Université

de Strasbourg

igbmc
institut de génétique et de
biologie moléculaire et cellulaire

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

Institut de Génétique et de Biologie Moléculaire et Cellulaire

Université de Strasbourg, UMR 7104, U 964

THÈSE présentée par : **Marina Jacinta VITORIA GOMES**

Soutenue le : 06 octobre 2021

Distinct conformational dynamics of human Cohesin SMC1A and SMC3 ATPase domains

Pour obtenir le grade de : Docteur de l'université de Strasbourg

Discipline/ Spécialité : Sciences de la vie et de la Santé

THÈSE DIRIGÉE PAR :

Dr. ROMIER Christophe   DR1 CNRS, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch-Graffenstaden – France

RAPPORTEURS :

Pr. Dr. HÄRING Christian Faculty of Medicine, Würzburg – Allemagne

Dr. TIMMINS Joanna DR2 CNRS, Institut de Biologie Structurale, Grenoble – France

EXAMINATEUR INTERNE :

Dr. LEBARS Isabelle CR CNRS, Institut de Biologie Moléculaire et Cellulaire, Strasbourg – France

EXAMINATEUR EXTERNE :

Dr. BECKOUËT Frédéric CR CNRS, Laboratoire de Biologie Moléculaire Eucaryote (LBME), Toulouse – France

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

A	Alanine
Å	Angström
ABC	ATP-binding cassette
ADP	Adenosine 5'-diphosphate
ADP-AlF₃	ADP-aluminum fluoride
ADP-BeF₃	ADP-beryllium fluoride
ADP-VO₄	ADP-orthovanadate
A. thaliana	*Arabidopsis thaliana*
ATP	Adenosine 5'-triphosphate
ATPγS	Adenosine 5'-O-(3-thio)triphosphate
BMOE	Bismaleimidoethane
C	Cysteine
°C	Degree celcius
CC	Coiled-coil
CCsh	Shortened coiled-coil
C. elegans	*Caenorhabditis elegans*
CdLS	Cornelia de Lange Syndrome
CID	Chromatin Interacting Domain
Cryo-EM	Cryo-electron microscopy
C-ter, C-terminal	Carboxy-terminus
D	Aspartate
C. thermophilum	*Chaetomium thermophilum*
D. melanogaster	*Drosophila melanogaster*
DNA	Deoxyribonucleic acid
DSB	Double-strand break
DTT	Dithiothreitol
DMSO	Dimethyl sulfoxide
dsDNA	Double-stranded DNA

E	Glutaname
EM	Electron microscopy
EQ-mutant	Glutamate to glutamine mutant
F	Phenylalanine
G	Glycine
G0	Gap 0
G1	Gap 1
G2	Gap 2
GD	Globular domain
H	Histidine
HAWK	HEAT-Repeat proteins Associated With Kleisins
HCl	Hydrochloric acid
HD	Head
HEAT	Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A), kinase TOR1
HR	Homologous recombination
H. sapiens	*Homo sapiens*
I	Isoleucine
ITC	Isothermal Titration Calorimetry
IPTG	Isopropyl β-D-1-thiogalactopyranoside
J	Joint
K	Lysine
KITE	Kleisin-Interacting Tandem winged-helix Element
kDa	Kilodalton
L	Leucine
LE	Loop extrusion
M	Methionine
mAU	Milli Absorbance Unit
Mbp	Megabase pair
MES	2-(N-morpholino)ethanesulfonic acid
MESG	2-amino-6-mercapto-7-methylpurine riboside

Mg^{2+}	Magnesium ion
MgCl$_2$	Magnesium chloride
ml	Millilitre
mM	Milimolar
MMT	DL-Malic acid, MES monohydrate, Tris
N	Asparagine
NaCl	Sodium chloride
NBD	Nucleotide-binding domain
NHEJ	Non-homologous end joining
nm	Nanometre
N-ter, N-terminal	Amino-terminus
P	Proline
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
P. furiosus	*Pyrococcus furiosus*
P$_i$	Inorganic phosphate
PNP	Purine nucleoside phosphorylase
Q	Glutamine
R	Arginine
rpm	Revolution per minute
RNA	Ribonucleic acid
S	Serine
S. cerevisiae	*Saccharomyces cerevisiae*
SMC	Structural Maintenance of Chromosomes
S-phase	Synthesis-phase
S. pombe	*Schizosaccharomyces pombe*
T	Threonine
TAD	Topologically Associating Domain
TCEP	Tris(2-carboxyethyl)phosphine
TMD	Transmembrane domain

Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
U	Enzyme unit
V	Valine
v/v	Volume by volume
W	Tryptophane
WHD	Winged-Helix Domain
WT	Wild type
w/v	Weight by volume
X. laevis	*Xenopus laevis*
Y	Tyrosine
Δ	Delta, deletion

SUMMARY OF THE THESIS WORK

Introduction

The DNA inside each eukaryotic cell nucleus is compacted into chromatin, a dynamic nucleic acid-protein structure that regulates the access to genetic information. Numerous epigenetic factors associate to chromatin and modulate its structure, having therefore a major impact on the control and on the regulation of nuclear processes. The cohesin complex is part of the family of Structural Maintenance of Chromososmes (SMC) protein complexes that associate with chromatin. SMC complexes contain a homodimer of SMC proteins in prokaryotes, and a heterodimer in eukaryotes, which have essential roles in the three-dimensional genome organization and in the maintenance of its stability and integrity.

Cohesin is involved in vital genome regulation processes, notably by playing a key role in sister chromatid cohesion throughout the cell cycle, in chromosome segregation, in DNA double-strand break repair, in chromatin structure organization, and in transcriptional regulation via the establishment of chromatin domains, known as Topological Associated Domains (TADs). Cohesin functions depend on its ability to establish topological links between two intra- or inter-molecular DNA segments, by bringing and holding them in close spatial proximity.

The core of the eukaryotic cohesin multiprotein complex is composed of a heterodimer of two structural maintenance of chromosomes (SMC) proteins, SMC1 and SMC3. SMC1 and SMC3 are each composed of N- and C-terminal globular regions that are extended by a long coiled region, in the middle of which is found a hinge domain. Each SMC protein folds upon itself, and the N- and C-terminal regions build together an ATP-binding head domain, connected to the hinge by a 50 nm long antiparallel coiled-coil. In humans, the core of the mitotic cohesin complex is composed of the SMC1A and the SMC3 proteins, which form a heterodimeric ATPase by interacting through their hinge domains and, upon ATP binding, through their ATPase head domains, thus forming a composite ATPase module. A third core subunit, the RAD21 kleisin, binds

to SMC1A and SMC3 ATPase heads via its C- and N-termini, respectively, to form a tripartite ring structure that is required to capture and entrap DNA.

Various other auxiliary subunits and regulators interact with the core cohesin subunits to build a fully functional cohesin complex and to modulate its activity. Cohesin is loaded onto chromatin before the start of DNA replication, as early as at the end of telophase. It subsequently entraps the newly synthesized sister DNAs during the synthesis phase, holding them together until mitosis. Loading of the Cohesin complex onto chromatin depends on its recruitment by the loading factor NIPBL-MAU2 and on the ATP binding and hydrolysis by the SMC1A and SMC3 ATPase heads. Cohesion establishment depends on the replacement of NIPBL by the regulator PDS5, on SMC3 acetylation by the ESCO1 and ESCO2 acetyltransferases and, in vertebrates, on the recruitment of stabilization factors such as SORORIN. During interphase, cohesin also uses its ATPase activity to extrude large loops of DNA, which notably have roles in controlling gene expression and in the 3D-organization of chromatin structure through the formation of TADs, together with the vertebrate transcription factor and TAD-insulator protein called CTCF.

In vertebrates, including human, cohesin is removed from chromatin in two distinct ways, at two different steps of the cell cycle. Firstly, cohesin is removed from DNA at the onset of mitosis in a process called the "prophase pathway", through the destabilization of the SMC3-RAD21 interface, also called DNA "exit-gate", by destabilizing factor WAPL that binds to PDS5. Secondly, in anaphase, RAD21 is cleaved by the protease Separase, leading to the release of sister DNAs from the remaining cohesin complexes and allowing the segregation of sister chromatids and the progression into cell division. The HDAC8 deacetylase further deacetylates SMC3 and allows the recycling of cohesin components for the next cell division cycle.

The functional roles of Cohesin have been shown to strongly rely on the ATPase activity of this highly dynamic molecular motor. ATP binding and hydrolysis by the composite ATPase formed by SMC1A and SMC3 are known to drive important structural changes within the cohesin complex, which allows cohesin dynamic interactions with chromatin and enables its translocation on the DNA molecule. The ATPase activity of cohesin plays therefore a central role in the functional roles of cohesin.

However, the molecular bases of ATP binding and hydrolysis by SMC1A and SMC3 ATPase heads and the subsequent structural changes that occur and lead to cohesin mechanisms remain poorly understood. Moreover, mutations in the components of the cohesin complex and its regulators are involved in various types of solid and hematologic cancers and in developmental disorders called cohesinopathies. To this date, there is no targeted therapy to the dysfunctional Cohesin complex, as the structural mechanisms of the ATPase activity that are impaired in disease also remain poorly understood.

Since elucidating the three-dimensional organization of a protein complex will help understand its mechanisms, the structural characterization of human cohesin is therefore an indispensable step that should allow the fundamental understanding of cohesin mechanisms and functions, as well as their deregulations in disease, and further enable the design of new targeted therapies.

During my thesis project, I aimed to structurally characterize by X-ray crystallography both human cohesin SMC1A and SMC3 ATPase domains individually, bound to the C- and N-terminal domains of RAD21, respectively. Secondly, I aimed to assess their ATP binding and hydrolysis properties, by using biochemical and biophysical methods such as isothermal titration calorimetry (ITC) and ATPase activity assays.

Results

Biochemical and biophysical characterization of human Cohesin ATPase domains

Since the eukaryotic cohesin evolved into a heterodimeric SMC1-SMC3 complex, as compared to the prokaryotic homolog SMC homodimer, questions arise concerning the differences between SMC1 and SMC3 and what makes each subunit unique and essential for proper eukaryotic cohesin function. Their potential asymmetry and differences in ATP binding and hydrolysis mechanism is one of the many aspects that remain largely unexplored. Therefore, to assess the ATP binding and hydrolysis properties of SMC1A and SMC3, ITC experiments and

ATPase activity assays were performed using individual protein constructs of SMC1A and SMC3 ATPases, with shortened coiled coils.

The results show that SMC1A and SMC3 have distinct ATP binding and hydrolysis dynamics, which can contribute to an asymmetric cohesin ATPase, and that both ATPase heads potentially have overall distinct conformational dynamics, the SMC1A ATPase displaying more flexibility than the SMC3 ATPase. To complete these observations and assess the underlying structural changes, both ATPase heads were further structurally characterized by X-ray crystallography.

Structural characterization of human Cohesin ATPase domains

The protein constructs of the human cohesin SMC1A and SMC3 ATPase domains with shortened coiled-coils, either wild type or bearing respectively E1157Q and E1144Q (EQ) mutations, were co-expressed with either the C- or N-terminal domains of RAD21. The mutations in glutamines of glutamates located in the ATP binding pocket of each ATPase head was used to reduce any residual ATPase activity, allowing the stabilization of a cohesin-nucleotide structure suitable for crystallography assays. Each ATPase head sub-complexes were purified by affinity and size exclusion chromatography, then crystallized in absence or in the presence of ADP or ATPγS, a weakly hydrolysable ATP analog. Crystals were obtained in various conditions for both SMC1A and SMC3 ATPase heads, which allowed the collection of X-ray diffraction data, after extensive rounds of crystallization conditions optimization and of protein constructs optimization through the modulation of the length of the coiled-coils. Structure determination was made by molecular replacement, using the only structures of cohesin that were then available, which were the yeast SMC1 and SMC3 ATPase head structures. I thus solved the structures of SMC1A and SMC3 ATPase domains individually, in the apo, ADP- and ATPγS-bound conformations for the EQ mutant SMC1A ATPase domain, in the ADP- and ATPγS- bound conformations for the EQ mutant SMC3 ATPase domain, and in the ADP-bound conformation for the wild type SMC3 ATPase. Additionally, trials to crystallize the SMC3 ATPase in its apo configuration yielded diffracting crystals for which the structures contained an SO_4^{2-} ion, from the crystallization condition, bound to the catalytic site of the EQ mutant SMC3 ATPase.

The analysis of the newly solved structures revealed major structural differences between SMC1A and SMC3, which can explain their distinct ATP binding and hydrolysis dynamics. Notably, the results allowed the identification of a region of the human SMC1A ATPase around which a hinge-like movement occurs during the ATPase cycle, while these movements seem to be less striking for SMC3, thus explaining the dynamic differences. Additionally, a conformation of the DNA exit gate until now undescribed was observed in the SMC3 ATPase structures, which could potentially be involved in the regulation of the cohesin ATPase activity through the SMC3 ATPase head.

Conclusion

The results obtained throughout my thesis work provide a new insight into the ATP binding and hydrolysis mechanisms by human cohesin, by enabling the identification of structural changes that occur during the ATP hydrolysis cycle and highlighting specific regions around which these changes occur. Recently, cryo-electron microscopy (cryo-EM) structures of the human cohesin ATPase bound to NIPBL and DNA were released to the research community by the Hongtao Yu research team, as well as the yeasts cohesin ATPase structures in their SMC1A-SMC3 hetero-dimerized form in absence or in presence of the yeast NIPBL homolog Scc2 and DNA, by the Daniel Panne, the Kim Nasmyth and the Frank Uhlmann research laboratories (Shi et al., 2020, Higashi et al., 2020, Collier et al., 2020, Muir et al., 2020). My solved SMC1A and SMC3 ATPase crystallographic structures are highly complementary to the newly released cryo-EM structures, as they notably contribute together to structurally explain the effects of NIPBL and DNA binding to the SMC1A and SMC3 ATPase heads, as NIPBL and DNA were shown to both act in promoting ATP hydrolysis by the cohesin ATPase (Kim et al., 2019, Davidson et al., 2019).

Altogether, the results obtained during my thesis work reveal specific biochemical differences of SMC1A and SMC3 ATPase heads that stem from structural differences observed between both ATPases. They also suggest that the cohesin ATP binding and hydrolysis cycle might be asymmetrical, and help explain how it could be regulated by cohesin interacting factors, such as

NIPBL and DNA. Ultimately, in addition to help in elucidating the mechanisms at the core of cohesin functions, these results will further contribute to the development of structure-based therapies targeting the dysfunctional Cohesin in human diseases.

RESUME EN FRANÇAIS

L'ADN à l'intérieur du noyau des cellules eucaryotes est compacté sous forme de chromatine. La chromatine est une structure hautement dynamique, composée d'acide nucléique et de protéines qui régulent l'accès à l'information génétique. De nombreux facteurs épigénétiques s'associent à la chromatine et modulent sa structure, ayant donc un impact majeur sur le contrôle et la régulation des processus nucléaires.

Le complexe cohésine fait partie de la famille des complexes protéiques de maintenance structurelle des chromosomes (SMC, Structural Maintenance of Chromosomes) qui s'associent à la chromatine. Les complexes SMC sont composés d'un homodimère de protéines SMC chez les procaryotes et d'un hétérodimère chez les eucaryotes, et jouent un rôle essentiel dans l'organisation tridimensionnelle du génome et dans le maintien de sa stabilité et de son intégrité.

La cohésine est impliquée dans des processus de régulation du génome vitaux, notamment en jouant un rôle clé dans la cohésion des chromatides sœurs tout au long du cycle cellulaire, dans la ségrégation des chromosomes, dans la réparation des cassures du double brin d'ADN, dans l'organisation tri-dimensionnelle de la structure de la chromatine et dans la régulation transcriptionnelle via l'établissement de domaines chromatiniens, connus sous le nom de domaines topologiques associés (TADs, Topologically Associating Domain).

Les fonctions de la cohésine dépendent de sa capacité à établir des liens topologiques entre deux segments d'ADN intra- ou inter-moléculaires, en les rapprochant et en les maintenant dans une proximité spatiale. Le coeur du complexe cohésine eucaryote est composé d'un hétérodimère formé par deux protéines SMC, SMC1 et SMC3. SMC1 et SMC3 sont chacune composées de régions globulaires N- et C-terminales prolongées par une longue région en hélice, au milieu de laquelle se trouve un domaine charnière (hinge). Chaque protéine SMC se replie sur elle-même et les régions N- et C-terminales s'assemblent en un domaine possédant un motif de liaison de l'ATP, dite tête ATPase, connectée à la charnière par une superhélice antiparallèle de 50 nm de long. Chez l'homme, le coeur du complexe mitotique de la cohésine est composé des protéines

SMC1A et SMC3, qui interagissent via leurs domaines charnières et, lors de la liaison de l'ATP, via leurs têtes ATPase, formant ainsi un module ATPase composite. La kleisine RAD1, qui est une troisième sous-unité du complexe cœur de la cohésine, se lie aux têtes ATPase de SMC1A et SMC3 via ses domaines C- et N-terminaux respectivement, et forme ainsi une structure en anneau qui est nécessaire pour la capture de l'ADN par la cohésine. De nombreuses autres sous-unités auxiliaires et régulatrices interagissent avec les sous-unités coeur de la cohésine, afin d'aboutir au complexe cohésine entièrement fonctionnel et de moduler son activité.

Chez les vertébrés, la cohésine est recrutée sur la chromatine avant le début de la réplication de l'ADN, dès la fin de la télophase. La cohésine capture ensuite les molécules d'ADN constituant les chromatides sœurs nouvellement synthétisées pendant la phase S du cycle cellulaire, et les maintient ensemble jusqu'à la mitose. Le recrutement du complexe de cohésine sur la chromatine dépend de son facteur de recrutement NIPBL, avec son partenaire moléculaire MAU2, et de la liaison et de l'hydrolyse de l'ATP par les têtes ATPase de SMC1A et SMC3. L'établissement de la cohésion dépend du remplacement de NIPBL par le régulateur PDS5, de l'acétylation de SMC3 sur une paire de lysines très conservées par l'acétyltransférase ESCO1/2, et, chez les vertébrés, du recrutement de facteurs de stabilisation tels que la Sororine.

Pendant l'interphase du cycle cellulaire, la cohésine utilise également son activité ATPase dans l'extrusion de larges boucles d'ADN, qui jouent notamment un rôle clé dans le contrôle de l'expression des gènes et dans l'organisation tri-dimensionnelle de la structure de la chromatine par la formation de TADs. Chez les vertébrés, la cohésine coopère avec CTCF afin d'isoler et de maintenir les TADs.

Chez les organismes vertébrés, y compris chez l'homme, la cohésine est éliminée de la chromatine par deux voies distinctes, a différentes étapes du cycle cellulaire. Tout d'abord, la cohésine localisée sur les bras des chromosomes est éliminée de l'ADN au début de la mitose dans un processus appelé « voie de la prophase », par la déstabilisation de l'interface SMC3-RAD21 par le déstabilisateur WAPL, qui se lie à PDS5. L'interface entre SMC3 et RAD21 est ainsi également appelée « porte de sortie » de l'ADN. Deuxièmement, pendant l'anaphase du cycle cellulaire, RAD21 est clivé par la protéase Separase, conduisant ainsi à la libération de l'ADN des

complexes cohésine restants, et permettant la ségrégation des chromatides sœurs et la progression vers la division cellulaire. La désacétylase HDAC8 désacétyle ensuite SMC3, et permet le recyclage des composants de la cohésine pour le prochain cycle de division cellulaire.

Il est établi que les rôles fonctionnels de la cohésine reposent fortement sur l'activité ATPase de ce moteur moléculaire hautement dynamique. La liaison et l'hydrolyse de l'ATP par l'ATPase composite formée par SMC1A et SMC3 est connue pour entraîner d'importants changements structurels au sein du complexe cohésine, ce qui permet les interactions dynamiques de la cohésine avec la chromatine et permet sa translocation sur la molécule d'ADN.

L'activité ATPase de la cohésine joue donc un rôle central dans les rôles fonctionnels de la cohésine. Cependant, les bases moléculaires de la liaison et de l'hydrolyse de l'ATP par les têtes ATPase de SMC1A et SMC3 et les changements structurels sous-jacents qui conduisent aux mécanismes de la cohésine restent mal compris.

De plus, des mutations dans les composants du complexe de la cohésine et de ses régulateurs sont impliquées dans divers types de cancers solides et hématologiques et dans des troubles neurodéveloppementaux appelés cohésinopathies. À ce jour, il n'existe pas de thérapie ciblée contre le complexe cohésine dysfonctionnelle, car les mécanismes structurels de l'activité ATPase qui sont altérés dans la maladie restent également mal compris. Puisque l'élucidation de l'organisation tridimensionnelle d'un complexe protéique aide à comprendre ses mécanismes, la caractérisation structurale de la cohésine humaine est donc une étape indispensable qui devrait permettre la compréhension fondamentale des mécanismes et fonctions de la cohésine, ainsi que leurs dérégulations dans la maladie, pour à terme permettre la conception de nouvelles thérapies ciblées.

Pendant mon projet de thèse, j'ai eu pour objectif de caractériser structuralement les domaines ATPase de la cohésine humaine, liées au domaine C- et N-terminal de RAD21, respectivement, par cristallographie aux rayons-X. Dans un second temps, j'ai analysé leurs propriétés de liaison et d'hydrolyse de l'ATP, par l'utilisation de méthodes biophysiques et biochimiques telles que la titration par calorimétrie isotherme (ITC) et des tests d'activité ATPase.

Caractérisation biochimique et biophysique des domaines ATPase de la cohésine humaine

Puisque la cohésine eucaryote a évolué en un complexe hétérodimèrique, en comparaison à son homologue prokaryote, le complexe homodimèrique Smc, de nombreuses questions se posent quant aux caractéristiques de la cohésine qui rendent chacune des sous unités unique et essentielle au bon fonctionnement du complexe. La potentielle asymétrie du cycle ATPase de la cohésine ainsi que ses propriétés de liaison de l'ADN demeurent des aspects encore peu explorés. Par conséquent, afin d'analyser les propriétés de liaison et d'hydrolyse de l'ATP par les domaines ATPase de SMC1A et SMC3 de la cohésine humaine, des expériences d'ITC ainsi que des tests d'activité ATPase ont été réalisés en utilisant des constructions des domaines ATPase avec leurs régions en superhélice raccourcies.

Les résultats montrent que SMC1A et SMC3 ont des dynamiques de liaison et d'hydrolyse de l'ATP bien distinctes, qui peuvent contribuer à un cycle ATPase asymétrique, et que les deux têtes ATPase ont globalement des dynamiques conformationnelles bien distinctes, le domaine ATPase de SMC1A montrant plus de fléxibilité que le domaine ATPase de SMC3. Afin de compléter ces analyses et comprendre ces observations, les domaines ATPase de SMC1A et SMC3 ont été caractérisés par cristallographie au rayons-X, individuellement.

Caractérisation structurale des domaines ATPase de la cohésine humaine

Les constructions protéiques des domaines ATPase de SMC1A et SMC3 de la cohésine humaine avec des régions en superhélice raccourcies, soit natives, soit portant respectivement la mutation E1157Q et E1144Q (EQ), ont été co-exprimées avec le domaine C-ou N-terminal de RAD21, respectivement. Les mutations en glutamine des glutamates catalytiques ont été utilisées, afin de réduire toute activité ATPase résiduelle, permettant de cette façon de stabiliser

une conformation de la cohésine liée à l'ATP qui conviendrait mieux à des essais de cristallisation. Chacune des têtes ATPase a été purifiée par plusieurs étapes de chromatographie d'affinité, puis par chromatographie d'exclusion de taille. Des essais de cristallisation ont ensuite été effectués sur les complexes purifiés, en absence et en présence d'ADP, ou d'ATPγS, un analogue faiblement hydrolysable de l'ATP. Des cristaux ont été obtenus dans différentes conditions, aussi bien pour le domaine ATPase de SMC1A que pour celui de SMC3. Après plusieurs étapes d'optimisation des cristaux ou bien des constructions des protéines si nécéssaire, la diffraction des rayons-X par les cristaux obtenus a permis de collecter plusieurs jeux de données de diffraction. La résolution des structures a été effectuée par remplacement moléculaire, en utilisant les seules structures de SMC1A et SMC3 alors connues, celles de la levure *Saccharomyces cerevisiae.* J'ai ainsi pu résoudre les structures des domaines ATPase de la cohésine humaine, sous une forme apo, liée à l'ATPγS et à l'ADP pour le mutant EQ de SMC1A, sous une forme liée à l'ADP et à l'ATPγS pour le mutant EQ de SMC3, et sous une forme liée à l'ADP pour le domaine ATPase de SMC3 natif. De plus, les essais de cristallisation du domaine ATPase de SMC3 sous une forme apo a uniquement donné des cristaux pour lesquels la structure résolue contenait un ion sulfate, provenant de la condition de cristallisation, lié au site catalytique du mutant EQ de SMC3.

L'analyse des structures nouvellement résolues a ainsi révélé des différences structurales majeures entre les domaines ATPase de SMC1A et de SMC3, qui peuvent justifier leur dynamique de liaison et d'hydrolyse de l'ATP distinctes. Notamment, les résultats ont permis l'identification d'une région du domaine ATPase de SMC1A humain autour duquel des mouvements de charnière ont lieu pendant le cycle ATPase, alors que ces mouvements sont moins marqués pour SMC3. De plus, une conformation jusqu'alors non décrite de la porte de sortie de l'ADN sur le domaine ATPase de SMC3 a pu être observée. L'ensemble de ces données appuyent ainsi le caractère distinct de la dynamique conformationnelle des domaines ATPase de SMC1A et de SMC3 de la cohésine humaine.

Conclusion

Les résultats obtenus pendant mon travail de thèse fournissent un nouvel aperçu du cycle de liaison et d'hydrolyse de l'ATP par la cohésine humaine, en permettant l'identification de changements structuraux qui ont lieu pendant le cycle ATPase et en permettant de distinguer les régions spécifiques sur les structures autour desquelles les changements de conformation ont lieu. Récemment, la structure du module ATPase de la cohésine humaine liée à son facteur de recrutement NIPBL et à l'ADN a été résolue par cryo-microscopie électronique (cryo-EM) par l'équipe de recherche de Hongtao Yu (Shi et al., 2020). En parallèle, les structures des domaines ATPase de la cohésine des levures *S. cerevisiae* et *Schizosaccharomyces pombe* ont également été résolues, en présence d'ADN et de Scc2, l'homologue chez la levure de NIPBL, également résolues par cryo-EM par les équipes de Daniel Panne, de Kim Nasmyth, et de Frank Uhlmann (Collier et al., 2020, Higashi et al., 2020, Muir et al., 2020). Les structures cristallographiques des domaines ATPase SMC1A et SMC3 humaines que j'ai pu résoudre sont hautement complémentaires aux nouvelles structures résolues par cryo-EM de la cohésine. Ces structures contribuent ensemble dans l'explication des effets de régulation positive de NIPBL et de l'ADN sur l'activité ATPase de la cohésine, puisque NIPBL et l'ADN peuvent favoriser et augmenter le taux d'hydrolyse de l'ATP par la cohésine (Kim et al., 2019, Davidson et al., 2019).

Dans leur ensemble, les résultats obtenus pendant mes travaux de thèse révèlent des différences biochimiques et biophysiques spécifiques entre le domaine ATPase de SMC1A et celui de SMC3 de la cohésine humaine. Ils suggèrent également que le cycle ATPase de la cohésine pourrait être asymétrique, et aident à comprendre comment ce cycle pourrait être régulé par des régulateurs de la cohésine comme NIPBL et l'ADN. Finalement, en plus d'aider à aider à comprendre les mécanismes au cœur de l'activité de la cohésine, ces résultats fournissent également les bases structurales qui seront nécessaires pour contribuer, à terme, à la mise au point de thérapies ciblées pour la cohésine dysfonctionnelle dans un contexte pathologique.

INTRODUCTION

I. DNA and genome organization: from the double helix to the X-shaped chromosome

A. Discovery of the DNA molecule, as the bearer of genetic inheritance

The DNA molecule was first identified and isolated in 1869 from the nucleus of leucocytes by the physiological chemist Friedrich Miescher, as a phosphate-rich molecule that he named "nuclein" (Dahm, 2005). Later in the 1870s, Walther Flemming, one of the founders of cell biology, observed that the nuclear substance from *Salamandra salamandra* cells could be distinctly stained by specific dyes. For this reason, Flemming named that nuclear substance ''chromatin'', derived from Greek *chroma*, ''color''. By selectively staining chromatin, Flemming carefully observed the cell division process and described that chromatin goes through a drastic transformation: it collapses into long thread-shaped bodies, which subsequently split in half longitudinally, shortly before cell division.

Flemming named this transformation process ''mitosis'', from the Greek term *mitos*, "warp thread", and the suffix -osis "process, action". Flemming's seminal works on the cell cycle, not only on mitosis but also later on meiosis, the cell division process of germinal cells, thereby paved the way to modern cell biology. The filamentous bodies observed by Flemming were later termed ''chromosomes'' by Wilhelm Waldeyer in 1888 (Paulson et al., 2021). In the early 1900s, the significance of chromosomes as carriers of the genetic information, from one dividing parent cell to its next generation daughter cells, was settled by Theodor Boveri and Walter Sutton through the fundamental ''chromosomal theory of inheritance'', thereby supporting the earlier described Mendelian heredity mechanisms (Crow and Crow, 2002; Dahm, 2005).

A century after the discovery of DNA, in 1953, geneticists James Watson and Francis Crick elucidated the structure of the double stranded DNA molecule, based on X-ray crystallography diffraction patterns obtained by their coworkers Rosalind Franklin and Raymond Gosling. Since then, the past decades have seen a gigantic leap forward in the characterization of DNA, and how this fascinating molecule holds and transmits the genetic information necessary to build every living organism. However, numerous relevant questions on the structural and molecular details of this vital process remain elusive.

From our current knowledge, if stretched out, the total length of DNA inside a human cell nucleus adds up to about 2 meters long. This doubles to 4 meters during the cell cycle, towards mitotic cell division. Knowing precisely how the cell manages to tightly pack the DNA molecule inside a 6 micrometers cell nucleus, while still granting access to the genetic information and remarkably avoiding an unmanageable entanglement, is one of the major riddles being addressed by modern biology.

It is now known that the chromatin three-dimensional organization and the maintenance of its stability require a few but extremely important macromolecular complexes that interact with DNA, among which the cohesin complex (Paulson et al., 2021). The following sections of this introductory chapter will shortly describe the mitotic cell-cycle process, then focus on our current knowledge of the different levels of DNA packaging inside eukaryotic mitotic cells. The subsequent sections will then describe the main protein actors that participate in the establishment and the regulation of this essential cellular process, and the major roles of the cohesin complex. Comparison with corresponding mechanisms in bacteria will be made whenever relevant.

B. Transmission of genetic information through the cell cycle

The cell-cycle is a series of cell growth and development processes, during which one cell duplicates its genetic material and then divides into two daughter cells bearing the same genetic information. In eukaryotic somatic cells, the cell cycle is composed of two main phases: the interphase and the mitotic phase. During interphase, the cell is in a proliferative state. It increases in size, replicates its DNA as sister chromatids, and prepares for cell division during three distinct stages, called G1 (Gap 1), S (synthesis phase), and G2 (Gap 2), respectively.

Once the DNA is replicated, the chromatin is condensed while the replicated sister chromatids are stably attached together through their entire length. The cell thus enters the M (mitotic) phase, composed of the mitosis and cytokinesis subphases. During mitosis, the cell positions each chromosome for segregation and cell division. The eukaryote mitotic phase is itself divided into distinct stages: prophase, metaphase, anaphase, and telophase (Figure 1).

 During prophase, the chromosome is extensively condensed, the sister chromatid arms are separated but still held together at their central constricted region, called centromere. A cytoskeletal structure, called mitotic spindle or spindle apparatus, which is required for cell division, begin to assemble, and positions itself at the opposite poles of the cell. During prometaphase, between the prophase and metaphase phases, the nuclear envelope is fragmented and allows the mitotic spindle to have access to the chromosomes. Microtubules emerging from the spindle apparatus attach at chromosome kinetochores, which are protein assemblies associated to the centromeres, and begin to pull the chromosomes in opposite directions towards each pole of the cell. Chromosomes are thus bi-orientated and aligned on the equatorial plate of the cell during metaphase. A thorough molecular spindle assembly checkpoint (SAC) ensures that the chromosomes are correctly positioned before entering the next phase of cell division. During anaphase, the sister chromatids composing each chromosome are abruptly pulled away from each other.

By the end of mitosis, one copy of each sister chromatid is thus distributed into each daughter cell and both cells separate at the outcome of cytokinesis. The new cells then enter the G0 phase, a short non-proliferative state, before entering the interphase for the start of a new cell cycle.

During the cell cycle, the DNA molecule therefore undergoes very different stages of morphological transformations, which compact and organize the genome, while still granting access to the underlying genetic information.
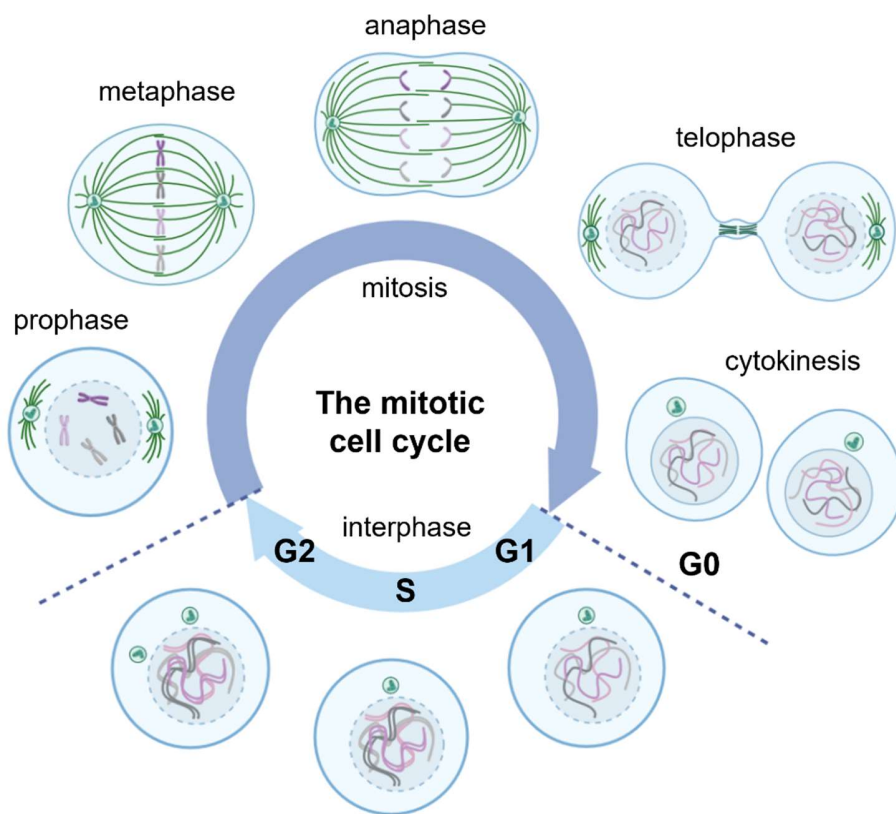
Figure 1: The mitotic cell cycle. Schematic representation of the eukaryotic cell division cycle during mitosis. The mitotic cell cycle is divided into two main phases, interphase, and mitosis. Mitosis subdivides mainly into prophase, metaphase, anaphase, and telophase phases. The cell division is completed by cytokinesis, during which the parent cell separates into two daughter cells.

C.   DNA organization during interphase: different levels of compaction

1.   DNA compaction into chromatin

Inside each cell nucleus, the DNA molecule is found associated with various proteins to form the nucleoprotein structure called chromatin. DNA is first compacted into a basic repeating chromatin folding unit, which was described by Ada and Donald Olins (Olins and Olins, 1974) and named "nucleosome" by Pierre Chambon and his collaborators (Oudet et al., 1975). First, four core histone proteins H2A, H2B, H3 and H4 assemble into H2A-H2B and H3-H4 heterodimeric pairs. Two copies of each histone pair then assemble to form a histone octamer, around which are wrapped around 147 base pairs of DNA, to build the nucleosome (Kornberg, 1977; Luger et al., 1997) (Figure 2A). The nucleosome is further stabilized by histone H1 that binds at the entry/exit sites of the DNA segment from the nucleosome, forming the chromatosome (Figure 2B).

Sequential nucleosomes on a DNA molecule build up the so-called 10 nm chromatin fiber (Figure 3). The next order of chromatin compaction has long been thought to be a 30 nm fiber, where the chromatin fiber would arrange in solenoidal arrays (Paulson et al., 2021). However, the existence of the 30 nm fiber is nowadays strongly challenged by new research. Notably, Ou et al., 2017 proposed that the DNA fiber could be folded into disordered and flexible chains of variable diameters, which would be further packed with variable densities inside the cell nucleus, regulating in this way the access to the underlying DNA. In bacteria, histone-like basic proteins also associate with DNA and contribute to bacterial genome folding (Wang and Maier, 2015).
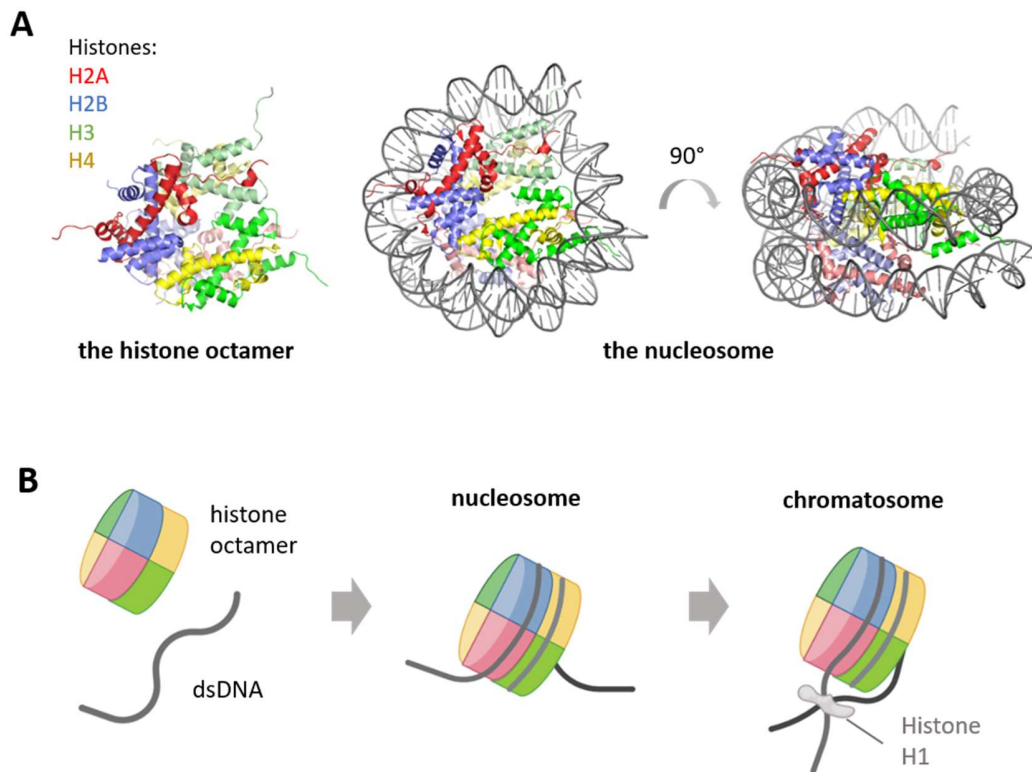
Figure 2: Compaction of the DNA molecule into the nucleosome and chromatosome. (A) The four core histones H2A, H2B, H3 and H4 assemble in pairs, then in a histone octamer, around which wraps a 147 bp DNA segment to build the basic chromatin folding unit, the nucleosome. (B) The DNA ends emerging from the nucleosome are further secured by the H1 histone, thus forming the chromatosome.

2.   Chromatin organization into loops and topological domains

The compacted DNA is further organized into chromatin loops and topological domains. Electron micrographs of the *Escherichia coli* chromosome showed for the first time the folding of chromatin into topological domains, consisting in a rosette of loops of supercoiled DNA, emerging from a central core that incorporates RNA (Kavenoff and Bowen, 1976).

These self-interacting chromosomal domains have now been detected genome-wide in a large range of organisms, from bacteria to fungi and metazoans. In higher eukaryotes, during interphase, it has been shown that the chromosomic arms are formed by extremely organized

cis-interaction domains that were discovered in 2012 by modern genome analysis tools (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012). Using high-throughput Chromosome Conformation Capture methods (Hi-C), seminal studies showed increased frequency of intra-molecular DNA contacts ranging from hundreds of kilobase pairs to one Megabase pairs (Mb) of DNA, which were named Topologically Associating Domains (TADs) by the work of Edith Heard's team in 2012 (Nora et al., 2012) (Figure 3). TADs are cell type variants: they represent functional domains of transcriptional regulation and participate in the formation of compartments inside each cell type, that relate to their gene expression status.

It is now known that the chromatin loops that give rise to TADs arise from an active loop extrusion process, which is mediated by the motor activity of specific multi-protein complexes called Structural Maintenance of Chromosome (SMCs) proteins.

3.   Chromatin compartmentalization and chromosome territories

Modulation of the structure of chromatin is a critical process for nuclear processes, notably transcriptional regulation, as it controls the access to the DNA. It is known that in higher eukaryotes, the DNA molecule is maintained in discrete regions that occupy a limited volume inside the cell nucleus, known as chromosome territories.

The use of Hi-C methods allowed the identification of chromatin contacts genome-wide and led to the identification of patterns of chromatin sub-compartments into the chromosome territories, with an enrichment in intra- and inter-contacts from the same type of compartment (Bonev and Cavalli, 2016; Dekker and Heard, 2015; Falk et al., 2019; Lieberman-Aiden et al., 2009; Sun et al., 2018). Over extended distances beyond the Megabase pairs of DNA, chromatin is distributed into two distinct and mutually excluded type of sub-compartments, which are clustered heterochromatic or euchromatic regions, called A and B compartments, respectively. A compartments form clusters of gene-rich and transcriptionally active loci, whereas B compartments, which are clustered near the nuclear periphery in the heterochromatic region,

are gene-poor and enriched in silenced chromatin, such as telomere and pericentromeric loci. More than 92% of the human genome is euchromatic (International Human Genome Sequencing Consortium, 2004).

Hi-C also shows that repressed or activated chromatin compartments associate in preference with domains having similar level of activity (Imakaev et al., 2012). Nuclear organization leads therefore to compartmentalization and regulation of the access to the underlying DNA molecule, and can directly correlate to cell type-specific expression of genes.

Towards the mitotic phase during the cell cycle, the chromosomal DNA is replicated and undergoes a dramatic morphological change to assemble into the discrete rod-shaped chromatids that build up the chromosomes.

D.   DNA organization during mitosis: the chromosome

During the mitotic cell cycle, two replicated sister chromatids are closely held together to form the higher order architecture called chromosome. In humans the entire genome is packed into 46 chromosomes that are organized into 23 pairs. Each chromosome is composed of two sister chromatids, each containing one copy of the same genetic information. Sister chromatids are notably tightly linked at a constricted region called centromere. The mitotic chromosome is organized by a central helical scaffold from which emerge chromatin loops, thus adopting a spiral "staircase" folding. The chromosome is thus the highest level of chromatin compaction, with an approximate 10^4-fold compaction level compared to the size of its DNA molecule (Paulson et al., 2021) (Figure 3).

As soon as the cell division is complete, the chromatids "dissolve" back into chromatin, and reappear as chromosomes at the next prophase step. This indicates that chromosomes must maintain a controlled physical continuity and unity during successive cell divisions, through a finely orchestrated process. The intricate task of keeping each cell genome tightly packed throughout the entire cell cycle, and yet free of any entanglement, is performed by specific

proteins that bind to DNA to organize and maintain the genome stability. Chromatin and chromosomes are shaped by a few but extremely important multiprotein complexes. Those include complexes formed by Structural Maintenance of Chromosome (SMC) proteins, which fold and organize chromatin through their ability to form loops of DNA (Figure 3).
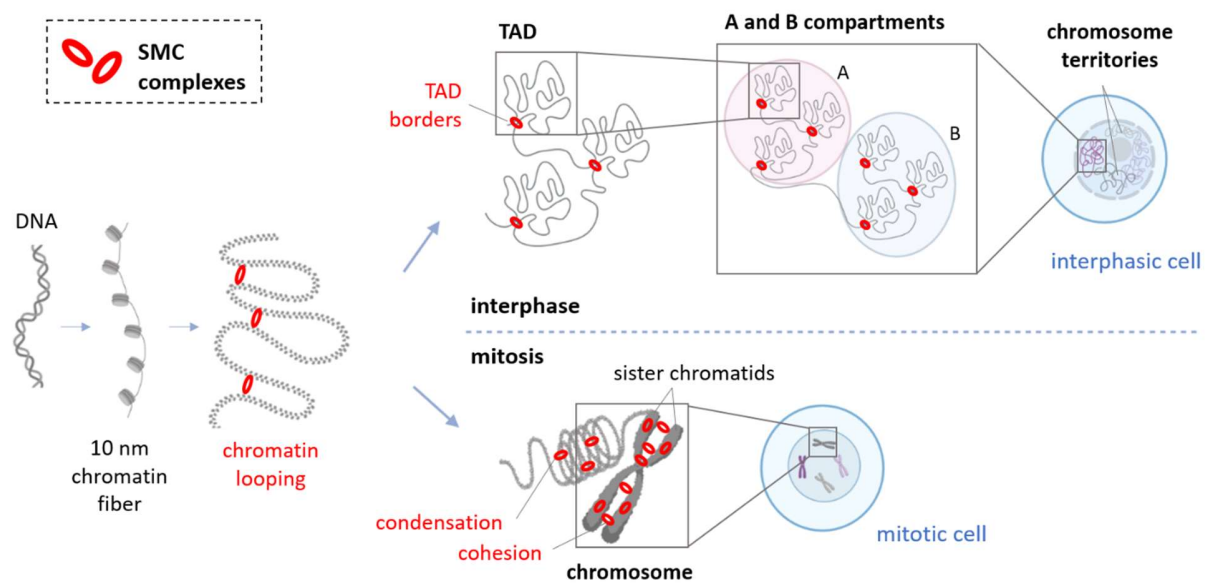
Figure 3: The different levels of DNA compaction and the role of SMC complexes. The sequential nucleosome positioning onto the DNA molecule compacts the DNA into the 10 nm chromatin fiber. The fiber is further compacted, with different levels of compaction depending on the state of the cell cycle. SMC complexes (red ring shapes) have essential functions in the chromatin organization processes.

II.   SMC-Kleisin complexes: discovery, architecture, and functions

A.   Discovery of the SMC proteins

From the late 1980s to the early 2000s, successive discoveries were made of homologous proteins interacting with chromatin, which were gradually shown to be crucial for proper genome organization and dynamics in all kingdoms of life.

In 1991, the Sota Hiraga team identified the *mukB* gene in *Escherichia coli*, and showed it to be required for chromosome partition (Niki et al., 1991). Prediction of the secondary structure of the *mukB* gene product suggested a globular N-terminal domain, containing an ATP binding motif, separated from a globular C-terminus by a long coiled region. Further work from the Hiraga team showed that the resulting MukB protein is able to dimerize by adopting a rod-like structure with two globular domains separated by a long coiled region, although it was not yet clear whether the two proteins would form intra- or intermolecular coiled coils (Niki et al., 1991).

Alongside, the *smc1* gene was discovered and cloned from the budding yeast *Saccharomyces cerevisiae*. The product protein Smc1 was shown to be required for chromosome segregation and proper cell division, hence the name originally standing for "Stability of Minichromosomes 1" (Larionov et al., 1985; Strunnikov et al., 1993). Nucleotide and protein sequence comparisons against existing databases showed that yeast Smc1 likely shares a similar architecture with the bacterial MukB, and with other related bacterial and eukaryotic proteins.

Successively, other homologous proteins to MukB and Smc1 were identified in all domains of life, from eukaryotes to prokaryotes and archaea. It was consequently proposed that these proteins would all belong to an ubiquitary superfamily of chromosomal proteins, the **Structural Maintenance of Chromosomes (SMC)** proteins, which share the same molecular architecture and have fundamental functions in chromatin organization and dynamics during the cell cycle in all domains of life (Hirano, 2005; Jessberger, 2003; Strunnikov and Jessberger, 1999).

B.   Assembly and architecture of the SMC-Kleisin complexes

1.   SMC proteins organization and architecture

How SMCs interact with the DNA molecule is a fundamental question, whose answer requires in the first place to explore the molecular and structural aspects of these intriguing proteins.

a)   *SMC sequence organization*

It was inferred from sequence analyses that each SMC protein is composed of a polypeptide chain of approximately 110-170 kilo Daltons (Harvey et al., 2002) that adopts a symmetrical domain organization (Figure 4). In the middle of the polypeptide sequence is found an ordered domain that is referred to as the "hinge". From either side of the hinge emerges two long regions that adopt a coiled secondary structure, which end up forming globular domains at the N- and C-termini of each SMC. The terminal domains contain specific ATP-binding and dimerization motifs that are shared with ATP-binding cassette (ABC) transporters (Higgins and Linton, 2004; Jones and George, 2004).

b)   *SMC three-dimensional folding*

By solving the crystallographic structure of the hinge domain from the *Thermotoga maritima* bacterium Smc, the Kim Nasmyth research team revealed that each SMC protein folds upon itself at their central hinge domain (Haering et al., 2002). The two coiled arms emerging on either side of the hinge swirl around each other into a 50 nm intramolecular and anti-parallel coiled-coil domain, which brings together the N- and C-termini regions to build a globular hemi-ATPase module, also called ATPase "head" domain. Each SMC organizes therefore into a rod-shaped structure, with a hinge domain at one end and a hemi-ATPase module at the other end, both separated by the long and flexible coiled-coil.

SMCs have two short discontinuity regions within their coiled-coil domain: a region proximal to the head module, called "joint" (Diebold-Durand et al., 2017), and an off-centered region towards the hinge domain referred to as the "elbow" (Figure 4). This latter region allows the SMC arm to adopt a kinked configuration (Bürmann et al., 2019), which can bring the hinge into spatial proximity with the ATPase head and is likely involved in major conformational switches that allow DNA transactions by SMCs (Bauer et al., 2021).

c) SMC-SMC dimerization

The structural characterization of *T. maritima* Smc hinge domain by Haering et. al 2002 additionally paved the way for the elucidation of the SMCs assembly, by revealing that two Smc proteins from *T. maritima* homodimerize through their hinge domains into a composite toroid-shaped structure. Additionally, Haering et al. 2002 showed by electron microscopy experiments that the Smc1 and Smc3 proteins from the yeast *S. cerevisiae* behave as monomers when expressed alone, but interestingly form heterodimers when co-expressed. This observation agreed with the heterodimerization observed for human and *Xenopus laevis* SMC proteins (Hirano et Mitchison, 1994; Anderson et al. 2002).

Moreover, the crystallographic structure of the archaea *Pyrococcus furiosus* Smc, in absence and in presence of ATP, revealed that ATP binding by two SMC proteins induces the dimerization of their hemi-ATPase domains (Lammens et al., 2004).

Therefore, it was concluded that in order to build a complete and functional ATPase module, SMC proteins must dimerize, either forming a homodimer in prokaryotes and archaea, or a heterodimer in eukaryotes. The dimerization thus occurs at two interfaces. Firstly, at their hinge domains, to form a V-shaped SMC-SMC dimer (Haering et al., 2002). Secondly, SMCs transiently dimerize through their ATPase heads upon ATP binding (Lammens et al., 2004), thus building a composite functional ATPase module where two ATP molecules are sandwiched in a head-to-head manner at the dimerization interface (Figure 4).

2.   SMC-interacting Kleisin proteins

A large family of SMC-interacting proteins, called Kleisins, were progressively shown to be specific partners of SMCs (Schleiffer et al., 2003) and to be as essential as the SMCs for vital processes, including chromosome segregation and condensation (Soppa et al., 2002). Kleisins are composed of conserved globular N- and C-terminal domains that asymmetrically bridge two SMC hemi-ATPase modules, thereby closing a large SMC-Kleisin ring-shaped structure (Figure 4) (Gruber et al., 2003; Schleiffer et al., 2003; Haering et al., 2004, 2002).

Specifically, the kleisin N-terminus is folded into a helix-turn-helix motif (HTH) that forms a three α-helix bundle with the proximal coiled-coil region (neck) of one SMC ATPase head, whereas the C-terminus folds into a winged-helix domain (WHD) that binds to the bottom side (cap) of the second SMC ATPase. The Kleisin N- and C-terminal domains are separated by a poorly ordered linker domain of variable length and sequence that constitutes a binding platform for numerous other accessory and regulatory subunits of SMC-Kleisin complexes.

Notably, two structurally distinct classes of accessory subunits dynamically interact with the central domain of Kleisins: KITE proteins (Kleisin interacting tandem winged-helix elements of SMC complexes) in bacterial and archaeal Smcs as well as the eukaryotic SMC5/SMC6 complex, and HAWK proteins (HEAT-repeat-containing proteins associated with Kleisins) in the eukaryotic condensin and cohesin complexes (Baxter et al., 2019; Haering and Gruber, 2016; Palecek and Gruber, 2015) (Figure 5).

3.   The SMC-Kleisin complexes are vital molecular motors

The core of SMC complexes is therefore built by two SMC proteins and a Kleisin subunit. The tripartite ring-like structure it provides allows the SMC complexes to embrace and entrap one or two DNA segments (Gruber et al., 2003; Schleiffer et al., 2003; Haering et al., 2004, 2002). Additionally, SMC complexes are known for their ability to extrude large loops of DNA, which

supports many essential cellular processes, including chromosome segregation, chromatin compaction and organization, as well as DNA repair and transcriptional regulation.

The dynamic interaction of SMC complexes with DNA and their DNA translocase activity is allowed and precisely orchestrated by the ATP binding and hydrolysis cycles of their ATPase head domains, thereby making the ATPase module the central motor of these intriguing molecular machines (Figure 4). Moreover, the regulatory subunits that bind to the core SMC complex throughout the cell cycle all act in the vicinity of, or directly on the ATPase module, to regulate its ATPase activity. However, although the SMC research field has tremendously advanced our understanding of the function of many of the regulatory partners of SMC complexes, most of their precise modes of interaction with the core complex and the detailed molecular mechanisms of the ATP binding and hydrolysis cycle by the ATPase module remain to be fully characterized.

The biochemical and structural characterization of SMC complexes ATPase module is therefore a primordial step towards the full understanding of the molecular mechanisms that underlie the vital cellular functions of SMC complexes.
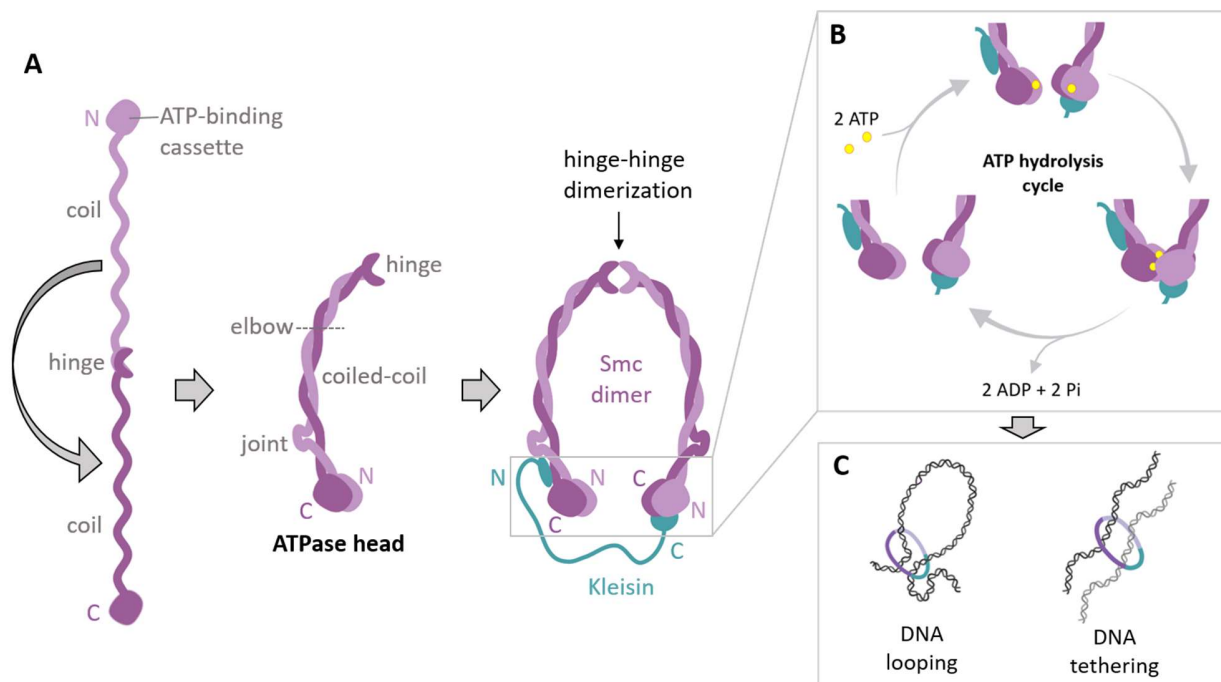
Figure 4: The molecular architecture of SMC complexes. SMC proteins are composed by two N-ter and C-ter globular domains separated by a long coiled region, in the middle of which is found a hinge region. SMCs fold back upon themselves, forming a globular ATPase head separated from the hinge by a long intra-molecular coiled-coil. SMC proteins dimerize at their hinge regions, and, upon ATP binding, at their ATPase heads. Both ATPase heads are connected by a Kleisin protein. The ATP binding and hydrolysis into ADP + Pi cycle of SMC complexes provides the chemical and mechanistic basis for DNA transactions by these molecular machines.

C.   Emergence and evolution of eukaryotic SMC complexes

1.   Emergence and evolutionary aspects

All SMC-Kleisin ternary complexes are therefore comparable biological systems, which dynamically organize chromatin and chromosomes throughout the different stages of the cell cycle, in all living cells. Although sharing sequence features and a common global architecture, SMC complexes are however distinguished by specialized functions in chromosome dynamics in different organisms, and at different steps of the cell cycle.

While prokaryotes usually have one SMC complex that performs several roles in chromosome organization and segregation, in eukaryotes, however, multiple SMC complexes are found and have distinct and specialized functions as well as complexified ways of regulation (Figure 5).

One can speculate that SMCs evolved in the perspective of a gain of complexity, fundamentally due to the evolution of the eukaryotic chromosome architecture. Chromosomes are for example larger in size in eukaryotes, as compared to prokaryotes, and might therefore require enhanced and specialized organizational tools, which would be provided by the differentiation of SMC complexes. Phylogenetic analyses of SMC proteins indeed suggest that throughout evolution, the duplication event of a single bacterial *smc* gene, which led to the apparition of different paralogous SMCs in eukaryotes, is likely to have occurred early during the emergence of eukaryotes (Cobbe and Heck, 2004) that happened about 2.7 billion years ago.

Consequently, even though a similar mode of action can be delineated for all SMCs, detailed characterization of each SMC complex is fundamental (i) to understand the molecular mechanisms of each complex and of its regulators, (ii) to explain what makes them unique in performing their specific functions, and (iii) to assess whether there are organism-dependent differences between members within each paralogous group, and how these differences can affect their functions.

2.   The eukaryotic SMC complexes

In eukaryotes, SMCs are distributed into six paralogous protein families, ranging from SMC1 to SMC6, which are paired into the core of three functional complexes (Figure 5). Each pair is formed by a κ-SMC and ν-SMC dimer, which asymmetrically interacts with the Kleisin subunit (κ, "kappa" standing for cap-interacting, and ν, "nu" for neck-interacting, with the Kleisin) (Bürmann et al., 2013), into the following complexes: SMC1-SMC3 in the cohesin complex, SMC4-SMC2 (mostly designated as SMC2-SMC4 in the literature) in the condesin complex and SMC5-

SMC6 in the SMC5-SMC6 complex (Wu and Yu, 2012) (Figure 5). Each eukaryotic SMC complex is regulated by a distinct set of regulatory subunits that bind to the SMC-Kleisin core and that are specialized to the function of their specific complex (summarized in Table 1).
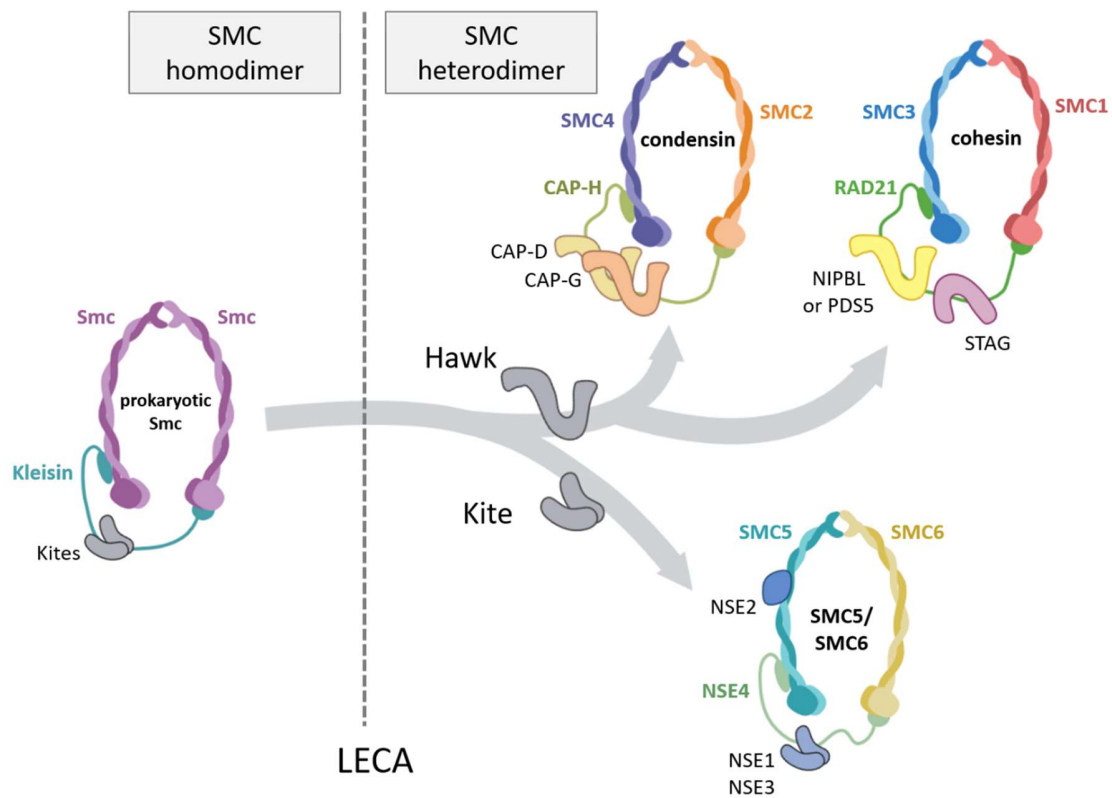
Figure 5: Kleisin-binding molecular partners of SMC complexes. Throughout evolution, SMC complexes subunit-composition has diversified. An homodimer is found in prokaryotes, whereas eukaryotes possess the three homolog heterodimeric SMC complexes cohesin, condensin and the SMC5/SMC6 complex. The Kleisin-binding regulatory subunits have also evolved: the Kite proteins were replaced by Hawks in the cohesin and condensin complexes. LECA : Last Eukaryotic Common ancestor. This figure was adapted from Wells et al., 2017.

a)   SMC2-SMC4: the condensin complex

The condensin complex was first identified as required for the condensation of mitotic chromosomes, from *X. laevis* eggs extract (Hirano and Mitchison, 1994; Hirano et al., 1997).

Condensin facilitates the topological resolution of sister chromatids resolution during mitosis and is required for condensation of chromosomes arms in higher eukaryotes (Hirano and Mitchison, 1994). Most eukaryotic species possess two types of condensin complexes, the canonical condensin I and the condensin II, whereas some species such as yeasts only possess the equivalent of condensin I. Both condensin I and II share the same SMC2-SMC4 heterodimer but differ by having specific Kleisins and HAWK regulators (Figure 5, Table 1). Condensin II is nuclear during interphase and accumulates on chromosomes at the onset of prophase, and extrudes large loops of DNA. Condensin I is however cytoplasmic during interphase. Once the nuclear envelope has dissolved at the end of prophase, condensin I relocates to chromosome arms, where it forms small loops into the larger condensin II loops. Condensin is therefore a key factor in the organization of the mitotic chromosome structure into the spiral "staircase" architecture (Davidson and Peters, 2021; Gibcus et al., 2018; Hirano, 2012; Ono et al., 2003; Uemura et al., 1987).

b) The SMC5-SMC6 complex

The SMC5-SMC6 complex was first identified through a genetic screen in *Schizosaccharomyces pombe*, which showed its implication in post replicative DNA damage repair (Lehmann et al., 1995). Conserved in other eukaryotes, this complex promotes genomic stability by acting as a DNA compaction device and by participating in DNA repair processes during interphase (Aragón, 2018; Gallego-Paez et al., 2014). Various non-SMC proteins, identified as Nse (for NonSmc Elements), interact with the core of SMC5-SMC6 and confer additional biochemical activities to the complex. Notably, the Nse and Nse2 subunits provide ubiquitin and SUMO ligase activities, respectively (Gallego-Paez et al., 2014) (Figure 5, Table 1). Although the SMC5-SMC6 complex is rather elaborated, it is only named after its two SMCs, as its modes of action remain either poorly understood or unknown.

Table 1: The eukaryotic SMC complexes and their regulators.

		S. cerevisiae (Budding yeasts)	S. pombe (Fission yeasts)	A. thaliana (Plants)	C. elegans (Nematodes)	D. melanogaster (Insects)	H. sapiens (Human)
Cohesin							
Subunits	κ-SMC	Smc1	Psm1	SMC1	HIM-1	Smc1	SMC1A, SMC1B*
	ν-SMC	Smc3	Psm3	SMC3	SMC-3	Smc3	SMC3
	α-kleisin	Scc1, Rec8*	Rad21, Rec8*	SYN2, SYN3, SYN4, SYN1*	SCC-1, COH-1, REC-8*, COH-3/4*	Rad21, C(2)M?*	RAD21, REC8*, RAD21L*
	HEAT-A	Pds5	Pds5	PDS5	EVL-14	Pds5	PDS5A, PDS5B
	HEAT-B	Scc3	Psc3, Rec11*	SCC3	SCC3-3	SA, SA2*	SA1, SA2, SA3*
Regulators	Kollerin (loading complex)	Scc2	Mis4	SCC2	PQN-85	Nipped-B	NIPBL
		Scc4	Ssl3	-	MAU-2	Mau-2	MAU2
	Acetyl-transferase	Eco1	Eso1	ECO1	F08F8.4	Eco + San	ESCO1, ESCO2
	Deacetylase	Hos1	-	-	-	-	HDAC8
	Stabilizer	-	-	-	-	Dalmatian	Sororin
	Destabilizer	Wpl	Wpl1	WAPL	WAPL-1	Wapl	WAPL
	Separase	Esp1	Cut1	ESP	1-Sep	Sse + Thr	ESPL1
	Shugoshin-phosphatase complex	Sgo1	Sgo1	SGO1	-	Mei-S332	SGOL1
		PP2A	PP2A	-	-	-	PP2A
Condensin							
Subunits	κ-SMC	Smc4	Cut3	SMC4A	SMC-4, DPY-21**	Smc4	SMC4
	ν-SMC	Smc2	Cut14	SMC2A/B	MIX-1	Smc2	SMC2
Condensin I — γ-kleisin		Brn1	Cnd2	CAP-H	DPY-26	Barren	CAP-H
Condensin I — HEAT-IA		Ycs4	Cnd1	CAP-D2	DPY-28	Cap-D2	CAP-D2
Condensin I — HEAT-IB		Ycg1	Cnd3	CAP-G	CAPG-1	Cap-G	CAP-G
Condensin II — β-kleisin		-	-	CAP-H2	KLE-2	Cap-H2	CAP-H2
Condensin II — HEAT-IIA		-	-	CAP-D3	HCP-6	Cap-D3	CAP-D3
Condensin II — HEAT-IIB		-	-	CAP-G2	CAPG-2	-	CAP-G2
Regulators	Cyclin-dependent kinase	Cdc28	Cdc2	-	-	-	CDK1
	Aurora B kinase	Ipl1	Ark1	-	AIR-2	Aurora B	Aurora B
	Polo-like kinase	Cdc5	-	-	-	-	PLK1
SMC5/SMC6							
Subunits	κ-SMC	Smc5	Smc5	SMC5	SMC-5	Smc5	SMC5
	ν-SMC	Smc6	Smc6	SMC6A/B	SMC-6	Smc6	SMC6
	Kleisin	Nse4	Nse4	NSE4A/B	-	Nse4	NSE4A, NSE4B
	Tandem-WHD E3 ligase	Nse1	Nse1	NSE1	-	Nse1	NSE1
	Tandem-WHD	Nse3	Nse3	NSE3	-	Mage	MAGE-G1
	Sumo ligase	Mms21	Nse2	NSE2	-	Quijote, Cervante	NSE2
Regulators	Recruitment	Nse5	Nse5	-	-	-	-
		Kre29	Nse6	-	-	-	SLF2
		Rtt107?	Brc1?	-	-	-	SLF1

Subunit composition of the three homolog eukaryotic SMC complexes, cohesin, condensin and the SMC5/SMC6 complex in model organisms. At the top right corner, components of the human cohesin complex are framed in red. The three core components of the human mitotic cohesin complex, which are studied in the present work, are underlined in black. Meiotic variant subunits are indicated by an asterisk (*). Two asterisks (**) indicate a C. elegans specific condensin subunit involved in dosage compensation. This table was adapted from Haering et Gruber, 2016.

III.   The cohesin complex: composition and architecture

Cohesin is the third eukaryote-specific SMC complex that is required for genome organization and stability. As for the condensin and the SMC5-SMC6 complexes, cohesin can control the DNA topology through DNA tethering and looping. However, the cohesin complex is unique in its central role in the cohesion of sister chromatids of dividing eukaryotic cells. The budding yeast *S. cerevisiae* cohesin was first identified as being essential for the cohesion of replicated sister chromatids and to prevent their premature segregation (Michaelis et al., 1997), but cohesin has since been described in most eukaryotes, including human (Peters et al., 2008).

Interestingly, however, important differences in the regulatory mechanisms have emerged throughout evolution, especially between metazoans and lower eukaryotes such as fungi. Notably, gene duplications and additional regulatory subunits have been identified in metazoans (Ladurner et al., 2016; Nishiyama et al., 2010; Rankin et al., 2005; Rubio et al., 2008). This raises important questions on whether the differences between lower and higher eukaryotes extend beyond the regulatory subunits, into the core components of the cohesin complex, and what the consequences at a molecular level could be for the functional roles of cohesin. A thorough examination of the cohesin complex composition and three-dimensional organization is therefore essential, as its architecture underlies the important functions of cohesin.

The current literature allows to outline a general description of the cohesin complex, its specialized functions, and mechanisms, which we will address in the following chapter. The differences observed between species will be highlighted when significant. For more clarity,

cohesin core and regulatory subunits (Table 1) will be named after the human protein designations approved by the HGNC (Human Genome Organization (HUGO) Gene Nomenclature Committee). When differing from the human nomenclature, the protein names from other species will be indicated in superscript.

A. Composition of the core cohesin complex

1. SMC1A, SMC3 and RAD21

The core of the cohesin complex is composed of three main subunits: the two SMC proteins SMC1 and SMC3, and the Kleisin RAD21. In humans, SMC1A is a 1233 amino acid protein located on chromosome Xp21.1 (Rocques et al., 1995), and SMC3 is a 1217 amino acid protein located to 10q25.2. Originally, SMC3 was isolated as ''Bamacan'', a proteoglycan protein found to bind chondroitin sulfate and to be exported to the basal membrane of mammalian cells (Ghiselli et al., 1999; Wu and Couchman, 1997). However, this function of SMC3 remains unexplored and poorly understood.

SMC1 and SM3 associate through their hinge domains to form the characteristic V-shaped SMC-SMC dimer. SMC1 and SMC3 also transiently dimerize through their ATPase domains upon ATP binding, thereby forming the composite ATPase module of the cohesin complex.

