How does the loading complex interact with cohesin?



Jean Metson
Corpus Christi College

Department of Biochemistry
University of Oxford

Thesis submitted for the degree of
Doctor of Philosophy
Hilary Term 2021

Acknowledgements

I wish to thank:

Professor Kim Nasmyth for the opportunity of undertaking a D. Phil in his laboratory.

Professor David Sherratt for helpful discussions, the reading this thesis and for moral support when my spirits were low.

To everyone in the Nasmyth lab over the last sixteen years, for invaluable help and discussions, but mostly for trying to instil some confidence in me.

Special thanks to: Dr. Naomi Petela, for her endless patience whilst teaching me the ChIP sequencing protocol, the use of her 'spare pair of hands' during many experiments, but mostly for reminding me that I am capable of so much more.

Finally, to Dr. Martin Houlard, Miss Olivia Metson and Miss Zoe Metson, whose endless love and support have got me through this enormous challenge, and to whom I wish to dedicate this thesis.

Abstract

Cohesin is an SMC complex essential for genome organisation and the accurate segregation of sister chromatids in mitosis. The activity of this ATP dependent ring shape trimer composed of two SMC proteins and a kleisin subunit relies on two additional complexes. While the loading complex made of Scc2 and Scc4 is essential for the recruitment of cohesin at the loading sites, the releasing complex comprised of Pds5, Scc3 and Wapl is necessary to adjust the turnover of cohesin on DNA. During replication, Smc3 is acetylated and cohesion is established between the two newly replicated DNA molecules. Scc2 is essential for the ATPase activity of cohesin, its loading on chromatin and the genome reorganisation by DNA loop extrusion.

In order to better understand the biological role of Scc2, the interaction between Scc2 and the different subunits of the cohesin complex was investigated using genetic mutations combined with *in vivo* and i*n vitro* protein crosslinking. Scc2 makes extensive contacts with at least three regions of Scc1 whose deletion cause defects in loading and translocation of cohesin. Scc2 competes with Pds5 in order to stimulate the ATPase activity of the complex, but the mutation of the binding site in Scc1, common to Pds5 and Scc2, is not sufficient to abolish cohesin loading. By using different crosslinking strategies, accurate binding sites could be identified between Scc2 and Scc1. Scc2 was also found to interact with a conserved loop in the head of Smc1 that is essential for its engagement with Smc3's head and ATP hydrolysis. Finally, crosslinking experiments confirmed the existence *in vivo* of the folded configuration of cohesin that brings the hinge into contact with the coiled coils. This configuration was observed throughout the cell cycle, independently of cohesin acetylation, suggesting that Scc2 is not involved in its stabilization.

Contributions

The work in this thesis is my own, with the exception of the following:

Chapter two: The ATPase assays were conducted by Dr. Thomas Gligoris and Mr. Menelaos Voulgaris (Nasmyth lab, Oxford University) and the GST pull down experiments were undertaken by Dr. Satoro Kikuchi (Yu lab, University of Texas).

Chapter four: The project to find and test a new cysteine pair between the Smc1 hinge/Smc3 coiled coil was completed by Mr. Antonio Ferrier-Harding with my direction and supervision and included the G1 vs G2 experiment. The Cryo EM structure of the folded conformation of cohesin was used with permission from Mr. Andres Gonzalez-Llamazares (Löwe lab, LMB, Cambridge)

List of Abbreviations

°C	degrees Celsius
Å	Angström
5-FOA	5-Fluoroorotic acid
µL	microliters
µg	microgram
AID	Auxin inducible degron
APC/C	Anaphase Promoting Complex/Cyclosome
ATP	Adenosine 5’ Triphosphate
ADP	Adenosine 5’ Diphosphate
B. subtilis	*Bacillus subtilis*
bp	base pair
BPA	*para*-benzoyl-L-phenylalanine
C. glabrata	*Candida glabrata*
C. thermophilum	*Chaetomium thermophillum*
C- (terminal)	Carboxyl-
CDC (cdc)	Cell Division Cycle
CDE	Centromere DNA Element
Cdh	Cdc20 homologue
CEN	Centromere
CDK	Cyclin dependent kinase
CdLS	Cornelia de Lange Syndrome
ChIP	Chromatin Immunoprecipitation
CTCF	CCCTC binding factor

ddH2O	Double distilled water
DDK	Dbf4-dependent kinase
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
E. coli	*Escherichia coli*
ECO/ESCO (Eco)	Establishment of cohesion
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol-bis(beta-aminoethylther)-N,N,N,N'-tetraacetic acid
FKBP12	FK506 binding protein 12
FRB	FKBP-rapamycin binding domain of mTOR
G1/G2 phase	Gap phase 1/Gap phase 2
G2/M	Gap phase 2/Mitosis
GAL	Galactose-inducible promoter
HA	Hemagglutinin
HEAT	Huntingtin, Elongation factor 3, Protein Phosphatase 2A, Tor1
HEPES	4-(2-hydroxylethyl)-1-piperazineethanesulfonic acid
HIS (his)	Histidine requiring
HRP	Horse Radish Peroxidase
IP	Immunoprecipitation
kb	kilobase
kDa	kilodaltons
LEU (leu)	Leucine requiring

M phase	Mitosis
M. musculus	*Mus Musculus*
MET	Methionine
MgCl2	Magnesium Chloride
mL	millilitres
mM	millimolar
Myc	Myelocytomatosis
N- (terminal)	Amino-
NaCl	Sodium Chloride
NBD	Nucleotide Binding Domain
NIPBL	Nipped-B- Like
nM	nanomolar
nm	nanometre
OD_{600}	Optical density at 600nm
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction
Pds	Precocious dissociation of sisters
PK	small epitope on the P/V proteins of the paramyxovirus simian virus 5
PMSF	Phenylmethylsulfonyl fluoride
PVDF	Polyvinylidene fluoride
RAD (Rad)	Radiation Sensitive
RBS	Roberts Syndrome
rpm	revolutions per minute

S phase	Synthesis
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Schizosaccharomyces pombe*
SAC	Spindle assembly checkpoint
SCC (Scc, *scc*)	Sister Chromatid Cohesion
Scp	Segregation and condensation protein
SDS	Sodium dodecyl sulfate
SMC (SMC, Smc, *smc*)	Structural Maintenance of Chromosomes
TCA	Trichoroacetic acid
TE	Tris EDTA
tetO	Tetracycline Operator
TetR	Tetracycline Repressor
TPR	Tetratricopeptide repeat
tRNA	Transfer ribonucleic acid
TRP (trp)	Tryptophan requiring
ts	temperature sensitive
URA (ura)	Uracil requiring
UV	Ultra Violet
v/v	volume/volume
w/v	mass/volume
WCE	Whole cell extract
YEP(D)	Yeast Extract Peptone (Dextrose)
2xTY	2 times tryptone yeast extract

Contents

Acknowledgements……2

Abstract……3

Contributions……4

List of Abbreviations……5

Chapter 1: Introduction

The cell ……11

The cell cycle and its regulation……11

The challenges of chromosome segregation……14

Chromosome condensation……14

Chromosome cohesion……15

The segregation of sister chromatids……16

Structural Maintenance of Chromosome complexes……18

The cohesin cycle and its regulatory subunits ……22

Topological entrapment of DNA……29

Cohesin loading: the entry gate……33

SMC chambers and DNA……35

The loop extrusion model……38

The topology of Chromatid cohesion: one ring or two?……42

Localisation of cohesin with chromatin……42

Cohesinopathies……45

Cohesin and Cancer……46

Thesis Objective……48

Chapter 2: Identification of Scc2 binding motifs in the N-terminal region of Scc1

Introduction……49

2.1 Deletions within the Pds5 binding region of Scc1 cause lethality in the absence of WT protein……52

2.2 Mutations in the Pds5 binding site are defective in pericentromeric accumulation……57

2.3 Wild type cohesin can modify the activity of mutant complexes……60
2.4 Scc2 creates multiple contacts across the N-terminal half of Scc1……66
2.5 Cohesin loading can be abolished by combining small deletions of Scc1……72
2.6 Residues within SCC1 crosslink to SCC2 using BPA……78
2.7 Scc1 E203C crosslinks to Scc2 A709C……81
Discussion……83
Chapter 3: Analysis of SCC2 interaction with the SMC1 head
Introduction……89
3.1 Mutation of multiple residues in the F-loop of SMC1 causes lethality……91
3.2 Mutations in the F-loop of SMC1 do not suppress the deletion of Eco……93
3.3 Mutations in the F-loop of SMC1 impact on cohesin-DNA interaction……96
3.4 The ATPase activity of cohesin is inhibited by mutations in the SMC1 F-loop……97
3.5 The triple mutation in SMC1 F-loop is defective in head engagement……101
3.6 Analysis of the slow growth phenotype of K1121D in SMC1 F-loop……101
3.7 Identification of an essential region in the C-terminus of Scc2……105
3.8 Identification of interacting residues between the Smc1 head and Scc2……108
Discussion……110
Chapter 4: Analysis of the folded conformation of cohesin
Introduction……113
4.1 Predicting interacting residues between the Smc3 coiled coils and the Smc1 hinge……117
4.2 The SMC1 hinge can crosslink to multiple residues in the SMC3 coiled coils……123
4.3 Cohesin in the folded state occurs throughout the cell cycle……125
4.4 The folded conformation is compatible with acetylation……128
Discussion……130
Chapter 5: Summary and conclusions……135
Chapter 6: Materials and methods……137
References……163
Appendix……176

Introduction

The cell

For millions of years, life on our planet has slowly evolved into the vast array of species we see today. From a seemingly simple, single cell organism to a complex multicellular animal, the fundamental need for growth, repair, and reproduction is paramount for the continuity of life. In order to accomplish this, a faithful cell division is required. Each 'mother' cell must accurately copy its genetic information and divide to create two 'daughter' cells. Whilst some mutations in the genome may be tolerated, they may even be advantageous, however, some changes can have more serious consequences. The loss of an essential gene will lead to lethality in a single cell organism. The loss or gain of a chromosome, better known as aneuploidy, may result in cell death, or more frequently developmental disorders or cancer in multicellular organisms. As a result, the mechanisms controlling cell division are tightly regulated (Brooker & Burkovich, 2014).

The cell cycle and its regulation

The eukaryotic cell cycle refers to a set of distinct phases (G1, S, G2 and M), through which a cell must pass in order to grow, replicate its genome and divide into two daughter cells (Figure 1-1). The timing of the cell cycle varies from one cell type to another, but the process is regulated by a series of conserved checkpoints, cyclins and cyclin dependant kinases (CDK). Specific Cyclin-CDK complexes are triggered at each phase to promote the continuation of the cycle. As cell size is an important factor

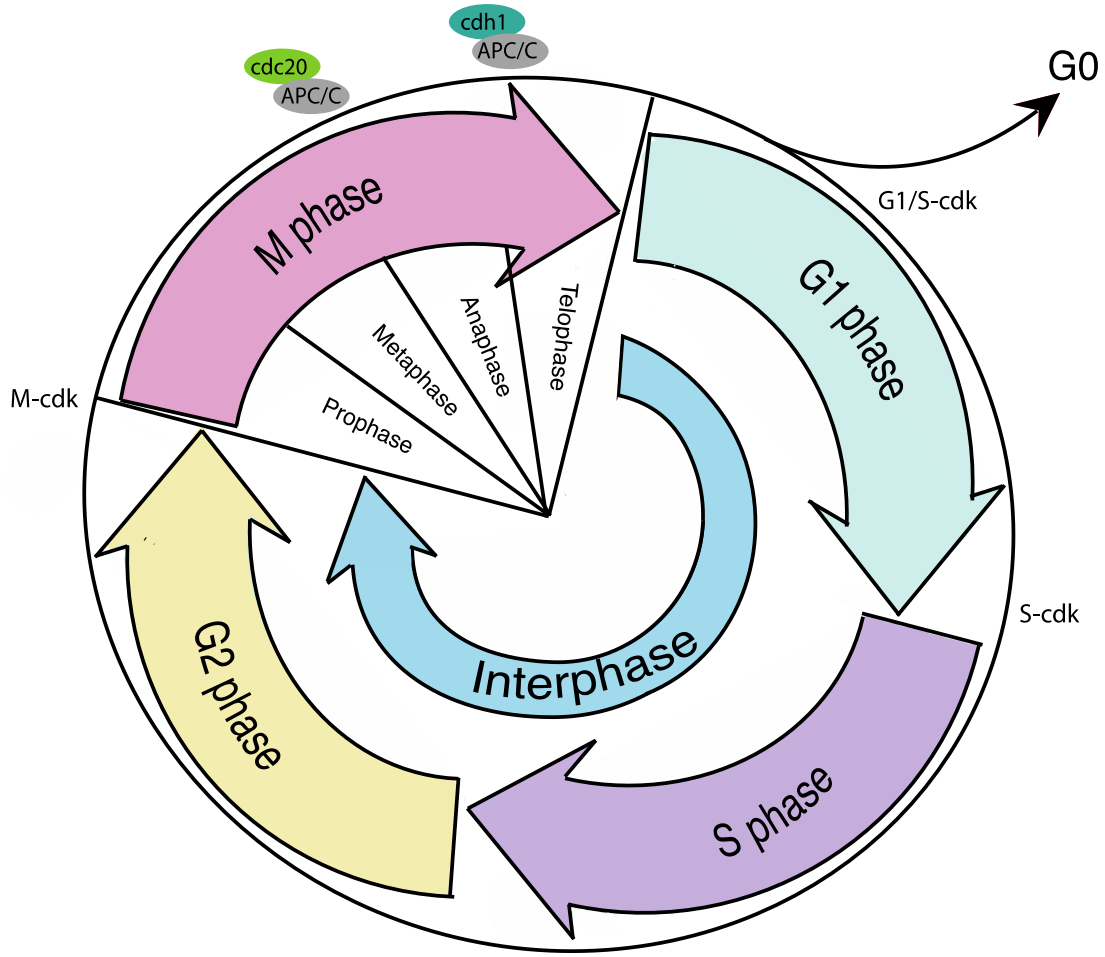


Figure 1-1. The cell cycle

The cell cycle is comprised of four phases: G1, S, G2 and M. CDK complexes trigger movement in a unidirectional manner. G1 allows the cell to mature and move on to S phase or to leave the cell cycle, to a quiescent stage known as G0. During S phase the genome will be precisely replicated. In G2 the cell continues to grow ready for division. M phase or mitosis, can be further split into four stages: prophase, metaphase, anaphase and telophase. The APC/C is regulated by cdc20 and cdh1, which trigger the separation of genetic material equally partitioned between the two daughter cells.

of the cell cycle, gap phase 1 (G1) allocates time for the cell to grow by producing enough cellular proteins and organelles to share between the future daughter cells. The environment and the intracellular content are continuously checked in order for the cell to proceed to the next stage only when favourable conditions are met. If these requirements are not satisfied, the cell can enter a quiescence phase known as gap 0 (G0). Once the cell has reached the prerequisite conditions to enter the cycle, the G1/S checkpoint is satisfied, and the G1/S-CDK is activated to enable the cell to proceed to the synthesis phase (S). This is, however, a one-way ticket: once committed the cell must complete the cycle. Once the S-CDK is activated, the precise duplication of the genome takes place prior to the cell entering gap phase 2 (G2), where replication is checked and completed as the cell continues to grow. Interphase, the collective name for the first three phases G1, S and G2, has now been completed. M-CDK is then activated to permit progression through the G2/M checkpoint. Mitotic phase (M) is the fourth phase and is further sub divided into four stages: Prophase, Metaphase, Anaphase and Telophase. Broadly, it begins with the condensation of DNA into chromosomes (Prophase). Then the nuclear envelope breaks down and the spindle forms from the centrioles present at opposite poles of the cell and connects to a single chromatid of each duplicated chromosome (Metaphase). Correct connection and tension allows the chromosomes to be aligned at the equator of the cell. Once complete, anaphase is triggered, allowing the sister DNAs to be separated, and pulled to opposing poles of the cell. Telophase and cytokinesis are the final stages, where the cell closes around each pool of genetic material, creating two daughter cells (Morgan, 2007).

The challenges of chromosome segregation

In order to segregate chromosomes, the cell faces two main challenges. The first, consists in compacting the chromatin fibres into small individualised chromosomes with appropriate physical properties, size, volume and rigidity. The second challenge entails the separation of the two sets of chromosomes equally in anaphase, the two sister chromatids of each pair must segregate, one into each daughter cell. Both issues are solved by members of the Structural Maintenance of Chromosome family, SMC: condensin and cohesin (Nasmyth, 2001).

Chromosome condensation

During interphase, DNA has a basic level of compaction mediated by a scaffold of proteins: the nucleosome, comprising histone proteins H2A, H2B, H3 and H4, arranged into an octamer around which the DNA is wrapped, and inserted at intervals along the genome. The nucleosome organisation creates an 11nm wide fibre (Morgan, 2007). Further levels of compaction are created by histone H1 which neutralises the DNA charges and non-histone proteins which interact with histone tails, protruding from the nucleosomes.

In mitosis, an estimated 10,000-fold compaction is needed to pack the eukaryotic genome into mitotic chromosomes that can be disentangled from one another, physically moved and separated during cell division (Morgan, 2007). This is mediated by the condensin complex, a member of the SMC family. Condensin is activated by M cyclin-CDK1 phosphorylation and in association with Topoisomerase II, is responsible for the compaction of chromosomes and the establishment of their

rigidity (Hirano, 2005). Loss of condensin function leads to insufficient rigidity, which results in the missegregation of the chromosomes as they are unable to resist the pulling forces of the spindle (Houlard *et al.*, 2015).

Chromosome cohesion

Another vital issue consists in the ability to identify and recognise the individual copies of each duplicated chromosome, in order to segregate one copy to each daughter cell. This problem is solved by another SMC complex, cohesin, which embraces the two sister chromatids together (Guacci *et al.*, 1997; Michaelis *et al.*, 1997). This allows the spindle to sense tension between the two copies as they are pulled in opposite directions, signalling to the cell that a pair of chromatids is ready to segregate (Pinksy *et al.,* 2005). The tension sensing mechanism involves a specific region of the chromosome: the centromere. Centromeric chromatin in yeast is a 125bp sequence consisting a single nucleosome that contains a specific variant of histone H3 called Cse4 (Furuyama & Biggins, 2007; Henikoff & Henikoff, 2012). It is at this centromeric region that a multiprotein complex called the Kinetochore is assembled, to which the microtubules are attached in mitosis (Furuyama & Biggins, 2007). In order to segregate accurately in anaphase, each sister chromatid kinetochore must be attached to microtubules emanating from opposite poles. The link created by cohesin between sister chromatids enables the cell to sense any lack of attachment or incorrect attachment of the microtubules to the kinetochores: either two microtubules to a single chromatid (monotelic), microtubules from both poles to a single chromatid (merotelic) or two microtubules from the same pole attaching to

both chromatids (syntelic). In these examples, no tension would be sensed by the microtubule pulling forces and the chromosome would not segregate accurately if not corrected (Figure 1-2) (Morgan, 2007).

The Spindle Assembly checkpoint (SAC) is responsible for sensing any lack of attachment or tension between sisters and as a consequence, delays the cell transition into anaphase. A single unattached kinetochore is enough to signal to the cell to induce the detachment of the microtubules that can then reattach in another attempt to achieve amphitelic attachment enabling the chromatids to bi-orient (Brooker, & Berkowitz, 2014). Cohesin is responsible for holding sister chromatids together and creating tension across the sister centromeres that can resist the pulling forces of the mitotic spindle, thereby signalling to the cell that the chromatids are properly attached and oriented (Kelly & Funabiki, 2009). Once the attachment and the correct bi-orientation of all chromosomes is reached, the spindle assembly checkpoint (SAC) is satisfied and allows the cell cycle to proceed through anaphase (Musacchio & Salmon, 2007).

The segregation of sister chromatids

The cell commitment into anaphase is regulated by a large multiprotein complex called the Anaphase Promoting Complex/Cyclosome: APC/C. This large ubiquitin conjugating complex (1.2MDa, 14 subunits) regulates the succession of steps necessary to go through the cell cycle by interacting with two coactivator proteins: CDC20 and CDH1 (Yamano *et al.,* 2019). During mitosis, when all the chromatids are properly attached and oriented, the spindle checkpoint is inactivated and APC/CDC20

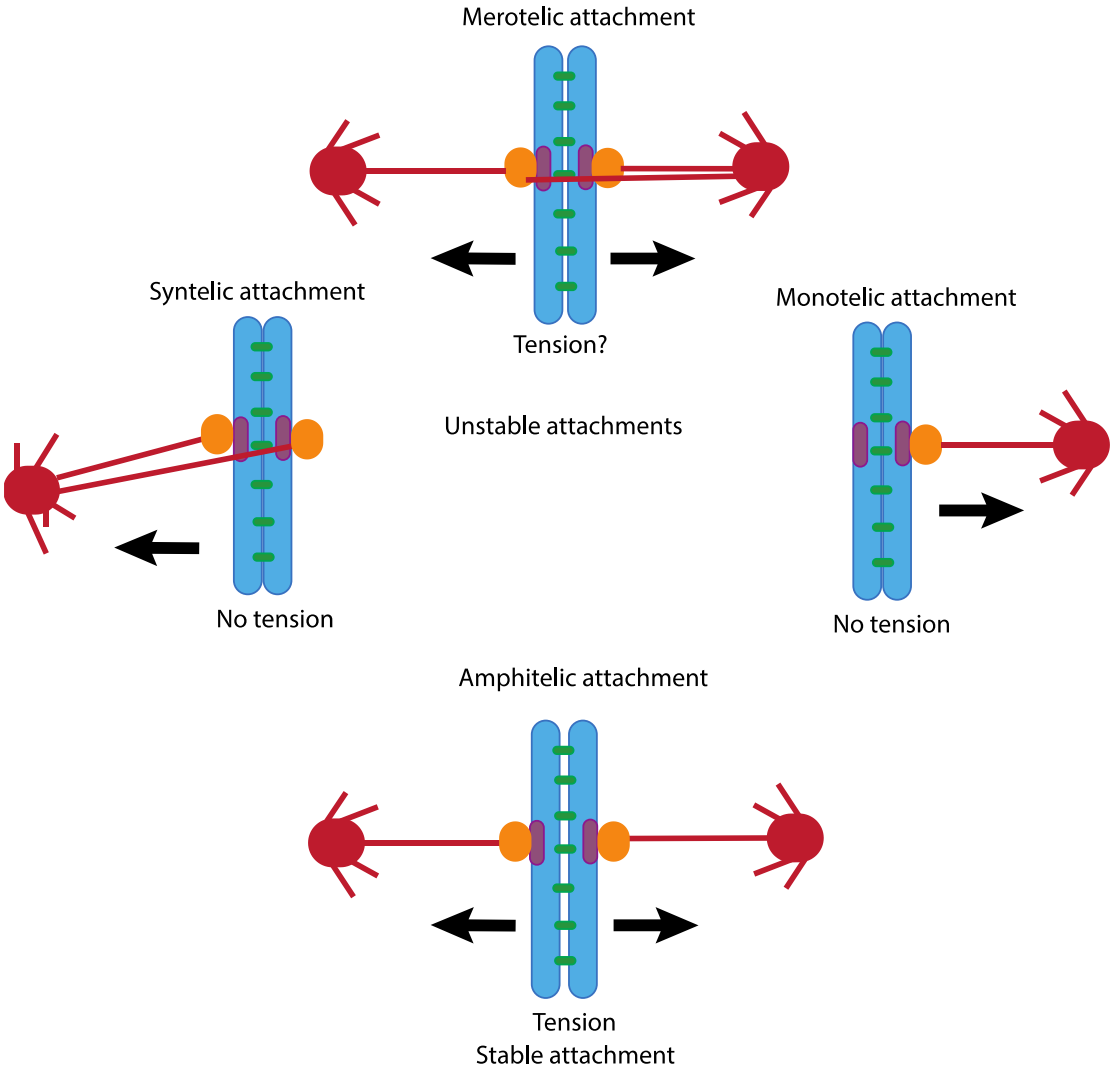


Figure 1-2. Microtubule-kinetochore attachment to chromatids
Incorrect and unstable attachments: monotelic, merotelic and syntelic. Correct, stable, amphitelic attachment creates tension between sisters due to cohesion, when pulling forces are exerted.

triggers the ubiquitination of both mitotic cyclin B and Securin, leading to their destruction (Musacchio & Salmon 2007). Securin associates and inhibits a protease called Cut2/Separase that is responsible for the proteolytic cleavage of the cohesin complex (Funabiki *et al.,* 1996; Ciosk *et al*., 1998). In metazoans, Separase is also inhibited by CDK1 phosphorylation (Musacchio & Salmon, 2007). Once mitotic Cyclin B and Securin are degraded, Separase rapidly cleaves cohesin releasing the two sister chromatids. This facilitates the abrupt separation of sisters, to opposite poles of the cell, by the microtubules.

Structural Maintenance of Chromosome complexes

The Structural Maintenance of Chromosome (SMC) family of proteins are highly conserved throughout evolution, from bacteria to humans (Cobbe & Hecke, 2002). SMC proteins contain 45nm long antiparallel coiled coil domain separating a dimerization domain, or hinge, at one end and an ATP binding cassette (ABC)-like ATPase head domain on the other end (Haering *et al.,* 2002). This SMC structure is in fact the result of the folding of the SMC protein on itself at the hinge region, that allows the two alpha helices domains to wrap around each other creating an intramolecular antiparallel coiled coil. This tight association of the coils brings the N- and C-terminal head domains into close contact, creating the ABC like ATP binding cassette (Figure 1-3).

The ATPase head domains contain specific functional motifs. The N-terminal Walker A motif (GXXXXGKS/T) contains a lysine residue that is crucial for nucleotide binding. It is associated with the C-terminal region containing the signature motif

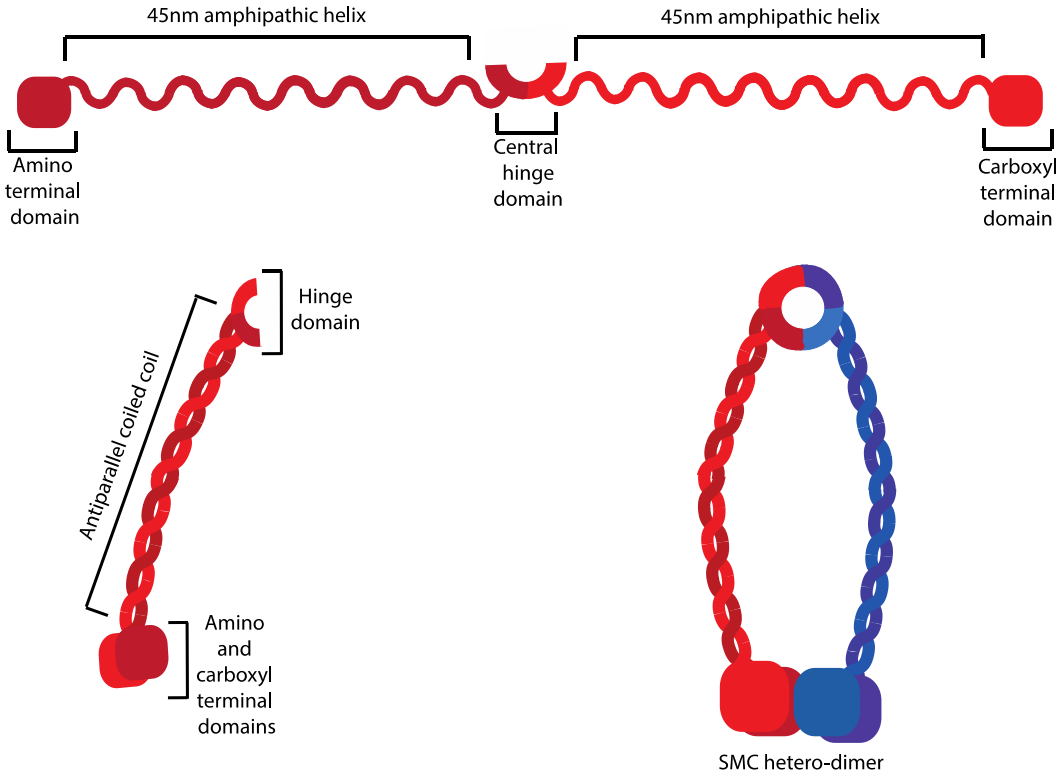


Figure 1-3. SMC protein structure

Top, linear representation of a single SMC protein with N and C terminal domains, the central hinge region and the 45nm long helices. Bottom left, a single SMC protein folded at the hinge region, with its antiparallel coiled coils and the interacting N and C-terminal head domains. Right, SMC hetero-dimer, two SMC subunits interacting through their hinge domains, allowing the heads to interact.

(LSGGQ/E) and the Walker B motif (hhhhDE, where h denotes a hydrophobic amino acid). In the Walker B motif, the aspartate residue coordinates a magnesium ion, which is required to stabilise nucleotide binding while the glutamate residue is essential for ATP hydrolysis (Gaudet & Wiley 2001; Verdon *et al.,* 2003). When the Walker B motifs of each SMC head are brought opposite each other and the signature motif of one head domain is aligned to the Walker A motif of the second head domain, the SMC dimer is capable of binding two molecules of ATP (Arumugam *et al.,* 2003) (Figure 1-4).

SMC proteins interact at their hinge regions to form a V-shaped hetero dimer: Smc1 and Smc3 for eukaryotic cohesin, Smc2 and Smc4 for condensin, and Smc5 with Smc6 (Figure 1-3). In Bacteria, the Smc-ScpAB complex is formed by a homodimer of SMC proteins (Haering & Gruber, 2016). These dimers of SMC proteins are then completed by a Kleisin subunit to form a topological ring structure: Scc1 for cohesin, Brn1/Ncaph1/2 for condensin, Nse4 for SMC5/6 complex (Haering & Gruber, 2016). The kleisin binds asymmetrically to the SMC heterodimer. For cohesin, the N-terminus of Scc1 binds just above the Smc3 ATPase head, to the base of the coiled coil, forming a four-helix bundle (Gligoris *et al*., 2014). The C-terminal of Scc1 forms a winged helix domain, which interacts with the base of the Smc1 ATPase head (Haering *et al.,* 2004). One bacterial SMC-like complex, MukBEF, is different as the kleisin, MukF, is dimeric, which could create higher order complexes of the complex, making dimers of dimers (Rajasekar et al, 2019). Specific variants of the cohesin subunits are involved during meiosis: Smc1 beta in mammalian cells and Rec8 in eukaryotes.

Initially, SMC complexes have been described as a ring. However, atomic force microscopy (AFM) and cryo-EM/negative stain electron microscopy (EM) of different

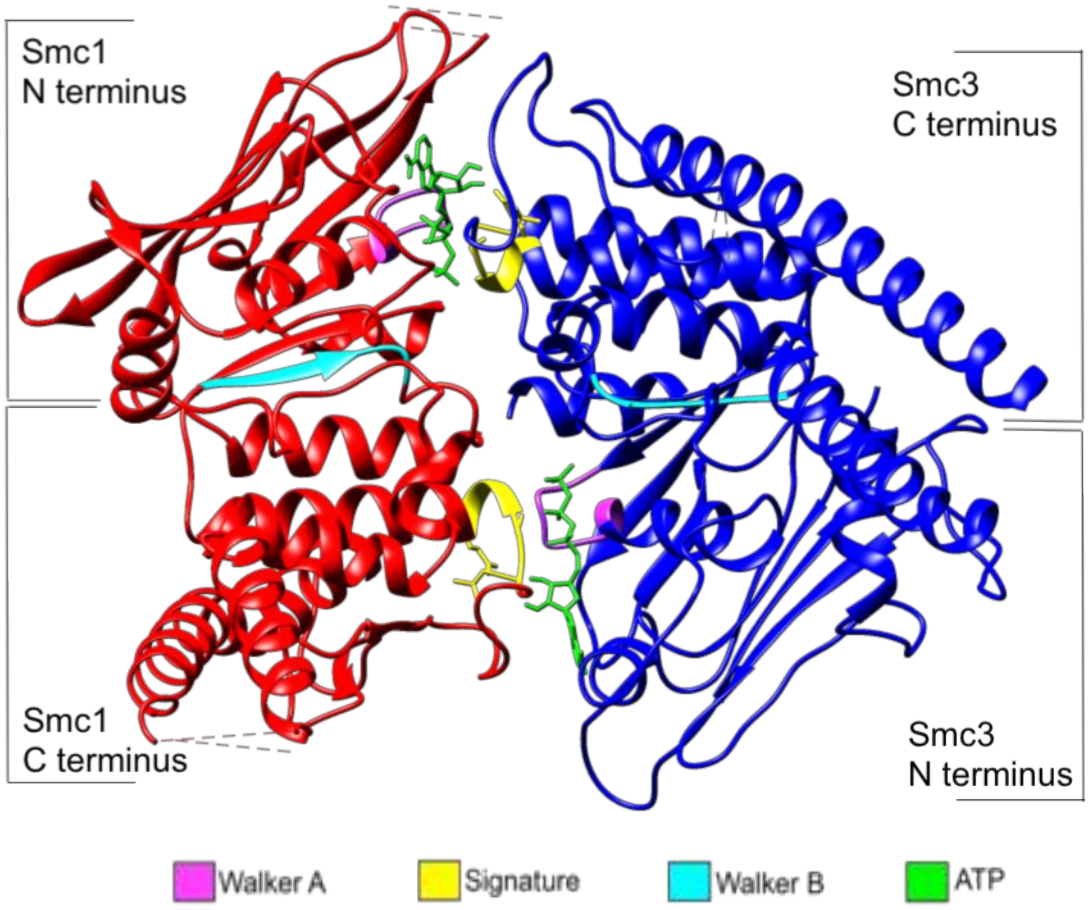


Figure 1-4. Structure of the cohesin head domain bound to two molecules of ATP
S. cerevisiae Smc1 and Smc3 head domains bound to two molecules of ATP (green). Smc1 (red) N-terminal Walker-A motif (magenta) opposite the Smc3 (blue) C terminal signature motif (yellow) and Smc3 (blue) N-terminal Walker-A motif (magenta) opposite Smc1 (red) C- terminal signature motif (yellow) and two Walker B motifs (cyan) opposite each other. PDB 6ZZ6.

SMC complexes, have isolated several different conformations; a ring, a straight rod and a folded rod (Melby *et al*., 1998, Anderson *et al*., 2002; Haering *et al*., 2002; Fuentes-Perez *et al*., 2012; Burmann *et al*, 2019; Tatebe *et al*., 2020). These conformations are possible due to specific sites of flexibility within the coiled coils. Crystal structure analysis has identified regions within the coiled coils, such as the joint and the elbow in bacterial SMC (Diebold-Durrant *et al.,* 2017), features that are common to all SMC homologues. These studies also infer that the SMC coiled coils can be in close proximity (Figure 1-5). Crosslinking and mass spectrometry experiments have identified a series of contacts along the coiled coil, reinforcing the idea that the coiled coils can be 'zipped up' along their length, from the hinge to the heads (Soh *et al.,* 2015; Burmann *et al.,* 2019).

The physiological relevance of these structures obtained from *in vitro* purified complexes remains to be determined *in vivo*. However, recently published structures of cohesin with DNA show the hinge near the joint region of the coils, in the folded formation (Higashi *et al.,* 2020; Shi *et al.,* 2020). When DNA is in contact with the heads, the region of the coiled coils between the hinge and elbow remains zipped up (Collier *et al.,* 2020) (Figure 1-5).

The cohesin cycle and its regulatory subunits

Formation of the cohesin complex is dependent on the levels of Scc1 protein which progressively accumulate from early to late G1. This is due to both transcriptional repression and separase activity, which prevent the production of the Scc1 protein before late G1 (Guacci *et al.,* 1997; Michaelis *et al.,* 1997).

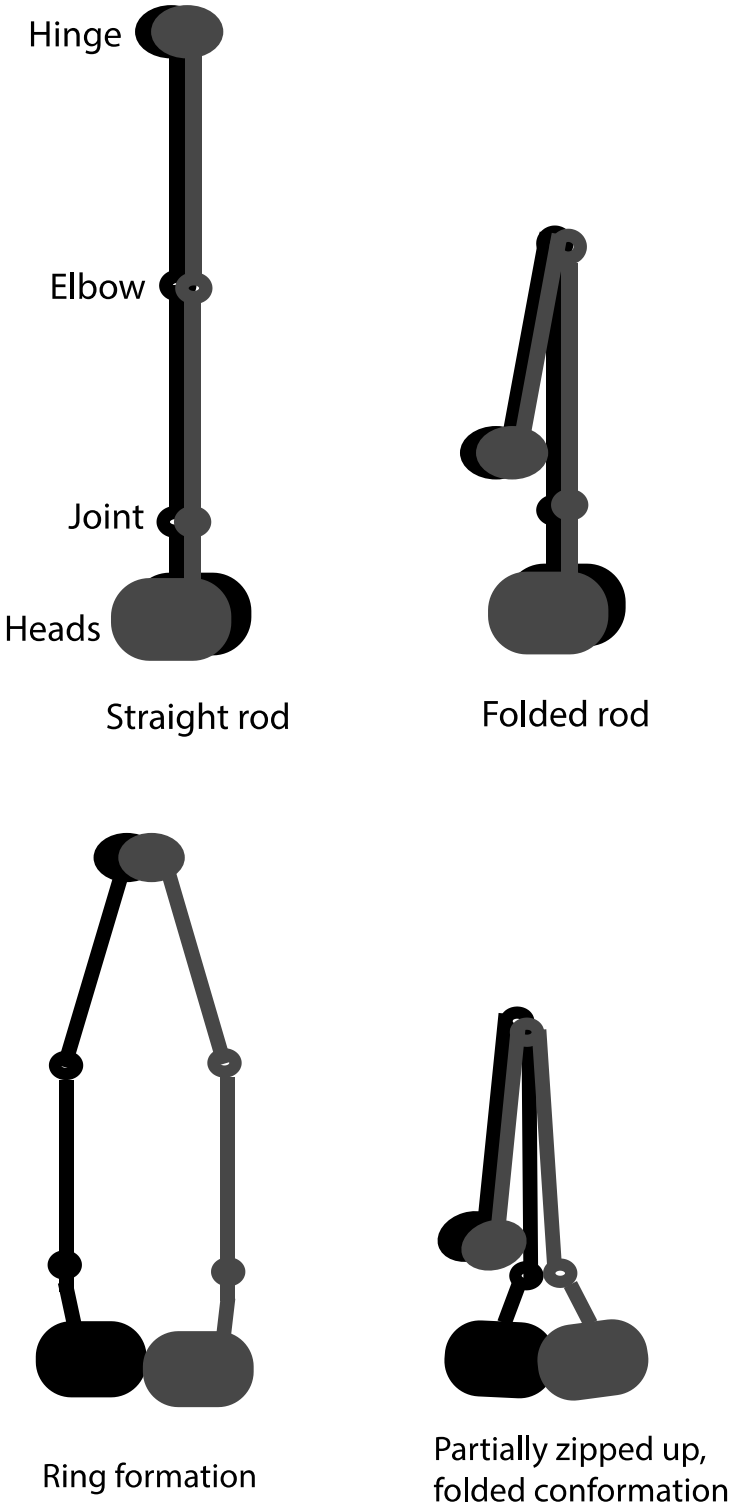


Figure 1-5. Models of SMC complex organisation
Illustrations of the different SMC conformations visualised by AFM/Cryo/negative stain EM (straight rod, folded rod and ring). The partially zipped up conformation or 'clamped/gripping state' was observed by cryo EM.

The activity of cohesin is regulated by three HEAT (Huntingtin, Elongation factor 3, protein phosphatase 2A, TOR1) repeat containing proteins (Andrade & Bork, 1995). These three proteins Scc3, Pds5 and Scc2 in yeast (and SA1/SA2, Pds5A/Pds5B and NipBL, in mammals), have all been shown to interact with Scc1 (Figure 1-6) (Toth *et al.,* 1999; Ciosk *et al.,* 2000; Hartman *et al.,* 2000). These HAWKs (Heat repeat proteins Associated With Kleisins) share a similar overall shape, a hook with a handle, in particular, the C-terminal hook of the HAWKs shows a high degree of similarity (Kikuchi *et al.,* 2016). The arrangement of helices permits a high degree of flexibility, which allows them to change their conformation. Structural comparisons between Pds5, Scc3 and Scc2 suggests that they share a common ancestor (Wells *et al.,* 2017).

During the G1 phase of the *S. cerevisiae* cell cycle, cohesin has a very dynamic interaction with chromatin (Figure 1-7) (Hu *et al.,* 2011). Cohesin cycles between loading events, governed by the loading complex (Scc2 and Scc4) and releasing events, controlled by the releasing complex consisting of Pds5, Scc3 and another heat repeat containing protein: Wapl (Rad61 in yeast) (Skibbens *et al*., 2009; Sutani *et al.,* 2009; Chan *et al.,* 2012). Wapl, initially isolated in Drosophila (Wings apart Like), binds to Scc3 and the N-terminal of Pds5 to form the releasing complex. Little is known about the structure and mode of action of Wapl and only a partial structure of the C-terminus has been resolved (Ouyang *et al.*, 2013). In mammalian cells WAPL is an essential gene, whilst in yeast it is dispensable. Wapl is essential for the releasing activity, that facilitates the opening of the interface between the Smc3 coiled coil and the Scc1 N-terminus called the exit gate (Chan *et al.,* 2012; Eichinger *et al*., 2013; Buheitel & Stemmann, 2013).

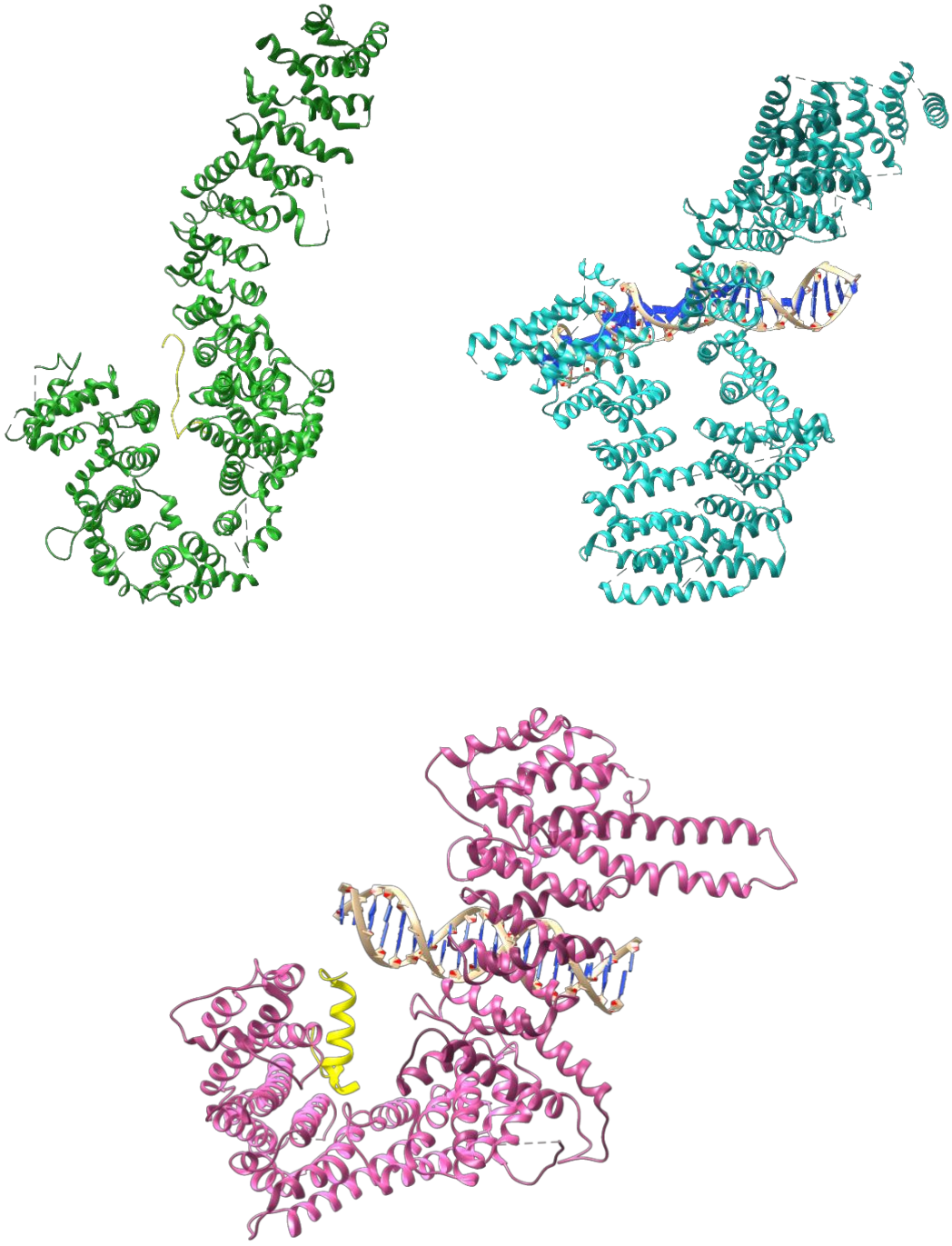


Figure 1-6. HEAT repeat proteins associated with Kleisins (HAWKs)
Left: Pds5 (green) with unstructured section of Scc1 (yellow) from *L. thermotolerans*, PDB:5F0O.
Right: Scc2 (light blue) with DNA from *S. cerevisiae*, PBD:6ZZ6.
Bottom: Scc3 (pink) with a helix of Scc1 (yellow) and DNA from *S. cerevisiae*, PDB:6H8Q.

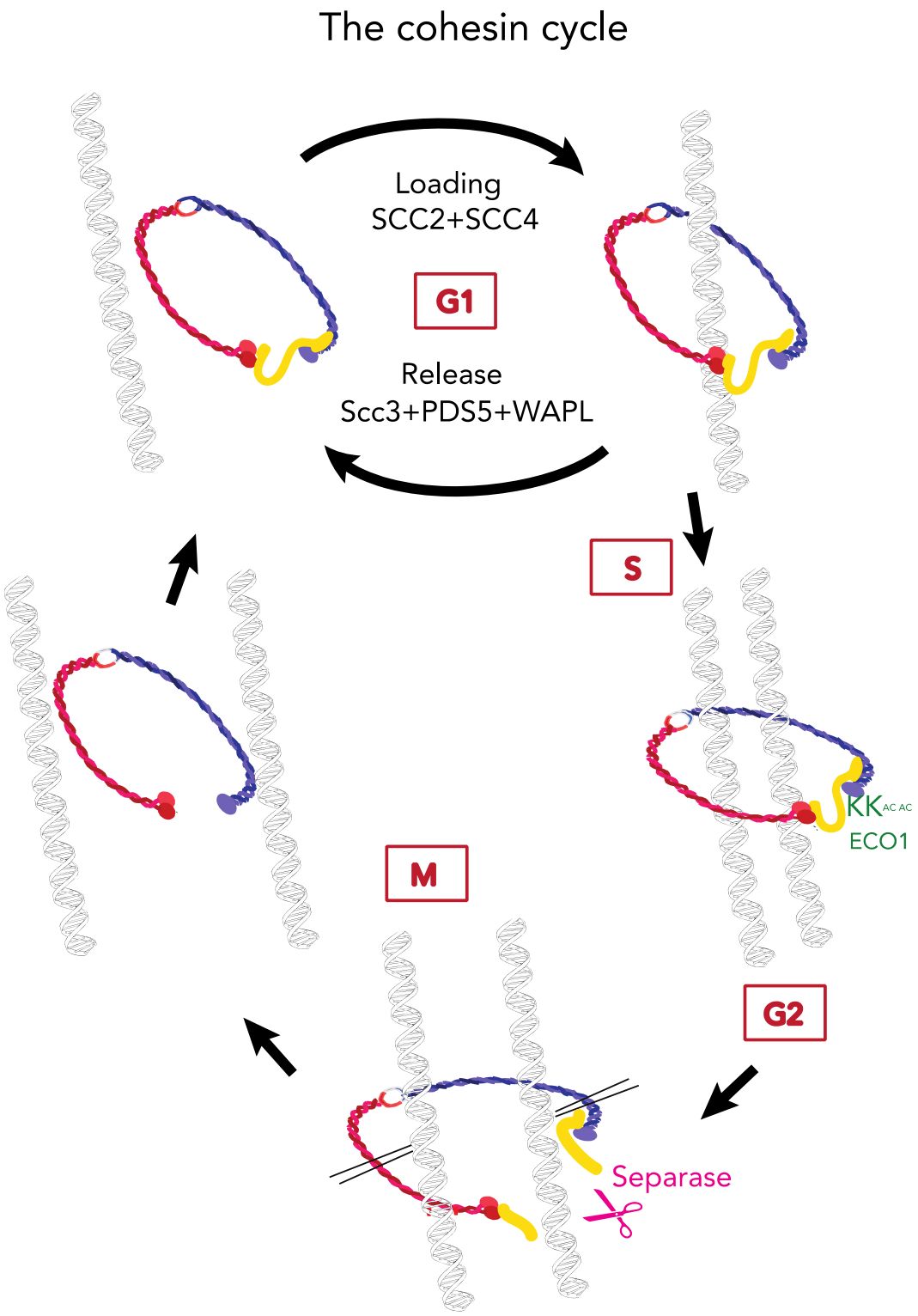


Figure 1-7. The cohesin cycle
The dynamic association between cohesin and DNA is regulated in G1 by the loading complex Scc2/Scc4 and releasing complex Pds5/Wapl/Scc3. After replication cohesin is stabilised on DNA by Eco1 acetylation establishing cohesion. During Mitosis, Separase cleaves Scc1 and sister chromatids are separated to opposite poles.

At 133kDa, Scc3 is the smallest of the HAWKs in *S. cerevisiae*. Scc3 has long been considered the forth constituent protein of the cohesin ring, as it seems to interact with Scc1 from G1 until anaphase (Haering *et al.*, 2002). Two sites were identified in Scc1 (residues 319-327 and 343-393) as binding to Scc3 (Roig *et al.,* 2014). This Scc1 binding region was confirmed by the recent crystal structure of *S. cerevisiae* Scc3, which showed that the inside of the C-terminal hook region (819-1054), binds to a small helix towards the C-terminus of Scc1 (360-385aa). This study also showed that Scc3 was able to bind to DNA in a central grove on its handle (Li *et al.*, 2018). Microscopy analysis has shown that cohesin is unable to associate with DNA in the absence of Scc3 (Hu *et al.*, 2011). Further evidence for its importance was confirmed using calibrated ChIP sequencing, showing the association of Scc3 as part of the pre-loading complex (Petela *et al.,* 2018). The role of Scc3 in the cohesin cycle seems to be more complicated than initially thought, as it is necessary for loading and interacting with DNA but also in the releasing activity.

