



HAL
open science

Structural analysis of the cohesin complex

Yan Li

► **To cite this version:**

Yan Li. Structural analysis of the cohesin complex. Structural Biology [q-bio.BM]. Université Grenoble Alpes, 2019. English. NNT : 2019GREAV012 . tel-03187406

HAL Id: tel-03187406

<https://theses.hal.science/tel-03187406>

Submitted on 1 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ GRENOBLE ALPES



THESIS / THÈSE

To obtain the title of / Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Discipline / *Spécialité* : Structure Biology and Nanotechnology

Arrêté ministériel : 7 août 2006

Presented by / Présentée par

Yan LI

Thesis supervisor / *Thèse dirigée par* **Dr. Daniel Panne**

Thesis prepared at / *Thèse préparée au sein du*

European Molecular Biology Laboratory (EMBL), Grenoble/

In / Dans l'École Doctorale de Chimie et Science du Vivant

Structural analysis of the Cohesin Complex / Analyse structurale du complexe de la cohésine

Public defense on / Thèse soutenue publiquement le **01/04/2019**

Jury members / devant le jury composé de :

Dr. Benjamin ROWLAND

Reviewer / Rapporteur

Dr. Camilla BJORKEGREN

Reviewer / Rapporteur

Dr. Ramesh PILLAI

President / Président

Dr. Joanna TIMMINS

Examiner / Examineur



Table of Contents

| | |
|---|----|
| Table of Contents | 2 |
| List of Abbreviations | 9 |
| List of Figures / Tables | 14 |
| Genome cohesion and condensation during the cell cycle | 15 |
| Role of CTCF in the regulation of paternal and maternal imprinting | 15 |
| Model for CTCF-mediated cohesin binding and for the generation of TADs by convergently oriented CTCF sites | 16 |
| Abstract | 18 |
| Résumé..... | 21 |
| Introduction..... | 24 |
| 1.1 The cell cycle | 26 |
| 1.2 Chromosomes are organised by SMC-kleisin enzymes | 27 |
| Figure 1.1 Genome cohesion and condensation during the cell cycle. | 28 |
| 1.3 Organisation of the core cohesin complex | 29 |
| Figure 1.2 Structure of the cohesin core subunits and domain boundaries. | 32 |
| Figure 1.3 Structure of the human SA2-Scc1 complex and domain boundaries..... | 34 |
| 1.4 Overview of the cohesin cycle | 34 |
| Figure 1.4 Cartoon illustration of the ring-shape cohesin and its accessory factors. | 36 |
| Table 1.1 Cohesin and its related factors in different species | 37 |
| 1.5 Cohesin ATPase dimerization | 37 |
| Figure 1.5 Cartoon of cohesin functional cycle and regulators..... | 40 |
| 1.6 The cohesin loading complex..... | 40 |
| 1.7 Cleavage-independent cohesin unloading | 41 |
| Figure 1.6 Structure of the yeast Pds5-Scc1 complex and domain boundaries..... | 42 |
| Figure 1.7 Structure of the human Wapl-C domain and domain boundaries. | 43 |
| 1.8 SMC3 acetylation (establishment of cohesion)..... | 43 |
| Figure 1.8 Structure of human Escp1 and domain boundaries..... | 44 |

| | |
|--|----|
| 1.9 Cohesin stabilization | 45 |
| 1.10 Cohesin release at chromosome arms | 45 |
| 1.11 Cohesion protection at the centromere | 45 |
| Figure 1.9 Structure of the Sgo1-PP2A complex and domain boundaries..... | 47 |
| 1.12 Cohesin cleavage at the centromere | 48 |
| Figure 1.10 Structure of the yeast Separase-Securin complex and domain boundaries...49 | |
| 1.13 Cohesin beyond cohesion..... | 49 |
| 1.14 CTCF..... | 50 |
| Figure 1.11 Model of CTCF zinc fingers 2-9 in complex with DNA..... | 51 |
| Figure 1.12 Role of CTCF in the regulation of paternal and maternal imprinting..... | 52 |
| Table 1.2 CTCF interaction partners | 53 |
| 1.15 Cohesin and CTCF | 54 |
| 1.16 Loop extrusion model and different levels of the mammalian genome folding | 55 |
| Figure 1.13 Levels of Genome folding..... | 57 |
| Scientific aims..... | 59 |
| Materials and Methods..... | 61 |
| Matériels et méthodes | 62 |
| 2.1 Construct Cloning | 64 |
| 2.2 Protein expression | 64 |
| 2.2.1 Protein expression | 64 |
| 2.2.2 Selenomethionine-labelled protein expression..... | 65 |
| 2.3 Protein purification..... | 65 |
| 2.3.1 Purification with His or His-GST tag..... | 65 |
| 2.3.2 Purification with GST tag..... | 66 |
| 2.3.3 Ion-exchange purification..... | 66 |
| 2.3.4 Size exclusion..... | 66 |
| 2.4 Crystallisation..... | 67 |

| | |
|---|----|
| Table 2.1 Scc3T-Scc1K crystallisation screen conditions. | 67 |
| 2.4.1 Crystals of ctSmc1hd-CScc1 complex | 67 |
| 2.4.2 Crystals of the Scc3-Scc1-DNA complex | 68 |
| 2.4.3 Crystal seeding | 68 |
| 2.4.4 Crystals of the SA2T-Scc1H complex | 68 |
| 2.5 Structure determination | 69 |
| 2.5.1 Structure of the ctSmc1-CScc1 complex..... | 69 |
| 2.5.2 Structure of the Scc3-Scc1-DNA complex..... | 69 |
| 2.5.3 Structure of the SA2-Scc1-CTCF complex..... | 70 |
| 2.6 Bioinformatic methods..... | 71 |
| 2.6.1 Sequence alignment..... | 71 |
| 2.7 Protein biochemistry | 71 |
| 2.7.1 In vitro pull down | 71 |
| 2.7.2 Electrophoretic Mobility Shift Assays (EMSA)..... | 71 |
| 2.7.3 Fluorescence Polarization (FP)..... | 72 |
| 2.7.4 Isothermal Titration Calorimeter (ITC)..... | 72 |
| 2.8 <i>In vivo</i> -Yeast methods..... | 73 |
| 2.8.1 Tetrad analysis..... | 73 |
| 2.8.2 Chromatin immunoprecipitation and ChIP-qPCR..... | 73 |
| 2.9 Peptide array..... | 73 |
| Result_part 1 Assembling the cohesin complex in <i>E.Coli</i> ('Clink') | 75 |
| Résultats_partie 1 | 76 |
| 3.1 Expression of cohesin complex in <i>E.coli</i> | 78 |
| 3.2 Expression trials of Smc1 head domain of different species in <i>E.coli</i> | 78 |
| Figure 3.1 Smc1 head boundaries and purification..... | 79 |
| 3.3 Structure of the <i>Chaetomium thermophilum</i> Smc1 head in complex with <i>Saccharomyces cerevisiae</i> Scc1 (ctSmc1hd-CScc1)..... | 80 |

| | |
|--|-----|
| Figure 3.2 Structure of the ctSmc1 head domain in complex with yeast CSc1..... | 80 |
| Table 3.1 X ray data collection, phasing and refinement statistics | 81 |
| 3.4 ctSmc1/Smc3 head assembly via NScc1 linker -“Clink” | 81 |
| Figure 3.3 Purification and DNA binding of Clink115/120/125..... | 84 |
| 3.5 Complex assembly trials of Clink159 in complex with Pds5 and Scc3..... | 84 |
| Figure 3.4 Purification and EMSA of Clink159 in complex with either Pds5, Scc3, or both..... | 85 |
| 3.6 Crystallization trials of the Clink complex. | 86 |
| Result_part2 Structural basis for Scc3-dependent cohesin recruitment to chromatin | 87 |
| Résultats_partie 2..... | 88 |
| 4.1 Investigating DNA binding by the cohesin complex | 90 |
| Figure 4.1 SDS-PAGE analysis of the purified cohesin components and DNA binding analysis by EMSA. | 91 |
| Figure 4.2 Cohesin subunits that lack DNA binding affinity..... | 92 |
| Figure 4.3 DNA binding by the Scc3-Scc1 subcomplex..... | 93 |
| Table 4.1 DNA oligos used for crystallization, EMSA and Fluorescence Polarization (FP)..... | 94 |
| 4.2 Crystal structure of the Scc3-Scc1 in complex with a 19 bp DNA..... | 95 |
| Table 4.2 Data collection and refinement statistics..... | 96 |
| Figure 4.4 Structure of the Scc3-Scc1 subcomplex bound to DNA..... | 98 |
| Figure 4.5 Electron density for the DNA molecule bound to the Scc3-Scc1 subcomplex. | 99 |
| Figure 4.6 Biochemical analysis of Scc3-Scc1 subcomplexes and Scc3 conservation alignment. | 102 |
| Figure 4.7 DNA binding affinity test of both Scc3T-Scc1K wild type and mutants via EMSA..... | 103 |
| 4.3 <i>In vivo</i> characterization of the Scc3-Scc1-DNA structure | 103 |

| | |
|---|-----|
| Figure 4.8 A conserved DNA binding domain in the Scc3-Scc1 subcomplex is required for cohesin association with chromatin. | 104 |
| Figure 4.9 In vivo analysis of SCC3 mutants..... | 105 |
| Table 4.3 ChIP-qPCR primer sequences (5'→3')..... | 106 |
| Table 4.4 Yeast genotypes..... | 107 |
| Figure 4.10 Purification of selenomethionine-labeled Scc3T-Scc1K and different Scc3-Scc1 truncations used for crystallization and FP trials..... | 108 |
| 4.4 Screening for possible protein interactions among cohesin subcomplexes | 109 |
| Figure 4.11 Screening for possible protein interactions among cohesin subcomplexes. | 110 |
| Discussion_part 2..... | 111 |
| Discussion_partie 2..... | 112 |
| Figure 4.12 A conserved DNA binding interface in cohesin and condensin. | 116 |
| Result_part 3 Principles of cohesin regulation on chromatin via its SA2-Scc1 subunits..... | 118 |
| Résultats_partie 3..... | 119 |
| 5.1 Motivation to work on human SA2..... | 121 |
| Figure 5.1 Yeast Pds5 and Scc3 do not bind human CTCF..... | 121 |
| 5.2 Mapping SA2-Scc1 binding regions on CTCF | 122 |
| Figure 5.2 SA2-Scc1 constructs, purification and EMSA analysis..... | 123 |
| Figure 5.3 Biochemical analysis of CTCF binding to SA2-Scc1..... | 124 |
| 5.3 Crystallization trials of the SA2-Scc1-CTCF ternary complex | 125 |
| Figure 5.4 Purification of SA2-Scc1H mutants, CTCF-N6 mutants and CTCF covering both SA2-Scc1 binding region and zinc finger 1-7 (CTCF-ZF7) | 126 |
| 5.4 SA2-Scc1-CTCF ternary complex is sensitive to DNA binding..... | 126 |
| Figure 5.5 Co-expression and co-purification of the SA2T-Scc1H-CTCF complex and DNA binding assay..... | 127 |
| 5.5 Structure of SA2-Scc1 in complex with a CTCF peptide | 128 |
| Figure 5.6 Structure of hSA2-Scc1 bound to a fragment of CTCF..... | 130 |

| | |
|---|-----|
| 5.6 Mutagenesis analysis of the SA2-Scc1-CTCF ternary complex | 131 |
| Figure 5.7 Mutagenesis and in vitro analysis of the SA2-Scc1-CTCF ternary complex. | 132 |
| Figure 5.8 Analysis of the SA2-Scc1-CTCF structure..... | 133 |
| Figure 5.9 Model for CTCF-mediated cohesin binding and for the generation of TADs by convergently oriented CTCF sites..... | 134 |
| 5.7 Analysis of additional cohesin ligands containing Y/FxF motifs | 135 |
| Figure 5.10 Y/FxF motifs are conserved among CTCF, Wapl, Sororin and Shugoshin. | 137 |
| Figure 5.11 Sgo1 interacts with SA2-Scc1..... | 138 |
| Figure 5.12 Structure of the SA2-Scc1 and Sgo1 ternary complex. | 139 |
| Table 5.1 X-ray data collection and refinement of SA2-Scc1-CTCF and SA2-Scc1-Sgo1 complex. | 140 |
| 5.8 Bioinformatic prediction of cohesin binding factors..... | 140 |
| Figure 5.13 Novel SA2-Scc1 binding partners screened via peptide array..... | 142 |
| Figure 5.14 Model of SA2-Scc1 subcomplex bound with different Y/FxF-containing proteins | 143 |
| Discussion_part 3..... | 144 |
| Discussion_partie 3..... | 145 |
| Acknowledgement | 149 |
| Appendix..... | 152 |
| Table S1 Timeline of all cohesin and cohesin cycle-related structures..... | 153 |
| Table S2 Constructs table..... | 154 |
| References..... | 158 |

List of Abbreviations

| | |
|------------------|---|
| 3C | chromatin conformation capture |
| 3D | three dimensions |
| 6-FAM | 6-Carboxyfluorescein |
| A | Alanine / compartment A |
| Å | Ångstrom |
| ABC | Adenosine triphosphate binding cassette |
| Ac | Acetylation |
| Ac-CoA | Acyl-coenzyme A |
| ag | <i>Aphis gossypii</i> |
| APC/C | Anaphase promoting complex / cyclosome |
| AT-hook | Adenine-Thymine hook, a DNA binding motif |
| ATP | Adenosine triphosphate |
| ATP- γ -S | Adenosine triphosphate gamma S |
| B | Bound / compartment B |
| bp | Basepairs |
| Brn | Barren |
| C2H2 | Cysteine 2-Histidine 2 |
| Cdc | Cell-division cycle protein |
| Cdk | Cyclin-dependent kinase |
| Cen | Centromeric |
| CENPU | Centromere Protein U |
| CES | conserved essential surface |
| ChIP-qPCR | Chromatin immunoprecipitation sequencing-quantitative Polymerase Chain Reaction |
| Clink | Linked cohesin |
| Co | Cobalt |
| CoA | Coenzyme A |
| ct | <i>Chaetomium Thermophilum</i> |
| CTCF | CCCTC-binding factor |
| Ctf | chromosome transmission fidelity |
| D | Acide aspartique |
| DNA | Deoxyribonucleic acid |
| dsDNA | Double strand deoxyribonucleic acid |

| | |
|----------------|--|
| E | Glutamic acid / Enhancer |
| <i>E.Coli</i> | <i>Escherichia coli</i> |
| Eco/Esco/Eso | Establishment of cohesion |
| EDTA | Ethylenediaminetetraacetic acid |
| EM | Electron microscopy |
| EMBL | European Molecular Biology Laboratory |
| EMSA | Electrophoretic mobility shift assay |
| Esp | Extra Spindle Pole bodies |
| F | Phenylalanine |
| FGF | Phenylalanine Glycine Phenylalanine |
| fl | Full-length |
| FP | Fluorescence Polarization |
| G | Glycine |
| GNAT | Gcn5-related N-acetyltransferase |
| GST | Glutathione S-transferase |
| HA | Human influenza hemagglutinin |
| hd | Head |
| HEAT | Huntington, EF3, PP2A, TOR1 |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| Hi-C | High-throughput sequencing-based chromosome conformation capture |
| His-tag | Polyhistidine-tag |
| HTX | High-throughput crystallography |
| I | Input / Isoleucine |
| ICR | Imprinting Control Region |
| IMAC | Immobilised metal ion chromatography |
| IP6 | inositol hexa-kis-phosphate |
| IPTG | Isopropyl-- β -thio-galactoside |
| ITC | Isothermal Titration Calorimeter |
| K | Lysine |
| K _d | dissociation constant |
| L | Leucine |
| LB | Lysogeny Broth |
| M | Methionine / Protein Marker |

| | |
|----------------|--------------------------------------|
| mAU | Milliabsorbance units |
| Mcd | Mitotic chromosome determinant |
| MCM | Minichromosome maintenance |
| Me | Methylation |
| MES | 2-(N-morpholino) ethanesulfonic acid |
| μl | Microlitres |
| mg | Milligram |
| ml | Millilitres |
| μM | MicroMolar |
| mM | MilliMolar |
| Myc | Myelocytomatosis |
| N | Asparagine |
| N ₂ | Nitrogen |
| NaCl | Sodium chloride |
| nm | nanometres |
| OD | Optical density |
| ORF | Open reading frame |
| PBS | Phosphate-buffered saline |
| PCR | Polymerase chain reaction |
| PDB | Protein databank |
| Pds | Precocious dissociation of sisters |
| PEG | Polyethylene Glycol |
| pH | Potential Hydrogen |
| p / pho | Phosphorylated |
| PK | V5 epitope |
| Plk | Polo-like kinase |
| PP2A | Protein phosphatase 2A |
| PRC | polycomb-group repressive complex |
| precen | pericentromeric |
| R | Arginine |
| RBR | RNA-binding region |
| RNA | Ribonucleic acid |
| S | Serine |

| | |
|-------------------|--|
| SA | Stromal Antigen |
| SAC | spindle assembly checkpoint |
| Scc | Sister chromatid cohesion |
| SDS-PAGE | Sodium dodecyl sulphate-polyacrylamide gel electrophoresis |
| Se / SeMet | Selenomethionine |
| Sgo | Shugoshin |
| Smc | Structural maintenance of chromosomes |
| sp | <i>Schizosaccharomyces pombe</i> |
| STAG | Stromal antigen |
| SYCP | Synaptonemal complex protein |
| T | Threonine |
| TAD | Topologically Associating Domain |
| TBS | Tris-buffered saline |
| TBS-T | Tris-buffered saline with Tween 20 |
| TCEP | Tris(2-carboxyethyl)phosphine |
| TEV | Tobacco etch virus |
| TRIS | Tris(hydroxymethyl)aminomethane |
| V | Valine |
| W | Tryptophan |
| Wapl | Wings apart-like protein homologue |
| WHD | winged helix domain |
| w/w | Weight by weight |
| Y | Tyrosine |
| YB | Y box Binding |
| Ycg | Yeast Cap G |
| YSR | Tyrosine-Serine-Arginine |
| ZF | Zinc finger |
| ZGPAT | Zinc finger CCCH-type with G patch domain-containing protein |
| ZnCl ₂ | Zinc Chloride |

List of Figures / Tables

| | | |
|-------------|--|-----|
| Figure 1.1 | Genome cohesion and condensation during the cell cycle | 27 |
| Figure 1.2 | Structure of the cohesin core subunits and domain boundaries | 31 |
| Figure 1.3 | Structure of the human SA2-Scc1 complex and domain boundaries | 34 |
| Figure 1.4 | Cartoon illustration of the ring-shape cohesin and its accessory factors | 36 |
| Table 1.1 | Cohesin and its related factors in different species | 37 |
| Figure 1.5 | Cartoon of cohesin functional cycle and regulators | 40 |
| Figure 1.6 | Structure of the yeast Pds5-Scc1 complex and domain boundaries | 42 |
| Figure 1.7 | Structure of the human Wapl-C domain and domain boundaries | 43 |
| Figure 1.8 | Structure of human Escol1 and domain boundaries | 44 |
| Figure 1.9 | Structure of the Sgo1-PP2A complex and domain boundaries | 47 |
| Figure 1.10 | Structure of the yeast Separase-Securin complex and domain boundaries | 49 |
| Figure 1.11 | Model of CTCF zinc fingers 2-9 in complex with DNA | 51 |
| Figure 1.12 | Role of CTCF in the regulation of paternal and maternal imprinting | 52 |
| Table 1.2 | CTCF interaction partners | 53 |
| Figure 1.13 | Levels of Genome folding | 57 |
| Table 2.1 | Scc3T-Scc1K crystallisation screen conditions | 67 |
| Figure 3.1 | Smc1 head boundaries and purification | 79 |
| Figure 3.2 | Structure of the ctSmc1 head domain in complex with yeast CScc1 | 80 |
| Table 3.1 | X ray data collection, phasing and refinement statistics | 81 |
| Figure 3.3 | Purification and DNA binding of Clink115/120/125 | 84 |
| Figure 3.4 | Purification and EMSA of Clink159 in complex with either Pds5, Scc3, or both | 85 |
| Figure 4.1 | SDS-PAGE analysis of the purified cohesin components and DNA binding analysis by EMSA | 91 |
| Figure 4.2 | Cohesin subunits that lack DNA binding affinity | 92 |
| Figure 4.3 | DNA binding by the Scc3-Scc1 subcomplex | 93 |
| Table 4.1 | DNA oligos used for crystallization, EMSA and Fluorescence Polarization (FP) | 94 |
| Table 4.2 | Data collection and refinement statistics | 96 |
| Figure 4.4 | Structure of the Scc3-Scc1 subcomplex bound to DNA | 98 |
| Figure 4.5 | Electron density for the DNA molecule bound to the Scc3-Scc1 subcomplex | 99 |
| Figure 4.6 | Biochemical analysis of Scc3-Scc1 subcomplexes and Scc3 | 102 |

| | | |
|-------------|--|-----|
| | conservation alignment | |
| Figure 4.7 | DNA binding affinity test of both Scc3T-Scc1K wild type and mutants via EMSA | 103 |
| Figure 4.8 | A conserved DNA binding domain in the Scc3-Scc1 subcomplex is required for cohesin association with chromatin | 104 |
| Figure 4.9 | In vivo analysis of SCC3 mutants | 105 |
| Table 4.3 | ChIP-qPCR primer sequences (5'→3') | 106 |
| Table 4.4 | Yeast genotypes | 107 |
| Figure 4.10 | Purification of selenomethionine-labeled Scc3T-Scc1K and different Scc3-Scc1 truncations used for crystallization and FP trials | 108 |
| Figure 4.11 | Screening for possible protein interactions among cohesin subcomplexes | 110 |
| Figure 4.12 | A conserved DNA binding interface in cohesin and condensin | 116 |
| Figure 5.1 | Yeast Pds5 and Scc3 do not bind human CTCF | 121 |
| Figure 5.2 | SA2-Scc1 constructs, purification and EMSA analysis | 123 |
| Figure 5.3 | Biochemical analysis of CTCF binding to SA2-Scc1 | 124 |
| Figure 5.4 | Purification of SA2-Scc1H mutants, CTCF-N6 mutants and CTCF covering both SA2-Scc1 binding region and zinc finger 1-7 (CTCF-ZF7) | 126 |
| Figure 5.5 | Co-expression and co-purification of the SA2T-Scc1H-CTCF complex and DNA binding assay | 127 |
| Figure 5.6 | Structure of hSA2-Scc1 bound to a fragment of CTCF | 130 |
| Figure 5.7 | Mutagenesis and in vitro analysis of the SA2-Scc1-CTCF ternary complex | 132 |
| Figure 5.8 | Analysis of the SA2-Scc1-CTCF structure | 133 |
| Figure 5.9 | Model for CTCF-mediated cohesin binding and for the generation of TADs by convergently oriented CTCF sites | 134 |
| Figure 5.10 | Y/FxF motifs are conserved among CTCF, Wapl, Sororin and Shugoshin | 137 |
| Figure 5.11 | Sgo1 interacts with SA2-Scc1 | 138 |
| Figure 5.12 | Structure of the SA2-Scc1 and Sgo1 ternary complex | 139 |
| Table 5.1 | X-ray data collection and refinement of SA2-Scc1-CTCF and SA2-Scc1-Sgo1 complex | 140 |
| Figure 5.13 | Novel SA2-Scc1 binding partners screened via peptide array | 142 |
| Figure 5.14 | Model of SA2-Scc1 subcomplex bound with different Y/FxF-containing | 143 |

| | | |
|----------|--|-----|
| | proteins | |
| Table S1 | Timeline of all cohesin and cohesin cycle-related structures | 153 |
| Table S2 | Constructs table | 154 |

Abstract

The cohesin complex is required for numerous chromosomal transactions including sister chromatid cohesion, DNA damage repair, transcriptional regulation and control of 3D chromatin architecture. The basic subunits of cohesin, Smc1, Smc3, Scc1 assemble a ring-shaped complex via connection of the heterodimeric ‘hinge’ domains contributed by Smc1 and Smc3, and through linkage of the SMC ATPase domains by Scc1. Additional accessory factors play important roles in different aspects of cohesin function, such as Scc3, the loading and unloading complexes, Scc2-Scc4 and Pds5-Wapl, respectively, responsible for cohesin loading and its disassociation from chromatin. To establish cohesion, an acetyltransferase called Eco1 acetylates the ATPase domain of Smc3, which may recruit sororin to stabilize cohesion by competing off Wapl and by stabilizing Pds5 binding. In metazoans, cohesin is released from chromosomes in two major steps. One step involves cohesin release at chromosome arms triggered by cohesin phosphorylation and the second step involves centromere cohesin cleavage by separase. Beyond cohesion, cohesin is also involved in a number of additional biological functions including, by cooperating with CTCF, organization of mammalian 3D genome architecture.

To investigate the cohesin ATPase module, I sought to generate a series of expression constructs in *E.coli*. As only Smc1 proved difficult to express, I screened a series of Smc1 heads from different species. I was able to produce the Smc1 head from *Chaetomium Thermophilum* Smc1 (ctSmc1) and to obtain a crystal structure in complex with a fragment of yeast Scc1 (CScc1) at 2.1 Å resolution. This allowed me to express and co-purify an assembled cohesin head Smc1-Smc3-Scc1 heterotrimer by fusing NScc1 to the C terminal end of ctSmc1 together with the remaining head module components.

To understand how cohesin engages DNA, I determined a crystal structure of the Scc3-Scc1 subcomplex bound to double-stranded DNA, showing that the Scc3-Scc1 subcomplex

engages double-stranded DNA through a conserved, positively charged surface. We found that this surface is essential for the enrichment of cohesin complexes on chromosomes.

To investigate the molecular basis for the collaboration between cohesin and CTCF in defining the 3D architecture of the genome, I identified the critical region of CTCF required for binding to the SA2-Scc1 subcomplex. I was able to determine the structure of the ternary complex by x-ray crystallography. The structure revealed the molecular mechanism for CTCF-cohesin interaction. A similar motif is present in Wapl, the cohesin release factor, which binds to the same surface on the SA2-Scc1 complex, as well as in Shugoshin and in a number of established and novel cohesin clients. I was able to determine a crystal structure of the relevant motif of Shugoshin in complex with SA2-Scc1, which confirmed that it binds to the same conserved surface on SA2-Scc1. As binding of each of the factors investigated is mutually exclusive, competition for the conserved surface might be a required aspect of cohesin function in different biological contexts, for example, to counteract Wapl-mediated release. My results thus provide fundamental insights into the molecular mechanism that enables regulation of cohesin function and dynamic chromatin folding.

Résumé

Le complexe de la cohésine est requis pour de nombreuses transactions chromosomiques, notamment la cohésion des chromatides soeurs, la réparation des dommages à l'ADN, la régulation de la transcription et le contrôle de l'architecture tridimensionnelle de la chromatine. Les sous-unités de base de la cohésine, Smc1, Smc3, Scc1 assemblent un complexe en forme d'anneau via la connexion des domaines «hinge» hétérodimériques fournis par Smc1 et Smc3, et par la liaison des domaines SMC ATPase par Scc1. D'autres facteurs accessoires jouent un rôle important dans différents aspects de la fonction de la cohésine, tels que Scc3, les complexes de chargement et de déchargement, Scc2-Scc4 et Pds5-Wapl, respectivement, responsables du chargement de la cohésine et de sa dissociation de la chromatine. Pour établir la cohésion, une acétyltransférase appelée Eco1 acétyle le domaine ATPase de Smc3, qui peut recruter la sororine pour stabiliser la cohésion en faisant concurrence à Wapl et en formant un complexe de stabilisation avec Pds5. Chez les métazoaires, la cohésine est libérée des chromosomes en deux étapes principales. Une étape implique la libération de la cohésine au niveau des bras chromosomiques déclenchée par la phosphorylation de la cohésine et la seconde étape comprend le clivage de la cohésine centromérique par la séparase. Au-delà de la cohésion, la cohésine est également impliquée dans un certain nombre de fonctions biologiques supplémentaires, y compris, en coopérant avec la CTCF, l'organisation de l'architecture tridimensionnelle du génome mammifère.

Pour étudier le module ATPase de la cohésine, j'ai généré une série de constructions pour expression dans *E. coli*. Comme seul Smc1 était difficile à exprimer, j'ai criblé une série de constructions de Smc1 de différentes espèces. J'ai pu produire la tête de Smc1 de *Chaetomium Thermophilum* (ctSmc1) et obtenir une structure cristallographique en complexe avec le fragment de Scc1 de levure (CScc1) à une résolution de 2,1 Å. Cela m'a permis d'exprimer et de co-purifier un hétérotrimère de la tête de cohésine assemblée Smc1-Smc3-

Scc1 en fusionnant NScc1 à l'extrémité C-terminal de CtSmc1 en même temps que les autres composants du module de la tête.

Pour comprendre comment la cohésine engage l'ADN, j'ai déterminé une structure cristallographique du sous-complexe Scc3-Scc1 lié à l'ADN double brin, montrant que le sous-complexe Scc3-Scc1 engage l'ADN double brin à travers une surface conservée chargée positivement. Nous avons trouvé que cette surface est essentielle pour l'enrichissement de complexes de la cohésine sur les chromosomes.

Pour étudier les bases moléculaires de la collaboration entre la cohésine et la CTCF dans la définition de l'architecture tridimensionnelle du génome, j'ai identifié la région critique du CTCF nécessaire à la liaison au sous-complexe SA2-Scc1. J'ai pu déterminer la structure du complexe ternaire par cristallographie aux rayons X. La structure a révélé le mécanisme moléculaire de l'interaction CTCF-cohésine. De manière frappante, un motif similaire est présent dans Wapl, le facteur de libération de la cohésine, qui se lie à la même surface sur le complexe SA2-Scc1, ainsi que dans Shugoshin et dans un certain nombre de établis et nouveaux clients de la cohésine. J'ai pu déterminer une structure cristallographique du motif pertinent de Shugoshin dans un complexe avec SA2-Scc1, ce qui a confirmé qu'il se lie à la même surface conservée sur SA2-Scc1. Comme la liaison est mutuellement exclusive, la compétition pour la surface conservée pourrait être un aspect requis pour la fonction de la cohésine dans différents contextes biologiques. Mes résultats fournissent des informations fondamentales sur le mécanisme moléculaire qui permet la régulation de la fonction de la cohésine et le repliement dynamique de la chromatine.

Introduction

Le complexe de la cohésine est requis pour de nombreuses transactions chromosomiques, notamment la cohésion des chromatides sœurs, la réparation des dommages de l'ADN, la régulation de la transcription et le contrôle de l'architecture tridimensionnelle de la chromatine. La manière dont la cohésine engage la chromatine est restée une question majeure. Les sous-unités de base de la cohésine, Smc1, Smc3, Scc1 assemblent un complexe en forme d'anneau via la connexion des domaines hétérodimères SMC fournis par Smc1 et Smc3, et par la liaison des domaines SMC ATPase par Scc1. D'autres facteurs accessoires jouent un rôle important dans différents aspects de la fonction de la cohésine, tels que Scc3, qui favorise l'association de la cohésine à l'ADN, les complexes de chargement et de déchargement, Scc2-Scc4 et Pds5-Wapl respectivement, responsables du chargement de la cohésine et de sa dissociation de la chromatine. Au cours de la phase S, une acétyltransférase appelée Eco1 acétyle le domaine ATPase de Smc3 et déclenche la stabilisation ou l'établissement de la cohésine. Pour augmenter davantage la cohésion, un facteur métazoaire supplémentaire, la sororine forme un complexe avec Pds5 pour empêcher la liaison de Wapl. Pendant la métaphase, la cohésine centromérique est protégée par le complexe shugoshin-PP2A. Chez les métazoaires, la cohésine est libérée des chromosomes en deux étapes principales. La première nécessite la phosphorylation de la cohésine et permet à Wapl de se lier à nouveau à Pds5 afin d'assurer la médiation de la libération de la cohésine indépendante du clivage à partir des bras des chromosomes. La seconde se produit lors de la réalisation de l'assemblage de la broche et nécessite l'activation d'une protéase appelée séparase, ce qui entraîne le clivage de la Scc1, libérant ainsi des chromatides sœurs qui se séparent dans des cellules filles. Au-delà de la cohésion, il devient également évident que la cohésine joue des rôles plus divers en interagissant avec une multitude d'autres facteurs, notamment la CTCF, une protéine à doigt de zinc connue comme un isolant, dont il a été rapporté qu'elle collabore à la détermination de la structure tridimensionnelle du génome.

1.1 The cell cycle

Cell division has been a wide-spread and fundamental study for biologists since the development of the microscope permitted the direct observation of the cell. Genetic material must be faithfully duplicated and divided into daughter cells. From an evolutionary perspective, from bacteria to eukaryotes, this genetic 'pool' expands considerably, ranging from megabase to gigabase, which makes it more and more complicated. Unravelling the mechanisms by which genomes are accurately maintained, duplicated, and readily segregated into daughter cells has eluded researchers for over a century, since Walther Flemming described the congression and longitudinal splitting of chromosomes, or 'mitosis', a Greek word meaning 'thread', during cell division.

In Bacteria, which lack a nucleus, genome replication and segregation are more tightly and temporally coupled, as compared to eukaryotes. The bacterial cell cycle is divided into three periods: B, C and D. B marks the initiation of chromosome replication, which proceeds during C, and terminates with the end of replication and execution of division in period D. The process of cell division is complex and involves many regulatory steps. Most of these steps are regulated by reversible protein phosphorylation events, which are thought to facilitate crucial checkpoints throughout the cell cycle. In eukaryotes, the cell cycle is mainly divided into two periods: interphase and mitotic phase (Figure 1.1). Interphase consists of G₁, G₂ and S phases. During G₁ phase cells grow and prepare for DNA replication which takes place during S phase. During G₂ phase, the cell prepares to divide. During the mitotic (M) phase, several steps to organise and divide the replicated DNAs take place, namely prophase, metaphase and anaphase. There are three main checkpoints that are important either for entering the cell cycle (**G₁ checkpoint**, which makes the key decision on whether the cell should divide), or entry into the proliferative M phase (**G₂ checkpoint**, wherein the nuclear envelop breaks down and individual chromosomes are resolved into distinct, rod-

shaped masses). Finally, all chromosomes must be aligned along the metaphase plate, and kinetochore biorientation and attachment to spindle microtubules must be achieved to satisfy the **Metaphase checkpoint**, which is also known as the **spindle assembly checkpoint (SAC)** (Figure 1.1).

1.2 Chromosomes are organised by SMC-kleisin enzymes

Chromosomal DNA is typically organised within cellular, or subcellular compartments of dimensions considerably smaller than its linear length (Ou, et al. 2017). Chromatin must therefore be organised and stored in a fashion which permits its enclosure within the cell, without restricting its accessibility. Genome maintenance, replication, transcription, and segregation hence present a significant topological problem.

An ancient family of enzymes, SMC (structural maintenance of chromosomes)-kleisins (from the Greek for ‘closure’), facilitates chromatin organisation from bacteria to humans. In eukaryotes, three evolutionarily conserved SMC-kleisin complexes have arisen from a single prokaryotic precursor and fulfil multiple distinct roles within the cell cycle. These complexes are called cohesin (Smc1 and Smc3), condensin (Smc2 and Smc4) and Smc5/6. Together they predominantly associate with sister chromatin cohesion, chromosome condensation, and DNA damage repair, respectively (Fujioka, et al. 2002; Hirano, et al. 1997; Losada, et al. 1998).

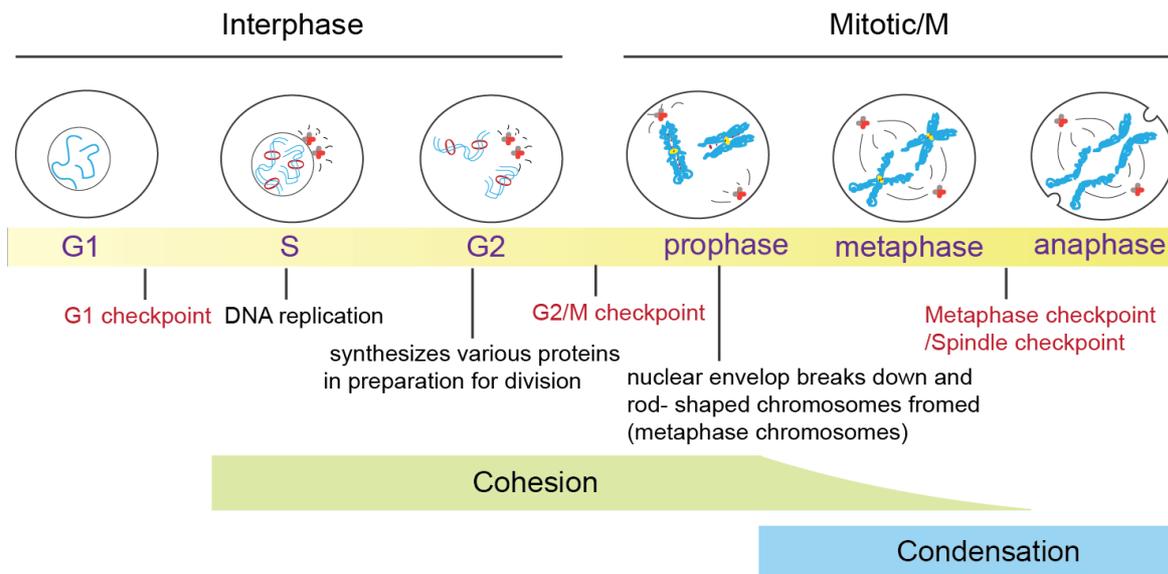


Figure 1.1 Genome cohesion and condensation during the cell cycle. Cohesin is loaded onto DNA during G1 phase and stays attached until the onset of anaphase. The bulk of cohesin is released from chromatin to produce the typical ‘X’ shaped chromosomes that are aligned along the metaphase plate. In animal cells, condensin binds to chromosomes at about the same time that most cohesin dissociates: between prophase and prometaphase (Losada, et al. 1998).

A fundamental and conserved property of SMC-kleisin enzymes is their capacity to topologically engage DNA molecules (Cuylen, et al. 2011; Haering, et al. 2008; Kanno, et al. 2015; Wilhelm, et al. 2015). Once engaged, these enzymes are thought to regulate higher-order genome architecture through coordinated cycles of DNA entrapment and release (Hassler, et al. 2018; Hirano 2016).

As this thesis is primarily concerned with the structural biochemistry of cohesin, I will elaborate predominantly on the function of this complex and the underlying mechanisms, with some reference and comparison to other SMC-kleisin proteins where appropriate.

Due to significant temporal separation between genome duplication and segregation in eukaryotes, the sister chromatids produced during S phase need to be paired from the moment they are generated until their equal division into two daughter cells. The co-entrapment of

sister DNA strands by cohesin allows this complex to establish chromosome cohesion, a fundamental prerequisite for the equal distribution of genetic information to daughter cells during eukaryotic cell division (Nasmyth 2011).

Concurrent with the cell cycle, cohesin undergoes its own functional cycle. Cohesin is recruited to chromatin during G1 phase, and loaded on to chromatin by a dedicated loading apparatus; cohesion is then established in S phase during the synthesis of sister chromatids. In metazoans, cohesin is released in two steps. Most cohesin along chromosome arms is removed during prophase, allowing the emergence of X-shaped chromosomes. Finally, residual centromeric cohesin, which is necessary for kinetochore biorientation and satisfaction of the SAC, is released at the metaphase-to-anaphase transition by separase-mediated proteolysis. The concept of a 'cohesin cycle' is elaborated on further in section 1.4.

1.3 Organisation of the core cohesin complex

The core cohesin complex consists of an Smc1-Smc3 heterodimer bound to the kleisin protein Scc1 (sister chromatid cohesion 1) (Gligoris, et al. 2014; Haering, et al. 2002; Haering, et al. 2004), which in turn recruits the essential HEAT (Huntingtin elongation factor 3, A subunit of protein phosphatase 2A, TOR) repeat protein Scc3.

Smc1 and Smc3 feature two globular ABC-type ATPase 'head' domains assembled by their N- and C-termini. Each Smc head contains canonical Walker A, Walker B and Signature motifs. Dimerisation of the heads results in the formation of two composite ATPase sites, and is essential for ATP hydrolysis. The remainder of each Smc consists of an extended anti-parallel coil-coiled domain of 300-400 amino acids, with a length of approximately 50 nm that intersects the head domains. The central region of these coiled-coils folds into an SMC 'hinge' domain of ~150 amino acids, through which the Smc proteins heterodimerise.

The Scc1 subunit (also called Mcd1, mitotic chromosome determinant 1, or Rad21, radiation-sensitivity 21), forms a bridge between the Smc3 and Smc1 ATPase domains with which it interacts through its N-terminal helix-turn-helix and C-terminal winged-helix domains respectively. Thus, the ring is completed, and is able to wrap around sister chromatids. In addition to sealing the complex, Scc1 also provides docking sites for additional regulatory factors which are essential for the proper function and regulation of the complex.

Like other SMC-kleisins, cohesin complexes organise as large individual ring-like assemblies. The ring model was initially developed by visualization of the cohesin holocomplex by electron microscopy and through mapping interactions between its core subunits (Anderson, et al. 2002; Haering, et al. 2002) (Figure 1.4b). Visualisation by EM suggests an annular geometry, while biochemical experiments indicate that a full complement of inter-subunit contacts are essential as the individual disruption of any of the SMC-kleisin interfaces (Figure 1.2a) results in lethality (Arumugam, et al. 2006; Gligoris, et al. 2014; Haering, et al. 2004; White, et al. 2013; Xu, et al. 2010). Hence, the formation of a fully closed cohesin ‘ring’ is indispensable for its function. Recent evidence strongly supports the hypothesis that individual cohesin complexes are able to encircle individual DNA molecules as well as pairs of sister DNAs by topological embrace (Gligoris, et al. 2014; Huis in 't Veld, et al. 2014).

