Probing the cohesin loading reaction using forward genetics and quantitative ChIP-sequencing

Naomi Petela
Corpus Christi College

Department of Biochemisty
University of Oxford

Thesis submitted for the degree of
Doctor of Philosophy
Trinity Term 2016

To Mum and Dad

Probing the cohesin loading reaction using forward genetics and quantitative ChIP-sequencing

Naomi Petela
Corpus Christi College

Submitted for the degree of Doctor of Philosophy
Trinity Term 2016

Abstract

Cohesion between sister chromatids that are newly formed in S-phase is crucial to ensure the fidelity of chromosome segregation. The cohesin complex mediates cohesion and is comprised of two SMC proteins, Smc1 and Smc3, with a kleisin subunit, Scc1. The SMCs interact via their hinge domains to form a v-shaped heterodimer, with Scc1 forming a bridge between their head domains to produce a ring structure that topologically entraps sister chromatids. In order to confer cohesion, cohesin needs to be loaded onto DNA, a process that is poorly understood. It is known that cohesin loading depends on a loading complex comprised of Scc2 and Scc4, ATP hydrolysis and presumably opening of at least one of the interfaces of the cohesin ring.

Using the recently developed technique, quantitative ChIP-seq, we show that Pds5, a cohesin-associated protein, is displaced at centromeres from the cohesin ring during the loading reaction. This event is accompanied by the engagement of SMC ATPase heads in the presence of ATP. Upon ATP hydrolysis, the loading reaction completes with DNA entrapment, translocation to the pericentromeric sequences and the displacement of Scc2 by Pds5.

We also describe mutations in Scc2, Smc1 and histone proteins H2A and H2B that enable loading in the absence of Scc4. The mutations in Scc2 alter the dynamics of the association between the loading complex and cohesin, reducing dissociation during the translocation step. All of the mutations are able to restore loading along chromosome arms but not at centromeres where Scc4 has a particular role. However, the Smc1 mutation hinders loading at the centromere, even in the presence of Scc4, due to a defect in the second step of the loading reaction, namely translocation driven by ATP hydrolysis. We suggest that this discrepancy in functionality at the arms and centromere is due to changes in the hinge accompanying both the first and second steps. The rate limiting reaction is the first step at arms and the second at the centromeres. The histone mutants show that nucleosome occupancy plays an important role, at least on chromosome arms, but we find no evidence that this involves the RSC complex.

Contributions

The work in this thesis stems from a screen to identify suppressors of the thermosensitive allele *scc4-4* preformed by Dr Maria Demidova that identified a mutation in the hinge domain of Smc1. Dr Sarah Dixon characterised this mutation prior to this thesis. The technique of quantitative ChIP-seq was developed in collaboration with Dr Bin Hu and Dr Alexander Kurze. In addition, Dr Bin Hu obtained the ChIP-sequencing data used in Figures II-6A and II-10, and Dr Christophe Chapard obtained the ChIP-sequencing data used in Figure II-4D. The screen to identify mutations able to rescue the absence of Scc4 was preformed by Wentao Chen. The experiments investigating the role of the RSC complex were preformed in collaboration with Dr Sarah Dixon. The structure of Scc2 from *Chaetomium thermophilum* was used with permission from Dr Hongtao Yu.

Acknowledgements

This work was generously funded by the Medical Research Council.

I am very grateful to Kim Nasmyth for the opportunity to work in his lab, many invaluable discussions, and guidance. I am also hugely grateful to Sarah Dixon who paved the way for this project and continued to support me, even from another country. Her thoughts, comments, and patience were very much appreciated.

Thanks go to those I've collaborated with: Bin Hu, Christophe Chapard, Wentao Chen and Stefanos Skylakakis. I'd also like to thank everyone who lent a helping hand during the course of my DPhil. In particular, I would like to thank Johanna Scheinost for her patience, advice, and amazing proofreading.

A special thanks goes to Jean Metson who has been my "lab mum", supporting me both technically and emotionally. I hope to return the favour.

I also would like to thank the many friends I've made over the years. A massive thank you to Katarzyna Kozyrska for sharing my struggles, cups of tea, and many, many bottles of wine. Also to Cai Marshall and Adam Povey for putting up with me and keeping me sane.

Most importantly, I am extremely thankful to my family for all their unconditional love and support - even just by being there at the other end of many a long phone call. In particular, to my Mum and Dad for helping me get this far. Finally to Peter West, both for his general encouragement and more importantly, for his patience teaching me to code.

Abbreviations

ABC	ATP-binding cassette
AcK	Acetylated lysine
APC/C (Apc)	Anaphase-Promoting Complex/Cyclosome
ATP	Adenosine-5'-triphosphate
B. subtilis	*Bacillus subtilis*
BSA	Bovine serum albumin
C- (terminal)	Carboxy-
C. glabrata	*Candida glabrata*
$CaCl_2$	Calcium chloride
CAR	cohesin-associated regions
CDC (Cdc)	Cell Division Cycle
CDE	Centromere DNA Element
Cdk	Cyclin Dependant Kinase
CdLS	Cornelia de Lange Syndrome
CEN	Centromere
ChIP	Chromatin immunoprecipitation
CTCF	CCCTC-binding factor
ddH_2O	Double-distilled water
DDK	Dbf4-dependent kinase
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	*Escherichia coli*
ECO/ESCO (Eco)	Establishment of cohesion
EDTA	Ethylenediaminetetraacetic acid
EGTA	ethylene glycol-bis(β-aminoethylether)-N,N,N',N'-tetraacetic acid
FACS	Fluorescence-activated cell sorting
G_1/ G_2 (phase)	Gap 1/2
H(1, 2A, 2B, 3, 4)	Histone
HEAT	Huntingtin, Elongation factor 3, Protein Phosphatase 2A, TOR1

HEPES	4-(2-hydroxylethyl)-1-piperazineethanesulfonic acid
HRP	Horseradish peroxidase
IP	Immunoprecipitation
IR	infrared
KOAc	Potassium Acetate
KOH	Potassium hydroxide
LiCl	Lithium chloride
LiOAc	Lithium acetate
M (phase)	Mitosis
Mad	Mitotic arrest deficient
MET	Methionine
$MgCl_2$	Magnesium chloride
M. musculus	*Mus musculus*
MNase	Micrococcal nuclease
N- (terminal)	Amino-
NaCl	Sodium Chloride
NBD	Nucleotide Binding Domain
NIPBL (NIPBL)	Nipped-B-like
OD_{600}	Optical density at 600 nanometres
OR	Occupancy Ratio
PBS	Phosphate-buffered saline
PBST	PBS + 0.05% TWEEN® 20
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase Chain Reaction
Pds	Precocious dissociation of sisters
PEG	poly(ethylene glycol)
PK	small epitope on the P/V proteins of the paramyxovirus simian virus 5
PMSF	phenylmethylsulfonyl fluoride
PP2A	Protein Phosphatase 2A

qPCR	quantitative Polymerase Chain Reaction
RAD (Rad)	Radiation-sensitive
RBS	Roberts syndrome
rDNA	ribosomal DNA
RNA	Ribonucleic acid
RT	Room temperature
S (phase)	Synthesis
S. cerevisiae	*Sacharromyces cerevisiae*
S. pombe	*Schizosaccharomyces pombe*
SAC	Spindle-assmebly checkpoint
SCC (Scc, *scc*)	Sister Chromatid Cohesion
SDS	Sodium dodecyl sulfate
SGD	Saccharomyces Genome Database
sgo	Shugoshin
SMC (SMC, Smc, *smc*)	Structural Maintenance of Chromosomes
SNPs	Single Nucleotide Polymorphisms
SPB	Spindle pole body
TE	Tris EDTA
TES	Transcription end site
TEV	Tobacco Etch Virus
TOR	Target of rapamycin
TPR	Tetratricopeptide repeat
ts	temperature sensitive
TSS	Transcription start site
URA (*ura*)	Uracil requiring
UV	Ultraviolet
v/v	volume/volume
w/v	mass/volume
WCE	Whole cell extract
YEP	Yeast Extract Peptone
YPD	Yeast Extract Peptone Dextrose

Contents

Introduction **11**
The cell **11**
The basic unit of life 11
The cell cycle and its regulation 12
Entry into Mitosis: Preparation of DNA for division 14
Exit from Mitosis: Sister chromatid segregation 19
SMC containing complexes **20**
A family of SMC-containing complexes 20
The structure of SMC proteins 21
The cohesin complex **24**
A topological device 24
Cohesin loading onto chromosomes 29
Localisation of cohesin and of its loading sites on chromatin 34
Cohesin removal from chromosomes 37
Cohesin establishment and maintenance 40
Cohesinopathies 42
Thesis Objective **45**

Quantitative ChIP-sequencing reveals novel insights into the cohesin loading reaction **46**
Introduction **46**
Results **50**
Measuring the localisation of cohesin by quantitative ChIP-seq. 50
Cohesin localisation on chromosomes depends on the loading complex. 56
The cohesin-loading complex primarily associates with centromeres. 58
Loaded cohesin spreads from centromeric to pericentromeric regions during the cell cycle. 61
Occupancy of the loading complex at centromeres depends on Scc1. 65
Pds5 is excluded from the “pre-loading” complex that forms at centromeres. 68
Cohesin accumulates around transcription end sites during the cell cycle 72
Scc2 is excluded from TSS and TES but is enriched at tRNAs 75
Discussion **79**

A mutation in the hinge domain of Smc1 bypasses the requirement for Scc4 but is defective in centromeric loading **84**
Introduction **84**
Results **87**
smc1(D588Y) suppresses the loading defect of scc4-4 on chromosome arms but not around centromeres 87
smc1(D588Y) is defective in a late step in the loading reaction. 95
Discussion **98**

Mutations in Scc2 enable loading of cohesin in the absence of Scc4 ..... **105**
Introduction **105**
Results **106**
Mutations in the N-terminal domain of Scc2 are able to suppress the temperature sensitive phenotype of scc4-4. 106

Mutations in Scc2 are able to bypass the requirement for the essential protein Scc4. 107
Scc2 residues E822 and L937 are both highly conserved residues. 109
scc2(E822K, L937F) is not upregulated in the absence of Scc4. 112
scc2(E822K, L937F) suppresses the loading defect of scc4-4 on chromosome arms but not at the centromere 112
scc2(E822K, L937F) alters the distribution of Scc2 and smc3(E1155Q) occupancy at the centromere 121
Discussion 125

Mutations in histone proteins H2A and H2B are able to bypass the requirement for Scc4 129
Introduction 129
Results 130
Mutations in histone proteins H2A and H2A are able to suppress the temperature sensitive phenotype of scc4-4. 130
Inactivation of the RSC complex has little or no effect on cohesin loading. 135
Inactivation of the RSC complex or the cohesin loading complex has no effect on nucleosome occupancy. 139
Discussion 141

Materials and methods 143
Yeast strains and growth conditions. 143
Fluorescence-activated cell sorting. 144
Site-directed mutagenesis. 144
Yeast transformation. 145
Genomic integration. 146
Genomic DNA preparation from yeast. 146
Tetrad dissection. 147
Screening for suppressors of scc4-4. 147
Preparation of native yeast cell extracts. 148
Quantification of protein concentration in crude extracts. 148
Protein gel electrophoresis and Western blotting 149
Multiple sequence alignment. 150
Quantitative ChIP-seq. 150
Extraction of yeast DNA for deep sequencing. 152
MNase-seq. 152
Preparation of sequencing libraries. 153
Data analysis, alignment and production of BigWigs. 154
Visualisation of ChIP-seq profiles. 155
Identification of mutations from whole genome sequencing. 156
List of strains 157

References 159

Introduction

The cell

The basic unit of life

At the heart of all species on this Earth are cells: some species (unicellular organisms) take the form of single cells while others (multicellular organisms) are composed of multitudes of different types of cells working together. Thus, a fundamental requirement for all organisms is the ability to produce new cells, a process not only essential for the continuation of the species but also important for growth, regeneration, and repair of tissues in multicellular organisms.

Production of new cells occurs by cell division, whereby a 'mother' cell will grow, replicate its contents and then divide to produce two 'daughter' cells. This process of duplication and division must be done with accuracy and precision to make it a robust and reliable process over multiple generations. This is particularly true of the replication and segregation of the genetic information stored in a cell's DNA, whose almost perfect inheritance by the daughter cell will determine whether or not the cell can function as required. Errors in cell division often lead to an abnormal numbers of chromosomes in cells (aneuploidy), which can have a range of deleterious effects. For single-celled organisms, aneuploidy can be lethal while in multicellular organisms it can also lead to oncogenesis and developmental disorders. Since robust cell division is so fundamental to the survival and wellbeing of all organisms it is not surprising that the mechanisms involved in this phenomenon are so precisely controlled.

The cell cycle and its regulation

The process of cell division occurs by the cell progressing through a set of well-defined phases that together are termed the cell cycle (Figure I-1). The transition between these phases is controlled by a series of molecular switches that ensure timely and sequential progression through the cell cycle in a unidirectional manner. These switches are composed of two types of proteins: cyclins and cyclin-dependent kinases (Cdks). Changes in the levels of different cyclins during the cell cycle lead to oscillations in the activity of Cdks that become active upon formation of Cdk-cyclin complexes. When active, these Cdk-cyclin complexes phosphorylate their substrates to initiate a cascade of events that are required for progression to the next phase of the cell cycle (Morgan, 2006). The ability of the system to exist in two steady states or 'bistability' of these switches ensures that, once all conditions for the progression into the next cell phase are met, there is no intermediate step and the cell transitions into the next phase in an ordered, unidirectional and irreversible manner (Novak, Tyson, Gyorffy, & Csikasz-Nagy, 2007).

The progression through the cell cycle is also subjected to an additional level of regulation in the form of checkpoints. Checkpoints monitor the state of the cell and allow the progression through the cell cycle to be influenced by various intracellular and extracellular signals. This is particularly important to ensure the completion of the events of one phase before the cell enters the next.

The cell cycle begins in the first gap phase (G_1) where the cell is highly metabolically active and grows. It is near the end of G_1 that a cell encounters its first checkpoint known as Start or the G_1/S checkpoint, where it will check that extracellular conditions are favourable for cell division. If conditions are

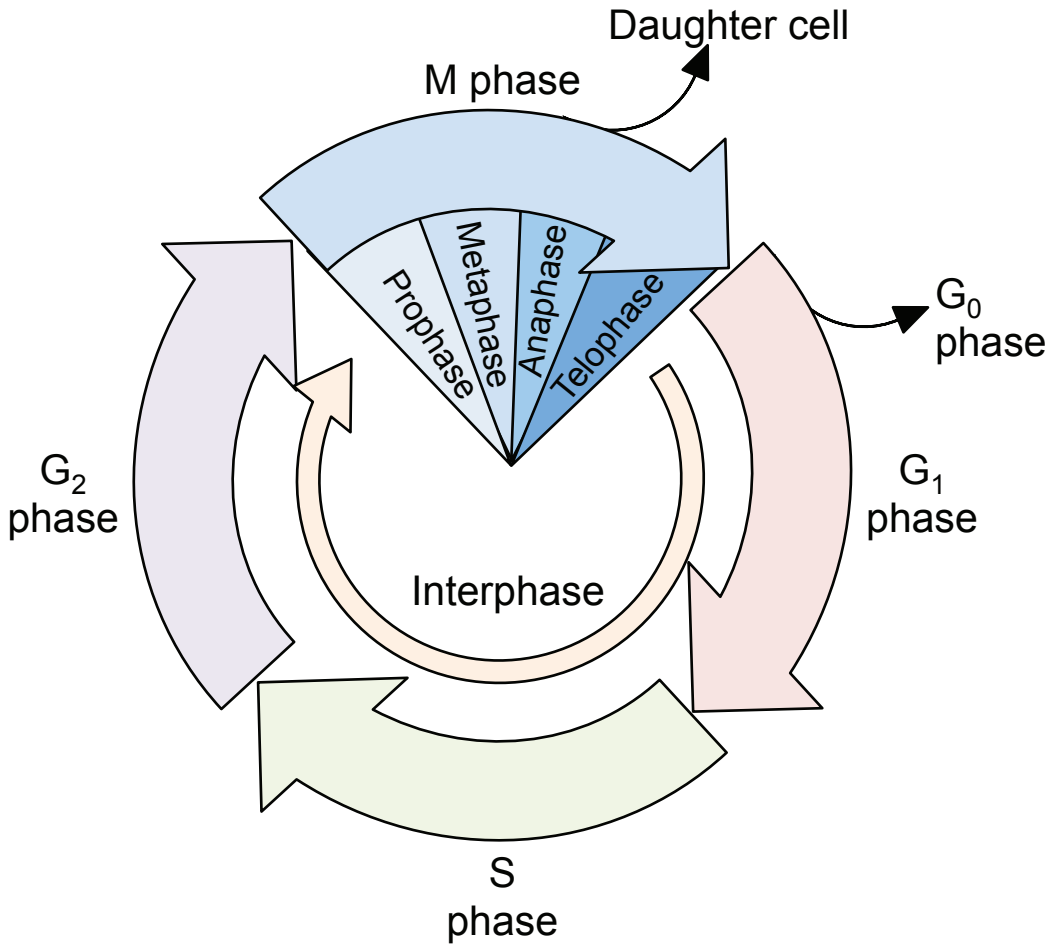


Figure I-1
The cell cycle.
The cell cycle can be divided into four phases: G_1, S, G_2 and M. During G_1 the cell grows and prepares for cell cycle progression. If conditions are not favourable, the cell will enter a resting state (G_0). Otherwise, it will continue onto S phase, where the genome will be replicated. On completion of S phase the cell will enter G_2 where it will grow and prepare for M phase. M phase can be further divided into prophase, metaphase, anaphase and telophase. During these phases the cell will divide its contents equally amongst the two resulting daughter cells, completing the cycle.

disadvantageous, the cell will delay cell cycle progression and may even enter Gap 0 (G_0), a specialised resting state. Otherwise, in favourable conditions, the cell will commit to progression into synthesis (S) phase and replicate its entire genome. After S phase, the cell will enter a second gap phase (G_2) where it will continue to grow and synthesise the proteins required for mitotic phase (M). Together G_1, S and G_2 are termed interphase. Following interphase, M phase can be further subdivided into four phases: prophase, metaphase, anaphase and telophase (Morgan 2006).

At the entry into mitosis, there is a second major checkpoint, the G_2/M checkpoint. This verifies that the replication of the genome is complete before the cell continues into M phase. During M phase, the cell will undergo the remarkably complicated process of dividing its contents, most notably the huge amount of genetic material, into the two daughter cells. It is during this phase that the final major checkpoint, the spindle assembly checkpoint (SAC), is encountered. This checkpoint ensures that all the chromosomes are ready and capable of being segregated equally between the daughter cells before the division takes place (Morgan 2006).

Entry into Mitosis: Preparation of DNA for division

Entry into M phase is preceded by an increase in concentration of the mitotic cyclin (M cyclin) which binds to Cdk1, but the resulting complex is kept inactive by inhibitory phosphorylation by the Wee1 kinase family. Upon activation of the Cdc25 phosphatase, the inhibitory phosphorylation is removed activating the M cyclin-Cdk complexes (M-Cdk), which then induce chromosome

condensation and the assembly of the mitotic spindle (Morgan 2006). Up until this point, the DNA existed without clear structure or separation within the nucleus. Thus, in order to separate and equally divide the DNA, a mechanism to re-organise and condense the genetic material into discrete entities has evolved. This mechanism of condensation occurs in prophase and gives rise to visible discrete rod-like structures we recognise as condensed chromosomes.

To reach this level of condensation, multiple levels of compaction are required and are achieved by a variety of proteins. At the most basic level, DNA is bound by nucleosomes, which are comprised of two of each of the general histone proteins: H2A, H2B, H3 and H4. Together they form a histone octamer, which is positively charged and can bind 146 base pairs (bp) of the negatively charged DNA helix to produce a nucleosome (Figure I-2A) (Luger, Mäder, Richmond, Sargent, & Richmond, 1997; T. J. Richmond, Finch, Rushton, Rhodes, & Klug, 1984; White, Suto, & Luger, 2001). By repeatedly wrapping the DNA around several histone octamers, one after the other, a string of nucleosomes spaced by small linker regions is produced, condensing the DNA into an 11 nm-wide fibre (Figure I-2B) (Thoma, Koller, & Klug, 1979). Histone H1 binds to the linker region between these histone octamers and is thought to play a role in chromatin arrangement and compaction (Simpson, 1978). It has been hypothesised that the additional compaction mediated by histone H1 gives rise to a 30 nm-wide fibre, which has been observed by electron microscopy (Marsden & Laemmli, 1979), but the existence of this structure *in vivo* is disputed (Grigoryev & Woodcock, 2012; Maeshima, Hihara, & Eltsov, 2010; Nishino et al., 2012).

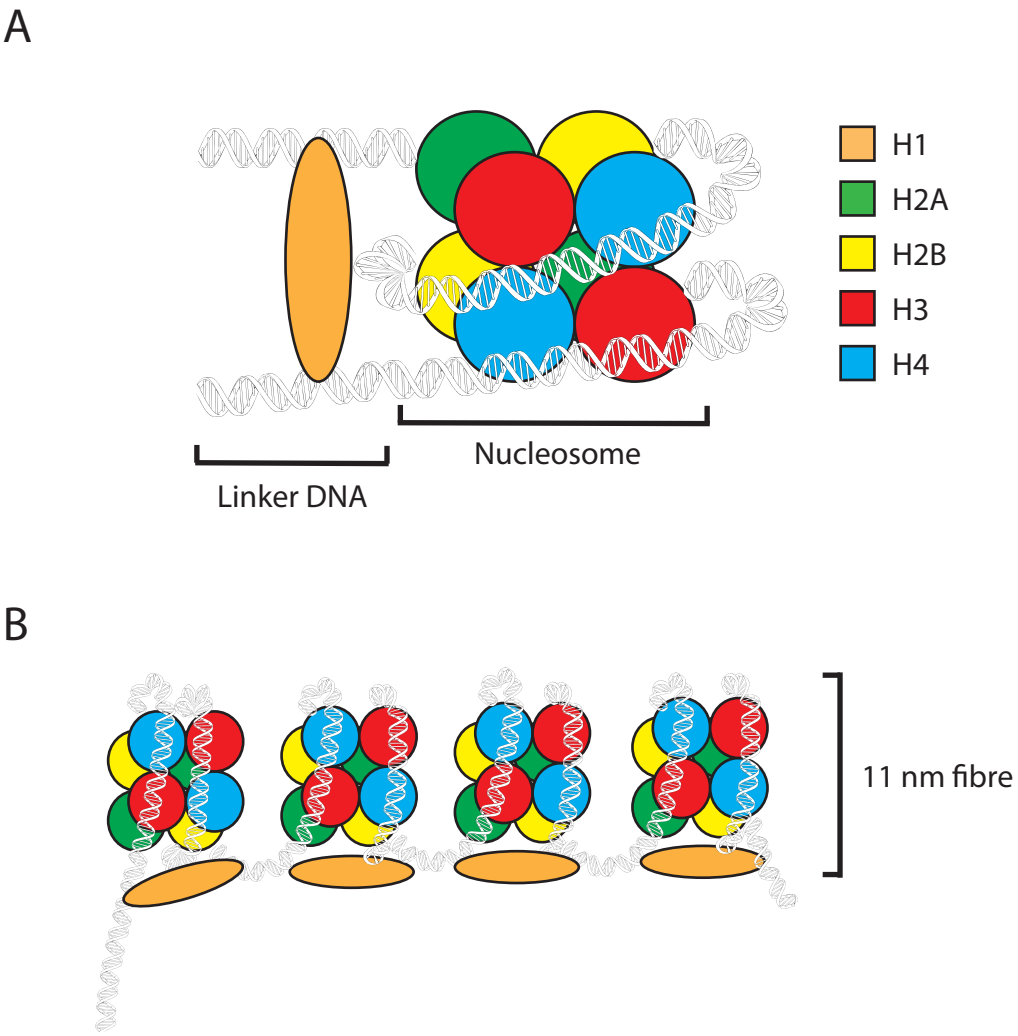


Figure I-2

Nucleosomes comprising of histone proteins produce the most basic level of DNA compaction.

A - Two of each of the histone proteins H2A, H2B, H3 and H4 form an octamer that binds 146 bp of DNA to form a nucleosome. Histone H1 binds to the linker region. B - Repeatedly wrapping DNA around histone octamers produces an 11 nm wide fibre.

Chromatin is further compacted during mitosis by the condensin complex, a member of a family of protein complexes that contain Structural Maintenance of Chromosomes (SMC) proteins that will be discussed in greater detail later. The activation of condensin is stimulated by phosphorylation of condensin subunits by M-Cdk and, once active, condensin begins to compact DNA into threads (Hirano, 2005). In doing so, condensin confers three vital physical properties to chromosomes: compaction, intra-chromatid cohesion to generate sufficient rigidity to withstand tension and a mechanism to promote sister chromatid disentanglement (Houlard et al., 2015).

It is thought that condensin also promotes resolution of catenation, the intertwining of sister chromatids that physically links them together (Charbin, Bouchoux, & Uhlmann, 2014). This catenation is an inherent result of DNA replication and needs to be resolved to allow sister chromatid segregation. This is achieved by topoisomerase II, which makes a double strand break in one DNA strand, passes the second strand through the break and then repairs the first DNA strand to remove the catenation (Goto & Wang, 1982).

Mitosis would not be possible if condensation were the only major form of higher order chromosome organisation. Once disentangled, freely floating sister chromatids completely separated from one another would be difficult to divide equally between daughter cells. How does the cell ensure that each daughter will receive a single copy of each chromosome? The answer is that sister chromatids are not completely separated. They are held together by cohesin, another SMC protein-containing complex, that holds together sister chromatids together from their initial formation in S phase (Farcas, Uluocak, Helmhart, & Nasmyth, 2011; Guacci, Koshland, & Strunnikov, 1997; Michaelis,

Ciosk, & Nasmyth, 1997). The sister chromatid cohesion conferred by the cohesin complex is vital for the accurate division of sister chromatids, as it enables them to be manipulated as pairs after the resolution of catenation by topoisomerase II.

At metaphase, sister chromatids align on the metaphase plate, a process that relies on the mitotic spindle. The spindle begins to form after breakdown of the nuclear envelope (a feature not observed in yeast) in a manner dependent on the activity of M-Cdk complexes (Nigg, 2001). Microtubules usually originating from the centrioles or spindle-pole bodies (SPBs) located at opposite sides of the cell attach to the pairs of chromatids via their kinetochores. These kinetochores are multi-subunit protein complexes that assemble at the centromere (CEN), a single region of chromatin in each chromosome that is specialised for binding to the kinetochore complex (Cho et al., 2010). It is at this point that each sister must be attached to microtubules emanating from opposing centrioles (amphitelic attachment) to ensure accurate and equal inheritance of chromosomes by the daughter cells (Mitchison & Kirschner, 1984).

To ensure that all chromatids are attached to microtubules in an amphitelic manner, the cell cannot continue through the cell cycle until the SAC is satisfied. Only in the case of amphitelic attachment will the poleward forces from the shrinkage of microtubules pull against the cohesion of sister chromatids conferred by cohesin to generate a state of tension that stabilises the microtubule-kinetochore attachments. If both sisters are pulled in the same direction by the microtubules from only one side (syntelic attachment), or if one kinetochore is unbound by the microtubules (monotelic attachment), then no tension can be generated. In such cases the kinetochore-associated kinase

Aurora B destabilises the attachments, thereby promoting error correction until bi-orientation is achieved (D. Liu & Lampson, 2009). The SAC can only be satisfied and the cell cycle allowed to progress when all sister chromatid pairs have been bi-orientated by the mitotic spindle and exhibit tension.

Exit from Mitosis: Sister chromatid segregation

Once the sister chromatids have bi-orientated, loss of sister chromosome cohesion in anaphase is permitted through activation of the anaphase-promoting complex or cyclosome (APC/C). The APC/C and its activator Cdc20 (APC^{Cdc20}) are kept inactive prior to bi-orientation by the sequestration of Cdc20 by Mad2. Upon activation, the APC/C polyubiquitinates securin and the mitotic cyclin (cyclin B), amongst other targets, marking them for degradation by the proteasome (Peters, 2006). Destruction of securin liberates separase, a thiol protease enzyme that proteolytically cleaves a subunit of the cohesin complex. Separase, however, is also kept inactive by inhibitory phosphorylation by M-Cdk (Stemmann, Zou, Gerber, Gygi, & Kirschner, 2001). Thus, the APC/C mediated destruction of M-Cdk prevents the phosphorylation of separase enabling its activation as well as releasing it from the inhibition mediated by securin. Activation of separase and the cleavage of cohesin causes the dissolution of sister chromatid cohesion and allows sister chromatids to be pulled to opposite poles of the cell by the mitotic spindle (Uhlmann, Lottspeich, & Nasmyth, 1999; Uhlmann, Wernic, Poupart, Koonin, & Nasmyth, 2000).

Although the SAC regulates the timing of anaphase initiation by the APC/C in most cells, it is not strictly necessary for the orderly activation of the APC/C in

many cells such including those of yeast, flies, worms and even some mammalian cells. This is because the APC/C bound to Cdc20 is only active upon phosphorylation of the APC/C subunits Apc1 and Apc3 by Cdk1. Therefore Cdk1 not only triggers entry into mitosis but also initiates its own demise (Qiao et al., 2016; S. Zhang et al., 2016).

With the sister chromatids divided equally between the two poles of the cell, the mitotic spindle is dismantled, signifying the beginning of telophase. The chromosome condensation is dismantled as the nuclear envelope reappears to create two daughter nuclei. Finally, the cell itself divides along the spindle midpoint to produce two daughter cells through cytokinesis (Morgan, 2006).

SMC containing complexes

A family of SMC-containing complexes

Given the huge amount of DNA contained within each cell, mechanisms to physically manage the DNA are required. This not only applies to the condensation and segregation of replicated DNA as previously discussed but also to the management of DNA that has been damaged, the regulation of gene expression and the modulation of chromatin architecture.

Complexes that contain members of the Structural Maintenance of Chromosomes (SMC) family of proteins are evolutionarily conserved, from bacteria to humans (Cobbe & Heck, 2006). In eukaryotes the primordial SMC gene has undergone duplications to give rise to paralogs and orthologs that have become the basis of three identified SMC containing complexes present in all eukaryotes (Gligoris & Löwe, 2016). These are cohesin, condensin, and the

Smc5/6 complex. As previously mentioned, cohesin (Smc1 and Smc3) confers sister chromatid cohesion and condensin (Smc2 and Smc4) mediates sister chromatid compaction and segregation. The Smc5/6 complex is involved in the dissolution of the replication fork as well as DNA repair and contains Smc5 and Smc6. A fourth complex, the dosage compensation complex, is involved in heterochromatin formation but only exists in certain organisms such as *Caenorhabditis elegans.* All of these complexes share a characteristic design that allows them to perform these varied roles and at the heart of this design are the SMC proteins themselves (Jessberger, 2002).

The structure of SMC proteins

SMC proteins share a generic structure (Strunnikov, Larionov, & Koshland, 1993). At the amino (N) and carboxy (C)- termini are large globular domains, each containing highly conserved nucleotide binding motifs. The N-terminus contains a Walker A box whereas the C-terminus contains a DA box with Walker B-like sequence (Walker, Saraste, Runswick, & Gay, 1982). At the center of the protein lies a globular hinge domain, separated from the N- and C-terminal domains by 50 nm long amphipathic α-helices (Figure I-3A). The protein folds at the hinge to bring the flanking α-helices together to form an antiparallel intramolecular coiled coil domain This also brings together the N- and C- terminal domains to form an ATP binding cassette (ABC)-like ATPase domain (Figure I-3B) (Haering, Löwe, Hochwagen, & Nasmyth, 2002).

Two SMCs associate via the hinge region to produce V-shaped molecules that can be observed with electron microscopy. When in this form, the ABC-like

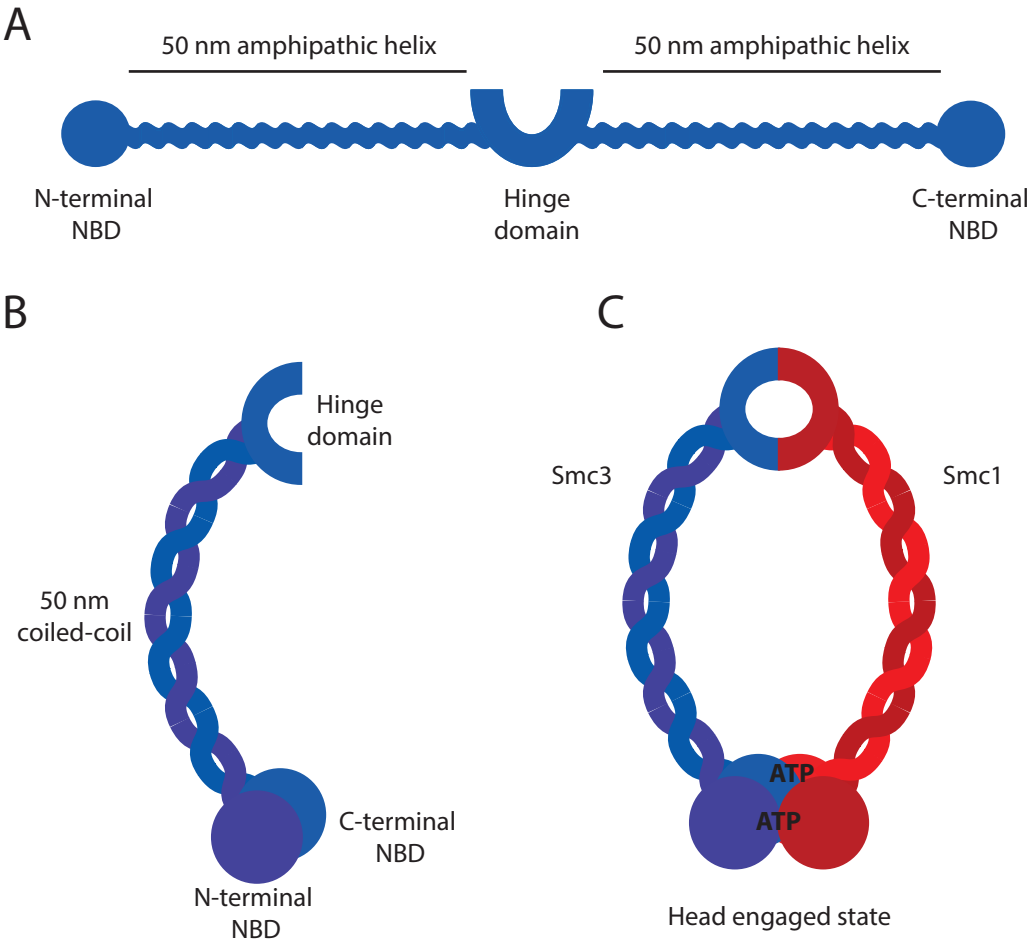


Figure I-3

SMC structure and cohesin.

A - Schematic of an SMC protein prior to folding. The central hinge domain is flanked by two 50 nm amphipathic helices each ending with a globular NBD.

B - The SMC proteins fold at the hinge domain bringing the flanking helices together to form an antiparallel intramolecular coiled coil and the NBDs together to form an ABC-like ATPase domain.

C - Two SMCs associate via the hinge domain enabling the two NBDs to interact by sandwiching two molecules of ATP between them.

ATPase domains also known as nucleotide binding domains (NBDs) transiently come together and interact by sandwiching two molecules of ATP between them, each bound by one Walker A/Walker B pair (Figure I-3C). The ATP hydrolysis activity of the SMCs is enhanced upon addition of DNA but it is still unclear on how this effect is mediated (Haering et al., 2002; Hirano, 2006).

The hinge domains of SMC dimers themselves also have several features that are very well conserved across species throughout evolution. Interaction of two SMC domains forms a central pore which is invariably positively charged, a characteristic that seems to be necessary for the complex to confer cohesion as cohesin hinges that have this positive charge neutralised load normally onto chromosomes but are unable to confer sister chromatid cohesion (Kurze et al., 2010). The high level of conservation between SMCs also means that the points of contact above and below the pore are almost identical making it even more striking that despite this homology across SMCs, there is sufficient sequence diversity to ensure that each monomer is only able to interact with its partner (Griese, Witte, & Hopfner, 2010). Thus cohesin's Smc1 is only able to interact with Smc3 and not with Smc2, 4, 5 or 6.

It is worth noting at this point that in some species, such as *Escherichia coli* (*E. coli*) and some thermophilic bacteria, it has not been possible to identify a canonical SMC protein. Despite this, functional homologs have been identified though their features vary slightly. The first SMC protein discovered was MukB in *E. coli*, mutants of which showed abnormal chromosome partitioning (Niki, Jaffé, Imamura, Ogura, & Hiraga, 1991). It differs to canonical SMCs in that MukB homodimers do not have a central pore in the hinge domain (Ku, Lim, Shin, Shin, & Oh, 2010). More divergent still in the hinge domain is Rad50, which dimerises

via a zinc hook motif that chelates a Zn^{2+} ion between two cysteine residues from each monomer (Anderson, Trujillo, Sung, & Erickson, 2001).

The cohesin complex

A topological device

SMC containing complexes are formed upon the binding of additional subunits interacting with the core SMC heterodimer. A third subunit known as the kleisin closes the V-shaped formation by bridging the two NBDs to produce a basic ring to which additional regulatory subunits bind (Schleiffer et al., 2003). It is these additional specialised subunits that play an important role in defining the roles of each of the SMC complexes.

In the case of cohesin, the α-kleisin Scc1 (also known as Rad21, Rec8 and Mcd1) bridges the end of the V-shaped heterodimer of Smc1 and Smc3 (Figure I-4) (Gruber, Haering, & Nasmyth, 2003). The globular winged helix motif at the C-terminus of Scc1 binds to the base of the NBD of Smc1 (Haering et al., 2004). However the N-terminus of Scc1, despite the similarity between Smc1 and Smc3, makes an asymmetrical contact with Smc3. Instead of binding towards the bottom of the NBD, it forms a four-helix bundle with the coiled coil emerging from the Smc3 NBD (Gligoris et al., 2014). Interestingly, the binding of the C-terminal domain of Scc1 is required for binding of the N-terminal domain of Scc1 to Smc3 but not *vice versa* (Haering et al., 2004).

Often the structure of a protein complex can give key insights into the mechanism of its function. In the case of cohesin, which needs to be able to hold sister chromatids together until anaphase, its ring structure provides some

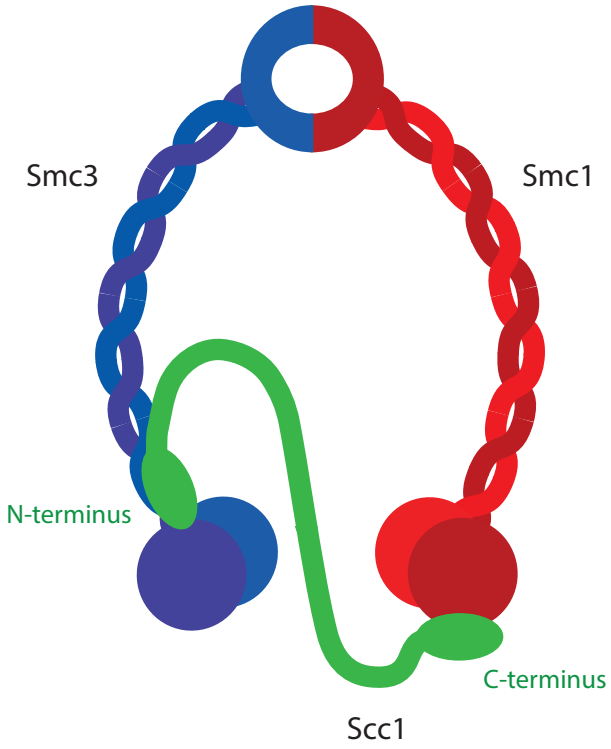


Figure I-4
The cohesin ring.
Smc3 and Smc1 associate via the hinge domain, and the cohesin ring is closed by the association of the kleisin subunit, Scc1, in an asymetric manner.

interesting possibilities. Since the predicted diameter of the ring is 40 nm (Haering et al., 2002), sufficient to comfortably hold two 11 nm chromatin fibres, one of the simplest explanations would be that both sister chromatids are trapped within a single cohesin ring. This is the most basic topological hypothesis known as the ring model (Figure I-5A) (Gruber et al., 2003; Haering et al., 2002). Slightly more complicated is the handcuff model where each sister is entrapped within a cohesin ring and these two rings interact topologically with each other to produce cohesion (Figure I-5B). A variation of this model proposes that a bridge between kleisin subunits physically holds cohesin rings that topologically entrap DNA together (Figure I-5C) (Ivanov & Nasmyth, 2005; N. Zhang et al., 2008b). The final topological model, the dimeric ring model, suggests that two cohesin rings multimerise in order to entrap sister chromatids within a larger ring (Figure I-5D). Of course there is also the possibility that the ring structure of cohesin is not inherent to its function and that it merely physically contacts both sisters, holding them together (Figure I-5E).

Consistent with the notion of a topological embrace is the evidence that proteolytic cleavage of the cohesin destroys cohesion between sister chromatids, both *in vivo* and *in vitro*, and linearisation of a circular minichromosome releases cohesin molecules that were previously bound (Gruber et al., 2003; Ivanov & Nasmyth, 2005; Oliveira, Hamilton, Pauli, Davis, & Nasmyth, 2010; Uhlmann et al., 1999; 2000). Although the former experiments do not rule out a purely physical connection, the latter implies that, once linearised, cohesin is able to slide off the DNA. This characteristic is incompatible with the idea of a purely physical contact between cohesin and the DNA and is corroborated by a similar

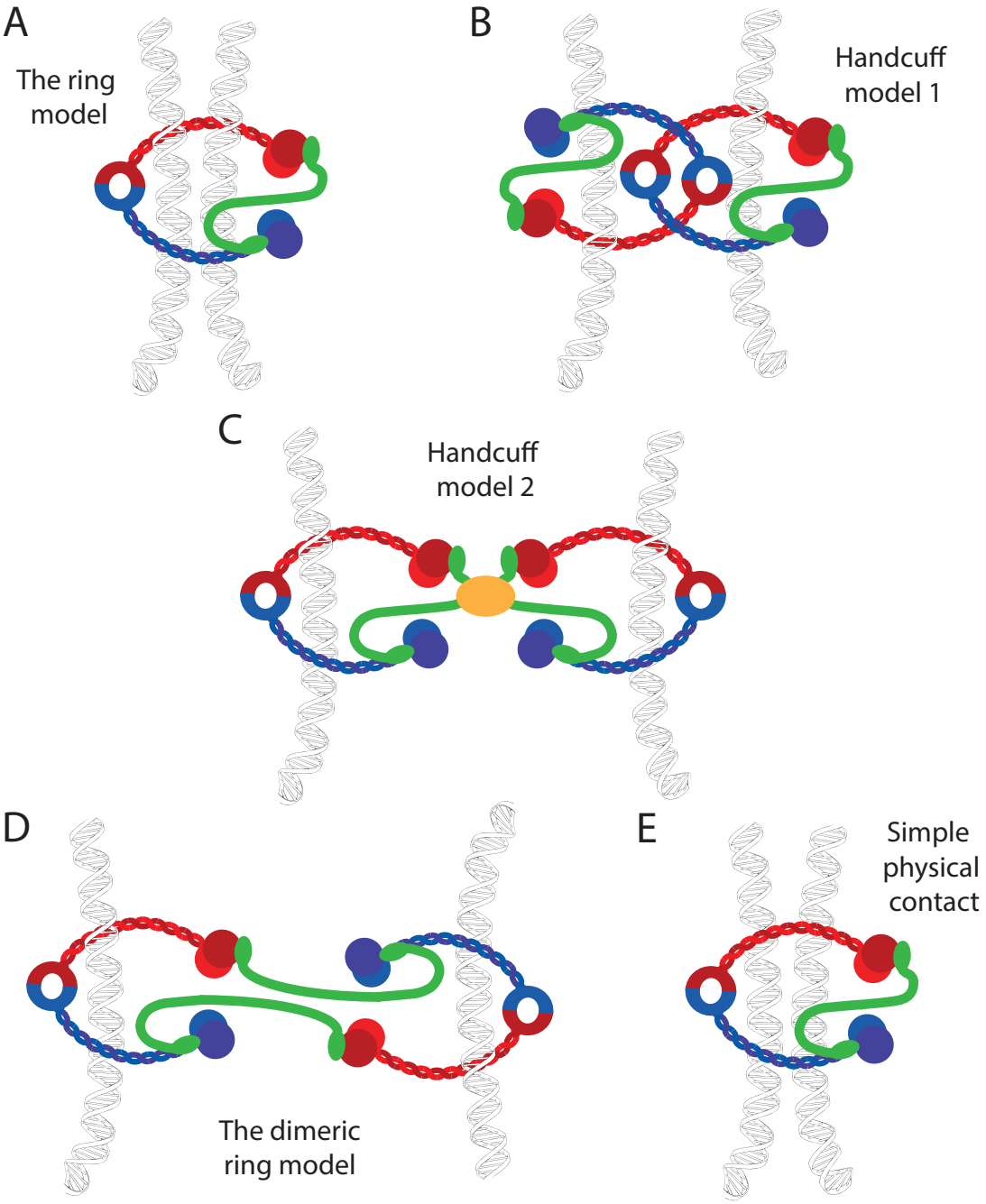


Figure I-5

Models of the mechanism of sister chromatid cohesion.