Pds5 is 147 kDa in *S. cerevisiae* and the crystal structure from *L. thermotolerans* shows that it binds to the N terminal of Scc1 (125-140aa), inside its hook region (from 400aa) (Lee *et al.,* 2016). There is no evidence that Pds5 can interact with DNA directly. Pds5 appears to have a dual role in cohesin regulation as it is involved in both facilitating the release of the complex but also its stabilisation on DNA. The binding of Pds5 to Wapl is essential for the releasing activity and the regulation of cohesin turn over on the DNA.

During replication in yeast, cohesin complexes are acetylated by the acetyl transferase Eco1 at two conserved lysine residues (K112, K113) that are situated within the Smc3 N-terminal head domain. Acetylation of these residues creates a

stable association between cohesin and DNA, and establishes cohesion between the sister chromatids as cohesin becomes resistant to the activity of the releasing complex after acetylation (Rolef Ben-Shahar *et al.,* 2008; Unal *et al*, 2008; Zhang *et al.,* 2008a; Rowland *et al.,* 2009). Pds5 is essential for the maintenance of cohesion from S phase until anaphase by regulating cohesin acetylation of Smc3 by Eco1 and protecting the complex from deacetylation by Hos1 (Beckouet *et al.,* 2016, Minamino *et al*., 2015; Hartman *et al.*, 2000, Panizza *et al.,* 2000). Once all of the chromosomes are properly attached and oriented, Separase cleaves Scc1 allowing the two copies of the genome to segregate (Uhlmann *et al.,* 2000).

The deacetylation of Smc3 by Hos1 occurs directly after anaphase, and drives the removal of the N-terminal fragment of Scc1 from Smc3. This allows the recycling of Smc3 for the next cell cycle (Beckouet *et al.,* 2010; Borges *et al.,* 2010).

In mammalian cells, acetylation is mediated by two different acetyl transferase Esco1 and Esco2 (Hou & Zou, 2005). Smc3 acetylation takes place even in G1 and is not synonymous of cohesion (Minamino *et al.,* 2015). Cohesion involves another protein called Sororin that protects cohesin from Wapl mediated release by preventing the interaction between Wapl and Pds5 (Nishiyama *et al*., 2010). Moreover, a supplementary step occurs during prophase when the releasing complex removes most of the cohesin present on the arms of the chromosomes without any intervention of Separase. The cohesive fraction present at the centromere is protected by two proteins: Sgo1 and PP2A (Losada *et al.,* 1998, Gandhi *et al.,* 2006; Kueng *et al.*, 2006).

Scc2 is the largest of the HAWK subunits in *S. cerevisiae* at 171 kDa. A recent cryo-EM structure indicates that Scc2 binds DNA in a central groove, comparable to

Scc3 (Collier et al, 2020). Scc2 stimulates the ATPase activity and is essential for loading and translocation, but also for protecting the complex from release in G1 (Petela *et al*., 2018, Srinivasan *et al*., 2019). However, Scc2 is not required in G2 for yeast to maintain cohesion (Ciosk *et al.,* 2000).

The N-terminus of Scc2 binds to Scc4 (Chao *et al.*, 2015; Hinshaw *et al.*, 2015), together they form the loading complex that is required for cohesin's association with chromosomes. The crystal structure of Scc4 (70kDa) has been solved with the unstructured N-terminal of Scc2 (134aa) and a small section of the kinetochore protein Ctf19 (Hinshaw *et al.*, 2017). Scc4 is largely composed of tetratricopeptide (TPR) repeats which form a hollow tube that sheathes the N terminal of Scc2 (Hinshaw *et al.*, 2015). Scc4 is essential for the recruitment of the complex at the centromere through its interaction with Ctf19 (Hinshaw *et al.*, 2017). However, it has been shown *in vitro* that the cohesin loading activity of Scc2/4 only requires Scc2 (Murayama & Uhlmann, 2014). Scc4 is also dispensable for arm loading with *smc1D588Y*, a suppressor mutation of *scc4Δ*, found in the *SMC1* hinge (Dixon, 2013).

Topological entrapment of DNA

Several models have been suggested to explain how SMC proteins interact with DNA. These vary from a simple physical contact between cohesin and DNA, through to multiple interlinking rings entrapping chromatin (Haering *et al.*, 2002; Invanov *et al.*, 2005; Zhang *et al.*, 2008b). The simplest is the 'ring model' which suggests the entrapment of two molecules of DNA inside one SMC ring (Gruber *et al.*, 2003; Haering *et al.*, 2002). Given the internal diameter of a cohesin ring is predicted

to be 40nm, it seems plausible that two 11nm fibres would fit inside the fully open lumen of one trimeric ring (Haering *et al.*, 2002). However, other models involving the interaction between rings, each entrapping a single molecule of DNA have been suggested to explain cohesion (Figure 1-8).

Evidence supporting the topological entrapment of DNA comes from experiments demonstrating that the proteolytic cleavage of one of the cohesin trimer subunits, *in vitro* and *in vivo,* always triggers release of DNA (Uhlmann *et al.*, 1999; Uhlmann *et al.*, 2000, Oliviera *et al.*, 2010 and Tachibana-Konwalski *et al.*, 2010). Furthermore, the linearization of a circular mini chromosome, allows non-cleaved cohesin to slide off DNA (Ivanov & Nasmyth, 2005). These experiments suggest the DNA must be trapped within the cohesin ring, not merely interacting with it. Similar work on condensin also suggested a topological embrace model (Cuylen & Haering, 2011).

The crystal structures of the three cohesin trimer interfaces identified the residues involved in each interaction. This allowed the design of chemical crosslinking leading to the closure of the different interfaces. This technique relies on the mutagenesis of two amino acids into cysteine. The residues selected should face each other in the 3D structure of the complex and be within a suitable distance. In the presence of a bifunctional maleimide crosslinker that reacts with the thiol group, such as BMOE or dBBr, the two cysteines can be covalently linked to each other (Dhar *et al.*, 2004). The analysis of the crosslink efficiency in different contexts or using different mutated proteins can be used to measure protein-protein interaction *in vivo*. In the case of the cohesin complex, this technique allows the closure of the three interfaces of the trimer. This creates a closed ring, which is resistant to denaturation

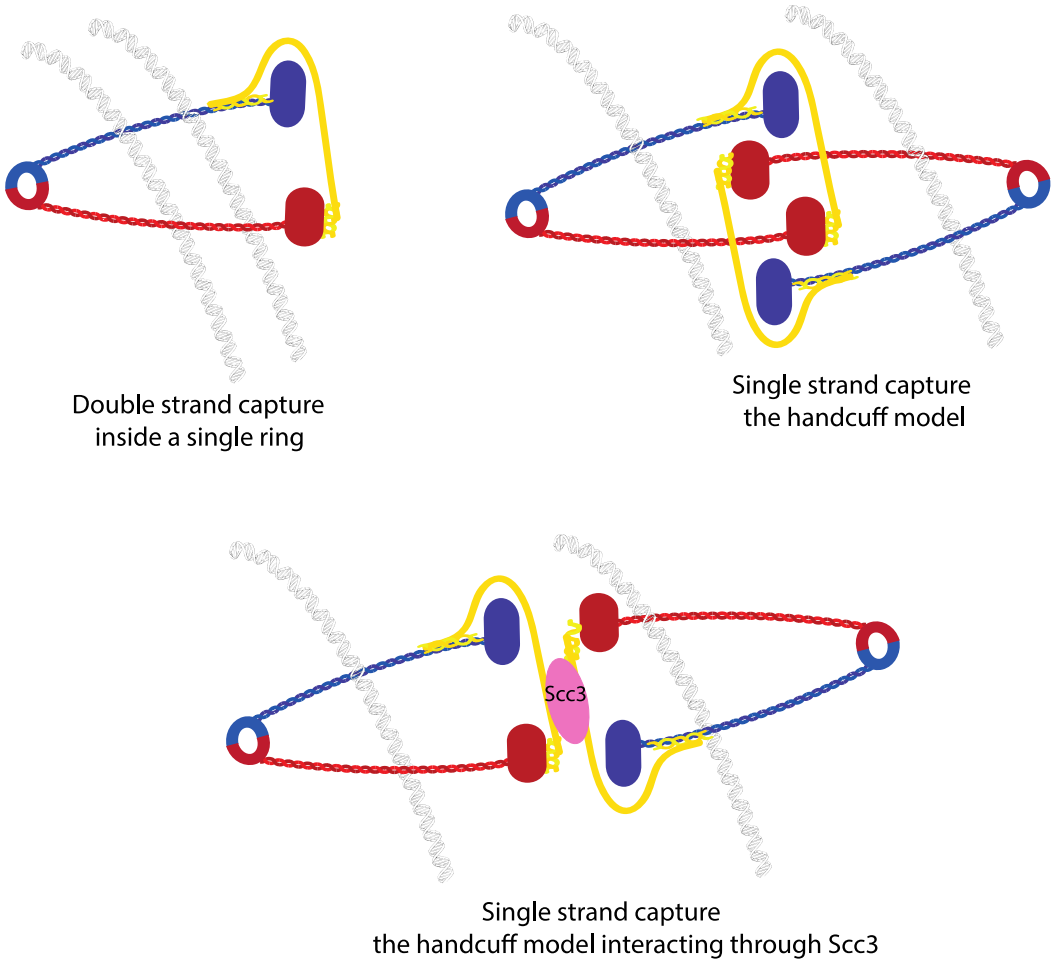


Figure 1-8. Illustrations of possible models for cohesion by topological entrapment
Illustrations of a single cohesin complex 'the ring model' for double strand entrapment (top left), a single DNA entrapped in two interlinked cohesin complexes 'the handcuff model' (top right) and the 'alternative handcuff model' with Scc3 connecting the two cohesin rings (bottom).

by SDS and enables the topological entrapment of circular plasmids.

In an initial experiment, a cross-linkable cysteine pair was inserted into the Smc1/Smc3 hinge interface and a second pair into the Scc1/Smc1 interface. Initially, as no structural information was available for the Scc1/Smc3 interface, the design of cysteine pairs was impossible and a viable fusion protein was used instead. Yeast cells expressing these modified subunits of cohesin and carrying a 2.3kb circular mini-chromosome, were arrested in G2. The complex could then be purified, and covalently closed by chemical crosslinking, denatured by SDS and the DNA visualised by southern blotting. The circular mini-chromosomes could be seen as dimeric bands, resistant to SDS treatment, suggesting that both DNA circles were topologically trapped inside the cohesin ring. A final control was completed by migrating these crosslinked concatenates in a 2D gel using proteinase K. This showed that these concatenated structures are due to a ring of cohesin that links two DNA circles and not merely knots between the two DNA molecules (Haering *et al*., 2008).

More recently, the structure of the third interface, namely Scc1/Smc3, was obtained allowing the design of a third pair of cysteines that enabled the crosslinking of all three interfaces *in vivo*. This approach which locks *in situ* the complex onto the DNA, a circular 2.3kb mini-chromosomes introduced in the cell, revealed that both concatenated monomers (single DNA) and concatenated dimers (replicated DNA) could be detected topologically entrapped inside the cohesin ring (Gligoris *et al*., 2014). Only concatenated monomers were obtained from G1 cells, whereas both monomers and dimers were obtained from G2 cells (Srinivasan et al., 2018). Since this assay was developed, several known mutations of the cohesin complex have been analysed. This showed an exact correlation between the ability of the cohesin complex

to load on the DNA and to form CMs, or their ability to mediate cohesion, with the formation of CDs (Srinivasan *et al.,* 2018). To date there is only one mutant for which DNA binding ability does not correlate with topological entrapment as CM or CDs: this mutant is neutralising the charges in the lumen of the hinge, the *smc1K554D, K661D, smc3R665A, K668A, R669A* mutations (Kurze *et al*., 2011; Srinivisan *et al*., 2019). These mutant complexes, which do not sustain life, still seem to associate with chromosomes, as seen by ChIP sequencing, but do not entrap circular mini-chromosomal DNA. This suggests a non-topological interaction of this mutant with DNA (Srinivasan *et al,* 2018).

Cohesin loading: the entry gate

The ring model demonstrates the ability of cohesin to entrap sister chromatids, but leads to the question of how it is able to do so. In order to mediate cohesion and topologically entrap DNA, as with the mini-chromosome assay, one of the three connecting interfaces of cohesin needs to open to allow DNA inside the ring. Alternatively, the ring could form around chromatin, *de novo* or the strand of DNA could be broken, moved into the ring and repaired. Given the importance of maintaining the integrity of the genome, it would be a risky strategy to induce such breakages, making the latter an unlikely scenario. Furthermore, evidence from electron micrographs suggests cohesin forms a ring before encountering DNA (Anderson *et al*., 2002), suggesting the transient opening of the ring is more likely. Experiments involving closing the three interfaces biochemically, also suggests complete cohesin rings exist without DNA interaction (Gruber *et al*., 2003).

Experiments in yeast using fusion proteins either Scc1-Smc1 or Scc1-Smc3 showed permanent closure of these interfaces did not lead to lethality, suggesting that the opening of these individual interfaces is not necessary for the entry of DNA. Due to the position of the hinge, in the centre of the protein, it could not be expressed as a fusion protein. To overcome the problem, two human proteins FRB and FKBP12 were introduced into the hinge domains of Smc1 and Smc3 respectively. FRB can only interact with FKBP12 when the latter has bound to rapamycin. These insertions were viable and revealed that hinge closure on addition of rapamycin in G1, before cohesion establishment, was lethal, whereas in G2, no impact was seen. This suggests that the hinge could be the DNA 'entry gate' (Gruber *et al.*, 2006). A similar experiment in human cells also found the hinge was required for loading from telophase into G1 (Buheitel & Stemmann, 2013). Even If the mechanism by which DNA could enter through the hinge is still unknown, the crystal structure of the Smc1/Smc3 hinge from both *T. maritima* and *H. sapiens*, gave a possible clue by describing a half open hinge suggesting a possible entry mechanism for DNA (Haering *et al*., 2002; Shi *et al*., 2020). More recently a mutation in the hinge *smc1D588Y*, was found to supress *scc4*Δ, emphasising the importance of the hinge for cohesin's association with DNA. However more direct evidence is needed to determine which of the cohesin's interfaces is the entry gate.

The transient 'on and off' association of cohesin with DNA, independent of Separase cleavage, also requires an exit gate. A growing body of evidence has suggested this to be the Smc3-Scc1 interface. The release of cohesin is abolished by the fusion of the Smc3 C-terminal to the N-terminus of Scc1 (Chan *et al.*, 2012). The removal of the N terminal fragment of Scc1, created by Separase cleavage from the

Smc3 coiled coil, is abolished by all mutations known to affect release in Wapl, Scc3 Pds5 or Smc1 (Beckouet *et al.,* 2016; Elbatsh *et al.,* 2016; Muryama *et al.,* 2015, Rowland *et al.,* 2009). Until recently, Wapl and the releasing complex have been considered as the main mechanism responsible for the releasing activity. However, recent studies in yeast have shown that in G1, even in the absence of the releasing complex, Scc2 is essential to prevent the cohesin complex from being released (Srinivasan *et al.,* 2019). This suggests that the different HAWKs are regulators of the exit gate opening.

SMC chambers and DNA

As previously mentioned, the lumen of the ring is thought to be large enough to hold two molecules of DNA, but recent experiments have suggested that the cohesin ring may in fact be subdivided into chambers. Using cysteine crosslinking, it has been possible to identify different conformations of the complex and propose where within the cohesin ring the DNA molecules could be held. In particular, how the changing position of the SMC heads throughout the ATPase cycle, correlates with the localisation of DNA with the ring. During the ATPase cycle, the heads have been described in three different configurations: open, engaged or juxtaposed (Arumugam *et al*., 2003; Chapard *et al*., 2019; Diebold *et al*., 2017). The 'E' or engaged state sandwiches two ATP molecules between the SMC heads, and can be measured by cysteine cross-linking *in vivo* (Chapard *et al*., 2019). An in-between state was first observed in *B. subtilis* Smc-Scp complexes *in vitro* (Diebold *et al*., 2017). This 'J' or juxtaposed state was then measured in both *S. cerevisiae and B. subtilis, in vivo* by a

cross-linkable cysteine pair (Chapard *et al*., 2019; Vazquez Nunez *et al*., 2019).

The Smc-Kleisin or S-K chamber, represents the fully open complex. The engaged head-Smc or E-S compartment is the lumen created between the coiled coils from the hinge to the engaged heads. The engaged heads -kleisin or E-K compartment refers to the loop created by the kleisin (Scc1), when the heads are engaged (Figure 1-9). Using cysteine pairs specific to the engaged state or the juxtaposed state combined with pairs in the other interfaces, it was possible to lock the interfaces of the two chambers and address where the DNA was held using the mini chromosome assay. It appeared that *in vivo*, DNA could only be detected in the chamber between the heads and the kleisin, as a monomer or as a dimer. These experiments also showed that the complexes that coentrap two minichromosomes between the juxtaposed heads and the kleisin are enriched in Smc3 acetylation, suggesting they had established cohesion (Chapard *et al*, 2019). A similar finding in *B. Subtilis* SMC suggests a unified mechanism between SMC complexes (Vazquez Nunez *et al*, 2019).

More recently the cysteine crosslinking approach was used in an *in vitro* entrapment assay using plasmid DNA. This assay that does not measure cohesion but only the topological entrapment of a plasmid inside the cohesin ring, could only detect the DNA in the Smc chamber (E-S). The advantage of this in vitro assay is to test the contribution of each subunit of the complex by adding them one at a time to the entrapment reaction. This lead to the characterization of a “transition state” in which the DNA slides into the complex between the heads, merely pushing the kleisin upwards (Collier *et al.,* 2020). In this transition state, that is dependent on Scc2 but not Scc3, the DNA is detected in the ES and EK compartment but not in the SK

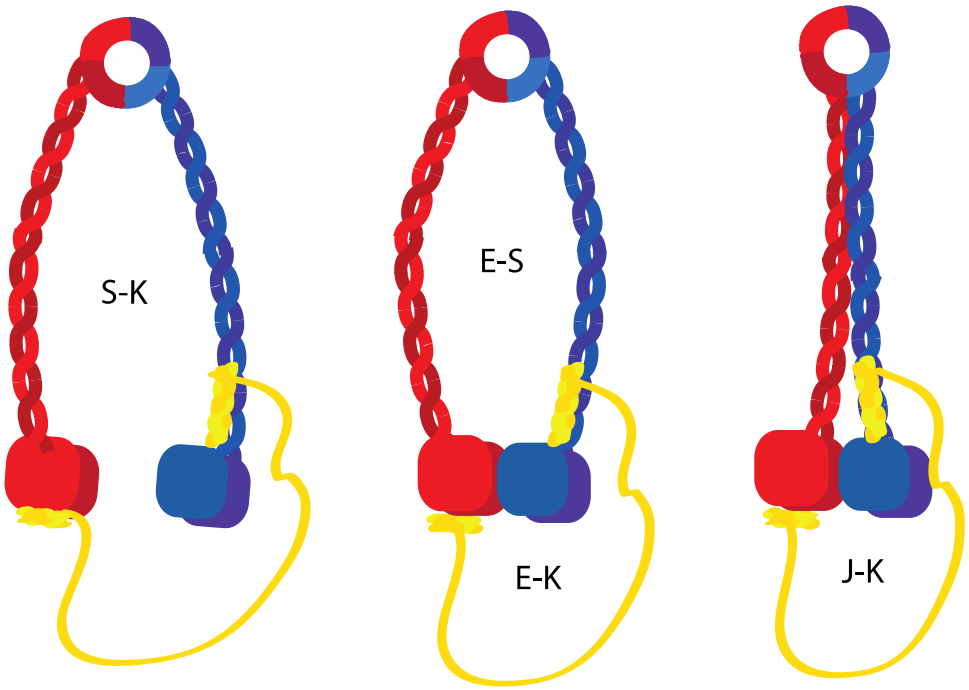


Figure 1-9. SMC chambers for entrapment of DNA
Left: Fully open complex S-K or SMC-kleisin.
Middle: Open SMC compartment closed by engaged heads E-S, or a bottom chamber-E-K, between the engaged heads and the kleisin.
Right: The bottom chamber between juxtaposed heads and the kleisin.

compartment. Entrapment in the S-K compartment, however, requires both Scc2 and Scc3 (Collier *et al*, 2020). In parallel to this study, two different cryo-EM structures have been published describing how this "transition" or "clamp" state binds to DNA (Higashi *et al*., 2020; Shi *et al*., 2020).

The loop extrusion model

The role of cohesin in the maintenance of chromosome cohesion, ensuring their proper segregation, was the first function identified for cohesin. Since, it has become more and more apparent, especially through the work accomplished in mammalian cells, that cohesin also has an essential role in organising chromatin 3D structure in interphase. It has been hypothesised that through this organisation of the chromatin by cohesin that enhancers distant to promoters in a linear sequence can be brought into proximity in a 3D space to promote transcriptional events (Busslinger *et al*., 2017; Rao *et al*., 2017; Schwarzer *et al*., 2017). However, cohesin depletion has only a mild effect on transcription. The most direct implication of LE in a biological process has recently been demonstrated during VDJ recombination in lymphocytes (Dai *et al*., 2021).

In 2001, it was proposed that chromosome condensation could be achieved by condensin complexes through the extrusion of loops of DNA within their rings, a mechanism called loop extrusion (Nasmyth, 2001). The hypothesis was that an SMC ring could associate with a molecule of DNA, slide over it and then extrude a loop (Figure 1-10). If many of these rings did the same, a long array of loops would be created, emanating from contiguous rings. This mechanism could explain the loops

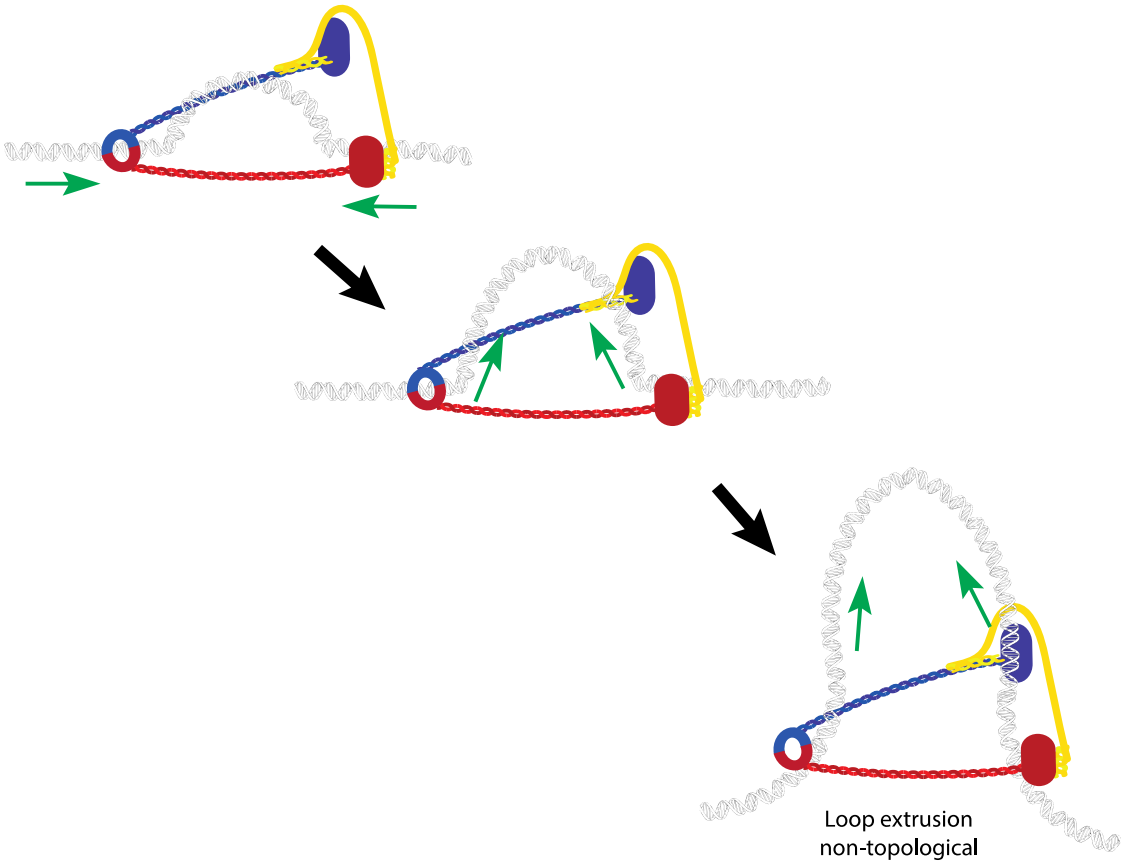


Figure 1-10. Model for loop extrusion (LE)
The non-topological loop extrusion model, with a single Cohesin complex. Loop begins to form and passes twice through the ring. DNA feeds into the ring making a more extensive loop.

observed by Marsden and Laemmli, that were emanating from a proteinaceous scaffold (Marsden & Laemmli, 1979). Moreover, the compaction would be physically maintained by the DNA integrity not by the resulting protein axis, thus also consistent with the observation that condensed chromosome structures can be broken by digestion of the DNA (Marko, 2008). Therefore, the loop extrusion model can explain the condensation process and the localisation of condensin all along the chromosomes axis.

It is only recently that a new technique designed to create maps of contact probability between genomic regions could reinforce the loop extrusion model and explain the mechanism used by cohesin to organise the genome during interphase. Hi-C data shows the probability of intra-chromosomal interactions which, when plotted, reveal the presence of high contact probability domains within chromosomal regions called Topologically Associated Domains (TADs). Initially performed in mammalian cells, these contact probability maps showed that these domains are dependent on cohesin and are between convergent CTCF sites, a DNA binding protein which acts as a road block for cohesin. When a loop is formed between two CTCF sites it grows until these two sites are brought in close proximity, resulting in a more stable loop that appears on the plot as a corner peak. Accordingly, it has been shown by ChIP sequencing, that cohesin peaks correspond to the beginning and the end of each TAD, at convergent CTCF sites. Furthermore, the rapid degradation of scc1, using an auxin inducible degron (AID) showed that all TADs disappear swiftly showing the essential role of cohesin in TAD organisation (Rao *et al.,* 2017). However, the mechanism by which the loops are stabilised at CTCF sites remains elusive.

More recently the organisation of the yeast genome was analysed by Hi-C and showed that *S. cerevisiae* chromosomes are also organised by cohesin mediated loops. However, the mechanism is different as there is no CTCF in yeast. Loop extension is controlled by a Pds5 mediated balance between Wapl and Eco1 (Dauban *et al.,* 2020). The organisation of the centromere involves both cohesion and loop extrusion. After loading at the CEN and translocating away, cohesin complexes trapped between convergent genes at the border of the pericentric region organise the two regions on both sides of the CEN by loop extrusion. When the pulling forces are applied by the spindle, the cohesive fraction present at these borders keeps the two sisters together while the centromeres are pulled apart by extending the pericentric region enabling the establishment of biorientation (Paldi *et al.,* 2020).

Finally, the loop extrusion model found a more recent validation with the development of single molecule techniques visualising loop extrusion *in vitro*. Initially set up with condensin but more recently confirmed with cohesin, this technique involves using real-time microscopy to track a piece of DNA tethered to a glass slide, after addition of an Smc complex produced *in vitro* (Banigan *et al*., 2020; Kim *et al*., 2019; Ganji *et al*., 2018). This enabled the direct observation of ATP dependent loop extrusion for both condensin and cohesin suggesting that the mechanism could be general to all Smc complexes. More importantly, locking the three interfaces of the cohesin trimer using a combination of peptidic fusion and chemical cross linking revealed that a completely circular molecule could extrude loops *in vitro*. This opens new investigation on the topology of the Smc complexes entrapment of DNA.

The topology of Chromatid cohesion: one ring or two?

The question of whether the two sister DNA molecules go through a single ring or multiple rings in order to establish cohesion, was initially addressed using the same strategy by closing the three interfaces of the cohesin trimer. This was tested in a heterozygous yeast strain in which one of the two alleles can be cleaved by the Tobacco Etch Virus (TEV) protease. Mini-chromosomes topologically entrapped by a 50/50 mix of cohesin TEV cleavable or not were isolated and analysed with or without TEV cleavage. The result was that 50% of the mini-chromosomes were lost after TEV treatment suggesting that only a single ring holds two sisters. A higher percentage of mini-chromosome loss after TEV treatment would be expected if multiple rings were necessary for cohesion, such as the handcuff model (75%) (Haering *et al.,* 2008).

In a second experiment using a diploid cell with two differently tagged Smc3 molecules, the cohesin complexes were circularized by *in vivo* crosslinking and immunoprecipitated using one of the tags. The purified complexes were then analysed using the second tag, but no signal could be detected and no higher molecular weight species were seen. This suggests that only one heterotrimeric cohesin ring holds two DNA molecules together (Gligoris *et al.,* 2014).

Localisation of cohesin on chromatin

Genome wide studies using ChIP have identified sites of cohesin interaction throughout chromosomes (Blat & Kleckner 1999). These cohesin associated regions (CARs) exist, irregularly spaced throughout chromosome arms (2-35kb), they are

located between convergent genes (Lengronne *et al.,* 2004). A predominant site for cohesin loading in yeast, is a 20-50kb region which covers the centromere DNA element (CDE) (Lengronne *et al.,* 2004). The Ctf19/CCAN complex has been associated with loading at the centromere (Eckert *et al.,* 2007; Fernius & Marston, 2009; Hinshaw *et al.,* 2017). The absence of either Ctf19 or Chl4 reduces the recruitment of Dbf4-dependent kinase (DDK) to the centromere, this in turn affects the recruitment of the loading complex (Natsume *et al*., 2013). DDK is essential to activate the Mcm2-7 helicase for the replication of DNA, it is recruited in G1 by Ctf3, which is bound to Ctf19. DDK phosphorylates Ctf19, allowing it to bind to Scc4, thus recruiting the loading complex to centromeres (Hinshaw *et al.,* 2017).

The analysis of cohesin complexes localised on DNA is technically challenging. Mini-chromosome assays combined with the closure of the complex interfaces using cross linking remains the only way to characterise topological entrapment, that is resistant to SDS treatment. However, it is not informative on the genomic localisation of the complex in cells. A genome wide ChIP sequencing approach is more informative in that respect, but does not distinguish between topological entrapment and DNA interaction. However, ChIP sequencing profile analysis of cohesin mutants has brought some insights on the way the complex interacts with DNA. Mutant complexes that are unable to bind ATP, Smc1K39I and Smc3K38I, are also unable to interact with DNA (Hu *et al.,* 2011). Mutations in the ATPase heads Smc1 E1158Q or Smc3 E1155Q that can bind ATP but cannot hydrolyse it, can localise at the centromere as a single peak where they accumulate (Hu *et al.,* 2015; Petela *et al.,* 2018). However, the mini chromosome assay revealed that these mutations are unable to entrap DNA

topologically (Srinivasan *et al*., 2018) Together, these experiments showed that the binding of ATP is crucial for the recruitment of the complex to the DNA, and that the topological entrapment of DNA requires hydrolysis.

In this context, it is reasonable to consider the transition state described above as an intermediate state of the loading reaction. In this intermediate state, Scc2 and ATP binding are needed to entrap the DNA in the ES and EK chambers without SK entrapment. In a later step, ATP hydrolysis and Scc3 are essential for the SK entrapment of DNA (Collier *et al.,* 2020). Accordingly, the loading complex Scc2-Scc4, is essential for the recruitment of cohesin to both the centromeres and the arms in yeast (Ciosk *et al.,* 2000). Scc3 has been shown to interact with cohesin during the loading reaction *in vitro* in fission yeast (Muryama *et al.,* 2014) and in *S. cerevisiae in vivo* (Petela *et al.,* 2018). A cycle of ATP hydrolysis, driven by the SMC heads, is also required during loading (Arumugam *et al.,* 2003). All these data lead to the current model of the loading reaction: first the Scc2/4 recruits cohesin to the DNA, a cycle of ATP hydrolysis must then be completed. Cohesin is then able to translocate away from where it was loaded. Translocation away from the loading site at the yeast centromere has also been demonstrated *in vivo*: the calibrated ChIP sequencing analysis of synchronised cells sampled every 15 minutes shows the movement of cohesin throughout the centromeric region (Fernius *et al* 2013, Hu *et al.,* 2015; Petela *et al.,* 2018). Translocation of the human cohesin complex along linear DNA *in vitro*, has been demonstrated using TIRF microscopy. By attaching a linear molecule of DNA to a glass slide and flowing in a mixture of purified cohesin, loading complex and ATP, cohesin can load onto DNA and move along it (Kanke *et al.,* 2016; Davidson *et al.,*

2016).

Finally, several studies have shown a tight link between transcription and cohesin positioning. In mammalian cells deleted for both CTCF and Wapl, cohesin was found to accumulate at the converging ends of transcribed genes, showing that transcription can indeed relocate cohesin (Busslinger *et al.,* 2017). In Yeast, which does not have CTCF, similar positioning of cohesin dependant on transcription has been described (Bausch *et al.,* 2007; Hinshaw *et al.,* 2017; Ocampo-Hafalla *et al.,* 2016; Paldi, *et al.,* 2020). Once on DNA, cohesin is able to translocate from these loading sites, driven by active transcription. Cohesin has been shown to accumulate at sites of convergent transcriptional termination (Davidson *et al*., 2016; Glynn *et al.,* 2004; Lengronne *et al.,* 2004; Ocampo-Hafalla *et al.,* 2016).

Cohesinopathies

The term cohesinopathies relates to a group of genetic diseases caused by a wide range of mutations and deletions, which occur within human cohesin and its regulatory proteins. These syndromes result in a range of developmental delay/learning difficulties, distinct facial features, growth retardation and limb/organ abnormalities. They have a varying degree of severity, often relating to which gene carries the mutation. These diseases include Roberts syndrome (RBS) and Cornelia de Lange syndrome (CdLS).

RBS is a rare autosomal recessive congenital disorder, caused by mutations in the acetyltransferase gene Esco2. The mitotic cells of patients exhibit a severe defect

in chromosome segregation, linked to the inability of Esco2 to stabilise sister chromatid cohesion (Colombo *et al.,* 2019).

CdLS affects as many as 1 in 10,000 births. Many studies of CdLS patients have identified point mutations, insertions and deletions in Smc1A, Smc3, Rad21, NipBL and HDAC8 (Kleine, 2018). Up to 60% of mutations are found in NipBL (Scc2) and cause the most severe defects (Bose *et al.,* 2010; Remeseiro *et al.,* 2013). One recent study has also implicated other cohesin regulatory genes Stag1, Stag2, Pds5A, and Wapl (Yuan *et al.,* 2019). Other non-Cohesin genes have been identified, involved in transcription regulation and chromatin remodelling (Avagliano *et al.,* 2020). Whilst mutations found in CdLS patients do not tend to cause problems with cohesion, it is thought defects are related to abnormal transcription throughout the genome (Kawauchi *et al*., 2009; Krantz *et al*., 2004; Tonkin *et al.,* 2004). A recent study has also shown *SMC1A* and *NIPBL* are involved in V(D)J recombination, suggesting a possible role in long range gene interactions or DNA repair (Björkman et al, 2018). The different actions of cohesinopathies demonstrate the diversity of cohesin's roles in the cell.

Cohesin and cancer

The first somatic mutations implicated in colon cancer, were isolated in 2008. These were heterozygous missense mutations in one or more, of the cohesin sub units *SMC1A, SMC3, NIPBL*, and *STAG3* (Barber, 2008, cited in Hill, 2016). Further studies also identified deletions of *RAD21* and *STAG2* (Rocquain *et al.,* 2011, cited in Hill *et al.*, 2016). Since then, many other cancers have been linked to mutations or truncations of cohesin genes. Mutations in *STAG2*, which is present on the X chromosome, have a

high prevalence in cancer, and usually result in a truncation of the protein. *STAG2* mutations have been used to identify roles in anaphase integrity, chromosome stability and sister chromatid cohesion (Hill *et al.,* 2016). A reduction in the overall levels of core cohesin proteins has also been observed in cells with mutated *STAG2* (Kim *et al.*, 2016 cited in Hill *et al.*, 2016).

Thesis Objective

In order to better understand the biological role of Scc2, the aim of this thesis is to investigate the interaction between Scc2 and the different subunits of the cohesin complex using genetic mutations combined with *in vivo* and i*n vitro* protein crosslinking. Upon identification of interaction sites, their functional relevance would further be investigated in relation to ATP hydrolysis and cohesin occupancy on DNA.

Chapter 2: Identification of Scc2 binding motifs in the N-terminal region of Scc1

Introduction

At the beginning of this project, little was known about how and where Scc2 binds to the cohesin complex. Identifying the points of contact between Scc2 and the different subunits of the cohesin complex is crucial to understand how it regulates its activity: conformation, ATPase activity and translocation on the DNA.

An early study in yeast had co-immunoprecipitated Scc2 with all four of the cohesin subunits (Arumugam *et al.*, 2003). Similar findings from a study undertaken in fission yeast using peptide arrays, suggested a complex mode of interaction between Scc2 and the four other subunits of the cohesin complex (Muryama *et al.,* 2014). In 2016, the crystal structures of Scc2 from *A. gossypii* (Chao *et al.,* 2017) and *C. thermophilum* (Kikuchi *et al.*, 2016) were published, describing a HEAT repeat protein similar to Scc3 and Pds5 with a hook and a handle. In *C. thermophilum,* pull down assays described the binding of Scc2 to the N-terminal half of Scc1. However, it had been shown by a genetic deletion screen (Chan *et al.,* 2013) and also in the description of the crystal structure, that Pds5 also interacts with the N-terminus of Scc1 (Lee *et al.,* 2016; Muir *et al.,* 2016). This led to the hypothesis that Pds5 and Scc2 could substitute each other for Scc1 N-terminal binding (Kikuchi, *et al.,* 2016).

Two different functional assays provided further evidence for the potential switch between Scc2 and Pds5 binding on Scc1. First, an *in vitro* ATPase assay showed that, when tested with the cohesin tetramer (Smc1, Smc3, Scc1 and Scc3) in the presence of DNA, Scc2 stimulates the ATPase activity. However, Pds5 not only failed

to induce ATPase activity but could inhibit the stimulating effect of Scc2, when added in excess (Petela *et al.,* 2018). Second, the localisation of both Scc2 and Pds5 was analysed by ChIP sequencing. In a wild type situation, pulling on Scc1 shows the normal localisation of the cohesin complex which consists of one peak either side of the centromere and a peak at each peri-centromic region (Figure 2-1A). Looking at the loading complex by pulling on Scc2 gives a sharp peak directly at the centromere, showing that the centromere is the major loading site and that Scc2 does not translocate with the cohesin complex (Figure 2-1B). This was confirmed by immunoprecipitating cohesin complexes containing an ATP hydrolysis mutant (Smc3E1155Q) in a heterozygous strain, that accumulates at the centromeric loading site, but cannot translocate away (Figure 2-1C). Conversely, immunoprecipitating Pds5 gives a WT cohesin profile, and does not colocalise with mutant complexes at the centromere (Figure 2-1D). This suggests that Pds5 is associated with the cohesin complex after, not during the loading reaction.

In the context of this hypothesis, a swap between Scc2 and Pds5 would fit with several experimental observations. First, after the initial recruitment mediated by the Scc2/Scc4 complex via its interaction with Ctf19/Chl4, Scc2 has a stimulating effect on the ATP hydrolysis activity of cohesin that is essential for the loading on the DNA and its translocation away from the centromere. In vitro translocation assays, using single molecule tracking, showed that translocation depends on ATP hydrolysis (Davidson et al, 2019; Gutierrez-Escribano *et al* 2019; Kim *et al*, 2019). Once the cohesin complex is loaded, Scc2 appears to be very dynamic, as observed by FRAP in yeast in which Scc2 has a residence time of 2 to 4 seconds (Hu *et al*., 2011). The same rapid 'On and Off' rate, of NIPBL binding to the chromatin, was also observed in mammalian cells

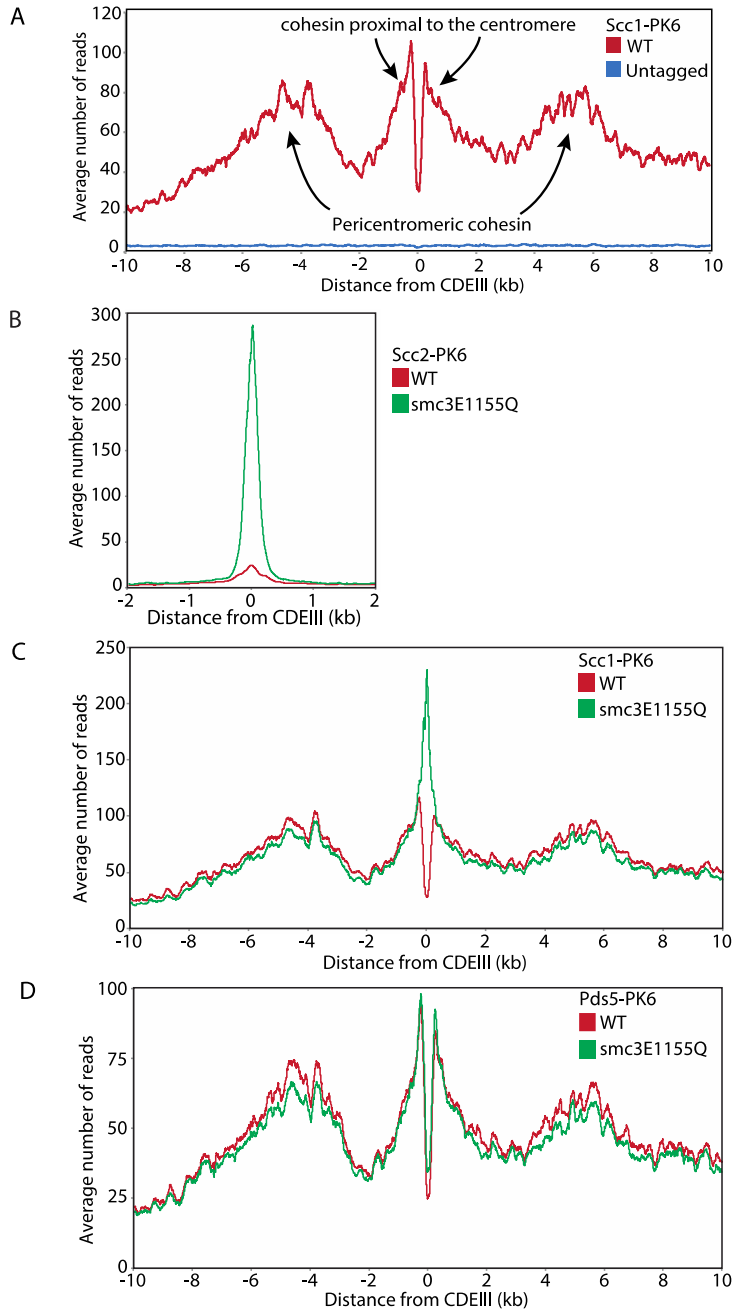


Figure 2-1. Calibrated ChIP-sequencing profiles showing average distribution of cohesin subunits around centromeres

A. Calibrated ChIP-sequencing profile demonstrating the average occupancy of WT cohesin. Showing cohesin occupancy proximal to either side of the centromere and at the peri-centromeric regions, around 5kb from the centromere (figure from Petela, 2016). B. Pulling on Scc2-PK6, calibrated ChIP-sequencing profile demonstrating the occupancy an Smc3E1155Q mutant, defective in ATP hydrolysis and translocation from the centromere compared to WT. C. Pulling on Scc1-PK6, calibrated ChIP-sequencing profile demonstrating occupancy of an Smc3E1155Q mutant, compared to WT. D. Pulling on Pds5 shows a WT like read, suggesting it does not localise at the centromere itself. B, C and D expressed both WT Smc3 and Smc3E1155Q as the latter is lethal (figures from Petela *et al.*, 2018).

(Rhodes *et al.,* 2017). The fact that Scc2 only binds to chromatin very transiently is possibly why the traditional methods of fixation used in ChIP sequencing cannot detect Scc2 association with DNA, except at the main centromeric loading site.

If Pds5 was competing for Scc1 binding it could regulate cohesin activity: Pds5 is essential for either releasing the complex via Wapl or mediating cohesion via Eco1 (Chan *et al* 2012; Chan *et al* 2013). Therefore, a swap of Scc2 for Pds5 could adjust the activity of cohesin in loop extrusion or cohesion establishment.

In order to investigate the functional significance of the competition for binding to Scc1, it is essential to fully characterise where these proteins bind. Disruption of either the Scc2 or Pds5 binding site on Scc1 should be lethal. Furthermore, defects in binding to Scc2 would be expected to cause a reduction in loading and therefore calibrated ChIP sequencing could be used to measure this. Using such an approach, can I define discrete binding sites for Pds5 and Scc2 on Scc1?

2.1 Deletions within the Pds5 binding region of Scc1 cause lethality in the absence of WT protein

A library of small deletions within Scc1 had previously been created to study the interaction between Scc1 and Pds5. These deletions were successfully used to identify the binding sites to Pds5 in the N-terminal region of Scc1 (Chan *et al*., 2013). However, it has been shown in fission yeast, *S. pombe*, that the N-terminal region of Rad21 (Scc1) also binds to Mis4 (Scc2): residues 145-152aa (Muyama *et al.,* 2014). Finally, in

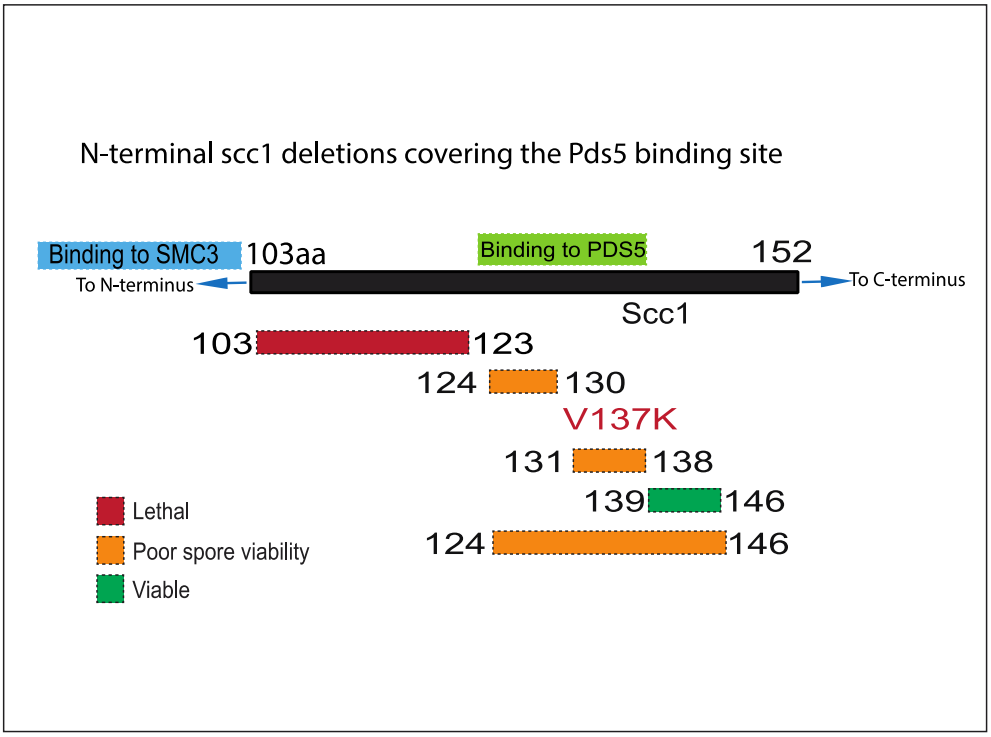


Figure 2-2. Schematic of PDS5 binding region of Scc1
Schematic of the N-terminal of the Scc1 protein 103-152aa (Black), with small deletions created to find the binding of Pds5. Colouring indicates viability (see key). Data from Dr. Chris Chan, personal communication.

C. thermophilum, using a GST pulldown assay, there was a reduction in the amount of Scc1 pulled down by Scc2 in the presence of increasing quantities of Pds5, suggesting a competition between Pds5 and Scc2 for the binding to Scc1 N-terminal region (Kikuchi *et al.,* 2016). It followed that some of the existing Scc1 deletion strains could be used to look for an interaction with Scc2.

The Scc2/Scc4 complex is essential for the interaction between DNA and cohesin, therefore, a deletion that disrupts this interaction would be expected to cause a defect in cohesin loading and would most likely cause sickness or lethality of the cell. However, when previously tested, most of the strains deleted at the N-terminus of Scc1 were found to have poor spore viability, but were still viable (Figure 2-2). One possible explanation came from the fact that the endogenous *SCC1* gene was placed under the control of a repressible promoter in order to analyse the potential lethality of the different deletions within *scc1* introduced at an ectopic locus. The wild type *SCC1* gene had been placed under the Gal promoter, which can be repressed in the presence of glucose. However, this promoter is known to be leaky and could potentially still express very low levels of Scc1 even in the presence of glucose. Hence a small number of colonies carrying the mutation were able to grow, suggesting a viable yet slow growth phenotype. The first aim of the project was to re-assess the viability of a series of deletions in the Pds5 binding region of Scc1.