Several structures from different species illustrate the molecular determinants of cohesin ring assembly, in which the subunits form a closed circuit connected via the SMC hinge, and bridging of the head domains via the separate termini of Scc1. The structure of the *M.musculus* hinge domain, consisting of Smc1 and Smc3, is toroidal (Figure 1.2b, Table S1), harbours a positively charged channel between Smc1 and Smc3, and is thought to be required for DNA entry (Gruber, et al. 2006; Srinivasan, et al. 2018). SMC heterodimerization is conferred principally by the hinge domain, but the head domains are able to dimerize

transiently via bound nucleotide (as modelled in figure 1.2c). The structure of the Smc3-Scc1 complex shows that the N terminus of Scc1 assembles into two antiparallel helices that pack against the coiled-coil of Smc3 (Figure 1.2c, Table S1 (Gligoris, et al. 2014)). Disconnection of the Smc3-NScc1 interface is thought to release cohesin from chromosomes (Chan, et al. 2012; Eichinger, et al. 2013). The structure of the Smc1-Scc1 complex shows that the C-terminal winged helix domain (WHD) of Scc1 binds to the N-lobe of the Smc1 ATPase (Figure 1.2c).

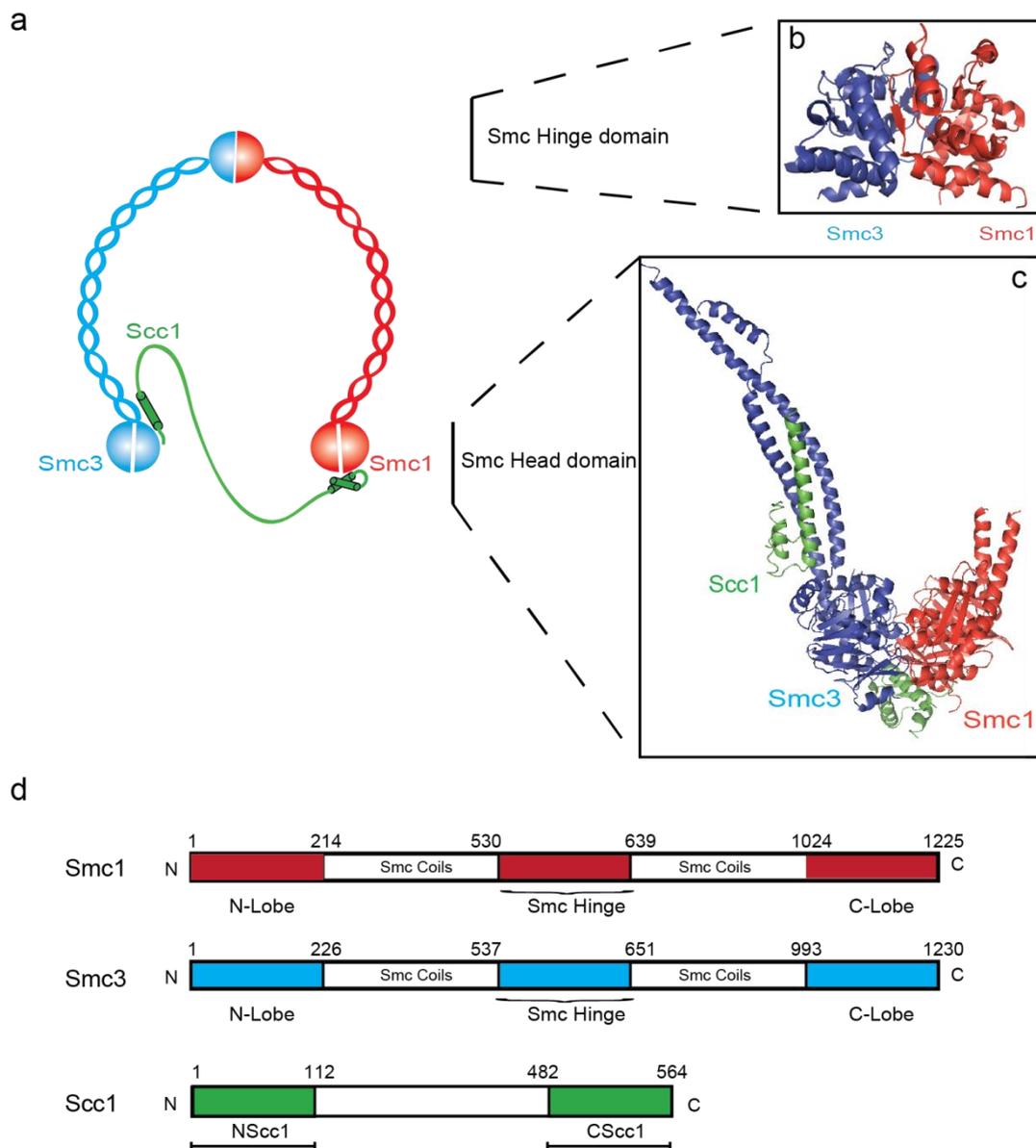


Figure 1.2 Structure of the cohesin core subunits and domain boundaries. (a) Model of the cohesin core subunits assembled as a ring-shaped complex. Heterodimerization of Smc3 (blue) and Smc1 (red) leads to a V-shaped conformation. (b) The structure of the hinge domain heterodimer. (c) Structural model of the assembled Smc1-3 ATPase heterodimer. Smc1-CScc1 (red-green) and Smc3-NScc1 (blue-green). (d) Cartoon showing domain boundaries of the Smc1, Smc3 and Scc1 subunits.

The essential HEAT-repeat subunit, Scc3, is recruited through the central region of Scc1. Since its discovery, Scc3 has been implicated in both cohesin loading and release, and is known as an essential cohesion maintenance factor (Hara, et al. 2014; Hu, et al. 2011; Murayama and Uhlmann 2014; Roig, et al. 2014; Rowland, et al. 2009).

Whilst yeast contains only one Scc3 isoform, there are three Scc3 paralogs in metazoans: SA1-SA3 (Stromal Antigen) (Sumara, et al. 2000). Evidence indicates that SA1 and SA2 have different roles in cohesin biology, and SA3 is a meiosis-specific cohesin subunit: SA1 associates with DNA at the telomeres and SA2 mostly at the centromeres. SA1 binds directly to telomeric DNA via its unique AT-hook which is not found in SA2 (Bisht, et al. 2013). SA1-containing cohesin has been implicated in gene regulation and replication of telomeres (Remeseiro, et al. 2012) and its deficiency drives aneuploidy and tumorigenesis in mouse due to its impaired replication of telomeres (Remeseiro et al. 2012). SA2, the dominant form in somatic cells, is mainly responsible for sister chromatid cohesion at the centromere (Remeseiro et al., 2012). Although both SA1 and SA2 are found at CCCTC-binding factor (CTCF) sites, a distinct population of the SA2-cohesin is linked to tissue-specific transcriptional regulation (Kojic, et al. 2018). It was also reported that downregulation of these two SA paralogs have different consequences for gene expression and chromosome structure: Downregulation of SA1 mainly impacts topologically associating domain (TAD)

boundaries whereas SA2 downregulation mainly impacts cell-type-specific enhancer-promoter contacts (Kojic, et al. 2018). SA3 (STAG3) is a germline-cell specific protein and functions during meiosis I (Pezzi, et al. 2000) and together with meiosis-specific cohesin, it influences meiotic chromosome structure and recombination (Adelfalk, et al. 2009; Ding, et al. 2006; Revenkova, et al. 2004).

Scc3 associates with chromatin during G1/S, and dissociates, with cohesin, in metaphase after being phosphorylated by Plk1. Plk1, a polo-like kinase, has been reported to be required for regulating a variety of mitotic and meiotic events, such as the metaphase/anaphase transition, cytokinesis and the onset of mitosis and spindle formation (Golsteyn, et al. 1995; Goto, et al. 2006; Randall, et al. 2007; van Vugt, et al. 2004). Plk1 is thought to phosphorylate cohesin subunit SA1/2 and Scc1 directly during prophase which results in the dissociation of the bulk of cohesin from chromosome arms (Figure 1.5) (Hauf, et al. 2005; Losada, et al. 1998; Sumara, et al. 2000; Sumara, et al. 2002).

The structure of the human SA2-Scc1 complex was determined in 2015, showing that SA2 and Scc1 form a 1:1 heterodimer. Scc1 interacts extensively with a highly conserved surface along the inner hook of SA2 (Hara, et al. 2014). This paper further identified an interaction between SA2-Scc1 and shugoshin1 (Sgo1), a cohesin protector at the centromere (Figure 1.3, Table S1). It was also shown that shugoshin and Wapl, the cohesin release factor, can compete for the same binding interface on SA2-Scc1 (Hara, et al. 2014). It is likely that interaction of Scc3 with cohesin regulators is an integral, albeit not completely understood aspect of Scc3 function (Beckouet, et al. 2016; Hara, et al. 2014; Roig, et al. 2014).

The binding of Scc3 to cohesin has also been shown to be essential for efficient loading of the complex onto chromatin (Hara, et al. 2014; Hu, et al. 2011; Murayama and Uhlmann 2014; Orgil, et al. 2015). The mechanism by which Scc3 promotes the association of cohesin with chromosomal DNA will be explored extensively in later chapters of this thesis.

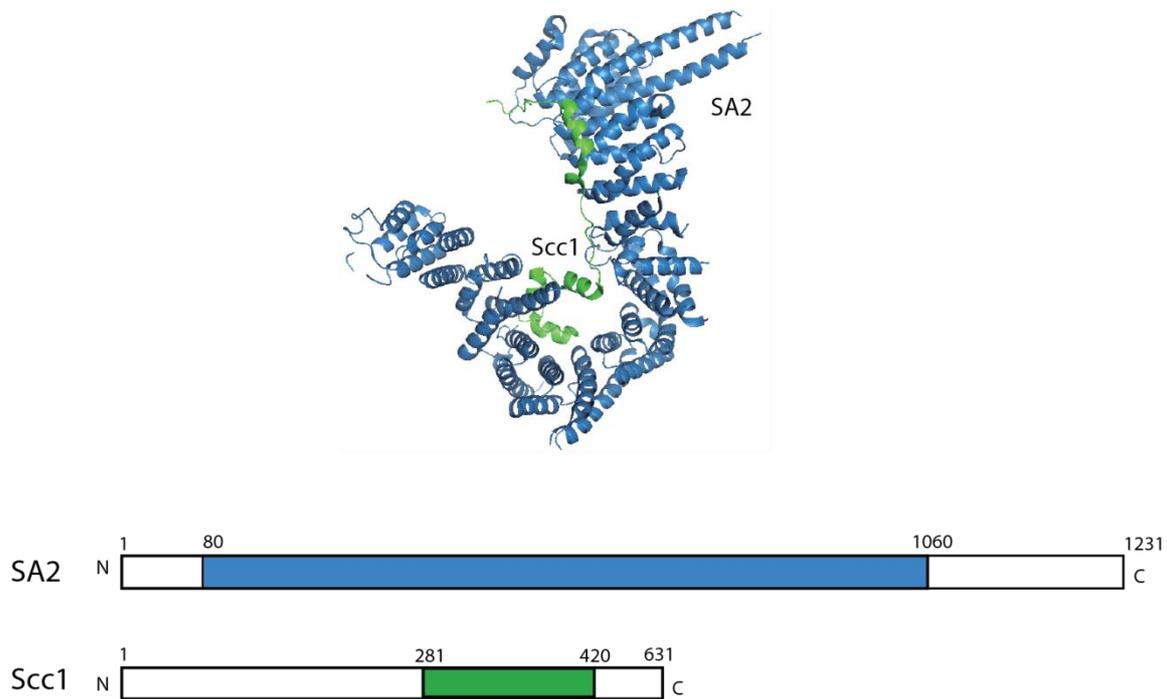


Figure 1.3 Structure of the human SA2-Scc1 complex and domain boundaries.

Scc1 (green) interacts with SA2 (blue) extensively along the cradle of the SA2 hook (PDB: 4PJU). A binding pocket which requires both SA2 and Scc1 is essential for shugoshin (Sgo1) association, thus competing with Wapl for cohesin binding. Phosphorylation of Sgo1 at residue T346 regulates Sgo1 association (Hara, et al. 2014).

1.4 Overview of the cohesin cycle

Following the cell cycle, cohesin also has its own cycle. During G1, cohesin is loaded onto chromatin with the help of the Scc2-Scc4 ‘loading’ complex, and is released again by the recruitment of a dissociation (unloading) factor, Wapl, which interacts with Pds5 and Scc3 to open the Smc3-Scc1 interface. Following loading, and during genome replication, an acetyltransferase, Eco1, acetylates the ATPase head domain of Smc3 to trigger the stabilization of cohesion (Rolef Ben-Shahar, et al. 2008; Unal, et al. 2008). In vertebrates, acetylated cohesin

recruits an additional factor, sororin, which further stabilizes cohesin during the maintenance phase by inhibiting the interaction of Wapl and Pds5 (Nishiyama, et al. 2010). In metazoans, cohesin is removed from sister chromatids in two steps. In the first step, at the beginning of prophase, cohesin is phosphorylated by mitotic kinases such as CDK1, Plk and aurora B, which displaces sororin, allowing Wapl to re-bind Pds5 and to release cohesin from chromosome arms (Dreier, et al. 2011; Hauf, et al. 2005). To permit assembly of the mitotic spindle, residual centromeric cohesion is protected by Sgo1, which counteracts cohesin phosphorylation by recruiting PP2A, and directly inhibits the association of Wapl with Scc3. Complete dissociation does not occur until all sister kinetochores have been attached to opposite poles of the mitotic spindle (Ramos, et al. 2010), via a second protease-dependent pathway. The attainment of this state satisfies the spindle assembly checkpoint, thus activating separase through APC-mediated degradation of its inhibitor securin (Cohen-Fix, et al. 1996; Yamamoto, et al. 1996). In this second step, at the metaphase-to-anaphase transition, centromeric cohesin is cleaved by the separase protease at two cleavage sites on Scc1 (Uhlmann, et al. 1999) releasing sister chromatids and allowing segregation into daughter cells (Nasmyth and Haering 2009; Peters, et al. 2008) (Figure 1.4, Table 1.1, Figure 1.5).

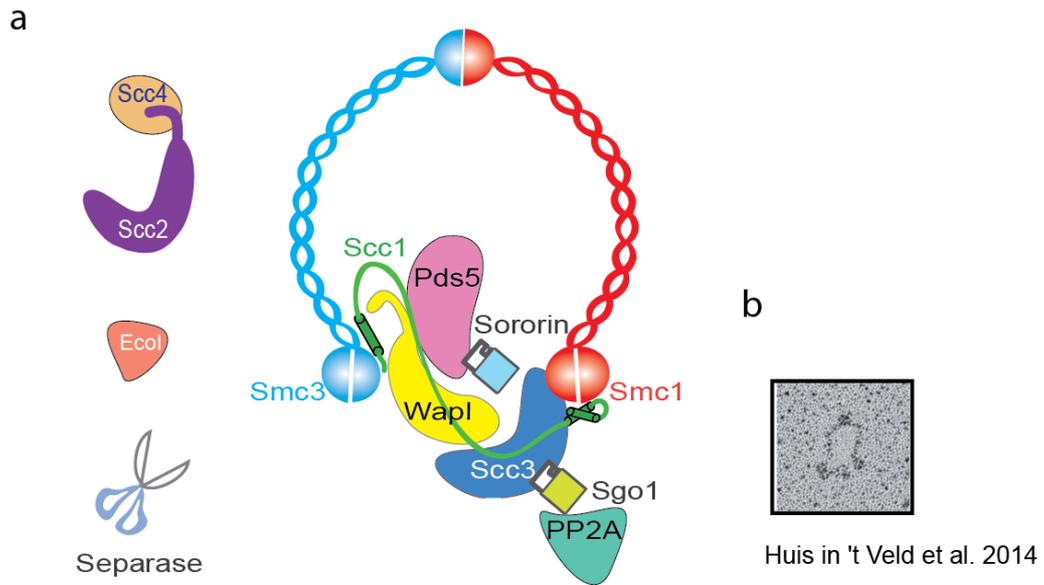


Figure 1.4 Cartoon illustration of the ring-shape cohesin and its accessory factors. (a) Cartoon of the cohesin core complex and its accessory factors. **(b)** Electron micrograph of the ring-shaped cohesin complex.

| | <i>Saccharomyces cerevisiae</i> | <i>Schizosaccharomyces pombe</i> | <i>Drosophila melanogaster</i> | <i>Xenopus laevis</i> | Human |
|-------------------------|---------------------------------|----------------------------------|--------------------------------|--------------------------|---------------------------|
| Cohesin subunits | SMC1 | PSM1 | SMC1 | SMC1 | SMC1 |
| | SMC3 | PSM3 | SMC3 | SMC3 | SMC3 |
| | MCD1 (SCC1) | RAD21 | RAD21 | RAD21 | RAD21 (SCC1) |
| | IRR1 (SCC3) | PSC3 | SA | SA1, SA2 | SA1(STAG1),SA2 (STAG2) |
| Loading | SCC2 | MIS4 | NIPBL | SCC2 | NIPBL |
| | SCC4 | SSL3 | N/C | xSCC4 | MAU2 (hSCC4) |
| Establishment | ECO1 (CTF7) | ESO1 | San, Deco | XECO1, XECO2 | EFO1(ESCO1), EFO2 (ESCO2) |
| Maintenance | PDS5 | PDS5 | PDS5 | PDS5A, PDS5B | PDS5A, PDS5B |
| | RAD61 | WPL1 | WAPL | N/C | WAPL |
| Dissolution | PDS1 | CUT2 | PIM | Securin | Securin (PTTG) |
| | ESP1 | CUT1 | Separase (SSE) | Separin | Separin (ESPL1) |
| | CDC5 | PLO1 | POLO | PLX1 | Plk1 |
| | SGO1 | SGO1, SGO2 | SGO1(Mei-S332) | Shugoshin-like 1 (xSGO1) | Shugoshin (hSGO1) |

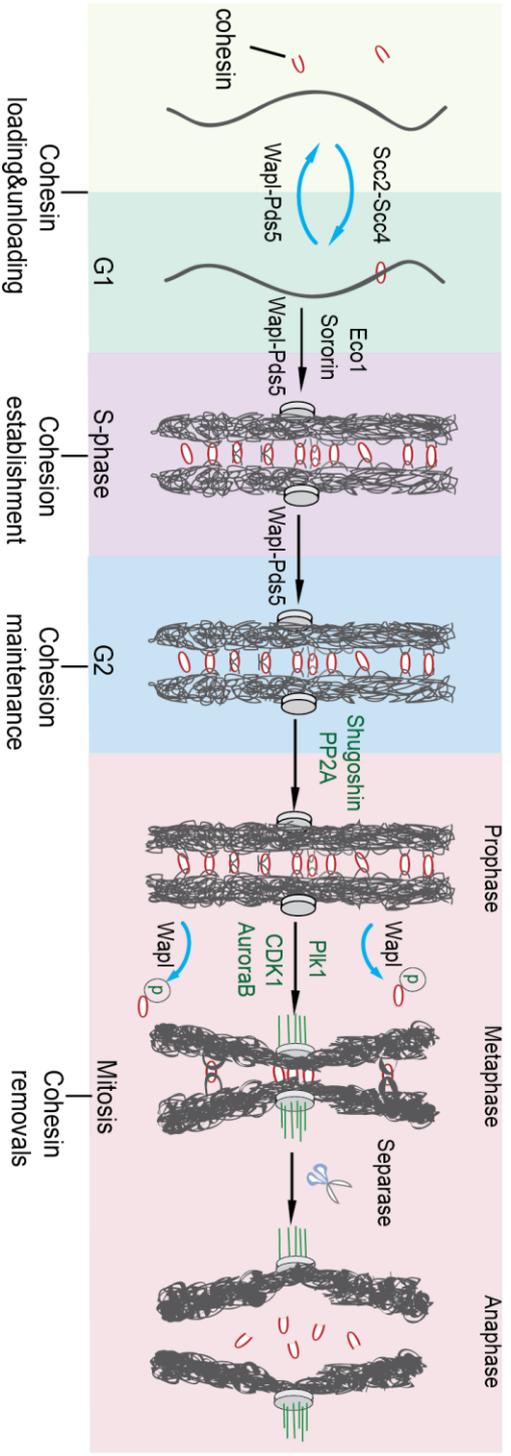
Table 1.1 Cohesin and its related factors in different species (adapted from (Onn, et al. 2008))

1.5 Cohesin ATPase dimerization

The ATPase domain of cohesin is evolutionarily related to ABC transporters (ATP-binding cassette transporters), which are typically membrane transporters that utilize the energy of adenosine triphosphate (ATP) binding and hydrolysis enable translocation of substrates across membranes. In cohesin, the substrate is chromosomal DNA, suggesting that the ATPase domain of Smc1 and Smc3 provide the motor activity (Arumugam, et al. 2003; Weitzer, et al. 2003) for DNA transport as well as loading of DNA into the cohesin ring to enable association with chromosomes (Arumugam, et al. 2003).

Essential ATP or DNA binding residues are predicted based on the structure of another ABC ATPase called Rad50, which also contains a Walker A-B motif, and shares similar architecture (Liu, et al. 2016). Further, mutation of amino acid residues in the head domain (SMC1-K39, D1157; SMC3-K38, D1154) abolishes ATP binding while mutation of residues SMC1-E1158 and SMC3-E1155 abolishes ATP hydrolysis and SMC1-S1130 and SMC3-S1128 impair ATP-dependent dimerization (Arumugam, et al. 2003). Also, DNA loading is thought to occur through transient opening of the hinge domain as artificial linking of the SMC1 and SMC3 hinge abolishes cohesin loading (Gruber, et al. 2006).

a



b

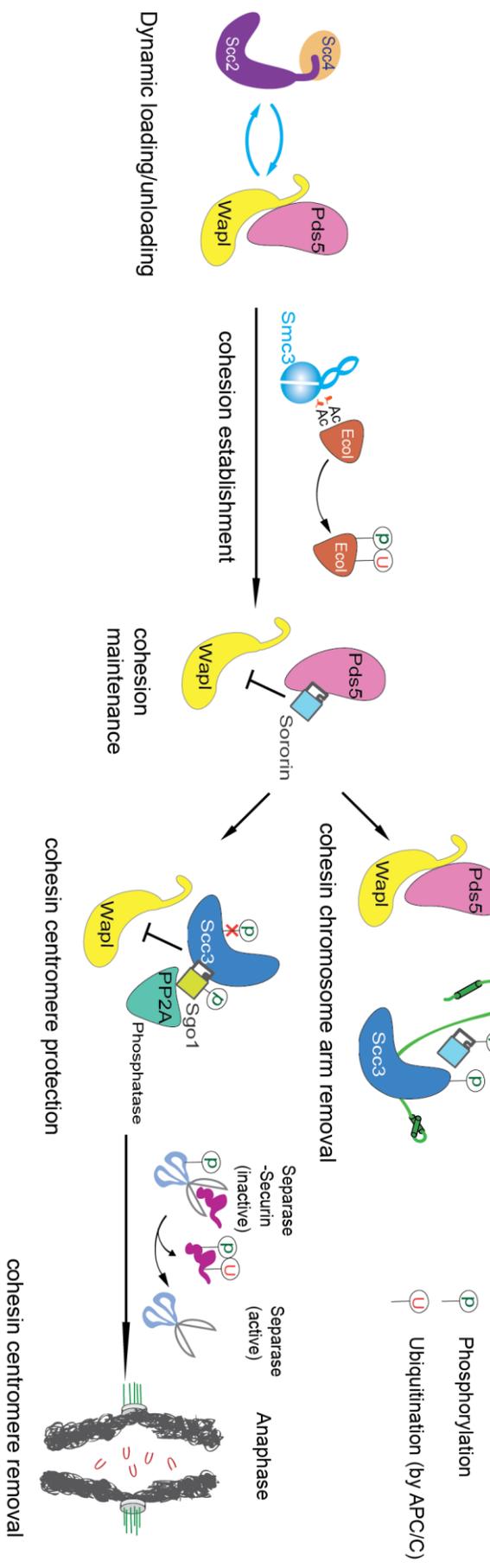


Figure 1.5 Cartoon of cohesin functional cycle and regulators. (a) From left to right: dynamic cohesin during G1 phase is regulated by the loading/unloading complex (Scc2-Scc4/Pds5-Wapl). Once the head domain of Smc3 is acetylated by Eco1, it triggers the establishment of cohesion; sororin antagonizes Wapl and binds to Pds5 to strengthen cohesion. From prophase to metaphase, cohesin subunits Scc1, sororin and Scc3, located at the chromosome arms, are phosphorylated, which allows the Wapl-Pds5 unloading complex to dissolve cohesion along chromosome arms. During the metaphase-to-anaphase transition, Scc1 is cleaved by separase which permits the segregation of sister chromatids into two daughter cells. (b) Cartoon of cohesin regulators. The cohesin cycle is dependent on cell cycle-dependent kinases and also the APC/C complex for appropriate temporal regulation. Figure adapted from (Haarhuis, et al. 2014).

1.6 The cohesin loading complex

The ability of cohesin to bind chromosomes depends on the Scc2–Scc4 complex, which is the loading factor for cohesin onto DNA (Ciosk, et al. 2000), and its deletion causes premature sister separation and abnormal chromosome segregation. Scc2-Scc4 forms a separate complex independently of cohesin (Kogut, et al. 2009). Scc2-Scc4 associates with cohesin during G1 through Scc1 and becomes dispensable during S and G2 phase when cohesin establishment and maintenance take place respectively. Scc2 contains a hook-shaped C terminal HEAT repeat domain, and a disordered N-terminal domain that snakes through the Scc4 subunit (Chao, et al. 2015; Hinshaw, et al. 2015; Kikuchi, et al. 2016). In budding yeast, cohesin is loaded at particular locations but quickly relocalizes to new locations, and it is supposed that cohesin could slide along the chromatin fiber to another site after loading (Glynn, et al. 2004; Lengronne, et al. 2004; Ocampo-Hafalla and Uhlmann 2011).

1.7 Cleavage-independent cohesin unloading

Dynamic cohesin release via Smc3-Scc1 gate opening is generally considered to depend on a ‘releasing’ subcomplex formed by Pds5 and Wapl (Gandhi, et al. 2006; Kueng, et al. 2006; Murayama and Uhlmann 2015; Ouyang, et al. 2016; Rowland, et al. 2009; Sutani, et al. 2009).

Pds5 (Precocious dissociation of sisters 5, BimD/Spo76), is an evolutionarily conserved from fungi to human (van Heemst, et al. 1999), containing a HEAT repeat domain. In vertebrates, deletion of Pds5 results in no discernible defects in chromosome arm cohesion but loosened cohesion at the centromere while overall cohesin levels are not affected. Thus Pds5 might play both positive and negative roles in sister chromatid cohesion (Losada, et al. 2005). Aside from forming the unloading complex with Wapl, Pds5 is also required to establish and maintain cohesion (Chan, et al. 2013; Hartman, et al. 2000).

Pds5 forms an elongated HEAT repeat protein that binds to Scc1 via a conserved surface patch and this interface is important for Pds5 recruitment to cohesin and mutations of this binding interface results in loss of sister chromatid cohesion and cell viability (Figure 1.6, Table S1) (Muir, et al. 2016).

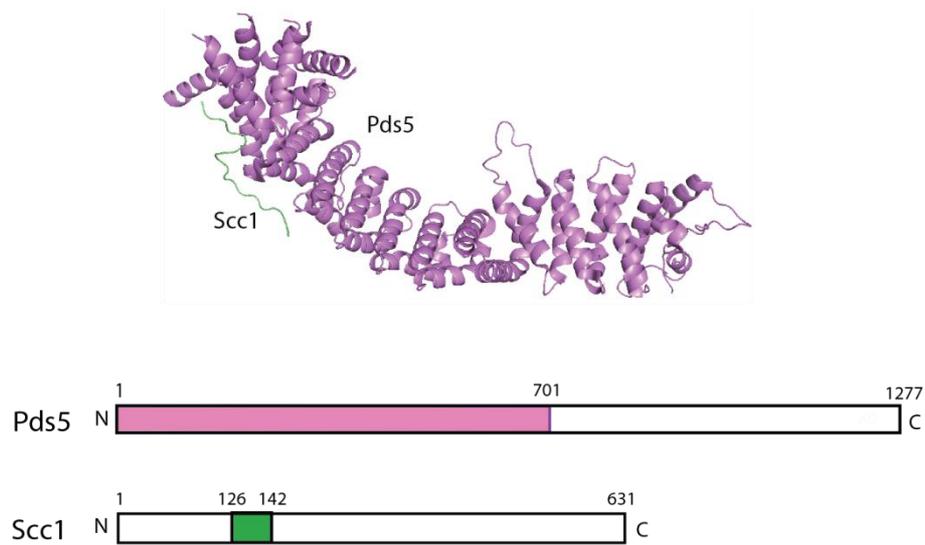


Figure 1.6 Structure of the yeast Pds5-Sccl complex and domain boundaries. Pds5 (purple) interacts with a discrete binding module on Sccl (green) (PDB: 5FRP), and this interaction is essential to maintain sister chromatid cohesion (Muir, et al. 2016).

As a subunit of the cohesin releasing complex, Wapl (wings apart-like) was originally identified as a gene product that regulates heterochromatin organization in *Drosophila* (Verni, et al. 2000). Its relevance for cohesion was discovered in 2006 (Gandhi, et al. 2006; Kueng, et al. 2006) when it was reported to promote the release of cohesin by direct interaction with Pds5 and Sccl3. Therefore the countervailing influences of Wapl-Pds5 and Sccl2-Sccl4, result in dynamic cohesin association (Kueng, et al. 2006). Further studies suggest that the well-conserved FGF motifs at the N-terminus of Wapl are responsible for the interaction between Pds5 and Sccl3 in vertebrates, but not in yeast (Shintomi and Hirano 2009).

Wapl proteins from different species have a very divergent N terminal domain with variable lengths, but conserved C terminal domains. The structure of the human Wapl C terminal domain shows an elongated shape containing HEAT repeats. Mutation of surface amino acids shows that Wapl is essential for cohesin release. It is also been reported that mutation of the FGF motifs of Wapl reduces its binding to Sccl3-Sccl1. However, truncation of the entire

N terminus comprising the FGF motifs (amino acids 1-500) still retained Scc3-Scc1 binding activity indicating that aside from the FGFs, there are additional regions that interact with Scc3-Scc1 (Figure 1.7, Table S1) (Ouyang, et al. 2013).

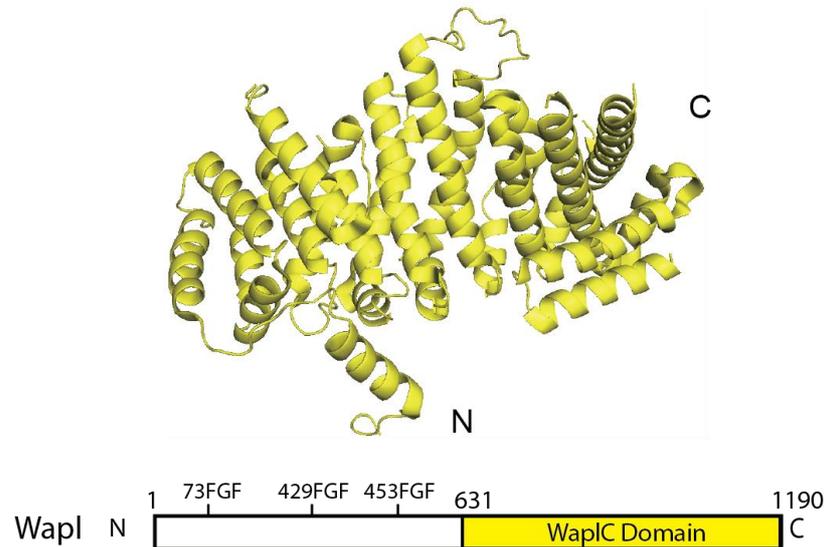


Figure 1.7 Structure of the human Wapl-C domain and domain boundaries. Wapl contains a flexible N terminal region and it forms extensive interactions with Pds5 and SA2-Scc1. The three FGF motifs distributed in the Wapl N-terminus contribute to, but are not strictly required for the Wapl-cohesin interaction. (Ouyang, et al. 2013).

1.8 SMC3 acetylation (establishment of cohesion)

To establish cohesion, the head domain of Smc3 is acetylated by Eco1 acetyltransferase at two evolutionarily conserved residues, Smc3-K112 and K113 in yeast, K105 and K106 in human, and these modifications arise only during S phase (Rolef Ben-Shahar, et al. 2008; Rowland, et al. 2009; Skibbens, et al. 1999; Toth, et al. 1999; Unal, et al. 2008; Zhang, et al. 2008). Eco1, also called Ctf7 (chromosome transmission fidelity) in yeast and ESCO1 and 2 in humans, was first identified for its involvement in cohesion in 1999 (Skibbens, et al. 1999) and directly interacts with components of the replication machinery (Kenna and Skibbens 2003). In Eco1 mutants, cohesin still associates with chromosomes but fails to build

cohesion, suggesting that Eco1 functions in the establishment of cohesion. Eco1 contains a zinc finger in the N-terminus and an acetyltransferase domain at the C-terminus (Toth, et al. 1999). Eco1 not only acetylates Smc3 during S phase, but *in vitro* evidence shows that Eco1 could also acetylate itself, Scc1 and Pds5 (Heidinger-Pauli, et al. 2010; Ivanov, et al. 2002; Ivanov, et al. 2018).

To limit cohesion establishment to S phase, Eco1 is phosphorylated by Cdk1 and thus becomes the substrate of the APC/C complex targeting for it for degradation (Lyons and Morgan 2011).

The structure of human Esco1 contains a catalytic acetyltransferase domain, which is shared with other member of the Gcn5-related N-acetyltransferase family (GNAT), and a N terminal zinc finger. An additional ‘loop insert’ (residues 727-766) is thought to bind Smc3 (Figure 1.8, Table S1).

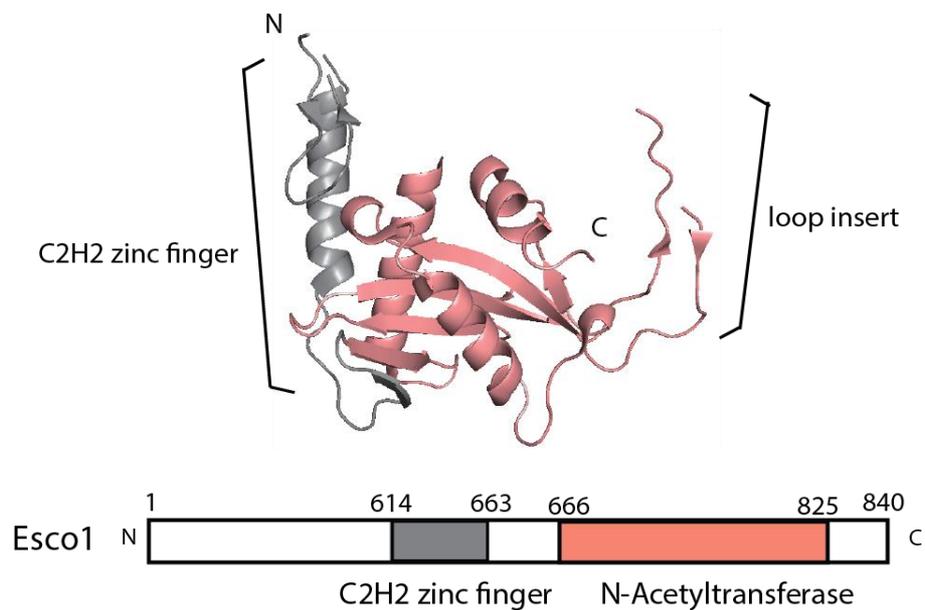


Figure 1.8 Structure of human Esco1 and domain boundaries. Esco1 acetylates Smc3 specifically at K105 and K106. Ac-CoA is bound to Esco1 (PDB:5T53) (Rivera-Colon, et al. 2016).

1.9 Cohesin stabilization

To stabilize cohesion, sororin (p35), is recruited via Smc3 acetylation at the DNA replication fork and stays connected throughout S and G2 phase (Schmitz, et al. 2007). To maintain cohesion, sororin forms a stabilizing complex with Pds5 during G2 phase (Figure 1.5) and it is found that sororin shares a FGF-like motif with Wapl (Wu, et al. 2011) and is supposed to be essential for association with Pds5. An additional Tyrosine-Serine-Arginine (YSR) motif found in both Wapl and sororin is also reported to bind to Pds5 at the same conserved site, thus explaining Wapl-sororin antagonism in cohesin regulation (Ouyang, et al. 2016).

1.10 Cohesin release at chromosome arms

Upon entry into mitosis, phosphorylation of sororin by Cdk1/CyclinB abolishes its ability to inhibit Wapl so that the Wapl-Pds5 complex can form again, allowing cohesin removal in prophase (Figure 1.5). Phosphorylated sororin then interacts with Plk1 via its Plk1 binding motif. Plk1 is thought to also regulate the dissociation of cohesin by phosphorylating Scc3 (Kuang, et al. 2006; Zhang, et al. 2011).

1.11 Cohesion protection at the centromere

In the Separase cleavage-independent release pathway, the centromeric cohesin is protected by another complex, formed by shugoshin and PP2A (Kitajima, et al. 2006; Riedel, et al. 2006) (Figure 1.5). Shugoshin, (a Japanese word for ‘guardian spirit’) is a centromere-associated protein conserved from yeast to humans and acts as a critical regulator of kinetochore assembly and centromeric cohesion (Kitajima, et al. 2004; Pouwels, et al. 2007; Watanabe and Kitajima 2005). Shugoshin is mostly localized at the centromeres by the conserved spindle checkpoint kinase Bub1, and Haspin, which install specific

phosphorylation marks on pericentric nucleosomes (Kerrebrock, et al. 1995; Kitajima, et al. 2005; Marston, et al. 2004; Tang, et al. 2004; Yamagishi, et al. 2010) (Figure 1.5). Shugoshin contains two highly-conserved domains, an N-terminal coiled coil and a C-terminal basic region. Shugoshin forms homo-oligomers through the coiled coil domain and recruits PP2A, which in turn dephosphorylates Scc3 at the centromere and prevents Plk1-dependent removal of shugoshin (Clarke, et al. 2005; Tang, et al. 2006). Phosphorylation of the human shugoshin (Sgo1) at residue T346 also allows to Sgo1 directly compete with Wapl to protect Scc3 (Hara, et al. 2014; Liu, et al. 2013).

The PP2A holo-enzyme contains a catalytic C, a scaffold A subunit, which binds to C directly, and a more varied B subunit. The structure of the Sgo1-PP2A complex reveals that the coiled coil domain of Sgo1 binds to C and B subunits of PP2A, and it turns out to be essential for centromere cohesion protection (Figure 1.9, Table S1).

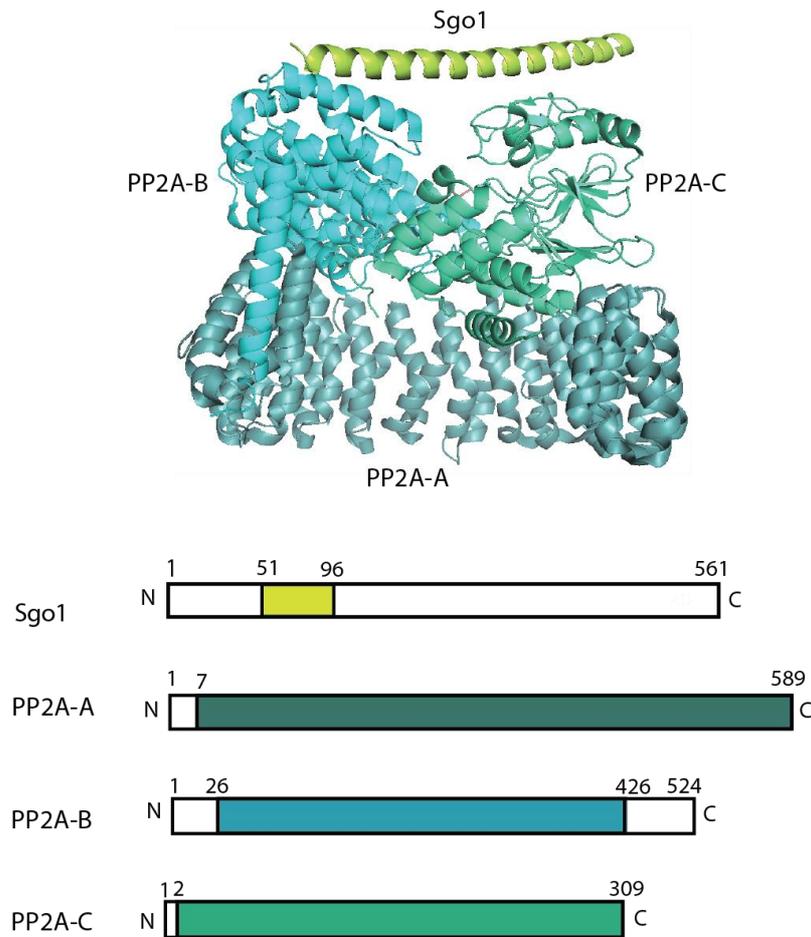


Figure 1.9 Structure of the Sgo1-PP2A complex and domain boundaries. Shugoshin (Sgo1) prevents precocious dissociation of cohesin from centromeres, PPA2, which prevents cleavage by separase via dephosphorylating cohesin. Sgo1 (yellow) forms a homodimeric parallel coiled-coil (not shown) that docks onto PP2A complex (multiple green colors) (PDB: 3FGA), and their interaction is essential for both separase cleavage protection and supporting spindle assembly checkpoint (Xu, et al. 2009).