The human RAD21 Kleisin subunit is a 631 amino acids protein, encoded by a gene located to chromosome 8q24.11. RAD21 binds asymmetrically to SMC3 and SMC1 ATPase domains via its N- and C-termini, respectively, and bridges the two heads to close the tripartite ring-like structure required to entrap DNA (Haering et al., 2002; Schleiffer et al., 2003). The ring structure thus provides three "gate" interfaces, of which two have the potential to act as a DNA entry gate into and/or exit gate from the cohesin ring lumen: the SMC1-SMC3 hinge dimerization interface which is thought to act as the entry gate (Buheitel and Stemmann, 2013; Gruber et al., 2006), and the SMC-RAD21 interface that is identified as the exit gate (Beckouët et al., 2016) (Figure 6).

In vertebrates, including human, three versions of the core cohesin complex exist that contain different SMC1 and RAD21 isoforms. The SMC1A-SMC3-RAD21 cohesin complex, which is specific to mitosis, and the SMC1B-SMC3-REC8 or SMC1B-SMC3-RAD21L complexes, which are dedicated to meiotic processes (Ishiguro, 2019; Parisi et al., 1999; Revenkova et al., 2001). All descriptions made in this chapter for vertebrate cohesin depict the mitotic complex containing SMC1A and RAD21, unless stated otherwise.

2. Regulators and auxiliary subunits

Additional regulatory subunits bind to the core cohesin complex, and dynamically regulate cohesin's activities in chromatin dynamics and organization (Figure 5, Figure 10, Table 1) (Haering and Gruber, 2016). In most eukaryotes, cohesin regulators include at least the HAWK proteins STAG, NIPBL and PDS5, the acetyltransferase ESCO, the deacetylase HDAC8 and the accessory subunit WAPL. Many cohesin features and mechanisms are thought to be well conserved from lower to higher eukaryotes. However, evolution has often led to gene duplications in higher eukaryotes for many of the cohesin core subunits and regulators, and to the emergence of new regulatory subunits. This is notably the case in most metazoans, especially in vertebrates, where as a consequence, duplicated proteins (STAG1 and STAG2, ESCO1 and ESCO2, PDS5A and PDS5B) and new regulatory subunits (SORORIN, CTCF) can support extended functions and additional or alternative regulatory pathways specific to vertebrate cohesin.

a) STAG

The regulatory HAWK subunit Stromal Antigen (SA or STAG) is a major cohesin regulatory subunit. It is often considered as the fourth subunit of the core cohesin complex, as the corresponding yeast ortholog STAG^{Scc3} was seen to co-purify with the other three core subunits in a stoichiometric way and to form a stable complex with them (Toth et al., 1999). Furthermore, unlike other regulatory elements of cohesin, STAG orthologs are present in all eukaryotes (Roig

et al., 2014). STAG binds to the central disordered region of RAD21 and also binds directly to DNA (Li et al., 2018; Shi et al., 2020), thus acting as an anchor for cohesin onto DNA.

In vertebrate cells are found two mitotic STAG variants STAG1 and STAG2, and a meiotic variant STAG3 (Arruda et al., 2020; Cuadrado and Losada, 2020; Losada et al., 2000; van der Lelij et al., 2017). In the mitotic cohesin complex, STAG1 and STAG2 are both mutually exclusive, having redundant as well as distinct effects on cohesin localization onto chromatin. Notably, while STAG1 and STAG2 are both required for cohesin loading, stabilization and unloading from chromatin, they are also functionally different, as STAG2 is key for cohesion at the chromosome centromere, whereas STAG1 is essential for cohesion at the telomeres (Canudas et al., 2007; Canudas and Smith, 2009). In human cells, STAG1 binds directly to TRF1 and TRF2, two DNA-binding proteins that are required for telomere replication and that participate in shielding of the linear chromosome ends.

Moreover, STAG1 and STAG2 have distinct contributions to development and definition of cell identity through the DNA looping by cohesin. The stable STAG1-bound cohesin complex is abundant at TAD boundaries, whereas the STAG2-bound cohesin is mainly found at dynamic intra-TAD regulatory loop borders, as it displays a more transient residency time on DNA (Cuadrado et al., 2019; van der Lelij et al., 2017).

b) NIPBL and MAU2

The HAWK NIPBL[Scc2] together with its molecular partner MAU2[Scc4], form a heterodimeric loading complex (also called kollerin, from the greek *kollao*, to glue) which mediates cohesin loading onto DNA (Bermudez et al., 2012; Ciosk et al., 2000; Watrin et al., 2006). MAU2 participates in cohesin initial association with chromatin in interphase, through associating with the unstructured N-terminal domain of NIPBL (Watrin et al., 2006). NIPBL in turn recruits the cohesin complex by binding to an unstructured N-terminal fragment of RAD21, downstream and in proximity to the PDS5 binding site. Additionally, NIPBL interacts with STAG, with the DNA molecule, and binds directly to the SMC1A and SMC3 ATPase coiled coils and ATPase heads (Shi et al., 2020; Higashi et al., 2020; Collier et al., 2020). In collaboration with DNA, NIPBL brings

SMC1A and SMC3 ATPase heads into close spatial proximity and promotes their engagement in the presence of ATP, thereby presumably promoting and enhancing the cohesin ATPase activity (Collier et al., 2020; Higashi et al., 2020; Shi et al., 2020; Kim et al., 2019; Davidson et al., 2019).

c) PDS5

The HAWK PDS5 (Precocious Dissociation of Sisters 5) is a versatile accessory subunit of the cohesin complex, which has a dual role in regulating cohesin association with chromatin, depending on the other regulatory subunits it recruits to the complex (Carretero et al., 2013; Lee et al., 2016; Zhang et al., 2021). On one hand, it promotes cohesin stable association with chromatin, by recruiting other regulators such as the ESCO acetyltransferase (Chan et al., 2013), the transcription factor CTCF in metazoans, and the cohesin auxiliary subunit SORORIN in vertebrates. On the other hand, PDS5 promotes cohesin removal from DNA when recruiting the releasing factor WAPL, which competes with SORORIN for binding to PDS5 (Nishiyama et al., 2010; Ouyang et al., 2013). PDS5 is recruited to the cohesin complex via interaction with an approximately 20 amino acid region of RAD21, which locates nearly immediately after the N-terminal helix that contributes to the DNA exit gate at the SMC3-RAD21 interface. The PDS5 binding site is located upstream and in proximity to the NIPBL binding site, thus making both HAWKs mutually exclusive in binding to the core complex.

Two PDS5 paralogs exist in metazoans: PDS5A and PDS5B (Losada et al., 2005; Sumara et al., 2000). Both versions of PDS5 have overlapping and distinct roles in cohesin stabilization and removal from DNA. Although the differences in functional roles of PDS5A and PDSB remain poorly understood, examples of their distinct roles have been reported in the literature. For instance, unlike PDS5A, PDS5B is specifically required for cohesion at the chromosome centromere (Carretero et al., 2013) and is also able to directly interact with BRCA2 and RAD51, which are key players in the repair of DNA damage (Couturier et al., 2016; Morales et al., 2020). Interestingly, while PDS5A and PDS5B share an N-terminal and central regions, PDS5B harbors an insertion at its C-terminus that is absent from PDS5A. It is an interesting possibility that this additional region could account for the differences in functions between both PDS5 paralogs (Zhang et al., 2021).

d)   WAPL

The WAPL (wings apart-like) protein dynamically controls the association of cohesin with chromatin (Kueng et al., 2006). By being recruited to the complex through binding to PDS5 N-terminal domain, WAPL drives the release of cohesin from chromatin by dissociating the SMC3-RAD21 interface. It thereby opens a DNA exit gate, through which the DNA molecule escapes from the cohesin ring (Beckouët et al., 2016). However, the mechanism by which WAPL operates is yet to be unraveled.

e)   ESCO and HDAC8

The ESCO acetyltransferase participates in the stable association of cohesin onto DNA by acetylating two conserved lysines present in the SMC3 ATPase domain, namely K105 and K106 in human (Ajam et al., 2020; Chao et al., 2017; Kouznetsova et al., 2016; Ladurner et al., 2014; Zhang et al., 2008). SMC3 acetylation by ESCO inhibits DNA-dependent ATPase activation and hampers the DNA translocase activity of cohesin (Dauban et al., 2020; Murayama and Uhlmann, 2015).

In metazoans, a gene duplication led to the emergence of two variants, ESCO1 and ESCO2, which target cohesin in different functional contexts during the cell cycle. While both variants contribute to cohesin stabilization onto DNA, unlike ESCO2, ESCO1 contains a PDS5 binding domain at its N-terminus (Minamino et al., 2015), acetylates cohesin throughout the cell cycle and promotes cohesin role in shaping the chromatin structure (Alomer et al., 2017). ESCO2, however, mostly acetylates cohesin during the S phase and is essential for the cohesion establishment between sister chromatids (Alomer et al., 2017). At the end of the cell cycle, and once cohesin is removed from chromatin, the two acetylated SMC3 lysines are deacetylated by the histone deacetylase HDAC8, thereby allowing the recycling of the cohesin subunits.

In addition to gene duplications, new cohesin auxiliary and regulatory subunits have emerged in higher eukaryotes throughout evolution.

Notably, in most metazoans, the CCCTC-binding transcription factor CTCF mediates the cohesin anchoring onto DNA and the formation of TAD borders, by binding to the STAG and PDS5 subunits (Filippova et al., 1996; Li et al., 2020; Nora et al., 2020; Pugacheva et al., 2020; Rubio et al., 2008; Wutz et al., 2017).

Additionally, the SORORIN subunit (named after the latin word *soror*, sister) is present exclusively in vertebrate cells, where it is required for cohesin stable association with chromatin. SORORIN antagonizes the effect of WAPL, by competing for a shared binding region onto the PDS5 N-terminal domain. SORORIN thus prevents the opening of the DNA exit gate at the SMC3-RAD21 interface, although its mode of action remains unknown (Ladurner et al., 2014; Ouyang et al., 2013; Peters and Nishiyama, 2012; Rankin et al., 2005).

All cohesin accessory subunits thus dynamically interact with the core complex and with chromatin, thereby assisting cohesin in its diverse functional roles.

A. Global architecture of the cohesin complex

The topology and the first structures of the core cohesin complex have been characterized primarily in yeast and orthologous SMC complexes (Gligoris et al., 2014; Haering et al., 2004, 2002). However, electron microscopy (EM) imaging analyses have shown that all eukaryotic cohesin complexes might adopt a similar global architecture (Anderson et al., 2002; Bürmann et al., 2019; Collier et al., 2020; Higashi et al., 2020; Hons et al., 2016; Huis in 't Veld et al., 2014; Shi et al., 2020) (Figure 6). Specifically, cohesin can adopt a variety of conformations due to the inherent flexibility conferred by the SMC1A and SMC3 coiled coils. Among these conformations, cohesin has been reported to adopt an open ring conformation, where SMC1 and SMC3 coiled coils are away from each other, as well as a closed rod conformation, where the coiled coils are

brought together (Skibbens, 2019). Additionally, in both the ring and the rod conformations, the coiled coils from SMC1A and SMC3 have been shown to be able to fold at their elbow regions into additional kinked conformations, where the hinge domain is brought to the vicinity of the ATPase heads (Shi et al., 2020; Xiang and Koshland, 2021) (Figure 6).

The cohesin regulatory and auxiliary subunits appear to play a role in modulating cohesin's architecture and opening of the DNA entry and/or exit gates, with implications for the different functions of the complex and for its transactions with DNA. However, the mode of interaction of cohesin with most of its regulatory subunits and the dynamics of these interactions are still unknown or poorly characterized. Moreover, the physiological significance and the functional relevance of each one of the cohesin conformations observed experimentally still remain to be linked to each step of the molecular mechanisms underlying the cohesin functions.
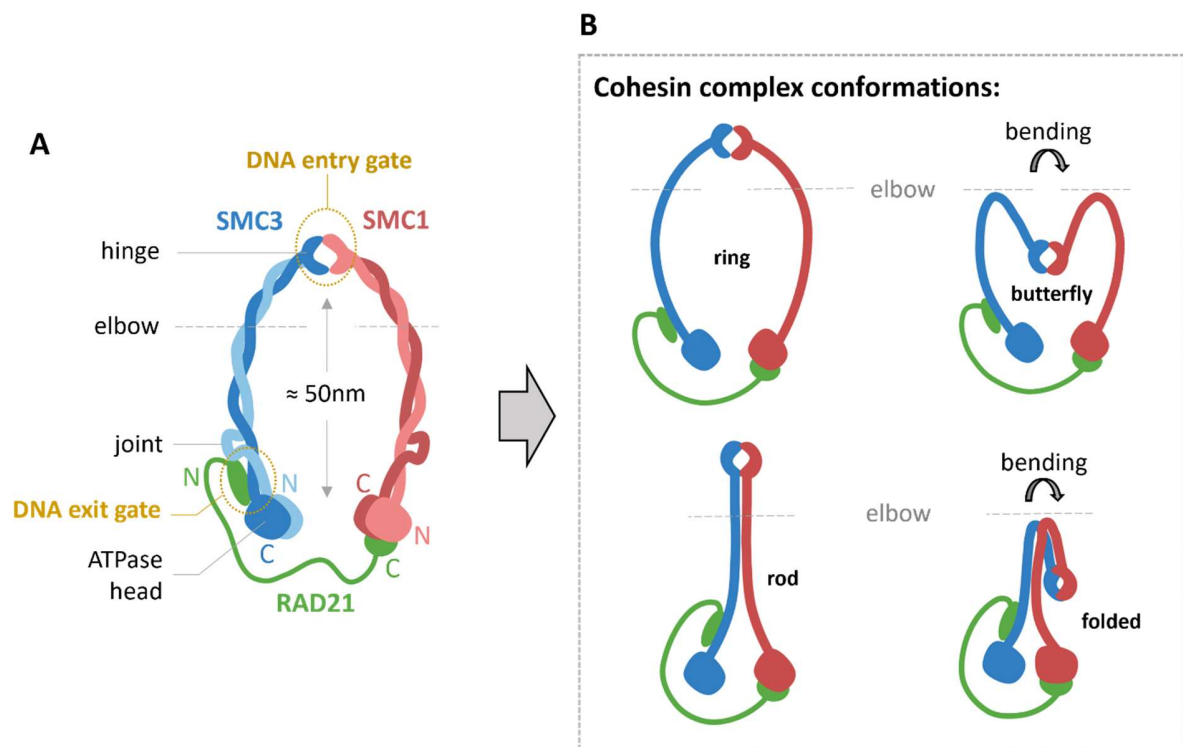
Figure 6: Overall molecular architecture of the cohesin complex. (A) Architectural feautures of the cohesin complex. (B) Different conformations that can be adopted by the cohesin complex, due to the flexibility conferred by the coiled-coils.

IV.   Functional roles of cohesin

The cohesin complex was first identified for its central role in the establishment and maintenance of sister chromatids in eukaryotic dividing cells. Cohesin is now known as participating in numerous other cell events that are essential for chromatin architecture and dynamics. Notably, cohesin has key roles in DNA damage repair, in chromatin three-dimensional organization, in the regulation of gene expression and, in vertebrates, in the recombination processes of immunoglobulin genes (Birkenbihl and Subramani, 1992; Eser et al., 2017; Michaelis et al., 1997; Mizuguchi et al., 2014; Nora et al., 2012; Parelho et al., 2008; Peters, 2021; Szabo et al., 2019; Vian et al., 2018) (Figure 7).
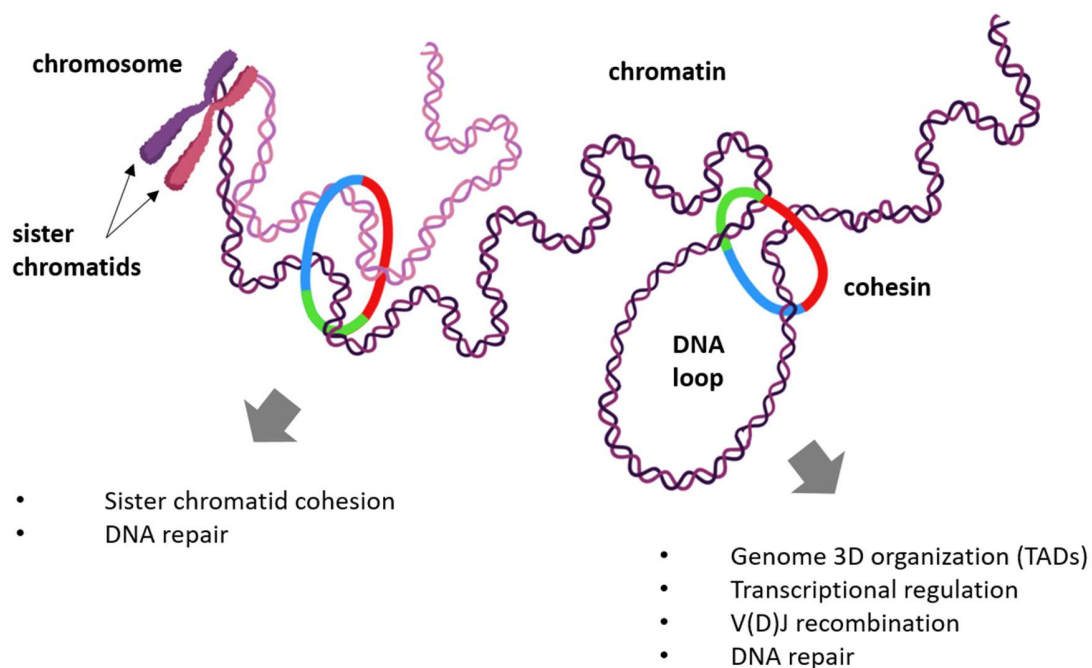
Figure 7: Key cellular roles of the cohesin complex. Cohesin interacts with the DNA molecule by either topologically embracing sister chromatids inside its ring lumen, or in a pseudo-topological manner where it can extrude large loops of DNA. Maintenance of sister chromatids into close spatial proximity allows sister chromatid cohesion during the cell cycle, and sister chromatid cohesion-mediated DNA repair processes. Loops of DNA formed by cohesin serve many cellular processes, including genome organization, gene expression regulation, immunoglobulin recombination and, in DNA repair.

A. Role in sister chromatid cohesion

Inside a dividing cell, the genomic DNA is duplicated during the S-phase by the replication process. The duplicated genetic material is further condensed into a pair of sister chromatids, which are distributed into each daughter cell through cell division. Sister chromatids must therefore be paired and held closely together, in order to be properly bioriented on the mitotic spindle. The pulling force of microtubules from opposite cell poles will distribute one sister chromatid into each daughter cell. To avoid premature separation, sister chromatids need a cohesive force that holds them together.

In eukaryotes, sister chromatid cohesion is fulfilled by the cohesin complex. The cohesin ring mediates sister chromatids cohesion by encircling the two sister DNAs inside its lumen (Gruber et al., 2003; Haering et al., 2008; Ivanov and Nasmyth, 2005). Sister chromatid cohesion is closely associated to DNA replication during the S-phase, although *de novo* cohesion can also be established during G2, in association with DNA-damage repair processes.

Cohesin therefore establishes and maintains sister chromatid cohesion from the S-phase through mitosis and is only removed once the sister chromatids are securely attached to the bipolar spindle and correctly bioriented. Cohesion is consequently crucial for faithful chromosome segregation and the distribution of a copy of the genetic information into each daughter cell.

In higher eukaryotes, at the onset of mitosis, a consequent pool of cohesin is removed from the chromosome arms and only a small subset persists at centromeres, which maintains the sister chromatids in proximity until the metaphase-to-anaphase transition. Cohesin from chromosome arms is replaced by the condensin complex, which takes over the structuration and compaction of the mitotic chromosome arms. In yeast however, there is a significant difference: cohesin is responsible for most of chromosome arms condensation from the S-phase through the entire mitosis process, while condensin is mostly present at specific locus such as the pericentromeric region (Schalbetter et al., 2017).

B. Role in DNA repair

Cells from all living organisms are continuously exposed to DNA damage, either provoked by exogenous agents (ionizing radiation, genotoxic chemical agents) or endogenous factors (including reactive oxygen species that can be by-products of metabolic processes). When damage occurs on DNA, the cell must immediately respond by efficiently detecting, signaling, and initiating the repair of that lesion. All living organisms have evolved a DNA damage response, which makes use of intricate signaling and repair pathways, to counteract the deleterious lesions of DNA. To prevent the transmission of damaged genetic information to the next cell generation, the cell cycle is temporarily halted by cellular DNA damage checkpoints until the damage is repaired. If left unrepaired, DNA lesions can lead to missense mutations up to chromosome loss and chromosomal aberrations, which ultimately contribute to the emergence of genetic diseases such as cancer, or lead to cell death (Mehta and Haber, 2014).

DNA double stand breaks (DSBs) are among the major lesions that have significant cytotoxic effects. Eukaryotic cells have evolved two main DSB repair pathways: non-homologous end-joining (NHEJ) and homologous recombination (HR).

1. Non-homologous end-joining (NHEJ)

The NHEJ pathway is active throughout the cell cycle but is predominant during the G1 phase, when chromatids are yet un-replicated. The NHEJ pathway rejoins two free DNA ends derived from a DNA DSB, independently of any sequence homology. By directly rejoining two free DNA ends with minimal DNA strand processing, this DNA repair route is prone to errors and can possibly result in unwanted mutations, insertions and deletions in the genome and therefore lead to loss of genetic information (Weterings and van Gent, 2004).

2. Homologous recombination (HR)

The HR pathway, in contrast to NHEJ, uses a homologous and unaltered DNA template, such as the sister chromatid, to repair a DNA DSB. This process requires more extensive processing of the broken DNA end to faithfully restore the damaged loci. It ensures thereby that no loss of genetic information occurs (Jasin and Rothstein, 2013; Wyman and Kanaar, 2006). HR is a major DNA repair pathway in mammalian cells and is predominantly operational in S- and G2-phases, during which the newly synthesized sister chromatid becomes available to be used as repair template.

3. Contribution of the cohesin complex to HR

During DNA damage repair by HR, the broken DNA helix is unwound and brought into close spatial proximity to the homologous template. For this to happen, the two sister chromatids must be held closely together. This prerequisite is ensured by the cohesion of sister chromatids that is established by the cohesin complex.

The implication of cohesin in the cellular response to DNA damage was first suggested by Birkenbihl and Subramani, 1992, with the identification of the *rad21* gene from *S. pombe* as being involved in DNA DSB repair and whose mutations caused cell sensitivity to radiation. Additionally, other cohesin subunits have since then been identified as necessary for the DNA DSB response, including subunits that are required for proper sister chromatids cohesion. Notably, SMC1 and SMC3 are phosphorylated by the S-phase checkpoint protein kinase Ataxia Telangiectasia Mutated (ATM). SMC1 and SMC3 phosphorylation thus contributes to regulation of the cell cycle checkpoint in response to DNA damage, through the mobilization of cohesin at DSB sites (Bauerschmidt et al., 2011; Kim, 2002; Yazdi, 2002). Moreover, the Cohesin regulator PDS5B has been shown to be required for proper DNA repair by HR, by physically interacting with the tumor

suppressor BRCA2 (Morales et al., 2020) and with the RAD51 recombinase (Couturier et al., 2016), thereby promoting RAD51-mediated invasion of the DNA repair template.

Interestingly, it was shown that cohesin also promotes DSB repair throughout the cell cycle with a mechanism distinct from its cohesive activity of sister chromatids. Notably, in human cells, cohesin was shown to mediate repression of transcription at DSBs during interphase (Meisenberg et al., 2019). Additionally, cohesin-dependent genome organization within TADs, by formation of loops of DNA, was recently shown to have extended functions in the response to DNA damage, by promoting the formation of DNA damage repair foci (Arnould et al., 2021).

C. Role in 3D genome organization

The three-dimensional folding of eukaryotic interphase chromatin is an extremely dynamic process. It ensures proper genome organization and compaction, while granting access to the underlying DNA molecule, for the fulfillment of vital biological processes. Chromosome conformation capture methods, including Hi-C, have allowed the identification of long-range chromatin *cis*-interactions mediated by large loops of DNA (Dixon et al., 2012; Lieberman-Aiden et al., 2009; Nora et al., 2012), which have been termed TADs by seminal studies in vertebrate cells (Dixon et al., 2012; Nora et al., 2012).

Cohesin is an essential factor to TAD formation in vertebrate cells. The complex uses its ability to extrude large loops of DNA, until it encounters two convergently-oriented CTCF proteins (Li et al., 2020). CTCF mediates the arrest of loop expansion by cohesin, thus acting together with cohesin as TAD insulators, as they stabilize the loops at their boundaries (Hansen, 2020; Pugacheva et al., 2020).

Cohesin-dependent chromatin-organizing loops have also been identified in *S. cerevisiae* and *S. pombe* yeasts, but might however be regulated by different mechanisms, since CTCF is absent from lower eukaryotes (Dauban et al., 2020; Eser et al., 2017; Mizuguchi et al., 2014).

D. Role in transcriptional regulation

In addition to genome organization, cohesin is involved in transcriptional regulation (Perea-Resa et al., 2020). Cohesin implication in transcriptional regulation has been observed in a variety of species, from vertebrates to drosophila and yeasts, but appear to be managed differently depending on the organism (Dheur et al., 2011; Dorsett, 2007; Dorsett and Merkenschlager, 2013; Dorsett and Ström, 2012; Gullerova and Proudfoot, 2008; Lin et al., 2011; Peric-Hupkes and van Steensel, 2008).

In vertebrates, cohesin-dependent transcriptional regulation is carried out by the internal organization of the CTCF anchored TADs and the formation of smaller loops within the TADs. Those cohesin-dependent sub-loops are extremely dynamic and allow the regulation of gene expression by bringing together distant enhancers and promoters of transcriptionally active genes. The transcription-regulating sub-loops can also be anchored by other transcription factors that interact with cohesin such as the Yin and Yang 1 factor (YY1), which promotes enhancer-promoter interactions (Weintraub et al., 2017).

In yeasts however, cohesin influences the transcription initiation, elongation and termination processes mediated by the RNA polymerase (Perea-Resa et al., 2020; Peric-Hupkes and van Steensel, 2008). The precise underlying mechanisms are not yet fully characterized, and it is not yet known to what extent the cohesin mechanisms in transcriptional regulation are conserved in different species.

E. Role in recombination

In vertebrates, T-cell receptors and immunoglobulins produced by B-cells are composed by different domains, whose assembly is directed by specific locus rearrangements. Gene recombination consequently leads to a large diversity of T-cell receptors and immunoglobulins, which are necessary to recognize and protect the organism against a plethora of exogenous

antigens including allergens, bacteria, viruses, and parasites. Vertebrate cohesin is required for B- and T-cell recombination processes, through forming the loops of DNA that bring together distal loci for recombination (Ba et al., 2020; Peters, 2021).

Cohesin functional roles are supported by and regulated through distinct molecular mechanisms, which will be described in the following chapter.

V. Molecular mechanisms and regulation of the cohesin functions

A. Cohesin loading onto and interaction with DNA

All functional roles of the cohesin complex depend on its ability to establish topological as well as non-topological links between two intra- or inter-molecular DNA segments, by bringing and holding them together in close spatial proximity (Figure 7). To perform these functions, cohesin must initially interact with and be loaded onto DNA.

1. Loading onto DNA by NIPBL-MAU2

Cohesin dynamically associates to and dissociates from the chromatin in a way that is dependent on the cell cycle. In metazoans, cohesin is loaded onto chromatin as early as at the end of telophase, whereas in budding yeast this happens in the G1 phase, before DNA replication (Haarhuis et al., 2014).

Cohesin is loaded onto DNA by the loading factor NIPBL and its molecular partner MAU2 (Figure 8) (Bermudez et al., 2012) by a mechanism that is dependent on ATP binding and hydrolysis by the cohesin ATPase module (Arumugam et al., 2003; Gutierrez-Escribano et al., 2020). The cohesin complex is recruited at promoters of active genes, at chromosome centromeres, at DNA replication forks, and at sites of DNA damage (Bot et al., 2017; Kagey et al.,

2010; Muñoz et al., 2020). In budding yeast, it was also shown that the chromatin remodeler RSC (Remodels Structure of Chromatin) complex plays a role in cohesin loading onto chromosomes (Lopez-Serra et al., 2014; Muñoz et al., 2020) by acting as a receptor of the cohesin loader complex NIPBL^{Scc2}-MAU2^{Scc4} and by providing a region depleted of nucleosomes for the loading of cohesin. In humans, NIPBL has been shown to be enriched at active promoters and enhancers (Hua et al., 2021), although the precise loading requisites remain less clear.
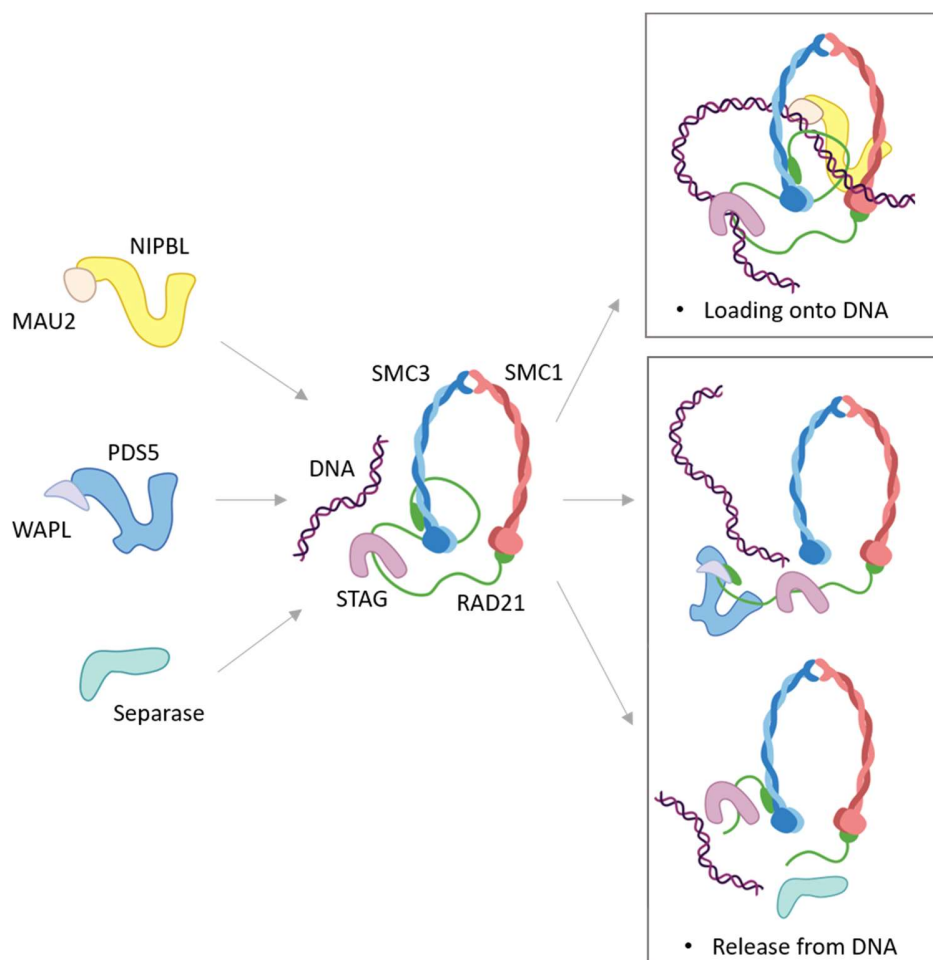
Figure 8: Schematic representation of the cohesin loading onto and release from DNA. Cohesin is loaded onto DNA by the loader NIPBL and its chromatin-binding molecular partner MAU2. Cohesin is released from DNA either during the prophase pathway (vetebrates) by the destabilizing action of PDS5-WAPL on the DNA exit-gate, or during anaphase where RAD21 is cleaved by Separase to release the DNA molecule from the ring lumen.

2. Two modes of interaction with DNA: non-topological and topological

The cohesin complex has been shown to interact with both double stranded (ds) and single stranded (ss) DNA, through the hinge in the latter case, and to potentially entrap the sister chromatid during replication by capturing the nascent sister chromatid ssDNA (Murayama and Uhlmann, 2015; Shi et al., 2020). Interestingly, experiments with the budding yeast *S. pombe* Smc3[Psm3] ATPase head showed that it can bind ssDNA with a higher affinity than to dsDNA (Kurokawa and Murayama, 2020), although the interaction and physiological relevance of ssDNA binding by cohesin remains poorly understood or unassessed in other eukaryotes. Cohesin is thought to interact with the dsDNA molecule through two distinct modes, non-topological (or pseudo-topological) and topological, depending on the functional context (Figure 9).

Topological loading onto DNA requires the entry of DNA into the cohesin ring lumen. This leads to the entrapment and the embrace of one or more DNA segments, such as the sister chromatids, inside the cohesin ring (Haering et al., 2008). On another hand, in the non-topological interaction, the DNA molecule is not stably entrapped inside the cohesin ring. Non-topological loading is therefore less stable than the topological loading mode, as it was additionally shown to be sensitive to increasing salt concentrations (Davidson et al., 2019).
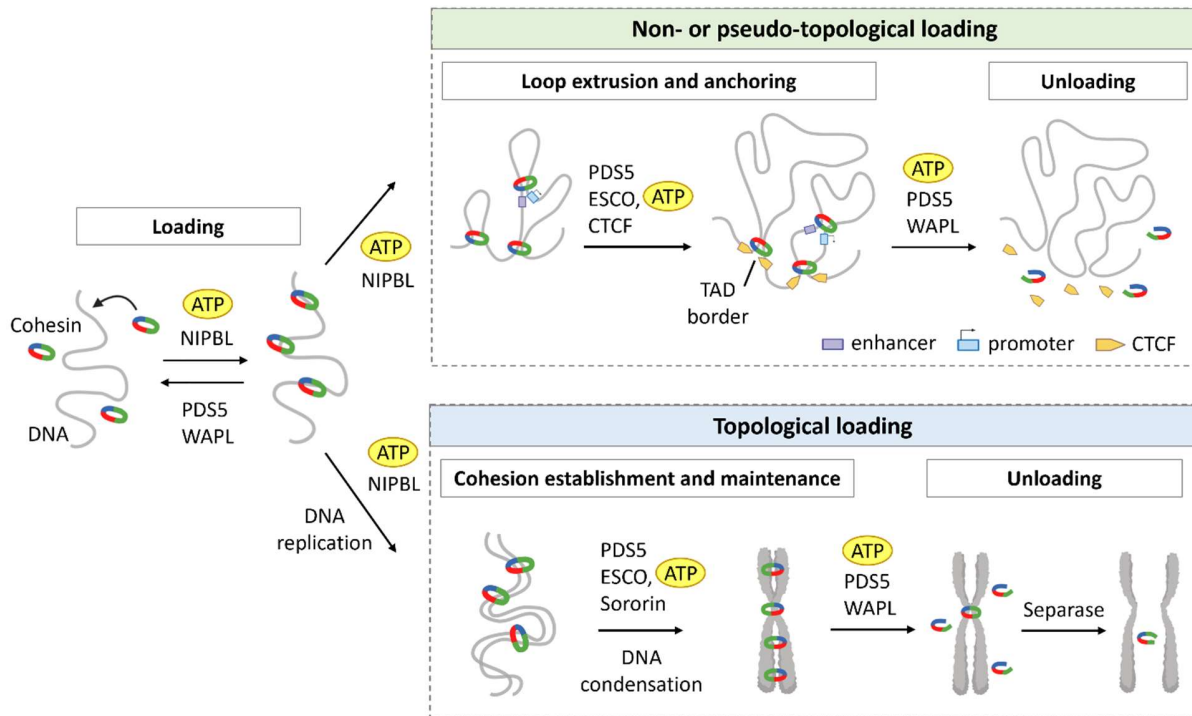
Figure 9: The cycle of vertebrate cohesin interaction with chromatin. Cohesin is first loaded onto DNA. After DNA replication, cohesin establishes the cohesion of sister chromatids. Cohesion is maintained from interphase through mitosis. During interphase, cohesin also froms large loops of DNA, by DNA loop extrusion, thus participating in TADs formation and in gene regulation, by interacting with transcription factors and TAD insulators such as CTCF. In vertebrates, cohesin is released from chromatin in two distinct ways: by opening of the DNA-exit gate at the SMC3-RAD21 interface (prophase pathway), or by proteolytic cleavage of RAD21 by the protease separase (during anaphase). ATP binding and/or hydrolysis by the cohesin ATPase module is essential at most steps of the cycle, by mediating cohesin loading, stabilization, and removal from chromatin, as well as cohesin functions in DNA looping.

3. DNA enters the cohesin ring by topological entrapment

The topological entrapment of two DNA molecules inside the cohesin ring mediates sister chromatid cohesion, which also has implications for DNA damage repair by holding the damaged DNA strand and the repair template in spatial proximity. To be topologically captured by cohesin, DNA must enter the ring by the transient opening of one of the possible entry gates. It has been proposed that DNA enters the ring through the transient opening of the SMC1-SMC3 hinge interface in *S. cerevisiae* and in humans (Buheitel and Stemmann, 2013; Gruber et al., 2006) or of the Smc3-RAD21[Scc1] interface in *S. pombe* (Uhlmann, 2016), although it is not excluded that

the DNA entry gate could differ depending on the organism. Although multiple models for sister chromatid entrapment have been proposed (Haering et al., 2002; Lengronne et al., 2006; Rhodes et al., 2017; Zhang et al., 2008), the detailed molecular mechanisms of DNA entrapment by cohesin remain unknown. It is however known that the entrapment process requires the ATPase activity of cohesin (Weitzer et al., 2003).

B. DNA loop formation

The DNA loop extrusion mechanism accounts for the functional roles of cohesin in chromatin organization, in transcriptional regulation, in V(D)J recombination, and in DNA damage repair by formation of repair foci. Cohesin uses its motor activity to reel in the DNA molecule in a processive manner, which progressively enlarges the loop of DNA (Haarhuis et al., 2014).

The ability of the cohesin complex to extrude loops of DNA was concomitantly confirmed by the works of Jan-Michael Peters and Hongtao Yu teams (Davidson et al., 2019; Kim et al., 2019). By single molecule studies *in vitro*, these works provided the first evidence for the ability of human cohesin to extrude loops of DNA, both on naked DNA and on nucleosome-containing DNA. This process does not require the opening of any of the three interfaces of the cohesin ring, showing that the loop extrusion process requires a non-topological loading of cohesin onto DNA. It was shown that this process is symmetric and bidirectional (Kim et al., 2019) in contrast to the condensin complex which has been shown to extrude loops unidirectionally (Ganji et al., 2018). Although many models for the loop extrusion process have been proposed (Hassler et al., 2018; Higashi et al., 2020; Marko et al., 2019; Nichols and Corces, 2018; Ryu et al., 2020) and even the possibility that two cohesin macromolecules could be involved in the extrusion of one loop (Kim, 2020; Kim et al., 2019), the molecular details of this process remain elusive. However, it was shown that the loop domains formed by cohesin are based on an ATPase dependent process (Vian et al., 2018). Single molecule studies confirmed that cohesin can indeed extrude loops of DNA in an ATP-hydrolysis dependent manner, which is enhanced by the concerted action of

NIPBL-MAU2 and DNA (Davidson et al., 2019; Kim et al., 2019). The loop enlargement process is dynamically regulated by a turnover established with the releasing activity of PDS5 and WAPL (Liu et al., 2021).

C. Cohesin stabilization onto DNA

Two distinct pools of cohesin were observed onto chromatin in the G1-phase of the cell cycle: one pool that is stably bound, and one main pool for which is observed a dynamic exchange, called cohesin turnover. It was shown by (Gerlich et al., 2006) using live-cell experiments from rat cells, that after the S-phase of the cell cycle, about one third of the dynamic pool became more stably bound to chromatin until the onset of mitosis. This is explained by the fact that once loaded onto chromatin, stable association with the DNA molecule is required for the durable maintenance of sister chromatid cohesion from interphase to anaphase, and sometimes for longer periods of time up to several years in oocytes in case of the meiotic cohesin complex (Cheng and Liu, 2017; Tsutsumi et al., 2014; Wutz et al., 2020), and for the maintenance of DNA loops. Cohesin is therefore stabilized onto DNA by two processes, through sister chromatid cohesion and through DNA loop anchoring, both dependent on the replacement of NIPBL by PDS5 and the acetylation of the SMC3 ATPase by ESCO.

1. Maintenance of sister chromatid cohesion

After establishing the cohesion of sister chromatids, cohesin stabilization onto chromatin for cohesion maintenance depends on PDS5 (Panizza et al., 2000), on ESCO, and on the vertebrate subunit SORORIN. As soon as the sister chromatids are entrapped into the cohesin ring, the conserved lysine pair of at the SMC3 ATPase (K105 and K106 in humans) is acetylated, mainly by the ESCO2 acetyltransferase. In vertebrate cells, SMC3 acetylation triggers the recruitment of

SORORIN, which ensures the stable cohesion of sister chromatids, by antagonizing the releasing effect of WAPL (Lafont et al., 2010; Nishiyama et al., 2010; Zhang and Pati, 2012).

SORORIN does not exist in yeasts, and it is not known if another protein performs SORORIN-like functions in those organisms, where Smc3 acetylation by ESCO[Eco1] seems to be sufficient for cohesion establishment and protection against cohesin release from DNA (Rowland et al., 2009).

2. Stability and anchoring of DNA loops

Specifically in vertebrates, during the processes requiring loop extrusion, the composition in accessory subunits of the cohesin complex varies at the loop borders. The stable STAG1-bound cohesin is more abundant at TAD boundaries and can persist for hours, whereas STAG2-cohesin is mainly found at the more dynamic intra-TAD regulatory loop borders and has a more transient residency on DNA (Cuadrado et al., 2019; van der Lelij et al., 2017). STAG1-cohesin is less sensitive to WAPL releasing activity than STAG2-cohesin. This could explain the difference in STAG1/2 composition and the higher stability of cohesin at TAD borders, compared to cohesin at intra-TAD loop borders, although the molecular basis for this difference in stability are not known (Wutz et al., 2020).

The arrest of the DNA loop extrusion process by cohesin is initiated when cohesin encounters two convergent DNA-bound CTCF proteins. By hindering the loop expansion process, CTCF participates to the insulation of TAD borders together with cohesin (Li et al., 2020) (Figure 9). CTCF binds directly to STAG and to PDS5, which replaces NIPBL (Nora et al., 2020; Li et al., 2020). Moreover, PDS5 is also essential to TAD boundary formation (Wutz et al., 2017), as it promotes the SMC3 ATPase acetylation by recruitment of ESCO1 (Chan et al., 2013; Vaur et al., 2012). Cohesin acetylation by ESCO1 at TAD boundaries protects the complex from the DNA releasing activity of WAPL (Wutz et al., 2020). SMC3 acetylation participates in the loop extrusion arrest, supposedly by inhibiting the DNA-dependent ATPase activation (Murayama and Uhlmann,

2015; Wutz et al., 2020), thus stabilizing the cohesin loop. SMC3 acetylation by ESCO has been shown to be dependent on the ATP binding and hydrolysis activity by the cohesin ATPase (Ladurner et al., 2014).

In *S. cerevisiae*, chromatin folding by cohesin-mediated loop extrusion is controlled by PDS5, WAPL and ESCO[Eco1]. In the budding yeast, the latter is sufficient to inhibit the cohesin DNA-dependent translocase activity, by acetylating the Smc3 ATPase head (Dauban et al., 2020).

D. Cohesin removal from DNA

Cohesin establishes and maintains chromosome dynamics and architecture during interphase. In metazoans, once the sister chromatids are securely attached to the bipolar spindle and correctly oriented at the onset of mitosis, cohesin must be gradually removed from the sister chromatids so they can be separated during chromosome segregation.

Dissociation of cohesin from chromosomes happens in two distinct steps, which occur during prophase and during anaphase (Figure 9). Firstly, cohesin is removed from the chromosome arms during prophase, via the "prophase pathway" (Sumara et al., 2000) once the chromosomes are attached to the mitotic spindle. SORORIN and STAG2 are phosphorylated by the Cdk1, Plk1 and Aurora B kinases (Liu et al., 2013; Nishiyama et al., 2013; Zhang et al., 2011). This notably inhibits the association of SORORIN with PDS5, and SORORIN is therefore replaced by the releasing factor WAPL (Nishiyama et al., 2013, 2010). WAPL dissociates the SMC3-RAD21 or DNA exit gate interface, thus opening the cohesin ring and releasing the two sister chromatids (Buheitel and Stemmann, 2013; Eichinger et al., 2013). The prophase pathway keeps each of the cohesin components intact for further reuse, as it does not require any proteolysis event.

However, until the proper segregation by condensin (Hirano, 2012) and the correct biorientation on the metaphase plate, in preparation for chromosome segregation, sister chromatids still need to resist the pulling force of the spindle apparatus and remain in close spatial proximity. A small subset of cohesin remains therefore bound to the chromosomes, at

their centromeres, which gives to the chromosomes their characteristic X shape. Centromeric cohesin is protected by the regulatory subunit Shugoshin 1 (from Japanese "shugoshin", guardian spirit). Shugoshin recruits the phosphatase PP2A to centromeric cohesin, which prevents SORORIN phosphorylation during prophase and avoids its replacement by WAPL (Kitajima et al., 2006). At anaphase onset, once each sister chromatids kinetochores are correctly attached to the microtubules from the spindle apparatus, the anaphase promoting complex (APC) activates the protease Separase, which then specifically cleaves RAD21. The degradation of RAD21 thus opens the cohesin ring and the centromeric sister DNAs are released (Hauf, 2001; Tomonaga, 2000). The pulling force from the mitotic spindle finally separates the sister DNAs and distributes them into the daughter cells.

In yeasts however, there is no prophase pathway and the majority of cohesive cohesin is removed in anaphase by Separase cleavage of Scc1 and its degradation.

Once cohesin is removed from DNA at the end of the cell cycle, SMC3 is deacetylated by the HDAC8 deacetylase (Hos1 in yeasts). SMC3 deacetylation allows the recycling and reusing of cohesin for the next cell division cycle. Howewer, since yeasts do not possess a prophase pathway and yeast cohesin is removed from DNA mainly through RAD21[Scc1] degradation, this signifies that yeast cells must reconstitute a new pool of cohesin in the next cell cycle before they can be loaded onto DNA in the G1 phase, while metazoan cells can reuse their recycled cohesin from the prophase pathway as soon as the telophase ends.

Since the cohesin complex performs vital functions throughout the entire division cycle of eukaryotic cells, it is apparent that the fulfillment of cohesin roles relies on the integrity of the complex, of its ATPase module, and of its regulatory subunits, for which most directly interact with the ATPase (Figure 10). Dysfunctions are deleterious to the finely orchestrated cohesin mechanisms and are involved in the emergence and development of disease in humans.
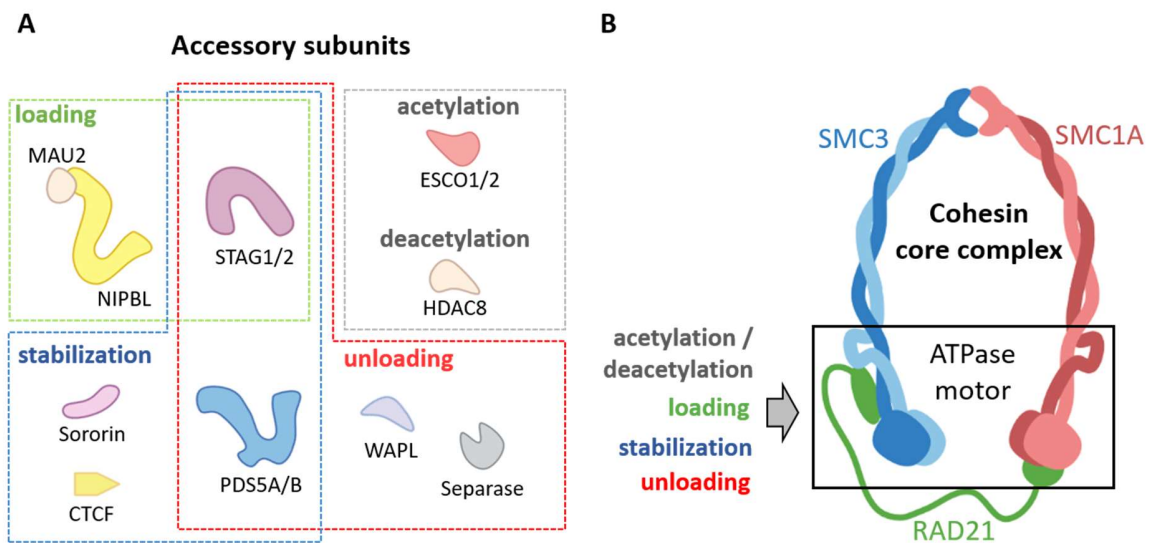
Figure 10: Functional relevance of the cohesin ATPase. (A) Various accessory subunits regulate cohesin mechanisms and functions, through mediating the complex loading, stabilization, and removal from DNA. (B) The accessory subunits all interact directly with or in the vicinity of the cohesin ATPase module.

VI. Cohesin implications in human disease

In humans, it has been shown that mutations in the components of the cohesin complex and its regulators impairs the cohesin mechanisms and normal functions. Cohesin mutations have been shown to be involved in various types of solid and hematologic cancers and in a spectrum of developmental disorders called cohesinopathies, among which the best characterized is the Cornelia de Lange Syndrome (CdLS) (Figure 11).

A. Mutant cohesin in the Cornelia de Lange Syndrome

The Cornelia de Lange syndrome, or Brachmann de Lange Syndrome, was first described by Winfried Brachman in 1916 and Cornelia de Lange in 1933. CdLS is a rare disease which globally affects 1:10,000 to 1:30,000 live births and which impairments range in a spectrum from severe to mild. CdLS is characterized by developmental anomalies, including defective limb morphogenesis, hearth defects, facial dysmorphism, gastroesophageal dysfunctions and growth and mental retardation. Germline mutations in cohesin subunits that lead to CdLS account for most of CdLS cases (Liu and Krantz, 2009). 50 to 60% of CdLS cases present mutations in NIPBL, while mutations on HDAC8, SMC1A, SMC3 and RAD21 account collectively to 15% of CdLS cases (Online Mendelian Inheritance in Man (OMIM) entries #122470, #300590 and #60759, Deardorff et al., 2012; Parenti et al., 2020). CdLS mutations have been shown to affect cohesin binding to DNA (Revenkova et al., 2009), although the sister chromatid cohesion function of cohesin is not thought to be affected (Sarogni et al., 2019). Mutations in the STAG1 and STAG2 subunits are also involved in X-linked cohesinopathies (Mullegama et al., 2019; Soardi et al., 2017).

B. Mutant cohesin in cancer

Somatic mutations in cohesin subunits have been shown to lead to the emergence and the development of several types of solid and hematologic malignancies. Mutations in the cohesin core subunits have been identified with high frequency, occurring in 21% of cases of bladder cancers, 19% of glioblastomas, 10% of melanomas and up to 46% of myeloid malignancies (Figure 11) (Mintzas and Heuser, 2019; Waldman, 2020). STAG2 is the cohesin subunit which is most frequently found mutated in cancer cells (Mintzas and Heuser, 2019). Moreover, upregulated and downregulated expression levels of cohesin core subunits have been identified in numerous types of cancers, including colorectal and prostate cancers, and are linked to cancer prognosis (Liu and Krantz, 2009; Rhodes et al., 2011; Yadav et al., 2019, 2013). The cohesin-interacting CTCF protein is also frequently mutated in various tumor types, and promotes cell proliferation and resistance to chemotherapy in colorectal cancer (Lai et al., 2020).

The pathogenicity of cohesin mutants involved in cancer and cohesinopathies including CdLS are poorly understood, but it is likely that the defects observed stem from a dysfunctional regulation of gene expression, which leads to genome instability (Heimbruch et al., 2021; Hill et al., 2016; Waldman, 2020). Mutations leading to disease are distributed all over cohesin subunits, including the ATPase module (Figure 12). Moreover, additional subunits such as STAG1/2, NIPBL and PDS5A/B bind in proximity to the ATPase and recruit other regulators that dynamically regulate cohesin's ATPase activity, thus modulating cohesin functions in capturing, holding, and translocating DNA segments. Since many cohesin regulators appear to bind to or in the vicinity of the ATPase module, it is crucial to characterize their interactions, to identify the dysfunctions that lead to disease. To this date, there is no targeted therapy to the dysfunctional cohesin complex, as the molecular mechanisms of the ATPase activity that are impaired in disease are poorly understood. Therefore, biochemical, and structural characterization of the human cohesin complex, especially of its ATPase motor, is not only essential for the fundamental understanding of cohesin function, but also for understanding the dysfunctions that lead to disease and for further therapeutic applications.
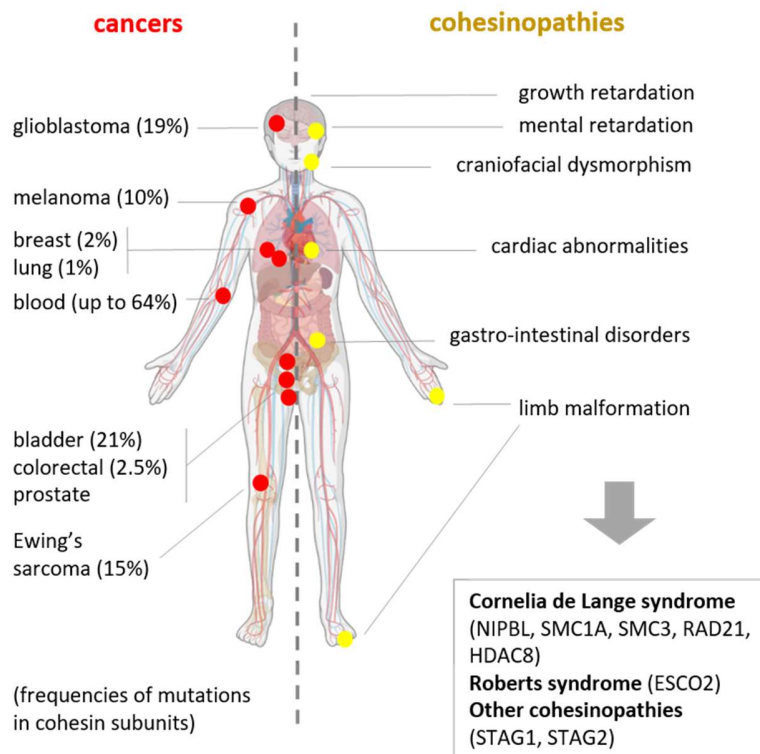
Figure 11: Effects of mutations on cohesin subunits on human health. Mutations on the cohesin subunits are involved in various types of solid and haematologic cancers, and in neurodevelopmental disorders called cohesinopathies. Into parentheses, the frequency of mutations in cohesin subunits that lead to cancer, reviewed by Mintzas et Heuser, 2018.
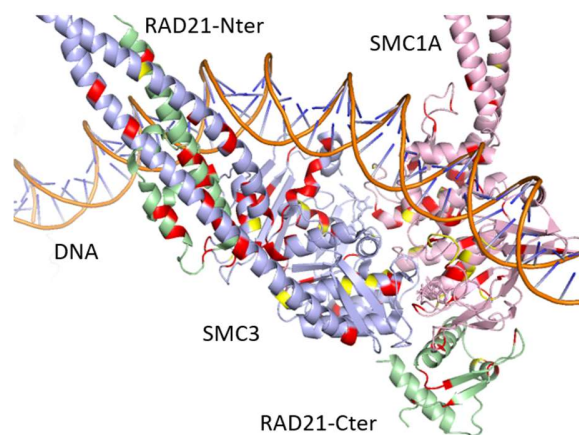
Figure 12: Distribution of cohesin disease mutations on the structure of human cohesin. Cancer somatic mutations (red) and CdLS mutations (yellow) are found overall the Cohesin ATPase (PDB: 6wg3, 6wge), and could be disturbing functionnaly relevant interfaces, such as at the DNA exit-gate or at the DNA-binding interfaces.

VII.   The cohesin ATPase module

As exposed on the previous chapters, it becomes apparent that regulation of cohesin mechanisms is mainly achieved by its ATPase cycle (Figure 9). The ATPase module is thus responsible for most functions of cohesin, and mutations have deleterious consequences that lead to disease. Cohesin cyclic ATPase activity is promoted by NIPBL together with DNA, while it is precluded by SMC3 ATPase head acetylation, and by PDS5 and the regulators that it recruits to remove or to stabilize cohesin onto DNA. To this day, the mechanisms of the cohesin ATPase module remain poorly characterized. However, the cohesin ATPase has been shown to share some sequence and structural features with a superfamily of ATPases, the ABC proteins, which have been extensively studied and characterized (Hirano, 2002). Therefore, in this chapter, we will first outline a general description of the highly conserved features and structural organization of the Nucleotide Binding Domain (NBD) of ABC ATPases, which might be conserved in cohesin, followed by the possible mechanisms that underlie their enzymatic action. This aims to draw an overall picture of the mechanisms that could potentially be preserved, at least in part, in cohesin, and which are to be kept in mind when exploring its ATPase module.

A. The ATPase module of ABC-proteins

1. The ABC-proteins ATPase module organization

SMC proteins ATPase head harbors conserved structural features and motifs that are typically found in ABC proteins, such as ABC transporters (Figure 13) (Hirano, 2002). ABC transporters are molecular transporting systems that channel molecules and ions across the cell membrane. They form one of the largest and likely most ancient protein super-families and are involved in a wide range of biological pathways across all domains of life. ABC transporters are dimeric nucleotide hydrolases, which use the energy of ATP binding and hydrolysis to perform

their mechanical functions. Each monomer composing an ABC ATPase consists in two distinct domains: a Nucleotide Binding Domain (NBD) located on the cytoplasmic side of the cell membrane, which dimerizes with another NBD in a head-to-tail manner, and a Transmembrane Domain (TMD) (Linton and Higgins, 2007; Oswald et al., 2006; Rees et al., 2009).

The NBD is composed of two distinct lobes. The first lobe is a structurally conserved RecA-like catalytic core domain where an ATP binding cassette is found, and the second lobe is defined as an α-helical subdomain that is specific to ABC ATPases. The TMD is composed of a bundle of α-helices that are inserted into the cell membrane bilayer and that can bind a large variety of molecular substrates (Linton and Higgins, 2007; Zolnerciks et al., 2011). In ABC transporters, substrate binding induces conformational changes in the TMD, which are transmitted through coupling helices located at the TMD-NBD interface, to the NBD. Consequently, ATP hydrolysis is triggered, which induces conformational changes both in the NBD and the TMD, and the substrate is translocated from one side of the membrane to the other side (Linton and Higgins, 2007).

ABC protein

SMC protein

ABC protein ATPase dimer

ABC protein active site

72
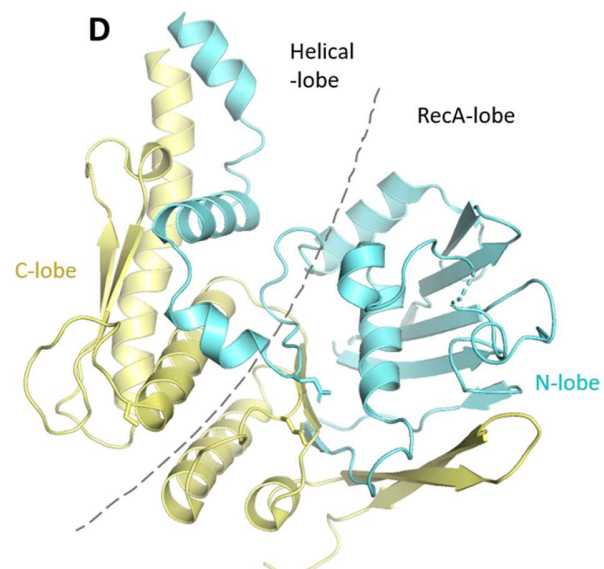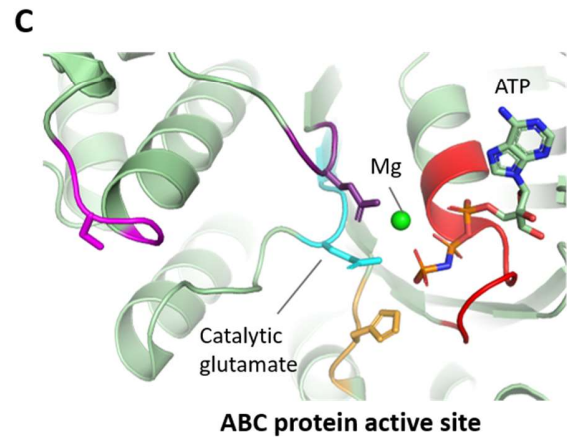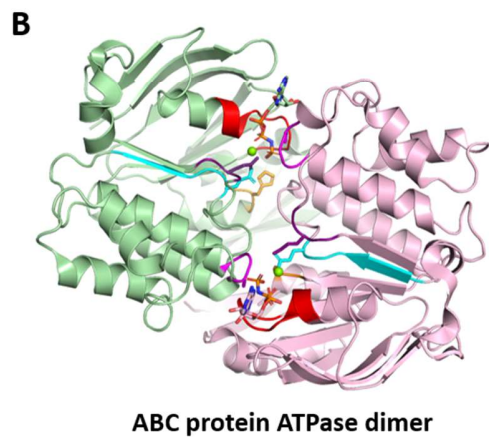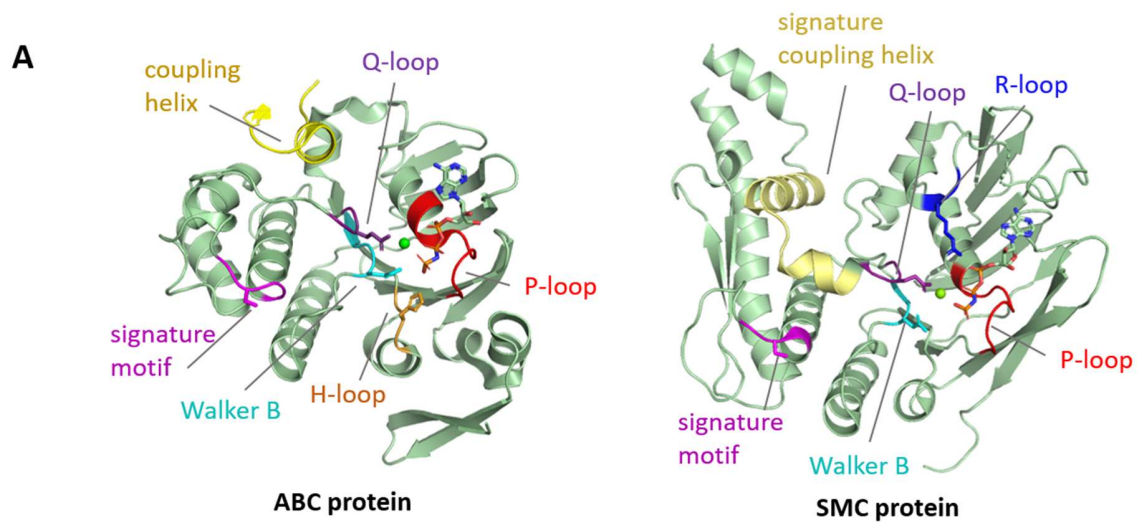
Figure 13: Cohesin ATPase domains resemble that of ATP-binding cassette (ABC) proteins. (A) Comparison of the three-dimensional organization and the motifs harbored by the ATPase domain of the *Escherichia coli* vitamin B12 transporter BtuCD-F (left panel; PDB: 4fi3) and by the ATPase domain of human SMC1A (right panel; PDB: 6wge, 6wg3). (B) Example of the dimerization properties of ABC-proteins. (C) Zoom into the ATP binding site of ABC-proteins, and position of the conserved catalytic glutamate. (D) The ABC-ATPase is built by the N- and C-termini globular domains of the protein, respectively N-lobe and C-lobe. The global architecture is divided into two distinct domains, the RecA-like lobe and the alpha-helical lobe, which can potentially move relatively to each other during the ATPase cycle.

2. How ATP binding at the NBD leads to an active ATPase module

The NBDs Waker A motif is often named P-loop, for phosphate-loop, and has a consensus sequence of GXXXXGK(T/S) (where X is any amino acid) (Figure 13) (Oswald et al., 2016). The P-loop frames a binding pocket that positions the tri-phosphate moiety of the ATP molecule into the RecA-like lobe of the NBD (Hopfner, 2016). The P-loop stabilizes ATP binding into the catalytic site by contacting the oxygen atoms of the β- and γ-phosphates through hydrogen bonding (Hopfner, 2016). Additionally, ATP is bound to a cofactor cation Mg^{2+}, which is required for efficient ATP hydrolysis. Mg^{2+} is coordinated by the negatively charged oxygen atoms from the β- and γ-phosphate groups, and by the serine residue that borders the C-terminal end of the P-loop (Oswald et al., 2006; Hopfner, 2016).

The Walker B motif (consensus φφφDE, where φ is a hydrophobic residue) is another conserved, which is located downstream of the P-loop (Zolnerciks et al., 2011; Hopfner, 2016). Its participation to the ATP and Mg^{2+} binding is not clear. However, the conserved glutamate residue has been shown to act as the main catalytic residue, by activating and polarizing the lytic water molecule towards the γ-phosphate of the ATP (Figure 13, Figure 14) (Hopfner, 2016).

The NBD of ABC ATPases often contain two other motifs that stabilize ATP binding to the RecA lobe and potentially participate in ATP hydrolysis: the A-loop, containing a conserved aromatic residue, which makes a π-π stacking interaction with the aromatic cycle of the ATP

adenosine moiety, and the H-loop, containing a conserved histidine residue, which is thought to help position the lytic water molecule (Oswald et al., 2006).

Structural studies of many ABC ATPases have shown that ATP binding to the NBD results in a rigid-body rotational movement of the helical domain towards the RecA-like domain (Oswald et al., 2006). Consequently, the side chain of a conserved glutamine, located at the tip of a protruding loop from the helical domain, reaches out and coordinates the Mg^{2+} ion and, in some cases, the γ-phosphate as well of the bound ATP molecule. This loop, named Q-loop after the conserved glutamine (also called the γ-phosphate switch) it contains, is thought to connect the helical and the RecA lobes of the NBD through the Mg^{2+} coordination, which generates a tightened conformation of each ABC monomer that renders them prone to dimerization (Higgins and Linton, 2004; Hopfner, 2016; Oswald et al., 2006).

Once ATP-Mg^{2+} is bound to the catalytic pocket and the NBD is ready for dimerization, another conserved motif (consensus LSGGQ) from the other ABC monomer, identified as the signature motif of ABC ATPases, reaches out to the ATP bound to the opposite ABC partner and coordinates its β- and γ-phosphate groups via the hydroxyl group of its serine residue. The signature motif acts as a "sensor" for the nucleotide-binding status of the opposite ABC and allows NBD-NBD dimerization in a head-to-tail manner (Figure 13B). The D-loop (consensus SALD) located downstream of the Walker B motif contains a conserved aspartate which further contributes to the NBDs dimerization (Figure 13) (Oswald et al., 2006).

Two ATP-Mg^{2+} are thus "sandwiched" at the interface of an NBD-NBD dimer, each one between the Walker A motif/P-loop of one NBD and the signature motif of the opposite partner (Figure 13B). This yields a complete and functional ATPase module with two active sites. After ATP hydrolysis, ADP and phosphate are released, and the NBDs dissociate back into their open conformation, resetting the ATPase for a new ATP binding and hydrolysis cycle. ATP-Mg^{2+} binding to the active site is thought to be the "power stroke" of ABC ATPases that permits substrate translocation, while ATP hydrolysis resets the ATPase to its basal state for a new cycle (Higgins and Linton, 2004; Oswald et al., 2006; Hopfner 2016).

3. A basic view of the catalytic mechanism of ABC ATPases NBD

Common sequence and structural features of ABC ATPases indicate that they share a similar mechanism in ATP binding and hydrolysis. Once ATP-Mg^{2+} is stabilized into the composite active site following NBD-NBD dimerization, the lytic water molecule, which is positioned between the γ-phosphate and the catalytic glutamate, is polarized, and activated by the glutamate. The catalytic glutamate acts as a general base by accepting a proton from the water molecule. The activated water proceeds to a nucleophilic attack on the γ-phosphate, which in turn gives away its spare electron to the neighboring β-oxygen, thus breaking the β- and γ-phosphate bond (Figure 14).

The Mg^{2+} cofactor plays a key role in promoting the hydrolysis of ATP. By attracting the electron clouds of the β- and γ-phosphates, thus rendering the phosphorous cores more electrophilic, Mg^{2+} facilitates the nucleophilic attack on the γ-phosphate, the stabilization of a pentavalent transition-state intermediate, and ultimately, the breakage of the β- and γ-phosphate bond (Figure 14). The hydrolysis of ATP yields ADP and HPO_4^{2-}. It has been proposed that the protonated catalytic glutamate could transfer its hydrogen back to the leaving HPO_4^{2-}, thus yielding $H_2PO_4^-$ and resetting the catalytic residue before the start of a new hydrolysis cycle (Krishnan et al., 2020).

Although this described mechanism could potentially apply to the cohesin basal catalytic activity, many aspects of the exact and complete mechanism of cohesin ATP binding and hydrolysis remain unknown. Indeed, ABC ATPases being one of the largest protein super-families, and despite sharing structural and basal mechanistic elements, the residues involved more broadly in the ATPase cycle may have evolved in different ABC ATPases and in different species. Good examples are the H-loop and A-loop motifs, which are absent from cohesin, while the H-loop motif is found in bacterial Smcs. In the same way, other specific residues of cohesin could directly participate in the modulation of the mechanism described above, or indirectly by binding regulatory subunits and via post-traductional modifications, thus regulating the cohesin ATPase activation.
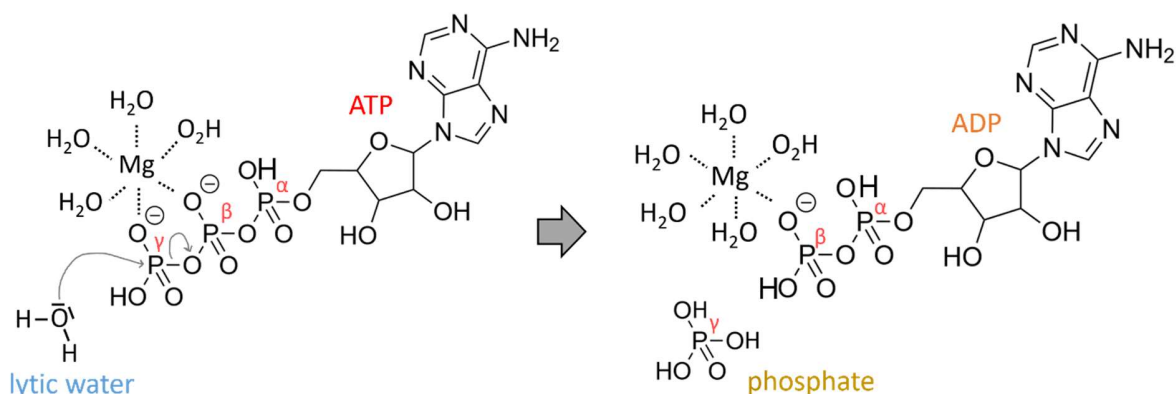
Figure 14: Mechanism of ATP hydrolysis into ADP and P_i. The gamma phosphate of the ATP molecule is attacked by an activated water molecule, which promotes de brakage of the gamma phosphate-beta oxygen bond and the release of phosphate. The cofactor Mg^{2+} ion assists and promotes the hydrolysis reaction. In ATPase proteins, the lytic water molecule is activated by a catalytic residue, usually a glutamate, and some of the hydrogen bonds formed by the water molecules coordinating the Mg^{2+} are replaced by the protein's residues to stabilize the ATP-Mg^{2+} inside the active site.

B. Structural insights into the cohesin ATPase module

1. Similarities and differences between the cohesin ATPase and the ABC-transporters ATPase

The first structures of the cohesin ATPase module were solved by X-ray crystallography, and revealed the three-dimensional organization of the *S. cerevisiae* Smc1 and Smc3 ATPase heads, bound to the slowly hydrolysable ATP analog ATPγS (Figure 15) (Haering et al., 2004; Glioris et al., 2014). These structures were recently followed by a composite Smc1 ATPase head from *S. cerevisiae* and *Chaetomium thermophilum* in its apo form, and an engineered *S. cerevisiae-C. thermophilum* chimeric Smc1-Smc3 heterodimer, in the presence of ATPγS (Muir et al., 2020) (Figure 15). Additionally, through the past year, the structures solved by cryo-EM of the human, *S. cerevisiae* and *S. pombe* cohesin ATPase modules were released, in the presence of

NIPBL[Scc2], DNA, ATP or ATP analogs. One of the two similar structures of the human ATPase module also included STAG1, bound to NIPBL and contacting the cohesin hinge (Figure 16) (Collier et al., 2020; Higashi et al., 2020; Shi et al., 2020).