To address this, all of the *scc1* deletion strains were back crossed to WT to remove the WT allele with the GAL promoter (Figure 2-3, top panel). A spore from each cross containing the WT *SCC1* at the endogenous locus and mutant *scc1* at the LEU2 locus was isolated. These spores were then crossed to a strain with *scc1Δ::KanMX4* at the endogenous locus and a marked copy of *SCC1* at the URA3 locus

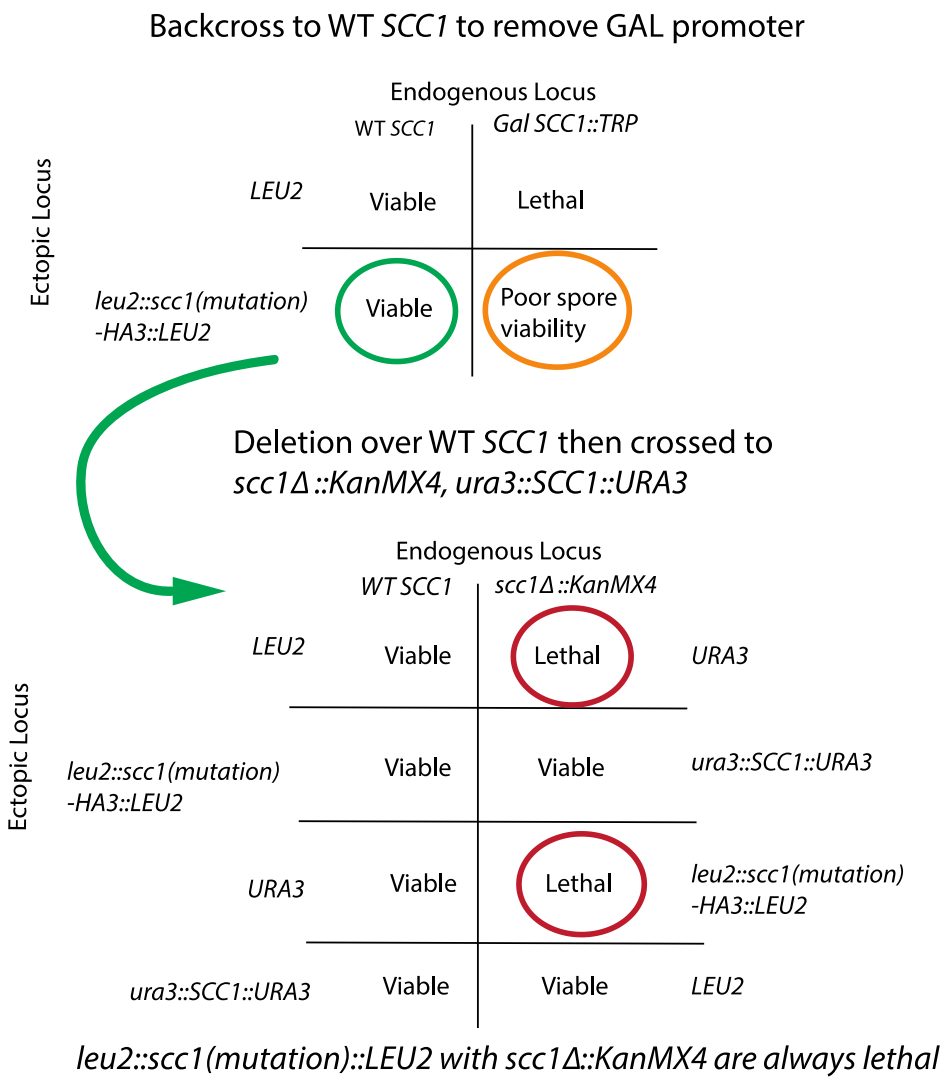


Figure 2.3. Genetic crosses show lethality of scc1 deletions after removal of GAL promotor on *SCC1*

Top panel: To remove the Gal promoter, the scc1 deletion strain was backcrossed to WT (K699), the resulting diploid was then dissected.

Bottom panel: A spore was taken from the backcross containing both endogenous WT *SCC1* and the scc1 deletion at the LEU2 locus (ectopic). This was then crossed to a strain containing *scc1Δ::KanMX4* at the endogenous locus and a marked copy of WT *SCC1* at the *URA3* locus (K13898). The empty *LEU2* and *URA3* loci are written for clarity of the cross.

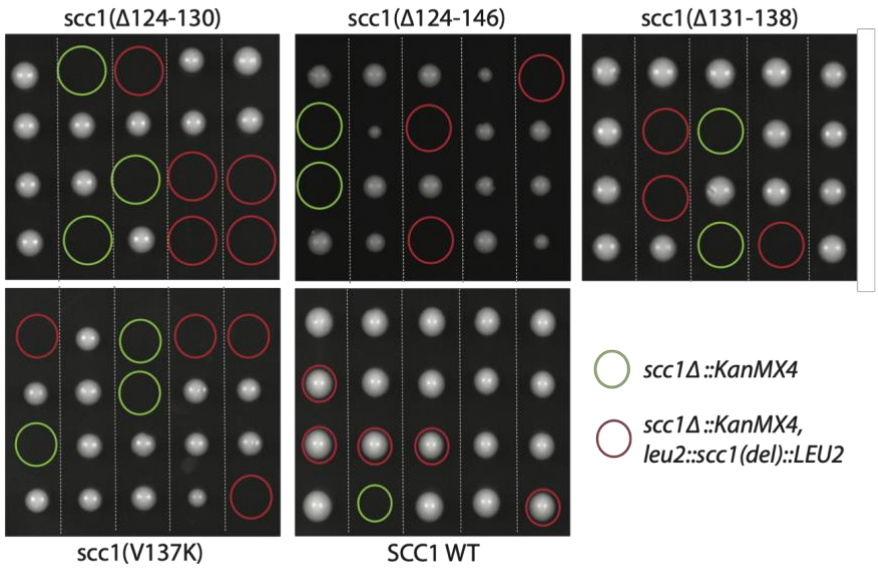


Figure 2-4. Scc1 deletion strains in the Pds5 region are lethal
Example of 5 dissected tetrads from four scc1 deletion strains.
K13898 x K26120, K26142, K26144, K17189 and K17184
Spores containing *scc1Δ::KanMX4* are circled green.
The spores containing the different scc1 deletions as the only source of scc1 are all lethal and circled in red (*scc1Δ::KanMX4, leu2::scc1(del)-HA3::LEU2*).

for viability (Figure 2-3, bottom panel). The crosses were then placed on starvation media, and the subsequent tetrads dissected. Once grown the genotypes of the resulting viable spores were analysed by replica plating. Two complete plates of 20 tetrads were dissected for each strain to confirm the result (Figure 2-4).

No spores expressing the mutant form of Scc1 as the unique source of Scc1 *(scc1Δ::KanMX4, leu2::scc1(del)::LEU2*) were viable, as depicted by the red circles in Figure 2-4. This shows that the series of deletions in the Pds5 interacting domain of Scc1 that were initially found to induce poor spore viability were in fact lethal. Given that Pds5 is an essential gene, and that this region of Scc1 is crucial for binding Pds5, this result appears more consistent than the initial finding.

2.2 Mutations in the Pds5 binding site are defective in pericentromeric accumulation

The initial *SCC1* screen identified several small lethal deletions within the N terminal region of Scc1 between amino acids 103 and 138. The latter part of this region 125-138aa was previously identified as interacting with Pds5. A single point mutation of Scc1 had also been isolated that was shown to abolish Pds5 interaction, valine 137 to a lysine: V137K. This point mutation was initially shown to induce slow growth but after removing the *GAL-SCC1*, *scc1V137K* appears to be lethal (Figure 2.4). At the time, the impact of this mutation on cohesin localisation on the genome was addressed by traditional ChIP sequencing and showed a slightly reduced distribution along chromosome arms compared to WT. In this experiment Scc1V137K did not seem to affect loading at centromeres (Chan *et al.,* 2013). However, If Pds5 and Scc2 were

binding to the same site on Scc1, we would expect to see a reduction in loading around the centromere when testing Scc1 deletions in this region. The Scc1V137K mutant was re-investigated using calibrated ChIP sequencing in order to measure the genuine genomic occupancy from background noise. Other Scc1 deletions previously characterised in the Pds5 binding region were also investigated for their interaction between DNA and cohesin.

As all of these Scc1 deletions cause lethality, all strains used for the following calibrated ChIP sequencing experiments had WT Scc1 expressed from the endogenous locus, and the deletion or mutation expressed from an ectopic locus, tagged with HA3. The immunoprecipitation was performed by pulling on the HA3 tag, so only the ectopic copy was assessed in the assay. Cells were grown to log phase and their cell cycle status was verified by flow cytometry. This showed that the cell cycle profile of the mutant strains was unaltered (Figure 2-5A). Analysis by western blot showed that the Scc1 protein expression levels were similar to WT (Figure 2-5B).

Both Scc1V137K and the Scc1Δ124-130 showed similar levels to WT either side of the centromere, however, both mutants have a slimmer profile at the base of the central peak, between 1-2kb (Figure 2-5C). At the pericentromeric region, the amount of cohesin is reduced by half and has a slight reduction along the arms. The fact that both Scc1V137K and Scc1Δ124-130 containing complexes are found at the proximal centromere indicates that these cohesin complexes can still hydrolyse ATP otherwise they would be expected to accumulate as a large peak directly at the CEN, as seen with the smc3E1155Q mutant. This was confirmed with Scc1V137K, which was still able to hydrolyse ATP *in vitro* both with and without the addition of DNA. The rate of

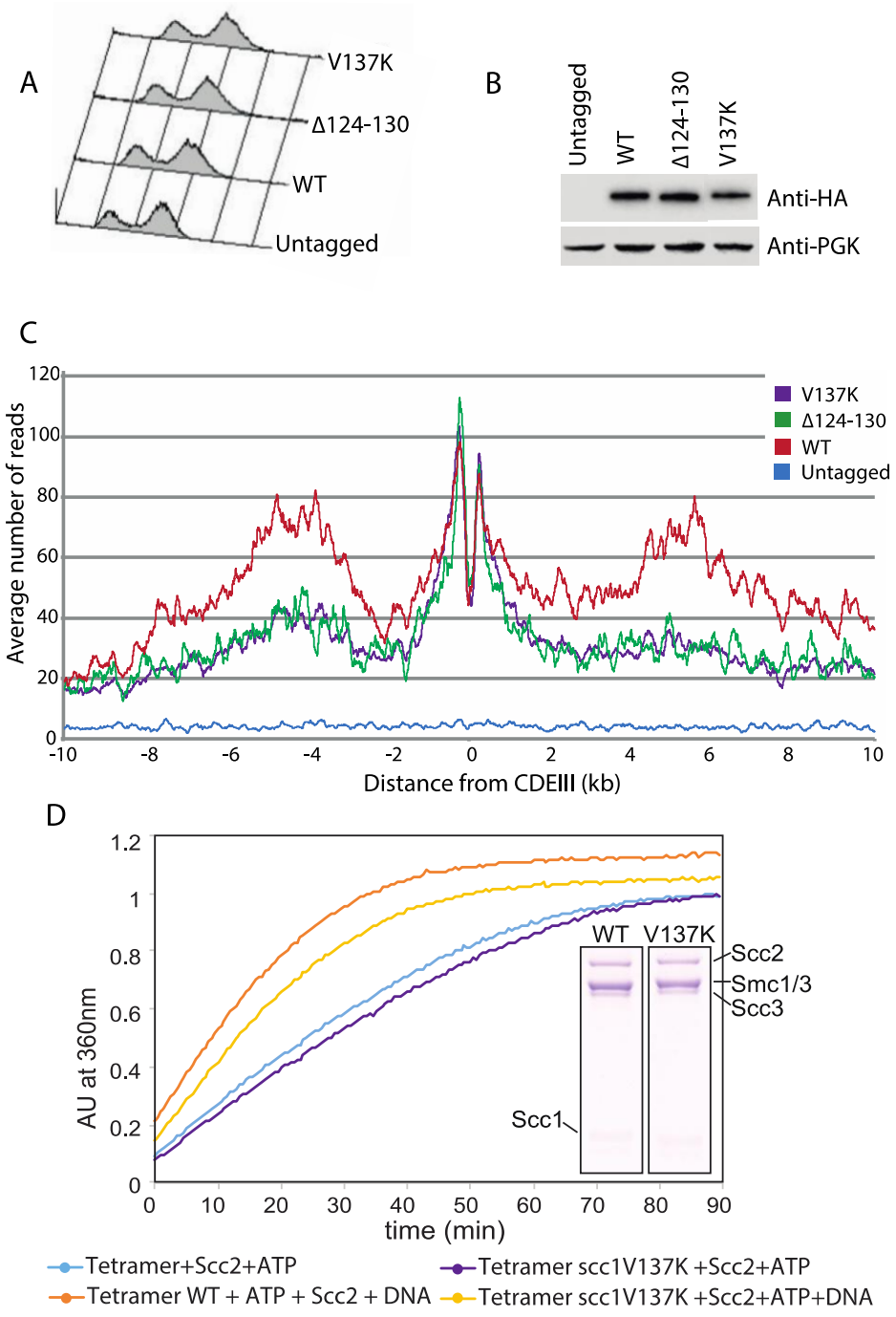


Figure 2-5. The Scc1 mutation V137K and deletion of 124-130 are defective in pericentromeric accumulation

A. Flow cytometry profile of cycling cells used for experiment. B. Expression levels of Scc1V137K-HA3, and Scc1Δ124-130-HA3, SCC1-HA3 and untagged control. Pgk1 was used as a loading control. C. Calibrated ChIP sequencing profile showing average distribution of cohesin around centromeres, 10 kb either side of the CDEIII strains: K17189, K26120, K17184, K699). Cycling cells expressing scc1(mutation)-HA3, were grown at 25°C in YPD. D. Effect on ATPase activity of wild type and Scc1V137K tetramers. This experiment was performed by Dr. T. Gligoris/Mr. M. Voulgaris, repeated x3).

hydrolysis was slightly lower in Scc1V137K when compared to WT (Figure 2-5D). However, these mutations prevent accumulation of the complex at the pericentromeric regions. This suggests three possibilities: either these complexes are not able to translocate from the centromere to the pericentromere; they cannot stop at the pericentromere; or they are not stable enough to accumulate there.

In order to further dissect the binding regions of Scc1 N-terminus, an additional three lethal strains carrying different deletions covering Scc1 amino acids 103 to 146 were analysed by ChIP sequencing.

Three Scc1 deletions were analysed: Δ103-123, Δ124-146 and Δ131-138. The ChIP sequencing profile analysis showed that these mutants have lower pericentromeric peaks in respect to their centromeric peak, when compared to the WT (Figure 2-6A). The strain containing Scc1Δ124-146 seems to show levels above WT proximal to the centromere, and similar loading at the pericentromeric region. However, this deletion mutant is expressed at higher levels compared to wild type (Figure 2-6B), which could account for the higher peaks. Following the same logic, the expression levels of Scc1Δ131-138 is much lower than wild type, which is also reflected in the number of reads on the graph. This suggests that there is a correlation between the protein expression level and the number of complexes observed bound to the chromatin. The cell cycle profile of these different strains was unchanged (Figure 2-6C). Nevertheless, the profile of genomic occupancy across the centromeric region show that all three deletions Δ103-123, Δ124-146 and Δ131-138 present a pattern similar to V137K and Δ124-130 (compared to Figure 2-5A). This suggests the

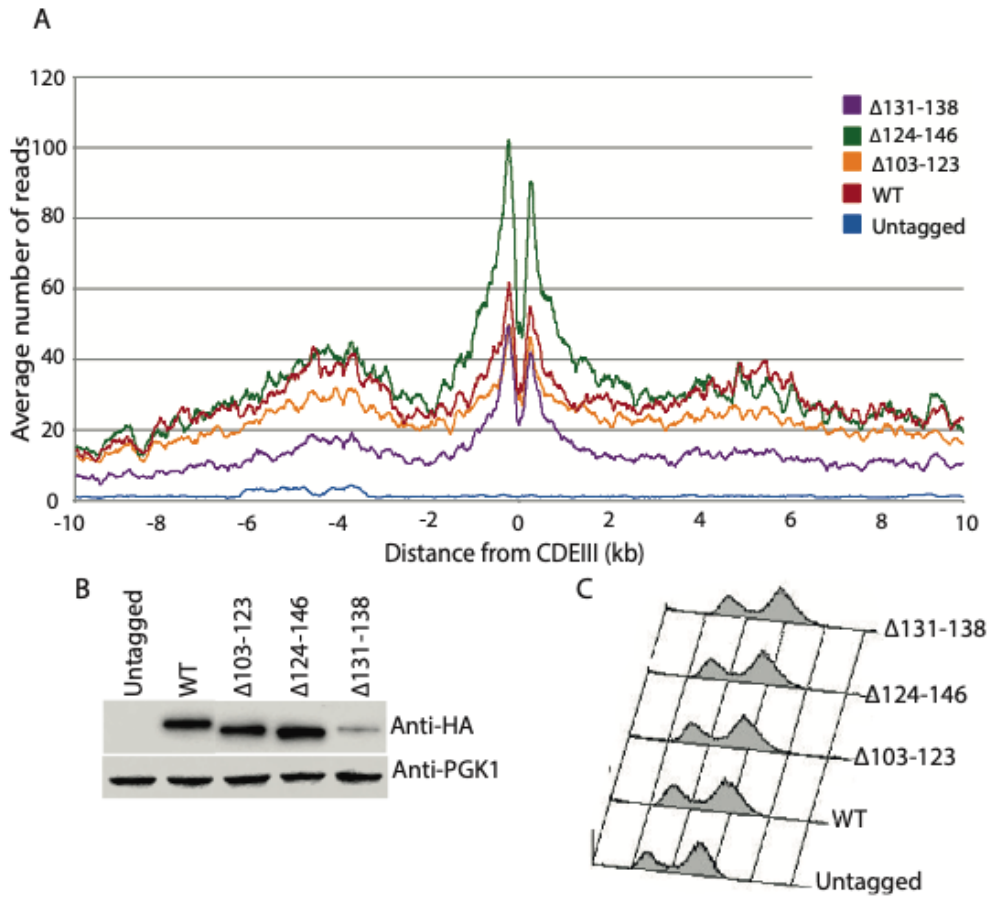


Figure 2-6. Deletions in the Pds5 binding region of Scc1 are defective in pericentromeric accumulation

Cycling cells expressing scc1-HA3, were grown at 25°C in YPD, strains K26140, K26142, K26144, K17184, K699

A. Calibrated ChIP sequencing profile showing average distribution of cohesin around centromeres, 10 kb either side of the CDEIII, repeated x2.

B. Expression levels of strains carrying Scc1-HA3, PGK was used as a loading control.

C. Flow cytometry profile of cycling cells used for the experiment.

mutant proteins are able to load, but all have a similar defect probably due to the lack of Pds5 binding.

2.3 Wild type cohesin can modify the activity of mutant complexes.

The analysis of the Scc1 mutations tested so far were obtained from cycling cells that were kept alive by the presence of wild type Scc1 at the endogenous locus.

In order to determine if the wild type cohesin complexes could have an impact on the Scc1 deleted mutants, endogenous Scc1 was placed under the MET promoter with a C-terminal Myc18 tag. This allows the inhibition of WT Scc1 expression in the presence of methionine. Cells were grown in –Met media, moved to YPD + methionine, pheromone arrested in G1, then released into YPD + methionine (Fernius *et al,* 2013). Samples were taken every 15 minutes for flow cytometry, which showed that 60 minutes post release, the cells had entered G2 (Figure 2-7A). However, they had not entered anaphase according to the counting of binucleates (Figure 2-7B). Samples were taken 60 minutes post release and processed for calibrated ChIP sequencing. A western blot probing for the Myc18 tag on Scc1, was carried out to check for the endogenous Scc1 protein levels in cycling cells, at release and 60 minutes post release, and confirmed the full depletion of the endogenous WT protein (data not shown). Two different Scc1 mutations were analysed using this strategy: Scc1V137K and Scc1Δ131-138.

In absence of wild type protein, both Scc1V137K and Scc1Δ131-138 show a peak at the proximal centromere similar to wild type (Figure 2-7C). However, the cohesin peak at the pericentromeric regions is almost completely lost. Furthermore,

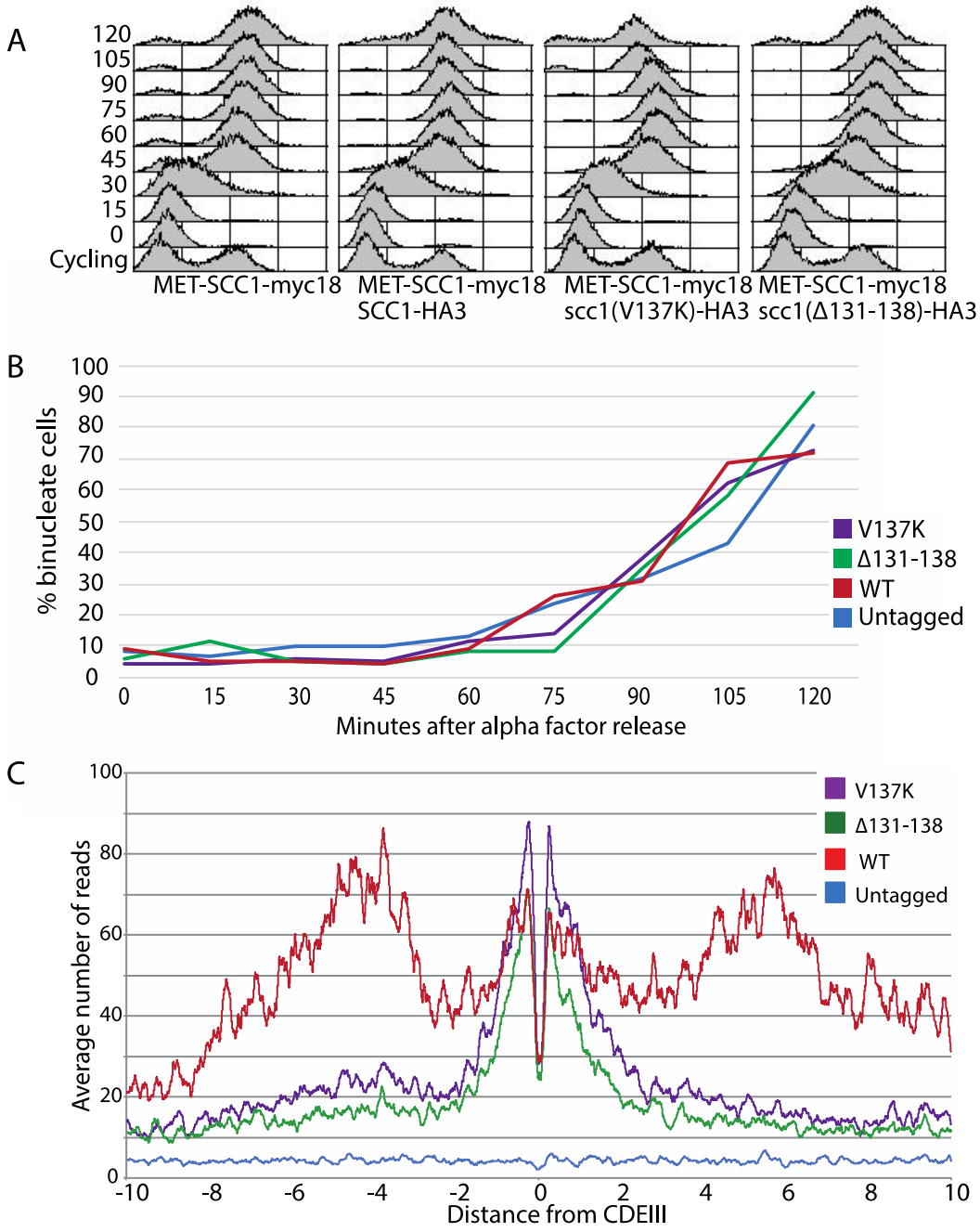


Figure 2-7. Defects in pericentromeric accumulation of Scc1 mutants is more pronounced without the WT copy.

Cells were grown in –MET media at 25°C, arrested with alpha factor and released into YPD + methionine. Samples were taken 60 minutes post release.

A. Calibrated ChIP sequencing profile showing average distribution of cohesin around centromeres, 10 kb either side of the CDEIII for Scc1 mutants strains *K26248, K26251, K26489 and K25222.* B. FACS showing cycling cells, G1 arrest and release into G2. C. Percentage of binucleates in cells.

this reduction is much more pronounced than in a strain in which both Scc1V137K and WT Scc1 are expressed (Figure 2-5). Finally, the Cohesin levels on the arms are reduced by about two-fold compared to WT. These results suggest that the Scc1V137K mutation does not modify the localisation of cohesin in direct proximity to the centromere. However, at the pericentromeric region, cohesin is almost completely lost if Scc1V137K is the only source of Scc1 in the cell. This result suggests that the endogenous wild type Scc1 appears to somehow help the localisation of the mutated complex at the pericentromere.

As mentioned earlier, the fact that these mutant cohesin rings can undergo the first step of the loading process and be recruited to the centromere and then proceed to translocate to the proximal centromere, suggest that they are able to interact with Scc2 and hydrolyse ATP. Nevertheless, the fact that they are enriched at the proximal centromere suggests that they might be unable to translocate from the proximal centromere to the pericentromere. Suggesting that in the presence of a wild type version of the cohesin complex in the same cell, would help the mutated complexes to translocate from the proximal centromere to the pericentromere by an unknown mechanism.

A second explanation for this ChIP sequencing profile could be an instability of the mutated cohesin complexes at certain sites of the genome. Cohesin complexes could load and translocate but would not be stabilised at the pericentromeric regions, as these mutant cohesin rings can no longer bind Pds5. A similar phenotype is observed after degradation of Pds5 by an auxin inducible degron, which induce both the loss of pericentromeric cohesin peak and cohesion (Petela *et al.,* 2018). This

instability was also observed in mammalian cells in which the double deletion of Pds5A and B triggers the destabilisation of the TADs corner peaks in Hi-C (Wutz *et al.,* 2017; Dauban *et al*., 2019).

A recent study in *S. cerevisiae* shows that cohesin accumulation at the pericentromeric regions happens at specific border sites that coincide with convergent transcription sites (Paldi *et al.*, 2020). In the case of Scc1V137K, which abolishes Pds5 binding, one possibility could be that cohesin can interact with Scc2 and as a consequence, load and translocate, but cannot accumulate at the pericentromeric borders where the convergent genes are. In order to further analyse this possibility, the ChIP sequencing profile of Scc1V137K was analysed in detail at these specific convergent genes reported as borders in the literature. It appears that the Scc1V137K cohesin peaks are consistently reduced at these sites (Figure 2-8). In the context of this second hypothesis, the presence of wild type cohesin in the cell can interact with Pds5 and accumulate at the pericentromeric borders. This could prevent the mutant complexes from translocating further and lead to this partial accumulation.

The fact that Pds5 and Scc2 compete for Scc1 binding suggested that both HAWKs could share the same recruitment motifs in Scc1. However, these results suggest that if the Pds5 binding region of Scc1 is also involved in Scc2 recruitment, it is not the only region of Scc1 necessary for Scc2 binding. Even if the defects observed after deleting this region was due to a defect in translocation, cohesin is still able to load, therefore either Scc2 does not bind in this region, or multiple contacts are created throughout the N-terminal half of Scc1.

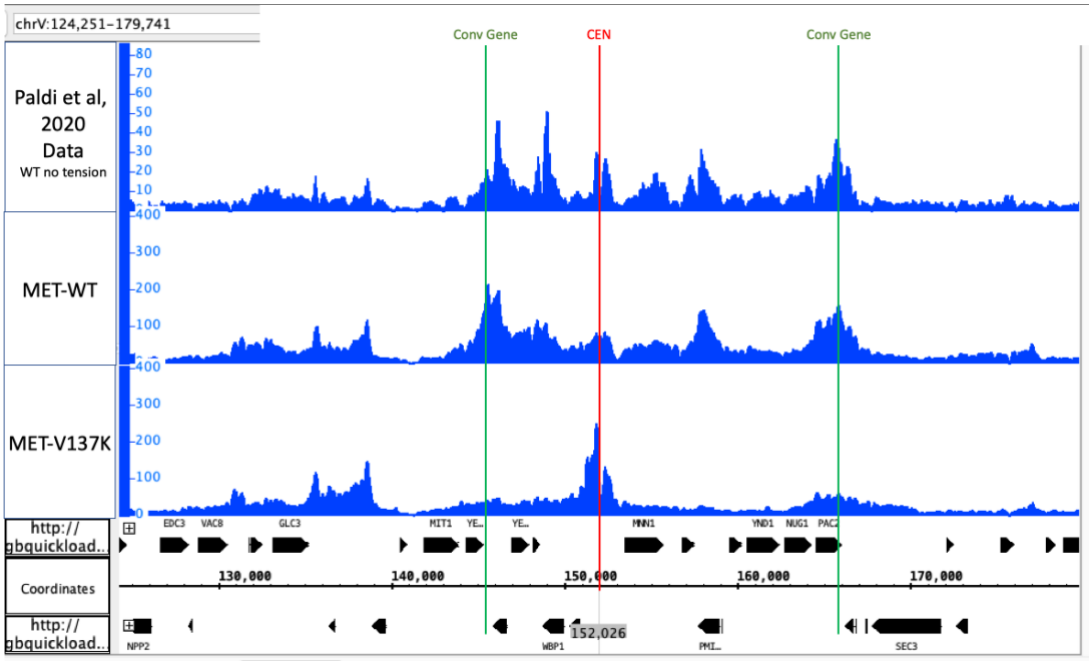


Figure 2-8. Reduction of cohesin occupancy observed in scc1V137K at the convergent genes, situated at the pericentromeric border.
Integrated Genome Browser alignment of ChIP sequencing data for Chromosome V, 124,251 to 179,741, highlighting the convergent genes at the pericentromere. Top panel: WT data without tension (GSM2789274_WTChip-ten.bw) (Paldi *et al* 2020). Middle: *MET-SCC1-myc18/leu2::SCC1-HA3::LEU2(K26489 bottom: MET-SCC1-myc18/leu2::scc1(V137K)-HA3::LEU2(K26248).*

2.4 Scc2 creates multiple contacts across the N-terminal half of Scc1

As none of the deletions analysed in the N-terminal region of Scc1 abolished the loading of cohesin, several Scc2 binding sites in Scc1 must exist. Therefore, different regions of Scc1 were investigated other than those already known to bind to the other members of the cohesin complex: Smc1(Scc1483-566aa), Smc3(Scc11-103aa) Scc3(Scc1319-393aa) or Pds5(Scc1124-140aa) (Figure 2-9). It had previously been observed that a large deletion Scc1Δ152-256, in the central region of Scc1 was lethal, but when split into two, both halves were viable (Scc1Δ152-206 and Scc1 Δ207-256aa). This could suggest that either, the removal of a large section of Scc1 makes it too short to function, or alternatively, multiple contacts are made between Scc1 and Scc2 throughout this central region.

In order to characterise if the lethality caused by this large deletion was due to the loss of function of Scc1, we analysed the impact of this deletion on *in vivo* crosslinking using BMOE to assess the ability of the larger deletions to form cohesin rings. The strains carrying Scc1Δ152-256 and Scc1V137K, were crossed to introduce *smc3K1032C–myc9,* which has been shown to crosslink to a natural cysteine (C56) in Scc1 (Gligoris *et al*., 2014). The Smc1/Scc1 interface is believed to bind before the Smc3/Scc1 interface (Haering *et al.*, 2002). Therefore, after the addition of BMOE in cycling cells, if the mutated Scc1 is still able to form a complete ring, the two cysteine residues in Smc3 and Scc1(Δ152-256aa) should crosslink. After immuno-precipitation of Scc1-HA3, the presence of the high molecular weight crosslinked protein was analysed by Western blot. The mutation V137K was also analysed, in parallel.

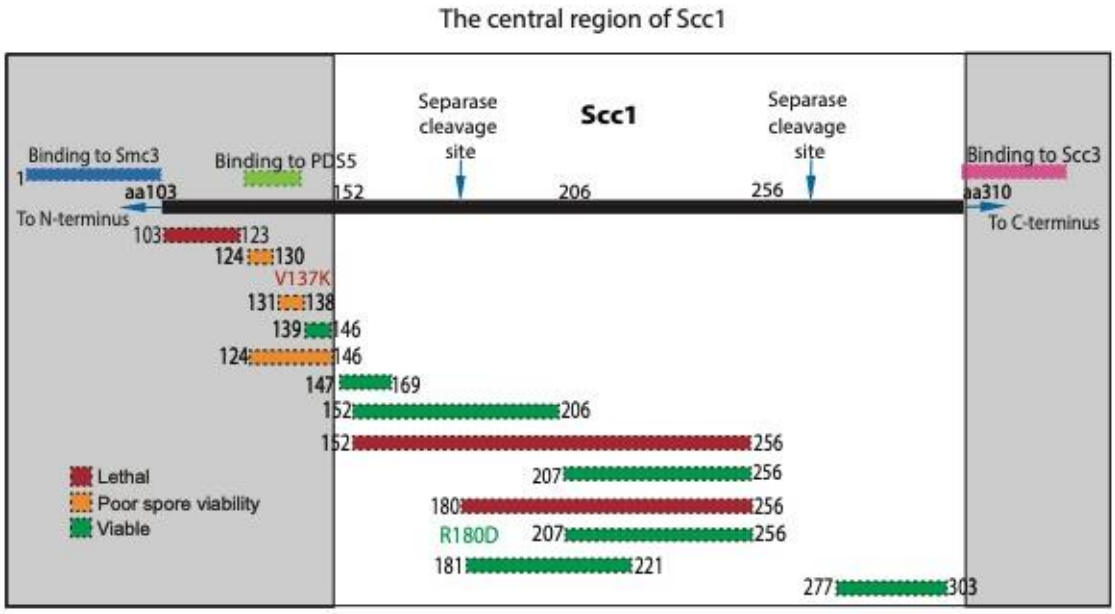


Figure 2-9. Schematic of scc1 deletions previously tested in the central region of SCC1

Summary of previous viability study of the central region of Scc1 Deletions to be investigated for Scc2 binding. (Data from Dr. Chris Chan, Personal communication)

Higher molecular weight cross linking was observed only in the presence of both cysteines and BMOE, in Scc1V137K and Scc1Δ152-256 (Figure 2-10A). This suggests that the large deletion of Scc1 does not impede on the closure of Smc3-Scc1 interface and therefore, probably does not induce lethality merely because of Scc1 length constraint.

Following this result, ChIP sequencing was used to analyse the interaction between cohesin containing the long central deletion in Scc1 and DNA. This showed a severe reduction in the amount of DNA precipitated by Scc1Δ152-256 suggesting a defect in loading. As previously mentioned, whereas the large deletion of Scc1Δ152-256 is lethal, the two halves of this region are viable when deleted. Deletions Δ152-206 and Δ207-256 both showed increased levels of cohesin interaction with DNA compared to the WT (Figure 2-10B). However, both deletions also showed higher levels of protein expression than WT, which could account for the increase observed in ChIP sequencing (Figure 2-10C). This implies that neither Scc1Δ152-206 or Scc1Δ207-256 is individually required, but both may contain important interaction sites. These results suggest that Scc2 creates multiple contacts across the N-terminal half of Scc1 and the large deletion Scc1Δ152-256 severely reduced most of these interactions. This prevents Scc2 binding and leads to the observed severe loading defect.

With the hypothesis of multiple binding sites throughout the N-terminal half of Scc1 in mind, a further deletion strain was designed, covering both the Pds5 binding region and the first half of the large central deletion. Whilst Scc1Δ124-206 was expected to be lethal and have similar defect to V137K, ChIP sequencing showed

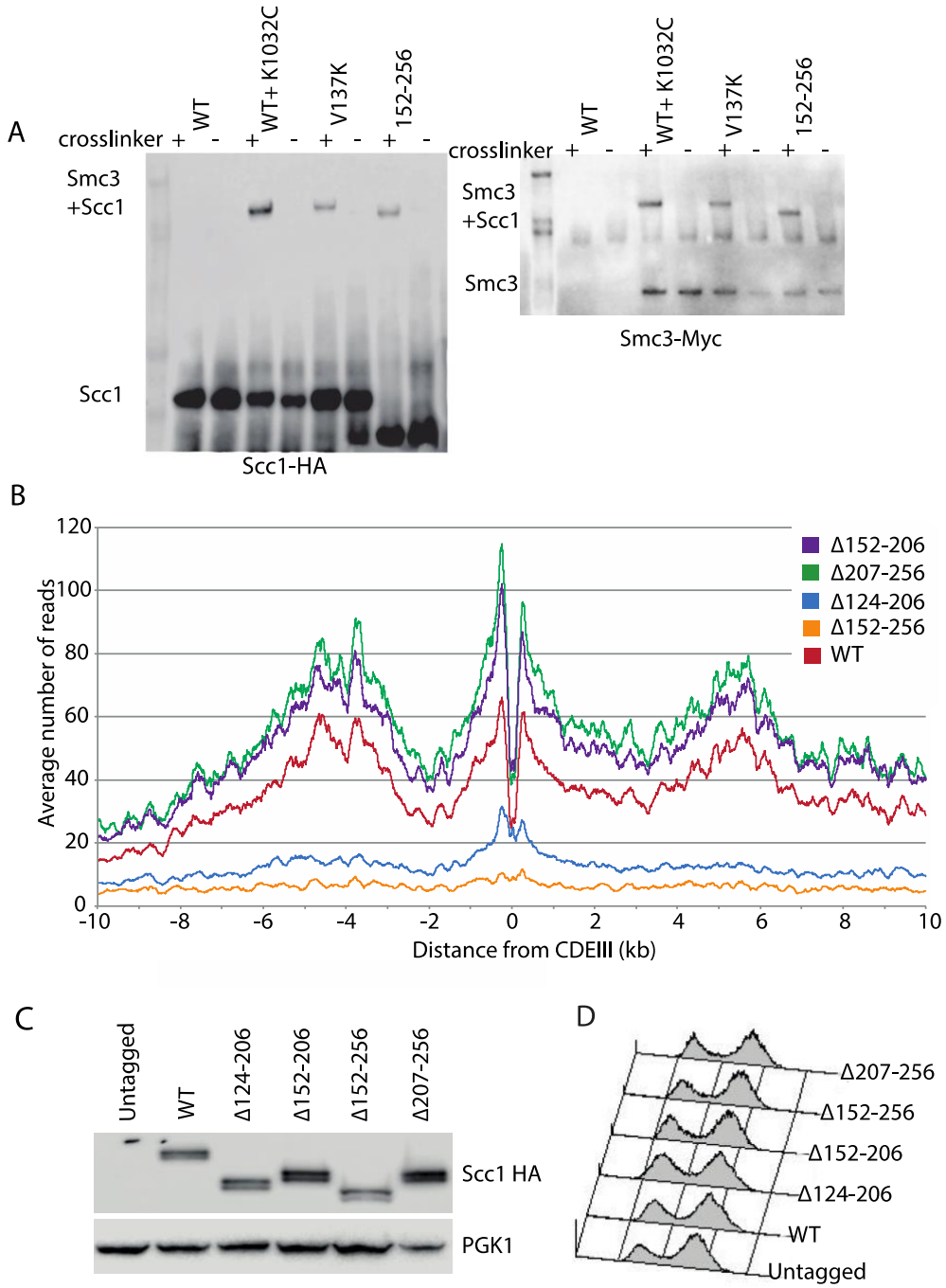


Figure 2-10. Cohesin with the large deletion of SCC1 is able to form a ring, but shows a reduced binding to SCC2

A. *In vitro* cysteine-cysteine cross linking between Smc3(K1032C) and Scc1(C56) using BMOE. Cycling cells grown in YPD at 25°C. Immunoprecipitation of HA tagged scc1. *SCC1-HA* (17184), K1032C (26189), V137K (26191) and 152-256 (26149), (repeated x2). B. Calibrated ChIP sequencing profile showing average distribution of cohesin around centromeres, 10 kb either side of the CDEIII. Strains K25995, K25997, K25896, K26103, K17184, (repeated x2). C. Expression levels of scc1 mutants tagged with HA. Pgk1 was used as a loading control. D. Flow cytometry analysis of cycling cells expressing scc1-HA3, grown at 25°C in YPD.

severely reduced levels of cohesin interaction with DNA that only retains a small amount of loading proximal to the centromere (Figure 2-10B). Given the aforementioned experiment performed with the deletion of Scc1Δ124-146 which showed wild type levels around the centromere (figure 2-6), this suggests that the region between Δ146-206aa is very important for loading.

Given the essential role of Scc2 in ATPase activity stimulation, the large central deletion of Scc1 (152-256) was tested and showed a reduced ATPase activity especially in the absence of DNA (Figure 2-11A). Finally, the ability of Scc1Δ152-256 to interact with Scc2 was then addressed by GST pull down assay using *in vitro* purified proteins. This showed that the interaction of Scc2 with Scc1 was greatly reduced after deletion of amino acids 152-256 (Figure 2-11B).

2.5 Cohesin loading can be abolished by combining small deletions of Scc1

The fact that the two deletions Scc1Δ152-206 and 207-256, are viable on their own but lethal when combined suggests that several essential scc2 binding motifs are present in each half of this region. In line with this hypothesis, the next step was to identify conserved motifs with in the Scc1 152-206 and 207-256 domains, that were lethal once removed respectively from Scc1Δ207-256 and Scc1Δ152-206.

With the exception of the N and C termini, Scc1 is poorly conserved throughout evolution. However, using an alignment of Scc1 throughout yeasts and ascomycetes, it becomes apparent that there are several highly conserved domains. Two of these domains are 164-172 aa and 244-251 aa (Figure 2-12). These regions were combined with Δ207-256 and Δ152-206 respectively.

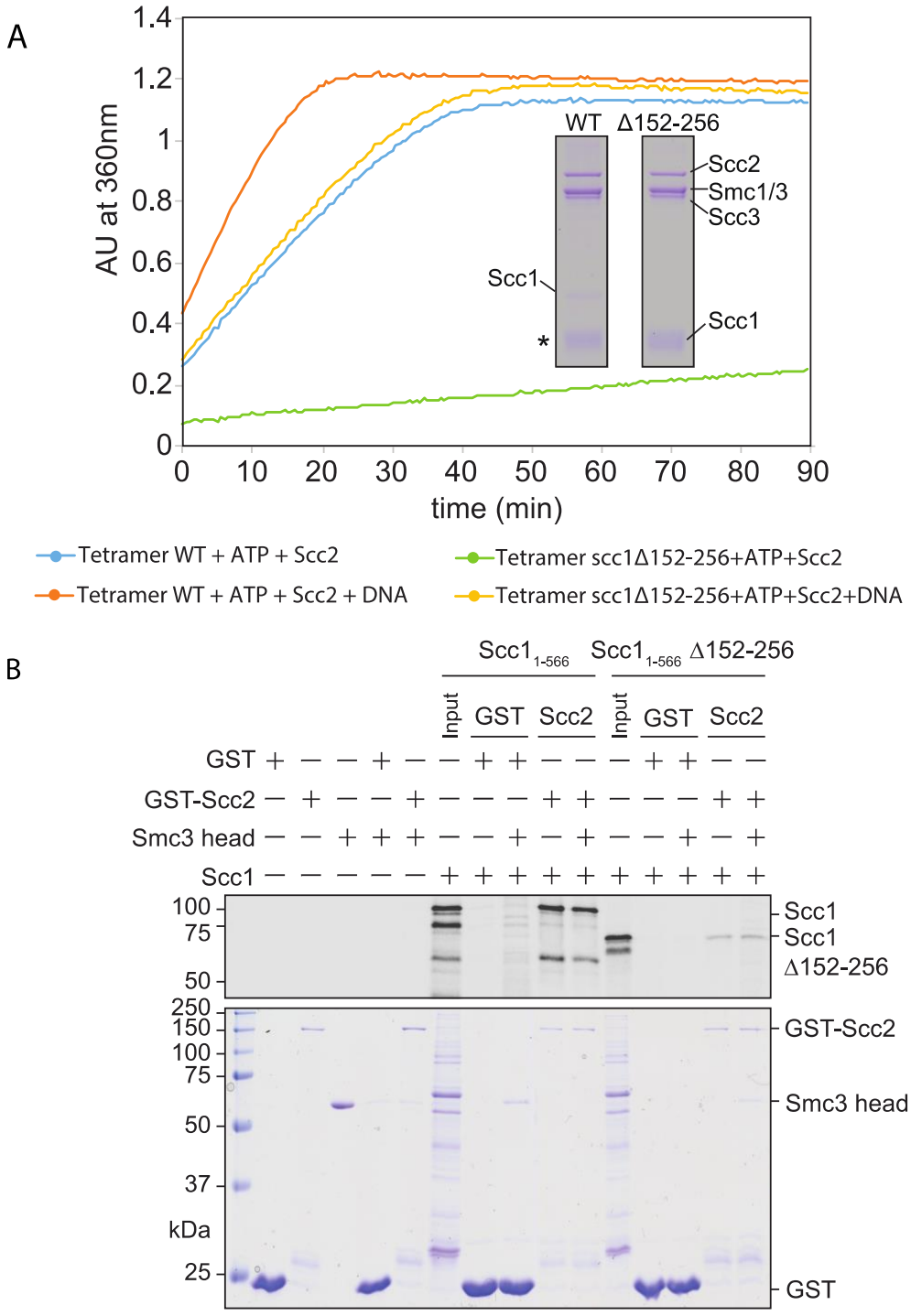


Figure 2.11. Cohesin with the large deletion of SCC1 abolishes its interaction with DNA

A. Effect of Scc1Δ152-256 on cohesin tetramer ATPase activity (repeated x3) experiment performed by Dr T Gligoris and Mr M Voulgaris.

B. Glutathione S-transferase (GST) or GST-Scc2$^{171–1504}$ proteins were immobilized on Glutathione Sepharose beads. Beads were incubated with 35S-labeled WT Scc1$^{1–566}$ or Scc1$^{\Delta152–256}$ in the presence or absence of Smc3head. Input and bound proteins were separated by SDS-PAGE, stained with Coomassie (bottom), and analysed with a phosphorimager, experiment performed by Dr S Kikuchi.

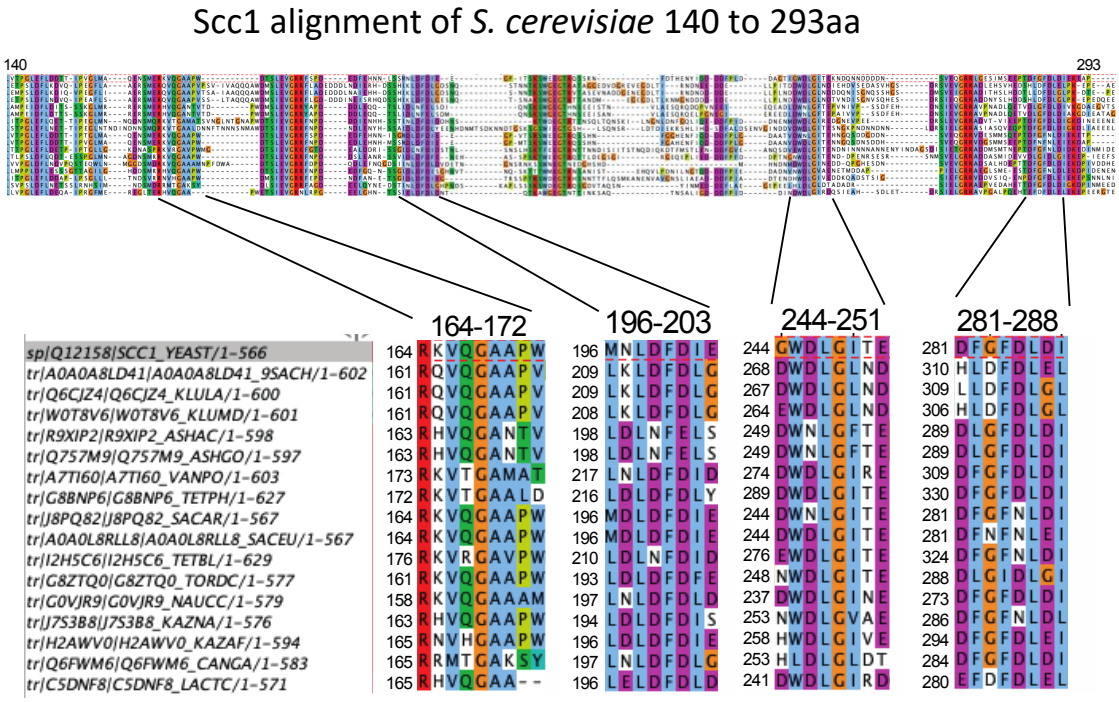


Figure 2-12. Alignment of Scc1 highlights conserved motifs

Top: Multiple sequence alignment showing the conservation of Scc1 among yeasts with point centromeres, residues 140-293aa. Bottom: Conserved residues in yeasts 164-172aa, 196-203aa, 244-251aa, 281-288aa of *S. cerevisiae*.

Scc1 Δ164-172+Δ207-256 was found to be viable and loaded cohesin at similar levels to WT, suggesting that these regions are not essential (Figure 2-13A). However, Scc1Δ152-206 + Δ244-251 was lethal and showed reduced levels of cohesin occupancy on DNA with very low levels proximal to the centromere and at the pericentromeric region (Figure 2-13A & C). Expression of both Scc1 deletions were comparable to WT levels as was the flow cytometry analysis of cycling cells (Figure 2-13B & D). This data suggests that 244-251 is one of the essential motifs present in the second half deletion 207-256 of Scc1, that is crucial for cohesin loading on to DNA. The ATPase activity of this mutant was tested and was shown to be at a similar level to the large deletion (Figure 2-13E).

Two other conserved motifs similar in sequence to 244-251aa were identified: 196-204aa and 280-288aa (Figure 2-12). Both show a similar repeating pattern of hydrophobic/acidic residues, similar to the lethal mutation Scc1Δ244-251.