A - The ring model where one ring topologically entraps both sister chromatids.
B - The handcuff model where each sister is topologically entrapped within a single ring which interact topologically to produce cohesion.
C - A variation of the handcuff model in which cohesin rings topologically entrapping a single sister chromatid are bridged via their kleisin subunits to produce cohesion.
D - The dimeric model where two cohesin rings multimerise to topologically entrap sister chromatids within a larger ring.
E - Simple physical interactions between cohesin and DNA hold both sister chromatids together

experiment performed with the condensin ring suggesting that a topological embrace may be inherent to all SMC complexes (Cuylen, Metz, & Haering, 2011).

To address which of the topological models is correct; a modified version of the cohesin ring wherein all of the interfaces could be covalently closed was generated. This was done by expression of Smc3 and Scc1 as a fusion protein and engineering the Smc1/Smc3 (hinge) and the Smc1/Scc1 (head) interfaces to contain cysteine residues. Upon addition of a thiol-reactive crosslinker, the pairs of cysteines that have been engineered form a covalent bond between them thus covalently closing all three interfaces. This enabled the purification of dimeric circular minichromosomes linked by covalently closed cohesin rings and led to the discovery that these dimers persist even under denaturing conditions (Farcas et al., 2011; Haering, Farcas, Arumugam, Metson, & Nasmyth, 2008). This implies that cohesin must confer cohesion by a topological embrace, ruling out the version of the handcuff model that suggests a protein bridge between cohesin rings.

Furthermore, by expressing two types of the Smc3-Scc1 fusion protein where one, but not the other, can be cleaved on addition of Tobacco Etch Virus (TEV) protease it is possible to distinguish between the topological models. Should the one ring model be the case *in vivo*, on addition of TEV half of the cohesin rings should be cleaved whereas the handcuff or dimeric models should give a higher occurrence of cleavage and thus a lower percentage of TEV resistance should be observed. Since half the dimeric minichromosomes were resistant to TEV cleavage, the most convincing model of cohesion is one ring entrapping both sister chromatids (Haering et al., 2008).

Cohesin loading onto chromosomes

The topological embrace model of cohesin binding to DNA, while providing a mechanism for sister chromatid cohesion, leads to the question of how sister chromatids become trapped inside the ring. Both biochemical data and electron micrographs of the cohesin complex indicate that cohesin forms a ring even in the absence of DNA (Anderson, Losada, Erickson, & Hirano, 2002; Gruber et al., 2003). This implies that cohesin first encounters DNA as a pre-formed ring. In other words, congregation of cohesin's individual subunits around DNA followed by ring formation is unlikely to provide the mechanism for entrapping DNAs. This means that either the DNA itself must be transiently broken to allow cohesin to load onto it by sliding over the newly formed end before the DNA is repaired, or the cohesin ring itself must be transiently opened to enable the DNA to enter the ring. Such large-scale breakage of the DNA seems implausible in a system that aims to maintain the integrity of the genetic material, which leaves the opening of the cohesin ring itself. Most logically this would occur through transient dissociation of one of its three interfaces.

To address which of the interfaces is the "entry gate", fusion proteins of Smc3-Scc1 or Scc1-Smc1 were expressed in yeast in the absence of the wild type proteins. Since these fusions resulted in functional cohesin rings, neither of these interfaces can be the sole entry gate. As the hinge interface cannot be fused in the same manner, owing to it being at the middle of each protein, an alternative method of using dimerisation domains at the hinge was used. Replacing the hinge domains with p14 and MP1 dimerisation domains could not support growth in the absence of the wild type proteins and prevented cohesin loading as measured by microscopy, qPCR and chromatin immunoprecipition.

Furthermore, rapamycin-dependent dimerisation domains FRB and FKBP12 were introduced into flexible loops of the wild type hinges and the effect of dimersation assessed before and after S phase. This showed that dimerisation by addition of rapamycin only prevents sister chromatid cohesion when added to cells prior to S phase and addition after S phase had no effect (Gruber et al., 2006). Both of these experiments are consistent with the entry gate being the hinge and provide evidence that the hinge plays an important role in the loading reaction besides Smc1/Smc3 dimerisation. However, definitive evidence that the hinge is the entry gate is still lacking. For example, the hinge might have a role in cohesin loading besides acting as a DNA entry gate. In *B. subtilis*, the BsSMC hinge domain has been shown to be required for a physical interaction with DNA and also may be involved in stimulating the ATPase activity of the NBDs (Hirano, 2006). Although there is no direct evidence of this in cohesin, these experiments provide an explanation of the importance of the hinge without it having role as the entry gate for DNA.

An alternative possibility is that there is in fact no single entry gate. In other words, there is redundancy between the interfaces. This is a slightly less convincing hypothesis; as opening of more than one interface would leave one of the subunits unbound to the complex, presumably free to diffuse away. To make this work, either the missing subunit would have to bind *de novo* to complete the ring or some mechanism to tether the subunit to the cohesin ring would be necessary. As there is currently no evidence for either of these models, at the moment, the field favours the hinge entry gate model.

In order to load cohesin onto DNA a separate loading complex, sometimes referred to as kollerin, is required. This complex is comprised of two

evolutionarily conserved proteins, Scc2 and Scc4 (Bernard et al., 2006; Ciosk et al., 2000; Furuya, Takahashi, & Yanagida, 1998; Gillespie & Hirano, 2004; Michaelis et al., 1997; Watrin et al., 2006). Scc2 is the larger of the subunits (approximately 170 kDa in *S. cerevisiae*) and is predicted to contain a number of α-helical HEAT (Huntingtin, Elongation factor 3, protein phosphatase 2A, TOR1) repeat motifs towards its C-terminal domain with a more unstructured N-terminus (Andrade-Navarro & Bork, 1995; Chao et al., 2015; Kikuchi, Borek, Otwinowski, Tomchick, & Yu, 2016; Neuwald & Hirano, 2000). It is the first 181 residues of this more unstructured N-terminal domain that is sufficient to bind Scc4 (approximately 70 kDa in *S. cerevisiae*). A recent structure of the N-terminus of Scc2 bound to Scc4 revealed that Scc4 forms a superhelical array of 13 TetratricoPeptide Repeat (TPR) modules interrupted with a beta ribbon insertion between repeats 6 and 7. The very N-terminal residues (10-50) of Scc2 are embedded within the inner cavity of Scc4, emerging at both ends but with residues 58-132 making extensive contacts with the external surface of Scc4 (Chao et al., 2015; Hinshaw, Makrantoni, Kerr, Marston, & Harrison, 2015). This implies that, since the concave surface of Scc4 is entirely enclosed, in order to dissociate from Scc2 either proteolysis or unfolding of one of the two proteins would be required.

How the loading complex interacts with the cohesin ring is not entirely understood. A weak/transient interaction is indicated from data acquired by mass spectrometry of purified loading complexes and also from co-immunoprecipition studies (Arumugam et al., 2003; Tóth et al., 1999). Recently, it has been shown that Scc2 interacts with the N-terminal region of Scc1 and

competes for binding with Pds5, a structurally related protein that will be discussed in more detail later (Kikuchi et al., 2016).

A further requirement of cohesin loading is Scc3, a cohesin-associated protein that is stably and stoichiometrically bound to the cohesin ring (Chan et al., 2012; Haering et al., 2002). The structure of Scc3 reveals that either half of the protein is composed of different elements. The N-terminal half of the protein begins with 9 irregularly distributed α-helices followed by a helix-loop-helix that partially protrudes from the bulk of the protein and 3 HEAT repeats, whereas the C-terminal half is comprised of a tandem array of anti-parallel α-helices that resemble HEAT repeats and is separated from the N-terminal half of the protein by 4 inter-crossed α-helices that act to change the axis between the two halves. Scc3 binds to the C-terminal end of Scc1 through several contacts within the inner surface of its curved backbone (Roig et al., 2014). This interaction is required for cohesin to associate with DNA but it is important to note that there is currently no evidence that Scc3 plays a role in the loading reaction *per se* (Hu et al., 2011).

The final requirement for cohesin loading is a complete cycle of ATP hydrolysis at the NBDs of both Smc1 and Smc3. Mutations that prevent ATP binding are lethal and cause cohesin to no longer associate with DNA. Similarly, mutations that prevent the hydrolysis of ATP bound to either of the SMC NBDs are lethal and, although cohesin is still recruited to centromeres, the stable association of cohesin with chromatin is abolished indicating that ATP hydrolysis defective cohesin is unable to complete the loading reaction (Figure I-6) (Arumugam et al., 2003; Hu et al., 2011; Weitzer, Lehane, & Uhlmann, 2003).

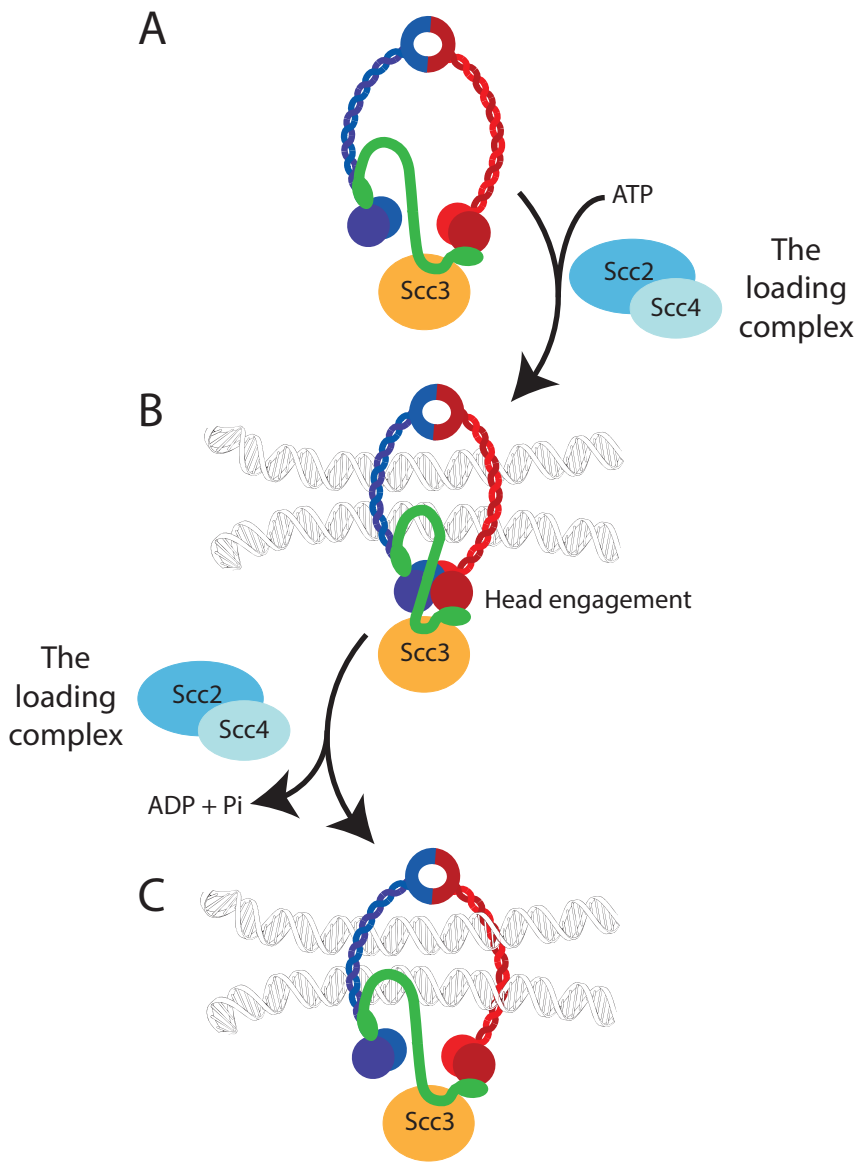


Figure I-6

The loading reaction.

A -The SMC NBDs come together to bind ATP. Cohesin associates with the DNA with the loading complex.

B - Cohesin loads onto DNA in a manner dependent on the loading complex, Scc3 and ATP hydrolysis.

C - Cohesin topologically entraps sister chromatids.

Localisation of cohesin and of its loading sites on chromatin

The localisation of cohesin on chromosomes has been described by many genome-wide studies using chromatin immunoprecipitation techniques (Blat & Kleckner, 1999; Laloraya, Guacci, & Koshland, 2000; Megee, Mistrot, Guacci, & Koshland, 1999; Tanaka, Cosma, Wirth, & Nasmyth, 1999). Cohesin is present throughout the genome, although not uniformly, with cohesin-associated regions (CARs) spaced at 2-35 kb intervals over the chromosome and tending to correlate with the intergenic space between convergent genes (Filipski & Mucha, 2002; Glynn et al., 2004; Lengronne et al., 2004). The predominant site of cohesin enrichment is a 20-50 kb region surrounding the centromere (Blat & Kleckner, 1999). This is particularly important in yeast as it defines the pericentromere, which lacks heterochromatin unlike pericentromeres of other eukaryotes. Pericentric cohesion facilitates biorientation of sister chromatids on the metaphase plate possibly by generating an appropriate geometry for the interaction with the spindle (Ng, Waples, Lavoie, & Biggins, 2009).

In mammalian cells, it has been observed that cohesin frequently co-localises with CCCTC-binding factor (CTCF) (Wendt et al., 2008). CTCF is a site-specific zinc finger DNA binding protein that has been suggested to define the boundaries between active and inactive chromatin domains (Merkenschlager & Nora, 2016). It was observed that cohesin enrichment depended on CTCF but not vice versa and that CTCF depletion did not reduce the amount of cohesin on chromosomes (Parelho et al., 2008; Wendt et al., 2008). This implies that CTCF does not actively promote cohesin loading but does affect its localisation, causing it to accumulate at CTCF sites. It is thought that CTCF may direct cohesin to regulate transcription by the formation of loops with cohesin entrapping distant

chromosome segments within its ring (Guillou et al., 2010; Holwerda & de Laat, 2013; Merkenschlager & Nora, 2016).

Sites of cohesin loading have also been described by genome-wide mapping of both components of the loading complex. This revealed sites distinct from that of cohesin enrichment but which do correlate with the regions that cohesin is first detected at upon cell cycle entry (Lengronne et al., 2004). Similarly, cohesin mutants that are unable to hydrolyse ATP, and are therefore unable to complete the loading reaction, also co-localise with the loading complex (Hu et al., 2011). Thus it is thought that cohesin translocates away from these loading sites and accumulates at CARs. How cohesin translocates across the chromosome is not well understood but transcription has been shown to alter the distribution of cohesin, suggesting it might be involved (Lengronne et al., 2004).

In the case of the centromere, the loading complex associates with the core (~125 bp) centromere, in yeast conferred by three small centromere DNA elements (CDEI, CDEII, CDEIII). These centromeric sequences confer high rates of loading throughout a large window of approximately 30 kb either side of the CDEIII. They also to promote the loading of cohesin onto minichromosomes and relocation of these sequences to a chromosome arm eliminates cohesin enrichment at the endogenous centromere and promotes its enrichment at the ectopic site (Tanaka et al., 1999; Weber et al., 2004). This implies that the centromeric sequences are necessary to produce these highly enriched regions at the pericentromere.

The Ctf19 subcomplex, a component of the kinetochore, has also been identified as important specifically for centromeric cohesin enrichment (Eckert,

Gravdahl, & Megee, 2007; Fernius & Marston, 2009; Hinshaw et al., 2015). Since the loading complex fails to associate with the centromere in mutants that lack components of the Ctf19 complex but cohesin enrichment can be partially rescued upon artificial tethering of Scc4 to the centromere (Fernius et al., 2013), its stands to reason that the role of the Ctf19 complex is to direct the loading complex to centromeres. The mechanism of the Ctf19 complex action is still unknown but it has been shown to enable Dbf4-dependent kinase (DDK) to associate with the kinetochore from telophase to G_1. DDK, when associated with kinetochores, also attracts the loading complex to centromeres in late G_1 and therefore it is possible that DDK creates a binding site for Scc4 by phosphorylating a component of the Ctf19 complex (Natsume et al., 2013). A conserved patch on Scc4 itself is also required for the targeting of the loading complex to centromeres (Hinshaw et al., 2015). Although its interaction partners are still unknown, this patch could be the interface for an interaction with Ctf19.

Although the centromere is the most studied cohesin loading site, these studies have also revealed that it is not essential. Mutations that abolish centromeric enrichment are not lethal and do not effect cohesin loading on the arms (Fernius & Marston, 2009; Hinshaw et al., 2015). This implies that two modes of cohesin loading may be occurring; one that depends on the conserved patch in Scc4 that happens at centromeres, and a mode that occurs everywhere on the chromosome and is sufficient to support viability.

Cohesin removal from chromosomes

As previously discussed, cohesion is rapidly lost upon activation of the APC/C at the metaphase-to-anaphase transition. Once activated, the APC/C polyubiquitinates securin and cyclin B, marking them for degradation. This activates the protease separase, which then cleaves Scc1 and thus irreversibly opens the cohesin ring, destroying cohesion and allowing the sister chromatids to be pulled towards opposite poles (Musacchio & Salmon, 2007; Uhlmann et al., 1999; 2000).

In fact, a second cohesin removal pathway exists, called the prophase pathway. This pathway, as the name suggests, operates in prophase before activation of the APC/C and does not depend on cleavage by separase. Instead it requires a regulatory subunit associated with cohesin called Wapl, and removes cohesin from the DNA in a cleavage-independent manner (Gandhi, Gillespie, & Hirano, 2006; Kueng et al., 2006). Although budding yeast doesn't remove cohesin via the prophase pathway in the same way, it does use the same mechanism to remove cohesin from the DNA throughout the cell cycle (Chan et al., 2012).

Wapl interacts with the cohesin ring though Pds5, another cohesin-associated protein, and Scc3 (Gandhi et al., 2006; Kueng et al., 2006). Pds5 is another large, HEAT repeat-containing protein (147 kDa in *S. cerevisiae*) that binds directly to the cohesin ring via Scc1 and is essential for sister chromatid cohesion (Ouyang, Zheng, Tomchick, Luo, & Yu, 2016; Panizza, Tanaka, Hochwagen, Eisenhaber, & Nasmyth, 2000). It is now understood that Wapl binding to Pds5 and Scc3, and therefore to the cohesin complex, induces dissociation of the N-terminus of Scc1 from Smc3 thus providing an exit gate for

the DNA (Beckouët et al., 2010; Ouyang et al., 2016). This is also a useful mechanism for cohesin rings that have been cleaved by separase as, in the absence of such Wapl-mediated release, the N-terminus of Scc1 remains bound to the Smc3 head. In such a situation, the heterodimer could not be reused in the next cell cycle, as it would be unable to bind the newly synthesised, uncleaved Scc1. This indicates that this release pathway has a role in freeing the SMC heterodimer from Scc1 for reuse in the next cell cycle (Beckouët et al., 2016).

It is interesting that Wapl-mediated release requires at least part of an ATP hydrolysis cycle. Recent studies have described mutants in the Smc1 ATPase (D1164E and L1129V) that are defective in ATP hydrolysis but are viable, associate stably with DNA and confer sister chromatid cohesion. This is in stark contrast to the previously described hydrolysis mutants (*smc3(E1155Q)* and *smc1(E1158Q)*), which are lethal and do not stably associate with DNA. Thus, they do not confer sister chromatid cohesion. Furthermore these viable Smc1 ATPase mutants stabilise cohesin on chromatin by preventing Wapl mediated release. This is again contrary to the EQ mutants that block hydrolysis but do not prevent release. Since this property is not shared across all hydrolysis mutants, it implies that ATP hydrolysis *per se* is not required for release but that perhaps an earlier step in the ATP hydrolysis cycle is. It has been suggested that engagement of the SMC heads, required for both loading and release, may have more than one state with each state promoting either loading or release (Figure I-7) (Beckouët et al., 2016; Elbatsh et al., 2016). This release engaged state might be defective in these Smc1 ATPase mutants preventing Wapl mediated release.

It has been suggested that Wapl-mediated release of the N-terminal part of Scc1 from Smc3 may also create the entry gate during the cohesin loading

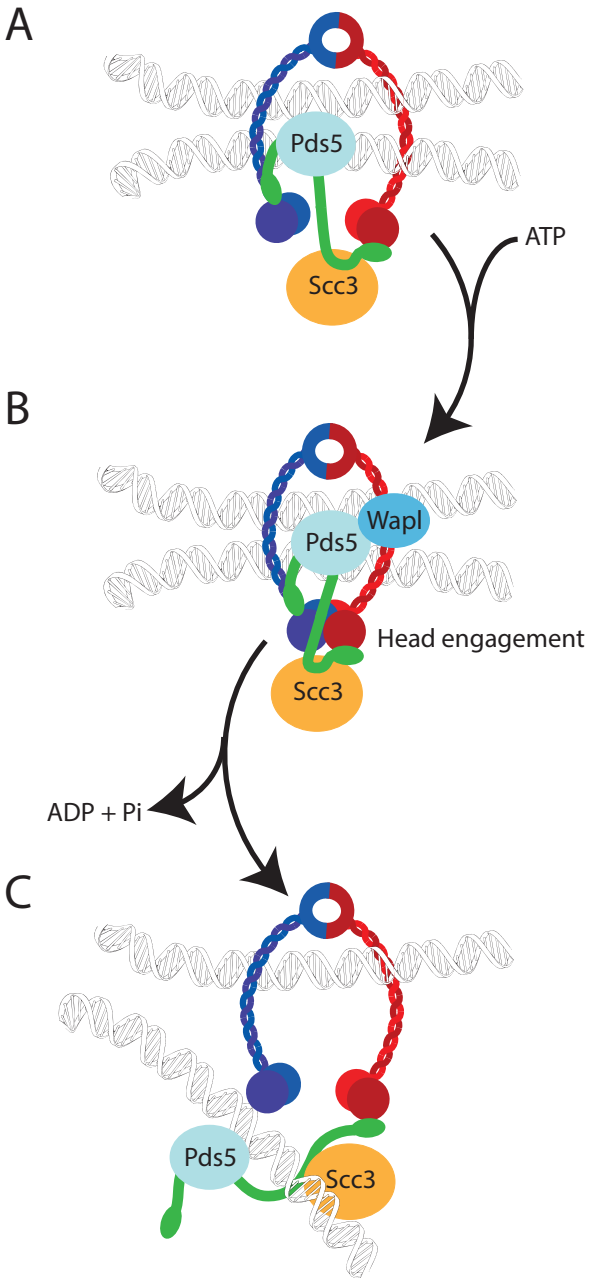


Figure I-7

Wapl dependent removal of cohesin from DNA.

A - Once cohesin is loaded onto DNA, Pds5 maintains their topological embrace.

B - Upon ATP binding the SMC NBDs engage and Wapl binds to the complex via Pds5.

C - Wapl binding induces dissociation of the N-terminus of Scc1 from Smc3 providing an exit gate for DNA.

reaction. *In vitro* studies using cohesin proteins purified from *S. pombe* have demonstrated that loading can occur in the absence of the loading complex but in the presence of Wapl (Murayama & Uhlmann, 2013; 2015). However, Wapl is not essential for loading. If it plays such an important role in the loading reaction, why then is it not essential? To explain this, it has been claimed that spontaneous opening of the Smc3-Scc1 interface can complete the reaction in the absence of Wapl, but at lower rate (Murayama & Uhlmann, 2015). Thus also attempting to explain why there are lower levels of cohesin on DNA in the absence of Wapl. But this explanation fails to adequately account for the observation that covalently fusing the Smc3 and Scc1 interfaces together does not prevent loading (Gruber et al., 2006). Thus further experiments are needed to explore any potential role of Wapl in loading *in vivo*.

Cohesin establishment and maintenance

As cohesin is constantly turning over in G_1 due to the opposing actions of the loading complex and Wapl mediated release, how then does it become stably bound before M phase in order to provide sister chromatid cohesion? This occurs through the action of the acetyl transferase Eco1, which has been shown to acetylate two adjacent and highly conserved lysines in the N-terminal head domain of Smc3 (Rolef Ben-Shahar et al., 2008; Rowland et al., 2009; Ünal et al., 2008; J. Zhang et al., 2008a). Acetylation at these sites after DNA replication prevents dissociation of the N-terminus of Scc1 from Smc3 and thereby stabilises cohesin on DNA (Beckouët et al., 2016). After the loss of cohesion at anaphase, Hos1 deacetylates Smc3 enabling turnover in the next cell cycle. Thus, the timing

of the acetylation state of cohesin reflects the transitions between stable and unstable cohesin populations (Gerlich, Koch, Dupeux, Peters, & Ellenberg, 2006). The coupling of Eco1 activity to replication has been suggested to occur via a physical interaction with a component of the replication fork, proliferating cell nuclear antigen (PCNA), but this is not well understood (Moldovan, Pfander, & Jentsch, 2006).

In vertebrates, an additional level of protection from Wapl activity exists in the form of sororin, another target of the APC/C that is essential for sister chromatid cohesion. Sororin, like Wapl, binds to the cohesin complex through Pds5. In fact, both Wapl and sororin contain YSR motifs and compete for the same binding site on Pds5 in the N-terminal domain. Thus sororin antagonises the action of Wapl by displacing it from Pds5. Binding of sororin depends on DNA replication and acetylation of Smc3 by Eco1, but phosphorylation by Cdk1 and Aurora B destabilises sororin's interaction with Pds5. This enables regulation of cohesin protection in a cell cycle dependent manner (Ladurner et al., 2016; Nishiyama et al., 2010; Ouyang et al., 2016; Rankin, Ayad, & Kirschner, 2005; N. Zhang, Panigrahi, Jiang, Rao, & Pati, 2010). Although no sororin candidate has been identified in yeast, distant orthologs exist in many metazoan species such as Dalmatian in *Drosophila melanogaster* (Nishiyama et al., 2010).

Regulation of cohesin release by sororin is particularly important in the prophase pathway. In vertebrates, removal of cohesin in mitosis occurs in two steps, first the majority of the cohesin on the arms is removed via the prophase pathway leaving a small amount at the centromeres that requires the timely cleavage by separase. But what then prevents centromeric cohesin from being removed during prophase? Cell cycle directed binding of Sgo1 and PP2A to

centromeric cohesin causes the local dephosphorylation of sororin and therefore stabilisation of cohesin at centromeres (H. Liu, Rankin, & Yu, 2012).

Cohesinopathies

Cohesin is well known for its role in providing sister chromatid cohesion during the cell cycle and, as previously mentioned, its malfunction can have severe consequences. Errors can be lethal or cause aneuploidy to occur. Aneuploidy in somatic cells can be a hallmark of oncogenesis, and in the germline, aneuploidy can cause developmental disorders in offspring. Genes encoding cohesin subunits or associated proteins are somatically mutated in a wide range of human cancers (Hill, Kim, & Waldman, 2016). However, mutations within cohesin and its associated proteins give rise to additional problems including Roberts Syndrome (RBS) and Cornelia de Lange Syndrome (CdLS). Collectively these are known as cohesinopathies.

RBS is a rare autosomal recessive disorder characterised by growth retardation leading to limb deformities, craniofacial defects and mental deficiency. Metaphase spreads of RBS cells distinctly show premature centromere separation indicating that sister chromatid cohesion is compromised. RBS is caused by mutations in the human ortholog of Eco1, ESCO2, which promotes the unstable deacetylated state of cohesin and thus compromises cohesion (B. Xu, Lu, & Gerton, 2014).

CdLS is more frequent, estimated to occur once in every 10,000 births. It is characterised by pre- and post-natal growth delays, facial dysmorphism, cognitive impairment and malformation of limbs and internal organs (Dorsett &

Krantz, 2009). CdLS is caused by heterozygous dominant mutations in the genes encoding the human versions of Smc1, Smc3, Hos1, Scc1 and Scc2. Mutations in *SMC1* or *SMC3* are always missense mutations and give much milder phenotypes (Deardorff et al., 2007). Nonsense or missense mutations in *HDAC8*, the human ortholog of *HOS1*, result in hyperacetylation of Smc3 promoting the retention of the cohesin complex during mitosis (Deardorff et al., 2012). However over 55% of CdLS cases are linked to mutations in the human *SCC2* ortholog, *NIPBL*. In these cases, more severe phenotypes are seen with frameshift or nonsense mutations that result in NIPBL haploinsufficiency (Krantz et al., 2004; Tonkin, Wang, Lisgo, Bamshad, & Strachan, 2004). Interestingly, CdLS patients do not present any obvious sister chromatid cohesion defects. These observations suggest that cohesin has another role as well as its role in sister chromatid cohesion. Patient lymphoblasts as well as tissues and cells from a Nipbl-mutant mouse have shown that the modest drop in the level of Nipbl is accompanied by small but widespread changes in gene expression (Kawauchi et al., 2009; J. Liu et al., 2009). This suggests that the phenotypes observed in CdLS patients may be due to a defect in cohesin-mediated transcription regulation.

There have been many studies into another role of cohesin in the regulation of transcription. But the discovery that cohesin inactivation in post-mitotic mushroom-body γ-neurons in *Drosophila* causes pruning defects clearly demonstrated that cohesin has non-mitotic functions. Furthermore, the finding that these pruning defects are partly due to a lack of expression of the ecdysone receptor indicated a role in the regulation of transcription (Pauli et al., 2008; Schuldiner et al., 2008).

As well as effects on transcriptional regulation, an increased sensitivity to DNA damage has also been reported in CdLS patients and seems to be a general feature of cohesinopathies (Vrouwe et al., 2007). The role of cohesin in DNA damage repair was actually discovered before its role in sister chromatid cohesion. Scc1 in *S. pombe* was first identified as Rad21 for radiation sensitive as its mutation caused cells to be hypersensitive to ultraviolet radiation (UV) or ionizing radiation (IR) (Birkenbihl & Subramani, 1992). Further studies revealed a particular role of cohesin in the repair of double strand breaks by homologous recombination. Cohesin is thought to confer sister chromatid cohesion at or near double stand breaks in order to promote strand invasion required for homologous recombination (Schär, Fäsi, & Jessberger, 2004; Sjögren & Nasmyth, 2001).

Together cohesinopathies highlight the importance of cohesin not only in its role in sister chromatid cohesion but also in its other functions including transcriptional regulation and DNA damage repair.

Thesis Objective

The aim of this thesis is to investigate the mechanism of the cohesin loading reaction. Despite many recent advances in the cohesin field, this aspect is still poorly understood. Chromatin immunoprecipitation-sequencing (ChIP-seq) has often been utilized to investigate the loading reaction but this method does not allow quantitative comparison between samples and therefore identified changes across samples can be misrepresentative. A recently developed method of quantitative ChIP-seq has provided an opportunity to revisit the current understanding of the loading reaction obtained from conventional ChIP-seq. A quantitative approach may reveal aspects of the loading previously overlooked due to the limitations of a non-quantitative approach. In addition, a genetic screen performed Dr Maria Demidova in our lab identified a mutation within the cohesin complex that is able to bypass the requirement for the essential part of the cohesin loading complex Scc4. Further characterisation of this mutation and subsequent screens to identify additional suppressors are discussed and provide novel insights into the cohesin loading reaction.

Quantitative ChIP-sequencing reveals novel insights into the cohesin loading reaction

Introduction

The cohesin complex has been implicated in many of the cell's basic functions including chromosome segregation, transcriptional regulation and DNA repair, and is thought to use the same fundamental mechanism of topological entrapment in all of its roles. Such topological entrapment logically would involve opening of at least one of the three interfaces that are created on formation of the trimeric ring structure. However the mechanism of the loading reaction is very poorly understood.

A technique that is often used to investigate the loading reaction *in vivo* is Chromatin Immunoprecipitation (ChIP) coupled with whole genome sequencing (ChIP-seq). Briefly, this involves purification of short fragments of DNA that have been chemically crosslinked to DNA-associated proteins and thus can be co-immunoprecipitated with an antibody specific to a protein of interest. Sequencing the DNA fragments enables identification of the binding sites of these DNA-associated proteins genome-wide *in vivo*. A genome-wide map of global binding sites can be very informative when studying chromatin-associated proteins, and ChIP-seq has been used extensively to map histone modifications and transcription factors (Barski et al., 2007; Johnson, Mortazavi, Myers, & Wold, 2007; Mikkelsen et al., 2007; Robertson et al., 2007). In the case of cohesin, it can identify the localisation on the DNA of the cohesin ring itself as well as of its associated regulatory proteins, including the loading complex. This can provide information on where the loading reaction occurs throughout the genome and,

by assessing co-localisation, also determine which other proteins are present. Furthermore ChIP-seq can be used to map the effects of mutations within the cohesin ring or its associated factors on their distribution, providing additional information on the requirements for completion of the loading reaction.

Despite its many advantages, ChIP-seq has three major limitations. Firstly, the use of formaldehyde to crosslink protein to DNA within the cell means that the resulting data is prone to artifacts. This is because formaldehyde crosslinking of double stranded DNA to protein is much less efficient than crosslinking single stranded DNA to protein, inferring an inherent bias in the system (Hoffman, Frey, Smith, & Auble, 2015). Unfortunately, this problem is most likely insurmountable. Thus, where possible, additional techniques that allow the observation of proteins within cells, such as fluorescence microscopy, are still needed to verify findings obtained by ChIP-seq. Secondly, ChIP-seq cannot reveal actual occupancy of proteins over the genome; it can only provide an estimate of occupancy and the relative occupancy between regions. It should be noted that the troughs between peaks of occupancy do not necessarily mean that the protein is not present at these sites, but that it may be below the detection limit. As the mechanism of formaldehyde crosslinking is very poorly understood it is unlikely that this problem can be addressed either. The final limitation occurs when attempting to compare ChIP-seq profiles between samples. These might be from the same strain in different experimental conditions or from different strains that, for example, contain different mutations. Although ChIP-seq can reveal changes across samples in terms of the genome-wide distribution of a protein of interest, it is unable to quantitatively detect global changes in protein occupancy. This means that if a set of conditions or mutations alters the amount

of the protein that associates with DNA genome-wide but not its distribution, no difference would be observed between the samples. This limitation can be overcome to some extent with quantitative PCR (qPCR); however, this technique can only sample specific loci and does not give a genome-wide comparison. Therefore any change detected by qPCR might reflect a change in distribution but not in the efficiency of loading *per se*. Furthermore, controlling for variation between ChIP samples is extremely difficult, given that the efficiency of each step in the protocol, such as lysis, sonication and immunoprecipitation, will always vary slightly between samples. These variations will accumulate throughout the experiment and can lead to subtle differences in the resulting profiles. Thus an identified change in occupancy across samples can be misrepresentative. This is a major problem as experiments that compare different cell states or assess the effect of different mutations are the most informative. Lack of quantification has also led to a difficulty in distinguishing signal from noise; the analysis of ChIP-seq data often therefore focuses on clear peaks in the signal while somewhat ignoring the signal between peaks.

To address the lack of quantitation a new method of performing ChIP-seq has been developed that enables quantitative comparison across ChIP-seq profiles (Bonhoure et al., 2014; Hu et al., 2015; Orlando et al., 2014). The method uses a calibration species that is 'spiked in' to the experimental species as an internal control that allows normalisation across samples with respect to the calibration species. For this to work, the calibration species must (I) have a genome that can be distinguished from the experimental species, (II) express the same epitope as the experimental species, and (III) be sufficiently similar in physiology that the processes of fixation, lysis, sonication and

immunoprecipitation can be preformed in the same way. Thus by introducing cells of one sample of the calibration species in a single state to all the samples of experimental species, the sequences from the calibration genome can be used as an internal control to calibrate the experimental samples. This is because the calibration profile should be the same in every sample it is added to and therefore, any changes detected between calibration profiles would then be due to differences in efficiency in steps of the ChIP-seq protocol and thus can be used to calibrate the experimental profile.

For studying cohesin loading in *S. cerevisiae*, the haploid yeast *Candida glabrata* was identified as a suitable calibration genome (Hu et al., 2015). *C. glabrata* is closely related to *S. cerevisiae* and so can be manipulated in the same way in terms of growth and lysis of the cells. Importantly, the genome of *C. glabrata* is sufficiently different to that of *S. cerevisiae* that sequencing reads of over 50 bp can generally be uniquely assigned to one of the two genomes. The only exceptions are sequences that encode the rDNA locus as well as some highly conserved genes involved in translation and metabolism. However, removal of those sequences that align to both genomes does not significantly alter the final profile obtained. Furthermore, addition of *C. glabrata* does not affect the occupancy profile observed from *S. cerevisiae*, and the relative percentage of the spiked sample has no effect on the profile but, as expected, the number of reads that align to *S. cerevisiae* increases linearly as the percentage of *C. glabrata* decreases (Hu et al., 2015).

This chapter aims to use this new method of quantitative ChIP-seq in *S. cerevisiae* both to validate the method based on what is currently known about

the cohesin loading reaction, and to see what novel observations can be obtained by using this quantitative approach.

Results

Measuring the localisation of cohesin by quantitative ChIP-seq.

To investigate the localisation of cohesin by calibrated ChIP-seq, a *S. cerevisiae* strain in which endogenous *SCC1* is tagged with PK_6 (K22005) was spiked with a *C. glabrata* strain in which endogenous *SCC1* is tagged with PK_9 (K23308). To control for background signal, an *S. cerevisiae* strain that has no PK epitope (K699) was also used. All three strains were grown to logarithmic phase, then 12.5 OD_{600} units of each *S. cerevisiae* strain were mixed with 5 OD_{600} units of the *C. glabrata* culture prior to fixation.

Once sequenced, each sample consists of a mix of reads from both *S. cerevisiae* and *C. glabrata*. In order to separate the reads into those uniquely aligning to *S. cerevisiae* and those uniquely aligning to *C. glabrata*, first, any reads shorter than 50bp were removed and the remaining reads aligned to one of the genomes. The reads that failed to align can then be aligned to the other genome to give the reads that uniquely align to the second genome. A method is then needed to calibrate the resulting ChIP-seq profile of aligned reads to give a quantitative profile. This is done by calculating the occupancy ratio (OR), a ratio that measures the probability of an epitope being associated with the experimental genome compared to the probability of the same epitope associated with the calibration genome (Figure II-1).

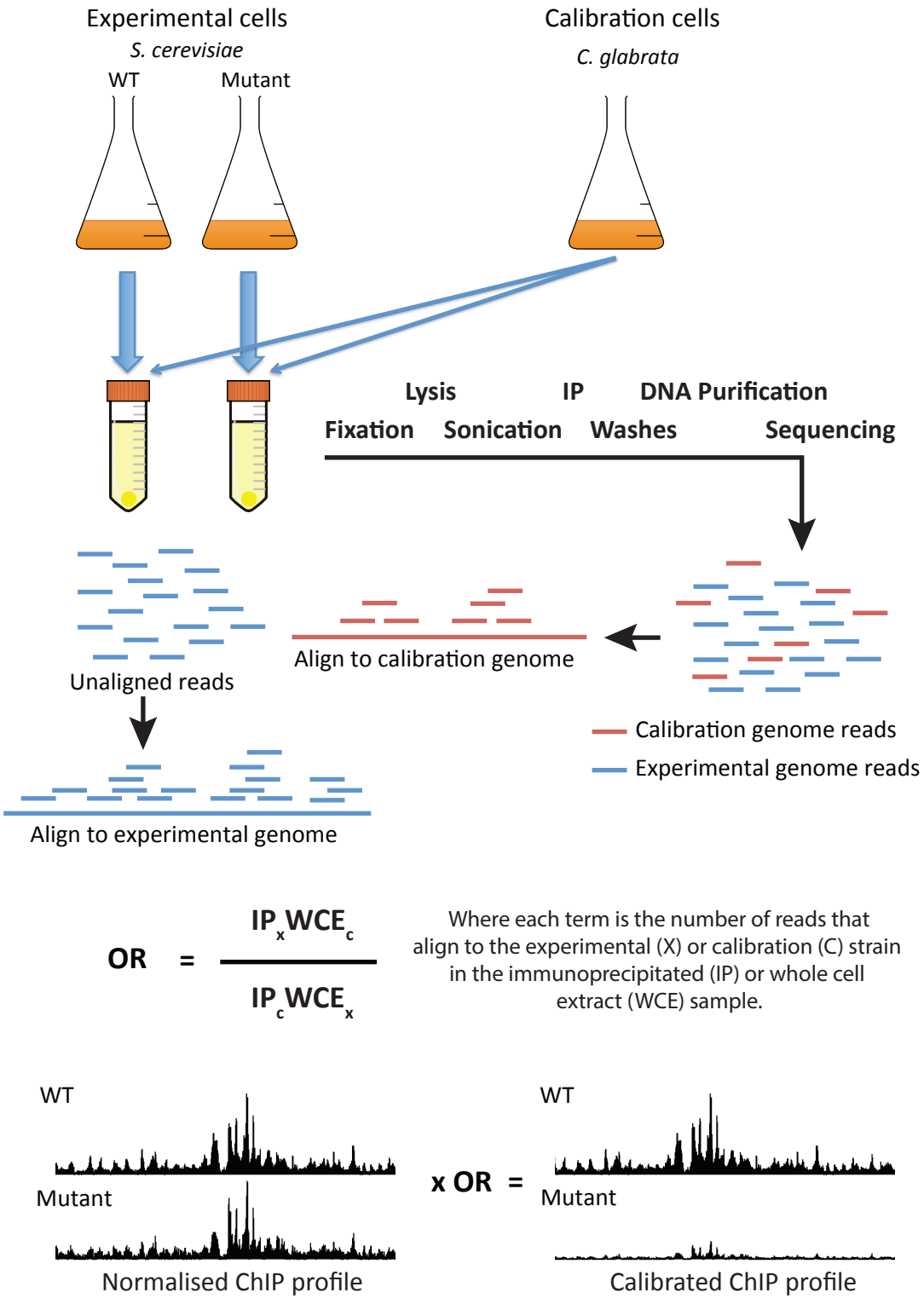


Figure II-1

Schematic of the quantitative ChIP-sequencing protocol.

