Dynamique conformationnelle du module ATPase de la Cohésine humaine

Pauline Landwerlin

▶ **To cite this version:**

Pauline Landwerlin. Dynamique conformationnelle du module ATPase de la Cohésine humaine. Biologie structurale [q-bio.BM]. Université de Strasbourg, 2022. Français. NNT : 2022STRAJ096 . tel-04043322

HAL Id: tel-04043322
https://theses.hal.science/tel-04043322

Submitted on 23 Mar 2023

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

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

Université de Strasbourg

UNIVERSITÉ DE STRASBOURG

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE

Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC)

UMR 7104 - U1258

Département de Biologie Structurale intégrative

THÈSE présentée par : **Pauline LANDWERLIN**

Soutenue le : **20 octobre 2022**

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

Discipline/ Spécialité Sciences de la Vie et de la Santé

Dynamique conformationnelle du module ATPase de la Cohésine humaine

THÈSE dirigée par :

Dr. ROMIER Christophe	Directeur de recherche, IGBMC, Illkirch, France

RAPPORTEURS :

Dr. PETOSA Carlo	Directeur de recherche, Institut de Biologie Structurale (IBS), Grenoble, France
Dr. BERNARD Pascal	Directeur de recherche, Ecole normale supérieure, Lyon, France

EXAMINATEUR INTERNE :

Dr. DANTZER Françoise	Directrice de recherche, Ecole Supérieure de Biotechnologie (ESBS), Illkirch, France

EXAMINATEUR EXTERNE :

Dr. LENGRONNE Armelle	Directrice de recherche, Institut de Génétique Humaine (IGH), Montpellier, France

MEMBRE INVITÉ :

Dr. DURAND Marie-Laure	Chargée de recherche, IGBMC, Illkirch, France

Remerciements

Je tiens tout d'abord à remercier les membres de mon jury de thèse, Dr. Carlo Petosa, Dr. Pascal Bernard, Dr. Françoise Dantzer, Dr. Armelle Lengronne d'avoir accepté d'évaluer mon travail. Merci pour votre disponibilité et votre bienveillance.

Je remercie tout particulièrement le Dr. Christophe Romier pour m'avoir accueil au sein de son équipe de recherche et de m'avoir accompagné dans cette aventure. Merci Christophe d'avoir cru en moi dès mon arrivée dans le laboratoire en master 2 et pour la patience, le soutien et les conseils prodigués tout au long de ma thèse. Cela a été une grande expérience d'apprentissage fort passionnante, la biologie structurale n'étant pas mon domaine d'étude initial.

Je tiens à remercier la Fondation ARC pour le soutien financier apporté pour mes trois premières années, ainsi que la Fondation pour la Recherche Médicale, qui m'a permis de réaliser ma dernière année de thèse.

Je remercie également mon comité de suivi de thèse : Dr. Eric Ennifar, Dr. Helgo Schmidt, Dr. Catherine-Laure Tomasetto pour leurs recommandations et suggestions pour mon travail de thèse.

Ce projet est le fruit d'un travail collaboratif et n'aurait pu être réalisé sans l'implication et la participation de nombreuses personnes. Je remercie nos collaborateurs, Dr. Eric Ennifar, Dr. Karl Brillet et Dr. Christelle Golzio, pour leurs expertises en ITC et sur le modèle animal du poisson zèbre.

J'aimerais également remercier tous les membres de mon équipe d'accueil pour la bonne ambiance, pour les discussions scientifiques et papotages personnels, qui ont été accompagnés d'innombrables gâteaux et tartes, …

Merci Marie-Laure pour ton aide, tes précieux conseils et de m'avoir fait bénéficier de ton expertise. Cela a été un vrai plaisir de travailler avec toi.

Merci Edouard pour ta loquacité et tous tes efforts pour les clonages récalcitrants, ainsi que tes associations improbables de gâteaux et compotes.

Merci Pierre pour tes suggestions et conseils avisés, notamment pour la rédaction des demandes de financement.

Je tiens également à remercier les membres actuels et anciens de mon équipe d'accueil, Hugo, Elizabeth, Pernelle, Marina, Natalia et Tajith pour vos conseils et votre bonne humeur au quotidien.

Merci à tous les membres de la plateforme de Biologie Structurale Intégrée et à l'ensemble de membres du département de Biologie Structurale Intégrative, en particulier à Catherine Birck, Alastair Mcewen, Pierre Poussin-Courmontagne, Alexandre Durand et Nils Marechal pour leurs aides et connaissances pour l'AUC, la cristallographie et la cryoEM.

Merci également à mes collèges ingénieurs, doctorants et post-doc du CBI avec qui j'ai pu partager des moments de convivialité et qui pour certain(e)s sont devenus des amis après ces quatre années. Merci à Camille et Anna (les doudous), Léa, Charlotte, Marlène et Christophe pour avoir égayé mes déjeuners et certaines soirées et pour m'avoir fait découvrir de nombreux jeux de société.

Petite mention pour Pernelle, la meilleure collègue possible. Merci pour ton éternelle bonne humeur, ton soutien sans faille et nos (nombreux) déjeuners à la table de l'Ile.

Pour finir, je voudrais remercier ma famille, sans qui tout cela n'aurait pas été possible. Merci de tout cœur à mes parents, Nathalie et Thierry, à ma sœur Charlène et à ma grand-mère Nicole, pour vos encouragements et soutiens après toutes ces années.

Enfin, je remercie celui qui partage ma vie et qui m'a courageusement épaulée durant cette thèse. Merci à toi, Valentin, pour ton soutien sans faille et tout simplement pour ta présence à mes côtés qui embellit mes journées. Merci également à tes parents Martine et Francis, ainsi qu'à ton frère Antoine pour leur gentillesse.

Merci à tous !

Table des matières

REMERCIEMENTS **3**

TABLE DES MATIERES **5**

LISTE DES ABREVIATIONS **10**

RESUME DE THESE **13**

I. RECONSTITUTION DU MODULE ATPASE ENGAGE DE LA COHÉSINE HUMAINE 15

II. CARACTERISATION STRUCTURALE DU MODULE ATPASE ENGAGE 15

III. RECONSTITUTION DU COMPLEXE QUADRIPARTITE (MODULE ATPASE COUPLÉ À STAG1) 17

CHAPITRE 1 – INTRODUCTION **18**

I. LE CYCLE CELLULAIRE EUCARYOTE 18

1.1. L'interphase *19*

1.2. La mitose *20*

1.3. Variations entre les eucaryotes inférieurs et les mammifères *21*

II. ARCHITECTURE DU GENOME EUCARYOTE : DE LA DOUBLE HELICE D'ADN AU CHROMOSOME 24

2.1. Découverte de la double hélice d'ADN *24*

2.2. L'organisation de la chromatine au cours du cycle cellulaire *25*

2.2.1. La structure de la chromatine lors de l'interphase 26

2.2.1.1. Du nucléosome aux nanodomaines de chromatine 27

2.2.1.2. Les boucles de chromatine 28

2.2.1.3. Domaines d'association topologique 29

2.2.1.3.1. Chez les mammifères *29*

2.2.1.3.2. Différences entre les mammifères, la drosophile et les levures *30*

2.2.1.4. Les compartiments A/B 30

2.2.1.5. Territoire chromosomique 31

2.2.2. Conformation des chromosomes dans les cellules mitotiques 32

III. LES COMPLEXES SMC : ORGANISATEURS DE LA CHROMATINE 36

3.1. Découverte des protéines SMC *36*

3.2. Architecture générale des complexes SMC *38*

3.2.1. Les protéines SMC 39

3.2.2. La protéine kleisine, une sous-unité cœur non-SMC régulatrice 40

3.2.3. L'assemblage des complexes SMC 41

3.2.4. Les protéines auxiliaires et régulatrices 44

3.3. Les différents complexes SMC : fonctions, compositions, architectures et spécificités *45*

3.3.1. Les complexes procaryotes 45

3.3.1.1. Le complexe SMC-ScpAB 45
3.3.1.2. Le complexe MukBEF 46
3.3.1.1. Le complexe MksBEF 50
3.3.2. Les complexes eucaryotes : Condensine, Complexe SMC5/SMC6 et Cohésine 51
3.3.2.1. La Condensine 51
3.3.2.2. Le complexe SMC5/SMC6 58
3.3.2.3. La Cohésine 60
IV. LE COMPLEXE COHESINE 65
4.1. Les fonctions de la Cohésine 65
4.1.1. Cohésion des chromatides sœurs (rôle canonique) 65
4.1.2. Réplication de l'ADN 67
4.1.3. Réparation des dommages de l'ADN par recombinaison homologue 67
4.1.4. Organisation 3D du génome et de la chromatine 69
4.1.5. Régulation de la transcription 71
4.1.6. Recombinaison V(D)J et commutation de classes des immunoglobulines 72
4.2. Composition de la Cohésine 73
4.2.1. L'anneau tripartite, le complexe cœur de la Cohésine 73
4.2.2. Les protéines auxiliaires et régulatrices 75
4.2.2.1. La protéine STAG : protéine auxiliaire essentielle pour la Cohésine 75
4.2.2.2. Les protéines régulatrices 77
4.2.2.2.1. Le complexe NIPBL/MAU2 : complexe de chargement 77
4.2.2.2.2. Le complexe PDS5 / WAPL / SORORINE : complexe d'association/dissociation 78
4.2.2.2.3. Les protéines ESCO et HDAC8 : stabilisation de l'interface Cohésine-ADN et recyclage 79
4.2.2.2.4. Le complexe SHUGOSHIN/PP2A : complexe protecteur 80
4.2.2.2.5. La protéine CTCF : protéine d'ancrage 80
4.3. Le cycle de la Cohésine 81
4.3.1. Modes d'interaction entre la Cohésine et l'ADN 81
4.3.2. Mécanisme impliqué dans la cohésion des chromatides sœurs 81
4.3.2.1. Chargement de la Cohésine sur l'ADN 83
4.3.2.2. Établissement de la cohésion des chromatides sœurs 83
4.3.2.3. Retrait de la Cohésine et dissolution de la cohésion 84
4.3.3. Mécanisme impliqué dans l'extrusions des boucles de chromatine 85
4.4. État de l'art : les structures de la Cohésine connues à ce jour 86
4.5. Implication de la Cohésine dans les pathologies humaines 93
4.5.1. Cohésinopathies 93
4.5.1.1. Le syndrome de Cornelia de Lange 95
4.5.1.2. Les autres cohésinopathies 95
4.5.2. Cancers 96
4.6. Le module ATPase de la Cohésine, acteur clé de son mode d'action 98
4.6.1. Composition du module ATPase de la Cohésine 98