The crystallographic and cryo-EM structures showed that both SMC1A and SMC3 ATPase domains share a similar structural organization as ABC transporters NBD, and contain most of their highly conserved sequence motifs (Figure 13) which are involved in ATP binding and heads dimerization. These conserved motifs include the Walker A/P-loop and the Walker B motifs located in the RecA-like domain, the ABC signature motif and the D-loop located in the α-helical domain, and the Q-loop that connects both domains.
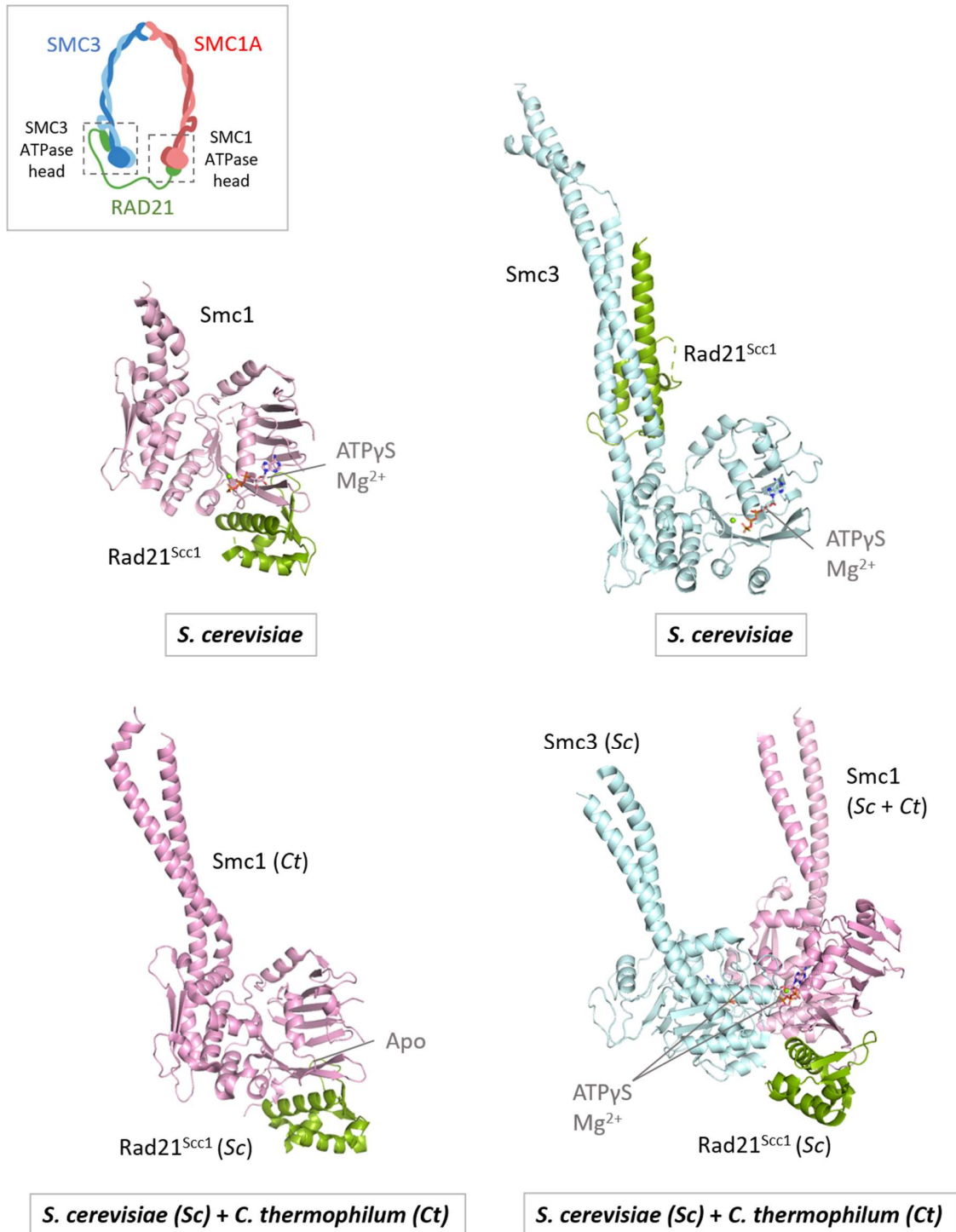
Figure 15: Crystallographic structures of the cohesin ATPAse heads. Smc1 ATPase from *S. cerevisiae*, bound to ATPγS, PDB:1w1w. Smc3 ATPase from S. cerevisiae, bound to ATPγS, PDB:4ux3. Chimeric Smc1 ATPase from *S. cerevisiae* and *C. thermophilum*, apo, PDB:6pqp. Smc1 and Smc3 engineered and chimeric structure, from *S. cerevisiae and C. thermophilum*, PDB:6qpw.
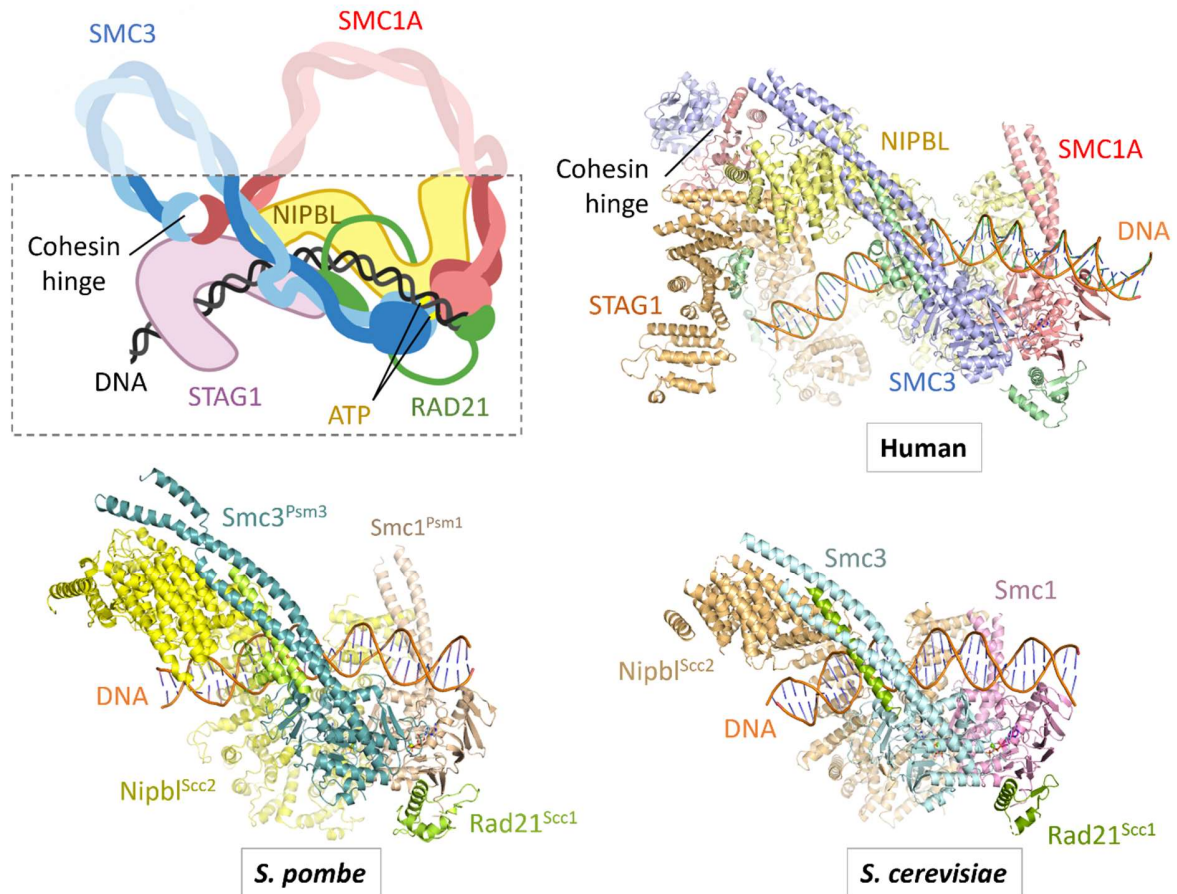
Figure 16: Structural insights into the architecture of the eukaryotic cohesin ATPase module, in the folded-ring conformation, bound to DNA and cohesin regulators. Top panel left: schematic representation of the cohesin ATPase module features, bound to DNA, NIPBL and STAG, which have recently been revealed by cryo-EM (dashed square). Top panel right, bottom panel left and bottom panel right, respectively: Cryo-EM structure of human (Shi et al., 2020; PDB: 6wge, 6wg3), *S. pombe* (Higashi et al., 2020; PDB: 6yuf) and *S. cerevisiae* (Collier et al., 2020; PDB: 6zz6) cohesin ATPase module, bound to DNA and the loader NIPBL[Scc2]. The human cohesin structure was also solved in the presence of STAG1.

Notably, in cohesin P-loop, the conserved lysine residue is especially important in contacting the ATP γ-phosphate, as a lysine-to-alanine mutant protein is unable to bind ATP (Arumugam et al., 2003). Interestingly, in the SMC2 condensin subunit, which is the v-SMC equivalent of SMC3 in cohesin, the P-loop has been shown to display flexibility and to adopt a closed conformation which hampers ATP-Mg^{2+} binding into the active pocket (Hassler et al.,

2019). However, the functional relevance of this observed flexibility must be further characterized. It is indeed a quite intriguing event, which raises the question whether it could be present in other SMCs, including cohesin, and be involved in the nucleotide binding dynamics.

Cohesin does not harbor the A- and H-loop conserved motifs of the NBD of ABC enzymes, whereas, as for all SMC ATPases, cohesin includes a characteristic R-loop spatially positioned above the ATP binding site. A conserved arginine in the R-loop (R57 in both human SMC1A and SMC3) has been proposed to help stabilize ATP in the catalytic pocket and to promote activation of the ATPase by DNA (Lammens et al., 2004). Additionally, crystal structures of nucleotide-bound Smcs show that the R-loop has a high degree of flexibility in yeast Smc1 and *P. furiosus* Smc, as it is not defined in the electron density of these solved structures (Haering et al., 2004; Lammens et al., 2004; Muir et al., 2020), while the yeast Smc3 R-loop is positioned away from the active site (Gligoris et al., 2014) (Figure 17). However, in the recently solved structures of human and yeasts cohesin bound to NIPBL and DNA, the SMC1 R-loop appears as ordered under the interaction with DNA, while the SMC3 R-loop keeps its position away from the active site (Collier et al., 2020; Higashi et al., 2020; Shi et al., 2020) (Figure 17).
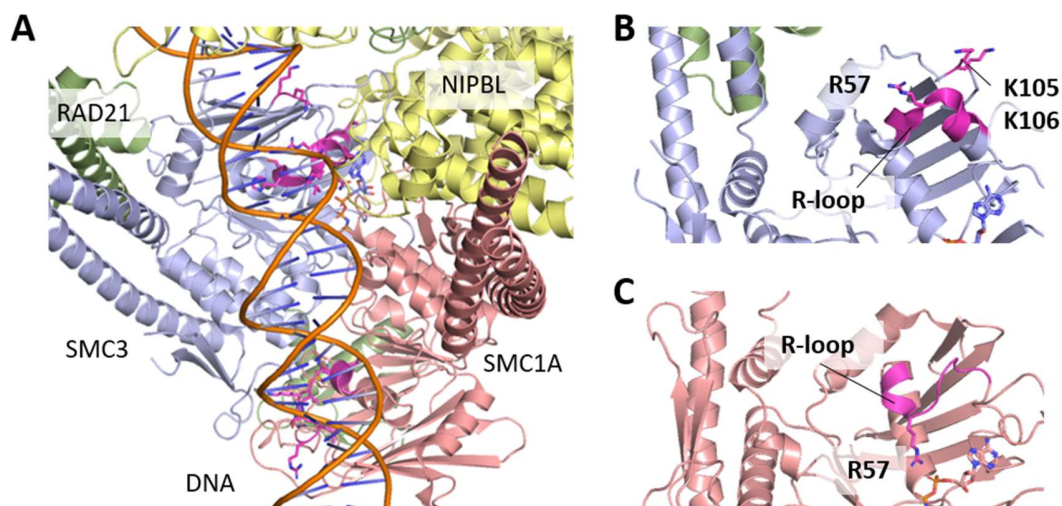
Figure 17: Positioning of the SMC1A and SMC3 conserved arginine-loops (R-loops) onto the DNA-bound human cohesin. (A) DNA can contact SMC1A and SMC3 R-loop residues (R-loop residues highlighted in magenta; PDB: 6wg3, 6wge). (B) Position of the human SMC3 R-loop, facing upwards in the opposite direction of SMC3 active site. The SMC3 R-loop is in proximity to the acetylated K106 and K106 lysisnes. (C) Position of the human SMC1A R-loop, facing inwards towards the nucleotide bound to the active site.

One significant difference with ABC transporters is that SMC complexes, including cohesin do not possess a TMD domain. Instead, their NBD is coupled to the long coiled-coil domain through an alpha helical domain located at the base of the coils. Consequently, SMC1A and SMC3 do not possess an equivalent of the coupling helix. However, SMC1A and SMC3 harbor a "signature coupling helix" located at the alpha helical domain, which links the coiled-coil domains to the NBD via the Q-loop (Hopfner, 2016), and which could potentially allow an allosteric coupling between the DNA-binding and ATPase activity of the complex (Figure 13).

As with ABC ATPases, cohesin SMC1A and SMC3 dimerize upon ATP-Mg^{2+} binding, to form a functional ATPase module. However, cohesin ATPase mechanisms and NBD movements that occur upon ATP binding and hydrolysis are not yet characterized at all steps of the ATPase cycle. As a matter of fact, the only available structures of cohesin ATPase heads are several yeasts and one human structures, stabilized by ATP analogs bound to their ATPase active sites, and a fungus Smc1 structure in its apo state (Figure 15, Figure 16). Therefore, these cohesin structures currently allow a limited description and comprehension of cohesin ATPase mechanisms, notably for the events happening before ATP binding and after its hydrolysis, which makes further characterization required.

2. SMC1A and SMC3 ATPase heads display a potential asymmetry

Human cohesin SMC1A and SMC3 ATPase heads are structurally very similar, both sharing sequence and structural features with ABC ATPases. However, when comparing both structures, it becomes apparent that there are significant differences and a certain degree of asymmetry at the structural and molecular level between both ATPase heads.

Firstly, because of the asymmetrical binding mode of RAD21 to SMC1A and SMC3, the latter interface having a key role in DNA release and possibly entry into the cohesin ring. Secondly, in the human cohesin cryo-EM structure, a major difference is observed between the SMC1A and SMC3 ATPase R-loops. Surprisingly, in SMC1A, the conserved arginine at the position

57 (R57) is directed towards and contacts the bound ATP to stabilize into the active site, whereas the equivalent R57 in Smc3 is positioned the opposite way, towards the bound DNA (Figure 17).

Another aspect of asymmetry is brought by the fact that SMC3 has key lysine residues whose acetylation is essential for stabilization of cohesin onto DNA by controlling the DNA-driven ATPase activation and the recruitment of stabilizing factors. Additionally, several studies have shown in humans and in yeasts that the cohesin ATPase possibly displays an asymmetric activity and stated the hypothesis that the SMC1 and SMC3 ATPases could potentially act differentially in DNA binding, tethering and release (Çamdere et al., 2015; Elbatsh et al., 2016). Nevertheless, it remains poorly understood whether this catalytic asymmetry could stem from the observed structural asymmetry.

Since the prokaryotic SMC complexes are fully functional being homodimers and the eukaryotic SMC complexes, including cohesin, evolved into heterodimers, questions arise concerning the differences between the SMC1A and SMC3 cohesin subunits and what makes each subunit unique and essential for proper cohesin function. Notably, how does the observed structural differences between SMC1A and SMC3 account for mechanistic and functional differentiation within the cohesin complex? In addition, the potential differences in ATP binding and hydrolysis contributions of SMC1A and SMC3, which could have direct consequences in the regulation of cohesin mechanisms, is one of the many aspects that remain largely unexplored.

C. The direct regulation of cohesin mechanisms through its ATPase module

The recently published cryo-EM structures cohesin bound to NIPBL[Scc2] and a DNA segment showed that NIPBL[Scc2] extensively interacts with SMC1 and SMC3 ATPase heads and their proximal coiled coils. NIPBL thus promotes the SMC1 and SMC3 ATPase heads heterodimerization in the presence of ATP, into a "productive" conformation of the cohesin ATPase, which is ready to hydrolyse ATP (Collier et al., 2020; Higashi et al., 2020; Shi et al., 2020). In humans and yeast, NIPBL[Scc2] has been shown to stimulate the ATP hydrolysis activity of

cohesin, which is further enhanced in the presence of the DNA molecule (Davidson et al., 2019; Kim et al., 2019; Petela et al., 2018). Interestingly, in the *Pyrococcus furiosus* archaea, the presence of DNA alone is sufficient to significantly increase the Smc ATP hydrolysis rate (Lammens et al., 2004), while for the human and yeast cohesin the presence of both NIPBL and DNA is required to enhance the ATPase activity, as DNA alone barely has an effect (Davidson et al., 2019; Kim et al., 2019).

The cryo-EM structures of cohesin ATPase modules show that the DNA molecule is positioned right above SMC1 and SMC3 ATP binding sites and binds through electrostatic interactions to a basic patch located into the V-shaped cleft of the heterodimeric ATPase. It appears that the R-loop of both SMC1 and SMC3 are major sites for DNA binding to cohesin. Interestingly, the conserved arginine residue of the R-loop, which is conserved in all SMCs, has been shown in *P. furiosus* to be directly involved in the activation of the ATPase activity by DNA (Lammens et al., 2004), though it is not known if the R-loop could have conserved such a similar activating effect in cohesin. The mechanisms by which DNA participates in the modulation of cohesin ATPase activity in concert with NIPBL[Scc2] thus remains unclear and should be further investigated.

Another means of directly regulating cohesin ATPase is by acetylation by ESCO of the two conserved lysines located on SMC3 ATPase (Figure 17). The SMC3 acetylation process is strongly dependent on ATP binding and hydrolysis by cohesin (Ajam et al., 2020; Ladurner et al., 2014). Interestingly, both lysines are in spatial proximity to DNA and NIPBL, and it has been shown that both lysines acetylation hampers the stimulatory effect of DNA on the cohesin ATPase (Murayama and Uhlmann, 2015). Murayama et Uhlmann thus proposed that the conserved lysines could possibly act as a DNA "sensor", which would trigger the ATP hydrolysis that leads to entrapment of the DNA molecule and stabilization of cohesin onto chromatin. The lysines acetylation involves the neutralization of their positive charges. Therefore, acetylation might be inhibiting the DNA-driven ATPase activation by hindering a potential interaction of the negatively charged DNA backbone with the acetyl-lysines. Moreover, it is interesting to note that acetylation of one or both lysines could also be possibly hampering the interaction of NIPBL with the cohesin

ATPase module, which would be blocking the enhancement of the ATPase activity that requires both NIPBL and DNA.

The cohesin ATPase is not only required for DNA entrapment into the cohesin ring and cohesin stabilization onto DNA (Ajam et al., 2020; Ladurner et al., 2014; Murayama and Uhlmann, 2015), it is also required for DNA release out of the ring (Elbatsh et al., 2016). It has been shown that in yeasts ATP binding and hydrolysis triggers the dissociation of the Smc3-Rad21[Scc1] interface by a process that could be blocked by Smc3 acetylation (Beckouët et al., 2016; Higashi et al., 2020), although it is not known if ATP-dependent opening of the exit gate also applies to human cohesin. Moreover, the releasing factor WAPL has been shown to possibly interact with the SMC3 ATPase head (Chatterjee et al., 2013). WAPL, which is also recruited by PDS5, participates in the destabilization of the DNA exit gate interface (Chan et al., 2012; Kueng et al., 2006), including in humans, thus controlling cohesin residency on chromatin.

A significant part of what is currently known about cohesin and its ATPase module regulation stems from research that has been performed on yeast model organisms. Many cohesin features and mechanisms are thought to be well conserved from lower to higher eukaryotes. However, evolution has often led to gene duplications in higher eukaryotes for many of the cohesin core subunits and regulators, and to the emergence of new regulatory subunits. This is notably the case in most metazoans, or more specifically in vertebrates, where duplicated and new regulatory subunits support extended cohesin functions and additional regulatory pathways. Therefore, questions arise on whether the differences in cohesin regulation between lower and higher eukaryotes possibly extend to the core components of cohesin. It is for example relevant to ask whether SMC1A or SMC3 from yeast bind and hydrolyse ATP in the same manner as their human homologs, or whether they present distinct mechanistic features.

AIMS OF THE THESIS WORK

The cohesin complex has a crucial role in chromatin organization and stability during cell division, and alterations in its functions are associated with genome instability that lead to disease, among which cohesinopathies and cancer. The past twenty years have seen increasingly rapid advances in the understanding of cohesin functions in genome organization and stability, through exploration of cohesin structure, mechanisms, and regulatory partners. It is now well established that cohesin dynamics and functions depend strongly on its ATPase motor. However, many questions remain partially or fully unanswered concerning the molecular basis at the ATPase module that supports cohesion functions. It is therefore essential to explore the structural and mechanistic features of the cohesin ATPase module and its activity, in order to understand cohesin biological functions whose deregulations lead to disease.

The first published structure of the human cohesin ATPase, bound to the major regulators STAG1, NIPBL and DNA provided a significant glimpse into what might be one of the critical steps during which the DNA molecule interacts with and is captured by cohesin. However, the molecular details of how this entrapment is achieved and regulated remain unknown, and little is known to date about the precise molecular mechanisms of cohesin ATPase which underlie all cohesin important functions. The available human cohesin structure represent a snapshot of the ATP-bound cohesin at a specific step of the cohesin cycle, when both SMC1A and SMC3 are engaged and ready to hydrolyse ATP. However, many important questions remain to be answered as to what happens at other steps of the ATPase cycle, notably before and after SMC1A and SMC3 ATPase heads bind and hydrolyse ATP, and what are the structural and mechanistic implications of these events.

Questions additionally arise following the emergence of the two SMC1A and SMC3 homologous proteins in eukaryotic cohesin, including in human. What are the implications for the ATPase mechanism? Do SMC1 and SMC3 bind ATP in the same manner, with the same functional purposes? What are the structural and mechanistic consequences of ATP binding and

hydrolysis by both ATPase heads? Are there additional levels of ATPase regulation within the SMC1A and SMC3 subunits?

All these questions are among the many missing pieces to the puzzle of the cohesin ATPase cycle, which once elucidated will provide a better insight into cohesin mechanisms and help answering other challenging questions on the functional roles of cohesin, including in pathological contexts. To answer these questions, it is needed to investigate what precisely happens at the molecular level during the ATPase cycle. A consequent amount of research in the current literature has been performed with yeasts cohesin. It is important to note that despite sharing major functions and mechanisms, vertebrate cohesin, including human, has adapted, and dramatically evolved novel different regulatory subunits and regulation pathways as compared to lower eukaryotes, over the more than 2 billion years spanning between the emergence of eukaryotes and the emergence of vertebrates. Furthermore, it is not known whether these differences extend to the core components of the cohesin complex, which highlights the importance of exploring the human cohesin, especially for further therapeutical considerations.

Starting my thesis project four years ago, the only available structures of core cohesin were yeast Smc1 and Smc3 ATPase heads bound to ATP analogues. No structures of human cohesin core subunits were available to the research community. However, my research team, aiming to contribute to the understanding of human cohesin functions and mechanisms and its dysregulation in pathologies, had started solving the structure of the human SMC1A ATPase head in its apo conformation.

For my thesis work, I subsequently took up the challenge to further investigate the molecular mechanisms and structural consequences of ATP binding and hydrolysis to the SMC1A and SMC3 ATPase heads. I sought to structurally characterize both independent human cohesin SMC1A and SMC3 ATPases by using X-ray crystallography, in their apo, ATP- and ADP-bound forms, as well as the ATPase intermediate forms. Furthermore, to this date no previous research in the literature has investigated the ATP binding properties of both SMC1A and SMC3, which I aimed to access using biophysical and biochemical methods such as isothermal titration calorimetry (ITC) and ATPase activity assays.

EXPERIMENTS OUTLINE AND METHODS

I.   Structural characterization of the SMC1A and SMC3 ATPase heads

A.   Preliminary data

The structural characterization of the human cohesin ATPase module is a fundamental step, as the three-dimensional organization of the cohesin motor domain components will help understanding its mechanisms and how they account for the diverse cohesin functions.

Before the start of my thesis work, my research team had started the structural investigation of the human SMC1A and SMC3 ATPase modules. Constructs for both ATPase domains were previously cloned by Dr. Tajith Shaik. The constructs contained short coiled coils, bridged by a peptide linker and a thrombin cleavage site (ESSKHPASLVPRGS) (Figure 18). The long SMC1A and SMC3 coiled coil regions were partially removed under and above the joint region, respectively (Figure 18, CC, J), in order to facilitate the protein crystallization trials, as the intrinsic flexibility of the coil would hinder protein crystallization. Both SMC1A and SMC3 ATPase constructs were either wild type, or the catalytic residues were mutated from glutamate to glutamine, into E1157Q and E1144Q in SMC1A and SMC3, respectively. The mutations in glutamine (EQ) of the catalytic glutamate located in the ATP binding pocket of each ATPase domain is intended to reduce their ATPase activity, thus allowing the stabilization of a complex-ligand structure suitable for crystallography. To complete both ATPase modules and improve their stability, constructs of the N- and C-terminal RAD21 fragments were also made. A 10-Histidine purification tag was incorporated at the C-terminal end of both RAD21 N- and C-terminal fragments, to facilitate the purification of the SMC1A-AD/RAD21C and SMC3-AD/RAD21N complexes by affinity chromatography on TALON resin.

The structure of WT SMC1A-CC/RAD21C in its apo conformation was obtained by Dr. Tajith Shaik during his PhD. In addition, he also obtained initial preliminary ITC data on the ATP-

binding properties of the SMC1A-CC/RAD21C and SMC3-J/RAD21N complexes, in collaboration with Pauline Landwerlin. The ITC results remained however to be further optimized.

Upon joining the research team, I further characterized structurally the SMC1A and SMC3 ATPases into their apo and nucleotide-bound forms, to further investigate their molecular mechanisms of ATP binding and hydrolysis. The following section describe the work that I have done during my PhD thesis.
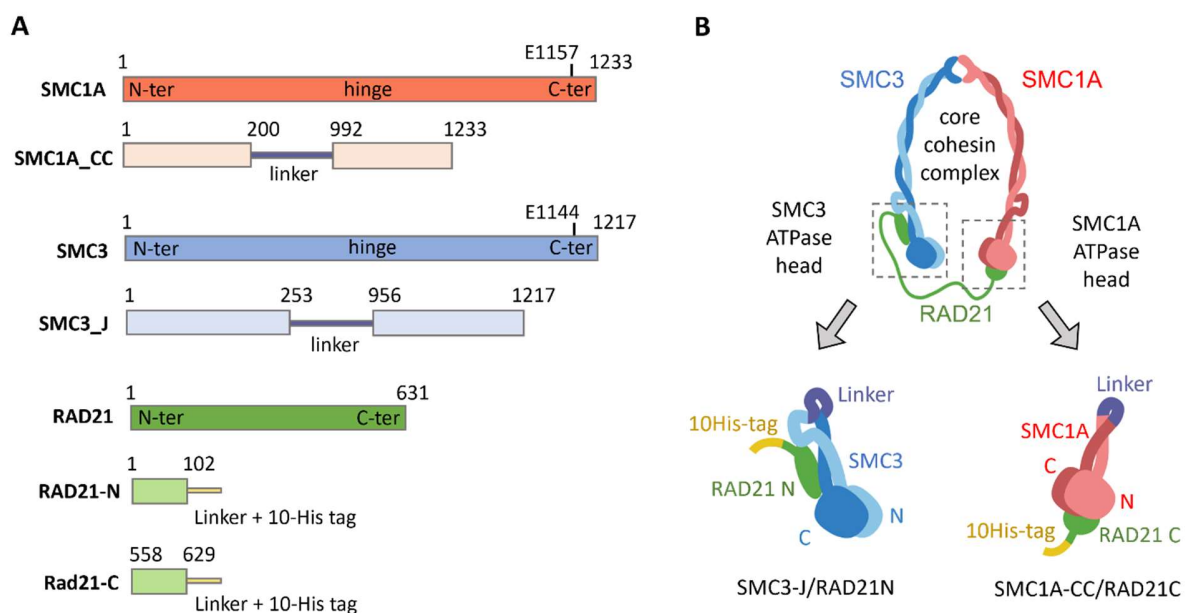
Figure 18: Protein constructs used for the human cohesin SMC1A and SMC3 ATPase characterization. (A) Sequence boundaries for the initial protein constructs, SMC1A-CC, SMC3-J, RAD21N and RAD21C, in comparison with the the full-length SMC1A (UniProtKB: Q14683), SMC3 (UniProtKB: Q9UQE7) and RAD21 (UniProtKB: O60216) reference sequences. (B) Schematic representation of the core cohesin complex and of the initial constructs used to characterize the SMC1A and SMC3 ATPase heads.

B.   Expression and purification of the SMC1A and SMC3 ATPase heads

Following the purification protocols already established in the team, I expressed and purified separately the SMC1A-CC-EQ/RAD21C and SMC3-CC-EQ/RAD21N sub-complexes by using TALON resin affinity chromatography, followed by removal of the purification tag, and size exclusion chromatography. I used the following methods for the large-scale overproduction and purification of all protein constructs, unless stated otherwise.

The plasmids coding for the SMC1A or the SMC3 ATPase heads constructs were respectively co-transformed with those coding for the RAD21C or RAD21N constructs into chemically competent *Escherichia coli* BL21(DE3) cells. The co-transformed cells were selected using antibiotics and grown for 6 hours into 1 L batches of 2x LB medium, at 37 °C, and under agitation at 190 revolutions per minute (rpm). Protein expression was induced at 25°C by the addition of a final concentration of 0.7 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG), and the cells were grown overnight under agitation. The culture media was removed by ultracentrifugation at 4,000 rpm for 30 minutes, and the bacterial pellets were resuspended with 30 ml of lysis buffer containing 200 mM (for SMC3 constructs) or 500 mM NaCl (for SMC1A constructs) and 10 mM Tris-HCl pH 8. The resuspended pellets were stored at -20°C until further use.

To extract the produced proteins, the cell pellets were thawed, and then lysed by sonication during 3 x 90 seconds at 40% amplitude and 0.5 second pulse cycles, using an ultrasonic processor (Labsonic P, Sartorius). The lysate was clarified by ultracentrifugation at 18,000 rpm for 1 hour, at 4°C. The recombinant SMC1A/RAD21C and SMC3/RAD21N protein complexes were then purified via the 10-Histidine purification tag on RAD21, by incubating the cleared lysates with TALON Metal Affinity Resin (Takara Bio). The purification tag was then removed by overnight 3C protease digestion into 200 mM NaCl and 10 mM Tris-HCl pH8, under mild agitation, at 4°C. The protein complexes were further purified by size exclusion chromatography using a 16/60 Superdex 200 column (GE Healthcare) linked to an Äkta system (GE Healthcare), in a buffer containing 200 mM NaCl, 10 mM Tris-HCl pH 8, 2 mM $MgCl_2$ and

either 1 mM TCEP (samples for crystallization assays and ITC) or 2 mM DTT (samples for the ATPase assays). For the medium-scale purifications, the size exclusion chromatography step was performed into an analytical Superdex 200 10/300 GL column (GE Healthcare). To assess the protein sample purity, an aliquot of each peak fractions was diluted with the same volume of 2x denaturing Lämmli buffer, then analyzed by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate (SDS-PAGE), using 15 % acrylamide gels. The proteins of interest were detected by staining the gels using Coomassie brilliant blue.

The main peak fractions with the protein complexes were pooled and concentrated with AMICON Ultra concentrator filters (Merck Milipore). The samples concentration was assessed using a NanoDrop spectrophotometer (Thermo Fisher Scientific). The concentrated proteins were used immediately or frozen in liquid nitrogen then stored at -80°C for later use. During all the cell lysis and the protein purification and concentration steps, the samples were always carefully kept on ice or stored at 4°C, in order to avoid degradation of the proteins of interest. The protein complex yields after the purification and protein concentration processes were satisfying for further structural and biochemical analysis, from 1 to 4 mg of protein per liter of bacterial culture. Final yields for the SMC1A-CC/RAD21C were usually higher than for the SMC3-J/RAD21N constructs.

C.   Crystallization trials, crystals optimization and X-ray diffraction

X-ray protein crystallography is an experimental method that allows to determine the spatial organization of the atoms building a protein structure, and thus determine the protein 3D structure. This result is achieved in 4 main steps, after the cloning, the expression and the purification of the protein or protein complex of interest (Figure 19): (1) first, the protein of interest is crystallized. The protein is exposed to various precipitating agents through crystallization screens, while the parameters of the crystallization conditions are varied (the pH, the temperature, and the ionic strength for example). This process is one of the major bottlenecks of the method, as it requires to obtain diffraction-quality crystals. However, it is often not the case at the first crystallization trials, as the protein crystallization time is not predictable,

and the crystals obtained initially can require optimization before proceeding further with the step (2), the exposition of the crystals to X-rays and the collection of diffraction patterns. (3) The data are then processed, to calculate an electron density map, followed by (4) the building, refinement, and validation of the protein structure model that fits into the map.
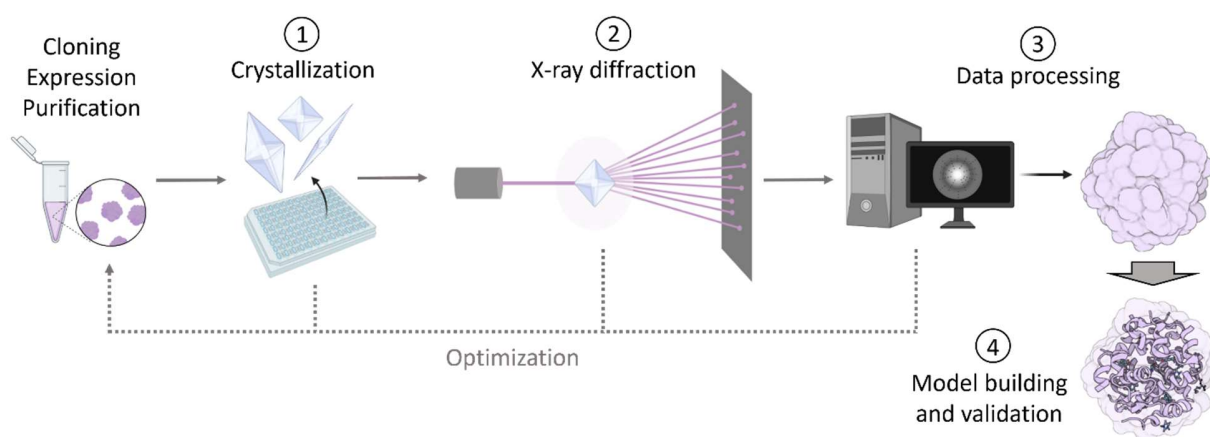
Figure 19: The protein X-ray crystallography method. After the cloning, expression, and purification into a pure and homogenous sample of the protein of interest, (1) the protein is crystallized. (2) The crystals are exposed under X-rays, and the diffraction patterns are used for (3) data processing and calculation of an electron density map. (4) A model of the protein 3D structure is iteratively built and refined into the electron density map. Protein sample preparation, and all of the following steps can be rate limiting steps: for example, the protein might not be expressed or soluble, might not crystallize or give poor-quality crystals, in which cases the experiment requires optimization to obtain high-quality diffracting crystals.

To structurally characterize the SMC1A and SMC3 ATPase heads, I carried out crystallization trials starting with the commercial crystallization screens PACT, JCSG+, Classics, WIZARD I and II, SaltRx, BCS, TOP 96, the MPD, the PEGs, LFS, Cations and AmSO$_4$, in the absence and in the presence of 2 mM of nucleotide, either ATPγS (a slowly hydrolysable ATP analog) or ADP. In order to capture the structures of the ATPase intermediate states, I also carried out crystallization assays in the presence of 2 mM of the structural analogues of the transition states of the ATPase cycle, ADP-BeF$_3$, ADP-VO$_4$ and ADP-AlF$_3$, corresponding to the ground state, the transition state and the post-hydrolysis state, respectively. Crystals were grown using the sitting-

drop vapor-diffusion method, at 4, 20 and 27 °C. 200 nl of 4 to 15 mg/ml protein sample in the absence or in the presence of the ligand was mixed with 200 nl of reservoir solution, onto Swissci 96-Well 3-Drop MRC crystallization plates (Molecular Dimensions). The selected crystals were cryo-protected with 20% (v/v) glycerol or with 20% (v/v) PEG200, then flash-cooled and stored into liquid nitrogen until data collection at synchrotron facilities, at the Source Optimisée de Lumière d'Énergie Intermédiaire du LURE (SOLEIL, France) and Swiss Light Source Synchrotron (SLS, Switzerland).

1.   SMC1A-CC-EQ/RAD21C crystallization

Several conditions from the commercial crystallization screens yielded protein crystals within a few days up to several weeks. SMC1A-CC-EQ/RAD21C crystals grew at 20°C in conditions found in JCSG+, WIZARD I and II and PACT, containing 0.1 M MMT buffer or 0.1 to 0.2 M various salts, mostly sodium and potassium salts, and 20 to 25% w/v high molecular weight PEGS (1500 to 6000) (Figure 20). The SMC1A-CC-EQ/RAD21C crystals grew in conditions with a pH of 6.5 to 8.5, although the pH did not seem to have an impact on crystal growth as some conditions without a pH-adjusting buffer also yielded crystals. Diffraction data were collected for several SMC1A-CC-EQ/RAD21C crystals at high resolution, from 1.7 to 2.5 Å.
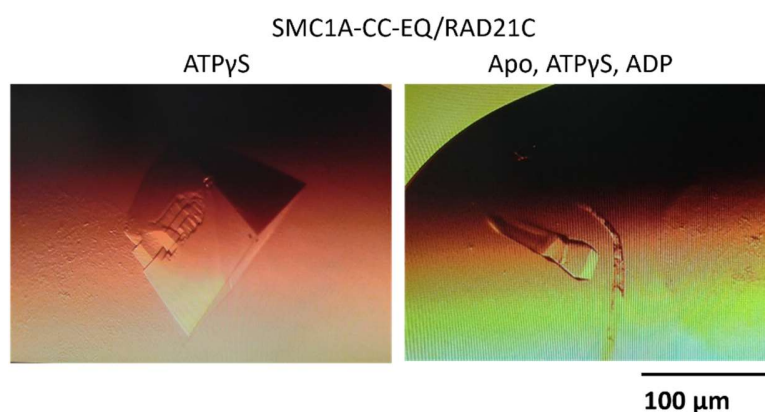
Figure 20: SMC1A-CC-EQ/RAD21C crystals. Examples of SMC1A-CC-EQ/RAD21C crystals grown in the presence of ATPγS, in the PACT Suite screen, at 20°C.

2.   SMC3-J-EQ/RAD21N crystallization

SMC3-J-EQ/RAD21C crystals grew at 4, 20 and 27°C, in conditions found in JCSG+, Classics and SaltRx. The crystals mainly grew in conditions with a high concentration (0.7 to 1.2 M) of organic acids, in a pH range from 7 to 9.5, and in the presence of low concentrations (0.5 to 1%) of PEGS (Jeffamine ED 2003, PEG 2000).

The process to obtain diffracting SMC3-J-EQ/RAD21C crystals was extremely challenging. In the initial screens, showers of small needle crystals were obtained and were optimized in order to increase their size and to obtain crystals suitable for X-ray diffraction. Larger three-dimensional rectangular shaped crystals were obtained after extensive optimization of the smaller crystals, by varying the crystallization conditions, the drop setting (sitting or hanging drop), and the crystallization temperature (4, 20 and 27°C) (Figure 21). Nevertheless, none of the optimized crystals diffracted.
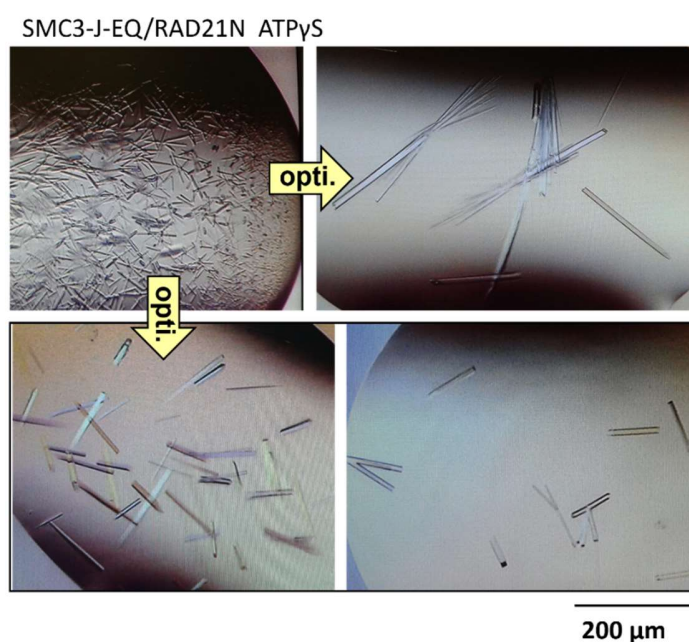
Figure 21: Optimization of SMC3-J-EQ/RAD21N crystals. The initially obtained showers of needle-shaped crystals (top left panel) were optimized into bigger 3D crystals (top right and bottom panels) suitable for diffraction experiments. The crystals were better optimized into 1.1-1.2 M Sodium malonate, 0.1 M Bis-tris propane pH 9-9.5, and 0.15-0.2% Jeffamine ED 2003 (bottom panels).

However, it appeared that sample contaminations could potentially be affecting the crystals diffraction. Indeed, after purification, the SMC3-J-EQ/RAD21N samples always displayed a ratio of absorbance at 260 nm and 280 nm near to 1.0, and the analysis of the protein samples on agarose gel electrophoresis in the presence of BET revealed the presence of co-purified contaminant nucleic acids. Therefore, I further optimized the initial protein purification protocol by adding an additional Heparin affinity purification step, with the aim to remove any residual nucleic acid contaminants from the SMC3 ATPase sample that could possibly be hindering the crystals diffraction (Figure 22, Figure 23). Heparin is a negatively charged and linear polysaccharide. It mimicks the structure and anionic charges of the DNA molecule, and can therefore interact with DNA-binding proteins and displace the bound DNA. Since the heparin-protein interaction is weakened by an increase in ionic strength, heparin affinity is therefore used as a protein purification method.

To purify SMC3-J/RAD21N by heparin affinity chromatography, after the removal of the 10Histidine-tag, the NaCl concentration of the protein samples was diluted to 50 mM, and the complexes were loaded onto a 1 ml or 5 ml HiTrap Heparin column (GE Healthcare) and eluted using a NaCl gradient from 50 mM to 1 M. The peak fractions containing the protein complexes were pooled and further purified as cited above in the protocol for large scale protein purification.
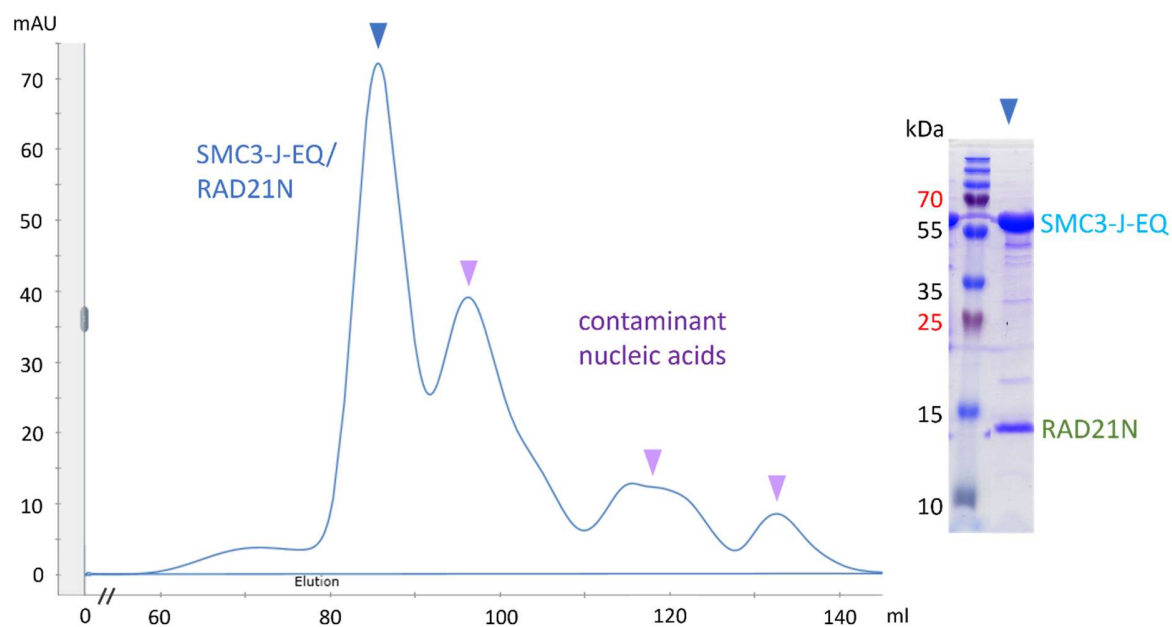
Figure 22: Size exclusion chromatography profile of the SMC3-J-EQ/RAD21N complex. The protein purified from 3 liters of bacterial culture was used. The peak fractions (blue arrow) were concentrated and used for the initial crystallization experiments. The following peak fractions (purple arrows) were shown later to be contaminant nucleic acids, through optimization of the purification conditions. Y axis: mili-Absorbance Units (mAU) at 280 nm; X axis: elution volume. On the right, the migration profile of SMC3-J-EQ/RAD21N by SDS-PAGE.

A

mAU

180
160
140
120
100
80
60
40
20
0

Without Heparin
— UV at 280 nm
···· UV at 260 nm

SMC3-J-EQ/RAD21N

contaminant nucleic acids

8   10   12   14   16   18   20   22   24   ml

B

mAU

160
140
120
100
80
60
40
20
0

— UV at 280 nm
— UV at 260 nm

1 M

NaCl

SMC3-J-EQ/RAD21N

Nucleic acids

50 mM

6.A.3   Waste(Frac)   Waste(Frac)   5.A.6   5.A.10   5.B.6   5.B.10   5.C.2   5.C.6   5.C.10   5.D.2   5.D.6   5.D.10   5.E.2

Sample Application   Column Wash   Elution

0   10   20   30   40   50   60   70   80   ml

C

mAU

100
90
80
70
60
50
40
30
20
10
0

With Heparin
— UV at 280 nm
···· UV at 260 nm

SMC3-J-EQ/RAD21N

8   10   12   14   16   18   20   22   24   ml

96

Figure 23: Optimization of the SMC3-J-EQ/RAD21N complex purification. (A) Analytical size exclusion chromatography profile of the SMC3-J-EQ/RAD21N complex. Blue arrow: the absorbance at 280 nm higher than at 260 nm indicates the presence of the protein complex. Purple arrows: the absorbance at 260 nm is higher than at 280 nm, indicating that the additionnal peaks contain contaminant nucleic acids. (B) Affinity purification using a 1 ml HiTrap Heparin column. Purification steps: sample application, column wash, and elution (with a 50 mM to 1M NaCl gradient; green curve). The contaminant nucleic acids (purple arrows) are either discarded in the sample application waste (nucleic acids not binding to the column), or through the elution step (nucleic acids bound to the column with the protein, and discarded by the salt gradient elution). The protein complex of interest is eluted as a single peak (black arrow), free from contaminants. (C) Analytical size exclusion chromatography profile of the SMC3-J-EQ/RAD21N complex, after the Heparin purification step. The complex elutes as a single uncontaminated peak (green arrow). Y axis: mili-Absorbance Units (mAU) at 280 nm; X axis: elution volume.

With the optimized protein samples, I obtained new SMC3-J-EQ/RAD21N crystals, which were exposed to X-ray radiation at synchrotron facilities. Unfortunately, the optimization experiments remained unsuccessful: most of the crystals did not diffract or diffracted anisotropically at very low resolutions. In-house data reprocessing indicated an 8 Å resolution, unsuitable for structure solution by molecular replacement.

Another main reason to why the SMC3-J-EQ/RAD21N crystals would not diffract could be explained by the possible flexibility of the long SMC3 coiled coil, which could be interfering in the crystal lattice formation, thus in the diffraction quality. This possibility was therefore explored as a new direction for optimization of the SMC3 ATPase head crystals.

D.   Optimization of the protein constructs and of the crystallization conditions

1.   Constructs optimization

In an attempt to stabilize the complex into the crystals and improve X-ray diffraction, and since the remaining flexible coiled region emerging from the SMC3 ATPase head could be hindering its crystallization, I designed and cloned new constructs of shortened coiled coils of the SMC3 ATPase (Figure 24). The same method was applied to the SMC1A ATPase, in an effort to

stabilize the complex and promote its crystallization in presence of the ATPase cycle transition state analogs (Figure 24). The new constructs were made based on sequence alignments of SMC1A and SMC3 sequences from different organisms, in order to select the new shortened coiled coils boundaries at variable regions, and avoid affecting conserved regions that could potentially be essential for the correct coiled-coil folding and stability (Appendix 1, Appendix 2). The protein sequences were aligned using Clustal Omega (Goujon et al., 2010), and the results were visualized and analyzed using the Jalview multiple sequence alignment analysis tool (Waterhouse et al., 2009). The shorter SMC1A ATPase construct was also designed taking into account the boundaries of the coiled-coil segment seen in the crystal structure of SMC1A-CC-EQ bound to ATPγS. Additionally, in the new constructs, I replaced the variable residues from the peptide linker between the coils by GS peptide sequences (from ESSKHPASLVPRGS to GSGSLVPRGSGS), as the linker composition, flexibility and stability can affect the protein stability in solution and crystallization (Chen et al., 2013). I prepared the N- and C-terminal end fragments of each construct using polymerase chain reaction (PCR), then assembled and ligated the products into native bacterial expression vectors using the Gibson assembly method. All expression vectors were verified for the new clones by sanger sequencing (Eurofins Genomics).
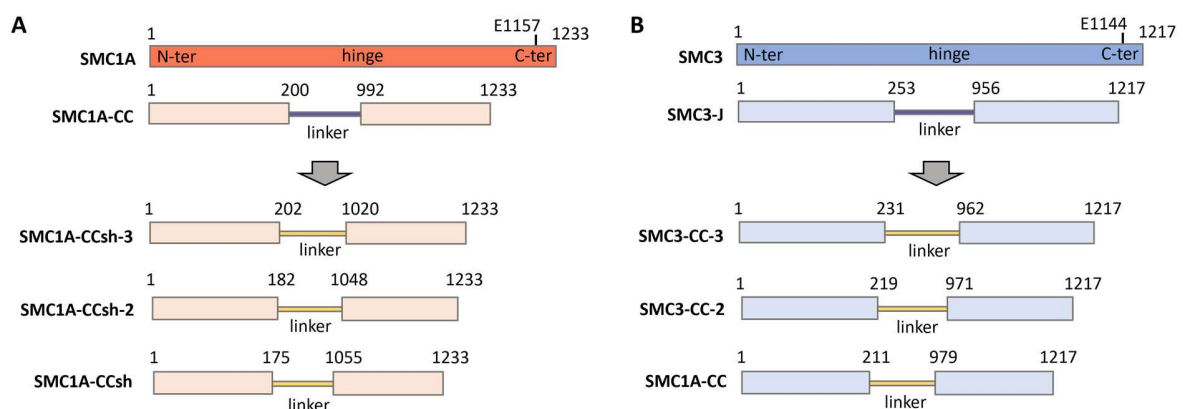
Figure 24: Optimized SMC1A and SMC3 ATPase constructs. Sequence boundaries of the optimized SMC1A and SMC3 constructs, by shortening their coiled coil.

2.   Small-scale and medium-scale expression and purification tests

I co-expressed the new SMC1A and SMC3 ATPase constructs with their respective 10His-tagged RAD21 binding fragments, and their stability was analyzed through small-scale expression tests. The technique, developed by my research team along with protein co-expression in bacterial cells, allows to assess the proteins expression and solubility while testing multiple protein buffers at the same time, thus providing a high-throughput and fast way to test and optimize purification conditions (Vincentelli and Romier, 2016).

For the small-scale expression and purification tests, I expressed each protein complex into 4 ml of cell culture, using the same conditions as cited above for the large scale, except using 1 mM IPTG for protein expression induction. The culture media was discarded by centrifugation at 4,000 rpm, during 15 min, at 4°C. The cells were resuspended into 1.2 ml of buffer containing 200 mM NaCl and 10 mM Tris-HCl pH8, then lysed using a 4-tip ultrasonic processor (Vibracell 75043, Bioblock Scientific), during 40 seconds, at 40% amplitude and 0.2 second pulse cycles.

The protein complexes were then purified on TALON affinity resin (Takara Bio) via the 10-Histidine tag combined to RAD21. The protein complexes were eluted from the TALON resin by adding 2x denaturing Lämmli buffer, followed by analysis by SDS-PAGE on 15 % acrylamide gels. All the new SMC1A and SMC3 ATPase head constructs were then expressed at medium scale, to assess the protein yields. The size exclusion chromatography step was performed into an analytical Superdex 200 10/300 GL column (GE Healthcare). All protein constructs were soluble and stable after the analytical size exclusion chromatography into 200 mM NaCl, 10 mM Tris-HCl pH 8, 2 m M MgCl$_2$ and 1 M TCEP, except the SMC3-CC-3 construct that showed strong precipitation upon concentration beyond 3 or 4 mg/ml. The shorter construct of SMC1A (CCsh) displayed better stability than both longer constructs (CC-2 and CC-3), despite its slightly lower protein yield. For SMC3 however, the shorter construct displayed the better stability and final yield. Thus, the SMC1A-CCsh and SMC3-CC constructs were selected for further crystallization assays (Figure 25, Figure 26).
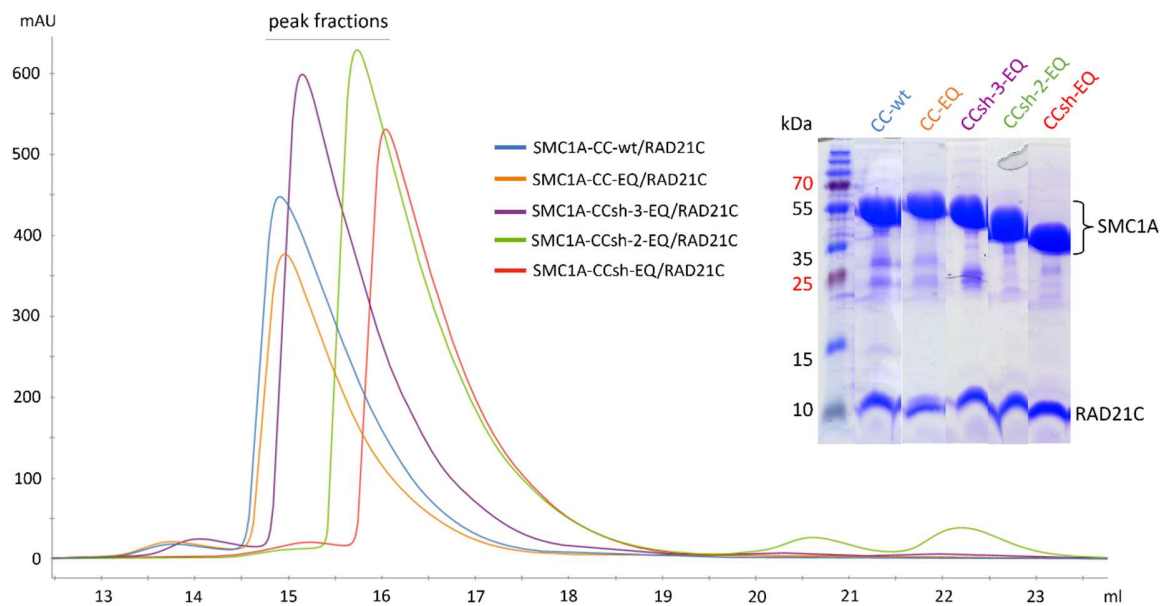
Figure 25: Analytical size exclusion chromatography of the SMC1A ATPase constructs. Each construct was purified purified from 1 liter of bacterial culture. Y axis: mili-Absorbance Units (mAU) at 280 nm; X axis: elution volume. On the right, the SDS-PAGE migration profiles for the peak fraction of each construct.
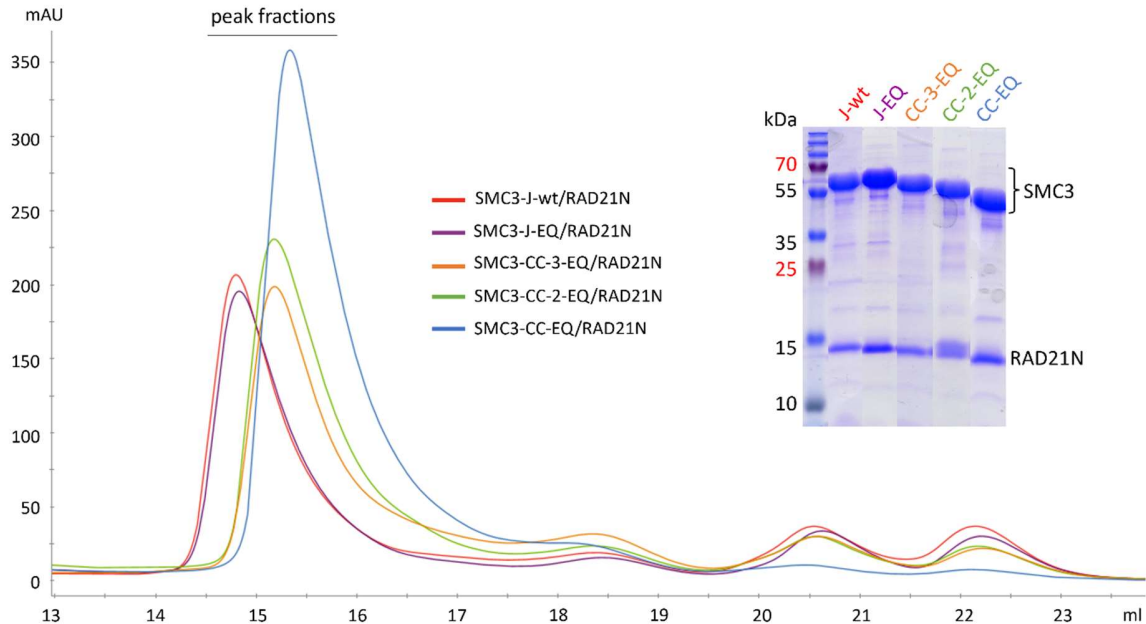
Figure 26: Analytical size exclusion chromatography of the SMC3 ATPase constructs. Each construct was purified purified from 1 liter of bacterial culture. Y axis: mili-Absorbance Units (mAU) at 280 nm; X axis: elution volume. On the right, the SDS-PAGE migration profiles for the peak fraction of each construct.

3.   Large scale expression and purification of the optimized constructs

Since all new protein constructs were still co-purifying with nucleic acids, I expressed and purified the SMC1A-CCsh/RAD21C and SMC3-CC/RAD21N complexes in larger scale, while including the Heparin affinity purification step in all of the following purification protocols, to remove the contaminants that could potentially interfere in the further biochemical and structural studies if not removed (Figure 27, Figure 28).
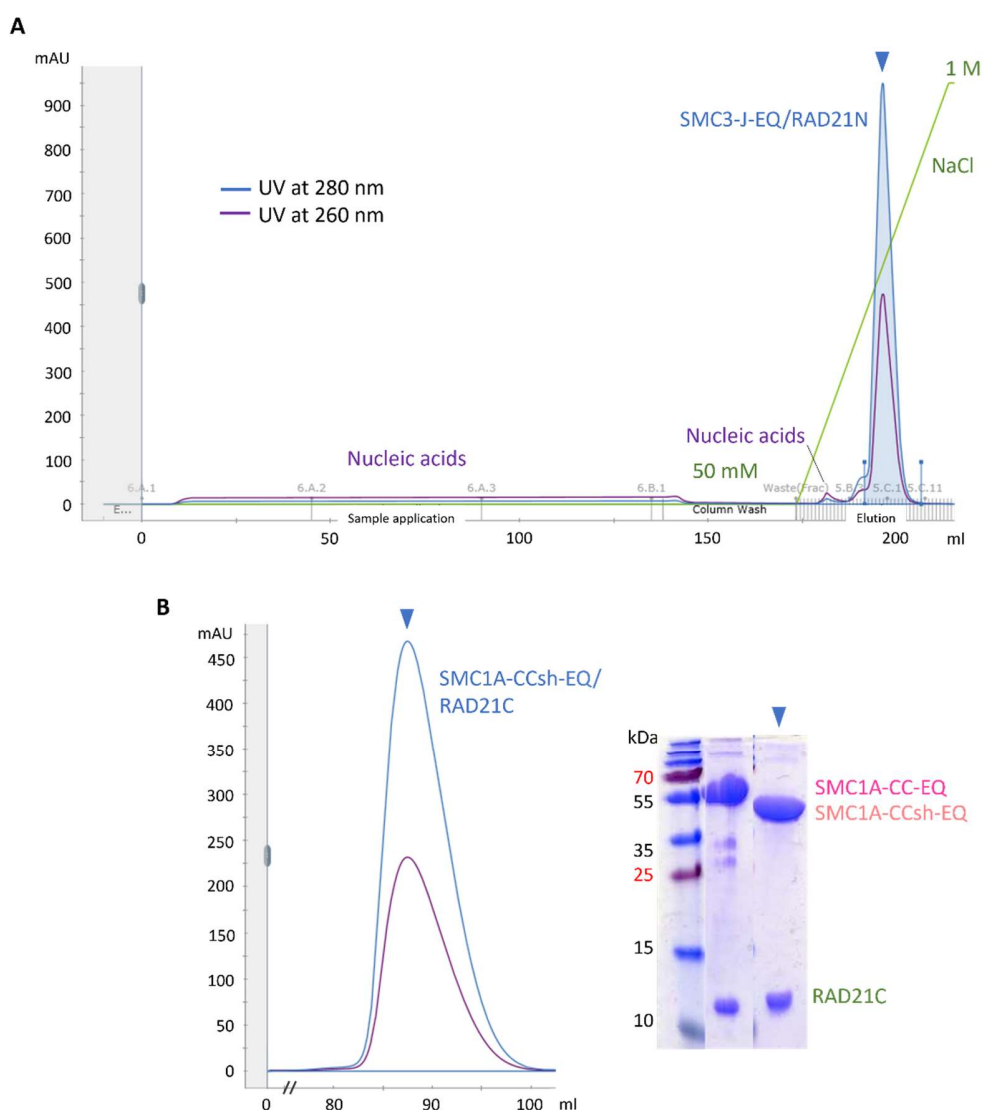
Figure 27: SMC1A-CCsh-EQ/RAD21C purification profiles. Purification of the SMC1A-CCsh-EQ/RAD21C construct for crystallization experiments, by (A) Heparin affinity chromatography following by (B) size

exclusion chromatography. The protein purified from 6 liters of bacterial culture was used. Y axis: mili-Absorbance Units (mAU) at 280 nm; X axis: elution volume. On the right of the panel (B), the SDS-PAGE migration profile of SMC1A-CCsh-EQ/RAD21C compared to SMC1A-CC-EQ/RAD21C. The new construct displays no degradation and is stable.
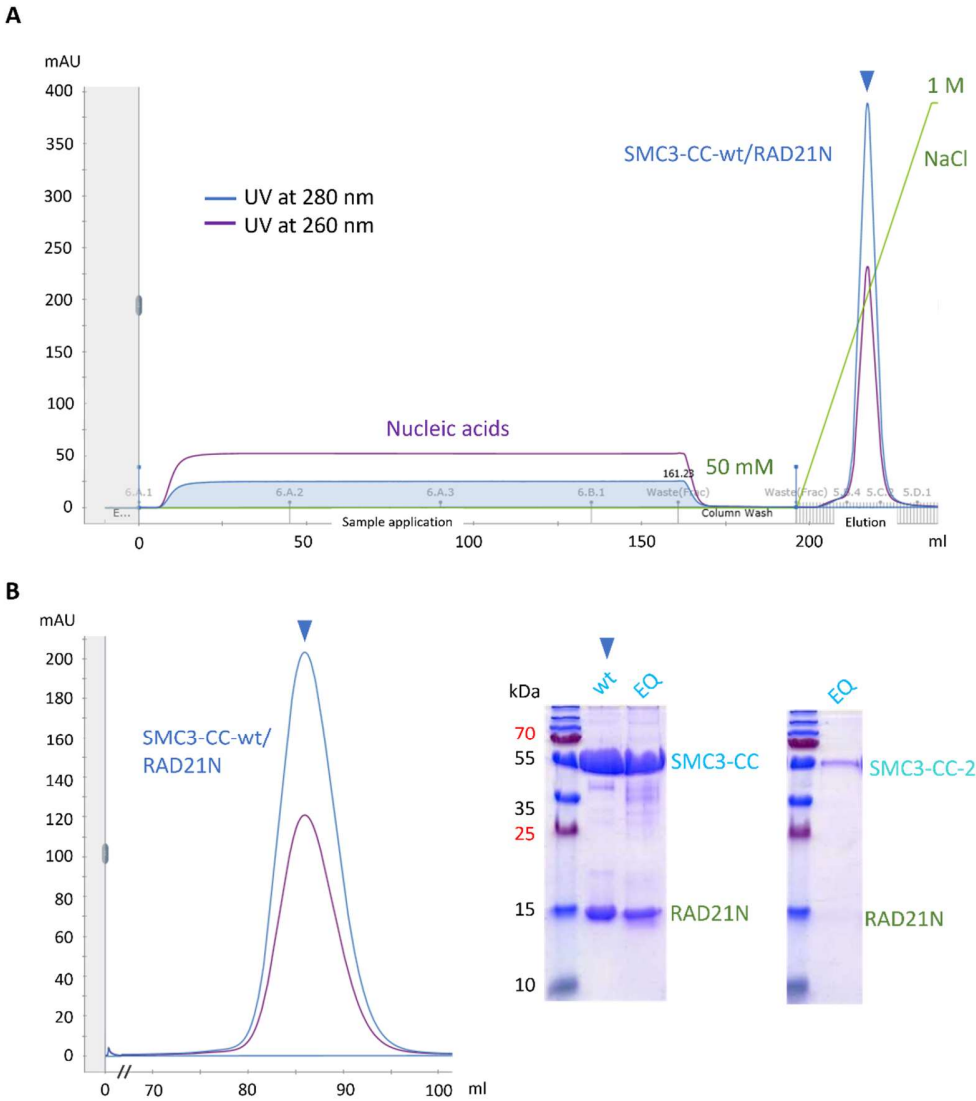
Figure 28: SMC3-CC-WT/RAD21N purification profiles. Purification of the SMC3-CC-WT/RAD21N construct, by (A) Heparin affinity chromatography following by (B) size exclusion chromatography. The protein purified from 12 liters of bacterial culture was used. Y axis: mili-Absorbance Units (mAU) at 280 nm; X axis: elution volume. On the right of the panel (B), the SDS-PAGE migration profile of SMC3-CC-WT/RAD21C. The SMC3-CC-EQ/RAD21N and SMC3-CC-2-EQ/RAD21N constructs have similar purification and SDS-PAGE migration profiles.

4. SMC1A-CCsh-EQ/RAD21C crystallization

The SMC1A-CCsh-EQ/RAD21C complex crystallized impressively fast (within a few hours) at 20°C, usually into large and isolated crystals or into a shower of smaller 3D crystals, in a significant number of crystallization conditions of the PACT Premier, BCS, LFS and JCSG+ commercial screens (Figure 29). Many SMC1A-CCsh-EQ/RAD21C crystals grew in conditions similar to SMC1A-CC-EQ/RAD21C crystals, although significantly faster and in higher numbers, in 0.2 M of various salts, 20% PEG3350 or PEG6000, and pH 6.0 to 8.5 or no adjusted pH.

The SMC1A-CCsh-EQ/RAD21C crystals grew equally well in the absence or in the presence of ADP, ATPγS and ATPase cycle transition state analogs, and most of them diffracted the X-rays. It is of interest to note that the cryoprotectant used, either 20 % v/v glycerol or 20 % v/v PEG200 into the reservoir solution, had a significant impact on the quality of the SMC1A-CCsh-EQ/RAD21C crystals diffraction data: the crystals cryoprotected with PEG200 displayed excellent diffraction quality, up to 1.3 Å, whereas the crystals cryoprotected into glycerol showed lower diffraction, inconsistencies in the unit cell parameters, and poor electron map density after data processing, indicating damage of the crystals by glycerol.
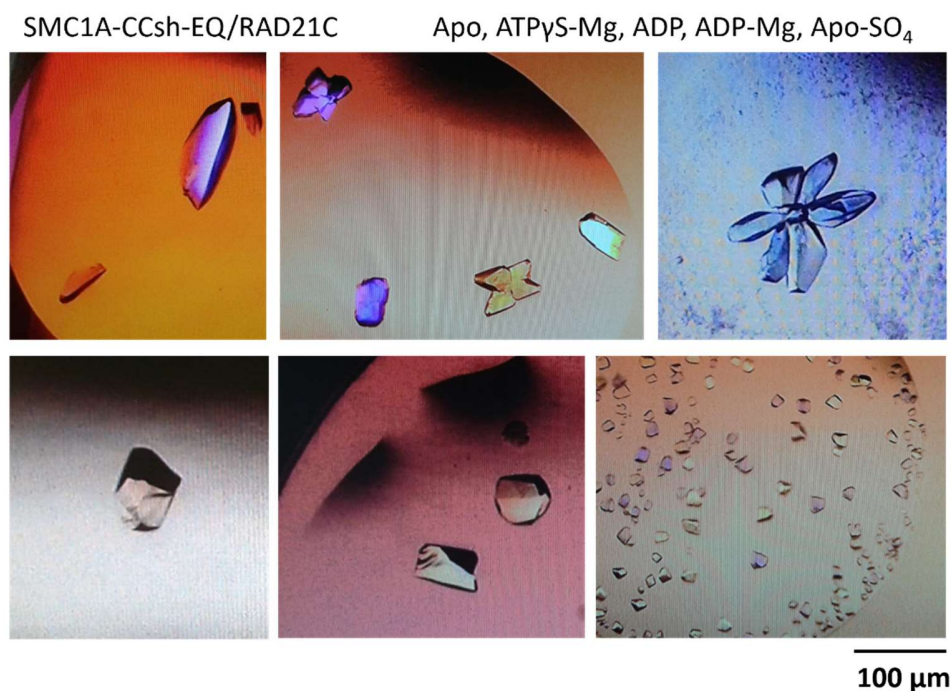
Figure 29: SMC1A-CCsh-EQ/RAD21C crystals. Typical crystal shapes obtained for SMC1A-CCsh-EQ/RAD21C, in multiple conditions of the PACT Premier, LFS and BCS screens. Optimization of the crystallization conditions gave clusters of large 3D crystals in some cases, for example the flower-shaped cluster at the top right panel.

5.   SMC3-CC-EQ/RAD21N and SMC3-CC-2-EQ/RAD21N crystallization

In addition, I finally obtained crystals of the SMC3 ATPase with the WT or EQ SMC3-CC/RAD21N constructs, in the presence of ADP and ATPγS, usually within 1 day to 2 weeks at 20°C. SMC3-CC/RAD21N crystals initially grew in many conditions of the BCS and Classics commercial screens, usually containing 10 to 20% of high molecular weight PEGS and at pH 6 to 9. Some of the obtained crystals were initially very small and not suitable for X-ray diffraction. I thus optimized the crystals, and larger diffracting crystals could be obtained in the presence of ADP and ATPγS (Figure 30).

In the absence of any ligand, the crystallization of the SMC3 ATPase remained challenging. The only crystals obtained of the SMC3-CC-EQ/RAD21N construct were clustered and flat needles

(Figure 30), unsuitable for crystallization assays, and the attempts to optimize them were unsuccessful. Therefore, I used the slightly longer construct SMC3-CC-2, which finally crystallized with RAD21N in the absence of ligand (Figure 30). The obtained small tetrahedral crystals were surprisingly similar to the WT and EQ SMC3-CC/RAD21N crystals in the presence of ADP, and were optimized into larger X-ray diffracting crystals.