These three small regions in Scc1 were created as scc1 deletions independently and tested for viability. All three were viable as single mutations (data not shown). When combined into double mutations, two were viable (Scc1Δ196-203+Δ281-288 and Scc1Δ244-251+Δ281-288) but one combination showed poor spore viability and slow growing colonies (Scc1Δ196-203+Δ244-251). Combining all three deletions caused lethality (Figure 2-14A). Both the double and triple deletion showed comparable expression levels to each other, and higher than WT, though in this instance the expression of the WT seemed lower than expected (Figure 2-14B). The slow growing double (Scc1Δ196-203+Δ244-251) and triple deletion (Scc1Δ196-203+Δ244-251+Δ281-288) were then tested for their ability to load cohesin (Figure 2-

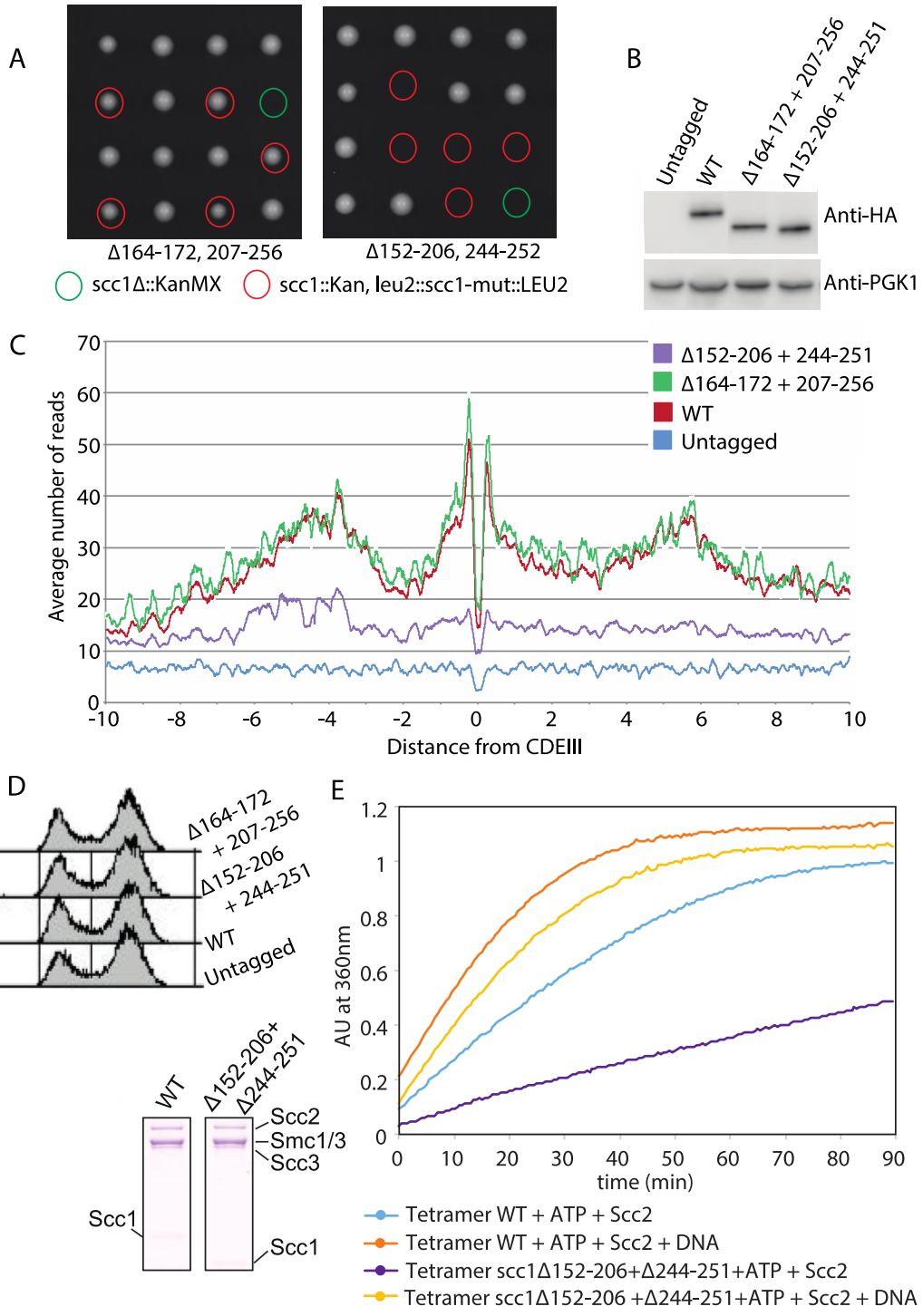


Figure 2-13. A conserved region 244-251 is important for the interaction with SCC2
A. Dissection of tetrads showing strain viability strains K26360 and K26362 x K13898. B. Expression levels scc1 deletions tagged with HA3. Pgk1 was used as a loading control. C. Calibrated average centromere profile covering 10 kb either side of CDEIII for strains K26360, K26361, K26362, K17184 and K699. Cycling cells grown at 25°C in YPD. D. Flow cytometry profile of cycling cells used for the ChIP sequencing experiment. E. Effect on ATPase assay of Scc1Δ152-206+Δ244-251. This experiment performed by T. Gligoris/M. Voulgaris, repeated x3).

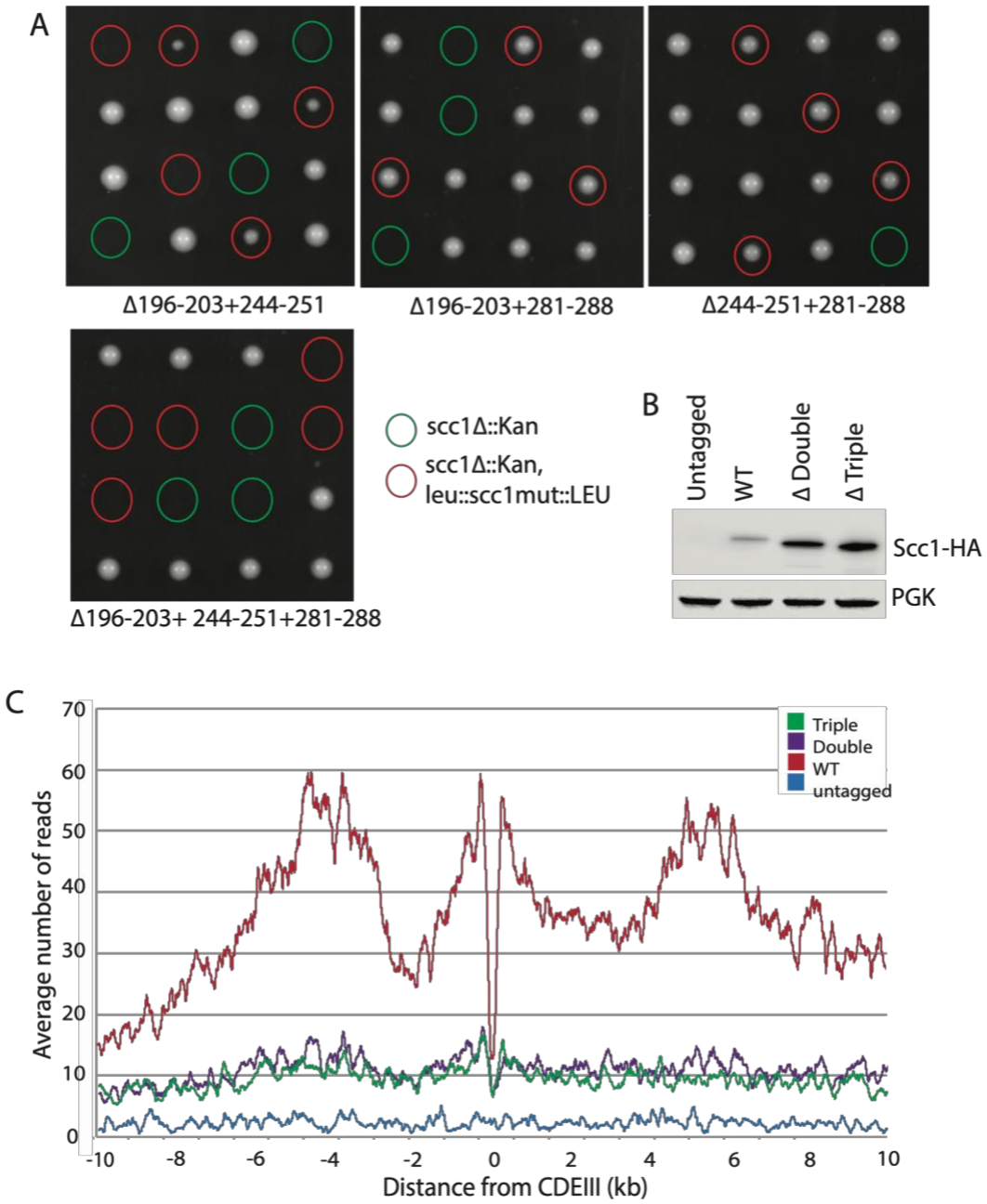


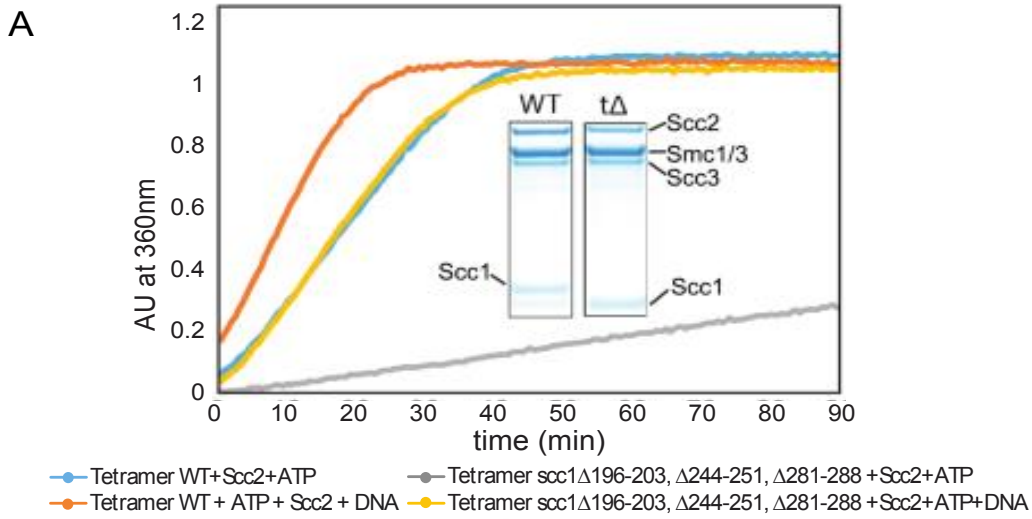
Figure 2-14. Combining the deletions of Scc1 motifs abolishes cohesin loading
A. Dissection plates of double and triple deletions showing viability. Strains K27115, K26989, K26991, K26992 x K13688. B. Expression test of double and Triple deletions. C. Calibrated average ChIP-seq profile of cycling cells K27115, K26992, K17184, K699 (experiment repeated x2).

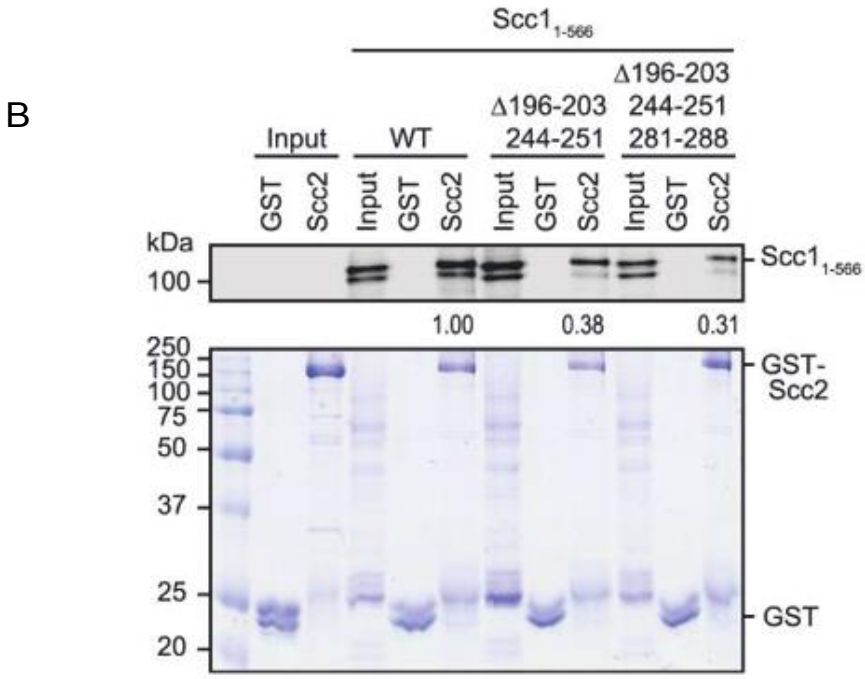
14C). Both double and triple mutations showed severe loading defects, with very little loading proximal to the centromere, but still above background, confirming that Scc2 must have multiple binding sites across the N-terminal half of Scc1. The triple deletion also showed a reduction in Scc2 dependant ATPase activity *in vitro* (Figure 2-15A).

Working with our collaborator, these small deletions were then tested for *in vitro* binding to Scc2 by GST pull down. The double mutant (Scc1Δ196-203+Δ244-251) reduced binding by 62%. The triple deletion (Scc1Δ196-203+Δ244-251+Δ281-288) further reduced binding to 69% compared to WT (Figure 2-15B). This clearly shows that Scc2 binds across a wide region of Scc1. Interestingly, it was noted when comparing the sequences of *S. cerevisiae* and *C. theromophilum* that a similar, but longer sequence of alternating hydrophobic residues can be identified in the latter organism, residues (Rad21L189-D200) (Figure 2-15C). GST pulldown of Scc2 revealed a reduction in binding to Scc1 suggesting that this longer LD motif is essential for an interaction between Scc1 and Scc2 in *C. thermophilum (Data not shown) (Satoro Kikuchi, personal communication).*

2.6 Residues within Scc1 crosslink to Scc2 using BPA

As previously shown above, *in vivo* crosslinking can be used to study protein-protein interactions and their ability to bind in different conditions such as with different mutations. It allows the analysis of binding compatibility between different protein partners, for example the competition between Scc2 and Pcc5 for Scc1 binding.





C Scc1 *S. cerevisiae* 196-203 - MNLDFDIE
244-251 - GWDLGITE
281-288 – DFGFDLDI
Rad21 *C. thermophilum* -L189-D200- LGLELDFGMDLD

Figure 2-15. Functional characterisation of the triple deletion in SCC1
A. Effect of the triple deletion *scc1Δ196-203, Δ244-251, Δ281-288* on the ATPase activity This experiment performed by T. Gligoris/M. Voulgaris, repeated x3 . B. GST or GST-Scc2$^{171–1504}$ proteins were immobilized on Glutathione Sepharose beads. Beads were incubated with 35S-labeled Scc11–566 double mutant or Scc1$^{1–566}$ triple mutant obtained through in vitro translation. Bound proteins were separated by SDS-PAGE, stained with Coomassie (bottom panel), and analysed with a phosphorimager (top panel). The relative binding intensities of Scc1 were quantified and indicated below the autoradiograph. This experiment performed by S. Kikuchi. C. Motifs identified in the three tested regions in *S. cerevisiae* with the similar, but longer motif from *C. thermophilum.*

Having identified three small regions within Scc1 that were shown to reduce loading, these regions were investigated further to identify specific residues that interact with Scc2. However, whilst this study had provided residues within Scc1 which could be used for crosslinking it did not address the identification of potential sites within Scc2. In the absence of any structural information on the interaction between Scc1 and Scc2 it was impossible to design a cysteine pair; therefore, p-Benzoyl-L-Phenylalanine (BPA) *in vivo* crosslinking was chosen.

The use of BPA crosslinking in Scc1 had successfully been used before, to identify binding to other subunits of the cohesin complex. An amber stop codon (TAG) was incorporated into the specific site of Scc1, using site directed mutagenesis (SDM) on an episomal plasmid, thus replacing the natural residue at the desired position. When used alongside a tRNA synthetase, which is able to read through the stop codon, and on addition of benzyl phenylalanine, an unnatural amino acid is created that can be photo-crosslinked by UV light (Chin *et al*., 2002; Chen *et al.,* 2007). BPA is able to crosslink residues that are up to 12 angstroms apart.

Each amber codon mutation was introduced in a strain containing a wild type Scc1 as an extra copy on a plasmid. The strains were first tested for viability by plating on 5-FOA media + BPA, this removes the URA plasmid containing the WT *SCC1* gene, leaving only the *scc1BPA* (Figure 2-16A). As Scc1 is essential, the cells will only support growth if the incorporation of the BPA produces a viable protein. This was the case for all the Scc1 BPA mutant strains.

To test which of the BPA residues could form a crosslink to Scc2, the experiment was performed with cycling cells grown in the presence of BPA and crosslinking was achieved by UV photo activation. The cohesin complex was

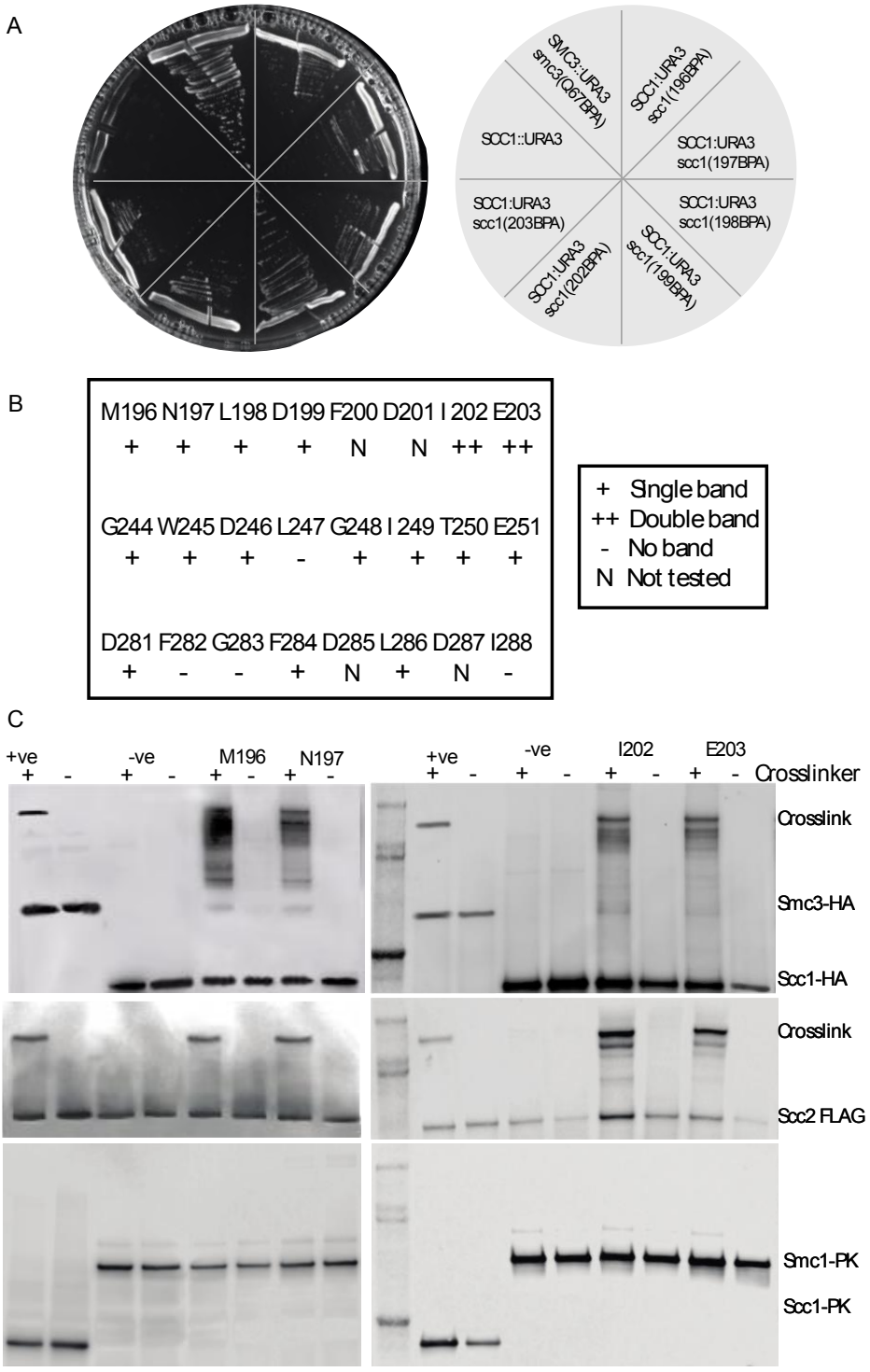


Figure 2-16. BPA crosslinking between Scc1 and Scc2

A. Region 1(196-203aa) residues as BPA, tested for viability on 5-FOA plates, strains B1821, JM002, JM003, JM004, JM005, JM006, JM007, JM022. B. Schematic of BPA crosslink results within the three Scc1 regions (Strains JM002 to JM021). C. Western blots of BPA residues *scc1M19*6 and *scc1N197* (single band for Scc2-FLAG) *scc1I202* and *scc1E203* (double band for Scc2-FLAG). Strains B1821, JM001, JM002, JM003, JM006, JM007.

immunoprecipitated using Smc1-PK and analysed by western blotting probing for scc1BPA-HA3 and Scc2-FLAG. The amber stop codon (TAG) was introduced into all positions of regions 1 (196-203), 2 (244-251) and 3(281-288). For some of the positions the mutation of the plasmid could not be obtained, after several attempts of SDM, otherwise all strains tested were viable. A summary of the results can be seen in Figure 2-16B.

Residues within all three regions showed some crosslinking to Scc2, validating the binding of these regions. The strongest individual residues were found in the 196-203 region, residues I202 and E203 (Figure 2-16C). Interestingly, these two amino acids gave two crosslinking bands of different sizes when probed with anti-FLAG Scc2.

This suggests that these residues in Scc1 may be binding to two different sites within Scc2. The two potential binding sites on Scc2 would have to be far apart to explain the difference in electrophoretic mobility. Similar differences in the size of the crosslinked bands were previously observed between Scc1 and Pds5 with BPA residues that were three residues apart in Scc1 (Chan *et al.,* 2013). This was later explained by the crystal structure that showed where Scc1 residues were binding. Two residues were binding on one side of the hook of Pds5 and the other residue, only three amino acids further along the sequence of Scc1, was binding to the opposite side of the hook. These two binding sites, whilst being very close within the Scc1 sequence were binding to very distant regions of Pds5, giving two different crosslinked species with different electrophoretic mobility. Following this hypothesis, the two crosslink bands observed with Scc1-BPA residues I202 and E203 suggest potential binding to either side of the hook of Scc2.

2.7 Scc1 E203C crosslinks to Scc2 A709C

The previous experiments identified the residues of Scc1 involved in the binding to Scc2. In order to identify the regions of Scc2 involved in this interaction, the mutations mapped in Cornelia de Lange syndrome (CdLS) were analysed. Specific positions in Scc2 are mutated in CdLs and some are known to reduce Scc1 binding (Kikuchi *et al,* 2016). The residues were mapped onto the structure of Scc2 and lay mostly on the internal edge of the protein. Using this information and supported by the idea that Scc1E203BPA could bind to either side of the hook, a small number of outward facing residues were identified in Scc2 to mutate to cysteine (Figure 2-17A). These changes were all found to be viable. Scc1E203C was also viable and was crossed to each of the Scc2 cysteine strains. These strains were tested to address the possibility of crosslinking between Scc1E203 and the different positions in Scc2. The only positive result found was Scc2A709C (Figure 2-17B). The crosslinking efficiency of this interaction was between 8-11% of the non-crosslinked band, making it too weak to be used for further experiments. Immunoprecipitation of Scc1-HA3 did not improve the efficiency. However, Scc2A709 is positioned directly opposite the C-terminus of Scc2 on the other side of the hook. This raises the possibility that Scc1E203C could crosslink to either A709 or the C-terminus of Scc2. This could explain the two different crosslinked species observed, as the two species would be differently shaped and migrate separately. However, this would require further investigation. This crosslink does however, give us an exact position where both proteins interact with each other.

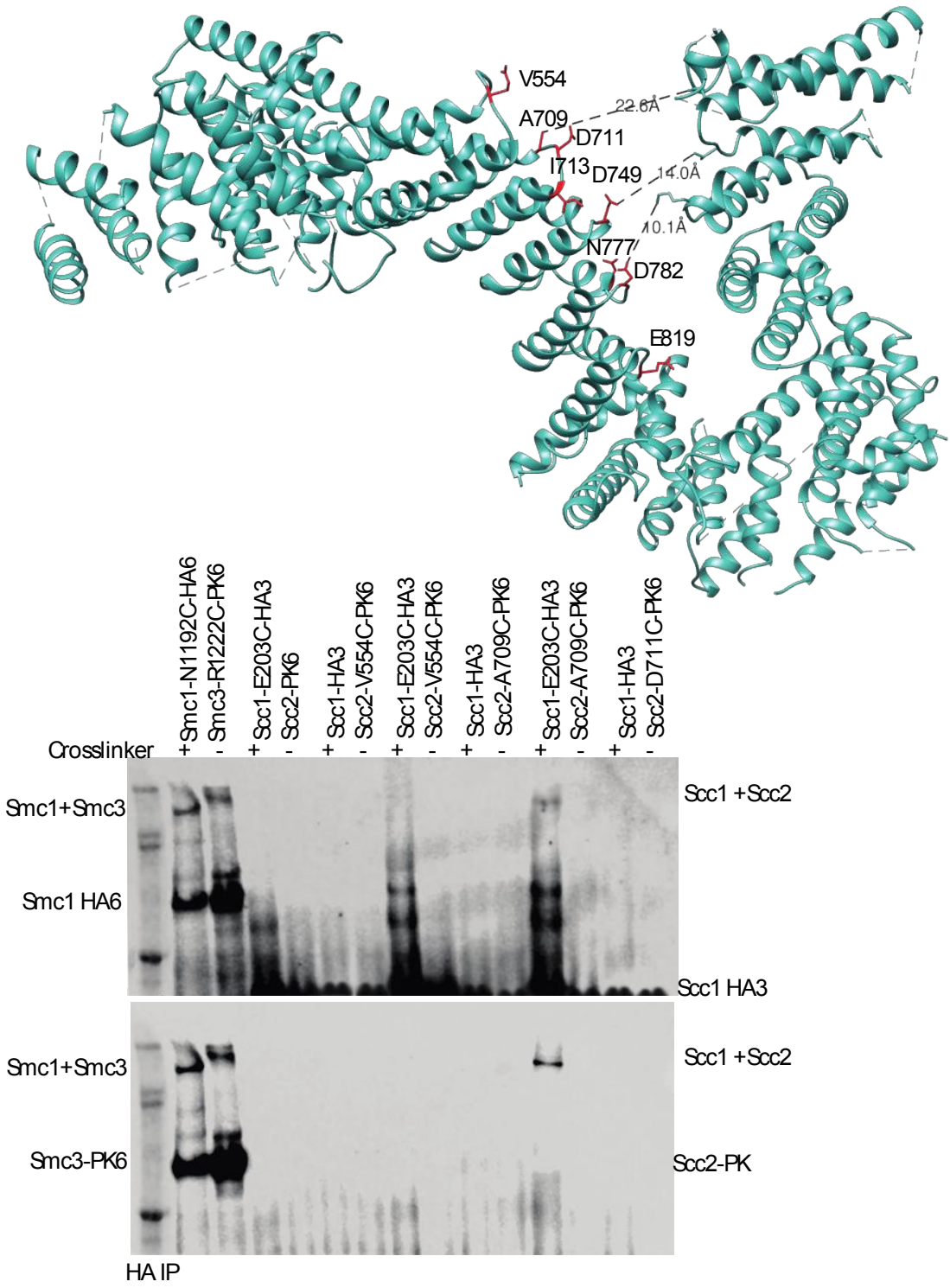


Figure 2-17. A cysteine-cysteine crosslink pair between Scc1 and Scc2
Top: Structure of Scc2 with residues to be changed to cysteine and test for crosslinking (red) and distances across the hook (black with dotted line between residues). Middle: Western blot using HA3 tag for IP, then probed for Scc1-HA3 (positive control Smc1-HA6). Bottom: Blot stripped and re-probed with anti PK, for Scc2 (positive control Smc3-PK6), Strains shown: K24455, JM031, JM033, JM032, JM035, JM034, JM037.

Discussion

Analysing how Scc2 interacts with the cohesin ring is a necessary step to understand the mechanism of loading, ATPase stimulation and translocation. This study has identified three small regions of Scc1 that are required for Scc2 binding. The deletion of Scc1 residues that form these three motifs (196-203, 244-251, 280-288) severely reduces cohesin occupancy on DNA. The interaction of this triple mutant with Scc2 as measured by GST pull down is also reduced down to 30%. The Scc2 dependant ATPase activity is also reduced, however, it is still stimulated on addition of DNA albeit, to a lower level than WT. This is suggestive that the regions identified may not be the only contacts within Scc1. Combining these three regions causes lethality, even though they do not completely deplete binding or ATPase activity. Further sequence analysis of the Pds5 binding region shows a small repeat motif containing leucine-aspartate (LD) at position 144-148 aa, this could be a candidate for binding as deleting the 124-206 severely reduced the occupancy of cohesin.

These sites identified in the central region of Scc1 do not overlap with the previously identified Pds5 binding sites. The results from Scc1V137K show that this residue of Scc1 is essential for Pds5 recruitment on cohesin but it does not mean that Pds5 does not interact with other regions of Scc1 like the central region, as this was never tested. However, it is unlikely that Pds5 can bind across the central region of Scc1, as it would potentially block the separase cleavage sites at positions 180 and 268 which would prevent cohesin cleavage in anaphase.

The Scc1V137K mutant, which does not interact with Pds5 is not detected at the convergent genes in the pericentromeric regions where cohesive sites have been

described. This suggests that during the process of loop extrusion the replacement of Scc2 by Pds5 is crucial for loop stabilization and cohesion. Surprisingly, the presence of a wild type allele of Scc1 in the cell, can somehow stabilise the cohesin complex associated with Scc1V137K at the pericentromere. This indicates that during the loop extrusion process, a transcomplementation mechanism can happen, between two cohesin complexes. However, this conclusion was based on two experiments, the first completed in cycling cells (Scc1V137K over wild type) and the second in G2 cells (MET-Scc1, Scc1V137K). This leads to the presence of two variables: the presence or absence of WT Scc1 and the cell cycle stage. Based on the analysis of other mutants (*Wapl⊿, Pds5-AID*), it has been shown that the amount of Scc1 detected on the DNA can vary considerably between G1 and G2 (Petela *et al.,* 2018 Chapard *et al.,* 2019). Therefore, these experiments should be repeated comparing G2 cells only. If this result can be confirmed, this suggests that Scc1V137K mutant cannot be stabilised at the pericentromeric sites due to its inability to interact with Pds5. In the presence of Scc1 wild type complexes, the stable structures they generate could somehow act as a road block, triggering the enrichment of Scc1V137K. However, the nature of these stable 'road blocks' would need further investigation to understand whether they are due to the formation of stable loops or due to cohesive complexes. Interestingly, a similar complementation effect was observed for cohesion. In vivo, cohesed minichromosomes detected as CDs could be mediated by Scc1V137K only in the presence of a wild type Scc1 without which only CM could be observed (Srinivasan *et al*., 2018). This implies that the enrichment of Scc1V173K in the presence of WT, is due to the collaboration of rings during the establishment of cohesion. However,

further investigation would be needed to look at the interplay between the two different types of rings.

Sequence comparison in yeast, of the three central binding motifs in Scc1 showed a similar composition of LD repeats. On further investigation of the conservation of the N-terminal region, it was observed that similar motifs could be identified in two patches just outside the Pds5 binding region. As more Scc2 binding sites are predicted to exist, addressing the functional implication of these additional sites might enable the complete abolition of Scc2 binding and ATPase activity.

Interestingly, a lysine-lysine crosslinking study in *S. cerevisiae* identified Scc1 amino acid position 165 as an interaction site with Scc2 (Burmann *et al*., 2019). This motif aligns with the binding region previously identified in *S. pombe* 145-152aa (Muryama *et al.,* 2014). The role of this sequence, that does not contain an LD motif, was tested but did not appear to be essential. However, it was only addressed in combination with 207-256 that does not include the first and last of the three LD regions identified in this study and should be readdressed in combination with the three central LD motifs.

The design of cross linking positions based on the three motifs identified should help to develop new tools to analyse Scc2 interaction with Scc1. The first approach based on the insertion of BPA residues confirmed the interaction between Scc1 and Scc2 throughout the three individual regions identified. The second approach using cysteine-cysteine showed only a weak crosslink between Scc2 binding to Scc1. A recently identified double mutation (*scc2E822K, L937F*), was shown to increase the loading activity and seemed to be a good candidate that could improve the crosslink

reaction efficiency between Scc2 and Scc1 (Petela *et al*., 2018). However, no modification in the crosslinking efficiency was observed (data not shown).

There could be two reasons to explain the weakness of this crosslinking reaction. Either the amino acid positions designed are not strictly facing each other in the 3D structure of the complex or are too far apart to crosslink efficiently. This seems unlikely given the distance and residue position observed in the crystal structure. An alternative explanation could be due to the apparent transient nature of Scc1-Scc2 interaction, based on ChIP sequencing data and FRAP experiments in yeast which showed that the residence time of Scc2 on DNA was measured at 2-4 seconds (Hu *et al.,* 2011). The dynamic behaviour of Scc2 suggests a possible transient interaction with the complex, it is therefore possible that the two proteins cannot be crosslinked efficiently using this approach. A last explanation could be that the residues chosen were not ideal to measure this interaction, as no crystal structure was available for Scc1 at the time. However, the recent publication of the *H. sapiens* cryo EM structure has shown density for a small section of Rad21 with NipBL. In *S. cerevisiae* residue Scc2A709 aligns with *H. sapiens* NipBLG1782. This structure reveals this residue could interact with Rad21 I154-E156. However, as there is little conservation through the evolution of Scc1, beyond the N and C termini, it is difficult to reconcile the difference in the distance between *H. sapiens* Rad21 (I154- E156) and *S. cerevisiae* Scc1E203. Further analysis would be required to identify potential residues with a higher crosslink efficiency.

During the characterization of the Scc1 motifs essential for Scc2 binding, three conserved LD motifs were identified. Curiously, while the ChIP sequencing profile of the lethal triple LD mutant shows a drastic reduction in loading, the ATPase activity

appears to be severely altered only in absence of DNA. In the presence of DNA, the ATPase activity of the triple mutant is similar to the level of Wild type cohesion without DNA. This suggests that the binding of DNA could stabilize the interaction of Scc2 with the cohesin complex. In that respect, it could be interesting to measure the crosslinking efficiency between Scc1 and Scc2 *in vitro* with and without DNA. It is however surprising that this triple mutant, which has a substantial stimulating effect on the ATPase activity in the presence of DNA, is not able to sustain life. One possibility could be that the weakening of Scc2 binding with the triple Scc1 mutant, despite being stabilized by DNA could be more sensitive to Pds5 competition. It would be interesting to measure the Pds5 inhibitory effect on the ATPase activity of the triple mutant.

According to previous reports, Scc2 can interact with all four subunits of the cohesin complex. The combination of the three mutations of Scc1 induced a reduction in the ATPase activity of the complex, but did not abolish it. This suggests that Scc2 interaction with the complex either requires further motifs in Scc1 or that as previously suggested in the literature, Scc2 interacts with other cohesin subunits of the trimer. Understanding the mode of interaction between Scc2 and the cohesin complex in the different conformations would bring further information on how this specific subunit stimulates the ATPase activity of the complex.

Chapter 3: Analysis of Scc2 interaction with the Smc1 head

Introduction

A recent study showed that in condensin, the head of the Smc4 can interact with the C terminus of Ycs4 (Hassler *et al.,* 2019). In this study, the authors show that the binding of ATP to Smc4 releases the interaction with Ycs4 enabling the Smc2 and Smc4 heads to engage. This interaction is mediated by a specific loop in Smc4 head that contains a tryptophan residue (W-loop) and a loop in Ycs4 called the KG-loop. Mutations in the W-loop were shown to inhibit ATP hydrolysis and abolish condensin association with DNA. The SMC proteins show a high level of conservation through all species and the similarity between Smc1 (cohesin) and Smc4 (condensin) is particularly high (Hassler *et al.,* 2019). This could suggest that one of cohesins HAWK subunits may bind to the Smc1 head.

The condensin complex has two HAWK subunits, Ycs4 and Ycg1, whereas cohesin has three, Scc2, Pds5 and Scc3 (Wells *et al.*, 2017). This is compatible with the notion that the kleisin subunit can only bind to two HAWKs at a time, due to the mutual exclusivity between Scc2 and Pds5. Despite showing very similar structures, the HAWKs of cohesin and condensin complexes do not share conservation when aligned (Figure1). The Ycg1 subunit in condensin has been previously shown to bind to DNA in the curve in its backbone and it has been shown that Ycg1 interacts with the condensin Kleisin, Brn1 towards the C terminal (Kschonsak *et al*., 2017). This is similar to what was shown for Scc3.

Ycs4 has been shown to bind to Brn1 towards the N-terminal, and it has been demonstrated as having a stimulating impact on the ATPase activity of condensin

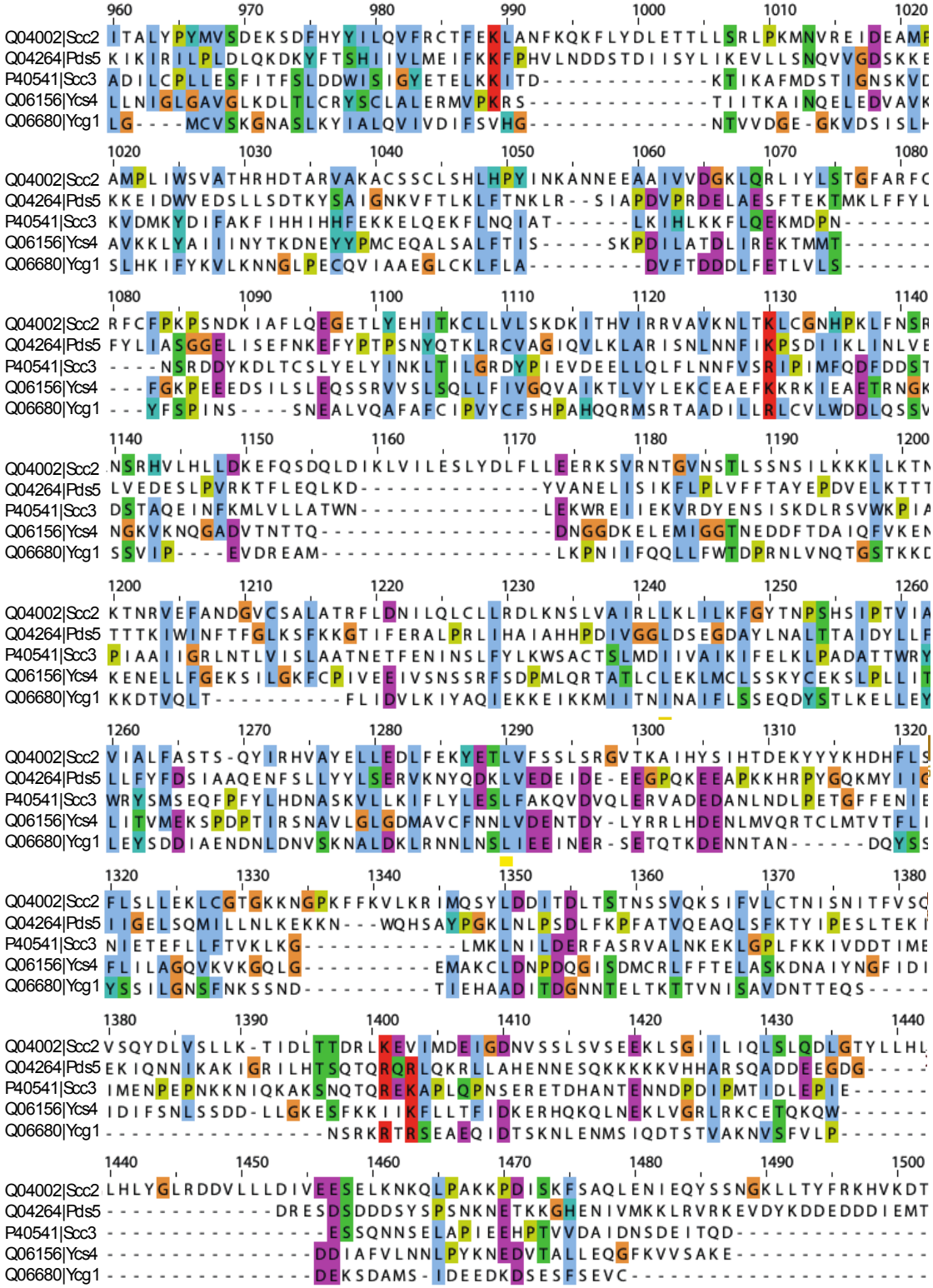


Figure 3-1. Alignment of the HAWK subunits show little conservation.
Alignment of the C-terminal half of the HAWK proteins in S. cerevisiae. Approximate Scc2 numbers used as guide. All sequences from Uniprot.org (entry name Scc2-Q04002, Pds5-Q04264, Scc3-P40541, Ycs4-Q06156, Ycg1-Q06680)

(Hassler *et al.,* 2019). This is similar to what was described for Scc2 and cohesin. Pds5, however has no stimulating effect on cohesin's ATPase activity but was shown to compete with Scc2 and inhibit ATP hydrolysis (Petela *et al.,* 2018).

More recently, our collaborators in Cambridge obtained a cryo-EM structure from *S. cerevisiae*, in which the position of Smc1 and Scc2 are similar to that of Smc4 and Ycs4 (Andres Gonzalez, personal communication). This suggests that the interaction in cohesin could be similar to that of condensin. The potential conservation of this interaction in the cohesin complex was investigated further.

3.1 Mutation of multiple residues in the F-loop of Smc1 causes lethality

To identify the corresponding residues of the F-loop in the Smc1 head, we analysed the alignment between the W-loop of Smc4 and the F-loop of Smc1 from several organisms. Each loop comprises seven amino acids, in Smc1 these are P1118, P1119, L1120, K1121, R1222, F1123 and K1124 (Figure 3-2). Based on their position on the structure, the latter five amino acids were mutated to alter either their charge or size. The viability of the corresponding strains was then assessed by yeast genetics. The F-loop mutations were introduced into the *SMC1* gene by site directed mutagenesis. The plasmid DNA was then transformed into a WT diploid yeast strain, at the endogenous locus. Single colonies were isolated and tested for integration of the mutation. Tetrads were sporulated, dissected and replica plated, to assess the viability of the mutations. All strains with the single mutations were also streaked at 37°C, to test for temperature sensitivity.

Within the condensin Smc4 head, two residues in the W-loop S1316D and

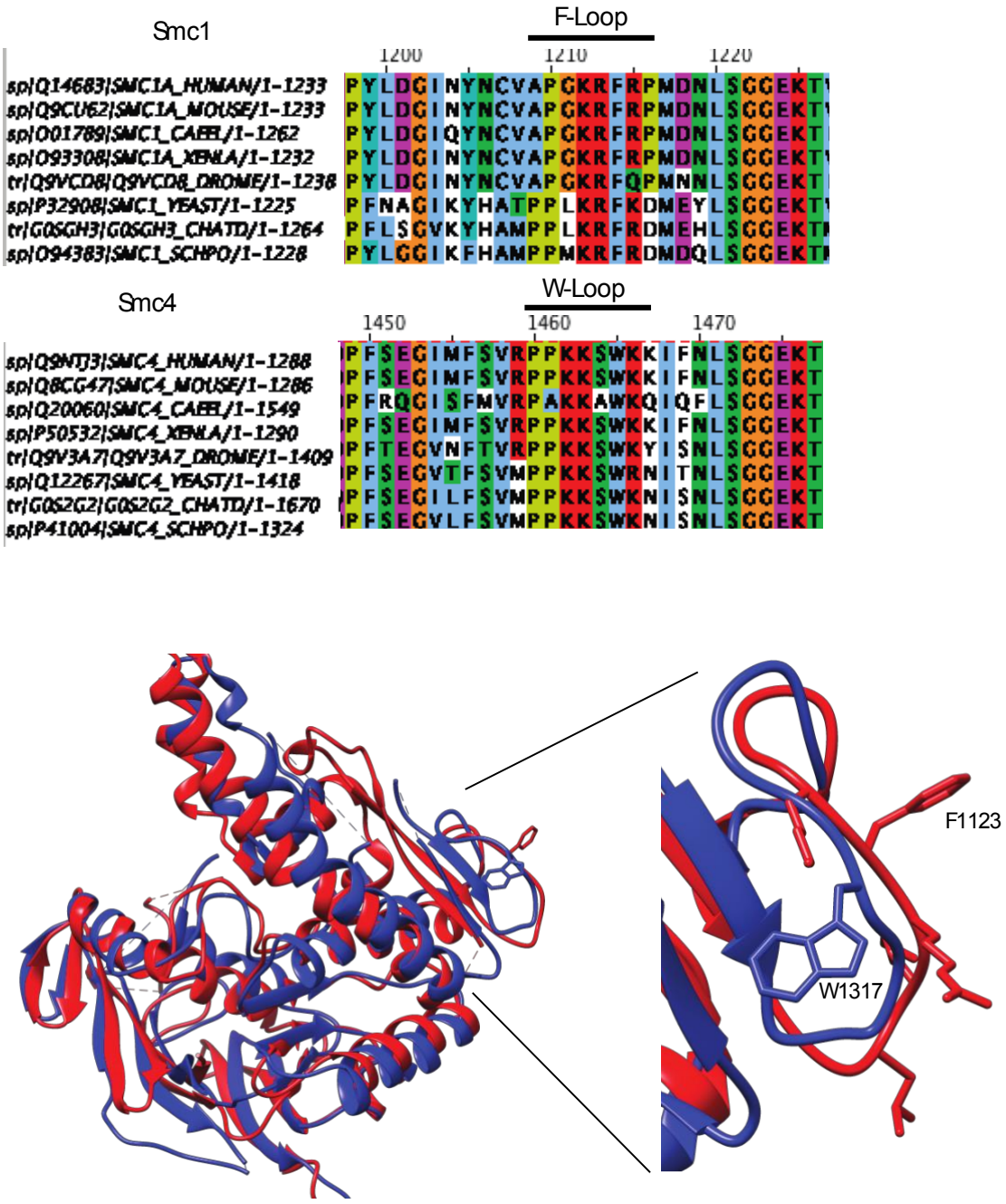


Figure 3-2. Similarity between the F-loop in Smc1 and the W-loop in Smc4
Top: Smc1 and Smc4 alignments showing conservation of F and W loops respectively. Bottom: Overlay of head structures Smc1 from cohesin in *S. cerevisiae* (in red; PDB 1w1w), and Smc4 from condensin in *C. thermophilum* (in blue; PDB 6QJ4). Right: F and W loops expanded.

W1317A were found to be lethal after charge reversal (Hassler *et al*., 2019). These are the equivalent position to R1122 and F1123 in Smc1. Unlike in the W-loop of condensin, no single mutation in the F-loop of cohesin was found to be lethal. This was unexpected given the structural similarity between the two proteins. However, the mutation of Smc1K1121D gave a phenotype of small, slow growing spores, as shown by 2:2 segregation (Figure 3-3, top middle). All other strains grew as per the WT control (Figure 3-3, top left). None of the mutations showed any temperature sensitivity at 37^{o}C data not shown). Single mutations were then combined into pairs to test for any further growth phenotype, as seen in the W-loop of condensin. The combination of Smc1K1121D with either Smc1F1123A or Smc1K1124D was lethal, as was the pairing of F1123A with K1124D (Figure 3-3, bottom panel). The combination of other F-loop mutations with K1121D did not induce any further deleterious effect other than the slow growth observed for K1121D a single mutation. (Figure 3-3, A top right and B). These results suggest that the residues in the F-loop of Smc1 are essential for its activity and these mutations were analysed further.

3.2 Mutations in the F-loop of Smc1 do not suppress the deletion of ECO1

The position of the F-loop was intriguing as three mutations in the head of Smc1 were identified during a screen looking for supressors of *eco1-1* (Elbatsh, 2016). These three mutations in the Smc1 head, are located close to the F-loop. Two of these mutations are found in the signature motif and one in the D-loop (Figure 3-4A). To address if the mutations in the F-loop are also capable of suppressing the deletion of *ECO1*, the Smc1 F-loop mutations were crossed to a strain containing *eco1Δ* and a

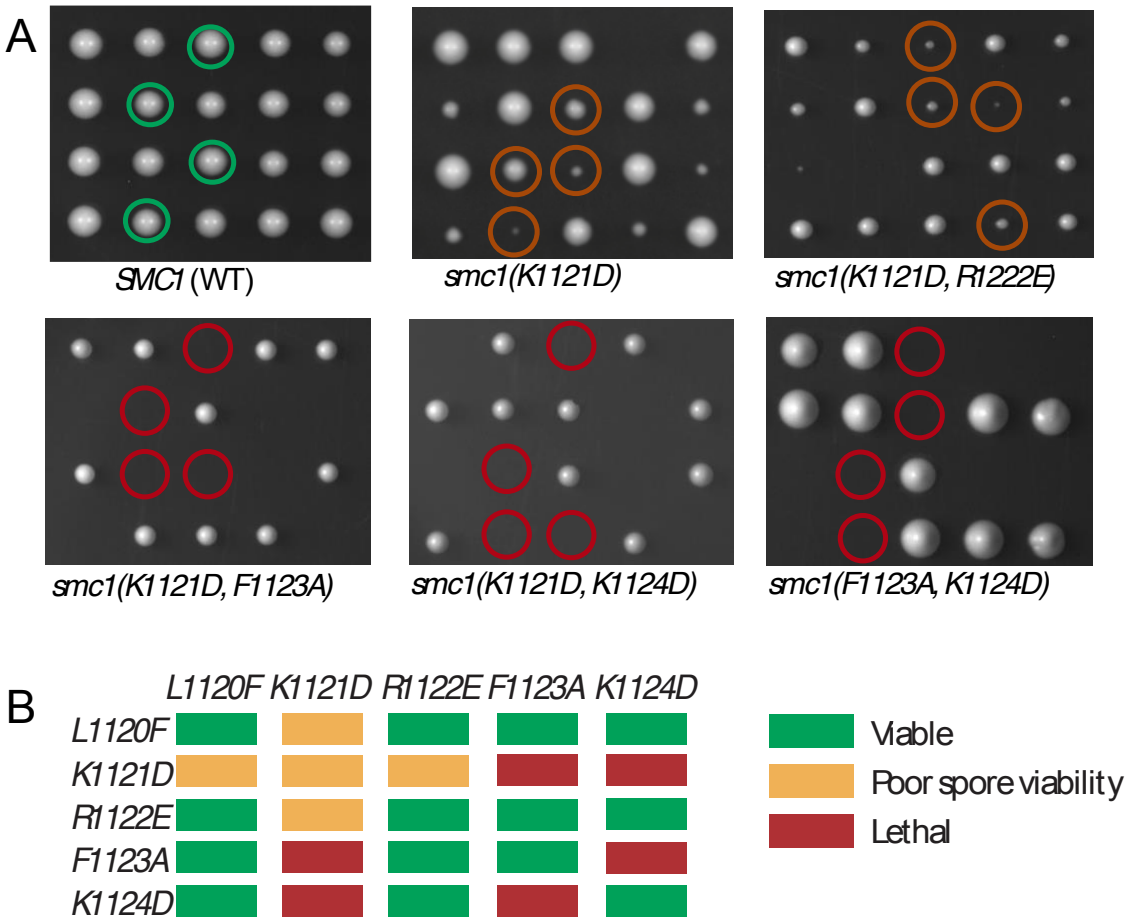


Figure 3-3. Mutating multiple F-loop residues induces lethality
A. Top: Tetrad dissections of WT (left), *smc1K1121D* mutation showing pore spore viability (middle), double mutation *smc1K1121D, R1122E* showing pore spore viability (top, right). Bottom: Double mutations showing lethality, *smc1K1121D, F1123A* (left), *smc1K1121D, K1124D* (middle) and *smc1F1123A, K1124D* (right).