Aside from this main function of shugoshin protection at the centromere, there is still a small, yet substantial, amount of cohesin bound at chromosome arms, which are protected by Sgo1. This small fraction of cohesin can only be removed by separase in order to completely remove cohesin (Nakajima, et al. 2007). The chromosome arms might comprise a mixture of cohesin complexes that require different mechanisms for their dissociation and thus coordinating chromosome separation and spindle elongation.

1.12 Cohesin cleavage at the centromere

Separase is a caspase-family protease that was initially identified as Esp1 in budding yeast and Cut1 in fission yeast (McGrew, et al. 1992; Uzawa, et al. 1990). Separase activation requires both its phosphorylation by Cdk1 (Stemmann, et al. 2001) and targeted-degradation of securin (Pds1/Cut2). Securin acts as a substrate mimic to inhibit separase activity (Boland, et al. 2017; Luo and Tong 2017). Separase associates with securin to stay inactive until the initiation of anaphase, then securin is removed through the ubiquitin protein ligase APC/C, releasing separase to allow cleavage of Scc1. Cohesin cleavage is also further activated by Cdc5, a Plk kinase, which phosphorylates Scc1 and thus enhances the cleavability by separase (Alexandru, et al. 2001; Clyne, et al. 2003; Hornig and Uhlmann 2004; Lee and Amon 2003; Lin, et al. 2016). The structure of the separase-securin complex from *Saccharomyces cerevisiae* shows a tight and extensive interaction, and residues 258-269 of securin occlude the separase active site, thus explaining how securin inhibits separase activity (Figure 1.10, Table S1).

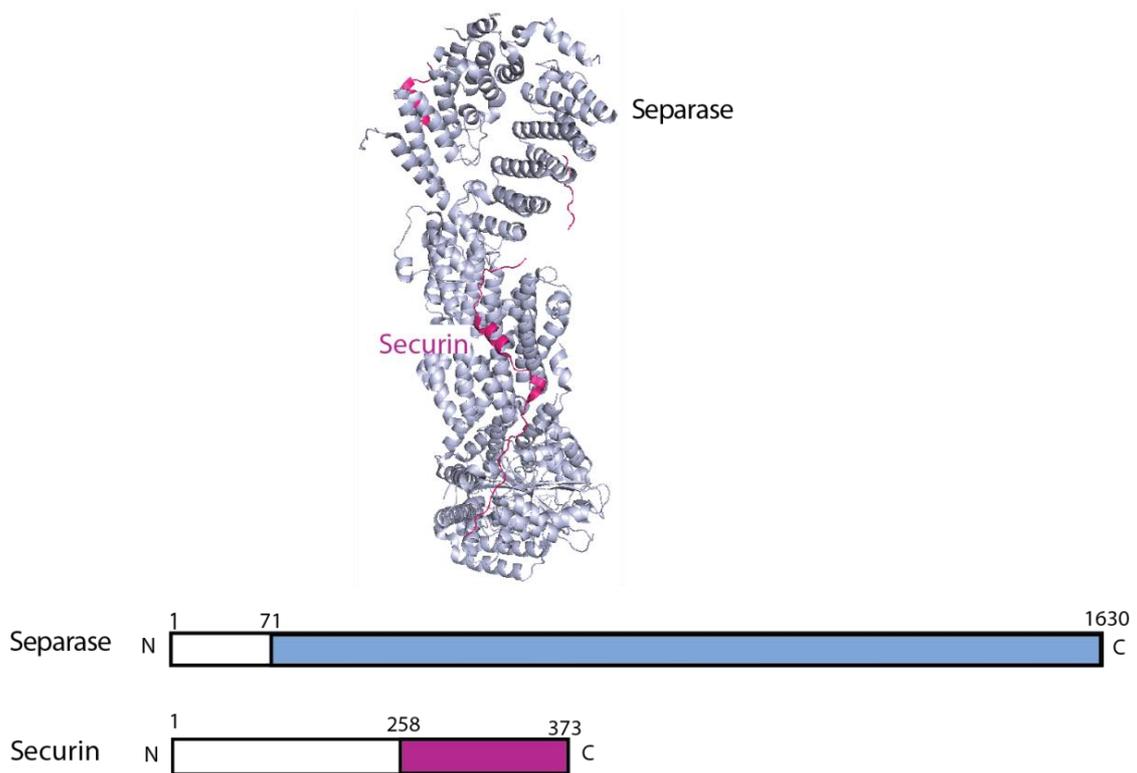


Figure 1.10 Structure of the yeast Separase-Securin complex and domain boundaries.

Separase cleavage activity on Scc1 is restricted by its inhibitor, securin (purple), which traverses the entire length of separase and blocks the active site (light blue) (PDB: 5U1T) (Luo and Tong 2017).

1.13 Cohesin beyond cohesion

As discussed, the co-entrapment of sister chromatids by cohesin allows the complex to establish sister chromatid cohesion (Nasmyth 2011). Beyond this activity, cohesin is also thought to regulate higher-order genome architecture through coordinated cycles of intra-chromatid DNA loop extrusion (Hassler, et al. 2018; Hirano 2016).

Several studies have shown that during interphase, cohesin might have additional divergent roles as a regulator of gene expression in different cell types. Cohesin also interacts with additional factors such as CTCF, the most well-known and most intensively studied insulator in vertebrates (Bell, et al. 1999). CTCF is reported to associate with cohesin through Scc3

(Xiao, et al. 2011). Additional cohesin binding partners include Mediator, a transcriptional coactivator that occupies enhancer and promoter regions at active genes sites and interacts with the cohesin loading factor Scc2 (Kagey, et al. 2010), and PRC1 (polycomb-group repressive complex 1), which is recruited by PRC2 to H3K27 methylated sites for gene silencing and it is proposed to interact with Scc1 (Strubbe, et al. 2011) to downregulate transcription.

1.14 CTCF

CTCF (CCCTC-binding factor) is a transcriptional insulator containing a tandem array of 11 zinc fingers and is universally conserved among Bilateria (Figure 1.11). It has sequence-specific DNA binding activity and plays a critical role in organizing genome structure and establishing gene expression patterns. CTCF was originally identified as a negative regulator of Myc expression (Klenova, et al. 1993; Lobanenko, et al. 1990) and subsequently shown to function more widely as an insulator protein. CTCF insulation function in general can be divided into two classes: enhancer-blocking insulators which, when placed between enhancers and promoters, prevent distant enhancers from activating a promoter (Filippova, et al. 2001; Hark, et al. 2000; Tanimoto, et al. 2003; Weth and Renkawitz 2011), or barrier insulators which appear to block heterochromatin from spreading (Cuddapah, et al. 2009; Downen, et al. 2014).

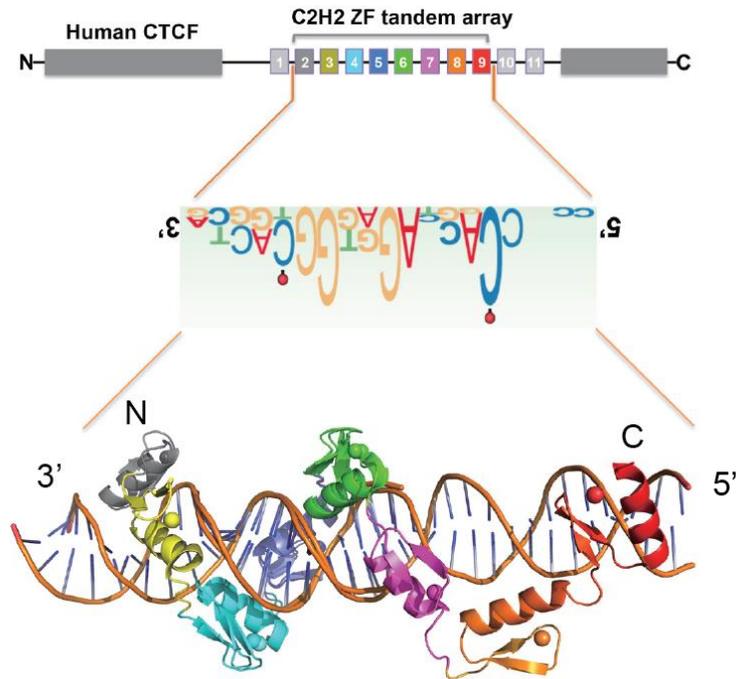


Figure 1.11 Model of CTCF zinc fingers 2-9 in complex with DNA. A tandem array of 11 zinc fingers of CTCF recognises one strand of double-stranded DNA from 3' to 5'. Zinc fingers ZF3 to ZF7 bind in the major groove of the 15bp core CTCF sequence motif. CTCF is methylation-sensitive at position 2 (close to 3' end, marked with red spot) of the core sequence, but insensitive at position 12 (close to 5' end, marked with red spot). ZF3, but not ZF8, contributes to DNA sequence-specific readout (Hashimoto, et al. 2017).

As an insulator, CTCF is involved in a number of different aspects of genome regulation. One such example is imprinting, in which genes show parent-of-origin-specific monoallelic expression, regulated by an imprinting control element called ICR (imprinting control region) (Ripoche, et al. 1997). The maternal chromosome ICR can bind to CTCF, which mediates silencing of the imprinted gene such as the *Igf2* and *H19* gene, but in the paternal chromosome, ICR is methylated which in return inhibits CTCF binding. The methylated ICR lacks insulator activity (Tremblay, et al. 1997; Xiao, et al. 2011) (Figure 1.12).

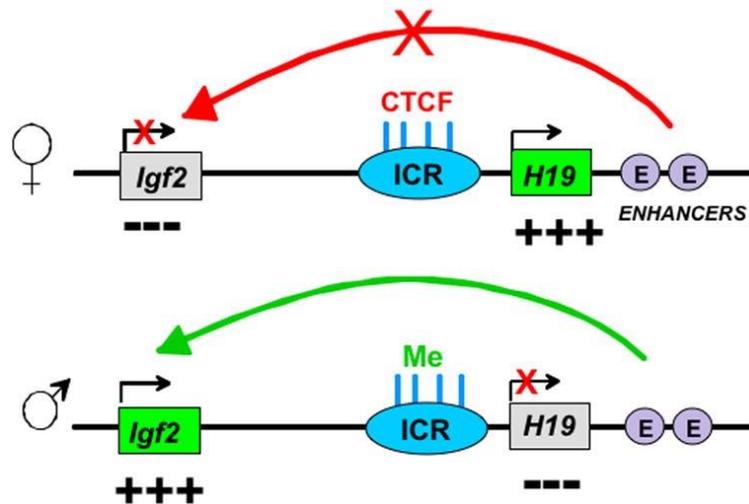


Figure 1.12 Role of CTCF in the regulation of paternal and maternal imprinting. In regulation of the *Igf2*-*H19* gene, the maternal chromosome ICR (imprinting control region) recruits CTCF, thus mediating silencing of the imprinting gene *Igf2* by bringing enhancers ('E') closer. Whereas in the paternal chromosome, the ICR is highly methylated ('Me') thus abolishing CTCF binding and insulation function (Wallace and Felsenfeld 2007).

CTCF is also able to recruit other protein factors. CTCF interacts with CHD8, a chromatin helicase protein, through its zinc finger domains, and this interaction is required for the insulator function (Ishihara, et al. 2006; Yusufzai, et al. 2004). It was also reported that both Sin3, a transcriptional repressor, and YB-1, a DNA/RNA-binding factor, also bind the zinc finger domain, indicating that the CTCF zinc finger domain is not just a DNA binder, but also promotes protein-protein interactions (Table 1.2).

| Protein name | Protein function | CTCF interaction domain |
|---|---|--|
| DNA-binding proteins and transcription factors | | |
| YB-1 | DNA/RNA-binding factor | Zinc fingers |
| Kaiso | Zinc-finger transcription factor | C-terminus |
| Yy1 | Zinc-finger transcription factor | All, but highest affinity for N-terminus |
| Histones and histone-modifying proteins | | |
| Sin3 | Transcriptional repressor, associated with histone deacetylases | Zinc fingers |
| CHD8 | Chromodomain helicase family member | Zinc fingers |
| Taf-1/Set | Molecular chaperone, associated with inhibition of histone acetyltransferases | N/A |
| H2A/H2A.Z | Nucleosome component | N/A |

Table 1.2 CTCF interaction partners (Wallace and Felsenfeld 2007).

Recent work identified RNA-binding regions (RBR) at ZF1, ZF10, as well as the C terminus of CTCF, showing that RBR mediates CTCF clustering. Loss of the RBRs appears to disrupt CTCF-mediated chromatin loops that are involved in the establishment of cell-type specific topology of chromatin, as well as gene expression profiles (Hansen, et al. 2018; Saldana-Meyer, et al. 2019).

1.15 Cohesin and CTCF

In addition to its crucial role in sister chromatid cohesion, cohesin is also involved in gene regulation (Wendt, et al. 2008). It was discovered that most cohesin binding sites in mammalian genomes are located proximal to CTCF motifs (Parelho, et al. 2008; Wendt, et al. 2008). CTCF is required for cohesin recruitment to CTCF sites, which in turn is required for formation of higher-order chromatin structure, thus indicating that cohesin has non-canonical functions that go beyond sister chromatid cohesion (de Wit, et al. 2015; Lengronne, et al. 2004). Much of our knowledge concerning long-range chromatin interactions comes from chromatin conformation capture (3C) (de Wit and de Laat 2012). 3C is the founding technique of a set of methods used to measure the spatial organization of chromatin, and it quantifies the number of interactions between genomic loci that are close to each other in 3D space but otherwise distant in sequence (Hakim and Misteli 2012). 3C and its derivatives have been applied to characterize many aspects of genome regulation, such as promoter-enhancer interactions. The main difference between 3C and its derivatives is their scope, either for quantifying two specific sequence fragments, such as the original 3C method (Dekker, et al. 2002), or to quantify interaction frequencies between all possible pairs of interactions simultaneously, such as Hi-C (high-throughput sequencing) (Hakim and Misteli 2012; Lieberman-Aiden, et al. 2009).

The topologically associating domains (TADs), which can be considered primary units of chromosome folding, are defined as spatially coincident regions of chromatin, as detected by 3C-based techniques. 3D genome architecture is then thought to arise as an emergent property of the organisation of many such TADs.

CTCF associates with chromosomes both at arms and centromeres and directly interacts with the Scc3/SA subunit (Rubio, et al. 2008). Most cohesin binding sites in lower organisms are found at regions of convergent transcription, but in many higher organisms possessing CTCF,

cohesin-CTCF loops can occur when CTCF binding sites are convergent (Figure 1.13a, b), both in the topological domain borders and the loops within such domains (Lengronne, et al. 2004; Phillips-Cremins and Corces 2013). While genome binding by CTCF is determined by the recognition of specific DNA sequences (Hou, et al. 2010), the positions of cohesin-CTCF defined loops tend to vary according to cell type. Depletion of CTCF reduces the enrichment of cohesin to specific sites but does not disrupt the presence of cohesin on chromatin (Hoque and Ishikawa 2002; Parelho, et al. 2008).

In terms of evolution, CTCF has only been identified in bilateria but cohesin is conserved from yeast to vertebrates, and is required for genome organisation in *Schizosaccharomyces Pombe* (Mizuguchi, et al. 2014), for example. Cohesin does not exhibit sequence-specific DNA-binding but it does colocalize with CTCF, which presumably confers sequence specificity by proxy. (Parelho, et al. 2008; Wendt, et al. 2008). Hence, it is probable that CTCF serves to position cohesin at specific regions within the genome, directing cohesin to modulate higher order chromatin architecture.

1.16 Loop extrusion model and different levels of the mammalian genome folding

Currently, the most prevalent model for genome organisation by SMC-kleisins is that these enzymes exploit their ATPase activity to structure chromatin by a process of DNA loop extrusion (Ganji, et al. 2018). The loop extrusion model has been further studied in the condensin system demonstrating that this complex is able to enlarge DNA loops *in vitro* in an asymmetric and processive fashion (Ganji, et al. 2018). Notably, this process was found to be dependent on the SMC ATPase and on a DNA anchoring function facilitated by the Ycg1, a paralogue of Scc3, HEAT-repeat subunit of condensin. As these molecular features are well

conserved, it is likely that the paralogous cohesin system, and perhaps indeed all SMC-kleisins, operate by similar principles.

According to the loop extrusion model, cohesin spontaneously extrudes DNA with energy from ATP binding and hydrolysis by the Smc3 and Smc1 subunits (Fudenberg, et al. 2016; Sanborn, et al. 2015). Such loop-extruding cohesin complexes accumulate at inward facing CTCF motifs thus forming a loop domain between convergently oriented CTCF motif pairs. Deletion of cohesin leads to disappearance of TADs (Busslinger, et al. 2017; Schwarzer, et al. 2017) but has no apparent effect on CTCF binding to chromatin (Rao, et al. 2017). Conversely, loss of CTCF causes TADs to disappear but for different reasons: without CTCF, cohesin extrusion continues but is no longer restricted to convergent CTCF binding sites (Wutz, et al. 2017). Thus CTCF must function upstream of cohesin to influence its localisation throughout the genome.

During interphase, de-condensed chromatin is organised into nuclear territories with characteristic gene densities and transcriptional activity. Euchromatin tends to cluster towards the center of the nucleus whereas heterochromatin accumulates towards the inner nuclear membrane and is associated with the nuclear lamina via CTCF (Cremer, et al. 2006; Reddy, et al. 2008). Such segmentation divides chromosomes into two distinct compartments which were first discovered in Hi-C studies: compartment A encompasses transcriptionally active regions, whereas compartment B contains inactive genes (Bickmore and van Steensel 2013; Fortin and Hansen 2015; Lieberman-Aiden, et al. 2009). These compartments can encompass TADs that are either on the same chromosome (*cis*) (Figure 1.13) or different chromosomes (*trans*) (Simonis, et al. 2006). Although both cohesin and CTCF can direct TAD formation, neither cohesin nor CTCF alone affect compartmentalization (Seitan, et al. 2013; Sofueva, et al. 2013; Zuin, et al. 2014). It could only be disrupted by depletion of both CTCF and

cohesin, which causes chromatin compaction (Haarhuis and Rowland 2017; Tark-Dame, et al. 2014).

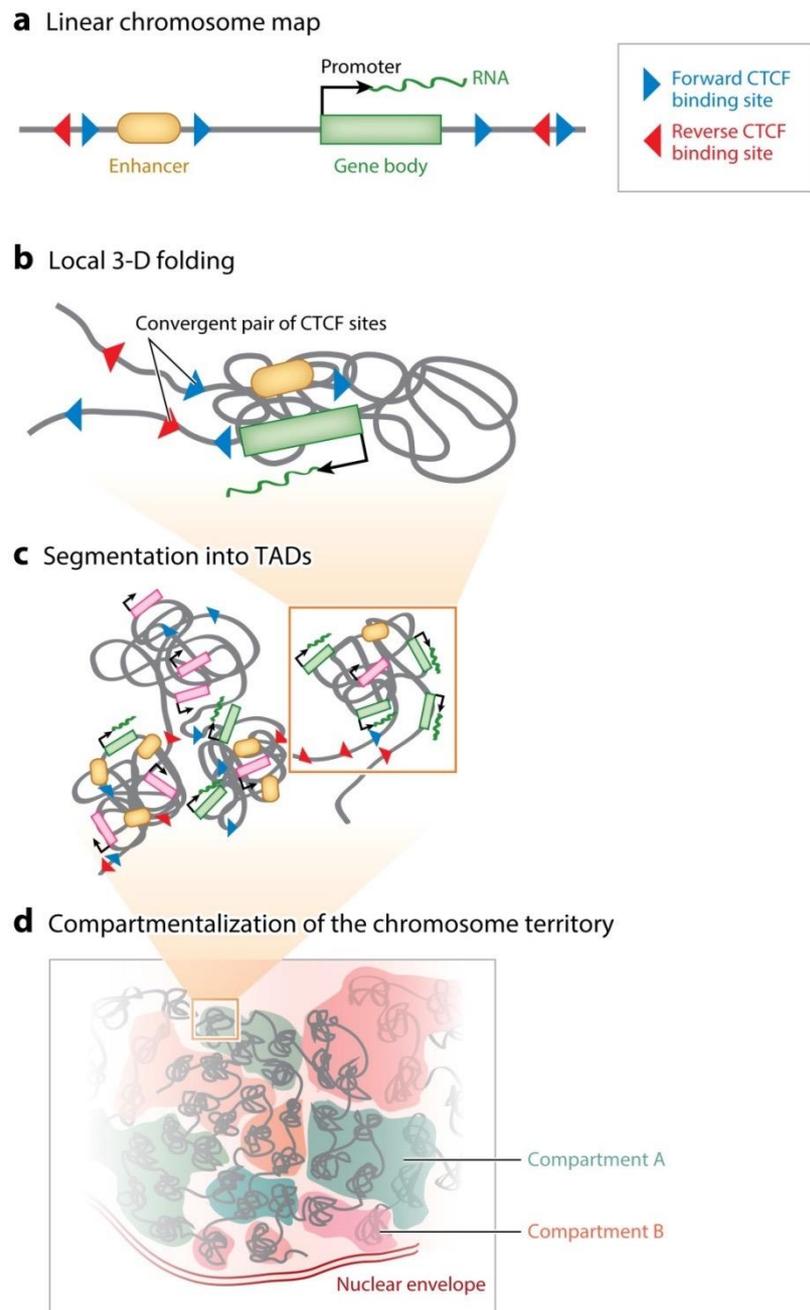


Figure 1.13 Levels of Genome folding. (a) Elements involved in cohesin-CTCF looping along chromosome. (b) CTCF helps form loops in a convergent orientation. (c) Segmentation of loops into larger TADs, which is defined by 3-C methods. (d) Active and inactive genes tend to be connected thus forming compartments (compartment A: active gene, tend to locate nuclear center; compartment

B: inactive genes tend to be located toward the periphery of the nucleus associated with the nuclear lamina) (Merkenschlager and Nora 2016).

Recent studies have provided greater mechanistic detail, suggesting that cohesin is required for the formation of TADs and loops but restricts compartmentalization. Wapl and Pds5 antagonize these functions: Wapl and Pds5 determine the size and the boundary of the DNA loops, and also facilitate the proper orientation of CTCF sites. The loading complex Scc2-Scc4 in turn promotes the extension of the loops which makes loop formation dynamic but more elaborate (Haarhuis, et al. 2017; Wutz, et al. 2017).

Scientific aims

When I began my Phd work, structural information of cohesin was available for the core subunits and some accessory factors. Several models have been proposed for how cohesin rings might hold sister chromatids together, yet the precise topology and mechanism of action remains unclear. Aside from the ring shaped particles that could be observed in EM images, virtually nothing is known about structural arrangement of the holo-cohesin complex or how cohesin coordinates its activity with its accessory factors throughout the cell cycle. Therefore, the aim of my thesis project was to assemble the holo-cohesin complex and to characterize it structurally and biochemically. It is known that cohesin exhibits a weak affinity for single and double stranded DNA yet at the same time does not show any sequence specificity (Fernius and Marston 2009; Murayama, et al. 2018; Sakai, et al. 2003). As there is no structural information explaining how cohesin engages DNA or nucleosomes directly, I also set out to investigate how cohesin structurally associates with DNA.

Materials and Methods

Matériels et méthodes

L'objectif de cette thèse était de tirer des informations structurales et biochimiques sur le complexe de la cohésine. Dans cette section, je décris les matériels et les méthodes relatifs à cet objectif. Pour faciliter l'isolement des protéines et des complexes protéiques produites de manière recombinants, des amorces ont été synthétisées pour amplifier les gènes d'intérêt par PCR, puis les fragments d'ADN digérés par des enzymes de restriction ont été insérés dans des plasmides cibles par ligation enzymatique. Pour obtenir les protéines d'intérêts, l'expression et la purification des protéines, ont été effectuées, les protéines sont ensuite concentrées et congelées dans de l'azote liquide pour le stockage. Pour déterminer la structure des molécules biologiques d'intérêt, des essais de cristallisation des protéines ainsi que plusieurs cycles d'optimisation des différentes conditions et de marquage avec de la sélénométhionine ont été réalisés. Les techniques utilisées pour comprendre les interactions protéine-protéine et protéine-ADN, y compris la co-expression/co-purification de protéines, des expériences de pull-down, des peptides arrays, des expériences EMSA et la polarisation de fluorescence (FP) sont résumées. Afin de comprendre et d'identifier les résidus essentiels pour les interactions protéine-protéine ou ADN-protéine, nous avons effectué de la mutagenèse dirigée. La caractérisation in vivo des mutants déterminés à partir des structures obtenues, notamment les analyses de ChIp-qPCR et de 'tetrad', ont été réalisées en collaboration avec d'autres équipes.

2.1 Construct Cloning

Most yeast genes were amplified from yeast Genomic DNA (Millipore) by PCR, unless specified otherwise. Codon optimised (for *E.coli*) genes comprising *A. gossypii* (ag), *S. pombe* (sp), and *C. thermophilum* (ct) Smc1head, and the ySmc3 head were produced by gene synthesis (Thermofisher), Pds5 and Wapl were cloned and expressed as described previously (Muir, et al. 2016), human SA2 inserted into pGEX-6p vector was provided by Hongtao Yu (UT Southwestern, Dallas) and is codon optimized for expression in *E.coli*. Full length human CTCF was ordered from Addgene, and different constructs amplified by PCR. The gene encoding human Shugoshin was synthesised and codon optimised for expression in *E.coli* (Thermofisher). Vectors containing cDNAs encoding human Scc1 and Wapl were a gift from Jan-Michael Peters (IMP, Vienna); constructs arising were subcloned into pACYC and pGEX-6p respectively.

For a list of all vectors, tags, linkers used and construct boundaries or mutants generated, please see Table S2. Site-directed mutagenesis, where required, was performed with the QuikChange® Lightning kit (Agilent Technologies).

2.2 Protein expression

2.2.1 Protein expression

All proteins were expressed in *E.coli* BL21(DE3). Expression was induced by auto-induction (Studier 2005). Cells were grown at 37°C until $OD_{600nm} = 0.6$ and then shifted to 18°C for 16 hours. Cells were harvested in a JLA-8.1 (Beckman) centrifuge at 4000xg and washed once with ice-cold PBS buffer. For protein containing His tags pellets were resuspended in buffer 1 (40 mM TRIS, pH 7.5, 500 mM NaCl, 20 mM imidazole, 0.5 mM TCEP) containing one tablet of complete mini EDTA-free protease inhibitors (Roche) and loaded on Co^{2+} beads

(IMAC Sepharose Fast Flow 6 resin (GE Healthcare), For GST-tagged proteins pellets were resuspended in buffer S (same as buffer 1 but without imidazole). For the expression and purification of Pds5fl, media and buffers were supplemented with 20 μ M or 5 μ M of inositol hexa-kis-phosphate (IP6), respectively (Ouyang, et al. 2016). CTCF constructs containing Zinc fingers were expressed in media supplemented with 50 μ M $ZnCl_2$. The purification buffers were supplemented with 5 μ M $ZnCl_2$.

2.2.2 Selenomethionine-labelled protein expression

Vectors containing Scc3T-Scc1K constructs were transformed into BL21(*DE3*), cultured in 2 ml LB until $OD_{600nm} = 0.6$. Cells were harvested and washed twice with M9 media, then transferred into M9 media supplemented with trace elements and vitamins as preculture (modified from (Van Duyne, et al. 1993)), and grown overnight at 37 °C. Cells were then transferred into 1L cultures of M9 (supplemented with trace elements and vitamins), and grown at 37 °C until $OD_{600nm} = 0.6$, amino acids were added to induce feedback inhibition of methionine synthesis, then IPTG with final concentration of 0.5 mM and 2.5 mM selenomethionine were added for the expression of the seleno- labelled Scc3T-Scc1K.

2.3 Protein purification

Cells were lysed using a microfluidiser (Microfluidics) and centrifuged at 15000 rpm for 1h using a JA-20/14 rotor (Beckman). Lysis procedures were identical for all protein purifications.

2.3.1 Purification with His or His-GST tag

The supernatant was loaded on 5 ml Co^{2+} -conjugated IMAC beads (GE healthcare) by using a peristaltic pump (GILSON). The column was washed with 10 column volumes of buffer 1 and the protein was eluted with buffer 2 (40 mM TRIS, pH 7.5, 300 mM NaCl, 300 mM

imidazole, 0.5 mM TCEP). The His-GST tag was cleaved by addition of His-tagged TEV protease (1:100 w/w) during overnight dialysis against 40 mM TRIS, pH 7.5, 300 mM NaCl, 0.5 mM TCEP at 4°C. For Smc3-NSccl, Smc1-CSccl, and the Smc3-Smc1 hinge, this cleavage step was bypassed. The His-GST tag, protease and uncleaved protein were removed by passing this mixture over Co²⁺ IMAC resin, and target proteins recovered from the flow-through.

2.3.2 Purification with GST tag

Supernatant was loaded on to 5 ml Glutathione Sepharose 4 Fast Flow resin (GE Healthcare), washed with 10 column volumes of buffer S, and protein eluted with GST elution buffer (40 mM HEPES pH 7.5, 0.3 M NaCl, 10 mM -Glutathione and 0.5 mM TCEP). The GST tag was cleaved by His-tagged 3C protease (provided by Protein Expression and Purification Core Facility, EMBL-Heidelberg) during overnight dialysis against 40 mM TRIS, pH 7.5, 300 mM NaCl, 0.5 mM TCEP at 4°C. Protease and uncleaved protein were removed by passing this mixture over Glutathione Sepharose 4 Fast Flow resin.

2.3.3 Ion-exchange purification

The flow-through was concentrated using an Amicon Ultra -15 concentrator (Millipore) and applied onto a MonoQ 5/50 GL column (GE healthcare) in buffer 3 (150 mM NaCl, 40 mM TRIS, pH 7.5, 0.5 mM TCEP). Proteins were eluted using a linear gradient in buffer 4 (1 M NaCl, 40 mM TRIS, pH 7.5, 0.5 mM TCEP). Proteins with a tendency to precipitate (ctSmc1head, all Clink complexes) were treated differently by diluting them in 250 mM NaCl, 40 mM TRIS, pH 7.5, 0.5 mM TCEP prior to loading onto a MonoQ column.

2.3.4 Size exclusion

The final purification step was performed by using a HiLoad 16/60 Superdex 75/200 prep-grade column in buffer 5 (150 mM NaCl, 20 mM TRIS, pH 7.5, 0.5 mM TCEP). Exceptions

were ctSmc1hd-CSccl1 and all Clink complexes which were eluted with buffer containing 300 mM NaCl, 20 mM TRIS, pH7.5, 0.5 mM TCEP, and human SA2T-Sccl1H with buffer containing 300 mM NaCl, 20 mM TRIS, pH7.7, 5 mM TCEP. Purified proteins were concentrated and flash frozen in liquid N2 for storage at -80 °C. The integrity and purity of protein samples were validated by SDS-PAGE.

2.4 Crystallisation

All initial crystallisation trials except for Sccl3T-Sccl1K-DNA complexes were performed by the EMBL Grenoble HTX facility with standard commercial screens. Sccl3T-Sccl1K-DNA samples were screened using a custom plate setup, optimized for protein-DNA complexes (Table 2.1).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | |
|--|------------------|---|---|---|---|---|-----------------|---|---|----|----|----|----------------------------|
| tri-Sodium citrate (100mM, pH5.6) | PEG 3350 (5-25%) | | | | | | PEG 3350(5-25%) | | | | | | HEPES (100mM, pH7.5) |
| | PEG 4000(5-25%) | | | | | | PEG 4000(5-25%) | | | | | | |
| | PEG 6000(5-25%) | | | | | | PEG 6000(5-25%) | | | | | | |
| | PEG 8000(5-25%) | | | | | | PEG 8000(5-25%) | | | | | | |
| MES (100mM, pH6.5) | PEG 3350(5-25%) | | | | | | PEG 3350(5-25%) | | | | | | TRIS (100mM, pH8.5) |
| | PEG 4000(5-25%) | | | | | | PEG 4000(5-25%) | | | | | | |
| | PEG 6000(5-25%) | | | | | | PEG 6000(5-25%) | | | | | | |
| | PEG 8000(5-25%) | | | | | | PEG 8000(5-25%) | | | | | | |

Table 2.1 Sccl3T-Sccl1K crystallisation screen conditions.

2.4.1 Crystals of ctSmc1hd-CSccl1 complex were obtained at 4 °C by hanging-drop vapour diffusion, after mixing equal volumes of protein at 15 mg ml⁻¹ and precipitant containing 0.1 M TRIS (pH8.5), 30% PEG300. During harvesting, crystals were transferred into well solution containing 15% Glycerol and flash frozen in liquid nitrogen.

2.4.2 Crystals of the Scc3-Scc1-DNA complex were obtained by mixing 8 mg ml⁻¹ protein with DNA at 1:1.1 ratio. 1 µl of the protein: DNA complex was mixed with 1 µl 10% PEG 8000, 0.1 M Bis-TRIS, pH 6.5 crystallization buffer and equilibrated against the crystallization buffer at 4 °C by hanging-drop vapour diffusion. Initial crystals with a 17 bp DNA substrate were obtained after 5 days. These crystals were used as seeds for crystallization of Scc3-Scc1 bound to a 19 bp DNA fragment using the same crystallization condition.

2.4.3 Crystal seeding

For crystal seeding, 10-20 crystals containing the Scc3-Scc1 complex bound to 17bp DNA were washed and cracked into small pieces within the crystallization buffer in a volume of around 40 µl. Then 0.4 µl of the seeds were mixed with 0.8 µl Scc3-Scc1 with a DNA oligonucleotide of different length and 0.8 µl of crystallization solution followed by equilibration against crystallization buffer at 4 °C.

2.4.4 Crystals of the SA2T-Scc1H complex were grown by hanging-drop vapour diffusion at 20 °C by mixing equal volumes (1 µl) of protein at 8 mg ml⁻¹ and crystallization solution (1 µl) containing 0.06 M Morpheus Divalents mix, 0.1 M Morpheus buffer system2 and 48% Morpheus EOD_P8K. Crystals were soaked for 24-48 h with peptides (peptid.de) spanning amino acid residues 220-230 of CTCF (Uniprot ID: Q8NI51, sequence: DVSVYDFEEE) or 331-341 of shugoshin (Uniprot ID: Q5FBB7; SNDAYNFNLEE). The crystals were cryo protected by adding 15% glycerol to the crystallization buffer followed by flash freezing in liquid nitrogen.

2.5 Structure determination

2.5.1 Structure of the ctSmc1-CSc1 complex

Diffraction data of the ctSmc1-CSc1 crystals were recorded with a Pilatus3 2M detector at the European Synchrotron Radiation facility on beamline ID30A-1/MASSIF-1 (Bowler, et al. 2015) using an X-ray wave-length 0.966 Å at 100 K. Location and optimal centering of crystals were determined automatically as described previously (Svensson, et al. 2015). The beam diameter was selected automatically to match the crystal volume of highest homogeneous quality (Svensson, et al. 2018). Data were processed with XDS (Kabsch 2010) and imported into CCP4 format using AIMLESS (Winn, et al. 2011).

Phases were determined using the previously published ySmc1-CSc1 complex (Haering, et al. 2004) as a search model for molecular replacement in Phaser (McCoy, et al. 2007). Iterative manual model building and refinement in Coot and PHENIX, respectively, were pursued to derive a final model. The final CtSmc1-Sc1 model was refined to 2.09 Å resolution with an R_{work} and an R_{free} of 21% and 24%, respectively (Table 3.1). Analysis of the refined structure by MolProbity showed that there are no residues in disallowed regions of the Ramachandran plot. The MolProbity all atom clash score was 2.5.

2.5.2 Structure of the Sc3-Sc1-DNA complex

Diffraction data for all native and selenomethionine-derivatised (Sc3-Sc1-DNA) crystals were collected at 100 K at an X-ray wavelength of 0.966 Å at beamline ID30A-1/MASSIF-1 (Bowler, et al. 2015) of the European Synchrotron Radiation Facility, with a Pilatus3 2M detector using automatic protocols for the location and optimal centring of crystals (Svensson, et al. 2015). The beam diameter was selected automatically to match the crystal volume of highest homogeneous quality (Svensson, et al. 2018). Access to a fully automatic beamline was important as only a few crystals in a hundred would diffract sufficiently. Diffraction from the best native crystal was anisotropic, with the best direction extending to a

minimal Bragg spacing of ~ 3.6 Å and only ~ 5.7 Å in the worst direction. At this resolution we had to confirm the sequence register by crystallisation of Selenomethionine-derivatized protein in order to use the anomalous signal to assign the sequence register. This required further automatic screening on MASSIF-1 to collect the best data possible. In total, I estimate that over 800 crystals were required to produce the final model of the Scc3-Scc1-DNA complex. Data were processed with XDS (Kabsch 2010) and imported into CCP4 format using AIMLESS (Winn, et al. 2011).

2.5.3 Structure of the SA2-Scc1-CTCF complex

Diffraction data for SA2-Scc1-CTCF crystals were collected at 100 K at an X-ray wavelength of 0.966 Å at beamline ID30A-1/MASSIF-1 (Bowler, et al. 2015) of the European Synchrotron Radiation Facility, with a Pilatus3 2M detector using automatic protocols for the location and optimal centering of crystals (Svensson, et al. 2015). The beam diameter was selected automatically to match the crystal volume of highest homogeneous quality (Svensson, et al. 2018). Data were processed with XDS (Kabsch 2010) and imported into CCP4 format using AIMLESS (Winn, et al. 2011).

The structure was determined by molecular replacement using Phaser. A final model was produced by iterative rounds of manual model building in Coot and refinement using PHENIX. The CTCF-containing model was refined to 2.6 Å resolution with an R_{work} and an R_{free} of 25% and 27%, respectively (Table 5.1). Analysis by MolProbity showed that there are no residues in disallowed regions of the Ramachandran plot and the all atom clash score was 7.2.

2.6 Bioinformatic methods

2.6.1 Sequence alignment

Collections of orthologous proteins were aligned on the EMBL-EBI ClustalOmega webserver (Sievers and Higgins 2014) and visualised using JalView.

2.7 Protein biochemistry

2.7.1 In vitro pull down

All *in vitro* pull down and competition assays were performed similarly.

2.7.1.1 GST pull-down

GST tagged and target proteins were mixed at a final concentration of 2.5 μ M, in binding buffer (150 mM NaCl, 20 mM TRIS, pH 7.5, 0.5 mM TCEP, 0.1% Tween20), and in a final volume of 50 μ l containing 25 μ l GST Sepharose beads per reaction. Reactions were incubated at 4 °C for 1 h, then 25 μ l of the reaction were taken as input ('I'), the remaining 25 μ l were washed and spun down 5 times with 500 μ l of the binding buffer. Samples were boiled for 5 min before loading on a gradient 4-20% SDS-PAGE (GenScript, Cat: M42015). Exception: for pulling down SA2-Scc1 with CTCF or shugoshin (Sgo1), 10 μ M of each GST-tagged protein was used in order to detect the interaction.

2.7.1. 2 His-tag pull-down

His-tag pull-down experiments were carried out essentially identically to GST pull-downs, except with a different buffer (150 mM NaCl, 20 mM TRIS, pH 7.5, 0.5 mM TCEP, 20 mM Imidazole, 0.1% Tween20), using Co^{2+} -conjugated IMAC beads.

2.7.2 Electrophoretic Mobility Shift Assays (EMSA)

For analysis of DNA-binding by EMSA and for co-crystallization, DNA substrates were generated by annealing complementary oligonucleotides (MWG Eurofins) at a final

concentration of 1 mM in 20 mM TRIS, pH 7.5, 150 mM NaCl. Successful annealing and purity of the oligonucleotides was confirmed by native PAGE on a 6% gel.

For EMSA experiments, varying concentrations of protein were incubated at the indicated molar ratios with 1 μ M of different lengths of DNA oligos or 0.35 μ M nucleosome (provided by Amede Larabi) in 150 mM NaCl, 20 mM TRIS, pH 7.5, 0.5 mM TCEP. Samples were incubated on ice for 30 min. Glycerol was added to a final concentration of 5% and the samples were analysed on a pre-run 6% native 1x TRIS-Glycine (25 mM TRIS, 250 mM glycine, pH 8.3, 5% Glycerol) polyacrylamide gel using 1x TRIS-Glycine running buffer. Gels were stained with SYBR Safe (Thermo Fisher Scientific) to visualize DNA-bound complexes or with Coomassie Blue to stain protein.