4.6.1. Importance du module ATPase dans les fonctions de la Cohésine 100

OBJECTIFS DU TRAVAIL DE THESE 103

CHAPITRE 2 - MATERIELS ET METHODES 105

I. BIOLOGIE MOLECULAIRE - CLONAGES 105

1.1. Vecteurs d'expression bactériens de la série pET-MCN 105

1.2. Méthodes de clonages 107

1.2.1. Clonage par assemblage de Gibson 107

1.2.2. Assemblage des constructions des têtes ATPases des protéines SMC1A et SMC3 et de RAD21 par PCR nichée 108

1.2.2.1. Les têtes ATPases de SMC1A et SMC3 109

1.2.2.2. Les constructions de RAD21 110

1.2.3. Concaténations des vecteurs pour l'expression de complexes protéiques 111

1.2.4. Mutagénèse dirigée par réplication circulaire du plasmide ou « Rolling circle » 114

1.2.5. Récapitulatif des constructions des protéines utilisées pendant ma thèse 115

II. EXPRESSION ET PURIFICATION DE PROTEINES 116

2.1. Test d'expression et de purification à petite échelle 116

2.1.1. Expression 116

2.1.2. Purification 118

2.2. Expression et purification de protéines recombinantes à grande échelle 119

2.2.1. Production 119

2.2.2. Purification 120

2.2.2.1. Lyse des cellules 120

2.2.2.2. Chromatographie d'affinité 121

2.2.2.3. Purification par héparine 122

2.2.2.4. Purification par échangeuse d'ions 123

2.2.2.5. Chromatographie d'exclusion stérique 124

III. CARACTERISATION PHYSICO-CHIMIQUE 125

3.1. Diffusion dynamique de la lumière 125

3.2. SEC-MALS 126

3.3. Spectrométrie de masse 126

3.4. Titration calorimétrique isotherme 127

3.5. Expériences de réticulation in vitro (crosslinking) 128

3.6. Test enzymatique 130

3.7. Reconstitution du module ATPase 131

3.7.1. Par méthode biochimique : SEC 131

3.7.2. Par méthode biophysique : ultracentrifugation analytique 132

IV. CARACTERISATION STRUCTURALE 134

4.1. Essai de cristallisation pour la cristallographie par diffraction aux rayons X 134

4.2. Diffusion des rayons X aux petits angles 136

4.3. Cryo-microscopie électronique ... 136
4.3.1. Coloration négative ... 137
4.3.2. Cryo-EM et analyse des particules isolées ... 138
4.3.2.1. Préparation de l'échantillon protéique ... 138
4.3.2.2. Préparation des grilles et congélation ... 138
4.3.2.3. Criblage des grilles et collecte des données ... 140
4.3.2.4. Traitement et affinement des données ... 142
4.3.2.5. Construction et validation du modèle du module ATPase de la Cohésine ... 145
4.3.2.6. Optimisations nécessaires à l'amélioration de l'échantillon et à la qualité des données ... 145

CHAPITRE 3 - RESULTATS ... 147

I. ETUDE DU MODULE ATPASE, ACTEUR CLE DANS LE CYCLE ATPASE DE LA COHESINE ... 147
1.1. Introduction & description des travaux ... 147
1.2. Discussion ... 223
1.2.1. Caractéristique et rôle de l'engagement des têtes ATPases humaines de la Cohésine ... 223
1.2.1.1. L'engagement des têtes humaines nécessite la contribution des deux sites ATPases et est dépendant de l'ATP ... 223
1.2.1.2. L'engagement des têtes ATPases induit une flexibilité de la porte de sortie SMC3/RAD21 chez l'homme 225
1.2.1.3. La porte de sortie de l'ADN dans l'interface SMC3/RAD21 semble avoir un comportement différent selon les espèces ... 226
1.2.2. Proposition d'un modèle décrivant les premières étapes du cycle de la Cohésine humaine, allant de la liaison de l'ATP à l'interaction de l'ADN et NIPBL ... 228
1.2.2.1. La liaison de l'ATP maintient les têtes ATPases de SMC1A et SMC3 dans un état de repos, mais propre à chaque tête ... 228
1.2.2.2. L'engagement des têtes entraine la semi-activation des têtes et la flexibilité de l'interface SMC3/RAD21 ... 230
1.2.2.3. La fixation de l'ADN et de NIPBL entraine l'activation complète des têtes ATPases ... 232
II. RECONSTITUTION DU MODULE ATPASE ASSOCIE A STAG1/STAG2 ... 235
1. Choix des constructions de RAD21 et de STAG1 et STAG2 ... 235
1.2. Constructions de RAD21 ... 235
1.3. Constructions de STAG1 et STAG2 ... 236
2. Reconstitution des complexes quadripartites de la Cohésine ... 239
2.1. Première approche : purifications indépendantes du module ATPase et de STAG1/STAG2, puis assemblage du complexe quadripartite ... 239
2.1.1. Expression et purification de STAG1 ... 239
2.1.2. Expression et purification de STAG2 ... 241
2.1.3. Expression et purification du complexe SMC1A-HD/SMC3-HD/RAD21 ... 243
2.1.4. Deuxième approche : expression et purification des quatre partenaires de la Cohésine ... 248
2.2. Discussion ... 252

2.2.1. Optimisation du protocole d'expression et de purification du complexe quadripartite SMC1ACC/SMC3CC/RAD21/STAG1-1 252
2.2.2. Caractérisation future du complexe quadripartite de la Cohésine 253
2.2.3. Le système *E. coli* et les tests à petite échelle sont des outils efficaces pour tester les interactions entre les complexes tripartite/quadripartite de la Cohésine et ses autres protéines auxiliaires 254
2.2.4. STAG1-STAG2 : vers une recherche appliquée 255

CHAPITRE 4 - CONCLUSION ET DISCUSSION GENERALE 257

I. CONCLUSION 257
II. PERSPECTIVES 259

BIBLIOGRAPHIE 263

LISTE DES FIGURES ET TABLEAUX 290

I. Liste des figures 290
II. Liste des tableaux 292

Liste des abréviations

3C :	¨Chromosome Conformation Capture ¨
3D :	Tridimensionnel
4C :	¨Chromosome Conformation Capture on chip ¨
5C :	¨Chromosome Conformation Capture carbon copy ¨
A :	Adénosine
ABC :	¨ATP-Binding Casette¨
AFM :	Microscopie à force atomique
AMP.PNP :	Adénlyly-imidodiphosphate
ATP :	Adénosine triphosphate
AUC	Ultracentrifugation analytique
AUC-SV :	AUC à vitesse de sédimentation
AUC-SE :	AUC à l'équilibre de sédimentation
B. subtilis :	*Bacillus subtilis*
BER :	Réparation par excision de base
BMOE	Bismaléimidoéthane
bp :	Paire de bases
C :	Cytosine
C. elegans :	*Caenorhabditis elegans*
C. thermophilum :	*Chaetomium thermophilum*
CDK :	Kinase dépendante des cyclines
CdLS :	Syndrome de Cornelia de Lange
CHAPS :	(3-((3-cholamidopropyl) diméthylammonio)-1-propanesulfonate)
ChIP-chip :	¨Chromatin ImmunoPrecipitation on Chip¨
chromEMT :	¨ChromEM tomography¨
CIP :	¨Calf-intestinal alkaline phosphatase¨
cryoEM :	Cryo-microscopie électronique
CSR :	Commutation de classe
CTCF :	¨CCCTC-binding factor¨
CTF :	Fonction de transfert de contraste
D. melanogaster :	*Drosophila melanogaster*
DDM :	n-dodecyl-b-D-maltoside
DSL	Diffusion dynamique de la lumière
E. coli :	*Escherichia coli*
FISH :	Hybridation in situ en fluorescence
G :	Guanine
G. gallus :	*Gallus gallus*