Experimental cells are mixed with calibration cells prior to the ChIP protocol. Samples are then sequenced and the reads aligned uniquely to each genome by subsequent alignments. The resulting profiles are calibrated by muliplying by the occupancy ratio (OR) which is derived from the number of reads aligning to each genome.

The occupancy ratio can be derived mathematically (Hu et al., 2015) as follows:

Let,

N_x or N_c = the number of cells of the experimental (X) or calibration (C) strain.

IP_x or IP_c = the number of reads co-immunoprecipitated from either strain.

O_x or O_c = the occupancy of the epitope averaged over either genome.

E_x or E_c = the combined efficiency of the ChIP-seq protocol.

F_x or F_c = the fraction of reads uniquely aligning to the experimental or calibration genome.

Then the number of reads immunoprecipitated from either strain should be equal to the number of cells multiplied by the occupancy and efficiency:

$$IP_x = N_x O_x E_x$$

$$IP_c = N_c O_c E_c$$

The fraction of reads co-immunoprecipitated from the experimental genome is equal to the number of reads co-immunoprecipitated from the experimental strain divided by the total number of reads co-immunoprecipitated:

$$F_x = \frac{IP_x}{IP_x + IP_c}$$

Substituting gives:

$$F_x = \frac{N_x O_x E_X}{N_x O_x E_x + N_c O_c E_c}$$

This can be arranged to give:

$$\frac{O_x E_x}{O_c E_c} = \frac{F_x N_c}{N_x(1 - F_x)}$$

E_x and E_c would vary across two different samples; however, as the strains are mixed at the very beginning of the protocol, the variation in efficiency will be identical for both strains. Also as the calibration cells in each sample are grown under constant conditions, O_c will be constant. Thus $O_c E_c / E_x$ will be a constant that we will call α. Thus:

$$O_x = \alpha \frac{F_x N_c}{N_x(1 - F_x)}$$

This shows that O_x is directly proportional to the fraction shown above which is the occupancy ratio (OR). This occupancy ratio can be measured as the ratio of calibration cells to experimental cells (N_c/N_x) and is also given by the ratio of the number of reads from the calibration cells to the number of reads from the experimental cells in a sample that has not been subjected to immunoprecipitation (whole cell extracts; WCE). F_x can be derived from the sequencing data, as the number of reads aligning to each genome is defined by the IP samples (IP_x and IP_c). Also 1-F_x can be substituted for F_c, which is also measurable. Therefore if WCE_x and WCE_c are the number of reads uniquely aligning to each genome from the WCE samples then the OR can be simplified to become:

$$OR = \frac{IP_x WCE_c}{IP_c WCE_x}$$

Thus the sequencing profiles can be calibrated by multiplying the profile by the OR and then normalising to 1 million reads.

This approach was applied to the analysis of Scc1-PK_6 and untagged strains in logarithmic growth phase. As previously observed by conventional ChIP-seq, cohesin is enriched around the centromere and is also present at a significant level throughout chromosome arms when compared to the background signal seen with the untagged strain (Figure II-2A). This shows that cohesin is distributed throughout the genome, being enriched at some sites on the arms and, most notably, around centromeres.

To better characterise cohesin occupancy around centromeres, a profile detailing the average occupancy of cohesin around a centromere was generated. To do this, all 16 chromosomes were aligned by their CDEIII elements, one of the highly conserved centromere DNA elements in *S. cerevisiae*, and the number of reads averaged at each base pair across the chromosome 10 kb either side of CDEIII. This process revealed four distinct peaks, two directly flanking CDEIII and two 5 kb away from CDEIII (Figure II-2B). These peaks at ± 5 kb are relatively broad, covering approximately 4 kb and represent a large fraction of total cohesin associated with DNA around centromeres, implying they might correspond to the 'barrel' structure observed by high resolution microscopy in biorientated cells expressing fluorescently labeled cohesin (Hu et al., 2015; Yeh et al., 2008). Similarly, the proximity of the other two peaks to the CDEIII element suggests that these correspond to cohesin that can be seen very close to kinetochores marked with fluorescently labeled Mtw1. In this thesis the 5 kb peaks will be referred to as pericentromeric cohesin and the peaks directly

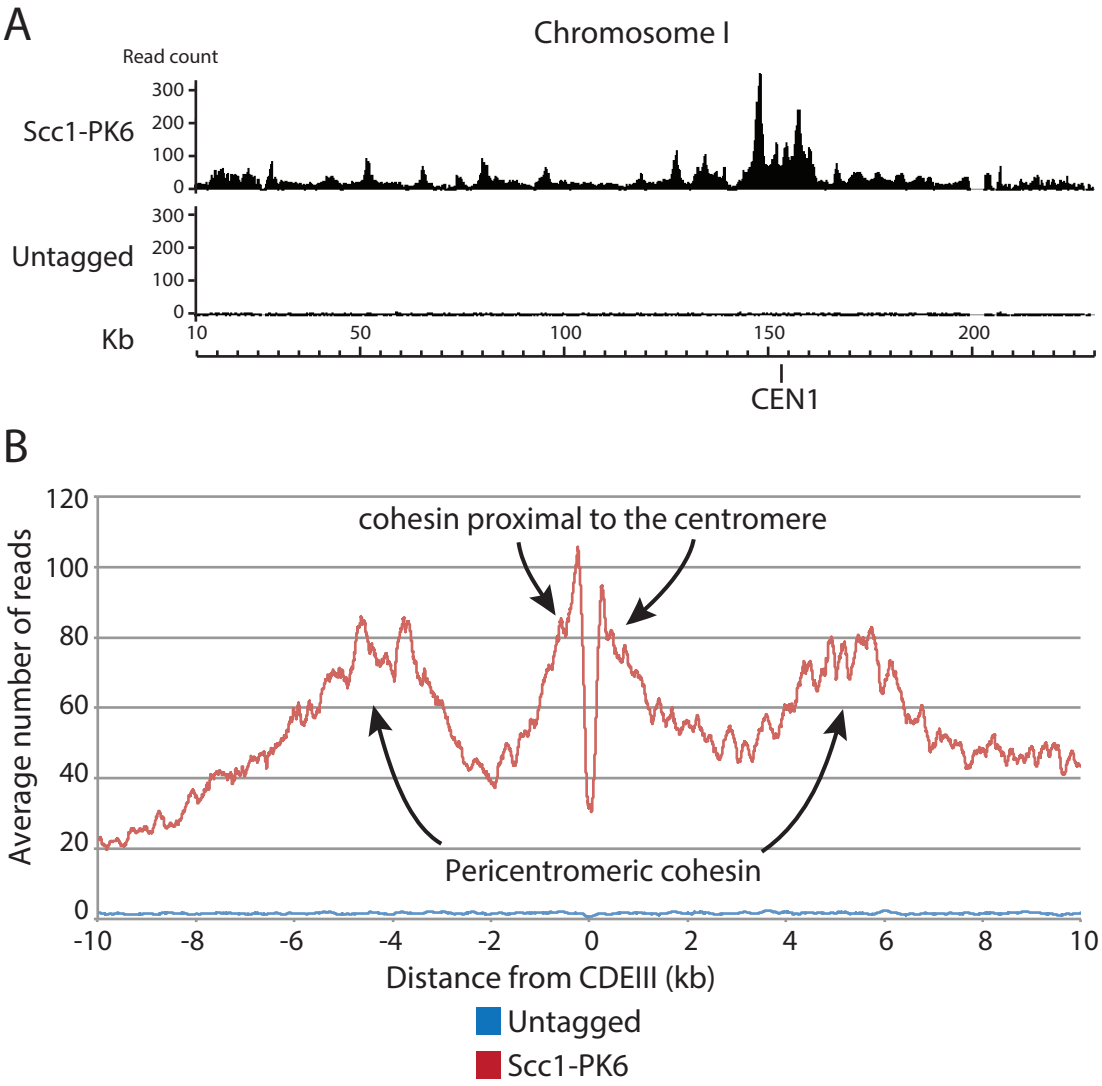


Figure II-2

Cohesin is distributed throughout the chromosome but is highly enriched at centromeres.

Cultures of cycling cells at 25°C expressing endogenous *SCC1* either tagged with PK6 (K22005) or untagged (K699) were analysed by quantitative ChIP-seq.

A - Quantitative ChIP-seq profile across chromosome I. The Y axis indicates the calibrated number of reads and the X axis, the base pair position across chromosome I.

B - The number of reads at each base pair in the 10 kb region either side of the CDEIII element of each chromosome were averaged over all 16 chromosomes to give a calibrated average cohesin distribution around centromeres. The y axis is the calibrated average number of reads and the x axis the position away from the centromere.

flanking CDEIII will be referred to as cohesin proximal to the centromere. Interestingly, cohesin occupancy at the CDEIII element itself is relatively low.

Cohesin localisation on chromosomes depends on the loading complex.

Quantitative ChIP-seq is a suitable technique to identify cohesin's global distribution on chromatin. However, it does not allow discrimination between stably bound cohesin and freely diffusing molecules of Scc1/cohesin that happen to be close enough to DNA to crosslink to it. Therefore it is necessary to determine what proportion of the cohesin association observed by quantitative ChIP-seq is loaded cohesin.

As discussed in Chapter I, the cohesin ring requires a separate complex in order to load onto DNA. To assess if the observed cohesin occupancy across the genome is due to loaded cohesin or also due to soluble cohesin rings, cells containing temperature sensitive (ts) alleles of the loading complex were analysed at the restrictive temperature at which the protein is inactive and cohesin should therefore not be loaded onto DNA. Thus if the cohesin occupancy observed is due to cohesin loaded onto DNA, then cohesin should not be present in cells containing ts alleles of the loading complex, whereas the occupancy of soluble cohesin would be unaffected.

Cells expressing the ts alleles of Scc2 (*scc2-4* or *scc2-45*) or Scc4 (*scc4-4*) and *SCC1-PK$_{6/9}$* were arrested in G_1 with α-factor and released into media pre-warmed to the restrictive temperature (37°C). The cell cycle progression of each strain was measured by FACS (Figure II-3B and 3D) and ChIP-seq samples taken when the cells had entered G_2. In all cases the cells containing ts alleles of

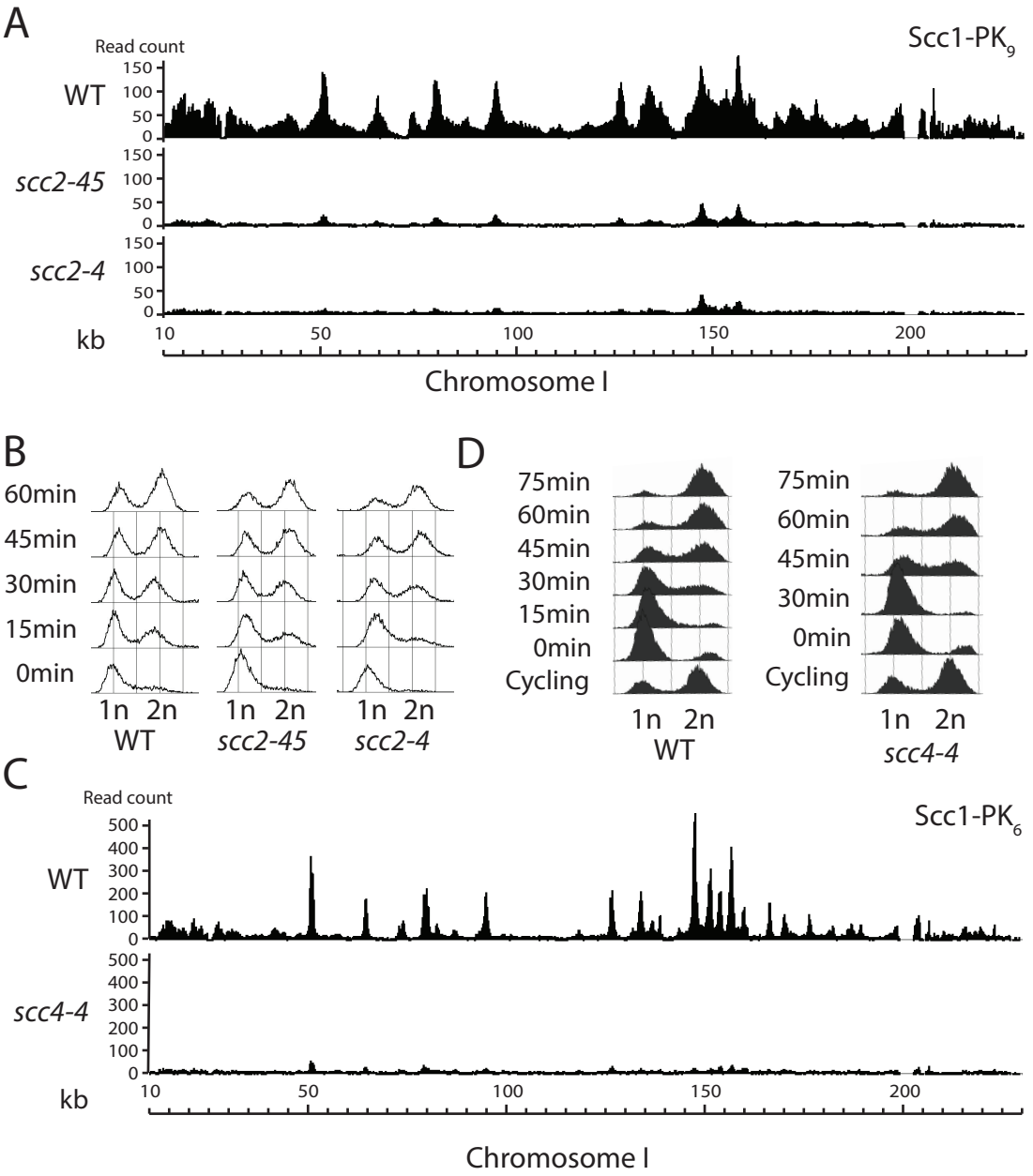


Figure II-3

Cohesin localisation on chromosomes depends on the loading complex.

Cells expressing endogenously $PK_{6/9}$-tagged *SCC1* and either WT or temperature sensitive alleles of the loading complex proteins Scc2 and Scc4 were arrested in alpha-factor and released into media preheated to 37°C. Samples were analysed by quantitative ChIP-seq.

A - Calibrated profile of chromosome I comparing cells expressing *SCC1-PK9* in the presence of WT *SCC2*, *scc2-45* or *scc2-4* at 60 min after release from alpha-factor into nocodazole. (K14601, K22390 and K14710, respectively)

B - Cell cycle progression during the experiment shown in A as measured by FACS.

C - Calibrated profile of chromosome I comparing cells expressing *SCC1-PK6* in the presence of WT *SCC4* or *scc4-4* at 75 min after release from alpha factor into nocodazole. (K22005 and K22001, respectively).

D - Cell cycle progression during the experiment shown in C as measured by FACS.

the loading complex proteins had drastically reduced cohesin loading compared to WT throughout the genome (Figure II-3A and 3C) implying that cohesin observed in WT cells is loaded cohesin. However, in the absence of Scc2/Scc4 function, some occupancy could still be seen in the vicinity of the centromere and at regions of cohesin enrichment on the chromosome arms albeit not as much as in WT. As it is unknown if the activity of the loading complex is completely abolished at the restrictive temperature, it is difficult to determine if these small peaks are some small amount of loaded cohesin. It is unlikely though that these peaks correspond to completely soluble cohesin, as different mutant alleles of Scc2 have different occupancies, which can be observed on the arms of chromosome I (Figure II-3A), suggesting it may be a reflection of incomplete inactivation of Scc2.

In conclusion, the vast majority of the cohesin associated with DNA as observed by quantitative ChIP-seq does depend on the loading complex and therefore represents loaded cohesin.

The cohesin-loading complex primarily associates with centromeres.

Analyses of ChIP-seq profiles of Scc2 have led to the suggestion that arm cohesin is loaded at transcriptional start sites of various genes, including ribosomal protein genes and genes that are characterised by broad and shallow nucleosome-free regions (Lopez-Serra, Kelly, Patel, Stewart, & Uhlmann, 2014). It has also been proposed that cohesin loaded at these sites subsequently slides away or is possibly pushed by RNA polymerase causing cohesin to accumulate at sites of convergent transcription (Lengronne et al., 2004).

Enrichment of cohesin along the arms and around centromeres could be due to a property of these sites that inhibits cohesin sliding or the loss of a driving force such as RNA polymerase causing cohesin to accumulate. Alternatively, these sites could be loading sites if sliding is rate limiting or simply doesn't occur therefore causing cohesin rings to accumulate at loading sites. To determine which of these models is correct, it is crucial to determine where cohesin is loaded relative to sites of cohesin enrichment. To address this question, the loading complex proteins themselves were PK-tagged and analysed by quantitative ChIP-seq.

Quantitative ChIP-seq profiles obtained from yeast strains containing PK_6-tagged Scc2, grown either in asynchronous logarithmic phase cultures or arrested in G_1 by pheromone, were compared to that of an untagged control (Figure II-4A and 4B). This revealed that the levels of Scc2 on DNA are much lower than that of cohesin (note the lower read count). This could be due to lower levels of Scc2 occupancy than cohesin occupancy. However, as Scc2, unlike cohesin, does not stably associate with chromatin (Hu et al., 2011) and turns over very rapidly, it is likely that it is poorly cross-linked to chromatin by formaldehyde causing lower levels to be observed.

Scc2 associates predominantly at the centromere with some protein occupying the region within 1 kb of the centromere. Although there may be some Scc2 association throughout the chromosome arms, it is difficult to distinguish peaks from background signal due to the low overall signal. This is contrary to previously published experiments that measure Scc2 peaks on chromosome arms (Kogut, Wang, Guacci, Mistry, & Megee, 2009; Lopez-Serra et al., 2014).

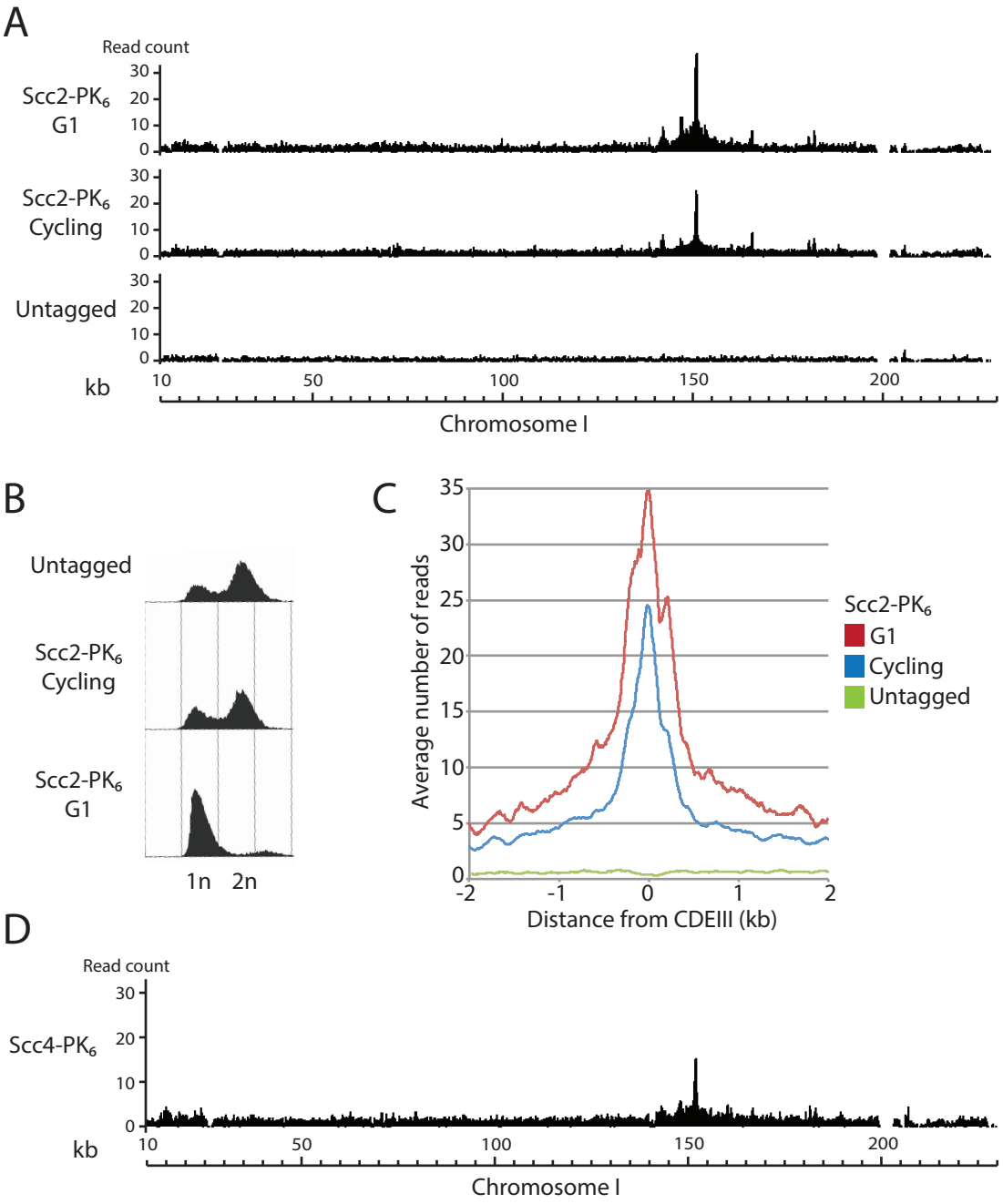


Figure II-4

The cohesin loading complex primarily associates with core centromeres.

Cells expressing endogenously PK_6-tagged *SCC2* or *SCC4* were analysed by quantitative ChIP-seq.

A - Calibrated profile of chromosome I comparing cells expressing endogenous *SCC2-PK$_6$* arrested in alpha factor to cycling cells and an untagged control (K21388 and K699, respectively).

B - Cell cycle state of samples in A as measured by FACS.

C - Calibrated average centromere profile of the experiment shown in A.

D - Calibrated profile of chromosome I of cycling cells expressing endogenous *SCC4-PK$_6$* (K22827).

However, these previously published experiments were not calibrated and therefore are not able to distinguish signal from noise.

A profile of the averaged centromeres reveals a defined peak directly at the CDEIII element (Figure II-4C). This is dissimilar to the localisation of cohesin itself, which occupies two peaks either side of CDEIII, with substantially less, though still a significant amount, actually at the centromere. This could indicate that cohesin is loaded onto DNA at the centromere before sliding away to form the peaks proximal to the centromere, and possibly the pericentromeric peaks as well. Also the occupancy of Scc2 at the centromere seems to be higher in G_1 than in cycling cells. This is also contrary to published results that reported that Scc2 did not associate with DNA in G_1 but at later time points (Fernius et al., 2013). A time course tracking the levels of Scc2 during the cell cycle would reveal the variation in occupancy in more detail, but the data presented herein suggests that Scc2 is present at the centromere in G_1 and cycling cells.

The occupancy of Scc4 in cycling cells was also measured (Figure II-4D, experiment performed by Dr Christophe Chapard) and the localisation of Scc4 shown to be very similar to that of Scc2, which is unsurprising given that the two proteins are known to form a complex.

Loaded cohesin spreads from centromeric to pericentromeric regions during the cell cycle.

Given that the localisation of cohesin is dissimilar to that of the loading complex, at least around the centromeres, it is possible that cohesin translocates away from its initial loading site. If this were the case, newly loaded cohesin would be expected to co-localise with the loading sites before moving to the

described enriched sites. A previous study into the localisation of cohesin during the cell cycle showed a large net increase in cohesin loading 15-30 min after release from pheromone arrest (Hu et al., 2015). Therefore careful examination of the localisation of cohesin around the centromere at these early time points may reveal translocation of newly loaded cohesin.

To examine changes in occupancy of the cohesin complex around the centromere during the cell cycle, data obtained from an experiment performed by Dr Bin Hu was reanalysed. The cells had been released from α-factor-induced G_1 arrest and samples taken at 15 min intervals for 120 min. An average profile was generated for the region 10 kb either side of CDEIII for time points 15, 30, 45 and 60 min, which correspond to the cells going through the first cell cycle (Figure II-5A). In G_1 (15 min) the level of cohesin on DNA is very low, with the pericentromeric peaks just beginning to form. However, a significant amount of cohesin occupancy can be seen in peaks proximal to the centromere. Entrance into S phase around 30 min is accompanied by an approximately twofold increase in the proximal peaks and a three- to fourfold increase at the pericentromeric loci. Interestingly, as the cells reach late S phase at 45 min post-release, the occupancy of cohesin at the centromere itself drops significantly, although the proximal peaks remain high. Furthermore the occupancy in the region between the proximal and the pericentromeric peaks also decreases whereas the pericentromeric peaks themselves seem to increase. When the cells are in G_2, at the 60 min time point, the proximal peaks decrease, again while the pericentromeric peaks increase. One explanation for this could be that cohesin is loaded at the centromere and, once loaded, translocates away to form the pericentromeric peaks. The sharp decrease in cohesin occupancy at the

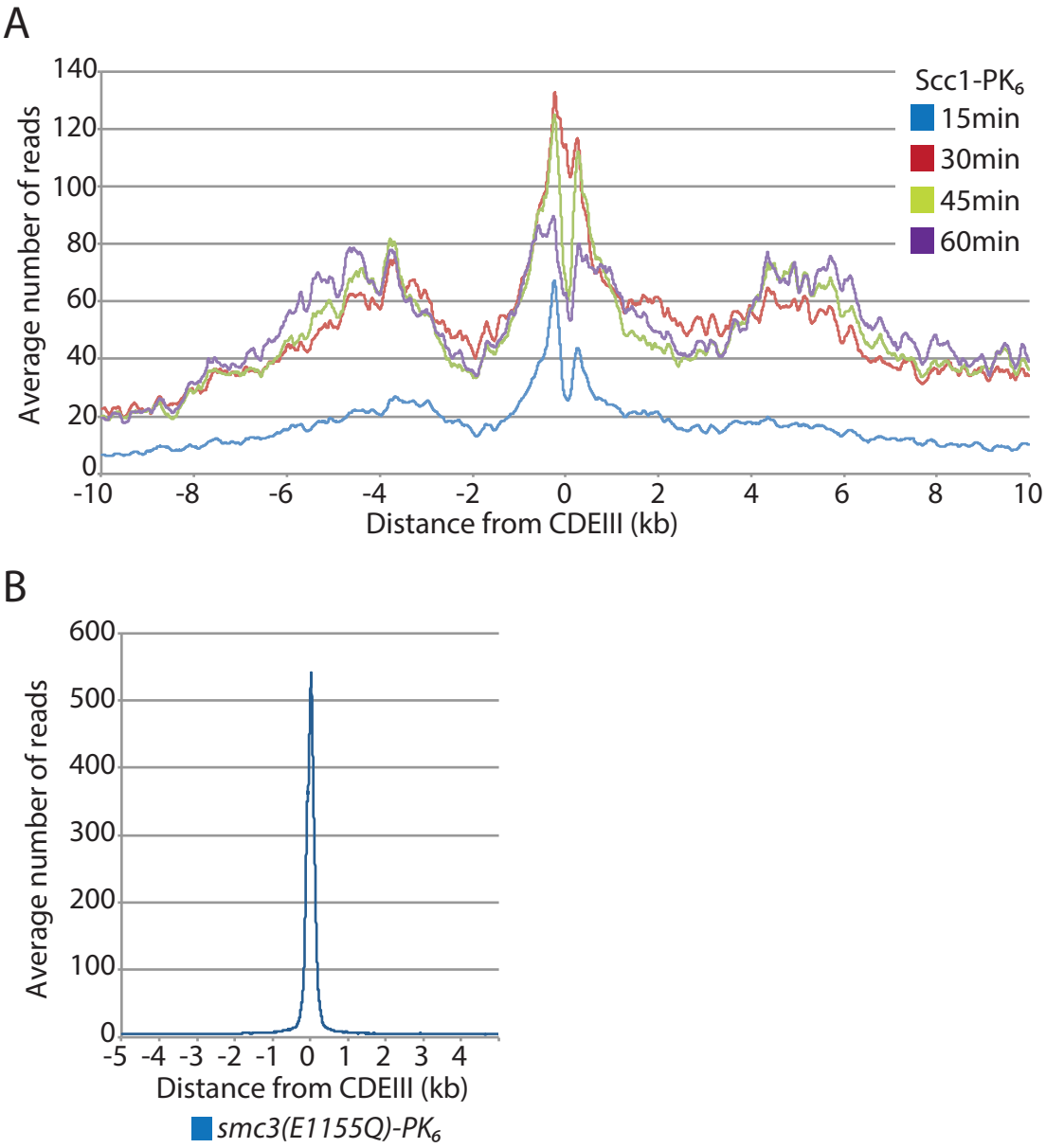


Figure II-5

Cohesin spreads from the centromeric to the pericentromeric regions during the cell cycle in a manner dependant on ATP hydrolysis.

A - Cells expressing endogenous *SCC1-PK$_9$* (K14601) were synchronised in G_1 before release at 25°C with samples taken every 15 min. Calibrated average centromere profile comparing 15, 30, 45 and 60 min after release from alpha-factor.

B - Cycling cells expressing *smc3(E1155Q)-PK$_6$* (K17409) from an ectopic locus were analysed by quantitative ChIP-seq. Calibrated average centromere profile covering 5 kb either side of CDEIII.

centromere in late S phase is consistent with the observation that Scc2 occupancy here is also decreased (Figure II-6). However, whereas Scc2 occupancy increases again in G_2, cohesin occupancy does not. This could be indicative of a change in system as the cell enters G_2, at which point cohesin should be stably associated with DNA and the loading complex is no longer essential (Bernard et al., 2006; Ciosk et al., 2000).

High-resolution fluorescence microscopy showed that cohesin rings that are unable to bind ATP are not recruited to DNA, whereas cohesin complexes that are able to bind ATP but are hydrolysis deficient associate with kinetochores, but cannot stably interact with DNA (Arumugam et al., 2003; Hu et al., 2011; Weitzer et al., 2003). Given that Scc2 also associates with kinetochores, this suggests that ATP hydrolysis is required before translocation can occur. Thus cohesin that is recruited to DNA but cannot complete the loading reaction should occupy the site of its initial loading. To determine if hydrolysis is required for translocation, the effect of the Smc3 hydrolysis mutant E1155Q on cohesin localisation was investigated by quantitative ChIP-seq.

Since hydrolysis mutants in either of the SMC subunits are lethal, an ectopic copy of *SMC3-PK9* containing the ATP hydrolysis defective mutation (*smc3(E1155Q)*) was used. As only the mutant copy of Smc3 was PK9-tagged analysis of the mutant using quantitative ChIP-seq can be performed even in the presence of the WT protein. As expected from previous studies, which demonstrated that the hydrolysis mutant associates with kinetochores (Hu et al., 2011), the Smc3 hydrolysis mutant forms a peak directly at the centromere, with very little signal on chromosome arms (Figure II-5B). This strongly supports the notion that cohesin is loaded at the centromere and translocates away.

Furthermore, this suggests that ATP hydrolysis is required to complete a late step in the loading reaction after association with DNA to permit cohesin translocation. What remains to be understood is what prevents the movement of cohesin further along the chromosome arms and what causes the troughs between the peaks proximal to the centromere and those at the pericentromere. Presumably this is due to other factors on the DNA that may interfere with the ability of cohesin to translocate further; however, the identities of such factors are currently unknown.

Occupancy of the loading complex at centromeres depends on Scc1.

It has been claimed that Scc2 association with centromeres depends on Scc1 (Fernius et al., 2013). Because the amount of cohesin on chromosomes is low in pheromone arrested cells due to low rates of Scc1 synthesis in G_1, persistent separase activity and Wapl-mediated turnover, it would be expected that Scc2 association with centromeres would be very low. This is consistent with the published finding that Scc2 is not recruited to centromeres in G_1 (Fernius et al., 2013). However, quantitative ChIP-seq of Scc2 revealed the opposite, namely that Scc2 is recruited to centromeres in G_1. Given this discrepancy, the question of the requirement of Scc1 for Scc2 localisation was re-examined.

A strain with *SCC1* under a methionine repressible promoter that also expressed *SCC2-PK*$_6$ was generated. Cells were synchronised in G_1 in the presence of methionine to prevent Scc1 synthesis, then released from the arrest into methionine-containing media. Samples were taken at 30, 45 and 60 min

after release from G_1 arrest and analysed by quantitative ChIP-seq. At all time points, Scc2 occupancy at the centromere was greatly reduced in the absence of Scc1 (Figure II-6A) implying that Scc2 localisation at the centromere depends on Scc1. Depletion of Scc1 was verified by Western blot looking at the levels of Scc1-myc_9 at 15 min intervals after release from α-factor compared to an asynchronous culture grown in the absence of methionine (Figure II-6B). The Western blot confirmed the absence of Scc1 at all time points after release from G_1. FACS samples were also taken at 15 min intervals to confirm that both strains progressed through the cell cycle at similar rates (Figure II-6C). During this experiment, cells progressed from G_1 to G_2 (t=60 min). As the cells entered S phase (30 min sample) Scc2 occupancy (in the presence of Scc1) is high but drops towards the end of S phase (45 min) before rising again to a higher level in G_2. This dip in Scc2 levels during S phase may indicate that the occupancy of Scc2 at the centromere is decreased during DNA replication, or it may due to the duplication of the centromeric sequences causing the occupancy to be halved.

In conclusion, Scc2 association does depend on Scc1. However, this observation is not inconsistent with high levels of Scc2 at centromeres in G_1 shown by quantitative ChIP-seq. Small amounts of cohesin are observed in the early stages of the cell cycle despite the persistence of separase and Wapl-mediated turnover. Therefore, it is plausible that cohesin is being loaded by Scc2, but is rapidly removed following loading, preventing its accumulation on DNA, explaining the low levels of Scc1 and high levels of Scc2 observed with quantitative ChIP-seq.

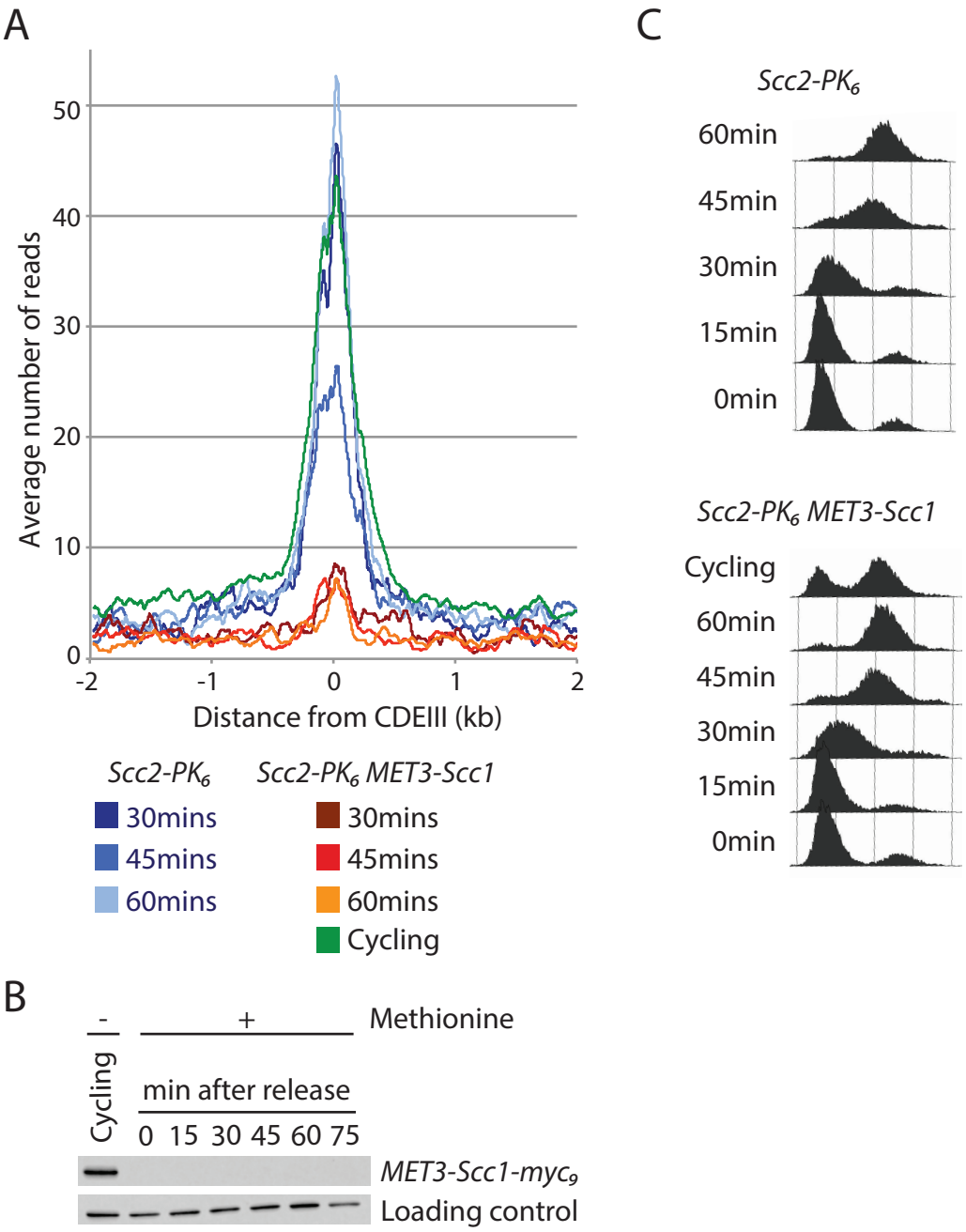


Figure II-6

Association of the cohesin loading complex at core centromeres depends on Scc1

Comparison on cells expressing endogenous *SCC2-PK$_6$* and endogenous *SCC1* either under the *MET3* promoter (K25222) or the endogenous Scc1 promoter (K21388). Cells were grown in methionine-containing media containing alpha factor to arrest them in G_1, then released into methionine-containing media at 25°C with samples taken at 30, 45 and 60 min after release.

A - Calibrated average centromere profile.

B - Western blot to show depletion of Scc1 in cells expressing *SCC1* under the *MET* promoter in the presence (in the time points after release) or absence of methionine (in cycling cells).

C - Cell cycle progression during the experiment shown in A as measured by FACS.

Pds5 is excluded from the "pre-loading" complex that forms at centromeres.

By using SMC hydrolysis mutants to prevent completion of the loading reaction it becomes possible to determine which other proteins are present at an early stage of the loading reaction, wherein cohesin rings have formed a "pre-loading" complex at centromeres but are unable to undergo the following steps. Proteins that are present during the loading reaction should be bound to cohesin in the "pre-loading" complex and thus also form a distinctive peak at centromeres. By introducing various PK-tagged proteins into strains with either *smc3(E1155Q)* or *SMC3* at an ectopic locus, a mixed profile should be observed. Having an extra ectopic WT copy of Smc3 should produce an occupancy profile similar to WT. However, with an ectopic E1155Q (EQ) copy of Smc3 there will be a mixture of WT and mutant reads. If the tagged protein forms a complex with the EQ copy of Smc3, this should lead to a WT profile in combination with the increased binding at the centromere that comes from the EQ profile. If the protein does not form a complex with the EQ mutant, this centromeric peak would be absent and only the WT profile would be observed.

To validate this approach, the localisation of Scc1 at centromeres in the presence and absence of the Smc3 EQ mutant was analysed. Cycling cells expressing *SCC1-PK*$_6$ in the presence of an ectopic copy of WT *SMC3* or the Smc3 hydrolysis mutant (E1155Q) were analysed by quantitative ChIP-seq (Figure II-7A). In the presence of the extra WT copy of Smc3, the profile is similar to that of WT Scc1 (blue trace, Figure II-2B), with peaks proximal to the centromere and at the pericentromere. However, as expected, in the presence of the EQ mutant copy of Smc3, the trough between the two peaks proximal to the centromere is

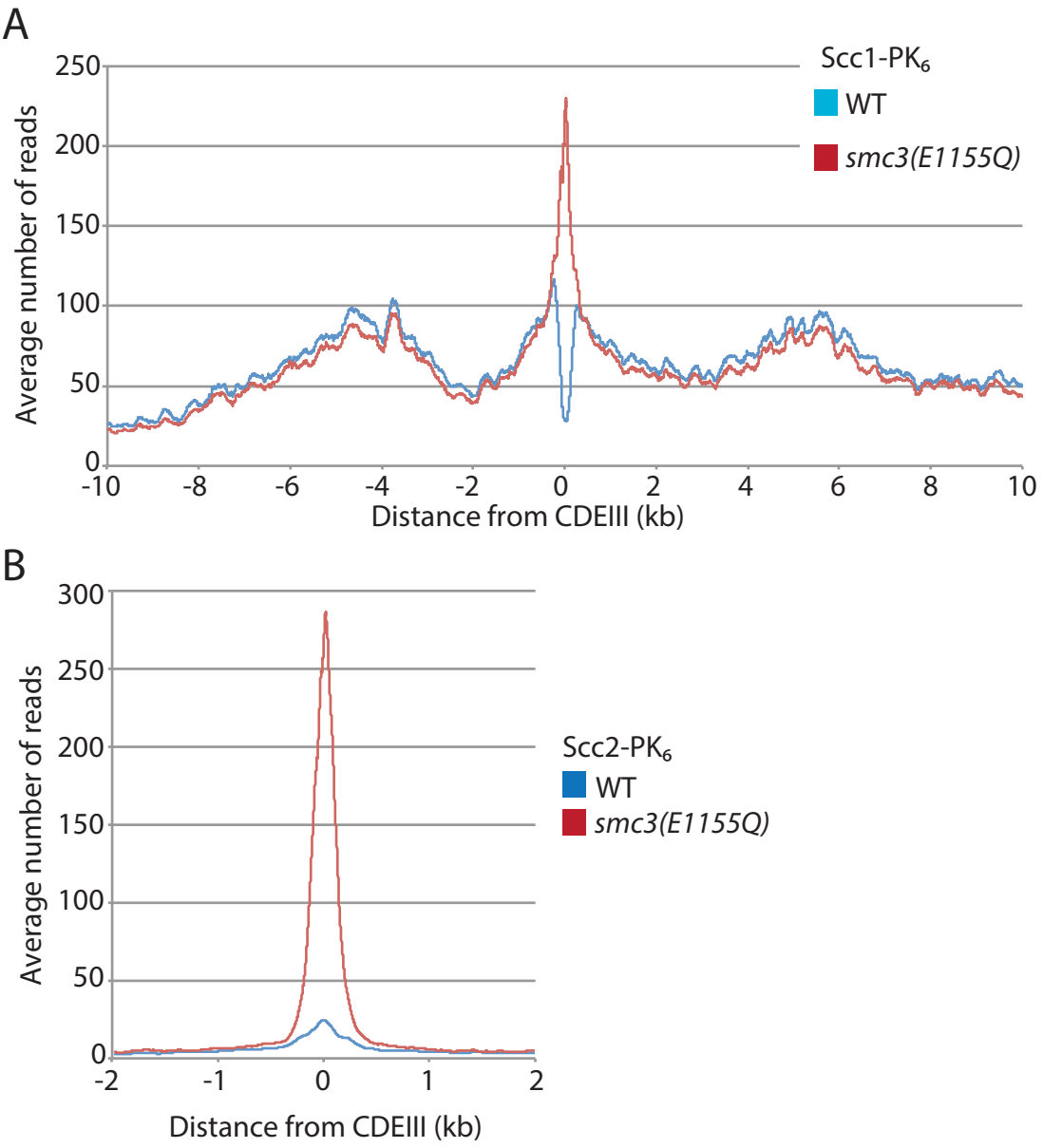


Figure II-7

Scc1 and Scc2 are part of a 'pre-loading' complex that forms at centromeres prior to ATP hydrolysis.