2.7.3 Fluorescence Polarization (FP)

32bp 6-FAM labelled DNA was prepared by annealing two complementary DNA strands, essentially as described for the crystallisation of DNA complexes, albeit under low-light conditions (Table 4.1). Fluorescence polarisation assays were conducted in a buffer containing 50 mM TRIS pH 7.5, 150 mM NaCl, 0.1% Tween20 and 0.5 mM TCEP. A series of protein concentrations, ranging from 0.5 to 25 μ M, were incubated in the presence of 50 nM DNA for 30 minutes at room temperature in order to attain equilibrium. Immediately thereafter, fluorescence polarization was recorded using 485 nm and 520 nm excitation and emission filters, respectively (CLARIOstar, BMG Labtech, Germany). The change in fluorescence polarization was then plotted as mean values of three independent replicates and the dissociation constant for each complex determined.

2.7.4 Isothermal Titration Calorimeter (ITC)

ITC was performed using a MicroCal iTC 200 (GE Healthcare) at 25 °C in 20 mM TRIS, pH7.7, 150 mM NaCl, 0.5 mM TCEP (buffer 6). For each titration, 300 μ l of 50 μ M SA2-

Scc1 was added to the calorimeter cell. The synthetic CTCF and Sgo1 peptides, at a concentration of 500 μ M, dialysed into buffer 6, were injected as 18 x 2 μ l syringe fractions. Results were analysed and displayed using the Origin software package supplied with the instrument.

2.8 *In vivo*-Yeast methods

All yeast work was performed by the Haering Laboratory at EMBL Heidelberg.

2.8.1 Tetrad analysis

Wild-type or mutant alleles of *SCC3* fused to a C-terminal PK₆ epitope tag were integrated into the *ura3* locus of a *SCC3/scc3::HIS3* heterozygous diploid yeast strain (C1073). Correct integration was confirmed by PCR (Table 4.4). Following sporulation, strains were tetrad dissected and cultured on YPAD media for 3 days at 30 °C before genotyping by replica plating.

2.8.2 Chromatin immunoprecipitation and ChIP-qPCR

Chromatin immunoprecipitation followed by quantitative PCR. ChIP-qPCR was performed from asynchronous yeast cell cultures as described (Cuylen and Haering 2011), except that sonication was performed with a Bioruptor Plus (Diagenode) at 4 °C using 6 cycles of 30 s on, 60 s off at “high” level. Quantitative PCR was performed with primers listed in Table 4.3.

2.9 Peptide array

Peptide arrays, with an area of 3cm², were obtained from Rudolf Volkmer (immunologie.charite.de). For each step, 5 ml of the required reagents were used. Arrays were washed with 100% ethanol for 5 minutes on a shaker at 21 °C, followed by 3 washes,

for a total of 10 minutes in TBS-T buffer (50 mM TRIS pH 7.5, 150 mM NaCl, 0.05% Tween-20). For the blocking step, arrays were incubated in 1x blocking buffer (Sigma B6429) for 3 h at 21 °C, followed by 3 washes in TBS-T for a total of 10 minutes. SA2-Scc1 and SA2 (F371A)-Scc1H were added to 1x blocking buffer at a final concentration of 1.2 μ M and incubated with the array overnight at 4 °C under gentle agitation. The membrane was washed 3 times (1x 30s, then 2x 5 minutes) at 21 °C. The anti 6xHistidine antibody was diluted 1:2000 in 1x blocking buffer and incubated with the arrays for 1h at 21 °C. The array was washed 3 times (1x 30s, then 2x 5 minutes) and developed by addition of 3,3'-Diaminobenzidine (Sigma D4293) for 1 minute followed by quenching in deionized H₂O. To measure non-specific binding of the anti 6x Histidine antibody, all steps were identical except that no SA2-Scc1 protein solution was added to 1x blocking buffer during the overnight incubation step. Arrays were imaged with a BioRad Gel Doc XR+ Documentation system.

Result_part 1 Assembling the cohesin complex in

***E.Coli* ('Clink')**

Résultats_partie 1

Afin d'étudier le module ATPase de la cohésine et ses facteurs associés, j'ai généré une série de constructions pour expression dans *E. Coli*. Comme seul Smc1 s'est avéré difficile à exprimer, j'ai criblé une série de "têtes" de Smc1 d'espèces différentes. J'ai ensuite obtenu la structure cristalline de de la tête de Smc1 (ctSmc1) de *Chaetomium Thermophilum* en complexe avec le domaine C-terminal de Scc1 (CScc1) de la levure à une résolution de 2,1 Å. Cela m'a permis d'assembler un hétérotrimère de cohésine Smc1-Smc3-Scc1 appelé "Clink" (cohésine "liée") en co-exprimant une fusion de NScc1 au segment C terminal de ctSmc1 avec les composants de module de tête restants (CScc1, le ctSmc1 N-terminal et le complexe NScc1-Smc3) dans *E. Coli*. En étendant Scc1, j'ai également pu intégrer les facteurs accessoires Pds5 et Scc3.

3.1 Expression of cohesin complex in *E.coli*

It has been proposed that cohesin contains both a DNA entry and exit gate and that these are located at the SMC hinge and Smc3-NScc1 interfaces, respectively. It has been argued that such physical separation of entry and exit gates allows for independent regulation of sister chromatid entrapment and their subsequent release (Chan, et al. 2012; Gruber, et al. 2006). Acetylation of the Smc3 head domain by Eco1 is an essential step to establish cohesion (Rolef Ben-Shahar, et al. 2008; Rowland, et al. 2009; Unal, et al. 2008; Zhang, et al. 2008). Hence, it is important to understand how cohesin heterodimerizes at its head domains and how the interactions at this interface are fine-tuned by regulatory factors (Eco1, releasing factors Wapl, Pds5) to enable controlled opening and closure of the exit gate. Assembling the Smc1/Smc3 head domain then is a prerequisite for elucidating these questions.

As previously reported, Smc3hd-NScc1, and Smc3hd-Pds5-NScc1 subcomplexes can be produced recombinantly in *E.coli* (Gligoris, et al. 2014; Muir, et al. 2016). We were able to co-express yeast Smc3-Scc1 in *E. coli*, but not the ySmc1hd-CScc1 complex. We also have previously found that yeast Smc1 cannot be solubly expressed in *E.coli* (unpublished data). We therefore decided to perform expression trials of Smc1hd orthologues from different species in order to screen for soluble constructs suitable for structural characterization.

3.2 Expression trials of Smc1 head domain of different species in *E.coli*

First, I tested expression of Smc1 head constructs from three different species: *A. gossypii* (ag), *S. pombe* (sp) and *C. thermophilum* (ct) (Figure 3.1a). As seen on Figure 3.1.b, no band for spSmc1 is visible, possibly because it cannot be expressed in *E.coli* under these conditions. AgSmc1 appears to be degraded, judged by the distribution of bands on the gel. Only the ctSmc1 head (ctSmc1hd) construct remains intact and soluble (Figure 3.1b). However, I observed that the ctSmc1hd precipitates during purification in absence of an

interaction partner. I therefore increased the salt concentration at the size-exclusion chromatography step from 150 mM to 300 mM and co-expressed it with a C-terminal fragment of budding yeast Scc1 (CSc1) (Figure 3.1c).

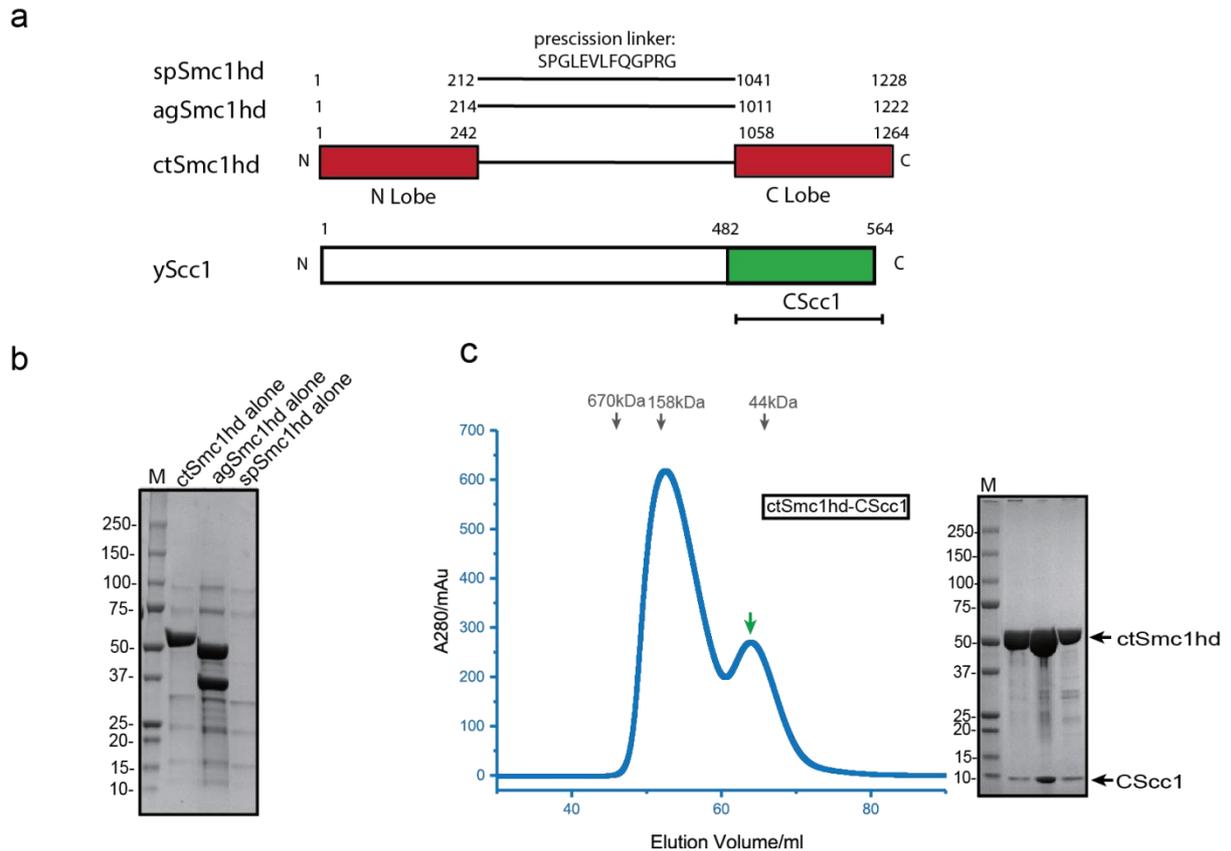


Figure 3.1 Smc1 head boundaries and purification. (a) Schematic representation of Smc1 and Scc1 head complex boundaries in different species (sp: *S. pombe*, ag: *A. gossypii*, ct: *C. thermophilum*, y: *S. cerevisiae*), Smc1 head is fused using a precision protease cleavable linker. (b) Purification of ctSmc1hd / agSmc1hd / spSmc1hd head alone. (c) Purification of ctSmc1 head in complex with C-terminus of yeast Scc1 (CSc1), green arrow indicates the elution volume of the complex.

3.3 Structure of the *Chaetomium thermophilum* Smc1 head in complex with *Saccharomyces cerevisiae* Scc1 (ctSmc1hd-CScc1)

The ctSmc1hd-CScc1 complex was set up for crystallization in the presence of ATP- γ -S (1 mM) and Mg²⁺ (2 mM). Crystals were obtained using the method described in section 2.4.1. Initial hits were further refined using manual setup of gradient concentration of the precipitant PEG300 ranging from 20% - 35% and TRIS buffer, from pH 8.0 - pH 9.0. We determined the 2.1Å X-ray crystal structure of CtSmc1 bound to yeastCScc1 domain by molecular replacement, using the yeast ortholog (Haering, et al. 2004) and found, as expected, high structural conservation (RMSD=0.98 Å²) (Figure 3.2c, Table 3.1). The structure showed more residues of the coiled-coil arm compared to the previously determined ySmc1-Scc1 complex (Figure 3.2c). Interestingly, in contrast to the yeast Smc1-Scc1 structure, we could not find any ATP- γ -S, Mg²⁺ bound to the ctSmc1 head domain.

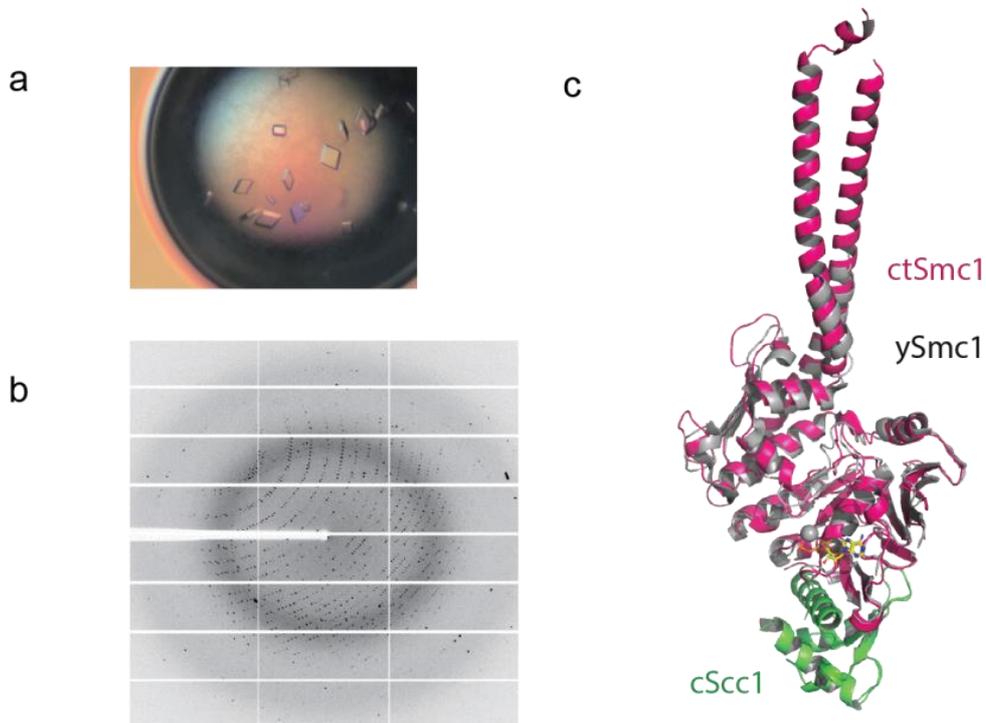


Figure 3.2 Structure of the ctSmc1 head domain in complex with yeast CScc1. (a) Image of the crystal of ctSmc1hd-CScc1 complex. (b) Diffraction pattern of ctSmc1hd-CScc1 crystals, best

resolution at 2.09 Å. (c) Cartoon representation of the ctSmc1hd-CSc1 structure and structural overlay with ySmc1. The ctSmc1hd (red) shows an extended coiled-coil as compared to the published yeast Smc1 (grey). The two proteins superpose with an RMSD of 0.98Å² showing, as expected, high structural conservation.

| ctSmc1-CSc1 | |
|--|---|
| Data collection | ESRF MASSIF ID30A-1 |
| Space group | P2 ₁ 2 ₁ 2 ₁ |
| Cell dimensions | |
| <i>a, b, c</i> (Å) | 80.7, 111.1, 166.1 |
| Resolution (Å) | 45.70–2.09 |
| No. reflections | 82453 (12513) |
| <i>R</i> _{sym} or <i>R</i> _{merge} | 6.2 (121)* |
| <i>I</i> / σ <i>I</i> | 13.4 (1.1)* |
| <i>CC 1/2</i> | 0.99 (0.49) |
| Completeness (%) | 98.8 (94.2)* |
| Redundancy | 4.5 (4.4)* |
| Refinement | |
| Resolution (Å) | 45.70–2.09 |
| <i>R</i> _{work} / <i>R</i> _{free} | 0.21 / 0.24 |
| No. atoms | 15801 |
| Smc1 [#] | 6431 |
| Sccl1 [#] | 1302 |
| <i>B</i> -factors (mean; Å ²) | |
| Smc1 [#] | 63.7 |
| Sccl1 [#] | 69.9 |
| R.m.s deviations | |
| Bond lengths (Å) | 0.002 |
| Bond angles (°) | 0.46 |

*Values in parentheses are for highest-resolution shell.

#Averages are shown for the two copies in the asymmetric unit

Table 3.1 X ray data collection, phasing and refinement statistics

3.4 ctSmc1/Smc3 head assembly via NScc1 linker -“Clink”

Previously, we have attempted to generate engineered head modules for structural studies by fusing the Sccl1 C-terminus to Smc3, which would serve as bait to capture Smc1. Ultimately this was unsuccessful (Muir 2016). To obtain a heterotrimeric Smc3-Smc1-Sccl1 ATPase

head module, I fused *Saccharomyces Cerevisiae* (hereafter ‘yeast’) NScc1 to the C-lobe of the ctSmc1hd (as described below) and co-expressed this construct in *E.coli* with the remaining head module constituents: the yeast Smc3 ATPase, the N-lobe ctSmc1 and its binding partner, yeast CSc1 (Figure 3.3a). Based on this idea, I generated three different constructs, which contained different lengths of the NScc1 linker: Scc1 (1-115), (1-120), and (1-125) followed by the prescission cleavage sites (SPGLEVLFGPRG) (Figure 3.3a, b). Co-purification of the Smc3-ctSmc1-Scc1 head module, termed c-link (‘linked’ cohesin via a His-tag on the ctSmc1 C-terminus,), showed that a direct physical interaction of NScc1 with the coiled coil of Smc3 is maintained (Figure 3.3c-e, Table S2). Clink complexes containing longer NScc1 (Scc1 1-125) tended to have higher yield in protein expression when compared to Clink115 and Clink120 (Figure 3.3c-e) and higher DNA binding affinity as shown by Electrophoretic mobility shift assay (EMSA), (Figure 3.3f-h).

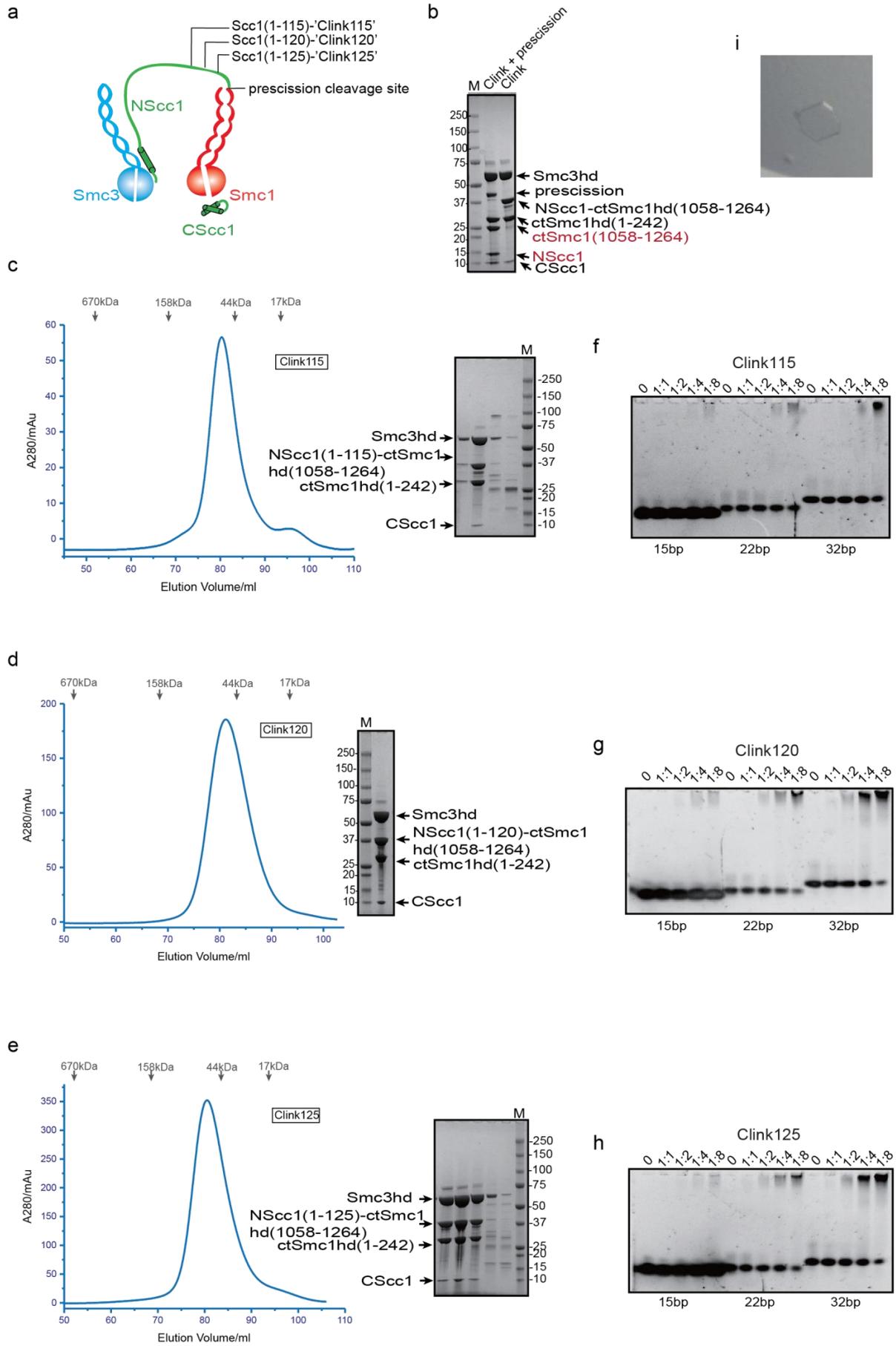


Figure 3.3 Purification and DNA binding of Clink115/120/125. (a) Cartoon model of the Clink complex assembling Smc1/Smc3 head modules via the NScc1 linker with three different lengths (NScc1 (1-115)-Clink115, NScc1(1-120)-Clink120, NScc1(1-125)-Clink125). The precision cleavage site adjacent to the Smc1 head is kept intact. (b) Clink is precision cleavable. Incubation with precision protease results in cleavage of the NScc1-ctSmc1 fusion protein (red bands). (c-e) Purification of the Clink115/120/125 complex, longer linkers gave a higher yield. (f-h). EMSA assay of Clink115/120/125 in complex with three different lengths of DNA oligonucleotides (15/22/32bp) (Table 4.1). (i) Crystal of Clink125 obtained in a crystallization screen including *in situ* chymotryptic digestion (1:10000 w/w) after 33 days.

3.5 Complex assembly trials of Clink159 in complex with Pds5 and Scc3

To incorporate additional cohesin regulators into the complex, I designed a longer NScc1 fusion which extended to the Pds5 binding region (1-159) (Muir, et al. 2016) for co-expression or co-purification with either truncated or full length Pds5 (T5) (Figure 3.4a-d).

To include Scc3 in the Clink complex, I made a longer CScc1 covering the Scc3 binding region (309-564). Unfortunately the longer CScc1 was degraded, which did not allow us to co-express Clink159 with Scc3. Clink159 variants with two truncations lacking residues 400-475 and 400-480 on the extended CScc1 were purified (data not shown) and only the one lacking 400-480 (called CScc1 Δ or Clink159-Scc3T Δ) showed less or no degradation (Figure 3.4 c, d).

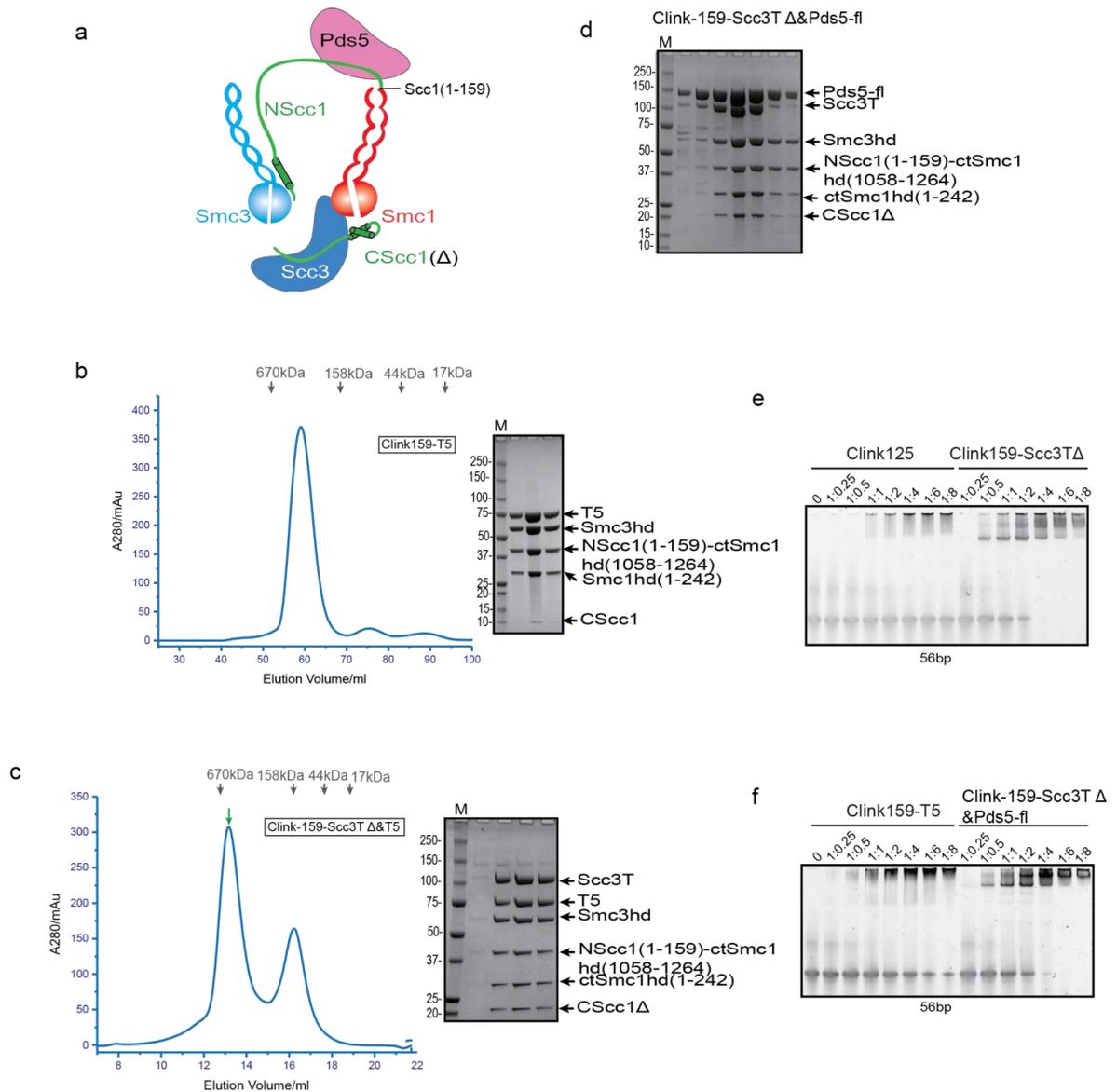


Figure 3.4 Purification and EMSA of Clink159 in complex with either Pds5, Scc3, or both. (a) Cartoon model of the Clink159 in complex with Pds5 and Scc3, CSc1 (Δ) spanning 309-564 covering both Smc1 and Scc3 binding regions but with an internal deletion spanning residues 400-480. (b) Purification of Clink159 in complex with Pds5 (T5). (c) Purification of Clink159 in complex with Scc3T and Pds5-T5, green arrow indicates the complex elution volume. (d) Purification of Clink159 in complex with Scc3T and Pds5-fl. (e) & (f). EMSA of the Clink125, Clink159-Scc3T Δ , Clink159-Pds5-T5 and Clink159-Scc3T Δ -Pds5-fl complex with a 56 bp DNA oligonucleotide (Table 4.1).

3.6 Crystallization trials of the Clink complex.

Crystallization at both 4°C and 20 °C, crystallisation with different lengths of DNA oligonucleotides, with/without ATP- γ -S, *in situ* digestion etc was attempted, but no crystals were obtained, except a single non-reproducible crystal which grew following *in situ* chymotryptic digestion (1:10000 w/w) (Figure 3.3i).

**Result_part2 Structural basis for Scc3-dependent
cohesin recruitment to chromatin**

Résultats_partie 2

Pour comprendre comment la cohésine engage l'ADN, j'ai étudié les propriétés de liaison à l'ADN de sous-complexes globulaires précédemment identifiés. En déterminant une structure cristalline de la protéine Scc3 de levure liée à un fragment de la sous-unité Scc1 kleisin et de l'ADN, j'ai pu démontrer que Scc3 et Scc1 forment un module d'interaction composite de l'ADN. Le sous-complexe Scc3-Scc1 engage un ADN double brin à travers une surface conservée, chargée positivement. Nous démontrons que ce domaine conservé est requis pour la liaison à l'ADN par Scc3-Scc1 in vitro, ainsi que pour l'enrichissement de la cohésine sur les chromosomes et pour la viabilité cellulaire. Ces résultats suggèrent que l'interface de liaison à l'ADN de Scc3-Scc1 joue un rôle central dans le recrutement de la cohésine sur les chromosomes et donc que cette dernière exécute fidèlement ses fonctions lors de la division cellulaire.

4.1 Investigating DNA binding by the cohesin complex

To investigate the DNA-binding properties of cohesin, I co-expressed and purified defined globular domains and subcomplexes of the *Saccharomyces cerevisiae* cohesin (Figure 4.1 & 4.2 & 4.3e). These encompassed Smc3hd-NSccl, ctSmc1hd-CSccl (Figure 4.1a, Figure 4.2a), as well as the Sccl3-Sccl1 subcomplex (Sccl3T-Sccl1K) (Figure 4.3a,b). In addition, we produced a Smc1-Smc3 hinge heterodimer (Figure 4.1b), Pds5 bound to an Sccl1 fragment (Muir, et al. 2016) with full-length Pds5 (Pds5-fl) or as a truncated variant (Pds5-T5); as well as full-length Wapl (Wapl-fl) or truncated variant (Wapl-C) (Figure 4.2b, c), and a GST tagged Eco1 full-length (Figure 4.1d). We found that the Smc3hd-NSccl module and the Smc1-Smc3 hinge heterodimer, consistent with a prior study (Murayama and Uhlmann 2014), bound DNA, as seen by the appearance of slower-migrating species in EMSA assays (Figure 4.1a, b). As expected for non-sequence specific DNA binding factors, longer DNA fragments (>21 bp) bound more efficiently than shorter DNA duplexes (15 bp). Isolated Smc3hd also interacted with DNA but with relatively weak affinity under these conditions (Figure 4.1c). Eco1, containing a C2H2 zinc finger domain, also bound to DNA directly (Figure 4.1d). Conversely, the Pds5 subunit, the ctSmc1hd-CSccl subcomplex or the Wapl subunit did not interact with DNA in this assay (Figure 4.2). The Sccl3T-Sccl1K subcomplex showed most robust DNA binding activity (Figure 4.3e). I also observed that DNA binding depended on the presence of Sccl1K as in its absence Sccl3T showed greatly reduced DNA binding affinity (Figure 4.3e). As cohesin interacts with chromosomal templates I also assessed if the Sccl3T-Sccl1K subcomplex interacts directly with nucleosomes. I observed robust nucleosome binding indicating that the Sccl3T-Sccl1K subcomplex not only accommodates regular B-DNA substrates but also apparently the altered DNA geometry of the nucleosome core particle (Figure 4.3f).

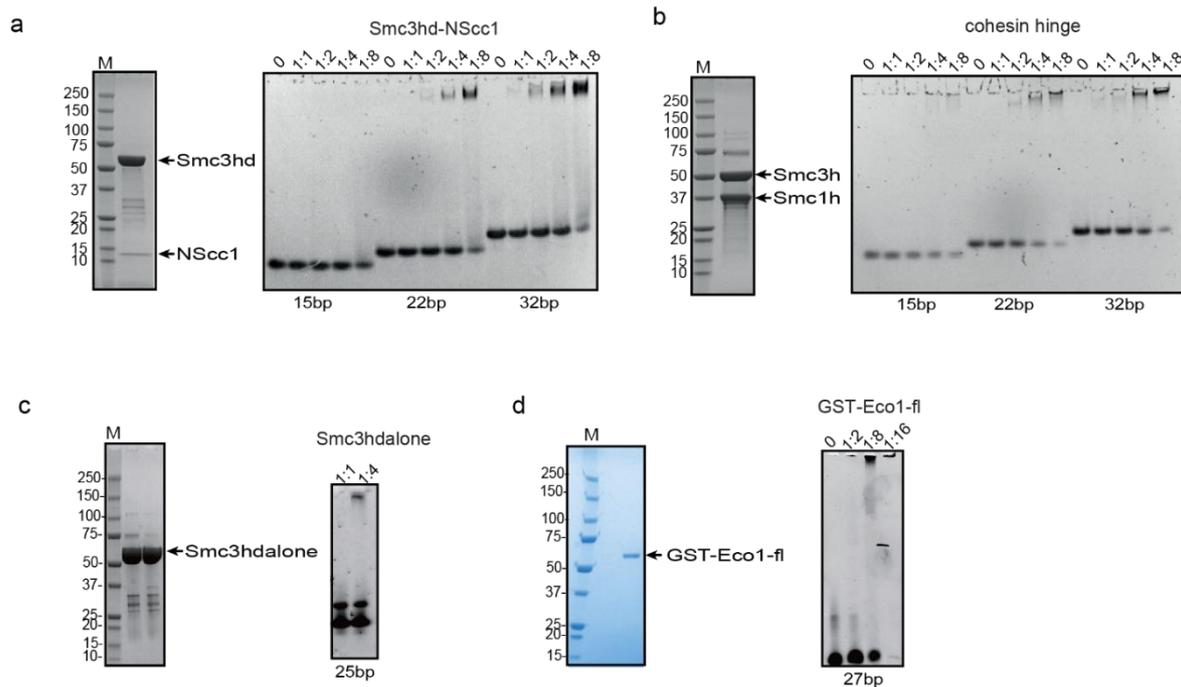


Figure 4.1 SDS-PAGE analysis of the purified cohesin components and DNA binding analysis by EMSA. (a) SDS-PAGE and DNA binding analysis of the Smc3hd-NScc1 complex. **(b)** SDS-PAGE and DNA binding analysis of the Smc3/Smc1 hinge heterodimer. **(c)** SDS-PAGE of the purified Smc3 head alone and DNA binding analysis. **(d)** SDS-PAGE and DNA binding analysis of GST-Eco1-fl. Concentration of the DNA duplex in the EMSA analysis was 1 μ M. The DNA was incubated at different molar ratios, as indicated on top of each lane, with the indicated protein preparation. Native EMSA PAGE gels were stained with Sybr Safe Nucleic Acid Gel Stain. SDS-PAGE gels were stained with Coomassie.

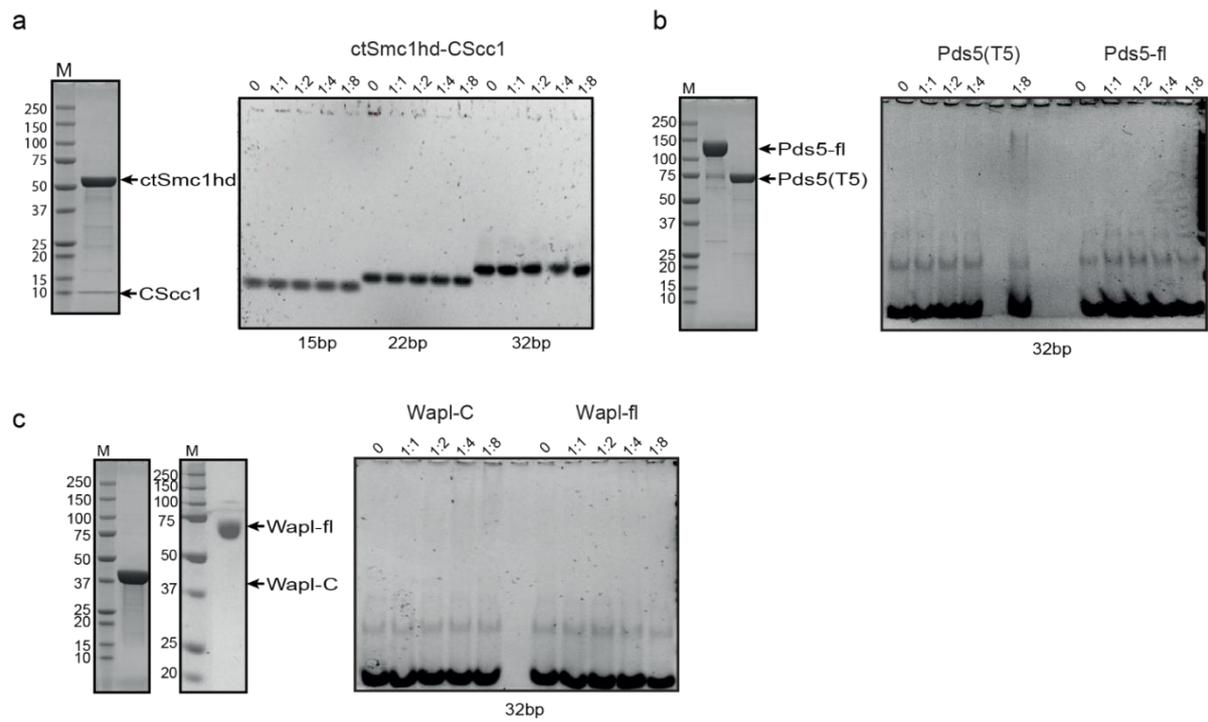


Figure 4.2 Cohesin subunits that lack DNA binding affinity. (a-c), ctSmc1hd-CScc1, Pds5 and yeast Wapl did bind double strand DNA.

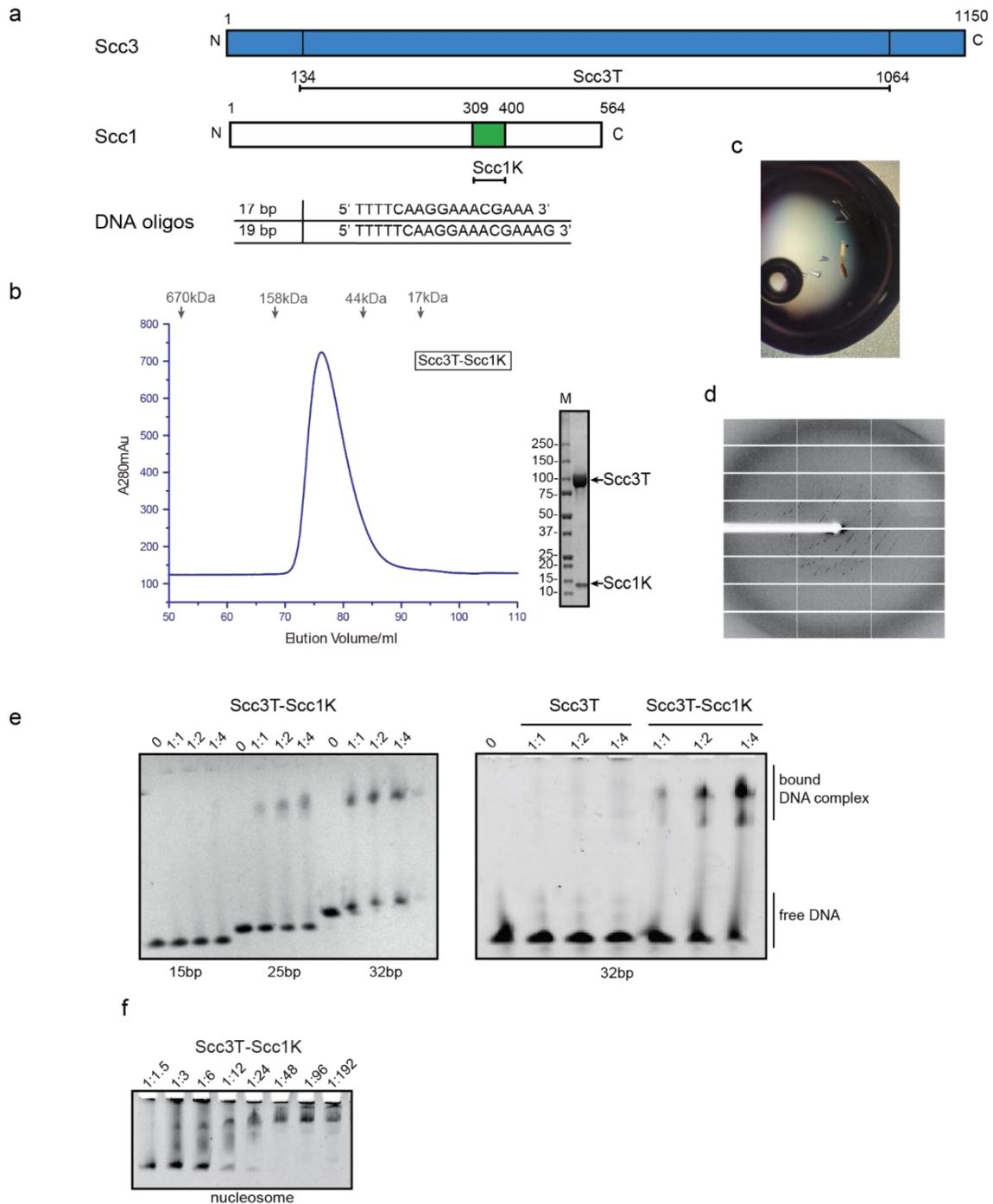


Figure 4.3 DNA binding by the Scc3-Scc1 subcomplex. (a) Domain structure of Scc3 and Scc1 and DNA oligonucleotides used for crystallization. Construct boundaries used and their acronyms are shown below. (b) Purification of Scc3T-Scc1K complex. (c) Crystals of the Scc3T-Scc1K in complex with the 19 bp DNA. (d) Diffraction pattern of the crystal. (e) DNA binding analysis by EMSA showing that Scc3t-Scc1k binds to longer DNA more efficiently as compared to

shorter DNA. The DNA binding capability of Scc3T is enhanced by Scc1K. (f) Scc3T-Scc1K also binds to nucleosomes.

| Length of DNA duplex | Sequence* |
|----------------------|--|
| 15 bp | 5' TAAACGAAAGTGAAC 3' |
| 17 bp | 5' TTTTCAAGGAAACGAAA 3' |
| 19 bp | 5' TTTTCAAGGAAACGAAAG 3' |
| 21bp | 5' TCTTTTCAAGGAAACGAAAGT 3' |
| 22bp | 5' TTCAAGGAAACGAAAGTGAAC 3' |
| 32 bp | 5' **TGGAAGCCTTTTCAAGGAAACGAAAGTGAAC 3' |
| 56 bp | 5'CCCCTTGCGGTTAAAACGCGGGGGACAGCGCGTACGTG CGTTTAAGCGGTGCTAG3' |
| 62 bp | 5'ATTGCAGTGCCCACAGAGGCCAGCAGGGGGCGTAGT GAGGCCTTTTCAAGGA AACGAAAGTG 3' |

* dsDNA containing a single 5' T (top strand) and 5' A (bottom strand) overhang were used. Only top DNA strand is indicated.