G. stearothermophilus :	*Geobacillus stearothermophilus*
GST :	Glutathion-S-Transférase (GST)
HAWKs:	¨HEAT proteins associated with kleisins¨
HDAC8 :	Histone désacétylase 8
HEAT :	¨Huntington/EF3/PP2A/Tor1¨
HEPES :	Acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique
HR :	Recombination homologue
IPTG :	Isopropyl β-D-1-thiogalactopyranoside
Kb :	Kilobase
kDa :	Kilodalton
KITE :	¨Kleisine interacting winged-helix tandem elements¨
L. thermotolerans :	*Lachancea thermotolerans*
LB :	Luria-Bertani
Mb :	Mégabase
MCM :	Maintenance de minochromosome
MESG :	2-amino-6-mercapto-7-méthylpurine riboside
MMR :	Réparation des mésappariements
NaCl	Chlorure de sodium
NEB :	New England BioLabs
NER :	Réparation par excision de nucléotide
NHEJ	Jonction des extrémités non homologue
Nse :	¨Non-SMC Elements¨
OG :	Octyl-b-glucoside
PCR :	Réaction en chaîne par polymérase
PDS5 :	¨Precocious dissociation of sister protein¨
PIC :	cOmplete™ Protease Inhibitor
PNP :	Purine nucléoside phosphorylase
PTM :	Modification post-traductionnelle
RPE :	Résonance paramagnétique électrique
RSS :	Séquences signal de recombinaison
S. cerevisiae :	*Saccharomyces cerevisiae*
S. pombe :	*Schizosaccharomyces pombe*
SA ou STAG :	*¨Stromal antigene¨*
SAXS	Diffusion des rayons X aux petits angles
SDS :	Dodécylsulfate de sodium
SEC	Chromatographie d'exclusion de taille
SGO	SHUGOSHINE
SMC :	Maintenance structurale des chromosomes
T :	Thymine
TAD :	Domaines d'association topologique

TB :	Terrific Broth
Tris-HCl	Chlorhydrate de tris(hydroxyméthyl)méthylamine
TRP :	Répétition de tétratricopeptides
V(D)J :	Variables, diversité et jonction
WAPL :	¨Wings-apart like¨
WH :	Hélice ailée ou ¨winged helix¨
WHD :	Domaine à hélice ailée
X. laevis :	*Xenopus laevis*

Résumé de thèse

La chromatine, forme condensée et dynamique de l'ADN, régit l'accès à l'information génétique. La modulation de cette structure par les effecteurs épigénétiques orchestre l'ensemble des processus nucléaires tout au long du cycle cellulaire. Cette modulation épigénétique est donc clé pour l'homéostasie et le développement cellulaires, et a des conséquences directes sur la santé et les maladies. La dérégulation de ces mécanismes épigénétiques est impliquée dans diverses pathologies, comme dans les maladies auto-immunes, les maladies neurologiques, ainsi que dans de nombreux cancers. Il est donc essentiel de caractériser les mécanismes épigénétiques pour comprendre comment leur altération participe à l'initiation et au développement de pathologies afin de pouvoir à terme développer des approches thérapeutiques.

Parmi les effecteurs épigénétiques, la Cohésine humaine exerce un rôle majeur dans de nombreux processus nucléaires essentiels à la stabilité génomique. Elle participe notamment à la cohésion des chromatides sœurs, à la ségrégation des chromosomes, à la réplication et la réparation de l'ADN, à la régulation de la transcription, ainsi qu'à l'organisation tri-dimensionnelle du génome. La Cohésine remplit ses fonctions par sa capacité d'interagir avec l'ADN de manière topologique ou non-topologique, lui permettant ainsi de former des boucles de chromatine et des domaines d'association topologique (TADs).

La Cohésine humaine se compose de quatre protéines cœur : de deux protéines de Maintenance Structurelle des Chromosomes, SMC1A et SMC3, de la kleisine RAD21 et d'une protéine de répétitions HEAT (¨Huntington/EF3/PP2A/Tor1¨), STAG1 ou STAG2. Ce complexe quaternaire adopte une structure en anneau, lui permettant de capturer l'ADN.

Les protéines SMC1A et SMC3 possèdent chacune une région charnière, ainsi qu'un domaine globulaire ATPase, appelé tête ATPase, qui est formé par leurs extrémités N- et C-terminales. Ces deux domaines sont séparés par une région organisée en superhélice de 50 nm de long. SMC1A et SMC3 forment un hétérodimère par dimérisation de leurs régions charnières et interagissent avec les régions N- et C- terminales de RAD21 au niveau de leurs têtes ATPases. STAG1 ou STAG2 se fixe ensuite à RAD21. Les têtes ATPases de SMC1A et SMC3 peuvent également interagir l'une avec l'autre en présence d'ATP pour constituer le module ATPase de la Cohésine. La fixation et l'hydrolyse de l'ATP par ce module induisent des

changements de conformation de l'ensemble de la Cohésine essentiels à ses fonctions, notamment pour son interaction avec l'ADN et pour la formation de boucles de chromatine par extrusion.

L'association stable de la Cohésine avec l'ADN nécessite également le recrutement de nombreuses protéines auxiliaires et régulatrices. La Cohésine est chargée sur l'ADN grâce au complexe de recrutement NIPBL/MAU2. L'interaction Cohésine-ADN est ensuite stabilisée par le recrutement des protéines auxiliaires PDS5A/B, SORORINE, CTCF, ainsi que par l'acétylation de SMC3 par les acétyltransférases ESCO1 ou ESCO2. La dissociation de la Cohésine de l'ADN fait intervenir WAPL ou la Séparase, permettant ensuite le recyclage de la Cohésine, grâce à la désacétylation de SMC3 par l'histone désacétylase HDAC8. Ainsi, la Cohésine joue un rôle primordial dans l'intégrité du génome et contribue à la régulation de l'ensemble des processus nucléaires.

Des mutations observées sur les sous-unités cœurs de la Cohésine et sur ses protéines auxiliaires sont impliquées dans un grand nombre de cancers solides et leucémies, ainsi que dans certaines maladies neurodéveloppementales, appelées cohésinopathies.

Malgré l'importance de la Cohésine en physiopathologie humaine, nos connaissances sur la structure et sur la fonction de ce complexe humain restent incomplètes et ne permettent pas encore le développement de thérapies ciblant la Cohésine. Notamment, les bases moléculaires et structurales de l'activité ATPase de la Cohésine impliquée dans l'ensemble des fonctions de ce complexe restent à établir. Ceci permettra à terme de mieux appréhender les dysfonctionnements de la Cohésine dans les maladies et développer des stratégies thérapeutiques.

L'objectif de mon projet de thèse a été d'étudier la dynamique conformationnelle de la Cohésine humaine, et plus particulièrement le rôle de la liaison et de l'hydrolyse de l'ATP dans cette dynamique. Pour cela, j'ai reconstitué et caractérisé, par des approches biophysiques, biochimiques et structurales, le module ATPase de la Cohésine en présence d'ATP. Ce module ATPase est composé des têtes de SMC1A et SMC3, ainsi que des régions N- et C-terminales de RAD21. J'ai ensuite complété cette étude par l'association de la quatrième sous-unité de la Cohésine, STAG1, à ce module ATPase, afin de poser les bases pour la caractérisation du rôle de STAG1 sur ses partenaires et sur la dynamique de la Cohésine.

I. Reconstitution du module ATPase engagé de la Cohésine humaine

J'ai débuté mes travaux de thèse par la reconstitution du module ATPase humain, constitué des têtes ATPases de SMC1A et SMC3, ainsi que des régions N- et C-terminales de RAD21 qui se fixent sur ces domaines globulaires. La présence d'ATP est nécessaire pour la formation de ce module ATPase, qui reste extrêmement instable. De ce fait, il s'agit d'un complexe très difficile à étudier et pour lequel sa structure et sa dynamique restaient à caractériser précisément.

J'ai tout d'abord déterminé les conditions expérimentales permettant de stabiliser ce module ATPase, en testant de nombreuses conditions expérimentales avec notamment l'utilisation de plusieurs mutants de SMC1A et SMC3 ayant une activité enzymatique plus faible, l'utilisation de l'ATP ou d'analogues de l'ATP faiblement hydrolysables (ATP-γ-S, AMP-PNP), ainsi que l'utilisation de diverses conditions de purifications (tampon, etc…).

Cela m'a permis de trouver les conditions optimales dans lesquelles le module ATPase humain est stable. J'ai notamment utilisé les mutants SMC1A-E1157Q et SMC3-E1144Q, que j'ai exprimés et purifiés séparément puis mélangés en présence d'ATP. Ces mutations de glutamines en glutamates sont localisées dans les sites actifs de SMC1A et SMC3 et réduisent l'activité ATPase résiduelle du module ATPase, tout en permettant la fixation de l'ATP et donc la formation de ce complexe. Cette formation d'un complexe stable a été confirmée par chromatographie d'exclusion de taille et par ultracentrifugation analytique.

II. Caractérisation structurale du module ATPase engagé

J'ai ensuite entrepris la caractérisation structurale de ce module ATPase, en utilisant trois approches structurales : la diffusion de rayons X aux petits angles (SAXS), la cristallographie aux rayons X et la cryo-microscopie électronique (cryo-EM). La première m'a permis d'obtenir une enveloppe à basse résolution du complexe, révélant une première organisation 3D de ce complexe. La seconde approche a été rendue impossible par l'incapacité à obtenir des cristaux du complexe ayant une diffraction suffisante pour la résolution de structure. Enfin, la troisième approche par cryo-EM m'a permis d'obtenir une reconstruction précise du module ATPase humain à une résolution de 4,0 Å.

Mes données apportent des informations novatrices et importantes sur la dynamique du module ATPase humain, notamment d'une part au niveau des sites actifs de SMC1A et de SMC3 et d'autre part au niveau de la région en superhélice de SMC3, qui interagit avec l'extrémité N-terminale de RAD21 et présente une flexibilité, qui, à ce jour, n'a jamais été mise en évidence.

En effet, en comparant cette structure aux autres structures des têtes seules de SMC1A et SMC3 sous forme native ou liées à l'ATP-γ-S ou l'ADP, obtenues au sein de mon laboratoire d'accueil, ainsi qu'à la structure de la Cohésine liée à l'ADN et NIPBL, nous observons des différentes structurales nous permettant d'établir le cycle de liaison/hydrolyse de l'ATP par la Cohésine, ainsi que les changements opérés lors du chargement de la Cohésine sur l'ADN.