B. Schematic of the viability of all single and combined F-loop mutations.

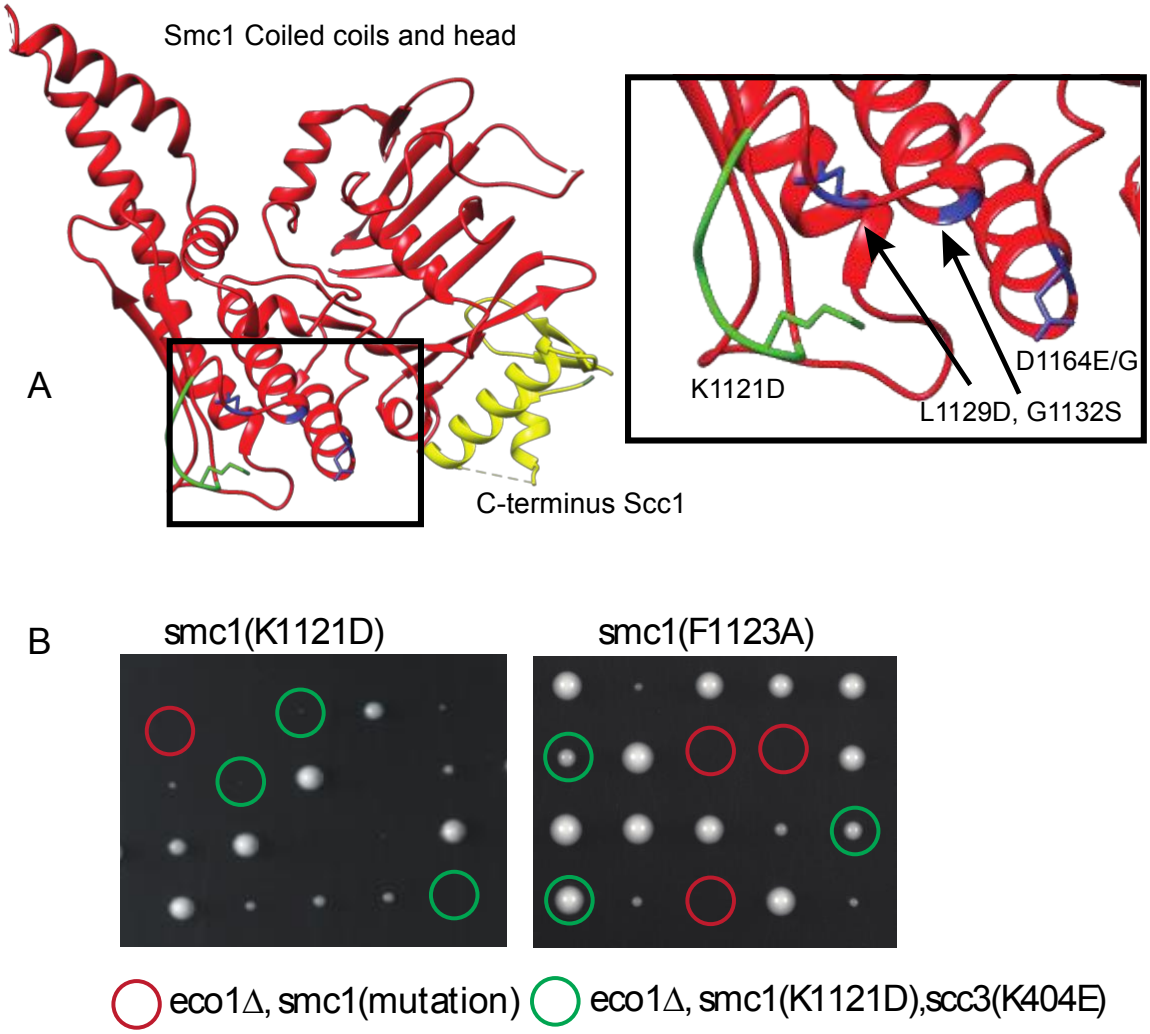


Figure 3-4. Mutations in the F-loop are not suppressors of *eco1Δ*

A. The structure of the Smc1 head of *S. cerevisiae* (in red) showing positions of known suppressor mutations of *ECO1* (in purple) and the position of the F-loop (in green). (PDB: 1w1w).

B. Plate dissection of *smc1K1121D* and *smc1F1123A* showing lack of growth where the genotype would carry both *eco1Δ* and the smc1 mutation (in red circle). Colonies also carrying the rescue mutation of *eco1Δ*: *scc3K404E* are marked with a green circle. Strains K28051 and K28053 x K24422

mutation that rescues the deletion of *ECO1, scc3K404E.* Whilst *scc3K404E* is able to restore viability to cells with *eco1Δ*, the spores grow more slowly than a WT strain. After sporulation, dissection and replica plating, the viability of spores was assessed.

None of the F-loop mutations were able to supress the lethality of *eco1Δ*, however the growth defect of *smc1K1121D* further compromised the growth of colonies carrying *eco1Δ* and *scc3K404E* (3-4B). These results show that despite the proximity of the F-loop mutations to the mutations previously identified in the signature motif and the D-loop, the F-loop mutations are not able to suppress the deletion of *ECO1*.

3.3 Mutations in the F-loop of Smc1 impact on cohesin-DNA interaction

In condensin, the mutations in the W-loop of the Smc4 head cause a severe reduction in chromosome association of condensin by ChIP sequencing. If the same is true for cohesin, one would expect the mutations in the F-loop of Smc1 to impede cohesin's association with DNA. In order to address this, the impact of the Smc1 F loop mutations on the genome occupancy of cohesin was analysed by ChIP sequencing. As two of the tested strains were lethal, all mutations and controls were introduced as a tagged copy at the *LEU2* locus. The endogenous *SMC1* gene was fused to a PK6 tag and an inducible degron (mini AID) (Nishimura *et al*., 2009). This enables the degradation of the WT protein to only visualise the effect of the F-loop mutants. The strains were grown in YPD at 25°C and synchronised in early G1 using alpha factor. Cells were released into media containing auxin and nocodazole. Samples for flow

cytometry were taken every 15 minutes from release to ensure all strains had correctly arrested, released, completed S phase, and entered G2 (Figure 3-5A). An anti PK western was performed on samples taken at release from G1 and 75 minutes after release, to ensure the WT Smc1 protein had been degraded (Figure 3-5B). Experimental cells were taken at 75 minutes post release.

A ChIP sequencing profile showing the average occupancy of cohesin around centromeres was built by aligning all 16 chromosomes by their CDEIII elements. The number of reads averaged at each base pair 60kb either side of the centromere can be visualised (Figure 3-5C). The single mutant Smc1K1121D shows a similar profile to the WT strain, though at a slightly decreased level. The decrease was more pronounced around the centromere. When the results are plotted as the percentage of WT the quantification of the decrease shows a reduction of 10% on the arms and 20% at the centromeric region (Figure 3-5D). The two lethal double mutations (K1121D, F1123A and K1121D, K1124D) have very little interaction with chromatin. A loss of around 70% on the arms and 90% at the centromeric region. The small peaks around the centromere suggests the F-loop double mutations are still able to load, and translocate away from the centromere, but at an extremely low level that is insufficient for viability. These results suggest that the F-loop is essential for cohesin's loading and interaction with chromatin.

3.4 The ATPase activity of cohesin is inhibited by the mutations in the Smc1 F-loop

If loading is defective , what aspect of the mechanism is affected? The cohesin

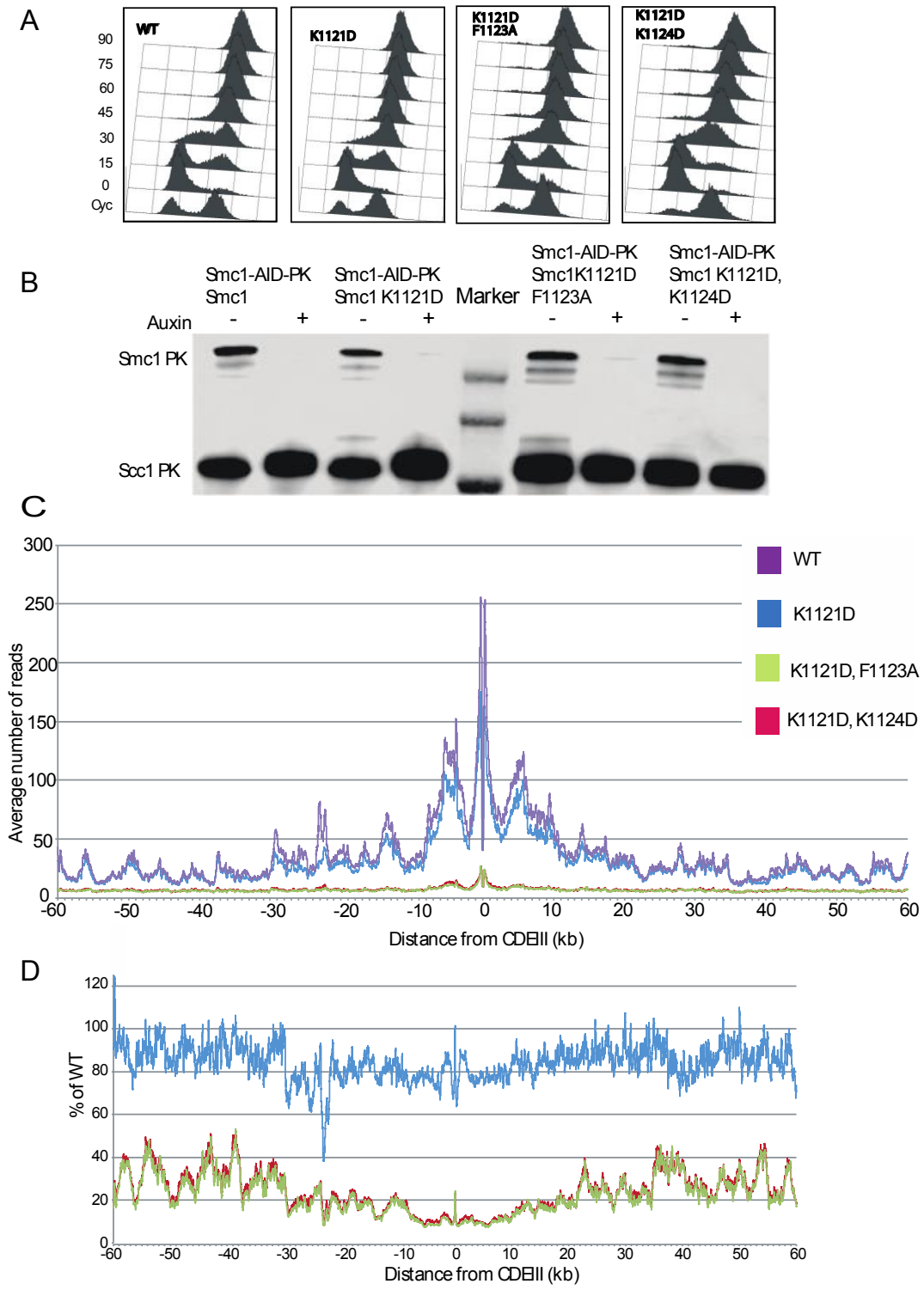


Figure 3-5. Double mutations in the F-loop abolish cohesin's interaction with DNA

A. Flow cytometry analysis of G1 arrest and release into G2 with nocodazole. B. Western analysis showing depletion of Smc1-PK6-AID +/- addition of auxin. Loading control Scc1-PK6. C. Average calibrated ChIP sequencing profile of single and double F-loop mutations. Cells were released into YPD+auxin+nocodazole after pheromone arrest in early G1. Cells were taken at 60 minutes post release. D. Percentage of WT graph of ChIP-sequencing data. (K28211, K28212, K28213, K28214).

loading reaction relies upon the engagement of the SMC heads, the ability to trap two molecules of ATP between them and then hydrolyse them. Cohesin's ATPase activity is driven by Scc2, and thus, relies on the ability of Scc2 to interact with the complex. If the binding of Scc2 to the SMC heads was compromised it may expected that ATP hydrolysis is also affected, thus producing a loading defect. The next step was then to test the effect of the F-loop mutations on the ATP hydrolysis stimulated by Scc2.

Two different mutated versions of the Smc1 F-loop were analysed: a triple mutant (K1121D, F1123A and K1124D) and the single mutant K1121D mutant. Cohesin trimers (Smc1, Smc3, and Scc1) were expressed in insect cells using the *Baculovirus* expression system and purified. Full length Scc3 and a shortened version of Scc2 lacking its N-terminal domain ($Scc2^{133–1493}$) were purified separately. This truncated version of Scc2, while no longer capable of binding to Scc4, is sufficient to activate cohesin's ATPase activity (Petela *et al*., 2018).

The ATPase activity of the mutated trimer protein was measured *in vitro*, with Scc2, with or without Scc3 or DNA. The single mutant, Smc1K1121D caused a reduction in ATP hydrolysis compared to WT, especially in the absence of DNA. The triple mutant however, was almost completely defective in ATP hydrolysis in all conditions tested (Figure 3-6).

These results reflect the ChIP sequencing results for the Smc1K1121D mutant which showed reduced levels of ATPase activity and a small reduction in loading in ChIP sequencing. The combined mutations are defective in both the loading and ATPase activity of the complex. Three possible explanations could be envisioned for

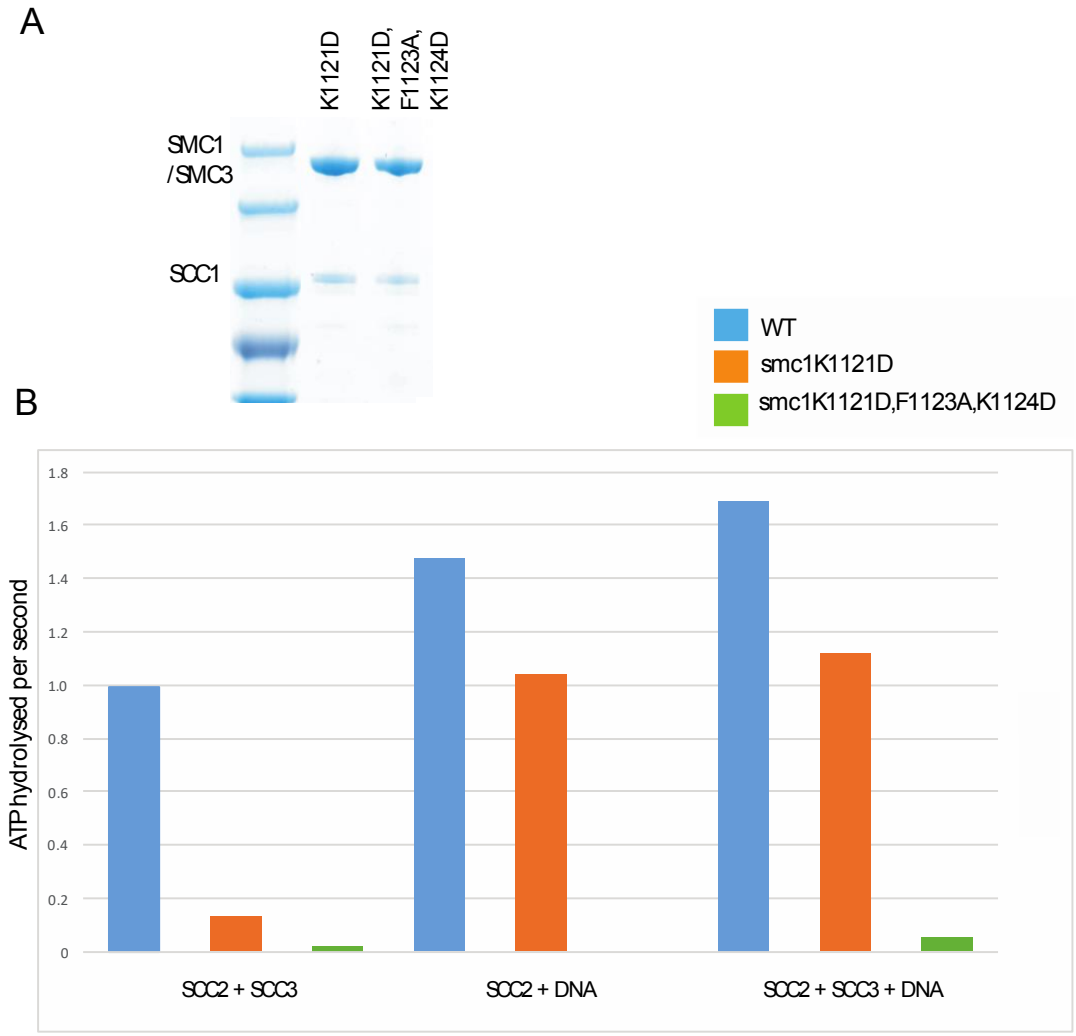


Figure 3-6. Cohesin's ATPase activity is abolished by multiple F-loop mutations
A. Purification of the cohesin trimer containing Smc1K1121D and Smc1K1121D, F1123A, K1124D mutations.
B. ATPase activity of WT cohesin, Smc1K1121D and smc1K1121D, F1123A, K1124D cohesin trimer with Scc2 and Scc3, Scc2 and DNA and Scc2, Scc3 and DNA, repeated x2.

3.5 The triple mutation in the Smc1 F-loop is defective in head engagement

In order to address if the F-loop mutations can affect head engagement, the triple mutation was combined with a cysteine pair that can be used to measure head engagement as previously described: Smc1N1192C and Smc3 R1222C (Figure 3-7A). The proteins were expressed and purified from insect cells (Figure 3-7B). The WT and mutant proteins were crosslinked *in vitro* using BMOE in the presence or absence of Scc2.

As represented in Figure 3-7, the wild type complex can adopt the head engaged state, which is enhanced by the addition of Scc2, as detected by the crosslink band. The triple mutant of Smc1 is severely defective in crosslinking (Figure 3-7C). However, the small amount of protein able to crosslink is still stimulated by the addition of Scc2 as far as we can conclude from the low levels of crosslinking. This result suggests the lethal phenotype of the F-loop triple mutation is due to a defect in head engagement, rather than its inability to interact with Scc2. This does not however, exclude the possibility of a defect in nucleotide binding, which was not tested.

3.6 Analysis of the slow growth phenotype of K1121D in Smc1 F-loop

The single mutation of Smc1K1121D is still able to hydrolyse ATP and load onto DNA. It is therefore unlikely that the severe growth defects of the cells could be due

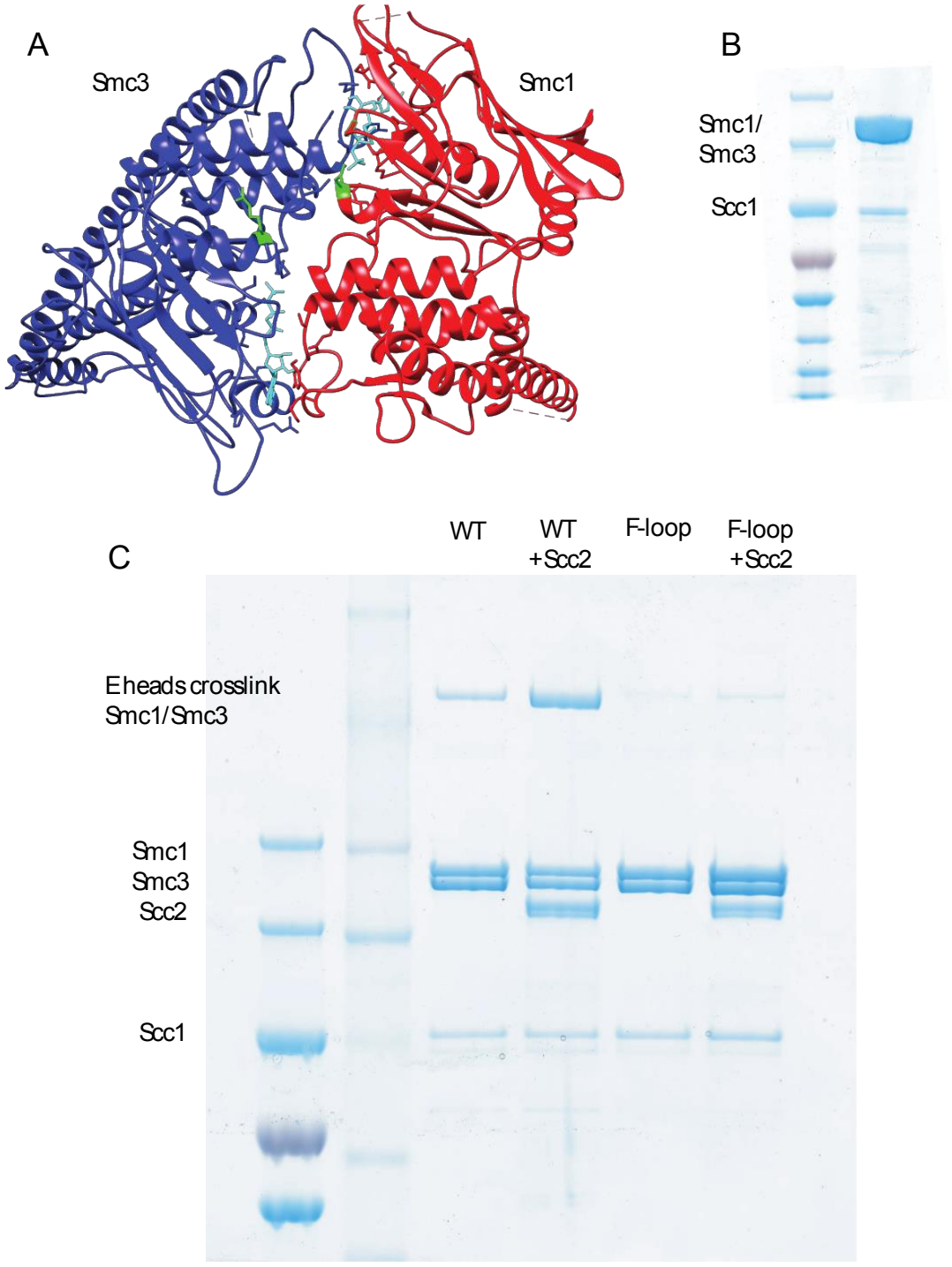


Figure 3-7. Head engagement is impeded by multiple F-loop mutations
A. The structure of Smc1 (in red) and Smc3 (in blue) heads in the engaged state with cysteine pair Smc1 N1192C and Smc3 R122C (in green) ATP molecules are highlighted in light blue) (*S. cerevisiae,* PDB 6ZZ6). B. Purification of cohesin trimer containing Smc1K1121D, F1123A, K1124D and N1192C and Smc3 R1222C mutations. C. Head engagement of WT and Smc1K1121D, F1123A, cohesin trimer containing smc1N1192C and smc3 R1222C. Tested with and without Scc2 in the presence of Mg2+ and ATP, repeated x2.

to head engagement. In order to analyse the slow growth phenotype of this mutant the cell cycle progression of the mutated strain was analysed. Flow cytometry analysis of cells expressing Smc1K1121D shows an increase in the amount of cells in G2 compared to WT which could be due to cohesion defects. (Figure 3.8A).

To investigate this potential cohesion defect, the split dot assay was used (Michaelis et al., 1997; Silver, 2009). A new strain was created containing the *smc1K1121D* mutation, Pds1-Myc9, a recombinant tet repressor protein coupled to GFP (TetR:-GFP), and a multiple integration of tet operators (tetOx336) at the URA3 locus, 36kb from the centromere of chromosome 5 (CEN5). Using immunofluorescence, the GFP signal can be followed as the tetR-GFP binds to the tetO array and appears as a fluorescent dot. After replication, in securin positive cells, the two copies of the tetO array can be followed as two GFP fluorescent dots. In WT cells a single GFP dot is observed indicating the tight embrace of cohered sister chromatids. However, in cells that are defective in cohesion the two sister chromatids are not held together as tightly. As the spindle forces pull towards the poles of the cell, two dots are frequently observed. The higher the percentage of two or 'split dots', the more severe the defect is considered to be. Pds1 (securin) can be used as marker to identify cells that have not yet gone through anaphase, as securin prevents cleavage by separase, thus the sister DNAs have not separated.

A preliminary experiment using live cell imaging instead of immunofluorescence showed that the mutant Smc1K1121D presents some split dots in early budded cells and multiple dots in some cases, whereas wild type cells contained only a single GFP dot (Figure 3.8B). As only a small number of cells were

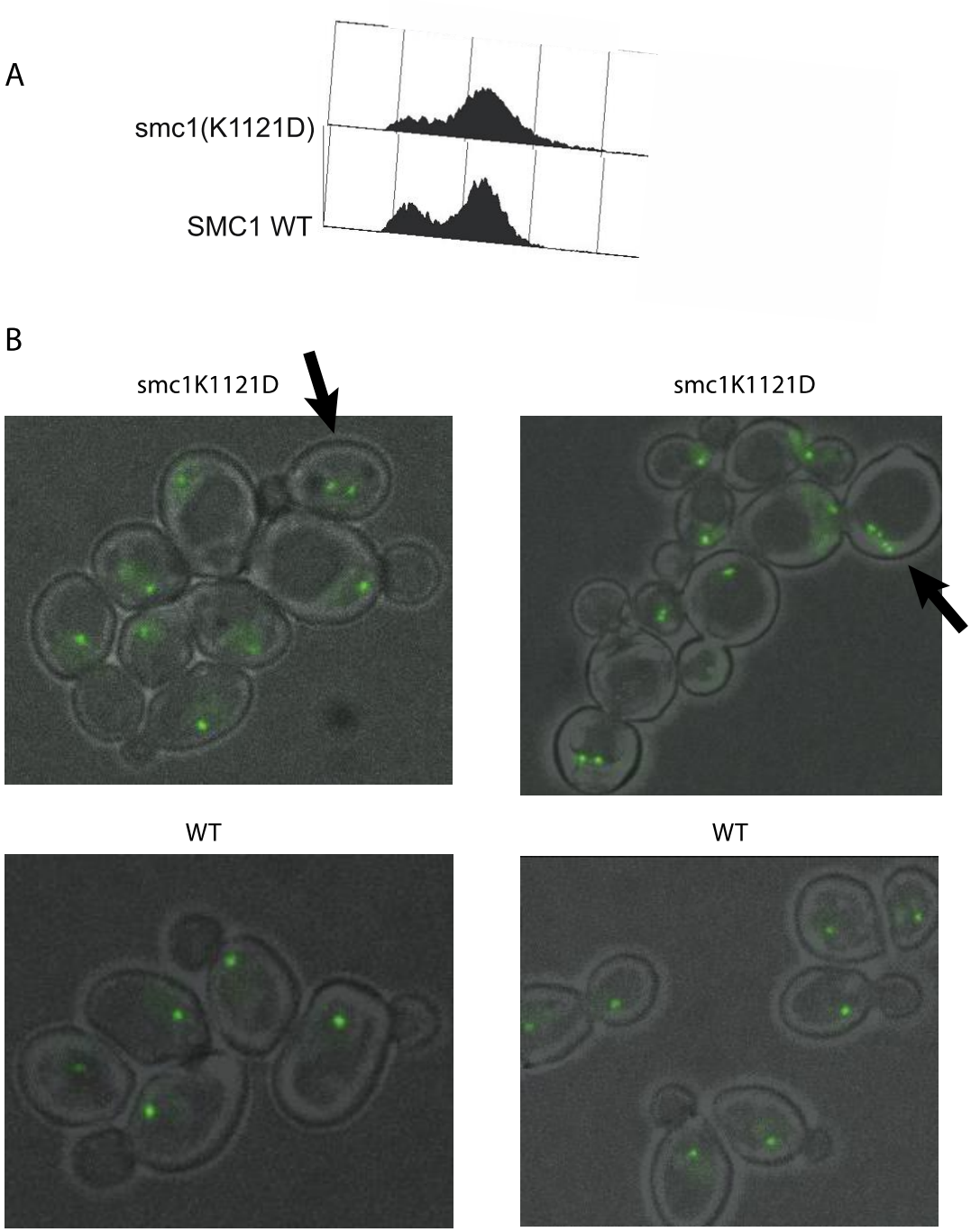
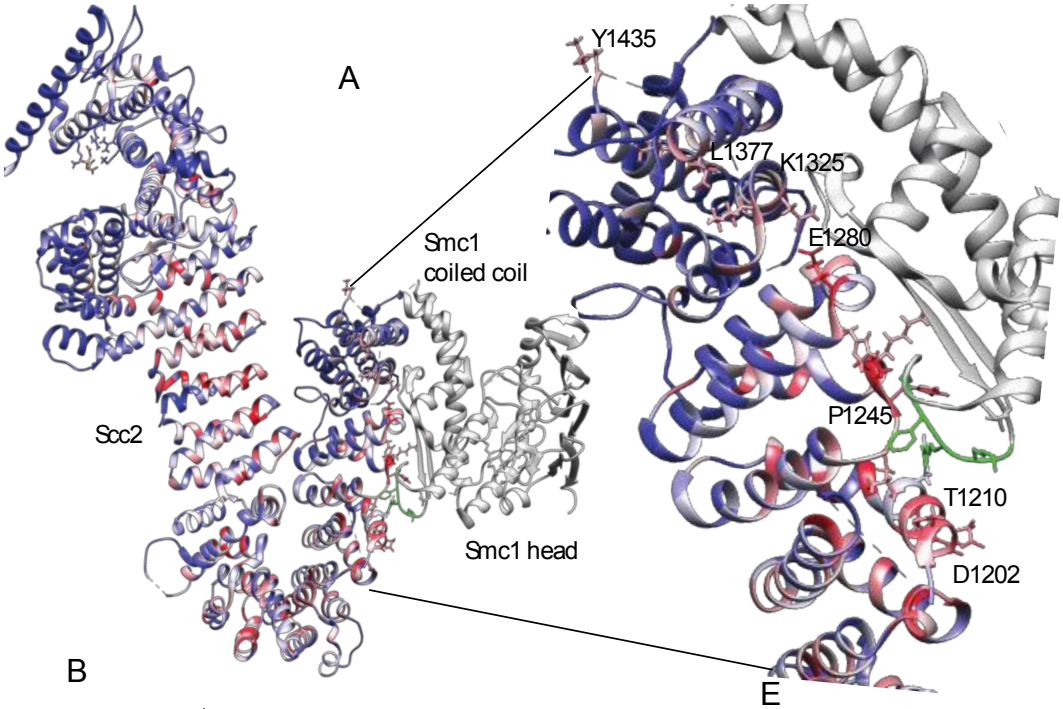


Figure 3-8. Single mutation smc1K1121D has a cell cycle defect
A. FACS analysis shows mutant cells are slow to complete the cell cycle, with a shift towards G2 compared to WT (Strains: K28051, K28401)
B. Split dot analysis shows multiple dots in a single cell and split dots in an early budding cell, suggesting a defect. (Strains: K28419, K10100)

visualised, further work using immunofluorescence is necessary to confirm these results.

3.7 Identification of an essential region in the C-terminus of Scc2

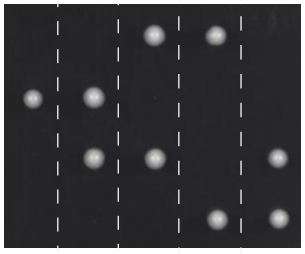
In order to identify the residues in Scc2 involved in the interaction with the Smc1 head, the residue conservation in Scc2 was analysed amongst *Ascomycetes.* The position of the conserved residues obtained from this analysis were then analysed on the crystal structure of Scc2 from *C. thermophilum*. Several conserved positions appeared to be laying on the external edge of the hook and could potentially bind Smc1 F-loop (Figure 3-9A). As mentioned above, our collaborator in Cambridge had preliminary cryo-EM results of Scc2 bound to the cohesin trimer of *S. cerevisiae*. After fitting the crystal structure of *C. thermophilum* Scc2 onto the new yeast cryo-EM model it was confirmed that the residues identified could bind to the head of Smc1. Some of the C terminal residues of Scc2 were identified as potential interaction sites with the first part of the coiled coil of Smc1 emanating from the head. Even if these residues interact with the coiled coil and not the head, their conservation suggested that they should be investigated further. All the conserved residues were mutated as pairs to either alanine or glutamate, as this was expected to highlight the most important positions in the protein. The mutated plasmid was integrated into a diploid yeast strain, sporulated, dissected and replica plated. The viability and thermosensitivity of the mutants was then analysed. The residues and their viability results are summarised in figure 3-9B.



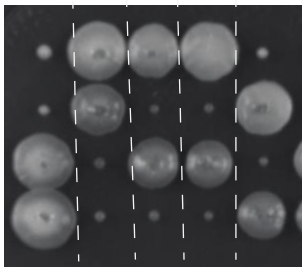
B

Paired mutations in *SCC2*	Viability
N1245E, P1245G	- Lethal
P1245G, S1247E	- Lethal
T1210A, R1211A	- Lethal
D1202A, G1203A	- Viable (TSat 37)
E1280A, T1281A	- Viable
K1325E, K1329E	- Viable
L1377A, D1376A	-Viable
Y1434E, G1435E	- Lethal
Single mutation in *SCC2*	Viability
N1245E	- Viable
P1246G	- PSV*
S1247E	- Viable

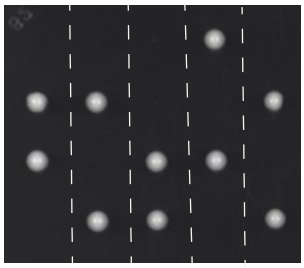
*Poor Spore Viability
- visible after 5 days



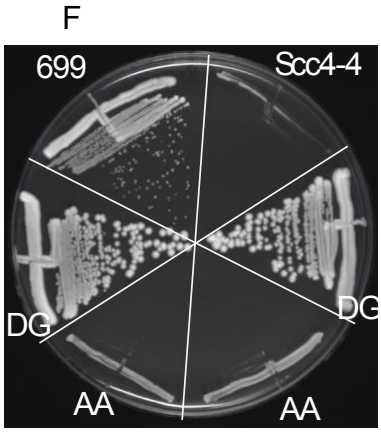
Y1434E, G1435E



P1246G



T1210A, R1211A



D1202A/G1203A

Figure 3-9. The SMC1 F-loop interacts with Scc2

A. Modelled placement of the Smc1 head (grey, PDB 1W1W) with F-loop (green) and Scc2 showing residue conservation (red highly conserved to blue no conservation, PDB 5T8V). Mutations showing side chain along the C-terminal edge, facing the Smc1 head. B. Summary of viability of conserved C terminal residues (strains: K28501, K28502, K28503, K28504, K28506, K28509, K28511, K28513. C. and D. Lethal double mutations. *scc2Y1434E,G1435E* and *scc2T1210A, R1211A*. E. *scc2P1246G*, mutant spores grew after 5 days on YPD at 25°CF. TS mutation *scc2D1202A, G1203A*, the two spores containing the AA mutation do not grow at 37°C

Out of the eight pairs tested, three showed no phenotype, one was thermosensitive and four were lethal. The first of these lethal pairs, *scc2Y1435E, G1436E* (Figure 3-9C) is situated at the very C-terminal of Scc2. Using structural analysis, these residues were initially identified as binding sites with the coiled coil of Smc1. This confirmed that these residues play an essential role in Scc2's function despite the fact that this region of Nipbl was never described in CdL syndrome. However, due to the fact that they don't interact with SMC1 F-loop, they were not analysed further in this study but could be informative in the future.

The second lethal pair, *scc2T1210A,R1211A* (Figure 3-9D) is situated further down the side of the hook, from the model shown in figure 3-9A. This pair was initially designed based on its high conservation and didn't seem to be interacting with the SMC1 F-loop based on our initial modelling. As a consequence, this pair was not analysed further. However, the new cryo EM of *S. cerevisiae* Smc1 head bound to Scc2 in the transition state suggests that these residues may interact with the Smc1 head and should be investigated in more detail (Collier *et al.*, 2020).

The remaining two lethal pairs of lethal residues, consisting of the combination of three amino acids (*scc2N1245E, P1246G* and *scc2P1246G, S1247E*), were located in a small loop of Scc2 sitting opposite to the F-loop of Smc1 on both the initial model and the new cryo EM. This suggested that these two pairs of lethal residues were the most promising candidates for the interaction between Scc2 and the F-loop of SMC1 and therefore were investigated further. The three residues that comprised these two pairs were mutated and integrated as single mutations. *scc2N1245E* and *scc2S1247E* were both viable, however, *scc2P1246G* was found to have a slow growth phenotype after five days (Figure 3-9E).

Interestingly the double mutation of two conserved residues *scc2D1202A, G1203A* was found to be thermosensitive at 37°C (Figure 3.9F). The first residue *scc2D1202* is located on a loop missing from the yeast structure, but the second residue *scc2G1203* is located on the structure and faces the interaction point of the Smc1 and Smc3 heads, when they are engaged. It is therefore possible that these residues play a role in head engagement or ATP hydrolysis. This mutation is interesting because thermosensitive mutants can be used for suppressor screens. Given the position of the residues in Scc2, any suppressors identified in the SMC heads could advance our understanding of the role of Scc2 in the ATPase activity of the complex.

3.8 Identification of interacting residues between the Smc1 head and Scc2

Interestingly the thermosensitive pair of mutations (Scc2D1202A, G1203A, identified above) are on a loop, nine amino acids away from Scc2K1193, a residue that was previously identified during a Lysine-Lysine crosslinking screen as potentially interacting with the Smc1 F-loop position K1124 (Burmann *et al.,* 2019). These amino acids were good candidates to design as a cysteine crosslinking pair, however, on mutation to cysteine, residue Scc2K1193 appeared to be lethal. The second amino acid Smc1K1124C, was viable and could be used for crosslinking providing an alternative cysteine viable partner could be identified in Scc2.

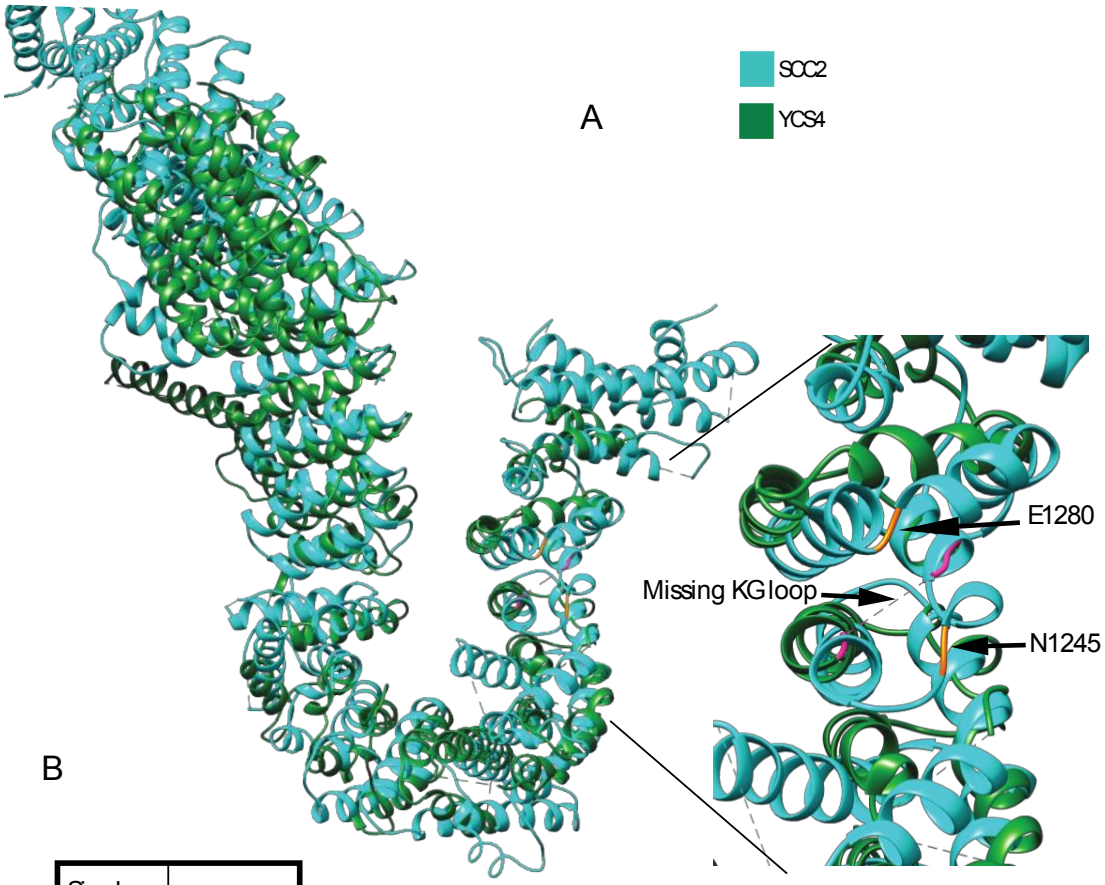
As previously mentioned, the Smc4 W-loop has been shown to interact with the KG-loop of Ycs4. When the structures from *C. thermophilum* Ycs4 (PDB: 6QJ4) and Scc2 (PDB:5tv8) are overlaid, the KG-loop of Ycs4 sits between two loops in the structure of Scc2 (Hassler *et al*., 2019; Kikuchi *et al*., 2016). These loops can be

identified as (Scc2N1245/P1246/S1247) a lower loop and (Scc2E1280/T1281) an upper loop (Figure 3-10A).

Residues in both loops (S. cerevisiae Scc2: Y1243, T1244, N1245, P1246, S1247, T1248, E1280 and T1281) were mutated to cysteine and tested for their viability as described in figure 3-10B. Each Scc2 cysteine was then crossed to Smc1K1124C and tested for *in vivo* crosslinking using BMOE in cycling cells. The western blot detected a signal between one cysteine pair. A faint crosslink band is visible between Smc1K1124C and Scc2N1245C (Figure 3.10C). These cysteine pairs were tested in whole cell extract, whilst this confirms the interaction of Smc1 F-loop with the C-terminus of Scc2, the crosslinking efficiency could be analysed after immunoprecipitation pulling on Scc1, in order to enrich for fully formed cohesin rings. The interaction between this cysteine pair is supported by the structural model, but is not strong enough to use for further studies. Additional investigation could eventually identify more efficient residues for crosslinking.

Discussion

These results confirm the essential role of the Smc1 F-loop for cohesin function, similar to what has been observed in the W-loop is for the condensin complex. The combined lethal mutations of the F-loop in Smc1 triggered a complete loss of cohesin activity; affecting both the interaction between DNA and the complex and its ATPase activity, both of which are mediated by Scc2. However, the fact that these mutations, which are very close to the signature motif, also induce the loss of head engagement suggests that they might just prevent the two heads from coming



Single mutation	Viability
Y1243C	- Lethal
T1244C	-lethal
N1245C	- Viable
P1246C	- Viable
S1247C	- Viable
L1192C	-Viable
K1193C	-Lethal
E1198C	-Viable
E1280C	- Viable
T1281C	- Viable

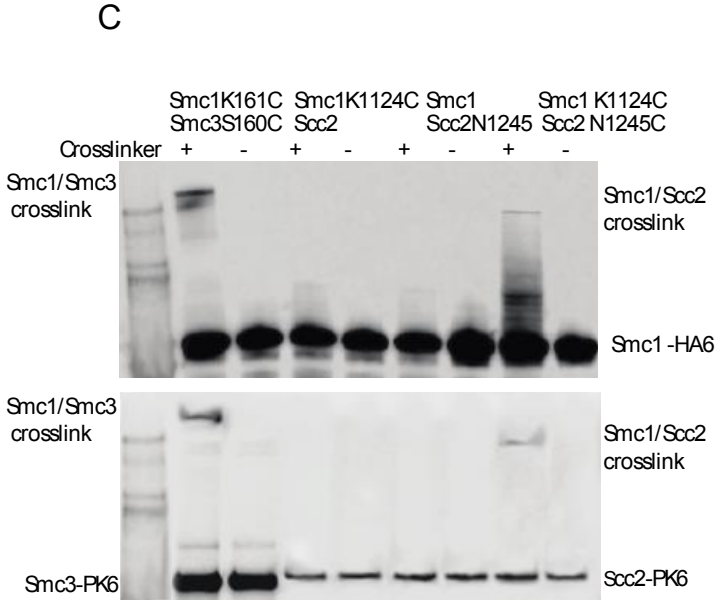


Figure 3-10. Scc2 and Ycs4 show structural similarity between their C-terminal hook regions

A. Overlay of Ycs4 (in green; PDB 6QJ4) highlighting the missing K-G loop (in Pink, dashed line), Scc2 (in cyan; PDB 5T8V) highlighting loops at E1280aa and N1245aa (orange) above and below the site of the KG-loop. Structures from *C. thermophilum.* B. Viability results of cysteine mutations. C. Crosslink between Smc1K1124C and Scc2N1245C. Strains: K28343, K28174, K28344, K27359 repeated x2).

together. The mutation of *smc1 K1121D* does, however, suggest a more complicated picture. This mutation has only a modest effect on DNA dependant ATP hydrolysis and on the genome occupancy of the cohesin complex. However, the growth defect detected in the cells is severe and the cells appear to accumulate in G2/M. Preliminary results suggest that this may be due to a cohesion defect. This will require further investigation to determine the nature and the severity of the defects.

In parallel to the work in this thesis, a BPA crosslinking study of the Smc1 head identified two residues which are located in the Beta-sheets just above the F-loop, as interacting with Scc2 (Bin Hu, personal communication). This data alongside the aforementioned Lysine-Lysine crosslinking pair between Smc1K1124 and Scc2K1193 found by Burman *et al.*, reinforces the idea that the F-loop of Smc1 is involved in the binding to the C-terminal region of Scc2. The position of both of these interactions agrees with the cysteine crosslink pair identified here between Smc1K1124 in the F-loop and Scc2N1245C. The identification of the pair of thermosensitive residues in Scc2 (D1202A, G1203A) in a region very close to Scc2 N1245C (the residue able to crosslink to Smc1) presents the opportunity to screen for genetic suppressors of this mutation, that could be used to further dissect the loading mechanism.

Further evidence for the interaction between Scc2 and Smc1 was recently provided with the publication of a cryo EM structure of yeast cohesin with the heads engaged, described as clamping the DNA (PDB:6zz6). In this structure, described above as the transition state, DNA is held in two compartments created between the engaged heads and the coiled coil, and the engaged heads and the kleisin. Using an *in vitro* entrapment assay, it was shown that Scc2 plays an essential role in this

entrapment and creates a sub compartment by associating with the engaged heads (Collier *et al.*, 2020). Despite the fact that in the *S. cerevisiae* cryo EM structure, the Scc2 loop identified by Burmann *et al*., as binding to Smc1 is missing, the position and proximity of the two proteins is compatible with the interaction detected by the different crosslinking experiments and is confirmed by the human cryo EM structure (Shi *et al.,* 2020)

In the transition state cryo EM structure, the Smc1 and Smc3 heads are engaged while Scc2 is bound to Smc1 head close to the F-loop. The same observation was made in the cryo EM structure of human cohesin. This result differs from what was suggested for condensin. The condensin structure (Hassler et al, 2019) describes the head of Smc4 interacting with Ycs4 but Smc2 is absent. Experimental evidence suggested in the same study, that this binding was incompatible with head engagement due to a sterical clash. Moreover, addition of ATP or ADP seemed to disrupt Smc4 binding to Ycs4. This suggests a potential difference of behaviour between condensin and cohesin.

Chapter 4: Analysis of the folded conformation of cohesin

Introduction

Three different activities have been described for SMC complexes: DNA entrapment, translocation and loop extrusion. The way DNA is "handled" by the complex and how it is topologically entrapped inside the different chambers of the SMC complexes is essential to understand their function in chromosome organisation. SMC complexes have previously been described as an open ring through which the DNA strands can pass. This hypothesis is supported by some early electron microscopy data showing SMC complexes as a fully open ring with the coiled coils separated. However, atomic force microscopy (AFM) and cryo/negative stain electron microscopy (EM) of different SMC complexes, have isolated several different conformations: a ring, a straight rod and a folded rod (Melby *et al.,* 1998, Anderson *et al.,* 2002; Haering *et al*., 2002; Fuentes-Perez *et al.,* 2012; Burmann *et al,* 2019). In fact, recent Cryo-EM analysis of SMC complexes from different organisms suggest that the SMC complexes exist mainly as a rod shape structure in which the two coiled coils are zipped up from the hinge down towards the heads.