Cycling cells expressing either *SCC1-PK$_6$* or *SCC2-PK$_6$* in the presence of an extra ectopic copy of Smc3 that is either WT or contains a mutation that prevents ATP hydrolysis (E1155Q) were analysed by quantitative ChIP-seq.

A - Calibrated average centromere profile comparing Scc1-PK$_6$ in the presence or absence of *smc3(E1155Q)* (K22005 and K25370, respectively)

B - Calibrated average centromere profile comparing Scc2-PK$_6$ in the presence or absence of *smc3(E1155Q)* (K21388 and K25467, respectively)

replaced by a large peak which corresponds to the profile expected from cohesin complexes that include *smc3(E1155Q)*.

This approach was repeated for *SCC2-PK$_6$* (Figure II-7B). Since the presence of Scc2 at centromeres depends on Scc1 (Figure II-5A) and *vice versa* (Figure II-3A), both proteins be should present with the EQ copy of Smc3. If Scc2 forms a complex with cohesin, then it would be expected that the levels of Scc2 at the centromere should be increased since there is much more EQ cohesin at centromeres than WT cohesin. Cycling cells in the presence of an extra copy of WT Smc3 showed the expected small peak at the centromere that was previously observed (Figure II-4C). In the presence of *smc3(E1155Q)*, the peak is over ten times bigger. Crucially, this implies that Scc2 forms a complex with rings containing the EQ mutant and does not just co-localise with cohesin on the chromosome. Thus both Scc1 and Scc2 are present in the "pre-loading" complex that forms at centromeres before the completion of loading and subsequent translocation, but what about the other cohesin regulatory proteins Scc3 and Pds5?

Cycling cells expressing *SCC3-PK$_6$* in the presence of an ectopic copy of either WT *SMC3* or *smc3(E1155Q)* were analysed by quantitative ChIP-seq in the same manner (Figure II-8A). This revealed that, in the presence of the EQ mutant, a peak appears at the centromere. Contrary to the observation with *SCC1-PK$_6$*, this peak is relatively small, comparable with those seen proximal to the centromere in the WT strain. This suggests that Scc3 is present in the preloading complex but not to the same extent as Scc1 or Scc2, perhaps suggesting that, during ATP hydrolysis, cohesin enters a state that destabilises

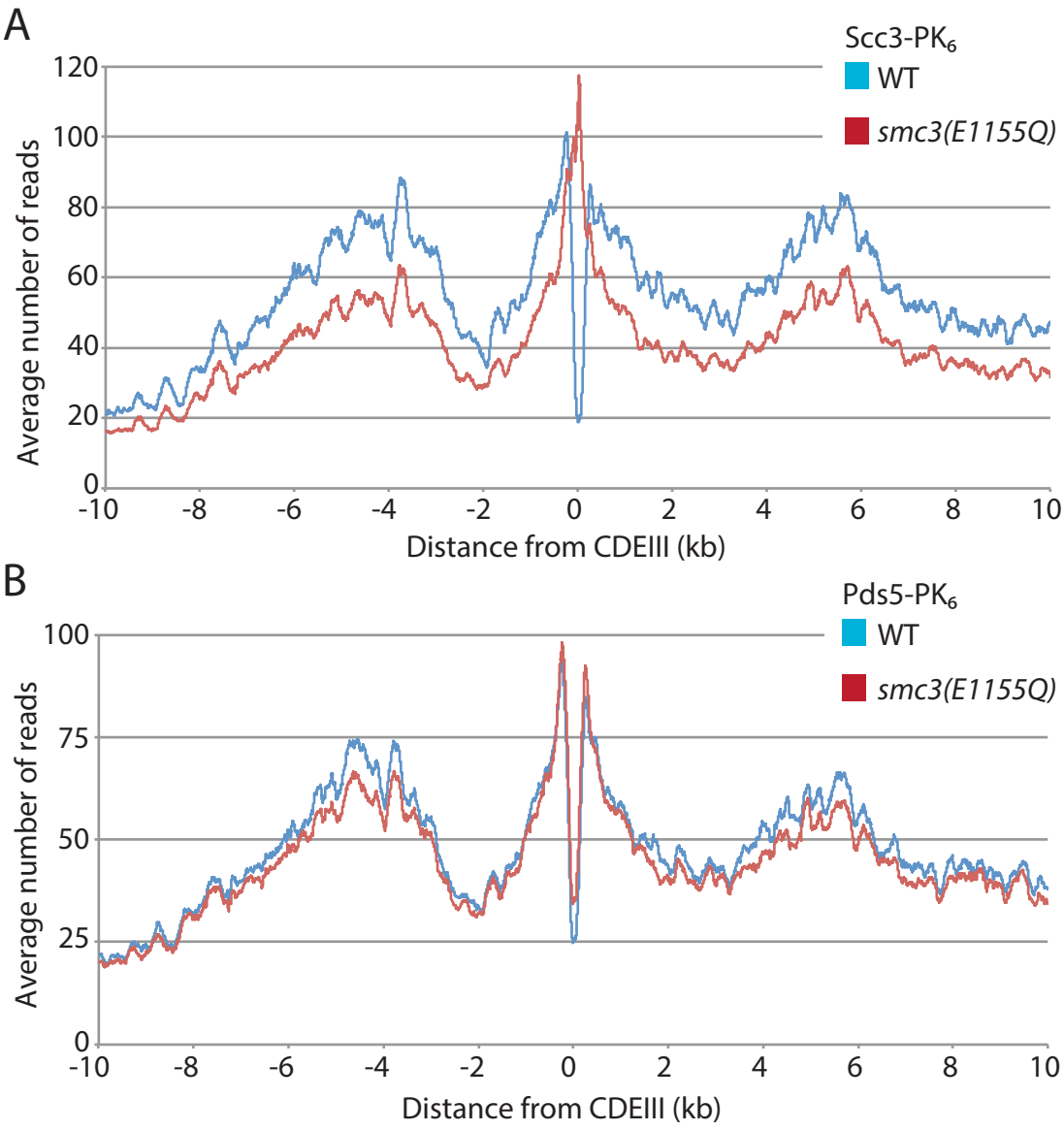


Figure II-8

Pds5 is excluded from the 'pre-loading' complex that forms at centromeres prior to ATP hydrolysis.

Cycling cells expressing either *SCC3-PK$_6$* or *PDS5-PK$_6$* in the presence of an extra ectopic copy of *SMC3* that is either WT or contains a mutation that prevents ATP hydrolysis (E1155Q) were analysed by quantitative ChIP-seq.

A - Calibrated average centromere profile comparing Scc3-PK$_6$ in the presence or absence of *smc3(E1155Q)* (K17438 and K25373, respectively).

B - Calibrated average centromere profile comparing Pds5-PK$_6$ in the presence or absence of *smc3(E1155Q)* (K19012 and K25376, respectively).

Scc3 binding. It should also be noted that the overall occupancy of Scc3 on DNA in the presence of *smc3(E1155Q)* is reduced compared to *SMC3*.

In contrast, cells expressing *PDS5-PK*$_6$ showed a similar profile both in the presence of ectopic WT *SMC3* or *smc3(E1155Q)* (Figure II-8B). The lack of a peak at the centromere indicates that Pds5 is excluded from the "pre-loading" complex. Pds5 and Scc2 share structural similarities in that they are both large HEAT-repeat containing proteins that form a hook-like shape. It has been shown that Scc2 binds the Scc1 in the vicinity of the Pds5 binding site and that they compete for binding to Scc1 (Kikuchi et al., 2016). Thus the presence of Scc2 in the pre-loading complex may prevent the binding of Pds5. Therefore, this would imply that Pds5 could only bind to the complex once Scc2 has been displaced. Pds5 has three roles, firstly in stabilising cohesin on DNA by promoting acetylation of Smc3 and preventing de-acetylation, secondly in promoting release of DNA from cohesin's topological embrace through Wapl binding, and finally an undefined role in maintaining cohesion during G_2. Thus the presumed later association of Pds5 with the cohesin complex might signify the progression from loading to a maintenance state, in which Pds5 acts to either keep cohesin on DNA or allow its controlled release where required.

Cohesin accumulates around transcription end sites during the cell cycle.

Given the evidence that cohesin is loaded at the centromere and translocates to the pericentromere, it is plausible that cohesin loaded on chromosome arms also translocates towards sites of enrichment. Published studies have indicated that transcription may play a role in cohesin localisation

on chromosome arms (Lopez-Serra et al., 2014). At sites with active transcription, cohesin could be pushed along the actively transcribed genes to form peaks at termination sites, and especially between convergent genes. If this is the case, it should be possible to visualise the accumulation of cohesin at termination sites as the cell progresses through the cell cycle.

To produce profiles of cohesin localisation around transcription start and end sites, genes longer than 2 kb were aligned by their annotated transcription start sites (TSS) and transcription end sites (TES) as documented by (Z. Xu et al., 2009). The calibrated profiles were then averaged covering 1 kb either side of the TSS or TES. This was performed for each time point obtained from the 120 min G_1-release experiment performed by Dr Bin Hu (Figure II-6A). At 15 min, very little cohesin is associated around the TSS with a slight dip in the profile approximately 100 bp upstream of the TSS, corresponding to the promoter region (Figure II-9A). As the cells progress through the cell cycle, cohesin occupancy increases in the gene body. However, at the promoter region although the occupancy does increase, the relative drop in occupancy becomes increasingly prominent. Towards the end of the cell cycle (60-90 min), cohesin occupancy begins to drop as separase is activated, before increasing again at the beginning of the next cell cycle (105 min). This suggests that cohesin is generally excluded from gene promoter regions. Similarly, around TESs cohesin also accumulates in line with cell cycle state but, rather strikingly, forms two peaks flanking the TES (Figure II-9B). It should be noted that the occupancy of cohesin around the TES is approximately double the occupancy at the TSS at time points at which cohesin occupancy is at its maximum (e.g. 120 min).

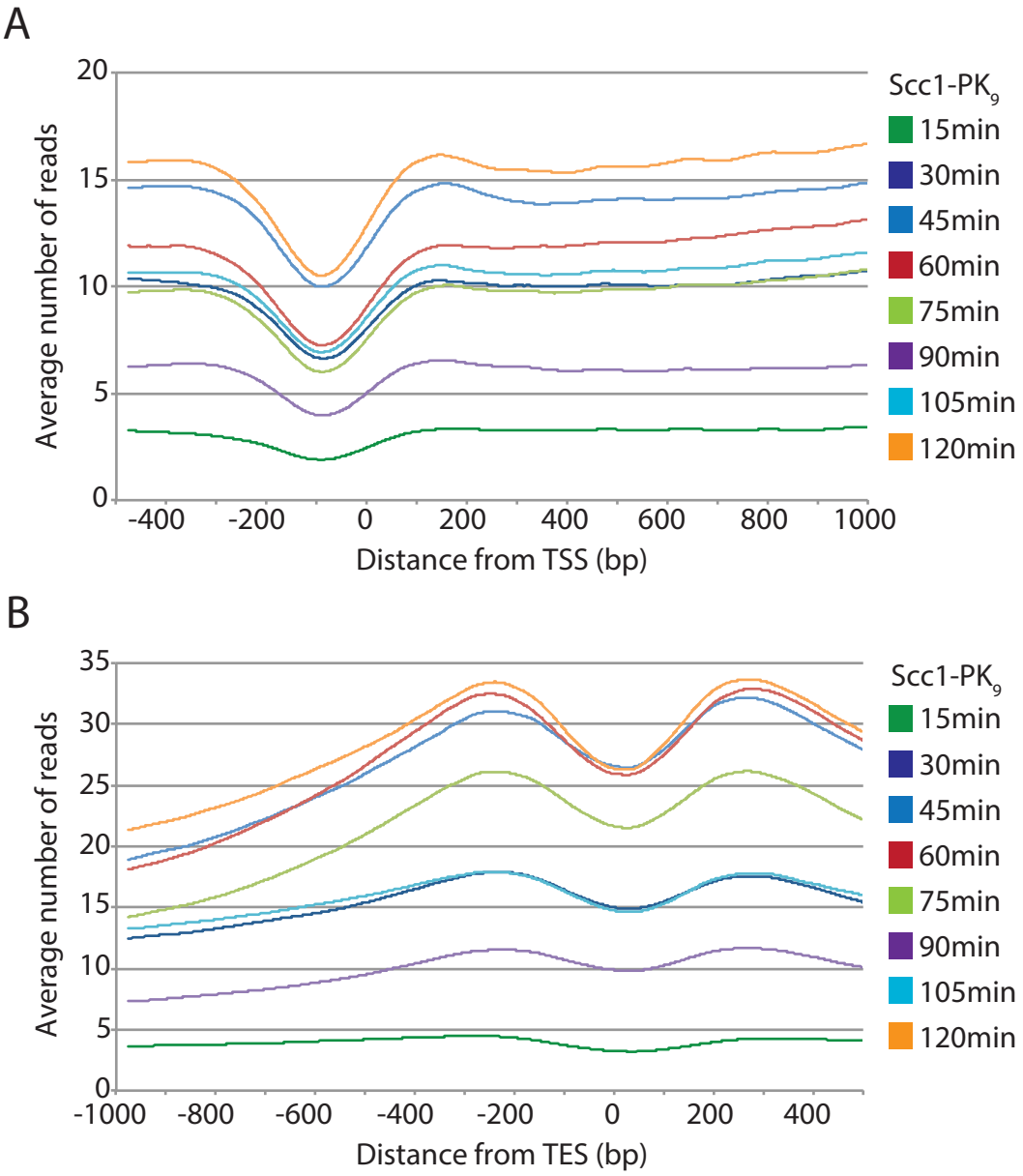


Figure II-9

Cohesin accummulates around TES sites during the cell cycle.

Cells expressing *SCC1-PK9* were synchronised in G_1 before release at 25°C with samples taken every 15 min as in figure II-5.

A - Calibrated average profile around the TSS sites of genes longer than 2 kb.

B - Calibrated average profile around the TES sites of genes longer than 2 kb.

Low levels of cohesin in genes at early time points indicate that cohesin is initially found throughout the gene. Cohesin then accumulates towards the 3' ends of all genes (not just convergent genes (Filipski & Mucha, 2002; Glynn et al., 2004; Lengronne et al., 2004)) as the cell progresses through the cell cycle, a previously undocumented phenomenon. This is consistent with the model of cohesin being pushed by the transcription machinery towards the TES. However, at the TES itself cohesin occupancy appears to decrease. The TES is a nucleosome-free region, perhaps this is suggestive of a mechanism that keeps this region free of obstructions, which goes some way to preventing accumulation at this site but does not halt translocation completely, as a significant amount must be able to pass through in order to form the second peak after the TES. Then, since transcription has terminated, perhaps the absence of a driving force past the TES causes the accumulation of cohesin after the TES rather than its continued translocation across the chromosome. This differs from the current hypothesis that suggests that convergent transcription causes cohesin accumulation in the regions between convergent genes due to the opposing movement of the polymerases (Glynn et al., 2004; Lengronne et al., 2004). Although this may be the case, it is important to note that cohesin enrichment occurs at the end of all genes and therefore convergent transcription is not necessary for cohesin enrichment to occur.

Scc2 is excluded from TSS and TES but is enriched at tRNAs.

To understand cohesin loading and its subsequent translocation, it is necessary to know where cohesin is loaded first on chromosome arms.

Quantitative ChIP-seq has shown that Scc2 primarily localises at the centromere but distinct peaks on the arms are difficult to distinguish from background noise. To overcome this difficulty, different types of chromosomal loci, in particular transcription start sites and transcription end sites were grouped, aligned and Scc2 occupancy was examined. An untagged control was again included to control for background signal. Average profiles around the TSS and TES of genes greater than 2 kb (as detailed earlier) showed Scc2 above background signal throughout the gene bodies with decreased levels at the promoter region and around the TES, the decrease at the promoter region being slightly more pronounced (Figure II-10A and 10B). This trend is also visible in the untagged control albeit to a lesser extent, indicating an inherent bias. This cannot, however, completely explain the more pronounced decrease seen in the tagged sample. Overall this suggests that cohesin loading takes place throughout gene bodies as well as in the flanking regions.

Scc2 has been reported to be enriched at tRNA genes and ribosomal protein genes (Lopez-Serra et al., 2014); therefore, these groups were also included in the analysis. The average profile around the start of tRNA genes and ribosomal protein genes revealed that at both sets of genes the level of Scc2 was greater than the average gene. (Figure II-10C and 10D). The signal of Scc2 occupancy is higher than the background signal in all of these cases. This indicates that, like Scc1, Scc2 is found throughout all genes but is enriched at the 5' end of tRNAs and more modestly at ribosomal protein genes. Therefore the data suggests that loading can occur everywhere in the genome but has some preference for specific sites, most notably at the centromere. This is contrary to a previous study that described specific sites of cohesin loading (Lengronne et

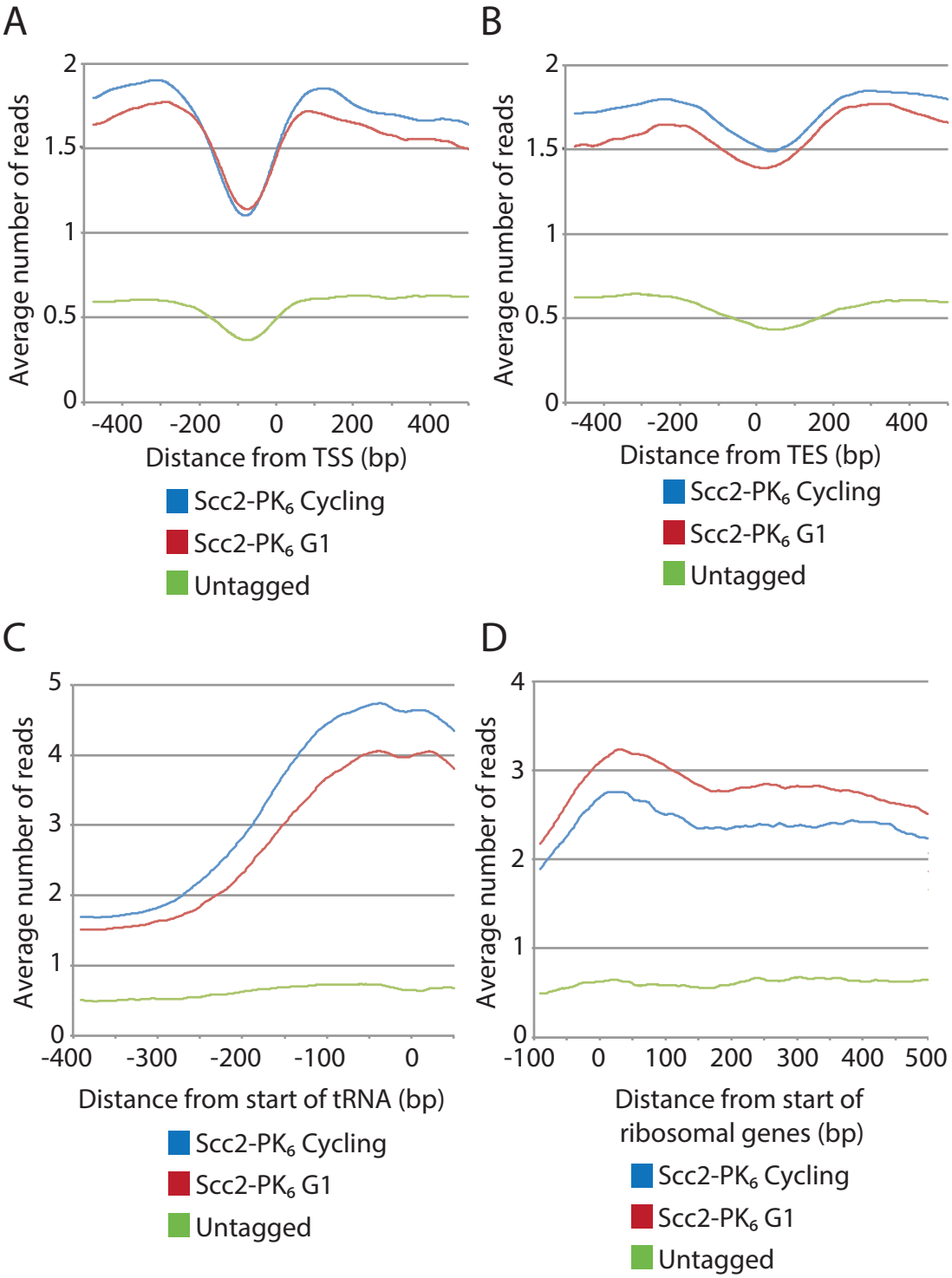


Figure II-10

Scc2 is excluded from TSS and TES sites but enriched at tRNAs.

Cells expressing *SCC2-PK6* as in figure 3.

A - Calibrated average profile of Scc2 localiastion around the TSS site of genes longer then 2 kb.

B - Calibrated average profile of Scc2 localiastion around the TES site of genes longer then 2 kb.

C - Calibrated average profile of Scc2 localisation around tRNAs.

D - Calibrated average profiles of Scc2 localisation around ribosomal genes.

al., 2004). However, their study was performed with conventional ChIP-seq, which is unable to reliably distinguish signal from noise.

Discussion

Despite many advances in the cohesin field, the cohesin loading reaction remains poorly understood. A popular technique used to probe the loading reaction is ChIP-sequencing but this technique has several limitations. One of these limitations is that the method is not quantitative and therefore different samples cannot be directly compared. To address this, a new method of quantitative ChIP-sequencing was developed. This method involves spiking each sample of the experimental cells with calibration cells that are in a defined and reproducible state. The reads obtained from the calibration cells are then used to calibrate the experimental profile to obtain a quantitative profile. This chapter describes the use of this new technique to validate what is currently known about the loading reaction and to potentially reveal novel observations.

It was observed that cohesin associates with DNA genome-wide at levels that are above background signal, and that it is enriched at specific sites along chromosome arms and at the centromere. Although enriched sites have been well documented previously (Blat & Kleckner, 1999; Glynn et al., 2004; Lengronne et al., 2004), the finding that the signal of cohesin occupancy is above background signal even between previously documented sites of enrichment implies that a significant proportion of cohesin occupancy has been largely overlooked. The finding that cohesin occupancy between peaks of enrichment also depends on the presence of the loading complex Scc2/Scc4 indicates that cohesin occupancy even in the troughs, is due to loaded cohesin complexes on DNA and may be as functionally significant as the observed peaks.

Like cohesin, Scc2 was observed to be present throughout the genome at levels greater than background; however, specific sites of enrichment across the chromosome arms were difficult to identify. This is in contrast to many studies that have documented specific Scc2 peaks on chromosome arms (Fernius et al., 2013; Kogut et al., 2009; Lengronne et al., 2004). Given that these studies did not use a quantitative approach it would be pertinent to determine if their findings could be observed above background level. Nevertheless the observation that Scc2 is enriched at tRNA and ribosomal genes could be verified by quantitative ChIP-seq, showing that Scc2 does not associate uniformly across the chromosome arms. What drives the enrichment of at these sites remains to be identified. The loading complex was also shown to predominantly localise to the centromere. This is a well-characterised property of the loading complex that has been reported to depend on kinetochore proteins and Scc4 (Blat & Kleckner, 1999; Eckert et al., 2007; Fernius & Marston, 2009; Hinshaw et al., 2015).

The level of Scc2 occupancy was also shown to change during the cell cycle. Previously published experiments indicated that Scc2 association with DNA is dependent on Scc1 and that Scc2 does not associate with DNA in G_1 (Fernius et al., 2013). It was claimed that this absence of Scc2 association in G_1 was therefore due to low levels of Scc1, due to the continued presence of separase and low levels of Scc1 transcription. In the present study, however, quantitative ChIP-seq showed that Scc2 is present at centromeres in G_1. Further experiments verified that Scc2 occupancy at centromeres does depend on Scc1. Therefore Scc1 must also be present with Scc2 in G_1 so it is likely that the Scc2 observed at centromeres in G1 reflects cohesin loading intermediates. However, the low levels of Scc1 occupancy on DNA in G_1 suggests that loaded complexes

are rapidly removed from DNA due to high rates of cohesin release mediated by both Wapl and separase.

Quantitative ChIP-seq profiles detailing the changes in cohesin occupancy during the cell cycle show that cohesin first accumulates at peaks immediately either side of the centromere, with those further away at the pericentromere forming later. Furthermore, during the early time points when the bulk of cohesin is being loaded, there are also high levels of cohesin at the CENs themselves. The findings that the loading complex associates with the centromere and that cohesin rings that are unable to complete the loading reaction (i.e. the Smc3 hydrolysis mutant E1155Q) also associate with centromeres, are consistent with cohesin being loaded at the centromere and very quickly translocating away from its loading site to sites proximal to the centromere and later, the pericentromere. The idea that cohesin is able to translocate along chromosome arms has been put forward in numerous papers but has not been formally proven. Although the data presented here also falls short of proving the existence of cohesin translocation, it is consistent with this concept and may provide the clearest evidence so far. It therefore raises the possibility that other SMC complexes have a similar property that may help to explain the mechanism of their function.

The utilisation of a mutation in Smc3 that severely impairs ATP hydrolysis (*smc3(E1155Q)*) enabled the visualisation of a transient state in the loading reaction. As this state can be seen in greater abundance and less transiently than Scc2, it can be used to test which proteins are present during the loading reaction. This is because the signal from those proteins in the act of loading can now be distinguished from those that have completed loading. This

revealed a complete exclusion of Pds5 from the pre-loading complex and much less Scc3 present than would be expected. Given the structural similarity of Scc2 and Pds5, the exclusion of Pds5 could be explained by Scc2 competing with Pds5 for binding to the cohesin ring as previously described (Kikuchi et al., 2016). If this were the case, what would cause the dissociation of Scc2 to enable Pds5 to bind? One possibility is the state of the SMC heads as they go through the ATP hydrolysis cycle. Perhaps, the engaged state of the heads in the presence of ATP induces a conformational change that supports Scc2 binding instead of Pds5. Then potentially, on completion of the hydrolysis cycle, the heads will enter a different conformation that causes Scc2 to dissociate and Pds5 to preferentially bind to the complex. However, as head engagement is also required for release (Beckouët et al., 2010; Elbatsh et al., 2016), a process that requires Pds5 binding (Gandhi et al., 2006; Kueng et al., 2006), it is difficult to envisage how the cell could distinguish between cohesin complexes in their engaged state that are involved in loading and those involved in release. Therefore head engagement *per se* cannot drive the displacement. It has been suggested that more than one state of head engagement exists (Beckouët et al., 2010), which would go some way to explaining this problem but further proof of such states would be required.

The finding that Scc3 is present in the "pre-loading" complex but not to the same extent as Scc1 or Scc2 is also surprising. Scc3 has been reported to be required to complete the loading reaction (Haering et al., 2002) and Scc3 binding is required for EQ mutants to concentrate at centromeres (Hu et al., 2011); however, its role remains unclear. One explanation of its low levels in the "pre-loading" complex, is that Scc3 association may be made unstable by the

formation of an intermediate during the ATP hydrolysis reaction. This would cause Scc3 binding to turn over more rapidly and therefore its occupancy to be reduced.

Altogether, these findings can be formulated into a revised model of the loading reaction wherein the loading complex (Scc2/Scc4) first forms a preloading complex with the cohesin ring (Smc1/Smc3/Scc1) that can associate with DNA upon head engagement in the presence of ATP. Scc2 binding to the cohesin ring in this manner prevents Pds5 binding to Scc1 and promotes cohesin loading onto DNA. Subsequently the loading mechanism occurs in a manner dependent on ATP hydrolysis, which may cause transient destabilisation of Scc3 binding. Only when the loading reaction is complete will Scc2 dissociate enabling Pds5 to bind to the cohesin complex. Perhaps also at this point, cohesin will start to translocate away from the loading site. Many aspects of this model are highly speculative and the binding states described would be extremely transient, but biochemical studies into the interactions between these proteins may provide some evidence of these binding states.

A mutation in the hinge domain of Smc1 bypasses the requirement for Scc4 but is defective in centromeric loading

Introduction

Another method that can be used to gain insights into the mechanism of the loading reaction is forward genetics. Yeast genetics is a very powerful tool that can be used to identify components involved in specific biological processes. To determine factors involved in the cohesin loading reaction, a genetic screen was previously performed to identify suppressors of temperature sensitive alleles of the loading complex proteins. For Scc2, *scc2-4* was used but no extragenic suppressors were isolated (Dr M. Demidova, personal communication). For Scc4, twelve independent suppressors of the *scc4-4* allele were isolated and genetically linked to *SMC1* (Dr M. Demidova, personal communication). The mutation was identified as *smc1(D588Y)* and was also independently isolated from similar screens using the temperature sensitive alleles *scc4-2* and *scc4-6* (K. Kozyrska, personal communication). The mutation was confirmed by reconsitituting it *de novo* and, strikingly, it was demonstrated to rescue not only the temperature sensitivity of *scc4-4* cells but also the deletion of Scc4, thus it is able to bypass the requirement of Scc4 for viability. However, the mutation was unable to bypass the requirement for Scc2 but did allow cells expressing the temperature sensitive allele *scc2-4* to grow at a higher temperature of 30°C (Dr M. Demidova, personal communication).

Smc1 D588 is a well conserved residue that is situated within the central hinge dimerisation domain of Smc1 (Figure III-1A) (Kurze et al., 2010). Various

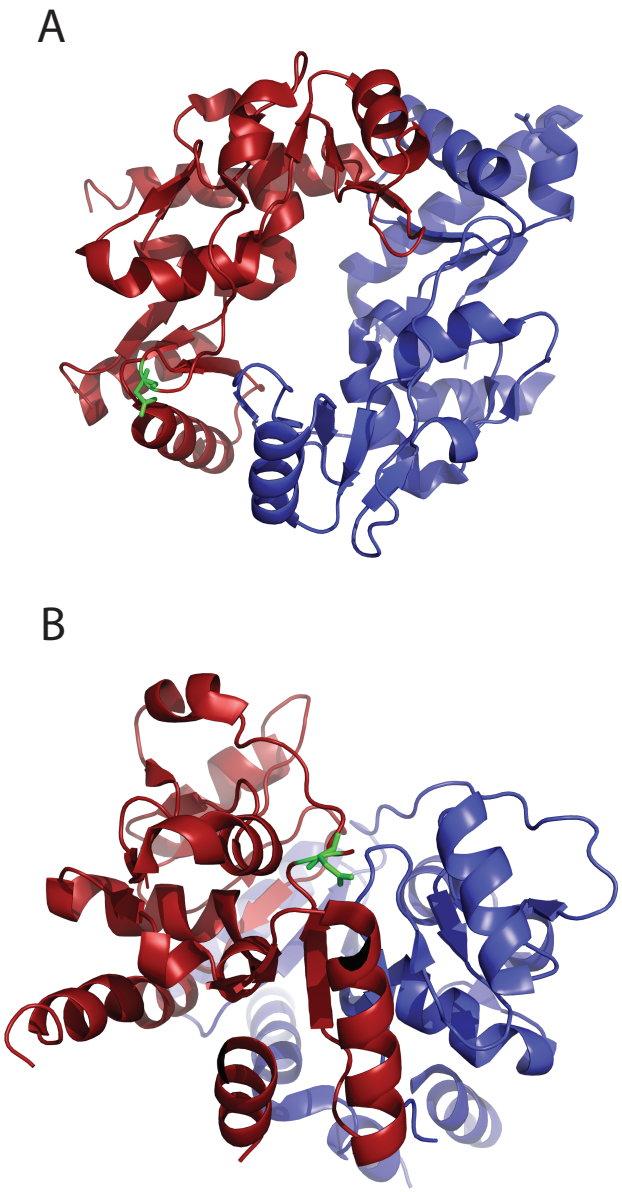


Figure III-1
***smc1(D588Y)* is located in the hinge domain of the coheisn ring.**
Crystal structure of the hinge domain of cohesin from *M. musculus* (PDB 2WD5). Smc1 is coloured in red and Smc3 in blue. Residue D588 in Smc1 is coloured in green.
A - View looking down onto the top of the hinge.
B - Side view of the hinge

substitutions of residue D588 revealed that other uncharged aromatic residues were also able to bypass the requirement for Scc4 but substitutions of the corresponding residues in the Smc3 hinge domain had no effect (Dixon, 2013). Crystalisation of the mouse hinge domain containing the corresponding mutation (*smc1(D574Y)*) revealed that the mutation has no effect on the structure of the hinge in isolation (Dr Byung-Ha Oh and Heekyo Seo, personal communication).

To address why this residue has such a drastic effect on cohesin loading, its effect on hinge dimerisation was assessed by *in vitro* co-immunoprecipitation (Dixon, 2013). This showed that uncharged aromatic residues at position 588 weaken the hinge interface. However, if this weakening alone were to be the mechanism by which *smc1(D588Y)* is able to bypass Scc4 then other mutations that weaken the interface should presumably also have an effect. Several mutations in the hinge domain that weaken the interface were tested but none were able to bypass the requirement of Scc4 (Dixon, 2013). Furthermore, it is striking that, despite the inherent bias towards single base pair mutations, no other mutation within either Smc1 or Smc3 was identified through the screen, again suggesting that the suppression is not due to such a generic effect as weakening the hinge interaction. Thus a biochemical explanation for the suppression of *scc4Δ*-related lethality remains to be identified.

The effect of *smc1(D588Y)* on cohesin loading was further characterised using both fluorescence microscopy and ChIP-seq (Dixon, 2013). This revealed the absence of pericentromeric enrichment of cohesin in cells lacking Scc4. Cohesin association on chromosome arms was unaffected. However, it should be noted that conventional ChIP-seq was used and therefore quantitative changes in

cohesin occupancy could not be described. Scc2 association at centromeres was also abolished in the absence of Scc4. These observations verified that at least one of the roles of Scc4 is to target loading to the centromere, but this is not the essential function. Furthermore, it demonstrated that Scc2 is capable of completing the loading reaction without Scc4 on chromosome arms. This suggests that loading at chromosome arms occurs by a different mechanism to loading at centromeres.

Although ChIP-seq and fluorescent microscopy enabled the visualisation of the distribution of cohesin, it is not able to determine relative differences in terms of the quantity across samples, as a quantitative method was not used. In this chapter the effect of the *smc1(D588Y)* mutation on the loading reaction will be readdressed using quantitative ChIP-seq. This will reveal any changes in the level of occupancy in a quantitative manner which may give additional insights into the mechanism by which *smc1(D588Y)* affects the loading reaction.

Results

smc1(D588Y) suppresses the loading defect of scc4-4 on chromosome arms but not around centromeres

To assess how well *smc1(D588Y)* suppresses the thermosensitive phenotype of *scc4-4* cells, strains expressing *SCC1-PK*$_6$ in the presence of WT *SMC1* or *smc1(D588Y)* and either WT *SCC4* or *scc4-4* were analysed by quantitative ChIP-seq at the restrictive temperature. The strains were synchronised in G_1 and released into media containing nocodazole prewarmed to

37°C. Samples were taken after 75 min, when all strains had completed S phase and were in G_2 (Figure III-2B).

As previously described (Figure II-3C) at the restrictive temperature, *scc4-4* greatly reduced loading throughout the genome (Figure III-2A). *smc1(D588Y)* restores loading in *scc4-4* cells to approximately WT levels on chromosome arms but not around centromeres (Figure III-2A). Interestingly, *smc1(D588Y)* in the presence of Scc4 also appeared to have decreased cohesin occupancy around the centromere. To investigate the effect around the centromere in greater detail, average chromosome profiles were generated covering 10 kb either side of the CDEIII (Figure III-3A). This revealed that *smc1(D588Y),* in the presence of *scc4-4,* has very slightly increased levels around the centromeres when compared to *scc4-4* cells. Strikingly, *smc1(D588Y)* even in the presence of WT Scc4 has decreased cohesin occupancy around centromeres, suggesting that *smc1(D588Y)* causes a defect in loading even in the presence of Scc4. To quantify the changes in cohesin levels with respect to the WT level, an average profile stretching 60 kb either side of the CDEIII was generated for each strain and plotted as a percentage of the reads obtained in the WT sample (Figure III-3B). This showed that *scc4-4* reduced cohesin association along chromosome arms (40-60 kb away from the centromere) to about 20-40% of WT levels and at centromeres to approximately 10%. *smc1(D588Y)* restored cohesin levels on the chromosome arms to 70-100% of WT levels, but the levels proximal to the centromere was only modestly increased to 15-20% of WT. *smc1(D588Y)* in the presence of Scc4 had reduced centromeric loading, at a minimum of 40% of WT levels compared to around 100% on chromosome arms. It should be noted that despite generally reduced levels of cohesin proximal to the CEN compared

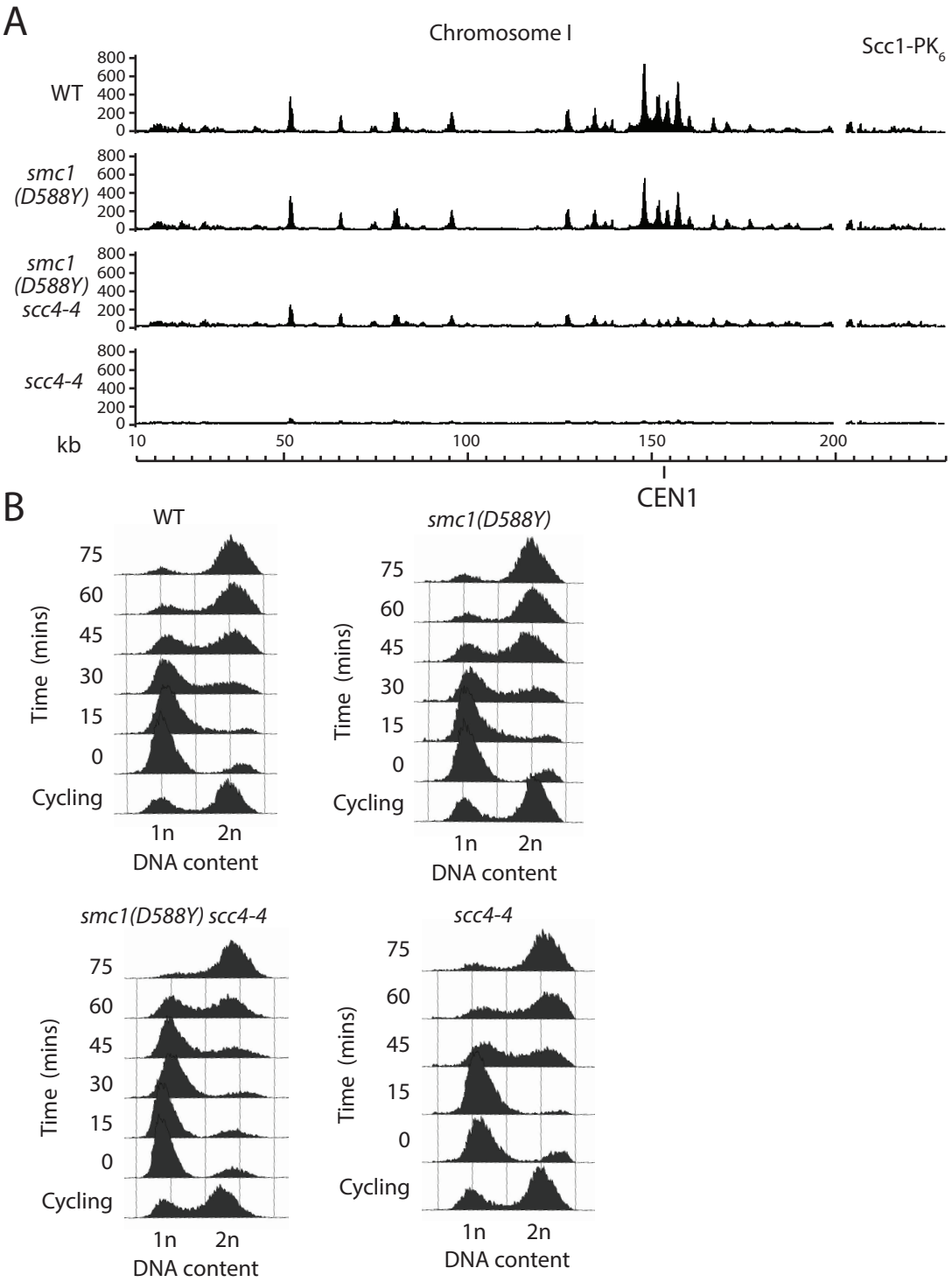


Figure III-2

***smc1(D588Y)* suppresses the loading defect of *scc4-4* at the arms but not around the centromere.**

Strains expressing endogenous *SCC1-PK$_6$* were arrested with alpha factor before release into nocodazole containing medium at 37 °C. Samples were taken at 75 min.

A - Quantitative ChIP-seq profile across chromosome I. The y axis indicates the calibrated number of reads and the x axis, the base pair position across chromosome I. (K22005, K22009, K21999, K22001 from top to bottom)

B - Cell cycle progression during the experiment shown in A as measured by FACS.

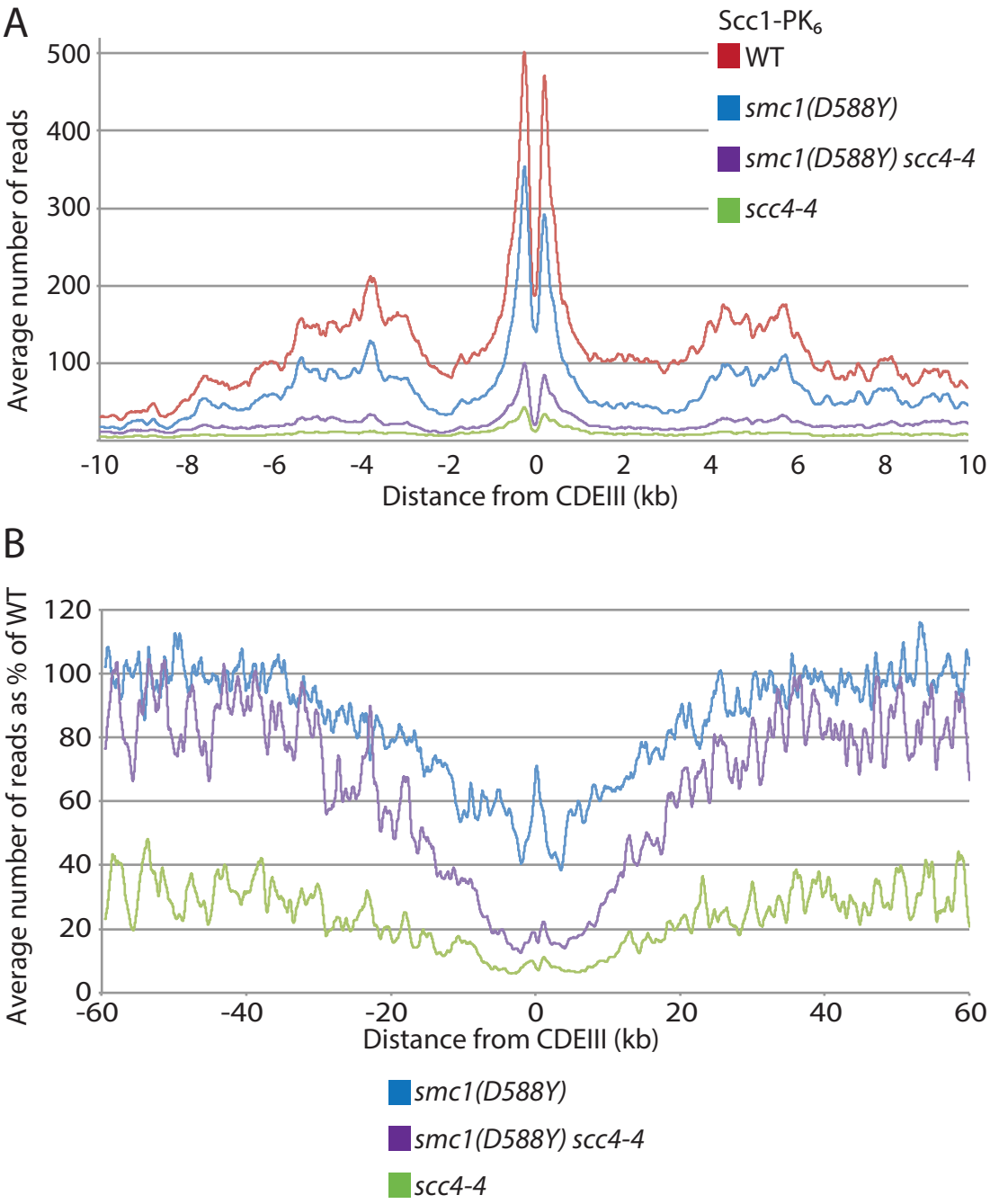


Figure III-3

***smc1(D588Y)* restores cohesin loading in cells containing *scc4-4* on chromosome arms but not around the centromere.**

A - Calibrated average centromere profile of the experiment detailed in figure III-2. B - The number of reads at base pair in the 60 kb region either side of the CDEIII element of each chromosome were averaged over all 16 chromosomes and plotted as a percentage of the number of reads observed in WT. The y axis indicates the calibrated average number of reads as a percentage of WT and the x axis, the position away from the centromere.

to WT, the level at the centromere itself (i.e. at the CDEIII element) is relatively high reaching a

maximum of 75% of WT. This higher occupancy exactly at the centromere was not observed in strains containing *scc4-4*, indicating that it depends on Scc4 function. This is consistent with the requirement for Scc4 in cohesin loading at the centromere. Futhermore, the analysis shows that the mutation in Smc1 is defective in centromeric loading even in the presence of functional Scc4. However, the high levels directly at the centromere may suggest that recruitment of cohesin to the centromere may not be the defective step in the loading reaction.