La fixation de l'ATP entraîne peu de changements conformationnels, les têtes ATPases de SMC1A et SMC3 restant inactives. Par contraste, l'engagement ATP-dépendant de ces deux têtes entraine la formation d'un complexe asymétrique : SMC1A reste inactif, mais SMC3 adopte une conformation productive. L'engagement des têtes ATPases n'est donc pas suffisant pour activer complètement le module ATPase. La fixation de l'ADN et de NIPBL entraîne des réarrangements structuraux, qui conduisent à l'alignement correct des deux têtes ATPases, ainsi qu'à la stabilisation des superhélices de SMC1A et de SMC3, permettant alors d'obtenir un complexe pleinement actif. Concernant la flexibilité de la superhélice de SMC3, elle se traduit par l'absence de densité observée dans la structure cryo-EM du module ATPase en comparaison aux autres structures. Cela a été confirmé par des expériences de réticulation *in vitro* (crosslinking). Cette flexibilité joue un rôle important dans l'activité et la dynamique de la Cohésine. En effet, l'engagement des têtes facilite les mouvements de la superhélice de SMC3, sans toutefois permettre d'atteindre la conformation active de SMC3, état atteint grâce à l'ancrage de NIPBL sur le module ATPase.

Mes résultats démontrent ainsi qu'il est possible de moduler la dynamique du module ATPase pour inhiber ou favoriser l'hydrolyse de l'ATP, permettant d'agir sur les changements de conformation de la Cohésine et donc sur ses fonctions. Toutefois, une question concernant la dernière sous-unité de la Cohésine, STAG1/STAG2, demeure. Les données récentes sur la Cohésine liée à NIPBL et à l'ADN montrent une grande proximité de STAG1 avec la superhélice de SMC3. Les protéines STAG1 et STAG2 pourraient potentiellement stabiliser le module ATPase.

III. Reconstitution du complexe quadripartite (module ATPase couplé à STAG1)

J'ai mené des essais de reconstitution à grande échelle du complexe quaternaire, en exprimant et purifiant d'une part le complexe SMCC1A/SMC3/RAD21 et d'autre part STAG1 seul, puis en mélangeant l'ensemble des protéines et suivant la formation du complexe par chromatographie par exclusion de taille. Différentes constructions de RAD21 ont été réalisées et comprennent les sites de liaison aux têtes ATPases de SMC1A et de SMC3, mais aussi le site de liaison à STAG1. Toutefois, du fait d'une forte dégradation de RAD21, cette approche n'a pas été concluante et j'ai alors choisi de co-exprimer et co-purifier les quatre partenaires. Les résultats à petite échelle ont semblé prometteurs, mais ceux à grande échelle sont plus mesurés. Pour la poursuite du projet, d'autres approches doivent être envisagées.

Mes résultats de thèse, combinés à ceux obtenus au sein de mon équipe d'accueil, apportent des informations novatrices et essentielles sur l'architecture et la dynamique du module ATPase de la Cohésine humaine. Cela nous permet d'avoir une meilleure compréhension de l'activité ATPase de la Cohésine et de sa modulation tout au long du chargement de l'ADN au sein de ce complexe. Toutefois, des études supplémentaires sont nécessaires afin de compléter nos connaissances sur ce complexe, son activité et ses fonctions.

Chapitre 1 – Introduction

I. Le cycle cellulaire eucaryote

Le cycle cellulaire est l'ensemble des étapes permettant d'aboutir à la formation de deux cellules filles à partir d'une seule cellule mère, par la duplication du génome et des organites de la cellule mère et la répartition précise et équitable de ce matériel entre ses deux cellules filles.

La notion que les nouvelles cellules proviennent de la division de cellules préexistantes date de la deuxième moitié du XIX^{e} siècle et a été proposée par Rudolph Virchow, plaçant ainsi la cellule comme unité fondamentale et autonome de la vie (Virchow, 1859). Par la suite, Schneider a découvert les étapes composant la division cellulaire, avec la transformation du noyau en structures en forme de bâtonnets. Cela a également été confirmé dans les années 1870, par Walther Flemming, qui, grâce à ses travaux sur des cellules de salamandre, a observé la réorganisation du matériel nucléaire en ¨fils¨, qu'il nomme chromatine, ainsi que la séparation de ces fils en deux groupes. Il a esquissé les étapes entrainant les arrangements de la chromatine, et les a regroupé en un terme, la mitose (Flemming, 1882). En 1888, Heinrich Wilhelm Gottfried Waldeyer-Hartz a renommé ces fils observés par Flemming, les chromosomes (Waldeyer, 1888). Les travaux de Flemming ont ainsi posé les bases pour les études du cycle cellulaire et plus particulièrement pour la mitose.

Il faudra ensuite attendre les années 1950-60 pour que les différentes étapes du cycle cellulaire soient déterminées, avec notamment la caractérisation de l'interphase par Alma Howard et Stephen Pelc, grâce à leurs travaux sur l'incorporation de phosphate^{32} sur des cellules de fève *Vicia faba* (Howard, 1953).

L'ensemble de ces travaux a permis de définir le cycle cellulaire eucaryote, qui se divise en deux étapes : l'interphase, où l'ADN est répliqué en deux copies identiques de chaque chromosome, appelé chromatides sœurs, et la mitose qui se traduit par la séparation des chromatides sœurs et leur répartition dans deux cellules filles, permettant ainsi de maintenir la stabilité de l'information génétique. Le cycle cellulaire est un processus finement régulé afin de garantir que les chromosomes incomplets ou endommagés ne soient pas transmis aux cellules filles.

1.1. L'interphase

L'interphase est caractérisée par la décondensation des chromosomes, organisées alors en territoire chromosomique. Ces territoires chromosomiques sont distribués de telle sorte que le noyau apparaît morphologiquement uniforme. L'interphase se décompose en 3 phases : la phase G1, la phase S et la phase G2 (Figure 1).

La phase G1 correspond à l'intervalle entre la mitose et l'initiation de la réplication de l'ADN et se caractérise par un accroissement du volume cellulaire et par l'accumulation d'éléments constitutifs de l'ADN chromosomique et protéine associées. Un point de contrôle se produit entre cette phase et la phase suivante, la phase S. Si les conditions de croissance ne sont pas propices, la cellule entre dans une phase quiescente, appelée phase G0. Selon les espèces, les facteurs contrôlant ce passage sont différents. Les cellules animales sont notamment régulées par des facteurs de croissance extracellulaires, ainsi que par la présence de dommages à l'ADN et leurs réparations. Ainsi, lors de la phase G1 la cellule croît, exposée à un ensemble de signaux et son ADN, si nécessaire, est réparé. Si l'ensemble de ces points de contrôle sont satisfaisants, la cellule progresse vers la phase S. (Cooper and Hausman, 2000).

La phase S (phase de synthèse de l'ADN) est marquée par la réplication de l'ADN, entrainant la formation, pour chaque chromosome, de deux chromatides sœurs qui ont un contenu génétique identique. Les centrioles, éléments clés pour la ségrégation chromosomique, sont également dupliqués lors de cette phase. Les chromatides sœurs sont maintenues à proximité l'une de l'autre, grâce à la Cohésine, qui par sa structure en anneau, peut capturer l'ADN (Uhlmann and Nasmyth, 1998). Il s'agit de la cohésion des chromatides sœurs, phénomène crucial pour l'organisation de l'information génétique jusqu'à la mitose. La Cohésine et son rôle au sein du cycle cellulaire seront développés dans la partie IV. Le complexe Cohésine.

Lors de la phase G2, les chromatides sœurs sont attachées fermement l'une à l'autre, grâce à la Cohésine, et ne seront séparées qu'en mitose (Haarhuis et al., 2014). La phase G2 est également impliquée dans l'initiation des processus impliqués dans la division cellulaire. Cette phase comprend un point de contrôle, appelé point de contrôle des dommages de l'ADN, empêchant le passage vers la mitose, tant que la réplication de l'ADN n'est pas achevée et que les lésions de l'ADN ne sont pas réparées (Cooper and Hausman, 2000; Yang and Sheridan, 2014).

Ainsi l'interphase se compose d'une progression étape par étape avec plusieurs points de contrôle, assurant une synthèse d'ADN de bonne qualité. De nombreuses protéines participent à cette régulation, telles que les kinases dépendantes des cyclines (CDK), les cyclines et les inhibiteurs de CDK, qui permettent de réguler positivement ou négativement le cycle cellulaire (Brooker and Berkowitz, 2014; Verhest and Heimann, 2008).

1.2. La mitose

Lors de la mitose, la cellule se divise donnant naissance à deux cellules filles identiques. Au cours de cette étape, l'ADN est condensé, séparé et distribué dans les deux cellules filles. Il s'agit de la ségrégation chromosomique. La mitose se décompose en plusieurs étapes : la prophase, la prométaphase, la métaphase, l'anaphase, la télophase et la cytokinèse, aboutissant à la division cellulaire (Figure 1) (Lian and Chircop, 2016; Scholey et al., 2003).

La prophase est caractérisée par la condensation de la chromatine dupliquée en chromosomes. Ce phénomène nécessite l'intervention de la Condensine, un complexe appartenant à la même famille que la Cohésine. Ses fonctions seront abordées dans la partie III. Les complexes SMC : organisateurs de la chromatine. Concernant la Cohésine, elle se dissocie des bras des chromosomes, excepté au niveau des centromères, où elle poursuit le maintien des deux chromatides sœurs jusqu'à l'anaphase (Waizenegger et al., 2000). Les chromosomes adoptent alors leur structure caractéristique en X. Lors de cette phase, les centrosomes, où émergent les microtubules, débutent également leur déplacement vers les pôles opposés de la cellule, initiant la formation du fuseau mitotique (Lian and Chircop, 2016; Scholey et al., 2003).