The zipped-up conformation has been confirmed by crosslinking experiments showing the proximity of the two coiled coils along their entire length (Soh *et al.*, 2015; Diebold-Durand *et al*., 2017; Burmann *et al*, 2019). However, to what extent the coiled coils are zipped up during head engagement is still under investigation, the juxtaposed state seems to be compatible with fully zipped up coiled coils. The ATP binding and head engagement would seem to trigger divergent angles of the coiled coils emanating from the heads that could induce their unzipping (Figure 4.0). From two

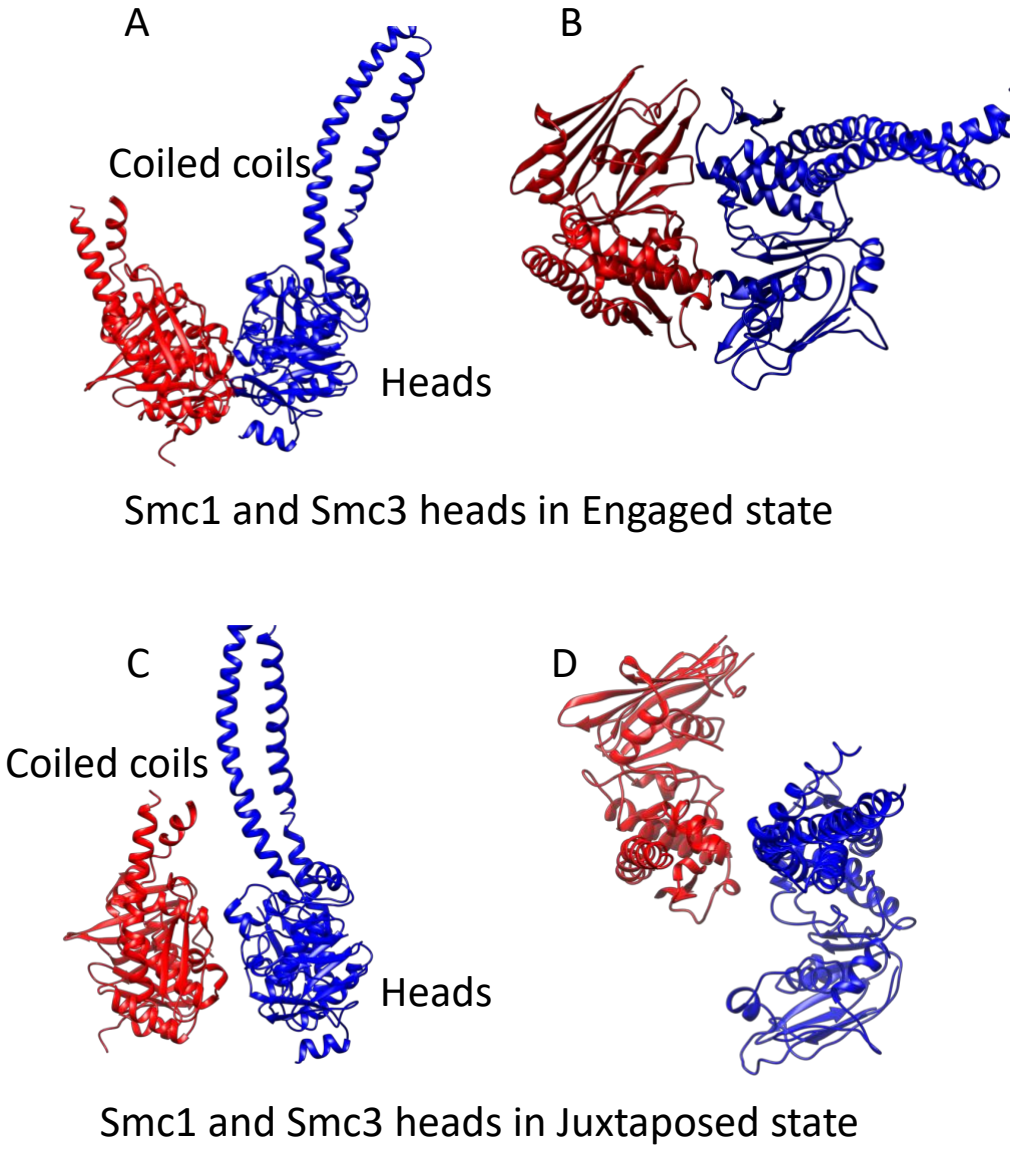


Figure 4-0. The angle of the coiled coils differs on head positioning
Top: Models of engaged heads; outward angle of coiled coils. A. View from front and B. view from above
Bottom: Model of juxtaposed heads; inward angle of coiled coil allowing full length zipping of the coiled coils. C. View from front and D. View from above. Models made using PDB 1w1w and 4ux3, based on Chapard *et al,* 2019. Smc1 in red, Smc3 in blue.

fully disengaged heads to a juxtaposed position and a fully engaged status upon ATP binding, the mechanical strain applied on the coiled coils may adjust their level of zipping thereby regulating the opening of the SMC chamber, compatible with DNA entrapment (Diebold-Durand, 2017). Analysis of a recent Cryo EM structure shows that when the heads are engaged, the coiled coils are separated for at least four turns of the alpha helices (Muir *et al.,* 2020). Collier *et al.*, also showed that the Scc2 dependent entrapment of DNA in the transition state is accompanied by the unzipping of the coiled coil that prevented the crosslinking between Smc1K210C and Smc3K198C. This partial unzipping of the coiled coil would create a positively charged chamber just above the heads, compatible with the DNA binding previously described for the Rad50 complex (Liu *et al*., 2016).

More recent analysis of SMC complex structures highlights the existence of another type of conformation: the folding of the coiled coils. Recent advances in cryo EM techniques have produced high quality images of such structures in which the coiled coils are bent at a region that contains a disruption of the coils, called the elbow. This action brings the hinge down towards the heads, in a hair-pin fold (Burmann *et al*., 2019). This conformation brings many questions about its possible involvement in the ability of cohesin to entrap DNA, translocate and extrude loops. By alternating the cohesin conformation from extended to a folded state, it is possible that the hinge could capture the DNA and that the subsequent folding could bring the DNA close to the heads and their associated HAWKs. In line with this hypothesis, a recent analysis has shown that Scc2 can bind to the hinge (Smc1K620bpa) in *S. cerevisiae* (Bin Hu, personal communication). Furthermore, Burmann *et al.,* showed using a BPA crosslinking screen that Pds5 can bind to the same residue in the hinge of yeast

cohesin (*smc1K620bpa*). This crosslink was abrogated in an Scc1V137K mutant showing that it depends on the assembly of the complex.

This hypothesis could have major functional implications as several pieces of experimental evidence have involved the hinge in the loading of cohesin. The hinge was shown to interact with DNA (Chui *et al*., 2004). As previously mentioned, the induction in G1 phase of the hinge closure using the FRB-FKBP system is lethal, suggesting a potential role for this interface as an entry gate (Gruber *et al*., 2006). In addition, different mutations in the hinge were shown to affect loading (Mishra *et al.*, 2010; Kurze *et al.*, 2011; Petela, 2016). Charge reversal mutation of the residues exposed in the lumen of the hinge localises across the genome in a manner similar to WT complexes as measured by ChIP sequencing but cannot entrap mini-chromosomes topologically (Srinivasan *et al.*, 2018). Another mutation at the base of the hinge (*SMC1D588Y*) can bypass the requirement of *SCC4* for loading (Dixon, 2013). All these observations strongly suggest that the hinge is involved in loading via a possible interaction with DNA. In the context of the folded conformation, this role of the hinge could be included in a more dynamic picture of how SMC complexes interact with DNA (Dixon, 2013; Petela, 2016).

As the folded conformation of the SMC coiled coils has thus far only been observed in vitro, the first essential question to address was whether this folding occurs *in vivo*.

4.1 Predicting interacting residues between the Smc3 coiled coils and the Smc1 hinge

The cryo EM structure recently obtained by our collaborator (Mr. Andres Gonzalez-Llamazares, personal communication) shows a full-length yeast cohesin heterodimer in a folded configuration (Figure 4-1, left). This structure is of high enough resolution to identify the individual coiled coils. Therefore, it was possible to determine that the hinge of Smc1 could interact with the coil of Smc3 (Figure 4-1, right). However, a more precise analysis of both potentially interacting regions was necessary in order to identify specific residues that could crosslink after mutagenesis into cysteine.

The SMC proteins are highly conserved throughout evolution. It was therefore possible to use the existing crystal structure of the hinge region of *M. musculus* (PDB:2wd5, Kurze *et al.,* 2011). This structure was modelled over the newly resolved cryo EM structure of the folded conformation of cohesin from *S. cerevisiae* (Figure 4-1, right). This enabled the identification of two conserved small alpha helices in the Smc1 hinge that are in close proximity to the coiled coil of Smc3 and may therefore interact (Figure 4-2, top).

Based on the estimated proximity of these two helices of the Smc1 hinge with the Smc3 coil, it was predicted that the lower helix of Smc1 (residues 505-511 in *M. musculus*) was slightly further away and to the side of the Smc3 coil, it was therefore less likely to interact than the second helix (558-564 in *M. musculus*) which was closer and sat directly between the coiled coils, thus has a higher probability of interaction with Smc3. These correspond to Smc1 residues 520-526 and 572-578, respectively, in

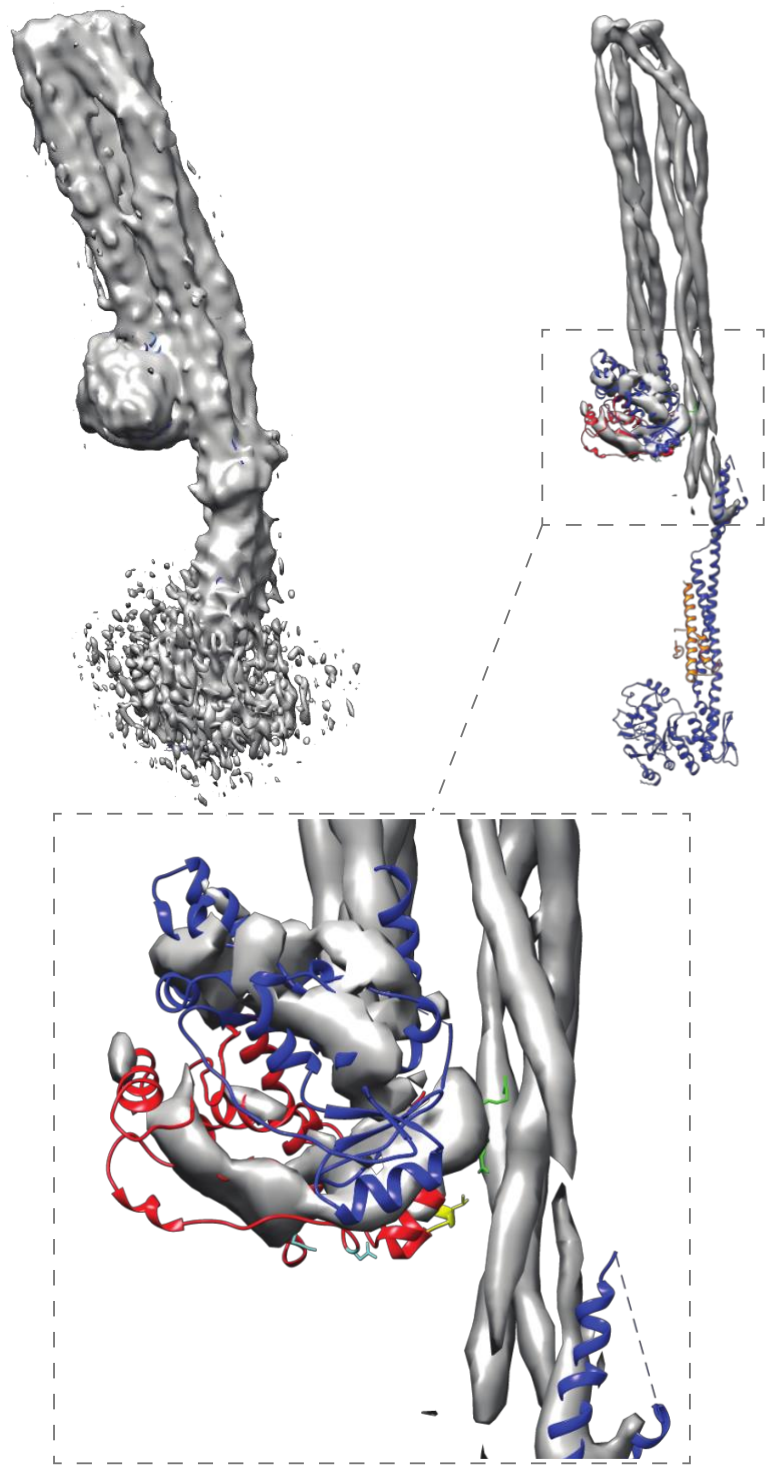


Figure 4-1. The smc1 hinge can interact with the smc3 coiled coil in the folded conformation

Top left: Cryo-EM of full length cohesin in the folded conformation (Data used with permission Mr. Andres Gonzalez-Llamazares (Löwe lab, LMB, Cambridge).Top right: Modelled hinge of *M. musculus* (Smc1 in red, Smc3 in blue, PDB:2wd5), and Smc3 head to the joint of the coiled coil of *S.* cerevisiae (Smc3 in blue, Scc1 in orange, PDB: 4ux3). Bottom: Magnification of interacting region. Side chains in green and yellow.

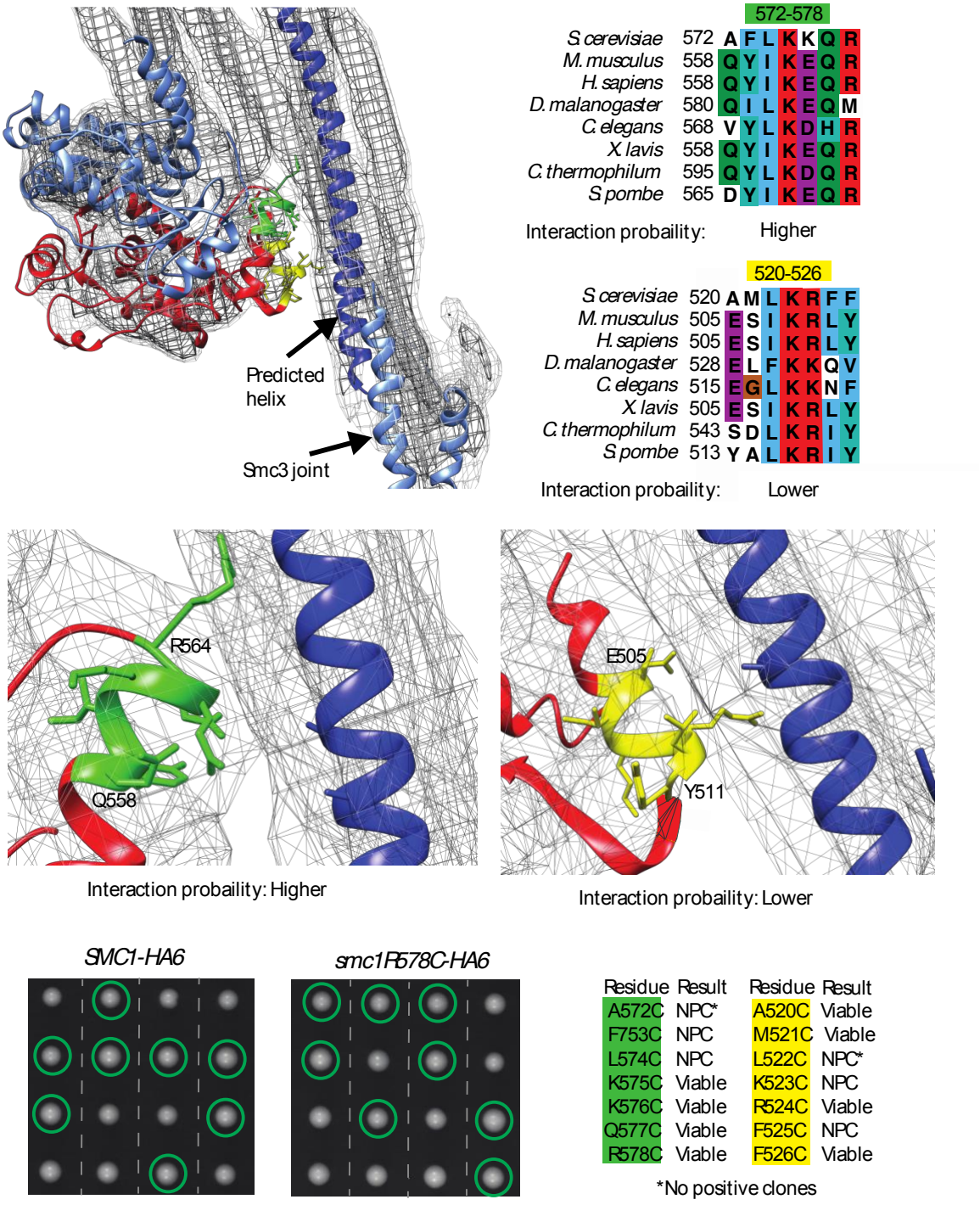


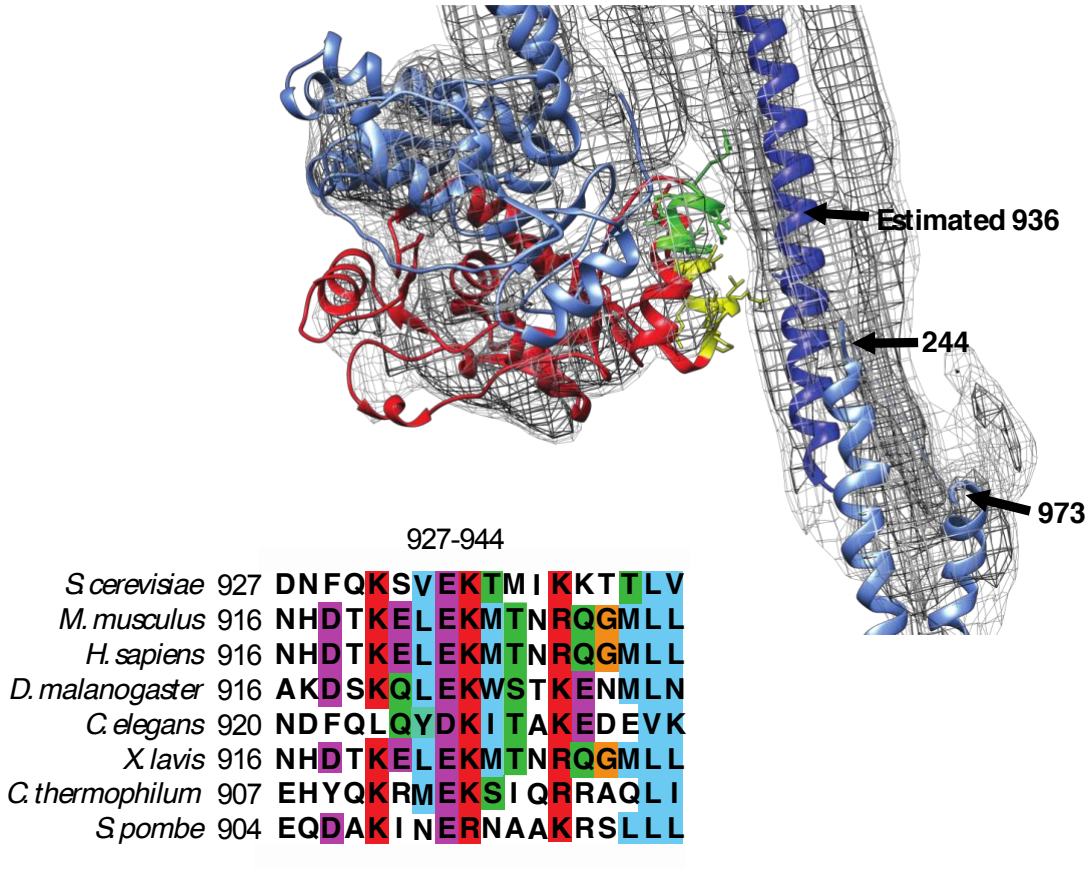
Figure 4-2. Predicting interactions in the Smc1 hinge

Top left: Cryo-EM (grey) with modelled hinge (Smc1 red, Smc3 light blue, PDB:2wd5), predicted helix (dark blue) and joint of Smc3 (light blue, 4ux3). Right: Alignment of Smc1 evolution for higher (green) and lower (yellow) probability predicted interactions. Middle: Enlargement of Smc1 predicted interactions; left: Higher probability predictions (green helix) right: Lower probability predictions (yellow helix). Bottom: left, Examples of viable plate dissections WT control *(SMC1-HA6)* and *smc1R578C-HA6*, green circles depict the HIS3 marker that carries the mutation. Right, summary of results for higher (green) and lower (yellow) probability residues.

S. cerevisiae (Figure 4-2 middle).

The fourteen residues identified in the Smc1 hinge (572-578 and 520-526) were mutated to cysteine using SDM on a plasmid containing the *SMC1* gene carrying an HA6 tag, and marked by the *HIS3* gene. The HA6 tag would be used to visualise the resulting protein and potential crosslinking bands. All resulting plasmids were transformed into a WT diploid yeast strain to recombine with the endogenous locus and dissected after meiosis to test viability of the resulting spores containing the mutated gene. Of the fourteen designed Smc1 hinge mutations, eight were successfully integrated in a first round of transformation and were all found to be viable after dissection (Figure 4-2, bottom).

Having identified potential binding residues of the hinge, a similar approach was used to predict its potential binding site in the coiled coil of Smc3. However, the crystal structure available only contained a short fragment of the coiled coil: from Smc3 head up to the joint. This crystal structure was modelled over the new cryo EM structure. However, because this Smc3 crystal structure did not cover enough of the coiled coil region, it was necessary to model how many turns of the alpha helix would be needed to reach the motif predicted to interact with the hinge. This led to the estimation that the hinge would contact Smc3 around the position 936 of its coiled coil. In order to design possible crosslinking residues an interval of an initial estimate three turns of helix above and below position 936 of Smc3 spanning from amino acid 927 to 944 was used (Figure 4-3, top). The conservation of theses seventeen residues is indicated in Figure 4-3 (middle).



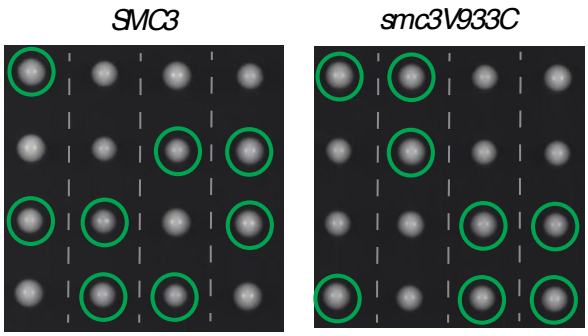


Figure 4-3. Predicting interactions in the Smc3 coiled coil

Top: Cryo-EM (grey mesh) with modelled mouse hinge (Smc1 red, Smc3 light blue, PDB:2wd5), predicted helix (dark blue) and modelled joint of Smc3 (light blue 4ux3). Middle: Alignment of Smc3 evolution for selected residues. Bottom: Examples of viable plate dissections WT *SMC3* control and *smc3V933C*, green circles depict the KanMX4 marker that carries the mutation.

a plasmid containing the *SMC3* gene marked by the kanamycin resistance gene (KanMX4). All 17 residues were successfully transformed into yeast and were viable after dissection (Figure 4-3, bottom).

The next step was to create strains in which each individual cysteine mutation in Smc1 would be combined with each cysteine position in Smc3. However, due to the large number of possible combinations that these crosses would create, it was not feasible to undertake. Therefore, further modelling of the residues in the Smc1 hinge was undertaken looking at the possible positions that each amino acid can adopt within the helix (572-578). The possible angles that the side chain of each residue can adopt was determined by rotamer analysis. This identified residue Smc1Q564C in the *M. musculus* hinge analogous to *smc1R578C* in *S. cerevisiae* as the most likely to interact. However, this analysis could only be undertaken in the two helices of Smc1 hinge as no crystal structure was available for the Smc3 coiled coil. Without a crystal structure, the orientation that each residue faces cannot be predicted. It was therefore, not possible to reduce the number of residues in Smc3, and all seventeen sequential cysteine residues (927-944) had to be tested in combination with *smc1R578C*. All of the seventeen strains containing a cysteine in both Smc1 and Smc3 were found to be viable (data not shown).

4.2 The Smc1 hinge can crosslink to multiple residues in the Smc3 coiled coils

The *in vivo* crosslinking of these different cysteine pairs was analysed. This was undertaken using the BMOE crosslinker in cycling cells. An existing strain with a cysteine pair in the Smc1/3 hinge was used as a positive control (Smc1K639C with Smc3E570C). This cysteine pair gives a crosslinked product of 460 kDa. The crosslinked product of the folded conformation was expected to give a similar size. Each of the first eleven strains tested (927-937) gave a higher molecular weight band of the correct estimated size (Figure 4-4). These bands were only observed in the presence of the BMOE crosslinker. This result was surprising given the conformation of the helix, raisng the concern of unspecific crosslinking to Smc1. It would be expected that at least one residue within a helical turn would not be able to crosslink as it would face the wrong direction. However, some residues, approximately every third, appeared to give a stronger signal, suggesting that these may lie in closer proximity as they show more efficient crosslinking. The remaining six strains were not tested due to time constraints.

The crosslinking pair Smc1R578C with Smc3V933C, gave the highest efficiency of 53% when probing for Smc1HA6. However, due to the lack of tag on Smc3, it was not possible to address whether the crosslink band was due to the introduction of both cysteines. Therefore, a PK6 tag was introduced onto Smc3 to be able to address the specificity. In parallel the single cysteine strains were analysed to ensure that the crosslinking band was specific to the presence of both cysteines and that there was no unspecific crosslink giving background signal. The new strains were viable the

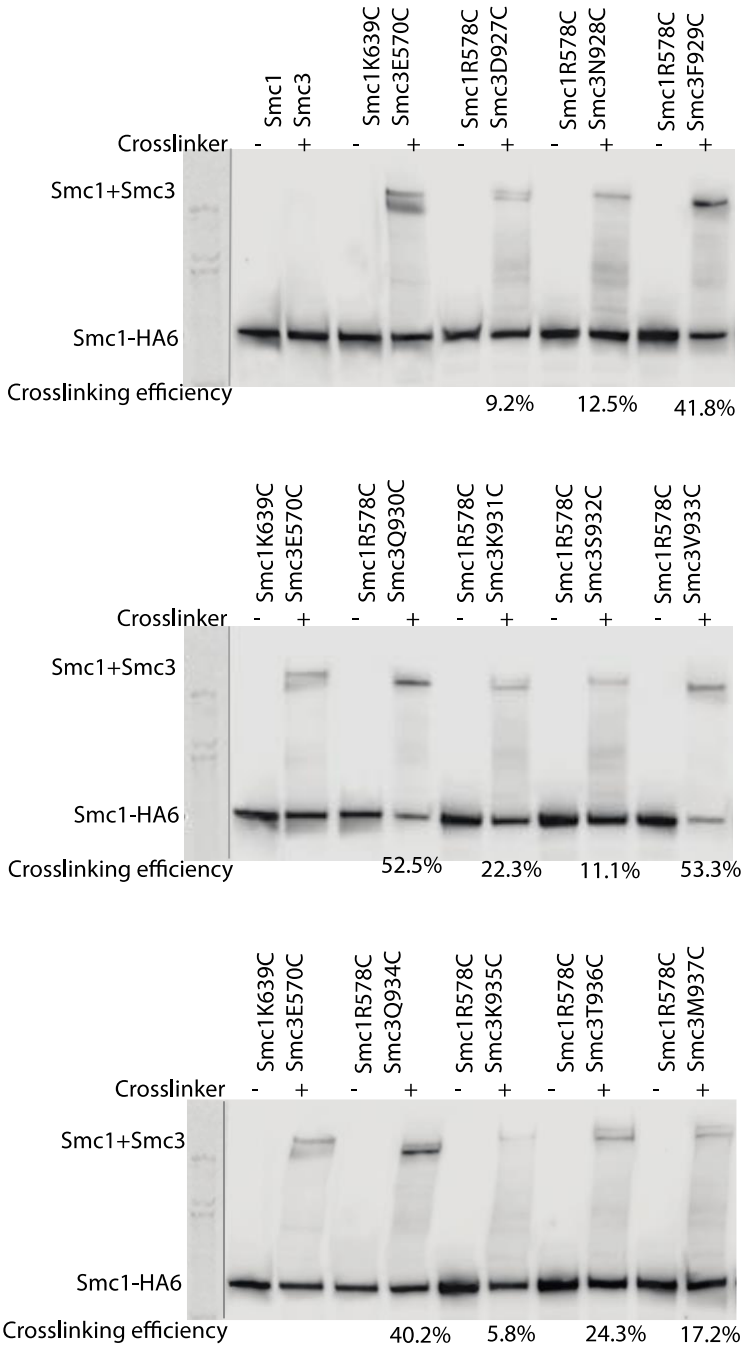


Figure 4-4. The Smc1 hinge can form multiple crosslinks with the Smc3 coiled coil
Western blot analysis testing the crosslinking efficiency of the folded state using Smc1R578C with seventeen cysteine residues in Smc3 in cycling cells (K28413, K28379 K28381, K28383, K28422, K28424, K28528, K28389 K28426, K28387, K28400). A strain containing no cysteines was used as a tag only control (K28401) and a strain containing Smc1K637C with Smc3E570C was used as a positive crosslinking control (K26081). The blot was probed with anti HA antibody and the efficiency of the crosslink was calculated (Experiments performed by Mr. Antonio Ferrer-Harding).

single cysteines gave no crosslink bands and the cysteine pair gave the expected crosslink band visualised by both antibodies. For this experiment the juxtaposed heads cysteine pair (*smc1S161C and smc3K160C*) was used as a positive control as it contained both Smc1-HA6 and Smc3-PK6. This shows that the crosslink of *smc1R578C* with *smc3V933C* is both efficient and specific (Figure 4-5).

4.3 Cohesin in the folded state occurs throughout the cell cycle

This new cysteine pair, Smc1R578C with Smc3 V933C is a useful tool to analyse the possible link between the conformation of the cohesin complex and its activity *in vivo*. Is this conformation involved during the process of loop extrusion in G1 or more specifically after acetylation of the complex and cohesion establishment in G2? In order to test whether the folded conformation is specific to certain phases of the cell cycle, the crosslinking efficiency of Smc1R578C with Smc3V933C was analysed in different stages of the cell cycle.

The strain containing the cysteine pair Smc1R578C with Smc3V933C was synchronised in early G1 using pheromone arrest and released in the presence of nocodazole, to arrest in G2. Samples were taken at 0 and 90 minutes post release and a control of cycling cells was also analysed. For each of the three conditions, the crosslink was induced *in vivo* by addition of BMOE before cell lysis and analysis by western blot.

This analysis confirmed the presence of the expected crosslink band in both G1 and G2 stages of the cell cycle (Figure 4-6). Both phases showed crosslinking efficiencies comparable to that of cycling cells. This suggests that the folded

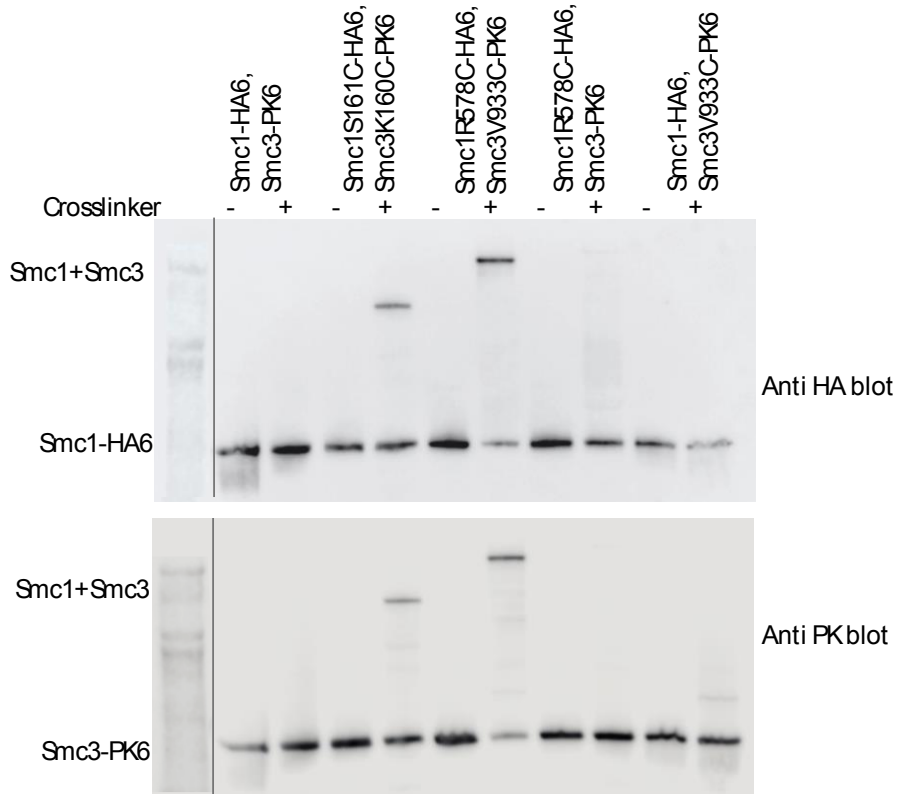


Figure 4-5. The *smc1R578C* with *smc3V933C*, hinge-coiled coil crosslink is specific
Western blot analysis of the crosslinking specificity of the folded state in cycling cells using negative control (Smc1-HA6, Smc3-PK6), positive control juxtaposed heads crosslink (Smc1S161C-HA6 and Smc3K160C-PK6). The specificity of the folded state was addressed using strains containing only single cysteines (Smc1R578C-HA6 with Smc3-PK6) and (smc1-HA6 with Smc3V933C-PK6) (Strain numbers: K27359, K28401, K28585, K28546 and K28583). The blot was probed with antibody and the efficiency of the crosslink was calculated. Top: Probed with HA antibody for Smc1. Bottom: Probed with PK6 antibody for smc3 (Experiments performed by Mr. Antonio Ferrer-Harding).

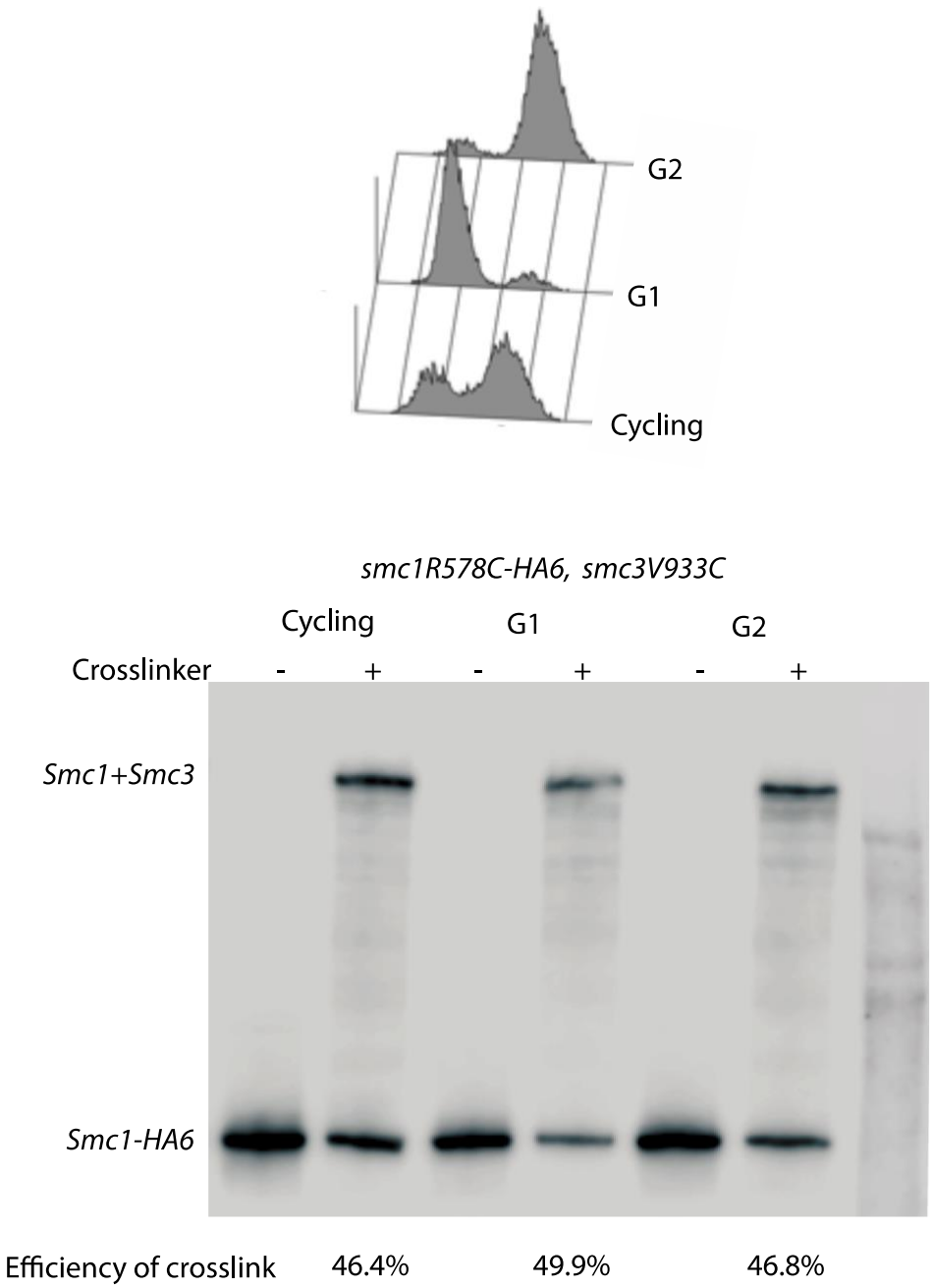


Figure 4-6. The folded form of cohesin occurs throughout the cell cycle
Top: FACS analysis of cells containing *smc1R578C* and *smc3V933C*. cells were cycling, G1 arrested or arrested and released into G2 with nocodazole.
Bottom: Western blot analysis of crosslinking measuring the folded state in cycling cells and arrested in G1 or G2. (Strain: K28586). The blot was probed with anti HA antibody and the efficiency of the crosslink was calculated (experiment performed by Mr. Antonio Ferrer-Harding).

conformation occurs throughout the cell cycle with no significant change to crosslinking efficiency in G1 or G2.

4.4 The folded conformation is compatible with acetylation

If the folded conformation was occurring throughout the cell cycle, we would expect it to be found with both single and replicated DNA as suggested by the G1 vs G2 experiment, but is this configuration compatible with the cohesive fraction? To test this hypothesis, the compatibility of hinge-coiled coil crosslink pair was tested with the acetylation of Smc3. The Smc1R578C with Smc3V933C crosslink pair was compared to the hinge-hinge (Smc1K639C with Smc3E570C) crosslink in cycling cells. This crosslinking pair had previously been characterised as showing a similar crosslinking efficiency before and after acetylation (Chapard *et al.,* 2019).

As represented in Figure 4-7, the acetylated cohesin complex present a similar crosslinking efficiency between the hinge and the coiled coils as the general population of cohesin. The crosslinking efficiency of the acetylated complex, calculated as the ratio of the crosslinked band intensity to the free Smc3 detected by the anti-acetyl Smc3 antibody is similar to the crosslinking efficiency obtained with an anti-tag antibody. As a negative control, in G1 arrested cells, the crosslink between Smc1 and Smc3 hinge regions did not show any acetylation. This suggest that the folding of the cohesin complex happens independently to the acetylation status of Smc3.

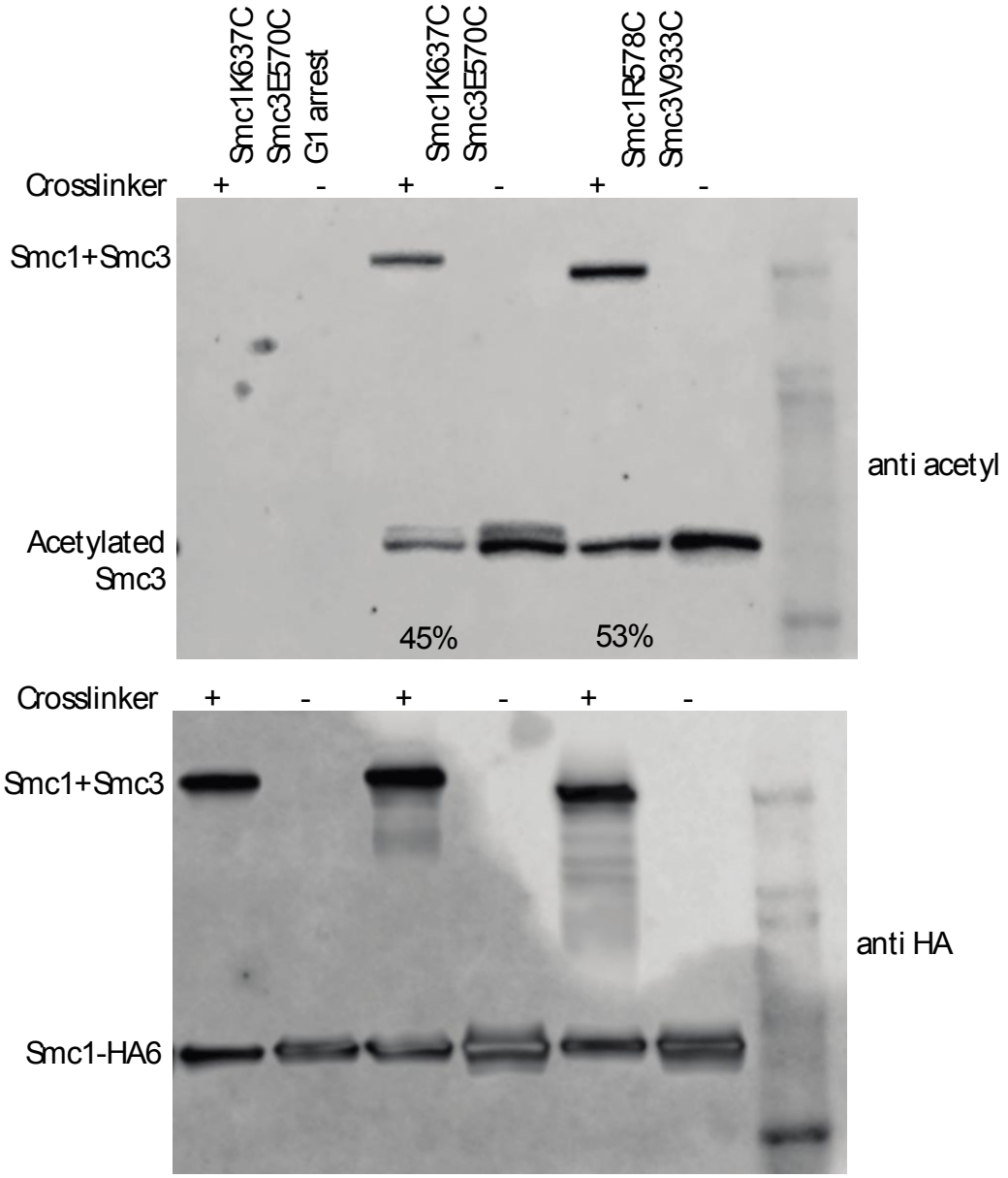


Figure 4-7. The folded form of cohesin can be acetylated

Western blot analysis of crosslinking measuring the folded state in cycling cells using Smc1R578C and Smc3V933C. A strain expressing the SMC hinge crosslinking pair Smc1K637C with Smc3E570C was used as a positive control. This strain was arrested in G1 providing a negative control for acetylation. (K28585, K26081). Top: Probed with anti Smc3 acetylation antibody. Bottom: Probed with anti HA antibody.

Discussion

Several structural studies have previously described the different conformations of the cohesin complex. The folding of the coiled coils at the elbow bringing the hinge into contact with the coiled coil has long been observed but the existence of this configuration has never been addressed *in vivo*. The results obtained here show for the first time that the folded state of cohesin exists *in vivo*. Furthermore, the observation that the crosslinking efficiency corresponding to this conformation appears unchanged during the cell cycle suggest that it is a major state of the complex that might be involved at different stages of the cohesin cycle. The physiological relevance of this state remains to be investigated.

The fact that the crosslinking efficiency between the hinge and the coiled coil is not affected by the acetylation of Smc3 suggests that this folding is not specific to any functional state of cohesin: cohesive or non-cohesive. This question should be addressed further using the mini-chromosome assay which has previously shown for all the cohesin mutants tested a perfect correlation between their ability to form cohesion *in vivo* and their ability to entrap mini-chromosome dimers after crosslinking (Srinivasan *et al*., 2018). Creating a 6C strain using cysteine pairs at both Smc-kleisin interfaces (Smc1/Scc1 and Smc3/Scc1) and the new pair between the Smc1 hinge and Smc3 coiled coil, this should enable the topological entrapment of DNA. If this conformation is a general feature of cohesin, as the compatibility with acetylation suggests, both CMs and CDs should be detected.

The hinge region has previously been described as a possible DNA binding region. (Chiu *et al.,* 2004). This could suggest that the folding of the complex could

bring the DNA bound to the hinge close to the heads and the associated HAWKs, feeding the DNA for the complex to extrude loops (Figure 4.8). Initial *in vitro* experiments using purified proteins have shown that the folded configuration can occur independently of any interaction with the HAWKs. More recently, two different studies have shown that the N-terminus of Scc2 could interact with the hinge. First, our collaborator obtained a cryo EM structure of the folded conformation including Scc2. In this structure the C-terminus of Scc2 is bound to the Smc1 head while its N terminus is close to the hinge region (Andres Gonzalez, personal communication). In a separate study, it was demonstrated that the N-terminal region of Scc2 can crosslink to the hinge of Smc1K620*, in vivo* (Bin Hu, personal communication). It was also shown in this thesis that Scc2 could interact with the Smc1 head *in vivo*.

The interaction of Scc2 N terminus with the hinge raises the possible involvement of the HAWKs in the stabilization of the folded state. However, the crosslinking efficiency between the hinge and the coiled coil is not modified in early G1 compared to G2 or cycling cells (this study). In early G1 Scc1 is in low abundance in the cell after its cleavage in the previous cell cycle, and therefore needs to be resynthesized (Ciosk *et al,* 2000). As a consequence, at this stage of the cell cycle, the association of the HAWKs is reduced as their binding to cohesin requires Scc1(Ciosk *et al,* 2000). If Scc2 had any contribution to the stability of the folded conformation, one would expect the crosslinking efficiency to be reduced in early G1. This suggests that the HAWKs are not involved in the stability of the folded conformation. However, this is only an indirect observation that would need further investigation to address the crosslinking efficiency in the presence or absence of Scc2 and the levels of Scc1

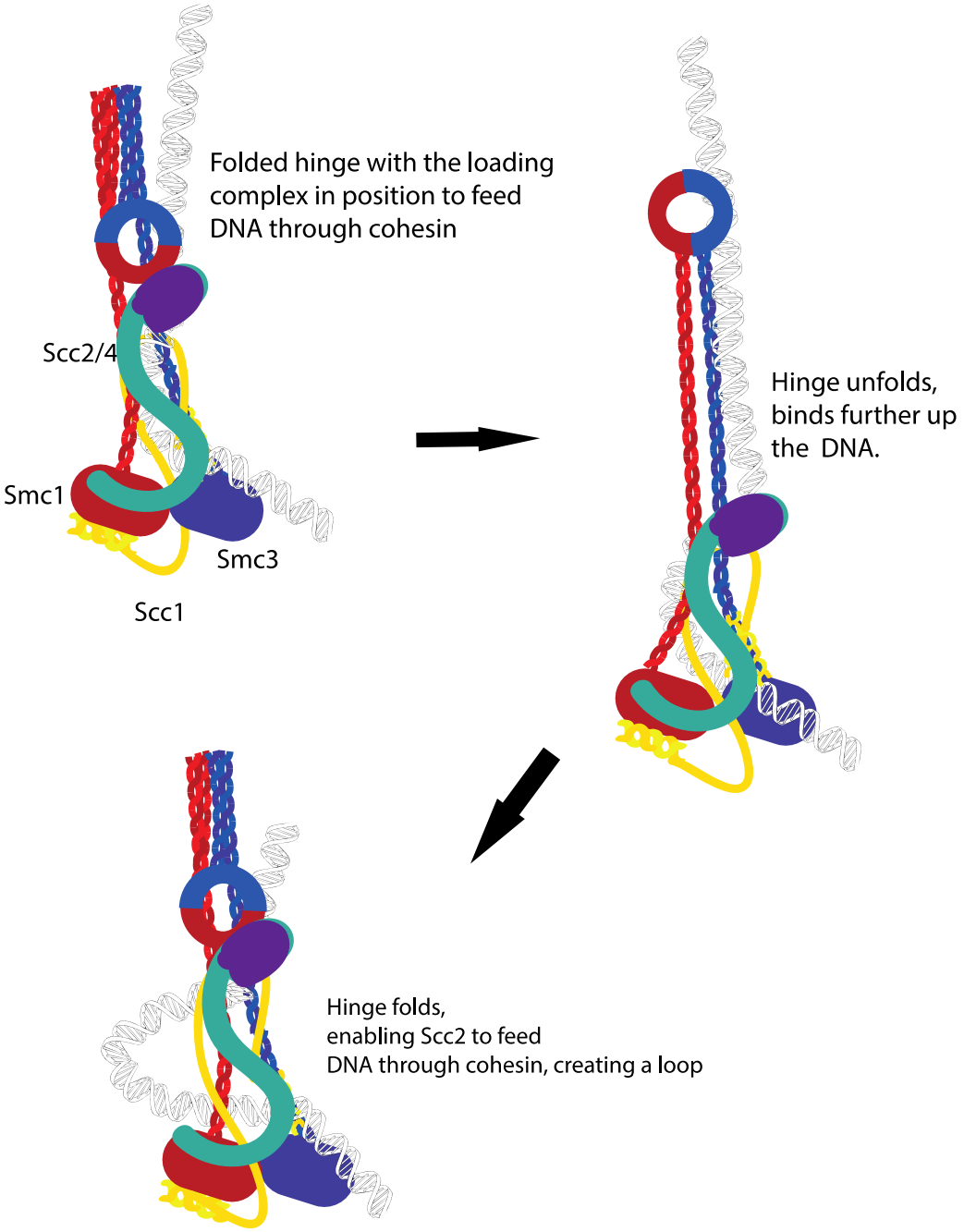


Figure 4-8. Model of the cohesin trimer interacting with Scc2/4 (the loading complex): How the hinge could be involved in loop extrusion.
Smc1 in red, Smc3 in blue, Scc1 in yellow, Scc2 in turquoise and Scc4 in purple.

present.

The folded state has been observed in many SMC complexes (MukBEF, cohesin and condensin). However, structural analysis of condensin revealed that in this folded state, the hinge is unable to reach as far down the complex when compared to cohesin (Kong *et al.,* 2020, Lee *et al*., 2020). The fact that the condensin hinge fails to reach as far down, suggests that it is unable to interact with the N-terminus of the HAWKs. This implies that even if the different SMC complexes can bend, the contact between the hinge and the HAWKs could be specific to cohesin.

The binding of the hinge to Scc2 suggests that this interaction could be involved during the loading process. The amino acid in Smc1 hinge (K620bpa) found to crosslink to Scc2 is very close to a previously identified mutation that bypasses the requirement for Scc4: Smc1D588Y. Furthermore, these two amino acids (Smc1K620 and Smc1D588) are in close proximity to the position found to crosslink with the coiled coil (Smc1R578). Knowing that Scc2 can bind to this part of the hinge may give an insight into how the Smc1D588Y mutation is able to affect Scc2 binding and its ability to load in the absence of Scc4.