To investigate the effect of the *smc1D588Y* mutation at more physiological temperatures and in cells that completely lack Scc4, a second experiment was performed to assess cohesin occupancy in cells that have a deletion of the *SCC4* gene (*scc4Δ*). The strains were again synchronised in G_1 and now released into medium at 25°C containing nocodazole to prevent entry into anaphase. Samples were taken at 60 min after release, at which point all cells had reached G_2 as determined by measuring DNA content using flow cytometry (Figure III-4B).

Chromosomal profiles again showed that *smc1(D588Y)* rescued cohesin loading on the arms but not around the centromere in the presence of Scc4 deletion (Figure III-4A). However, the level of suppression on the arms seemed to be better at 25°C with the chromosome arm profile looking very similar to WT levels. Strikingly, cohesin occupancy on chromosome arms in *smc1(D588Y)* cells with WT Scc4 appeared to be higher than WT levels, but still reduced at centromeres. Again, to better visualise the effect at the centromere, an average profile was generated (Figure III-5A). This showed that cohesin levels are

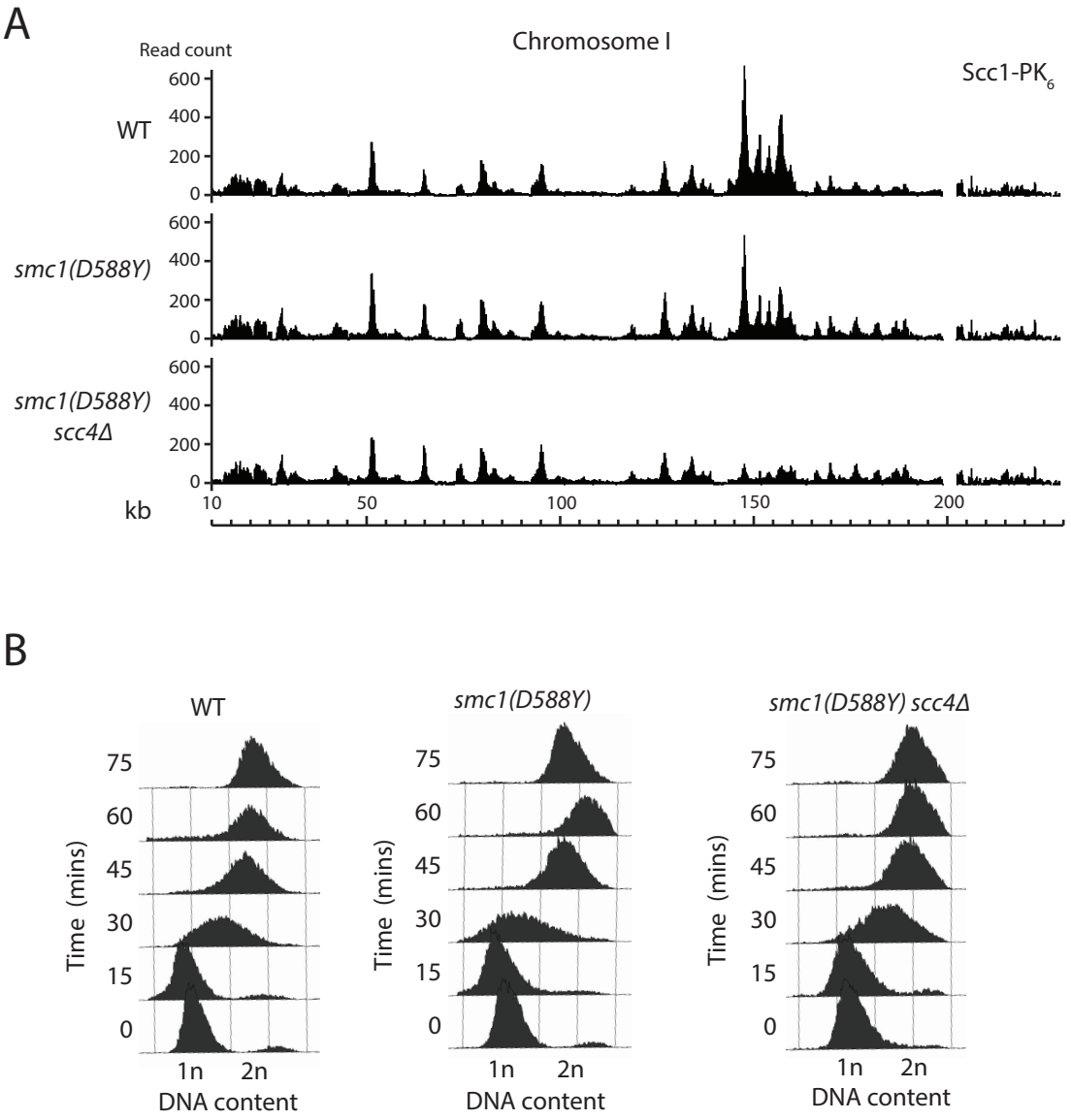


Figure III-4

***smc1(D588Y)* restores cohesin loading along chromosome arms but not around centromeres in *scc4Δ* cells.**

Cells expressing *SCC1-PK*$_6$ were synchronised in G_1 with alpha factor before release into nocodazole containing medium at 25 °C with samples taken at 60 min after release.

A - Quantitative ChIP-seq profile across chromosome I comparing WT cells to cells expressing *smc1(D588Y)* in the presence and absence of Scc4. (K22005, K22009 and K19624 from top to bottom)

B - Cell cycle progression during the experiment shown in A as measured by FACS.

decreased at around the centromere in *smc1(D588Y)* cells and further decreased in the absence of Scc4. The average profile was extended to include some of the chromosome arms by covering 60 kb either side of the CDEIII element and each profile plotted as a percentage of WT levels (Figure III-5B). This revealed that *smc1(D588Y)* in both the presence and absence of Scc4 increases cohesin levels on the arms to about 120-150% of WT cells. However, at centromeres in the presence of Scc4, cohesin levels are reduced to approximately 60% of WT levels. This is further reduced in the absence of Scc4 to approximately 20%. This suggests that *smc1(D588Y)* increases cohesin loading on the arms while decreasing loading at centromeres, both effects occurring in a manner independent of Scc4.

It should also be noted that, as in the previous experiment, the levels of cohesin at the CDEIII element in cells containing *smc1(D588Y)* is higher than the levels around the centromere. Surprisingly, at 25°C, cohesin levels at the centromere even reaches approximately 110% of WT levels. This would suggest that *smc1(D588Y)* is still recruited to the centromere. Therefore the mutation doesn't prevent formation of the preloading complex at centromeres but is defective in a downstream step in the loading reaction, preventing completion required for subsequent translocation away from the centromere.

A simple explanation for how *smc1(D588Y)* is able to increase loading on chromosomes arms in the absence of Scc4 is that, due to the defect in loading at centromeres, there is more cohesin available to load on chromosome arms. However, if this were to be the case, then mutations that abolish centromeric loading, such as *chl4Δ* (Fernius et al., 2013), would also be able to rescue *scc4Δ*. To test this, strains containing *chl4Δ* and *scc4Δ* rescued by an ectopic copy at the

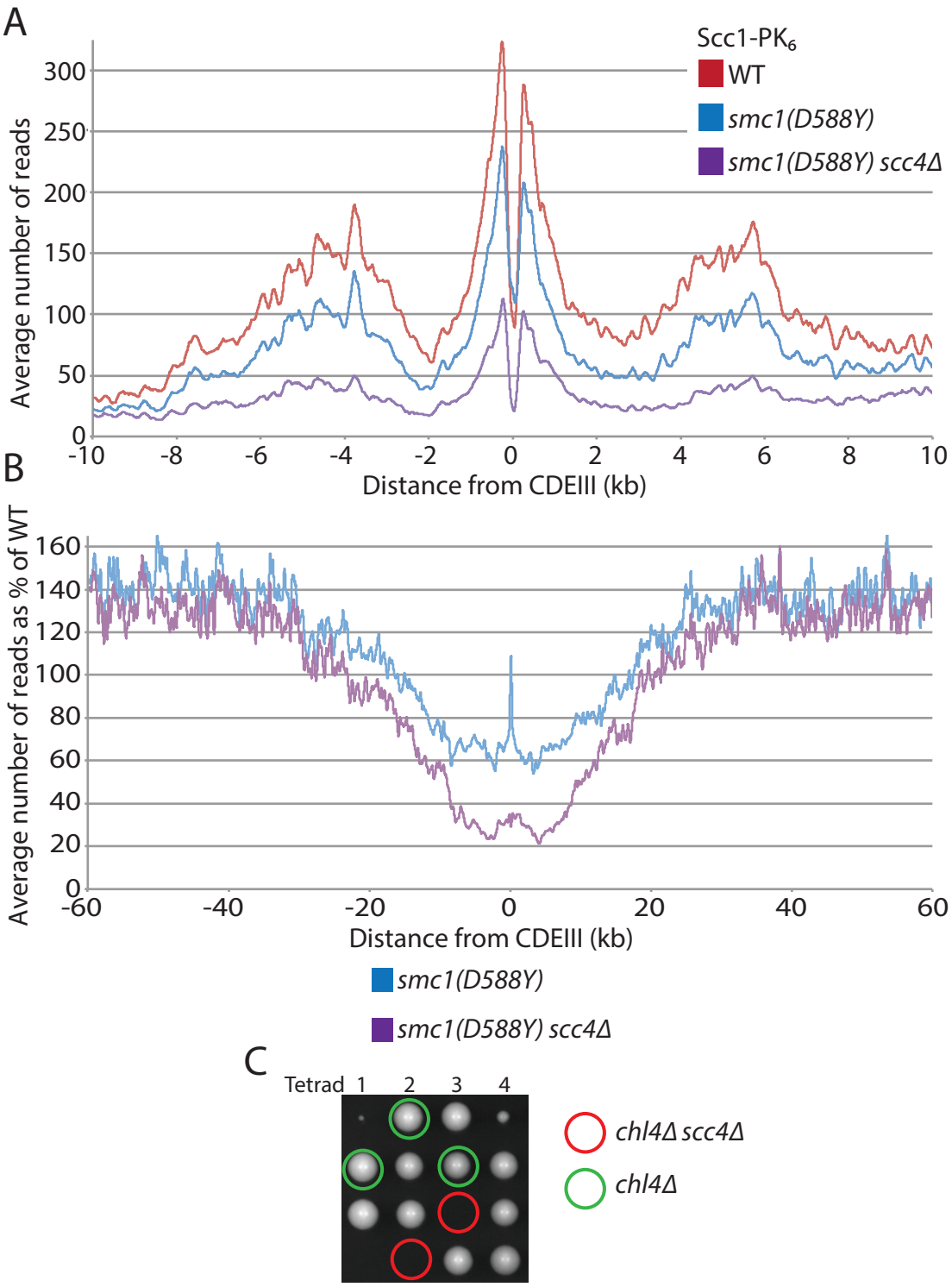


Figure III-5

***smc1(D588Y)* increases cohesin loading on chromosome arms compared to WT but is defective in centromeric loading.**

A - Calibrated average centromere profile of the experiment detailed in figure III-4.

B - Calibrated average chromosome plotted as a percentage of the number of reads obtained for WT cells.

C - Tetrad dissection of diploid strains containing CHL4/*chl4Δ*, *SCC4/scc4Δ*, *LEU2/leu2::scc4-4::LEU2*. Spores that contained the combination of *chl4Δ scc4Δ LEU2* are circled in red.

LEU locus were crossed and dissected. Spores that contained both *chl4Δ* and *scc4Δ* in the absence of an ectopic copy of WT *SCC4* were unable to grow (Figure III-5C). This shows that increased loading of cohesin on chromosome arms in the *smc1(D588Y)* mutant is not due to more cohesin availability due to the inability to load at centromeres.

smc1(D588Y) is defective in a late step in the loading reaction.

To address which step in the loading reaction is affected by *smc1(D588Y)*, it is necessary to breakdown the loading reaction into more discrete steps. Three key observations support the hypothesis that cohesin loaded at the centromere translocates to the pericentromere: firstly, point centromeres (CENs) and centromere-associated proteins are required for pericentromeric enrichment (Eckert et al., 2007; Fernius & Marston, 2009; Hinshaw et al., 2015; Tanaka et al., 1999; Weber et al., 2004); secondly, the loading complex associates with the centromere at high levels (Figure II-4A and 4C), presumably representing complexes that are engaged in the loading reaction; and finally, cohesin complexes that can bind but not hydrolyse ATP, and therefore cannot complete the loading reaction also associate at high levels with the centromere (Figure II-6B) (Hu et al., 2011). Collectively, these observations suggest that cohesin loading at centromeres can be split into two major steps. First the loading complex must form a preloading complex with the cohesin ring and associate at CEN loading sites. This step depends on binding but not hydrolysis of ATP (Hu et al., 2011). Crucially, the pre-loading complex can also be visualised by imaging GFP-tagged Scc2 in living cells, which reveals that its association with CENs is

very unstable (Hu et al., 2011). Subsequently, ATP hydrolysis triggers SMC head disengagement and translocation to the pericentromeric sequences.

The preferential recruitment of cohesin at CENs relative to the pericentromeric sequences in *smc1D588Y* cells suggest that's the second but not the first step is defective. If so, recruitment of Scc2 and *smc3(E1155Q)* to CENs should not be reduced. To test this, the occupancy of Scc2 and the cohesin hydrolysis mutant, *smc3(E1155Q)* in *smc1(D588Y)* cells at CENs were examined.

Remarkably, the levels of both Scc2 and *smc3(E1155Q)* occupancy at CENs is significantly increased in the presence of *smc1(D588Y)* compared to WT Smc1 (Figure III-6A and 6B). This implies that the Smc1 mutation adversely affects the second step and not the first. In other words, *smc1D588Y* adversely affects translocation of preloading complexes at CENs into pericentromeric sequences. This is consistent with the relatively high levels of Scc1 observed at the centromere that presumably represents complexes that are stuck in this very late stage of loading and are therefore unable to translocate away to form the proximal and pericentromeric peaks observed in WT cells.

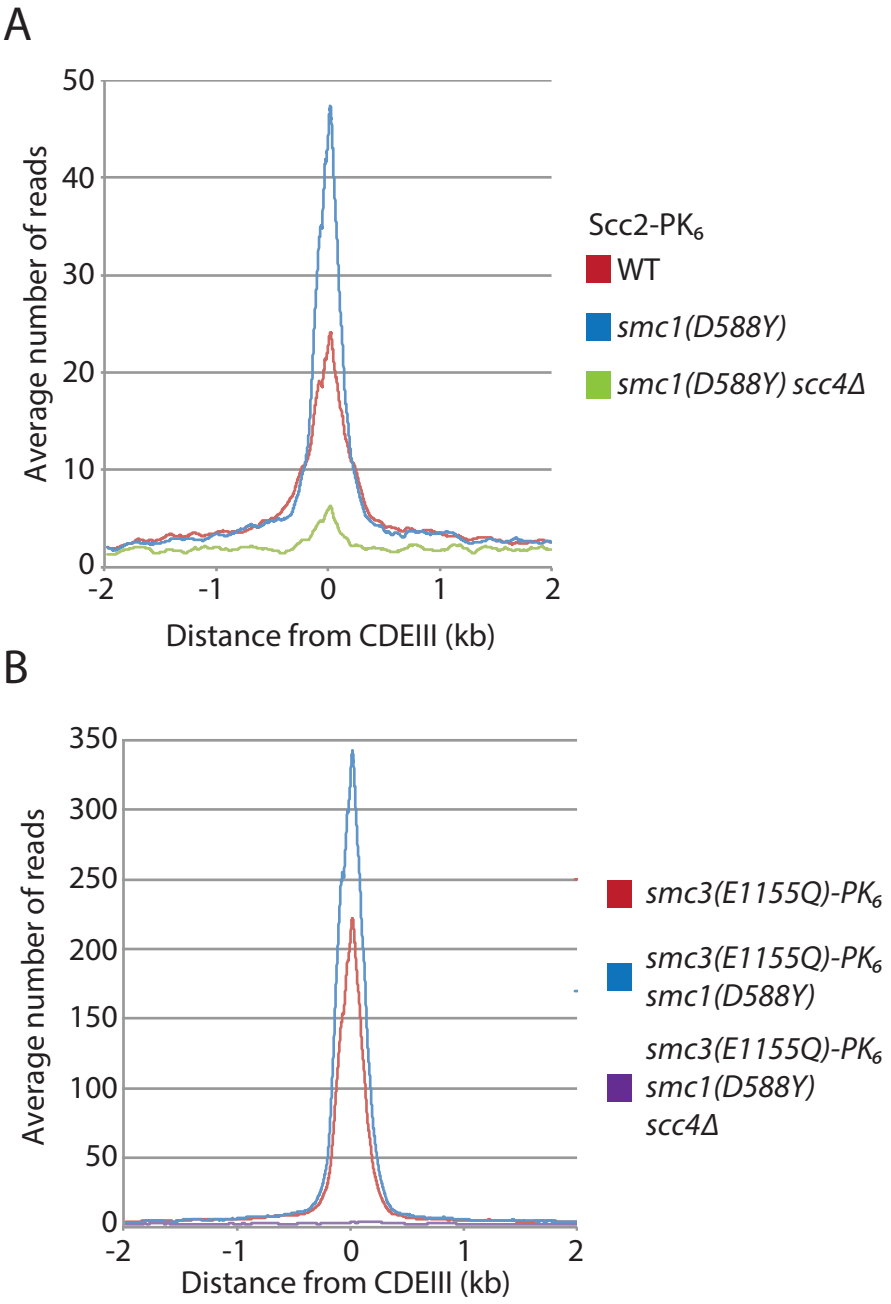


Figure III-6

***smc1(D588Y)* increases Scc2 and *smc3(E1155Q)* localisation at centromeres.**

A - Calibrated average centromere profile comparing cycling cells expressing *SCC2-PK$_6$* with cells also expressing *smc1(D588Y)* in the presence and absence of Scc4 (K21388, K24680 and K24678, from top to bottom).

B - Calibrated average centromere profile comparing cycling cells expressing an ectopic copy of *smc3(E1155Q)-PK$_6$* with cells also expressing *smc1(D588Y)* in the presence and absence of Scc4 (K24562, K24689 and K24564, from top to bottom).

Discussion

Although it has been shown that the Scc2/4 loading complex requires Scc1 for association with DNA and vice versa (Fernius et al., 2013; Michaelis et al., 1997), the specific role of Scc2 and Scc4 and how they interact with the cohesin ring, as well as the changes that may occur within ring itself to enable loading, are very poorly understood.

To try to identify links between the complexes in order to better understand the loading reaction, a screen for suppressors of the thermosensitive phenotype of the *scc4-4* allele was performed (Dr M. Demidova, personal communication). This revealed a mutation in the hinge domain of Smc1 that can suppress the temperature sensitivity of *scc4-4* and also bypass the requirement for the otherwise essential protein altogether. This finding not only revealed a genetic link between the cohesin ring and the loading complex but to a specific interface within the ring, which interestingly is the interface that has been suggested to be the DNA entry gate.

It was found that substitutions to tyrosine, tryptophan and phenylalanine at position 588 could also bypass the requirement for Scc4 but no other residue in either Smc1 or Smc3 was found to do so (Dixon, 2013). Although these suppressor mutations were observed to cause weakening of the interaction between the Smc1 and Smc3 hinge domains, this doesn't account for the mechanism of suppression, as other mutations that are known to weaken the hinge were unable to suppress *scc4-4* related lethality (Dixon, 2013). This implies that the mechanism of suppression by the D588 mutations is not due to a generic change to the hinge such as weakening the Smc1/Smc3 interface.

It was also previously found that *smc1(D588Y)* was able to restore cohesin loading on chromosome arms but not at the centromere (Dixon, 2013). This demonstrated that one of the roles of Scc4 is to recruit Scc2 to the centromere but also that this role is not essential as *scc4Δ* can be rescued without restoring centromeric loading. However the ChIP-seq experiments used were not calibrated and therefore the results obtained could not be quantitatively compared. This chapter aimed to repeat the ChIP-seq experiments in a quantitative manner to see what further insights could be elucidated from the *smc1(D588Y)* mutant.

The effect of *smc1(D588Y)* was measured in terms of suppression of both *scc4-4* and *scc4Δ*. In both cases, as in prior experiments, loading was restored on chromosome arms but not at the centromere. However, it was also revealed that at 25°C, cohesin levels were not only restored but were restored to levels greater than in WT cells. Furthermore, it was observed that *smc1(D588Y)* even in the presence of Scc4, is defective at loading at centromeres. This defect was found to occur in a late step of the loading reaction, as recruitment of Scc2 and association of the hydrolysis mutant, *smc3(E1155Q)*, at centromeres, were also higher in the presence of *smc1(D588Y)*. This indicated that the defect occurs in the second loading step, translocation from CENs to pericentromeric sequences upon ATP hydrolysis.

Why then does the mutation restore loading in *scc4-4* cells along chromosome arms but not at centromeres? It is already known that there are differences between loading on chromosome arms compared to loading at centromeres. Centromeric loading depends on centromeric sequences and on kinetochore-associated proteins but this does not imply that the mechanism of

loading is fundamentally different *per se*. However, the behavior of the *smc1D88Y* mutant suggests that it is. *smc1D588Y* is able to increase loading on chromosome arms but not at centromeres and this increase on chromosome arms is not merely due to more cohesin availability for loading at the arms due to the inability to load to the centromere. If this were the case, mutations that reduce loading at centromeres, such as *chl4Δ*, would also be able to rescue the lack of Scc4 but this is not the case. Therefore *smc1D588Y* has different effects on chromosome arms than at centromeres indicating that there is a different mechanism of loading on chromosome arms than at centromeres.

The observation that the loading complex and cohesin rings containing the *smc3(E1155Q)* mutation and so unable to complete the loading process, associate with the centromere and not the surrounding region, together with the observation that cohesin association starts at sites proximal to the centromere before accumulating at the pericentromere suggest that cohesin is first loaded at the centromere and then translocates away to the pericentromere. These findings are also consistent with two steps in the loading reaction. First, that cohesin binds ATP causing the Smc1 and Smc3 heads to engage. The loading complex binds also binds to the cohesin ring and this "pre-loading" complex associates with the centromere and does so in the absence of ATP hydrolysis. In the second step, ATP hydrolysis occurs, completing the loading reaction before subsequent translocation to the pericentromeric sequences.

Although this currently cannot be observed on chromosome arms, there is no reason to believe that the same basic mechanism doesn't occur there. But why then, does the same mutation increase loading on the arms but inhibit loading at centromeres? This paradox can be explained with a single cohesin

loading cycle. If the loading reaction involves cohesin going through a cycle of changes that include a single cycle of ATP binding and hydrolysis linked to cycle of changes in the hinge domain then the following cycle becomes apparent. Soluble cohesin with heads unengaged form a complex with the loading complex. In this complex the heads become engaged. On completion of the loading reaction, cohesin rings are released from the loading complex and the heads return to an unengaged state. Similarly the hinge domain undergoes a cycle where soluble cohesin hinges are in one state, we shall call this the "D" state, and cohesin bound to the loading complex enter another state, which we will call the "Y" state. Thus the cycle of head engagement can be accompanied by a cycle in the state of the hinge domain.

The cycle therefore becomes, unengaged "D" state cohesin binds to the loading complex and transitions to engaged "Y" state cohesin. Loading is completed upon ATP hydrolysis and cohesin returns to the unengaged "D" state (Figure III-7A). This would mean that the complexes observed at centromeres would represent the engaged "Y" state. If the transition from the unengaged "D" state to the engaged "Y" state was rate-limiting on the arms but the transition from the engaged "Y" state to the unengaged "D" state was rate limiting at centromeres, then we have introduced a key difference between the chromosome arms and the centromere. Note that this assumption is consistent with the observation of high levels of the loading complex at the centromere but very low levels at the arms. The rate-limiting step at centromeres would be the first step in the loading reaction causing the dissociation of Scc2 from centromeres to be slower than at chromosome arms. This would cause high

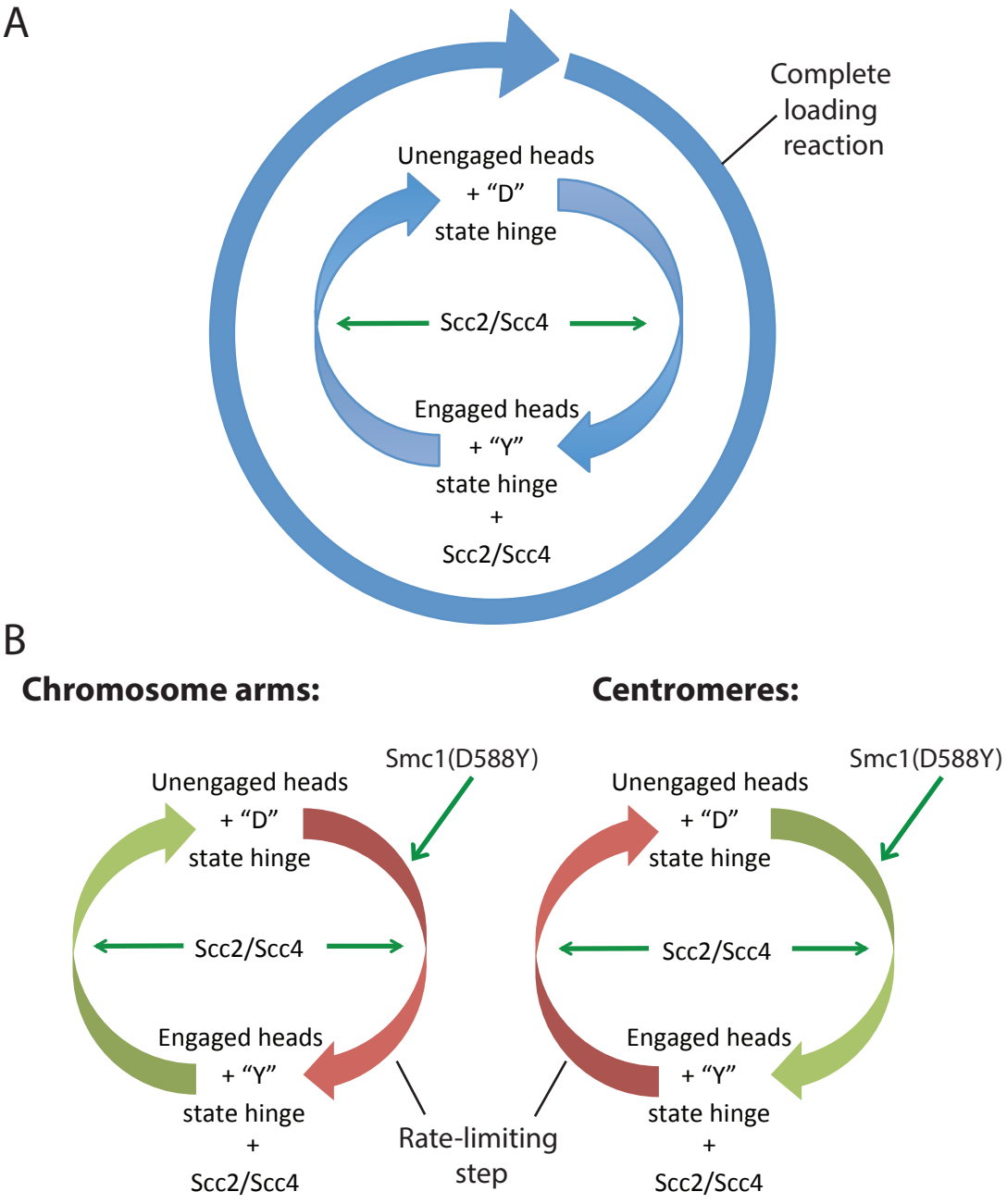


Figure III-7

A model of the loading reaction

A - Transition from the unengaged "D" state cohesin to the engaged "Y" state cohesin that is bound to the loading complex and back again is required to complete the loading reaction.

B - The rate-limiting step in this reaction is different on chromosome arms to centromeres causing *smc1(D588Y)* to increase loading at the arms but hinder loading at centromeres.

levels of Scc2 occupancy to be observed at centromeres than at the arms, which is observed. Therefore this assumption is not altogether unreasonable.

Given this cycle, if *smc1(D588Y)* stabilised the "Y" state, then it would facilitate the transition from the unengaged "D" state to the engaged "Y" state. At the arms, this would accelerate the rate-limiting step enabling increased levels of loading. However at centromeres, this would promote the engaged "Y" state and thereby hinder the already rate limiting step of transitioning from the engaged "Y" state to the unengaged "D" step thereby impeding the loading reaction. This could then explain the different effect of the same mutation on loading at the centromeres and chromosome arms (Figure III-7B).

The exact nature of "D" and "Y" states is a key question that remains to be answered. One possibility is that *smc1(D588Y)* alters opening of the hinge domain to allow DNA entry into the cohesin ring. However, as previously discussed it is unlikely that the D588Y mutation merely weakens the interface as many other mutations have a similar effect but are unable to bypass Scc4. Loss of function effects observed from locking the hinge domain suggested that the hinge could be the entry gate for DNA (Gruber et al., 2006). However, there is still no direct proof that the hinge acts as an entry gate for DNA. Furthermore, models that include more than one interface of the cohesin ring opening to allow DNA to enter the ring have not been satisfactorily eliminated. Logically, more than one interface appears unlikely, as presumably another mechanism would be required to maintain the released protein with the complex. Scc2 has been shown to bind Scc1 (Kukuchi et al. in press). If the Smc1-Scc1 and Smc3-Scc1 interfaces both opened to enable DNA entry, perhaps Scc1 remains bound to the complex via Scc2. This Scc1/Scc2 subcomplex would still need to interact

elsewhere with the cohesin ring to remain bound. Perhaps then, Scc2 also interacts with the hinge domain. This would suggest that the role of *smc1(D588Y)* is to strengthen the interaction between the hinge and Scc2. Although highly speculative, this would explain the importance of the hinge in the loading reaction without requiring it to be the entry gate.

Irrespective of the mechanism by which *smc1D588Y* acts, the mutation shows that the hinge has an intimate role during the loading reaction. Until now, the evidence for a role of the hinge in loading has stemmed from loss of function effects and therefore any observed effect on loading could be indirect. The observed differential effects of *smc1(D588Y)* on loading that depend on precisely where the loading reaction is taking place cannot be explained by indirect effects of the hinge mutation on the loading reaction. Therefore *smc1(D588Y)* provides the clearest evidence that the hinge has a role in the loading reaction.

Mutations in Scc2 enable loading of cohesin in the absence of Scc4

Introduction

The screen used to identify suppressors of the temperature sensitive phenotype of *scc4-4* led to the identification of the first genetic link between the loading reaction and the hinge domain of cohesin. This discovery also demonstrated for the first time that it was possible to bypass the requirement for Scc4 for viability. These findings have subsequently enhanced our understanding of the loading reaction as they have demonstrated that the hinge has a key role in the loading reaction. However, there are still many aspects that remain a mystery, including the precise role of the hinge and the function of the loading complex.

Further screening has the potential to elucidate other factors involved in the loading reaction. Given that *smc1(D588Y)*, the hinge mutant that is able to bypass the requirement for Scc4, was isolated from 12 independent clones of the parental strain grown at the restrictive temperature, it stands to reason that other mutations that could arise were not identified due to the predominance of *smc1(D588Y)*. Thus the original screen may not have been saturated.

To identify additional suppressors the original screen would need to be modified in order to reduce the probability of Smc1 D588 mutating to any of the residues that were identified as suppressors, namely tyrosine, tryptophan and phenylalanine. This led to two subsequent screens, one performed in the same manner as the original screen that was used to identify suppressors of *scc4-4* and

one modified to allow for the swift identification of suppressors able to bypass the requirement for Scc4 entirely.

This chapter aims to outline these screens and characterise the identified mutants, namely suppressors of thermosensitive Scc4 in Scc2.

Results

Mutations in the N-terminal domain of Scc2 are able to suppress the temperature sensitive phenotype of scc4-4.

To identify additional factors involved in the cohesin loading reaction, a strain containing a temperature sensitive allele of Scc4, *scc4-4* as well as *smc1(D588E)* at the endogenous loci was generated. Glutamate was chosen to substitute aspartic acid in Smc1 at position 588 as with a base sequence of GAG at least two point mutations would be required to change the residue to either tyrosine, tryptophan or phenylalanine. Furthermore, *smc1(D588E)* had previously been shown to be unable to rescue *scc4-4* but able to completely rescue *smc1Δ* (Dr S. Dixon, personal communication). As two point mutations randomly occurring in the same codon is unlikely, this should effectively remove *smc1(D588Y/W/F)* from the possible suppressors. Forty independent clones of the resulting strain were grown at the restrictive temperature of 35.5°C until naturally arising mutants that permitted growth at high temperature could be isolated.

A total of 388 suppressor colonies were picked, but the vast majority were either revertants of the *scc4-4* allele or did not show 2:2 segregation of the suppressor phenotype and were therefore discarded. Whole genome sequencing

revealed 5 suppressors with mutations in Scc2. These mutations were V36L, S170L and L937F, with S170L occurring once and the others twice.

Residues V36 and S170 are both located in the N-terminal domain of Scc2 (Figure IV-1) (Hinshaw et al., 2015). Recently this domain was co-crystalised with Scc4 (Chao et al., 2015; Hinshaw et al., 2015), this showed that Scc4 wraps around the largely unstructured N-terminus of Scc2. Although the structure doesn't reach S170, it seems likely that both of these mutations stabilise the interaction between Scc2 and Scc4 enough to reverse the effects of Scc4(Y40N) which causes the temperature sensitivity of *scc4-4*.

Mutations in Scc2 are able to bypass the requirement for the essential protein Scc4.

To identify other mutations that not only suppress *scc4-4* but also rescue *scc4Δ*, a second screen was designed. The endogenous copy of *SCC4* was deleted and rescued by a *URA3*-containing plasmid containing the *scc4-4* allele. As in the previous screen, the endogenous copy of *SMC1* was mutated to *smc1(D588E)* to reduce the probability of isolating the known suppressors in Smc1. Again 40 independent clones of the resulting strain were grown at the restrictive temperature (35.5°C) until suppressors could be isolated. Once isolated, the suppressors were allowed to recover at 25°C before they were streaked onto medium containing the drug 5-FOA. 5-FOA is toxic to strains containing the *URA3* gene and so selects for cells that have naturally lost the plasmid due to imperfect segregation. This means that all strains that are viable on 5-FOA containing media must also contain a mutation that enables the cells to grow in the absence of Scc4 (screen performed by W. Chen). Whole genome sequencing of the

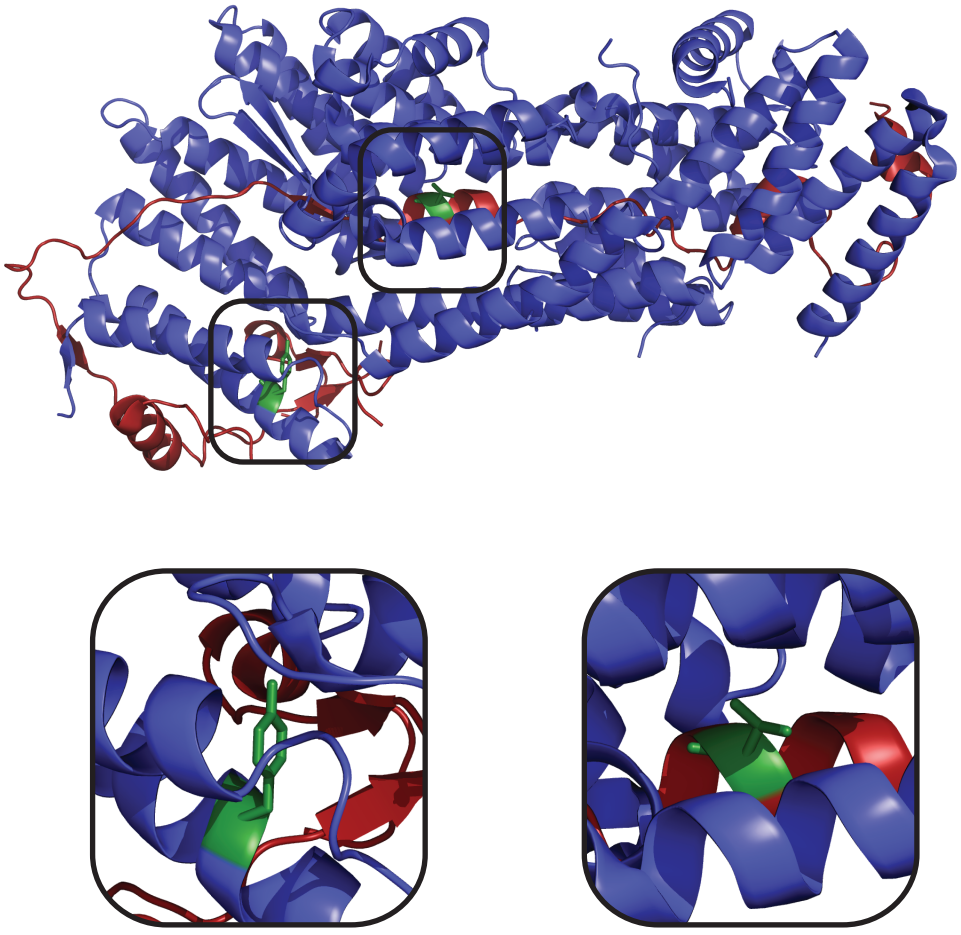


Figure IV-1

V36 is located in the Scc4 binding region of Scc2.

Crystal structure of the Scc2$^{1-181}$/Scc4 complex with Scc2 in red and Scc4 in blue (PDB 4XDN). Scc4 Y40 that causes temperature sensitivity when mutatated to asparagine and the Scc2 residue V36 that suppresses the temperature sensitivity when mutated to leucine are shown in green.

suppressors revealed 10 that contained mutations in Scc2. Six of these contained the mutation E822K, two L937F and two L937W.

To verify that the mutations are the cause of the suppression, the mutations were recreated *de novo* by integrating the mutated copy of Scc2 into the endogenous locus. The resulting strain was then crossed to a strain with endogenous *SCC4* deleted but rescued by an ectopic copy of *scc4-4*. The resulting tetrads confirmed viability even in the absence of *SCC4* (Figure IV-2A, B). Notably cells expressing *scc2(E822K)* grew much better than those expressing *scc2(L937F)* but neither grew to WT levels. Interestingly the combination of the two mutations gave spores that grew as well as WT spores (Figure IV-2C) so the effect of this double mutation was studied further.

Scc2 residues E822 and L937 are both highly conserved residues.

A multiple sequence alignment of Scc2 sequences from yeasts through to mouse shows that both residues are very highly conserved and lie within a highly conserved region of Scc2 (Figure IV-2D). This suggests these residues may be required for a fundamental part of Scc2's role in loading.

A crystal structure of Scc2 has recently been obtained from *Chaetomium thermophilum* (Kikuchi et al., 2016). The shape of Scc2 resembles a "handled hook" with H1-7 forming the handle and H8-24, the hook. This reveals both E822 and L937 to be in the hook domain of the protein with E822K in the region described to interact with Scc1, pointing out from the surface of the protein (Figure IV-3).