Lors de la prométaphase, l'enveloppe nucléaire se rompt, exposant ainsi l'ensemble du contenu du noyau au cytoplasme. Le fuseau mitotique se forme, avec la fixation des microtubules au niveau des kinétochores des chromosomes, permettant d'orienter ces derniers, afin que toutes les chromatides sœurs soient attachées par leurs kinétochores aux microtubules des pôles opposés. Il s'agit de la bi-orientation des chromosomes (Lian and Chircop, 2016; Scholey et al., 2003).

La métaphase est définie par l'alignement des chromosomes, au niveau de l'équateur de la cellule, appelé ¨plaque de métaphase¨, à mi-chemin entre les deux pôles de la cellule et perpendiculaire à l'axe du fuseau mitotique. La cohésion des chromatides sœurs est encore

maintenue au niveau des centromères par la Cohésine, les chromosomes étant donc toujours composés de deux chromatides sœurs. Un point de contrôle, appelé ¨point de contrôle de l'assemblage du fuseau¨ se produit à cette étape. Il permet de s'assurer que la bi-orientation de tous les chromosomes, ainsi que leur déplacement sur la plaque de métaphase soient appropriés, avant le passage à la prochaine étape (Lian and Chircop, 2016; Scholey et al., 2003).

L'anaphase est marquée par la séparation des chromatides sœurs, grâce au clivage de la Cohésine au niveau de leurs centromères (Uhlmann et al., 1999; Waizenegger et al., 2000). Chaque chromatide migre vers le centrosome, situé à l'un des pôles de la cellule, du fait du raccourcissement des fibres du réseau mitotique. Un allongement de la cellule est également observé.

Lors de la télophase, les chromosomes sont parvenus aux pôles de la cellule et se décondensent, l'enveloppe nucléaire se reforme autour de chaque ensemble de chromosome et le fuseau mitotique est désassemblé (Lian and Chircop, 2016; Scholey et al., 2003).

Le cycle cellulaire s'achève par la cytokinèse, déclenchée par la construction de la membrane cellulaire et caractérisée par la séparation physique des composants cytoplasmique, permettant la division entière de la cellule mère en deux cellules filles. Les chromosomes se refondent à nouveau dans un réseau chromatinien morphologiquement uniforme (Scholey et al., 2003; Verhest and Heimann, 2008; Lian and Chircop, 2016).

1.3. Variations entre les eucaryotes inférieurs et les mammifères

Afin de transmettre leur matériel génétique à la génération suivante, les cellules présentent un cycle cellulaire divisé en plusieurs étapes détaillées dans la partie ci-dessus. Toutefois, le cycle tel qu'il est décrit est le cycle observé chez les cellules animales. Certaines différences sont observées entre les eucaryotes supérieurs et les eucaryotes inférieurs, comme les levures *S. cerevisiae* et *S. pombe*, dont le cycle cellulaire entre ces deux espèces diverge. Notamment, *S. pombe* se divise par fission, c'est-à-dire qu'un septum se forme au centre de la cellule mère, la divisant en deux cellules filles, alors que la division de *S. cerevisiae* se fait par bourgeonnement, où une grande cellule mère génère un bourgeon fille (Oliva et al., 2005).

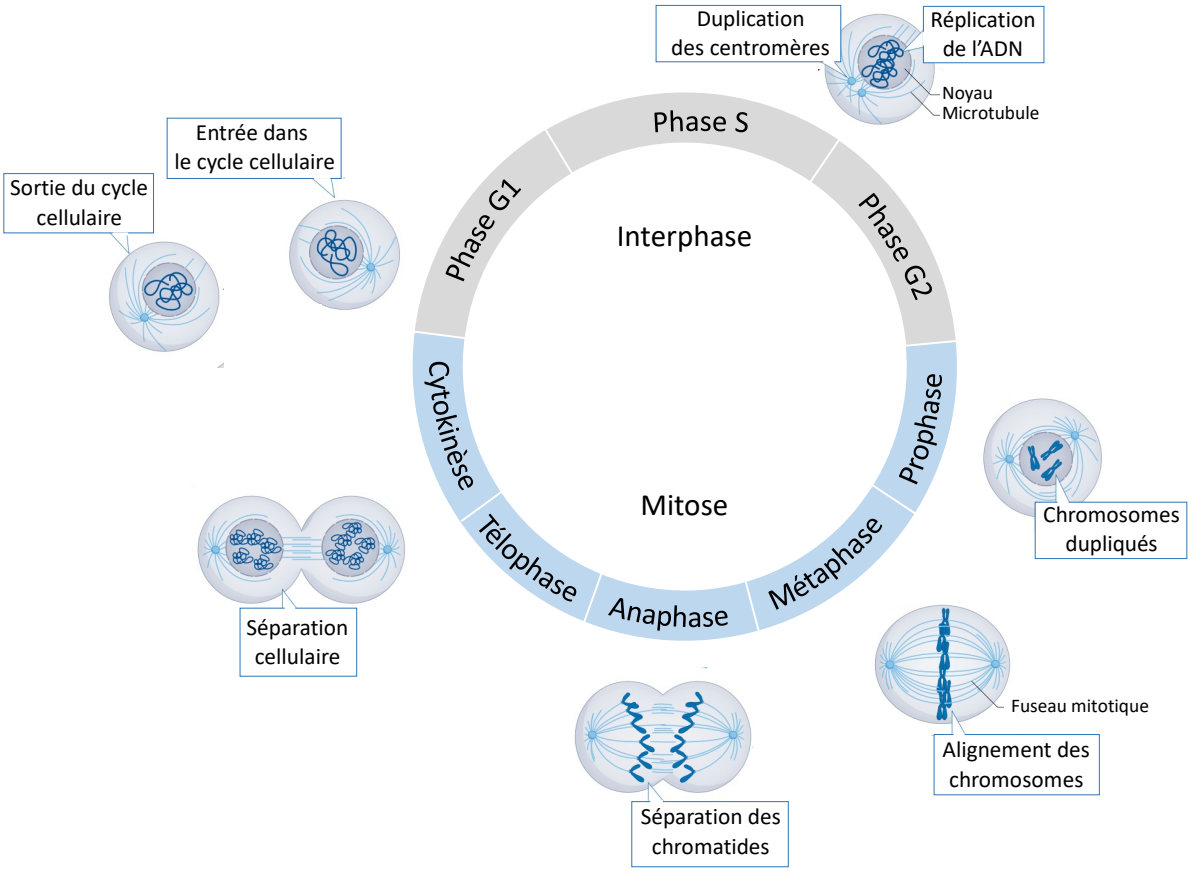


Figure 1 : Représentation schématique des différents étapes du cycle cellulaire eucaryote.
Le cycle cellulaire se divise en deux étapes principales : l'interphase au cours de laquelle l'ADN est répliqué et la mitose caractérisée par la ségrégation chromosomique. Enfin, la cytokinèse achève ce cycle et permet la formation de deux cellules filles identiques. Adapté de (Matthews et al., 2022).

Une autre différence notoire est la durée nécessaire à la division cellulaire. En moyenne, 90 min sont nécessaires pour que des cellules de *S. cerevisiae* se divisent et 2 à 3 h pour des cellule de S. *pombe*, alors qu'en moyenne 24h sont nécessaire pour des cellules humaines, avec environ une phase G1 de 11 h, une phase S de 8 h, une phase G2 de 4 h et une phase de mitose d'1 h (Cooper and Hausman, 2000). La régulation du cycle cellulaire par les CDK, les kinases et les phosphatases s'est complexifié au cours de l'évolution, avec notamment la présence de plusieurs homologues ayant des spécificités fonctionnelles propres (Harashima et al., 2013)

De plus, il existe plusieurs types de mitose. Celle décrite dans la partie précédente avec la rupture de l'enveloppe nucléaire est appelée mitose ouverte et est observée chez les eucaryotes supérieurs. Au contraire, de nombreux eucaryotes inférieurs, notamment *S. cerevisiae* et *S. pombe*, subissent une mitose dite fermée, où l'enveloppe nucléaire est

maintenue. La mitose est alors réalisée dans le noyau (De Souza and Osmani, 2007; Boettcher and Barral, 2013). De plus, dans certaines circonstances, *S. pombe* peut se diviser en absence de fuseau mitotique, ce processus est alors appelé fission nucléaire (Castagnetti et al., 2010).

Une autre différence s'observe également lors de la mitose entre les mammifères et *S. cerevisiae* : une partie de la Cohésine de mammifère peut être dissociée des chromatides sœurs dès la prophase, alors que la Cohésine de levure est associée aux chromatides sœurs jusqu'à l'anaphase, où elle est clivée (Uhlmann et al., 1999; Sumara et al., 2000).

Ainsi, il existe certaines disparités du cycle cellulaire au sein des eucaryotes, autant dans l'interphase que dans la mitose. Pourtant, les comportements morphologiques de l'ADN sont similaires avec une alternance des états condensé/décondensé tout au long du cycle cellulaire. Cette organisation de l'ADN, essentielle au maintien et à la transmission de l'information génétique, est un processus très complexe impliquant plusieurs degrés d'organisation, qui varient selon les différentes étapes du cycle cellulaire.