The functional implication of the proximity between the hinge and Scc2 will need further investigation. Scc2 is crucial to stimulate the ATPase activity mediated by cohesin head engagement. The previous section of this thesis showed the possible implication of Scc2 binding to the head of Smc1. If, as suggested by the cryo EM structure, Scc2 can bind to both Smc1 head and the hinge simultaneously when cohesin is in the folded state, it would be interesting to analyse the status of head

engagement in this conformation. *In vivo* crosslinking using a combination of cysteine pairs will be the best tool to answer this question by addressing if the crosslink of the hinge with the coiled coil is compatible with the crosslink between the heads in an engaged or juxtaposed state.

The characterisation of the different conformations of cohesin including the effects of HAWK binding on the folded state, head engagement and coiled coil unzipping, will be key to understanding the mechanistic basis of how cohesin organises the genome.

Chapter 5: Summary and conclusions

The association of the cohesin tripartite ring with the loading complex is essential for the interaction with chromatin during the cell cycle. In this thesis, the interaction between Scc2 and cohesin has been investigated using both genetic and biochemistry techniques in a yeast system. I have shown that Scc2 interacts with Scc1 across multiple binding sites in the N terminal half of the protein. Interestingly, as it has been shown that Scc2 and Pds5 compete for binding to cohesin, yet these two proteins do not share a physical binding site within Scc1. A discrete binding site has been identified by introduction of a cysteine crosslinking pair between Scc2A709C and Scc1E203C.

By reinvestigating the Pds5 binding site on Scc1 I could show that this interaction is crucial for the stabilisation of cohesin at the pericentromeric borders which align with convergent genes (Paldi *et al* 2020).

The SMC heads are capable of binding and hydrolysing two molecules of ATP during each cycle, this ATPase activity is driven by Scc2 in an unknown mechanism. In this thesis I have shown that Scc2 interacts with a loop in Smc1 head that is essential for both head engagement and cohesin's interaction with chromatin. This interaction, that is conserved in Condensin will need further investigation to address its role in cohesin regulation.

The shape of the cohesin trimer has been visualised by several different *in vitro* techniques, here for the first time, the folded conformation of cohesin has been demonstrated *in vivo* using BMOE chemical crosslinking. This interaction occurs throughout the cell cycle, in both acetylated and non-acetylated cohesin and suggest

that the HAWKs may not be involved in the regulation of this configuration. However, several experiments have shown that the N-terminal of Pds5, Scc2 and Scc3 interact with the hinge questioning the role of these interaction in the regulation of the cohesin cycle (Burman *et al* 2019; Bin Hu, personal communication).

Taken together these results show that the interaction between Scc2 and the cohesin trimer is complex, with several binding motifs in the kleisin subunit and also the F-loop in the Smc1 head. The different cysteine pairs identified here will be a valuable tool to address the configuration of the cohesin complex during its cycle and how Scc2 regulates its activity.

Chapter 6: Materials and Methods

Yeast strains and growth conditions

All strains were derived from W303 and were grown at 25°C in rich growth medium (YPD), unless otherwise stated. YEP media was supplemented with 2% glucose (YPD) before use. Cultures were agitated at 200rpm (Multitron standard, Infors HT).

Where cells were arrested in early G1, the pheromone α-factor was added to a final concentration of 2mg/L/h, every 30 minutes for 2.5 hrs (Bücking-Throm, E. Duntze, W., Hartwell, L.H., and Manney, T. R, 1973). Cells were checked by microscopy for Shmoo cells >90%. Release from G1 (t=0) was achieved by filtration of the cells captured on 1.2 µm filter paper (Whatman ® GE Healthcare), washed with 1L of YPD and resuspended in YPD (or appropriate growth media). Where cells were arrested in G2, nocodozole (Sigma) was added at a final concentration of 10 µg/mL. Cells were incubated until synchronisation was achieved and >90% cells formed large buds as confirmed microscopically.

To create cells deficient in WT Scc1, the gene was placed under the MET3 repressible promoter. Cultures were grown in minimal media supplemented with 2% glucose and 1% -methionine drop out solution overnight. Cultures were diluted to OD600=0.2 and grown to OD600=0.4. Cells were then collected by filtration, washed with 1L YPD supplemented with 8mM methionine and arrested in early G1 (as above). Cells were then collected by filtration, washed with 1L YPD supplemented with 8mM methionine released into YPD supplemented with 8mM methionine.

Inactivation of scc1-73 was achieved by G1 arrest, and release into pre-warmed media at to 37°C for 60 minutes.

Flow Cytometry

1ml of culture was harvested by centrifugation at 4°C for 1minute at 13,300rpm. The cell pellet was resuspended in 1ml of cold 70% (v/v) ethanol and stored overnight at 4°C. Cells were centrifuged at 7,000rpm for 5 minutes and the cell pellet was resuspended in 1ml of 50mM Tris-HCL, pH 7.5 supplemented with 200ug/mL RNase A (Roche) and incubated for 2 hours at 37°C with agitation at 900rpm. The cells were centrifuged and resuspended in 400µL FACS buffer (180mM Tris-HCL, pH 7.5; 70mM $MgCl_2$; Propidium Iodide 50µg/mL; NaCl 190mM). Each sample was sonicated using a Vibracell^{TM} VCX 130FSJ (Sonics and Materials inc.), for 5 seconds at 40% output, to ensure dispersal. Cells were diluted 1:200 in 1mL of 50mM Tris-HCL, pH 7.5 in 5mL FACS tubes and analysis was performed by measuring 10,000 events per sample using a FACSCalibur^{TM} flow cytometer (BD Biosciences).

Site directed mutagenesis

A plasmid containing the gene of interest flanked by a marker cassette, was used as a template for the site directed mutagenesis PCR. Forward and reverse primers were designed with 9-12 base pairs before the mutation change, and 20-25 bp after. The annealing temperature was designed to be 63-65°C. A 50ul reaction was used with Phusion DNA polymerase (New England Biolabs) as per manufacturer's instructions. Reactions were subjected to the following thermocycling conditions:

Initial denaturation 98°C 2 minutes
Denaturation 98°C for 30 seconds
Annealing 65°C * for 30 seconds (*dependant on primer design)
Extension 72°C for 6:30 minutes (up to 9kb plasmid)
(Repeat for 35 cycles)
72°C for 10 minutes
Cool to 10°C

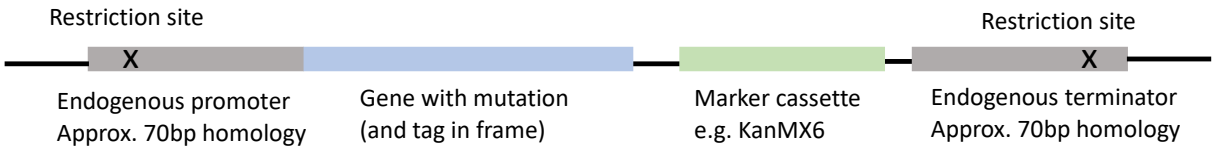
On completion of thermo cycling, 1µl DpnI enzyme (New England Biolabs) was added and samples incubated for 3 hours or overnight at 37°C. 10µl of the reaction was transformed into competent XL1 blue E. coli cells and plated on 2xTY media supplemented with ampicillin and grown overnight at 37°C. Resulting colonies were grown overnight in 5mL 2xTY media with ampicillin 1000X. The DNA was extracted using a miniprep kit (Qiagen) and the mutations were confirmed by sequencing.

Gibson assembly

Forward and reverse oligos were designed using the NEBuilder Assembly tool® with up to 25 base pairs of homology either side of the deletion or mutation. A Q5 high fidelity polymerase (New England Biolabs) was used following a standard protocol to amplify the fragments. Reactions were analysed on a 1.5% agarose gel to check fragment size. Gel bands were excised and extracted using a Zymo™ gel extraction kit. DNA concentration was measured and fragments assembled as per instructions in the Gibson Assembly kit® (New England Biolabs). 50°C for up to one hour. The plasmid was then transformed using the cells supplied, and plated on 2xTY media containing antibiotic (according to the resistance cassette).

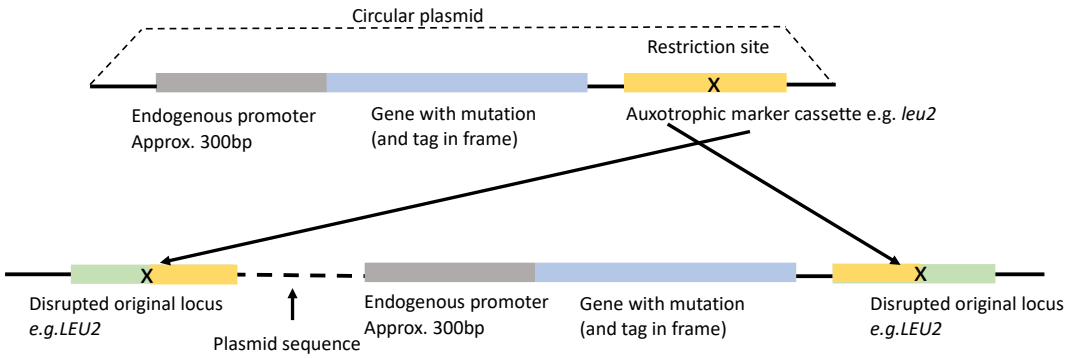
Genomic integration

To integrate mutant genes at the endogenous locus, a vector carrying the mutated gene and a marker cassette between the gene and the flanking terminator sequence was digested with appropriate enzymes leaving homology to both the endogenous gene promoter and terminator regions.



Correct integration was analysed by PCR on isolated genomic DNA using primers that flank the insertion sites in the endogenous promoter and terminator.

To integrate at an ectopic locus, mutated integrative plasmids (YIplac 128) were linearized by restriction enzyme digest at a single site within the selective auxotrophic marker (*leu2*) prior to transformation.



Integration into the ectopic loci (*LEU2)* was verified by PCR using specific primer sets designed in the plasmid that test for single or multiple copy number, or no integration.

Digests were analysed for correct size by gel electrophoresis. The DNA was used for transformation into yeast.

Yeast genomic DNA preparation

A toothpick of fresh cells was resuspended in 100µl of lysis buffer (1% SDS; 200mM Lithium acetate). Cell were incubated at 70 °C for 15 minutes with agitation at 900rpm. 300µl of 100% Ethanol was then added, the cells were mixed by inversion. Samples were then centrifuged for 3 minutes 13000rpm. The supernatant was completely removed and the pellet was resuspended in 100µl TE (100mM Tris; 50mM EDTA pH8.0). The cell debris was pelleted by a brief centrifugation and 1µl of the supernatant was used for a PCR reaction of 50µl.

Yeast transformation

Salmon sperm DNA (10mg/mL) was heat inactivated at 95°C for 5 minutes and cooled on ice before use. 50ml of exponentially growing cells OD_{600}=0.4-0.6 were harvested at 3,500 rpm for 5 minutes by centrifugation. Cells were washed twice in 100mM LiOAC and re-suspended in 200µl 100mM LiOAc. 50µl aliquots, one per reaction, were placed into 2ml round bottomed Eppendorf tubes and incubated at 30°C for 30 minutes. The following were added to each cell aliquot; 5ul of DNA (approximately 200ng-2µg), 45µl ddH2O, 36µl 1M LiOAc, 25µl and 240µl 50% (w/v) PEG 2000. Tubes were inverted to mix contents and incubated for 30 minutes at 30°C. Cells were then heat shocked at 42°C in a water bath for 20 minutes. Cells were centrifuged at 6,000rpm for 3 minutes, re-suspended in 8ml YPD and recovered for 2-

3hours at 25°C. Cells were collected by centrifugation 3,500rpm for 3 minutes, and re-suspended in 100µl YPD. Cells were then plated on to selective media or on YPD with antibiotic according to the drug resistance cassette. After 2-days, resulting colonies were streaked for single colonies on appropriate media. Single colonies were then patched on YPD and tested by PCR and sequencing for the correct mutation.

Tetrad dissection

Diploid yeast strains were placed onto nitrogen deficient media plates (SpoVB) to induce meiosis. Sporulation was assessed after 48hours at 25°C by microscopy, visualising tetrad formation. Sporulated cells were spheroplasted in 300 µl of dissection buffer (1M Sorbitol; 1.5mg/ml zymolase® T20 (Amsbio)). Cells were incubated for 30 minutes at room temperature without shaking. 10ul of cell mix was placed onto a YPD plate (or appropriate media plate) and dissected using a dissection microscope (MSM300, Singer instruments) and incubated for 2-3 days at 25°C. Growth and viability was assessed. Mutations were followed by replica plating onto a selection of appropriate media plates. All plates were grown for 24-48hrs and scored according to growth.

Preparation of native yeast cell extracts

Cells grown to exponential phase in liquid culture were harvested by centrifugation 3,500rpm for 5minutes and washed with 1xPBS. Cell pellets were resuspended in 200 µl of lysis buffer (150mMNaCl; 50mM Tris pH 7.5; 0.5% (v/v) NP-40; 5mM EDTA; 1.25mM DTT; 1mM PMSF; 1 tablet/50mL protease inhibitor cocktail (Roche)). 200 µl of acid washed beads (425-600 µm Sigma) was added. Cells were

lysed using a Fastprep®-24 benchtop homogeniser (MP Biomedicals) at 4°C (3x 60 seconds at 6.5m/s or until >90% cell lysis was achieved, assessed by microscopy). The insoluble fraction was pelleted by centrifugation at 13,300rpm at 4°C for 10 minutes, the soluble fraction was isolated.

Quantification of protein concentration in crude cell extracts

The Bradford protein assay was used to determine the concentration of samples (Bradford, 1970). Bradford reagent (dye concentrate, Bio-Rad) was diluted 1:4 and 1mL aliquoted into cuvettes. Serial diluted BSA was used to generate a standard curve and 1 µl sample was added to each 1mL aliquot of Bradford dye. Absorbance was measured at 595nm using an Ultraspec 3,300 pro spectrophotometer (Amersham Bioscience) for each sample relative to a blank control containing lysis buffer in diluted Bradford dye. The sample concentrations were calculated from the standard curve created.

Trichloroacitic acid (TCA) preparation of extracts

Cells were grown in log phase to OD_{600} = 0.4-0.6. 25 OD_{600} unit of cells were resuspended in 240 µl of pre-cooled 20% TCA (w/v) in a 2ml tube and kept on ice, an equal volume of acid washed beads (425-600 µm Sigma) was added. Cells were lysed using a Fastprep®-24 benchtop homogeniser (MP Biomedicals) at 4°C (3x 60 seconds at 6.5m/s or until >90% cell lysis was achieved, assessed by microscopy). The bottom of each 2ml tube was pierced with a hot needle and placed into a 5ml FACS tube. The supernatant was then collected by centrifugation at 3,500rpm at 4°C for 5 minutes in a sorvall centrifuge. The beads were washed with 1ml of cold 5% TCA. Samples were

transferred to a 2ml round bottomed tube and pelleted by centrifugation at 13,000rpm for 10 minutes at 4°C. The supernatant was discarded and the pellet was thoroughly resuspended in 200 µl of running buffer (100mM Tris HCl pH 8.8, 4% SDS, 20% Glycerol, 200mM DTT, 0.04% bromophenol blue), turning the sample yellow. 40 µl of 1M Tris HCl, pH 9.5 was added then added to the sample restoring its colour, this was briefly mixed by vortex spun. Samples were heated to 80°C for 5 minutes before loading onto a gel.

Protein gel electrophoresis and western blotting

Samples were mixed with 4X LDS sample buffer (NuPAGE® Life Technologies) and denatured at 95°C for 5 minutes and centrifuged for 5 minutes @13,300 rpm. Samples were loaded onto a 3-8% Tris acetate gel (NuPAGE® Life Technologies), with the appropriate protein ladder, the proteins were separated as per manufacturers recommendation. The proteins were then transferred to either 0.2µm nitrocellulose or PVDF membrane using Trans-blot® Turbo transfer pack for the Trans-blot® Turbo system (BioRad). The membrane was then blocked with 5% skimmed milk powder (w/v) in PBS with 0.05% Tween 20 for 30min. Incubation of the membrane with primary antibody (dilutions below) was achieved using 5% skimmed milk powder in PBS with 0.05% Tween 20 for 1.5h at RT or overnight at 4°C.

Primary antibody	Immunogen	Origin	Dilution
3F10 (Roche)	HA	Rat	1:1000
PGK1	PGK1	Mouse	1:20000
V5 (AbD Serotec)	PK	Mouse	1:2500
C-Myc (Santa Cruz)	Myc	Mouse	1:10000

Flag	Flag	Mouse	1:10000
*Smc3 acetylation	Acetyl	Mouse	1:1000

The membrane was washed in PBS+ 0.05%Tween 20 (2x 10 minutes). Secondary antibody was added to 5% skimmed milk powder (w/v) in PBS+ 0.05%Tween 20 (see dilution below) and incubated for 45 minutes.

Secondary antibody	Conjugate	Dilution
Goat anti mouse (Millipore)	HRP	1:5000
Goat anti Rat (Millipore)	HRP	1:5000

The membrane was then washed in PBS+ 0.05%Tween 20 (3x 10 minutes) and once in PBS. Millipore ECL was then used to detect protein using the ODYSSEY® Fc Imaging System (LI-COR).

For the anti Smc3 acetylation* western blot, PVDF membrane was used Trans-blot® Turbo transfer pack was used for transfer. The membrane was washed three times in 1 x buffer A (100mM Tris, 300mM NaCl adjusted to pH7.4 with HCl), before blocking with 5% Milk prepared in buffer A overnight at 4°C. The membrane was washed in 5% milk with buffer A +0.05% Tween 20 for 10 minutes, then briefly washed once in buffer A + 0.05% Tween 20 to remove excess milk and then incubated for 2-3hours at room temperature with anti-acetylation antibody, diluted in buffer A + 0.05% Tween 20. The membrane was then washed 1x10 minutes in buffer A and 2x 10 minutes in buffer A + 0.05%Tween 20. Secondary antibody was diluted in 5% milk prepared in buffer A + 0.05%Tween 20 and incubated for 1hour at room temperature. The membrane was then washed 1x10 minutes in buffer A and 1x 15 minutes in buffer A + 0.05%Tween and 1x15 minutes in buffer A. Millipore ECL was then used to detect protein using the ODYSSEY® Fc Imaging System (LI-COR).

In vivo cross-linking with BPA

A liquid culture was grown for 8 hours in –LEU-TRP media with BPA at 25°C. Cells were diluted to OD_{600}=0.05 in 150ml of –LEU-TRP media with BPA and grown overnight to $OD_{600=}$0.4-0.6. 2x 15 OD of cells were harvested by centrifugation at 3,500rpm for 5 minutes at 4°C. The pellet was washed twice in 20ml of PBS and collected by centrifugation at 3000rpm for 3 minutes at 4°C. The cells were resuspended in 1ml of ice cold PBS. One sample was transferred to a 2ml bead beater tube and placed on Ice. The second sample was transferred into one well of a 4-well dish, placed on ice and photo crosslinked by UV light at 360nm (3x, 300 seconds UV, then 5 minutes rest, on ice) using a Spectrolinker™ (Spectronics Corporation). The cells in the dish were then resuspended and transferred to a 2ml bead beater tube. All samples were collected by brief centrifugation and resuspended 300ul of lysis buffer as per native yeast cell extracts, above. An IP was then performed as below.

Immunoprecipitation

Sample concentrations were adjusted up to 750 µL with lysis buffer and 5µg of primary antibody was added. Samples were rotated at 4°C for 1hour before 30ul of washed protein-G dynabeads were added. Samples were rotated for 1 hour at 4°C and then pelleted by a brief centrifugation. Using a magnetic rack, the beads were collected and washed twice with lysis buffer. The supernatant was discarded and 50 µL of 1X sample buffer was added to the beads. Samples were boiled for 5 minutes, and the supernatant was moved to a new tube to remove the beads.

In vivo cysteine cross-linking using BMOE

15 OD units of cells grown in exponential phase were washed with ice-cold PBS and kept on ice throughout the experiment. Cells were resuspended in 500µL cold PBS and 300µL was added to 2 x 2mL bead beater tubes. 12.5µL BMOE (125mM in DMSO to a final concentration of 5mM) or 12.5µL DMSO was added before incubating on ice for 6 min. Cells were washed twice with 1mL cold PBS containing 5mM DTT. Cells were resuspended in 300 µL lysis buffer (25mM Hepes; 50mM KCl; 10mM MgSO4; 10mM Tri-Sodium Citrate; 25mM Sodium Sulfite; 0.25% Triton-X; 1mM PMSF; 2mM DTT; 0.11 tablet/50mL protease inhibitor cocktail (Roche); 0.1mg/mL RNase A (Roche)). The samples were then treated as native yeast cell extracts, above.

Multiple sequence alignment

Multiple sequence alignments were made using Clustal Omega (Sievers, 2011) and were processed using JalView (Waterhouse *et al*., 2009).

Structural analysis

Analysis of crystal and cryo-EM structures was performed using Chimera (Pettersen *et al.,* 2004)

Virus generation and protein expression

DNAs were first transformed into DH10Bac (Thermo Fisher) cells and bacmids containing the expression vector screened for by blue-white selection. DNA was then

extracted and 2 mg of bacmid DNA was transfected into 2 ml *S. frugiperda,* Sf9 cells (Thermo Fisher) at a cell density of 1 x10^6 cells ml^{-1} using FuGENE HD reagent (Promega), grown in Sf900 II SFM media (Thermo Fisher). These were then incubated at 27˚C for 5 days to create the P1 virus. P2 virus was then amplified by infecting 50 ml Sf9 cells at a density of 2 x10^6 cells ml^{-1} with 500 µL of P1 virus and incubating in the dark at 27˚C for 3 days with shaking at 100 rpm. P2 virus was then harvested by pelleting cells by centrifugation at 4000 g and decanting supernatant supplemented by 5% FBS (Sigma), and then stored in the dark at 4˚C. Proteins were then expressed by adding 5 ml of P2 virus to 500 ml Sf900 cells at a density of 2 x10^6 cells ml^{-1} and incubating in the dark at 27˚C for 2 days with shaking at 100 rpm. Cells were then harvested by centrifugation at 1000 g, washed with PBS, frozen in liquid nitrogen and stored at -80˚C.

Protein purification

Cell pellets were thawed in Buffer A (50 mM HEPES pH 7.5; 150 mM NaCl; 1 mM TCEP (Thermo Fisher); 5% glycerol) supplemented with 1 tablet of Complete Protease Inhibitor (EDTA-free) (Roche), 70 mg RNAse A (Roche), and 100 U ml^{1} Supernuclease (Sino Biological). Cells were then and then lysed by sonication and cell lysate was supplemented with 1 mM PMSF (Sigma). Proteins were then purified via a three-step purification protocol. First, proteins were purified via affinity pulldown of their StrepII tags using a StrepTrap HP column (Fisher Scientific) and eluted into Buffer A supplemented with 2.5 mM desthiobiotin (Fisher Scientific). Proteins were then further purified by anion exchange chromatography using a 5 ml HiTrap Q HP column (GE Healthcare) across a gradient of 100 mM to 1M NaCl. Finally, proteins were

purified via size exclusion chromatography using a Superose 6 increase 10/300 GL column (VWR) for cohesin trimers.

ATPase assay

ATPase activity was measured by using the EnzChek phosphate assay kit (Invitrogen) by following the protocol provided. Cohesin in various complexes and its subunits was added to a final concentration of 50nM (Cohesin trimer2.65µg; Scc2 1.2µg; Scc3 1.2µg) and carried out always under 50mM NaCl in the presence of 700nM 40bp dsDNA in those experiments testing the effect of duplex DNA. The reaction was started with addition of ATP to a final concentration of 1.3mM (final reaction volume: 150ul). After completion, a fraction of each reaction was run in SDS-PAGE and the gel stained with Coomassie brilliant blue in order to test that the complexes were intact throughout the experiment and that equal amounts were used when testing various mutants and conditions.

Crosslinking in Vitro

For protein crosslinking assays, 10 µl reactions were prepared containing 570 nM protein (buffered with 50 mM HEPES pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 1 mM TCEP, 5% glycerol). If added, ATP (Sigma) and pUC19 DNA were added to a concentration of 5 mM and 60 nM respectively. Reactions were first incubated on ice for 5 min before adding 1 µl BMOE (Thermo Fisher) to a final concentration of 0.64 mM. Reactions were then incubated on ice for 6 min. Samples were denatured by adding 4xLDS buffer and heating at 70°C for 10 min before being separated in 3–8%

Tris-acetate gels (Thermo Fisher) run at 150 V for 3 hr. Gels were stained with Quick Coomassie stain (Generon).

Live cell imaging

Exponentially growing cells were immobilised on 2% agarose pads mounted on glass slides. All imaging was performed at 25°C using a spinning disk confocal system controlled by the volocity^{TM} program (Perkin Elmer UltrVIEW) with an EMCCD camera (Hamamatsu) mounted on an Olympus IX8 microscope with an Olympus 100X 1.35 N.A. objective.

Indirect Immunofluorescence labelling of S. cerevisiae

Cells were grown asynchronously to OD_{600}=0.5. 10ml of cells were taken and fixed with 4% Formaldehyde (37%, Sigma) for 30 minutes at room temperature, with gentle mixing. Cells were pelleted by centrufigation at 800g for 3 minutes and washed with 1 mL of 0.1 M potassium phosphate, pH 6.5. and then twice with 1ml of 0.1 M potassium phosphate (pH 6.5); 1.2 M sorbitol and stored at 4°C. In situ immunofluorescence to detect MYC tagged Securin and GFP marked URA3 loci was carried out as described in (Silver et al., 2009). The fixed cells were spheroplasted by addition of 15 µL of 10 mg/mL Zymolyase T100 (Amsbio) and 5 µL beta-mercaptoethanol (Sigma) for 25 min at room temperature without agitation. Cells were collected by centrifugation at 800g and washed with 1ml of 0.1 M potassium phosphate (pH 6.5); 1.2 M sorbitol. The spheroplasts were resuspended in 1ml of with 1ml of 0.1 M potassium phosphate (pH 6.5); 1.2 M sorbitol and 25 µL was adhered onto a polylysine coated multiwell slide and permiabilised by incubation with 1% NP-

40 for 5 min. The slides were blocked with PBS (137 mM NaCl; 2.7 mM KCl; 10 mM Na_2HPO_4; 1.8 mM KH_2PO_4) containing 1% BSA (Sigma) and incubated overnight with PBS/1% BSA/anti MYC antibody at 4°C. The slides were washed 10 times with PBS/1% BSA and treated with fluorescently labelled secondary antibody for 2 hours at room temperature. The slides were mounted with DAPI containing mounting medium and observed using a spinning disk confocal system as above.

ChIP Sequencing

Cells were grown exponentially to OD_{600} = 0.5 and the required cell cycle stage where necessary. 15 OD_{600nm} units of *S. cerevisiae* cells were then mixed with 5 OD_{600nm} units of *C. glabrata* to a total volume of 45 mL and fixed with 4 mL of fixative (50 mM Tris-HCl, pH 8.0; 100 mM NaCl; 0.5 mM EGTA; 1 mM EDTA; 30% (v/v) formaldehyde) for 30 minutes at room temperature (RT) with rotation. The fixative was quenched with 2 mL of 2.5 M glycine at room temperature for 5 minutes with rotation. The cells were then harvested by centrifugation at 3,500 rpm for 3 minutes and washed with ice-cold PBS. The cells were then resuspended in 300 µL of ChIP lysis buffer (50 mM Hepes-KOH, pH 8.0; 140 mM NaCl; 1 mM EDTA; 1% (v/v) Triton X-100; 0.1% (w/v) sodium deoxycholate; 1 mM PMSF; 2X Complete protease inhibitor cocktail (Roche)) and an equal amount of acid-washed glass beads (425–600 µm, Sigma) added before cells were lysed using a FastPrep−24 benchtop homogeniser (MP Biomedicals) at 4°C (3 × 60 s at 6.5 m/s or until >90% of the cells were lysed as confirmed by microscopy).

The soluble fraction was isolated by centrifugation at 2,000 rpm for 3 minutes then sonicated using a bioruptor (Diagenode) for 30 minutes in bursts of 30 seconds

on/30 seconds off, at high level in a 4°C water bath to produce sheared chromatin with a size range of 200–1000 bp. After sonication, the samples were centrifuged at 13,200 rpm at 4°C for 20 minutes and the supernatant was transferred into 700 µL of ChIP lysis buffer. 30 µL of protein G Dynabeads (Invitrogen) were added and the samples were pre-cleared for 1 hour at 4°C. 80 µL of the supernatant was removed, this sample was termed 'whole cell extract' (WCE) and kept at 4°C. 5 µg of antibody (anti-PK (Bio-Rad) or anti-HA (Roche)) was added to the remaining supernatant which was then incubated overnight at 4°C. 50 µL of protein G Dynabeads were then added and incubated at 4°C for 2 hours before washing 2x with ChIP lysis buffer, 3x with high salt ChIP lysis buffer (50 mM Hepes-KOH, pH 8.0; 500 mM NaCl; 1 mM EDTA; 1% (v/v) Triton X-100; 0.1% (w/v) sodium deoxycholate; 1 mM PMSF), 2x with ChIP wash buffer (10 mM Tris-HCl, pH 8.0; 0.25 M LiCl; 0.5 % NP-40; 0.5% sodium deoxycholate; 1 mM EDTA; 1 mM PMSF) and 1x with TE pH7.5. The immunoprecipitated chromatin was then eluted by incubation in 120 µL TES buffer (50 mM Tris-HCl, pH 8.0; 10 mM EDTA; 1% SDS) for 15 min at 65°C and the collected supernatant termed 'IP sample'. The WCE samples were mixed with 40 µL of TES3 buffer (50 mM Tris-HCl, pH 8.0; 10 mM EDTA; 3% SDS) and all samples were de-crosslinked by incubation (static) at 65°C overnight. RNA was degraded by incubation with 2 µL RNase A (10 mg/mL; Roche) for 1 hour at 37°C and protein was removed by incubation with 10 µL of proteinase K (18 mg/mL; Roche) for 2 hours at 65°C. DNA was purified using ChIP DNA Clean and Concentrator kit (Zymo Research).

Preparation of sequencing libraries

Sequencing libraries were prepared using NEB Next Fast DNA Library Prep Set for Ion Torrent Kit (New England Biolabs) according to the manufacturer's instructions. Briefly, 10–100 ng of fragmented DNA was converted to blunt ends by end repair before ligation of the Ion Xpress Barcode Adaptors. Fragments of 300 bp were then selected using E-Gel Size Select 2% agarose gels (Life Technologies) and amplified with 6–8 PCR cycles. The DNA concentration was then determined by qPCR using Ion Torrent DNA standards (Kapa Biosystems) as a reference. 12–16 libraries with different barcodes could then be pooled together to a final concentration of 350 pM and loaded onto the Ion PI V3 Chip (Life Technologies) using the Ion Chef (Life Technologies). Sequencing was then completed on the Ion Torrent Proton (Life Technologies), typically producing 6–10 million reads per library with an average read length of 190 bp.

Data analysis, alignment and production of BigWigs

Data analysis was performed on the Galaxy platform. Quality of reads was assessed using FastQC (Galaxy tool version 1.0.0) and trimmed as required using 'trim sequences' (Galaxy tool version 1.0.0). Generally, this involved removing the first 10 bases and any bases after the 200th, but trimming more or fewer bases may be required to ensure the removal of kmers and that the per-base sequence content is equal across the reads. Reads shorter than 50 bp were removed using Filter FASTQ (Galaxy tool version 1.0.0, minimum size: 50, maximum size: 0, minimum quality: 0, maximum quality: 0, maximum number of bases allowed outside of quality range: 0, paired end data: false) and the remaining reads aligned to the necessary genome(s)

using Bowtie2 (Galaxy tool version 0.2) with the default (--sensitive) parameters (mate paired: single-end, write unaligned reads to separate file: true, reference genome: SacCer3 or CanGla, specify read group: false, parameter settings: full parameter list, type of alignment: end to end, preset option: sensitive, disallow gaps within *n*-positions of read: 4, trim *n*-bases from 5' of each read: 0, number of reads to be aligned: 0, strand directions: both, log mapping time: false). To generate alignments of reads that uniquely align to the *S. cerevisiae* genome, the reads were first aligned to the *C. glabrata* (CBS138, genolevures) genome with the unaligned reads saved as a separate file. These reads that could not be aligned to the *C. glabrata* genome were then aligned to the *S. cerevisiae* (sacCer3, SGD) genome and the resulting BAM file converted to BigWig (Galaxy tool version 0.1.0) for visualization. Similarly, this process was done with the order of genomes reversed to produce alignments of reads that uniquely align to *C. glabrata*.

Visualisation of ChIP-seq profiles

The resulting BigWigs were visualized using the IGB browser. To normalize the data to show quantitative ChIP signal the track was multiplied by the samples' occupancy ratio (OR) and normalized to 1 million reads using the graph multiply function. In order to calculate the average occupancy at each base pair up to 60 kb around all 16 centromeres, the BAM file that contains reads uniquely aligning to *S. cerevisiae* was separated into files for each chromosome using 'Filter SAM or BAM' (Galaxy tool version 1.1.0). A pileup of each chromosome was then obtained using samtools mpileup (Galaxy tool version 0.0.1) (source for reference list: locally cached, reference genome: SacCer3, genotype likelihood computation: false, advanced

options: basic). These files were then amended using our own script (chr_position.py) to assign all unrepresented genome positions a value of 0. Each pileup was then filtered using another in-house script (filter.py) to obtain the number of reads at each base pair within up to 60 kb intervals either side of the centromeric CDEIII elements of each chromosome. The number of reads covering each site as one successively moves away from these CDEIII elements could then be averaged across all 16 chromosomes and calibrated by multiplying by the samples' OR and normalizing to 1 million reads.

List of strains:

C. glabrata:

K23308	*MATa, SCC1-PK9::NatMX4*
K25532	*MATa, SCC1-HA6::NatMX4* (Gift from Wolfgang Zachariae)

S. cerevisiae:

K699	*MATa, W303, wildtype*
K700	*MAT∝, W303, wildtype*
K10110	*MATa, ura3:: tetOs3x112::URA3, his3::HIS3 tetR-GFP, Pds1myc18::TRP(K.Lactis)*
K13898	*MAT∝, scc1::KanMX4, ura3::SCC1::URA3*
K17184	*MATa, leu2::Pscc1-scc1-HA3::LEU2*
K17189	*MAT∝, leu2::Pscc1-scc1(V137K)-HA3::LEU2*
K18104	*MATa/∝ W303, wildtype*
K24422	*MATa eco1::NatMX4, scc3(K404E)::TRP1*
K25222	*MATa, scc1::MET-SCC1-MYC18::TRP, SCC2-PK6::KanMX6*
K25544	*MATa, smc1(N1192C)-HA6::HIS3, smc3(R1222C)-PK6::KanMX6*
K25896	*MAT∝, leu2::Pscc1-scc1(Δ152-256)-HA3::LEU2*
K25995	*MATa, leu2::Pscc1-scc1(Δ152-206)-HA3::LEU2*
K25997	*MATa, leu2::Pssc1-scc1(Δ207-256)-HA3::LEU2*
K26081	*MATa, smc1(K639C)-HA6::HIS3, smc3(E570C)::NatMX4*
K26103	*MAT∝, leu2::Pscc1-scc1(Δ124-206)-HA3::LEU2*
K26120	*MATa, leu2::Pscc1-scc1(Δ124-130)-HA3::LEU2*
K26140	*MATa, leu2::Pscc1-scc1(Δ103-123)-HA3::LEU2*
K26142	*MATa, leu2::Pscc1-scc1(Δ124-146)-HA3::LEU2*
K26144	*MATa, leu2::Pscc1-scc1(Δ131-138)-HA3::LEU2*
K26149	*MAT∝, his3:: smc3(K1032C)-MYC9::NatNT2::his3, leu2::Pscc1-scc1(Δ152-256)-HA3::LEU2*
K26189	*MAT∝, his3:: smc3(K1032C)-MYC9::NatNT2::his3, leu2::Pscc1-scc1-HA3::LEU2*
K26191	*MAT∝a, his3:: smc3(K1032C)-MYC9::NatNT2::his3, leu2::Pscc1-scc1(V137K)-HA3::LEU2*
K26248	*MATa, scc1::MET-SCC1-MYC18::TRP, leu2::SCC1-HA3::LEU2*

K26251	*MATa, scc1::MET-SCC1-MYC-18::TRP, leu2::Pscc1-scc1(V137K)-HA3::LEU2*
K26360	*MAT∝, leu2::scc1(Δ164-172, Δ207-256)-HA3::LEU2*
K26362	*MAT∝, leu2::Pscc1-scc1(Δ152-206, Δ244-251)-HA3::LEU2*
K26489	*MATa, scc1::MET-SCC1-MYC-18::TRP, leu2::Pscc1-scc1(Δ131-138)-HA3::LEU2*
K26986	*MATa, leu2::Pscc1-scc1(Δ244-251)-HA3::LEU2*
K26987	*MATa, leu2::Pscc1-scc1(Δ281-288)-HA3::LEU2*
K26988	*MATa*, *leu2::Pscc1-scc1(Δ196-203)-HA3::LEU2*
K26989	*MAT∝, leu2::Pscc1-scc1(Δ196-203, Δ281-288)-HA3::LEU2*
K26991	*MAT∝, leu2::Pscc1-scc1(Δ244-251, Δ281-288)-HA3::LEU2*
K26992	*MAT∝, leu2::Pscc1-scc1(Δ196-203, Δ244-251, Δ281-288)-HA3::LEU2*
K27115	*MATa*, *leu2::Pscc1-scc1(Δ196-203, Δ244-251)-HA3::LEU2*
K27321	*MAT∝, scc1::KanMX4, YCPlac33scc1::URA, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX6*
K27359	MATa, smc1(S161C)-HA6::HIS3, smc3(K160C)-PK6::KanMX4
K28051	*MATa, smc1(K1121D)-HA6::HIS3*
K28067	*MATa, smc1(L1120F)-HA6::HIS3*
K28068	*MATa, smc1(R1122E)-HA6::HIS3*
K28104	*MATa, smc1(F1123A)-HA6::HIS3*
K28105	*MATa, smc1(K1124D)-HA6::HIS3*
K28106	*MATa/∝, smc1(K1121D, R1122E)-HA6::HIS3*
K28107	*MATa/∝, smc1(R1122E, F1123A)-HA6::HIS3*
K28110	*MATa/∝, smc1(R1122E, K1124D)-HA6::HIS3*
K28111	*MATa/∝, smc1(F1123A K1124D)-HA6::HIS3*
K28167	*MATa/∝, smc1(L1120F, K1124D)-HA6::HIS3*
K28168	*MATa/∝, smc1(K1121D, F1123A)-HA6::HIS3*
K28169	*MATa/∝, smc1(K1121D, K1124D)-HA6::HIS3*
K28170	*MATa/∝, smc1(L1120F, F1123A)-HA6::HIS3*
K28171	*MATa/∝, smc1(L1120F, K1121D)-HA6::HIS3*
K28173	*MATa/∝, smc1(L1120F, R1122E)-HA6::HIS3*
K28174	*MATa, smc1 (K1124C)-HA6::HIS3*
K28211	*MATa, SMC1-PK3-aid::KanMX4, SCC1-PK6::KanMX6, leu2::SMC1::LEU2, , ura::ADH1 promoter-OsTIR1-myc9::URA3,*

K28212	*MATa, SMC1-PK3-aid::KanMX4, SCC1-PK6::KanMX6, leu2::Smc1(K1121D)::LEU2, , ura::ADH1 promoter-OsTIR1-9myc::URA3*
K28213	*MATa, SMC1-PK3-aid::KanMX4, SCC1-PK6::KanMX6, leu2::smc1 (K1121D,F1123A)::LEU2, ura::ADH1 promoter-OsTIR1-9myc::URA3*
K28214	*MATa, SMC1-PK3-aid::KanMX4, SCC1-PK6::KanMX6, leu2::smc1 (K1121D,K1124D)::LEU2, ura::ADH1 promoter-OsTIR1-9myc::URA3*
K28301	*MATa, smc1 (K1124C)-HA6::HIS3, scc2(D1202C)-PK6::NatMX4*
K28303	*MATa, smc1(K1124C)-HA6::HIS3, scc2(E1280C)-PK6::NatMX4*
K28304	*MATa, smc1 (K1124C)-HA6::HIS3, scc2(T1281C)-PK6::NatMX4*
K28341	*MATa, smc1 (K1124C)-HA6::HIS3, scc2(E1198C)-PK6::NatMX4*
K28344	MATa, *SMC1-HA6*::HIS3, *scc2(N1245C)-PK6::NatMX4*
K28343	*MATa, smc1 (K1124C)-HA6::HIS3, scc2(N1245C)-PK6::NatMX4*
K28349	*MATa smc1(A520C)-HA6::HIS3*
K28351	*MATa smc1(M521C)-HA6::HIS3*
K28353	*MATa smc1(R524C)-HA6::HIS3*
K28355	*MATa smc1(F526C)-HA6::HIS3*
K28357	*MATa smc1(K575C)-HA6::HIS3*
K28359	*MATa smc1(K576C)-HA6::HIS3*
K28361	*MATa smc1(Q577C)-HA6::HIS3*
K28363	*MATa smc1(R578C)-HA6::HIS3*
K28366	*MATα smc3(D927C)::KanMX4*
K28368	*MATα smc3(N928C)::KanMX4*
K28370	*MATα smc3(F929C)::KanMX4*
K28372	*MATα smc3(Q930C)::KanMX4*
K28374	*MATα smc3(V933C)::KanMX4*
K28376	*MATα smc3(E934C)::KanMX4*
K28378	*MATα smc3(T936C)::KanMX4*
K28379	*MATa smc1(R578C)-HA6::HIS3, smc3(N928C)::KanMX4*
K28381	*MATa smc1(R578C)-HA6::HIS3, smc3(F929C)::KanMX4*
K28383	*MATa smc1(R578C)-HA6::HIS3, smc3(Q930C)::KanMX4*
K28385	*MATa smc1(R578C)-HA6::HIS3, smc3(V933C)::KanMX4*
K28387	*MATa smc1(R578C)-HA6::HIS3, smc3(T936C)::KanMX4*
K28389	*MATa smc1(R578C)-HA6::HIS3, smc3(E934C)::KanMX4*
K28394	*MATα smc3(K931C)::KanMX4*

K28396	*MATα smc3(S932C)::KanMX4*
K28398	*MATα smc3(K935C)::KanMX4*
K28400	*MATα smc3(M937C)::KanMX4*
K28401	*MATa SMC1-HA6::HIS3, SMC3-PK6::KanMX6*
K28403	*MATa, smc1 (K1124C)-HA6::HIS3, scc2(L1192C)-PK6::NatMX4*
K28413	*MATa smc1(R578C)-HA6::HIS3, smc3(D927C)::KanMX4*
K28419	*MATa, ura3:: tetOs3x112::URA3, his3::HIS3 tetR-GFP, Pds1myc18::TRP(K.Lactis),smc1(K1121D)-HA6::HIS*
K28422	*MATa smc1(R578C)-HA6::HIS3, smc3(K931C)::KanMX4*
K28424	*MATa smc1(R578C)-HA6::HIS3, smc3(S932C)::KanMX4*
K28426	*MATa smc1(R578C)-HA6::HIS3, smc3(K935C)::KanMX4*
K28428	*MATa smc1(R578C)-HA6::HIS3, smc3(M937C)::KanMX4*
K28501	*MATa/∝, scc2(N1245E, P1246G)-PK6::NatMX4*
K28502	*MATa/∝, scc2(T1244C)-PK6::NatMX4*
K28503	*MATa/∝, Scc2(Y1435E G1436E)-PK6::NatMX4*
K28504	*MATa, Scc2(K1325E, K1329E)-PK6::NatMX4*
K28505	*MATa/∝, scc2(P1246G, S1247E)-PK6::NatMX4*
K28506	*MATa, Scc2(D1202A, G1203A)-PK6::NatMX4*
K28508	*MATa/∝, Scc2(T1210A, R1211A)-PK6::NatMX4*
K28509	*MATa/∝, Scc2(D1376A, L1377A)-PK6::NatMX4*
K28510	*MATa, Scc2(E1280R, T1281A)-PK6::NatMX4*
K28512	*MATa, smc1(K1124C)-HA6::HIS3, scc2(H1248C)-PK6::NatMX4*
K28546	*MATa smc1(R578C)-HA6::HIS3, SMC3-PK6::KanMX4*
K28576	*MATα smc3(V933C)-PK6::KanMX4*
K28583	*MATa SMC1-HA6::HIS3, smc3(V933C)-PK6::KanMX4*
K28585	*MATa smc1(R578C)-HA6::HIS3, smc3(V933C)-PK6::KanMX4*
K28614	*MATa, smc1(K1124C)-HA6::HIS3, scc2(P1246C)-PK6::NatMX4*
K28799	*MATa, smc1(K1124C)-HA6::HIS3, scc2(S1247C)-PK6::NatMX4*
K28830	*MATa/∝, scc2(S1193C)-PK6::NatMX4*
B1821	*MAT∝, smc3::KanMX4, SCC1-PK6::KanMX4, YCPlac33smc1::URA, SCC2-His6-FLAG6::KanMX4, pLH157, leu2::Psmc1-Smc3(Q67bpa)-HA3::Leu*
JM001	*MAT∝, scc1::KanMX4, SMC1-PK6::KanMX4, YCPlac33scc1::URASCC2-His6-FLAG6::KanMX4, pLH157 (tRNA synthetase),leu2::Pscc1-Scc1-HA3::Leu2*

JM002 *MAT∝, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(M196bpa)-HA3::leu2*

JM003 *MAT∝, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(N197bpa)-HA3::Leu2*

JM004 *MAT∝, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(L198bpa)-HA3::Leu2*

JM005 *MAT∝, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(D199bpa)-HA3::Leu2*

JM006 *MAT∝, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(I202bpa)-HA3::Leu2*

JM007 *MAT∝, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(E203bpa)-HA3::Leu2*

JM008 *MAT∝, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(G244bpa)-HA3::Leu2*

JM009 *MAT∝, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(W245bpa)-HA3::Leu2*

JM010 *MAT∝, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(D246bpa)-HA3::Leu2*

JM011 *MAT∝, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(L247bpa)-HA3::Leu2*

JM012 *MAT∝, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(G248bpa)-HA3::Leu2*

JM013 *MAT∝, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(I249bpa)-HA3::Leu2*

JM014 *MAT∝, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(T250bpa)-HA3::Leu2*

JM015 *MAT∝,, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(E251bpa)-*

	HA3::Leu2
JM016	*MATα,, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(D281bpa)-HA3::Leu2*
JM017	*MATα, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(F282bpa)-HA3::Leu2*
JM018	*MATα, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(G283bpa)-HA3::Leu2*
JM019	*MATα, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(F284bpa)-HA3::Leu2*
JM020	*MATα, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4,pLH157, leu2::Pscc1-scc1(L286bpa)-HA3::Leu2*
JM021	*MATα, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4, pLH157, leu2::Pscc1scc1(I288bpa)-HA3::Leu2*
JM022	*MATα, scc1::KanMX4, YCPlac33scc1::URA3, SMC1-PK6::KanMX4, SCC2-His6-FLAG6::KanMX4, pLH157*
JM030	*MATa, scc1::KanMX4, scc2::NatMX4, leu2::SCC1-HA3::LEU2, lys2::SCC2-PK6:: HphNT1::LYS2*
JM031	*MATa, scc1::KanMX4, scc2::NatMX4, leu2::scc1(E203C)-HA3::LEU2, lys::SCC2-PK6:: HphNT1::LYS2*
JM032	*MATa, scc1::KanMX4, scc2::NatMX4, leu2::scc1(E203C)-HA3::LEU2, lys::scc2(V554C)-PK6:: HphNT1::LYS2*
JM033	*MATa, scc1::KanMX4, scc2::NatMX4, leu2::SCC1-HA3::LEU2, lys::scc2(V554C)-PK6:: HphNT1::LYS2*
JM034	*MATa, scc1::KanMX4, scc2::NatMX4, leu2::scc1(E203C)-HA3::LEU2, lys::Scc2 (A709C)-PK6:: HphNT1::LYS2*
JM035	*MATa, scc1::KanMX4, scc2::NatMX4, leu2::SCC1-HA3::LEU2, lys::scc2(A709C)-PK6::HphNT1::LYS2*
JM036	*MATa, scc1::KanMX4, scc2::NatMX4, leu2::scc1(E203C)-HA3::LEU2, lys::scc2(D711C)-PK6:: HphNT1::LYS2*
JM037	*MATa, scc1::KanMX4, scc2::NatMX4, leu2::SCC1-HA3::LEU2, lys::scc2(D711C)-PK6:: HphNT1::LYS2*
JM038	*MATa, scc1::KanMX4, scc2::NatMX4, leu2::scc1(E203C)-HA3::LEU2, lys::scc2(I713C)-PK6:: HphNT1::LYS2*

JM039	*MATa, scc1::KanMX4, scc2::NatMX4, lleu2::SCC1-HA3::LEU2, lys::scc2(I713C)-PK6:: HphNT1::LYS2*
JM040	*MATa, scc1::KanMX4, scc2::NatMX4, leu2::scc1(E203C)-HA3::LEU2, lys::scc2(N777C)-PK6:: HphNT1::LYS2*
JM041	*MATa, scc1::KanMX4, scc2::NatMX4, leu2::SCC1-HA3::LEU2, lys::scc2(N777C)-PK6:: HphNT1::LYS2*
JM042	*MATa, scc1::KanMX4, scc2::NatMX4, leu2::scc1(E203C)-HA3::LEU2, lys::scc2(D782C)-PK6:: HphNT1::LYS2*
JM043	*MATa, scc1::KanMX4, scc2::NatMX4, leu2::SCC1-HA3::LEU2, lys::scc2(D782C)-PK6:: HphNT1::LYS2*
JM044	*MATa, scc1::KanMX4, scc2::NatMX4, leu2::scc1(E203C)-HA3::LEU2, lys::scc2(E819C)-PK6:: HphNT1::LYS2*
JM045	*MATa, scc1::KanMX4, scc2::NatMX4, leu2::SCC1-HA3::LEU2, lys::scc2(E819C)-PK6:: HphNT1::LYS2*