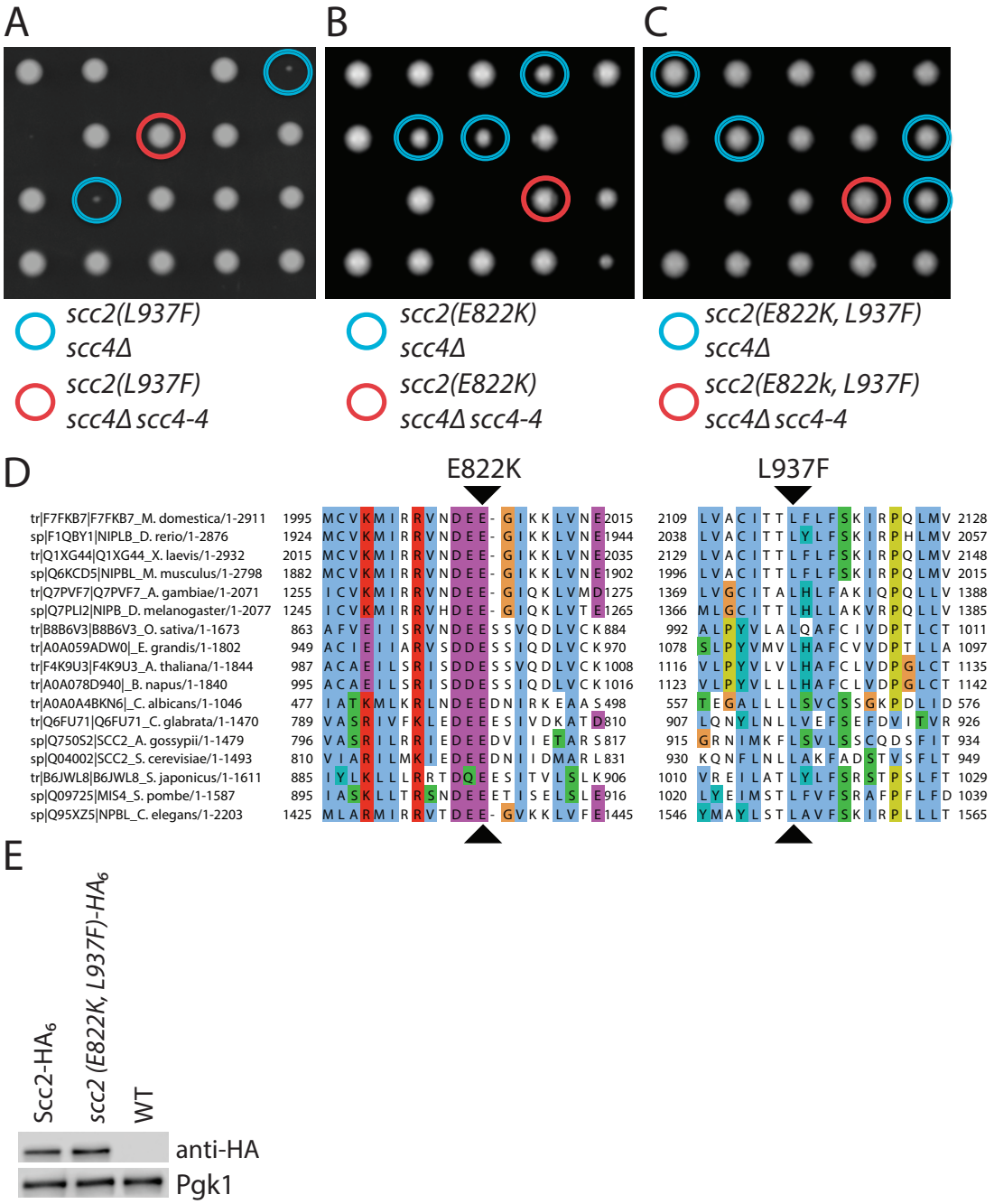


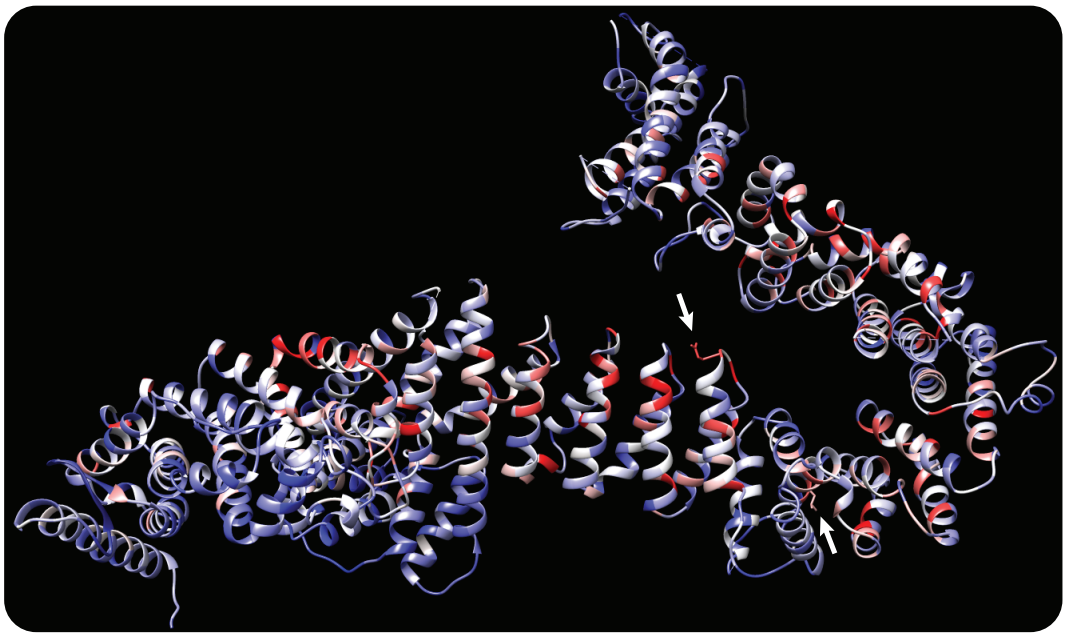
Figure IV-2

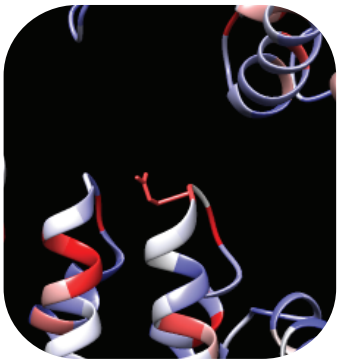
Suppressor mutations in conserved residues in Scc2 are able to rescue *scc4Δ*.

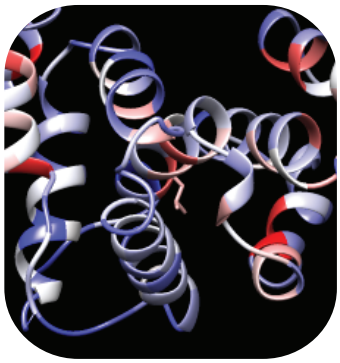
A-C - Tetrad dissection of diploid strains containing *SCC4/scc4Δ, LEU2/leu2::scc4-4::LEU2* with *SCC2/scc2(L937F)* (A), *SCC2/scc2(E822K)* (B), or *SCC2/scc2(E822K, L937F)* (C). Spores in which *scc4Δ* is rescued by the endogenous Scc2 mutants are circled in blue.

D - Multiple sequence alignment indicating the conservation of Scc2 residues E822 and L937 in *S. cerevisiae* across various other eukaryotes.

E - Western blot against HA_6-tagged Scc2 shows that the suppressor mutations have no effect on Scc2 protein level. Pgk1 was used as a loading control (K25054, K25053 and K699, from left to right).







E1125
(E822 in *S. cerevisiae*)

L1230
(L937 in *S. cerevisiae*)

Low High
Conservation

Figure IV-3

Location of the Scc2 suppressors in the Scc2 structure.

Crystal structure of Scc2$^{385-1840}$ from *Chaetomium thermophilum* coloured to depict the conservation of the residues (red - highly conserved, blue - less conserved). Both Scc2 residues that give rise to suppressor mutations are conserved and located within the hook domain of Scc2 with E1125 in *Chaetomium thermophilum (*E822 in *S. cerevisiae*) in the region described to interact with Scc1.

scc2(E822K, L937F) is not upregulated in the absence of Scc4.

A possible explanation as to how these mutations bypass the requirement for Scc4 is that they cause upregulation of Scc2 in order to compensate for the lack of Scc4. To test this, strains where either the WT Scc2 protein or the mutant protein were tagged with HA_6 were grown to log phase and protein levels were measured by Western blot (Figure IV-2E). This showed no difference between the protein levels and therefore the suppression observed is not due to upregulation of Scc2.

scc2(E822K, L937F) suppresses the loading defect of scc4-4 on chromosome arms but not at the centromere.

To investigate to what extent *scc2(E822K, L937F)* is able to suppress *scc4-4*, strains expressing *SCC1-PK*$_6$ in the presence of WT *SCC2* or *scc2(E822K, L937F)* with and without *scc4-4* were analysed by quantitative ChIP-seq. The strains were arrested in alpha factor to synchronise them before release into media pre-warmed to 37°C containing nocodazole to prevent entry into anaphase. Samples were taken at 75 min when all strains had reached G_2 (Figure IV-4B).

The chromosome profile of *scc2(E822K, L937F)* in the presence of *scc4-4* showed cohesin occupancy to be partially restored on the arms compared to *scc4-4* alone, but with very little cohesin associated around the centromere compared to WT (Figure IV-4A). *scc2(E822K, L937F)* itself however did not have any decrease in cohesin loading, but possibly an increased or distorted profile around the centromere. Average centromere profiles were plotted to investigate the difference in the region ±10 kb around the centromere in more detail (Figure

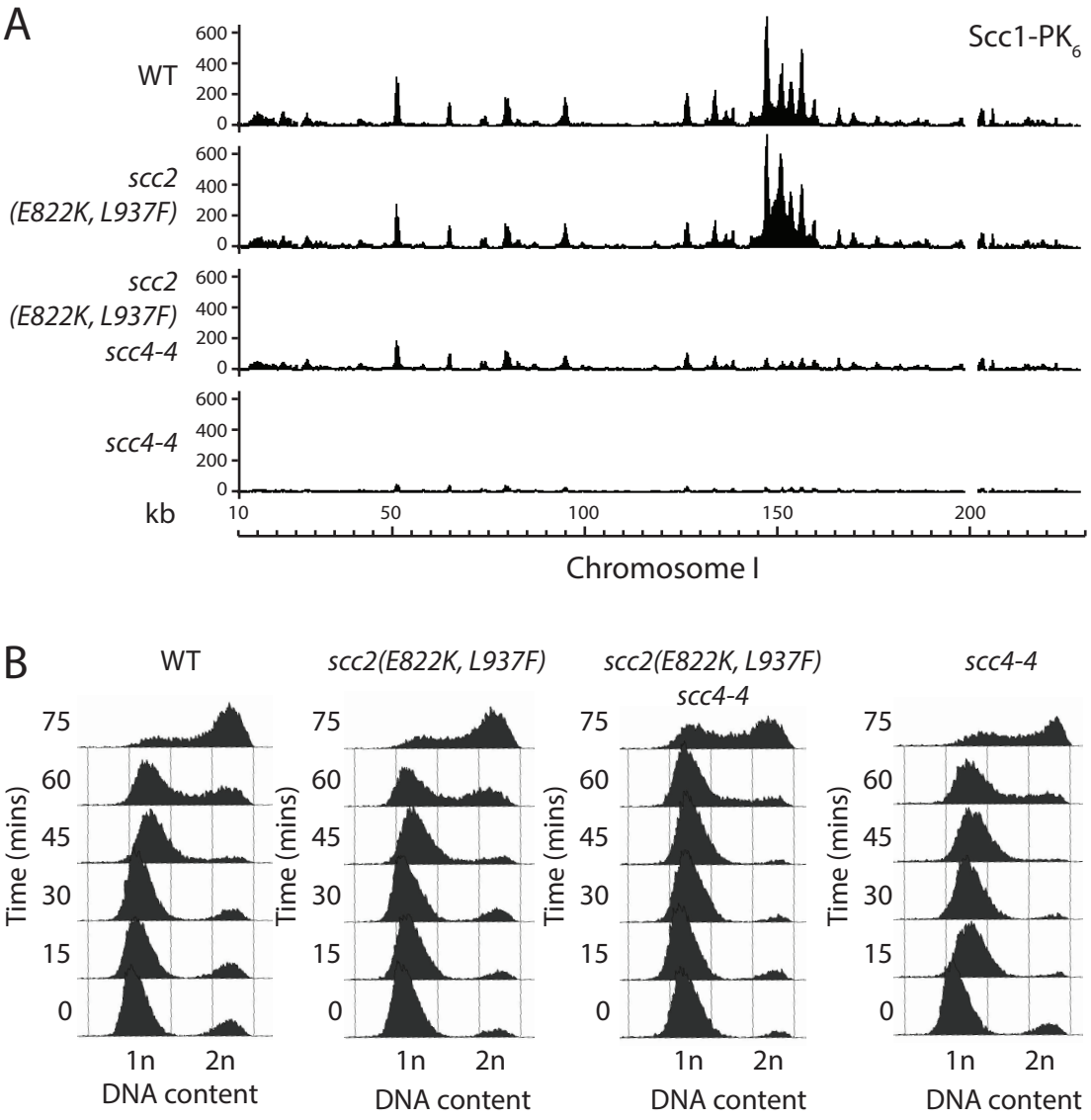


Figure IV-4

***scc2(E822K, L937F)* suppresses the loading defect of *scc4-4* at the arms but not around the centromere.**

Strains expressing endogenous *SCC1-PK$_6$* were arrested in alpha factor and released into nocodazole containing medium at 37°C. Samples were taken at 75 min after release.

A - Quantitative ChIP-seq profile across chromosome I. The y axis indicates the calibrated number of reads and the x axis the base pair position across chromsome I. (K22005, K24687, K24744 and K22001, respectively).

B - Cell cycle progression during the experiment shown in A as measured by FACS.

IV-5A). This showed very little recovery of cohesin occupancy around the centromere in suppressor cells expressing *scc4-4* but in the presence of WT Scc4, cohesin occupancy at the peaks proximal to the centromere and in the regions between the proximal centromeric peaks and the pericentromeric peaks is increased. This suggests that in the presence of Scc4, these mutations in Scc2 enable increased levels of loading at the centromere compared to WT. Furthermore, a greater fraction of cohesin loaded at CENs is found in the vicinity of the CEN loading sites than in WT cells suggesting that the process of translocation is also affected.

To assess the level of loading compared to WT throughout the genome, a calibrated average profile 60 kb either side of the centromere was generated for each strain and plotted as a percentage of the WT profile (Figure IV-5B). This showed the increase around the centromere reached over 200% of WT levels. The suppressor mutations also modestly increased the level of loading at the arm regions, 40-60 kb away from the centromere suggesting that these mutations enable Scc2 to preform better across the genome. This will be readdressed at a more physiological temperature (25°C) to confirm the effect. However, in the presence of *scc4-4*, although loading on the arms remains at 80-90% of WT, there is only a very slight increase at the centromere. Thus although the mutations may enable Scc2 to function better, it is still unable to bypass the role of Scc4 at centromeres.

To assess the effect of the mutations at more physiological temperatures and in the complete absence of Scc4, the effect of *scc2(E822K, L937F)* on cohesin loading was also analysed at 25°C with or without *scc4Δ*. Again strains were synchronised with alpha factor before release and samples taken when cells had

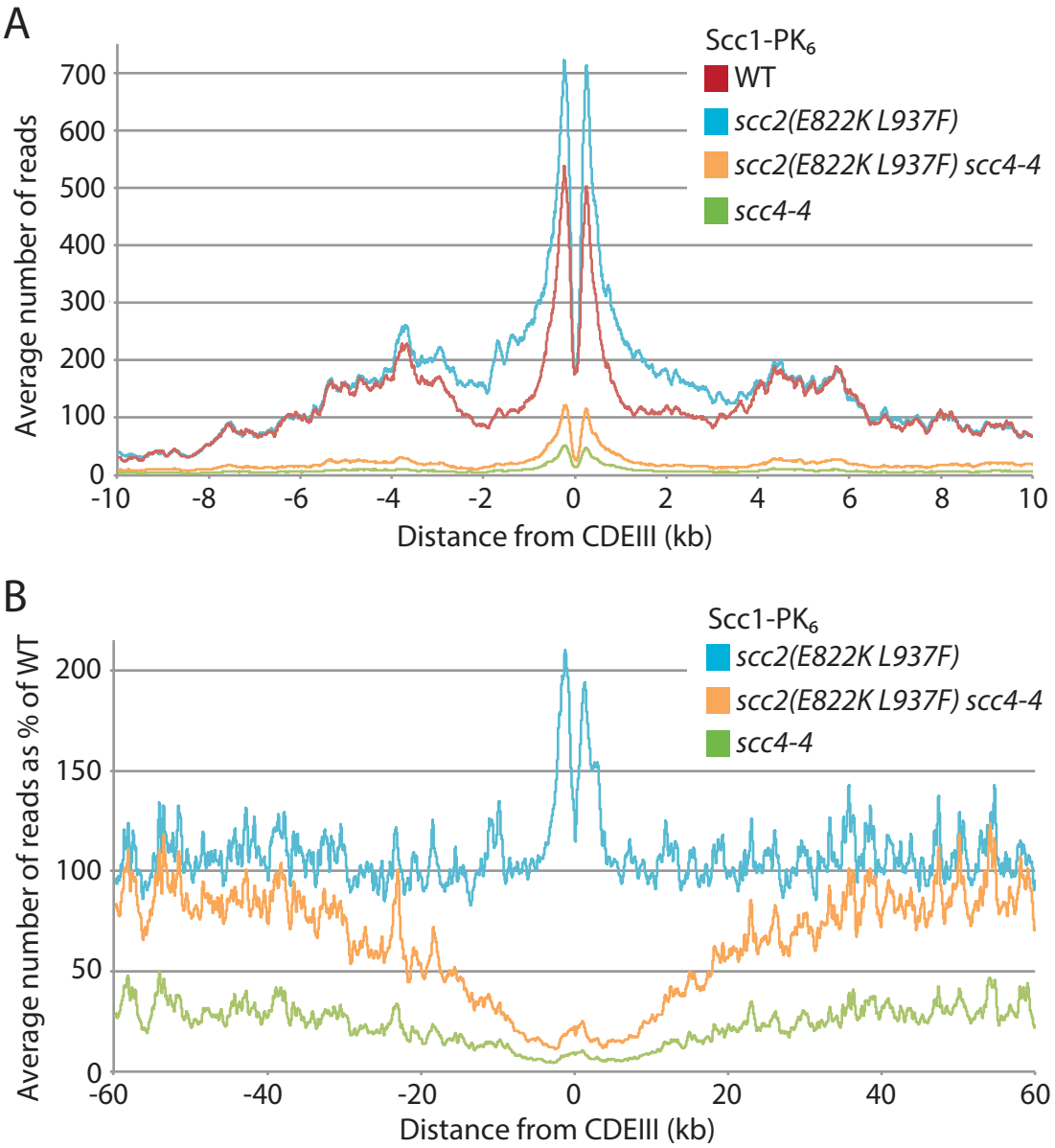


Figure IV-5

***scc2(E822K, L937F)* restores cohesin loading in cells containing *scc4-4* at the restrictive temperature on chromosome arms but not at the centromere.**

A - Calibrated average centromere profile covering 10 kb either side of CDEIII of the experiment detailed in figure IV-4.

B - Calibrated average chromosome covering 60 kb either side of CDEIII plotted as a percentage of the number of reads obtained for WT cells.

entered G_2 at 60 min (Figure IV-6B). In the presence of the double mutant and in the absence of Scc4, an essential gene in WT background, chromosome profiles revealed that cohesin was present at WT levels at the expected loci on the chromosome arms. Around the centromere, cohesin levels were significant, albeit lower than WT (Figure IV-6A). This is contrary to the previous experiment preformed at 37°C and with the equivalent experiment preformed with *smc1D588Y*. Average centromere profiles also show very significant levels of cohesin around the centromere and within the pericentric region in the absence of Scc4 (Figure IV-7A). When averaging all chromosomes 60 kb either side of the CDEIII this increase is quantified to be greater than 60% of WT levels (Figure IV-7B). Furthermore, loading on the arms was increased, reaching 110% of WT levels. The effect of the suppressor mutations in the presence of Scc4 was again seen in this experiment with increased levels of cohesin occupancy seen at both the centromere, reaching approximately 160% of WT, and at chromosome arms, at 110-120% of WT.

The striking recovery of cohesin loading at the centromere even in the absence of Scc4 suggests that these suppressor mutations are actually able to rescue partially the function of Scc4 at centromeres. However, in cells containing the two mutations in combination with the temperature sensitive allele *scc4-4* at restrictive temperature, cohesin levels at centromeric regions were very much reduced (compare figures IV-5B and IV-7B). This discrepancy could be an effect of *scc4-4* itself: Cells containing the combination of the suppressor mutations and *scc4-4* grew slightly slower than the other strains at restrictive temperature, possibly indicating that despite Scc4 being non-functional, it may still be able to bind Scc2 and inhibit the effect of the suppressor mutations at the centromere

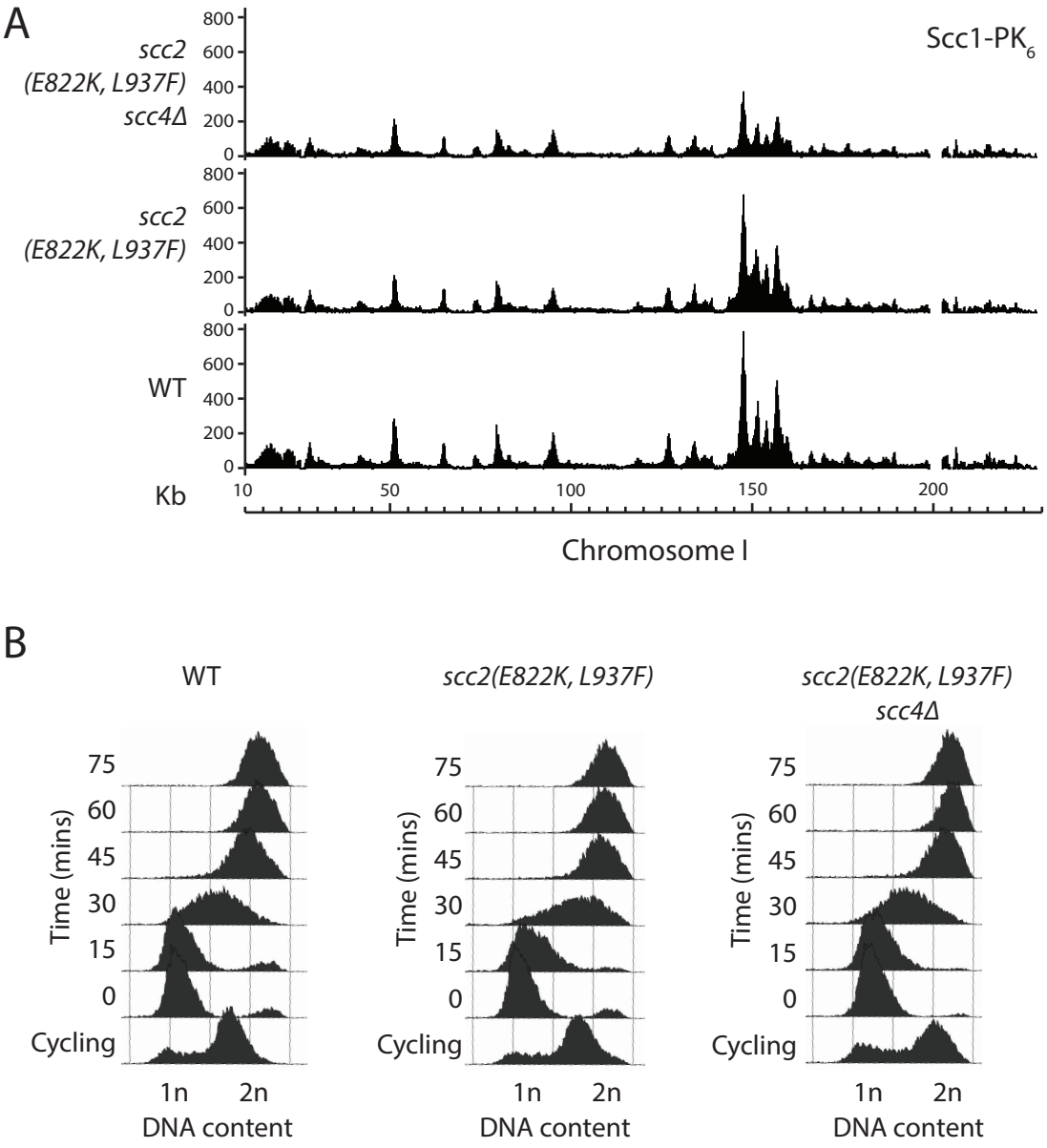


Figure IV-6

***scc2(E822K, L937F)* rescues cohesin loading on chromosome arms and partially at the centromere.**

Cells were synchronised in G_1 with alpha factor before being released into nocodazole containing medium at 25°C. Samples were taken 60 min after release.
A - Quantitative ChIP-seq profile across chromosome I comparing WT cells to cells expressing *scc2(E822K, L937F)* in the presence and absence of Scc4. The y axis indicates the calibrated number of reads and the x axis the base pair position across chromsome I. (K24745, K24687 and K22005 from top to bottom).
B - Cell cycle progression during the experiment shown in A as measured by FACS.

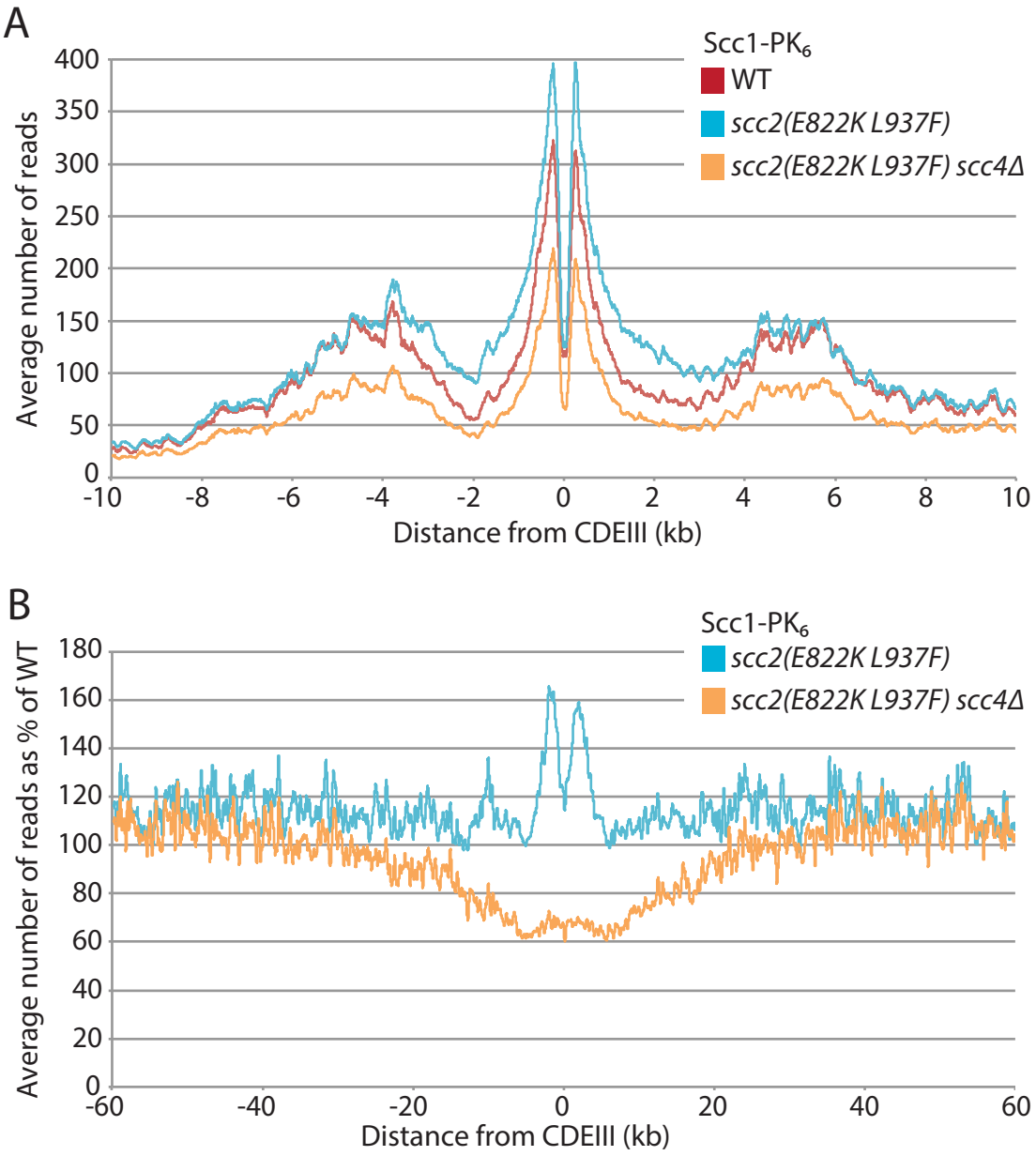


Figure IV-7

***scc2(E822K, L937F)* increases cohesin loading around the centromere in an Scc4 dependent mechanism.**

A - Calibrated average centromere profile covering 10 kb either side of CDEIII of the experiment detailed in figure IV-6.

B - Calibrated average chromosome covering 60 kb either side of CDEIII plotted as a percentage of the number of reads obtained for WT cells.

but not at the arms. Or the suppression of the centromeric loading in the absence of Scc4 could be due to the experimental protocol as these experiments were performed at different temperatures. To determine if the suppression is reproducible at 25°C, a repeat experiment was performed using cycling cells that are either WT or contain the Scc2 suppressor mutations as well as *scc4Δ*. A profile averaging centromeres and a 10 kb region either side showed very little loading around the centromere compared to WT cohesin occupancy (Figure III-8A), indicating that the Scc2 suppressors are unable to rescue cohesin loading at the centromere in the absence of Scc4. Why then was the opposite observed in the previous experiment using the same strains at the G_2 cell cycle stage (compare figures IV-7A and IV-8A)? Two changes were made in the setup between the two experiments: the use of cycling cells instead of cells entering G_2 and the use of nocodazole in the first experiment to prevent cells from entering anaphase. It is unlikely that the cell cycle state could cause such a drastic difference as the majority of cells in a cycling population would be in G_2. So although the cell cycle state might lead to slight differences in the profiles it would not account for such a large discrepancy. The only other plausible explanation is that the effect is due to the spindle poison nocodazole. To test this, average chromosome profiles from cells containing the suppressor mutations in Scc2 and *scc4Δ* that had either been grown in log phase or grown in the presence of nocodazole for two hours were compared. Although these profiles were from different experiments preformed on different days, due to the quantitative nature of the protocol and the fact the calibration cells added in both experiments were in the same state, the profiles should be comparable. The average profile for cells grown in the presence of nocodazole shows high levels

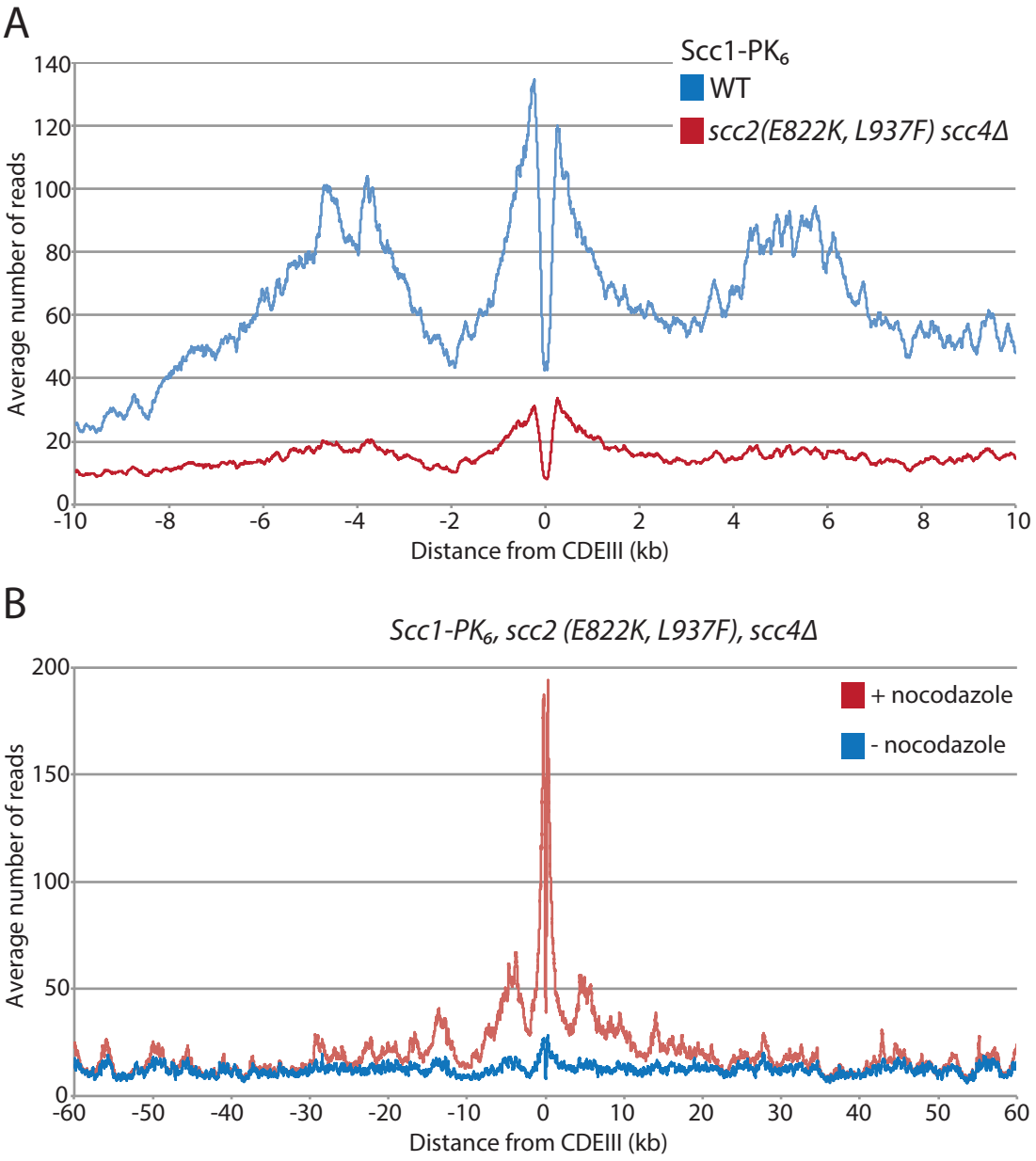


Figure IV-8

Centromeric loading is increased in the presence of nocodazole.

A - Average chromosome profile of cycling cells expressing *SCC1-PK$_6$* in a WT or *scc2(E822K, L937F) scc4Δ* background. (K22005 and K24745, respectively).

B - Average chromosome profiles showing the effect of nocodazole in cells expressing *SCC1-PK$_6$, scc2(E822K, L937F), scc4Δ* (K24687). Cells were either grown in log phase (- nocodazole) or in the presence of nocodazole (+ nocodazole) for 2 hours.

of cohesin around the centromeres and in the pericentromeric regions whereas centromeric loading is abolished in the absence of nocodazole (Figure IV-7B). This suggests that nocodazole treatment induces an Scc4-independent pathway that loads cohesin at centromeres. It is likely that this was not seen with the *smc1(D588Y)* mutation due to its defect in the loading reaction at centromeres. Thus this Scc4-independent pathway must still involve the hinge domain.

scc2(E822K, L937F) alters the distribution of Scc2 and smc3(E1155Q) occupancy at the centromere.

As previously discussed in Chapter II, the loading reaction can be broken down into two major steps. First, the loading complex Scc2/Scc4 binds to cohesin to form the "pre-loading" complex that then associates with centromeres. Then, in the second step, ATP hydrolysis occurs, after which cohesin can translocate away from the centromere. To investigate which of these steps is enhanced by the suppressor mutations in Scc2, the effect of these mutations on the occupancy of Scc2 and ATP hydrolysis deficient cohesin was observed by quantitative ChIP-seq.

Tagging *scc2(E822K, L937F)* with PK_6 and combining it with *scc4Δ* proved to be lethal by tetrad dissection, so an HA_6 tag was used instead. A *C. glabrata* strain with endogenous Scc1 tagged with HA_3 was used as the calibration strain (K25532). Log phase cells were analysed and revealed *scc2(E822K, L937F)* to be present on DNA at levels around 3-4 times higher than the WT protein (Figure IV-9A). Interestingly, the mutations not only increased the occupancy but also altered the localisation pattern of Scc2. Instead of a single peak at the centromere in a WT situation, two peaks either side of the centromere were

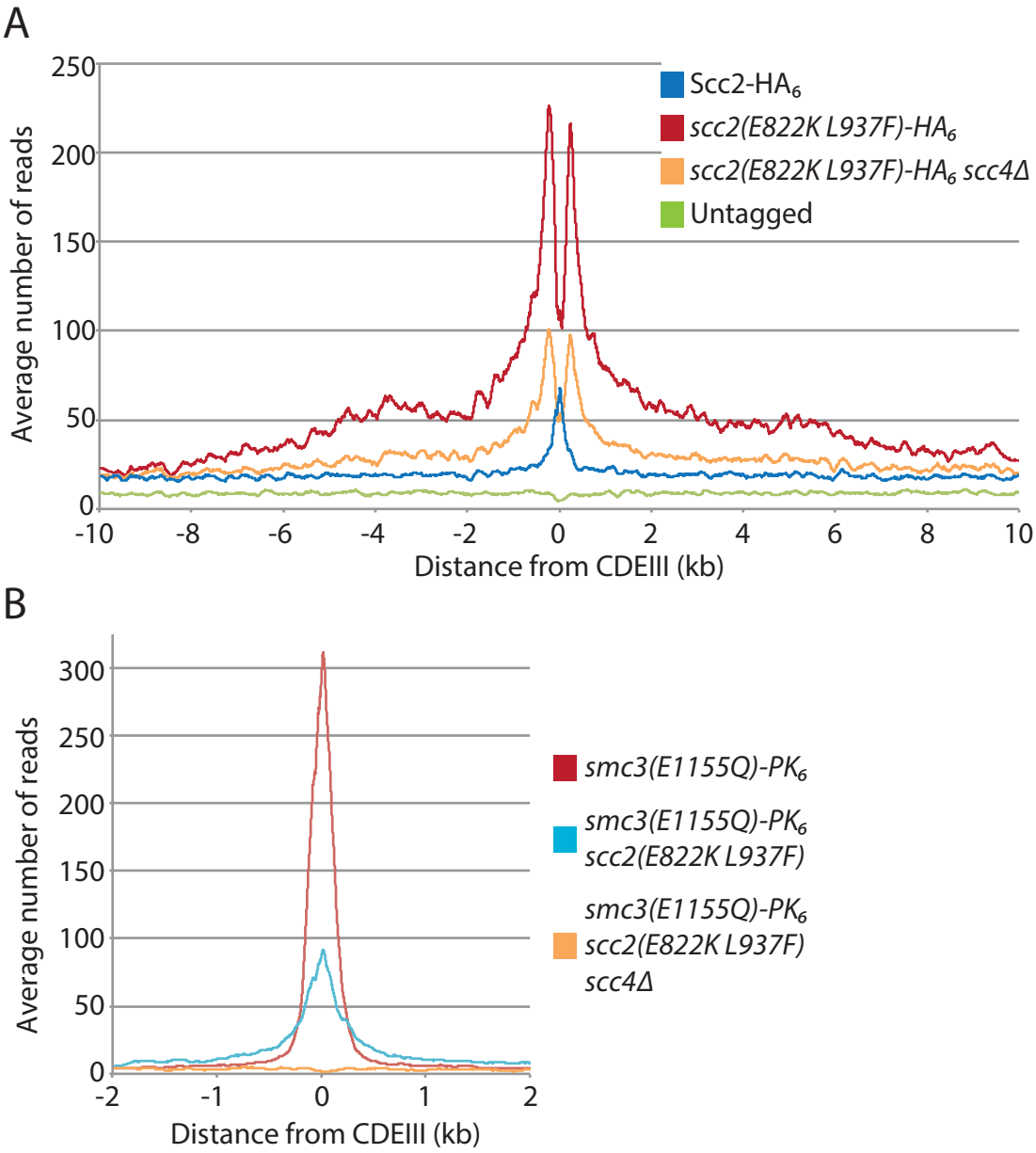


Figure IV-9

***scc2(E822K,L937F)* alters the distribution of Scc2 and Smc3(E1155Q) localisation at centromeres.**

A - Calibrated average centromere profile covering 10 kb either side of CDEIII comparing cycling cells expressing WT *SCC2-HA$_6$* with *scc2(E822K, L937F)-HA$_6$* in the presence or absence of Scc4. An untagged control was included to indicate background levels (K25054, K25053, K25418 and K699, respectively).

B - Calibrated average centromere profile covering 2 kb either side of CDEIII comparing cycling cells expressing *smc3(E1155Q)-PK$_6$* with cells also expressing *scc2(E822K, L937F)* in the presence or absence of Scc4 (K25063, K25065 and K25067, respectively).

present in the double mutant, much like the proximal peaks observed with cohesin. Furthermore, the spread of Scc2 localisation is much wider in the double mutant than in the WT, also resembling the pericentromeric peaks also observed with cohesin. This altered distribution is unchanged in the absence of Scc4 but the level of occupancy is significantly decreased, although still above WT. There are two possible explanations to this observation: either Scc2 now associates with regions of the DNA it was previously unable to, or the rate of dissociation of Scc2 from the cohesin ring is decreased, causing Scc2 to stay with the ring for longer, thereby moving along the chromosome with the cohesin ring as it begins to translocate.

To assess the other step in the loading reaction wherein ATP hydrolysis occurs, cycling cells expressing the Smc3 ATP hydrolysis mutant E1155Q in the presence and absence of *scc2(E822K, L937F)* and *scc4Δ* were assessed by quantitative ChIP-seq. As expected even in the presence of the Scc2 suppressor mutants, the lack of Scc4 completely abolished cohesin occupancy at the centromere, demonstrating that the suppressor mutants are not able to rescue the requirement of Scc4 for centromeric loading (Figure IV-9B). However it is strange that the absence of Scc4 does not abolish the association of *scc2(E822K, L937F)* with the CEN proximal sequences. This discrepancy could be due to the differences in tags causing the effect to be more drastic in the case of the EQ hydrolysis mutant. Though, both experiments did show a reduction in the absence of Scc4 when compared to *scc2(E822K, L937F)* alone. Interestingly, the suppressor mutations in the presence of Scc4 had reduced occupancy of hydrolysis-impaired cohesin at the centromere compared to WT Scc2. However, the centromeric peak is clearly broader. Moreover, measuring the areas under

the curves in Figure IV-9B up to 1 kb away from the centromere shows that they are in fact very similar. This indicates that the level of occupancy in both peaks is the same. This raises the possibility that *scc2(E822K, L937F)* somehow enables the hydrolysis mutant to translocate a limited distance away from the centromeric loading site. One way of explaining this is if the mutation causes a decrease in the dissociation rate of Scc2 from cohesin causing it to stay in the "pre-loading" complex for longer, providing a greater time frame for hydrolysis. The latter may be very much reduced but not completely abolished in the EQ mutant and by reducing the rate of Scc2 dissociation *scc2(E822K, L937F)* creates a longer window for hydrolysis. As a consequence of which, the cohesin ring is capable of a modest degree of translocation.

Discussion

In order for cohesin to topologically entrap DNA, a separate Scc2/4 loading complex is required. This complex consists of two essential proteins, Scc2 and Scc4. One role of Scc4 is to target the loading complex to the centromere causing enrichment at this site (previous sections and (Hinshaw et al., 2015)), but this is not its essential function as cells that cannot form this enrichment at centromeres are viable. The precise mechanistic role of Scc2 and the essential function of Scc4 are still unknown. A further requirement for the completion of the loading reaction is ATP hydrolysis but the exact mechanism of how the energy provided by this reaction is utilised in the loading reaction is also unknown. A screen to identify additional requirements of the loading identified the first genetic link between the loading reaction and the hinge domain of Smc1. However, the prevalence of this mutation may have masked mutations that could indicate other requirements for the loading reaction. To determine if other mutations are also able to bypass Scc4, two subsequent screens were performed in a manner that reduced the probability of the hinge mutation naturally arising. Both screens identified mutations within Scc2 that are also able to bypass not only the temperature sensitive allele of Scc4, *scc4-4*, but also the deletion.

The mutations were identified as E822K and L937F. Separately, each mutation is able to rescue Scc4 deletion but not to WT levels. However, the combination of these mutations enabled the cells to grow as well as WT cells. Both residues are highly conserved across eukaryotes and the recent crystal structure revealed that both mutations lie in the hook domain, which is thought to interact with Scc1 (Kikuchi et al., 2016).

To investigate how these mutations suppress the loss of Scc4 in the loading reaction, quantitative ChIP-seq was used to determine where cohesin occupancy is restored and to what extent. This revealed that in the absence of Scc4, the mutations were able to restore loading on chromosome arms to levels just above WT but only made a slight improvement on loading at the centromere, indicating the mutations are unable to bypass the requirement of Scc4 for centromeric loading. This was verified by examining the occupancy of the mutated form of Scc2 and cohesin rings defective in ATP hydrolysis. In both cases, depletion of Scc4 abolished the peak observed at centromeres in cells containing Scc4.