** For fluorescence polarization assays the 5' end of the top strand of the 32bp was labeled with 6-FAM.

Table 4.1 DNA oligos used for crystallization, EMSA and Fluorescence Polarization (FP). Oligos labeled with red color are used for Smc3hd-NSc1, Hinge, Scc3-Scc1 and all Clink crystallization trails, oligo 62 bp in bold indicates CTCF zinc finger binding motif (Figure 5.5).

4.2 Crystal structure of the Scc3-Scc1 in complex with a 19 bp DNA

To determine the molecular basis of how cohesin subunits interact with DNA, I screened all cohesin subcomplexes that interacted with DNA using commercial crystallisation screening conditions (EMBL Grenoble HTX facility), i.e. SMC3hd-NScc1, Smc3/Smc3 hinge, Scc3T-Scc1K in complex with different lengths of DNA oligonucleotides as indicated in table S2. None of the complexes crystallized. I then tried conditions which are typically suitable for crystallization of DNA complexes, based on the collation of conditions reported on the RSCB-PDB (Materials and methods, Table 2.1). I obtained crystals of the Scc3T-Scc1K in complex with a 17 bp DNA substrate. Despite testing hundreds of these crystals for diffraction at the automated Massif beamline (ESRF) the resolution was found to be severely limited. To improve the resolution, I tested different Scc1 truncations, containing Scc3T with either shorter (residues 309-388), or longer Scc1 (residues 273-400, 240-400 or 170-400). I also tested shorter DNA oligonucleotides ranging from 12 to 16 bp (data not shown). While I obtained crystals (data not shown), they did not show improved diffraction. I then used the crystals containing the 17 bp DNA as seeds and rescreened the crystallisation conditions but with different lengths DNA oligonucleotides. I was thus able to reproducibly obtain crystals of Scc3T-Scc1K in complex with a 19 bp DNA fragment.

The optimized crystals diffracted anisotropically to a minimum Bragg spacing of 3.9 Å in the best, and ~5.7 Å in the worst direction (Table 4.2). We determined the structure by molecular replacement, using the structures of a Scc3 ortholog from *Zygosaccharomyces rouxii* and a C-terminal fragment of *Saccharomyces cerevisiae* Scc3 as search models (Roig, et al. 2014). The resulting electron-density map provided a continuous trace of the polypeptide main chain, but with a limited level of detail owing to the anisotropy of the data. Despite these drawbacks, we successfully traced the amino-acid register using a selenomethionine

derivative and refined a model encompassing amino acid residues 134 to 1064 of Scc3 in complex with residues 309 to 400 of Scc1, bound to a 19 bp DNA molecule (Table 4.2).

| | Scc3T-Scc1K native | Scc3T-Scc1K SeMet |
|---|----------------------------------|----------------------------------|
| Data collection | | |
| Space group | P2 ₁ 2 ₁ 2 | P2 ₁ 2 ₁ 2 |
| Cell dimensions <i>a, b, c</i> (Å) | 109.9, 115.4, 295.6 | 109.9, 115.6, 296.2 |
| Wavelength (Å) | 1.282 | 1.282 |
| Resolution (Å) | 50–3.60 | 49.9–4.79 |
| No. reflections | 20963 | 10279 |
| <i>R</i> _{merge} | 5.8 (122.6)* | 4.6 (112.3)* |
| <i>I</i> / σI | 11.9 (1.6)* | 10.6 2(.1) |
| <i>CC 1/2</i> | 0.99 (0.56) | 0.99 (0.52) |
| Completeness (%) | 91.4 (63.5)* | 93.6 (71.2)* |
| Redundancy | 4.4 (6.0)* | 1.8 (1.8) |
| Refinement | | |
| Resolution (Å) | 50-3.60 | |
| <i>R</i> _{work} / <i>R</i> _{free} | 0.28 / 0.31 | |
| No. atoms | 16465 | |
| Protein | 14909 | |
| DNA | 1556 | |
| <i>B</i> -factors (mean) | | |
| Protein | 254.5 | |
| DNA | 266.4 | |
| R.m.s deviations | | |
| Bond lengths (Å) | 0.002 | |
| Bond angles (°) | 0.5 | |

*Values in parentheses are for highest-resolution shell.

Table 4.2 Data collection and refinement statistics

As seen in previous structures of Scc3, the protein is hook-shaped in the C-terminal section and contains an N-terminal 'nose' formed by a pair of extended antiparallel α -helices (Figure 4.4a). Similarly to the interaction of human Scc1 with Scc3, the yeast Scc1 in our structure binds along the convex surface of the hook-shaped HEAT-repeat subunit. We detected additional electron density corresponding to dsDNA within the cradle of this hook (Figure 4.5a, b). Whereas the DNA duplexes were aligned to form pseudocontinuous double helices throughout the crystal, the DNA duplex was slightly too short for tight end-to-end stacking (Figure 4.5c). As a result, the DNA density was only partially resolved, apparently due to rotational and translational disorder of the DNA in the binding cavity.

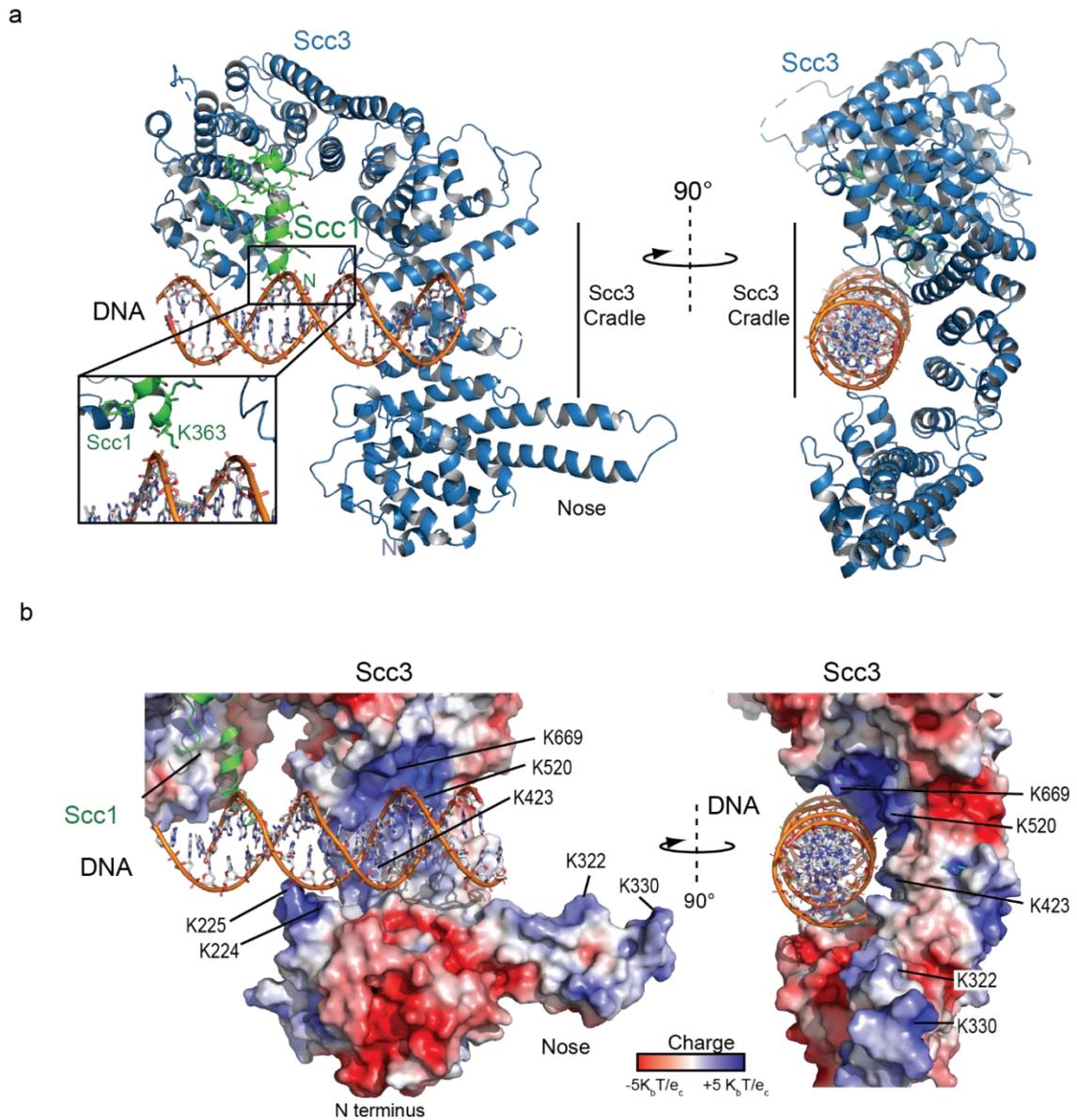


Figure 4.4 Structure of the Scc3-Scc1 subcomplex bound to DNA. (a) Cartoon representation of the Scc3-Scc1 complex bound to a 19-bp dsDNA substrate. The N- and C- termini of Scc3 (violet) and Scc1 subunits (green) are shown. The inset shows a close-up view of the Scc1 amino acid K363. (b) Electrostatic surface potential representation of the Scc3-Scc1 subcomplex with bound dsDNA (calculated with APBS and displayed with Pymol).

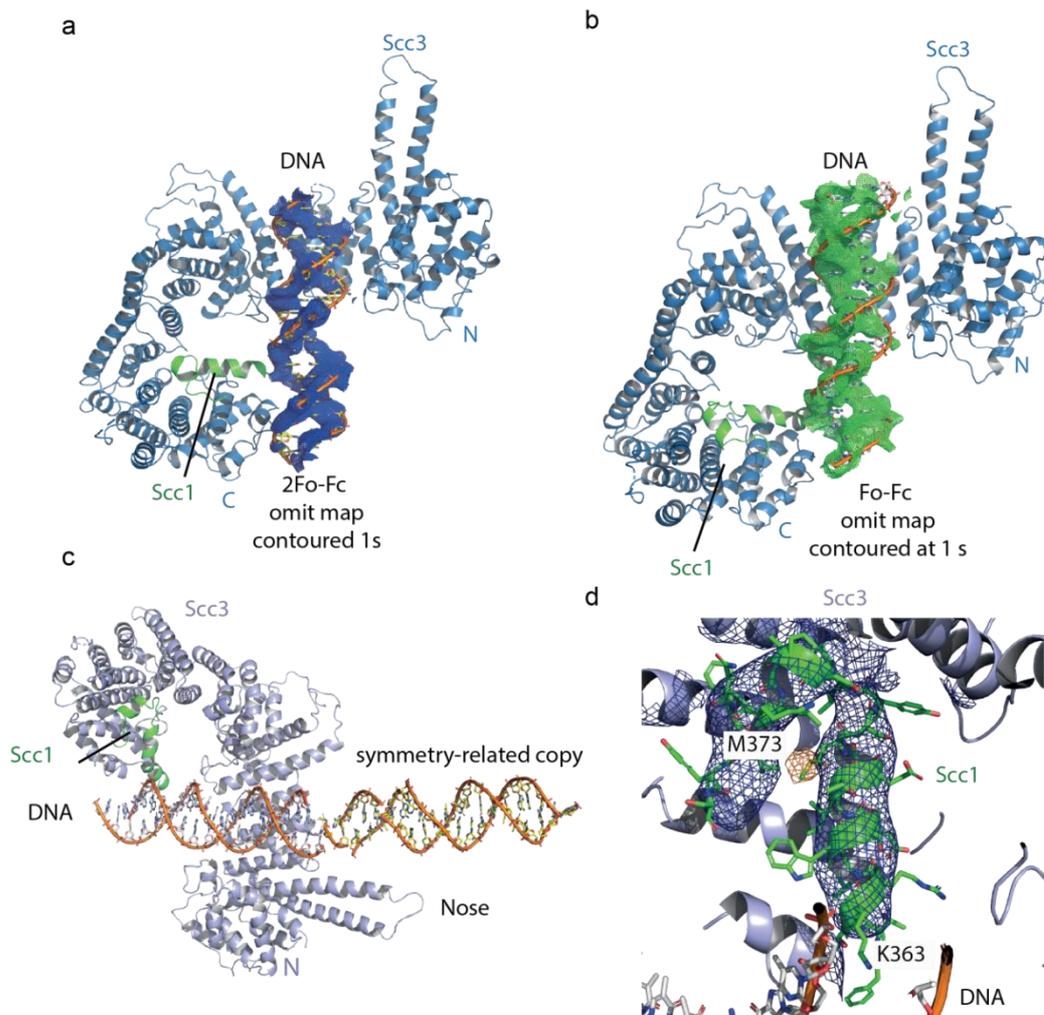


Figure 4.5 Electron density for the DNA molecule bound to the Scc3-Scc1 subcomplex.

(a) View of the DNA binding site of the Scc3-Scc1 subcomplex. DNA is shown as stick, along with its 2Fo-Fc electron density Fourier map (blue mesh) contoured at 1.0 s. (b) Fo-Fc omit electron density Fourier map (grey mesh) contoured at 1.0 s. (c) N-terminal ‘nose’ of Scc3 which interacts through a set of positively charged amino acid residues with the DNA of a neighbouring complex related by crystallographic symmetry. (d) Anomalous difference Fourier map (yellow mesh) contoured at 2.8 s of a selenomethionine–substituted crystal, confirming the location of Scc1 residue M373. Blue mesh: 2Fo-Fc electron density Fourier map around Scc1 contoured at 1.5 s. K363 of Scc1 is located in close proximity to the DNA phosphate backbone.

To identify amino acid residues potentially involved in DNA binding, we mapped the electrostatic surface potential onto the Scc3-Scc1 structure (Figure 4.4b). This revealed that DNA is nested within an extended cradle spanning the majority of Scc3-Scc1, lined by a set of positively charged residues that directly contact the DNA phosphate backbone. The DNA is aligned almost parallel to the N-terminal ‘nose’ of Scc3, which interacts through a set of positively charged amino acid residues with the DNA of a neighbouring complex related by crystallographic symmetry (Figure 4.5c). We observed no direct nucleotide base–amino acid interactions, which explains the apparent lack of DNA sequence specificity.

To ascertain the amino-acid register of Scc1, we used an anomalous difference map peak for M373 in the selenomethionine-derivative data (Figure 4.5d). The deduced register places Scc1 residue K363 in close proximity to the DNA (Figure 4.4a, Figure 4.5d). To assess if this region of Scc1 contributes to DNA binding, I truncated Scc1 amino acids 355-400, thus deleting the region that was not visible in our structure. This variant failed to co-purify with Scc3 in our expression system indicating that Scc1 (309-355) is essential for Scc3-Scc1 interaction (Figure 4.10e). Direct interactions between Scc1 and the DNA phosphate backbone potentially explain why the Scc3T-Scc1K subcomplex has greater DNA binding affinity than does isolated Scc3T (Figure 4.3e).

Mapping of sequence conservation onto the structure revealed that amino acid residues in the DNA binding groove are generally well conserved among yeast Scc3 orthologs (Figure 4.6c). In particular, amino acid residues that are located proximal to the DNA phosphate backbone showed strong conservation. To further evaluate the contributions made by individual segments of the DNA-binding surface, we subdivided participating residues into a series of three patches, based on their physical proximity to the DNA (Figure 4.8b), and subjected these patches to site-directed mutagenesis.

We measured the DNA equilibrium dissociation constant by fluorescence polarization (FP), using a 32-bp 6-FAM labelled DNA substrate (Table 4.1), which is sufficient to bridge the entire DNA binding surface present in the Scc3-Scc1 crystal structure (Figure 4.8c). Whereas the wild-type Scc3T-Scc1K complex bound this substrate with an equilibrium dissociation constant of 2.2 μM , charge-inversion mutations in patches 1 and 3, located at the crest of and within the Scc3 cradle, had only modest effects on affinity (equilibrium dissociation constants of 9.5 μM and 7.3 μM , respectively). Patch 2 mutants, residing in the Scc3 'nose', exhibited essentially unaltered DNA binding affinity. In contrast, the simultaneous mutation of all three patches (a heptamutant) reduced the binding affinity of the patch 1 or 3 mutants even further (to 29.3 μM , Figure 4.8c). The defect in binding DNA was not due to any impact of the mutations on the structural integrity of Scc3-Scc1, as all mutant complexes eluted indistinguishably from wild-type Scc3-Scc1 during size-exclusion chromatography (Figure 4.6a) and efficiently formed a complex with Scc1 *in vitro* (Figure 4.6b) or *in vivo* (see below). Similar results were observed via EMSA as well (Figure 4.7). We conclude that the positively charged Scc3 cradle comprises the major DNA-binding surface of Scc3 and that the positively charged amino acid residues located in patch 1 and 3 constitute a composite DNA binding surface. The distribution of these residues across an extended surface of Scc3 might explain why their significance has thus far eluded cell-biological and genetic characterization.

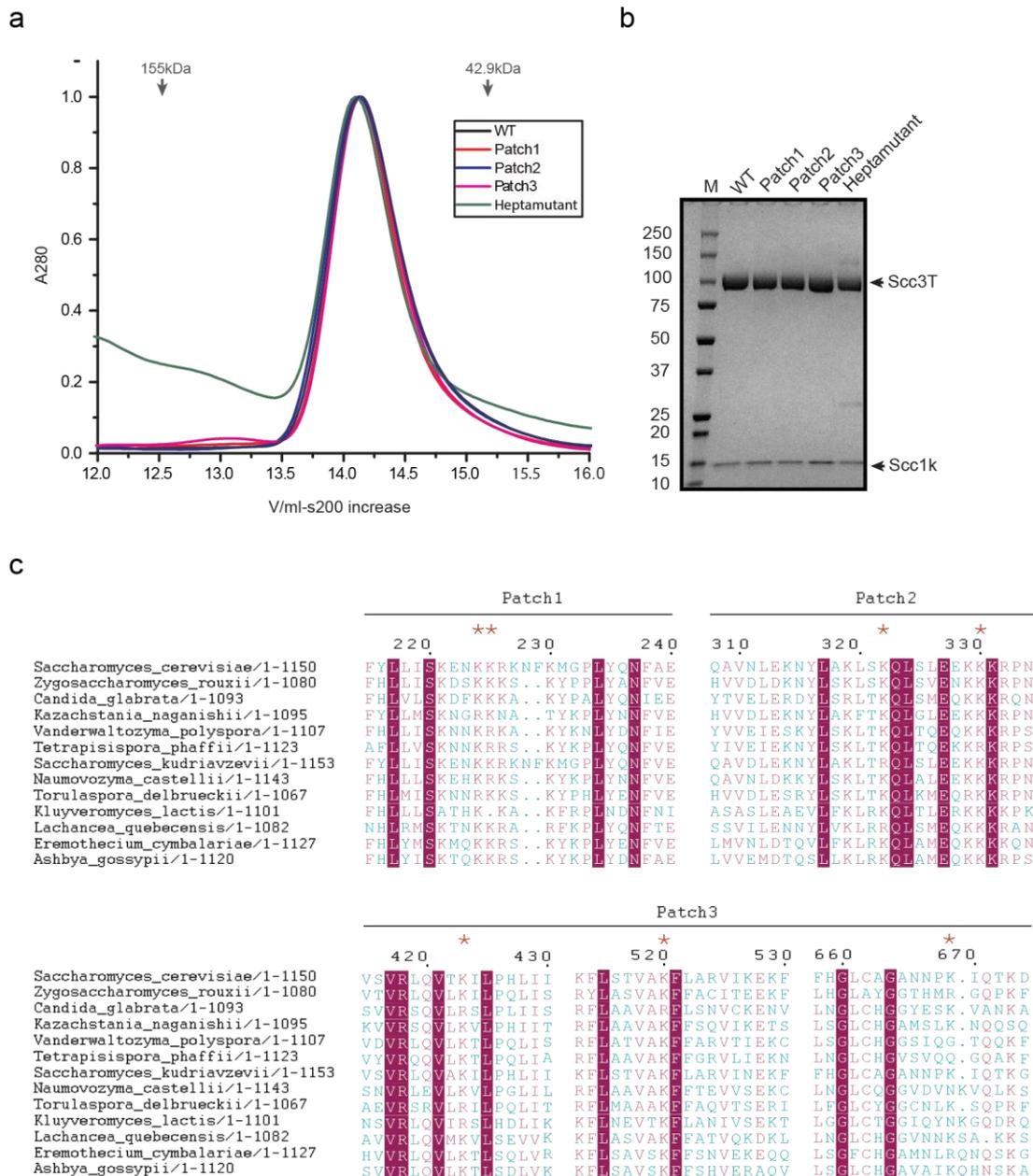


Figure 4.6 Biochemical analysis of Scc3-Scc1 subcomplexes and Scc3 conservation alignment. (a) Size-exclusion chromatography of Scc3-Scc1 complexes. Arrows indicate the elution volume of molecular weight size markers. (b) SDS-PAGE analysis of Scc3-Scc1 complexes from peak fractions in panel (A), showing that the mutants do not impact Scc1 binding in vitro. (c) Alignment of Scc3 amino acid sequences. Residues that were probed by mutagenesis are labeled with red stars (*). Magenta boxes indicate fully conserved amino acid residues while chemically conserved residues are colored magenta. Blue residues indicate divergent residues.

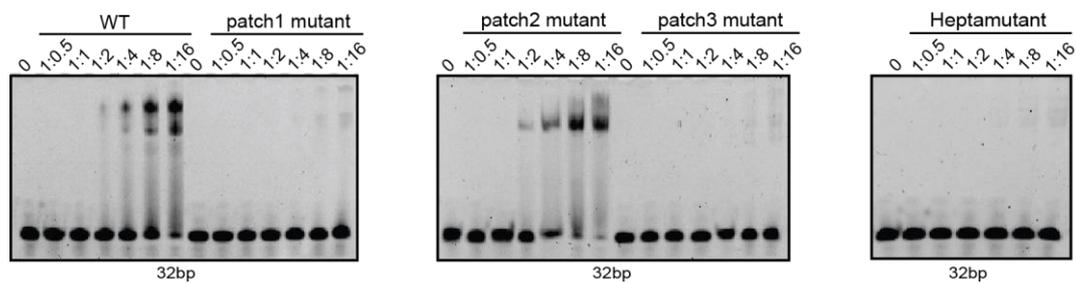


Figure 4.7 DNA binding affinity test of both Scc3T-Scc1K wild type and mutants via EMSA. The binding affinity difference is consistent with the fluorescence polarization experiment (Figure 4.8c).

4.3 *In vivo* characterization of the Scc3-Scc1-DNA structure

We then assayed whether DNA binding by Scc3 is important for cohesin function *in vivo*. We collaborated with Christian H Haering from EMBL Heidelberg. Wild-type or mutant versions of Scc3 were integrated into a diploid yeast strain in which one of the two endogenous *SCC3* genes had been deleted (Table 4.4). Using tetrad dissection, these newly generated yeast strains were then tested for their ability to compensate the loss of *SCC3* with the integrated mutant or wild type variant (Figure 4.8e). Whereas all three individual patch mutants could support cell proliferation, cells expressing the heptamutant version as their only source of Scc3 failed to divide, despite expressing the mutant Scc3 protein at wild-type levels (Figure 4.9a).

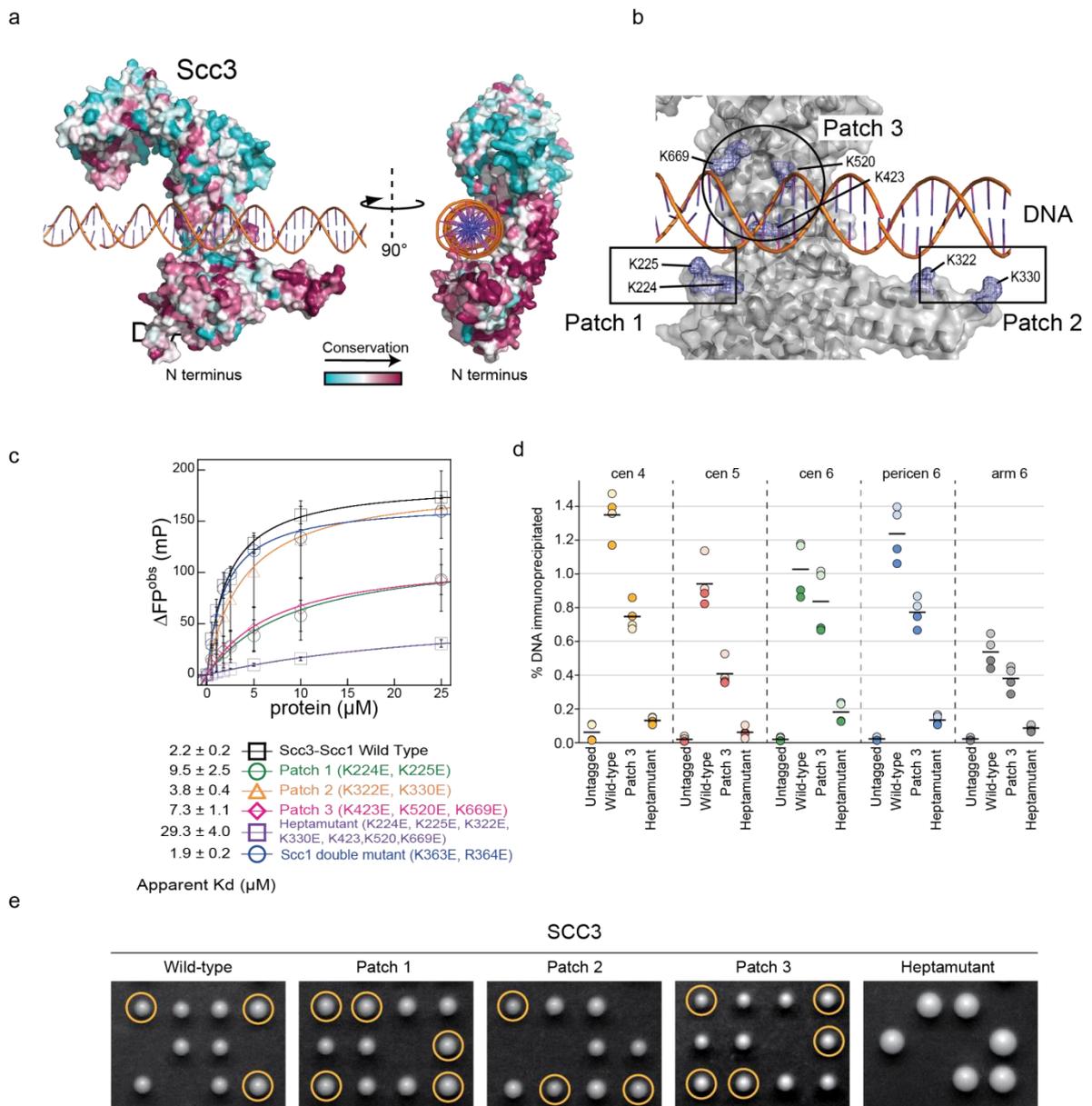


Figure 4.8 A conserved DNA binding domain in the Scc3-Scc1 subcomplex is required for cohesin association with chromatin. (a) Surface amino acid conservation of yeast Scc3. Residues in the DNA binding domain are well conserved. (b) DNA binding residues are located in three surface patches of Scc3. (c) DNA binding fluorescence polarization of 6-FAM labelled 32bp dsDNA by variants of the Scc3-Scc1 subcomplex. Data points corresponding to the average of three independent experiments were fitted to a standard binding equation assuming a single binding site using Kaleidagraph. Standard deviations are depicted as vertical error bars. Apparent K_d are noted below. (d) ChIP-qPCR analysis of binding to centromeric (cen), pericentromeric (pericen) or arm 6.

chromosome arm (arm) regions (chromosomes IV, V, and VI as indicated) of untagged (strain C3) or PK₆-tagged wild-type or mutant versions of Scc3 expressed from an ectopic locus under its endogenous promoter (strains C5013, C5043, C5033). The fractions of immunoprecipitated DNA relative to input DNA are plotted as circles for two biological repeats with two technical repeats each (same colour pairs). Mean values of all four data points are shown as lines. (e) Tetrad analysis of diploid budding yeast strains expressing ectopic wild-type or mutant versions of Scc3 under control of the endogenous promoter in an *SCC3/scc3Δ* background (strains C5013, C5014, C5015, C5043, C5033). Images were recorded after three days at 30 °C on rich media. Genetic marker analysis identified *Scc3(mutant)*, *scc3Δ* cells (circles).

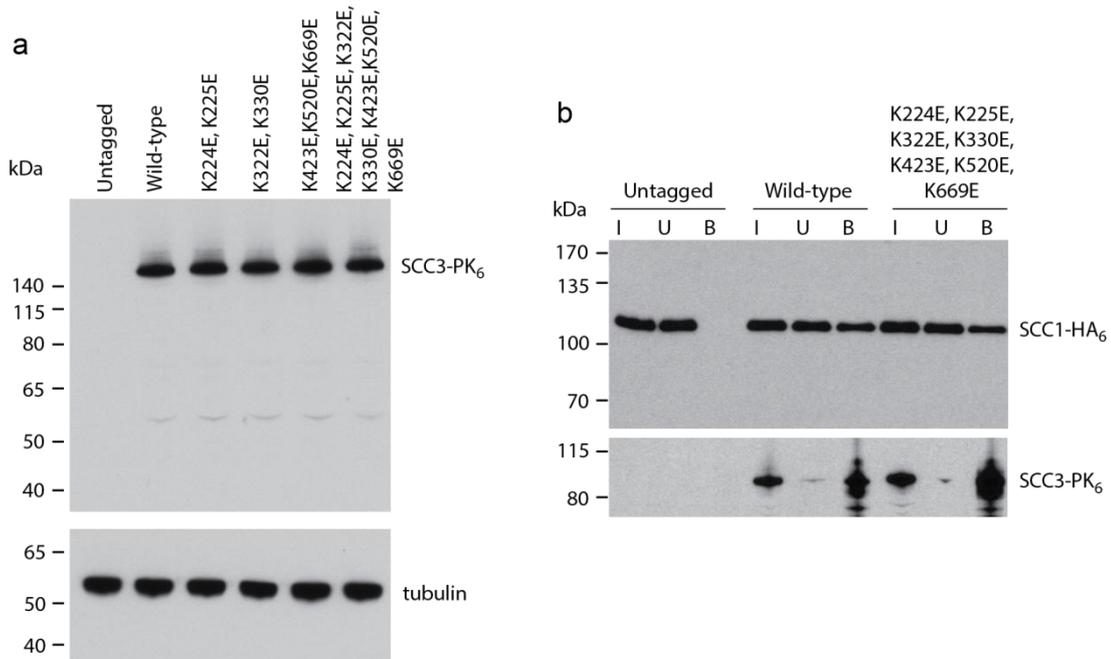


Figure 4.9 In vivo analysis of SCC3 mutants. (a) Protein levels of ectopically expressed wild-type and mutant versions of Scc3 tagged with PK6 (strains C1073, C5013, C5014, C5015, C5043, C5033) measured by western blotting of whole cell lysates against the PK tag (top) and a tubulin as loading control (bottom). (b) Co-immunoprecipitation of Scc1-HA6 with wild-type or mutant versions of PK6-tagged Scc3 expressed from an ectopic locus under the endogenous promoter (strains C501, C5165, C5166).

| | |
|-----------|----------------------------------|
| cen 4 | TGGTGTGGAAGTCCTAATATCG |
| | TGCATGATCAAAAGGCTCAA |
| cen 5 | AACCTTAACAAATGAAGTAAATTCAA AA |
| | TCAATGTGTTAGTTAAAGCAAAAAG AA |
| cen 6 | GGGCGATGGAAGAGGTAAAGT |
| | AGCATTAAACAACCTTCGACAGGT |
| pericen 6 | AAGAAGAATTTAGCGTGGTCAGA |
| | TCCTTTCTCTCGAGTTTCCGT |
| arm 6 | AGCAACGGATACCAGTCAACT |
| | TGCTGGTTAACTCGGACTTCA |

Table 4.3 ChIP-qPCR primer sequences (5'→3')

| | |
|-------|--|
| C3 | <i>MATa/a</i> |
| C501 | <i>MATa, SCC1-HA₆::HIS3</i> |
| C1073 | <i>MATa/a, smc3::HIS3/SMC3</i> |
| C5013 | <i>MATa/a, scc3::natMX4/SCC3, ura3::SCC3-PK₆::URA3/ura3</i> |
| C5014 | <i>MATa/a, scc3::natMX4/SCC3, ura3::SCC3(K224E, K225E)-PK₆::URA3/ura3</i> |
| C5015 | <i>MATa/a, scc3::natMX4/SCC3, ura3::SCC3(K322E, K330E)-PK₆::URA3/ura3</i> |
| C5016 | <i>MATa, scc3::natMX4, ura3::SCC3-PK₆::URA3</i> |
| C5018 | <i>MATa, scc3::natMX4, ura3::SCC3(K224E, K225E)-PK₆::URA3</i> |
| C5020 | <i>MATa, scc3::natMX4, ura3::SCC3(K322E, K330E)-PK₆::URA3</i> |
| C5033 | <i>MATa/a, scc3::natMX4/SCC3, ura3::SCC3(K224E, K225E, K322E, K330E, K423E, K520E, K669E)-PK₆::URA3/ura3</i> |
| C5043 | <i>MATa/a, scc3::natMX4/SCC3, ura3::SCC3(K423E, K520E, K669E)-PK₆::URA3/ura3</i> |
| C5054 | <i>MATa, scc3::natMX4, ura3::SCC3(K423E, K520E, K669E)-PK₆::URA3</i> |
| C5165 | <i>MATa, SCC1-HA₆::HIS3, ura3::SCC3-PK₆::URA3</i> |
| C5166 | <i>MATa, SCC1-HA₆::HIS3, ura3::SCC3(K224E, K225E, K322E, K330E, K423E, K520E, K669E)-PK₆::URA3</i> |

Table 4.4 Yeast genotypes.

To test whether the inability of the heptamutant version of Scc3 to support cohesin function was the result of a defect in the association of the mutant cohesin complex with chromosomes, Christian Haering's lab measured the levels of wild-type and mutant cohesin complexes at five independent binding sites in the budding yeast genome by chromatin

immunoprecipitation followed by quantitative PCR (ChIP-qPCR). The quantity of chromosomal DNA that co-immunoprecipitated with the Scc3 patch 3 mutant was on average 40% lower than that which co-immunoprecipitated with wild-type Scc3. The Scc3 heptamutant failed to bind chromatin entirely (Figure 4.8d), although it was incorporated normally into cohesin complexes (Figure 4.9b). We conclude that DNA binding by Scc3-Scc1 is essential for the stable association of cohesin complexes with chromosomes *in vivo* and hence an important determinant of cohesin function.

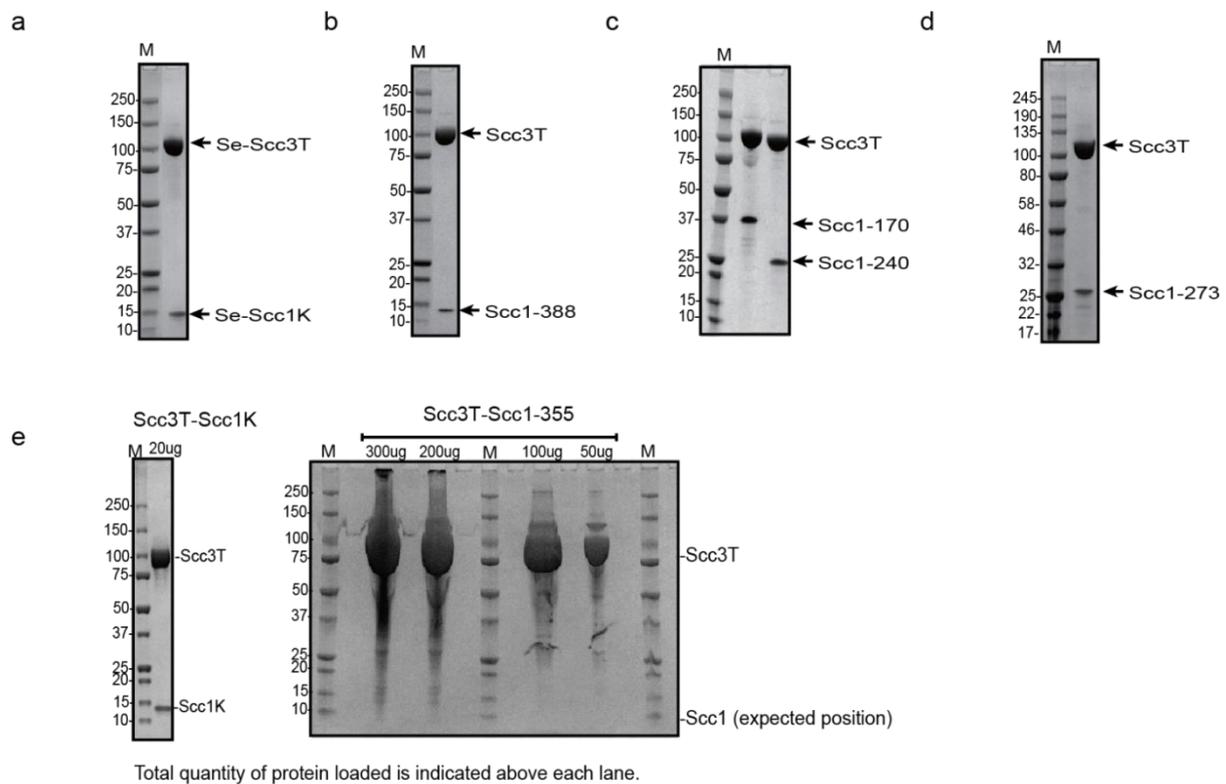


Figure 4.10 Purification of selenomethionine-labeled Scc3T-Scc1K and different Scc3-Scc1 truncations used for crystallization and FP trials. (a) Selenomethionine-labeled Scc3T-Scc1K used for structure optimization. (b-d) Scc3-Scc1 constructs used for crystallization trials. (e) Scc1(355-400) fails to form a complex with Scc3.

4.4 Screening for possible protein interactions among cohesin subcomplexes

In order to explore possible interactions between cohesin and its accessory factors, I performed pull-down assays with additional subunits available in the lab to identify potential physical interactions *in vitro*. I failed to detect an interaction between Smc3-Scc1, Smc1-Scc1 and Scc3-Scc1 subcomplexes in the presence of ATP- γ -S and / or a 31bp DNA substrate. (Figure 4.12a, b). I tested whether yeast Wapl interacts with other factors such as Smc3hd alone, Smc3hd-NScc1, ctSmc1hd-CScc1 Scc3T-Scc1K, or Clink125. For none of these components could I detect direct interactions under these conditions (Figure 4.12c, e). Pds5 full length might interact with Wapl, but with low affinity (Figure 4.12c).

The hinge domain is potentially a DNA entry gate for cohesin, and has been proposed to bend back onto the ATPase head domain, thus adopting a head-to-tail conformation which would deliver the required energy to open the hinge interface (Kurkcuoglu and Bates 2010). Such a folded conformation predicts that the hinge domain would directly interact with the cohesin head complex under certain conditions. As I was able to purify the assembled Smc1, Smc3, Scc1, Pds5 and Scc3 hexameric complex (Clink159-Scc3T-Pds5-fl), as shown in the previous section, I tested if the hinge heterodimer directly interacts with the Clink complex. However, in pull-down assays I could not detect direct binding (Figure 4.11d).