II. Architecture du génome eucaryote : de la double hélice d'ADN au chromosome

2.1. Découverte de la double hélice d'ADN

L'ADN a été découvert en 1868 par Friedrich Mischer lors de ses travaux sur la composition chimique des leucocytes. Il est parvenu à isoler cette substance jusqu'alors inconnue, riche en phosphate, localisée dans le noyaux et qui n'est ni une protéine, ni un lipide et l'a appelé nucléine (Dahm, 2008).

Le vingtième siècle a également apporté de nombreuses avancées majeures sur l'ADN. D'une part, dans les années 1920, le désoxyribonucléotide, l'unité de base de l'ADN, a été isolé et purifié par Phoebus Levene (Levene, 1919). Il se compose d'un sucre pentose, le désoxyribose, d'un groupement phosphate et d'une base azotée. Il existe 4 bases azotées : adénosine (A), thymine (T), guanine (G) et cytosine (C). Chargaff a démontré par la suite que la composition en bases azotées de l'ADN diffère d'un organisme à l'autre, mais qu'il y a autant d'adénosines que de thymines et autant de cytosines que de guanines (Chargaff, 1950).

D'autre part, les travaux de Oswald T. Avery, Colin MacLeod et Maclyn McCarty ont permis d'identifier le rôle de l'ADN en tant que support héréditaire de l'information génétique, confirmé quelques années plus tard par Alfred Hershey et Martha Chase lors de leurs travaux sur les phages T2 (Avery et al., 1944; Hershey and Chase, 1952). Ils ont marqué radioactivement l'ADN ou les protéines des phages et ont montré que le transfert d'ADN dans les bactéries infectées est nécessaire à la croissance des phages, tandis que les protéines ne le sont pas (Hershey and Chase, 1952).

En 1953, la première structure tridimensionnelle (3D) de l'ADN est résolue par James Watson et Francis Crick grâce aux clichées de cristaux d'ADN obtenus par diffraction des rayons X par Rosalind Franklin et Maurice Wilkins. Il s'agit ¨d'une structure à deux chaînes hélicoïdales enroulées chacune autour du même axe¨ (Watson and Crick, 1953), adoptant une structure en hélice droite avec des brins antiparallèles de 2 nm de diamètre, appelée conformation de type B. Les bases A sont appariées avec les bases T et les bases G avec les bases C, grâce à la formation de liaisons hydrogènes au centre de l'hélice. Les groupements phosphates sont disposés vers l'extérieur. Deux nucléotides adjacents sont séparés de 3,4 nm

et interagissent l'un avec l'autre via des liaisons phosphodiesters. Dix paires de bases (pb) sont nécessaires par tour de la double hélice d'ADN (Watson and Crick, 1953).

Les années 1960 marquent ensuite la découverte du code génétique, un système d'information universel et redondant déchiffré par Crick, Khorana et Nirenberg, établissant ainsi le lien entre l'ADN, l'ARN messager, l'ARNt et les protéines. Leurs travaux ont permis d'identifier les codons de nucléotides associés aux acides aminées (Crick, 1962; Nirenberg, 1963; Khorana, 1968). Les progrès techniques majeurs de ces dernières décennies ont permis de réaliser le premier séquençage du génome humain, qui a été débuté en 1990 et a été finalisé en 2004 (International Human Genome Sequencing Consortium, 2004). Depuis, de nombreux autres séquençages ont été menés afin d'obtenir des informations sur des populations multi-ethniques et aboutir à une cartographie génétique globale du génome humain (Durbin et al., 2010; Fakhro et al., 2016; Wall et al., 2019).

2.2. L'organisation de la chromatine au cours du cycle cellulaire

Le génome humain contient environ 3 milliards de pb, ce qui représente 2 m d'ADN linéaire réparti dans les 23 paires de chromosomes de chaque cellule. Cet ADN est organisé sous forme condensée, appelée chromatine, et est hébergé dans le noyau cellulaire, dont le diamètre est de l'ordre de 10 μm. Afin de pouvoir être maintenu dans cet espace très restreint, l'ADN est compacté de manière hiérarchique à différentes échelles, mais doit rester accessible pour maintenir l'ensemble de ses transactions (réplication, réparation, transcription). Cette organisation à la fois spatiale et temporelle a ainsi un impact direct sur les fonctions génomiques, notamment sur la régulation et l'activité des gènes et nécessite une régulation très fine et orchestrée tout au long du cycle cellulaire eucaryote.

La technologie de la capture de la conformation des chromosomes 3C et ses dérivées (4C, 5C, Hi-C, …), ainsi que le développement de la microscopie à haute résolution ont permis une avancée majeure dans la biologie de la chromatine, notamment pour la détermination des différents niveaux d'organisation, ainsi que du rôle de la topologie du génome dans l'ensemble des processus nucléaires. La chromatine est organisée avec différents degrés de condensation en fonction du cycle cellulaire.

2.2.1. La structure de la chromatine lors de l'interphase

Lors de l'interphase, l'ADN s'enroule autour d'octamères d'histone formant le nucléosome, qui est l'unité de base de la chromatine. Les nucléosomes forment ainsi la fibre de chromatine de 11 nm. Cette dernière s'organise ensuite en petites boucles de chromatine, appelées nanodomaines, puis en domaines de chromatine, tels que les domaines d'association topologique (TADs). Ces domaines sont ensuite dispersés dans des compartiments de chromatine, qui occupent différents territoires chromosomiques (Figure 2).

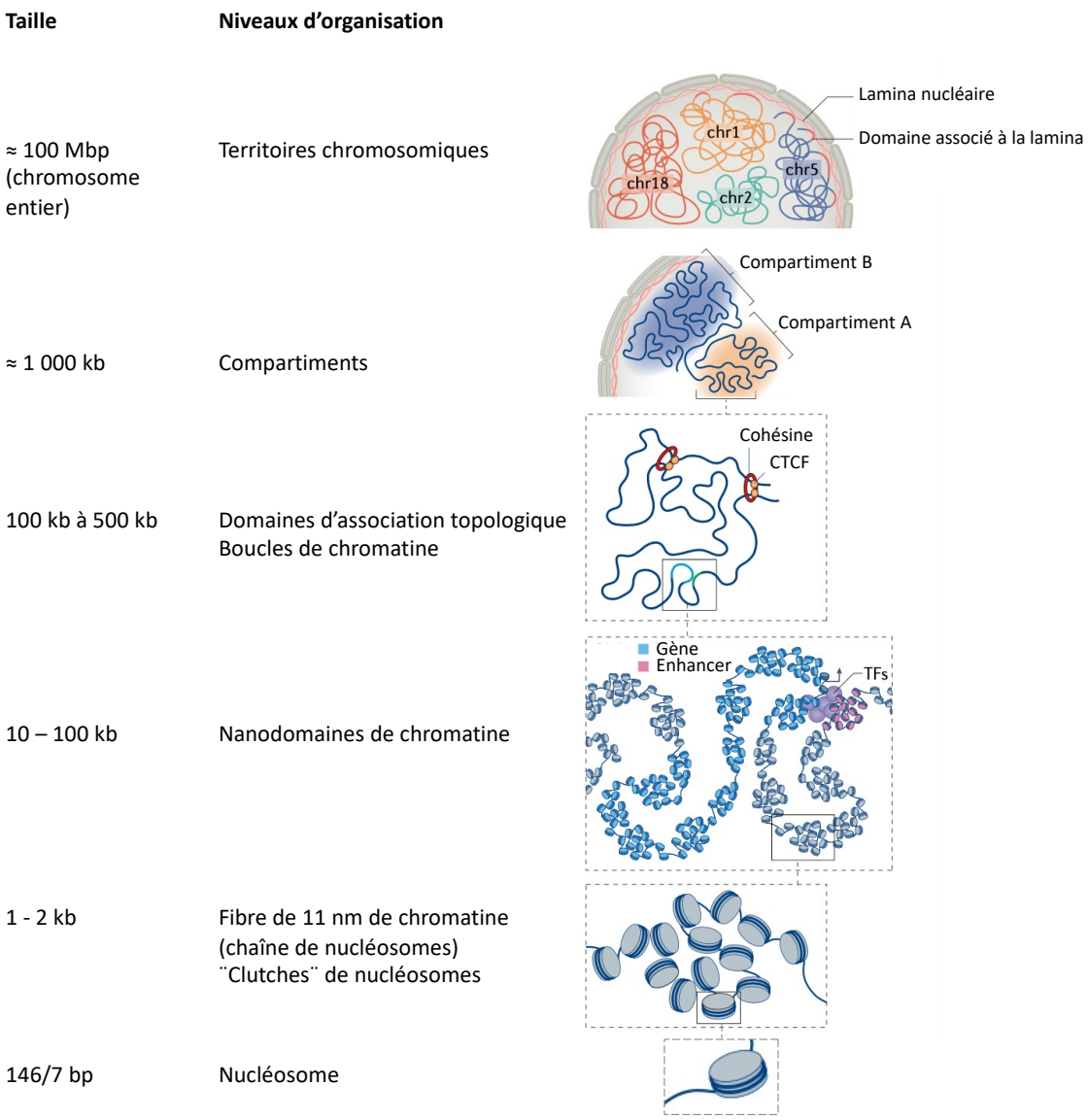


Figure 2 : Les différents niveaux d'organisation de la chromatine dans les cellules eucaryotes en interphase.
TFs : facteurs de transcription. Adapté de (Jerkovic´ and Cavalli, 2021).