Additionally, one crystal of SMC3-CC-EQ/RAD21N grew in the presence of ADP-BeF$_3$, a few months after the crystallization plate setup.
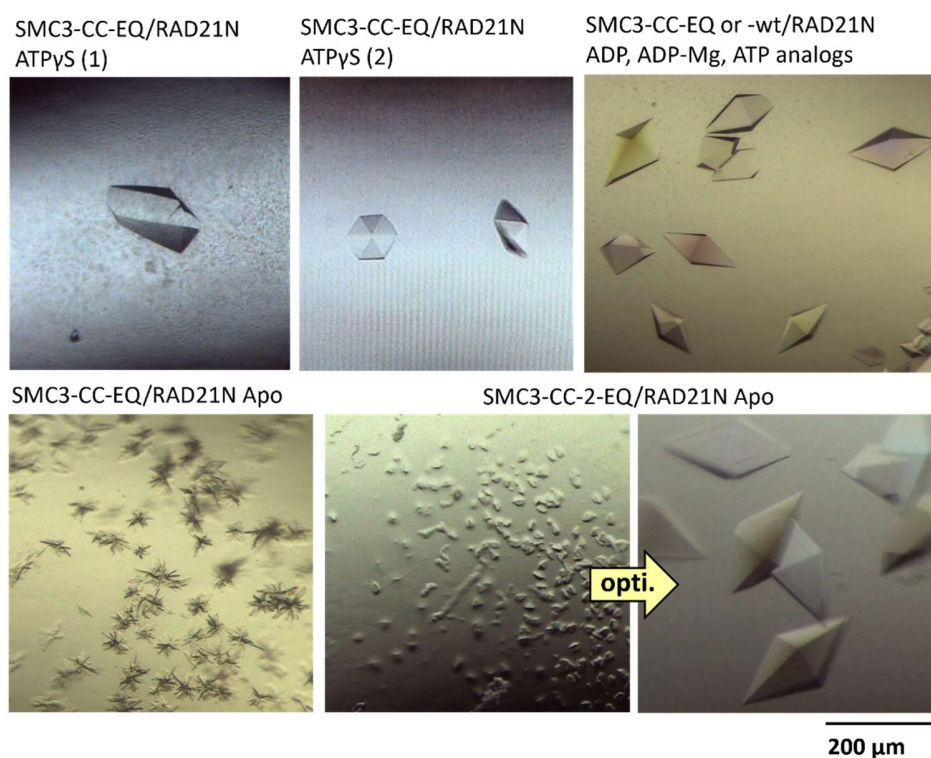
Figure 30: SMC3-CC-EQ/RAD21N and SMC3-CC-2-EQ/RAD21N crystals. Typical crystal shapes obtained for the SMC3-CC-EQ/RAD21N constructs in the resence of ADP, ATPγS, and the intermediate ATPase cycle analogs (ADP.BeF$_3$, ADP.VO$_4$ and ADP.AlF$_3$) (top and bottom left panels). SMC3-CC-2-EQ/RAD21N crystals were obtained in absence of ligand, and optimized into large 3D and X-ray diffracting crystals (bottom right).

E.   Structure solution of the SMC1A and SMC3 ATPase heads

Altogether, many of the crystals that I obtained throughout the optimization processes allowed the collection of diffraction data, at the SOLEIL and SLS synchrotrons. I obtained diffraction data between 1.36 and 3.11 angstroms for the various complexes, which I processed by indexation, integration, and scaling within the XDS program. The data were merged using Aimless from the CCP4 software suite. I then solved the structures of the SMC1A-HD/RAD21C and SMR3-HD/RAD21N complexes by molecular replacement using PhaserMR (CCP4), using as models either the SMC1A-CC-WT/RAD21C structure previously solved in the team, the yeast Smc1-HD/Scc1C structure (PDB: 1w1w; Haering et al., 2004) or the yeast Smc3-HD/Scc1N structure (PDB: 4ux3; Gligoris et al., 2014).

I built the structures and iteratively refined them using the Coot and Phenix refinement tools (CCP4), until suitable convergent R_{work} and R_{free} values were obtained. All crystallization conditions that yielded new crystallographic structures and the data processing statistics are provided in the Supplementary figures of the Results section. All structures overall conformations are presented in (Appendix 3, Appendix 4).

1.   SMC1A-CC-EQ/RAD21C structures

Altogether, from the collected X-ray diffraction datasets, I could solve the structures of the SMC1A ATPase domain, as SMC1A-CC-EQ/RAD21C in the apo, ADP- and ATPγS-bound conformations. A structure in an apo "loop" conformation was also solved, where the P-loop seems to adopt two distinct open and closed conformations. This structure is very similar to the ADP-bound structure, and was most probably initially crystallized with an ADP bound to the P-loop, which was possibly evicted during the cryoprotection step.

106

2.   SMC1A-CCsh-EQ/RAD21C structures

In addition, I solved the SMC1A-CCsh-EQ/RAD21C structure in the apo form or with ADP-, ADP-Mg-, ATPγS or SO$_4$ bound to the active site. The diffracting SMC1A-CCsh-EQ/RAD21N crystals grown in the presence of the ATPase cycle intermediate analogs only yielded ADP-bound structures. All SMC1A-CCsh-EQ/RAD21C structures were solved in the same I222 space group, with similar unit cell parameters. As a consequence, all of the SMC1A-CCsh-EQ/RAD21C structures are very similar to each other, whatever the nucleotide binding state. In addition, the tight crystal packing seems to have an impact in the observed overall tightened conformation of the SMC1A-CCsh-EQ/RAD21C structure. This is most probably due to the fast crystallization rate of this complex. In an attempt to decrease the SMC1A-CCsh-EQ/RAD21C nucleation and crystal growth rate, I made new crystal optimization assays, by decreasing the protein concentrations (from 10 to 3-4 mg/ml) and the crystallization temperature (from 20 to 4°C), but SMC1A-CCsh-EQ/RAD21C crystals still grew very fast and with unchanged physical properties.

3.   SMC3-CC-WT/EQ/RAD21N and SMC3-CC-2/RAD21N structures

I solved SMC3-CC/RAD21N structures in the ADP- and ATPγS- bound conformations for the EQ mutant, and in the ADP-bound conformation for the wild type SMC3 ATPase. Additionally, the trials to crystallize the SMC3 ATPase in its apo configuration yielded a structure containing an SO4^{2-} ion, from the crystallization condition, bound to the catalytic site of the EQ mutant SMC3 ATPase. The structure of SMC3-CC/RAD21N in presence of ADP-BeF$_3$ was solved in the same space group and with the same unit cell parameters as one of the SMC3-CC/RAD21N structures bound to ATPγS. Unfortunately, only the ADP was detected in the structure and not the BeF$_3$ moiety, most probably due to poor crystal diffraction stemming from a poor crystal cryoprotection, or from the eviction of the BeF$_3$ into the cryoprotant solution.

The SMC1A-CC-EQ/RAD21C and SMC3-CC-EQ/RAD21N crystallyzed as homodimers in the presence of ATPγS, either with a symmetric molecule or with a second molecule in the assymetric

unit of the crystal packing. It is of relevance to note that aspecific homodimerization *in vitro* of cohesin ATPase domain constructs was also reported in the literature (Haering et al., 2004; Gligoris et al., 2014), but it is most probably unlikely to happen *in vivo*, notably due to the asymmetrical binding of the SMC1A and SMC3 ATPase domains to their loader and activator NIPBL, which moreover enhances their heterodimerization (Shi et al., 2020; Higashi et al., 2020; Collier et al., 2020).

Analyses of all the obtained SMC1A and SMC3 ATPase head structures revealed structural differences between the apo, ATPγS- and ADP-bound structures, and are described throughout the Results section. Notably, the results show distinct conformational dynamics between the SMC1A and SMC3 ATPases and provide an insight into the cohesin ATPase cycle in its Apo, ATP- and ADP-bound states.

II.   Biochemical characterization of the SMC1A and SMC3 ATPase heads

A.   Isothermal Titration Calorimetry (ITC)

To assess the nucleotide binding properties of SMC1A and SMC3 ATPases and whether nucleotide binding mediates the differences in conformational dynamics observed in the crystallographic structures, the preliminary ITC experiments with the SMC1A and SMC3 ATPase heads were optimized.

ITC is a quantitative method for the analysis of molecular interactions. This technique measures the heat that is either released or absorbed during the interaction of two molecular partners, for example the binding of a ligand to its target protein, by the titration of one partner (in general the ligand, the most concentrated) into the second one (the protein). ITC directly gives access to the thermodynamic parameters of the molecular reaction, including the changes of

entropy (ΔS) and of enthalpy (ΔH), the free Gibbs energy (ΔG), as well as the dissociation constant (K$_d$) and the stoichiometry (N) of the binding event.

To assess the nucleotide binding properties of the SMC1A and SMC3 ATPase heads, I expressed and purified the SMC1A-CC/RAD21C and SMC3-CC/RAD21N complexes either WT or EQ, and prepared the nucleotides ATP, ATPγS or ADP at a final concentration of 4 mM into the protein gel filtration buffer (200 mM NaCl, 10 mM Tris-HCl pH 8, 2 mM MgCl$_2$ and 1 mM TCEP). The titration experiments and their processing were performed by the collaborative work with Dr. Eric Ennifar and Dr. Karl Brillet at the Institut de Biologie Moleculaire et cellulaire (IBMC; Strasbourg, France). After extensive optimization of the proteins and ligands concentrations, for the final experiments each nucleotide was injected into 189 to 330 µM of either wild type or EQ-mutant SMC1A-CC/RAD21C or SMC3-CC/RAD21N. ITC measurements were performed at 5 °C, using a PEAQ-ITC microcalorimeter (Malvern Panalytical). The data were then corrected for the dilution heats generated by the injection of the buffer into the protein sample and of the nucleotide sample into the buffer. The ITC data were then fitted with either one set of binding sites or two dependent binding sites models, using the AFFINImeter analysis software (Piñeiro et al., 2019).

The results, detailed in the Result section, reveal distinct modes of interaction with ATP and ADP between the SMC1A and SMC3 ATPases, which is in accordance with the structural observations: the SMC1A ATPase displays a higher degree of flexibility than the SMC3 ATPase, and the binding of ATP to SMC1A induces significant conformational changes within the structure.

B.   ATPase assays

I analyzed the ATPase activities of the purified wild type SMC1A-CC/RAD21C and SMC3-CC/RAD21N ATPases, through a spectrophotometric method using the EnzChek Phosphate Assay Kit (Thermo Fischer Scientific).

In this method, in the presence of inorganic phosphate (P_i) released by ATPases, the purine nucleoside phosphorylase (PNP) converts its substrate 2-amino-6-mercapto-7-methylpurine riboside (MESG) into the products the ribose 1-phosphate and 2-amino-6-mercapto-7-methylpurine. The formation of mercapto-7-methylpurine is monitored at 360 nm and allows to calculate the concentration of released P_i and thus assess the ATPase activity of the analyzed protein (Figure 31).
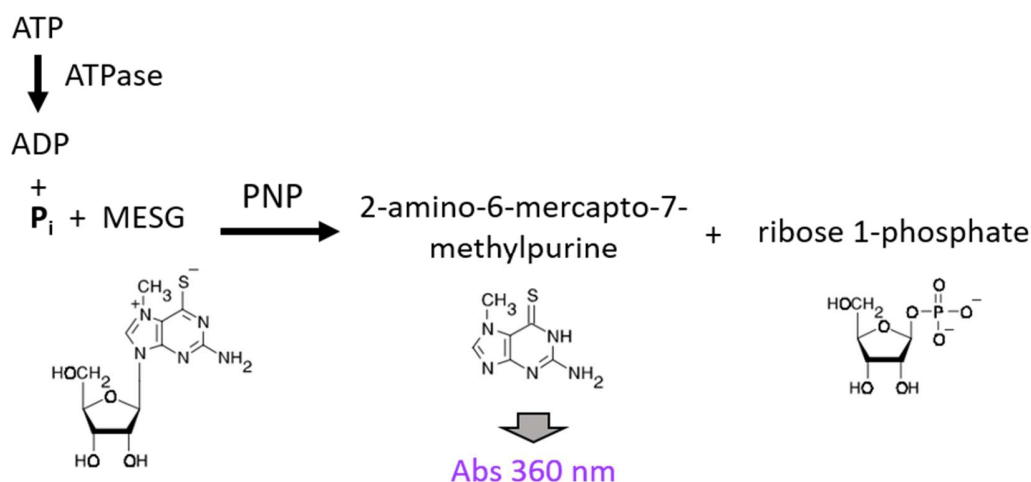
Figure 31: The EnzChek Phosphate Assay Kit method. The ATPase of interest hydrolyzes ATP into ADP and P_i. The PNP enzyme uses the released P_i to convert its substrate 2-amino-6-mercapto-7-methylpurine riboside (MESG) into the products the ribose 1-phosphate and 2-amino-6-mercapto-7-methylpurine. The formation of mercapto-7-methylpurine is monitored at 360 nm and allows to calculate the total concentration of released P_i, and thereby to assess the ATPase activity of the analyzed protein.

To assess the ATPase activity of the SMC1A and SMC3 ATPases, the wild-type proteins were purified into 200 mM NaCl, 10 mM Tris-HCl pH 8, 2 mM $MgCl_2$ and 2 mM DTT. The EnzChek kit protocol final reaction volumes were downsized to 150 µl reactions, containing 0.2 mM MESG, 0.15 U PNP, 30 µl of 5 x reaction buffer (1 M NaCl, 0.5 M Tris-HCl pH 8, 10 mM $MgCl_2$), the protein of interest, 1 mM ATP, and the needed amount of filtered and sterilized water for a final volume of 150 µl. ATP was always added last. The reactions were carried out into 96 well plates (Corning Costar). A final concentration of 10 µM of SMC1A-CC/RAD21C and 10 µM of SMC3-CC/RAD21N

was used in each reaction. In the case of SMC1A-CC/RAD21C or SMC3-CC/RAD21N alone with ATP, 20 μM of protein was used since the heads dimerize to hydrolyze ATP. After the addition of ATP, the ATP hydrolysis activity was immediately assessed at 30°C, by measuring the absorbance at 360 nm, each 42 seconds for 2 hours, using a spectrophotometer plate reader (TECAN). The experiments were performed as either duplicates or triplicates.

The results reveal that the residual individual SMC1A and SMC3 ATPase rates are extremely low, explained by the aspecific homodimerization in the experimental setup in the presence of ATP, while the ATPase activity of the SMC1A-SMC3 heterodimer is increased. The measured ATPase rates remain however at a basal level, under one molecule of ATP hydrolyzed per minute per dimer, in accordance with published results for the human cohesin ATPase (Figure 32) (Kim et al., 2019; Anderson et al., 2019). The published results additionally showed that the basal activity of the cohesin ATPase is significantly enhanced in the presence of the cohesin loader NIPBL and of DNA, and suggested that this enhancement is part of the normal cohesin function in DNA loop extrusion.
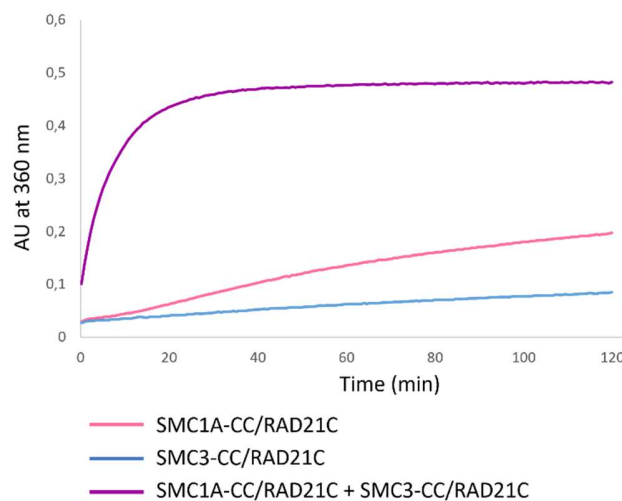
Figure 32: ATPase activity of the SMC1A and SMC3 ATPase heads. Comparison of the ATPase activity rates of the SMC1A-CC-WT/RAD21C and SMC3-CC-WT/RAD21N complexes and effect of their dimerization on the ATPase activity.

III.   Analysis of the interactions of SMC1A and SMC3 ATPases with DNA

All cohesin functions depend on its hability to interact with chromatin. Intriguingly, in the initial protein purification experiments, both SMC1A and SMC3 ATPases were always co-purified with contaminant nucleic acids, despite the two-step purification by TALON affinity and size exclusion chromatography.

Interestingly, the crystal structure of the ATPase module of Rad50, which is an SMC family member, was solved in several conformations in the presence of DNA (Liu et al., 2016; Rojowska et al., 2014; Seifert et al., 2016). The DNA fragment is either bound to a region of the coiled coil proximal to the ATPase module, or sitting above the ATPase module and contacting residues at the surface formed by the SMC-SMC dimer (PDB: 4w9m, 5dny, 5f3w, 5dac). Moreover, analysis of the *P. furiosus* ATPase module suggested that the conserved R-loop motif, protruding above the Smc ATPAse, could be implicated in the activation of the ATPase upon binding to DNA (Lammens et al. 2004). In addition, interestingly, it was also shown that in ABC transporters, which share a common NBD architecture with SMCs including cohesin, the substrate binding can be sufficient to stimulate the ATPase rate by promoting NBD-NBD dimerization (Doshi and van Veen, 2013).

Given these intriguing observations, it is therefore relevant to ask whether DNA also has a direct action on the human cohesin ATPase module and its activity, and to investigate the structural consequences of DNA binding. Therefore, I was curious to see wether some of these observations could apply to the human cohesin, and whether DNA could bind directly to the SMC1A and SMC3 ATPases and modulate their activity.

A.   Crystallization trials of SMC1A and SMC3 ATPases in the presence of DNA

Trials to solve the crystallographic structure of the cohesin ATPase heads in the presence of a 40 bp dsDNA fragment (5' TTAGTTGTTCGTAGTGCTCGTCTGGCTCTGGATTACCCGC 3', Vazquez Nunez

et al., 2019) were carried out, in the absence and in the presence of ATP and ATP analogs, including the ATPase cycle intermediate ADP-AlF$_3$ that was shown to help stabilize a DNA-bound conformation of cohesin (Camdere et al., 2018). The DNA fragment was added to the protein samples for a 1:2 protein:DNA concentration ratio, before the setup of the crystallization plates.

In many of the crystallization conditions of the PACT Premier and Nucleix commercial screens, the DNA organized into droplets of liquid-crystalline phases, characterized by a strong light polarization. A few birefringent 3D crystals were obtained with the SMC1A-CCsh-EQ construct, and were dense enough to be mounted onto cryo-loops for further X-ray diffraction experiments. However, the crystals either did not diffract or diffracted and the processed data revealed only the structure of SMC1A-CCsh-EQ, without any bound DNA. As the optimization of the crystallization conditions was being planned, the cryo-EM structure of the human cohesin ATPase module was released to the research community (Shi et al., 2020). The structure revealed the SMC1A and SMC3 ATPases, engaged in the presence of the ATP analog AMPPNP, with the DNA molecule stabilized above both heads by NIPBL and STAG1. I therefore decided not to investigate further these aspects.

B.   Characterization of the effect of DNA and R-loop mutants on the cohesin ATPase module activity

Interestingly, the released cryo-EM structure of the human cohesin ATPase module revealed that the DNA molecule binds above the SMC1A and SMC3 ATPase heads while contacting their R-loops. Mutations on the cohesin R-loops have been identified in disease, notably the V58-R62 deletion of SMC1A in the Cornelia de Lange Syndrome. Moreover, the conserved R59 residue in *P. furiosus* (human equivalent of R57) has been shown to be directly involved in the DNA-driven direct Smc ATPase activation (Lammens et al., 2004).

Therefore, I focused on analyzing the potential role of the cohesin R-loops in the ATPase cycle of the human cohesin, and whether the mechanism of ATPase activation by DNA through the conserved arginine could be conserved in humans. I cloned, expressed, and purified the

SMC1A-CC-R57A and SMC1A-CC-ΔV58-R62 mutants, then performed preliminary ATPase assays in the presence of the 40 bp dsDNA fragment (Figure 33, Figure 34). The experiments showed that the human SMC1A R57 does not seem to directly participate in the DNA-driven ATPase activity, in contrary to the archaeal Smc. However, the deletion the adjacent V58-R62 loop seems to hamper the ATPase activation in the presence of DNA, highlighting the functional importance of the SMC1A R-loop. Interestingly, however, the same experiments in the presence of SMC3 suggests that SMC3 could possibly have a rescuing effect of both SMC1A mutations, in the case of the heterodimer ATPase module.

To assess the effect of both mutations on the SMC1A ATPase, crystallization trials were performed with the mutant SMC1A-CC-EQ-R57A/RAD21C and SMC1A-CC-EQ-ΔV58-R62 but no crystals were obtained. Further analysis should elucidate the precise mechanisms through which the ATPase activation is achieved, and how it can be perturbed in R-loop and disease mutants.
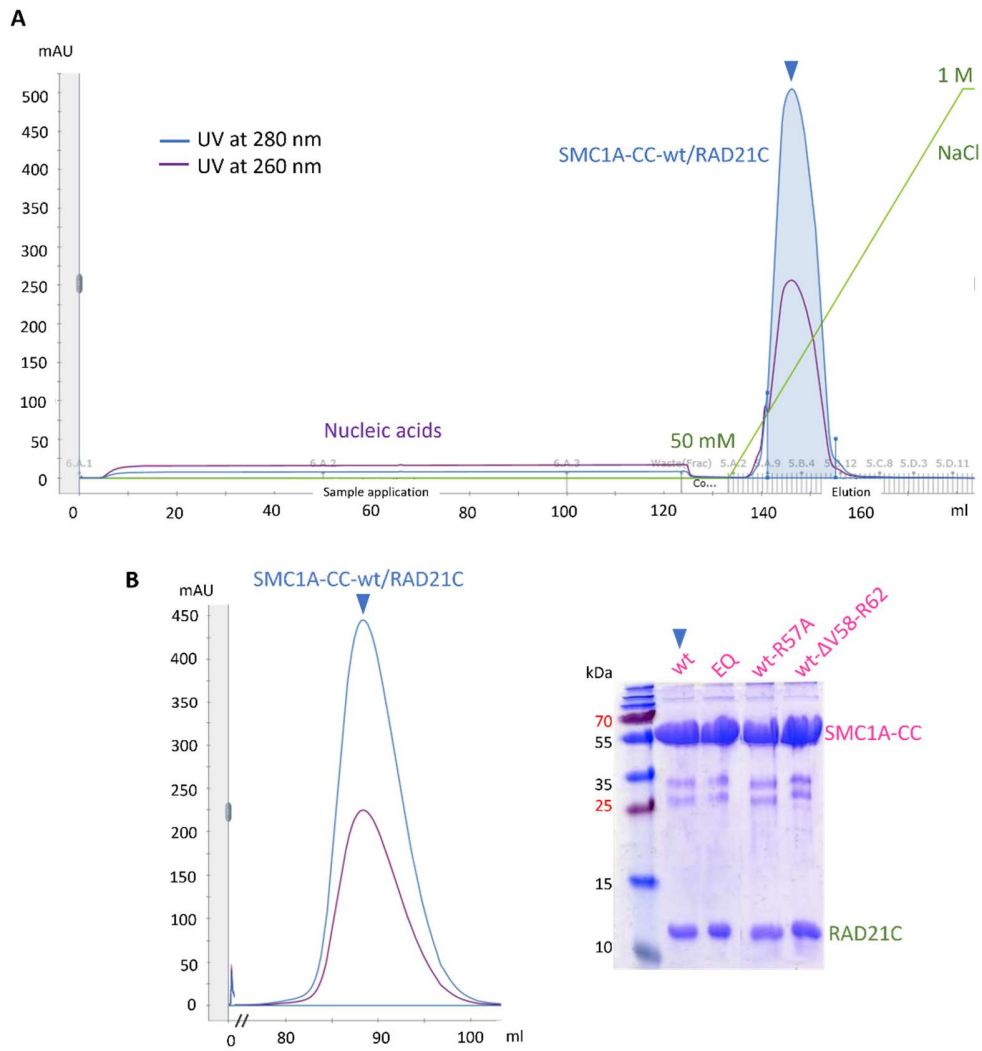
Figure 33: SMC1A-CC-WT/RAD21C purification profiles and comparison with the mutant complex. Purification of the SMC1A-CC-WT/RAD21C construct, by (A) Heparin affinity chromatography following by (B) size exclusion chromatography. The protein purified from 9 liters of bacterial culture was used. Y axis: mili-Absorbance Units (mAU) at 280 nm; X axis: elution volume. On the right of the panel (B), the SDS-PAGE migration profile of SMC1A-CC-WT/RAD21C. The SMC1A-CC-EQ/RAD21C, SMC1A-CC-R57A/RAD21C and SMC1A-CC-ΔV58-R62/RAD21C constructs have a similar purification and SDS-PAGE migration profiles as SMC1A-CC-WT/RAD21C.
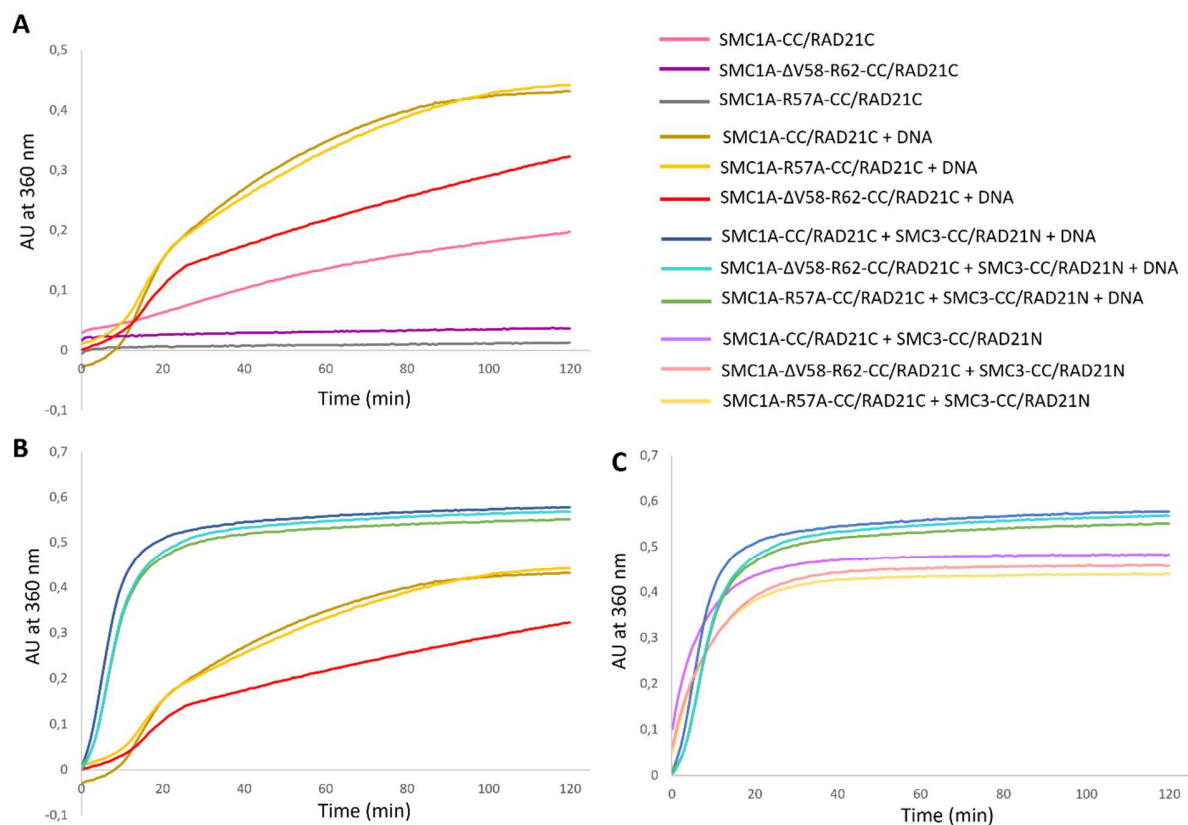
Figure 34: Effect of dsDNA on the cohesin ATPase activity. (A, B, C) DNA has a slightlly enhancing effect on the cohesin ATPase activity. The comparison of the effects of dsDNA on the ATPase activity rates of the SMC1A WT and R57A and ΔV58-R62 mutants suggest that (A) the V58-R62 loop could participate in DNA binding by the SMC1A ATPase; (B) SMC3 could potentially have a rescuing effect of mutations on the SMC1A R-loop.

IV. Analysis of the newly identified DNA exit gate conformation using *in vitro* chemical crosslinking

The crystal structures of the SMC3 ATPase module revealed a previously unknown conformation of the DNA exit gate, at the SMC3 and RAD21 interface. In all solved structures, from five different datasets and in three different space groups, the SMC3 coiled coil harbored a significant kinked conformation, which allows the RAD21 N-terminal globular domain to contact the RecA domain of SMC3. On the cohesin ATPase cryo-EM structure however, the SMC3 coil is in an extended position maintained by NIPBL and DNA (Shi et al., 2020). To analyze whether the kinked conformation of the DNA exit gate exists in solution and could possibly be mechanistically relevant, I performed chemical crosslinking assays between the RAD21N globular domain and the globular RecA domain of the SMC3 ATPase.

A. Cloning and small-scale production of the constructs for chemical crosslinking

The selection of pairing interface residues, and their mutation into cysteines for the crosslinking assays was performed with the help of Dr. Marie-Laure Durand and Edouard Troesch (Figure 35, Figure 36). The SMC3-D120/RAD21-K25 and SMC3-N119/RAD21-H22 pairs were selected, to be crosslinked in the kinked conformation and in the extended conformation of the DNA exit gate, respectively (Figure 36). Additionally, in order to avoid any aspecific crosslinking event, the endogenous cysteines from both SMC3 and SMC1A ATPases were mutated into serines (Figure 35). All point mutations were introduced into the constructions using rolling circle, gradient, or nested PCR. The mutated clones were then expressed and purified at a small scale. To eliminate contaminants from the protein samples, the protein-bound resin was washed with a buffer containing 200 mM NaCl, 10 mM Tris-HCl pH 8, 1 mM TCEP and 10 mM then 20 mM imidazole pH 8. Each complex was then eluted into 200 mM NaCl, 10 mM Tris-HCl pH 8, 0.5 mM TCEP and 250 mM imidazole pH 8.
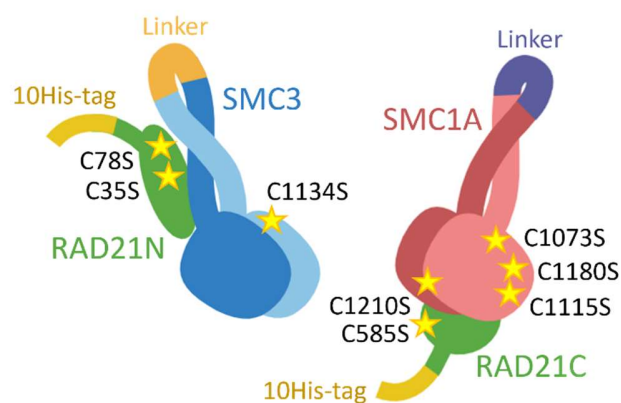
Figure 35: Cysteine less complexes used in the chemical crosslinking assays. All cysteine residues on the SMC1A-CC-EQ, the SMC3-CC-EA, the RAD21N and the RAD21C were mutated into serines, in order to avoid aspecific crosslinking. The position of the mutated cysteines on the structures are indicated by the yellow stars.
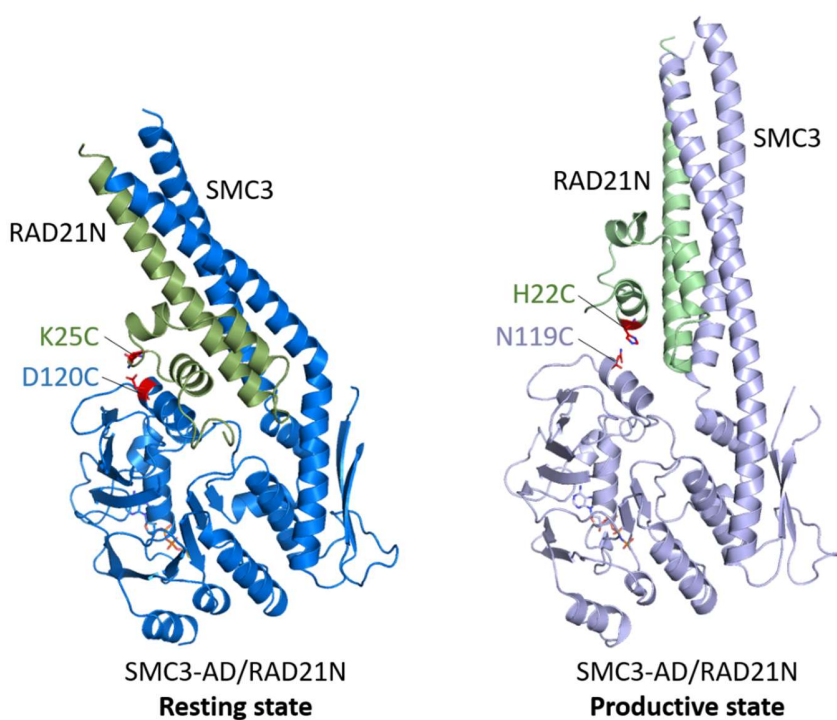
Figure 36: Position of the amino acid pairs used to analyze the observed conformations of the DNA exit gate. The SMC3-D120/RAD21-K25 and SMC3-N119/RAD21-H22 pairs were selected (colored in red) for chemical crosslinking assays, with the aim to capture the kinked conformation and the extended/productive conformation of the DNA exit gate, respectively. The corresponding residues were mutated into cysteines and used in BMOE crosslinking experiments.

B. Capture of the exit gate conformations by BMOE crosslinking assays

Bismaleimidoethane (BMOE) is a small chemical compound composed of two maleimide groups, linked by an 8 Å ethane spacer (Figure 37). BMOE is reactive towards the sulfhydryl (or thiol) group, which is notably present in the side chain of the cysteine amino acid. BMOE is therefore used in chemical crosslinking assays, where it allows the covalent linking of two cysteine residues located in spatial proximity, thus entrapping specific conformations of a target protein.
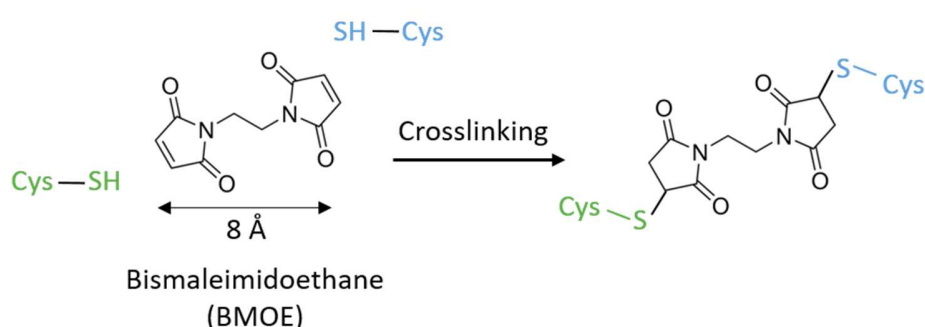
Figure 37: Mechanism of the chemical crosslinking by BMOE. The BMOE molecule reacts with the sulfidryl group (-SH) on the side chain of two cysteine amino acids (Cys), and forms a covalent link between them, thereby revealing their spatial proximity.

To perform the *in vitro* BMOE crosslinking assays, I adapted protocols available in the literature (Collier et al., 2020; Soh et al., 2015), with suggestions from Dr. Marie-Laure DURAND, and optimized them for the SMC3 ATPase in the absence or in the presence of the SMC1A ATPase and ATP. A final concentration of 20 µM of SMC1A-CC-EQ/RAD21C or SMC3-CC-EQ/RAD21N were incubated with 1 mM BMOE diluted in DMSO, into a reaction buffer containing 200 mM NaCl, 10 mM Tris-HCl pH 7.5 and 0.5 mM TCEP. Controls were set up in the same conditions, using DMSO without BMOE. All BMOE crosslinking experiments were performed at room temperature, in order to avoid significant pH changes that could affect the crosslinking reaction. The presence of imidazole in the samples did not show any alteration of the crosslinking efficiency. The reaction was then stopped for each sample by adding the same volume of 2x denaturing Lämmli buffer containing β-mercaptoethanol, then heated 5 minutes at 70°C before being analyzed by SDS-

PAGE on 15% acrylamide gels. The proteins were detected by staining the gels using Coomassie brilliant blue. The same experimental conditions were applied for crosslinking experiments of SMC3-CC-EQ/RAD21N in the presence of the SMC1A-CC-EQ/RAD21C complex and 4 mM ATP.

After incubation with BMOE, the results show that extended and kinked conformations of the DNA exit-gate can both exist in solution, including in the presence of SMC1A and ATP. The latter observation suggests that even in the context of the SMC1A-SMC3 ATPase heterodimer, the DNA exit gate conformation could be oscillating between the resting and the extended/productive states. However, since residual aspecific crosslinking is still seen in the experiments, they should be further optimized.

Altogether, the comparison between the crystal and the cryo-EM structures reveal that the SMC3 ATPase is in an inactive state when not engaged in the heterodimer and stabilized by NIPBL and DNA. The crosslinking assays show that the SMC3 coiled coil is indeed flexible and can oscillate between an inactive state of the SMC3 ATPase, hereafter named the resting state, and a productive state once bound to SMC1A, NIPBL and DNA in the presence of ATP. Together with the evidence of the cohesin ATPase activity enhancement in the presence of NIPBL and DNA (Davidson et al., 2019; Kim et al., 2019), this resting conformation of the DNA exit-gate could provide a molecular explanation to why the cohesin ATPase is intrinsically latent and needs activation by NIPBL and DNA, which reposition the SMC3 coiled coil into an extended/productive state. A model for the molecular mechanisms underlying activation of the cohesin ATPase by NIPBL and DNA is therefore proposed in the Discussion section.

RESUME EN FRANÇAIS DES EXPERIENCES REALISEES ET DES METHODES UTILISEES

I. Caractérisation structurale des domaines ATPase de SMC1A et de SMC3

A. Données préliminaires

La caractérisation structurale du module ATPase de la cohésine humaine est une étape fondamentale, puisque l'organisation tri-dimensionnelle des composants du domaine moteur de la cohésine va aider à comprendre ses mécanismes, et comment ils sont impliqués dans les diverses fonctions de la cohésine.

Avant le début de mon travail de thèse, mon équipe d'accueil avait débuté la caractérisation structurale des domaines ATPase de SMC1A et de SMC3 de la cohésine humaine. Des constructions de chacun des domaines ATPase ont été clonés par le Dr. Tajith Shaik. Chaque construction était soit native (wild type, WT), soit portant une mutation du résidu catalytique glutamate en glutamine, en E1157Q et E1144Q respectivement pour SMC1A et SMC3 (Figure 18). La mutation du glutamate en glutamine (EQ) a pour objectif de réduire l'activité ATPase de chaque complexe et de permettre leur stabilisation dans une structure complexe-ligand qui convient aux expériences de cristallisation. Les constructions du domaine ATPase de SMC1A et de SMC3 ont été respectivement co-exprimés et co-purifiés avec les domaines C- et N-terminal respectifs de liaison à RAD21, contenant une étiquette de purification 10 Histidine à leur extrémité C-terminale.

La structure de SMC1A-CC/RAD21C WT dans sa conformation en absence de ligand (apo) a été résolue par le Dr. Tajith Shaik pendant son travail de thèse. De plus, il a également obtenu des résultats préliminaires issus d'expériences d'ITC sur les complexes SMC1A-CC/RAD21C et

SMC3/RAD21N, qui avaient pour but d'analyser l'affinité de liaison de l'ATP à ces complexes. Les expériences d'ITC devaient néanmoins être poursuivies et optimisées.

Lorsque j'ai rejoint mon équipe de recherche pour mon travail de thèse, j'ai continué la caractérisation biochimique, biophysique et structurale des domaines ATPase de SMC1A et de SMC3 de la cohésine humaine sous leurs formes apo et liées à l'ADP et à l'ATPγS (analogue non hydrolysable de l'ATP, utilisé afin de stabiliser le complexe protéine-ligand). La section suivante décrit le travail de recherche que j'ai effectué pendant ma thèse.

B. Expression et purification des domaines ATPase de SMC1A et SMC3 de la cohésine humaine

En suivant les protocoles d'expression et de purification déjà établis dans mon équipe, j'ai exprimé et purifié séparément les complexes SMC1A-CC-EQ/RAD21C et SMC3-CC-EQ/RAD21N, par chromatographie d'affinité sur résine TALON, suivi de la digestion de l'étiquette de purification et d'une chromatographie par exclusion de taille.

C. Essais de cristallisation, optimisation des cristaux et diffraction de rayons-X

L'études des protéines par cristallographie à rayons-X est une méthode expérimentale qui permet de déterminer l'organisation spatiale des atomes constituant la protéine, et donc de résoudre sa structure tridimensionnelle (3D). Ce résultat est achevé en 4 étapes principales, après le clonage, l'expression et la purification de la protéine ou du complexe de protéines d'intérêt (Figure 19) : (1) tout d'abord, la protéine d'intérêt est cristallisée. La protéine est exposée à divers agents précipitants au travers de criblages des conditions de cristallisation, en variant d'autres paramètres physico-chimiques comme le pH des conditions, la température ou encore la force ionique. Ce processus est l'une des étapes limitantes majeures, car il requiert l'obtention de cristaux possédant de bonnes qualités de diffraction. Cependant, ce n'est pas

toujours le cas lors des premiers essais, notamment car le temps de cristallisation d'une protéine ne peut pas être prédit, ou bien car les premiers cristaux obtenus nécessitent souvent de longues étapes d'optimisation avant de pouvoir les utiliser dans l'étape suivante, (2) l'exposition des cristaux aux rayons-X dans des expériences de diffraction. Les clichés de diffraction qui peuvent ainsi être obtenus à cette étape sont ensuite (3) traitées afin de calculer une carte de densité électronique, moyennant la position dans l'espace de chaque atome composant la protéine. La carte obtenue permet ainsi de (4) construire, d'affiner et de valider un modèle de la structure 3D de la protéine analysée.

Afin de caractériser structuralement les domaines ATPase de SMC1A et de SMC3 de la cohésine humaine, j'ai effectué des tests de cristallisation de SMC1A-CC-EQ/RAD21C et de SMC3-J-EQ/RAD21N, initialement dans des criblages de conditions de cristallisation commerciaux, en absence ou en présence de 2 mM ATPγS ou d'ADP. Afin de capturer les états conformationnels des complexes pendant les étapes intermédiaires e leur cycle d'hydrolyse de l'ATP, j'ai également effectué les tests de cristallisation en présence de 2 mM de ADP-BeF3, ADP-VO4 et ADP-AlF3, des analogues de l'ATP correspondant respectivement à l'état pré-hydrolyse, l'état intermédiaire et l'état post-hydrolyse. J'ai effectué les essais de cristallisation avec des concentrations de protéine et à des températures variables. J'ai ainsi obtenu des cristaux de SMC1A-CC-EQ/RAD21C et de SMC3-J-EQ/RAD21N (Figure 20, Figure 21), qui ont été soumis à la diffraction de rayons-X aux synchrotrons SOLEIL (France) et SLS (Suisse).

Plusieurs cristaux de SMC1A-CC-EQ/RAD21C ont diffracté à de bonnes résolutions, entre 1.7 et 2.5 Å, et j'ai pu résoudre les structures dans la forme apo et en présence de ATPγS et d'ADP. En présence d'analogues de l'ATP pendant le cycle ATPase aucun cristal de SMC1A-CC-EQ/RAD21C n'a été obtenu.

Néanmoins, les cristaux de SMC3-J-EQ/RAD21N n'ont pas diffracté où ont diffracté à des résolutions très faibles ne permettant pas de résoudre la structure 3D. J'ai donc effectué l'optimisation extensive à la fois des conditions de purification et de cristallisation de la protéine. Notamment, la mesure du ratio d'absorbance à 260 nm et à 260 nm a montré que même après purification, les échantillons de SMC3-J-EQ/RAD21N étaient contaminés avec des acides

nucléiques, qui pouvaient possiblement de fixer à a protéine et ainsi avoir un impact négatif sur maille cristalline des cristaux et affecter de façon négative leur diffraction des rayons-X (Figure 22, Figure 23). Malgré ces étapes d'optimisation, les cristaux obtenus n'ont pas diffracté.

Un autre motif pour lequel les cristaux n'avaient pas une bonne diffraction où n'ont pas diffracté pouvait être dû à la taille du domaine superhélice de SMC3, dont la flexibilité pourrait affecter la cristallisation de la protéine et la maille des cristaux. Afin de stabiliser la construction du domaine ATPase de SMC3 et d'améliorer la qualité de diffraction des cristaux, j'ai donc conçu et effectué le clonage moléculaire de nouvelles constructions du domaine ATPase de SMC3, avec le domaine superhélice raccourci (Figure 24). J'ai appliqué la même méthode au domaine ATPase de SMC1A, dans le but de parvenir à cristalliser le complexe avec les analogues de l'ATP pendant le cycle ATPase (Figure 24).

Parmi ces constructions, SMC1A-CCsh et SMC3-CC étaient particulièrement bien exprimées et/ou stables en solution (Figure 25, Figure 26, Figure 27, Figure 28), et ont donc été utilisées dans de nouveaux essais de cristallisation. De nombreux cristaux ont ainsi été obtenus pour SMC1A-CCsh-EQ/RAD21C et pour le WT ou le EQ mutant SMC3-CC/RAD21N, en présence d'ATPγS, d'ADP, ADP-BeF$_3$, ADP-VO$_4$ et de ADP-AlF$_3$ (Figure 29, Figure 30). En absence de ligand, des cristaux de SMC1A-CCsh/RAD21C ont également été obtenus. Pour SMC3-CC/RAD21N cependant, les amas de petits cristaux obtenus en absence de ligand n'étaient pas utilisables pour des expériences de diffraction (Figure 30), et leur optimisation n'a pas abouti à des cristaux de meilleure qualité. J'ai donc testé une autre construction, SMC3-CC-2 pour des essais de cristallisation en absence de ligand. J'ai pu obtenir des cristaux de SMC3-CC-2-EQ/RAD21N, qui après optimisation ont diffracté les rayons-X.

D.   Résolution de la structure des domaines ATPase de SMC1A et de SMC3

Plusieurs cristaux que j'ai obtenus par les tests de cristallisation et par leurs optimisations ont permis la collecte de plusieurs jeux de données de diffraction différents. J'ai ainsi obtenu et

traité des jeux de données ayant diffracté a des résolutions entre 1.36 et 3.11 Å, et j'ai pu résoudre plusieurs structures des domaines ATPase de SMC1A et SMC3 humaines, par remplacement moléculaire. Les structures de SMC1A-CC-EQ/RAD21C et de SMC1A-CCsh-EQ/RAD21C ont été résolues sous forme apo et liées à l'ATPγS et à l'ADP. La structure du WT ou mutant EQ de SMC3-CC/RAD21N a été résolue sous forme liée à l'ATPγS et à l'ADP. L'une des structures SMC1A-CCsh-EQ/RAD21C ainsi que la structure de SMC3-CC-2/RAD21N ont été résolues avec une molécule de SO_4^{2-} liée au site actif, issu des conditions de cristallisation. Les conditions de cristallisation, les statistiques du traitement des données ainsi que la vue d'ensemble des structures résolues sont fournis dans la partie Résultats et dans Annexes 3 et 4.

L'analyse des structures obtenues des domaines ATPase de SMC1A et de SMC3 ont révélé des différences structurales entre les différents états, apo et lié à l'ATPγS et à l'ADP, dévoilant ainsi des dynamiques conformationnelles distinctes entre SMC1A et SMC3 pendant le cycle ATPase de la cohésine humaine.

II. Caractérisation biochimique des domaines ATPase de SMC1A et de SMC3 par titration calorimétrique isotherme et tests d'activité ATPase

Les expériences d'ITC préliminaires ont été optimisées, afin d'analyser les propriétés de liaison à l'ATP et à l'ADP des domaines ATPase de SMC1A et de SMC3, et de comprendre comment la liaison de ces nucléotides conduit aux dynamique conformationnelles observées sur les structures cristallographiques.

Les expériences d'ITC ont été réalisées et analysées en collaboration avec le Dr. Eric Ennifar et le Dr. Karl Brillet à Institut de Biologie Moleculaire et cellulaire (IBMC; Strasbourg, France). Les résultats révèlent des modes d'interaction distincts de l'ATP et de l'ADP avec les domaines ATPase de SMC1A et de SMC3, en accord avec les observations structurales : le domaine ATPase de SMC1A possède une plus grande flexibilité structurale que l'ATPase de SMC3,

et la liaison de l'ATP à SMC1A induit d'importants changements conformationnels au sein de la structure.

Additionnellement, des tests d'activité ATPase montrent que l'activité d'hydrolyse de l'ATP par la cohésine humaine est très faible (moins d'une molécule d'ATP par minute) (Figure 31, Figure 32), en accord avec des résultats publiés dans la littérature (Kim et al., 2019, Davidson et al., 2019), et suggèrent que cette activité nécessite probablement un effet d'activation.

III.   Analyse de l'interaction des domaines ATPase de SMC1A et de SMC3 avec l'ADN

Toutes les fonctions de la cohésine dépendent de sa capacité à interagir avec la chromatine. Curieusement, lors les expériences initiales de purification des complexes, les ATPases SMC1A et SMC3 ont toujours été co-purifiées avec des acides nucléiques contaminants, malgré la purification en deux étapes par chromatographie d'affinité et d'exclusion de taille.

De façon intéressante, la structure cristalline du module ATPase de Rad50, qui est un membre de la famille SMC, a été résolue dans plusieurs conformations en présence d'ADN (Liu et al., 2016 ; Rojowska et al., 2014 ; Seifert et al., 2016). Le fragment d'ADN est soit lié à une région du domaine superhélice proximale au module ATPase, soit lié au-dessus du module ATPase. De plus, l'analyse du module ATPase de l'archée *P. furiosus* a suggéré que le résidu R59 du motif conservé « R-loop » et localisée au-dessus du site de liaison de l'ATP pourrait être impliqué dans l'activation de l'ATPase dépendante de la liaison à l'ADN (Lammens et al. 2004).

Compte tenu de ces observations, il est donc pertinent de se demander si l'ADN a une action directe sur le module cohésine ATPase humaine et sur son activité, et d'étudier les conséquences structurelles de la liaison à l'ADN. Par conséquent, j'étais curieuse de voir si certaines de ces observations pouvaient s'appliquer à la cohésine humaine, et si l'ADN pouvait se lier directement aux ATPases SMC1A et SMC3 et moduler leur activité.

J'ai donc effectué des essais de cristallisation des domaines ATPase de SMC1A et de SMC3 en présence d'un court fragment d'ADN. Les cristaux obtenus n'ont cependant pas diffracté, où ont diffracté et la résolution de la structure a révélé seulement SMC1A-CCsh-EQ/RAD21C, en absence d'ADN. Alors que ces expériences de cristallisation étaient en cours d'optimisation, la structure de l'hétérodimère de SMC1A et SMC3 de la cohésine humaine, en présence d'un fragment d'ADN, stabilisé par NIPBL et par STAG1, résolue par cryo-microscopie électronique (cryo-EM), a été révélée à la communauté de recherche (Shi et al., 2020). J'ai donc décidé de ne pas approfondir ces aspects.

Par conséquent, je me suis concentrée sur l'analyse du rôle potentiel des R-loop de la cohésine et de leur impact dans le cycle ATPase, afin de déterminer si le mécanisme d'activation de l'ATPase par l'ADN via la R-loop pourrait être conservé chez l'homme. J'ai donc testé l'effet des mutants R57A et la délétion de la boucle V58-R62, qui est une mutation retrouvée dans des cas du syndrome Cornelia de Lange. Les résultats suggèrent que contrairement au Smc de *P. furiosus*, le résidu R57 de la cohésine humaine n'est pas suffisant pour augmenter significativement l'activité ATPase de la cohésine humaine, tandis que la boucle V58-R62 semble participer à la reconnaissance et/où à la fixation de l'ADN pour l'activation de l'ATPase de SMC1A par l'ADN (Figure 34).

IV.   Analyse d'une nouvelle conformation identifiée pour l'interface entre SMC3-RAD21

Les structures résolues par cristallographie du domaine ATPase de SMC3 ont révélé une conformation « pliée » du domaine superhélice jusqu'alors jamais observée, à l'interface entre SMC3 et RAD21N, soit à la « porte de sortie » de la molécule d'ADN du complexe. Afin d'analyser si la conformation pliée de l'interface SMC3-RAD21N existe en solution et pourrait éventuellement être pertinente d'un point de vue mécanistique, j'ai analysé les différentes conformations de l'interface par jonction chimique en présence de l'agent de réticulation

bismaleimidoethane (BMOE) (Figure 35, Figure 36, Figure 37). Ces expériences ont montré que l'état replié de l'interface entre SMC3 et RAD21N existe en solution, tout comme la conformation « relevée » ou « productive » observée sur la structure cryo-EM de Shi et al., 2020, suggérant ainsi que cette interface est flexible. Cela suggère également pourquoi l'activité ATPase de la cohésine humaine est basale en absence de l'ADN, mais surtout de NIPBL, qui maintient cette interface dans la conformation « poductive » et promeut l'activité ATPase.

RESULTS - THE DISTINCT CONFORMATIONAL DYNAMICS OF HUMAN SMC1A AND SMC3 ATPASE HEADS REVEAL A RESTING STATE FOR THE COHESIN DNA EXIT GATE

The following paper presenting the results is in preparation for publication.

During my thesis work, I made the following contributions to the project:

- Optimization of the protein purification conditions, in order to remove co-purified contaminant nucleic acids;

- Design of the constructs, molecular cloning, protein expression, purification, crystallization, and structure determination of SMR1ACCsh-EQ-apo, SMR1ACCsh-EQ-loop-SO$_4$, SMR1ACCsh-EQ-ADP, SMR1ACCsh-EQ-ADP-Mg, SMR1ACCsh-EQ-AGS-Mg, SMR3CC-EQ-ADP, SMR3CC-WT-ADP-Mg, SMR3CC-EQ-AGS1-Mg, SMR3CC-EQ-AGS2-Mg and SMR3CC_2-EQ-SO$_4$;

- Protein expression, purification, crystallization and structure determination of SMR1ACC-EQ-apo, SMR1ACC-EQ-loop, SMR1ACC-EQ-ADP and SMR1ACC-EQ-AGS-Mg;

- Protein expression and purification for ITC experiments;

- Protein expression and purification, followed by ATPase activity assays and BMOE crosslinking assays;

- Molecular cloning of the SMC1ACC R57A mutant.

The distinct conformational dynamics of human SMC1A and SMC3 ATPase heads reveal a resting state for the Cohesin DNA exit gate

Gomes, M.[1], Landwerlin, P. [1], Diebold-Durand[1], M.-L, Shaik[1], T.B., Durand, A.[1], Brillet, K.[2], Troesch, E. [1], Antony, P[1]., Ennifar, E.[2] and Romier, C. [1,§]

[1]Université de Strasbourg, CNRS, INSERM, Institut de Génétique et de Biologie Moléculaire et Cellulaire, UMR 7104, U 1258, 67400, Illkirch, France.

[2]Architecture et Réactivité de l'ARN, Institut de biologie moléculaire et cellulaire (IBMC), UPR 9002 du CNRS, Université de Strasbourg, 15 Rue René Descartes, 67084 Strasbourg Cedex, France.

[§]Correspondence should be addressed to Christophe Romier (romier@igbmc.fr).

Keywords: Genome organization, Chromatin, SMC complexes, Cohesin, ATPase activity, Coiled coils, Conformational changes, DNA exit gate.

Abstract

The Cohesin complex plays key roles in sister chromatid cohesion, chromosome segregation, DNA repair, transcription and during genome and chromatin structure organization by DNA loop extrusion. ATP binding and hydrolysis is key to the various functions of this complex. The molecular basis of the Cohesin ATPase activity, how this activity is enhanced or impaired by regulatory proteins and DNA, and how these influence Cohesin conformational dynamics and functions remains to be fully characterized. Here we show that the human Cohesin SMC1A and SMC3 ATPase heads have different conformational dynamics upon switching between the apo, ATP-bound and ADP-bound states, and during their engagement. Unexpectedly, our structures reveal that SMC3 can adopt an ATP hydrolysis inactive state caused by the resting conformation of the Cohesin DNA exit gate which is formed upon the interactions between the globular domains of the human SMC3 head and of the RAD21 N-terminal region. This resting conformation of the Cohesin DNA exit gate is compatible with SMC1A and SMC3 heads engagement but not with the formation of a productive SMC3 ATPase active site. Formation of the productive complex requires conformational changes of the SMC1A and SMC3 heads that can partially be induced by the heads engagement, but that need further amplification and stabilization by DNA and the Cohesin loader NIPBL. Collectively, our results decipher the molecular mechanisms supporting the ATPase activity of the human Cohesin ATPase module and identify the DNA exit gate as an important evolutionary mechanistic element of Cohesin.

Introduction

Structural Maintenance of Chromosome (SMC) complexes play key roles in the organization of genomes in all three kingdoms of life [1-9]. All SMC complexes share a common core architecture composed of two Smc proteins and a kleisin protein that form large tripartite structures that can associate with DNA. While prokaryotes have a single SMC complex composed of a homodimer of Smc proteins in addition to their kleisin subunit, eukaryotes have three SMC complexes that each accommodate a distinct Smc heterodimer and a specific kleisin [6,10]. Although the SMC complexes share similar mechanisms, they have evolved specific functional roles that are supported by distinct regulatory subunits.

Smc proteins harbour at each of their extremities a hinge domain and an ABC-like ATPase head domain (HD) that are separated by a long coiled coil [11-23]. Smc proteins dimerize through their hinge domains, while the kleisin subunit interacts in an asymmetric manner through its N- and C-terminal regions with both HDs, thus forming the tripartite ring characteristic of SMC core complexes [15,18,21,24,25].

In addition, Smc dimerization also occurs when Smc ATPase HDs bind two ATP molecules in a head-to-tail manner, a transient engagement which is lost upon ATP hydrolysis [26]. Importantly, ATP binding and hydrolysis as well as DNA binding can induce structural changes that contribute to the different conformations of the various SMC complexes, although the intrinsic flexibility between the ATPase modules and the arms of these complexes still questions on how these structural changes can be propagated [11-14,17,19,20,23,27-34].

Among the eukaryotic SMC complexes, Cohesin is acting in the cohesion of sister chromatids, the segregation of chromosomes, in DNA repair, in transcription regulation, and in the organization of the genome by forming chromatin loops and Topological Associated Domains (TADs) through a loop extrusion mechanism [1,2,4-9]. The core Cohesin complex is composed of the SMC1A[Smc1] and SMC3[Smc3] Smc subunits and of the RAD21[Scc1] kleisin subunit (human names used throughout unless species-specific proteins are discussed; *S. cerevisiae* names are hereafter given in superscripts when diverging significantly from the human names). The importance of human Cohesin is highlighted by the large number of mutations observed for its core and regulatory

subunits in numerous solid and haematologic cancers and in neurodevelopmental disorders termed cohesinopathies [35-38].

Formation of the Cohesin complex involves SMC1A and SMC3 heterodimerization through their hinge domains and the bridging by the RAD21[Scc1] kleisin subunit of SMC1A and SMC3 HDs [13,15,16,18,20,22,23,34]. Specifically, RAD21[Scc1] N-terminal region (RAD21N) is formed of a globular domain followed by a long α-helix. This latter helix interacts with the coiled coil (CC) emerging from the globular domain (GD) of SMC3 HD. In contrast, RAD21[Scc1] C-terminal region (RAD21C) folds into a Winged-Helix Domain (WHD) that directly binds to the globular domain of SMC1A HD, thus closing the Cohesin ring.

Cohesin displays a large conformational flexibility and has initially been observed in an open conformation with its two CCs dissociated from each other [11]. More recent observations have shown that Cohesin also adopts a rod conformation, where its CCs are interacting with each other by co-alignment [27,39]. SMC1A and SMC3 CCs can also bend sharply at their elbow elements, bringing the hinges closer to the ATPase HDs [12,13,20,23,34]. The molecular basis for Cohesin conformational changes is still matter of debate, but ATP binding and hydrolysis by this complex as well as its interaction with DNA appear major effectors of this conformational dynamics [13,20,23,26,27,34].

Specifically, in addition to the SMC1A-hinge/SMC3-hinge, SMC1A-HD/RAD21C and SMC3-HD/RAD21N interfaces that enable the formation of the Cohesin ring, ATP binding to the HDs causes their transient engagement and is preventing Cohesin CCs to interact with each other, at least in the vicinity of the HDs [13,20,23,26]. In contrast, in the absence of ATP and in the rod conformation, the interaction of the Cohesin CCs brings the two HDs in close vicinity but in a non-engaged, juxtaposed state [27,34]. These different conformational rearrangements create various compartments that can be used by the Cohesin complex for its interactions with DNA.

Binding of Cohesin to DNA can be achieved either in a topological or in a pseudo-topological manner. Specifically, Cohesin entraps DNA in a topological manner during sister chromatid cohesion [13,20,23,40-46]. In contrast, the formation of loops by the loop extrusion mechanism implies a pseudo-topological binding [47-49]. Entry of the DNA in the Cohesin ring during entrapment has been shown to occur through the SMC1A-hinge/SMC3-hinge interface (DNA entry gate) [46,50,51].

In contrast, DNA exit from the Cohesin ring has been shown to occur through the SMC3-HD/RAD21N interface (DNA exit gate) [51-54].

In vertebrates, opening of the DNA exit gate during prophase represents a major mechanism (prophase pathway) for Cohesin removal from the genome [55,56], while yeasts make principally use of the ubiquitous alternative removal mechanism based on the proteolytic cleavage of RAD21[Scc1] by Separase during anaphase [57-60].

Most SMC complexes interact with KITE (Kleisin-Interacting Tandem winged-helix Elements) proteins to regulate their functions. In contrast, Cohesin and its eukaryotic paralogue Condensin (SMC4/SMC2/CAP-H[Brn1]), which is involved in chromosome condensation, bind to HAWK (HEAT-repeat proteins Associated With Kleisin) regulatory proteins to perform their biological roles. Notably, Cohesin associates with three HAWKs: STAG1 or STAG2 (SA[Scc3]), NIPBL[Scc2], and PDS5A or PDS5B (PDS5) [6,10]. Specifically, SA[Scc3] binds to the second half of RAD21[Scc1] and is considered to stay mostly stably associated with Cohesin. SA[Scc3] is involved in both DNA loading and unloading by Cohesin and interacts directly with the DNA [23,44,56,61,62]. In addition, SA[Scc3] is able to bind to the Cohesin hinges and, in vertebrates, cooperates with the transcription factor/insulator CTCF (CCCTC-binding Factor) to organize chromatin loops and Topological Associated Domains (TADs) [23,63].

In contrast, the NIPBL[Scc2] and PDS5 HAWKs bind to the first half of RAD21[Scc1] in a mutually exclusive manner and play different functional roles. NIPBL[Scc2], in complex with MAU2[Scc4], is responsible for the loading of Cohesin onto the genome, a process requiring HDs engagement [64]. This initial loading appears non-topological. Cohesin can then either stably bind DNA topologically or translocate on the genome and form chromatin loops and TADs by loop extrusion (LE), both mechanisms requiring NIPBL[Scc2] [40,42,47,65-68]. How Cohesin switches between these two states remains unclear. While ATP hydrolysis is strictly required for LE, its requirement for entrapment is debated [43,47,49,65,66]. Importantly, NIPBL[Scc2] plays a unique and essential role in stimulating Cohesin ATPase activity [47,49,69]. This stimulation is higher in the presence of DNA although DNA alone barely shows a stimulatory activity.

The recent structures of *S. cerevisiae*, *S. pombe* and human Cohesins bound to DNA and to NIPBL[Scc2] C-terminal domain have shed light on this activation [13,20,23]. Specifically, NIPBL[Scc2] and

the DNA form a tight complex with the engaged SMC1A and SMC3 HDs. NIPBL[Scc2] is favouring HDs engagement, interacting with both SMC1A and SMC3 GDs and CCs, these three proteins firmly clamping the DNA between themselves. In addition, SMC1A and SMC3 hinges, NIPBL[Scc2] and STAG1[Scc3] are observed interacting together, albeit more loosely, STAG1[Scc3] also binding to the DNA in the human structure [13,20,23]. In these structures, the DNA is pseudo-topologically bound into the upper chamber formed by the engaged SMC1A and SMC3 HDs, but is prevented from entering the Cohesin ring by the ~60 amino acids region of RAD21[Scc1] that links its SMC3 CC-binding N-terminal helix and its NIPBL[Scc2]-interacting region. These structures, which have been obtained in the presence of SMC1A and SMC3 mutants having a low hydrolysis rate, have been described as a pre-ATP hydrolysis state that we hereafter refer to as productive state.

In contrast to NIPBL[Scc2], the PDS5 HAWK plays a dual role by stabilizing or removing Cohesin from the genome, depending on additional partners. In the case of Cohesin stabilization on the genome, this process is enhanced by the recruitment by PDS5 of ESCO1 or ESCO2 (ESCO[Eco1]) that acetylate SMC3 on two specific lysines, K105 and K106 (human numbering throughout). This acetylation (i) stabilizes the interaction of Cohesin with PDS5, (ii) prevents Cohesin interaction with NIPBL[Scc2], and (iii) prevents deacetylation of K105/106 by HDAC8[Hos1] [42,52,70-81]. The recent structure of yeast PDS5 bound to Cohesin in the absence of ATP and DNA shows that PDS5 interaction with the SMC3 HD occurs mostly in the region of these two lysines [34]. In vertebrates, PDS5 further recruits SORORIN that enhance the stabilization of Cohesin on DNA [82-87].

PDS5 also plays an opposite role in Cohesin unloading when complexed to WAPL, which, in vertebrates, competes with SORORIN for a unique binding site on PDS5 [56,84,85,88]. This unloading role of PDS5-WAPL requires SA[Scc3] and the opening of the DNA exit gate. If HDs engagement through ATP binding is required for unloading, ATP hydrolysis is not [71,89]. The mechanisms by which SORORIN and WAPL act to stabilize or unload Cohesin are not understood. In vertebrates, the ATP-dependent interplay between NIPBL[Scc2], PDS5, SORORIN and WAPL, together with CTCF, leads to the dynamic regulation of chromatin loops, TADs and transcription by Cohesin [63,90-101].

ATP binding and hydrolysis by Cohesin is thus central to the functions of this complex and its regulations. Understanding the molecular basis of Cohesin ATPase activity is therefore crucial to decipher the structure/function relationships of this complex. The current knowledge on the

Cohesin ATPase module remains however fragmented and requires further characterization. Although comparative analyses of the Cohesin and bacterial ATPase modules can yield invaluable information, this latter homodimeric module cannot fully account for the reported asymmetry of the Cohesin ATPase cycle [72,102-104]. Similarly, the comparison with the eukaryotic Condensin ATPase module is restricted due to the different regulations of this paralogous complex.

Here we have characterized the human Cohesin ATPase module by studying the SMC1A and SMC3 HDs independently in their apo, ATP-bound and ADP-bound states as well as upon their engagement. Our results show that the human SMC1A HD adopts an inactive relaxed conformation and displays a specific conformational dynamics in its different nucleotide binding states. In contrast, human SMC3 does not adopt a relaxed conformation and is locked in an ATP-hydrolysis inactive state due to a so far uncharacterized resting state adopted by its DNA exit gate. A stable productive conformation of the SMC1A and SMC3 HDs is not induced upon HD engagement but further requires structural changes caused by DNA and NIPBL[Scc2] binding. Our work reveals similarities but also unexpected differences between the ATPase modules of different organisms, supporting evolutionary divergence, notably at the DNA exit gate. Collectively, our results decipher the molecular mechanisms of the Cohesin ATPase module, including those leading to the activation of the Cohesin ATPase activity in the productive state. We further propose that the switch between the resting and the productive states of the human DNA exit gate contributes actively to the regulation of the interactions of Cohesin with the genome.

Results

Human SMC1A and SMC3 ATPase heads display distinct ADP and ATP binding properties
To study the human SMC1A and SMC3 ATPase heads (HDs), we created different constructs of both HDs by varying the length of the coiled coil (CC) emerging from their globular domain (GD) (Figure 1A; Supplementary Figure 1). The length of each CC was chosen to end up either above the joint element (joint, J), which breaks the CC shortly above the HD, or below the joint element

136

(coiled coils, CC). Shorter constructs of SMC3 were not made to avoid the dissociation of its complex with the RAD21[Scc1] N-terminal domain (RAD21N).

In contrast, since the SMC1A CC is not involved in binding to the RAD21[Scc1] C-terminal WHD (RAD21C), a shorter construct of SMC1A was also made for our structural analyses by removing additional residues in each coil compared to the CC construct (coiled coil short, CCsh) (Figure 1A). The RAD21N and RAD21C constructs were designed according to previous work (Figure 1A; Supplementary Figure 1) [15,18]. Our SMC1ACC, SMC3J, RAD21N and RAD21C constructs turned out to be highly similar to the same regions of SMC1A, SMC3 and RAD21 that could be built in the cryo-EM map of the recent structure of the human Cohesin complex in the productive state [23].

If independent expression of the SMC1A and SMC3 constructs led to insoluble proteins, co-expression of these proteins with their respective RAD21C and RAD21N partners however yielded soluble SMC1A-HD/RAD21C and SMC3-HD/RAD21N complexes. The complex containing the SMC1AJ construct was poorly soluble and could not be purified in large amounts, preventing its subsequent use. All other complexes could be purified in large scale using similar purification protocols. Initial crystallization attempts with the different complexes yielded various crystal forms but the SMC3J/RAD21N crystals repeatedly showed poor diffraction, possibly due to a certain degree of flexibility of the joint element, as already reported [23]. We therefore used the SMC3CC construct instead for our investigations.

We first assessed the ATPase activity of the SMC1ACC/RAD21C and SMC3CC/RAD21N complexes, either alone or when mixed together. The independent complexes displayed an extremely low ATPase activity. SMC1CC/RAD21C showed however a twofold higher activity than SMC3CC/RAD21N (Figure 1B). Upon mixing of the two complexes, a 2.5 fold increase in ATPase activity was observed compared to SMC1ACC/RAD21N alone, this activity remaining however very low, as previously reported [47,49,69] (Figure 1B). We next measured the Kd of the SMC1ACC/RAD21C and SMC3CC/RAD21N complexes for ADP and ATP using isothermal titration calorimetry (ITC). Only the WT complexes were used for measurements with ADP. However, due to the residual low ATPase activity of both complexes, Kd measurements for ATP were carried out both with the WT and the low ATPase activity EQ mutants (SMC1A-E1157Q and SMC3-E1144Q) complexes, in the presence of either ATP or the low hydrolysable ATPγS analogue.

Initial ITC measurements with the SMC1ACC/RAD21N complex showed that ATP binding can be observed at 5°C but not at 30°C. Therefore, all subsequent measurements were performed at 5°C. Both complexes behaved differently in these assays. In the case of ADP, while both complexes displayed a single binding event, the measured Kd for SMC1ACC/RAD21C (~28 μM) was more than 10 times higher than that for SMC3CC/RAD21N (~2 μM) (Figure 1C-D; Supplementary Figure 2A-B).

In contrast, in the presence of ATP and ATPγS, while the SMC1ACC/RAD21C complex showed a single binding event, the SMC3CC/RAD21N complex unambiguously showed two events (Figure 1C; Supplementary Figure 2A-B). Since ADP binding showed a single event, we reasoned that the second binding event could be due to a homodimerization of SMC3CC/RAD21N in the presence of ATP and ATPγS. We therefore processed the ITC data for SMC3 with a model with two consecutive binding events (SMC3-HD + ATP -> SMC3ATP + SMC3ATP -> SMC3ATP homodimer). Processing of the data with this model gave excellent fitting of the experimental curves and yielded different Kds for both events, one in the μM range and one in the low nM range (Figure 1C-D; Supplementary Figure 2B). Since the measured Kd of ADP for SMC3CC/RAD21N was in the μM range, we assumed that the Kd of ATP and ATPγS for the same complex corresponded to the higher measured Kd, whereas the low Kd would correspond to the homodimerization event.