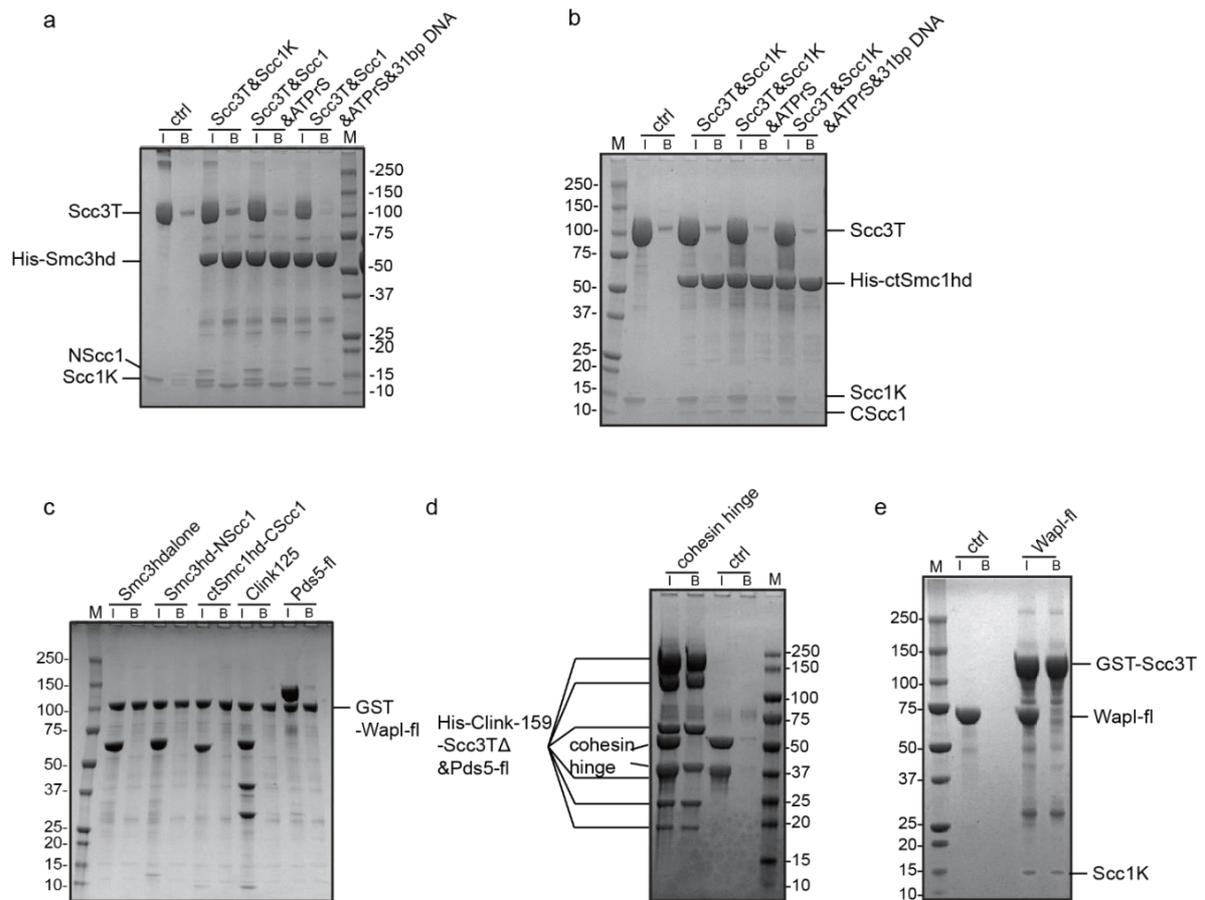


Figure 4.11 Screening for possible protein interactions among cohesin subcomplexes. (a) Scc3-Scc1 does not bind to Smc3hd-NScc1 or **(b)** ctSmc1hd-CSccl1 in the presence of ATP- γ -S +/- DNA. **(c)** yeast Wapl did not bind Smc3-Scc1, Smc3 head alone or ctSmc1-CSccl1 and Clink125, but might bind weakly to Pds5-fl. **(d)** The cohesin hinge did not bind Clink159-Scc3T Δ & Pds5-fl, the largest cohesin complex I could assemble. **(e)** Scc3T-Scc1K does not interact with Wapl.

Discussion_part 2

Discussion_partie 2

Dans cette étude, j'ai identifié un site de liaison à l'ADN dans la sous-unité Scc3 du complexe de la cohésine et j'ai déterminé la base moléculaire de son interaction avec l'ADN à une résolution quasi atomique. Ces découvertes fournissent la preuve d'un site de contact direct avec l'ADN dans le complexe de la cohésine, qui contribue vraisemblablement à l'étape initiale de piégeage de chromosome et / ou de translocation d'ADN. Je propose que le sous-complexe Scc3-Scc1 fournisse un point d'ancrage dynamique de l'ADN nécessaire au chargement et / ou au maintien efficace de la cohésine sur la chromatine. Ainsi, la liaison de l'ADN par le sous-complexe Scc3-Scc1 pourrait être la première étape du déplacement de l'ADN chromosomique dans le cycle de la cohésine.

Targeting of cohesin to the genome is essential for numerous aspects of chromosome biology, including sister chromatid cohesion, DNA damage repair, and transcriptional regulation (Uhlmann 2016). In this study, I identified a direct DNA-binding site in the Scc3 subunit of the cohesin complex and determined its interaction with DNA at near-atomic resolution. These findings provide evidence for a site of direct DNA contact in cohesin complexes, which presumably contributes to the initial step of chromosome entrapment and/or DNA translocation (Figure 4.11c).

We were able to determine the DNA-bound structure of Scc3 in complex with a minimal fragment of Scc1 and demonstrated that DNA binding depends on conserved positively charged residues within a composite Scc3-Scc1 interface. We used this structure to derive DNA-binding deficient Scc3 variants, which fail to support cohesin recruitment to chromatin and consequently cell division. In addition to providing a scaffold for the assembly of cohesin regulators (Hara, et al. 2014; Murayama and Uhlmann 2014; Roig, et al. 2014) and thereby participating in the maintenance of cohesion, the Scc3-Scc1 subcomplex also plays a key role in DNA substrate recognition, and hence the efficient association of the cohesin holocomplex with chromatin.

The cohesin ring has been proposed to topologically embrace chromosomal DNA, which requires that DNA be loaded into the ring complex during cohesin's catalytic cycle (Nasmyth 2011). As mutations in Scc3 that prevent DNA engagement by the Scc3-Scc1 subcomplex *in vitro* fail to stably bind to chromatin *in vivo*, it is likely that direct DNA-cohesin interactions contribute to such DNA capture (Figure 4.11c).

We propose that the Scc3-Scc1 subcomplex provides a dynamic DNA anchoring point that is required for the efficient loading and/or maintenance of cohesin on chromatin (Figure 4.11). Such a model for Scc3 is supported by previous data, which indicate that Scc3 contributes to

cohesin loading (Hu, et al. 2011; Murayama and Uhlmann 2014; Orgil, et al. 2015; Roig, et al. 2014). In agreement with this model, Scc1 deletion mutants that lack the sequence responsible for binding Scc3 fail to load onto yeast chromosomes (Hu, et al. 2011). Furthermore, Scc3 enhances topological DNA entrapment by cohesin in *in vitro* loading assays (Murayama and Uhlmann 2014). Thus, DNA binding by the Scc3-Scc1 subcomplex might be the first step in moving chromosomal DNA into the cohesin ring. The subsequent entrapment reaction is then presumably catalysed by the Scc2-Scc4 complex (Murayama and Uhlmann 2014; Stigler, et al. 2016).

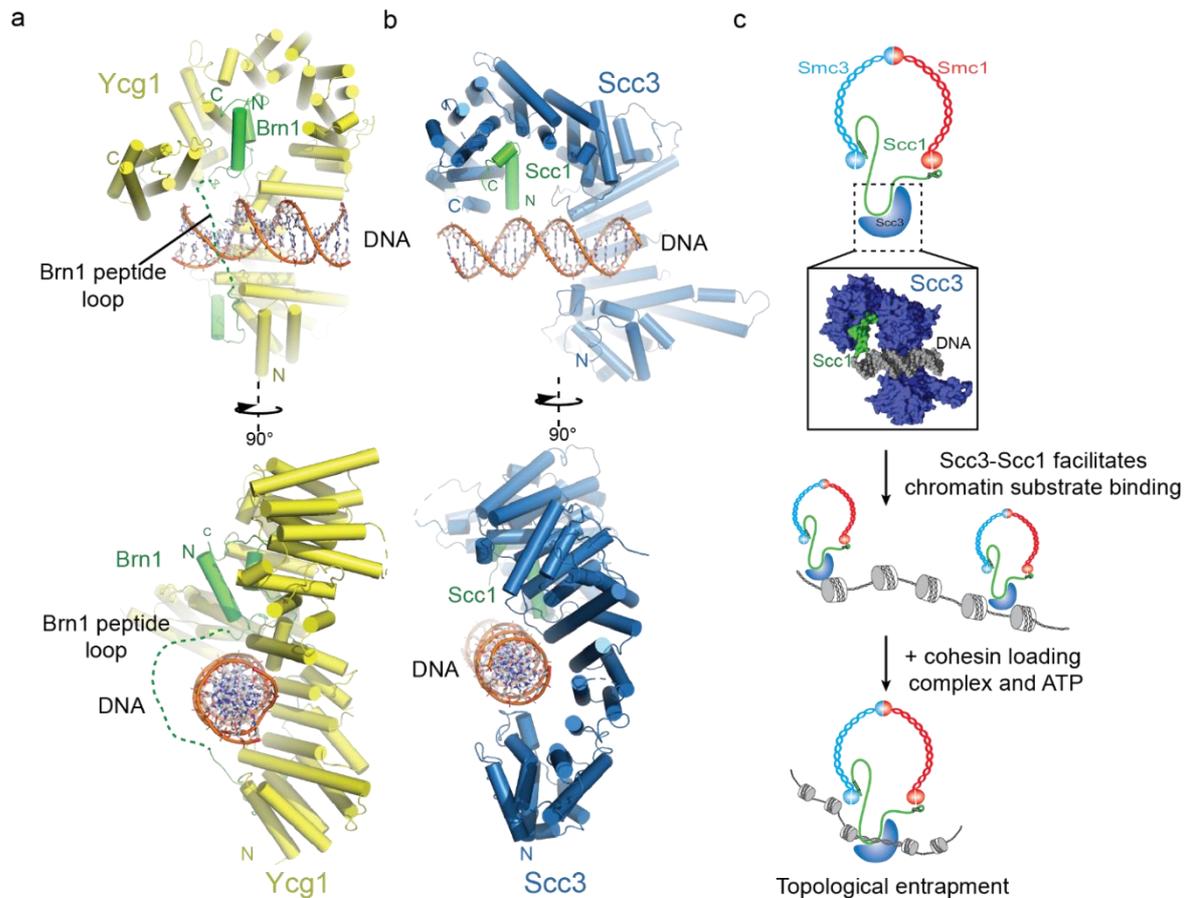


Figure 4.12 A conserved DNA binding interface in cohesin and condensin. (a) Structure of the DNA-bound Ycg1–Brn1 subcomplex from condensin. (b) Structure of the Scc3-Scc1 complex. In the condensin structure, a peptide loop (green dashed line) of the Brn1 kleisin subunit encircles the bound DNA and prevents its dissociation. Alignments were generated by secondary structure matching using C α atoms from the Scc3 HEAT-repeats and the structurally equivalent region of the condensin Ycg1 HEAT-repeat subunit. (c) Model for Scc3-mediated DNA binding by cohesin complexes and its possible role in sister chromatid cohesion or chromatin loop extrusion. Scc3-Scc1 enables direct chromatin binding. Cohesin is loaded by Scc2-Scc4 in an ATP dependent fashion.

Surface patch 3 mutations in Scc3, which partially ablate DNA binding by the Scc3-Scc1 subcomplex and reduce cohesin levels on chromatin, do not exhibit any obvious growth defects and are therefore presumably competent to establish sister chromatid cohesion. Indeed, partial depletion of cohesin does not seem to impact some of its core functions, including sister-chromatid cohesion and chromosome segregation (Elbatsh, et al. 2016; Heidinger-Pauli, et al. 2010). In contrast, even mutations that only slightly impair DNA binding of the equivalent HEAT-repeat / kleisin module of the condensin complex are sufficient to abolish stable chromatin association and to interfere with cell division (Kschonsak, et al. 2017). Such discrepancies might be due to alternate loading and/or maintenance mechanisms of cohesin and condensin. Whilst cohesin is loaded by the Scc2-Scc4 complex, no such independent loading factor has been identified for condensin thus far, which could explain why the latter depends more strongly on the direct DNA binding site formed by its HEAT-repeat and kleisin subunits.

Binding of condensin to DNA is further stabilized by the entrapment of the bound DNA helix within a kleisin peptide loop (Figure 4.11a, b) (Kschonsak, et al. 2017). The relevant section of Scc1 that would contribute to such topological DNA entrapment is not present in our structure. As Scc1 is clearly required for DNA binding of the Scc3-Scc1 subcomplex, it is possible that cohesin uses a similar mode of chromatin engagement. These findings thus point towards a conserved molecular mechanism that enables chromatin substrate engagement by condensin and cohesin. This mechanism potentially facilitates topological loading, chromatin looping and tracking along chromatin fibres by these SMC complexes (Ganji, et al. 2018).

**Result_part 3 Principles of cohesin regulation on
chromatin via its SA2-Scc1 subunits**

Résultats_partie 3

Pour étudier les bases moléculaires de la collaboration fonctionnelle signalée entre la cohésine et le CTCF dans la structure tridimensionnelle du chromosome, j'ai identifié et déterminé la structure d'un complexe ternaire composé de SA2 humain (un orthologue de Scc3), de Scc1 et de CTCF. La structure révélait un motif de liaison SA2-Scc1 très répandu qui était présent non seulement dans la CTCF, mais aussi dans d'autres facteurs connectés, tels que la shugoshin, Wapl, un autre facteur accessoire de la cohésine, soit pour la protection du centromère, soit la libération de la cohésine aux bras du chromosome. Les tests de compétition par "pull-down" ont indiqué que la liaison de ces facteurs à SA2-Scc1 était mutuellement exclusive, ce qui suggère fortement qu'ils interagissent avec la cohésine via des mécanismes similaires. Pour démontrer ce principe, j'ai pu déterminer une structure de shugoshin en complexe avec SA2-Scc1, ce qui a confirmé que la shugoshin et le CTCF se lient à la même surface conservée sur la cohésine.

Les résultats d'un "peptide array" contenant un motif de la cohésine impliquent potentiellement le site de liaison SA2-Scc1 plus largement dans la régulation des gènes, la cohésion centromérique, la formation d'éléments latéraux méiotiques, ainsi que la libération de la cohésine.

5.1 Motivation to work on human SA2

Following the yeast Scc3-Scc1-DNA structure, I wanted to understand how SA2, a human ortholog of yeast Scc3, functions in the regulation of chromosomal cohesin. I therefore co-expressed and purified the SA2-Scc1 complex from *E.coli* (Figure 5.2). The codon optimized gene corresponding to the SA2 construct (residues 80-1060, named 'SA2T'), originally from Hongtao Yu's lab (Hara, et al. 2014), was inserted into the pGEX-6p vector. An Scc1 construct spanning amino acids 281-420 was inserted into the pACYC-Duet vector (Figure 5.2a, c). The two expression plasmids were co-transformed into *E.coli* and the SA2-Scc1 complex was co-expressed and co-purified via GST affinity purification, anion exchange chromatography and a final size-exclusion chromatography step. The SA2T-Scc1-281 complex showed considerable degradation of Scc1 (Figure 5.2, c). A shorter construct of Scc1 (317-400) which lacked the disordered region of Scc1 as predicted from the published SA2-Scc1 structure (PDB: 4PJU) (Figure 5.2a, Table S1) was designed and co-expressed with SA2T. This truncated construct showed no degradation (Figure 5.2d). In order to test whether human SA2-Scc1 binds to DNA, EMSA assays were performed with a 32 bp DNA oligonucleotide. However, no DNA binding was detected under equivalent conditions which yield robust DNA binding of yeast Scc3-Scc1 (Figure 5.2b).

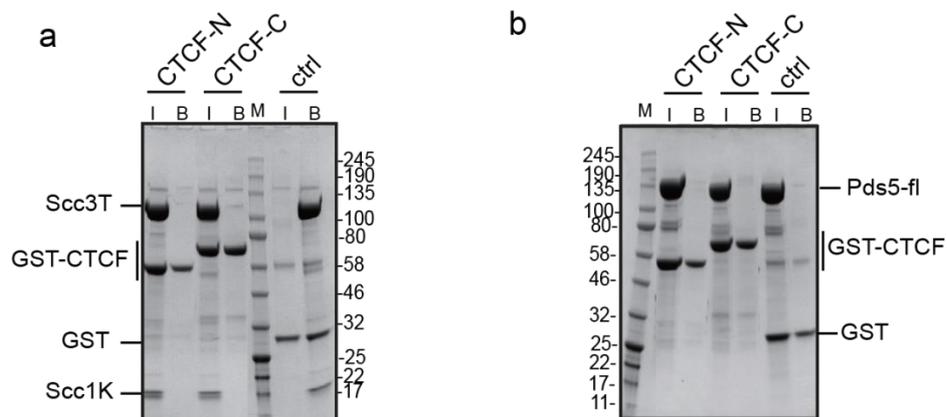


Figure 5.1 Yeast Pds5 and Scc3 do not bind human CTCF (a-b), indicating that the yeast cohesin lacks the essential residues that are responsible for CTCF binding.

5.2 Mapping SA2-Scc1 binding regions on CTCF

SA2 has been reported to interact with CTCF (Xiao, et al. 2011), a zinc finger protein with strong sequence-specific DNA binding affinity (Hashimoto, et al. 2017). As demonstrated in figure 5.2b, SA2-Scc1 does not interact directly with DNA. One possibility therefore is that human SA2-Scc1 associates with DNA through accessory factors such as CTCF (Figure 5.2, g), thus implying a physical relationship between cohesin and CTCF. To investigate the mechanism of SA2-Scc1 binding to CTCF, I generated different truncations of CTCF and tested in pull-down assays their ability to retain SA2-Scc1. I found that an N-terminal fragment, but not the C-terminal domain interacted robustly with SA2-Scc1 (Figure 5.3a, b). All CTCF constructs were designed and inserted into pGEX-6p vector and purified by GST affinity, anion exchange and size-exclusion chromatography (data not shown), the GST tag was either cleaved or not, depending on their intended purpose. Based on these results, I was able to narrow down the SA2T-Scc1-317 binding region to amino acids 227-235 of CTCF (Figure 5.3a, b).

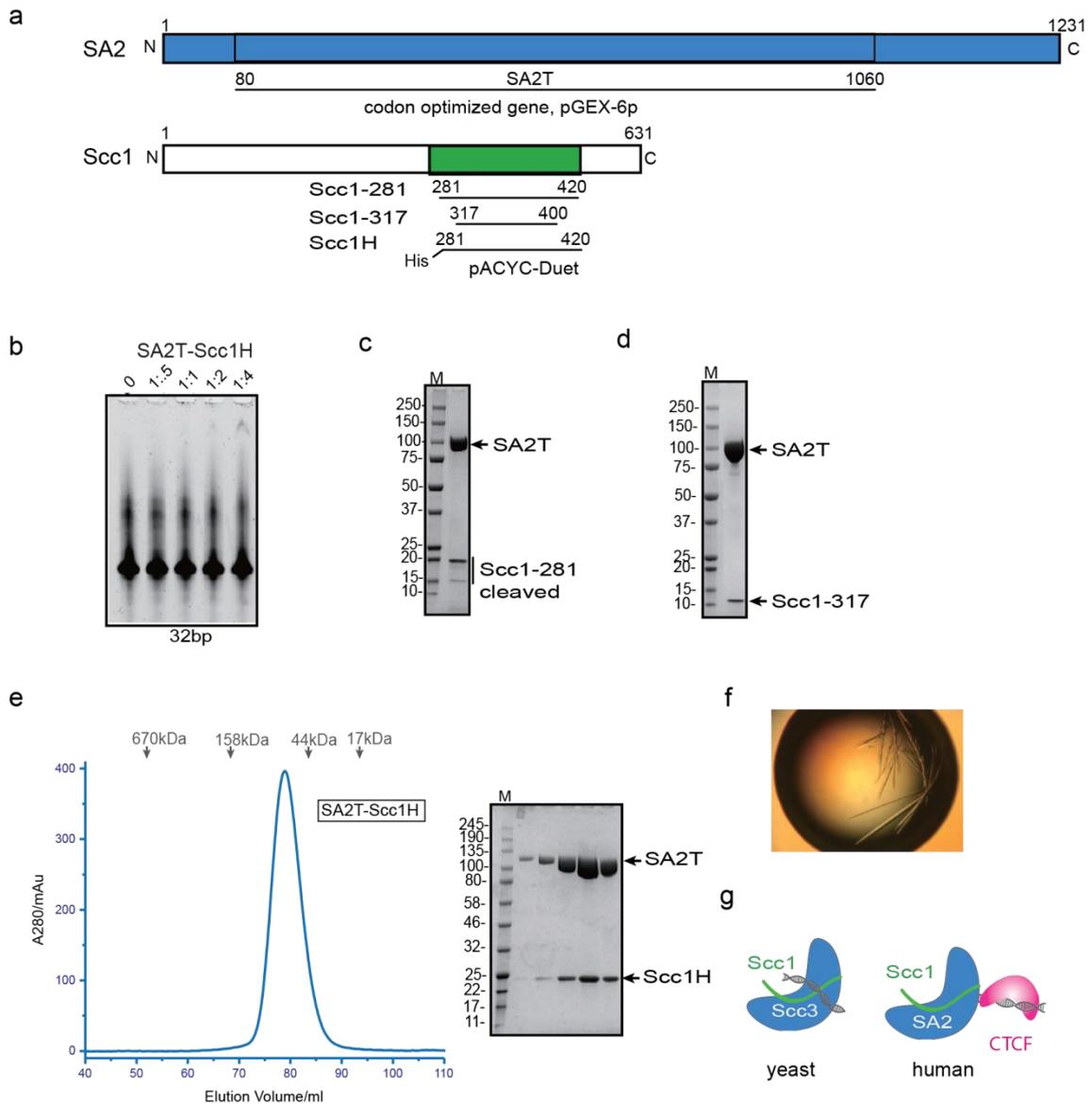


Figure 5.2 SA2-Scc1 constructs, purification and EMSA analysis. (a) Cartoon of human SA2-Scc1 constructs used for crystallization and interaction studies. (b) EMSA analysis of hSA2-Scc1 complex with a 32 bp dsDNA substrate (Table 4.1), showed no DNA binding for hSA2-Scc1. (c-d) SDS-PAGE analysis of hSA2 with different truncations of Scc1. The Scc1 construct (281-420) is partially degraded after expression/purification from *E.coli*. Construct (317-400), contains a truncation of the disordered regions. (e) Purification of SA2T-Scc1H complex used for crystallization and interaction analysis. (f) Crystal image of the SA2T-Scc1H complex used for peptide soaking. (g) Proposed model on how yeast and human Scc3 associate with DNA. In yeast, Scc3 binds with DNA

directly, but in human or higher vertebrates, Scc3 has lost the ability to bind to DNA, but has associates with DNA through CTCF mediated recruitment.

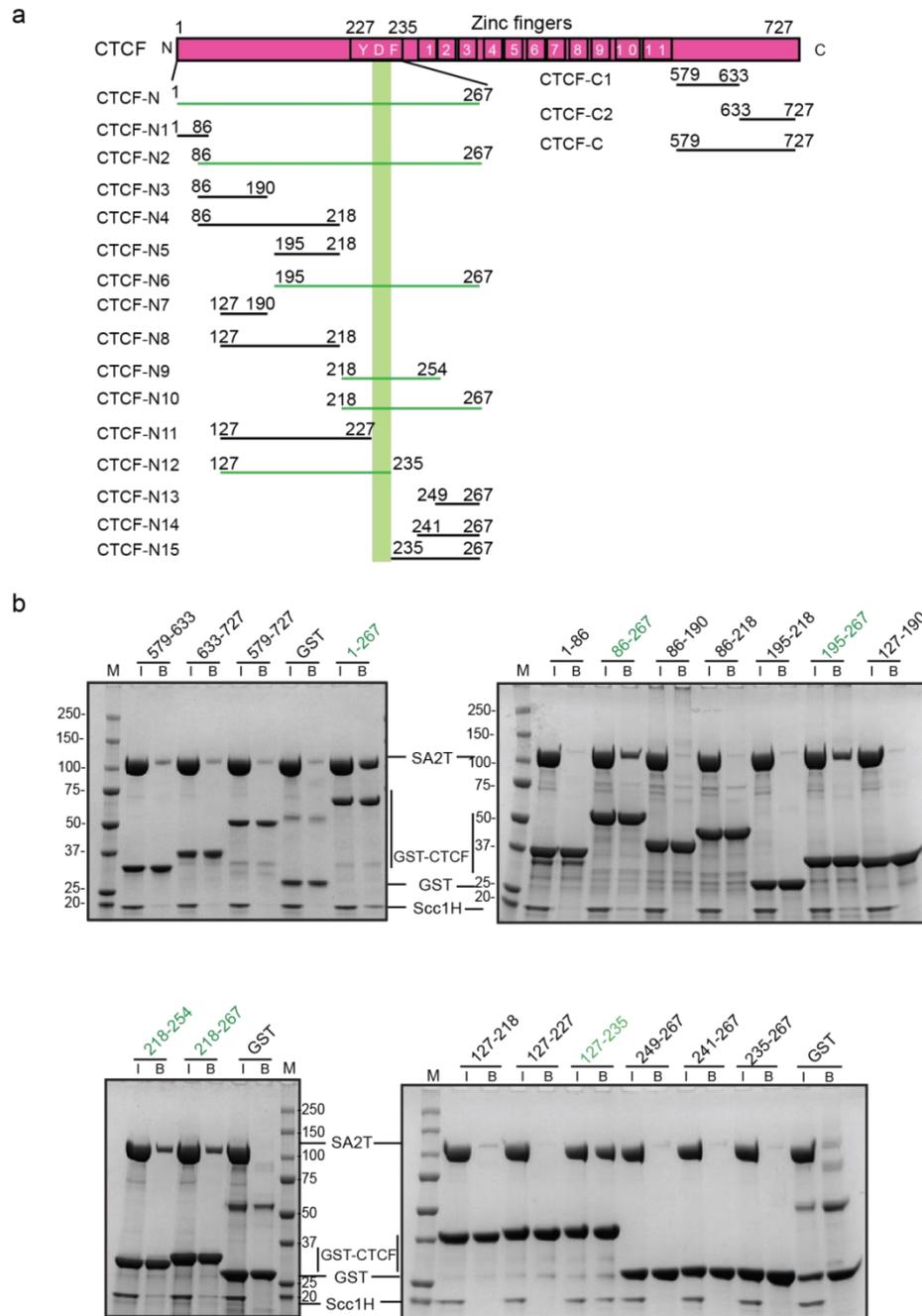


Figure 5.3 Biochemical analysis of CTCF binding to SA2-Scc1. (a) Domain architecture of CTCF. The CTCF fragments tested for SA2-Scc1 binding in GST pull-down analyses are indicated. Fragments that bind SA2-Scc1 are shown in green. (b) Summary of data showing results of GST

pulldowns. The input (I) and the bound (B) fraction were analysed by SDS-PAGE. CTCF fragments that bind SA2-Scc1 are shown in green.

5.3 Crystallization trials of the SA2-Scc1-CTCF ternary complex

In order to obtain the structure of the SA2-Scc1-CTCF complex, I first tried to co-express SA2T-Scc1-317 with a CTCF fragment spanning 1-267 (CTCF-N) that covers the binding region 227-235 of CTCF (Figure 5.5a). The complex of CTCF-N and SA2T-Scc1-317 dissociated during ion exchange chromatography (Figure 5.5a, left gel). To restore the ternary complex, fractions containing CTCF-N and SA2T-Scc1-317 were mixed again and diluted to lower salt concentration (150 mM NaCl), followed by size exclusion chromatography (Figure 5.5a, right gel). This showed that CTCF and SA2-Scc1 assembled into a ternary complex. As complex assembly was dependent on the ionic strength suggested a possible charge-based interaction between these proteins.

In another attempt to produce the ternary complex, CTCF (CTCF-N2, amino acids 86-267) and SA2T-Scc1-317 were purified separately and then mixed at a 3:1 ratio (Figure 5.5b) prior to size exclusion chromatography. The ternary complex was subjected to crystallization (by the HTX lab EMBL Grenoble), either directly or in the presence of small amounts of trypsin or chymotrypsin for *in situ* digestion (ratio protease: complex - 1 : 5000). However, no crystals were obtained, irrespective of the conditions. Isolated CTCF-N2 digested *in situ* with trypsin digestion yielded crystals, but they were not investigated further.

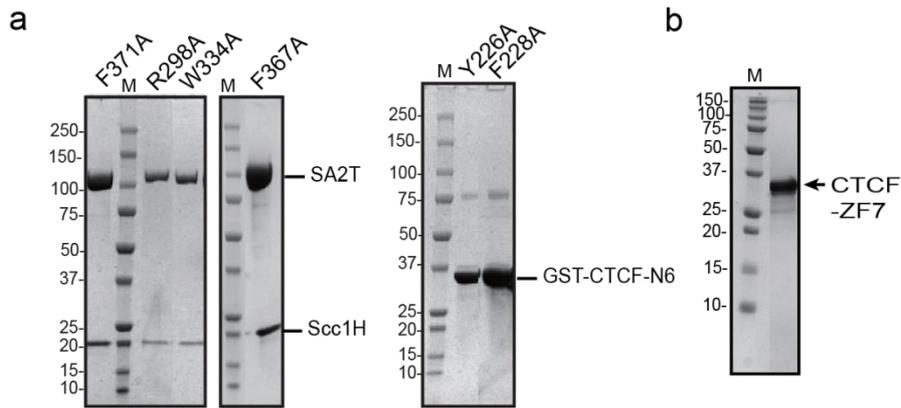


Figure 5.4 Purification of SA2-ScclH mutants, CTCF-N6 mutants and CTCF covering both SA2-Sccl1 binding region and zinc finger 1-7 (CTCF-ZF7), used for pull-down and EMSA.

5.4 SA2-Sccl1-CTCF ternary complex is sensitive to DNA binding

In order to investigate if the ternary complex consisting of SA2, Sccl1 and CTCF interacts with DNA similar to CTCF alone, SA2T-Sccl1H-CTCF-ZF7 was incubated with a 62 bp DNA oligonucleotide containing a consensus CTCF recognition site (labeled in bold) (Table 4.1, Figure 5.5) and the resulting complexes analyzed via EMSA. The CTCF-ZF7 construct covers both the SA2-Sccl1 binding region and zinc fingers through to zinc finger 7. The results showed that both CTCF alone as well as the ternary complex are able to interact with DNA (Figure 5.5c). I also observed that CTCF partially dissociated from the ternary complex when bound to its DNA substrate (Figure 5.5c).

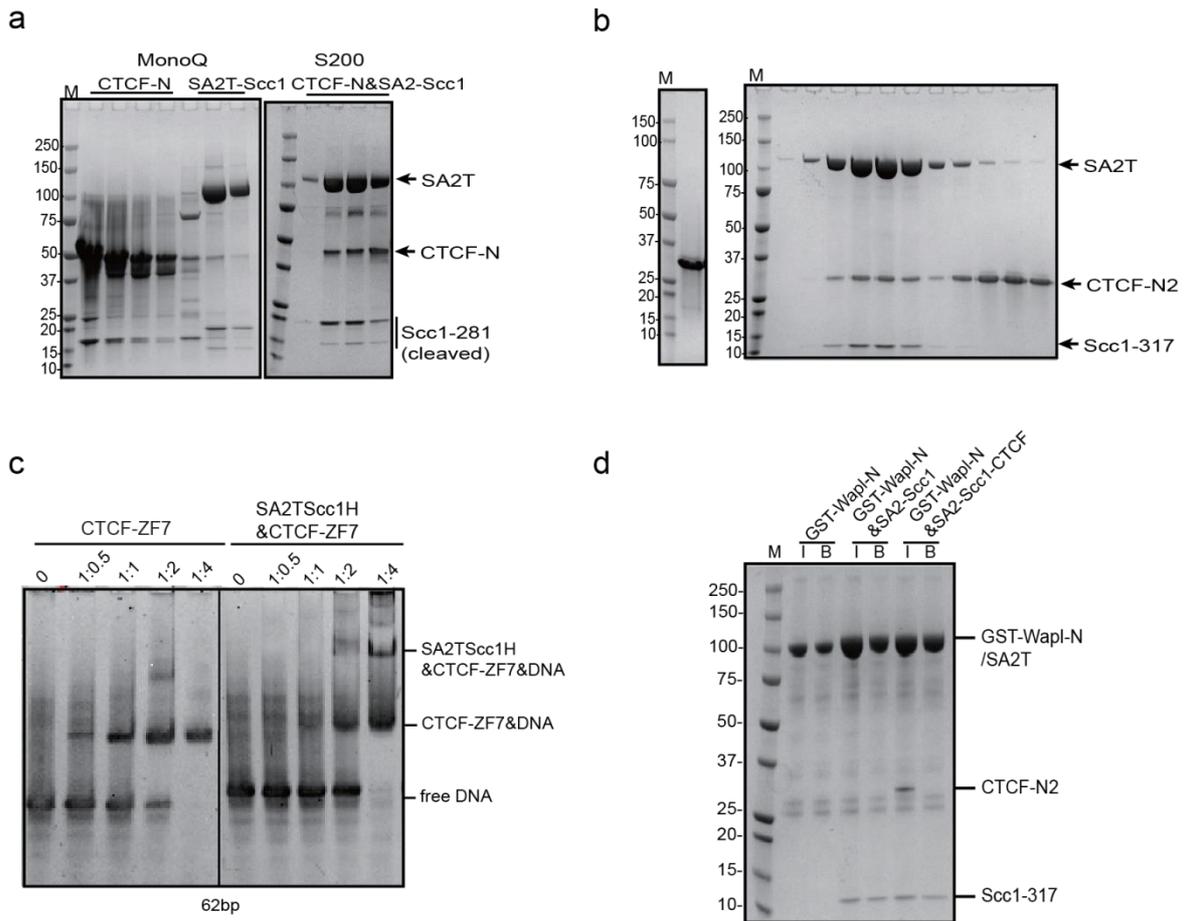


Figure 5.5 Co-expression and co-purification of the SA2T-Scc1H-CTCF complex and DNA binding assay. (a) When co-expressed together, CTCF-SA2T-Scc1H complex partially falls apart after anion exchange (left SDS-PAGE), but form a complex again at lower salt concentration (150 mM NaCl) as shown on the size-exclusion profile (right SDS-PAGE). (b) SA2T-Scc1H and CTCF-N2 are purified separately and then mixed at a 1:3 ratio prior to size exclusion (in low salt, 150 mM NaCl), and co-elute, demonstrating an interaction. (c) A specific 62 bp DNA oligonucleotide containing both CTCF (in bold) and a DNA sequence previously shown to interact with yeast Scc3 (5'-ATTG CAGT GCCC ACAG AGGC CAGC AGGG GCG TAGT GAGG CCTT TTCA AGGA AACG AAAG TG- 3') (Table 4.1) was used for an EMSA. A CTCF construct covering both the SA2-Scc1 interaction region and zinc fingers 1-7 (CTCF amino acids 218-460) was able to shift DNA (left). In the presence of SA2T-Scc1H, I observed a supershifted complex presumably corresponding to the SA2T-Scc1H-CTCF-ZF7 complex. Thus, while SA2T-Scc1H does not bind DNA strongly on its own, in contrast to yeast Scc3-Scc1, it does bind CTCF-DNA complexes. (d)

Competition between Wapl and CTCF for SA2-Scc1 binding. GST-Wapl was incubated alone, with SA2-Scc1 or with a complex of SA2-Scc1/CTCF N2. In the presence of Wapl, CTCF is displaced from SA2-Scc1. The input (I) and the bound (B) fraction were analysed by SDS-PAGE.

5.5 Structure of SA2-Scc1 in complex with a CTCF peptide

As SA2-Scc1 has previously been crystallized in isolation, I decided to reproduce the SA2-Scc1 crystals. SA2T-Scc1 spanning amino acids 317-400 was set up for crystallization but failed to crystallize. I then designed a construct spanning amino acids 281-420, closer to the published crystallization construct (Figure 5.2a), but with an additional uncleavable His tag at the N terminus (Scc1H), hereafter called either SA2T-Scc1H or SA2-Scc1. This allowed purification of SA2-Scc1 via metal affinity chromatography to separate it from the degraded Scc1 (the N-terminally degraded Scc1 does not retain the tag), followed by anion exchange and size exclusion chromatography.

My domain-mapping experiments implicated CTCF residues 227-235 as a minimal SA2-Scc1 interacting region. Thus, informed by these experiments and sequence conservation of CTCF (Figure 5.10), we synthesized a CTCF peptide (sequence: DSVYDFEEE) spanning amino acid residues 220-230 of CTCF. I further confirmed the interaction between the CTCF peptide and SA2T-Scc1-317 by ITC, and found the affinity to be comparable to that of the intact CTCF N2 construct. Crystals were obtained using protocol similar to that published previously (Hara, et al. 2014). I harvested around 80 crystals and sent them to MASSIF for automated screening/data collection. The best crystals diffracted to a minimal Bragg spacing of 2.6 Å and crystals of unliganded SA2-Scc1 to 2.4 Å (Table 5.1). We determined the structure by molecular replacement and an Fo-Fc omit map exhibited clear features for the CTCF peptide (Figure 5.7c) The refined model encompasses amino acid residues 223-231 of CTCF, 80-1048 of SA2 and 321-393 of Scc1.

The CTCF peptide is bound to the convex surface of SA2, in close proximity to the N-terminal 'nose' which is formed by a pair of extended antiparallel α 9-10 helices (Figure 5.6c, d) The CTCF binding surface is predominantly hydrophobic and composed of amino acids contributed by both SA2 and Scc1. The lead 'anchoring' amino acids of CTCF, which bury the largest solvent-accessible surface area upon binding, are Y226 and F228 (Figure 5.6d) F228 inserts into a pocket made up of amino acids from Scc1 (S334, I337, L341) and SA2 (Y297, W334) (Figure 5.6e) The hydroxyl group of Y226 hydrogen bonds with E326 of SA2 in a deep hydrophobic pocket lined by L329, F366 and F367 (Figure 5.6f) E229 and E230 of CTCF, constitute secondary anchoring residues which presumably contribute to binding specificity by forming salt bridges with R298 of SA2 and R338 of Scc1. Because CTCF engages a composite binding surface containing amino acids from Scc1 and SA2, prior mapping studies using isolated SA2 are likely to have been misleading (Xiao, et al. 2011).

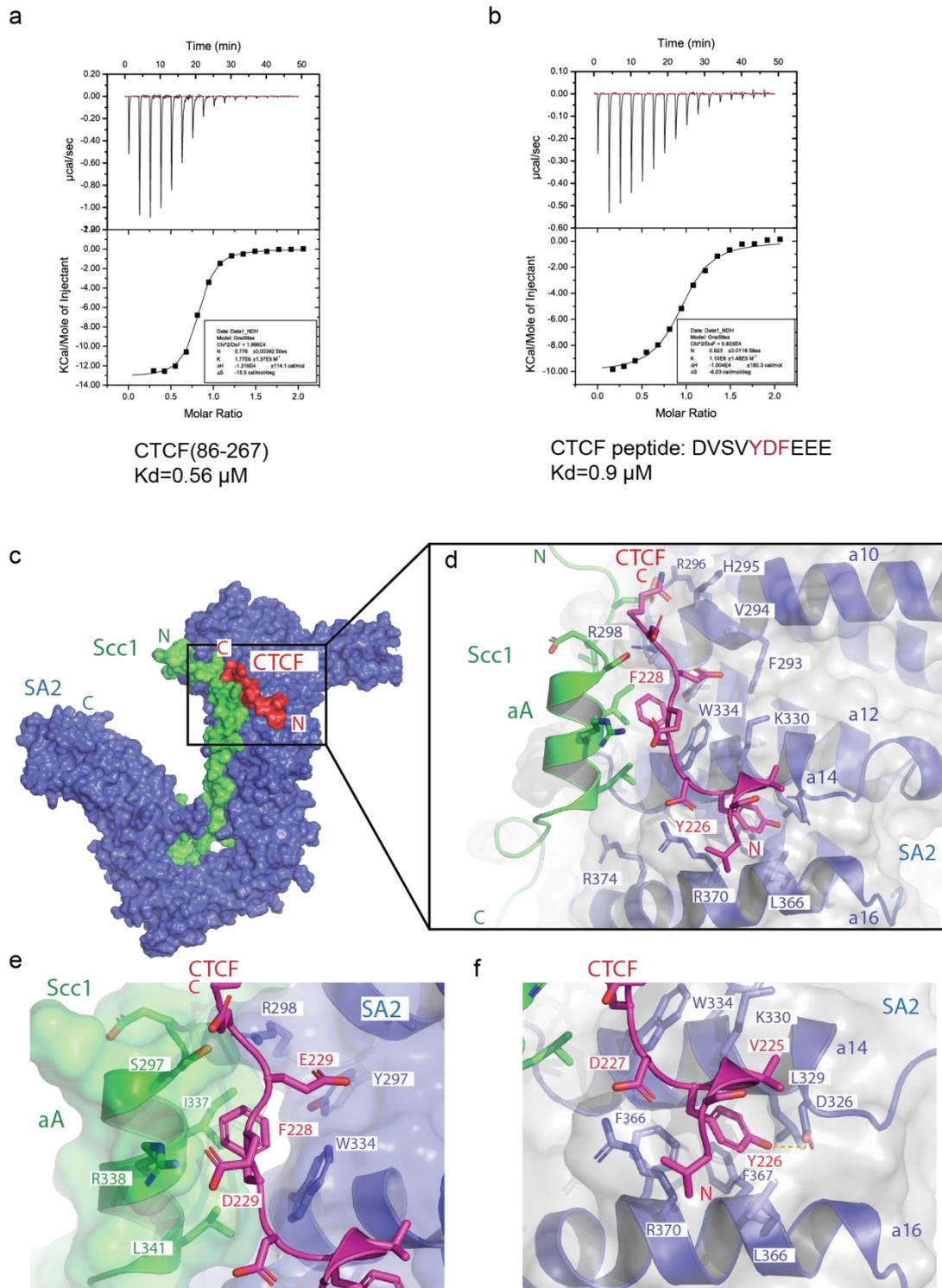


Figure 5.6 Structure of hSA2-Scc1 bound to a fragment of CTCF. (a-b) ITC progress curves of binding between SA2-Scc1 and CTCF 86-267 (left) or a synthetic CTCF peptide spanning 222-231 (right). The binding stoichiometry (N) and dissociation constant (K_d) are indicated. (c) Surface-rendered cartoon of the SA2-Scc1-CTCF complex colored in blue, green and magenta,

respectively. **(d)** Detailed view of the binding interface with SA2 residues in blue, Scc1 in green and CTCF in magenta. **(e)** Details of the composite binding pocket around F228 and **(f)** around Y226 of CTCF.