Plasmids for protein expression

pACEbac1 SMC1-8xHis-SMC3 / pIDC scc1-2xStrepII

pACEbac1 SMC1-8xHis-SMC3 / pIDC scc1-2xStrepII-SCC3

pACEbac1 SMC1-8xHis-SMC3 / pIDC scc1(V137K)-2xStrepII-SCC3

pACEbac1 SMC1-8xHis-SMC3 / pIDC scc1(Δ152-256) -2xStrepII-SCC3

pACEbac1 SMC1-8xHis-SMC3 / pIDC scc1(Δ152-206, Δ244-251) -2xStrepII-SCC3

pACEbac1 SMC1-8xHis-SMC3 / pIDC scc1(Δ196-203, Δ244-251)-2xStrepII-SCC3

pACEbac1 SMC1-8xHis-SMC3 / pIDC scc1(Δ196-203, Δ244-251, Δ281-288)-2xStrepII-SCC3

pACEbac1 SMC1(K1121D,F1123A,K1124D)-8xHis-SMC3 / pIDC scc1-2xStrepII

pACEbac1 SMC1(K1121D)-8xHis-SMC3 / pIDC scc1-2xStrepII

pACEbac1 SMC1(K1121D,F1123A,K1124D,N1192C)-8xHis-SMC3(R1222C) / pIDC scc1-2xStrepII

pACEbac1 SMC1(N1192C)-8xHis-SMC3(R1222C) / pIDC scc1-2xStrepII

pACEbac1 2xStrepII-SCC3

pACEbac1 Scc2$^{133-1493}$-2xStrepII

pGEX6p-1 GST-Scc2$^{171-1504}$

pCS2 Scc1$^{1-566}$

pCS2 Scc1$^{1-566}$ (Δ152-256)

pCS2 Scc1$^{1-256}$

References

Anderson, D. E., Losada, A., Erickson, H. P., & Hirano, T. (2002). Condensin and cohesin display different arm conformations with characteristic hinge angles. *The Journal of cell biology*, *156*(3), 419–424. https://doi.org/10.1083/jcb.200111002

Andrade, M. A., & Bork, P. (1995). HEAT repeats in the Huntington's disease protein. *Nature genetics*, *11*(2), 115–116. https://doi.org/10.1038/ng1095-115

Arumugam P, Gruber S, Tanaka K, Haering CH, Mechtler K, Nasmyth K (2003) ATP hydrolysis is required for cohesin's association with chromosomes. *Curr Biol* 13: 1941-1953

Avagliano, L., Parenti, I., Grazioli, P., Di Fede, E., Parodi, C., Mariani, M., Kaiser, F. J., Selicorni, A., Gervasini, C., & Massa, V. (2020). Chromatinopathies: A focus on Cornelia de Lange syndrome. *Clinical genetics*, *97*(1), 3–11. https://doi.org/10.1111/cge.13674

Banigan, E. J., van den Berg, A. A., Brandão, H. B., Marko, J. F., & Mirny, L. A. (2020). Chromosome organization by one-sided and two-sided loop extrusion. eLife, 9, e53558. https://doi.org/10.7554/eLife.53558

Bausch, C., Noone, S., Henry, J. M., Gaudenz, K., Sanderson, B., Seidel, C., & Gerton, J. L. (2007). Transcription alters chromosomal locations of cohesin in Saccharomyces cerevisiae. *Molecular and cellular biology*, *27*(24), 8522–8532. https://doi.org/10.1128/MCB.01007-07

Beckouet F, Srinivasan M, Roig MB, Chan KL, Scheinost JC, Batty P, Hu B, Petela N, Gligoris T, Smith AC, Strmecki L, Rowland BD, Nasmyth K (2016) Releasing Activity Disengages Cohesin's Smc3/Scc1 Interface in a Process Blocked by Acetylation. *Mol Cell* 61: 563-574

Beckouët, F., Hu, B., Roig, M. B., Sutani, T., Komata, M., Uluocak, P., Katis, V. L., Shirahige, K., & Nasmyth, K. (2010). An Smc3 acetylation cycle is essential for establishment of sister chromatid cohesion. *Molecular cell*, *39*(5), 689–699. https://doi.org/10.1016/j.molcel.2010.08.008

Björkman, A., Du, L., van der Burg, M., Cormier-Daire, V., Borck, G., Pié, J., Anderlid, B. M., Hammarström, L., Ström, L., de Villartay, J. P., Kipling, D., Dunn Walters, D., & Pan-Hammarström, Q. (2018). Reduced immunoglobulin gene diversity in patients with Cornelia de Lange syndrome. The Journal of allergy and clinical immunology, 141(1), 408–411.e8. https://doi.org/10.1016/j.jaci.2017.06.043

Borges, V., Lehane, C., Lopez-Serra, L., Flynn, H., Skehel, M., Rolef Ben-Shahar, T., & Uhlmann, F. (2010). Hos1 deacetylates Smc3 to close the cohesin acetylation cycle. Molecular cell, 39(5), 677–688. https://doi.org/10.1016/j.molcel.2010.08.009

Bose, T., & Gerton, J. L. (2010). Cohesinopathies, gene expression, and chromatin organization. *The Journal of cell biology*, *189*(2), 201–210. https://doi.org/10.1083/jcb.200912129

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248-254

Brooker, A. S., & Berkowitz, K. M. (2014). The roles of cohesins in mitosis, meiosis, and human health and disease. *Methods in molecular biology (Clifton, N.J.)*, *1170*, 229–266. https://doi.org/10.1007/978-1-4939-0888-2_11

Buheitel, J., & Stemmann, O. (2013). Prophase pathway-dependent removal of cohesin from human chromosomes requires opening of the Smc3-Scc1 gate. *The EMBO journal*, *32*(5), 666–676. https://doi.org/10.1038/emboj.2013.7

Bürmann, F., Lee, B. G., Than, T., Sinn, L., O'Reilly, F. J., Yatskevich, S., Rappsilber, J., Hu, B., Nasmyth, K., & Löwe, J. (2019). A folded conformation of MukBEF and cohesin. *Nature structural & molecular biology*, *26*(3), 227–236. https://doi.org/10.1038/s41594-019-0196-z

Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A., & Nasmyth, K. (2000). Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. *Molecular cell*, *5*(2), 243–254. https://doi.org/10.1016/s1097-2765(00)80420-7

Chan KL, Gligoris T, Upcher W, Kato Y, Shirahige K, Nasmyth K, Beckouet F (2013) Pds5 promotes and protects cohesin acetylation. *Proc Natl Acad Sci U S A* 110: 13020-13025

Chan KL, Roig MB, Hu B, Beckouet F, Metson J, Nasmyth K (2012) Cohesin's DNA exit gate is distinct from its entrance gate and is regulated by acetylation. *Cell* 150: 961-974

Chao, W. C., Murayama, Y., Muñoz, S., Jones, A. W., Wade, B. O., Purkiss, A. G., Hu, X. W., Borg, A., Snijders, A. P., Uhlmann, F., & Singleton, M. R. (2017). Structure of the cohesin loader Scc2. *Nature communications*, *8*, 13952. https://doi.org/10.1038/ncomms13952

Chapard, C., Jones, R., van Oepen, T., Scheinost, J. C., & Nasmyth, K. (2019). Sister DNA Entrapment between Juxtaposed Smc Heads and Kleisin of the Cohesin Complex. *Molecular cell*, *75*(2), 224–237.e5. https://doi.org/10.1016/j.molcel.2019.05.023

Chen, H. T., Warfield, L., & Hahn, S. (2007). The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nature structural & molecular biology*, *14*(8), 696–703. https://doi.org/10.1038/nsmb1272

Chin, J. W., Martin, A. B., King, D. S., Wang, L., & Schultz, P. G. (2002). Addition of a photocrosslinking amino acid to the genetic code of Escherichiacoli. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(17), 11020–11024. https://doi.org/10.1073/pnas.172226299

Chiu, A., Revenkova, E., & Jessberger, R. (2004). DNA interaction and dimerization of eukaryotic SMC hinge domains. *The Journal of biological chemistry*, *279*(25), 26233–26242. https://doi.org/10.1074/jbc.M402439200

Cobbe N, Heck MM (2006) The evolution of ATPase activity in SMC proteins. *Proteins* 63: 685-696

Cobbe, N., & Heck, M. M. (2000). Review: SMCs in the world of chromosome biology-from prokaryotes to higher eukaryotes. *Journal of structural biology*, *129*(2-3), 123–143. https://doi.org/10.1006/jsbi.2000.4255

Collier, J. E., Lee, B. G., Roig, M. B., Yatskevich, S., Petela, N. J., Metson, J., Voulgaris, M., Gonzalez Llamazares, A., Löwe, J., & Nasmyth, K. A. (2020). Transport of DNA within cohesin involves clamping on top of engaged heads by Scc2 and entrapment within the ring by Scc3. *eLife*, *9*, e59560. https://doi.org/10.7554/eLife.59560

Colombo, E. A., Mutlu-Albayrak, H., Shafeghati, Y., Balasar, M., Piard, J., Gentilini, D., Di Blasio, A. M., Gervasini, C., Van Maldergem, L., & Larizza, L. (2019). Phenotypic Overlap of Roberts and Baller-Gerold Syndromes in Two Patients With Craniosynostosis, Limb Reductions, and *ESCO2* Mutations. *Frontiers in pediatrics*, *7*, 210. https://doi.org/10.3389/fped.2019.00210

Dai, H. Q., Hu, H., Lou, J., Ye, A. Y., Ba, Z., Zhang, X., Zhang, Y., Zhao, L., Yoon, H. S., Chapdelaine-Williams, A. M., Kyritsis, N., Chen, H., Johnson, K., Lin, S., Conte, A., Casellas, R., Lee, C. S., & Alt, F. W. (2021). Loop extrusion mediates physiological Igh locus contraction for RAG scanning. *Nature*, *590*(7845), 338–343. https://doi.org/10.1038/s41586-020-03121-7

Dauban, L., Montagne, R., Thierry, A., Lazar-Stefanita, L., Bastié, N., Gadal, O., Cournac, A., Koszul, R., & Beckouët, F. (2020). Regulation of Cohesin-Mediated Chromosome Folding by Eco1 and Other Partners. *Molecular cell*, *77*(6), 1279–1293.e4. https://doi.org/10.1016/j.molcel.2020.01.019

Davidson, I. F., Bauer, B., Goetz, D., Tang, W., Wutz, G., & Peters, J. M. (2019). DNA loop extrusion by human cohesin. *Science (New York, N.Y.)*, *366*(6471), 1338–1345. https://doi.org/10.1126/science.aaz3418

Davidson IF, Goetz D, Zaczek MP, et al. Rapid movement and transcriptional re-localization of human cohesin on DNA. EMBO J. 2016 Dec 15;35(24):2671–2685. PubMed PMID: 27799150; PubMed Central PMCID: PMCPMC5167347.

Dhar, G., Sanders, E. R., & Johnson, R. C. (2004). Architecture of the hin synaptic complex during recombination: the recombinase subunits translocate with the DNA strands. *Cell*, *119*(1), 33–45. https://doi.org/10.1016/j.cell.2004.09.010

Diebold-Durand ML, Lee H, Ruiz Avila LB, Noh H, Shin HC, Im H, Bock FP, Burmann F, Durand A, Basfeld A, Ham S, Basquin J, Oh BH, Gruber S (2017) Structure of Full-Length SMC and Rearrangements Required for Chromosome Organization. *Mol Cell* 67: 334-347 e335

Dixon, S, E. (2013) Towards understanding the mechanism of cohesin loading. Thesis, University of Oxford.

Eckert, C. A., Gravdahl, D. J., & Megee, P. C. (2007). The enhancement of pericentromeric cohesin association by conserved kinetochore components promotes high-fidelity chromosome segregation and is sensitive to microtubule-based tension. *Genes & development*, *21*(3), 278–291.

Eichinger, C. S., Kurze, A., Oliveira, R. A., & Nasmyth, K. (2013). Disengaging the Smc3/kleisin interface releases cohesin from Drosophila chromosomes during

interphase and mitosis. The EMBO journal, 32(5), 656–665. https://doi.org/10.1038/emboj.2012.346

Elbatsh AMO, Haarhuis JHI, Petela N, Chapard C, Fish A, Celie PH, Stadnik M, Ristic D, Wyman C, Medema RH, Nasmyth K, Rowland BD (2016) Cohesin Releases DNA through Asymmetric ATPase-Driven Ring Opening. *Mol Cell* 61: 575-588

Fernius, J., & Marston, A. L. (2009). Establishment of cohesion at the pericentromere by the Ctf19 kinetochore subcomplex and the replication fork-associated factor, Csm3. *PLoS genetics*, *5*(9), e1000629. https://doi.org/10.1371/journal.pgen.1000629

Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA (2016) Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* 15: 2038-2049

Fuentes-Perez ME, Gwynn EJ, Dillingham MS, Moreno-Herrero F. Using DNA as a fiducial marker to study SMC complex interactions with the atomic force microscope. Biophys J. 2012 Feb 22;102(4):839-48. doi: 10.1016/j.bpj.2012.01.022. Epub 2012 Feb 21. PMID: 22385855; PMCID: PMC3283809.

Funabiki, H., Yamano, H., Kumada, K., Nagao, K., Hunt, T., & Yanagida, M. (1996). Cut2 proteolysis required for sister-chromatid seperation in fission yeast. Nature, 381(6581), 438–441. https://doi.org/10.1038/381438a0

Furuyama, S., & Biggins, S. (2007). Centromere identity is specified by a single centromeric nucleosome in budding yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(37), 14706–14711. https://doi.org/10.1073/pnas.0706985104

Gandhi, R., Gillespie, P. J., & Hirano, T. (2006). Human Wapl is a cohesin-binding protein that promotes sister-chromatid resolution in mitotic prophase. *Current biology : CB*, *16*(24), 2406–2417. https://doi.org/10.1016/j.cub.2006.10.061

Ganji, M., Shaltiel, I. A., Bisht, S., Kim, E., Kalichava, A., Haering, C. H., & Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. Science (New York, N.Y.), 360(6384), 102–105. https://doi.org/10.1126/science.aar7831

Gaudet, R., & Wiley, D. C. (2001). Structure of the ABC ATPase domain of human TAP1, the transporter associated with antigen processing. The EMBO journal, 20(17), 4964–4972. https://doi.org/10.1093/emboj/20.17.4964

Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W. J., & Nekrutenko, A. (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome research*, *15*(10), 1451–1455. https://doi.org/10.1101/gr.4086505

Gerlich, D., Koch, B., Dupeux, F., Peters, J. M., & Ellenberg, J. (2006). Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication. Current biology : CB, 16(15), 1571–1578. https://doi.org/10.1016/j.cub.2006.06.068

Gligoris T, Lowe J (2016) Structural Insights into Ring Formation of Cohesin and Related Smc Complexes. *Trends Cell Biol* 26: 680-693

Gligoris, T. G., Scheinost, J. C., Bürmann, F., Petela, N., Chan, K. L., Uluocak, P., Beckouët, F., Gruber, S., Nasmyth, K., & Löwe, J. (2014). Closing the cohesin ring:

structure and function of its Smc3-kleisin interface. *Science (New York, N.Y.)*, *346*(6212), 963–967. https://doi.org/10.1126/science.1256917

Andres Gonzalez (personal communication)

Gruber S, Arumugam P, Katou Y, Kuglitsch D, Helmhart W, Shirahige K, 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., & Nasmyth, K. (2003). Chromosomal cohesin forms a ring. *Cell*, *112*(6), 765–777. https://doi.org/10.1016/s0092-8674(03)00162-4

Guacci, V., Koshland, D., & Strunnikov, A. (1997). A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in S. cerevisiae. *Cell*, *91*(1), 47–57. https://doi.org/10.1016/s0092-8674(01)80008-8

Gutierrez-Escribano, P., Newton, M. D., Llauró, A., Huber, J., Tanasie, L., Davy, J., Aly, I., Aramayo, R., Montoya, A., Kramer, H., Stigler, J., Rueda, D. S., & Aragon, L. (2019). A conserved ATP- and Scc2/4-dependent activity for cohesin in tethering DNA molecules. Science advances, 5(11), eaay6804. https://doi.org/10.1126/sciadv.aay6804

Haering, C. H., & Gruber, S. (2016). SnapShot: SMC Protein Complexes Part I. *Cell*, *164*(1-2), 326–326.e1. https://doi.org/10.1016/j.cell.2015.12.026

Haering, C. H., Farcas, A. M., Arumugam, P., Metson, J., & Nasmyth, K. (2008). The cohesin ring concatenates sister DNA molecules. *Nature*, *454*(7202), 297–301. https://doi.org/10.1038/nature07098

Haering CH, Schoffnegger D, Nishino T, Helmhart W, Nasmyth K, Lowe J (2004) Structure and stability of cohesin's Smc1-kleisin interaction. *Mol Cell* 15: 951-964

Haering CH, Lowe J, Hochwagen A, Nasmyth K (2002) Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol Cell* 9: 773-788

Hartman, T., Stead, K., Koshland, D., and Guacci, V. (2000). Pds5p is an essential chromosomal protein required for both sister chromatid cohesion and condensation in Saccharomyces cerevisiae. J. Cell Biol. 151, 613–626.

Hassler, M., Shaltiel, I. A., Kschonsak, M., Simon, B., Merkel, F., Thärichen, L., Bailey, H. J., Macošek, J., Bravo, S., Metz, J., Hennig, J., & Haering, C. H. (2019). Structural Basis of an Asymmetric Condensin ATPase Cycle. *Molecular cell*, *74*(6), 1175–1188.e9. https://doi.org/10.1016/j.molcel.2019.03.037

Hassler M, Shaltiel IA, Haering CH (2018) Towards a Unified Model of SMC Complex Function. *Curr Biol* 28: R1266-R1281

Henikoff, S., & Henikoff, J. G. (2012). "Point" centromeres of Saccharomyces harbor single centromere-specific nucleosomes. *Genetics*, *190*(4), 1575–1577. https://doi.org/10.1534/genetics.111.137711

Higashi T, Eickhoff P, Simoes J, Locke J, Nans A, Flynn H, Snijders A, Papageorgiou G, O'Reilly N, Chen Z, O'Reilly F, Rappsilber J, Costa A, Uhlmann F (2020) A Structure-Based Mechanism for DNA Entry into the Cohesin Ring. *BioRXiv*

Hill VK, Kim JS, Waldman T (2016) Cohesin mutations in human cancer. *Biochim Biophys Acta* 1866: 1-11

Hinshaw, S. M., Makrantoni, V., Harrison, S. C., & Marston, A. L. (2017). The Kinetochore Receptor for the Cohesin Loading Complex. *Cell*, *171*(1), 72–84.e13. https://doi.org/10.1016/j.cell.2017.08.017

Hinshaw, S. M., Makrantoni, V., Kerr, A., Marston, A. L., & Harrison, S. C. (2015). Structural evidence for Scc4-dependent localization of cohesin loading. *eLife*, *4*, e06057. https://doi.org/10.7554/eLife.06057

Hirano T. (2005). Condensins: organizing and segregating the genome. *Current biology : CB*, *15*(7), R265–R275. https://doi.org/10.1016/j.cub.2005.03.037

Hou, F., & Zou, H. (2005). Two human orthologues of Eco1/Ctf7 acetyltransferases are both required for proper sister-chromatid cohesion. *Molecular biology of the cell*, *16*(8), 3908–3918. https://doi.org/10.1091/mbc.e04-12-1063

Houlard, M., Godwin, J., Metson, J., Lee, J., Hirano, T., & Nasmyth, K. (2015). Condensin confers the longitudinal rigidity of chromosomes. *Nature cell biology*, *17*(6), 771–781. https://doi.org/10.1038/ncb3167

Hu, B. (2019) Personal communication)

Hu, B., Petela, N., Kurze, A., Chan, K. L., Chapard, C., & Nasmyth, K. (2015). Biological chromodynamics: a general method for measuring protein occupancy across the genome by calibrating ChIP-seq. *Nucleic acids research*, *43*(20), e132. https://doi.org/10.1093/nar/gkv670

Hu B, Itoh T, Mishra A, Katoh Y, Chan KL, Upcher W, Godlee C, Roig MB, Shirahige K, Nasmyth K (2011) ATP hydrolysis is required for relocating cohesin from sites occupied by its Scc2/4 loading complex. *Curr Biol* 21: 12-24

Ivanov, D., & Nasmyth, K. (2005). A topological interaction between cohesin rings and a circular minichromosome. *Cell*, *122*(6), 849–860. https://doi.org/10.1016/j.cell.2005.07.018

Jessberger R. (2002). The many functions of SMC proteins in chromosome dynamics. *Nature reviews. Molecular cell biology*, *3*(10), 767–778. https://doi.org/10.1038/nrm930

Kawauchi, S., Calof, A. L., Santos, R., Lopez-Burks, M. E., Young, C. M., Hoang, M. P., Chua, A., Lao, T., Lechner, M. S., Daniel, J. A., Nussenzweig, A., Kitzes, L., Yokomori, K., Hallgrimsson, B., & Lander, A. D. (2009). Multiple organ system defects and transcriptional dysregulation in the Nipbl(+/-) mouse, a model of Cornelia de Lange Syndrome. *PLoS genetics*, *5*(9), e1000650. https://doi.org/10.1371/journal.pgen.1000650

Kanke M, Tahara E, Huis In't Veld PJ, et al. Cohesin acetylation and Wapl-Pds5 oppositely regulate translocation of cohesin along DNA. EMBO J. 2016 Dec15;35(24):2686–2698. PubMed PMID: 27872142; PubMed Central PMCID: PMCPMC5167340. [Crossref], [PubMed], [Web of Science ®], [Google Scholar]

Kelly, A. E., & Funabiki, H. (2009). Correcting aberrant kinetochore microtubule attachments: an Aurora B-centric view. *Current opinion in cell biology*, *21*(1), 51–58. https://doi.org/10.1016/j.ceb.2009.01.004

Kikuchi S, Borek DM, Otwinowski Z, Tomchick DR, Yu H (2016) Crystal structure of the cohesin loader Scc2 and insight into cohesinopathy. *Proc Natl Acad Sci U S A* 113: 12444-12449

Kim Y, Shi Z, Zhang H, Finkelstein IJ, Yu H (2019) Human cohesin compacts DNA by loop extrusion. *Science* 366: 1345-1349

Kline, A. D., Moss, J. F., Selicorni, A., Bisgaard, A. M., Deardorff, M. A., Gillett, P. M., Ishman, S. L., Kerr, L. M., Levin, A. V., Mulder, P. A., Ramos, F. J., Wierzba, J., Ajmone, P. F., Axtell, D., Blagowidow, N., Cereda, A., Costantino, A., Cormier-Daire, V., FitzPatrick, D., Grados, M., … Hennekam, R. C. (2018). Diagnosis and management of Cornelia de Lange syndrome: first international consensus statement. *Nature reviews. Genetics*, *19*(10), 649–666. https://doi.org/10.1038/s41576-018-0031-0

Kong, M., Cutts, E. E., Pan, D., Beuron, F., Kaliyappan, T., Xue, C., Morris, E. P., Musacchio, A., Vannini, A., & Greene, E. C. (2020). Human Condensin I and II Drive Extensive ATP-Dependent Compaction of Nucleosome-Bound DNA. *Molecular cell*, *79*(1), 99–114.e9.

Kschonsak, M., Merkel, F., Bisht, S., Metz, J., Rybin, V., Hassler, M., & Haering, C. H. (2017). Structural Basis for a Safety-Belt Mechanism That Anchors Condensin to Chromosomes. *Cell*, *171*(3), 588–600.e24. https://doi.org/10.1016/j.cell.2017.09.008

Krantz, I. D., McCallum, J., DeScipio, C., Kaur, M., Gillis, L. A., Yaeger, D., Jukofsky, L., Wasserman, N., Bottani, A., Morris, C. A., Nowaczyk, M. J., Toriello, H., Bamshad, M. J., Carey, J. C., Rappaport, E., Kawauchi, S., Lander, A. D., Calof, A. L., Li, H. H., Devoto, M., … Jackson, L. G. (2004). Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. *Nature genetics*, *36*(6), 631–635. https://doi.org/10.1038/ng1364

Kurze, A., Michie, K. A., Dixon, S. E., Mishra, A., Itoh, T., Khalid, S., Strmecki, L., Shirahige, K., Haering, C. H., Löwe, J., & Nasmyth, K. (2011). A positively charged channel within the Smc1/Smc3 hinge required for sister chromatid cohesion. *The EMBO journal*, *30*(2), 364–378. https://doi.org/10.1038/emboj.2010.315

Kueng, S., Hegemann, B., Peters, B. H., Lipp, J. J., Schleiffer, A., Mechtler, K., & Peters, J. M. (2006). Wapl controls the dynamic association of cohesin with chromatin. *Cell*, *127*(5), 955–967. https://doi.org/10.1016/j.cell.2006.09.040

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lee, B. G., Roig, M. B., Jansma, M., Petela, N., Metson, J., Nasmyth, K., & Löwe, J. (2016). Crystal Structure of the Cohesin Gatekeeper Pds5 and in Complex with Kleisin Scc1. *Cell reports*, *14*(9), 2108–2115. https://doi.org/10.1016/j.celrep.2016.02.020

Lee, B. G., Merkel, F., Allegretti, M., Hassler, M., Cawood, C., Lecomte, L., O'Reilly, F. J., Sinn, L. R., Gutierrez-Escribano, P., Kschonsak, M., Bravo, S., Nakane, T., Rappsilber, J., Aragon, L., Beck, M., Löwe, J., & Haering, C. H. (2020). Cryo-EM structures of holo

condensin reveal a subunit flip-flop mechanism. *Nature structural & molecular biology*, *27*(8), 743–751. https://doi.org/10.1038/s41594-020-0457-x

Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G. P., Itoh, T., Watanabe, Y., Shirahige, K., & Uhlmann, F. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature*, *430*(6999), 573–578. https://doi.org/10.1038/nature02742

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079.

Li, Y., Muir, K. W., Bowler, M. W., Metz, J., Haering, C. H., & Panne, D. (2018). Structural basis for Scc3-dependent cohesin recruitment to chromatin. *eLife*, *7*, e38356. https://doi.org/10.7554/eLife.38356

Liu Y, Sung S, Kim Y, Li F, Gwon G, Jo A, Kim AK, Kim T, Song OK, Lee SE, Cho Y. ATP-dependent DNA binding, unwinding, and resection by the Mre11/Rad50 complex. The EMBO Journal. 2016;35:743–758. doi: 10.15252/embj.201592462

Losada, A., Hirano, M., & Hirano, T. (1998). Identification of Xenopus SMC protein complexes required for sister chromatid cohesion. *Genes & development*, *12*(13), 1986–1997. https://doi.org/10.1101/gad.12.13.1986

Marko J. F. (2008). Micromechanical studies of mitotic chromosomes. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology, 16(3), 469–497. https://doi.org/10.1007/s10577-008-1233-7

Marsden, M. P., & Laemmli, U. K. (1979). Metaphase chromosome structure: evidence for a radial loop model. *Cell*, *17*(4), 849–858. https://doi.org/10.1016/0092-8674(79)90325-8

Melby, T. E., Ciampaglio, C. N., Briscoe, G., & 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. *The Journal of cell biology*, *142*(6), 1595–1604. https://doi.org/10.1083/jcb.142.6.1595

Michaelis C, Ciosk R, Nasmyth K (1997) Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell* 91: 35-45

Minamino, M., Ishibashi, M., Nakato, R., Akiyama, K., Tanaka, H., Kato, Y., Negishi, L., Hirota, T., Sutani, T., Bando, M., & Shirahige, K. (2015). Esco1 Acetylates Cohesin via a Mechanism Different from That of Esco2. *Current biology : CB*, *25*(13), 1694–1706. https://doi.org/10.1016/j.cub.2015.05.017

Mishra, A., Hu, B., Kurze, A., Beckouët, F., Farcas, A. M., Dixon, S. E., Katou, Y., Khalid, S., Shirahige, K., & Nasmyth, K. (2010). Both interaction surfaces within cohesin's hinge domain are essential for its stable chromosomal association. *Current biology : CB*, *20*(4), 279–289. https://doi.org/10.1016/j.cub.2009.12.059

Morgan, D. (2007) The Cell cycle, *New Science Press.*

Muir KW, Li Y, Weis F, Panne D (2020) The structure of the cohesin ATPase elucidates the mechanism of SMC-kleisin ring opening. *Nat Struct Mol Biol* 27: 233-239

Muir, K. W., Kschonsak, M., Li, Y., Metz, J., Haering, C. H., & Panne, D. (2016). Structure of the Pds5-Scc1 Complex and Implications for Cohesin Function. *Cell reports*, *14*(9), 2116–2126. https://doi.org/10.1016/j.celrep.2016.01.078

Murayama Y, Uhlmann F. Biochemical reconstitution of topological DNA binding by the cohesin ring. Nature. 2014 Jan 16;505(7483):367–371. PubMed PMID: 24291789; PubMed Central PMCID: PMCPMC3907785.

Musacchio, A., & Salmon, E. D. (2007). The spindle-assembly checkpoint in space and time. *Nature reviews. Molecular cell biology*, *8*(5), 379–393. https://doi.org/10.1038/nrm2163

Nasmyth K (2001) Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu Rev Genet* 35: 673-745

Nasmyth K (2002) Segregating sister genomes: the molecular biology of chromosome separation. *Science* 297: 559-565

Nasmyth K, Haering CH (2009) Cohesin: its roles and mechanisms. *Annu Rev Genet* 43: 525-558

Nishimura K, Fukagawa T, Takisawa H, Kakimoto T, Kanemaki M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat Methods. 2009 Dec;6(12):917-22. doi: 10.1038/nmeth.1401. Epub 2009 Nov 15. PMID: 19915560.

Nishiyama, T., Ladurner, R., Schmitz, J., Kreidl, E., Schleiffer, A., Bhaskara, V., Bando, M., Shirahige, K., Hyman, A. A., Mechtler, K., & Peters, J. M. (2010). Sororin mediates sister chromatid cohesion by antagonizing Wapl. Cell, 143(5), 737–749. https://doi.org/10.1016/j.cell.2010.10.031

Ocampo-Hafalla, M., Muñoz, S., Samora, C. P., & Uhlmann, F. (2016). Evidence for cohesin sliding along budding yeast chromosomes. *Open biology*, *6*(6), 150178. https://doi.org/10.1098/rsob.150178

Ouyang, Z., Zheng, G., Tomchick, D. R., Luo, X., & Yu, H. (2016). Structural Basis and IP6 Requirement for Pds5-Dependent Cohesin Dynamics. *Molecular cell*, *62*(2), 248–259. https://doi.org/10.1016/j.molcel.2016.02.033

Ouyang, Z., Zheng, G., Song, J., Borek, D. M., Otwinowski, Z., Brautigam, C. A., Tomchick, D. R., Rankin, S., & Yu, H. (2013). Structure of the human cohesin inhibitor Wapl. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(28), 11355–11360.

Paldi, F., Alver, B., Robertson, D., Schalbetter, S. A., Kerr, A., Kelly, D. A., Baxter, J., Neale, M. J., & Marston, A. L. (2020). Convergent genes shape budding yeast pericentromeres. *Nature*, *582*(7810), 119–123. https://doi.org/10.1038/s41586-020-2244-6

Panizza, S., Tanaka, T., Hochwagen, A., Eisenhaber, F., & Nasmyth, K. (2000). Pds5 cooperates with cohesin in maintaining sister chromatid cohesion. Current biology : CB, 10(24), 1557–1564. https://doi.org/10.1016/s0960-9822(00)00854-x

Petela & Gonzalez, 2020 (under review/in preparation)

Petela, N. J., Gligoris, T. G., Metson, J., Lee, B. G., Voulgaris, M., Hu, B., Kikuchi, S., Chapard, C., Chen, W., Rajendra, E., Srinivisan, M., Yu, H., Löwe, J., & Nasmyth, K. A. (2018). Scc2 Is a Potent Activator of Cohesin's ATPase that Promotes Loading by Binding Scc1 without Pds5. *Molecular cell*, *70*(6), 1134–1148.e7. https://doi.org/10.1016/j.molcel.2018.05.022

Petela, N. J, (2016) Probing the cohesin loading reaction using forward genetic and quantitative ChIP sequencing. Thesis, University of Oxford.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Rajasekar, K. V., Baker, R., Fisher, G., Bolla, J. R., Mäkelä, J., Tang, M., Zawadzka, K., Koczy, O., Wagner, F., Robinson, C. V., Arciszewska, L. K., & Sherratt, D. J. (2019). Dynamic architecture of the Escherichia coli structural maintenance of chromosomes (SMC) complex, MukBEF. *Nucleic acids research*, *47*(18), 9696–9707. https://doi.org/10.1093/nar/gkz696

Rao, S., Huang, S. C., Glenn St Hilaire, B., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K. R., Sanborn, A. L., Johnstone, S. E., Bascom, G. D., Bochkov, I. D., Huang, X., Shamim, M. S., Shin, J., Turner, D., Ye, Z., Omer, A. D., Robinson, J. T., Schlick, T., Bernstein, B. E., Casellas, R., … Aiden, E. L. (2017). Cohesin Loss Eliminates All Loop Domains. *Cell*, *171*(2), 305–320.e24. https://doi.org/10.1016/j.cell.2017.09.026

Remeseiro, S., Cuadrado, A., & Losada, A. (2013). Cohesin in development and disease. *Development (Cambridge, England)*, *140*(18), 3715–3718. https://doi.org/10.1242/dev.090605

Roig MB, Lowe J, Chan KL, Beckouet F, Metson J, Nasmyth K (2014) Structure and function of cohesin's Scc3/SA regulatory subunit. *FEBS Lett* 588: 3692-3702

Rolef Ben-Shahar, T., Heeger, S., Lehane, C., East, P., Flynn, H., Skehel, M., & Uhlmann, F. (2008). Eco1-dependent cohesin acetylation during establishment of sister chromatid cohesion. *Science (New York, N.Y.)*, *321*(5888), 563–566. https://doi.org/10.1126/science.1157774

Shi Z, Gao H, Bai XC, Yu H (2020) Cryo-EM structure of the human cohesin-NIPBL-DNA complex. *Science*

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, *7*, 539. https://doi.org/10.1038/msb.2011.75

Silver P. (2009). Indirect immunofluorescence labeling in the yeast Saccharomyces cerevisiae. *Cold Spring Harbor protocols*, *2009*(11), pdb.prot5317. https://doi.org/10.1101/pdb.prot5317

Skibbens R. V. (2009). Establishment of sister chromatid cohesion. *Current biology : CB*, *19*(24), R1126–R1132. https://doi.org/10.1016/j.cub.2009.10.067

Soh, Y. M., Bürmann, F., Shin, H. C., Oda, T., Jin, K. S., Toseland, C. P., Kim, C., Lee, H., Kim, S. J., Kong, M. S., Durand-Diebold, M. L., Kim, Y. G., Kim, H. M., Lee, N. K., Sato, M., Oh, B. H., & Gruber, S. (2015). Molecular basis for SMC rod formation and its dissolution upon DNA binding. *Molecular cell*, *57*(2), 290–303. https://doi.org/10.1016/j.molcel.2014.11.023

Srinivasan, M., Scheinost, J. C., Petela, N. J., Gligoris, T. G., Wissler, M., Ogushi, S., Collier, J. E., Voulgaris, M., Kurze, A., Chan, K. L., Hu, B., Costanzo, V., & Nasmyth, K. A. (2018). The Cohesin Ring Uses Its Hinge to Organize DNA Using Non-topological as well as Topological Mechanisms. *Cell*, *173*(6), 1508–1519.e18. https://doi.org/10.1016/j.cell.2018.04.015

Srinivasan, M., Petela, N. J., Scheinost, J. C., Collier, J., Voulgaris, M., B Roig, M., Beckouët, F., Hu, B., & Nasmyth, K. A. (2019). Scc2 counteracts a Wapl-independent mechanism that releases cohesin from chromosomes during G1. *eLife*, *8*, e44736. https://doi.org/10.7554/eLife.44736

Stigler J, Camdere GO, Koshland DE, et al. Single-Molecule imaging reveals a collapsed conformational state for DNA-Bound cohesin. Cell Rep. 2016May 3;15(5):988–998. PubMed PMID: 27117417; PubMed Central PMCID: PMCPMC4856582.

Strunnikov, A. V., Larionov, V. L., & 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. *The Journal of cell biology*, *123*(6 Pt 2), 1635–1648. https://doi.org/10.1083/jcb.123.6.1635

Sutani, T., Kawaguchi, T., Kanno, R., Itoh, T., & Shirahige, K. (2009). Budding yeast Wpl1(Rad61)-Pds5 complex counteracts sister chromatid cohesion-establishing reaction. Current biology : CB, 19(6), 492–497. https://doi.org/10.1016/j.cub.2009.01.062

Tatebe, H., Lim, C.T., Konno, H. et al. Rad50 zinc hook functions as a constitutive dimerization module interchangeable with SMC hinge. (2020). Nat Commun 11, 370 https://doi.org/10.1038/s41467-019-14025-0

Tóth, A., Ciosk, R., Uhlmann, F., Galova, M., Schleiffer, A., & Nasmyth, K. (1999). Yeast cohesin complex requires a conserved protein, Eco1p(Ctf7), to establish cohesion between sister chromatids during DNA replication. *Genes & development*, *13*(3), 320–333. https://doi.org/10.1101/gad.13.3.320

Tonkin, E. T., Wang, T. J., Lisgo, S., Bamshad, M. J., & Strachan, T. (2004). NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. *Nature genetics*, *36*(6), 636–641. https://doi.org/10.1038/ng1363

Unal, E., Heidinger-Pauli, J. M., Kim, W., Guacci, V., Onn, I., Gygi, S. P., & Koshland, D. E. (2008). A molecular determinant for the establishment of sister chromatid cohesion. *Science (New York, N.Y.)*, *321*(5888), 566–569. https://doi.org/10.1126/science.1157880

Uhlmann, F., Wernic, D., Poupart, M. A., Koonin, E. V., & Nasmyth, K. (2000). Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. *Cell*, *103*(3), 375–386. https://doi.org/10.1016/s0092-8674(00)00130-6

Uhlmann, F., Lottspeich, F., & Nasmyth, K. (1999). Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1. *Nature*, *400*(6739), 37–42. https://doi.org/10.1038/21831

Vazquez Nunez R, Ruiz Avila LB, Gruber S. 2019. Transient DNA occupancy of the SMC interarm space in prokaryotic condensin. *Mol. Cell* 75:209–23

Verdon, G., Albers, S. V., van Oosterwijk, N., Dijkstra, B. W., Driessen, A. J., & Thunnissen, A. M. (2003). Formation of the productive ATP-Mg2+-bound dimer of GlcV, an ABC-ATPase from Sulfolobus solfataricus. *Journal of molecular biology*, *334*(2), 255–267. https://doi.org/10.1016/j.jmb.2003.08.065

Walker, J. E., Saraste, M., Runswick, M. J., & 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. *The EMBO journal*, *1*(8), 945–951.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M. and Barton, G. J. (2009) "Jalview Version 2 - a multiple sequence alignment editor and analysis workbench" Bioinformatics25 (9) 1189-1191 doi: 10.1093/bioinformatics/btp033

Wells, J. N., Gligoris, T. G., Nasmyth, K. A., & Marsh, J. A. (2017). Evolution of condensin and cohesin complexes driven by replacement of Kite by Hawk proteins. *Current biology : CB*, *27*(1), R17–R18. https://doi.org/10.1016/j.cub.2016.11.050

Wilhelm L, Burmann F, Minnen A, et al. SMC condensin entraps chromosomal DNA by an ATP hydrolysis dependent loading mechanism in Bacillus subtilis. Elife. 2015 May 7;4:e06659. PubMed PMID: 25951515; PubMed Central PMCID: PMCPMC4442127.

Wutz, G., Várnai, C., Nagasaka, K., Cisneros, D. A., Stocsits, R. R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, M. J., Walther, N., Koch, B., Kueblbeck, M., Ellenberg, J., Zuber, J., Fraser, P., & Peters, J. M. (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *The EMBO journal*, *36*(24), 3573–3599. https://doi.org/10.15252/embj.201798004

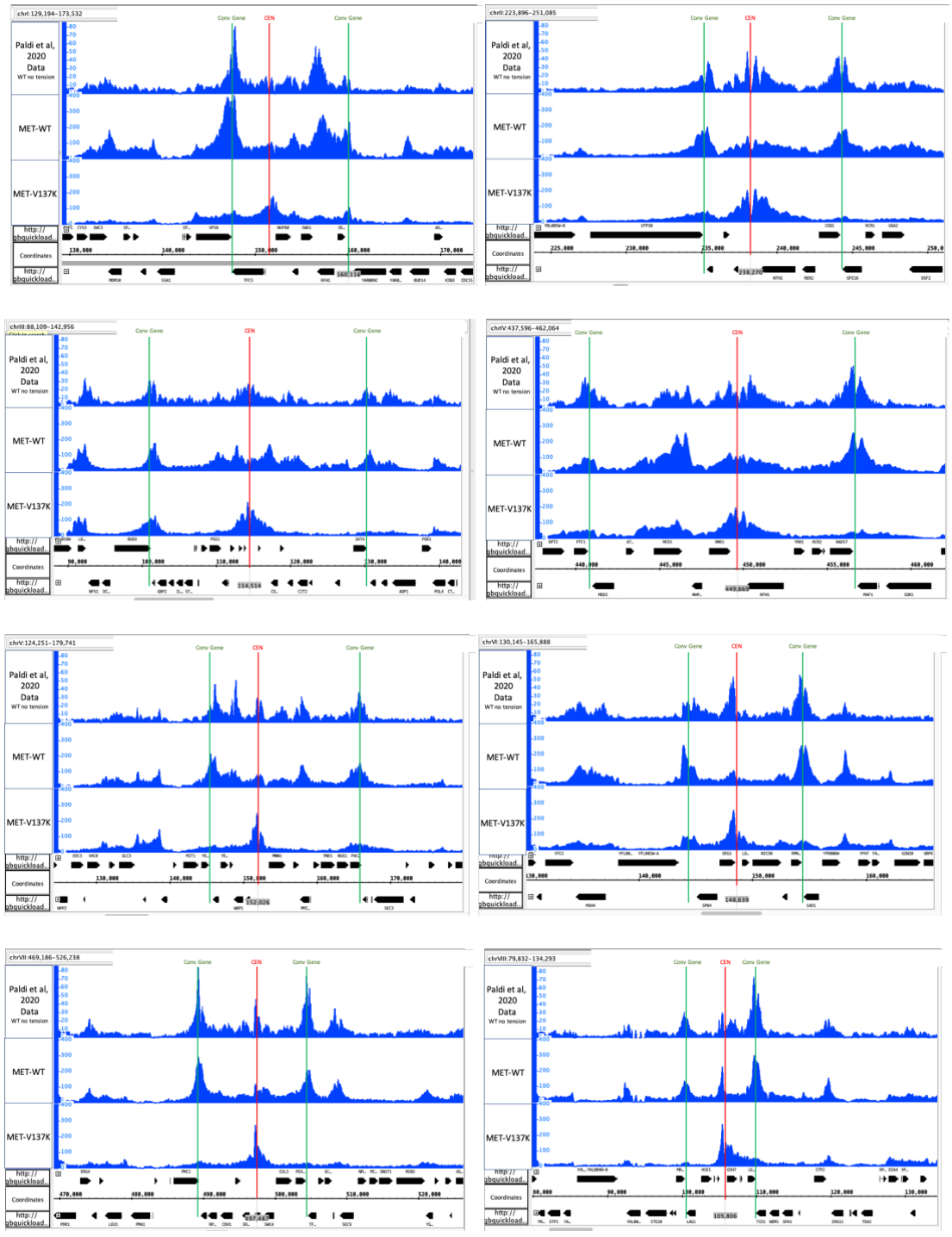
Yamano H. (2019). APC/C: current understanding and future perspectives. *F1000Research*, *8*, F1000 Faculty Rev-725. https://doi.org/10.12688/f1000research.18582.1

Yatskevich, S., Rhodes, J., & Nasmyth, K. (2019). Organization of Chromosomal DNA by SMC Complexes. *Annual review of genetics*, *53*, 445–482. https://doi.org/10.1146/annurev-genet-112618-043633

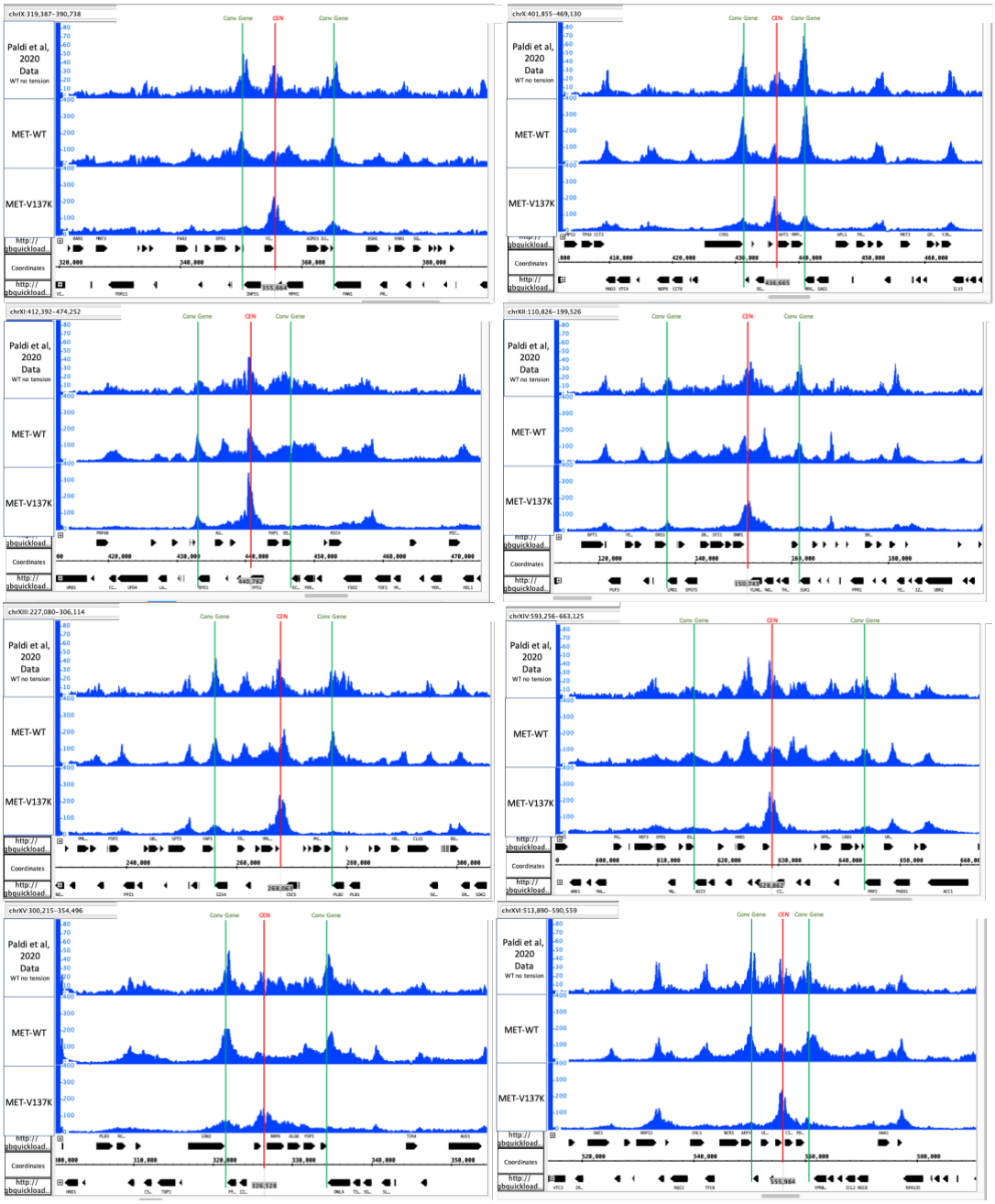
Yuan, B., Neira, J., Pehlivan, D., Santiago-Sim, T., Song, X., Rosenfeld, J., Posey, J. E., Patel, V., Jin, W., Adam, M. P., Baple, E. L., Dean, J., Fong, C. T., Hickey, S. E., Hudgins, L., Leon, E., Madan-Khetarpal, S., Rawlins, L., Rustad, C. F., Stray-Pedersen, A., … Liu, P. (2019). Clinical exome sequencing reveals locus heterogeneity and phenotypic variability of cohesinopathies. *Genetics in medicine : official journal of the American College of Medical Genetics*, *21*(3), 663–675. https://doi.org/10.1038/s41436-018-0085-6

Zhang, J., Shi, X., Li, Y., Kim, B. J., Jia, J., Huang, Z., Yang, T., Fu, X., Jung, S. Y., Wang, Y., Zhang, P., Kim, S. T., Pan, X., & Qin, J. (2008a). Acetylation of Smc3 by Eco1 is required for S phase sister chromatid cohesion in both human and yeast. *Molecular cell*, *31*(1), 143–151. https://doi.org/10.1016/j.molcel.2008.06.006

Zhang, N., Kuznetsov, S. G., Sharan, S. K., Li, K., Rao, P. H., & Pati, D. (2008b). A handcuff model for the cohesin complex. *The Journal of cell biology*, *183*(6), 1019–1031. https://doi.org/10.1083/jcb.200801157



Appendix I. Reduction in cohesin occupancy observed in Scc1V137K at the convergent genes, situated at the pericentromeric border.
Related to Chapter 2, figure 2-8. Integrated Genome Browser (IGB) alignment of Calibrated ChIP sequencing data for WT with no tension (Paldi *et al*., 2020) and Scc1V137K with and without WT. Green lines indicate convergent genes at the pericentromeric borders, red line indicates the centromere.
From top, left to right, chromosomes I to VIII.



Appendix I. Reduction in cohesin occupancy observed in Scc1V137K at the convergent genes, situated at the pericentromeric border.

Related to Chapter 2, figure 2-8. Integrated Genome Browser (IGB) alignment of Calibrated ChIP sequencing data for WT with no tension (Paldi *et al*., 2020) and Scc1V137K with and without WT. Green lines indicate convergent genes at the pericentromeric borders, red line indicates the centromere.
From top, left to right, chromosomes IX to XVI.