Comparative analysis of the different Kds showed that both ATP and ATPγS have Kds in a similar range when binding to SMC1ACC/RAD21C and SMC3CC/RAD21N (Figure 1D). The Kd of ATPγS is slightly lower than that of ATP for both complexes, while the WT complexes yield slightly lower Kds than the EQ mutants, these results most likely reflecting differences in physicochemical properties. In contrast, the homodimerization event of SMC3CC/RAD21N displays Kds in the low nM range, the strongest binding with a Kd of 1 nM being observed in the presence of the WT complex and of ATP (Figure 1C-D; Supplementary Figure 2B).

Comparison of the ITC-measured thermodynamic signatures of the SMC1ACC/RAD21C and SMC3CC/RAD21N complexes binding to ADP, ATP and ATPγS also showed significant differences between both complexes. Whereas the binding reaction for SMC1A/RAD21C is entropy-driven, suggesting an increase in solvent entropy, both binding events for SMC3/RAD21N are enthalpy

driven, with the strongest enthalpy difference observed for the homodimerization event (Figure 1D-E; Supplementary Figure 2C-D).

A closed conformation of the SMC1A HD P-loop does not prevent nucleotide binding

We next characterized structurally by x-ray crystallography the different complexes purified. Crystallization assays for the SMC1A HD were performed without nucleotide or in the presence of ADP or the low-hydrolysable ATPγS nucleotide, or in the presence of the different ATP-hydrolysis mimics ADP-BeF$_3$, ADP-Al-F$_4$ and ADP-VO$_4$. Crystallization attempts for both SMC1ACC/RAD21C and SMC1ACCsh/RAD21C complexes yielded crystals diffracting between 1.36 and 2.50 Å and belonging to different space groups (Supplementary Tables 1 and 2). These complexes could be crystallized without nucleotide or in the presence of ADP or ATPγS, but not with ADP-BeF$_3$, ADP-AlF$_3$ and ADP-VO$_4$. Structure determination was carried out by molecular replacement using the *S. cerevisiae* Smc1-HD/ATPγS structure as model (PDB code 1w1w).

Inspection of the electron densities at the nucleotide binding site for the various complexes revealed five binding states: apo, ADP-bound, ADP-Mg-bound, ATPγS-Mg-bound, and a fifth state that we termed loop, where the phosphate loop (Walker A motif/P-loop) was found in two alternative conformations: closed and open. In the apo structures, the P-loop adopts exclusively the closed conformation, which is incompatible with nucleotide binding, while in the nucleotide-bound structures, only the open conformation is observed (Figure 2A). In the loop structures, both conformations are found, which are caused by the low occupancy binding into the nucleotide-binding pocket of an ion, a cleaved phosphate or a bound ADP or ATPγS molecule. In addition, in one of the apo structures of the SMC1ACCsh/RAD21C complex, binding of a sulphate ion provided by the crystallization buffer causes the P-loop to adopt an open conformation.

A closed conformation has also been observed in the *Chaetomium thermophilum* Condensin Smc2 HD, which hampers ATP binding [19]. Our different structures and our ITC results show that the closed conformation in SMC1A cannot prevent ATP binding but can set a thermodynamic barrier to this binding. Notably, the closed conformation of the P-loop in human SMC1A is stabilized by a network of water molecules interacting with neighbouring residues (Figure 2B). Thus, the movement of the P-loop from a closed to an open conformation upon nucleotide

binding requires the release of this water network, in agreement with the observed entropy-driven binding of ADP and ATP to SMC1ACC/RAD21C. Interestingly, comparison with the apo structure of *C. thermophilum* Smc1 HD shows that, in this organism, the Smc1 P-loop is in an open conformation [105].

In the ADP-bound, ADP-Mg-bound and ATPγS-Mg-bound structures, the β- and γ-phosphate groups of the nucleotides and the Mg ion organize new water networks together with the Walker B and Q-loop active site elements that stabilize the binding of the nucleotides (Figure 2C). Specifically, in both structures with an Mg ion bound, this ion is coordinated by the hydroxyl of S39 and the β phosphate of the nucleotide. In the ATPγS-Mg-bound structure, the Mg ion is further coordinated by the γ group of the ATPγS nucleotide and by the side chain of Q137 from the Q-loop. Water molecules complete the hexameric coordination of the Mg ion in all structures. The ADP-Mg-bound most likely represents a post-hydrolysis state.

The R-loop in human SMC1A HD is poised for nucleotide binding

In its closed conformation, the human SMC1A P-loop is also stabilized by the formation of a hydrogen bond network formed between the carbonyl oxygen of S36, the hydroxyl of S39 and the side chain amide nitrogen of N40. The N40 side chain is also involved in a hydrogen bond network formed with the side chains of K13, S14, D43 and R57 (Figure 2D). The involvement of this latter residue is completely unexpected since, in bacteria and in fungi, the R-loops, including their arginine residues equivalent to R57, are not seen in density. Specifically, in these organisms, ordering of the R-loop occurs only upon ATPase heads engagement in the bacterial/archeal Smc homodimer, and upon formation of the productive complex in the presence of NIPBL[Scc2] and DNA in yeasts, suggesting that the R-loop helps sensing the formation of these complexes and transmits this information to the ATPase site [13,20,21,25].

In all our structures of the human SMC1A HD, the R-loop is defined in density, including the R57 side chain, and the aforementioned hydrogen bond network involving the N40 and R57 side chains is also conserved in these structures. In the ADP-bound and ATPγS-bound SMC1A HD structures, additional interactions are formed between the N40 side chain and the sugar and α-phosphate groups of these nucleotides. One exception concerns the side chain of S39 that rotates

away upon ATPγS-Mg binding and interacts with the Mg ion. This breaks the hydrogen bond formed between the R57 side chain and the S39 hydroxyl. The additional structural rearrangements induced by ATPγS binding (see below) actually bring the R57 side chain closer to the nucleotide, and the R57 guanidino group hydrogen bonds to the α phosphate of the nucleotide as well as to the signature motif of a symmetry-related complex, still maintaining its interactions with the N40 and D43 side chains (Figure 2E). These interactions are also observed in the human productive complex where the R-loop displays a similar conformation to that observed in our structures [23].

To further characterize the role of R57 in the human SMC1A HD, we have mutated this residue into alanine and analysed the ability of this mutant to hydrolyse ATP. When investigated on its own, the SMC1ACC-R57A/RAD21C mutant complex showed no ATP hydrolysis. However, this loss of enzymatic activity was minimal, compared to the WT complex, when both SMC1ACC-R57A/RAD21C and SMC3CC/RAD21N were mixed together (Figure 1B). In human SMC1A, R57 therefore plays an earlier role in organizing the nucleotide binding site, including by participating to the stabilization of the P-loop closed conformation, prior to HDs engagement. This contrasts to the case of fungi Smc1 where the R-loop appears to play a role once the DNA is bound to the engaged ATPase HDs.

ADP and ATPγS binding to SMC1A HD induce specific structural movements

We next analysed the global structural changes occurring upon ADP and ATPγS binding to the SMC1A HD using only the structures of the SMC1ACC/RAD21C complex since the SMC1ACCsh/RAD21C is locked into a single conformation, whatever its nucleotide binding state (see below). Superposition of the apo and ADP-bound structures, keeping RAD21C as fixed reference, revealed that ADP binding induces rotational movements within the SMC1A HD that are caused by the adjustment of the protein to the bound nucleotide (Figure 3A). Notably, a significant movement of the region of the RecA-lobe interacting with the adenine base is observed toward the RecA-lobe/helical-lobe interface. This in turn appears to force the helical-lobe to rotate away from the RecA-lobe (Figure 3A).

Upon ATPγS and Mg binding, the rotational movement of the RecA-lobe region binding to the adenine moiety toward the RecA-lobe/helical-lobe interface is also observed and even amplified, notably with the P-loop moving toward this interface compared to what is observed in the apo and ADP-bound structures where the P-loop remains identically positioned (Figure 3B). However, in contrast to ADP binding, ATPγS and Mg binding cause the helical-lobe to move toward the RecA-lobe. These two opposite movements bring the Q-loop near the nucleotide binding site, thus enabling the side chain of Q137 to contact the Mg ion (1.9 Å) and the γ group of the ATPγS nucleotide (2.6 Å) (Figure 3C).

The resulting shrinkage of the SMC1A HD upon ATPγS-Mg binding induces a tightening of the RecA-lobe/helical-lobe interface that only requires a few changes in side chain conformations rather than a major reorganization of this interface. Although a crystallographic homodimer of the SMC1ACC/RAD21C complex is formed in the ATPγS-bound crystals, the major structural changes observed are due to the effects caused by the presence of the Mg ion and of the γ group of the ATPγS nucleotide. Resulting from these tightening movements, the position of the ATPγS molecule is different from that of the ADP molecule when keeping RAD21C fixed (Figure 3D). This explains that the R57 side chain can interact with the α-phosphate of ATPγS.

Importantly, the amplitude of the movements caused by the binding of the nucleotides is larger further away from the nucleotide binding site. This is notably the case of the helical-lobe that moves as a rigid domain upon ADP and ATPγS binding. Specifically, we observe that the SMC1A CC rotates by ~10° in opposite directions upon the respective binding of ADP and ATPγS as compared to the central position of this CC observed in the apo structure (Figure 3E). Importantly, we observe that these rotational movements of the CC occur in a same plan that passes through Q137, which explains that this residue reaches in the nucleotide binding site upon ATPγS binding, whereas it moves away upon ADP binding (Figure 3F).

Comparative analysis of the *C. thermophilum* apo Smc1 HD with the *S. cerevisiae* Smc1 HD in its ATPγS-bound and Smc3-engaged forms upon keeping Scc1C fixed for the superposition of these structures similarly shows a planar ~10° rotational movement from a central position of the CC in the apo form to a shrinked ATPγS-bound form (Supplementary Figure 3A). We also observe a slight movement within the yeast RecA-lobe upon ATPγS binding but away from helical-lobe.

Nevertheless, the helical-lobe movement is sufficient to position the glutamine of the Q-loop in hydrogen bonding distance to the Mg ion (2.0 Å) and to the γ group of the nucleotide (2.7 Å) (Supplementary Figure 3B).

We next compared our SMC1ACC/RAD21C structures with that of the SMC1A HD in the productive complex, keeping RAD21C fixed as reference for superposition. We do not observe major structural changes within the RecA-lobe in the productive structure compared to our ATPγS-bound structure, with the exception of the P-loop which retains, in the productive state, its position as in our apo structure. In contrast, a major movement of the helical-lobe with respect to the RecA-lobe is observed between the two structures (Figure 4A). This movement, which has been termed "lever effect" in the bacterial case [21], induces a sliding of the helical-lobe along the RecA-lobe. The amplitude of this movement is very large in human SMC1A, as observed by the displacement (~7-8Å) of the second signature-coupling helix of SMC1A that precedes the N-terminal coil of this HD. Strikingly, this movement is almost perpendicular to the aforementioned planar rotational movement observed upon ADP and ATPγS binding to the SMC1A HD, showing that these two conformational changes are not related (Figure 4A-B).

Interestingly, upon leverage the SMC1A CC in the productive state is repositioned in a central position, as observed in our ATPγS-bound structure (Figure 4B). The lever effect observed in the productive mode is sufficient to bring Q137 in a productive distance (~2 Å) to the Mg ion. In the yeast case, a similar lever effect is observed but its amplitude is slightly smaller (~4-5Å) and the CC in the productive complex has the same position as in the ATPγS-bound and the Smc3-engaged Smc1-HD/Scc1N structures (Supplementary Figure 3C).

SMC1ACC adopts a relaxed state, whereas SMC1ACCsh resembles SMC1A HD in the productive state

Analysis of the interface between the RecA-lobe and the helical-lobe between the human SMC1ACC/RAD21C and productive SMC1A HD structures reveals that in order to accommodate this sliding between both lobes, a significant reorganization of the interface between these lobes is required. Since ADP and ATPγS binding are not sufficient to induce such a reorganization, we conclude that SMC1ACC adopts a stable relaxed conformation different from that of the

productive state. In the yeast case, a similar relaxed conformation is observed that is also not significantly modified upon ATPγS binding and HDs engagement. In yeast, the reorganization of the RecA-lobe/helical-lobe interface is not as extensive as in the human case due to the smaller lever effect required to reach the productive state.

Surprisingly, superposition of the SMC1ACCsh/RAD21C structures and of the productive state SMC1A HD structure reveals that the SMC1ACCsh/RAD21C structure is very similar to that of the productive state. Notably, only a small lever effect is observed between these two structures that also have a similar RecA-lobe/helical-lobe interface (Figure 4C). In addition, contrarily to the SMC1ACC/RAD21C complex that crystallized in different space groups depending on its nucleotide-binding state, the SMC1ACCsh/RAD21C complex crystallized in the same space group whatever the nucleotide-binding state, with only small variations in the unit cell, and no homodimerization in presence of ATPγS. Nucleotide binding does not induces large structural changes, although a slight shrinkage is induced upon ATPγS binding, caused by the rotation of the adenine-binding region of the N-lobe toward the N-lobe/C-lobe interface, including a slight movement of the P-loop. In contrast, no rotation of the C-lobe is observed. These different movements position Q137 close but not at an optimal distance to the Mg ion (2.75 Å), suggesting that SMC1ACCsh is locked into this specific productive-like conformation.

Inspection of the SMC1ACCsh/RAD21C complex crystal packing provides an explanation to this lock. Specifically, in the SMC1ACCsh/RAD21C crystals, two symmetry-related complexes interact in a head-to-tail manner, and the artificial linker built to bridge the two coils of one complex inserts itself into the groove formed between the RecA- and helical-lobes of the second complex (Supplementary Figure 4A), reminiscent of the interaction made by the coupling α-helices of ABC transporters that link the structural rearrangements of their ATP-binding domains with the movement of their trans-membrane domains [106]. Specifically, we observe a large number of hydrophobic and hydrogen bond interactions made by the SMC1ACCsh linker with the RecA- and helical-lobe of the symmetric complex that could explain the stabilization of a productive-like conformation (Supplementary Figure 4B).

Human SMC3CC/RAD21 is locked into an ATPase inactive state

Crystallization of the human SMC3CC/RAD21N complex was also performed in the absence of any nucleotide or in the presence of ADP, ATP and ATPγS as well as ADP-BeF$_3$, ADP-AlF3 and ADP-VO$_4$. Crystals belonging to different space groups and diffracting between 2.25 and 3.1 Å were obtained in apo, ADP-bound, ADP-Mg-bound and ATPγS-Mg-bound states (Supplementary Table 3). Crystals were also obtained with ADP-BeF$_3$ but diffracted only to 4 Å and were not fully refined. Structure determination was carried out by molecular replacement using the yeast Smc3-HD/ATPγS structure as model (PDB code 4ux3). No real apo structure was obtained since crystallization in the absence of a nucleotide gave crystals where a sulphate stemming from the crystallization buffer was bound to the P-loop at the position normally occupied by the β phosphate of the nucleotide.

In all the SMC3 structures, the P-loop S39 side chain is positioned by the Walker B D1143 side chain as for interaction with a Mg ion, in contrast to SMC1A where this side chain conformation is only stably observed upon ATPγS-Mg binding. The ADP-Mg-bound structures also represent a potential post-hydrolysis state, the Mg ion remaining positioned as in the ATPγS-Mg-bound structure. Further, the R-loop adopts a different conformation in SMC3 than in SMC1A, with R57 being exposed at the surface of the SMC3 HD N-lobe, as previously observed [13,15,20,23].

Despite very different space groups and crystal packings as well as heterodimerization in the presence of ATPγS and ADP-BeF$_3$, all the structures obtained show a high structural similarity, whatever the nucleotide binding state, in contrast to what is observed for the SMC1A HD (Figure 5A). Binding of a nucleotide, either ADP or ATPγS, does not induce significant structural changes to the SMC3 HD. Importantly, in all our structures the Q-loop is positioned far away (~9 Å) and the Q141 side chain is turned away from the nucleotide binding site (Figure 5B).

Comparison of our structures with the structure of the SMC3 HD in the productive complex reveals however significant structural changes (Figure 5C). Notably, in contrast with our observations on the SMC1A HD, these changes are similar to those described upon bacterial Smc HD homodimerization [21] with concerted conformational changes in the R-loop and in the wedge region and in the positioning and orientation of the signature-coupling helices. As for SMC1A, however, a lever effect is observed that moves the SMC3 second signature-coupling helix by ~4-

5 Å upon formation of the productive complex (Figure 5C). These different conformational changes drastically affect the positioning of the Q-loop in the productive complex where Q141 interacts productively with the Mg ion (1.9 and 2.1 Å) and the γ group of the nucleotide (3.0 Å) [23].

Comparison of the yeast published Smc3 structures [13,15,105] show that similar structural changes of the R-loop and the wedge region and a lever effect are required to form the productive state (Supplementary Figure 5A). Surprisingly, while the human SMC3 GD structure is relatively close to that in the productive state, we observe that this is not the case for yeast Smc3 that also adopts a relaxed conformation compared to the productive state (Supplementary Figure 5B).

A stable resting state is observed for the human DNA exit gate in SMC3CC/RAD21N

In fact, a major, specific and unexpected structural feature of the human SMC3CC/RAD21N complex is the observed ~50° kink of the CC emerging from the SMC3 GD in all our structures compared to that of SMC3 in the productive complex (Figure 6A). This kink, not seen in any yeast Smc3 structure, is due to a conformational change in both helices forming the CC and is located above the region of the CC interacting with the SMC3 GD but below the region of the CC that interact with RAD21N. This kink is observed in all our structures, whatever their nucleotide-binding state.

The flexibility of these short helical stretches of the human SMC3 CC that enable the kinking is further demonstrated by their two different conformations observed in our different structures, including between non-crystallographic symmetry mates in a single crystal form, where they slightly affect the position of the helical-lobe with respect to the RecA-lobe, but also the conformation of the wedge region, further demonstrating the intrinsic flexibility of this latter region (Figure 6A). Although the structure of RAD21N and its interaction with the CC are not changed by the kink, the new path of the CC positions differently the RAD21N globular domain at the interface between the SMC3 GD N-lobe and C-lobe (Figure 6B). The resulting complex thus stabilizes the DNA exit gate on the SMC3 GD. We have therefore named resting state this specific conformation of the human SMC3 HD/RAD21N DNA exit gate.

In the productive state, RAD21N only forms one direct contact with the SMC3 GD through interaction of the RAD21N R54 side chain with the main chain of SMC3 N123 (Figure 6C). The RAD21N globular domain itself is positioned above but does not interact with the SMC3 GD and is turned toward the DNA with which it interacts through electrostatic interactions, and through direct contacts such as those that can be formed by RAD21N K25 and K26 with the DNA phosphate backbone [23]. In the resting state, RAD21N residues 5 to 26 pack against the SMC3 GD forming stacking and hydrogen bonding contacts through main chain/main chain, side chain/main chain and side chain/side chain interactions (Figure 6D).

Notably, the first small α-helix of the RAD21N GD, composed of residues 13 to 23, packs against the SMC3 N-lobe but is also linked to the SMC3 CC through its perpendicular packing interaction with the long α-helix of RAD21N that runs parallel to the SMC3 CC, thus maintaining the structural integrity of the DNA exit gate (Figure 6B). While residues 13 to 17 of this small helix form main chain/main chain interactions with SMC3 GD, residues 18 to 23 are involved in main chain/side chain and side chain/side chain interactions with this GD. Following this helix, RAD21N K25 and K26 side chains are positioned in spatial proximity with SMC3 GD N-lobe D120 and D92 side chains (Figure 6D). Finally, R54 guanidino group makes several interactions with the SMC3 GD, notably with the main chain of V162 and the side chains of S131, S133 and D1135 (Figure 6D).

Importantly, the aforementioned V162 residue is located at the C-terminal end of the SMC3 second signature-coupling helix that precedes the N-terminal coil. In the productive state, this helix is moved forward upon leverage, in order for the SMC3 GD to adopt a productive conformation by bringing the Q-loop in contact with the nucleotide. Strikingly, the position of the RAD21N GD in the resting state and its stabilization by multiple interactions with the SMC3 GD is incompatible with the forward movement of the lever effect required to reach the productive state (Figure 7A). Therefore the resting state locks SMC3 into an inactive conformation by preventing the structural rearrangements required to reach the productive state.

Importantly, the resting state is observed in all our structures, whatever their nucleotide binding state. This indicates, similarly to what has been observed in the case of the SMC1A relaxed state, that ADP and ATPγS binding are not sufficient for relieving the inactive state imposed by the resting conformation of the DNA exit gate. In order to assess the flexibility of this gate, we have

147

complemented our structural analysis by cross-linking experiments in solution in the presence of bismaleimidoethane (BMOE) to look at the dynamics of the SMC3 CC in the SMC3CC/RAD21N complex in the absence or in the presence of SMC1ACC/RAD21C and of ATP. Based on our structures and the structure of the productive complex, we have tentatively identified pairs of residues in SMC3 HD and RAD21N that could help distinguish between both states upon their mutation in cysteines and their subsequent crosslinking. The relatively large distance between the RAD21N globular domain and the SMC3 GD in the productive state and the possibility of having SMC3 CC adopting an intermediate extended state without leverage complicated this task. We could however identify at least one pair (SMC3CC-D120C/RAD21N-K25C) that is clearly specific for the resting state. A pair (SMC3CC-N119C/RAD21N-H22C) was identified that should be specific for an extended/productive state. However, we cannot exclude that crosslinking for this pair (i) could also occur upon a small movement of the DNA exit gate away from its resting conformation or (ii) could even force the gate out of its resting state. All endogenous cysteines in the SMC1A, SMC3 and RAD21 constructs were mutated to serines to prevent unwanted crosslinking events. Our results showed that crosslinking occurs whatever the crosslinking pair and the complex studied, demonstrating that the resting state exists in solution, both in the non-engaged and in the engaged states (Figure 7B). These suggested also that the human DNA exit gate can potentially oscillate between the resting state and an extended or productive state.

The resting state positions the basic electrostatic patch of the human DNA exit gate on the outside of the DNA binding chamber

In the productive state, the DNA is kept tightly bound into a chamber formed by the GDs and CCs of the SMC1A and SMC3 engaged HDs as well as NIPBL[Scc2] [13,20,23]. Each protein component of the chamber displays a positively charged electrostatic potential that matches the negative electrostatic potential of the DNA (Figure 7C). The positive electrostatic potential of the chamber is particularly strong in the human complex, notably for the engaged SMC1A and SMC3 GDs onto which the DNA is lying in a straight conformation. The DNA is then bent at the DNA exit gate and then continues in a straight conformation toward SA1[Scc3] [23]. DNA bending is most likely induced by the positioning of NIPBL[Scc2]. However, the DNA exit gate and notably its RAD21N globular

domain also displays a highly positively charged patch which matches and contributes to the DNA curvature (Figure 7C). Such a large positively charged patch is not observed on SMC1A CC whose electrostatic contribution to DNA binding appear less important.

Since we have shown that the resting state is compatible with engagement, we have replaced the SMC3 HD in the productive complex by our resting SMC3 HD in the ATPγS-bound form (Figure 7D). Analysis of our model shows that the resting state is compatible with engagement provided slight structural rearrangements at the SMC1A/SMC3 interface. In fact, most of the significant structural differences observed between the two SMC3 conformations are located away from this interface. Notably, the kinked CC, the RAD21N GD in its resting position, and the Q-loop do not participate to this interface and, in fact, do not clash with the bound DNA observed in the productive complex, only the NIPBL[Scc2] head position being incompatible with the kinked SMC3 CC (Figure 7D). Similarly, a lever effect would not be necessary for engagement, in agreement with what has been observed in the yeast engaged complex (Supplementary Figure 5A) [105].

Importantly, one major feature of this model is that the basic patch formed by the RAD21N globular domain is not present anymore inside the engaged HDs chamber, but is now positioned on the outside of the engaged SMC3 HD (Figure 7E). While the residues composing the basic patch in the productive state belong mostly to RAD21N, those forming the basic patch in the resting state belong both to the RAD21N globular domain and to the SMC3 GD. Specifically, most of the positively charged residues of RAD21N participating to the basic patch in the productive state (K10, R11, K16, K25, K26, K29, H31, K48 and K50) also participate to the basic patch in the resting state. Collectively, our observations thus demonstrate of a conformational dynamics of the DNA exit gate of the human ATPase module that can significantly alter the conformational and physico-chemical properties of its SMC3-HD/RAD21N component.

Discussion

Deciphering the structure/function relationships of SMC complexes is essential to understand their biological functions. This however requires the delineation between shared mechanisms

and evolution-acquired specific mechanisms. Specifically, SMC complexes strongly rely on the binding and the hydrolysis of ATP for function. Here we have investigated the ADP and ATP binding, the engagement and the conformational dynamics of the human Cohesin SMC1A and SMC3 ATPase heads (HDs). Our results provide an unprecedented view on the evolution of these two HDs.

Notably, we show that each of these HDs has only retained a specific subset of the properties from the bacterial Smc ancestor. Specifically, the SMC1A HD has kept a bacterial-like conformation of the R-loop, notably with R57 reaching into the nucleotide binding site and contributing to ATP binding, as reported previously [13,18,20,21,23,25]. In contrast, although the conformational changes required for the SMC1A HD to adopt a productive conformation include a lever effect, these do not require a significant conformational change of the R-loop and of the adjacent wedge region. Conversely, the conformational changes required for the SMC3 HD to adopt a productive conformation require, beside a lever effect, bacterial-like structural changes of the R-loop and of its adjacent wedge region. In contrast, the R-loop itself is not involved in nucleotide binding site formation and R57 and is exposed at the surface of the N-lobe, as already observed [13,15,20,21,23,25].

Our structural analyses also reveal the conformational dynamics of the SMC1A HD between its apo, ATP-bound and ADP-bound forms, which is different from the lever effect required to reach the productive conformation. The movements observed between our structures represent conformational changes expected to occur in a non-engaged form, either before engagement or after ATP hydrolysis. Notably, the large movement of the helical-lobe between the ATPγS-bound and ADP-bound forms could displace significantly a HAWK regulatory protein bound to the SMC1A helical-lobe F-loop, thus coupling ATP binding and hydrolysis to Cohesin activity regulation. Comparison of our structures with the available previously published structures suggest that the yeast HDs share the same properties as those aforementioned for the human HDs, even though the absence of some yeast structures, notably in the presence of ADP, limit full comparative analysis.

However, our study also reveals unexpected significant differences between the yeast and human HDs. Specifically, human SMC1A HD P-loop adopts a closed conformation that is not

observed in yeast Smc1. A closed conformation for this loop has been observed in yeast Condensin Smc2, where it prevents ATP binding and supports an asymmetric ATPase cycle in Condensin [19]. Our different experiments reveal that the P-loop closed conformation does not prevent ATP binding but sets a thermodynamic barrier that requires desolvation of the nucleotide binding pocket upon ATP binding. The maintenance of a closed conformation seems to have induced a specific evolution of the human SMC1A HD, where the R-loop and notably R57 adopt a stable conformation, whatever the nucleotide binding state, and always contribute to the formation of the nucleotide binding site. Whether the SMC3 P-loop can also adopt a closed conformation remains unknown but is less likely considering the rather low Kd of ADP for binding to the SMC3 HD compared to the SMC1A HD.

The Cohesin ATPase cycle has been suggested to be asymmetric [72,102-104] but the molecular basis for this asymmetry remain poorly understood. We have shown that part of the mechanisms supporting the asymmetry observed for the Condensin complex [19] is not conserved in human Cohesin. Nevertheless, the various biochemical and structural reports on the Cohesin ATPase module, including this work, reveal very different biochemical and structural properties for the SMC1A and SMC3 HDs. These differences could certainly support the asymmetry of the Cohesin ATPase cycle. Interestingly, we show that the human HDs have diverging properties (e.g. closed P-loop, resting state) compared to their yeast orthologues that could contribute to a specific human asymmetric ATPase cycle.

To reach the productive state, two potentially independent lever effects on SMC1A and SMC3 are required. Specifically, that for SMC3 implies a switch of the human SMC3 CC from a resting state to an extended state and finally to a productive state. Furthermore, the lower Kd of ADP compared to ATP for the SMC3 HD suggests that the existence of a stable ADP-bound state could be observed for SMC3 but not for SMC1A, as proposed previously [104]. Collectively, these different properties, which appear independent from each other, could all contribute to an asymmetric ATPase cycle of Cohesin. It is however very likely that asymmetry is caused by a balance between all the positive and negative effects on ATP binding and hydrolysis induced by the different properties of the SMC1A and SMC3 HDs, and that these are further modulated by the various

Cohesin regulatory proteins. Nevertheless, our results provide the molecular basis to further investigate this asymmetry of the Cohesin ATPase cycle.

Formation of the Cohesin productive complex requires major structural changes, including leverage of both SMC1A and SMC3 CCs. In bacteria, HDs engagement appear sufficient to induce a lever effect in the Smc proteins [21]. Comparison of the structures of the various engaged HDs complexes [105] (and this work) and of the productive complexes [13,20,23] show that this is not the case for the Cohesin HDs. We therefore conclude that the action of NIPBL[Scc2] and DNA is the main driving force that brings these HDs to their stable productive conformations. This is in agreement with the activation of the Cohesin ATPase activity upon binding of these macromolecules to the engaged Cohesin HDs [47,49,69].

The required lever effects are not due to an interaction of NIPBL[Scc2] and/or the DNA with the N-lobe/C-lobe interface of SMC1A and SMC3, as observed in our SMC1ACCsh/RAD21C structures, although our work indicates that this interface could serve as a regulatory target. Importantly, if the structural work performed on the yeast HDs suggests that the formation of the yeast productive complex implies significant conformational changes of both HDs, our biochemical and structural analyses show that larger changes are required in the human case. Notably, we show that engagement does not lock the SMC3 CC in a pre-productive extended conformation. Our data and comparative analyses suggest an induced mechanism to reach the human productive state.

Specifically, the SMC3 resting state, by enlarging the DNA binding chamber, can ease the movements of the DNA into this chamber, prior to its binding on the electrostatic pit formed by the engaged HDs GDs. DNA can then attract electrostatically the RAD21N GD into this chamber, which would in turn stabilize DNA binding by restricting the size of the chamber and favour an extended conformation of the SMC3 CC. This would allow the correct productive positioning of NIPBL[Scc2] upon clamping the DNA, enabling it to interact with the SMC1A and SMC3 CCs and to lever them in order to stabilize them into their final productive conformation. Of note, in this mechanism NIPBL[Scc2] could already be bound to the SMC1A GD prior to DNA stabilization or even to HDs engagement, as shown recently [34], ATP binding, HDs engagement and DNA stabilization providing the required events to bring NIPBL[Scc2] in its productive position.

The presence of a resting state in human SMC3 and the observed flexibility of the human DNA exit gate raise the more general question on their role in the different stages of the human Cohesin cycle. Notably, the positioning of the RAD21N basic patch on the exterior of this HD in the resting state could promote specific interactions with proteins and/or DNA. In addition, the switch between a resting and an extended or productive conformation of the SMC3 CC could be used for various transactions between human Cohesin and its partners, as for instance for its stabilization on and unloading from the genome, which require the opening/closing of the Cohesin DNA exit gate.

In yeast, HDs engagement has been shown to be sufficient to open this gate [105]. Nevertheless, in human the additional interactions made between the SMC3 GD and the RAD21N globular domain in the resting state stabilize the closed form of this gate, and our various biochemical, biophysical, and structural characterizations do not show any opening of the human DNA exit gate upon ATP binding alone, neither in a non-engaged, nor in an engaged state. Notably, purification of the engaged complex in diluting conditions by size exclusion chromatography in an ATP-containing buffer does not cause the loss of RAD21N, and crosslinking between SMC3-HD and RAD21N occurs upon engagement in the presence of the SMC1A-HD/RAD21C complex and ATP. Thus, in the human case, additional mechanisms are required to open this gate.

In vertebrates a specific interplay between PDS5, SORORIN and WAPL participates to the switch between Cohesin stabilization on the genome by PDS5 and SORORIN, and Cohesin removal from the genome by PDS5 and WAPL. Interestingly, the fragment that links the C-terminal extremity of RAD21N to its PDS5 and NIPBL^{Scc2} binding regions is smaller in PDS5 (~10 residues expected in human) than in NIPBL^{Scc2} (~60 residues) (Supplementary Figure 1C) [85,107,108]. Forcing, by the PDS5 partners SORORIN and WAPL, the DNA exit gate to switch between a resting conformation and an extended or productive conformation could either stabilize or destabilize this gate. Notably, a pulling effect on the gate could be applied by WAPL upon increasing the distance between the RAD21N C-terminal extremity and the RAD21^{Scc1} PDS5-binding region. This could lead to opening of the gate, notably considering its sensibility to small perturbations [42]. Specifically, the required HDs engagement for WAPL-induced unloading of Cohesin from the genome could favour the flexibility of the gate, while the suggested sequestration of the RAD21N GD by PDS5 [85] could

facilitate the exit of the DNA from the Cohesin lumen. Interestingly, a similar mechanism relying on a switch from a resting state to an extended state has been observed for the RAD50/MRE11 complex, which shares many features with SMC complexes, forcing the MRE11 enzymatic subunit to move into a productive position [109].

Our observations raise the question whether similar mechanisms could also occur in yeasts. Interestingly, in the structure of the *S. pombe* productive state, the Rad21N globular domain resembles that of human, and this domain is also positioned to contact the DNA as in the human productive complex (Supplementary Figure 6A-B). This also suggests that a resting state exists for the *S. pombe* Psm3-HD/Rad21N interface, although the observed folding of the human SMC3 GD and of the RAD21N globular domain and their interface do not appear fully conserved in *S. pombe*. In contrast, the *S. cerevisiae* Scc1N globular domain does not have equivalent residues for K25 and K26 (Supplementary Figure 1C) and appears less structured [13,15].

These observations suggest that the DNA exit gate has evolved specific characteristics in eukaryotic Cohesin that could potentially support specific functions. It remains to be investigated whether such differences could explain the divergent reports on the Cohesin DNA entry and exit gates [13,43,44,51-54,110,111]. Our results and analyses however pinpoint, as suggested in previous reports [20,85], a specific role for the RAD21N globular domain of the Cohesin DNA exit gate, our work providing the molecular basis for its future investigation.

Collectively, our results provide novel knowledge on the vertebrate Cohesin ATPase module and its activity and on the potential mechanisms that can be used by vertebrate cells to regulate Cohesin functions. Our work also reveals unexpected properties that differentiate the yeast and human Cohesin core ATPase modules. This indicates that the increased mechanistic intricacy and the divergence/gain of functions of Cohesin along evolution has not only required the duplication and the apparition of novel regulatory subunits, but has also caused significant changes within the Cohesin core complex.

Material and Methods

Cloning

To generate the SMC1A and SMC3 ATPase head constructs with short coiled-coils, the sequences from the full length human *smc1a* and *smc3* genes coding for the hinges and those of the coiled-coils ranging from the hinge to the joint elements were replaced by sequences coding for a short protein linker (ESSKHPASLVPRGS or GSGSLVPRGSGS), as previously reported [15,18]. Point mutations were introduced into the constructions using rolling circle or nested PCR strategies. The constructs were then individually cloned by Gibson assembly into bacterial co-expression vectors [112-115] to yield proteins either native (SMC1A-HD and SMC3-HD constructs) or containing a C-terminal 10-histidine fusion tag preceded by a 3C protease cleavage site (RAD21N and RAD21C constructs).

Protein production and purification

The same large-scale overproduction and purification methods have been used for all protein constructs. SMC1A or SMC3 ATPase heads-coding plasmids were respectively co-transformed with RAD21C- or RAD21N-coding plasmids into chimiocompetent *Escherichia coli* BL21(DE3) cells. Co-transformed cells were selected using the appropriate antibiotics. Colonies were used to inoculate large cultures of 2x LB medium that were grown for 6 hours at 37 °C. Protein expression was induced at 25°C by the addition of a final concentration of 0.7 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG; Euromedex), and cells were further grown overnight at 25°C. Culture media was discarded after centrifugation and the bacterial pellets were resuspended in 30 ml lysis buffer containing 200 mM or 500 mM NaCl and 10 mM Tris-HCl pH 8. Pellets were stored at -20°C until further use.

After sonication of the cell pellets, lysate clarification was performed by centrifugation (1h at 45000 rpm). The recombinant SMC1A-HD/RAD21C and SMC3-HD/RAD21N protein complexes were then purified by affinity chromatography using the 10xHis purification tag on RAD21 by incubating the cleared lysates with TALON Metal Affinity Resin (Takara Bio). The purification tag was then removed on the affinity beads by overnight 3C protease digestion at 4°C. In order to

remove any remaining nucleic acid contaminants, the recovered protein complexes were further loaded onto a 1 ml or 5 ml HiTrap Heparin column (GE Healthcare) and eluted using a NaCl gradient from 50 mM to 1 M. Peak fractions containing the protein complexes were pooled and further purified by size exclusion chromatography using a 16/60 Superdex 200 column (GE Healthcare) equilibrated with a buffer containing 200 mM NaCl, 10 mM Tris-HCl pH 8, 2 mM $MgCl_2$ and either 1 mM TCEP (samples for crystallization and ITC assays) or 2 mM DTT (samples for ATPase assays). Sample purity was assessed by polyacrylamide gel electrophoresis in presence of sodium dodecyl sulfate at all purification stages. The main peak fractions with the protein complexes were pooled, concentrated with AMICON Ultra concentrator filters (Merck Milipore), and either used immediately or frozen in liquid nitrogen and stored at -80°C for later use.

ATPase assays

Purified wild type and mutant SMC1A-HD/RAD21C and SMC3-HD/RAD21N ATPase activities were assayed using the EnzChek Phosphate Assay Kit (Thermo Fischer Scientific). For these assays, a final concentration of 20 µM of a single protein complex or 10 µM of each protein complex upon mixing were incubated with 1 mM ATP. After addition of ATP, the ATP hydrolysis activity was immediately measured at 30°C, by measuring the absorbance at 360 nm every 42 seconds for 2 hours, using a spectrophotometer plate reader (BioTek). The experiments were performed either in duplicates or in triplicates.

Isothermal titration calorimetry measurements

For ITC measurements, ATP, ATPγS or ADP were dissolved at a final concentration of 4 mM into the protein gel filtration buffer (200 mM NaCl, 10 mM Tris-HCl pH 8, 2 mM $MgCl_2$ and 1 mM TCEP), then titrated at 5°C by sequential injections (usually 2 µL each) into wild type or the EQ-mutant SMC1ACC/RAD21C or SMC3CC/RAD21N complexes (between 200 and 300 µM). ITC measurements were performed using a PEAQ-ITC microcalorimeter (Microcal Malvern Panalytical). ITC data were then corrected for the dilution heats generated by the injection of the

buffer into the protein sample and of the nucleotide sample into the buffer. ITC data were processed using Microcal PEAQ-ITC Analysis and AFFINImeter softwares[116].

Cross-linking experiments

For crosslinking experiments, the SMC3CC/RAD21N complex, where all endogenous cysteines have been replaced by serines and where the cysteines used for crosslinking (see text) have been introduced by mutagenesis, was expressed in small scale using the same conditions as for the large-scale expression, and purified on TALON affinity resin (Takara Bio) by affinity chromatography. The protein-bound resin was washed with a buffer containing 200 mM NaCl, 10 mM Tris-HCl pH 8, 1 mM TCEP and 10 mM to 20 mM imidazole pH 8. The protein complex was then eluted from the resin with a similar buffer containing only 0.5 mM TCEP and 250 mM imidazole. For the crosslinking assays, a final concentration of 6 µM of the SMC3CC/RAD21N complex was incubated with 0.5 mM BMOE (bis-maleimidoethane), diluted in DMSO, into a reaction buffer containing 200 mM NaCl, 10 mM Tris-HCl pH 7.5 and 0.5 mM TCEP. Controls were set up in the same conditions, using DMSO without BMOE. All BMOE crosslinking experiments were performed at room temperature in order to avoid significant pH changes that could affect the crosslinking reaction. After 10 minutes, the crosslinking reactions were stopped by adding the same volume of 2 x denaturing Lämmli buffer containing β-mercaptoethanol, then heated 5 minutes at 70°C before analysis by polyacrylamide gel electrophoresis in presence of sodium dodecyl sulfate. Proteins were detected by staining the gels using Coomassie brilliant blue. The same experimental conditions were used for crosslinking experiments of the SMC3CC/RAD21N complex in the presence of the SMC1ACC/RAD21C complex, supplemented or not with 4 mM ATP, where all endogenous cysteines of this latter complex had also been mutated in serines.

Crystallizations

Initial crystallization assays were carried out with commercial crystallization screens in swissci 96-Well 3-Drop MRC crystallization plates (Molecular Dimensions). Crystals were grown using the sitting-drop vapor-diffusion method at 4, 20 and 27°C. Briefly, 200 nl of 4 to 15 mg/ml protein sample were mixed with 200 nl of reservoir solution. Several conditions from the commercial

crystallization screens PACT, JCSG+, Classics, WIZARD I and II, BCS and LFS yielded protein crystals within a few hours up to several weeks. Some of the crystals, especially those from SMC3-HD/RAD21N, required extensive optimization to improve their stability and diffraction limit. Selected crystals were cryo-protected with 20% (v/v) glycerol or 20% (v/v) PEG 200, then flash-cooled and stored into liquid nitrogen until data collection. Crystallization conditions are provided in the table below:

Dataset	Crystallisation condition
SMR1ACCsh	
SMR1ACCsh-EQ-apo	0.2 M ammonium nitrate pH 6.3; 20% w/v PEG 3350
SMR1ACCsh-EQ-loop-SO$_4$	0.2 M sodium sulfate; 20 % w/v PEG 3350
SMR1ACCsh-EQ-ADP	0.2 M sodium malonate; 0.1 M Bis Tris propane pH 6.5; 20 % w/v PEG 3350
SMR1ACCsh-EQ-ADP-Mg	0.2 M magnesium chloride; 0.1 M MES pH 6.0; 20 % w/v PEG 6000
SMR1ACCsh-EQ-AGS-Mg	0.2 M sodium formate; 0.1 M Bis Tris propane pH 6.5; 20 % w/v PEG 3350
SMR1ACC	
SMR1ACC-WT-apo	0.1 M Hepes pH 7.0; 10% w/v PEG 6000
SMR1ACC-EQ-apo	0.2 M sodium malonate dibasic monohydrate; 0.1 M Bis-Tris propane pH 6.5; 20% w/v PEG 3350
SMR1ACC-WT-loop	0.1 M Hepes pH 7.0; 10% w/v PEG 6000
SMR1ACC-EQ-loop	0.2 M sodium bromide; 0.1 M Bis-Tris propane pH 6.5; 20% w/v PEG 3350
SMR1ACC-EQ-ADP	0.2 M ammonium formate; 20% w/v PEG 3350
SMR1ACC-EQ-AGS-Mg	0.1 M MMT pH 7; 25% w/v PEG 1500
SMR3CC	

SMR3CC-EQ-ADP	0.1 M sodium/potassium phosphate pH 7.5; 0.1 M HEPES pH 7.5; 15% v/v PEG Smear High; 10% v/v ethylene glycol
SMR3CC-WT-ADP-Mg	0.075 M magnesium chloride; 0.075 M Sodium citrate tribasic dehydrate; 0.1 M Bis-Tris pH 6.0; 18% v/v PEG Smear Broad
SMR3CC-EQ-AGS1-Mg	0.3 M NaCl; 0.05 M L-Arginine; 0.1 M Tris pH7.5; 22.5 % v/v PEG Smear Broad; 0.05 M L-Glutamic acid monosodium salt hydrate
SMR3CC-EQ-AGS2-Mg	0.1 M sodium chloride; 0.1 M Bicine pH 9.0; 20% w/v PEG 550 MME
SMR3CC_2-EQ-SO$_4$	0.21 M sodium sulfate; 0.1 M Bis-tris propane pH7; 16 % PEG 3350

Crystallographic structure determination

X-ray diffraction data were collected at the synchrotron Source Optimisée de Lumière d'Énergie Intermédiaire du LURE (SOLEIL, France) and the Swiss Light Source Synchrotron (SLS, Switzerland). High to medium resolution diffraction data (ranging between 1.36 and 3.0 Å) were processed by indexation, integration, and scaling within the XDS program [117]. Data were merged using Aimless from the CCP4 software[118]. The various SMC1A-HD/RAD21C and SMC3-HD/RAD21N structures were solved by molecular replacement using PhaserMR [119] using respectively the yeast Smc1-HD/Scc1C (PDB: 1w1w) and Smc3-HD/Scc1N (PDB: 4ux3) structures as models. The initial models were subsequently iteratively built and refined using the Coot and Phenix refinement programs [120,121]. All refined models were verified with Molprobity [122] and showed good refinement statistics (Supplementary Tables 1-3).

Acknowledgments

Author contributions

MG, PL, MLDD, TBS and ET made the clones and purified the proteins. MG, PL, MLDD and TBS performed the biochemical and biophysical assays. KB and EE collected and analysed the isothermal titration calorimetry data. MG and TBS did the crystallographic work. All authors planned the experiments and analysed the data. CR wrote the manuscript, with input from MG.
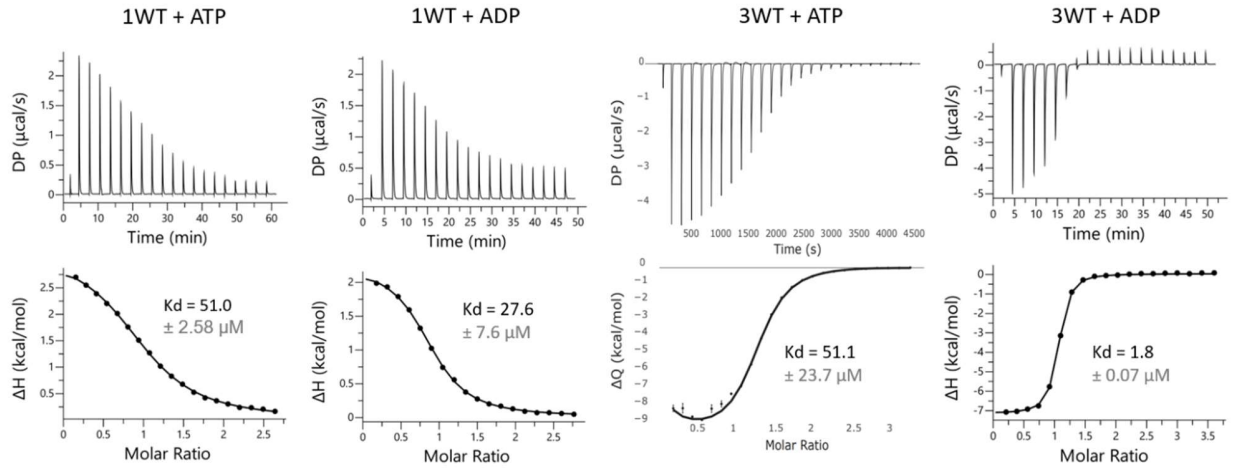
Competing financial interests

The authors declare no competing financial interests.

Figures

Figure 1.
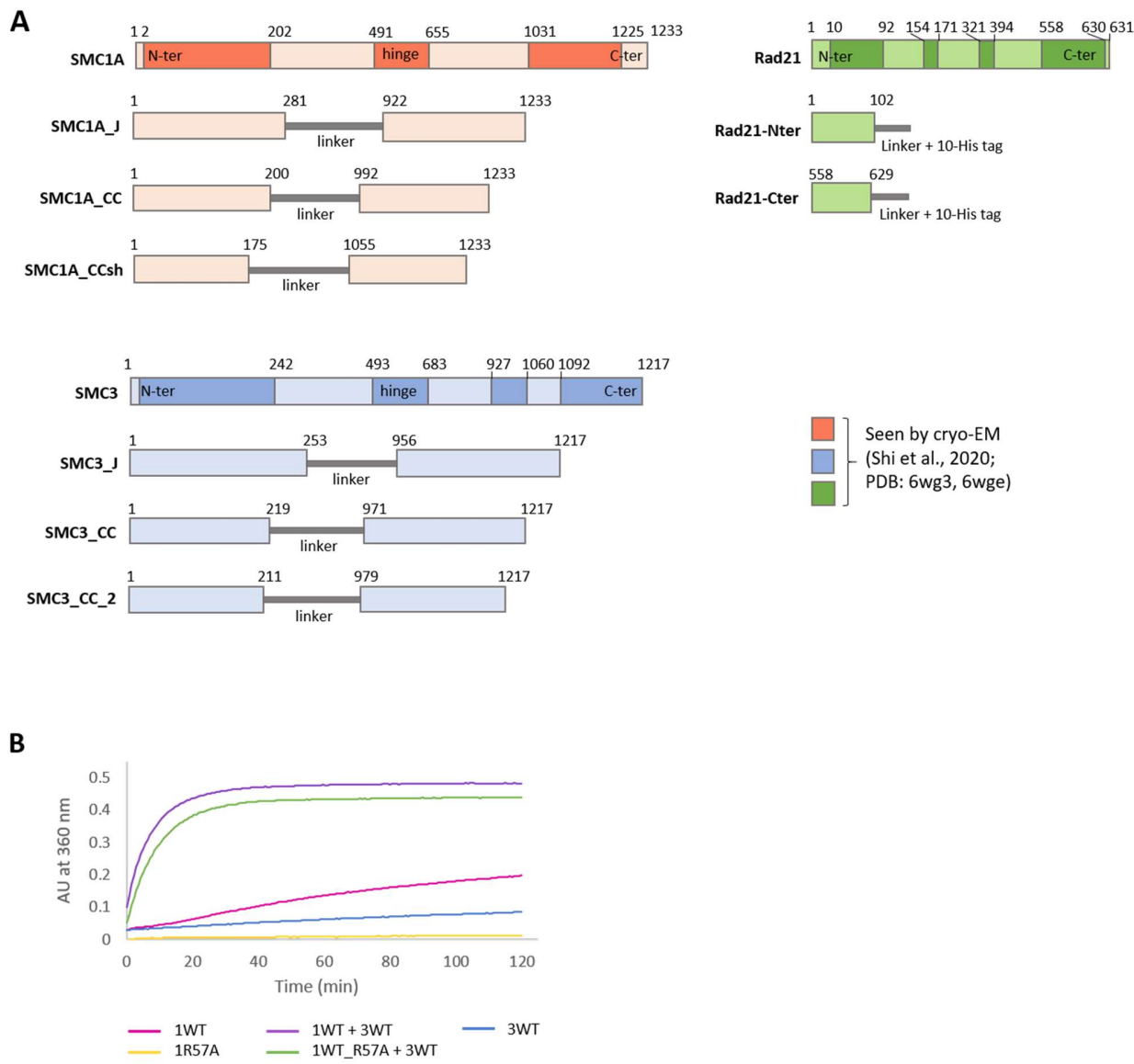
C

D

| Cell	Syringe	Fitting model: SMC + ATP → SMC-ATP			Fitting model: SMC + ATP → SMC-ATP + SMC-ATP → SMC-ATP homodimer			
		KD1 (µM)	ΔH1 (kcal/mol)	N1	KD1 (µM)	ΔH1 (kcal/mol)	KD2 (nM)	ΔH2 (kcal/mol)
1WT	ATP	51.0 ± 2.58	3.17 ± 0.048	1.00 ± 0.01	-	-	-	-
1WT	ATPγS	31.6 ± 1.00	1.89 ± 0.014	1.21 ± 0.01	-	-	-	-
1WT	ADP	27.6 ± 7.6	2.78 ± 0.19	0.85 ± 0.03	-	-	-	-
1EQ	ATP	55.3 ± 4.72	1.42 ± 0.038	1.05 ± 0.01	-	-	-	-
1EQ	ATPγS	43.0 ± 3.06	1.06 ± 0.021	1.01 ± 0.01	-	-	-	-
3WT	ATP	-	-	-	51.1 ± 23.7	- 2.7 ± 1.2	1.4 ± 0.7	- 18 ± 0.8
3WT	ATPγS	-	-	-	18.3 ± 1.5	- 6.0 ± 1.9	5.3 ± 0.6	- 7.4 ± 1.8
3WT	ADP	1.8 ± 0.07	- 7.19 ± 0.02	0.98 ± 0.02	-	-	-	-
3EQ	ATP	-	-	-	74.6 ± 33.8	- 5.2 ± 1.0	2.8 ± 1.8	- 14.9 ± 0.6
3EQ	ATPγS	-	-	-	35.5 ± 9.2	- 5.6 ± 0.8	3.8 ± 1.0	- 8.0 ± 0.7

E
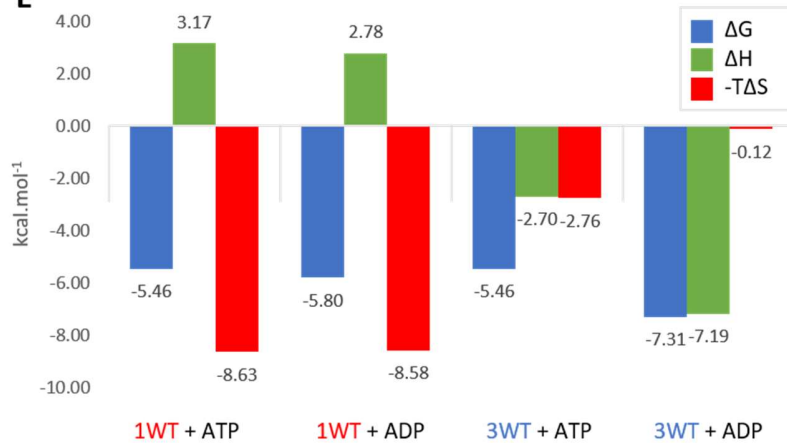
Figure 1. Distinct ADP and ATP binding properties of SMC1A and SMC3 ATPase heads.

(A) Protein constructs used for SMC1A and SMC3 ATPase heads and for RAD21 N-terminal and C-terminal domain. SMC1A and SMC3 constructs were made with shortened coiled coils, ending either above the joint domain (J) or below the joint (CC, CCsh). The truncated coiled coils are linked by a short linker, ESSKHPASLVPRGS or GSGSLVPRGSGS. SMC1A_CC and SMC3_CC constructs were used for the biochemical and biophysical analysis in the presented results.

(B) ATPase activity of SMC1ACC and SMC3CC. 1WT: SMC1ACC wild type; 3WT: SMC3CC wild type; 1R57A: SMC1ACC bearing the R57A mutation.

(C) ITC profiles of SMC1ACC wild type and SMC3CC wild type titrated with ATP or with ADP. 1WT: SMC1ACC wild type; 3WT: SMC3CC wild type.

(D) Table of all ITC results of SMC1ACC and SMC3CC, wild type (1WT and 3WT) or EQ mutants (1EQ and 3EQ). A simple one set of sites fitting model was used to fit SMC1ACC data, whereas for SMC3CC a model including dimerization upon ATP and ATPγS binding was used.

(E) Themodynamic parameters of the ATP or ADP binding event to SMC1ACC or SMC3CC.
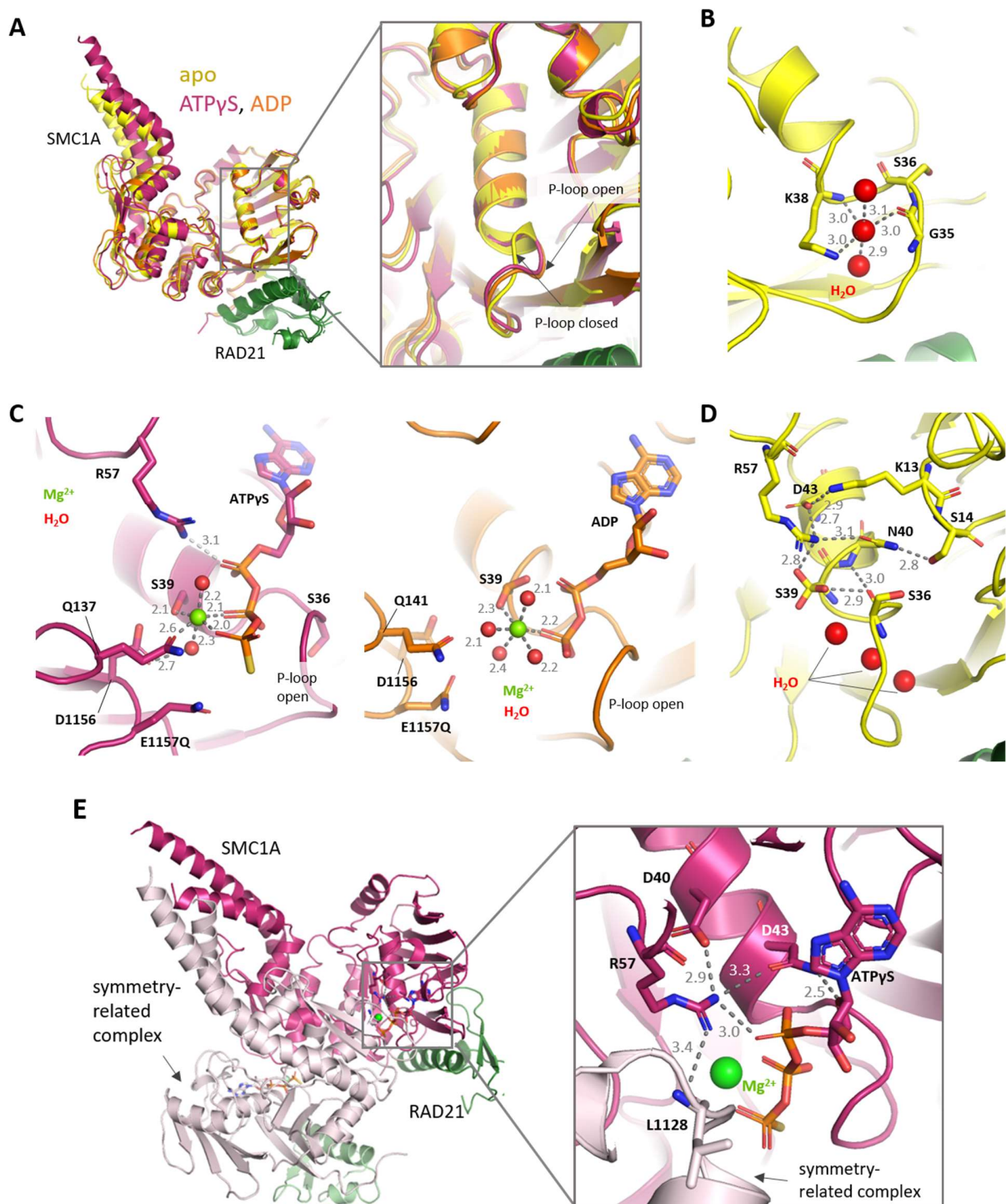
Figure 2.

Figure 2. Organization of the SMC1A ATP binding site by the P-loop and R-loop.

(A) Superposition of the crystallographic SMC1ACC structures, in the apo, ADP- and ATPγS-bound states (left). The superposition was made on RAD21C. Close up view of the P-loop reveals two possible conformations for the SMC1A P-loop: open, where it grants nucleotide access to the binding site, and closed, in which nucleotide binding is precluded.
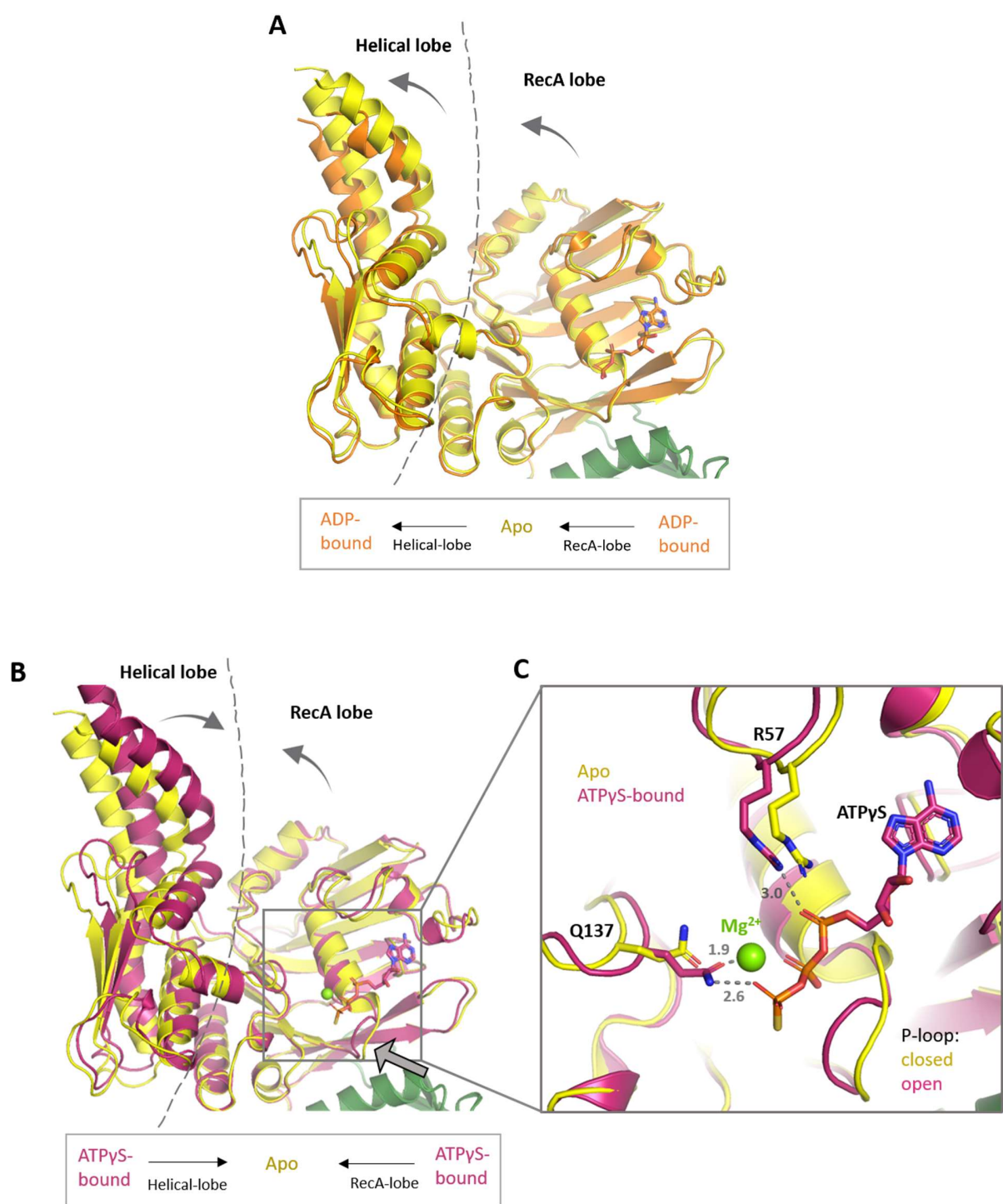
(B) Representation of the network of structural waters that interact with side chains of residues of the P-loop, and maintains the P-loop in a closed conformation in the SMC1ACC apo structure.

(C) Binding of ATPγS-Mg (left panel) or ADP-Mg (right anel) into SMC1ACC active site. Distances between the Mg^{2+} ion and the water molecules that coordinates the Mg^{2+} in both cases are shown in dashed grey lines. Distances are given in Angstrom.

(D) Side chain contribution of the catalytic site residues into a water network that keeps the P-loop in a closed position.

(E) Side chain network in which the R57 participates to stabilize ATPγS into the active site. R57 also participates in heads dimerization, as seen with a symmetry related molecule inside the crystal packing. Mg is shown as a green sphere. Water molecules are indicated as red spheres. Distances between atoms are shown as dashed gray lines. Distances are provided in Angstrom.

Figure 3.

D

RecA-lobe

Helical-lobe

R57

ATPγS

3.5 3.8

Q137

Mg²⁺

ADP

| ADP-bound | ← Helical-lobe | ATPγS-bound | RecA-lobe → | ADP-bound |

E

Helical-lobe movement

ADP ≈10° rotation ← Apo → ≈10° rotation ATPγS

F

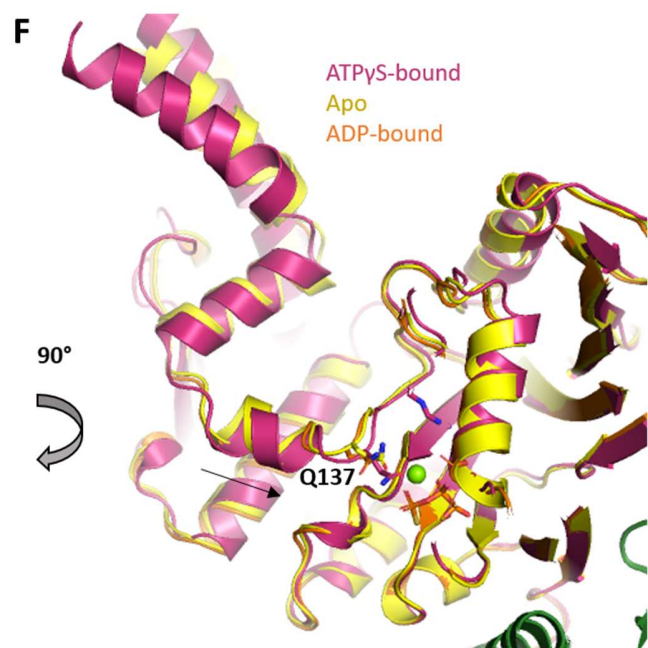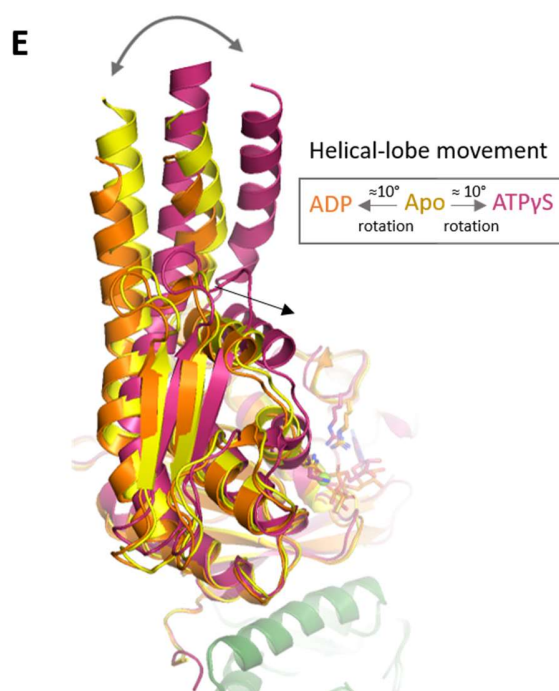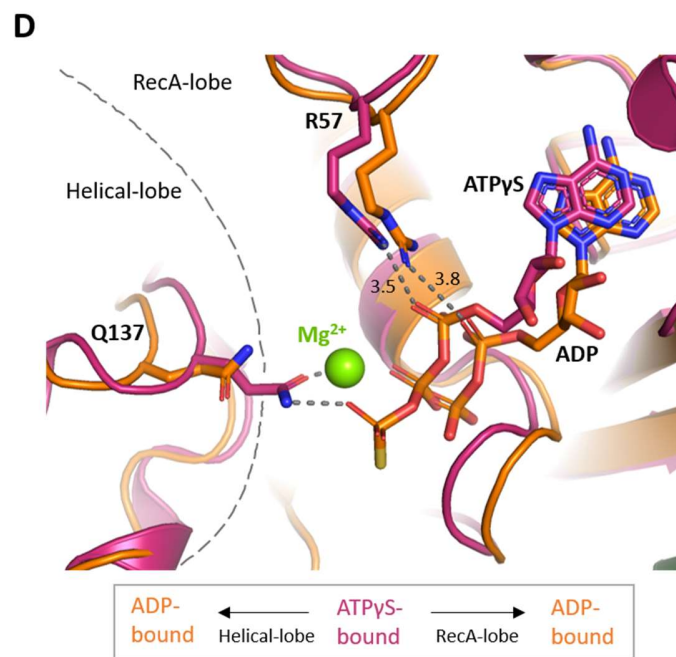ATPγS-bound
Apo
ADP-bound

90°

Q137

167

Figure3. Specific rotational movements of SMC1A helical- and RecA-lobes.

(A - B) Comparison of structures de SMC1ACC, aligned on RAD21-C, in the apo and ADP- and ATPγ-bound forms showing the rotational movements that occurs between both helical and RecA lobes.

(C) Constriction of the active site upon ATPγS binding. The P-loop opens to accommodate the nucleotide, and the conserved glutamine 137 of the Q-loop reaches out to coordinate the Mg ion bound to ATPγS.

(D) Relaxed state of the active site in the ADP-bound conformation, as compared with the ATPγS conformation.

(E) From the apo state, the nucleotide-induced conformational changes drive the rotation of the helical lobe from approximately 10°: inwards upon ATPγS binding and outwards when the active site is occupyied by ADP.

(F) View of the rotational coil movements in (E) at 90°. The rotational movement is planar, relative to the signature coupling helix of SMC1A.
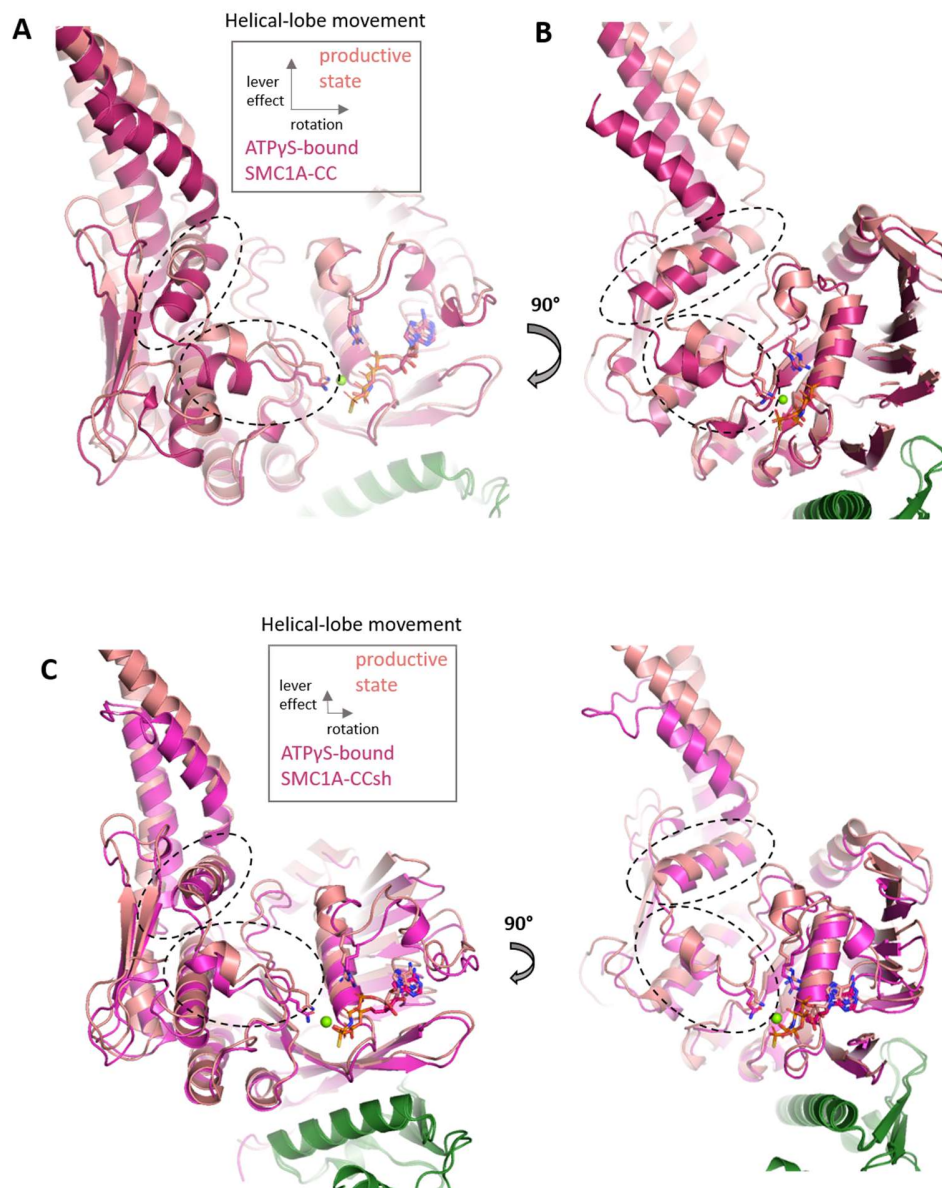
Figure 4.

Figure 4. SMC1A adopts a relaxed conformation different than from that in the productive mode. (A) Comparison of the ATPγS-bound SMC1ACC structure and the cryo-EM structure of SMC1A engaged into the productive complex (PDB: 6wg3). Structure alignment with RAD21 as a reference show that a lever effect occurs in the helical region which contacts NIPBL in the productive case.