5.6 Mutagenesis analysis of the SA2-Scc1-CTCF ternary complex

As expected, mutagenesis of Y226A or F228A in CTCF abolished SA2-Scc1 binding in a GST pulldown assay (Figure 5.7a). Likewise, mutation of critical amino acid residues in SA2, including W334A, F371A and F367A or in Scc1 I337A/L341A, abolished CTCF binding. SA2 contains a 86 amino acid motif, the so-called ‘conserved essential surface’ (CES) (Hara, et al. 2014; Roig, et al. 2014), which is conserved from fungi to mammals. For simplicity, henceforth we will refer to the SA2-Scc1 binding pocket as ‘CES’ collectively. Correspondingly, mapping sequence conservation onto the structure shows that the CTCF binding site is highly conserved (Figure 5.8) A series of missense mutations are found in SA2, Scc1 and CTCF in various cancer tissues (cancer.sanger.ac.uk; cosmic; (Forbes, et al. 2008)). Mapping of mutation frequencies onto the structure shows that R370 of SA2, L341 of Scc1 and the key anchoring residue Y226 of CTCF, which are all largely buried upon complex formation, are hotspots in cancer (Figure 5.8c) indicating that this interface is critical for normal cell physiology and that its disruption contributes to pathogenesis.

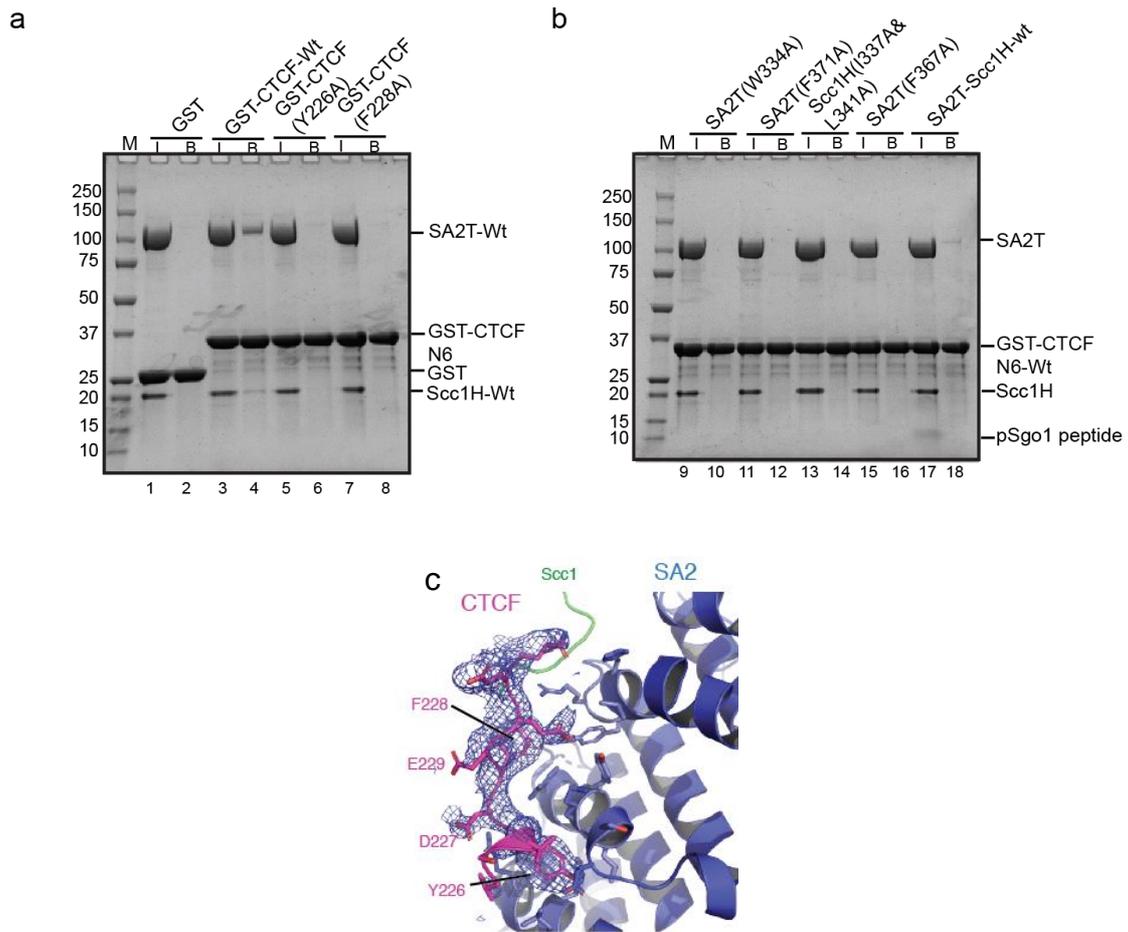


Figure 5.7 Mutagenesis and in vitro analysis of the SA2-Scc1-CTCF ternary complex.

(a) The indicated GST-CTCF variants were incubated with SA2-Scc1 and the input (I) and the bound (B) fraction analysed by SDS-PAGE. Mutation of Y226A or F228A of CTCF abolished SA2-Scc1 binding. (b) The indicated variants of SA2 or Scc1 were incubated with GST-CTCF wild-type and the input (I) and the bound (B) fraction analysed. Lane 9-10: Incubation with a synthetic pSgo1 peptide competes with CTCF for binding to SA2-Scc1. (c) Fo-Fc omit map contoured at 3 RMSD.

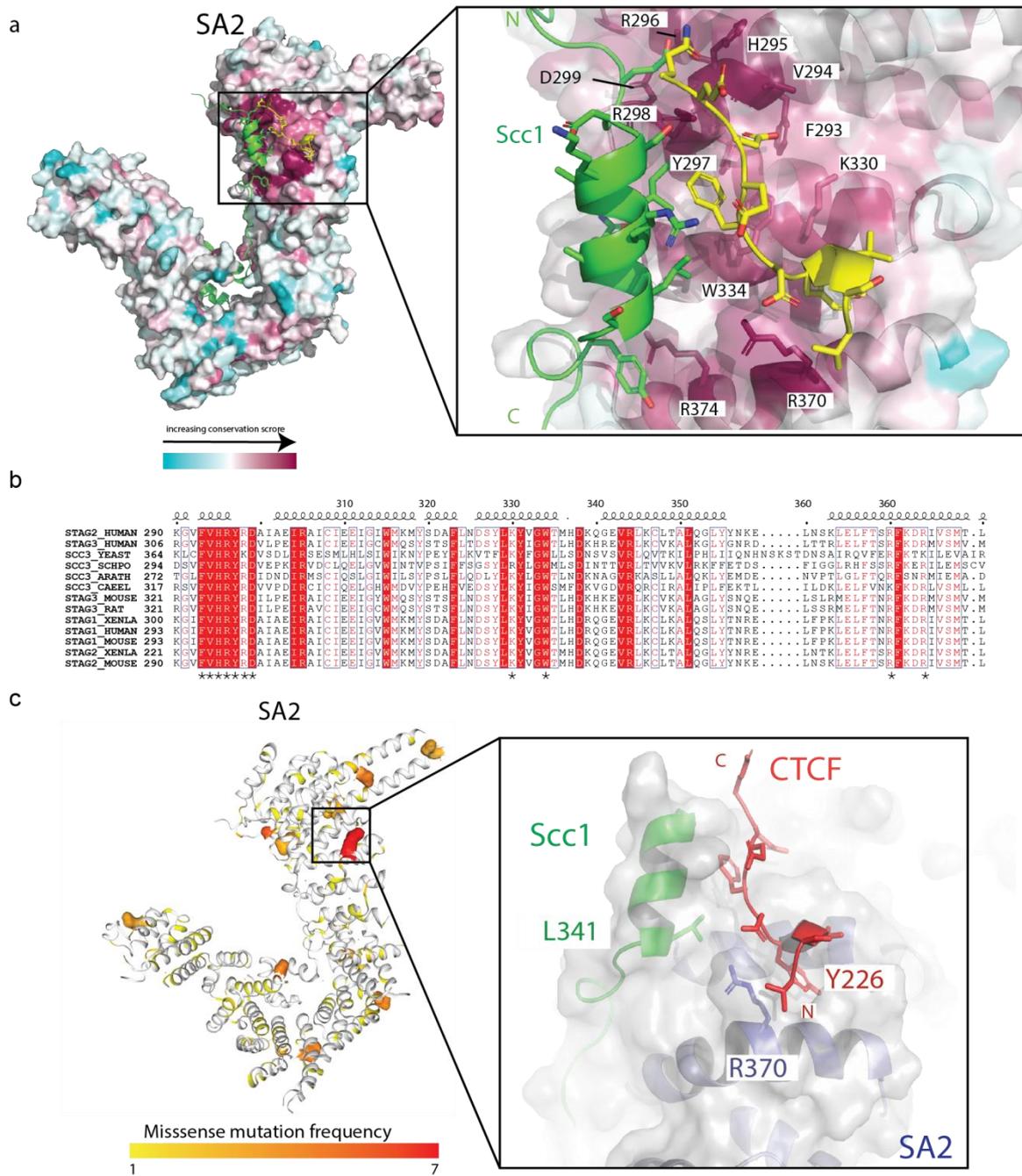


Figure 5.8 Analysis of the SA2-Scc1-CTCF structure. (a) Surface representation of the complex with SA2 surface rendered and colored according to sequence conservation. The inset shows a detailed view of conserved SA2 residues around the CTCF binding site. (b) Multiple sequence alignment of SA2 (STAG2) orthologs and paralogs. The key amino acid residues engaging CTCF are indicated by (*). (c) Missense mutation frequencies have been plotted onto the SA2 structure. The inset shows mutation hotspots R370 (SA2), Y226 and F228 (CTCF) and S334, L338 and L341 (Scc1).

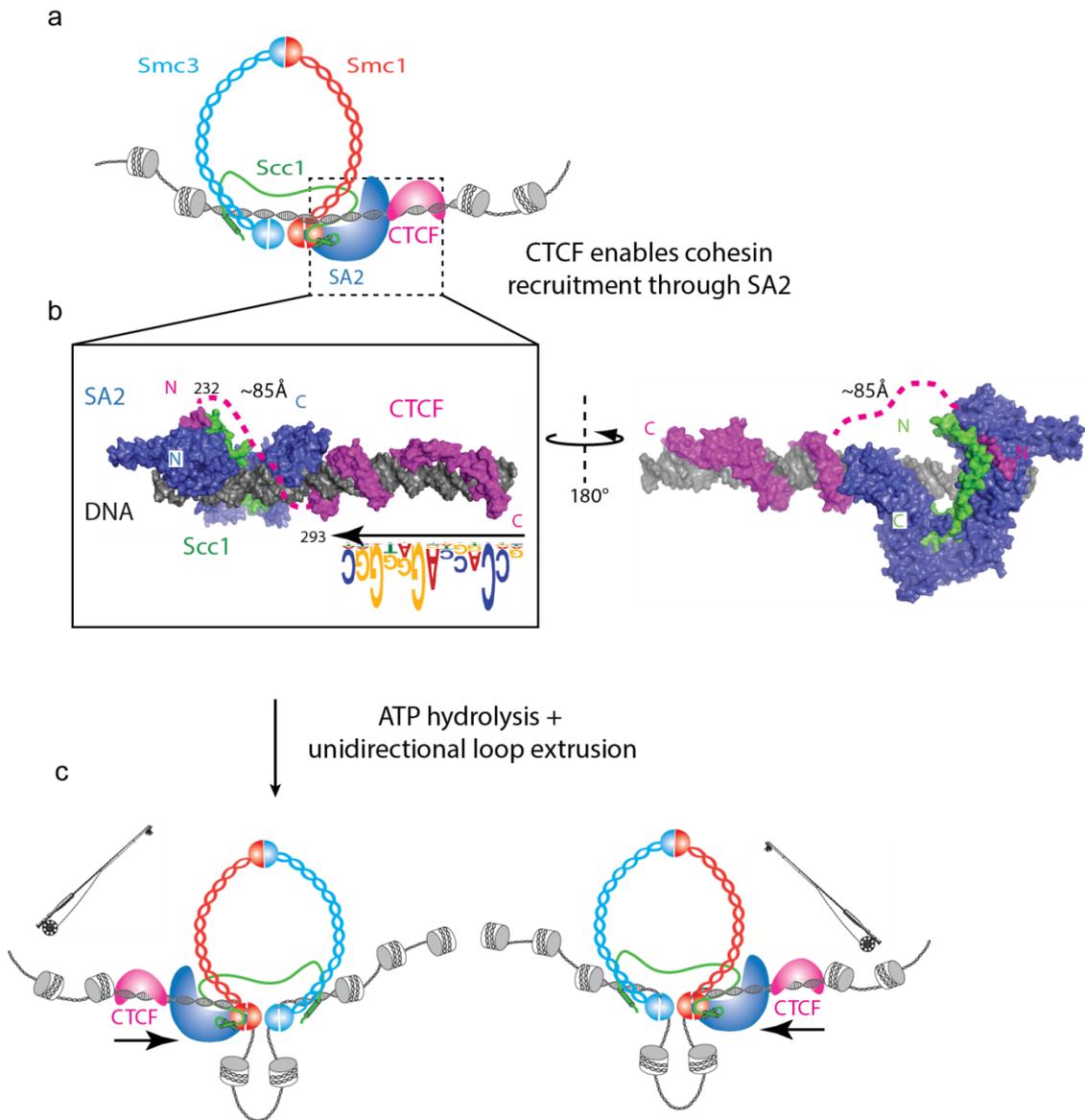


Figure 5.9 Model for CTCF-mediated cohesin binding and for the generation of TADs by convergently oriented CTCF sites. (a) CTCF engages cohesin through the SA2-Scc1 complex. **(b)** Molecular model of CTCF and SA2 bound to DNA. SA2-Scc1 is attached through an N-terminal Y/FxF motif to the N-terminus of CTCF. The N-terminal Y/FxF motif is separated by a flexible linker spanning residues 232-267 (magenta dotted line) to the C-terminal Zinc finger DNA binding domain of CTCF. This architecture positions SA2-Scc1, and by extension cohesin, on the downstream side of the core CTCF DNA binding motif. **(c)** The asymmetric attachment of cohesin to

the N-terminus of CTCF and ATP-fueled asymmetric DNA loop extrusion by cohesin results in loop formation and explains why CTCF sites on the genome are oriented convergently.

5.7 Analysis of additional cohesin ligands containing Y/FxF motifs

Previous data indicate that the SA2-Scc1 complex interacts with multiple cohesin regulators (Beckouet, et al. 2016; Hara, et al. 2014; Roig, et al. 2014). The opposing functions of two such factors, Wapl, the general cohesin inhibitor, and Shugoshin (Sgo1) is crucial for the protection of centromeric cohesion during mitosis (Gandhi, et al. 2006; Hara, et al. 2014; Kueng, et al. 2006; Liu, et al. 2013). This antagonism potentially arises as a result of direct competition for binding to SA2-Scc1 (Hara, et al. 2014). As mutants previously reported to interfere with both Sgo1 and Wapl binding cluster in the CTCF binding site on SA2-Scc1, we investigated whether these proteins are recruited to the CES by a comparable mechanism. In Sgo1, the reported CES-binding domain (amino acids 313- 353) contains a conserved Y/FxF motif which strongly resembles that of the CTCF peptide (Hara, et al. 2014). Vertebrate Wapl also contains several FGF motifs in its N-terminal region which are potentially involved in cohesin regulation (Ouyang, et al. 2016; Shintomi and Hirano 2009). A minimal fragment of Wapl capable of competing with Sgo1 for access to the CES (amino acids 410-590) contains two such FGF repeats (Hara, et al. 2014). To determine whether Wapl and Sgo1 are able to compete against CTCF for access to the SA2-Scc1, we performed GST-pulldown competition assays. As predicted, a GST-Wapl fusion protein spanning amino acids 1-600 bound to SA2-Scc1 and excluded CTCF from SA2-Scc1 (Figure 5.5d).

Similarly, titration of a Sgo1 phospho-T364 peptide, previously reported to preclude Wapl binding (Hara, et al. 2014), also displaced CTCF from SA2-Scc1 (Figure 5.7b, lane 17&18). To further understand this interaction, I synthesized a phosphorylated Sgo1 peptide (pSgo1,

331-349, sequence: SNDAYNFNLEEGVHLpTPFR) and was able to validate and quantify binding by ITC ($K_d = 2.32 \mu\text{M}$) (Figure 5.11a, right). I soaked the Sgo1 peptide into SA2-Scc1 crystals as done for CTCF, but crystals melted immediately after adding the peptide (Figure 5.11b). Based on the CTCF-liganded model of the SA2-Scc1, we predicted that the C-terminal phosphorylation motif would interfere with formation of a crystal contact (not shown). I therefore synthesized a shorter peptide lacking the phosphorylation site (Sgo1-shorter, 331-341, sequence: SNDAYNFNLEE) and tested the binding to SA2-Scc1 via ITC. This yielded a K_d of $13.5 \mu\text{M}$ (Figure 5.11a, left), thus indicating that the C-terminal section of the phosphopeptide contributes to the ~ six times higher binding affinity. The peptide was soaked into the SA2-Scc1 crystals similarly to CTCF peptide, but for much longer soaking time (7 days).

Crystals were sent to MASSIF for automated screening/data collection. The best crystals diffracted to a minimum Bragg spacing of $\sim 3.2 \text{ \AA}$ (Table 5.1) and we were able to determine the structure by molecular replacement (Figure 5.12a, b). The lead ‘anchoring’ residues on Sgo1 are Y335 and F337. As the phosphorylated Sgo1 (pSgo1) significantly increases the binding affinity (Figure 5.11a, b), there must be additional phosphorylation-mediated interactions that contribute the Sgo1 binding, thus explaining how Sgo1 phosphorylation enables protection against Wapl-mediated release at centromeres (Hara, et al. 2014).

As expected, although non-phosphorylated Sgo1 binds to SA2-Scc1 weakly, mutagenesis of Y335A or F337A in Sgo1-315 (regions of Sgo1-315-360) abolished SA2-Scc1 binding in a GST pulldown assay (Figure 5.12c). Hence we conclude that the CES of SA2-Scc1 is a general interaction hub for Y/FxF-motif containing cohesin regulators (Figure 5.10, Figure 5.14). The intermolecular competition reported between these factors here provides mechanistic insight into the previously observed antagonism of Sgo1 and CTCF function by

Wapl (Haarhuis and Rowland 2017; Hara, et al. 2014), and opens up the possibility that this may be a general phenomenon underpinning cohesin regulation and genome recruitment.

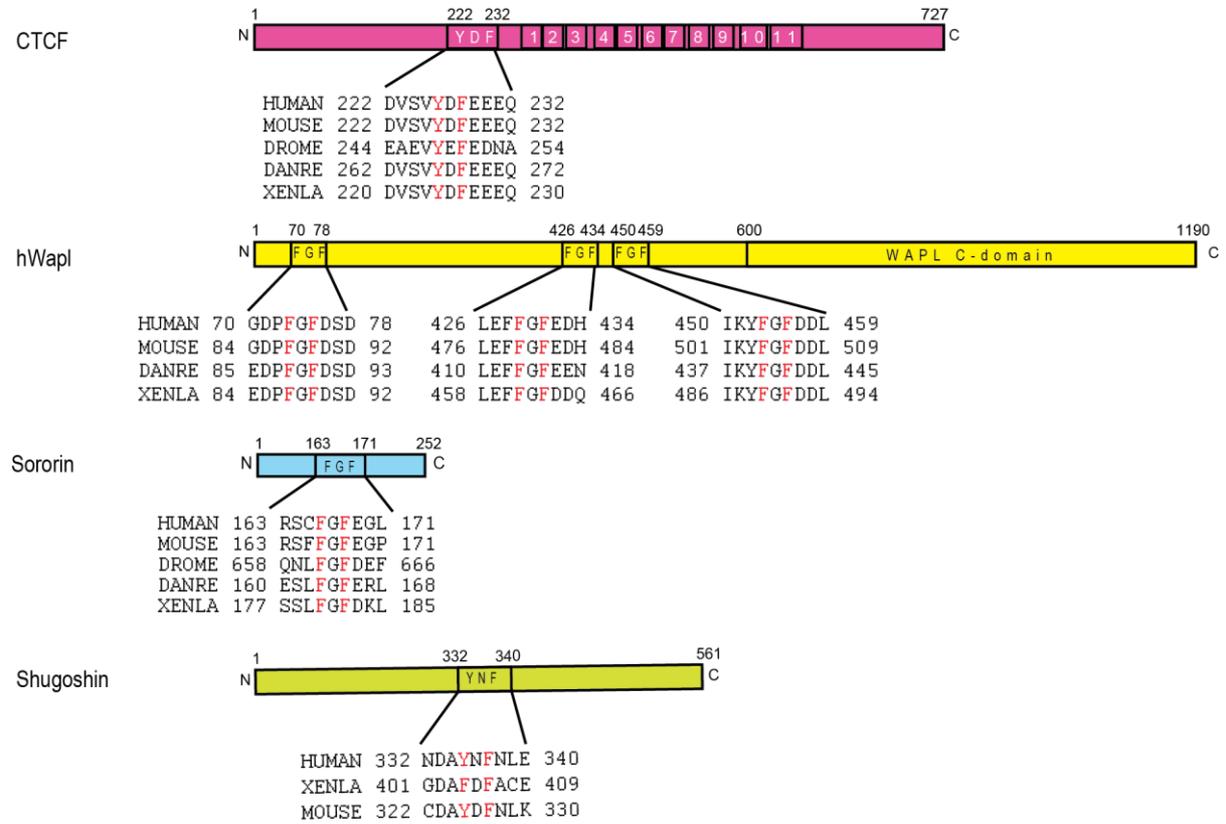
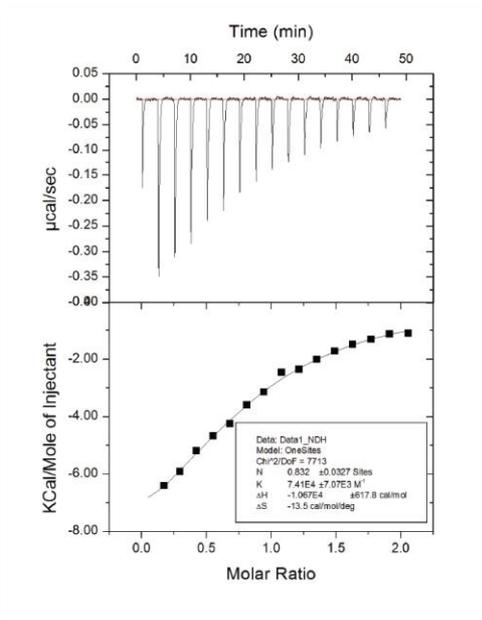
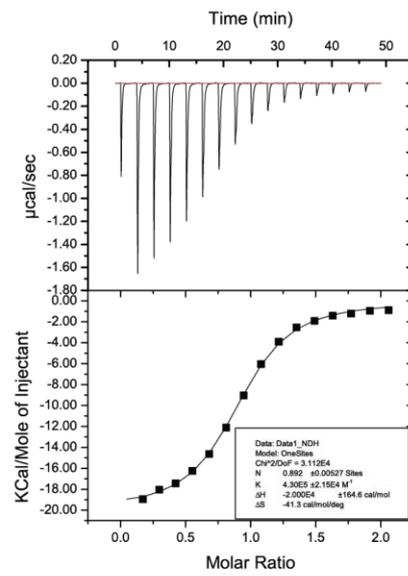


Figure 5.10 Y/FxF motifs are conserved among CTCF, Wapl, Sororin and Shugoshin.

a



Sgo1-shorter:SNDAYNFNLEE
Kd=13.5uM



pSgo1:SNDAYNFNLEEGVHLpTPFR
Kd=2.32uM

b

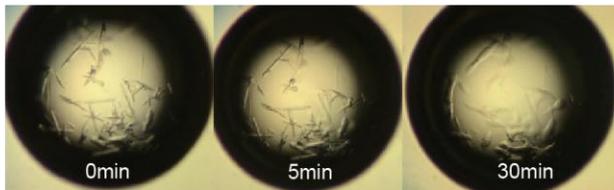


Figure 5.11 Sgo1 interacts with SA2-Scc1. (a) ITC showing binding isotherms of a truncated Sgo1 peptide (Sgo1-shorter) or of the extended phosphorylated Sgo1 peptide (pSgo1). While both substrates interact with SA2-Scc1, the phosphorylated peptide binds with approximately 6-fold higher binding affinity. (b) SA2-Scc1 crystals melted after soaking with pSgo1 peptide.

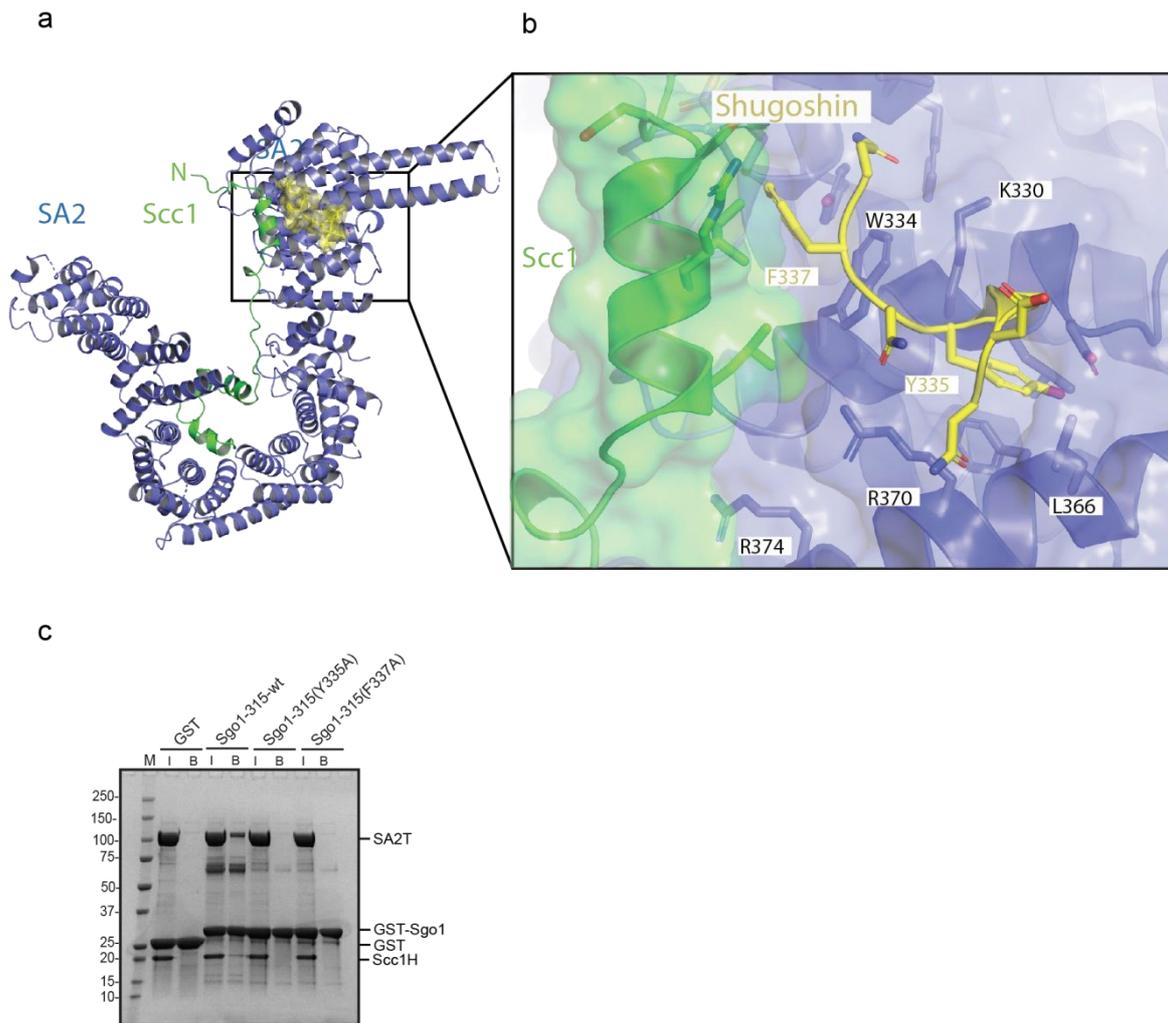


Figure 5.12 Structure of the SA2-Scc1 and Sgo1 ternary complex. (a) Model of the SA2-Scc1-Sgo1 complex colored in blue, green and yellow, respectively. (b) Detailed view of the binding interface with SA2 residues in blue, Scc1 in green and Sgo1 in yellow. (c) The indicated GST-Sgo1 variants were incubated with SA2-Scc1 and the input (I) and the bound (B) fraction analysed by SDS-PAGE. Mutation of Y335A or F337A of Sgo1 abolishes SA2-Scc1 binding.

| | SA2-Scc1-CTCF | SA2-Scc1-Sgo1 |
|--|---|----------------------------|
| Data collection | | |
| Space group | P2 ₁ 2 ₁ 2 ₁ | P2 ₁ |
| Cell dimensions | | |
| <i>a</i> , <i>b</i> , <i>c</i> (Å) | 79.0, 107.2, 176.4 | 78.8, 181.0, 111.3 |
| Resolution (Å) | 45.8–2.6 | 47.8–3.2 |
| No. reflections | 46759(4622) | 48453(4517) |
| <i>R</i> _{sym} or <i>R</i> _{merge} | 6.9 (175)* | 10.8 (120)* |
| <i>I</i> / σI | 12.0 (0.8)* | 8.1 (0.74)* |
| <i>CC</i> 1/2 | 0.99 (0.33) | 0.99 (0.45) |
| Completeness (%) | 99.6 (99.7)* | 97.9 (91.8)* |
| Redundancy | 4.4 (4.3)* | 2.7 (2.7)* |
| Refinement | | |
| Resolution (Å) | 45.8–2.6 | 47.8–3.2 |
| <i>R</i> _{work} / <i>R</i> _{free} | 0.24 / 0.28 | 0.31 / 0.34 |
| No. atoms | 16099 | 32295 |
| SA2 | 14793 | 29556 |
| Scc1 | 1249 | 2498 |
| Ligand | 157 _{CTCF} | 242 _{Shugoshin} |
| <i>B</i> -factors (mean; Å ²) | | |
| SA2 | 104.4 | 114.6 |
| Scc1 | 87.3 | 99.6 |
| Ligand | 119.6 _{CTCF} | 118.1 _{Shugoshin} |
| R.m.s deviations | | |
| Bond lengths (Å) | 0.005 | 0.003 |
| Bond angles (°) | 0.97 | 0.61 |

*Values in parentheses are for highest-resolution shell.

Table 5.1 X-ray data collection and refinement of SA2-Scc1-CTCF and SA2-Scc1-Sgo1 complex.

5.8 Bioinformatic prediction of cohesin binding factors

To investigate the prevalence of this linear interaction motif, we compiled an alignment from known cohesin partners (Figure 5.10) and derived a regular expression motif which we used to query the human and budding yeast proteomes (Krystkowiak and Davey 2017). From the

set of nuclear proteins arising from this search, we were able to identify known cohesin regulators as well as several novel potential binding factors. We generated peptide arrays containing these motifs and probed these with either wild-type SA2-Scc1 or a SA2 (F371A)-Scc1 mutant complex as a negative control (Figure 5.13). We observed robust signal for the CTCF peptide spanning 222-231. A CTCF Y226F mutant showed reduced binding indicating that the hydrogen bond between the hydroxyl group of Y226 and D326 of SA2 is required for higher affinity binding. Consistent with our pulldowns, the CTCF Y226A, F228A, and Y226A/F228A mutants all abolished binding to SA2-Scc1. The Wapl FGF motifs showed considerably weaker binding, while we could not detect binding for low binding affinity ligands such as Sgo1 (Figure 5.13, array position 12). Instead, robust binding was observed for Mcm3, a subunit of the replicative helicase (Figure 5.13, array position 13), SYCP3, a component of the synaptonemal complex (Figure 5.13, array position 17), ZGPAT, a transcriptional repressor (Figure 5.13, array position 23) and CENPU, a component of the inner kinetochore (Figure 5.13, array position 22). Thus, the conserved binding site on the SA2-Scc1 complex potentially effects cohesin recruitment in a number of different genome transactions.

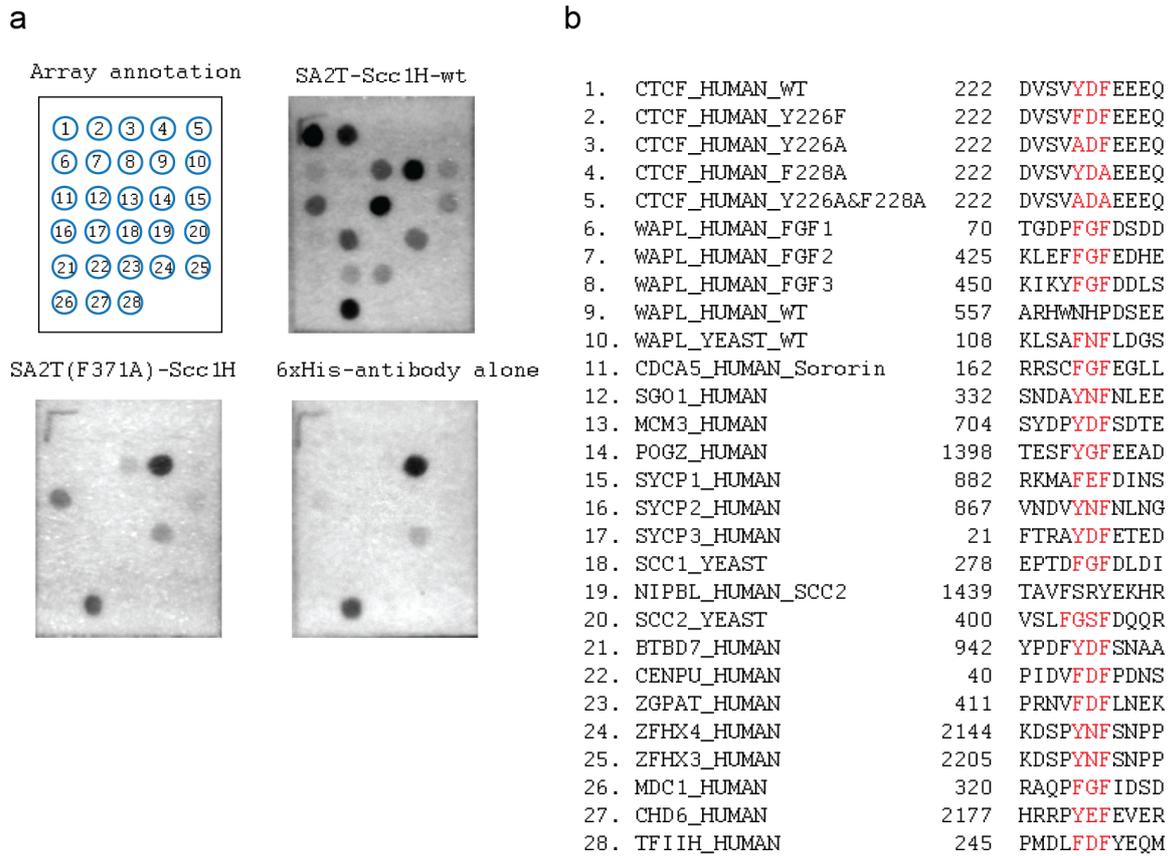


Figure 5.13 Novel SA2-Scc1 binding partners screened via peptide array. (a) Peptide array showing position of the spotted peptides (top left), SA2T-Scc1H binding (top right), binding of a SA2T (F371A)-Scc1H mutant complex (bottom left) and anti-6xHistidine antibody background control (bottom right). **(b)** Amino acid details of the peptide used. Predicted lead anchoring residues are colored in red.

Discussion_part 3

Discussion_partie 3

Dans cette section, nous avons déterminé deux complexes ternaires de CTCF ou de shugoshin liés au complexe SA2-Scc1 humain. Ces interacteurs partagent un motif de liaison commun qui fait partie de nombreux autres facteurs, tels que Wapl, MCM3, une hélicase impliquée dans l'initiation de la réplication du génome eucaryote, et plusieurs nouveaux facteurs qui n'ont pas été rapportés qu'ils interagissaient avec la cohésine. Par conséquent, mes travaux mettent en lumière un mécanisme apparemment répandu mais jusqu'à présent indéterminé sur la manière dont le CTCF et la cohésine se forment de manière convergente pour le bouclage d'ADN et sur la manière dont la cohésine est recrutée et participe à diverses transactions de la chromatine.

Our structural and previously published data (Hara, et al. 2014) support the notion that a number of cohesin ligands directly interact through a well-defined Y/FxF core motif with a conserved surface on the SA2-Scc1 complex. As the motif is present in a number of different nuclear proteins, binding through SA2-Scc1 is potentially a general mechanism for cohesin recruitment and regulation (Figure 5.14). Furthermore, since Scc3 and its orthologs, i.e. SA1, SA2 and SA3, are very well-conserved but function in different aspects of cohesin biology, e.g. either at chromosome arms or telomeres, or during meiosis, it will be interesting to further investigate how the ‘Y/FxF’ motif contributes to cohesin function either temporally or spatially through cell cycle.

Cohesin localisation to convergently oriented CTCF binding sites is a central event in TAD formation. An equilibrium between loop extrusion and release by cohesin at these sites gives rise to dynamic, tunable genome architecture. In this study, we show that CTCF binds cohesin via a short amino acid sequence in its N-terminus. Direct competition for this binding site by the Wapl likely explains how this cohesin release factor restricts TAD formation (Busslinger, et al. 2017; Haarhuis and Rowland 2017; Wutz, et al. 2017). Weakly binding motifs such as Shugoshin are tuned by addition of complementary negative phosphorylation charges around the motif core (Figure 5.12) (Hara, et al. 2014). Wapl Y/FxF motifs also contain conserved phosphorylation sites C-terminal to the Y/FxF motifs, suggesting that phosphorylation may be an important aspect in the regulation of competitive SA2-Scc1 ligand displacement.

The cohesin binding motif of CTCF is connected by a ~33 amino acid linker to the first of the 11 DNA binding zinc fingers suggesting that the interaction directly enables DNA recruitment of cohesin (Figure 5.9). In this thesis, I have shown that the concave side comprising the ‘nose’ of SA2 forms a non-specific DNA binding site in the budding yeast

ortholog Scc3 (Li, et al. 2018). We therefore propose that CTCF interacts with cohesin through the SA2-Scc1 interface to facilitate loop extrusion at CTCF binding sites in the genome. A model for how CTCF and SA2-Scc1 might co-associate on DNA is shown in Figure 5.9. The attachment of cohesin to the N-terminus of CTCF and the capacity of ATPase driven asymmetric DNA loop extrusion by SMC complexes, to ‘reel in’ DNA from one side (Ganji, et al. 2018), could explain how TADs are formed and why CTCF sites on the genome are oriented convergently (Fudenberg, et al. 2016). My work hence reveals a mechanism by which CTCF and cohesin interact, and it will be critical to assess what the effect is of disrupting this interaction in the cell.

Acknowledgement

First of all, I would like to thank EMBL and my supervisor, Dr. Daniel Panne, for giving me this opportunity to do my Ph.D. here in Grenoble, France. It has been very challenging but I do learn a lot, both on my scientific training and outlook of my life.

I would express my appreciation to Dr. Benjamin Rowland, Dr. Camilla Bjorkegren, Dr. Joanna Timmins and Dr. Ramesh Pillai for spending the time to evaluate my thesis. And also my thesis advisory committee (TAC) members, Dr. Stephen Cusack, Dr. Christian Haering, Dr. Carlo Petosa, thanks for all the valuable suggestions and kind help during my Ph.D.

As an excellent cohesin team, I would like to express my deepest appreciation to Dr. Kyle Muir who has invested a lot of effort in guiding the cohesin project in achieving the goal, sharing materials and ideas and it turns out extremely useful. We do spend time extensively working together for years.

Furthermore, I would also like to acknowledge with much appreciation on the crucial role of all members of the HTX lab for such effective support and MASSIF-1, especially Dr. Matthew Bowler, who helped test almost a thousand of crystals and this yields 5 essential structures in the cohesin field. It would not be possible for us to achieve this in such a short period without their professional and powerful support. I would also deeply thank Dr. Christian Haering and Dr. Benjamin Rowland for the professional collaboration, and also lab members, Jutta Metz from Haering at EMBL Heidelberg, Judith Haarhuis from Rowland lab at the Netherlands Cancer Institute.