Surprisingly, in the presence of Scc4, the suppressor mutations in Scc2 caused the cohesin occupancy around the centromere to increase, particularly at the peaks proximal to the centromere and in the region between the proximal peaks and the pericentromeric peaks. This suggests that, in the presence of Scc4, the mutations in Scc2 enable more cohesin loading at the centromere causing increased occupancy in the surrounding regions. Visualisation of the effect of the mutations on Scc2 occupancy showed a striking increased in the level of Scc2 associated with centromeres and more surprisingly, also altered the distribution from a single peak at the centromere to two broader peaks either side of the centromere, similar to the cohesin peaks proximal to the centromere. This suggested that either the mutations in Scc2 enable it to directly bind to regions it wouldn't normally associate indicating that more Scc2 is bound to the regions around centromere, or that the dissociation rate of Scc2 from the cohesin ring is reduced, causing Scc2 to remain bound for longer enabling it move with cohesin as the ring begins to slide away from the centromere. The observation that the

suppressor mutations caused the peak observed at the centromere in cells defective in ATP hydrolysis to become smaller and broader implies that the second hypothesis is more likely. If more Scc2 is binding to around the centromere as in the first hypothesis, then it would also be expected that more cohesin should be associated around the centromere in cells defective in ATP hydrolysis. However, the area under the curve revealed that amounts with and without the suppressor mutations to be similar implying that there is not more association but that it is displaced. Together, this suggests that in increased association time between Scc2 and the cohesin ring may keep the preloading complex at the centromere for longer providing more time for ATP hydrolysis to occur and therefore more cohesin beginning to translocate away. This might be expected to lead to a small increase in topologically bound cohesin in cells expressing both hydrolysis defective cohesin and the Scc2 suppressor mutations. Such an increase could potentially be measured using the mini chromosome assay, wherein cohesin rings loaded onto a circular minichromosome are chemically crosslinked and the resulting DNA-protein complexes formed from cohesin rings topologically entrapping monomeric or dimeric DNAs can be separated by electrophoresis and visualised by Southern blot. However, the increase in topologically bound cohesin may too small to detect.

It has been shown that Scc2 binds to the N-terminal region of Scc1 and competes with Pds5 for binding (Kikuchi et al., 2016). If the suppressor mutations in Scc2 stabilise Scc2 binding to Scc1 then this could be shown biochemically using purified proteins *in vitro*. Furthermore, the suppressor should also be able to displace Pds5 better than the WT Scc2, which also could be tested.

Another approach could be microscopy. If the dissociation rate of Scc2 is decreased the residence time of Scc2 at the DNA should be increased which could be measured by FRAP. Although this would be by far the quickest approach to determine strength of the hypothesis, a biochemical approach should not be ruled out as it could provide many more insights into the loading reaction.

Mutations in histone proteins H2A and H2B are able to bypass the requirement for Scc4

Introduction

The screen preformed by Dr. M Demidova to identify suppressors of the temperature sensitive phenotype of *scc4-4* led to the identification of the first genetic link between the loading reaction and the hinge domain of the cohesin ring. Subsequent screens detailed in Chapter 2 identified mutations in Scc2 that enable it to function in the absence of Scc4. Both of these screens also produced another class of mutations. These were identified as mutations in histone proteins H2A and H2B.

Nucleosome remodeling by the RSC complex has been suggested to play a key role of the loading reaction (Lopez-Serra et al., 2014). Namely, that the RSC complex catalyses the remodeling of nucleosomes to create a chromatin landscape conducive for cohesin loading. Furthermore, the study suggested that the loading complex has dual roles in catalysing the cohesin loading reaction and cooperating with the RSC complex to maintain nucleosome depletion at cohesin loading sites with the latter role most likely being performed by Scc4. This immediately raises the possibility that suppressor mutations in histone proteins are able to bypass the requirement for Scc4 by removing the requirement for the RSC complex.

This chapter aims to characterise the mutations in the histone proteins and investigate the potential link to nucleosome remodeling by the RSC complex.

Results

Mutations in histone proteins H2A and H2A are able to suppress the temperature sensitive phenotype of scc4-4.

As well as those identified in the original screen performed by Dr. M. Demidova, 50 independent suppressors across both subsequent screens, were found to contain mutations within *HTA1* (one of the two genes encoding the histone protein H2A). A single suppressor was also identified to contain a mutation in *HTB1* (one of the two genes encoding histone protein H2B). The mutations in H2A affected three residues, G30D, R31I/T/S/G, and R34I, and the mutation in H2B was identified as Y44D. The frequency of each mutation is detailed in the table below. All of the mutations permitted growth in the absence of Scc4 to varying extents with the mutation *hta1(R31I)* exhibiting the best growth (Figure V-1A). However, all of the mutations have a large growth defect and do not rescue the deletion of Scc4 as well as either *smc1(D588Y)* or *scc2(E882K, L937F)*.

Protein	Mutation	Number of times isolated
H2A	G30D	7
H2A	R31I	16
H2A	R31T	10
H2A	R31S	6
H2A	R31G	2
H2A	R34I	8
H2B	Y44D	1

All of the H2A mutations are located in a single patch of the histone surface that interacts with DNA and the single H2B mutation is also located

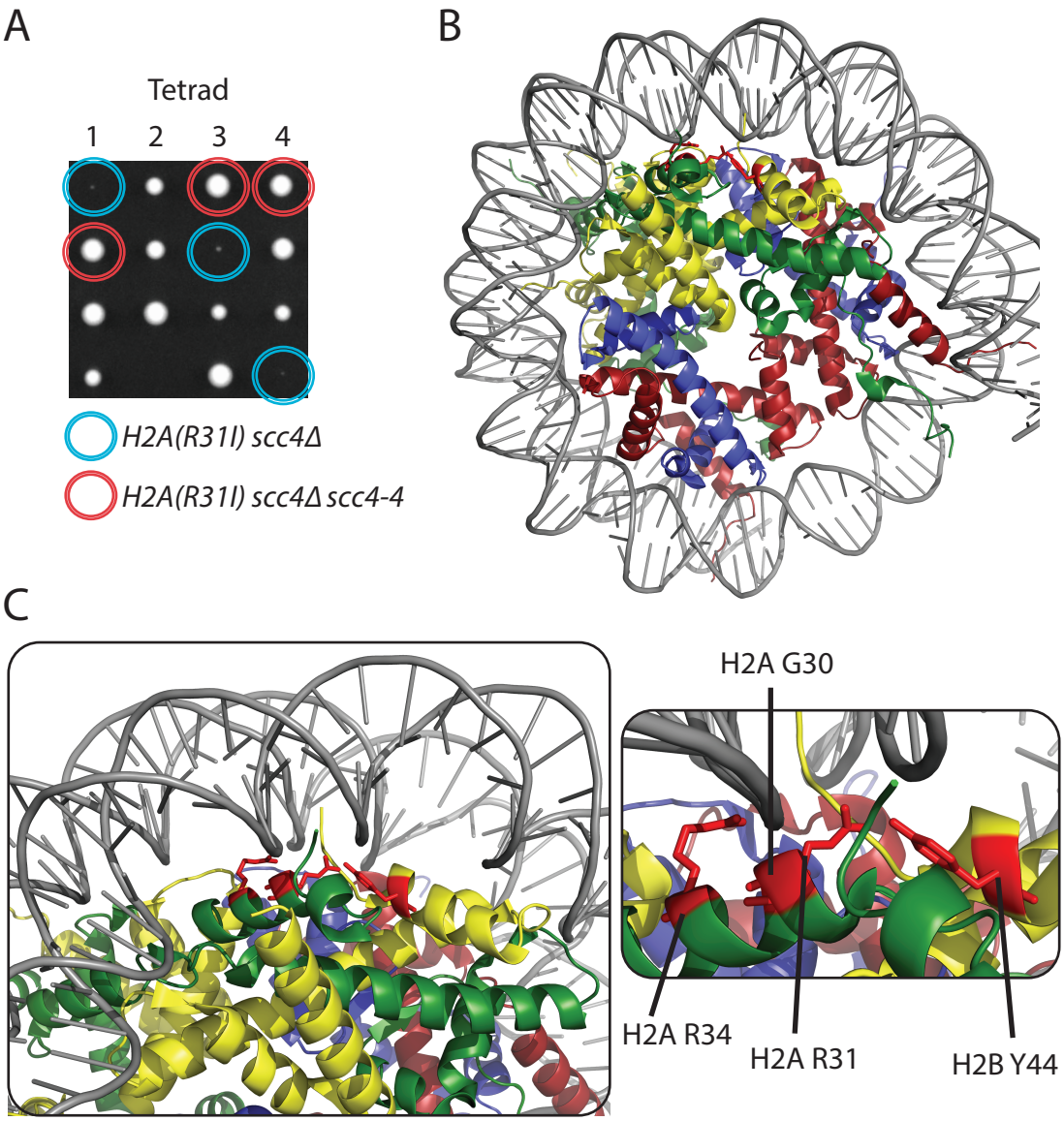


Figure V-1

Suppressor mutations in histone proteins H2A and H2B are able to rescue *scc4Δ*.

A- Tetrad dissection of diploid strains containing *SCC4/scc4Δ leu2/scc4-4* with *HTA1/hta1(R31I)* . Spores in which scc4Δ is rescued by the endogenous H2A mutant are circled in blue. The dissection was performed on YPD plates and were allowed to grow at 25°C for 3 days.

B - Structure of the yeast nucleosome (PDB 1ID3). H2A is shown in green and H2B in yellow. Suppressor mutations are shown in red.

C - Views of the region containing the syppressor mutations in greater detail.

nearby (Figure V-1B and 1C) (White et al., 2001). Although this could indicate that the role of these mutations is to weaken the interaction of the nucleosome with DNA, such a generic effect would suggest that mutations in other histone proteins that also weaken the interaction with DNA should have a similar effect. Given that only a specific patch gave rise to suppressors, weakening the interaction with DNA seems like an unlikely explanation.

To assess the effect of these mutations on cohesin loading, the healthiest suppressor, *hta1(R31I)*, was analysed by quantitative ChIP-seq in the presence and absence of *scc4-4*. A lower restrictive temperature (35.5°C) was necessary as *hta1(R31I)* cells are lethal at 37°C. Cells were arrested in G_1 before release into media pre-warmed to 35.5°C and samples taken when all strains had entered G_2 at 60 min post-release (Figure V-2B). The profile of cohesin occupancy across chromosome I showed that *hta1(R31I)* in the absence of Scc4, is unable to restore loading at the centromere but does produce a modest increase in cohesin loading on the arms (Figure V-2A). Furthermore, *hta1(R31)* itself does not appear to affect cohesin loading.

To analyse better the effect around the centromere, an average chromosome profile was generated for 10 kb region either side of CDEIII (Figure V-3A). This showed that *hta1(R31I)* has little, if any, effect at loading at the centromere in the presence of Scc4. In the absence of Scc4, *hta1(R31I)* was unable to restore loading indicating that the histone mutations are unable to bypass the role of Scc4 in centromeric loading. To quantify the effect of *hta1(R31)* at the centromere and at the arms, an average profile was generated for a region 60 kb either side of CDEIII and plotted as a percentage of the WT occupancy (Figure V-3B). As seen previously, suppression at the centromere was

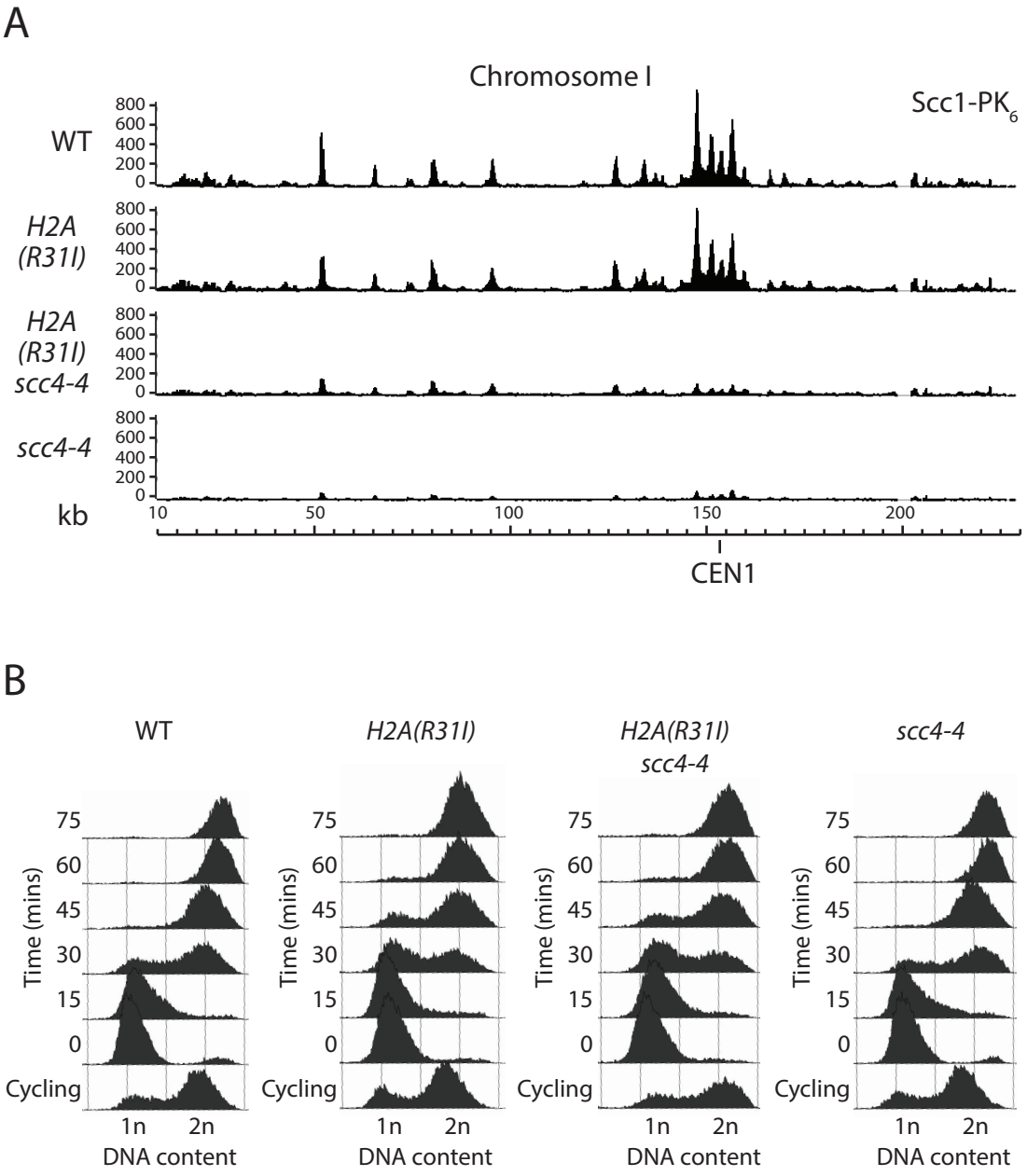


Figure V-2

***hta1(R31I)* suppresses the loading defect of *scc4-4* on chromosome arms but not at the centromere.**

Strains expressing *SCC1-PK$_6$* were arrestedin G_1 with alpha factor and released into nocodazole containing medium, pre-warmed to 35.5ºC. Samples were taken at 60 min post-release.

A - Quantitative ChIP-seq profile across chromosome I. The y axis indicates the calibrated number of reads and the x axis, the base pair position across chromosome I. (K22005, K24574, K24568 and K22001 from top to bottom)

B - Cell cycle progression during the xperiment shown in A as measured by FACS.

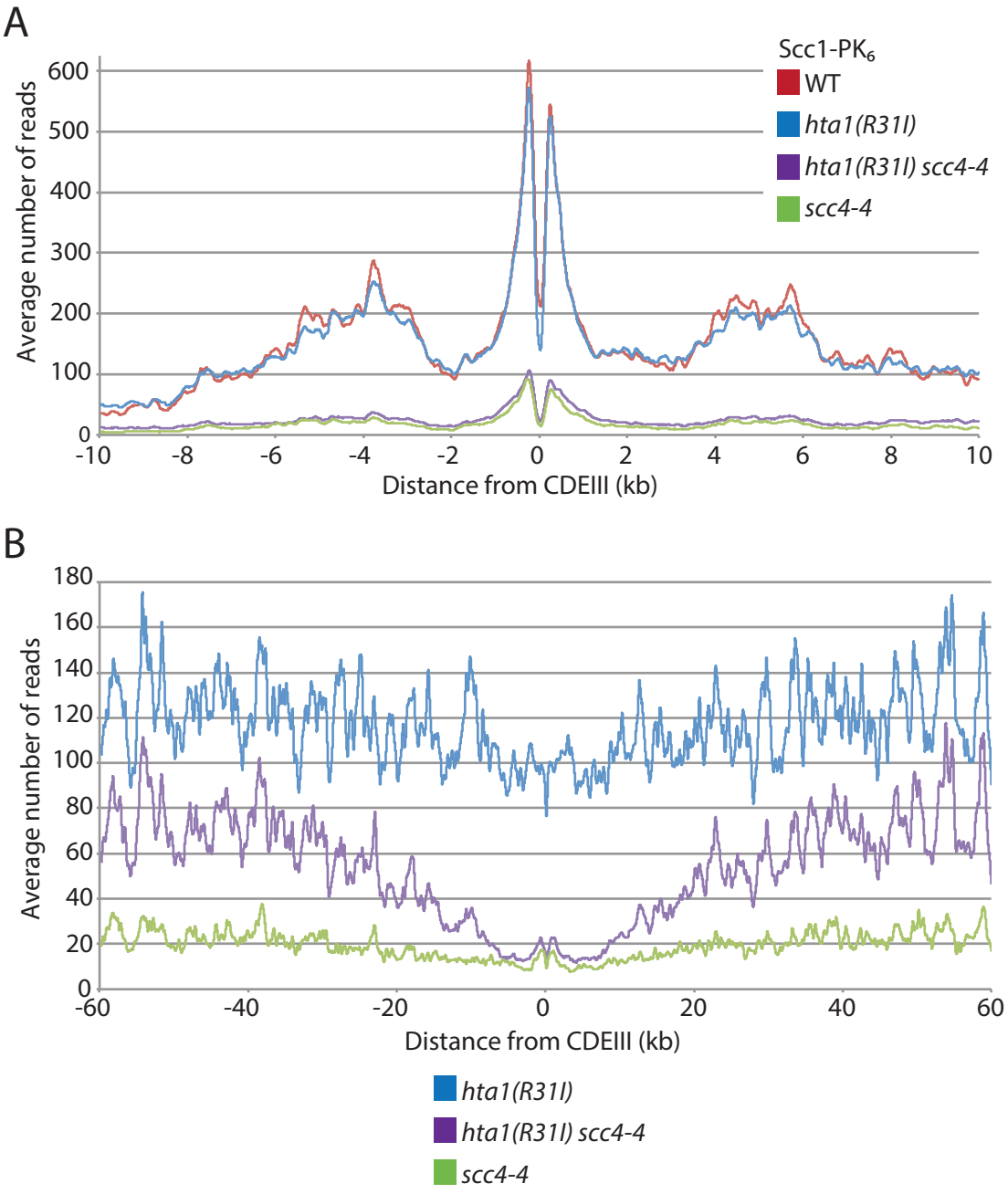


Figure V-3

***hta1(R31I)* restores cohesin loading in cells containing *scc4-4* on chromosome arms but not around centromeres.**

A - Calibrated average centromere profile of the experiement detailed in figure V-2.

B - Calibrated average chromosome plotted as a percentage of the number of reads obtained for WT cells.

minimal but at the arms, cohesin levels were raised to 60-80% of WT levels. This shows that the histone mutations suppress the *scc4-4* by increasing loading along chromosome arms. The fact that the suppression is reduced compared to that seen with either the hinge mutation or those in Scc2 might explain why the histone suppressors are unable to proliferate as well as the other suppressors. *hta1(R31I)* itself had very little effect on cohesin loading at the centromere but an increase of approximately 20% above WT levels on chromosome arms. This implies that nucleosomes do have an effect on cohesin loading at least on chromosome arms. Since the histone mutation does not reduce loading at the centromeres, unlike the hinge mutation, it is likely that *hta1(R31I)* and *smc1(D588Y)* affect different aspects of the loading reaction.

Inactivation of the RSC complex has little or no effect on cohesin loading.

As previously mentioned, it has been suggested that the RSC chromatin remodeling complex functions in the loading reaction to prepare a chromatin landscape that is nucleosome free to enable cohesin loading and that this is facilitated by the loading complex (Lopez-Serra et al., 2014). This model could explain how mutations in histone proteins could suppress the loading defects of *scc4-4* by removing the requirement for RSC.

sth1-3 is a temperature sensitive allele of the ATPase subunit of the RSC complex and has been reported to cause a marked loss of cohesin at both centromeres and along chromosome arms (Lopez-Serra et al., 2014). However, other studies have reported partial or even no effect of the inactivation of the RSC complex on cohesin loading (Baetz, Krogan, Emili, Greenblatt, & Hieter,

2004; Huang & Laurent, 2004). To verify the potential role of the RSC complex in cohesin loading, quantitative ChIP-seq was used to re-examine the loading defects previously observed in *sth1-3* cells.

Strains expressing *SCC1-PK$_6$* and either WT *STH1* or *sth1-3* were arrested in G_1 with α-factor before release into media containing nocodazole, pre-warmed to 37°C. Cells expressing *sth1-3* were delayed in both DNA replication and budding so samples were taken from both strains at 105 min post-release (Figure V-4A and 4B). At this time point most *sth1-3* cells had budded and undergone DNA replication whereas WT cells had reached this point much earlier. To get WT cells in a comparable cell cycle state to the *sth1-3* cells at 105 min, samples were also taken at 75 min post-release. A western blot to detect the levels of Scc1 at each time point also confirmed these time points to be similar (Figure V-4C). The profile of cohesin occupancy across chromosome I showed the cohesin occupancy in WT cells at 75 min to be very similar to that in *sth1-3* cells at 105 min with possibly a small decrease along chromosome arms (Figure V-4D). As expected, *sth1-3* cells at 75 min showed significantly less cohesin loading due to cell cycle delay. Similarly, WT cells at 105 min showed increased cohesin occupancy due to a longer time frame for cohesin loading to occur. On inspection of an average chromosome profile covering 60 kb either side of CDEIII, *sth1-3* cells show modest decrease in cohesin occupancy but otherwise a very similar profile to WT cells (Figure V-5A). To check that the cohesin occupancy observed in each strain is similar genome-wide, the number of reads at each base pair from one strain was plotted against the number of reads at the corresponding base pair in the other strain (Figure V-5B). As the varying number of copies of the ribosomal DNA cannot be accounted for, the

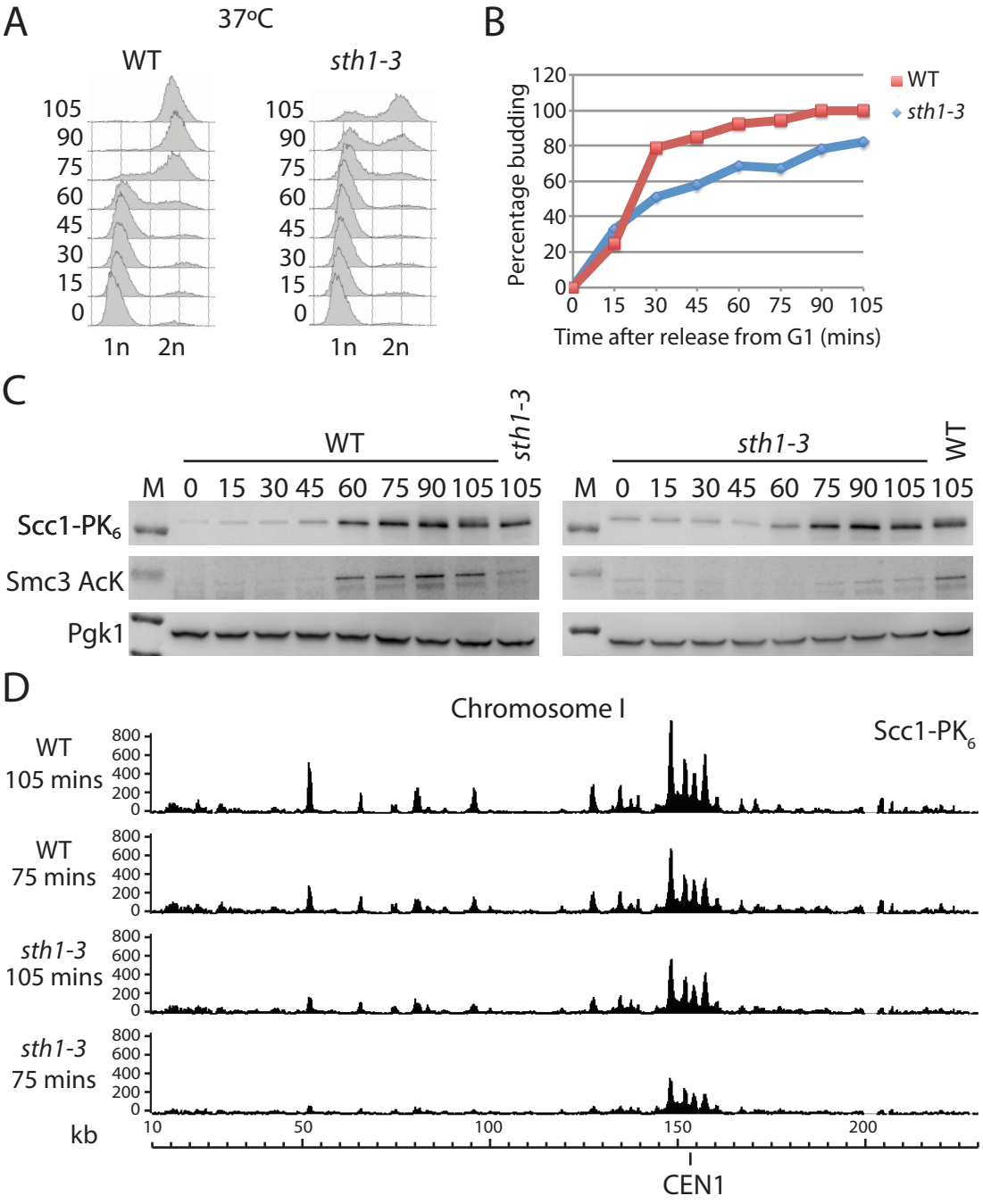


Figure V-4

Cells expressing *sth1-3* exhibit cell cycle delay.

Cells expressing *SCC1-PK$_6$* in the presence or absence of *sth1-3* (K23997 and K22005 respectively) were arrested in G_1 with alpha factor before release into nocodazole containing medium pre-warmed to 37°C. Samples were taken at 75 min and 105 min post-release.

A - Cell cycle progression as measured by FACS

B - The fraction of cells with buds (budding index)

C - Western blot to measure the levels of Scc1-PK$_6$ and acetylation of Smc3 at each time point.

D - Quantitative ChIP-seq profile across chromosome I comparing WT cells to cells expressing *sth1-3* at time points 75 mins and 105 mins post-release from G_1.

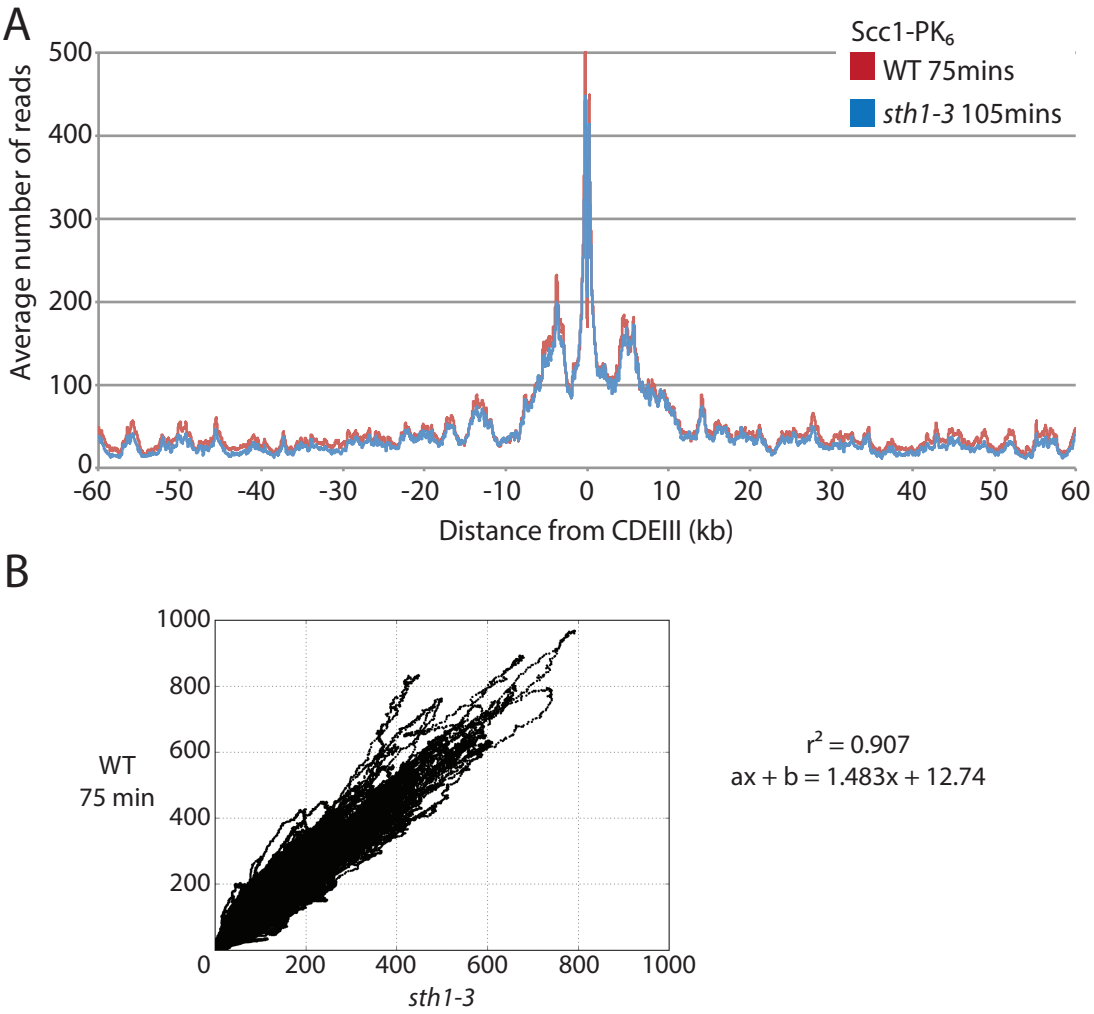


Figure V-5

Cohesin occupancy is unaffected by inactivation of the RSC complex.

A - Calibrated average chromosome profile covering 60 kb either side of CDEIII of the experiment detailed in figure V-4.

B - Scatterplot comparing the calibrated number of reads at each base pair along the chromosomes (with exception of chromosome XII) between WT at 75 min and *sth1-3* at 105 min post-release.

ribosomal DNA was excluded. The resulting scatterplot showed a very close correlation between WT and *sth1-3* cells fitting a curve with a gradient close to 1 with an r^2 of 0.907. Altogether this suggest that, if there is an effect of the RSC complex on cohesin loading, it is very modest or may not be so easily observed on a genome-wide scale.

Inactivation of the RSC complex or the cohesin loading complex has no effect on nucleosome occupancy.

Inactivation of Sth1, Scc2 and Scc4 was also claimed to affect nucleosome positioning at specific sites including the promoter region of *RPS8B* (Lopez-Serra et al., 2014). To see if the effect of the RSC complex can be observed at specific regions, the nucleosome occupancy was determined genome-wide in cells expressing WT or *sth1-3* at the restrictive temperature (37°C) by micrococcal nuclease sequencing (MNase-seq). Overall the nucleosome occupancy was very similar between the two strains and the previously reported increase in nucleosome occupancy at the promoter region of *RPS8B* could not be replicated (Figure V-6A). The same approach was also used to investigate the effect of the loading complex of nucleosome occupancy. Both *scc2-4* and *scc4-4* alleles were tested at the restrictive temperature but neither demonstrated the increase in nucleosome occupancy at the promoter region of *RPS8B* or any other large scale nucleosome remodeling (Figure V-6B and 6C). Therefore the claim that inactivation of Sth1, Scc2 or Scc4 increases nucleosome occupancy at the promoter region of *RPS8B* could not be replicated. This suggests that the RSC complex and the cohesin loading complex do not work together to promote a nucleosome free region that is conducive to cohesin loading.

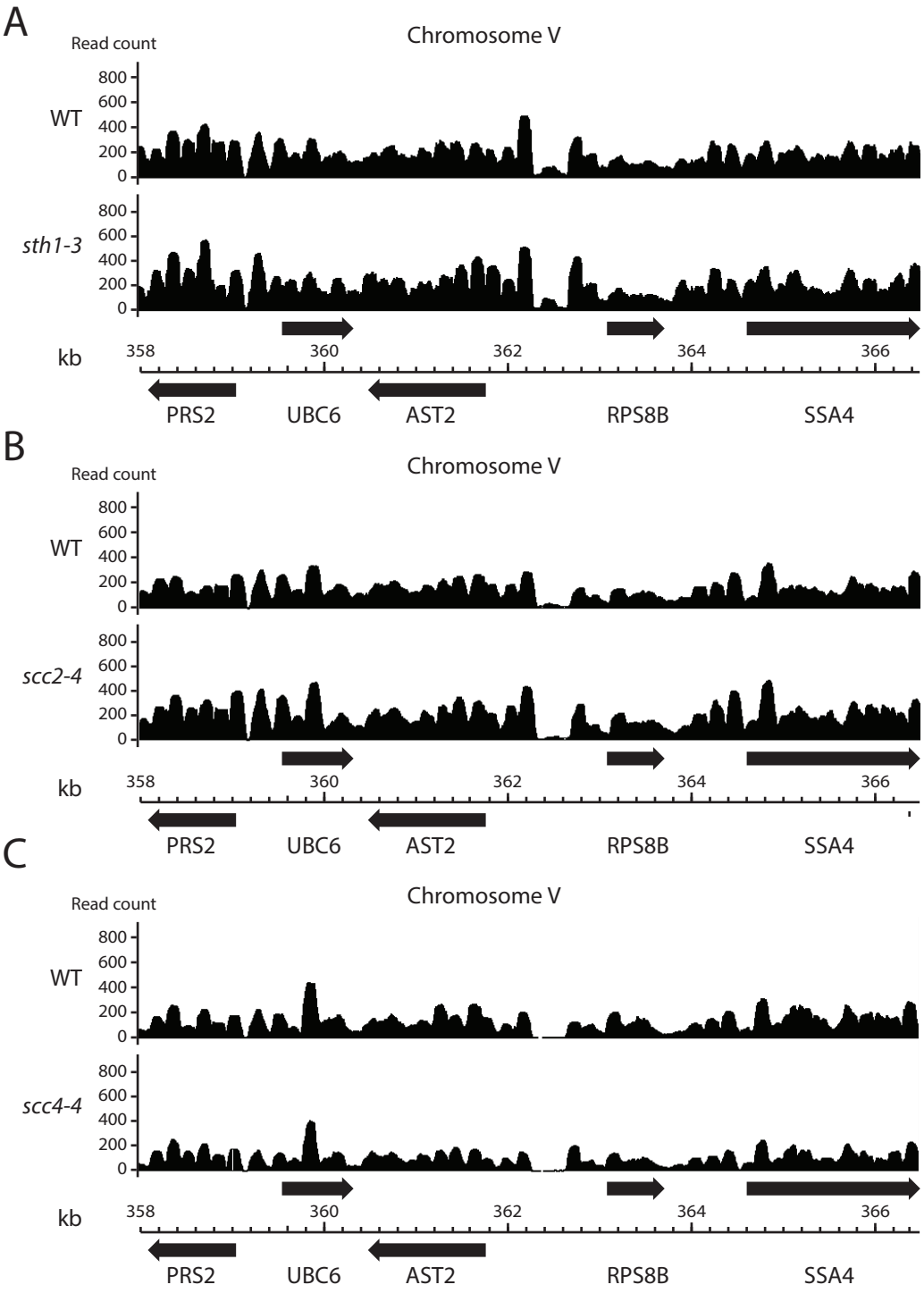


Figure V-6

***sth1-3, scc2-4* and *scc4-4* do not effect nucleosome occupancy.**

Logarithmic phase cells were shifted to 37°C for 2 hours prior to fixation, MNase digestion and sequencing.

A -C - Comparison of nucleosome occupancy across part of chromosome V between **A** - WT and *sth1-3,* **B** - WT and *scc2-4*, and **C** - WT and *scc4-4* (K699, K24032, K5828 and K8326).

Discussion

The finding that the effect of the loading complex on nucleosome occupancy could not be reproduced calls into question the proposed model for a dual role of the loading complex in catalysing the loading reaction as well as altering the chromatin landscape to promote cohesin loading. Although the data doesn't exclude the possibility that RSC has a modest effect on cohesin loading and this may be due to small changes in nucleosome occupancy, it is inconsistent with the hypothesis that one of the main roles of the loading complex is to promote the reorganisation of nucleosomes through the RSC complex to promote cohesin loading.

The discovery of several mutations in histone proteins that are able to rescue the loss of Scc4, clearly indicates an important role of nucleosomes in cohesin loading at least on chromosome arms. This could be due to the previously claimed role of the loading complex in altering nucleosome occupancy in addition to its role in the loading mechanism. Alternatively, the loading complex may have no effect on nucleosome occupancy at all, but perhaps nucleosomes inhibit the loading reaction, which is weakened but not eliminated in Scc4 mutants. Therefore, the histone suppressor mutations may alter the nucleosome so that it no longer inhibits the loading reaction making it possible for Scc2 to load cohesin onto chromosome arms in the absence of Scc4. It is conceivable that cohesin loading occurs genome-wide, but the chromatin landscape can create regions that increase or decrease the likelihood of cohesin loading in that area. Thus, mutations that globally increase the number of regions that promote cohesin loading would enable more loading to take place

just by removal of obstacles. This could also explain the modest effect of the inactivation of RSC on cohesin loading. It is, however, odd that only a small region of the whole nucleosome gave rise to suppressors if such a generic effect is enough to rescue lethality caused by the absence of functional Scc4. It would be expected that any mutation that would weaken the binding of the nucleosome to DNA would be able to restore viability in cells lacking Scc4. Nevertheless, it should also be noted that all of the suppressors found in histone proteins grew very slowly in the absence of Scc4, suggesting that they do not suppress the lack of Scc4 particularly well. Perhaps then it is only those mutations that had the greatest effect on nucleosome binding that produced enough of an effect genome-wide to restore viability. Alternatively, there must be something specific to this region of the nucleosome, possibly a binding site, that effects nucleosome occupancy or cohesin loading more directly that remains to be identified. It has been found that Spt16, a chaperone of the FACT nucleosome remodeling complex binds to the region of H2A and H2B that the suppressor mutations are located (Hondele et al., 2013). This suggests that these suppressor mutations may act to help FACT to destabilise the nucleosome binding to DNA and therefore aid in nucleosome removal which could indirectly create sites that favour cohesin loading.

Materials and methods

Yeast strains and growth conditions.

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

To arrest the cells in G_1, α-factor (Bücking-Throm, Duntze, Hartwell, & Manney, 1973) was added to a final concentration of 2 mg/L/h, every 30 min for 2.5 h. Release was achieved by filtration wherein cells were captured on 1.2 µm filtration paper (Whatman® GE Healthcare), washed with 1 L YPD and resuspended in the appropriate fresh media. To arrest the cells in G_2, nocodazole (Sigma) was added to the fresh media to a final concentration of 10 µg/mL and cells were incubated until the synchronization was achieved (>95% large-budded cells). To inactivate temperature sensitive alleles, fresh media was pre-warmed prior to filtration (Aquatron, Infors HT).

To produce cells deficient of Scc1, the gene was placed under the *MET3*-repressible promoter. Liquid cultures were grown in minimal media supplemented with 2 % glucose and 1 % -MET dropout solution overnight, diluted to OD_{600}=0.2 and allowed to grow to OD_{600}=0.4. Cells were then collected by filtration as described above, resuspended in YPD supplemented with 8 mM methionine and arrested in G1. Once arrested, the cells were collected by filtration, washed with YPD in the presence of 8 mM methionine and released into the same media.

Fluorescence-activated cell sorting.

1 mL of culture was harvested by centrifugation at 13,000 rpm for 1 min. The pellet was resuspended in 1 mL of 70 % (v/v) ethanol and stored at 4 °C. The samples were centrifuged at 7,000 rpm for 5 min and the cell pellet was then resuspended in 1 mL 50 mM Tris-HCl, pH 7.5 containing 200 µg/mL RNase A (Roche) and incubated at 37 °C at 950 rpm overnight. The cells were then collected by centrifugation at 7,000 rpm for 5 min and resuspended in 400 µL of FACS buffer (200 mM Tris-HCl pH 7.5; 211 mM NaCl; 78 mM $MgCl_2$; 50 µg/mL propidium iodide).

To ensure the cells were dispersed, samples were sonicated at 40 % output for 5 s (Vibra-Cell™ VCX 130FSJ, Sonics & Materials Inc.). Cells were then diluted 1:200 into 1 mL of 50 mM Tris-HCl, pH 7.5 in FACS tubes and 10,000 events per sample was measured using a FACSCalibur™ flow cytometer (BD Biosciences).

Site-directed mutagenesis.

Primers were designed such that there were 9-12 bases before the mutation followed by up to 32 bases trimmed to give an annealing temperature of 65 °C. The primers were used with Phusion DNA polymerase (New England Biolabs) in a mutagenesis PCR mix according to the manufacturer's instructions and subjected to the thermal cycling conditions below.

Cycling conditions

98 °C for 2 min

98 °C for 30 s

65 °C for 30 s

72 °C for 6 min 30 (or as required)

Repeat for 30 cycles

72 °C for 10 min

4 °C ∞

On completion, 1 µL of Dpn1 (New England Biolabs) was added (37 °C, overnight). 5 µL of the reaction was transformed into competent XL1 Blue *E. coli*. Mutations were confirmed by sequencing.

Yeast transformation.

Salmon sperm DNA (10 mg/mL) was heat-inactivated at 95 °C for 5 min then kept on ice prior to use. 50 mL of cells growing exponentially at approximately OD_{600}=0.5 were harvested at 3,500 rpm for 3 min. Cells were washed twice with 0.1 M LiOAc before being resuspended in 200 µL 0.1 M LiOAc and incubated for 30 min at room temperature (RT). 50 µL of this was used per reaction with 5 µL of DNA (200 ng to 2 µg), 240 µL 50 % (w/v) PEG2000, 36 µL 1 M LiOAc, 25 µL salmon sperm DNA and 45 µL water. Cells were then incubated for 30 min at RT, 20 min at 42 °C, then collected by centrifugation at 6,000 rpm for 3 min, resuspended in 8 mL YPD and allowed to recover for 3 h at 25 °C. After recovery cells were again collected by centrifugation at 3500 rpm for 3 min and

resuspended in 100 µL water before plating on selective dropout plates or in the case of a drug resistance cassette, onto YPD plates containing the appropriate antibiotic.

Genomic integration.

To integrate mutant genes at the endogenous loci, constructs that contained the gene of interest with a selection marker between the homology arms that flank the desired insertion site were used. These were modified as required using site directed mutagenesis then again digested with an appropriate restriction endonuclease prior to transformation as described above. Correct integration was confirmed by PCR on isolated genomic DNA using primer sets that flank the insertion site and check for the presence of the mutations introduced.

Genomic DNA preparation from yeast.

A culture of exponentially growing cells was harvested by centrifugation at 3,500 rpm for 3 min. The pellet was washed with ddH_2O before being resuspended in 200 µL of SCE buffer (1 M sorbitol; 100 mM sodium citrate; 60 mM EDTA).

50 µL of zymolase 20T (20 mg/mL; Amsbio) and 1.6 µL of β-mercaptoethanol were added and the cells incubated (37 °C, 950 rpm, 30 min). Cells were then centrifuged at 8,000 rpm for 30 s and resuspended in 200 µL of SDS lysis buffer (2 % SDS (w/v); 100 mM Tris-HCl, pH 8.8; 50 mM EDTA). The mixture was incubated for 5 min at 65 °C before 200 µL of 5 M KOAc was added

to precipitate the cell debris. The sample was centrifuged at 13,200 rpm for 10 min and the supernatant transferred into 400 µL of ice-cold 70 % (v/v) ethanol. The pellet was air dried before resuspension in 200 µL TE buffer, pH 7.5.