(B) The lever movement significantly displaces upwards the SMC1A signature coupling helix that is adjacent to the Q-loop. This movement is distinct from the rotational movements that occur upon nucleotide binding to SMC1ACC.

(C) Comparison of the ATPγS-bound SMC1ACCsh structure and the cryo-EM structure of SMC1A engaged into the productive complex (PDB: 6wg3). A lever effect is observed in SMC1ACCsh copared to SMC1A in the productive conformation, albeit with a less significant amplitude than in SMC1ACC shown in (B).
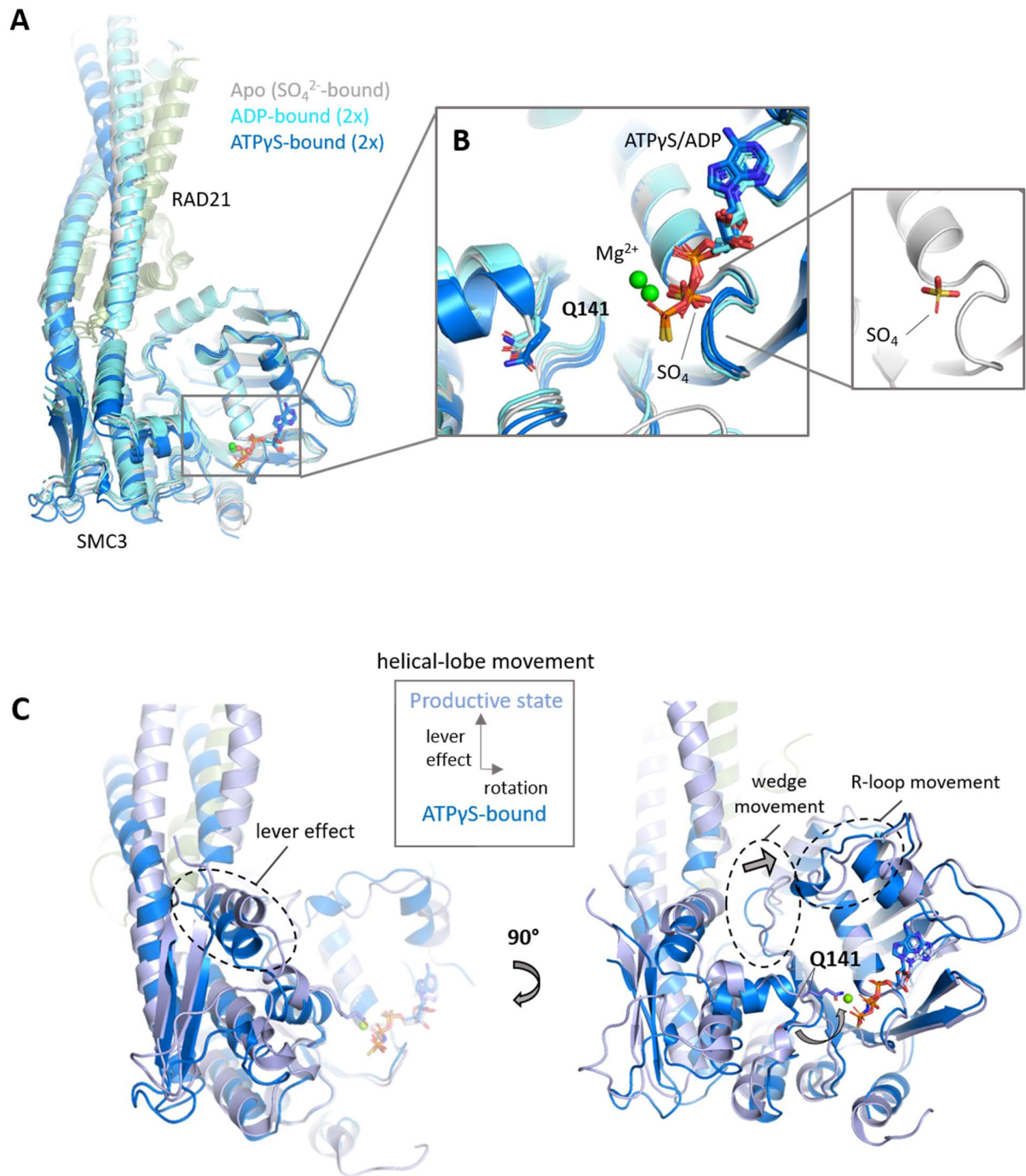
Figure 5.

Figure 5. SMC3 is precast for heterodimerization but is inactive.

(A) Structure comparison of all the five obtained crystallographic structures of SMC3CC, bound to ADP, to ATPγS or to a sulfate ion, keeping the helix containing the walker A in the RecA domain as a fixed reference.

(B) Comparison of the active sites of the SMC3CC crystallographic structures. They adopt similar conformation, with the conserved glutamine 141 from the Q-loop that looks away from the active site, thus adopting an inactive state.

(C) Comparison of the crystallographic SMC3CC structures with the cryo-EM SMC3 structure (PDB: 6wg3) showing the lever effect that occurs at the SMC3 helical region through the upward movement of the signature coupling helix, to achieve the productive conformation (left panel). Significant structural movements also occur at the wedge and P-loop regions, and, importantly, in the productive mode the Q-loop adopts a productive conformation, with the Q141 reaching out to coordinate the nucleotide ant the Mg ion bound at the active site (right panel).
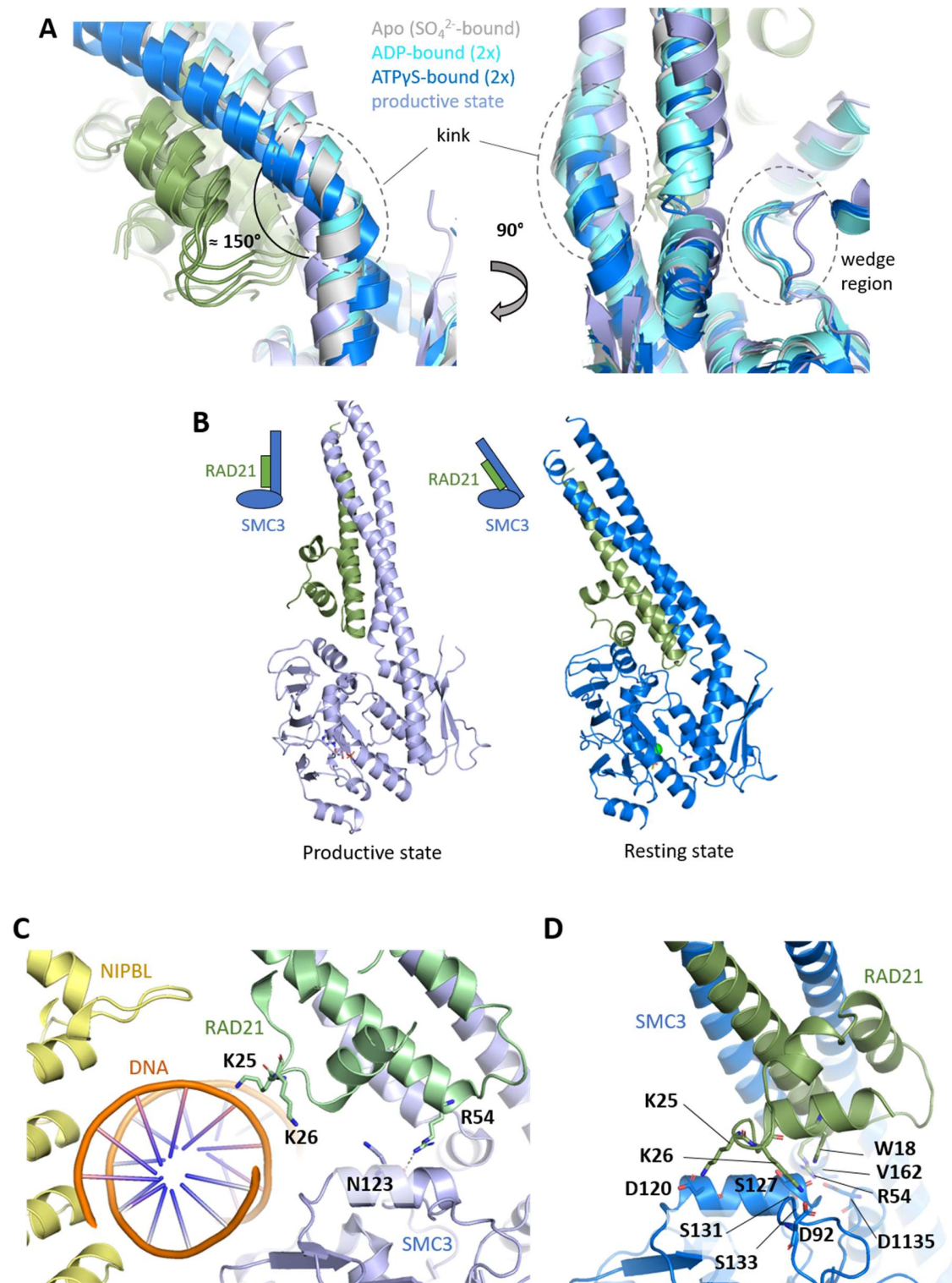
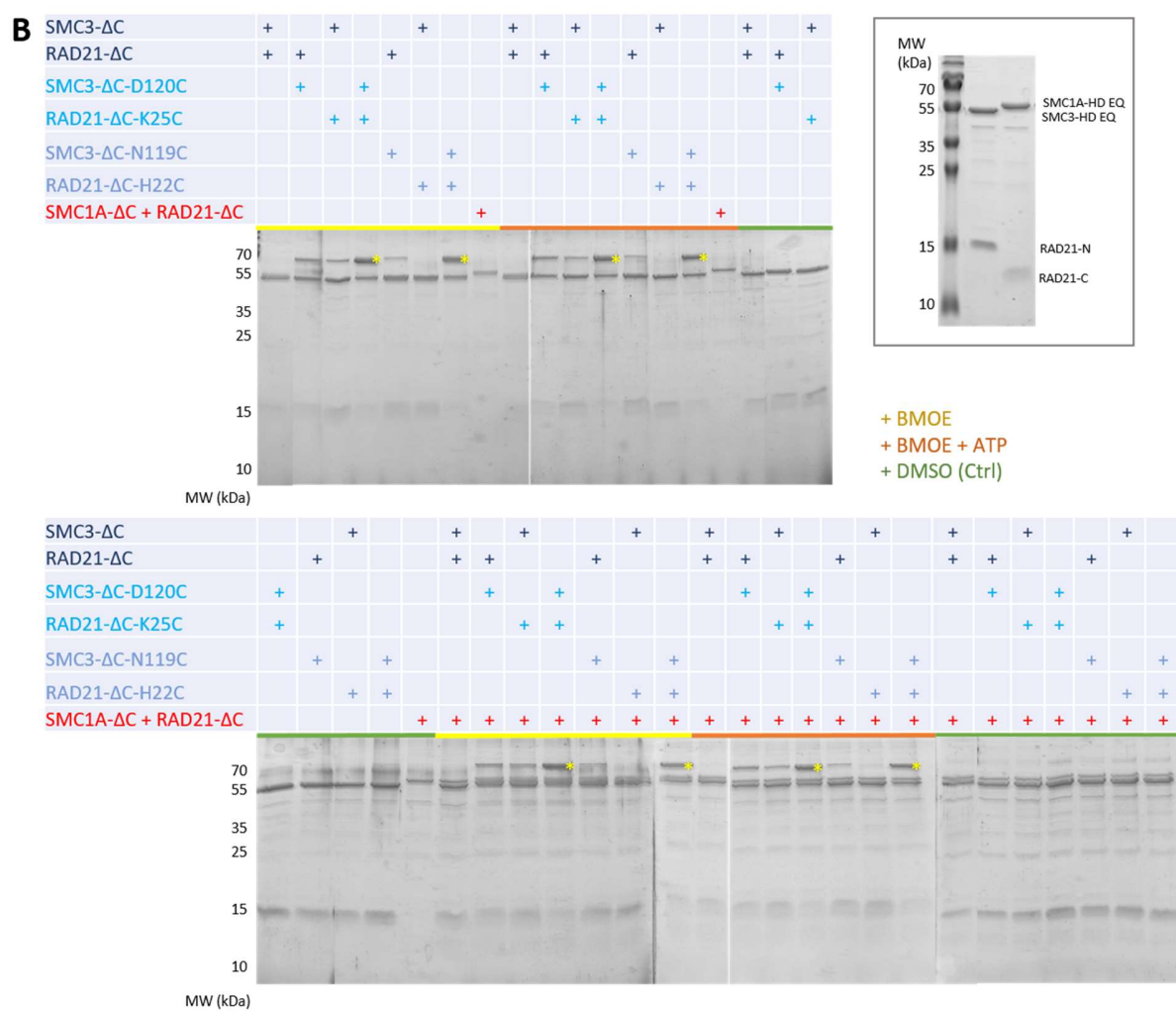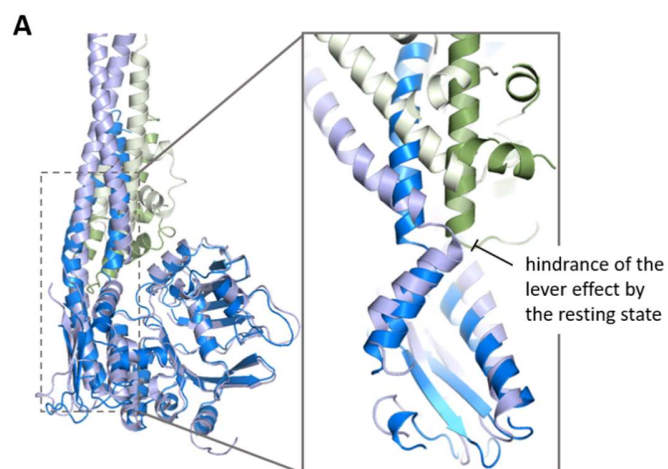**Figure 6.**

Figure 6. SMC3 adopts a resting state.

(A) Structure comparison of all the obtained crystallographic structures, with the cryo-EM structure of SMC3 in the productive mode (PDB: 6wg3). In the crystallographic structures SMC3 coiled-coil adopts a kinked conformation, which has direct consequences on the movement of the wedge region.

(B) Side-by-side comparison of the overall positioning of the RAD21 globular domain and of the SMC3 coiled coil, relative to the SMC3 globular domain, in the productive state (left) and in the resting state (right).
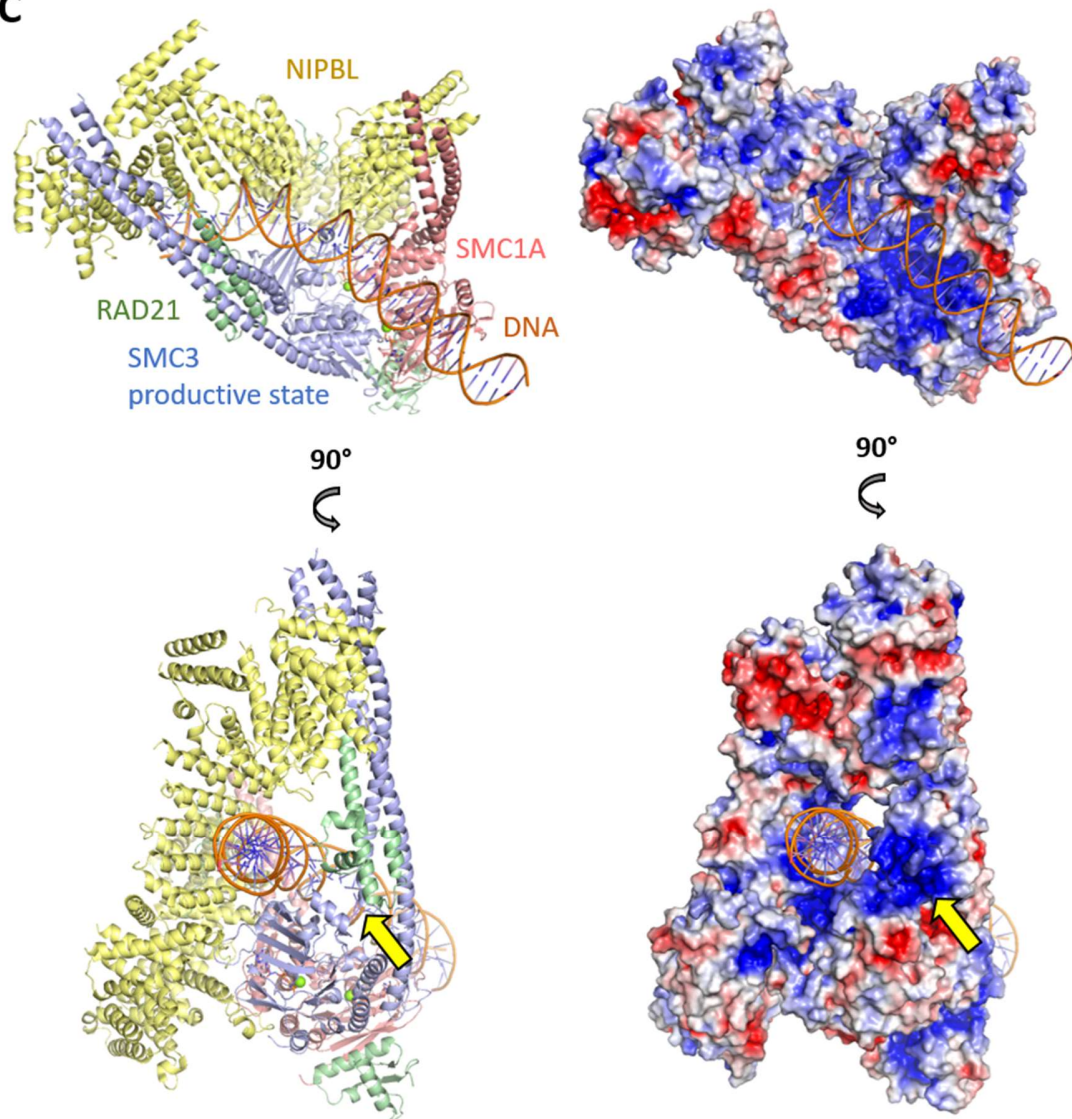
(C) Close-up view on the interactions of RAD21N globular region with DNA in the productive conformation of SMC3 (PDB: 6wg3).

(D) Same view as in (C), using a crystallographic ATPγS-bound SMC3CC structure. In this case RAD21N globular domain contacts the globular domain of SMC3.

Figure 7.

C

NIPBL

SMC1A

DNA

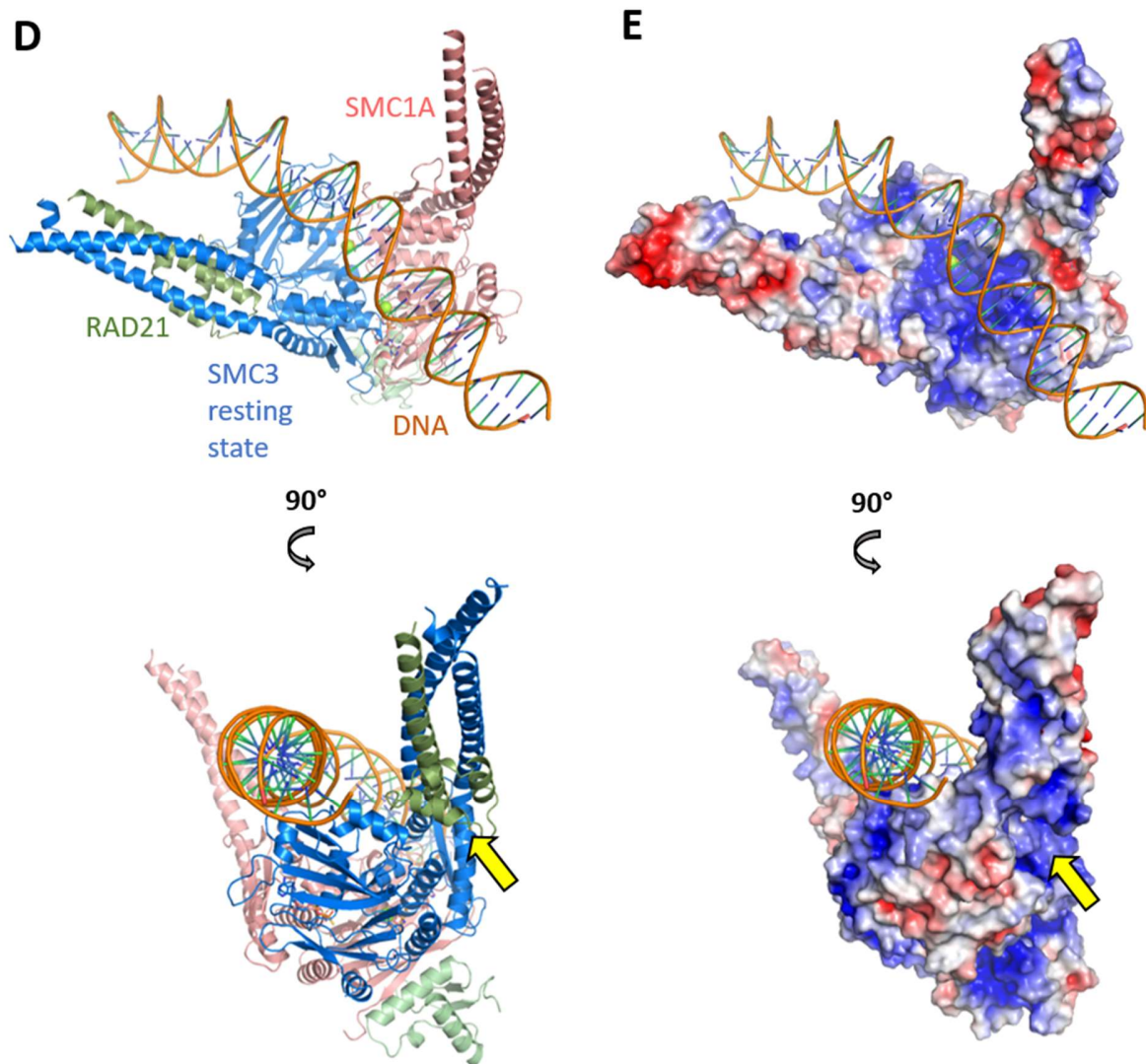RAD21

SMC3
productive state

90°

90°

Figure 7. SMC3 resting state does not prevent ATPase heads engagement.

(A) Structure comparison of SMC3 in the productive (PDB: 6wg3) and in the resting state showing that the resting state is incompatible with the lever effect that brings SMC3 to a productive state.

(B) BMOE crosslinking experiments made with pairs specific to either the resting state (SMC3CC-D120C and RAD21N-K25C, and SMC3CC-D119C and RAD21N-H22C, respectively). In yellow: experiment in presence of BMOE; in orange: in presence of BMOE and ATP; in green: controls (ctrl) in absence of BMOE and of ATP. Crosslinking can occur in both conformations in solution (yellow asterisk), and either in the SMC1A engaged or in the independent SMC3-HD. Some aspecific crosslinking can be observed in the control experiments, with a band appearing slightly higher than the actual crosslink. This is due to the solvent exposition of the mutated cysteines

which leads to aspecific SMC3-SMC3 or RAD21-RAD21 crosslinking. Top-right panel shows the SDS PAGE migration pattern of SMC1ACC, SMC3CC, RAD21N and RAD21C.

(C) Electrostatic potential of the roductive complex shows an acidic patch on RAD21N globular domain which interacts with DNA (bottom panels, indicated by a yellow arrow).

(D) Model of the engaged SMC1A and SMC3 heads with SMC3 in the resting state conformation, based on the cryo-EM cohesin ATPase structure (PDB: 4wge).

(E) In the resting state, the positively charged RAD21N patch is rotated away from the DNA molecule.

The distinct conformational dynamics of human SMC1A and SMC3 ATPase heads reveal a resting state for the Cohesin DNA exit gate

Gomes, M.[1], Landwerlin, P.[1], Diebold-Durand[1], M.-L, Shaik[1], T.B., Durand, A.[1], Brillet, K.[2], Troesch, E.[1], Antony, P[1]., Ennifar, E.[2] and Romier, C.[1,§]

[§]Correspondence should be addressed to Christophe Romier (romier@igbmc.fr).

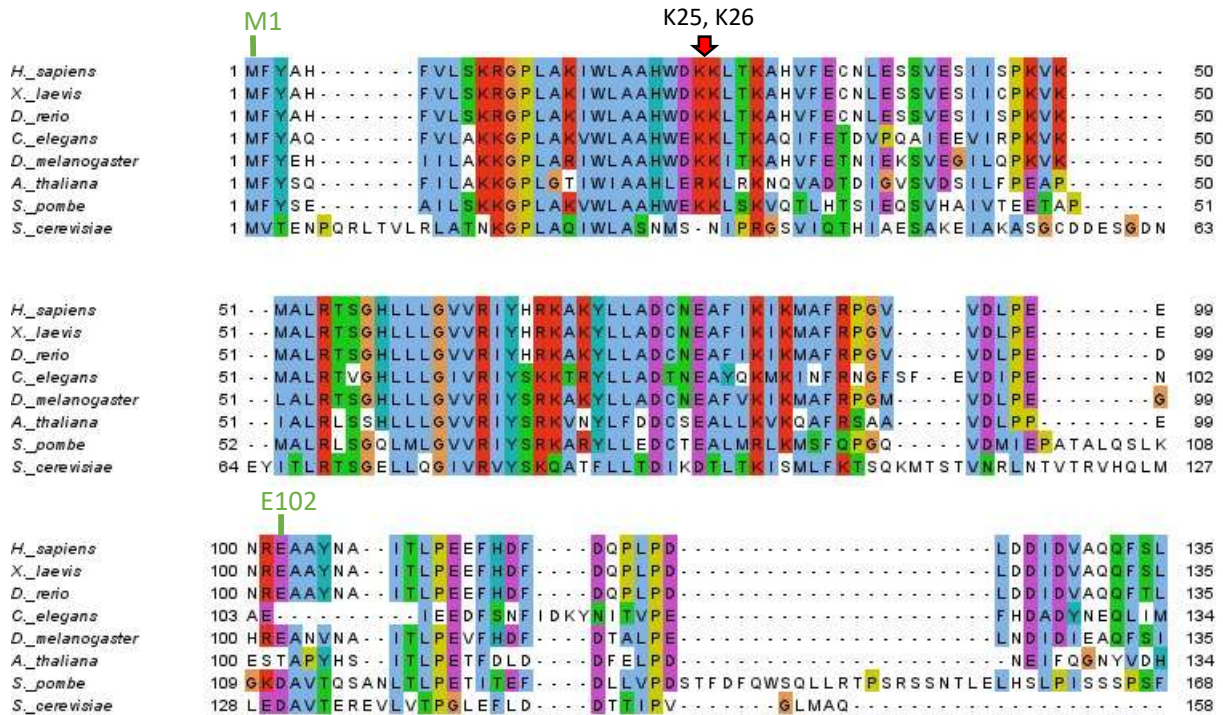**Figure S1.**

(A)

SMC1A-HD Nter

SMC1A-HD Cter

(B)

SMC3-HD Nter

SMC3-HD Cter

```
                                                             Q971      Y979
H._sapiens       931 MLLKKKEECMKKIRELGSLPQEAF-EKYQTLSLKQLFRKLEQCNTELKKYSHVNKKALDQFVNF  993
X._laevis        931 MLLKKKEECMKKIRELGSLPQEAF-EKYQTLSLKQLFRKLEQCNTELKKYSHVNKKALDQFVNF  993
D._rerio         931 MLLKKKEECMKKIRELGSLPQEAF-EKYQTLTLKQLFRKLEQCNTELKKYSHVNKKALDQFVNF  993
C._elegans       935 EVKQKREDSLKKMRLLGALPTDTF-SKWQNVKPRELEKKLLECVNELKKYENVNKKALDQYMTA  997
D._melanogaster  931 MLNEKIDECTEKIASLGAVPLVD--PSYTRMSLKNIFKELEKANQHLKKYNHVNKKALDQFLSF  992
A._thaliana      921 TLLAKQDEYTKKIRGLGPLSSDAF-DTYKRKNIKELQKMLHRCSEQLQQFSHVNKKALDQYVNF  983
S._pombe         919 LLLARKKECNEKIKSLGVLPEEAF-IKYVSTSSNAIVKKLHKINEALKDYGSVNKKAYEQFNNF  981
S._cerevisiae    942 TLVTRREELQQRIPEIGLLPEDALVNDFSDITSDQLLQRLNDMNTEISGLKNVNKRAFENFKKF 1005


H._sapiens       994 SEQKEKLIKRQEELDRGYKSIMELMNVLELRKYEAIQLTFKQVSKNFSEVFQKLVPGGKATLVM 1057
X._laevis        994 SEQKEKLIKRQEELDRGHKSIMELMNVLELRKYEAIQLTFKQVSKNFSEVFQKLVPGGKATLVM 1057
D._rerio         994 SEQKEKLIKRQDELDRGYKSIMELMNVLELRKYEAIQLTFKQVSKNFSEVFQKLVPGGKATLVM 1057
C._elegans       998 SSQKEELTKRMAEQKKSEDSIEELLKVLENRKYEAIDLTFKQVKKNFEQVFKQLVPHGRGKMQM 1061
D._melanogaster  993 SEQKEKLYRRKEELDIGDQKIHMLIQSLEMQKVEAIQFTFRQVAQNFTKVFKKLVPMGAGFLIL 1056
A._thaliana      984 TEQREELQNRQAELDAGDEKIKELITVLDQRKDESIERTFKGVAHHFRDVFSELVQDGYGNLII 1047
S._pombe         982 TKQRDSLLARREELRRSQESISELTTVLDQRKDEAIERTFKQVAKSFSEIFVKLVPAGRGELVM 1045
S._cerevisiae   1006 NERRKDLAERASELDESKDSIQDLIVKLKQQKVNAVDSTFQKVSENFEAVFERLVPRGTAKLII 1069


H._sapiens      1058 KKG-DVEGSQSQDEGEGSGESERGSGSQSSVPSVDQFTGVGIRVSFTGK--QGEMREMQQLSGG 1118
X._laevis       1058 KKG-DVEGSQSQDEGEGS--------TQSSVPSVDQFTGVGIRVSFTGK--QAEMREMQQLSGG 1110
D._rerio        1058 KKG-DTEGGQSQDEGEGG-DSERASSQSSVPSVDQFTGVGIRVSFTGK--QGEMREMQQLSGG 1117
C._elegans      1062 RAREQRDDEE----------------GINSVELYEGISVLVSFVSDDGDSETREMTQLSGG 1106
D._melanogaster 1057 KTK-DNEGDEMEKEVENS--------------DAFTGIGIRVSFTGV--EAEMREMNQLSGG 1101
A._thaliana     1048 MKKKDLDNDDEDDDDDDG-GREAV-----TEGRVEKYIGVKVKVSFTG---QGETQLMKQLSGG 1102
S._pombe        1046 NRRSELSQSIEQDISMDI-------DTPSQKSSIDNYTGISIRVSFNSK--DDEQLNINQLSGG 1100
S._cerevisiae   1070 HRKNDNANDHDESIDVDM-DAESN-ESQNGKDSEIMYTGVSISVSFNSK--QNEQLHVEQLSGG 1129


H._sapiens      1119 QKSLVALALIFAIQKCDPAPFYLFDEIDQALDAQHRKAVSDMIMELAVH--AQFITTTFRPELL 1180
X._laevis       1111 QKSLVALALIFAIQKCDPAPFYLFDEIDQALDAQHRKAVSDMIMELASH-AQFITTTFRPELL 1172
D._rerio        1118 QKSLVALALIFAIQKCDPAPFYLFDEIDQALDAQHRKAASDMIMELAGH--AQFITTTFRPELL 1179
C._elegans      1107 QKSLVALAIIFSIQKCDPAPFYLFDEIDQALDAQHRKSVADMIQSLSDQ--AQFVTTTFRPELL 1168
D._melanogaster 1102 QKSLVALALIFSIQKCDPAPFYLFDEIDQALDAMHRKAVANMIHELSDT--AQFITTTFRPELL 1163
A._thaliana     1103 QKTVVALALIFAIQRCDPAPFYLFDEIDAALDPQYRTAVGNLIRRLADDYGTQFITTTFRPELV 1166
S._pombe        1101 QKSLCALTLIFAIQRCDPAPFNILDECDANLDAQYRSAIAAMVKEMSKT--SQFICTTFRPEMV 1162
S._cerevisiae   1130 QKTVCAIALILAIQMVDPASFYLFDEIDAALDKQYRTAVATLLKELSKN--AQFICTTFRTDML 1191


                                                    G1217
H._sapiens      1181 ESADKFYGVKFRNKVSHIDVITAEMAKDFVEDDTTHG........................... 1217
X._laevis       1173 ESADKFYGVKFRNKVSHIDVITAEQAKDFVEDDTTHG........................... 1209
D._rerio        1180 ESADKFYGVKFRNKVSHIDVISAEQAKDFVEDDTTHG........................... 1216
C._elegans      1169 ATAEKFYGVRFRNKVSHIDSVTREQAYDFVEDDTTHG........................... 1205
D._melanogaster 1164 ENAHKFYGVRFRNKVSHIDCVTREEAKDFVEDDSTHA........................... 1200
A._thaliana     1167 RVADKIYGVFHKNRVSIVNVISKDQALDFIEKDQSHDT........................... 1204
S._pombe        1163 KVADNFYGVMFNHKVSTVESISKEEAMAFVEG............................... 1194
S._cerevisiae   1192 QVADKFFRVKYENKISTVIEVNREEAIGFIRGSNKFAE........................... 1229
```

183

(C)

RAD21-HD Nter

(D)

RAD21-HD Cter

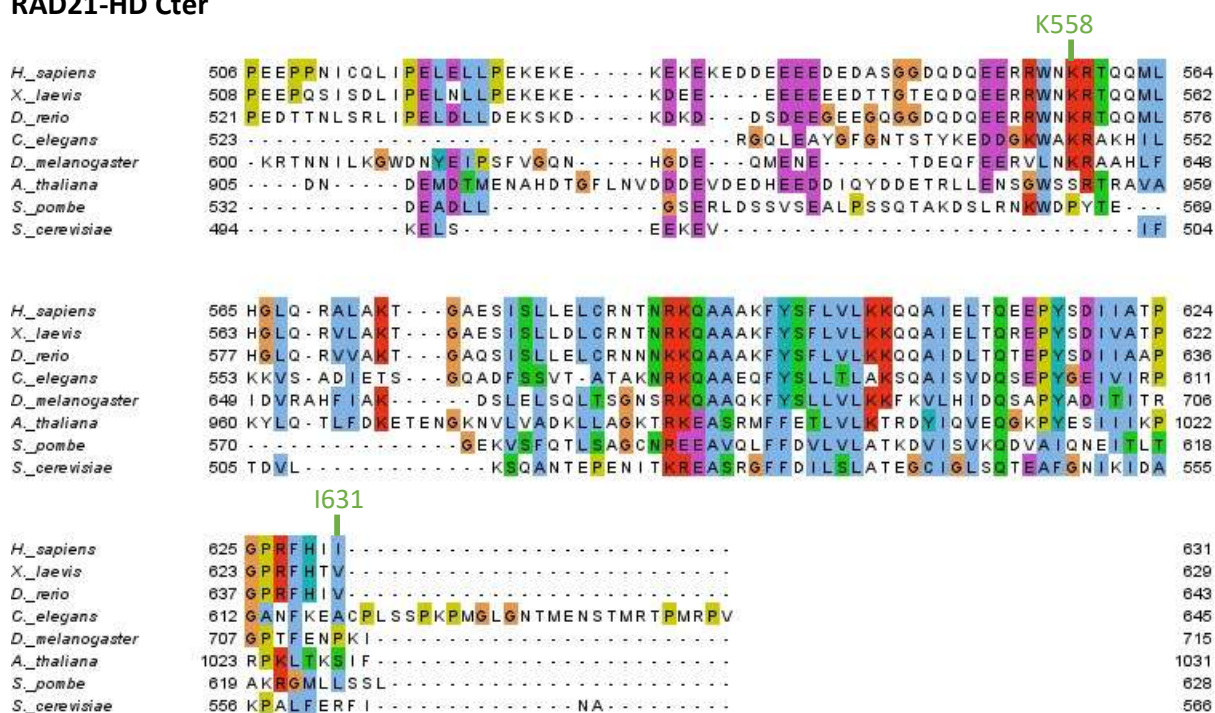**Supplementary Figure 1.** Sequence alignments of SMC1A HD, SMC3 HD, RAD21N and RAD21C of different model organisms.

The boundaries of the constructs used for crystallography experiments are annotated in pink for SMC1ACCsh, in red for SMC1A, in blue for SMC3CC, in gray for SMC3CC_2, in green for RAD21, and the shared boundaries of different constructs are annotated in black.

Figure S2.
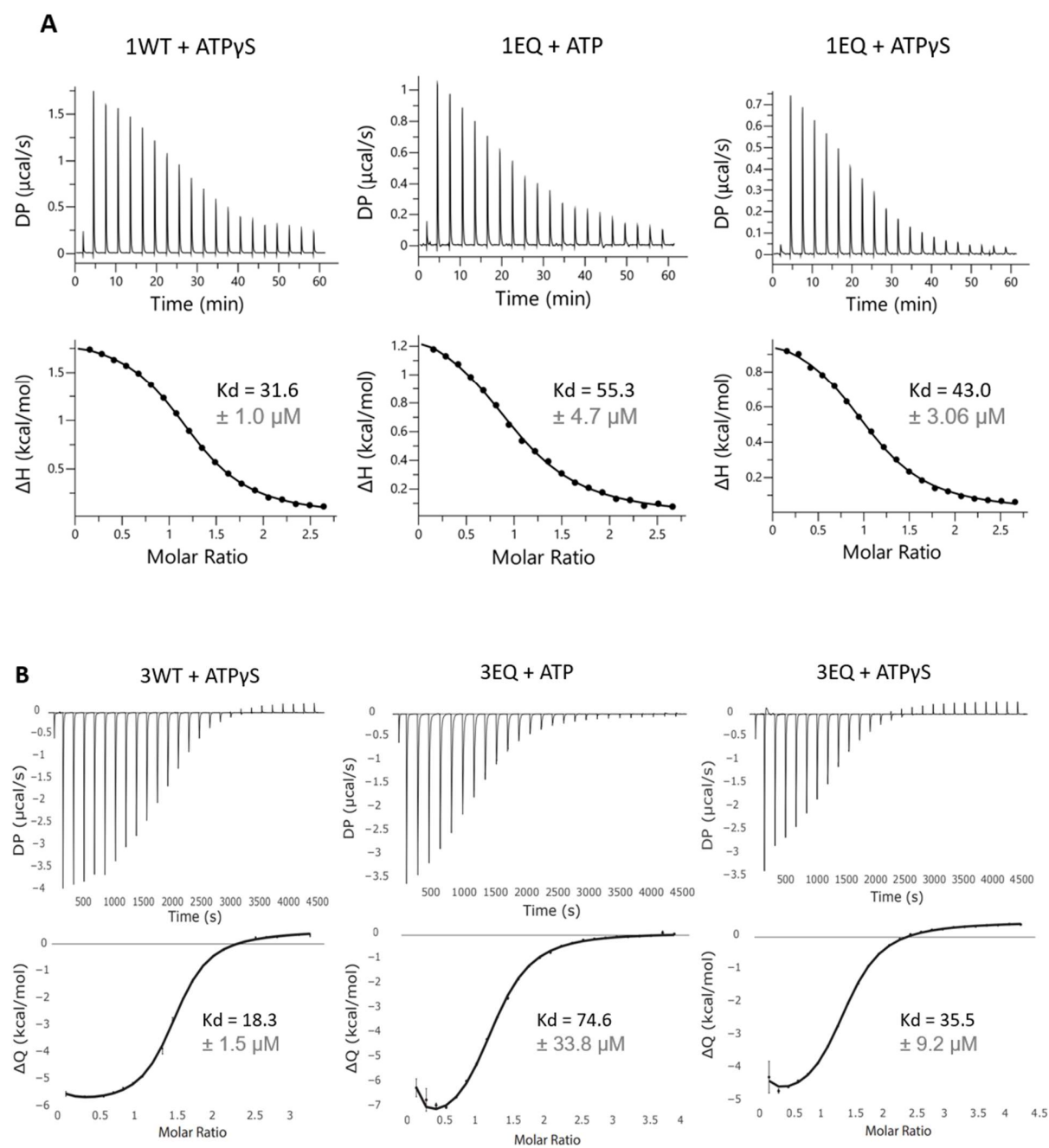
Supplementary figure 2. ITC of the binding of ATP and ATPγS to SMC1A and SMC3 ATPase heads.

(A) Binding of ATP and ATPγS to SMC1ACC wild type (1WT) and EQ mutant (1EQ).

(B) Binding of ATP and ATPγS to SMC3CC wild type (3WT) and EQ mutant (3EQ).
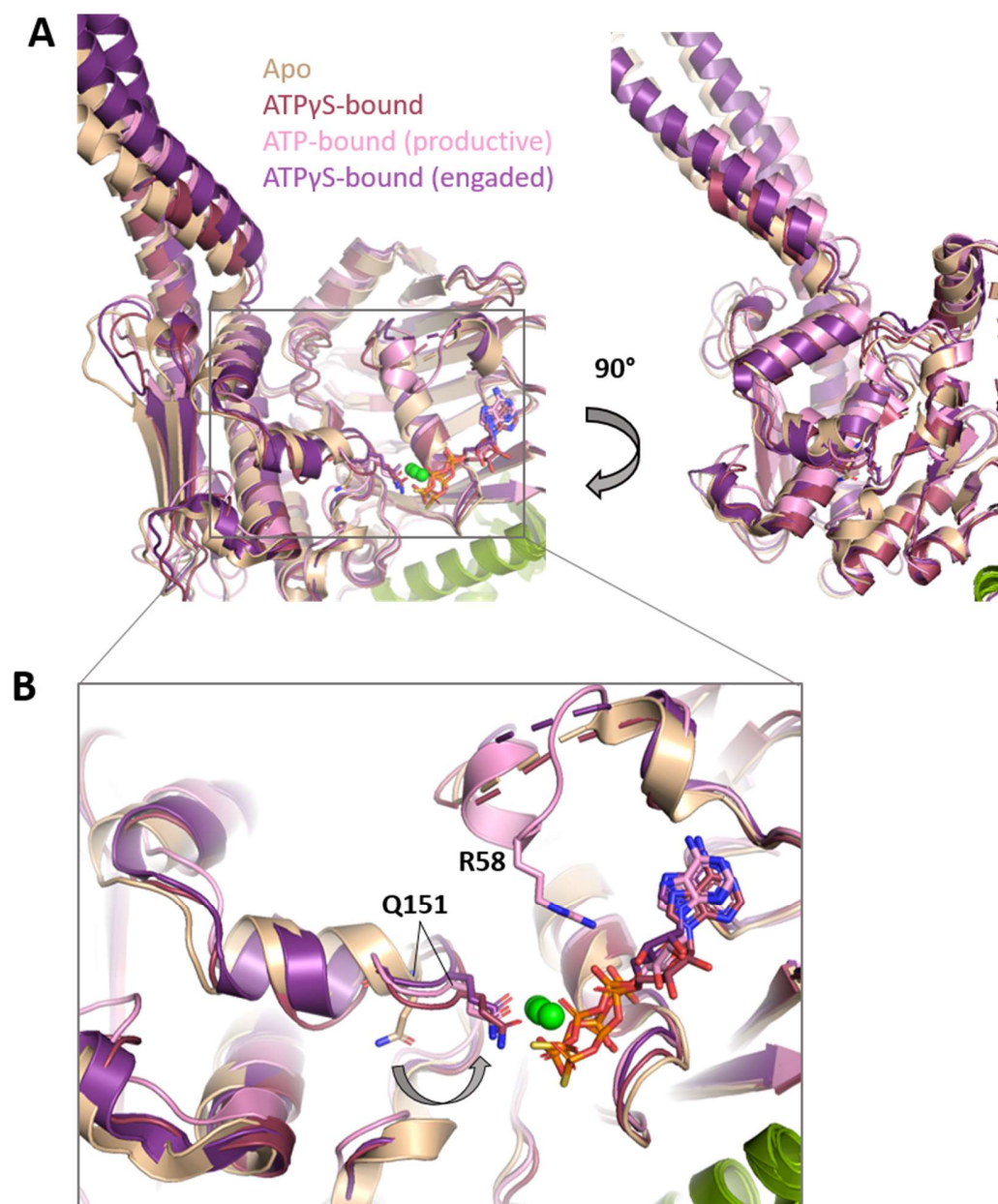
(C) Graphical representation of the thermodynamic parameters of ATP, ADP and ATPγS to SMC1ACC wild type (1WT) and EQ mutant (1EQ).

(D) Graphical representation of the thermodynamic parameters of ATP, ADP and ATPγS to SMC3CC wild type (3WT) and EQ mutant (3EQ).

ΔG: free energy, in blue; ΔH: enthalpy, in green; -TΔS: temperature times the change in entropy, in red.

Figure S3.
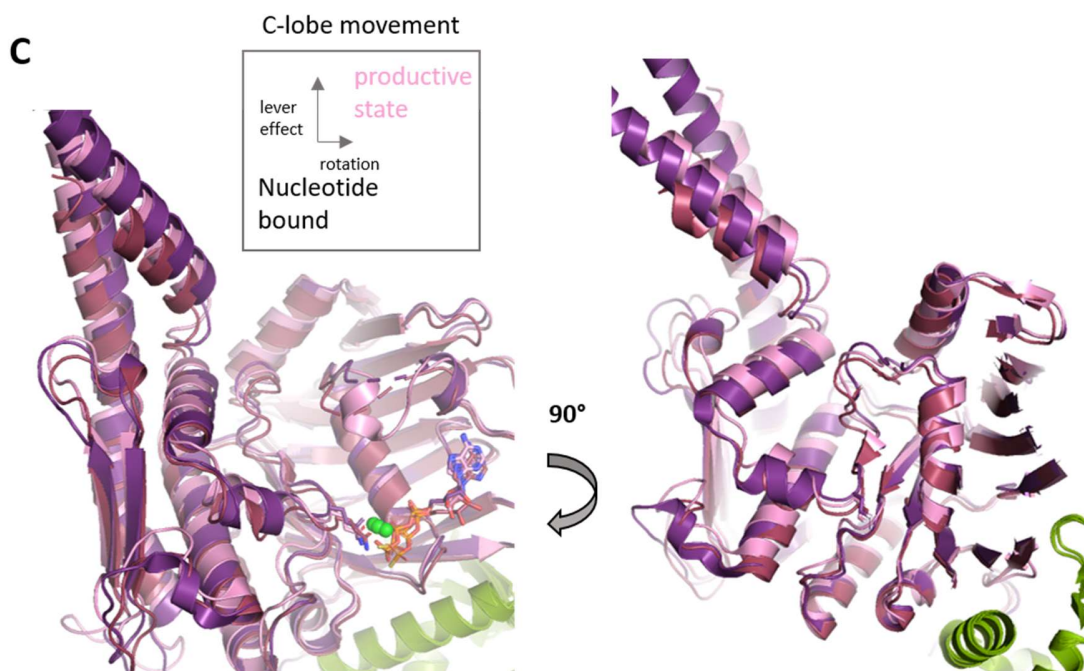
Supplementary Figure 3. Structural consequences of ATP binding to yeast SMC1.

(A) Structure comparison of the available yeast Smc1-HD structures, in an apo form (*C. thermophilum*, PDB: 6qpq), in an ATPγS-bound form (*S. cerevisiae*, PDB: 1w1w), in an ATPγS - bound and engaged with Smc3-HD (*S. cerevisiae*, PDB: 6qpw), and in the productive ATP-bound and engaged with Smc3, NIPBL[Scc2] and DNA form (*S. cerevisiae*, PDB: 6zz6).

(B) A movement of the helical domain positions the glutamine of the Q-loop in hydrogen bonding distance to the Mg ion and to the γ group of the nucleotide.

(C) A low amplitude lever effect is observed to reach the productive state.

Figure S4.
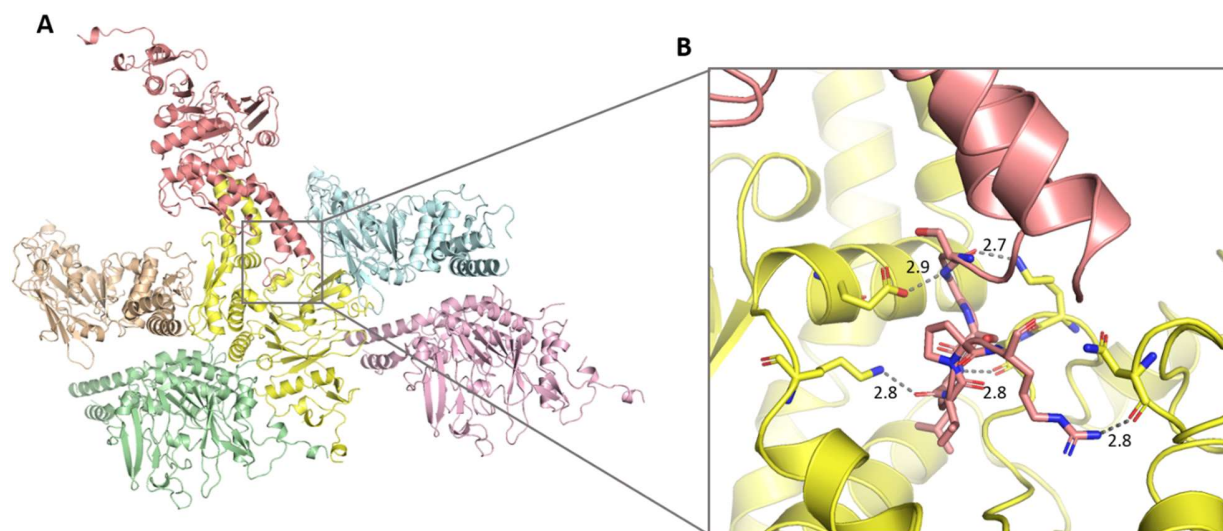
Supplementary Figure 4. Effect of the crystal packing on the structure of SMC1ACCsh.

(A) The molecule in the asymmetric unit is colored in yellow, and symmetric molecules with different symmetry operators are shown in salmon, pale cyan, pale pink, pale green and wheat. All SMC1ACCsh crystallographic structures shared a similar crystal packing.

(B) Close-up view of the hydrogen bond interactions made by a symmetric molecule linker with the helical- and RecA-lobes of SMC1ACCsh. Measurements between atoms are given in Angstrom.

Figure S5.
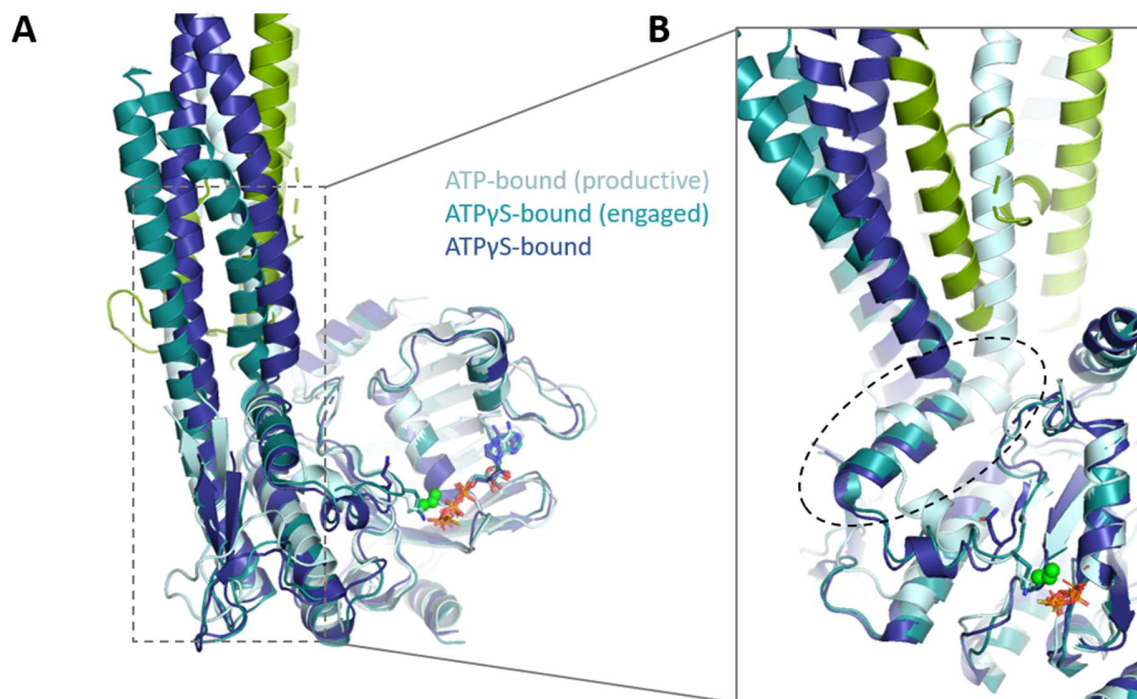
Supplementary Figure 5. SMC3 features in yeast.

(A) Comparison of the yeast published Smc3 structures of *S. cerevisiae* (PDB: 4ux3, 6qpw, 6zz6).

(B) Amplitude of the lever effect on the yeast Smc3.

Figure S6.
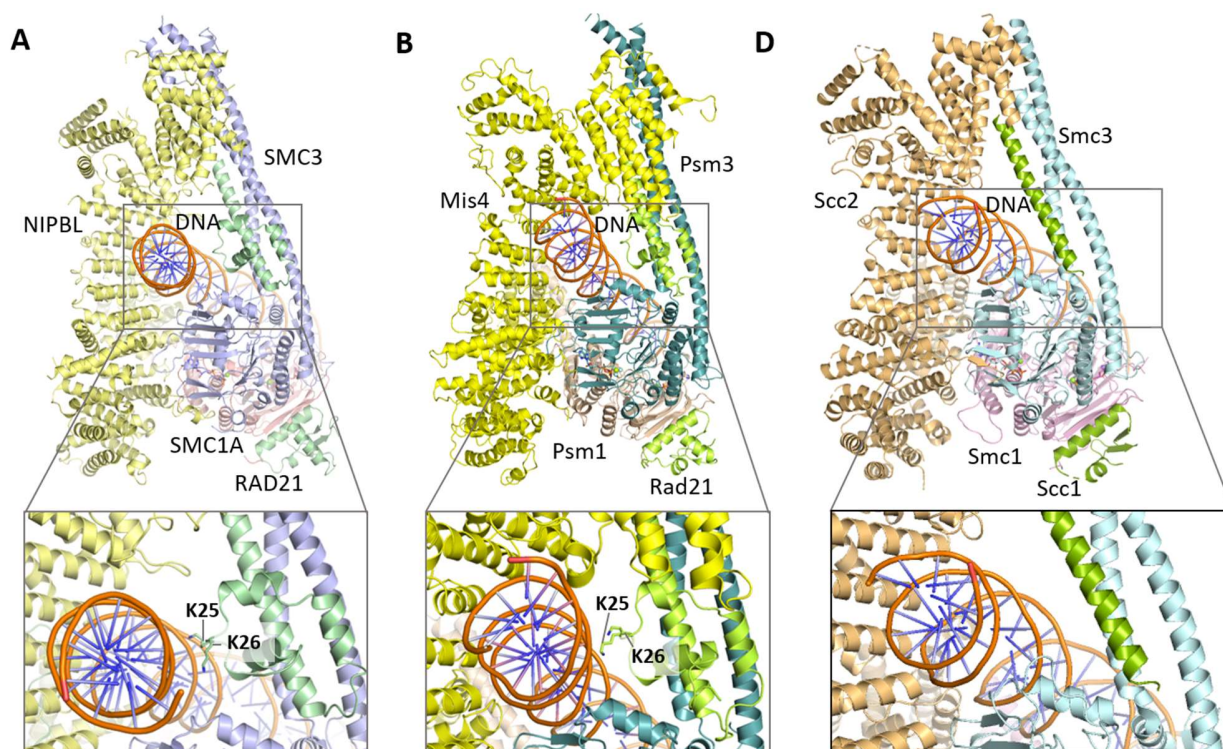
Supplementary Figure 6. The Cohesin DNA exit gate configuration in human versus in yeasts.

(A) Structure of the human cohesin DNA exit gate in the productive state (PDB: 6wge).

(B) Structure of the yeast S. pombe cohesin in the productive state (PDB: 6yuf).

(C) Structure of the yeast S. cerevisiae cohesin in the productive state (PDB: 6zz6).

Supplementary Table 1. Crystallographic statistics for SMC1ACC structures.

Data collection*	SMR1ACC-WT-apo	SMR1ACC-EQ-apo	SMR1ACC-WT-loop
Space group	C2	C2	C2
Cell dimensions			
a, b, c (Å)	187.93, 64.72, 47.34	158.55, 67.07, 51.63	187.60, 64.59, 47.13
α, β, γ (°)	90.00, 103.07, 90.00	90.00, 92.32, 90.00	90.00, 103.97, 90.00
Resolution (Å)	50.00 – 2.20 (2.34 – 2.20)	50.00 – 1.90 (2.02 – 1.90)	50.00 – 2.09 (2.22 – 2.09)
Rsym or Rmerge	0.072 (1.196)	0.091 (1.545)	0.06 (1.206)
I / σI	12.08 (1.05)	9.56 (1.02)	12.64 (1.32)
Completeness (%)	96.2 (87.3)	99.7 (98.3)	98.9 (95.5)
Redundancy	4.9 (4.4)	6.8 (6.9)	4.6 (4.6)
CC(1/2)	99.8 (51.3)	99.8 (71.9)	99.9 (72.0)
Refinement			
Resolution (Å)	46.12 – 2.20	44.05 – 1.90	32.30 – 2.09
No. reflections	27871	42461	32074
Rwork / Rfree	0.190 / 0.244	0.195 / 0.231	0.199 / 0.223
Number of atoms			
Protein	3248	3288	3136
Ligand/ion	-	-	-
Water	79	132	52
B-factors (Å^2)			
Protein	72.77	62.17	82.58
Ligand/ion	-	-	-
Water	58.40	53.06	58.55
R.m.s. deviations			
Bond lengths (Å)	0.008	0.006	0.008
Bond angles (°)	0.913	0.771	0.905
* Values in parentheses are for the highest-resolution shell.			

Data collection*	SMR1ACC-EQ-loop	SMR1ACC-EQ-ADP	SMR1ACC-EQ-AGS-Mg
Space group	C2	C2	$P4_1 2_1 2$
Cell dimensions			
a, b, c (Å)	187.92, 64.60, 89.95	189.31, 65.28, 105.76	68.92, 68.92, 215.33
α, β, γ (°)	90.00, 99.74, 90.00	90.00, 116.66, 90.00	90.00, 90.00, 90.00
Resolution (Å)	50.00 – 1.77 (1.87 – 1.77)	50.00 – 2.44 (2.58 – 2.44)	50.00 – 2.50 (2.65 – 2.50)
Rsym or Rmerge	0.055 (1.277)	0.120 (1.121)	0.126 (3.602)
I / σI	15.94 (1.05)	11.60 (1.26)	19.14 (0.84)
Completeness (%)	99.7 (98.7)	99.5 (98.3)	98.4 (98.5)
Redundancy	6.78 (6.2)	6.4 (5.2)	24.7 (24.9)
CC(1/2)	99.9 (58.0)	99.8 (57.9)	100.0 (53.8)
Refinement			
Resolution (Å)	44.63 – 1.77	47.80 – 2.44	48.74 – 2.50
No. reflections	103917	43312	18516
Rwork / Rfree	0.190 / 0.218	0.194 / 0.250	0.221 / 0.298
Number of atoms			
Protein	6489	6578	3312
Ligand/ion	19	54	32
Water	389	64	32
B-factors (Å²)			
Protein	47.11	60.76	93.06
Ligand/ion	62.46	57.85	67.92
Water	46.96	51.23	81.54
R.m.s. deviations			
Bond lengths (Å)	0.007	0.008	0.008
Bond angles (º)	0.886	0.965	1.024
* Values in parentheses are for the highest-resolution shell.			

Supplementary Table 2. Crystallographic statistics for SMC1ACCsh structures.

Data collection*	SMR1ACCsh-EQ-apo	SMR1ACCsh-EQ-loop-SO$_4$	SMR1ACCsh-EQ-ADP
Space group	I 2 2 2	I 2 2 2	I 2 2 2
Cell dimensions			
a, b, c (Å)	66.14, 114.04, 134.87	66.11, 113.84, 134.47	66.04, 113.84, 134.19
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	50.00 - 1.85 (1.96 – 1.85)	50.00 – 1.50 (1.59 – 1.50)	50.00 – 1.36 (1.44 -1.36)
Rsym or Rmerge	0.078 (1.308)	0.087 (2.382)	0.067 (2.855)
I / σI	21.22 (1.68)	17.87 (1.08)	20.31 (0.92)
Completeness (%)	99.9 (99.2)	99.9 (99.4)	99.9 (99.7)
Redundancy	12.7 (10.1)	13.2 (12.8)	13.2 (13.5)
CC(1/2)	100.0 (75.6)	100.0 (43.2)	100.0 (38.7)
Refinement			
Resolution (Å)	43.63 – 1.85	43.55 – 1.50	43.49 – 1.36
No. reflections	43937	81150	108065
Rwork / Rfree	0.181 / 0.222	0.198 / 0.232	0.204 / 0.219
Number of atoms			
Protein	3397	3313	3413
Ligand/ion	-	5	27
Water	257	360	301
B-factors (Å^2)			
Protein	41.05	32.12	31.19
Ligand/ion	-	22.86	47.72
Water	46.15	42.17	38.25
R.m.s. deviations			
Bond lengths (Å)	0.006	0.006	0.005
Bond angles (º)	0.809	0.831	0.833
Values in parentheses are for the highest-resolution shell.			

Data collection*	SMR1ACCsh-EQ-ADP-Mg	SMR1ACCsh-EQ-AGS-Mg
Space group	I 2 2 2	I 2 2 2
Cell dimensions		
a, b, c (Å)	66.45, 113.91, 133.84	66.22, 113.98, 134.07
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	50.00 – 1.65 (1.75 – 1.65)	50.00 – 1.94 (2.06– 1.94)
Rsym or Rmerge	0.113 (2.391)	0.166 (2.378)
I / σI	15.52 (1.05)	10.95 (1.06)
Completeness (%)	99.8 (99.1)	99.8 (99.0)
Redundancy	13.3 (13.2)	12.9 (11.7)
CC(1/2)	99.9 (42.0)	99.9 (36.6)
Refinement		
Resolution (Å)	43.57 - 1.65	43.54 - 1.94
No. reflections	61052	37968
Rwork / Rfree	0.196 / 0.228	0.205 / 0.253
Number of atoms		
Protein	3401	3289
Ligand/ion	28	32
Water	302	212
B-factors (Å^2)		
Protein	33.63	43.99
Ligand/ion	47.22	64.80
Water	42.81	46.12
R.m.s. deviations		
Bond lengths (Å)	0.007	0.007
Bond angles (º)	0.829	0.881
* Values in parentheses are for the highest-resolution shell.		

Supplementary Table 3. Crystallographic statistics for SMC3 structures.

Data collection*	SMR3CC-EQ-ADP	SMR3CC-WT-ADP-Mg	SMR3CC-EQ-AGS1-Mg
Space group	$P4_1 2_1 2$	$P4_1 2_1 2$	$C2 2 2_1$
Cell dimensions			
a, b, c (Å)	89.95, 89.95, 233.91	89.71, 89.71, 234.65	118.24, 154.46, 136.54
α, β, γ (°)	90.00, 90.00, 90.000	90.00, 90.00, 90.000	90.00, 90.00, 90.00
Resolution (Å)	50.00 – 2.45 (2.60 – 2.45)	50.00 – 3.00 (3.18 – 3.00)	50.00 – 2.25 (2.39 - 2.25)
Rsym or Rmerge	0.154 (2.621)	0.247 (3.684)	0.098 (2.233)
I / σI	22.60 (1.50)	16.53 (1.11)	20.32 (1.34)
Completeness (%)	99.9 (99.7)	99.9 (99.6)	98.0 (94.2)
Redundancy	26.1 (26.4)	26.0 (26.9)	13.6 (13.1)
CC(1/2)	99.9 (58.1)	99.9 (43.1)	100.0 (81.6)
Refinement			
Resolution (Å)	49.03 - 2.45	49.27 – 3.00	47.21 – 2.25
No. reflections	36209	19996	58103
Rwork / Rfree	0.195 / 0.239	0.196 / 0.264	0.202 / 0.253
Number of atoms			
Protein	3918	3931	7529
Ligand/ion	27	28	64
Water	110	4	90
B-factors (Å^2)			
Protein	74.80	96.65	75.14
Ligand/ion	55.69	79.86	47.68
Water	62.44	71.39	56.96
R.m.s. deviations			
Bond lengths (Å)	0.008	0.009	0.008
Bond angles (º)	0.940	1.080	0.927
Values in parentheses are for the highest-resolution shell.			

Data collection*	SMR3CC-EQ-AGS2-Mg	SMR3CC_2-EQ-SO$_4$
Space group	P6$_4$ 2 2	P4$_1$ 2$_1$ 2
Cell dimensions		
a, b, c (Å)	201.07, 201.07, 92.38	90.79, 90.79, 236.87
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.000
Resolution (Å)	50.00 – 3.11 (3.30 – 3.11)	50.00 – 2.60 (2.76 – 2.60)
Rsym or Rmerge	0.256 (4.814)	0.280 (3.842)
I / σI	17.00 (1.03)	14.93 (1.02)
Completeness (%)	99.5 (98.1)	99.9 (99.7)
Redundancy	39.1 (39.1)	26.2 (25.7)
CC(1/2)	99.9 (63.7)	99.9 (36.6)
Refinement		
Resolution (Å)	44.64 – 3.11	49.60 – 2.60
No. reflections	20146	31408
Rwork / Rfree	0.187 / 0.238	0.215 / 0.252
Number of atoms		
Protein	3955	4040
Ligand/ion	32	11
Water	-	9
B-factors (Å^2)		
Protein	112.95	71.43
Ligand/ion	93.80	75.57
Water	-	53.35
R.m.s. deviations		
Bond lengths (Å)	0.008	0.008
Bond angles (º)	0.997	0.906
* Values in parentheses are for the highest-resolution shell.		

References

1. Peters, J.M., Tedeschi, A. & Schmitz, J. The cohesin complex and its roles in chromosome biology. Genes Dev 22, 3089-114 (2008).

2. Nasmyth, K. & Haering, C.H. Cohesin: its roles and mechanisms. Annu Rev Genet 43, 525-58 (2009).

3. Hirano, T. Condensin-Based Chromosome Organization from Bacteria to Vertebrates. Cell 164, 847-57 (2016).

4. Uhlmann, F. SMC complexes: from DNA to chromosomes. Nat Rev Mol Cell Biol 17, 399-412 (2016).

5. Gruber, S. Shaping chromosomes by DNA capture and release: gating the SMC rings. Curr Opin Cell Biol 46, 87-93 (2017).

6. Yatskevich, S., Rhodes, J. & Nasmyth, K. Organization of Chromosomal DNA by SMC Complexes. Annu Rev Genet 53, 445-482 (2019).

7. Datta, S., Lecomte, L. & Haering, C.H. Structural insights into DNA loop extrusion by SMC protein complexes. Curr Opin Struct Biol 65, 102-109 (2020).

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

9. Davidson, I.F. & Peters, J.M. Genome folding through loop extrusion by SMC complexes. Nat Rev Mol Cell Biol 22, 445-464 (2021).

10. Haering, C.H. & Gruber, S. Snapshot: SMC protein complexes Part I. Cell 164, 326 (2016).

11. Anderson, D.E., Losada, A., Erickson, H.P. & Hirano, T. Condensin and cohesin display different arm conformations with characteristic hinge angles. J Cell Biol 156, 419-24 (2002).

12. Burmann, F. et al. A folded conformation of MukBEF and cohesin. Nat Struct Mol Biol 26, 227-236 (2019).

13. Collier, J.E. et al. Transport of DNA within cohesin involves clamping on top of engaged heads by Scc2 and entrapment within the ring by Scc3. Elife 9(2020).

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

15. Gligoris, T.G. et al. Closing the cohesin ring: structure and function of its Smc3-kleisin interface. Science 346, 963-7 (2014).

16. Griese, J.J., Witte, G. & Hopfner, K.P. Structure and DNA binding activity of the mouse condensin hinge domain highlight common and diverse features of SMC proteins. Nucleic Acids Res 38, 3454-65 (2010).

17. Haering, C.H., Lowe, J., Hochwagen, A. & Nasmyth, K. Molecular architecture of SMC proteins and the yeast cohesin complex. Mol Cell 9, 773-88 (2002).

18. Haering, C.H. et al. Structure and stability of cohesin's Smc1-kleisin interaction. Mol Cell 15, 951-64 (2004).

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

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

21. Kamada, K., Su'etsugu, M., Takada, H., Miyata, M. & Hirano, T. Overall Shapes of the SMC-ScpAB Complex Are Determined by Balance between Constraint and Relaxation of Its Structural Parts. Structure 25, 603-616 e4 (2017).

22. Kurze, A. et al. A positively charged channel within the Smc1/Smc3 hinge required for sister chromatid cohesion. EMBO J 30, 364-78 (2011).

23. Shi, Z., Gao, H., Bai, X.C. & Yu, H. Cryo-EM structure of the human cohesin-NIPBL-DNA complex. Science 368, 1454-1459 (2020).

24. Burmann, F. et al. An asymmetric SMC-kleisin bridge in prokaryotic condensin. Nat Struct Mol Biol 20, 371-9 (2013).

25. Lammens, A., Schele, A. & Hopfner, K.P. Structural biochemistry of ATP-driven dimerization and DNA-stimulated activation of SMC ATPases. Curr Biol 14, 1778-82 (2004).

26. Hassler, M., Shaltiel, I.A. & Haering, C.H. Towards a Unified Model of SMC Complex Function. Curr Biol 28, R1266-R1281 (2018).

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

28. Minnen, A. et al. Control of Smc Coiled Coil Architecture by the ATPase Heads Facilitates Targeting to Chromosomal ParB/parS and Release onto Flanking DNA. Cell Rep 14, 2003-16 (2016).

29. Soh, Y.M. et al. Molecular basis for SMC rod formation and its dissolution upon DNA binding. Mol Cell 57, 290-303 (2015).

30. Vazquez Nunez, R., Polyhach, Y., Soh, Y.M., Jeschke, G. & Gruber, S. Gradual opening of Smc arms in prokaryotic condensin. Cell Rep 35, 109051 (2021).

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

32. Xiang, S. & Koshland, D. Cohesin architecture and clustering in vivo. Elife 10(2021).

33. Lee, B.G. et al. Cryo-EM structures of holo condensin reveal a subunit flip-flop mechanism. Nat Struct Mol Biol 27, 743-751 (2020).

34. Petela, N.J. et al. Folding of cohesin's coiled coil is important for Scc2/4-induced association with chromosomes. Elife 10(2021).

35. De Koninck, M. & Losada, A. Cohesin Mutations in Cancer. Cold Spring Harb Perspect Med 6(2016).

36. Kline, A.D. et al. Diagnosis and management of Cornelia de Lange syndrome: first international consensus statement. Nat Rev Genet (2018).

37. Mintzas, K. & Heuser, M. Emerging strategies to target the dysfunctional cohesin complex in cancer. Expert Opin Ther Targets 23, 525-537 (2019).

38. Watrin, E., Kaiser, F.J. & Wendt, K.S. Gene regulation and chromatin organization: relevance of cohesin mutations to human disease. Curr Opin Genet Dev 37, 59-66 (2016).

39. Kulemzina, I. et al. A Reversible Association between Smc Coiled Coils Is Regulated by Lysine Acetylation and Is Required for Cohesin Association with the DNA. Mol Cell 63, 1044-54 (2016).

40. Haering, C.H., Farcas, A.M., Arumugam, P., Metson, J. & Nasmyth, K. The cohesin ring concatenates sister DNA molecules. Nature 454, 297-301 (2008).

41. Kurokawa, Y. & Murayama, Y. DNA Binding by the Mis4(Scc2) Loader Promotes Topological DNA Entrapment by the Cohesin Ring. Cell Rep 33, 108357 (2020).

42. Ladurner, R. et al. Cohesin's ATPase activity couples cohesin loading onto DNA with Smc3 acetylation. Curr Biol 24, 2228-37 (2014).