2.2.1.1. Du nucléosome aux nanodomaines de chromatine

Le nucléosome, l'unité répétitive de base de la chromatine, est un complexe nucléo-protéique de 10 nm de diamètre formé par l'enroulement de 146/147 pb d'ADN double brin autour d'un octamère d'histones (Olins Ada L. and Olins Donald E., 1974; Kornberg Roger D., 1974). La structure du nucléosome a été résolue à 2,8 Å, puis à 1,9 Å dans les années 2000 (Luger et al., 1997; Davey et al., 2002) (Figure 3A). L'octamère d'histones est constitué de deux hétérodimères H2A/H2B et d'un tétramère $(H3/H4)_2$. Les histones sont des protéines basiques de 15 kDa ayant un domaine central globulaire et des extrémités N- et C-terminales flexibles, appelées queues d'histone.

La formation et le positionnement des nucléosomes sont importants pour la régulation des processus génomiques, en médiant l'accessibilité à l'ADN et donc le recrutement et les interactions entre les protéines régulatrices et l'ADN. Cette régulation est contrôlée par plusieurs mécanismes épigénétiques, tels que la méthylation de l'ADN, les modifications post-traductionnelles (PTM) des histones, l'insertion de variants d'histones H2A, H2B et H3 spécifiques, comme H2A.X, H2A.Z, H3.3 et H2B.W, ainsi que le recrutement de facteurs de transcription et de remodelage de la chromatine (Strahl and Allis, 2000; Struhl and Segal, 2013).

La formation des nucléosomes permet de réduire de sept fois la longueur de l'ADN (Iashina et al., 2017). De plus, les nucléosomes sont liés les uns aux autres par des brins d'ADN de liaison de 20-75 bp, qui peuvent fixer l'histone H1 (Woodcock and Ghosh, 2010). Ces nucléosomes se regroupent en une chaîne de nucléosomes, la fibre de chromatine de 11 nm, adoptant une structure en collier de perles (Figure 3B).

Le niveau supérieur d'organisation de la fibre de chromatine reste controversé. En effet, pendant des décennies, un modèle de fibre de chromatine de 30 nm observée lors d'expérience *in vitro* de (cryo-)microscopie électronique faisait foi, mais est remis en question depuis quelques années, n'ayant jamais pu être confirmé *in vivo* (Finch J T and Klug A, 1976; Song Feng et al., 2014). Par la suite, des expériences d'imagerie à super-résolution et de microscopie électronique ont permis de démontrer que la fibre de chromatine présente un polymorphisme et qu'elle peut adopter plusieurs formes de l'ordre du kilobase avec des diamètres allant de 5 à 24 nm. Il s'agit de ¨clutches¨ de nucléosomes (Ricci et al., 2015; Ou et

al., 2017; Portillo-Ledesma et al., 2021) (Figure 3C). Ces derniers sont ensuite organisés en nanodomaines de chromatine d'une taille avoisinant les 100 kilobases (kb) (Szabo et al., 2020).

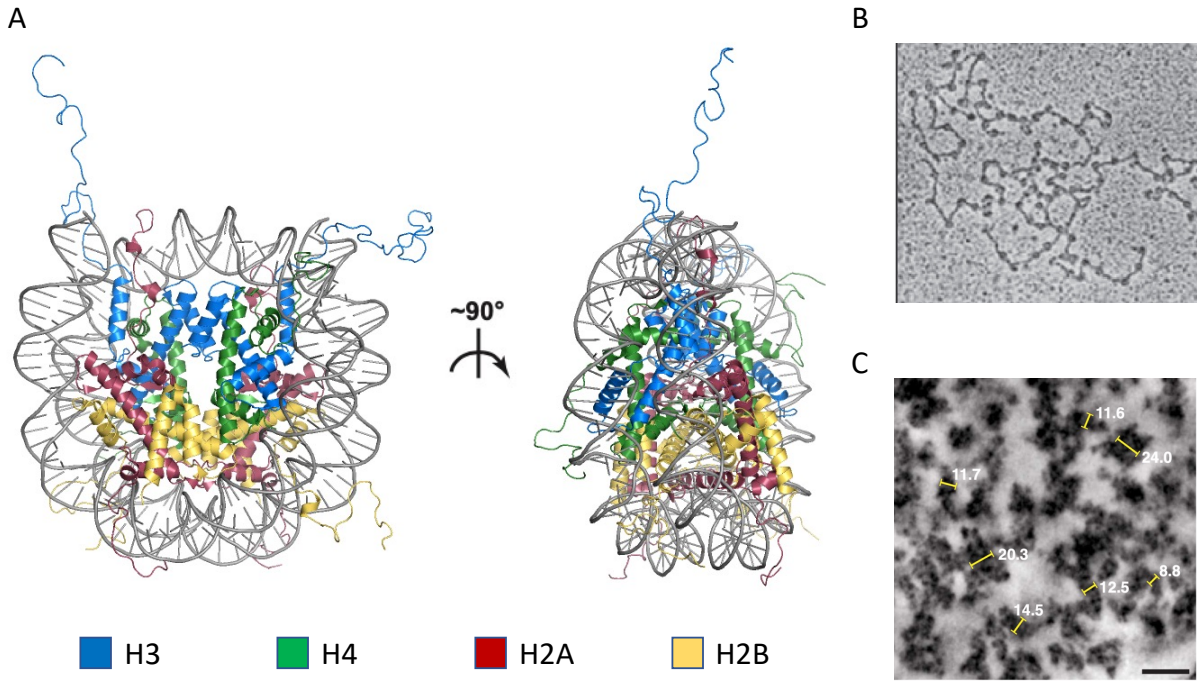


Figure 3 : Nucléosome et fibre de chromatine.
A. Structure cristallographique d'un nucléosome, composé de 146/147 bp d'ADN entourant un octamère d'histone (PDB : 1KX5).
B. Image de microscopie électronique d'une fibre de chromatine de 11 nm, caractérisée par une structure en collier de perles. Adapté de (Allan et al., 1982).
C. Coupe tomographique de cellules épithéliales humaines des petites voies respiratoires, issue d'une expérience de chromEMT, technique combinant la tomographie par microscopie électronique et une méthode de marquage par DRAQ5, permettant d'augmenter le contraste de l'ADN. Cette technique permet de visualiser les ¨clutches¨, fibres de chromatine de 5 à 24 nm. Les mesures de chromatine ont été faites manuellement. Barre d'échelle : 50 nm. Adapté de (Ou et al., 2017).

2.2.1.2. Les boucles de chromatine

Les nanodomaines de chromatine sont ensuite compactés en boucles de chromatine, apportant un cadre structurel plus stable, consolidées notamment par des protéines impliquées dans l'organisation 3D du génome, ainsi que par des facteurs de transcription. Ces boucles présentent des tailles différentes et peuvent atteindre des dizaines de kilobases. Grâce à la technique de Hi-C *in situ*, Rao et al ont établi un catalogue des boucles de chromatine à l'échelle du génome, comprenant environ 10 000 boucles dans des cellules lymphoblastoïdes (Rao et al., 2014). Ces dernières semblent conservées entre les types

cellulaires et entre la souris et l'homme. Ce sont des éléments structuraux clés dans les transactions génomiques, telles que la recombinaison de l'ADN et la régulation de l'expression des gènes (Greenwald et al., 2019). Ces boucles permettent notamment le rapprochement physique d'éléments régulateurs, tels que les enhancers et les promoteurs, qui sont distants généralement de 10 à plusieurs centaines de kb (Dekker and Misteli, 2015). Les mécanismes impliqués dans la formation des boucles font appel à la protéine CTCF et à la Cohésine et seront abordés plus en détail dans la partie IV. Le complexe Cohésine.

2.2.1.3. Domaines d'association topologique

2.2.1.3.1. Chez les mammifères

Les boucles de chromatine sont elles-mêmes organisées en domaines topologiquement associés de taille comprise entre 0,2 et 1 mégabase (Mb). Ce sont des domaines auto-associés et divisés par des régions de liaison. La découverte des TADs remonte à 2012 grâce aux développements des techniques de capture de conformation de chromosomes (Dixon et al., 2012; Nora et al., 2012). Il s'agit de domaines conservés à travers différents types cellulaires et espèces, notamment chez les vertébrés (Dixon et al., 2012; Li et al., 2022; Sauerwald and Kingsford, 2018). En utilisant la technique d'hybridation *in situ* en fluorescence (FISH) à haute résolution en cellule unique, Bintu et al ont déterminé les structures de TADs, et ont montré que ces domaines adoptent une conformation plutôt globulaire et présentent une séparation nette entre les domaines voisins (Bintu et al., 2018) (Figure 4A). Les interactions sont plus fréquentes au sein d'un même TAD qu'entre deux TADs différents et suggèrent une co-régulation entre les gênes d'un même TAD (Hsieh et al., 2020; Nora et al., 2012; Pope et al., 2014).

Tout comme les boucles de chromatine, les TADs participent à la régulation de l'expression des gènes, en assurant ou en empêchant des contacts entre les éléments régulateurs et les gènes cibles (Nora et al., 2012; Rao et al., 2014; Schoenfelder and Fraser, 2019). Les TADs sont divisés en deux types et peuvent se chevaucher. D'une part, nous retrouvons les TADs formés par des boucles de chromatine induites par le facteur de liaison CCCTC (CTCF) et par la Cohésine, d'autre part, les TADs dépourvus de CTCF et de Cohésine, appelés domaines compartimentaux. Des expériences de Hi-C sur des cellules dépourvues de

CTCF ou de RAD21, un des composants de la Cohésine, montrent une disparition des TADs CTCF-Cohésine dépendants, alors que les domaines compartimentaux sont maintenus (Nichols and Corces, 2021; Rao et al., 2017, 2014). Les mécanismes sous-jacents à la formation et au maintien des domaines compartimentaux ne sont pas connus, tandis que les mécanismes impliqués dans la formation et au maintien des TADs dépendants de CTCF et de la Cohésine seront abordés plus en détail dans la partie IV. Le complexe Cohésine.