Tetrad dissection.

Diploid strains were patched onto nitrogen-deficient medium agar plates to induce meiosis. Sporulation was assessed microscopically and when tetrads could be seen (after approximately 48 h at 25 °C) cells were spheroplasted in 300 µL tetrad dissection buffer (1 M sorbitol; 1.5 mg/mL zymolase® 20T (Amsbio)) for 20 min at RT. 10 µL of the suspension was transferred onto an appropriate plate and dissected on a dissection microscope (MSM 300, Singer Instruments).

Screening for suppressors of scc4-4.

Forty independent colonies of the parental strain were picked and grown overnight at 25 °C. Each was plated at 5 OD_{600} units per plate over 3 plates and incubated at 35.5 °C until colonies appeared. Up to 3 colonies were picked from each plate and streaked for single colonies at 25 °C before being retested for growth at 35.5 °C. Those that grew at 35.5 °C were checked by PCR from genomic DNA preparations for revertants of Scc4. Isolated suppressors that did not show revertant mutations were checked for 2:2 segregation and grouped into complementation groups prior to deep sequencing.

Preparation of native yeast cell extracts.

Cells grown in exponential phase in liquid culture were harvested by centrifugation 3,500 rpm for 3 min and washed with 1X PBS. The pellet was then resuspended in 200 µL of lysis buffer (150 mM NaCl; 50 mM Tris-HCl, pH 7.5; 0.5 % (v/v) NP-40; 5 mM EDTA; 1.25 mM DTT; 1 mM PMSF; 1 tablet/50 mL protease inhibitor cocktail (Roche)).

100 µL of acid-washed glass beads (425-600 µm Sigma) was added and the cells were lysed using a FastPrep®-24 benchtop homogeniser (M.P. Biomedicals) at 4 °C (3x 60 s at 6.5 m/s or until >90 % of the cells were lysed as confirmed by microscopy). The insoluble fraction was pelleted by centrifugation at 13,200 rpm for 15 min and the soluble fraction isolated.

Quantification of protein concentration in crude extracts.

The protein concentration of crude extracts was determined using the Bradford protein assay (Bradford, 1976). Briefly, Bradford® reagent (dye concentrate, Bio-Rad) was diluted 1:4 and added to diluted samples to a total volume of 1 mL. Serially diluted BSA was used to generate a standard curve. Absorbance was measured at 595 nm for each sample relative to a blank containing only lysis buffer using an Ultraspec 3 300 pro spectrophotometer (Amersham Biosciences) and the concentration calculated from the standard curve produced.

Protein gel electrophoresis and Western blotting.

The samples were mixed with 4X LDS sample buffer (NuPAGE® Life Technologies), loaded onto 3-8 % Tris-acetate gels (NuPAGE® Life Technologies) and the proteins separated using an appropriate current. The proteins were then transferred onto 0.2 µm nitrocellulose using Trans-blot® Turbo™ transfer packs for the Trans-blot® Turbo™ system (Bio-Rad). The membrane was then blocked with 5 % milk powder (w/v) in PBST for 30 min. Incubation with primary antibody was done to the dilutions in the table below in 5 % milk powder + PBST for 1.5 h at RT or overnight at 4 °C.

Primary antibody	Immunogen	Origin	Dilution
V5 (Bio-Rad)	PK	Mouse	1:1,000
3F10 (Roche)	HA	Rat	1:5,000
PGK1	PGK1	Mouse	1:5,000

The membrane was washed with PBST 3x for 5 min, then secondary antibody added using the dilutions in the table below in 5 % milk powder + PBST and incubated for 45 min at RT.

Secondary antibody	Dilution
Goat anti-mouse (Millipore)	1:5,000
Goat anti-rat (Millipore)	1:5,000

Again the membrane was washed with PBST 3x for 5 min and incubated with Immobilon™ Western Chemiluminescent HRP substrate (Millipore) before detection using an ODYSSEY® Fc Imaging System (LI-COR).

In the case of probing for acetylated Smc3, after transfer the membrane was washed 3x for 10 min in Buffer A (100 mM Tris-HCl, pH 7.4; 150 mM NaCl). The membrane was blocked with 5 % milk in Buffer A at 4 °C overnight, washed for 10 min in 5 % milk in Buffer A-T (1X Buffer A + 0.05 % TWEEN-20) and incubated with anti-Smc3AcK (1:1,000) in Buffer A-T for 3 h at RT. The membrane was washed for 10 min in Buffer A and 15 min in Buffer A-T before incubation with secondary antibody anti-rabbit HRP (1:20,000) in 5 % milk in Buffer A-T for 1 h at RT. Finally, the membrane was washed for 10 min in Buffer A, 15 min in Buffer A-T and 15 min in Buffer A before incubation with Immobilon™ Western Chemiluminescent HRP substrate (Millipore) and detection using an ODYSSEY® Fc Imaging System (LI-COR).

Multiple sequence alignment.

Multiple sequence alignments were created using Clustal Omega (Sievers et al., 2011).

Quantitative ChIP-seq.

Cells were grown exponentially to OD_{600}=0.5 and the required cell cycle stage where necessary. 15 OD_{600} units of *S. cerevisiae* cells were then mixed with 5 OD_{600} 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 min at RT with rotation.

Fixation was quenched with 2 mL of 2.5 M glycine incubated at RT for 5 min with rotation. The cells were then harvested by centrifugation at 3,500 rpm for 3 min and washed with ice-cold 1X 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; 1 tablet/25 mL 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 (M.P. Biomedicals) at 4 °C (3x 60s 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 min then sonicated using a bioruptor (Diagenode) for 30 min in bursts of 30 s on and 30 s off at high level in a 4 °C water bath to produce sheared chromatin with a size range of 200-1,000 bp. After sonication the samples were centrifuged at 13,200 rpm at 4 °C for 20 min 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 h at 4 °C. 80 µL of the supernatant was taken as the WCE and 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 h 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 pH 7.5. The immunoprecipitated chromatin was then eluted by incubation in 120 µL of 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 the IP sample. The WCE extracts 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 at 65 °C overnight. RNA was degraded by incubation with 2 µL RNase A (10 mg/mL; Roche) for 1 h at 37 °C and protein was removed by incubation with 10 µL of proteinase K (18 mg/mL; Roche) for 2 h at 65° C. DNA was purified using ChIP DNA Clean and Concentrator kit (Zymo Research).

Extraction of yeast DNA for deep sequencing.

Cultures were grown to exponential phase (OD_{600}=0.5). 12.5 OD_{600} units were then collected and diluted to a final volume of 45 mL before fixation as described in the protocol for ChIP-seq. The samples were treated as specified in the ChIP-seq protocol up to the completion of the sonication step whereby 80 µL of the samples were carried forward and treated as WCE samples.

MNase-seq.

Exponential phase cultures of 1 L were grown to OD_{600}=0.25. Cells were crosslinked by addition of 27 mL/L formaldehyde 37 % and incubated at RT for 10 min. The crosslinking reaction was quenched by addition of glycine to a final concentration of 50 mL/L and incubated for 5 min at RT. Cells were collected by centrifugation at 3,500 rpm for 5 min and washed 3x in ice-cold TBS. 10^{10} cells

were then collected, resuspended in 10mL spheroplasting buffer (4.77 mL ddH_2O; 5 mL Sorbitol (2 M); 200 µL zymolase T100 (10 mg/mL); 30 µL β-mercaptoethanol) and incubated at RT for 25 min with rotation. Spheroplasts were then collected by centrifugation at 5,000 rpm for 10 min and gently resuspended in 10 mL SDB buffer (1 M sorbitol; 50 mM NaCl; 10 mM Tris-HCl, pH 8.0; 5 mM $MgCl_2$; 1 mM $CaCl_2$; 1 mM β-mercaptoethanol; 0.075 % NP-40).

MNase was diluted 1:6 to give 2.5 U/µL, and serial aliquots from 30 µL to 135 µL were made. 1 mL of cells in SDB buffer was added to each MNase aliquot and incubated at 37 °C for exactly 10 min. Stop buffer (5% SDS, 0.25 M EDTA) warmed to 65 °C (100 µL) was then added to stop the digestion. Samples were incubated at 65 °C overnight to reverse the crosslinks, followed by incubations with 50 µL RNase A for 45 min at 42 °C and 100 µL proteinase K at 37 °C for 2 h. The DNA was then purified with ChIP DNA Clean and Concentrator kits (Zymo Research). 500 µg was then loaded onto E-Gel® SizeSelect™ 2 % Agarose gels (Life Technologies) to assess homogeneity and adequate digestion into mononucleosomes. The samples with sufficient digestion were then loaded onto E-Gel® SizeSelect™ 2 % Agarose gels (Life Technologies) and DNAs approximately 146 bp in length were purified.

Preparation of sequencing libraries.

Sequencing libraries were prepared using NEBNext® Fast DNA Library Prep Set for Ion Torrent™ Kit (New England Biolabs) to the manufacturers instructions. To summarise, 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® SizeSelect™ 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.

Unless otherwise specified, data analysis was preformed on the Galaxy platform (Giardine et al., 2005). Quality of the 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) (Langmead & Salzberg, 2012).

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 visualisation. 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 visualised using the IGB browser (Nicol, Helt, Blanchard, Raja, & Loraine, 2009). To normalise the data to show quantitative ChIP signal the track was multiplied by the samples occupancy ratio (OR) and normalised 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) (Li et al., 2009). 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 ‘chr_position.py’ to assign all unrepresented genome positions a value of 0. Each pileup was then filtered using ‘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 normalising to 1 million reads.

Identification of mutations from whole genome sequencing.

SNPs were called using command line on a local server. First a pileup was created using samtools mpileup (-v --skip-indels –f sacCer3.fa –o *sample name*.vcf *sample name*.bam), then SNPs called using bcftools call (-v –c –o *sample name*.bcf *sample name*.vcf). To find mutations unique to a suppressor strain, lists of SNPs from the parental strain or backcrossed clones of the suppressor strain were compared to the list of SNPs from the suppressor strain. In the case of parental strains, mutations that were present in both were removed and in the case of backcrossed clones of the suppressor strain, mutations that were present in both were kept in order to identify the mutation that caused the suppression phenotype. This was done using ‘MutationFinder.py’ and the resulting lists further narrowed using ‘yeastmine.py’ which searches the Saccharomyces Genome Database (SGD) for genes that correspond to the position of each mutation so that those that lie outside of genes could be removed. From this it was possible to identify the mutation in each suppressor that gave rise to the suppressor phenotype.

List of strains.

S. cerevisiae:

K699	*MATa, W303 wildtype*
K5828	*MATa, scc2-4*
K8326	*MATα, scc4Δ::HIS3, leu2::scc4-4::LEU2*
K14601	*MATa, SCC1-PK9::KanMX4*
K14610	*MATa, SCC1-PK9::KanMX4, scc2-4*
K17409	*MATa, trp1::smc3(E1155Q)PK6::TRP1*
K17438	*MATa, SCC3-PK6::KanMX6*
K19012	*MATa, PDS5-PK6::KanMX6*
K19624	*MATa, scc4Δ::HIS3, smc1(D588Y)::HphNT1, SCC1-PK6::KanMX4*
K21388	*MATa, SCC2-PK6::KanMX6*
K21999	*MATa, scc4Δ::HIS3, leu2::scc4-4::LEU2, smc1(D588Y)::HphNT1, SCC1-PK6*
K22001	*MATa, scc4Δ::HIS3, leu2::scc4-4::LEU2, SCC1-PK6::KanMX4*
K22005	*MATa, SCC1-PK6::KanMX4*
K22009	*MATa, SCC1-PK6::KanMX4, smc1(D588Y)::HphNT1*
K22390	*MATa, SCC1-PK9, scc2-45::NatMX4*
K22592	*MATa, sth1-3, smc1(D588Y)::HphNT1*
K22642	*MATa, SCC4-PK6::KanMX4*
K23997	*MATa, sth1-3 SCC1-PK6::KanMX4*
K24032	*MATa, sth1-3*
K24564	*MATa, trp1::smc3(E1155Q)::TRP1, ura3::smc1(D588Y)-myc9::URA3, scc4Δ::HIS3, smc1(D588Y)::HphNT1*
K24568	*MATa, scc4Δ::HIS3, leu2::scc4-4::LEU2, hta1(R31I)::KanMX4, SCC1-PK6::KanMX4*
K24574	*MATa, hta1(R31I)::KanMX4, SCC1-PK6::KanMX4*
K24678	*MATa, scc4Δ::HIS3, smc1(D588Y)::HphNT1, SCC2-PK6::KanMX6*
K24680	*MATa, smc1(D588Y)::HphNT1, SCC2-PK6::KanMX6*
K24687	*MATa, scc2(E822K, L937F)::NatMX4, SCC1-PK6::KanMX4*
K24689	*MATa, trp1::smc3(E1155Q)PK6::TRP1, ura3::smc1(D588Y)myc9::URA3, smc1(D588Y)::HphNT1*
K24744	*MATa, scc2(E822K, L937F)::NatMX4, SCC1-PK6::KanMX4, scc4Δ::HIS3, leu2::scc4-4::LEU2*
K24745	*MATa, scc2(E822K, L937F)::NatMX4, SCC1-PK6::KanMX4, scc4Δ::HIS3*
K24760	*MATa, smc1(D588Y)::HphNT1, scc2-4, SCC1-PK9::KanMX4*
K25053	*MATa, scc2(E822K, L937F)HA6::NatMX4*
K25054	*MATa, SCC2-HA6::NatMX4*
K25063	*MATa, trp1::smc3(E1155Q)PK6::TRP1*
K25065	*MATa, trp1::smc3(E1155Q)PK6::TRP1, scc2(E822K, L937F)::NatMX4*
K25067	*MATa, trp1::smc3(E1155Q)PK6::TRP1, scc2(E822K, L937F)::NatMX4, scc4Δ::HIS3*
K25222	*MATa, scc1::MET-SCC1-myc18::TRP, SCC2-PK6::KanMX6*
K25370	*MATa, leu2::smc3(E1155Q)HA3::LEU2, SCC1-PK6::KanMX6*
K25373	*MATα, leu2::smc3(E1155Q)HA3::LEU2, SCC3-PK6::KanMX6*
K25376	*MATa, leu2::smc3(E1155Q)HA3::LEU2, PDS5-PK6::KanMX6*
K25418	*MATα, scc2(E822K, L937F)HA6::NatMX4, scc4Δ::HIS3, leu2::scc4-4::LEU2*

K25467	*MATa, leu2::smc3(E1155Q)HA3::LEU2, SCC2-PK9::NatMX4*

C. glabrata:

K23308	*MATa, SCC1-PK9::NatMX4*
K25532	*MATa, SCC1-HA3::NatMX4 Gift from Wolfgang Zachariae*

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. http://doi.org/10.1083/jcb.200111002

Anderson, D. E., Trujillo, K. M., Sung, P., & Erickson, H. P. (2001). Structure of the Rad50 x Mre11 DNA repair complex from Saccharomyces cerevisiae by electron microscopy. *The Journal of Biological Chemistry*, *276*(40), 37027–37033. http://doi.org/10.1074/jbc.M106179200

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

Arumugam, P., Gruber, S., Tanaka, K., Haering, C. H., Mechtler, K., & Nasmyth, K. (2003). ATP Hydrolysis Is Required for Cohesin's Association with Chromosomes. *Current Biology*, *13*(22), 1941–1953. http://doi.org/10.1016/j.cub.2003.10.036

Baetz, K. K., Krogan, N. J., Emili, A., Greenblatt, J., & Hieter, P. (2004). The ctf13-30/CTF13 genomic haploinsufficiency modifier screen identifies the yeast chromatin remodeling complex RSC, which is required for the establishment of sister chromatid cohesion. *Molecular and Cellular Biology*, *24*(3), 1232–1244. http://doi.org/10.1128/MCB.24.3.1232-1244.2003

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., et al. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, *129*(4), 823–837. http://doi.org/10.1016/j.cell.2007.05.009

Beckouët, F., Hu, B., Roig, M. B., Sutani, T., Komata, M., Uluocak, P., et al. (2010). An Smc3 Acetylation Cycle Is Essential for Establishment of Sister Chromatid Cohesion. *Molecular Cell*, *39*(5), 689–699. http://doi.org/10.1016/j.molcel.2010.08.008

Beckouët, F., Srinivasan, M., Roig, M. B., Chan, K.-L., Scheinost, J. C., Batty, P., et al. (2016). Releasing Activity Disengages Cohesin's Smc3/Scc1 Interface in a Process Blocked by Acetylation. *Molecular Cell*, *61*(4), 563–574. http://doi.org/10.1016/j.molcel.2016.01.026

Bernard, P., Drogat, J., Maure, J.-F., Dheur, S., Vaur, S., Genier, S., & Javerzat, J.-P. (2006). A screen for cohesion mutants uncovers Ssl3, the fission yeast counterpart of the cohesin loading factor Scc4. *Current Biology : CB*, *16*(9), 875–881. http://doi.org/10.1016/j.cub.2006.03.037

Birkenbihl, R. P., & Subramani, S. (1992). Cloning and characterization of rad21 an essential gene of Schizosaccharomyces pombe involved in DNA double-strand-break repair. *Nucleic Acids Research*, *20*(24), 6605–6611.

Blat, Y., & Kleckner, N. (1999). Cohesins bind to preferential sites along yeast chromosome III, with differential regulation along arms versus the centric region. *Cell*, *98*(2), 249–259.

Bonhoure, N., Bounova, G., Bernasconi, D., Praz, V., Lammers, F., Canella, D., et al. (2014). Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization. *Genome Research*, *24*(7), 1157–1168. http://doi.org/10.1101/gr.168260.113

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of

microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, *72*, 248–254.
Bücking-Throm, E., Duntze, W., Hartwell, L. H., & Manney, T. R. (1973). Reversible arrest of haploid yeast cells in the initiation of DNA synthesis by a diffusible sex factor. *Experimental Cell Research*, *76*(1), 99–110.
Chan, K.-L., Roig, M. B., Hu, B., Beckouët, F., Metson, J., & Nasmyth, K. (2012). Cohesin's DNA Exit Gate Is Distinct from Its Entrance Gate and Is Regulated by Acetylation. *Cell*, *150*(5), 961–974. http://doi.org/10.1016/j.cell.2012.07.028
Chao, W. C. H., Murayama, Y., Muñoz, S., Costa, A., Uhlmann, F., & Singleton, M. R. (2015). Structural Studies Reveal the Functional Modularity of the Scc2-Scc4 Cohesin Loader. *Cell Reports*, *12*(5), 719–725. http://doi.org/10.1016/j.celrep.2015.06.071
Charbin, A., Bouchoux, C., & Uhlmann, F. (2014). Condensin aids sister chromatid decatenation by topoisomerase II. *Nucleic Acids Research*, *42*(1), 340–348. http://doi.org/10.1093/nar/gkt882
Cho, U.-S., Corbett, K. D., Al-Bassam, J., Bellizzi, J. J., De Wulf, P., Espelin, C. W., et al. (2010). Molecular structures and interactions in the yeast kinetochore. *Cold Spring Harbor Symposia on Quantitative Biology*, *75*(0), 395–401. http://doi.org/10.1101/sqb.2010.75.040
Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Tóth, 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.
Cobbe, N., & Heck, M. M. S. (2006). The evolution of ATPase activity in SMC proteins. *Proteins*, *63*(3), 685–696. http://doi.org/10.1002/prot.20795
Cuylen, S., Metz, J., & Haering, C. H. (2011). Condensin structures chromosomal DNA through topological links. *Nature Structural & Molecular Biology*, *18*(8), 894–901. http://doi.org/10.1038/nsmb.2087
Deardorff, M. A., Bando, M., Nakato, R., Watrin, E., Itoh, T., Minamino, M., et al. (2012). HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. *Nature*, *489*(7415), 313–317. http://doi.org/10.1038/nature11316
Deardorff, M. A., Kaur, M., Yaeger, D., Rampuria, A., Korolev, S., Pié, J., et al. (2007). Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation. *American Journal of Human Genetics*, *80*(3), 485–494. http://doi.org/10.1086/511888
Dixon, S. (2013) Towards understandingthe mechanism of cohesin loading. *Thesis*
Dorsett, D., & Krantz, I. D. (2009). On the molecular etiology of Cornelia de Lange syndrome. *Annals of the New York Academy of Sciences*, *1151*(1), 22–37. http://doi.org/10.1111/j.1749-6632.2008.03450.x
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. http://doi.org/10.1101/gad.1498707
Elbatsh, A. M. O., Haarhuis, J. H. I., Petela, N., Chapard, C., Fish, A., Celie, P. H., et al. (2016). Cohesin Releases DNA through Asymmetric ATPase-Driven Ring

Opening. *Molecular Cell*, *61*(4), 575–588. http://doi.org/10.1016/j.molcel.2016.01.025
Farcas, A.-M., Uluocak, P., Helmhart, W., & Nasmyth, K. (2011). Cohesin's concatenation of sister DNAs maintains their intertwining. *Molecular Cell*, *44*(1), 97–107. http://doi.org/10.1016/j.molcel.2011.07.034
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. http://doi.org/10.1371/journal.pgen.1000629
Fernius, J., Nerusheva, O. O., Galander, S., Alves, F. de L., Rappsilber, J., & Marston, A. L. (2013). Cohesin-dependent association of scc2/4 with the centromere initiates pericentromeric cohesion establishment. *Current Biology : CB*, *23*(7), 599–606. http://doi.org/10.1016/j.cub.2013.02.022
Filipski, J., & Mucha, M. (2002). Structure, function and DNA composition of Saccharomyces cerevisiae chromatin loops. *Gene*, *300*(1-2), 63–68.
Furuya, K., Takahashi, K., & Yanagida, M. (1998). Faithful anaphase is ensured by Mis4, a sister chromatid cohesion molecule required in S phase and not destroyed in G1 phase. *Genes & Development*, *12*(21), 3408–3418.
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. http://doi.org/10.1016/j.cub.2006.10.061
Gerlich, D., Koch, B., Dupeux, F., Peters, J.-M. 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. http://doi.org/10.1016/j.cub.2006.06.068
Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., et al. (2005). Galaxy: a platform for interactive large-scale genome analysis. *Genome Research*, *15*(10), 1451–1455. http://doi.org/10.1101/gr.4086505
Gillespie, P. J., & Hirano, T. (2004). Scc2 couples replication licensing to sister chromatid cohesion in Xenopus egg extracts. *Current Biology : CB*, *14*(17), 1598–1603. http://doi.org/10.1016/j.cub.2004.07.053
Gligoris, T. G., Scheinost, J. C., Bürmann, F., Petela, N., Chan, K.-L., Uluocak, P., et al. (2014). Closing the cohesin ring: structure and function of its Smc3-kleisin interface. *Science*, *346*(6212), 963–967. http://doi.org/10.1126/science.1256917
Gligoris, T., & Löwe, J. (2016). Structural Insights into Ring Formation of Cohesin and Related Smc Complexes. *Trends in Cell Biology*. http://doi.org/10.1016/j.tcb.2016.04.002
Glynn, E. F., Megee, P. C., Yu, H.-G., Mistrot, C., Ünal, E., Koshland, D. E., et al. (2004). Genome-wide mapping of the cohesin complex in the yeast Saccharomyces cerevisiae. *PLoS Biology*, *2*(9), E259. http://doi.org/10.1371/journal.pbio.0020259
Goto, T., & Wang, J. C. (1982). Yeast DNA topoisomerase II. An ATP-dependent type II topoisomerase that catalyzes the catenation, decatenation, unknotting, and relaxation of double-stranded DNA rings. *The Journal of Biological Chemistry*, *257*(10), 5866–5872.
Griese, J. J., Witte, G., & Hopfner, K.-P. (2010). Structure and DNA binding activity of the mouse condensin hinge domain highlight common and diverse

features of SMC proteins. *Nucleic Acids Research*, *38*(10), 3454–3465. http://doi.org/10.1093/nar/gkq038
Grigoryev, S. A., & Woodcock, C. L. (2012). Chromatin organization - the 30 nm fiber. *Experimental Cell Research*, *318*(12), 1448–1455. http://doi.org/10.1016/j.yexcr.2012.02.014
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*(3), 523–537. http://doi.org/10.1016/j.cell.2006.08.048
Gruber, S., Haering, C. H., & Nasmyth, K. (2003). Chromosomal cohesin forms a ring. *Cell*, *112*(6), 765–777.
Guacci, V., Koshland, D., & Strunnikov, A. V. (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.
Guillou, E., Ibarra, A., Coulon, V., Casado-Vela, J., Rico, D., Casal, I., et al. (2010). Cohesin organizes chromatin loops at DNA replication factories. *Genes & Development*, *24*(24), 2812–2822. http://doi.org/10.1101/gad.608210
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. http://doi.org/10.1038/nature07098
Haering, C. H., Löwe, J., Hochwagen, A., & Nasmyth, K. (2002). Molecular architecture of SMC proteins and the yeast cohesin complex. *Molecular Cell*, *9*(4), 773–788.
Haering, C. H., Schoffnegger, D., Nishino, T., Helmhart, W., Nasmyth, K., & Löwe, J. (2004). Structure and stability of cohesin's Smc1-kleisin interaction. *Molecular Cell*, *15*(6), 951–964. http://doi.org/10.1016/j.molcel.2004.08.030
Hill, V. K., Kim, J.-S., & Waldman, T. (2016). Cohesin mutations in human cancer. *Biochimica Et Biophysica Acta*, *1866*(1), 1–11. http://doi.org/10.1016/j.bbcan.2016.05.002
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. http://doi.org/10.7554/eLife.06057
Hirano, T. (2005). Condensins: organizing and segregating the genome. *Current Biology : CB*, *15*(7), R265–75. http://doi.org/10.1016/j.cub.2005.03.037
Hirano, T. (2006). At the heart of the chromosome: SMC proteins in action. *Nature Reviews. Molecular Cell Biology*, *7*(5), 311–322. http://doi.org/10.1038/nrm1909
Hoffman, E. A., Frey, B. L., Smith, L. M., & Auble, D. T. (2015). Formaldehyde crosslinking: a tool for the study of chromatin complexes. *The Journal of Biological Chemistry*, *290*(44), 26404–26411. http://doi.org/10.1074/jbc.R115.651679
Holwerda, S. J. B., & de Laat, W. (2013). CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1620), 20120369–20120369. http://doi.org/10.1098/rstb.2012.0369
Hondele, M., Stuwe, T., Hassler, M., Halbach, F., Bowman, A., Zhang, E. T., et al. (2013). Structural basis of histone H2A-H2B recognition by the essential chaperone FACT. *Nature*, *499*(7456), 111–114. http://doi.org/10.1038/nature12242

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. http://doi.org/10.1038/ncb3167

Hu, B., Itoh, T., Mishra, A., Katoh, Y., Chan, K.-L., Upcher, W., et al. (2011). ATP Hydrolysis Is Required for Relocating Cohesin from Sites Occupied by Its Scc2/4 Loading Complex. *Current Biology*, *21*(1), 12–24. http://doi.org/10.1016/j.cub.2010.12.004

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), gkv670–e132. http://doi.org/10.1093/nar/gkv670

Huang, J., & Laurent, B. C. (2004). A Role for the RSC chromatin remodeler in regulating cohesion of sister chromatid arms. *Cell Cycle*, *3*(8), 973–975.

Ivanov, D., & Nasmyth, K. (2005). A Topological Interaction between Cohesin Rings and a Circular Minichromosome. *Cell*, *122*(6), 849–860. http://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. http://doi.org/10.1038/nrm930

Johnson, D. S., Mortazavi, A., Myers, R. M., & Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science*, *316*(5830), 1497–1502. http://doi.org/10.1126/science.1141319

Kawauchi, S., Calof, A. L., Santos, R., Lopez-Burks, M. E., Young, C. M., Hoang, M. P., et al. (2009). Multiple organ system defects and transcriptional dysregulation in the Nipbl(+/-) mouse, a model of Cornelia de Lange Syndrome. *PLoS Genetics*, *5*(9), e1000650. http://doi.org/10.1371/journal.pgen.1000650

Kikuchi, S., Borek, D. M., Otwinowski, Z., Tomchick, D. R., & Yu, H. (2016). Crystal structure of the cohesin loader Scc2 and insight into cohesinopathy. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(44), 12444–12449. http://doi.org/10.1073/pnas.1611333113

Kogut, I., Wang, J., Guacci, V., Mistry, R. K., & Megee, P. C. (2009). The Scc2/Scc4 cohesin loader determines the distribution of cohesin on budding yeast chromosomes. *Genes & Development*, *23*(19), 2345–2357. http://doi.org/10.1101/gad.1819409

Krantz, I. D., McCallum, J., DeScipio, C., Kaur, M., Gillis, L. A., Yaeger, D., et al. (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. http://doi.org/10.1038/ng1364

Ku, B., Lim, J.-H., Shin, H.-C., Shin, S.-Y., & Oh, B.-H. (2010). Crystal structure of the MukB hinge domain with coiled-coil stretches and its functional implications. *Proteins*, *78*(6), 1483–1490. http://doi.org/10.1002/prot.22664

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

Kurze, A., Michie, K. A., Dixon, S. E., Mishra, A., Itoh, T., Khalid, S., et al. (2010). A positively charged channel within the Smc1/Smc3 hinge required for sister chromatid cohesion. *The EMBO Journal*, *30*(2), 364–378.

http://doi.org/10.1038/emboj.2010.315
Ladurner, R., Kreidl, E., Ivanov, M. P., Ekker, H., Idarraga-Amado, M. H., Busslinger, G. A., et al. (2016). Sororin actively maintains sister chromatid cohesion. *The EMBO Journal*, *35*(6), 635–653. http://doi.org/10.15252/embj.201592532
Laloraya, S., Guacci, V., & Koshland, D. (2000). Chromosomal addresses of the cohesin component Mcd1p. *The Journal of Cell Biology*, *151*(5), 1047–1056.
Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. http://doi.org/10.1038/nmeth.1923
Lengronne, A., Katou, Y., Mori, S., Yokobayashi, S., Kelly, G. P., Itoh, T., et al. (2004). Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature*, *430*(6999), 573–578. http://doi.org/10.1038/nature02742
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. http://doi.org/10.1093/bioinformatics/btp352
Liu, D., & Lampson, M. A. (2009). Regulation of kinetochore-microtubule attachments by Aurora B kinase. *Biochemical Society Transactions*, *37*(Pt 5), 976–980. http://doi.org/10.1042/BST0370976
Liu, H., Rankin, S., & Yu, H. (2012). Phosphorylation-enabled binding of SGO1-PP2A to cohesin protects sororin and centromeric cohesion during mitosis. *Nature Cell Biology*. http://doi.org/10.1038/ncb2637
Liu, J., Zhang, Z., Bando, M., Itoh, T., Deardorff, M. A., Clark, D., et al. (2009). Transcriptional dysregulation in NIPBL and cohesin mutant human cells. *PLoS Biology*, *7*(5), e1000119. http://doi.org/10.1371/journal.pbio.1000119
Lopez-Serra, L., Kelly, G. P., Patel, H., Stewart, A., & Uhlmann, F. (2014). The Scc2-Scc4 complex acts in sister chromatid cohesion and transcriptional regulation by maintaining nucleosome-free regions. *Nature Genetics*, *46*(10), 1147–1151. http://doi.org/10.1038/ng.3080
Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, *389*(6648), 251–260. http://doi.org/10.1038/38444
Maeshima, K., Hihara, S., & Eltsov, M. (2010). Chromatin structure: does the 30-nm fibre exist in vivo? *Current Opinion in Cell Biology*, *22*(3), 291–297. http://doi.org/10.1016/j.ceb.2010.03.001
Marsden, M. P., & Laemmli, U. K. (1979). Metaphase chromosome structure: evidence for a radial loop model. *Cell*, *17*(4), 849–858.
Megee, P. C., Mistrot, C., Guacci, V., & Koshland, D. (1999). The centromeric sister chromatid cohesion site directs Mcd1p binding to adjacent sequences. *Molecular Cell*, *4*(3), 445–450.
Merkenschlager, M., & Nora, E. P. (2016). CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annual Review of Genomics and Human Genetics*, *17*(1), annurev–genom–083115–022339. http://doi.org/10.1146/annurev-genom-083115-022339
Michaelis, C., Ciosk, R., & Nasmyth, K. (1997). Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell*, *91*(1), 35–45.
Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-

committed cells. *Nature*, *448*(7153), 553–560. http://doi.org/10.1038/nature06008
Mitchison, T., & Kirschner, M. (1984). Dynamic instability of microtubule growth. *Nature*, *312*(5991), 237–242.
Moldovan, G.-L., Pfander, B., & Jentsch, S. (2006). PCNA controls establishment of sister chromatid cohesion during S phase. *Molecular Cell*, *23*(5), 723–732. http://doi.org/10.1016/j.molcel.2006.07.007
Morgan, D. (2007). The Cell Cycle. *New Science Press*
Murayama, Y., & Uhlmann, F. (2013). Biochemical reconstitution of topological DNA binding by the cohesin ring. *Nature*, *505*(7483), 367–371. http://doi.org/10.1038/nature12867
Murayama, Y., & Uhlmann, F. (2015). DNA Entry into and Exit out of the Cohesin Ring by an Interlocking Gate Mechanism. *Cell*, *163*(7), 1628–1640. http://doi.org/10.1016/j.cell.2015.11.030
Musacchio, A., & Salmon, E. D. (2007). The spindle-assembly checkpoint in space and time. *Nature Reviews. Molecular Cell Biology*, *8*(5), 379–393. http://doi.org/10.1038/nrm2163
Natsume, T., Müller, C. A., Katou, Y., Retkute, R., Gierliński, M., Araki, H., et al. (2013). Kinetochores coordinate pericentromeric cohesion and early DNA replication by Cdc7-Dbf4 kinase recruitment. *Molecular Cell*, *50*(5), 661–674. http://doi.org/10.1016/j.molcel.2013.05.011
Neuwald, A. F., & Hirano, T. (2000). HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. *Genome Research*, *10*(10), 1445–1452. http://doi.org/10.1101/gr.147400
Ng, T. M., Waples, W. G., Lavoie, B. D., & Biggins, S. (2009). Pericentromeric sister chromatid cohesion promotes kinetochore biorientation. *Molecular Biology of the Cell*, *20*(17), 3818–3827. http://doi.org/10.1091/mbc.E09-04-0330
Nicol, J. W., Helt, G. A., Blanchard, S. G., Raja, A., & Loraine, A. E. (2009). The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics (Oxford, England)*, *25*(20), 2730–2731. http://doi.org/10.1093/bioinformatics/btp472
Nigg, E. A. (2001). Mitotic kinases as regulators of cell division and its checkpoints. *Nature Reviews. Molecular Cell Biology*, *2*(1), 21–32. http://doi.org/10.1038/35048096
Niki, H., Jaffé, A., Imamura, R., Ogura, T., & Hiraga, S. (1991). The new gene mukB codes for a 177 kd protein with coiled-coil domains involved in chromosome partitioning of E. coli. *The EMBO Journal*, *10*(1), 183–193.
Nishino, Y., Eltsov, M., Joti, Y., Ito, K., Takata, H., Takahashi, Y., et al. (2012). Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *The EMBO Journal*, *31*(7), 1644–1653. http://doi.org/10.1038/emboj.2012.35
Nishiyama, T., Ladurner, R., Schmitz, J., Kreidl, E., Schleiffer, A., Bhaskara, V., et al. (2010). Sororin Mediates Sister Chromatid Cohesion by Antagonizing Wapl. *Cell*, *143*(5), 737–749. http://doi.org/10.1016/j.cell.2010.10.031
Novak, B., Tyson, J. J., Gyorffy, B., & Csikasz-Nagy, A. (2007). Irreversible cell-cycle transitions are due to systems-level feedback. *Nature Cell Biology*, *9*(7), 724–728. http://doi.org/10.1038/ncb0707-724
Oliveira, R. A., Hamilton, R. S., Pauli, A., Davis, I., & Nasmyth, K. (2010). Cohesin cleavage and Cdk inhibition trigger formation of daughter nuclei. *Nature Cell*

Biology, *12*(2), 185–192. http://doi.org/10.1038/ncb2018
Orlando, D. A., Chen, M. W., Brown, V. E., Solanki, S., Choi, Y. J., Olson, E. R., et al. (2014). Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Reports*, *9*(3), 1163–1170. http://doi.org/10.1016/j.celrep.2014.10.018
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. http://doi.org/10.1016/j.molcel.2016.02.033
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.
Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell*, *132*(3), 422–433. http://doi.org/10.1016/j.cell.2008.01.011
Pauli, A., Althoff, F., Oliveira, R. A., Heidmann, S., Schuldiner, O., Lehner, C. F., et al. (2008). Cell-Type-Specific TEV Protease Cleavage Reveals Cohesin Functions in Drosophila Neurons. *Developmental Cell*, *14*(2), 239–251. http://doi.org/10.1016/j.devcel.2007.12.009
Peters, J.-M. M. (2006). The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nature Reviews. Molecular Cell Biology*, *7*(9), 644–656. http://doi.org/10.1038/nrm1988
Qiao, R., Weissmann, F., Yamaguchi, M., Brown, N. G., VanderLinden, R., Imre, R., et al. (2016). Mechanism of APC/CCDC20 activation by mitotic phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(19), E2570–8. http://doi.org/10.1073/pnas.1604929113
Rankin, S., Ayad, N. G., & Kirschner, M. W. (2005). Sororin, a substrate of the anaphase-promoting complex, is required for sister chromatid cohesion in vertebrates. *Molecular Cell*, *18*(2), 185–200. http://doi.org/10.1016/j.molcel.2005.03.017
Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D., & Klug, A. (1984). Structure of the nucleosome core particle at 7 A resolution. *Nature*, *311*(5986), 532–537.
Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., et al. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature Methods*, *4*(8), 651–657. http://doi.org/10.1038/nmeth1068
Roig, M. B., Löwe, J., Chan, K.-L., Beckouët, F., Metson, J., & Nasmyth, K. (2014). Structure and function of cohesin's Scc3/SA regulatory subunit. *FEBS Letters*. http://doi.org/10.1016/j.febslet.2014.08.015
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*, *321*(5888), 563–566. http://doi.org/10.1126/science.1157774
Rowland, B. D., Roig, M. B., Nishino, T., Kurze, A., Uluocak, P., Mishra, A., et al. (2009). Building Sister Chromatid Cohesion: Smc3 Acetylation Counteracts an Antiestablishment Activity. *Molecular Cell*, *33*(6), 763–774. http://doi.org/10.1016/j.molcel.2009.02.028
Schär, P., Fäsi, M., & Jessberger, R. (2004). SMC1 coordinates DNA double-strand

break repair pathways. *Nucleic Acids Research*, *32*(13), 3921–3929. http://doi.org/10.1093/nar/gkh716
Schleiffer, A., Kaitna, S., Maurer-Stroh, S., Glotzer, M., Nasmyth, K., & Eisenhaber, F. (2003). Kleisins: a superfamily of bacterial and eukaryotic SMC protein partners. *Molecular Cell*, *11*(3), 571–575.
Schuldiner, O., Berdnik, D., Levy, J. M., Wu, J. S., Luginbuhl, D., Gontang, A. C., & Luo, L. (2008). piggyBac-based mosaic screen identifies a postmitotic function for cohesin in regulating developmental axon pruning. *Developmental Cell*, *14*(2), 227–238. http://doi.org/10.1016/j.devcel.2007.11.001
Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(1), 539–539. http://doi.org/10.1038/msb.2011.75
Simpson, R. T. (1978). Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry*, *17*(25), 5524–5531.
Sjögren, C., & Nasmyth, K. (2001). Sister chromatid cohesion is required for postreplicative double-strand break repair in Saccharomyces cerevisiae. *Current Biology : CB*, *11*(12), 991–995.
Stemmann, O., Zou, H., Gerber, S. A., Gygi, S. P., & Kirschner, M. W. (2001). Dual inhibition of sister chromatid separation at metaphase. *Cell*, *107*(6), 715–726.
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.
Tanaka, T., Cosma, M. P., Wirth, K., & Nasmyth, K. (1999). Identification of cohesin association sites at centromeres and along chromosome arms. *Cell*, *98*(6), 847–858.
Thoma, F., Koller, T., & Klug, A. (1979). Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *The Journal of Cell Biology*, *83*(2 Pt 1), 403–427.
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. http://doi.org/10.1038/ng1363
Tóth, A., Ciosk, R., Uhlmann, F., Gálová, 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.
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. http://doi.org/10.1038/21831
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.
Ünal, 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*, *321*(5888), 566–569. http://doi.org/10.1126/science.1157880

Vrouwe, M. G., Elghalbzouri-Maghrani, E., Meijers, M., Schouten, P., Godthelp, B. C., Bhuiyan, Z. A., et al. (2007). Increased DNA damage sensitivity of Cornelia de Lange syndrome cells: evidence for impaired recombinational repair. *Human Molecular Genetics*, *16*(12), 1478–1487. http://doi.org/10.1093/hmg/ddm098

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.

Watrin, E., Schleiffer, A., Tanaka, K., Eisenhaber, F., Nasmyth, K., & Peters, J.-M. M. (2006). Human Scc4 Is Required for Cohesin Binding to Chromatin, Sister-Chromatid Cohesion, and Mitotic Progression. *Current Biology*, *16*(9), 863–874. http://doi.org/10.1016/j.cub.2006.03.049

Weber, S. A., Gerton, J. L., Polancic, J. E., DeRisi, J. L., Koshland, D., & Megee, P. C. (2004). The kinetochore is an enhancer of pericentric cohesin binding. *PLoS Biology*, *2*(9), E260. http://doi.org/10.1371/journal.pbio.0020260

Weitzer, S., Lehane, C., & Uhlmann, F. (2003). A model for ATP hydrolysis-dependent binding of cohesin to DNA. *Current Biology : CB*, *13*(22), 1930–1940.

Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*, *451*(7180), 796–801. http://doi.org/10.1038/nature06634

White, C. L., Suto, R. K., & Luger, K. (2001). Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *The EMBO Journal*, *20*(18), 5207–5218. http://doi.org/10.1093/emboj/20.18.5207

Xu, B., Lu, S., & Gerton, J. L. (2014). Roberts syndrome: A deficit in acetylated cohesin leads to nucleolar dysfunction. *Rare Diseases (Austin, Tex.)*, *2*(1), e27743. http://doi.org/10.4161/rdis.27743

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., et al. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature*, *457*(7232), 1033–1037. http://doi.org/10.1038/nature07728

Yeh, E., Haase, J., Paliulis, L. V., Joglekar, A., Bond, L., Bouck, D., et al. (2008). Pericentric chromatin is organized into an intramolecular loop in mitosis. *Current Biology : CB*, *18*(2), 81–90. http://doi.org/10.1016/j.cub.2007.12.019

Zhang, J., Shi, X., Li, Y., Kim, B.-J., Jia, J., Huang, Z., et al. (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. http://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. http://doi.org/10.1083/jcb.200801157

Zhang, N., Panigrahi, A. K., Jiang, Y., Rao, P. H., & Pati, D. (2010). Role of Sororin in sister chromatid cohesion and separation in vertebrates. *The FASEB Journal*, *24*(1_MeetingAbstracts), 8788.

Zhang, S., Chang, L., Alfieri, C., Zhang, Z., Yang, J., Maslen, S., et al. (2016).

Molecular mechanism of APC/C activation by mitotic phosphorylation. *Nature*, *533*(7602), 260–264. http://doi.org/10.1038/nature17973