I would also like to thank people who put a lot of effort in revising my thesis, Dr. Mariya Lukarska, Dr. Kyle Muir and Dr. Paul Sauer for the primary revision, Dr. Ziad Ibrahim, for the French abstracts and thank Dr. Daniel Panne for revising my thesis after.

I am grateful to all members of the Panne lab, Kyle Moo, Plauuu, Amede, Esther, Ziad and Silvia, thanks for all the support and discussion. And as well as Mathilde, Petra, Audrey, Sylvie, thanks for all the kind help.

在这里,我也要感谢我的父母及家人,感谢你们含辛茹苦的养育,感谢你们在我的人生道路上点亮的那盏灯.

--李岩

01/30/2019

Grenoble

Appendix

Table S1 Timeline of all cohesin and cohesin cycle-related structures.

| Complex | Species | Year | PDB code | Expression system |
|-----------------------|-------------------------|-------------|---|--------------------------|
| Smc1hd-CSccl | <i>S. Cerevisiae</i> | 2004 | 1W1W | <i>S. frugiperda</i> |
| Sgo1-PP2A | <i>H. Sapiens</i> | 2009 | 3FGA | <i>E.Coli</i> |
| Cohesin hinge | <i>m.musculus</i> | 2011 | 2WD5 | <i>E.Coli</i> |
| Wapl | <i>H. Sapiens</i> | 2013 | 4K6J | <i>E.Coli</i> |
| Wapl | <i>A. Gossypii</i> | 2013 | 3ZIL | <i>E.Coli</i> |
| SA2-Scc1 | <i>H. Sapiens</i> | 2014 | 4PJU | Trichoplusia ni |
| Smc3hd-NScc1 | <i>S. Cerevisiae</i> | 2014 | 4UX3 | <i>E.Coli</i> |
| Scc2-Scc4 | <i>A. Gossypii</i> | 2015 | 5C6G | <i>S. frugiperda</i> |
| Scc2-Scc4 | <i>S. Cerevisiae</i> | 2015 | 4XDN | <i>E.Coli</i> |
| Scc2C | <i>C.thermophilum</i> | 2016 | 5T8V | Trichoplusia ni |
| Scc2C | <i>A. Gossypii</i> | 2017 | 5ME3 | <i>S. frugiperda</i> |
| Scc2-Scc4-Ctf19 | <i>S. Cerevisiae</i> | 2017 | 5W94 | <i>E.Coli</i> |
| Pds5-Scc1 | <i>S. Cerevisiae</i> | 2016 | 5FRP | <i>E.Coli</i> |
| Pds5 | <i>L.thermotolerans</i> | 2016 | 5F0N | <i>Ecoli</i> |
| Pds5B-Wapl | <i>H. Sapiens</i> | 2016 | 5HDT | Trichoplusia ni |
| Esco1 | <i>H. Sapiens</i> | 2016 | 5T53 | <i>Ecoli</i> |
| Separase | <i>C.thermophilum</i> | 2016 | 5FBY | <i>E.Coli</i> |
| Separase-Securin | <i>S. Cerevisiae</i> | 2017 | 5U1T | Trichoplusia ni |
| CTCF-zinc fingers-DNA | <i>H. Sapiens</i> | 2017 | 5T0U, 5KKQ, 5T00, 5K5H, 5K5I, 5K5J, 5K5L, 5UND, | <i>E.Coli</i> |
| Scc3-Scc1-DNA | <i>S. Cerevisiae</i> | 2018 | 6H8Q | <i>E.Coli</i> |

Table S2 Constructs table.

| Construct | Organism | Residues | Vector | Tag | Variants |
|--------------------|--|--|----------------------|-------------------------------|--------------------------------------|
| cohesin | <i>S. Cerevisiae</i> | pETM 30-His-GST-Smc1-443-787 | | | Wt |
| hinge | | pACYC-Smc3-400-839 | | | |
| ctSmc1hd- CScc1 | <i>C.thermophilum</i> (<i>codon</i> <i>optimised</i>) <i>/S. Cerevisiae</i> | pRSF-His-ctSmc1-1-242-linker* 1058-1264; pACYC -yScc1-480-564 (no tag) | | | Wt |
| agSmc1hd | <i>A. Gossypii</i> (<i>codon</i> <i>optimised</i>) | 1-214- linker* 1011- 1222 | pRSF | Cter-His (uncleavabl e) | Wt |
| spSmc1hd | <i>S.Pombe</i> (<i>codon</i> <i>optimised</i>) | 1-212- linker* 1041- 1228 | pRSF | Cter- His(unclea vable) | Wt |
| Smc3hd- NScc1 | <i>S. Cerevisiae</i> (<i>SMC3codon</i> <i>optimised</i>) | pRSF-ORF1-Cter-His-Smc3-1-226- linker**-993-1230;ORF2-Scc1(1-112) | | | Wt |
| Pds5-fl | <i>S. Cerevisiae</i> | 1-1277 | pETM -11 | His | Wt |
| Pds5T(T5) | <i>S. Cerevisiae</i> | 1-701 | pETM-30/ pGEX -6p | His- GST/GST | Wt |
| Wapl-fl | <i>S. Cerevisiae</i> | 1-647 | pETM-11/ pETM -30 | His/His- GST | Wt |
| Wapl-C | <i>S. Cerevisiae</i> | 250-647 | pETM -30 | His-GST | Wt |
| EcoI-fl | <i>S. Cerevisiae</i> | Full length | pGEX -6T | GST | Wt |
| hWapl-N | <i>H. Sapiens</i> | 1-600 | pGEX -6p | GST | Wt,F73A/F75A/F42 9A/F431A/F453A/F |

| | | | | | |
|-----------|--|--|----------|------------------|--|
| | | | | | 455A |
| hWapl-423 | <i>H. Sapiens</i> | 423-463 | pGEX -6p | GST | Wt |
| Scc3T | <i>S. Cerevisiae</i> | 134-1064 | pETM -30 | His-GST | Wt,Patch1:K224E/K225E.Patch2:K322E/K330E.Patch3:K423E/K520E/K669E Heptamutant:K224E/K225E/K322E/K330E,K423E/K520E/K669E |
| Scc1K | <i>S. Cerevisiae</i> | 309-400 | pACYC | no | Wt,K363E/R364E(Scc1 double), L365M |
| Scc1-240 | <i>S. Cerevisiae</i> | 240-400 | pACYC | His(uncleavable) | Wt |
| Scc1-273 | <i>S. Cerevisiae</i> | 273-400 | pACYC | no | Wt |
| Scc1-170 | <i>S. Cerevisiae</i> | 170-400 | pACYC | His(uncleavable) | Wt |
| Scc1-388 | <i>S. Cerevisiae</i> | 309-388 | pACYC | no | Wt |
| Scc1-355 | <i>S. Cerevisiae</i> | 355-400 | pACYC | no | Wt |
| Clink-115 | <i>C.thermophilum</i> <i>/S. Cerevisiae</i> | pACYC-ORF1: ctSmc1-1-243, ORF2: ySccl-1-115-linker*-ctSmc1-1058-1264 pRSF-ORF1: Cter-His-ySmc3-1-260-linker**-971-1230, ORF2: ySccl-480-564 | | | Wt |
| Clink-120 | <i>C.thermophilum</i> <i>/S. Cerevisiae</i> | pACYC-ORF1: ctSmc1-1-243, ORF2: ySccl-1-120-linker*-ctSmc1-1058-1264 pRSF-ORF1: Cter-His-ySmc3-1-260-linker**-971-1230, ORF2: ySccl-480- | | | Wt |

| | | | | | |
|-------------------|--|---|----------|------------------|--------------------------------|
| | | 564 | | | |
| Clink-125 | <i>C.thermophilum</i> <i>/S. Cerevisiae</i> | pACYC:ORF1: ctSmc1-1-243, ORF2: yScc1-1-125-linker*-ctSmc1-1058-1264 | | | Wt |
| | | pRSF-ORF1: Cter-His-ySmc3-1-260-linker**-971-1230, ORF2: | | | |
| Clink-159 | <i>C.thermophilum</i> <i>/S. Cerevisiae</i> | pACYC-ORF1: ctSmc1-1-243, ORF2: yScc1-1-159-linker*-ctSmc1-1058-1264 | | | Wt |
| | | pRSF: ORF1: Cter-His-ySmc3-1-260-linker**-971-1230, ORF2: yScc1-480-564 | | | |
| Clink-159-T5 | <i>C.thermophilum</i> <i>/S. Cerevisiae</i> | pACYC-ORF1: ctSmc1-1-243, ORF2: yScc1-1-159-linker*-ctSmc1-1058-1264 | | | Wt |
| | | pRSF-ORF1: Cter-His-ySmc3-1-260-linker**-971-1230, ORF2: yScc1-480-564 | | | |
| | | pGEX -6p-GST-Pds5-1-701 | | | |
| Clink-159-Scc3T Δ | <i>C.thermophilum</i> <i>/S. Cerevisiae</i> | pRSF-ORF1:ctSmc1-1-243,ORF2: yScc1-1-159-linker*-ctSmc1-1058-1264 | | | Wt |
| | | pET-Duet: ORF1: Cter-His-ySmc3-1-260-linker**-971-1230, ORF2: yScc1-309-564(Δ400-480)(CSc1Δ) | | | |
| | | pETM-30-His-GST-yScc3(134-1064) | | | |
| SA2T | <i>H. Sapiens</i> (codon optimised) | 80-1060 | pGEX -6p | GST | Wt ,R298E, W334A, F367A, F371A |
| Scc1-281 | <i>H. Sapiens</i> | 281-420 | pACYC | no | Wt |
| Scc1H | <i>H. Sapiens</i> | 281-420 | pACYC | His(uncleavable) | Wt, I337A/L341A |

| | | | | | |
|-----------|-------------------|---------|----------|-----|------------------|
| Scc1-317 | <i>H. Sapiens</i> | 317-400 | pACYC | no | Wt |
| CTCF-ZF7 | <i>H. Sapiens</i> | 218-460 | pGEX -6p | GST | Wt |
| CTCF-ZF11 | <i>H. Sapiens</i> | 86-577 | pGEX -6p | GST | Wt |
| CTCF-N | <i>H. Sapiens</i> | 1-267 | pGEX -6p | GST | Wt |
| CTCF-N1 | <i>H. Sapiens</i> | 1-86 | pGEX -6p | GST | Wt |
| CTCF-N2 | <i>H. Sapiens</i> | 86-267 | pGEX -6p | GST | Wt |
| CTCF-N3 | <i>H. Sapiens</i> | 86-190 | pGEX -6p | GST | Wt |
| CTCF-N4 | <i>H. Sapiens</i> | 86-218 | pGEX -6p | GST | Wt |
| CTCF-N5 | <i>H. Sapiens</i> | 195-218 | pGEX -6p | GST | Wt |
| CTCF-N6 | <i>H. Sapiens</i> | 195-267 | pGEX -6p | GST | Wt, Y226A, F228A |
| CTCF-N7 | <i>H. Sapiens</i> | 127-190 | pGEX -6p | GST | Wt |
| CTCF-N8 | <i>H. Sapiens</i> | 127-218 | pGEX -6p | GST | Wt |
| CTCF-N9 | <i>H. Sapiens</i> | 218-254 | pGEX -6p | GST | Wt |
| CTCF-N10 | <i>H. Sapiens</i> | 218-267 | pGEX -6p | GST | Wt |
| CTCF-N11 | <i>H. Sapiens</i> | 127-227 | pGEX -6p | GST | Wt |
| CTCF-N12 | <i>H. Sapiens</i> | 127-235 | pGEX -6p | GST | Wt |
| CTCF-N13 | <i>H. Sapiens</i> | 249-267 | pGEX -6p | GST | Wt |
| CTCF-N14 | <i>H. Sapiens</i> | 241-267 | pGEX -6p | GST | Wt |
| CTCF-N15 | <i>H. Sapiens</i> | 235-267 | pGEX -6p | GST | Wt |
| CTCF-C | <i>H. Sapiens</i> | 579-727 | pGEX -6p | GST | Wt |
| CTCF-C1 | <i>H. Sapiens</i> | 579-633 | pGEX -6p | GST | Wt |
| CTCF-C2 | <i>H. Sapiens</i> | 633-727 | pGEX -6p | GST | Wt |
| Sgo1-315 | <i>H. Sapiens</i> | 315-360 | pGEX -6p | GST | Wt, Y335A, F337A |

*Smc1 linker sequence: SPGLEVLFGPRG

**Smc3 linker sequence: ESSKHPTSLVPRGS

References

- Adelfalk, C., et al.
2009 Cohesin SMC1beta protects telomeres in meiocytes. *J Cell Biol* 187(2):185-99.
- Alexandru, G., et al.
2001 Phosphorylation of the cohesin subunit Scc1 by Polo/Cdc5 kinase regulates sister chromatid separation in yeast. *Cell* 105(4):459-72.
- Anderson, D. E., et al.
2002 Condensin and cohesin display different arm conformations with characteristic hinge angles. *J Cell Biol* 156(3):419-24.
- Arumugam, P., et al.
2003 ATP hydrolysis is required for cohesin's association with chromosomes. *Curr Biol* 13(22):1941-53.
- Arumugam, P., et al.
2006 Cohesin's ATPase activity is stimulated by the C-terminal Winged-Helix domain of its kleisin subunit. *Curr Biol* 16(20):1998-2008.
- Beckouet, F., et al.
2016 Releasing Activity Disengages Cohesin's Smc3/Scc1 Interface in a Process Blocked by Acetylation. *Mol Cell* 61(4):563-574.
- Bell, A. C., A. G. West, and G. Felsenfeld
1999 The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98(3):387-96.
- Bickmore, W. A., and B. van Steensel
2013 Genome architecture: domain organization of interphase chromosomes. *Cell* 152(6):1270-84.
- Bisht, K. K., Z. Daniloski, and S. Smith
2013 SA1 binds directly to DNA through its unique AT-hook to promote sister chromatid cohesion at telomeres. *J Cell Sci* 126(Pt 15):3493-503.
- Boland, A., et al.
2017 Cryo-EM structure of a metazoan separase-securin complex at near-atomic resolution. *Nat Struct Mol Biol* 24(4):414-418.
- Bowler, M. W., et al.
2015 MASSIF-1: a beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules. *J Synchrotron Radiat* 22(6):1540-7.
- Busslinger, G. A., et al.
2017 Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* 544(7651):503-507.
- Chan, K. L., et al.
2013 Pds5 promotes and protects cohesin acetylation. *Proc Natl Acad Sci U S A* 110(32):13020-5.
- Chan, K. L., et al.
2012 Cohesin's DNA exit gate is distinct from its entrance gate and is regulated by acetylation. *Cell* 150(5):961-74.
- Chao, W. C., et al.
2015 Structural Studies Reveal the Functional Modularity of the Scc2-Scc4 Cohesin Loader. *Cell Rep* 12(5):719-25.
- Ciosk, R., et al.
2000 Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. *Mol Cell* 5(2):243-54.
- Clarke, A. S., et al.
2005 POLO kinase regulates the Drosophila centromere cohesion protein MEI-S332. *Dev Cell* 8(1):53-64.
- Clyne, R. K., et al.

- 2003 Polo-like kinase Cdc5 promotes chiasmata formation and cosegregation of sister centromeres at meiosis I. *Nat Cell Biol* 5(5):480-5.
- Cohen-Fix, O., et al.
1996 Anaphase initiation in *Saccharomyces cerevisiae* is controlled by the APC-dependent degradation of the anaphase inhibitor Pds1p. *Genes Dev* 10(24):3081-93.
- Cremer, T., et al.
2006 Chromosome territories--a functional nuclear landscape. *Curr Opin Cell Biol* 18(3):307-16.
- Cuddapah, S., et al.
2009 Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res* 19(1):24-32.
- Cuylen, S., and C. H. Haering
2011 Deciphering condensin action during chromosome segregation. *Trends Cell Biol* 21(9):552-9.
- Cuylen, S., J. Metz, and C. H. Haering
2011 Condensin structures chromosomal DNA through topological links. *Nat Struct Mol Biol* 18(8):894-901.
- de Wit, E., and W. de Laat
2012 A decade of 3C technologies: insights into nuclear organization. *Genes Dev* 26(1):11-24.
- de Wit, E., et al.
2015 CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell* 60(4):676-84.
- Dekker, J., et al.
2002 Capturing chromosome conformation. *Science* 295(5558):1306-11.
- Ding, D. Q., et al.
2006 Meiotic cohesins modulate chromosome compaction during meiotic prophase in fission yeast. *J Cell Biol* 174(4):499-508.
- Downen, J. M., et al.
2014 Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* 159(2):374-387.
- Dreier, M. R., M. E. Bekier, 2nd, and W. R. Taylor
2011 Regulation of sororin by Cdk1-mediated phosphorylation. *J Cell Sci* 124(Pt 17):2976-87.
- Eichinger, C. S., et al.
2013 Disengaging the Smc3/kleisin interface releases cohesin from *Drosophila* chromosomes during interphase and mitosis. *EMBO J* 32(5):656-65.
- Elbatsh, A. M. O., et al.
2016 Cohesin Releases DNA through Asymmetric ATPase-Driven Ring Opening. *Mol Cell* 61(4):575-588.
- Fernius, J., and A. L. Marston
2009 Establishment of cohesion at the pericentromere by the Ctf19 kinetochore subcomplex and the replication fork-associated factor, Csm3. *PLoS Genet* 5(9):e1000629.
- Filippova, G. N., et al.
2001 CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. *Nat Genet* 28(4):335-43.
- Forbes, S. A., et al.
2008 The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet* Chapter 10:Unit 10 11.
- Fortin, J. P., and K. D. Hansen
2015 Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *Genome Biol* 16:180.

- Fudenberg, G., et al.
2016 Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* 15(9):2038-49.
- Fujioka, Y., et al.
2002 Identification of a novel non-structural maintenance of chromosomes (SMC) component of the SMC5-SMC6 complex involved in DNA repair. *J Biol Chem* 277(24):21585-91.
- Gandhi, R., P. J. Gillespie, and T. Hirano
2006 Human Wapl is a cohesin-binding protein that promotes sister-chromatid resolution in mitotic prophase. *Curr Biol* 16(24):2406-17.
- Ganji, M., et al.
2018 Real-time imaging of DNA loop extrusion by condensin. *Science* 360(6384):102-105.
- Gligoris, T. G., et al.
2014 Closing the cohesin ring: structure and function of its Smc3-kleisin interface. *Science* 346(6212):963-7.
- Glynn, E. F., et al.
2004 Genome-wide mapping of the cohesin complex in the yeast *Saccharomyces cerevisiae*. *PLoS Biol* 2(9):E259.
- Golsteyn, R. M., et al.
1995 Cell cycle regulation of the activity and subcellular localization of Plk1, a human protein kinase implicated in mitotic spindle function. *J Cell Biol* 129(6):1617-28.
- Goto, H., et al.
2006 Complex formation of Plk1 and INCENP required for metaphase-anaphase transition. *Nat Cell Biol* 8(2):180-7.
- Gruber, S., et al.
2006 Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. *Cell* 127(3):523-37.
- Haarhuis, J. H., A. M. Elbatsh, and B. D. Rowland
2014 Cohesin and its regulation: on the logic of X-shaped chromosomes. *Dev Cell* 31(1):7-18.
- Haarhuis, J. H. I., et al.
2017 The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* 169(4):693-707 e14.
- Haarhuis, J. H., and B. D. Rowland
2017 Cohesin: building loops, but not compartments. *EMBO J* 36(24):3549-3551.
- Haering, C. H., et al.
2008 The cohesin ring concatenates sister DNA molecules. *Nature* 454(7202):297-301.
- Haering, C. H., et al.
2002 Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol Cell* 9(4):773-88.
- Haering, C. H., et al.
2004 Structure and stability of cohesin's Smc1-kleisin interaction. *Mol Cell* 15(6):951-64.
- Hakim, O., and T. Misteli
2012 SnapShot: Chromosome conformation capture. *Cell* 148(5):1068 e1-2.
- Hansen, Anders S, et al.
2018 An RNA-binding region regulates CTCF clustering and chromatin looping. [bioRxiv:495432](https://doi.org/10.1101/495432).
- Hara, K., et al.
2014 Structure of cohesin subcomplex pinpoints direct shugoshin-Wapl antagonism in centromeric cohesion. *Nat Struct Mol Biol* 21(10):864-70.
- Hark, A. T., et al.

- 2000 CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* 405(6785):486-9.
- Hartman, T., et al.
2000 Pds5p is an essential chromosomal protein required for both sister chromatid cohesion and condensation in *Saccharomyces cerevisiae*. *J Cell Biol* 151(3):613-26.
- Hashimoto, H., et al.
2017 Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA. *Mol Cell* 66(5):711-720 e3.
- Hassler, M., I. A. Shaltiel, and C. H. Haering
2018 Towards a Unified Model of SMC Complex Function. *Curr Biol* 28(21):R1266-R1281.
- Hauf, S., et al.
2005 Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2. *PLoS Biol* 3(3):e69.
- Heidinger-Pauli, J. M., et al.
2010 Systematic reduction of cohesin differentially affects chromosome segregation, condensation, and DNA repair. *Curr Biol* 20(10):957-63.
- Hinshaw, S. M., et al.
2015 Structural evidence for Scc4-dependent localization of cohesin loading. *Elife* 4:e06057.
- Hirano, T.
2016 Condensin-Based Chromosome Organization from Bacteria to Vertebrates. *Cell* 164(5):847-57.
- Hirano, T., R. Kobayashi, and M. Hirano
1997 Condensins, chromosome condensation protein complexes containing XCAP-C, XCAP-E and a *Xenopus* homolog of the *Drosophila* Barren protein. *Cell* 89(4):511-21.
- Hoque, M. T., and F. Ishikawa
2002 Cohesin defects lead to premature sister chromatid separation, kinetochore dysfunction, and spindle-assembly checkpoint activation. *J Biol Chem* 277(44):42306-14.
- Hornig, N. C., and F. Uhlmann
2004 Preferential cleavage of chromatin-bound cohesin after targeted phosphorylation by Polo-like kinase. *EMBO J* 23(15):3144-53.
- Hou, C., R. Dale, and A. Dean
2010 Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc Natl Acad Sci U S A* 107(8):3651-6.
- Hu, B., et al.
2011 ATP hydrolysis is required for relocating cohesin from sites occupied by its Scc2/4 loading complex. *Curr Biol* 21(1):12-24.
- Huis in 't Veld, P. J., et al.
2014 Characterization of a DNA exit gate in the human cohesin ring. *Science* 346(6212):968-72.
- Ishihara, K., M. Oshimura, and M. Nakao
2006 CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol Cell* 23(5):733-42.
- Ivanov, D., et al.
2002 Eco1 is a novel acetyltransferase that can acetylate proteins involved in cohesion. *Curr Biol* 12(4):323-8.
- Ivanov, M. P., et al.
2018 The replicative helicase MCM recruits cohesin acetyltransferase ESCO2 to mediate centromeric sister chromatid cohesion. *EMBO J* 37(15).
- Kabsch, W.
2010 Xds. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125-32.

- Kagey, M. H., et al.
2010 Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467(7314):430-5.
- Kanno, T., D. G. Berta, and C. Sjogren
2015 The Smc5/6 Complex Is an ATP-Dependent Intermolecular DNA Linker. *Cell Rep* 12(9):1471-82.
- Kenna, M. A., and R. V. Skibbens
2003 Mechanical link between cohesion establishment and DNA replication: Ctf7p/Eco1p, a cohesion establishment factor, associates with three different replication factor C complexes. *Mol Cell Biol* 23(8):2999-3007.
- Kerrebrock, A. W., et al.
1995 Mei-S332, a Drosophila protein required for sister-chromatid cohesion, can localize to meiotic centromere regions. *Cell* 83(2):247-56.
- Kikuchi, S., et al.
2016 Crystal structure of the cohesin loader Scc2 and insight into cohesinopathy. *Proc Natl Acad Sci U S A* 113(44):12444-12449.
- Kitajima, T. S., et al.
2005 Human Bub1 defines the persistent cohesion site along the mitotic chromosome by affecting Shugoshin localization. *Curr Biol* 15(4):353-9.
- Kitajima, T. S., S. A. Kawashima, and Y. Watanabe
2004 The conserved kinetochore protein shugoshin protects centromeric cohesion during meiosis. *Nature* 427(6974):510-7.
- Kitajima, T. S., et al.
2006 Shugoshin collaborates with protein phosphatase 2A to protect cohesin. *Nature* 441(7089):46-52.
- Klenova, E. M., et al.
1993 CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol Cell Biol* 13(12):7612-24.
- Kogut, I., et al.
2009 The Scc2/Scc4 cohesin loader determines the distribution of cohesin on budding yeast chromosomes. *Genes Dev* 23(19):2345-57.
- Kojic, A., et al.
2018 Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. *Nat Struct Mol Biol* 25(6):496-504.
- Krystkowiak, I., and N. E. Davey
2017 SLIMSearch: a framework for proteome-wide discovery and annotation of functional modules in intrinsically disordered regions. *Nucleic Acids Res* 45(W1):W464-W469.
- Kschonsak, M., et al.
2017 Structural Basis for a Safety-Belt Mechanism That Anchors Condensin to Chromosomes. *Cell* 171(3):588-600 e24.
- Kueng, S., et al.
2006 Wapl controls the dynamic association of cohesin with chromatin. *Cell* 127(5):955-67.
- Kurkcuoglu, O., and P. A. Bates
2010 Mechanism of cohesin loading onto chromosomes: a conformational dynamics study. *Biophys J* 99(4):1212-20.
- Lee, B. H., and A. Amon
2003 Role of Polo-like kinase CDC5 in programming meiosis I chromosome segregation. *Science* 300(5618):482-6.
- Lengronne, A., et al.

- 2004 Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* 430(6999):573-8.
- Li, Y., et al.
2018 Structural basis for Scc3-dependent cohesin recruitment to chromatin. *Elife* 7.
- Lieberman-Aiden, E., et al.
2009 Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326(5950):289-93.
- Lin, Z., X. Luo, and H. Yu
2016 Structural basis of cohesin cleavage by separase. *Nature* 532(7597):131-4.
- Liu, H., S. Rankin, and H. Yu
2013 Phosphorylation-enabled binding of SGO1-PP2A to cohesin protects sororin and centromeric cohesion during mitosis. *Nat Cell Biol* 15(1):40-9.
- Liu, Y., et al.
2016 ATP-dependent DNA binding, unwinding, and resection by the Mre11/Rad50 complex. *EMBO J* 35(7):743-58.
- Lobanenkov, V. V., et al.
1990 A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene* 5(12):1743-53.
- Losada, A., M. Hirano, and T. Hirano
1998 Identification of *Xenopus* SMC protein complexes required for sister chromatid cohesion. *Genes Dev* 12(13):1986-97.
- Losada, A., T. Yokochi, and T. Hirano
2005 Functional contribution of Pds5 to cohesin-mediated cohesion in human cells and *Xenopus* egg extracts. *J Cell Sci* 118(Pt 10):2133-41.
- Luo, S., and L. Tong
2017 Molecular mechanism for the regulation of yeast separase by securin. *Nature* 542(7640):255-259.
- Lyons, N. A., and D. O. Morgan
2011 Cdk1-dependent destruction of Eco1 prevents cohesion establishment after S phase. *Mol Cell* 42(3):378-89.
- Marston, A. L., et al.
2004 A genome-wide screen identifies genes required for centromeric cohesion. *Science* 303(5662):1367-70.
- McCoy, A. J., et al.
2007 Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658-674.
- McGrew, J. T., et al.
1992 Requirement for ESP1 in the nuclear division of *Saccharomyces cerevisiae*. *Mol Biol Cell* 3(12):1443-54.
- Merkenschlager, M., and E. P. Nora
2016 CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annu Rev Genomics Hum Genet* 17:17-43.
- Mizuguchi, T., et al.
2014 Cohesin-dependent globules and heterochromatin shape 3D genome architecture in *S. pombe*. *Nature* 516(7531):432-435.
- Muir, K. W., et al.
2016 Structure of the Pds5-Scc1 Complex and Implications for Cohesin Function. *Cell Rep* 14(9):2116-2126.
- Muir, Kyle
2016 Biochemical and biophysical characterisation of the Cohesin complex, Université Grenoble Alpes.

- Murayama, Y., et al.
 2018 Establishment of DNA-DNA Interactions by the Cohesin Ring. *Cell* 172(3):465-477 e15.
- Murayama, Y., and F. Uhlmann
 2014 Biochemical reconstitution of topological DNA binding by the cohesin ring. *Nature* 505(7483):367-71.
- 2015 DNA Entry into and Exit out of the Cohesin Ring by an Interlocking Gate Mechanism. *Cell* 163(7):1628-40.
- Nakajima, M., et al.
 2007 The complete removal of cohesin from chromosome arms depends on separase. *J Cell Sci* 120(Pt 23):4188-96.
- Nasmyth, K.
 2011 Cohesin: a catenase with separate entry and exit gates? *Nat Cell Biol* 13(10):1170-7.
- Nasmyth, K., and C. H. Haering
 2009 Cohesin: its roles and mechanisms. *Annu Rev Genet* 43:525-58.
- Nishiyama, T., et al.
 2010 Sororin mediates sister chromatid cohesion by antagonizing Wapl. *Cell* 143(5):737-49.
- Ocampo-Hafalla, M. T., and F. Uhlmann
 2011 Cohesin loading and sliding. *J Cell Sci* 124(Pt 5):685-91.
- Onn, I., et al.
 2008 Sister chromatid cohesion: a simple concept with a complex reality. *Annu Rev Cell Dev Biol* 24:105-29.
- Orgil, O., et al.
 2015 A conserved domain in the scc3 subunit of cohesin mediates the interaction with both mcd1 and the cohesin loader complex. *PLoS Genet* 11(3):e1005036.
- Ou, H. D., et al.
 2017 ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* 357(6349).
- Ouyang, Z., et al.
 2013 Structure of the human cohesin inhibitor Wapl. *Proc Natl Acad Sci U S A* 110(28):11355-60.
- Ouyang, Z., et al.
 2016 Structural Basis and IP6 Requirement for Pds5-Dependent Cohesin Dynamics. *Mol Cell* 62(2):248-259.
- Parelho, V., et al.
 2008 Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132(3):422-33.
- Peters, J. M., A. Tedeschi, and J. Schmitz
 2008 The cohesin complex and its roles in chromosome biology. *Genes Dev* 22(22):3089-114.
- Pezzi, N., et al.
 2000 STAG3, a novel gene encoding a protein involved in meiotic chromosome pairing and location of STAG3-related genes flanking the Williams-Beuren syndrome deletion. *FASEB J* 14(3):581-92.
- Phillips-Cremins, J. E., and V. G. Corces
 2013 Chromatin insulators: linking genome organization to cellular function. *Mol Cell* 50(4):461-74.
- Pouwels, J., et al.

- 2007 Shugoshin 1 plays a central role in kinetochore assembly and is required for kinetochore targeting of Plk1. *Cell Cycle* 6(13):1579-85.
- Ramos, C., et al.
2010 FGF-1 reverts epithelial-mesenchymal transition induced by TGF- β 1 through MAPK/ERK kinase pathway. *Am J Physiol Lung Cell Mol Physiol* 299(2):L222-31.
- Randall, C. L., M. E. Burkard, and P. V. Jallepalli
2007 Polo kinase and cytokinesis initiation in mammalian cells: harnessing the awesome power of chemical genetics. *Cell Cycle* 6(14):1713-7.
- Rao, S. S. P., et al.
2017 Cohesin Loss Eliminates All Loop Domains. *Cell* 171(2):305-320 e24.
- Reddy, K. L., et al.
2008 Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* 452(7184):243-7.
- Remeseiro, S., et al.
2012 Cohesin-SA1 deficiency drives aneuploidy and tumorigenesis in mice due to impaired replication of telomeres. *EMBO J* 31(9):2076-89.
- Revenkova, E., et al.
2004 Cohesin SMC1 beta is required for meiotic chromosome dynamics, sister chromatid cohesion and DNA recombination. *Nat Cell Biol* 6(6):555-62.
- Riedel, C. G., et al.
2006 Protein phosphatase 2A protects centromeric sister chromatid cohesion during meiosis I. *Nature* 441(7089):53-61.
- Ripoche, M. A., et al.
1997 Deletion of the H19 transcription unit reveals the existence of a putative imprinting control element. *Genes Dev* 11(12):1596-604.
- Rivera-Colon, Y., et al.
2016 Molecular Basis for Cohesin Acetylation by Establishment of Sister Chromatid Cohesion N-Acetyltransferase ESCO1. *J Biol Chem* 291(51):26468-26477.
- Roig, M. B., et al.
2014 Structure and function of cohesin's Scc3/SA regulatory subunit. *FEBS Lett* 588(20):3692-702.
- Rolef Ben-Shahar, T., et al.
2008 Eco1-dependent cohesin acetylation during establishment of sister chromatid cohesion. *Science* 321(5888):563-6.
- Rowland, B. D., et al.
2009 Building sister chromatid cohesion: smc3 acetylation counteracts an antiestablishment activity. *Mol Cell* 33(6):763-74.
- Rubio, E. D., et al.
2008 CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A* 105(24):8309-14.
- Sakai, A., et al.
2003 Condensin but not cohesin SMC heterodimer induces DNA reannealing through protein-protein assembly. *EMBO J* 22(11):2764-75.
- Saldana-Meyer, Ricardo, et al.
2019 RNA interactions with CTCF are essential for its proper function. [bioRxiv:530014](https://doi.org/10.1101/530014).
- Sanborn, A. L., et al.
2015 Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* 112(47):E6456-65.
- Schmitz, J., et al.
2007 Sororin is required for stable binding of cohesin to chromatin and for sister chromatid cohesion in interphase. *Curr Biol* 17(7):630-6.

- Schwarzer, W., et al.
2017 Two independent modes of chromatin organization revealed by cohesin removal. *Nature* 551(7678):51-56.
- Seitan, V. C., et al.
2013 Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Res* 23(12):2066-77.
- Shintomi, K., and T. Hirano
2009 Releasing cohesin from chromosome arms in early mitosis: opposing actions of Wapl-Pds5 and Sgo1. *Genes Dev* 23(18):2224-36.
- Sievers, F., and D. G. Higgins
2014 Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* 1079:105-16.
- Simonis, M., et al.
2006 Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* 38(11):1348-54.
- Skibbens, R. V., et al.
1999 Ctf7p is essential for sister chromatid cohesion and links mitotic chromosome structure to the DNA replication machinery. *Genes Dev* 13(3):307-19.
- Sofueva, S., et al.
2013 Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J* 32(24):3119-29.
- Srinivasan, M., et al.
2018 The Cohesin Ring Uses Its Hinge to Organize DNA Using Non-topological as well as Topological Mechanisms. *Cell* 173(6):1508-1519 e18.
- Stemmann, O., et al.
2001 Dual inhibition of sister chromatid separation at metaphase. *Cell* 107(6):715-26.
- Stigler, J., et al.
2016 Single-Molecule Imaging Reveals a Collapsed Conformational State for DNA-Bound Cohesin. *Cell Rep* 15(5):988-998.
- Strubbe, G., et al.
2011 Polycomb purification by in vivo biotinylation tagging reveals cohesin and Trithorax group proteins as interaction partners. *Proc Natl Acad Sci U S A* 108(14):5572-7.
- Studier, F. W.
2005 Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41(1):207-34.
- Sumara, I., et al.
2000 Characterization of vertebrate cohesin complexes and their regulation in prophase. *J Cell Biol* 151(4):749-62.
- Sumara, I., et al.
2002 The dissociation of cohesin from chromosomes in prophase is regulated by Polo-like kinase. *Mol Cell* 9(3):515-25.
- Sutani, T., et al.
2009 Budding yeast Wpl1(Rad61)-Pds5 complex counteracts sister chromatid cohesion-establishing reaction. *Curr Biol* 19(6):492-7.
- Svensson, O., et al.
2018 Multi-position data collection and dynamic beam sizing: recent improvements to the automatic data-collection algorithms on MASSIF-1. *Acta Crystallogr D Struct Biol* 74(Pt 5):433-440.
- Svensson, O., et al.
2015 Fully automatic characterization and data collection from crystals of biological macromolecules. *Acta Crystallogr D Biol Crystallogr* 71(Pt 8):1757-67.

- Tang, Z., et al.
2006 PP2A is required for centromeric localization of Sgo1 and proper chromosome segregation. *Dev Cell* 10(5):575-85.
- Tang, Z., et al.
2004 Human Bub1 protects centromeric sister-chromatid cohesion through Shugoshin during mitosis. *Proc Natl Acad Sci U S A* 101(52):18012-7.
- Tanimoto, K., et al.
2003 Human beta-globin locus control region HS5 contains CTCF- and developmental stage-dependent enhancer-blocking activity in erythroid cells. *Mol Cell Biol* 23(24):8946-52.
- Tark-Dame, M., et al.
2014 Depletion of the chromatin looping proteins CTCF and cohesin causes chromatin compaction: insight into chromatin folding by polymer modelling. *PLoS Comput Biol* 10(10):e1003877.
- Toth, A., et al.
1999 Yeast cohesin complex requires a conserved protein, Eco1p(Ctf7), to establish cohesion between sister chromatids during DNA replication. *Genes Dev* 13(3):320-33.
- Tremblay, K. D., K. L. Duran, and M. S. Bartolomei
1997 A 5' 2-kilobase-pair region of the imprinted mouse H19 gene exhibits exclusive paternal methylation throughout development. *Mol Cell Biol* 17(8):4322-9.
- Uhlmann, F.
2016 SMC complexes: from DNA to chromosomes. *Nat Rev Mol Cell Biol* 17(7):399-412.
- Uhlmann, F., F. Lottspeich, and K. Nasmyth
1999 Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1. *Nature* 400(6739):37-42.
- Unal, E., et al.
2008 A molecular determinant for the establishment of sister chromatid cohesion. *Science* 321(5888):566-9.
- Uzawa, S., et al.
1990 The fission yeast cut1+ gene regulates spindle pole body duplication and has homology to the budding yeast ESP1 gene. *Cell* 62(5):913-25.
- Van Duyne, G. D., et al.
1993 Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol* 229(1):105-24.
- van Heemst, D., et al.
1999 Spo76p is a conserved chromosome morphogenesis protein that links the mitotic and meiotic programs. *Cell* 98(2):261-71.
- van Vugt, M. A., et al.
2004 Polo-like kinase-1 is required for bipolar spindle formation but is dispensable for anaphase promoting complex/Cdc20 activation and initiation of cytokinesis. *J Biol Chem* 279(35):36841-54.
- Verni, F., et al.
2000 Genetic and molecular analysis of wings apart-like (wapl), a gene controlling heterochromatin organization in *Drosophila melanogaster*. *Genetics* 154(4):1693-710.
- Wallace, J. A., and G. Felsenfeld
2007 We gather together: insulators and genome organization. *Curr Opin Genet Dev* 17(5):400-7.
- Watanabe, Y., and T. S. Kitajima
2005 Shugoshin protects cohesin complexes at centromeres. *Philos Trans R Soc Lond B Biol Sci* 360(1455):515-21, discussion 521.
- Weitzer, S., C. Lehane, and F. Uhlmann

- 2003 A model for ATP hydrolysis-dependent binding of cohesin to DNA. *Curr Biol* 13(22):1930-40.
- Wendt, K. S., et al.
2008 Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451(7180):796-801.
- Weth, O., and R. Renkawitz
2011 CTCF function is modulated by neighboring DNA binding factors. *Biochem Cell Biol* 89(5):459-68.
- White, J. K., et al.
2013 Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell* 154(2):452-64.
- Wilhelm, L., et al.
2015 SMC condensin entraps chromosomal DNA by an ATP hydrolysis dependent loading mechanism in *Bacillus subtilis*. *Elife* 4.
- Winn, M. D., et al.
2011 Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):235-42.
- Wu, F. M., J. V. Nguyen, and S. Rankin
2011 A conserved motif at the C terminus of sororin is required for sister chromatid cohesion. *J Biol Chem* 286(5):3579-86.
- Wutz, G., et al.
2017 Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J* 36(24):3573-3599.
- Xiao, T., J. Wallace, and G. Felsenfeld
2011 Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol Cell Biol* 31(11):2174-83.
- Xu, H., et al.
2010 Rad21-cohesin haploinsufficiency impedes DNA repair and enhances gastrointestinal radiosensitivity in mice. *PLoS One* 5(8):e12112.
- Xu, Z., et al.
2009 Structure and function of the PP2A-shugoshin interaction. *Mol Cell* 35(4):426-41.
- Yamagishi, Y., et al.
2010 Two histone marks establish the inner centromere and chromosome bi-orientation. *Science* 330(6001):239-43.
- Yamamoto, A., V. Guacci, and D. Koshland
1996 Pds1p, an inhibitor of anaphase in budding yeast, plays a critical role in the APC and checkpoint pathway(s). *J Cell Biol* 133(1):99-110.
- Yusufzai, T. M., et al.
2004 CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol Cell* 13(2):291-8.
- Zhang, J., et al.
2008 Acetylation of Smc3 by Eco1 is required for S phase sister chromatid cohesion in both human and yeast. *Mol Cell* 31(1):143-51.
- Zhang, N., et al.
2011 Interaction of Sororin protein with polo-like kinase 1 mediates resolution of chromosomal arm cohesion. *J Biol Chem* 286(48):41826-37.
- Zuin, J., et al.
2014 Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A* 111(3):996-1001.