43. Minamino, M., Higashi, T.L., Bouchoux, C. & Uhlmann, F. Topological in vitro loading of the budding yeast cohesin ring onto DNA. Life Sci Alliance 1(2018).

44. Murayama, Y. & Uhlmann, F. Biochemical reconstitution of topological DNA binding by the cohesin ring. Nature 505, 367-71 (2014).

45. Murayama, Y. & Uhlmann, F. DNA Entry into and Exit out of the Cohesin Ring by an Interlocking Gate Mechanism. Cell 163, 1628-40 (2015).

46. Srinivasan, M. et al. The Cohesin Ring Uses Its Hinge to Organize DNA Using Non-topological as well as Topological Mechanisms. Cell 173, 1508-1519 e18 (2018).

47. Davidson, I.F. et al. DNA loop extrusion by human cohesin. Science 366, 1338-1345 (2019).

48. Golfier, S., Quail, T., Kimura, H. & Brugues, J. Cohesin and condensin extrude DNA loops in a cell cycle-dependent manner. Elife 9(2020).

49. Kim, Y., Shi, Z., Zhang, H., Finkelstein, I.J. & Yu, H. Human cohesin compacts DNA by loop extrusion. Science 366, 1345-1349 (2019).

50. Gruber, S. et al. Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. Cell 127, 523-37 (2006).

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

52. Chan, K.L. et al. Cohesin's DNA exit gate is distinct from its entrance gate and is regulated by acetylation. Cell 150, 961-74 (2012).

53. Eichinger, C.S., Kurze, A., Oliveira, R.A. & Nasmyth, K. Disengaging the Smc3/kleisin interface releases cohesin from Drosophila chromosomes during interphase and mitosis. EMBO J 32, 656-65 (2013).

54. Huis in 't Veld, P.J. et al. Characterization of a DNA exit gate in the human cohesin ring. Science 346, 968-72 (2014).

55. Haarhuis, J.H., Elbatsh, A.M. & Rowland, B.D. Cohesin and its regulation: on the logic of X-shaped chromosomes. Dev Cell 31, 7-18 (2014).

56. Kueng, S. et al. Wapl controls the dynamic association of cohesin with chromatin. Cell 127, 955-67 (2006).

57. Hauf, S., Waizenegger, I.C. & Peters, J.M. Cohesin cleavage by separase required for anaphase and cytokinesis in human cells. Science 293, 1320-3 (2001).

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

59. Uhlmann, F., Wernic, D., Poupart, M.A., Koonin, E.V. & Nasmyth, K. Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. Cell 103, 375-86 (2000).

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

61. Hara, K. et al. Structure of cohesin subcomplex pinpoints direct shugoshin-Wapl antagonism in centromeric cohesion. Nat Struct Mol Biol 21, 864-70 (2014).

62. Li, Y. et al. Structural basis for Scc3-dependent cohesin recruitment to chromatin. Elife 7(2018).

63. Li, Y. et al. The structural basis for cohesin-CTCF-anchored loops. Nature 578, 472-476 (2020).

64. Ciosk, R. et al. Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. Mol Cell 5, 243-54 (2000).

65. Arumugam, P. et al. ATP hydrolysis is required for cohesin's association with chromosomes. Curr Biol 13, 1941-53 (2003).

66. Dauban, L. et al. Regulation of Cohesin-Mediated Chromosome Folding by Eco1 and Other Partners. Mol Cell 77, 1279-1293 e4 (2020).

67. Hinshaw, S.M., Makrantoni, V., Kerr, A., Marston, A.L. & Harrison, S.C. Structural evidence for Scc4-dependent localization of cohesin loading. Elife 4, e06057 (2015).

68. Hu, B. et al. ATP hydrolysis is required for relocating cohesin from sites occupied by its Scc2/4 loading complex. Curr Biol 21, 12-24 (2011).

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

70. Beckouet, F. et al. An Smc3 acetylation cycle is essential for establishment of sister chromatid cohesion. Mol Cell 39, 689-99 (2010).

71. Beckouet, F. et al. Releasing Activity Disengages Cohesin's Smc3/Scc1 Interface in a Process Blocked by Acetylation. Mol Cell 61, 563-574 (2016).

72. Camdere, G., Guacci, V., Stricklin, J. & Koshland, D. The ATPases of cohesin interface with regulators to modulate cohesin-mediated DNA tethering. Elife 4(2015).

73. Chan, K.L. et al. Pds5 promotes and protects cohesin acetylation. Proc Natl Acad Sci U S A 110, 13020-5 (2013).

74. Heidinger-Pauli, J.M., Onn, I. & Koshland, D. Genetic evidence that the acetylation of the Smc3p subunit of cohesin modulates its ATP-bound state to promote cohesion establishment in Saccharomyces cerevisiae. Genetics 185, 1249-56 (2010).

75. Rolef Ben-Shahar, T. et al. Eco1-dependent cohesin acetylation during establishment of sister chromatid cohesion. Science 321, 563-6 (2008).

76. Rowland, B.D. et al. Building sister chromatid cohesion: smc3 acetylation counteracts an antiestablishment activity. Mol Cell 33, 763-74 (2009).

77. Sutani, T., Kawaguchi, T., Kanno, R., Itoh, T. & Shirahige, K. Budding yeast Wpl1(Rad61)-Pds5 complex counteracts sister chromatid cohesion-establishing reaction. Curr Biol 19, 492-7 (2009).

78. Tedeschi, A. et al. Wapl is an essential regulator of chromatin structure and chromosome segregation. Nature 501, 564-8 (2013).

79. Terret, M.E., Sherwood, R., Rahman, S., Qin, J. & Jallepalli, P.V. Cohesin acetylation speeds the replication fork. Nature 462, 231-4 (2009).

80. Unal, E. et al. A molecular determinant for the establishment of sister chromatid cohesion. Science 321, 566-9 (2008).

81. Zhang, J. et al. Acetylation of Smc3 by Eco1 is required for S phase sister chromatid cohesion in both human and yeast. Mol Cell 31, 143-51 (2008).

82. Ladurner, R. et al. Sororin actively maintains sister chromatid cohesion. EMBO J 35, 635-53 (2016).

83. Lafont, A.L., Song, J. & Rankin, S. Sororin cooperates with the acetyltransferase Eco2 to ensure DNA replication-dependent sister chromatid cohesion. Proc Natl Acad Sci U S A 107, 20364-9 (2010).

84. Nishiyama, T. et al. Sororin mediates sister chromatid cohesion by antagonizing Wapl. Cell 143, 737-49 (2010).

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

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

87. Schmitz, J., Watrin, E., Lenart, P., Mechtler, K. & Peters, J.M. Sororin is required for stable binding of cohesin to chromatin and for sister chromatid cohesion in interphase. Curr Biol 17, 630-6 (2007).

88. Gandhi, R., Gillespie, P.J. & Hirano, T. Human Wapl is a cohesin-binding protein that promotes sister-chromatid resolution in mitotic prophase. Curr Biol 16, 2406-17 (2006).

89. Camdere, G.O., Carlborg, K.K. & Koshland, D. Intermediate step of cohesin's ATPase cycle allows cohesin to entrap DNA. Proc Natl Acad Sci U S A 115, 9732-9737 (2018).

90. Dixon, J.R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376-80 (2012).

91. Haarhuis, J.H. & Rowland, B.D. Cohesin: building loops, but not compartments. EMBO J 36, 3549-3551 (2017).

92. Nora, E.P. et al. Molecular basis of CTCF binding polarity in genome folding. Nat Commun 11, 5612 (2020).

93. Parelho, V. et al. Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell 132, 422-33 (2008).

94. Rubio, E.D. et al. CTCF physically links cohesin to chromatin. Proc Natl Acad Sci U S A 105, 8309-14 (2008).

95. Stedman, W. et al. Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. EMBO J 27, 654-66 (2008).

96. Vian, L. et al. The Energetics and Physiological Impact of Cohesin Extrusion. Cell 175, 292-294 (2018).

97. Wendt, K.S. et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature 451, 796-801 (2008).

98. Wutz, G. et al. ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesin(STAG1) from WAPL. Elife 9(2020).

99. Wutz, G. et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. EMBO J 36, 3573-3599 (2017).

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

101. Nora, E.P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485, 381-5 (2012).

102. Arumugam, P., Nishino, T., Haering, C.H., Gruber, S. & Nasmyth, K. Cohesin's ATPase activity is stimulated by the C-terminal Winged-Helix domain of its kleisin subunit. Curr Biol 16, 1998-2008 (2006).

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

104. Huber, R.G. et al. Impairing Cohesin Smc1/3 Head Engagement Compensates for the Lack of Eco1 Function. Structure 24, 1991-1999 (2016).

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

106. Locher, K.P. Mechanistic diversity in ATP-binding cassette (ABC) transporters. Nat Struct Mol Biol 23, 487-93 (2016).

107. Lee, B.G. et al. Crystal Structure of the Cohesin Gatekeeper Pds5 and in Complex with Kleisin Scc1. Cell Rep 14, 2108-2115 (2016).

108. Muir, K.W. et al. Structure of the Pds5-Scc1 Complex and Implications for Cohesin Function. Cell Rep 14, 2116-2126 (2016).

109. Kashammer, L. et al. Mechanism of DNA End Sensing and Processing by the Mre11-Rad50 Complex. Mol Cell 76, 382-394 e6 (2019).

110. Guacci, V., Chatterjee, F., Robison, B. & Koshland, D.E. Communication between distinct subunit interfaces of the cohesin complex promotes its topological entrapment of DNA. Elife 8(2019).

111. Murayama, Y., Samora, C.P., Kurokawa, Y., Iwasaki, H. & Uhlmann, F. Establishment of DNA-DNA Interactions by the Cohesin Ring. Cell 172, 465-477 e15 (2018).

112. Diebold, M.L., Fribourg, S., Koch, M., Metzger, T. & Romier, C. Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer. J Struct Biol 175, 178-88 (2011).

113. Fribourg, S. et al. Dissecting the interaction network of multiprotein complexes by pairwise coexpression of subunits in E. coli. J Mol Biol 306, 363-73 (2001).

114. Romier, C. et al. Co-expression of protein complexes in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies. Acta Crystallogr D Biol Crystallogr 62, 1232-42 (2006).

115. Vincentelli, R. & Romier, C. Complex Reconstitution and Characterization by Combining Co-expression Techniques in Escherichia coli with High-Throughput. Adv Exp Med Biol 896, 43-58 (2016).

126. Piñeiro, Á. et al. AFFINImeter: A software to analyze molecular recognition processes from experimental data. Anal biochem 577, 117-134 (2019).

117. Kabsch, W. Xds. Acta Crystallogr D Biol Crystallogr 66, 125-32 (2010).

118. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50, 760-3 (1994).

119. McCoy, A.J. et al. Phaser crystallographic software. J Appl Crystallogr 40, 658-674 (2007).

120. Adams, P.D. et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D: Biological Crystallography 66, 213-221 (2010).

121. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. Acta Crystallographica Section D: Biological Crystallography 66, 486-501 (2010).

122. Williams, C.J. et al. MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci 27, 293-315 (2018).

DISCUSSION OF THE THESIS RESULTS AND PERSPECTIVES

Introduction

Cohesin is a key player into the organization of the genome and the maintenance of its stability. Through its mechanistic ability to tether one or two segments of DNA, as well as its DNA translocase ability, cohesin is involved in many vital cellular processes.

ATP binding and hydrolysis by cohesin is thought to lead to conformational changes, which are closely related to cohesin physiological functions. However, the precise molecular details of how human cohesin SMC1A and SMC3 ATPases bind and hydrolyze ATP, and of the underlying structural changes that support cohesin functions remained to be fully characterized.

Moreover, mutation-driven cohesin dysfunctions can lead to the emergence and development of a wide range of human malignancies and developmental disorders, including various types of cancers and cohesinopathies.

During my thesis work, I sought to characterize human cohesin SMC1A and SMC3 ATPase heads, by the means of biochemical, biophysical, and structural methods. The results show distinct conformational dynamics of SMC1A as compared to SMC3, supporting the molecular bases for human cohesin ATPase asymmetry.

The P-loop in controlling cohesin ATP binding and hydrolysis dynamics

Firstly, ATP binding to SMC1A ATPase active site seems to be possibly modulated by the SMC1A P-loop, which can adopt two distinct conformations: a closed conformation, in which the P-loop is stabilized by a highly ordered network of water molecules, and an open conformation,

generated by concomitant ATP binding and desolvation of the water at the binding site. Since SMC1A P-loop does not appear to participate in any crystallographic contact inside the crystal packing, its dual conformation appears to have functional purposes. It is most likely that SMC1A open and closed conformations have direct implications for nucleotide binding dynamics to the SMC1A ATPAse, such as modulating ATP binding and ADP release from the active site. An interesting observation is that in many conditions where SMC1A (CC and CCSh) was crystallized in presence of ATPγS or ADP, diffracting crystals yielded structures of SMC1A in an apo state with the P-loop in a closed conformation, whereas for SMC3, the diffracting crystals grown in presence of ATPγS or ADP always yielded structures with the nucleotides bound to an open P-loop. This general observation was made upon the solution of the structures from around a hundred diffracting crystals.

Crystals for SMC3 in absence of any nucleotide were however extremely hard to obtain. The only apo crystals that diffracted yielded a structure with a sulfate ion, from the crystallization solution, bound into the active site at the place that would be occupied by the beta phosphate of ATP. Since we do not have a strict apo SMC3 structure, we cannot say whether the P-loop of SMC3 can also adopt a closed conformation. However, an interesting observation makes it seem very likely that the SMC3 P-loop can only adopt one open conformation, precast to bind ATP, thus differing from that of SMC1A. If we compare SMC1A and SMC3 active sites, we can see that in SMC3 the bulky phenylalanine 1191 is positioned right above the serine 36 of the closed SMC3 P-loop and could be hindering its mobility. However, in SMC1A this residue is replaced by proline 1205, which could give more room for the P-loop to be mobile and free to adopt both close and open conformation. It would be interesting to analyze the SMC3 F1191P mutant, to see if F1191 is maintaining SMC3 P-loop in a permanent closed conformation.

Conformational changes upon ATP binding and hydrolysis by cohesin

The results also reveal significant conformational changes upon ATP binding to SMC1A ATPase head, as previously observed for other ABC proteins. SMC1A ATPase conformations range from a nucleotide-free relaxed state to a tightened productive state, which goes back to a relaxed state after ATP hydrolysis and ADP release. SMC1A ATPase head flexibility stems from a hinge-like and rotational movement that takes place between two distinct domains, the globular nucleotide-binding domain and the helical region from where stem the coiled-coils. Upon ATP-Mg^{2+} binding, the SMC1A Q-loop orchestrates the movement of the two domains compared to each other: as the conserved glutamine of the Q-loop reaches out to coordinate the ATP-Mg^{2+} bound to the active site, it brings the two domains together thus resulting in a closed/tightened SMC1A ATPase conformation which is suitable to dimerization with SMC3. This is in accordance with the observations made from the ITC experiments, where the strong desolvation of SMC1A upon ATP-Mg^{2+} addition suggests that the SMC1A ATPase is overall quite dynamic and significant conformational changes occur upon ATP-Mg binding, as compared to SMC3 ATPase which is less flexible.

On the contrary to SMC1A, SMC3 Q-loop does not reach out to the bound ATP-Mg^{2+}, unless SMC3 is engaged with SMC1A, NIPBL and DNA, as seen in the cohesin productive state structure solved by Shi et al., 2020. The unproductive configuration of the SMC3 Q-loop seems to be directed by a conformation of SMC3 exit gate, which was until then uncharacterized. In this conformation, which we call SMC3 resting state, the RAD21 globular domain contacts the globular domain of SMC3. When SMC3 is engaged in the productive state with SMC1A, NIPBL and DNA, both NIPBL and DNA favor the suppression of the resting state. Structural movements at the SMC3 signature-coupling helix then allow the SMC3 Q-loop to reach out to the bound ATP-Mg^{2+} in SMC3 active site, thus activating its ATPase. It can be thus hypothesized that the human SMC3 ATPase resting state could have a primordial role in the regulation of cohesin ATPase activity, thus participating in the modulation of cohesin interactions with chromatin such as the

prevention of futile ATPase cycles in absence of NIPBL or when cohesin is stably bound to chromatin.

Lastly, those observations provide the molecular explanations to how NIPBL and DNA enhance the cohesin ATPase activity, by showing that binding of NIPBL and DNA induce a lever movement and a coiled-coil stabilization in both SMC1A and SMC3 structures, which leads to their productive state.

Altogether these conformational differences between SMC1A and SMC3 explain structurally and mechanistically why the SMC1A-SMC3 composite ATPase alone has a low ATPase activity, which is greatly enhanced in presence of NIPBL and DNA. It would be interesting to use *in vivo* crosslinking experiments between SMC3 and the globular domain of RAD21N to see whether the resting state of cohesin occurs in the cellular context, and if it coincides with specific cell cycle events that require cohesin to be stabilized onto chromatin, such as sister chromatid cohesion and loop extrusion arrest.

Role of cohesin R-loops on the ATPase activity

Both SMC1A and SMC3 possess conserved R-loops that have been proposed to participate in sensing and binding the incoming DNA molecule, and transmit the information to the ATP hydrolysis sites of SMC complexes, thus leading to a DNA-driven activation of the ATPase (Lammens et al. 2004). Human SMC1A and SMC3 R-loops adopt structurally distinct conformations. Human SMC1A R-loop adopts an organized conformation that positions the conserved arginine 57 (R57) towards the bound nucleotide, which according to my results seem to be required for the proper ATP hydrolysis by SMC1A. However, the R-loop of SMC3 is directed in the opposite direction of the ATP binding site, and the conserved R57 reaches out to bind the DNA molecule that is positioned on top of the cohesin ATPase module. In my new solved crystallographic structures, both SMC1A and SMC3 R-loops keep the same conformation whether the complex is in an apo state or bound to ATP or ADP alone. This raises the important question

whether the human SMC3 R-loop is also involved in ATP binding and hydrolysis by cohesin, and whether SMC1A R-loop is involved in DNA sensing and binding in addition to participating to ATP binding and hydrolysis.

Preliminary experiments of ATPase assays performed in presence of DNA showed that SMC1A R57 does not seem to directly participate in the DNA-driven ATPase activity, in contrary to archaeal SMC (Lammens et al., 2004). However, the deletion of an adjacent loop containing the residues V58 to R62 seems to have a direct impact in an SMC1A ATPase activity driven by DNA binding.

This could be explained by the fact that the residues K59 and R62, shown to be directed towards DNA by Shi et al., 2020, could be ultimately the DNA sensing and binding residues of SMC1A, which then transmit this information to the ATPase active site through other adjacent residues. This is particularly interesting given that de deletion of the V58 to R62 residues has been found in cases of CdLS. In addition, while it strongly impairs the activity of the SMC1A homodimer, it is interesting to note that the V58-R62 deletion does not seem to have a strong deleterious effect on the SMC1A-SMC3 heterodimer ATPase activity. Interestingly, this observation (1) suggests that SMC3 could possibly have a compensatory effect on the deleterious effect of this mutation, and (2) suggests why the V58-R62 deletion is involved in a mild case of CdLS. This mutation should however be characterized more thoroughly using the entire complex and in the presence of cohesin regulators, such as NIPBL, including in the cellular context.

More detailed research is needed to understand the role of cohesin R-loops and their interactions with the DNA molecule, and which undoubtly are key elements in the molecular mechanism of cohesin ATP binding and hydrolysis cycle. For example, individual mutation of the putative DNA-binding residues instead of the full deletion of the V58-R62 segment, followed by DNA binding assays and ATPase activities of these mutants in presence and in absence of DNA and of cohesin regulators should bring more insight into their mechanistic role.

Position of SMC3 regulatory lysines at positions 105 and 105 and implications for ESCO binding and ESCO-dependent acetylation

It was previously shown by Ladurner et al., 2014 that cohesin ATPase activity was required in the coupling of cohesin loading onto DNA with SMC3 acetylation, and that only wild type cohesin was acetylated. This study also states that cohesin would associate with DNA, and subsequent ATP hydrolysis would lead to entrapment of the DNA molecule and convert SMC3 into a form that can be acetylated by ESCO acetyltransferases. Moreover, Ajam et al., 2020 showed that cohesin was optimally acetylated only in the presence of ATP and DNA, and Borges et al., 2010 showed that *in vivo* only chromatin-bound cohesin is acetylated, but not the free soluble pool.

Upon solving the human SMC3-RAD21N structures, I was curious to see if in the ADP-bound SMC3 ATPase structures we would observe a difference in the SMC3 K105 and K106 conformation, which could promote their acetylation. Surprisingly, in all crystallographic structures, as well as the cryo-EM structure from Shi et al. 2020, no difference in the position of both lysines was identified, apart from their shielding by NIPBL in the latter case. This question then arises: which mechanism of SMC3, which could be driven by the ATPase or by DNA, leads to ESCO binding and SMC3 acetylation? Answering this question is quite challenging. However, hypotheses can be made from a structural model, built by alignment of known vertebrate ESCO1 and ESCO2 bound to an SMC3 peptide containing both lysines (Figure 38A) that aligns well with the corresponding segment of the SMC3 structure. It shows that there could be a potential steric hindrance between a loop of ESCO neighboring its active site and the SMC3 R-loop (Figure 38B).
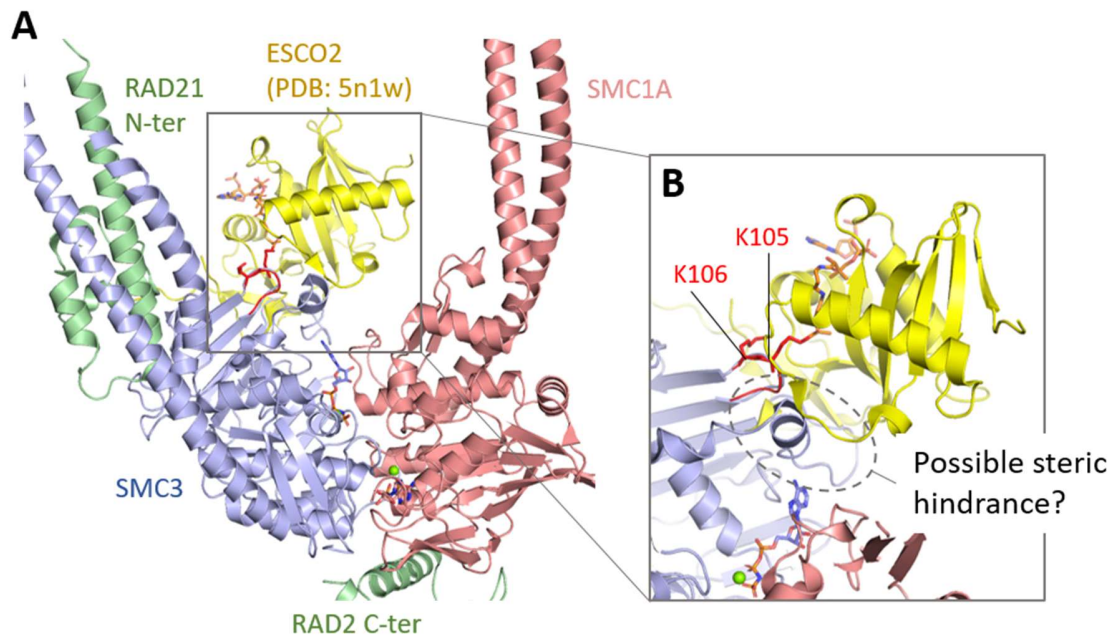
Figure 38: Molecular model of possible ESCO binding to SMC3. (A) The *X. laevis* ESCO2 structure bound to the K105-K106 peptide (PDB: 5n1w) was aligned to the same region in the human SMC3 productive state (PDB: 6wg3). (B) Close up view, showing a potential steric hindrance of SMC3 loops with ESCO loops.

Looking at this in the context of the NIPBL- and DNA-bound productive state of cohesin, and considering the facts that DNA is required for acetylation and that acetylation leads to DNA entrapment, the following hypothesis can be made. Concomitantly to the hydrolysis of ATP by both SMC1A and SMC3, the DNA molecule passes between both ATPase heads, as they separate after hydrolysis, in order to be entrapped. On its way to the other side, the DNA molecule, which we know contacts SMC3 R-loop, could be bringing the R-loop down towards SMC3 ATP-binding site (Figure 39). This would create an intermediate conformation of SMC3, in which DNA would push the R-loop downwards in a way that it would allow ESCO binding to SMC3 to acetylate both K105 and K106. To test this hypothesis, it could be interesting to perform acetylation assays on an SMC3 R-loop mutant, where some residues of its R-loop would be removed or replaced by less bulky residues, to see whether this facilitates SMC3 acetylation by ESCO in the absence of ATP and DNA.

Controlling SMC3 acetylation by the action of DNA would make sure that only the active cohesin on chromatin (about to entrap sister chromatids or engaged in loop extrusion) is acetylated, and not the free pool of cohesin. Although completely hypothetical, this could be a mechanistic explanation for Ladurner et al., 2014 and Ajam et al., 2020 observations on the ATPase coupling SMC3 acetylation with cohesin loading onto DNA, and the positive effect of DNA on SMC3 acetylation.
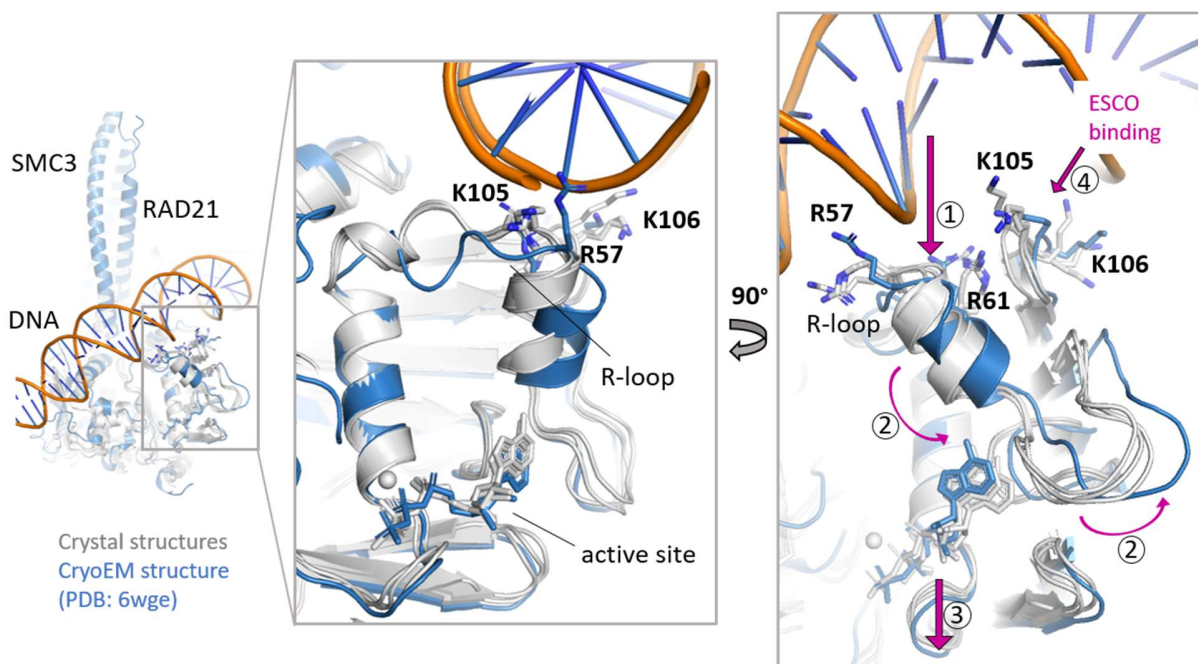
Figure 39: Potential effect of DNA on the nucleotide binding by the SMC3 ATPase. The comparison between the crystal structures (gray) and the DNA-bound cryo-EM structure (blue; PDB: 6wge) of the human SMC3 ATPase domain show that the DNA molecule could potentially have an effect on the SMC3 R-loop dynamics, and participate in the regulation of nucleotide binding and release from the SMC3 ATPase. The DNA molecule could be promoting ADP release after hydrolysis, by remodelling the active site through the following mechanism (pink arrows): (1) Following the ATP hydrolysis, DNA brings the R-loop down towards SMC3 ATP-binding site; (2) Remodelling of the R-loop allows (3) ADP release. While the DNA is moving downwards to pass through the SMC1A and SMC3 ATPas heads, it (4) makes room for ESCO to bind and acetylate K105 and K106.

How is ADP released after hydrolysis? Is it driven by an active mechanism?

The analysis of ATP binding to human SMC1A and SMC3 ATPases show that both bind ATP with a similar albeit low affinity. Surprisingly, ADP binds to SMC1A with a similar affinity as ATP, while it binds to SMC3 with at least a 10-fold higher affinity. In addition, ATPase activity assays show that the basal ATPase activity of SMC1A is higher than that of SMC3. These interesting results, especially for SMC3, could be explained in two non-mutually exclusive ways. Firstly, the SMC3 resting state, which keeps the SMC3 Q-loop in unproductive conformation, could be responsible for the low basal ATPase activity of SMC3, thus being responsible for a low ADP-ATP turnover by SMC3 ATPase, in the absence of NIPBL and DNA. But since the ADP affinity to SMC3 is still unexpectedly higher than that of ATP, the second explanation implies that ADP release at the end of one ATPase cycle could require an active mechanism, such as a DNA-driven ADP release. It has been proposed that ATP binding, hydrolysis and ADP release by ABC ATPases might be driven by binding of their substrates. Therefore, it is tempting to view cohesin as a potential "DNA-transporter", in and out of the cohesin ring for example, where DNA would act as the substrate. It would not be excluded in this case that the DNA would have direct actions on cohesin ATPase.

One experiment to test this hypothesis could be observing the ADP release by SMC3 in absence and in presence of DNA, through competition/nucleotide turnover experiments using fluorescence anisotropy. To do so, SMC3 would be firstly incubated with fluorescently labelled ADP. Then, unlabeled ADP or ATP would be added, in absence or in presence of DNA, and the fluorescence polarization by the labelled nucleotide would be measured. A fast fluorescence anisotropy decay in presence of DNA as compared with no added DNA would indicate that the DNA molecule could potentially have a role in promoting the release of ADP from cohesin ATPase, thus also promoting nucleotide turnover. This would also explain why the basal activity of both cohesin ATPases alone or together can be slightly enhanced by DNA (Figure 16, Davidson et al., 2019), albeit it is much less significant than in the presence of NIPBL. In this case, even if the DNA

could have a local impact on the cohesin ATPase activity, DNA alone is probably not sufficient to bring the heads spatially close, like NIPBL does, nor to induce the lever effect and stabilization of SMC3 and SMC1A coiled-coils into their full productive state that significantly promotes ATP hydrolysis.

Structurally, if we look again at the previous model of DNA-promoted SMC3 acetylation, the potential reorganization of the SMC3 R-loop by DNA towards the SMC3 active site could potentially promote ATP hydrolysis and subsequent ADP release. In this scenario, the SMC3 active site obstruction by ADP and the control of ADP release by DNA could help prevent futile ADP hydrolysis and help cohesin stabilization on chromatin, which is sometimes needed for long periods of time in the case of sister chromatid cohesion.

Intermediate states of the cohesin ATPase cycle

In an attempt to capture the intermediate steps of cohesin ATPase cycle, I additionally performed crystallization trials in presence of various ATP analogs: $ADP-BeF_3$, mimicking a pre-hydrolysis state, $ADP-AlF_3$ for a hydrolysis intermediate state, and $ADP-VO_4$ for a post-hydrolysis state. Several crystals were obtained in these conditions, among which some of them diffracted X-rays. Only apo (for SMC1A) ADP-bound (for SMC3) complexes could be detected upon the solution of most structures. However, one structure of SMC3-RAD21N in presence of $ADP-BeF_3$ was solved in the same space group and with the same unit cell parameters as one of the SMC3-RAD21N structures bound to ATPγS. Unfortunately, only the ADP was detected in the structure and not the BeF_3 moiety, most probably due to poor crystal diffraction stemming from a poor crystal cryoprotection.

Solving the structures of cohesin ATPases in presence of ATP analogs should be pursued, especially in presence of cohesin regulators and DNA. Indeed, it has been shown that the $ADP-P_i$ intermediate in the cycle of the cohesin ATPase, entrapped by $ADP-AlF_3$, allows the stable DNA

216

entrapment by cohesin (Camdere et al., 2018). The underlying molecular mechanisms of this entrapment remains to be fully characterized.

Further studies of disease mutants

The crystallographic structures of human SMC1A and SMC3 ATPase heads in their pre- and post-hydrolysis states, together with the cryo-EM structure in the productive state (Shi et al., 2020), reveal the molecular environment of mutations on the cohesin ATPase that lead to human disease. Many of these mutations are located on functionally relevant regions of the cohesin ATPase (Figure 12), such as on the DNA exit gate, near the ATP binding site, on the R-loops, and at the interface of NIPBL binding. Visualizing these mutations in their structural context in the various conformations of cohesin ATPase allows to make hypotheses on their possible deleterious effects on cohesin structure and mechanisms. To test these hypotheses, the mutant SMC1A and SMC3 should be further characterized *in vitro* (thermal stability, ATPase activity, ATP/DNA/cohesin subunits binding), following by analyses in the cellular context and *in vivo*.

It will be important to identify which mutations are deleterious to the cohesin structure, the enzymatic activity, or the binding of molecular partners. Notably, mutations that locally affect protein folding and its three-dimensional structure could possibly be rescued by using pharmacological chaperones. Those small molecules are designed to specifically assist a protein in its proper folding, so it is not discarded by the quality control of the cell, and it can join the pool of active protein. This method has been successfully used to treat other rare enzymatic diseases such as the Fabry disease, and has a huge advantage in that those small molecules can easily cross the blood-brain barrier, thus being suitable for the use in treating neurological impairments as it is often the case for cohesinopathies.

The solved crystallographic structures thus provide the structural bases that were until now missing and that will contribute to the structure-based drug design of new targeted therapies towards dysfunctional cohesin.

Proposed molecular model for the cohesin ATPase heads mechanisms

The crystal structures of the human SMC1A and SMC3 ATPases in their apo, ATPγS- and ADP-bound forms presented here are higly complementary to the cryo-EM structure of human cohesin bound to DNA and to its regulators NIPBL and STAG1 (referred to as the productive state). Together, these structures provide an insight into the ATP binding and hydrolysis cycle of human cohesin.

The model described here proposes that ATP binding to the SMC1A ATPase induces a hinge-like rotational conformation change, of the helical lobe and the RecA lobe towards each other. These structural changes in presence of ATP lead to a closed conformation of the SMC1A ATPase, which would be prone to dimerization with the SMC3 ATPase that is itself already in a dimerization-prone conformation. However, as shown by the reported low ATPase activity of the human SMC1A-SMC3 ATPases heterodimer, and as possibly explained by the novel observed resting state of the SMC3 coiled coil, the SMC1A-SMC3 ATPase heterodimer needs further regulation to be fully activated. It has been shown that the cohesin regulator NIPBL, together with the DNA molecule, actively enhances the cohesin ATPase activity. Apart from knowing that NIPBL brings the heads together in spatial proximity and possibly promote their dimerization, the exact molecular mechanism of this activation remains unknown.

The presented model (Figure 40) thus suggests that upon binding to the cohesin heterodimeric ATPase module, NIPBL induces a lever effect in the SMC1A ATPase, which leads to a fully activated closed conformation. More importantly, NIPBL could be suppressing the inhibitory effect of the SMC3 resting state on the SMC3 ATPase. By binding to SMC3 coiled-coils as shown in the productive state, it would be bringing them in a straightened conformation that is compatible with a lever effect, which allows the SMC3 Q-loop to reach the ATP-Mg bound to the SMC3 active site and thereby fully activate the cohesin ATPase. This could be especially important when the cohesin complex is being recruited and loaded onto chromatin by NIPBL and MAU2.
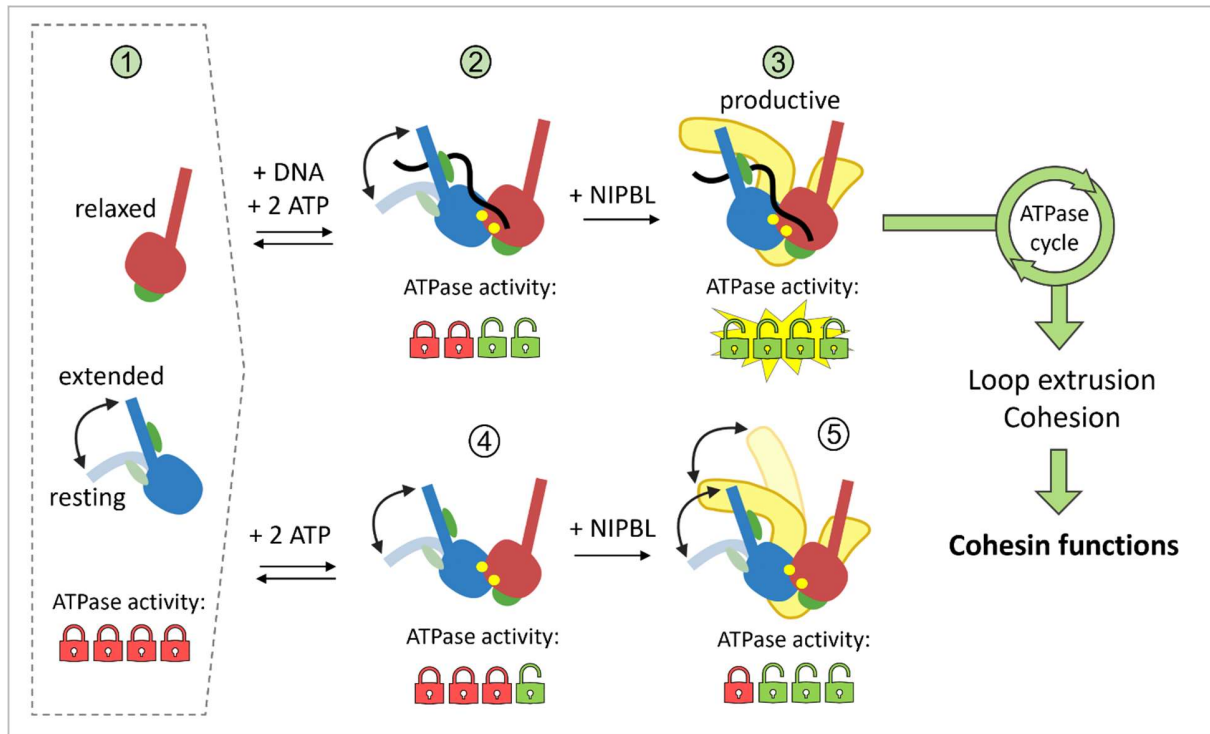
SMC1A ATPase, SMC3 ATPase, RAD21 N- and C-ter, ATP, NIPBL, DNA

(n) Stepwise activation of the Cohesin ATPase by ATP, DNA and NIPBL

Figure 40: Proposed mechanism of the cohesin ATPase activation by NIPBL and DNA. (1) Both SMC1A and SMC3 ATPases are inactive when not engaged in the presence of ATP; (2) Upon capture of a DNA segment and in the presence of ATP, both ATPase heads engage while the DNA molecule sits on top of the V-shaped heterodimer. Electrostatic interactions between the DNA and the RAD21N helps position the SMC3 coiled coil in an extended conformation; (3) NIPBL binding drives a lever effect on both SMC1A and SMC3 ATPases, and keeps the SMC3 coiled-coil in a productive conformation, leading to the full activated conformation of the cohesin ATPase that was solved using cryo-EM by Shi et al., 2020. (2), (4), (5) Alternative conformations of the complex could still provide a residual ATP hydrolysis activity, which is slightly enhanced in the presence of DNA or NIPBL alone, as described in in vitro experiments by Kim et al., 2019 and Davidson et al., 2020.

The further enhancing effect of DNA on the ATPase in the presence of NIPBL could be possibly due to the fact that the DNA molecule contacts the SMC1A and SMC3 R-loops. At least for SMC1A, the R-loop is directly linked to the ATP bound to the active site through the conserved R57, and could thus be directly transmitting the DNA binding event to the cohesin ATPase for its activation. The role of DNA binding to the SMC3 R-loop remains less clear, even if it can be

hypothesized that the SMC3 R-loop could participate in coupling DNA binding to SMC3 ATPase activation, or in the regulation of SMC3 acetylation on the conserved neighboring lysine pair.

While they provide an insight into the ATP binding and activation of cohesin ATPase, the solved crystallographic structures also allow to make alternative hypotheses on the mechanisms underlying the ATP hydrolysis, ADP release and the subsequent heads disengagement. Interestingly, if we align the helical domains of the SMC1ACC crystallographic structures, wathever the nucleotide binding state, to the helical domain of SMC1A in the SMC3-engaged productive state, we see that the helical domains can align well, without major structural changes towards the upper and central regions (Figure 41A). However, at the base of the helical domain, significant structural changes seem to occur at the loop including the signature motif and the conserved phenylalanine residue F1122, which could furthermore correspond to a state of SMC1 after hydrolysis of ATP at the SMC3 active site, leading to dissociation of the SMC1A signature motif from SMC3 and its repositioning into a specific "non-SMC3 engaged" state. Indeed, in the crystallographic SMC1A ATPase structures, the signature motif and its preceeding loop are displaced upwards by 2 to 4 Å as compared to the productive state of SMC1 (Figure 41B). This movement is enough to break the hydrogen bond formed by the SMC1A R1121 with the NIPBL Q2311, and to displace the conserved SMC1A F1122. The displacement of SMC1A F1122, which participates in SMC1A binding to NIPBL through hydrophobic interactions, creates a steric hindrance between this residue with the residues of NIPBL P2316 and H2315, thus potentially weakening the association of SMC1A with NIPBL at this region (Figure 41B). The conserved SMC1A F1122 is part of the SMC1A F-loop and could potentially be of significant relevance for SMC1A interaction with NIPBL, as suggested for the same conserved loop in the case of the condensin complex (Hassler et al., 2019).

If we look further to the structural changes that occur at this region of SMC1A, we see that the movement of the signature motif/F-loop has a significant impact on the adjacent helices, and, ultimately, on the whole SMC1A structure. Upward movement of the signature motif/F-loop leads to the creation of a hydrogen bond between the main chain -NH group of M1125 and the main chain CO of Y1113, ast their distance decreases from 6 to 2.8 Å (Figure 42A). By doing so,

M1125 creates a steric hindrance with A144, which is positioned on the signature coupling helix adjacent to the Q-loop. This causes A144 and the loop where it stands to further move upwards. This movement affects the residues at the C-terminal end of the loop with A144: the N147 residue, located at 3.9 Å from R152, comes into hydrogen bonding distance of R152 at 2.9 Å in the SMC1A ADP-bound form. This small movement has drastic consequences on the conformation of the adjacent signature coupling helix and on the Q-loop, which it rotates backwards, away from its initial position from 8 Å (Figure 42B). This movement occurring in the helical region leads to the rotation of the whole RecA-lobe, away from the signature motif of SMC3, thus allowing the wide opening of the SMC1A ATPase. This opening most likely leads to the hydration of SMC1A active site, and to subsequent changes in coordination of the bound Mg ion and ADP, as seen in the SMC1ACCsh ADP-Mg-bound structure (Figure 43A-B), which would further promote ADP-Mg release from SMC1A active site for the catalytic turnover (Figure 44), and the positioning of the SMC1A P-loop in its closed conformation by a new water network after ADP-Mg release (Figure 43C). The wide opening of SMC1A ATPase could most likely contribute to the DNA passage through the ATPase upon SMC1A and SMC3 disengagement.

It is also a possibility that these described mechanisms for SMC1A happen in the opposite way: following ATP hydrolysis at the SMC1A site into ADP and P_i, the electrostatic repulsion of the negatively charged P_i could promote the SMC1A-SMC3 dissociation at the SMC1A site. While the SMC1A rotates backwards, the cascade of events would lead to the upward movement of the SMC1A signature/F-loop. This would cause opening at the SMC3 site, and possible destabilization of the SMC1A-NIPBL interaction. It is also not excluded that these events could happen concomitantly, and lead to the simultaneous opening at the SMC1A and SMC3 sites after hydrolysis.

For the SMC3 ATPase, although a small rotational movement is seen between the different nucleotide-bound states, it is probable that the resting state of SMC3 could have a greater impact on its ATPase cycle: to open the SMC3 side of the ATPase, the coiled-coil of SMC3 could be pushing away NIPBL, and concomitantly disengage from DNA by rotating away the RAD21 positive patch. An alternative scenario would be that, while maintaining its coil bound to

NIPBL, SMC3 globular region could rotate away from SMC1A, thus displacing NIPBL and the DNA molecule towards the wide opening created between SMC1A and SMC3 ATPase heads, in order for DNA to pass through the heads (Figure 45). I the case of DNA loop extrusion, for example, once the DNA is displaced from the SMC1A and SMC3 ATPase heads, both heads could grab a further DNA segment and bring it towards the site where the loop is being extruded, thus enlarging the loop.
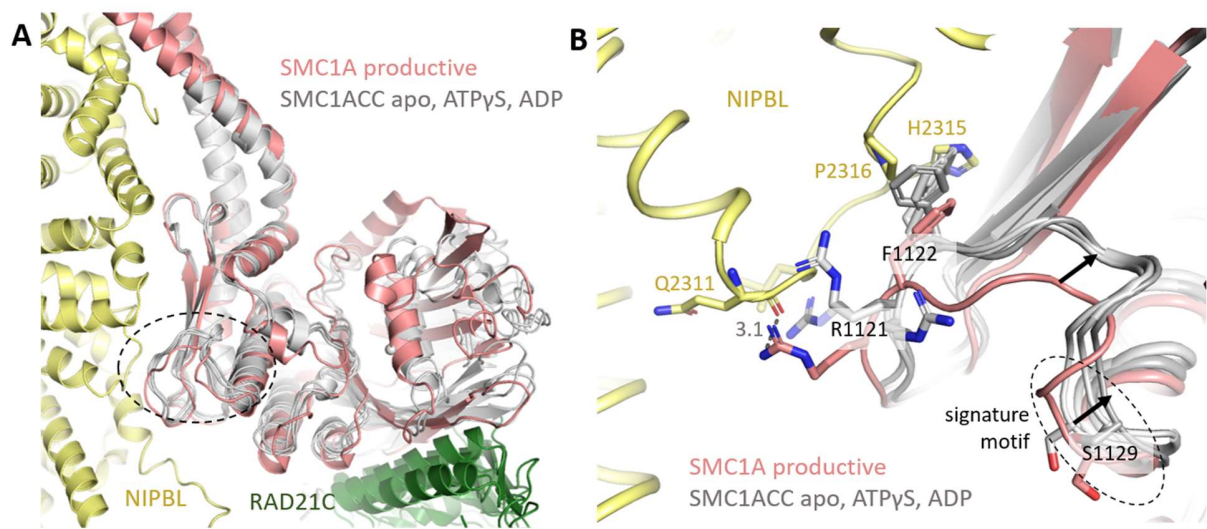

Figure 41: Structural changes that occur at the signature moti/F-loop region of SMC1A when in a non-SMC3-engaged state. (A) Structure comparison between SMC1ACC apo, ATPγS- and ADP-bound (in gray) and SMC1A in the productive state (in salmon). (B) Close-up view of the signature motif and F-loop movements (black arrows), and possible steric hindrance with NIPBL in the non-SMC3-engaged state.
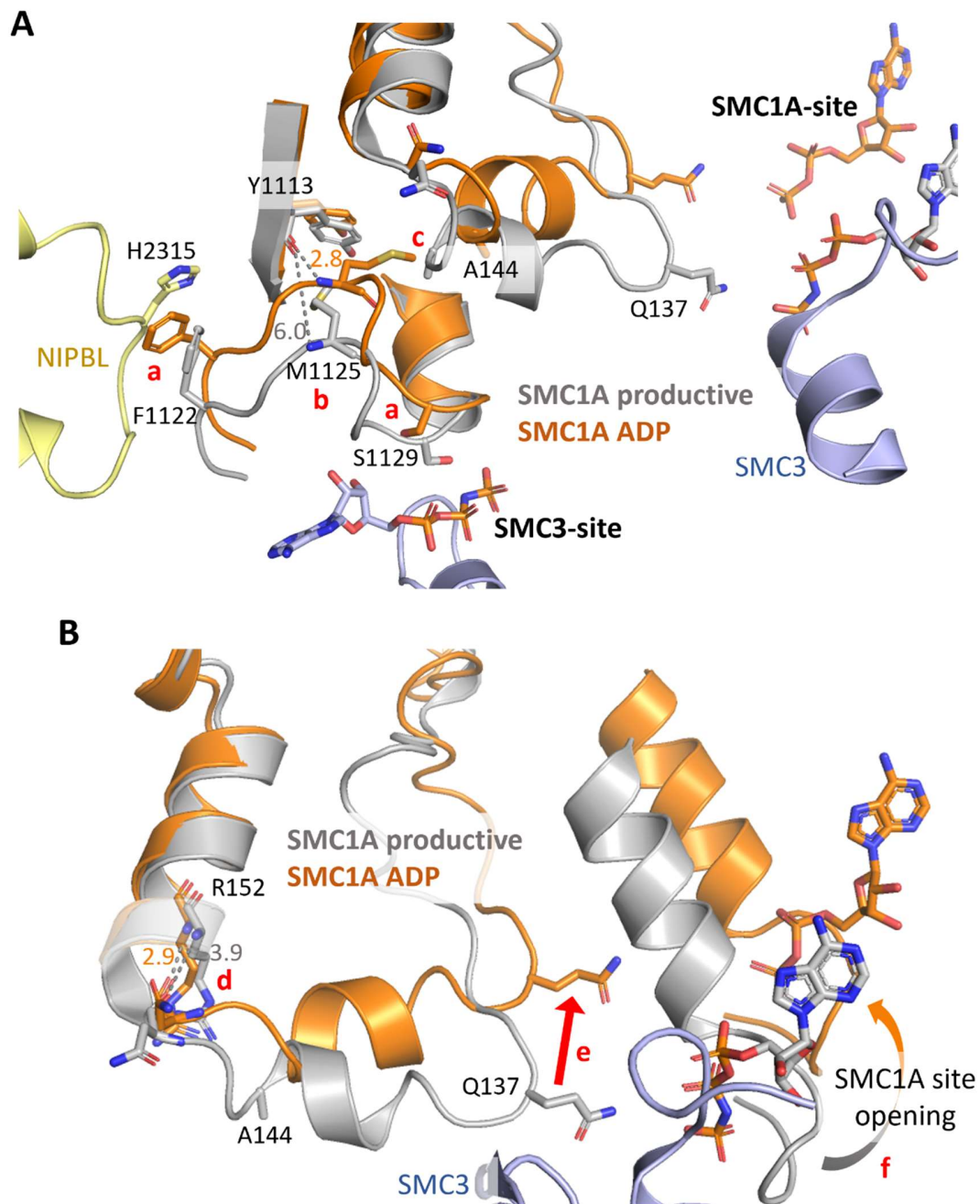
Figure 42: Structural movements possibly occurring at the SMC1A ATPase after ATP hydrolysis. (A) (a) The upward displacement of the SMC1A signature motif after ATP hydrolysis at the SMC3 site leads to (b) hydrogen bonds rearrangement of neighboring residues and (c) to upward displacement of the signature coupling helix adjacent to the Q-loop. (B) These movements further (d) induce hydrogen bonding rearrangements at the signature coupling helix, which ultimately allows (e) the significant displacement of the Q-loop and the whole SMC1A Rec-A domain, thus (f) opening the SMC1A site. This cascade of events could possibly instead occur in the reverse order: (f), (e), (d), (c), (b), (a). Distances shown in Å.
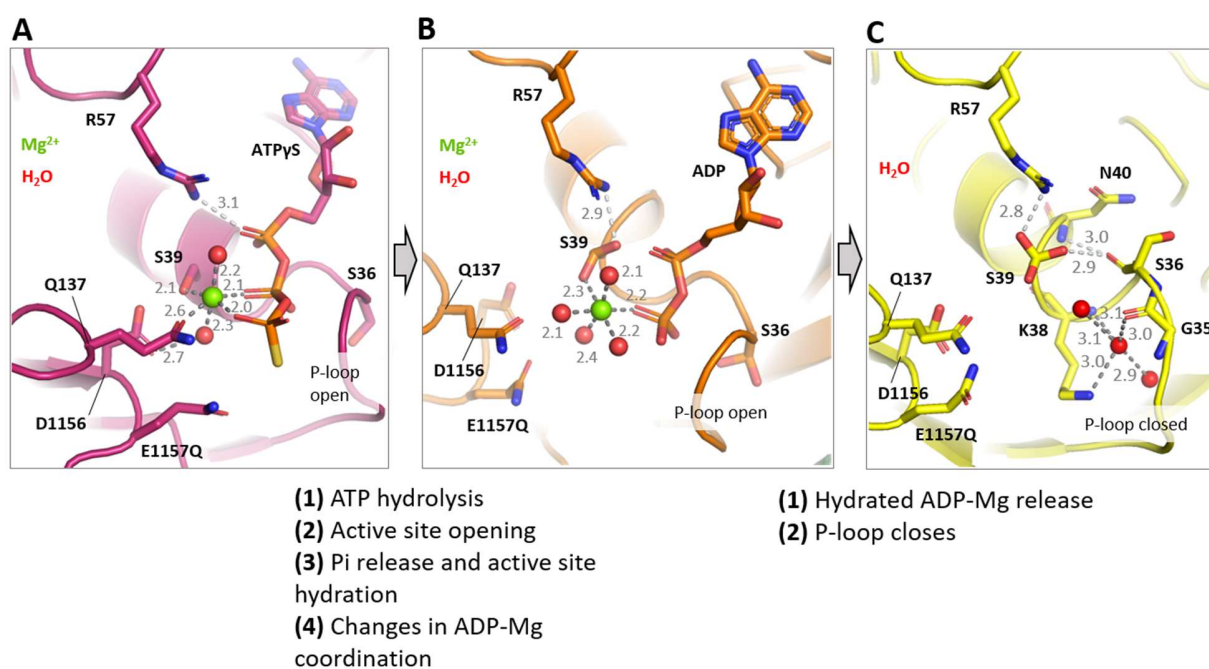
Figure 43: Close-up view of the active site of SMC1ACC in its (A) ATPγS-bound, (B) ADP-Mg-bound and (C) apo structures, and proposed steps for the nucleotide turnover. Changes in the SMC1A active site hydration state might control ATP binding, ADP release and the open/closed conformations of the P-loop.
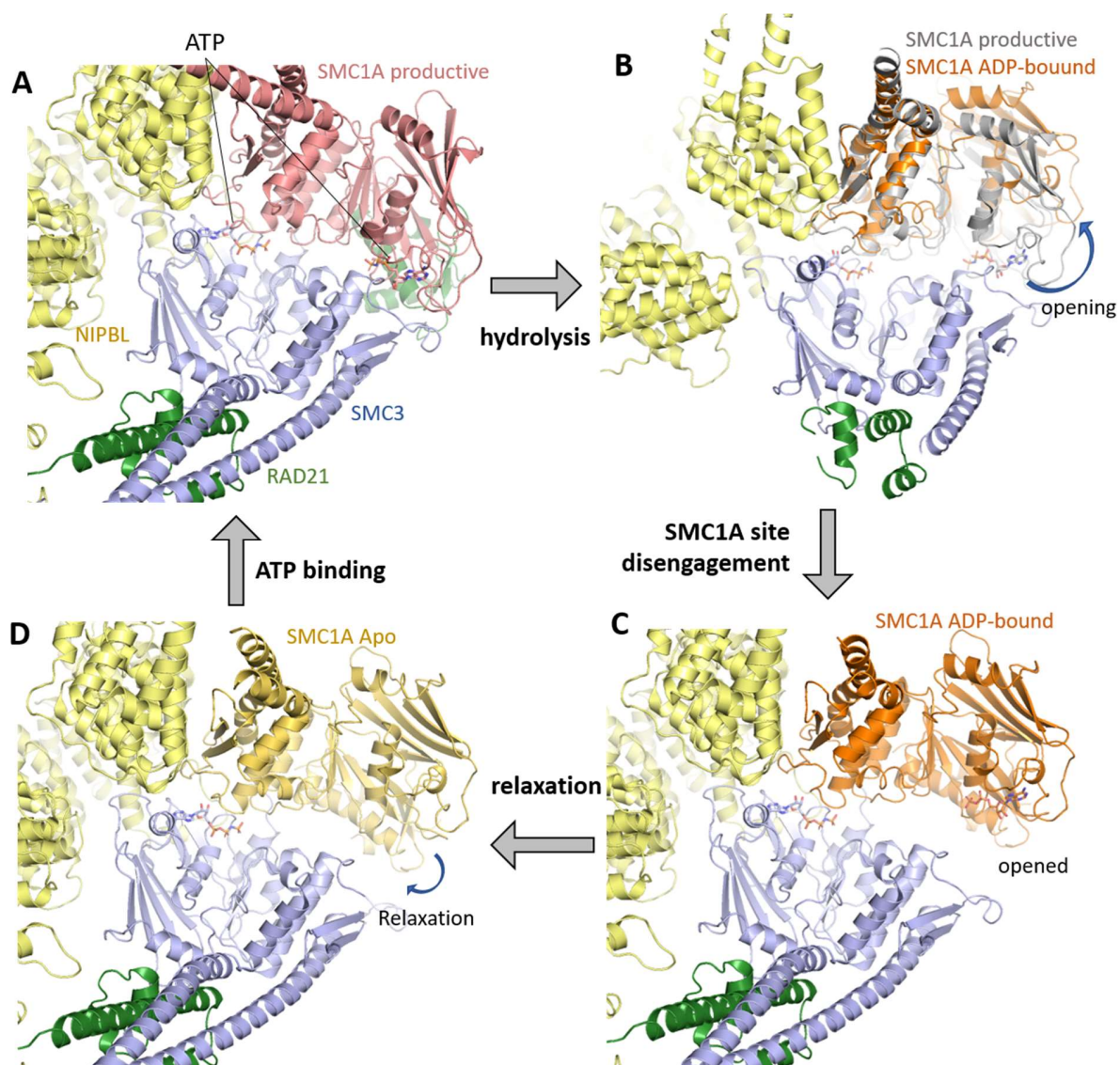
Figure 44: Possible ATPase mechanism at the SMC1A active site. (A) The productive engaged SMC1A and SMC3 hydrolyse ATP. (B-C) The dissociation of the SMC1A signature motif from SMC3 induces a cascade of structural changes, which ultimately promotes the wide opening of SMC1A active site, the dissociation of SMC1A active site from SMC3 and the hydration of the SMC1A active site. (D) After ADP release, SMC1A recovers a relaxed apo state, but slightly tighter that the ADP-bound structure, to allow the binding of a new ATP molecule, the SMC1A-SMC3 ATPase dimerization and the restart of the hydrolysis cycle.
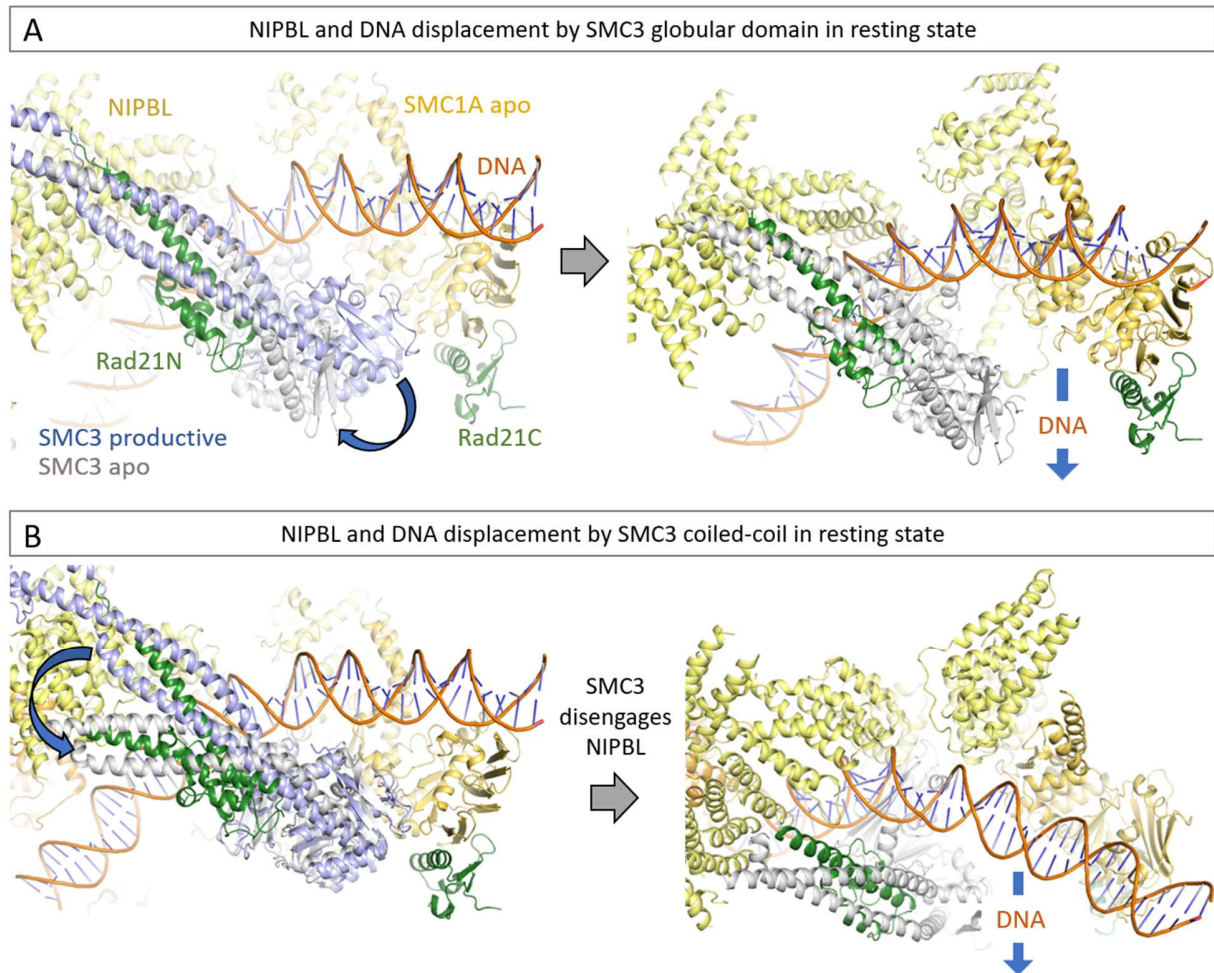
Figure 45: Possible mechanistic action of the SMC3 ATPase during the ATPase cycle. The disengagement of SMC1A and SMC3 and the wide opening of the interface between them to allow the passage of DNA through the heads, in order to be entrapped into the Rad21 compartment, could either happen by (A) the movement of the SMC3 globular region into the resting state, or (B) the movement of SMC3 coiled coils. Since SMC1A in its ADP-bound conformation is opended more widely than in its apo conformation, its ADP conformation could be the state of the ATPase which best favours the passage of DNA through the cohesin ATPase heads.

CONCLUSION

Despite huge progress in cohesin research in the past few years and the increasing emergence of new functional models, many questions remain to be answered concerning the structural and enzymatic mechanisms and the interaction with other factors, in order to validate the current models.

Together, the results obtained throughout my thesis work provide a new insight into ATP binding and hydrolysis by human cohesin, including with the obtention of the first structures of ADP-bound eukaryotic cohesin. The results show human SMC1A and SMC3 ATPases features which seem to play important roles in the molecular mechanism of ATP binding and hydrolysis by the human cohesin ATPase module, and highlights structural and conformational differences of SMC1A and SMC3 ATPases which can contribute to an asymmetric cohesin ATPase. The obtained crystallographic structures allow the identification of structural changes occurring during the ATP hydrolysis cycle of cohesin, and to highlight specific regions around which these changes occur.

Moreover, despite evolutionary conservation of some fundamental characteristics and key residues, cohesin developed throughout evolution new regulatory pathways at the cohesin ATPase. Notably, the novel DNA exit gate conformation observed in the solved human cohesin structures could have distinct functional consequences in humans and perhaps more largely in vertebrates, considering the fact that the DNA exit gate mechanism possess additional regulators and a specific prophase pathway in these organisms.

Comparisons of the SMR1 and SMR3 structures with existing structures from other organisms highlighted major structural differences that can explain the unique features of human and possibly vertebrate cohesin. The structural results are highly complementary to the recently released cryo-EM structure of human cohesin complex bound to DNA and its loading factor (Shi et al., 2020), and will help the cohesin research community in understanding the connection

227

between the human cohesin ATPase dynamics to the structural rearrangements which ultimately leads to cohesin functions.

Altogether, my thesis work brings new pieces to the puzzle whose progressive elucidation will help in understanding the mechanisms at the core of cohesin functions, and provides new structural bases for further development of structure-based therapies targeting dysfunctional cohesin in cancers and in cohesinopathies.

Ultimately, the more questions we answer on the mechanisms underlying cohesin functions, the more challenging the next questions that arise from the discoveries will become. Exciting challenges that lie ahead will be to finally elucidate how the structural changes induced by ATP binding and hydrolysis are coupled to and promote DNA transactions by cohesin: how does cohesin entrap one, then two sister chromatids? What precise mechanisms enable cohesin to extrude loops of DNA? What are the consequences of ATP binding and hydrolysis by the cohesin ATPase on the dynamics of regulatory subunits binding? Are there differences in the cohesin ATP binding properties between lower and higher eukaryotes? The use of complementary integrated structural biology methods, the observation of cohesin in the cellular and chromatin context, and importantly, the comparative approaches between different species will be pivotal to help precisely answer these exciting questions.

CONCLUSION EN FRANÇAIS

Des progrès considérables ont été accomplis ces dernières années dans le champ de la recherche sur la cohésine, et de nombreux nouveaux modèles ont émergé sur le mode de fonctionnement du complexe et de l'interaction avec ses régulateurs. Néanmoins, de nombreuses questions demeurent encore sans réponse, et nécessitent de poursuivre les efforts de recherche qui permettront de vérifier les modèles proposés.

Pendant mon travail de thèse, j'ai eu pour objectif de caractériser les domaines ATPase de SMC1A et de SMC3 de la cohésine humaine. L'ensemble de mes résultats fournit un nouvel aperçu des mécanismes de liaison et d'hydrolyse de l'ATP par les domaines ATPase de SMC1A et de SMC3 et sur les conséquences structurales de ces évènements, notamment sur l'asymétrie du domaine ATPase. Notamment, mes résultats présentent les premières structures résolues de domaines ATPase de la cohésine eucaryote liées à l'ADP. De plus, les résultats ont également révélé une nouvelle conformation de la porte de sortie qu'emprunte l'ADN pour être libéré du complexe pendant la mitose cellulaire.

Dans leur ensemble, les résultats obtenus pendant mon travail de thèse sont très complémentaires à la structure obtenue récemment par cryo-EM par Shi et al., 2020, du module ATPase de la cohésine humaine, révélant l'hétérodimère formé par SMC1A et SMC3 en présence d'un analogue de l'ATP, de NIPBL, de STAG1 et d'un fragment d'ADN. La comparaison des structures que j'ai résolues par cristallographie et de la structure cryo-EM résolue par Shi et al., 2020 permettent d'émettre des hypothèses et de proposer un modèle sur le mode d'activation du module ATPase de la cohésine humaine lors de la fixation de NIPBL et de l'ADN (Figure 40).

De plus, ces structures, analysées ensemble, vont permettre à la communauté de recherche de mieux comprendre les connexions entre la dynamique conformationnelle des domaines ATPase de SMC1A et de SMC3, l'interaction avec des régulateurs, et les fonctions de la cohésine humaine. Cela permettra également d'émettre des hypothèses sur les dysfonctionnements de la cohésine qui conduisent à l'émergence et au développement de

maladies, dont le cancer et les cohésinopathies, et de mettre en place de nouvelles stratégies thérapeutiques ciblant la cohésine dysfonctionnelle.

En conclusion, plus nous apportons de nouveaux éléments de réponse aux questionnements sur les modes de fonctionnement et les mécanismes de la cohésine, plus les nouvelles questions qui vont émerger de ces découvertes seront complexes et d'autant plus excitantes à investiguer. Par exemple, précisément comment est-ce que la cohésine utilise son domaine ATPase dans l'extrusion des boucles d'ADN, et dans l'établissement de la cohésion des chromatides sœurs ? Quelles sont les conséquences de la liaison et de l'hydrolyse de l'ATP sur la dynamique de liaison des régulateurs de la cohésine et de l'ADN ? Y a-t-il des différences entre les propriétés de liaison et d'hydrolyse de l'ATP par la cohésine entre les eucaryotes inférieurs, comme la levure, et supérieurs, comme l'humain ? L'utilisation de méthodes complémentaires de biologie structurale intégrative, l'observation du complexe cohésine dans le contexte cellulaire et, aussi important, les approches expérimentales comparatives entre différentes espèces seront cruciales pour aider à répondre précisément à ces questions de recherche particulièrement excitantes.
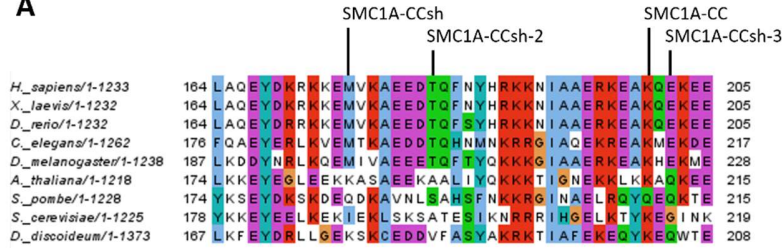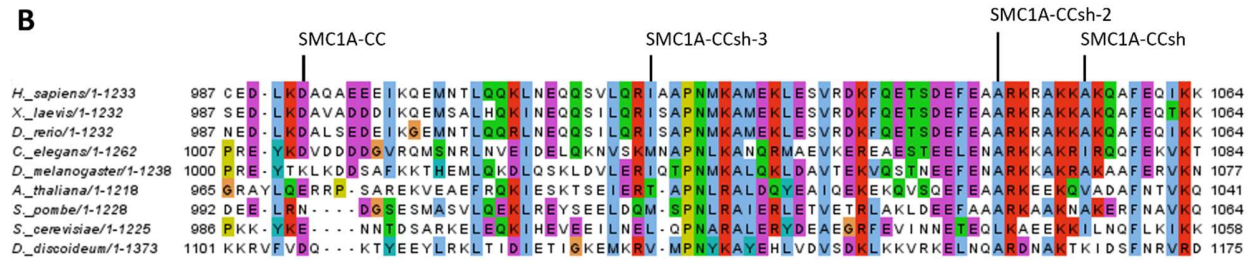
Appendix 1: Sequence alignment of the SMC1A from different organisms. The sequence boundaries of the human SMC1A ATPase head constructs are indicated, at (A) their N-terminal end and at (B) their C-terminal end.
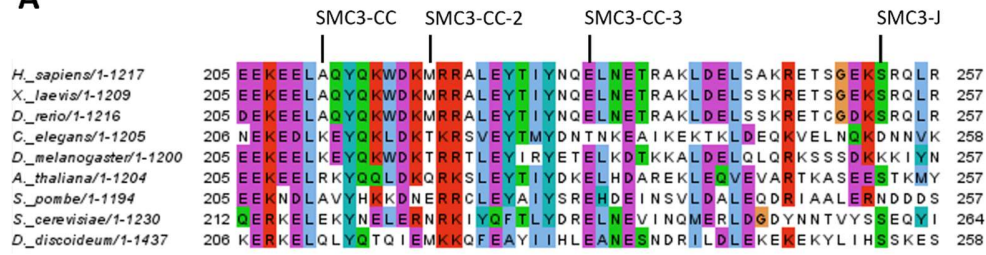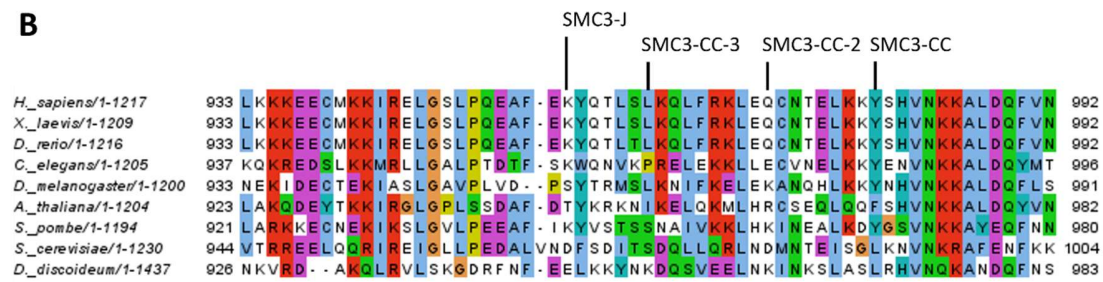
Appendix 2: Sequence alignment of the SMC3 from different organisms. The sequence boundaries of the human SMC3 ATPase head constructs are indicated, at (A) their N-terminal end and at (B) their C-terminal end.

