) Fanconi anemia group J mutation abolishes its DNA repair function by uncoupling DNA translocation from helicase activity or disruption of protein-DNA complexes, *Blood*, 116, 3780–3791.

Wu, Y., Sommers, J. A., Khan, I., De Winter, J. P. and Brosh, R. M. (2012) Biochemical characterization of Warsaw breakage syndrome helicase, *Journal of Biological Chemistry*, 287, 1007–1021.

Wu, Y., Shin-ya, K. and Brosh, R. M. (2008) FANCJ Helicase Defective in Fanconia Anemia and Breast Cancer Unwinds G-Quadruplex DNA To Defend Genomic Stability, *Molecular and Cellular Biology*, 28, 4116–4128.

Wu, Y., Suhasini, A. N. and Brosh, R. M. (2009) Welcome the Family of FANCJ-like Helicases to the Block of Genome Stability Maintenance Proteins, *Cellular and Molecular Life Sciences*, 66, 1209–1222.

Xu, H., Boone, C. and Brown, G. W. (2007) Genetic Dissection of Parallel Sister-Chromatid Cohesion Pathways, *Genetics*, 176, 1417–1429.

Ye, J., Osborne, A. R., Groll, M. and Rapoport, T. A. (2004) RecA-like motor ATPases - Lessons from structures, *Biochimica et Biophysica Acta* -

Bioenergetics, 1659, 1–18.

Yeeles, J. T. P., Deegan, T. D., Janska, A., Early, A. and Diffley, J. F. X. (2015) Regulated eukaryotic DNA replication origin firing with purified proteins, *Nature*, 519, 431–435.

Yeeles, J. T. P., Janska, A., Early, A. and Diffley, J. F. X. (2017) How the Eukaryotic Replisome Achieves Rapid and Efficient DNA Replication, *Molecular Cell*, 65, 105–116.

Yu, C., Gan, H., Han, J., Zhou, Z. X., Jia, S., Chabes, A., Farrugia, G., Ordog, T. and Zhang, Z. (2014) Strand-Specific Analysis Shows Protein Binding at Replication Forks and PCNA Unloading from Lagging Strands when Forks Stall, *Molecular Cell*, 56, 551–563.

Yuan, Z. et al. (2017) Structural basis of Mcm2-7 replicative helicase loading by ORC-Cdc6 and Cdt1, *Nature Structural and Molecular Biology*, 24, 316–324.

Yuan, Z., Georgescu, R., Santos, R. de L. A., Zhang, D., Bai, L., Yao, N. Y., Zhao, G., O'donnell, M. E. and Li, H. (2019) Ctf4 organizes sister replisomes and pol α into a replication factory, *eLife*, 8, 1–26.

Zabrady, K. et al. (2016) Chromatin association of the SMC5/6 complex is dependent on binding of its NSE3 subunit to DNA, *Nucleic Acids Research*, 44, 1064–1079.

Zegerman, P. and Diffley, J. F. X. (2010) Checkpoint-dependent inhibition of DNA replication initiation by Sld3 and Dbf4 phosphorylation, *Nature*, 467, 474–478.

Zhang, K. (2016) Gctf : Real-time CTF determination and correction, *Journal of Structural Biology*, 193, 1–12.

Zhao, X. and Blobel, G. (2005) A SUMO ligase is part of a nuclear multiprotein complex that affects DNA repair and chromosomal organization, *PNAS*, 102.

Zheng, S. Q., Palovcak, E., Armache, J.-P., Verba, K. A., Cheng, Y. and Agard, D. A. (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy, *Nature Methods*, 14, 331–332.

Zhou, Z. H., Hardt, S., Wang, B., Sherman, M. B., Jakana, J. and Chiu, W. (1996) CTF determination of images of ice-embedded single particles using a graphics interface, *Journal of Structural Biology*, 116, 216–222.

Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J. H., Lindahl, E. and Scheres, S. H. W. (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3, *eLife*, 7, 1–22.

Zou, L. (2013) Four pillars of the S-phase checkpoint, 227–233.

Zou, L. and Elledge, S. J. (2003) Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes, *Science*, 300, 1542–1548.