SMC1A-CC-EQ/RAD21C
Apo

SMC1A-CC-EQ/RAD21C
ATPγS-Mg

SMC1A-CC-EQ/RAD21C
ADP

SMC1A-CC-EQ/RAD21C
Apo ''loop''

SMC1A-CCsh-EQ/RAD21C
Apo

SMC1A-CCsh-EQ/RAD21C
Apo – SO$_4$
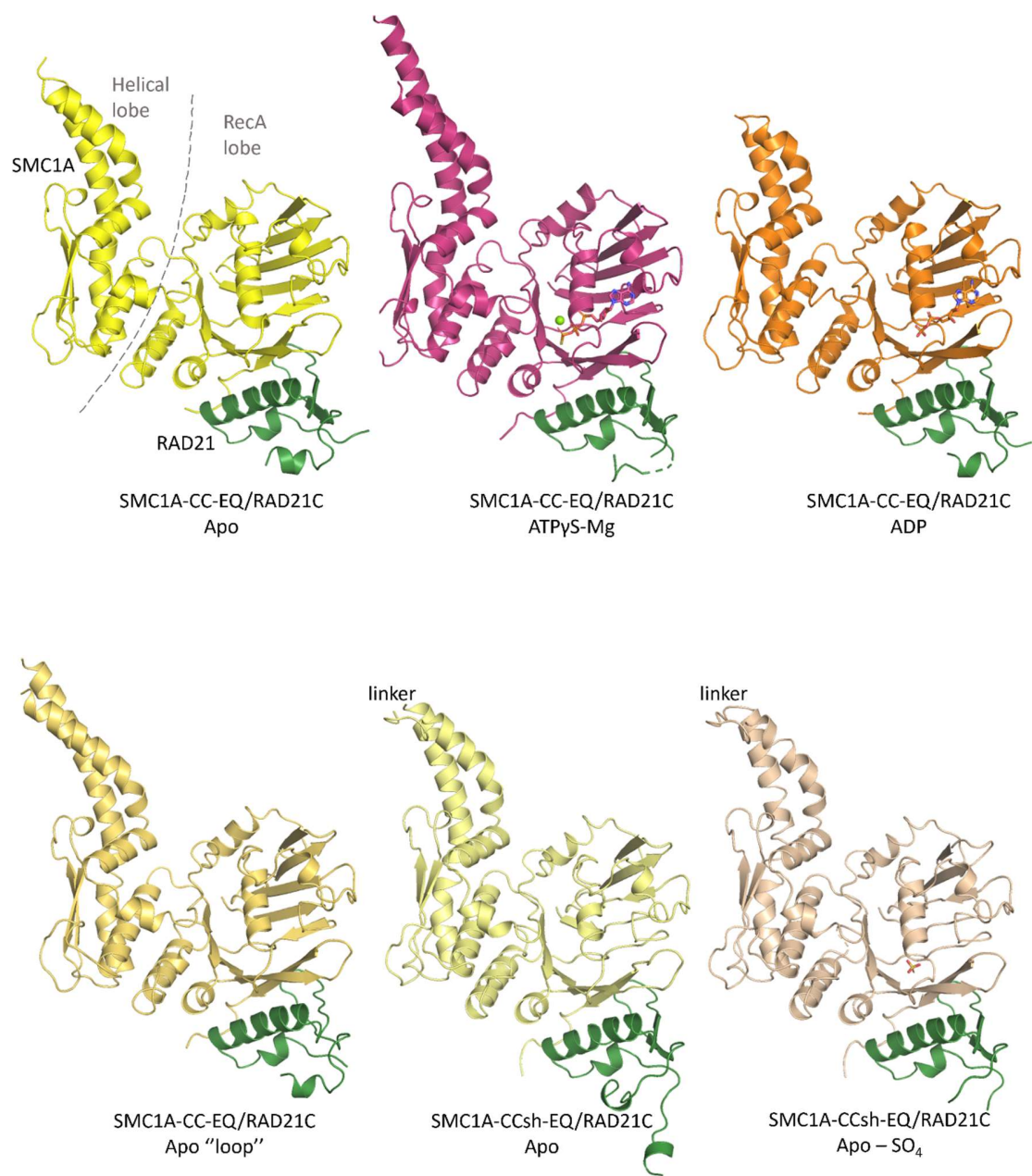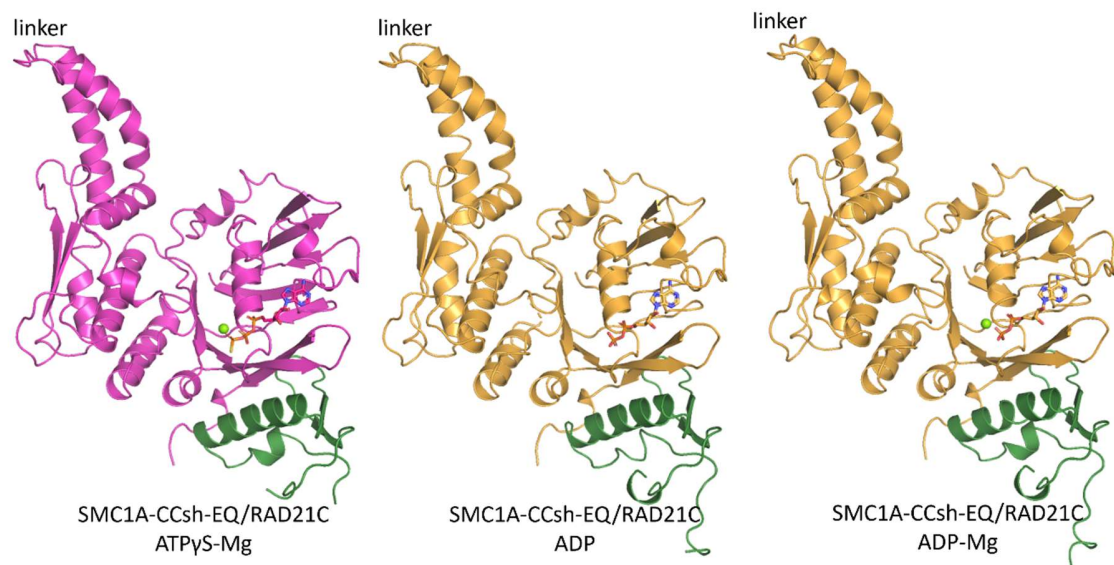
Appendix 3: Overall structures of the human cohesin SMC1A ATPase head, solved by X-ray crystallography. The constructs used and the ligand bound to the active site are indicated below each structure.
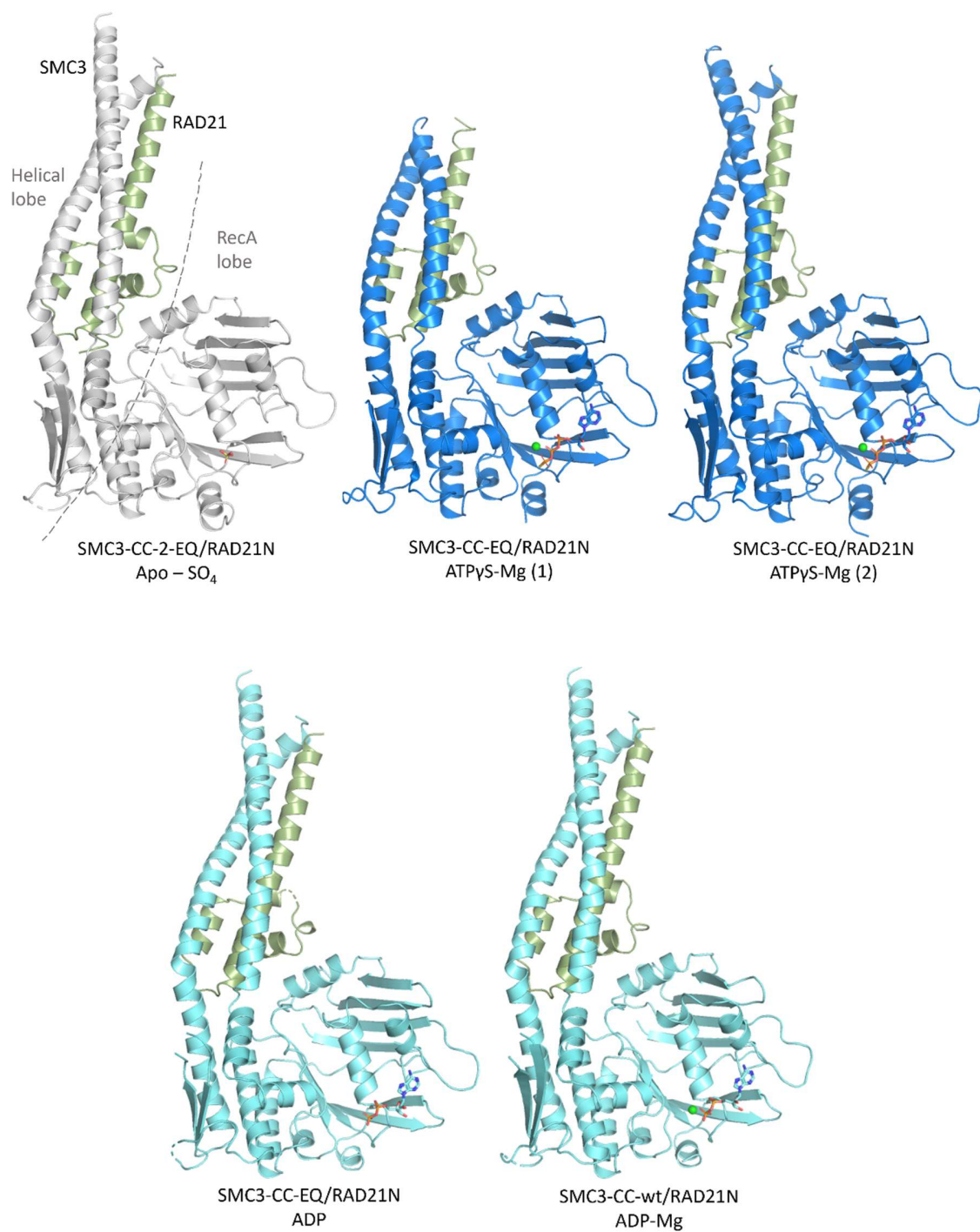
Appendix 4: Overall structures of the human cohesin SMC3 ATPase head, solved by X-ray crystallography. The constructs used and the ligand bound to the active site are indicated below each structure.

LIST OF FIGURES AND TABLES

REFERENCES

Figures were made using the PyMOL Molecular Graphics System, Version 2.3 Schrödinger, LLC., and BioRender.com.

Ajam, T., De, I., Petkau, N., Whelan, G., Pena, V., Eichele, G., 2020. Alternative catalytic residues in the active site of Esco acetyltransferases. Sci Rep 10, 9828. https://doi.org/10.1038/s41598-020-66795-z

Alomer, R.M., da Silva, E.M.L., Chen, J., Piekarz, K.M., McDonald, K., Sansam, C.G., Sansam, C.L., Rankin, S., 2017. Esco1 and Esco2 regulate distinct cohesin functions during cell cycle progression. Proc Natl Acad Sci USA 114, 9906–9911. https://doi.org/10.1073/pnas.1708291114

Anderson, D.E., Losada, A., Erickson, H.P., Hirano, T., 2002. Condensin and cohesin display different arm conformations with characteristic hinge angles. Journal of Cell Biology 156, 419–424. https://doi.org/10.1083/jcb.200111002

Aragón, L., 2018. The Smc5/6 Complex: New and Old Functions of the Enigmatic Long-Distance Relative. Annu. Rev. Genet. 52, 89–107. https://doi.org/10.1146/annurev-genet-120417-031353

Arnould, C., Rocher, V., Finoux, A.-L., Clouaire, T., Li, K., Zhou, F., Caron, P., Mangeot, Philippe.E., Ricci, E.P., Mourad, R., Haber, J.E., Noordermeer, D., Legube, G., 2021. Loop extrusion as a mechanism for formation of DNA damage repair foci. Nature 590, 660–665. https://doi.org/10.1038/s41586-021-03193-z

Arruda, N.L., Carico, Z.M., Justice, M., Liu, Y.F., Zhou, J., Stefan, H.C., Dowen, J.M., 2020. Distinct and overlapping roles of STAG1 and STAG2 in cohesin localization and gene expression in embryonic stem cells. Epigenetics & Chromatin 13, 32. https://doi.org/10.1186/s13072-020-00353-9

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

Ba, Z., Lou, J., Ye, A.Y., Dai, H.-Q., Dring, E.W., Lin, S.G., Jain, S., Kyritsis, N., Kieffer-Kwon, K.-R., Casellas, R., Alt, F.W., 2020. CTCF orchestrates long-range cohesin-driven V(D)J recombinational scanning. Nature 586, 305–310. https://doi.org/10.1038/s41586-020-2578-0

Bauer, B.W., Davidson, I.F., Canena, D., Wutz, G., Tang, W., Litos, G., Horn, S., Hinterdorfer, P., Peters, J.-M., 2021. Cohesin mediates DNA loop extrusion by a "swing and clamp" mechanism. Cell 184, 5448-5464.e22. https://doi.org/10.1016/j.cell.2021.09.016

Bauerschmidt, C., Woodcock, M., Stevens, D.L., Hill, M.A., Rothkamm, K., Helleday, T., 2011. Cohesin phosphorylation and mobility of SMC1 at ionizing radiation-induced DNA double-strand breaks in human cells. Experimental Cell Research 317, 330–337. https://doi.org/10.1016/j.yexcr.2010.10.021

Baxter, J., Oliver, A.W., Schalbetter, S.A., 2019. Are SMC Complexes Loop Extruding Factors? Linking Theory With Fact. BioEssays 41, 1800182. https://doi.org/10.1002/bies.201800182

Beckouët, F., Srinivasan, M., Roig, M.B., Chan, K.-L., Scheinost, J.C., Batty, P., Hu, B., Petela, N., Gligoris, T., Smith, A.C., Strmecki, L., Rowland, B.D., Nasmyth, K., 2016. Releasing Activity Disengages Cohesin's Smc3/Scc1 Interface in a Process Blocked by Acetylation. Molecular Cell 61, 563–574. https://doi.org/10.1016/j.molcel.2016.01.026

Bermudez, V.P., Farina, A., Higashi, T.L., Du, F., Tappin, I., Takahashi, T.S., Hurwitz, J., 2012. In vitro loading of human cohesin on DNA by the human Scc2-Scc4 loader complex. Proceedings of the National Academy of Sciences 109, 9366–9371. https://doi.org/10.1073/pnas.1206840109

Birkenbihl, R.P., Subramani, S., 1992. Cloning and characterization of *rad21* an essential gene of *Schizosaccharomyces pombe* involved in DNA double-strand-break repair. Nucl Acids Res 20, 6605–6611. https://doi.org/10.1093/nar/20.24.6605

Bonev, B., Cavalli, G., 2016. Organization and function of the 3D genome. Nat Rev Genet 17, 661–678. https://doi.org/10.1038/nrg.2016.112

Borges, V., Lehane, C., Lopez-Serra, L., Flynn, H., Skehel, M., Ben-Shahar, T.R., Uhlmann, F., 2010. Hos1 Deacetylates Smc3 to Close the Cohesin Acetylation Cycle. Molecular Cell 39, 677–688. https://doi.org/10.1016/j.molcel.2010.08.009

Bot, C., Pfeiffer, A., Giordano, F., Edara, D.M., Dantuma, N.P., Ström, L., 2017. Independent mechanisms recruit the cohesin loader protein NIPBL to sites of DNA damage. Journal of Cell Science jcs.197236. https://doi.org/10.1242/jcs.197236

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

Bürmann, F., Lee, B.-G., Than, T., Sinn, L., O'Reilly, F.J., Yatskevich, S., Rappsilber, J., Hu, B., Nasmyth, K., Löwe, J., 2019. A folded conformation of MukBEF and cohesin. Nat Struct Mol Biol 26, 227–236. https://doi.org/10.1038/s41594-019-0196-z

Bürmann, F., Shin, H.-C., Basquin, J., Soh, Y.-M., Giménez-Oya, V., Kim, Y.-G., Oh, B.-H., Gruber, S., 2013. An asymmetric SMC–kleisin bridge in prokaryotic condensin. Nat Struct Mol Biol 20, 371–379. https://doi.org/10.1038/nsmb.2488

Çamdere, G., Guacci, V., Stricklin, J., Koshland, D., 2015. The ATPases of cohesin interface with regulators to modulate cohesin-mediated DNA tethering. eLife 4, e11315. https://doi.org/10.7554/eLife.11315

Canudas, S., Houghtaling, B.R., Kim, J.Y., Dynek, J.N., Chang, W.G., Smith, S., 2007. Protein requirements for sister telomere association in human cells. EMBO J 26, 4867–4878. https://doi.org/10.1038/sj.emboj.7601903

Canudas, S., Smith, S., 2009. Differential regulation of telomere and centromere cohesion by the Scc3 homologues SA1 and SA2, respectively, in human cells. Journal of Cell Biology 187, 165–173. https://doi.org/10.1083/jcb.200903096

Carretero, M., Ruiz-Torres, M., Rodríguez-Corsino, M., Barthelemy, I., Losada, A., 2013. Pds5B is required for cohesion establishment and Aurora B accumulation at centromeres: Pds5B is required for cohesion establishment and Aurora B accumulation. The EMBO Journal 32, 2938–2949. https://doi.org/10.1038/emboj.2013.230

Chan, K.-L., Gligoris, T., Upcher, W., Kato, Y., Shirahige, K., Nasmyth, K., Beckouet, F., 2013. Pds5 promotes and protects cohesin acetylation. Proceedings of the National Academy of Sciences 110, 13020–13025. https://doi.org/10.1073/pnas.1306900110

Chan, K.-L., Roig, M.B., Hu, B., Beckouët, F., Metson, J., Nasmyth, K., 2012. Cohesin's DNA Exit Gate Is Distinct from Its Entrance Gate and Is Regulated by Acetylation. Cell 150, 961–974. https://doi.org/10.1016/j.cell.2012.07.028

Chao, W.C.H., Wade, B.O., Bouchoux, C., Jones, A.W., Purkiss, A.G., Federico, S., O'Reilly, N., Snijders, A.P., Uhlmann, F., Singleton, M.R., 2017. Structural Basis of Eco1-Mediated Cohesin Acetylation. Sci Rep 7, 44313. https://doi.org/10.1038/srep44313

Chatterjee, A., Zakian, S., Hu, X.-W., Singleton, M.R., 2013. Structural insights into the regulation of cohesion establishment by Wpl1. EMBO J 32, 677–687. https://doi.org/10.1038/emboj.2013.16

Chen, X., Zaro, J.L., Shen, W.-C., 2013. Fusion protein linkers: Property, design and functionality. Advanced Drug Delivery Reviews 65, 1357–1369. https://doi.org/10.1016/j.addr.2012.09.039

Cheng, J.-M., Liu, Y.-X., 2017. Age-Related Loss of Cohesion: Causes and Effects. IJMS 18, 1578. https://doi.org/10.3390/ijms18071578

Ciosk, R., Shirayama, M., Shevchenko, Anna, Tanaka, T., Toth, A., Shevchenko, Andrej, Nasmyth, K., 2000. Cohesin's Binding to Chromosomes Depends on a Separate Complex Consisting of Scc2 and Scc4 Proteins. Molecular Cell 5, 243–254. https://doi.org/10.1016/S1097-2765(00)80420-7

Cobbe, N., Heck, M.M.S., 2004. The Evolution of SMC Proteins: Phylogenetic Analysis and Structural Implications. Molecular Biology and Evolution 21, 332–347. https://doi.org/10.1093/molbev/msh023

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

Couturier, A.M., Fleury, H., Patenaude, A.-M., Bentley, V.L., Rodrigue, A., Coulombe, Y., Niraj, J., Pauty, J., Berman, J.N., Dellaire, G., Di Noia, J.M., Mes-Masson, A.-M., Masson, J.-Y., 2016. Roles for APRIN (PDS5B) in homologous recombination and in ovarian cancer prediction. Nucleic Acids Res 44, 10879–10897. https://doi.org/10.1093/nar/gkw921

Crow, E.W., Crow, J.F., 2002. 100 Years Ago: Walter Sutton and the Chromosome Theory of Heredity. Genetics 160, 1–4. https://doi.org/10.1093/genetics/160.1.1

Cuadrado, A., Giménez-Llorente, D., Kojic, A., Rodríguez-Corsino, M., Cuartero, Y., Martín-Serrano, G., Gómez-López, G., Marti-Renom, M.A., Losada, A., 2019. Specific Contributions of Cohesin-SA1 and Cohesin-SA2 to TADs and Polycomb Domains in Embryonic Stem Cells. Cell Reports 27, 3500-3510.e4. https://doi.org/10.1016/j.celrep.2019.05.078

Cuadrado, A., Losada, A., 2020. Specialized functions of cohesins STAG1 and STAG2 in 3D genome architecture. Current Opinion in Genetics & Development 61, 9–16. https://doi.org/10.1016/j.gde.2020.02.024

Dahm, R., 2005. Friedrich Miescher and the discovery of DNA. Developmental Biology 278, 274–288. https://doi.org/10.1016/j.ydbio.2004.11.028

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

Davidson, I.F., Bauer, B., Goetz, D., Tang, W., Wutz, G., Peters, J.-M., 2019. DNA loop extrusion by human cohesin. Science 366, 1338–1345. https://doi.org/10.1126/science.aaz3418

Davidson, I.F., Peters, J.-M., 2021. Genome folding through loop extrusion by SMC complexes. Nat Rev Mol Cell Biol. https://doi.org/10.1038/s41580-021-00349-7

Deardorff, M.A., Wilde, J.J., Albrecht, M., Dickinson, E., Tennstedt, S., Braunholz, D., Mönnich, M., Yan, Y., Xu, W., Gil-Rodríguez, M.C., Clark, D., Hakonarson, H., Halbach, S., Michelis, L.D., Rampuria, A., Rossier, E., Spranger, S., Van Maldergem, L., Lynch, S.A., Gillessen-Kaesbach, G., Lüdecke, H.-J., Ramsay, R.G., McKay, M.J., Krantz, I.D., Xu, H., Horsfield, J.A., Kaiser, F.J., 2012. RAD21 Mutations Cause a Human Cohesinopathy. The American Journal of Human Genetics 90, 1014–1027. https://doi.org/10.1016/j.ajhg.2012.04.019

Dekker, J., Heard, E., 2015. Structural and functional diversity of Topologically Associating Domains. FEBS Letters 589, 2877–2884. https://doi.org/10.1016/j.febslet.2015.08.044

Dheur, S., Saupe, S.J., Genier, S., Vazquez, S., Javerzat, J.-P., 2011. Role for Cohesin in the Formation of a Heterochromatic Domain at Fission Yeast Subtelomeres. Mol Cell Biol 31, 1088–1097. https://doi.org/10.1128/MCB.01290-10

Diebold-Durand, M.-L., Lee, H., Ruiz Avila, L.B., Noh, H., Shin, H.-C., Im, H., Bock, F.P., Bürmann, F., Durand, A., Basfeld, A., Ham, S., Basquin, J., Oh, B.-H., Gruber, S., 2017. Structure of Full-Length SMC and Rearrangements Required for Chromosome Organization. Molecular Cell 67, 334-347.e5. https://doi.org/10.1016/j.molcel.2017.06.010

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., Ren, B., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380. https://doi.org/10.1038/nature11082

Dorsett, D., 2007. Roles of the sister chromatid cohesion apparatus in gene expression, development, and human syndromes. Chromosoma 116, 1–13. https://doi.org/10.1007/s00412-006-0072-6

Dorsett, D., Merkenschlager, M., 2013. Cohesin at active genes: a unifying theme for cohesin and gene expression from model organisms to humans. Current Opinion in Cell Biology 25, 327–333. https://doi.org/10.1016/j.ceb.2013.02.003

Dorsett, D., Ström, L., 2012. The Ancient and Evolving Roles of Cohesin in Gene Expression and DNA Repair. Current Biology 22, R240–R250. https://doi.org/10.1016/j.cub.2012.02.046

Doshi, R., van Veen, H.W., 2013. Substrate Binding Stabilizes a Pre-translocation Intermediate in the ATP-binding Cassette Transport Protein MsbA. Journal of Biological Chemistry 288, 21638–21647. https://doi.org/10.1074/jbc.M113.485714

Eichinger, C.S., Kurze, A., Oliveira, R.A., Nasmyth, K., 2013. Disengaging the Smc3/kleisin interface releases cohesin from Drosophila chromosomes during interphase and mitosis. EMBO J 32, 656–665. https://doi.org/10.1038/emboj.2012.346

Elbatsh, A.M.O., Haarhuis, J.H.I., Petela, N., Chapard, C., Fish, A., Celie, P.H., Stadnik, M., Ristic, D., Wyman, C., Medema, R.H., Nasmyth, K., Rowland, B.D., 2016. Cohesin Releases DNA through Asymmetric ATPase-Driven Ring Opening. Molecular Cell 61, 575–588. https://doi.org/10.1016/j.molcel.2016.01.025

Eser, U., Chandler-Brown, D., Ay, F., Straight, A.F., Duan, Z., Noble, W.S., Skotheim, J.M., 2017. Form and function of topologically associating genomic domains in budding yeast. Proc Natl Acad Sci USA 114, E3061–E3070. https://doi.org/10.1073/pnas.1612256114

Falk, M., Feodorova, Y., Naumova, N., Imakaev, M., Lajoie, B.R., Leonhardt, H., Joffe, B., Dekker, J., Fudenberg, G., Solovei, I., Mirny, L.A., 2019. Heterochromatin drives compartmentalization of inverted and conventional nuclei. Nature 570, 395–399. https://doi.org/10.1038/s41586-019-1275-3

Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J., Lobanenkov, V.V., 1996. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol Cell Biol 16, 2802–2813. https://doi.org/10.1128/MCB.16.6.2802

Gallego-Paez, L.M., Tanaka, H., Bando, M., Takahashi, M., Nozaki, N., Nakato, R., Shirahige, K., Hirota, T., 2014. Smc5/6-mediated regulation of replication progression contributes to chromosome assembly during mitosis in human cells. MBoC 25, 302–317. https://doi.org/10.1091/mbc.e13-01-0020

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

Gerlich, D., Koch, B., Dupeux, F., Peters, J.-M., Ellenberg, J., 2006. Live-Cell Imaging Reveals a Stable Cohesin-Chromatin Interaction after but Not before DNA Replication. Current Biology 16, 1571–1578. https://doi.org/10.1016/j.cub.2006.06.068

241

Ghiselli, G., Siracusa, L.D., Iozzo, R.V., 1999. Complete cDNA Cloning, Genomic Organization, Chromosomal Assignment, Functional Characterization of the Promoter, and Expression of the Murine Bamacan Gene. Journal of Biological Chemistry 274, 17384–17393. https://doi.org/10.1074/jbc.274.24.17384

Gibcus, J.H., Samejima, K., Goloborodko, A., Samejima, I., Naumova, N., Nuebler, J., Kanemaki, M.T., Xie, L., Paulson, J.R., Earnshaw, W.C., Mirny, L.A., Dekker, J., 2018. A pathway for mitotic chromosome formation. Science 359, eaao6135. https://doi.org/10.1126/science.aao6135

Gligoris, T.G., Scheinost, J.C., Burmann, F., Petela, N., Chan, K.-L., Uluocak, P., Beckouet, F., Gruber, S., Nasmyth, K., Lowe, J., 2014. Closing the cohesin ring: Structure and function of its Smc3-kleisin interface. Science 346, 963–967. https://doi.org/10.1126/science.1256917

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., Lopez, R., 2010. A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Research 38, W695–W699. https://doi.org/10.1093/nar/gkq313

Gruber, S., Arumugam, P., Katou, Y., Kuglitsch, D., Helmhart, W., Shirahige, K., Nasmyth, K., 2006. Evidence that Loading of Cohesin Onto Chromosomes Involves Opening of Its SMC Hinge. Cell 127, 523–537. https://doi.org/10.1016/j.cell.2006.08.048

Gruber, S., Haering, C.H., Nasmyth, K., 2003. Chromosomal Cohesin Forms a Ring. Cell 112, 765–777. https://doi.org/10.1016/S0092-8674(03)00162-4

Gullerova, M., Proudfoot, N.J., 2008. Cohesin Complex Promotes Transcriptional Termination between Convergent Genes in S. pombe. Cell 132, 983–995. https://doi.org/10.1016/j.cell.2008.02.040

Gutierrez-Escribano, P., Hormeño, S., Madariaga-Marcos, J., Solé-Soler, R., O'Reilly, F.J., Morris, K., Aicart-Ramos, C., Aramayo, R., Montoya, A., Kramer, H., Rappsilber, J., Torres-Rosell, J., Moreno-Herrero, F., Aragon, L., 2020. Purified Smc5/6 Complex Exhibits DNA Substrate Recognition and Compaction. Molecular Cell 80, 1039-1054.e6. https://doi.org/10.1016/j.molcel.2020.11.012

Haarhuis, J.H.I., Elbatsh, A.M.O., Rowland, B.D., 2014. Cohesin and Its Regulation: On the Logic of X-Shaped Chromosomes. Developmental Cell 31, 7–18. https://doi.org/10.1016/j.devcel.2014.09.010

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

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

Haering, C.H., Löwe, J., Hochwagen, A., Nasmyth, K., 2002. Molecular Architecture of SMC Proteins and the Yeast Cohesin Complex. Molecular Cell 9, 773–788. https://doi.org/10.1016/S1097-2765(02)00515-4

Haering, C.H., Schoffnegger, D., Nishino, T., Helmhart, W., Nasmyth, K., Löwe, J., 2004. Structure and Stability of Cohesin's Smc1-Kleisin Interaction. Molecular Cell 15, 951–964. https://doi.org/10.1016/j.molcel.2004.08.030

Hansen, A.S., 2020. CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism. Nucleus 11, 132–148. https://doi.org/10.1080/19491034.2020.1782024

Harvey, S.H., Krien, M.J., O'Connell, M.J., 2002. Structural maintenance of chromosomes (SMC) proteins, a family of conserved ATPases 5. https://doi.org/10.1186/gb-2002-3-2-reviews3003

Hassler, M., Shaltiel, I.A., Haering, C.H., 2018. Towards a Unified Model of SMC Complex Function. Current Biology 28, R1266–R1281. https://doi.org/10.1016/j.cub.2018.08.034

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

Hauf, S., 2001. Cohesin Cleavage by Separase Required for Anaphase and Cytokinesis in Human Cells. Science 293, 1320–1323. https://doi.org/10.1126/science.1061376

Heimbruch, K.E., Meyer, A.E., Agrawal, P., Viny, A.D., Rao, S., 2021. A cohesive look at leukemogenesis: The cohesin complex and other driving mutations in AML. Neoplasia 23, 337–347. https://doi.org/10.1016/j.neo.2021.01.003

Higashi, T.L., Eickhoff, P., Sousa, J.S., Locke, J., Nans, A., Flynn, H.R., Snijders, A.P., Papageorgiou, G., O'Reilly, N., Chen, Z.A., O'Reilly, F.J., Rappsilber, J., Costa, A., Uhlmann, F., 2020. A Structure-Based Mechanism for DNA Entry into the Cohesin Ring. Molecular Cell S1097276520305037. https://doi.org/10.1016/j.molcel.2020.07.013

Higgins, C.F., Linton, K.J., 2004. The ATP switch model for ABC transporters. Nat Struct Mol Biol 11, 918–926. https://doi.org/10.1038/nsmb836

Hill, V.K., Kim, J.-S., Waldman, T., 2016. Cohesin mutations in human cancer. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer 1866, 1–11. https://doi.org/10.1016/j.bbcan.2016.05.002

Hirano, T., 2012. Condensins: universal organizers of chromosomes with diverse functions. Genes & Development 26, 1659–1678. https://doi.org/10.1101/gad.194746.112

Hirano, T., 2005. Condensins: Organizing and Segregating the Genome. Current Biology 15, R265–R275. https://doi.org/10.1016/j.cub.2005.03.037

Hirano, T., 2002. The ABCs of SMC proteins: two-armed ATPases for chromosome condensation, cohesion, and repair. Genes Dev. 16, 399–414. https://doi.org/10.1101/gad.955102

Hirano, T., Mitchison, T.J., 1994. A heterodimeric coiled-coil protein required for mitotic chromosome condensation in vitro. Cell 79, 449–458. https://doi.org/10.1016/0092-8674(94)90254-2

Hons, M.T., Huis in 't Veld, P.J., Kaesler, J., Rombaut, P., Schleiffer, A., Herzog, F., Stark, H., Peters, J.-M., 2016. Topology and structure of an engineered human cohesin complex bound to Pds5B. Nat Commun 7, 12523. https://doi.org/10.1038/ncomms12523

Hopfner, K., 2016. Invited review: Architectures and mechanisms of ATP binding cassette proteins. Biopolymers 105, 492–504. https://doi.org/10.1002/bip.22843

Hua, P., Badat, M., Hanssen, L.L.P., Hentges, L.D., Crump, N., Downes, D.J., Jeziorska, D.M., Oudelaar, A.M., Schwessinger, R., Taylor, S., Milne, T.A., Hughes, J.R., Higgs, D.R., Davies, J.O.J., 2021. Defining genome architecture at base-pair resolution. Nature. https://doi.org/10.1038/s41586-021-03639-4

Huis in 't Veld, P.J., Herzog, F., Ladurner, R., Davidson, I.F., Piric, S., Kreidl, E., Bhaskara, V., Aebersold, R., Peters, J.-M., 2014. Characterization of a DNA exit gate in the human cohesin ring. Science 346, 968–972. https://doi.org/10.1126/science.1256904

Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., Mirny, L.A., 2012. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods 9, 999–1003. https://doi.org/10.1038/nmeth.2148

International Human Genome Sequencing Consortium, 2004. Finishing the euchromatic sequence of the human genome. Nature 431, 931–945. https://doi.org/10.1038/nature03001

Ishiguro, K., 2019. The cohesin complex in mammalian meiosis. Genes Cells 24, 6–30. https://doi.org/10.1111/gtc.12652

Ivanov, D., Nasmyth, K., 2005. A Topological Interaction between Cohesin Rings and a Circular Minichromosome. Cell 122, 849–860. https://doi.org/10.1016/j.cell.2005.07.018

Jasin, M., Rothstein, R., 2013. Repair of Strand Breaks by Homologous Recombination. Cold Spring Harbor Perspectives in Biology 5, a012740–a012740. https://doi.org/10.1101/cshperspect.a012740

Jessberger, R., 2003. SMC Proteins at the Crossroads of Diverse Chromosomal Processes. IUBMB Life (International Union of Biochemistry and Molecular Biology: Life) 55, 643–652. https://doi.org/10.1080/15216540310001639661

Jones, P.M., George, A.M., 2004. The ABC transporter structure and mechanism: perspectives on recent research. Cellular and Molecular Life Sciences (CMLS) 61, 682–699. https://doi.org/10.1007/s00018-003-3336-9

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., Taatjes, D.J., Dekker, J., Young, R.A., 2010. Mediator and cohesin connect gene expression and chromatin architecture. Nature 467, 430–435. https://doi.org/10.1038/nature09380

Kavenoff, R., Bowen, B.C., 1976. Electron microscopy of membrane-free folded chromosomes from Escherichia coli. Chromosoma 59, 89–101. https://doi.org/10.1007/BF00328479

Kim, S.-T., 2002. Involvement of the cohesin protein, Smc1, in Atm-dependent and independent responses to DNA damage. Genes & Development 16, 560–570. https://doi.org/10.1101/gad.970602

Kim, Y., 2020. Shaping of the 3D genome by the ATPase machine cohesin. Molecular Medicine 7. https://doi.org/10.1038/s12276-020-00526-2

Kim, Y., Shi, Z., Zhang, H., Finkelstein, I.J., Yu, H., 2019. Human cohesin compacts DNA by loop extrusion. Science 366, 1345–1349. https://doi.org/10.1126/science.aaz4475

Kitajima, T.S., Sakuno, T., Ishiguro, K., Iemura, S., Natsume, T., Kawashima, S.A., Watanabe, Y., 2006. Shugoshin collaborates with protein phosphatase 2A to protect cohesin. Nature 441, 46–52. https://doi.org/10.1038/nature04663

Kornberg, R.D., 1977. Structure of Chromatin. Annu. Rev. Biochem. 46, 931–954. https://doi.org/10.1146/annurev.bi.46.070177.004435

Kouznetsova, E., Kanno, T., Karlberg, T., Thorsell, A.-G., Wisniewska, M., Kursula, P., Sjögren, C., Schüler, H., 2016. Sister Chromatid Cohesion Establishment Factor ESCO1 Operates by Substrate-Assisted Catalysis. Structure 24, 789–796. https://doi.org/10.1016/j.str.2016.03.021

Krishnan, A., Burroughs, A.M., Iyer, L.M., Aravind, L., 2020. Comprehensive classification of ABC ATPases and their functional radiation in nucleoprotein dynamics and biological conflict systems. Nucleic Acids Research 48, 10045–10075. https://doi.org/10.1093/nar/gkaa726

Kueng, S., Hegemann, B., Peters, B.H., Lipp, J.J., Schleiffer, A., Mechtler, K., Peters, J.-M., 2006. Wapl Controls the Dynamic Association of Cohesin with Chromatin. Cell 127, 955–967. https://doi.org/10.1016/j.cell.2006.09.040

Kurokawa, Y., Murayama, Y., 2020. DNA Binding by the Mis4 [scc2] Loader Promotes Topological DNA Entrapment by the Cohesin Ring. SSRN Journal. https://doi.org/10.2139/ssrn.3606296

Ladurner, R., Bhaskara, V., Huis in 't Veld, P.J., Davidson, I.F., Kreidl, E., Petzold, G., Peters, J.-M., 2014. Cohesin's ATPase Activity Couples Cohesin Loading onto DNA with Smc3 Acetylation. Current Biology 24, 2228–2237. https://doi.org/10.1016/j.cub.2014.08.011

Ladurner, R., Kreidl, E., Ivanov, M.P., Ekker, H., Idarraga-Amado, M.H., Busslinger, G.A., Wutz, G., Cisneros, D.A., Peters, J., 2016. Sororin actively maintains sister chromatid cohesion. EMBO J 35, 635–653. https://doi.org/10.15252/embj.201592532

Lai, Q., Li, Q., He, C., Fang, Y., Lin, S., Cai, J., Ding, J., Zhong, Q., Zhang, Y., Wu, C., Wang, X., He, J., Liu, Y., Yan, Q., Li, A., Liu, S., 2020. CTCF promotes colorectal cancer cell proliferation and chemotherapy resistance to 5-FU via the P53-Hedgehog axis. Aging 12, 16270–16293. https://doi.org/10.18632/aging.103648

Lammens, A., Schele, A., Hopfner, K.-P., 2004. Structural Biochemistry of ATP-Driven Dimerization and DNA-Stimulated Activation of SMC ATPases. Current Biology 5. https://doi.org/10.1016/j.cub.2004.09.044

Larionov, V.L., Karpova, T.S., Kouprina, N.Y., Jouravleva, G.A., 1985. A mutant of Saccharomyces cerevisiae with impaired maintenance of centromeric plasmids. Curr Genet 10, 15–20. https://doi.org/10.1007/BF00418488

Lehmann, A.R., Walicka, M., Griffiths, D.J., Murray, J.M., Watts, F.Z., McCready, S., Carr, A.M., 1995. The rad18 gene of Schizosaccharomyces pombe defines a new subgroup of the SMC superfamily involved in DNA repair. Mol Cell Biol 15, 7067–7080. https://doi.org/10.1128/MCB.15.12.7067

Lengronne, A., McIntyre, J., Katou, Y., Kanoh, Y., Hopfner, K.-P., Shirahige, K., Uhlmann, F., 2006. Establishment of Sister Chromatid Cohesion at the S. cerevisiae Replication Fork. Molecular Cell 23, 787–799. https://doi.org/10.1016/j.molcel.2006.08.018

Li, Y., Haarhuis, J.H.I., Sedeño Cacciatore, Á., Oldenkamp, R., van Ruiten, M.S., Willems, L., Teunissen, H., Muir, K.W., de Wit, E., Rowland, B.D., Panne, D., 2020. The structural basis for cohesin–CTCF-anchored loops. Nature 578, 472–476. https://doi.org/10.1038/s41586-019-1910-z

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

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., Dekker, J., 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. Science 326, 289–293. https://doi.org/10.1126/science.1181369

Lin, W., Wang, M., Jin, H., Yu, H.-G., 2011. Cohesin Plays a Dual Role in Gene Regulation and Sister-Chromatid Cohesion During Meiosis in *Saccharomyces cerevisiae*. Genetics 187, 1041–1051. https://doi.org/10.1534/genetics.110.122358

Linton, K.J., Higgins, C.F., 2007. Structure and function of ABC transporters: the ATP switch provides flexible control. Pflugers Arch - Eur J Physiol 453, 555–567. https://doi.org/10.1007/s00424-006-0126-x

Liu, H., Rankin, S., Yu, H., 2013. Phosphorylation-enabled binding of SGO1–PP2A to cohesin protects sororin and centromeric cohesion during mitosis. Nat Cell Biol 15, 40–49. https://doi.org/10.1038/ncb2637

Liu, J., Krantz, I., 2009. Cornelia de Lange syndrome, cohesin, and beyond. Clinical Genetics 76, 303–314. https://doi.org/10.1111/j.1399-0004.2009.01271.x

Liu, N.Q., Maresca, M., van den Brand, T., Braccioli, L., Schijns, M.M.G.A., Teunissen, H., Bruneau, B.G., Nora, E.P., de Wit, E., 2021. WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal gene regulation. Nat Genet 53, 100–109. https://doi.org/10.1038/s41588-020-00744-4

Liu, Y., Sung, S., Kim, Y., Li, F., Gwon, G., Jo, A., Kim, A., Kim, T., Song, O., Lee, S.E., Cho, Y., 2016. ATP - dependent DNA binding, unwinding, and resection by the Mre11/Rad50 complex. EMBO J 35, 743–758. https://doi.org/10.15252/embj.201592462

Lopez-Serra, L., Kelly, G., Patel, H., Stewart, A., Uhlmann, F., 2014. The Scc2–Scc4 complex acts in sister chromatid cohesion and transcriptional regulation by maintaining nucleosome-free regions. Nat Genet 46, 1147–1151. https://doi.org/10.1038/ng.3080

Losada, A., Yokochi, T., Hirano, T., 2005. Functional contribution of Pds5 to cohesin-mediated cohesion in human cells and *Xenopus* egg extracts. Journal of Cell Science 118, 2133–2141. https://doi.org/10.1242/jcs.02355

Losada, A., Yokochi, T., Kobayashi, R., Hirano, T., 2000. Identification and Characterization of Sa/Scc3p Subunits in the Xenopus and Human Cohesin Complexes. Journal of Cell Biology 150, 405–416. https://doi.org/10.1083/jcb.150.3.405

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J., 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389, 251–260. https://doi.org/10.1038/38444

Marko, J.F., De Los Rios, P., Barducci, A., Gruber, S., 2019. DNA-segment-capture model for loop extrusion by structural maintenance of chromosome (SMC) protein complexes. Nucleic Acids Research 47, 6956–6972. https://doi.org/10.1093/nar/gkz497

Mehta, A., Haber, J.E., 2014. Sources of DNA Double-Strand Breaks and Models of Recombinational DNA Repair. Cold Spring Harbor Perspectives in Biology 6, a016428–a016428. https://doi.org/10.1101/cshperspect.a016428

Meisenberg, C., Pinder, S.I., Hopkins, S.R., Wooller, S.K., Benstead-Hume, G., Pearl, F.M.G., Jeggo, P.A., Downs, J.A., 2019. Repression of Transcription at DNA Breaks Requires Cohesin throughout Interphase and Prevents Genome Instability. Molecular Cell 73, 212-223.e7. https://doi.org/10.1016/j.molcel.2018.11.001

Michaelis, C., Ciosk, R., Nasmyth, K., 1997. Cohesins: Chromosomal Proteins that Prevent Premature Separation of Sister Chromatids. Cell 91, 35–45. https://doi.org/10.1016/S0092-8674(01)80007-6

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

Mintzas, K., Heuser, M., 2019. Emerging strategies to target the dysfunctional cohesin complex in cancer. Expert Opinion on Therapeutic Targets 23, 525–537. https://doi.org/10.1080/14728222.2019.1609943

Mizuguchi, T., Fudenberg, G., Mehta, S., Belton, J.-M., Taneja, N., Folco, H.D., FitzGerald, P., Dekker, J., Mirny, L., Barrowman, J., Grewal, S.I.S., 2014. Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. Nature 516, 432–435. https://doi.org/10.1038/nature13833

Morales, C., Ruiz-Torres, M., Rodríguez-Acebes, S., Lafarga, V., Rodríguez-Corsino, M., Megías, D., Cisneros, D.A., Peters, J.-M., Méndez, J., Losada, A., 2020. PDS5 proteins are required for proper cohesin dynamics and participate in replication fork protection. Journal of Biological Chemistry 295, 146–157. https://doi.org/10.1074/jbc.RA119.011099

Muir, K.W., Li, Y., Weis, F., Panne, D., 2020. The structure of the cohesin ATPase elucidates the mechanism of SMC–kleisin ring opening. Nat Struct Mol Biol 27, 233–239. https://doi.org/10.1038/s41594-020-0379-7

Mullegama, S.V., Klein, S.D., Signer, R.H., UCLA Clinical Genomics Center, Vilain, E., Martinez-Agosto, J.A., 2019. Mutations in *STAG2* cause an X-linked cohesinopathy associated with undergrowth, developmental delay, and dysmorphia: Expanding the phenotype in males. Mol Genet Genomic Med 7, e00501. https://doi.org/10.1002/mgg3.501

Muñoz, S., Passarelli, F., Uhlmann, F., 2020. Conserved roles of chromatin remodellers in cohesin loading onto chromatin. Curr Genet 66, 951–956. https://doi.org/10.1007/s00294-020-01075-x

Murayama, Y., Uhlmann, F., 2015. DNA Entry into and Exit out of the Cohesin Ring by an Interlocking Gate Mechanism. Cell 163, 1628–1640. https://doi.org/10.1016/j.cell.2015.11.030

Nichols, M.H., Corces, V.G., 2018. A tethered-inchworm model of SMC DNA translocation. Nat Struct Mol Biol 25, 906–910. https://doi.org/10.1038/s41594-018-0135-4

Niki, H., Jaffé, A., Imamura, R., Ogura, T., Hiraga, S., 1991. The new gene mukB codes for a 177 kd protein with coiled-coil domains involved in chromosome partitioning of E. coli. The EMBO Journal 10, 183–193. https://doi.org/10.1002/j.1460-2075.1991.tb07935.x

Nishiyama, T., Ladurner, R., Schmitz, J., Kreidl, E., Schleiffer, A., Bhaskara, V., Bando, M., Shirahige, K., Hyman, A.A., Mechtler, K., Peters, J.-M., 2010. Sororin Mediates Sister Chromatid Cohesion by Antagonizing Wapl. Cell 143, 737–749. https://doi.org/10.1016/j.cell.2010.10.031

Nishiyama, T., Sykora, M.M., Huis in 't Veld, P.J., Mechtler, K., Peters, J.-M., 2013. Aurora B and Cdk1 mediate Wapl activation and release of acetylated cohesin from chromosomes by phosphorylating Sororin. Proceedings of the National Academy of Sciences 110, 13404–13409. https://doi.org/10.1073/pnas.1305020110

Nora, E.P., Caccianini, L., Fudenberg, G., So, K., Kameswaran, V., Nagle, A., Uebersohn, A., Hajj, B., Saux, A.L., Coulon, A., Mirny, L.A., Pollard, K.S., Dahan, M., Bruneau, B.G., 2020. Molecular basis of CTCF binding polarity in genome folding. Nat Commun 11, 5612. https://doi.org/10.1038/s41467-020-19283-x

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., Gribnau, J., Barillot, E., Blüthgen, N., Dekker, J., Heard, E., 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485, 381–385. https://doi.org/10.1038/nature11049

Olins, A.L., Olins, D.E., 1974. Spheroid Chromatin Units (ngr Bodies). Science 183, 330–332. https://doi.org/10.1126/science.183.4122.330

Ono, T., Losada, A., Hirano, M., Myers, M.P., Neuwald, A.F., Hirano, T., 2003. Differential Contributions of Condensin I and Condensin II to Mitotic Chromosome Architecture in Vertebrate Cells. Cell 115, 109–121. https://doi.org/10.1016/S0092-8674(03)00724-4

Oswald, C., Holland, I.B., Schmitt, L., 2006. The motor domains of ABC-transporters: What can structures tell us? Naunyn Schmied Arch Pharmacol 372, 385–399. https://doi.org/10.1007/s00210-005-0031-4

Ou, H.D., Phan, S., Deerinck, T.J., Thor, A., Ellisman, M.H., O'Shea, C.C., 2017. ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science 357, eaag0025. https://doi.org/10.1126/science.aag0025

Oudet, P., Gross-Bellard, M., Chambon, P., 1975. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. Cell 4, 281–300. https://doi.org/10.1016/0092-8674(75)90149-x

Ouyang, Z., Zheng, G., Song, J., Borek, D.M., Otwinowski, Z., Brautigam, C.A., Tomchick, D.R., Rankin, S., Yu, H., 2013. Structure of the human cohesin inhibitor Wapl. Proceedings of the National Academy of Sciences 110, 11355–11360. https://doi.org/10.1073/pnas.1304594110

Palecek, J.J., Gruber, S., 2015. Kite Proteins: a Superfamily of SMC/Kleisin Partners Conserved Across Bacteria, Archaea, and Eukaryotes. Structure 23, 2183–2190. https://doi.org/10.1016/j.str.2015.10.004

Panizza, S., Tanaka, T., Hochwagen, A., Eisenhaber, F., Nasmyth, K., 2000. Pds5 cooperates with cohesin in maintaining sister chromatid cohesion. Current Biology 10, 1557–1564. https://doi.org/10.1016/S0960-9822(00)00854-X

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., Cobb, B.S., Yokomori, K., Dillon, N., Aragon, L., Fisher, A.G., Merkenschlager, M., 2008. Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. Cell 132, 422–433. https://doi.org/10.1016/j.cell.2008.01.011

Parenti, I., Diab, F., Gil, S.R., Mulugeta, E., Casa, V., Berutti, R., Brouwer, R.W.W., Dupé, V., Eckhold, J., Graf, E., Puisac, B., Ramos, F., Schwarzmayr, T., Gines, M.M., van Staveren, T., van IJcken, W.F.J., Strom, T.M., Pié, J., Watrin, E., Kaiser, F.J., Wendt, K.S., 2020. MAU2 and NIPBL Variants Impair the Heterodimerization of the Cohesin Loader Subunits and Cause Cornelia de Lange Syndrome. Cell Reports 31, 107647. https://doi.org/10.1016/j.celrep.2020.107647

Paulson, J.R., Hudson, D.F., Cisneros-Soberanis, F., Earnshaw, W.C., 2021. Mitotic chromosomes. Seminars in Cell & Developmental Biology 117, 7–29. https://doi.org/10.1016/j.semcdb.2021.03.014

Perea-Resa, C., Bury, L., Cheeseman, I.M., Blower, M.D., 2020. Cohesin Removal Reprograms Gene Expression upon Mitotic Entry. Molecular Cell 78, 127-140.e7. https://doi.org/10.1016/j.molcel.2020.01.023

Peric-Hupkes, D., van Steensel, B., 2008. Linking Cohesin to Gene Regulation. Cell 132, 925–928. https://doi.org/10.1016/j.cell.2008.03.001

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

Peters, J.-M., 2021. How DNA loop extrusion mediated by cohesin enables V(D)J recombination. Current Opinion in Cell Biology 70, 75–83. https://doi.org/10.1016/j.ceb.2020.11.007

Peters, J.-M., Nishiyama, T., 2012. Sister Chromatid Cohesion. Cold Spring Harbor Perspectives in Biology 4, a011130–a011130. https://doi.org/10.1101/cshperspect.a011130

Peters, J.-M., Tedeschi, A., Schmitz, J., 2008. The cohesin complex and its roles in chromosome biology. Genes & Development 22, 3089–3114. https://doi.org/10.1101/gad.1724308

Piñeiro, Á., Muñoz, E., Sabín, J., Costas, M., Bastos, M., Velázquez-Campoy, A., Garrido, P.F., Dumas, P., Ennifar, E., García-Río, L., Rial, J., Pérez, D., Fraga, P., Rodríguez, A., Cotelo, C., 2019. AFFINImeter: A software to analyze molecular recognition processes from experimental data. Analytical Biochemistry 577, 117–134. https://doi.org/10.1016/j.ab.2019.02.031

Pugacheva, E.M., Kubo, N., Loukinov, D., Tajmul, M., Kang, S., Kovalchuk, A.L., Strunnikov, A.V., Zentner, G.E., Ren, B., Lobanenkov, V.V., 2020. CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. Proc Natl Acad Sci USA 117, 2020–2031. https://doi.org/10.1073/pnas.1911708117

Rankin, S., Ayad, N.G., Kirschner, M.W., 2005. Sororin, a Substrate of the Anaphase- Promoting Complex, Is Required for Sister Chromatid Cohesion in Vertebrates. Molecular Cell 18, 185–200. https://doi.org/10.1016/j.molcel.2005.03.017

Rees, D.C., Johnson, E., Lewinson, O., 2009. ABC transporters: the power to change. Nat Rev Mol Cell Biol 10, 218–227. https://doi.org/10.1038/nrm2646

Revenkova, E., Focarelli, M.L., Susani, L., Paulis, M., Bassi, M.T., Mannini, L., Frattini, A., Delia, D., Krantz, I., Vezzoni, P., Jessberger, R., Musio, A., 2009. Cornelia de Lange syndrome mutations in SMC1A or SMC3 affect binding to DNA. Human Molecular Genetics 18, 418–427. https://doi.org/10.1093/hmg/ddn369

Rhodes, J.D.P., Haarhuis, J.H.I., Grimm, J.B., Rowland, B.D., Lavis, L.D., Nasmyth, K.A., 2017. Cohesin Can Remain Associated with Chromosomes during DNA Replication. Cell Reports 20, 2749–2755. https://doi.org/10.1016/j.celrep.2017.08.092

Rhodes, J.M., McEwan, M., Horsfield, J.A., 2011. Gene Regulation by Cohesin in Cancer: Is the Ring an Unexpected Party to Proliferation? Mol Cancer Res 9, 1587–1607. https://doi.org/10.1158/1541-7786.MCR-11-0382

Rocques, P.J., Clark, J., Ball, S., Crew, J., Gill, S., Christodoulou, Z., Borts, R.H., Louis, E.J., Davies, K.E., Cooper, C.S., 1995. The human *SB1.8* gene (*DXS423E*) encodes a putative chromosome segregation protein conserved in lower eukaryotes and prokaryotes. Hum Mol Genet 4, 243–249. https://doi.org/10.1093/hmg/4.2.243

Roig, M.B., Löwe, J., Chan, K.-L., Beckouët, F., Metson, J., Nasmyth, K., 2014. Structure and function of cohesin's Scc3/SA regulatory subunit. FEBS Letters 588, 3692–3702. https://doi.org/10.1016/j.febslet.2014.08.015

Rojowska, A., Lammens, K., Seifert, F.U., Direnberger, C., Feldmann, H., Hopfner, K., 2014. Structure of the Rad50 DNA double-strand break repair protein in complex with DNA. EMBO J 33, 2847–2859. https://doi.org/10.15252/embj.201488889

Rowland, B.D., Roig, M.B., Nishino, T., Kurze, A., Uluocak, P., Mishra, A., Beckouët, F., Underwood, P., Metson, J., Imre, R., Mechtler, K., Katis, V.L., Nasmyth, K., 2009. Building Sister Chromatid Cohesion: Smc3 Acetylation Counteracts an Antiestablishment Activity. Molecular Cell 33, 763–774. https://doi.org/10.1016/j.molcel.2009.02.028

Rubio, E.D., Reiss, D.J., Welcsh, P.L., Disteche, C.M., Filippova, G.N., Baliga, N.S., Aebersold, R., Ranish, J.A., Krumm, A., 2008. CTCF physically links cohesin to chromatin. Proceedings of the National Academy of Sciences 105, 8309–8314. https://doi.org/10.1073/pnas.0801273105

Ryu, J.-K., Katan, A.J., van der Sluis, E.O., Wisse, T., de Groot, R., Haering, C.H., Dekker, C., 2020. The condensin holocomplex cycles dynamically between open and collapsed states. Nat Struct Mol Biol. https://doi.org/10.1038/s41594-020-0508-3

Sarogni, P., Palumbo, O., Servadio, A., Astigiano, S., D'Alessio, B., Gatti, V., Cukrov, D., Baldari, S., Pallotta, M.M., Aretini, P., Dell'Orletta, F., Soddu, S., Carella, M., Toietta, G., Barbieri, O., Fontanini, G., Musio, A., 2019. Overexpression of the cohesin-core subunit SMC1A contributes to colorectal cancer development. J Exp Clin Cancer Res 38, 108. https://doi.org/10.1186/s13046-019-1116-0

Schalbetter, S.A., Goloborodko, A., Fudenberg, G., Belton, J.-M., Miles, C., Yu, M., Dekker, J., Mirny, L., Baxter, J., 2017. SMC complexes differentially compact mitotic chromosomes according to genomic context. Nat Cell Biol 19, 1071–1080. https://doi.org/10.1038/ncb3594

Schleiffer, A., Kaitna, S., Maurer-Stroh, S., Glotzer, M., Nasmyth, K., Eisenhaber, F., 2003. Kleisins: A Superfamily of Bacterial and Eukaryotic SMC Protein Partners. Molecular Cell 11, 571–575. https://doi.org/10.1016/S1097-2765(03)00108-4

Seifert, F.U., Lammens, K., Stoehr, G., Kessler, B., Hopfner, K., 2016. Structural mechanism of ATP - dependent DNA binding and DNA end bridging by eukaryotic Rad50. EMBO J 35, 759–772. https://doi.org/10.15252/embj.201592934

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., Cavalli, G., 2012. Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. Cell 148, 458–472. https://doi.org/10.1016/j.cell.2012.01.010

Shi, Z., Gao, H., Bai, X., Yu, H., 2020. Cryo-EM structure of the human cohesin-NIPBL-DNA complex. Science 368, 1454–1459. https://doi.org/10.1126/science.abb0981

Skibbens, R.V., 2019. Condensins and cohesins – one of these things is not like the other! Journal of Cell Science 132, jcs220491. https://doi.org/10.1242/jcs.220491

Soardi, F.C., Machado-Silva, A., Linhares, N.D., Zheng, G., Qu, Q., Pena, H.B., Martins, T.M.M., Vieira, H.G.S., Pereira, N.B., Melo-Minardi, R.C., Gomes, C.C., Gomez, R.S., Gomes, D.A., Pires, D.E.V., Ascher, D.B., Yu, H., Pena, S.D.J., 2017. Familial STAG2 germline mutation defines a new human cohesinopathy. npj Genomic Med 2, 7. https://doi.org/10.1038/s41525-017-0009-4

Soh, Y.-M., Bürmann, F., Shin, H.-C., Oda, T., Jin, K.S., Toseland, C.P., Kim, C., Lee, H., Kim, S.J., Kong, M.-S., Durand-Diebold, M.-L., Kim, Y.-G., Kim, H.M., Lee, N.K., Sato, M., Oh, B.-H., Gruber, S., 2015. Molecular Basis for SMC Rod Formation and Its Dissolution upon DNA Binding. Molecular Cell 57, 290–303. https://doi.org/10.1016/j.molcel.2014.11.023

Strunnikov, A.V., Jessberger, R., 1999. Structural maintenance of chromosomes (SMC) proteins. Conserved molecular properties for multiple biological functions. Eur J Biochem 263, 6–13. https://doi.org/10.1046/j.1432-1327.1999.00509.x

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

Sumara, I., Vorlaufer, E., Gieffers, C., Peters, B.H., Peters, J.-M., 2000. Characterization of Vertebrate Cohesin Complexes and Their Regulation in Prophase. Journal of Cell Biology 151, 749–762. https://doi.org/10.1083/jcb.151.4.749

Sun, L., Yu, R., Dang, W., 2018. Chromatin Architectural Changes during Cellular Senescence and Aging. Genes 9, 211. https://doi.org/10.3390/genes9040211

Szabo, Q., Bantignies, F., Cavalli, G., 2019. Principles of genome folding into topologically associating domains. Sci. Adv. 5, eaaw1668. https://doi.org/10.1126/sciadv.aaw1668

Tomonaga, T., 2000. Characterization of fission yeast cohesin: essential anaphase proteolysis of Rad21 phosphorylated in the S phase. Genes & Development 14, 2757–2770. https://doi.org/10.1101/gad.832000

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

Tsutsumi, M., Fujiwara, R., Nishizawa, H., Ito, M., Kogo, H., Inagaki, H., Ohye, T., Kato, T., Fujii, T., Kurahashi, H., 2014. Age-Related Decrease of Meiotic Cohesins in Human Oocytes. PLoS ONE 9, e96710. https://doi.org/10.1371/journal.pone.0096710

Uemura, H., Shiba, T., Machida, M., Matsui, I., Jigami, Y., Tanaka, H., 1987. A Postitive Regulatoty Sequence of the Saccharomyuces crevisiae ENO1 Gene1. The Journal of Biochemistry 102, 181–189. https://doi.org/10.1093/oxfordjournals.jbchem.a122031

Uhlmann, F., 2016. SMC complexes: from DNA to chromosomes. Nat Rev Mol Cell Biol 17, 399–412. https://doi.org/10.1038/nrm.2016.30

van der Lelij, P., Lieb, S., Jude, J., Wutz, G., Santos, C.P., Falkenberg, K., Schlattl, A., Ban, J., Schwentner, R., Hoffmann, T., Kovar, H., Real, F.X., Waldman, T., Pearson, M.A., Kraut, N., Peters, J.-M., Zuber, J., Petronczki, M., 2017. Synthetic lethality between the cohesin subunits STAG1 and STAG2 in diverse cancer contexts. eLife 6, e26980. https://doi.org/10.7554/eLife.26980

Vazquez Nunez, R., Ruiz Avila, L.B., Gruber, S., 2019. Transient DNA Occupancy of the SMC Interarm Space in Prokaryotic Condensin. Molecular Cell 75, 209-223.e6. https://doi.org/10.1016/j.molcel.2019.05.001

Vian, L., Pękowska, A., Rao, S.S.P., Kieffer-Kwon, K.-R., Jung, S., Baranello, L., Huang, S.-C., El Khattabi, L., Dose, M., Pruett, N., Sanborn, A.L., Canela, A., Maman, Y., Oksanen, A., Resch, W., Li, X., Lee, B., Kovalchuk, A.L., Tang, Z., Nelson, S., Di Pierro, M., Cheng, R.R., Machol, I., St Hilaire, B.G., Durand, N.C., Shamim, M.S., Stamenova, E.K., Onuchic, J.N., Ruan, Y., Nussenzweig, A., Levens, D., Aiden, E.L., Casellas, R., 2018. The Energetics and Physiological Impact of Cohesin Extrusion. Cell 173, 1165-1178.e20. https://doi.org/10.1016/j.cell.2018.03.072

Vincentelli, R., Romier, C., 2016. Complex Reconstitution and Characterization by Combining Co-expression Techniques in Escherichia coli with High-Throughput, in: Vega, M.C. (Ed.), Advanced Technologies for Protein Complex Production and Characterization, Advances in Experimental Medicine and Biology. Springer International Publishing, Cham, pp. 43–58. https://doi.org/10.1007/978-3-319-27216-0_4

Waldman, T., 2020. Emerging themes in cohesin cancer biology. Nat Rev Cancer 20, 504–515. https://doi.org/10.1038/s41568-020-0270-1

Wang, G., Maier, R.J., 2015. Bacterial histone-like proteins: roles in stress resistance. Curr Genet 61, 489–492. https://doi.org/10.1007/s00294-015-0478-x

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

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

Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., Guo, Y.E., Hnisz, D., Jaenisch, R., Bradner, J.E., Gray, N.S., Young, R.A., 2017. YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell 171, 1573-1588.e28. https://doi.org/10.1016/j.cell.2017.11.008

Weitzer, S., Lehane, C., Uhlmann, F., 2003. A Model for ATP Hydrolysis-Dependent Binding of Cohesin to DNA. Current Biology 13, 1930–1940. https://doi.org/10.1016/j.cub.2003.10.030

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

Weterings, E., van Gent, D.C., 2004. The mechanism of non-homologous end-joining: a synopsis of synapsis. DNA Repair 3, 1425–1435. https://doi.org/10.1016/j.dnarep.2004.06.003

Wu, N., Yu, H., 2012. The Smc complexes in DNA damage response. Cell Biosci 2, 5. https://doi.org/10.1186/2045-3701-2-5

Wu, R.-R., Couchman, J.R., 1997. cDNA Cloning of the Basement Membrane Chondroitin Sulfate Proteoglycan Core Protein, Bamacan: A Five Domain Structure Including Coiled-Coil Motifs. Journal of Cell Biology 136, 433–444. https://doi.org/10.1083/jcb.136.2.433

Wutz, G., Ladurner, R., St Hilaire, B.G., Stocsits, R.R., Nagasaka, K., Pignard, B., Sanborn, A., Tang, W., Várnai, C., Ivanov, M.P., Schoenfelder, S., van der Lelij, P., Huang, X., Dürnberger, G., Roitinger, E., Mechtler, K., Davidson, I.F., Fraser, P., Lieberman-Aiden, E., Peters, J.-M., 2020. ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesinSTAG1 from WAPL. eLife 9, e52091. https://doi.org/10.7554/eLife.52091

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

Wyman, C., Kanaar, R., 2006. DNA Double-Strand Break Repair: All's Well that Ends Well. Annu. Rev. Genet. 40, 363–383. https://doi.org/10.1146/annurev.genet.40.110405.090451

Xiang, S., Koshland, D., 2021. Cohesin architecture and clustering in vivo. eLife 10, e62243. https://doi.org/10.7554/eLife.62243

Yadav, S., Kowolik, C.M., Lin, M., Zuro, D., Hui, S.K., Riggs, A.D., Horne, D.A., 2019. SMC1A is associated with radioresistance in prostate cancer and acts by regulating epithelial-mesenchymal transition and cancer stem-like properties. Molecular Carcinogenesis 58, 113–125. https://doi.org/10.1002/mc.22913

Yadav, Sushma, Sehrawat, A., Eroglu, Z., Somlo, G., Hickey, R., Yadav, Sailee, Liu, X., Awasthi, Y.C., Awasthi, S., 2013. Role of SMC1 in Overcoming Drug Resistance in Triple Negative Breast Cancer. PLoS ONE 8, e64338. https://doi.org/10.1371/journal.pone.0064338

Yazdi, P.T., 2002. SMC1 is a downstream effector in the ATM/NBS1 branch of the human S-phase checkpoint. Genes & Development 16, 571–582. https://doi.org/10.1101/gad.970702

Zhang, J., Shi, X., Li, Y., Kim, B.-J., Jia, J., Huang, Z., Yang, T., Fu, X., Jung, S.Y., Wang, Y., Zhang, P., Kim, S.-T., Pan, X., Qin, J., 2008. Acetylation of Smc3 by Eco1 Is Required for S Phase Sister Chromatid Cohesion in Both Human and Yeast. Molecular Cell 31, 143–151. https://doi.org/10.1016/j.molcel.2008.06.006

Zhang, N., Coutinho, L.E., Pati, D., 2021. PDS5A and PDS5B in Cohesin Function and Human Disease. IJMS 22, 5868. https://doi.org/10.3390/ijms22115868

Zhang, N., Panigrahi, A.K., Mao, Q., Pati, D., 2011. Interaction of Sororin Protein with Polo-like Kinase 1 Mediates Resolution of Chromosomal Arm Cohesion. Journal of Biological Chemistry 286, 41826–41837. https://doi.org/10.1074/jbc.M111.305888

Zhang, N., Pati, D., 2012. Sororin is a master regulator of sister chromatid cohesion and separation. Cell Cycle 11, 2073–2083. https://doi.org/10.4161/cc.20241

Marina Jacinta VITORIA GOMES

Distinct conformational dynamics of human Cohesin SMC1A and SMC3 ATPase domains

The cohesin complex is part of the family of Structural Maintenance of Chromosomes (SMC) protein complexes that associate with chromatin and have vital roles in 3D genome organization and in the maintenance of its stability and integrity. Cohesin is notably involved in major genome regulation processes such as sister chromatid cohesion, chromosome segregation, transcription regulation, and chromatin structure organization. The cohesin functions strongly depend on the ATP binding and hydrolysis activity of its composite ATPase module, which is composed of the ATPase heads of the SMC1A and SMC3 core subunits, bound to the C- and N-terminal domains of the kleisin RAD21, respectively. The cohesin ATPase activity is known to drive important structural changes in the complex architecture, thus enabling the dynamic association of cohesin with chromatin. However, the molecular basis of ATP binding and hydrolysis by SMC1A and SMC3 and their associated structural changes within cohesin remain poorly understood. Throughout my thesis work, I investigated the ATP binding and hydrolysis properties of the human cohesin SMC1A and SMC3 ATPase domains by using biochemical, biophysical, and structural methods. I solved the crystallographic structures of human SMC1A and SMC3 ATPase heads in their apo, ADP-bound, and an ATP analog, ATPγS-bound conformations, thus revealing specific structural changes upon ATP binding and its hydrolysis into ADP. Additionally, my results highlight the differences in ATP binding by SMC1A and SMC3 and their distinct conformational dynamics, and reveal a so far unidentified conformation of the DNA exit-gate from the cohesin ring. Altogether, my thesis results provide novel molecular insights into the ATP binding, hydrolysis, and release cycle of this essential molecular motor.

Keywords: Genome organization, Chromatin, SMC complexes, Cohesin, ATPase activity, Coiled coils, Conformational changes, DNA exit gate.

Le complexe de cohésine fait partie de la famille des complexes protéiques de maintenance structurelle des chromosomes (SMC), qui s'associent à la chromatine et jouent un rôle vital dans l'organisation 3D du génome et dans le maintien de sa stabilité et de son intégrité. La cohésine est notamment impliquée dans des processus majeurs de régulation du génome, tels que la cohésion des chromatides sœurs, la ségrégation des chromosomes, la régulation de l'expression des gènes et l'organisation de la structure de la chromatine. Les fonctions de la cohésine dépendent de l'activité de liaison et d'hydrolyse de l'ATP par son module ATPase, qui est composé des têtes ATPase des sous-unités cœur SMC1A et SMC3, liées respectivement aux domaines C- et N-terminaux de la kleisine RAD21. L'activité ATPase de la cohésine entraîne d'importants changements structurels au sein du complexe, qui permettent l'association dynamique de la cohésine avec la chromatine. Cependant, les bases moléculaires de la liaison et de l'hydrolyse de l'ATP par SMC1A et SMC3 et des modifications structurelles qui y sont associées restent mal comprises. Pendant mon travail de thèse, j'ai analysé la liaison et l'hydrolyse de l'ATP par les domaines ATPase SMC1A et SMC3 de la cohésine humaine, par des méthodes biochimiques, biophysiques et structurales. J'ai résolu des structures cristallographiques des domaines ATPase de SMC1A et de SMC3, sous leur forme non liée et liée à l'ADP et à l'ATPγS, un analogue de l'ATP. Mes résultats montrent les changements structuraux qui ont lieu lors de la fixation et de l'hydrolyse de l'ATP en ADP. De plus, mes résultats soulignent les différences dans la liaison et l'hydrolyse de l'ATP par SMC1A et SMC3, ainsi que leurs changements conformationnels distincts, notamment par la révélation d'une conformation jusqu'alors jamais observée de la porte de sortie de l'ADN du complexe cohésine. Dans leur ensemble, les résultats issus de mes travaux de thèse fournissent de nouvelles connaissances au niveau moléculaire dans le cycle ATPase de ce moteur moléculaire essentiel.

Mots clés : Organisation du génome, Chromatine, Complexes SMC, Cohésine, Activité ATPase, Domaine superhélice, Changements conformationnels, Porte de sortie de l'ADN.