2.2.1.3.2. Différences entre les mammifères, la drosophile et les levures

Les TADs ont également été identifiés chez *D. melanogaster*. Ils présentent toutefois une taille inférieure (taille médiane d'environ 100 kb), comparée à celle observée pour les mammifères (entre 0,1 et 2Mb) (Sexton et al., 2012). De plus, les TADs de *D. melanogaster* ne sont pas médiés par la protéine CTCF et la Cohésine, comme chez les mammifères (Rowley et al., 2017). Ils sont donc comparables au second type de TADs observé chez les mammifères, les domaines compartimentaux. L'organisation de ces TADs est directement induite par la ségrégation du génome en régions actives et silencieuses, correspondant aux domaines compartimentaux (Rowley et al., 2017)

De plus, une organisation semblable aux TADs de mammifères a également été observée chez de nombreux organismes modèles eucaryotes inférieures, tels que *S. pombe* et *S. cerevisiae.* Toutefois, ces organismes ne présentent pas d'homologues au facteur CTCF, la formation de ces TADs faisant donc appel à d'autres mécanismes que ceux retrouvés chez les mammifères (Mizuguchi et al., 2014; Hsieh et al., 2015).

Ainsi, les TADs se retrouvent chez de nombreuses espèces eucaryotes, mais présentent des mécanismes d'organisation sous-jacents différents et une complexité supérieure chez les mammifères.

2.2.1.4. Les compartiments A/B

Au-delà des TADs, le prochain niveau d'échelle distingue deux régions de plusieurs Mb : (i) le grand compartiment A, aussi appelé compartiment ouvert ou euchromatine, occupant les régions internes du noyau et comprenant des gènes transcrits de manière active et (ii) le compartiment B, aussi appelé compartiment fermé ou hétérochromatine, situé dans

la périphérie du noyau et dont les gènes sont inactifs (Lieberman-Aiden et al., 2009). Ces compartiments ont été observés pour la première fois en microscopie optique par Heitz en 1928 (Heitz, 1928).

Le compartiment A se compose de régions riches en gènes avec des marques d'amplification, telles que la mono-méthylation de la lysine 4 ou l'acétylation de la lysine 27 sur l'histone 3 (H3K4me1, H3K27ac) et des marques associées à la transcription, comme la tri-méthylation de la lysine 36 de l'histone 3 (H3K36me3) (Ouimette et al., 2019). Quant au compartiment B, il est caractérisé par des régions portant des marques répressives, telles que la tri-méthylation de la lysine 27 de l'histone 3 (H3K27me3) (Ouimette et al., 2019).

Les données de Hi-C obtenues par Rao et al, permettent de compléter cette classification en deux compartiments, qui peuvent être à leur tour divisés en plusieurs sous-compartiments selon certaines caractéristiques de la chromatine, telles que les modifications post-traductionnelles des histones, l'expression faible ou forte des gènes, la chronologie de la réplication (Rao et al., 2014). Récemment, Nichols et al ont suggéré une classification en trois compartiments distincts minimums chez l'homme, avec un compartiment actif et deux compartiments inactifs. Cette classification se base sur les PTMs des histones H3K27ac, H3K27me3 et H3K9me3, associées respectivement à la transcription, à l'extinction transcriptionnelle et aux séquences pour lesquelles la transcription est inactive (Nichols and Corces, 2021) .

La compartimentation de la chromatine semble ainsi bien plus complexe que la répartition binaire A/B et de nombreuses études sont requises pour mieux la caractériser. Les bases moléculaires impliquées dans la formation et le maintien des compartiments sont très peu connues. Il est toutefois intéressant de noter qu'à l'opposé des TADs, ces processus ne dépendent pas de la Cohésine et de CTCF (Nora et al., 2017; Rao et al., 2017).

2.2.1.5. Territoire chromosomique

Le dernier niveau d'organisation de la chromatine est le territoire chromosomique. Il s'agit de l'espace nucléaire occupé par un chromosome lors de l'interphase. L'idée d'une telle organisation dans une cellule en interphase a été émise vers les années 1900 par Carl Rabl (1887) et Theodor Boveri (1909) et est maintenant admise par l'ensemble de la communauté scientifique (Cremer and Cremer, 2010). Ces territoires ont pu être visualisés au sein du noyau

cellulaire par la technique de ReFISH, qui combine la technique de FISH et des sondes multiplex marquées avec plusieurs fluorochromes (Bolzer et al., 2005) (Figure 4B). En moyenne, les territoires chromosomiques de cellules de mammifères ont un diamètre de 2 à 3 µm et sont distribués de manière radiale et non aléatoire au sein du noyau. Les études sur la localisation des chromosomes 18 et 19 par des expériences de FISH ont montré que la localisation de ces territoires, ainsi que leur surface est fonction de leur nombre de gènes : plus un territoire chromosomique est riche en gènes, plus il est situé proche du centre nucléaire et inversement, un territoire chromosomique pauvre en gènes est situé dans une région plus périphérique du noyau (Croft et al., 1999; Cremer et al., 2001; Bolzer et al., 2005).

Ainsi la chromatine se définit par une organisation à plusieurs échelles, permettant une régulation fine de l'accès à l'information génétique et a donc un impact sur l'ensemble des processus biologiques opérés lors de l'interphase.

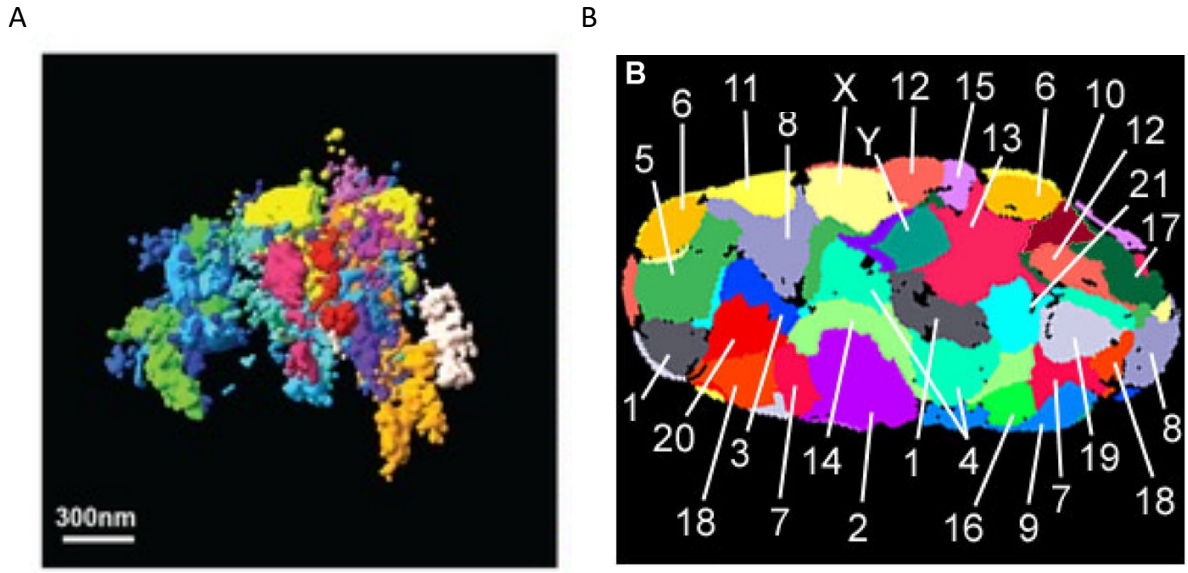


Figure 4 : Représentation des TADs (A) et des territoires chromosomiques (B)
A. Image 3D à super-résolution de la région chromatidienne de 1,2 Mb du chromosome 21 (Chr21:28Mb-29,2Mb) par la technique 3D STORM. Les TADs, représentés par des couleurs différentes, adoptent des structures globulaires et des limites franches. Adapté de (Bintu et al., 2018).
B. Représentation des paires de chromosomes dans le noyau d'un fibroblaste humain, obtenu par FISH (après classification avec le programme goldFISH). Adapté de (Bolzer et al., 2005).

2.2.2. Conformation des chromosomes dans les cellules mitotiques

A l'entrée en mitose, des changements majeurs dans la structure des chromosomes sont opérés. En effet, au contraire de l'interphase, où les chromosomes ne sont pas

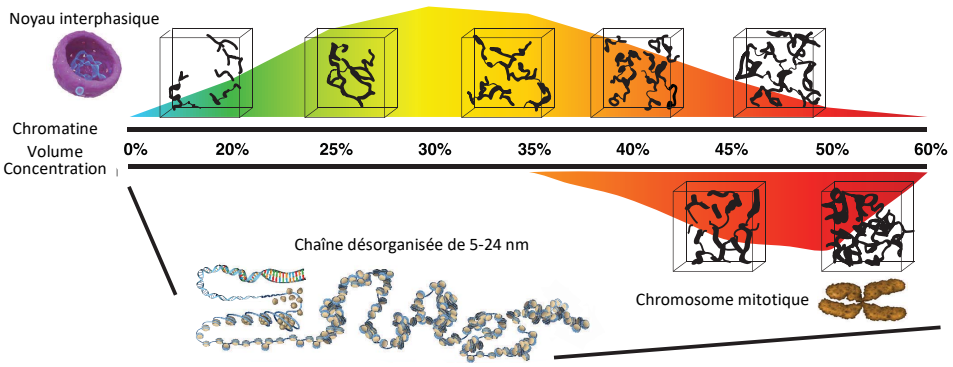
discernables individuellement, les chromosomes en mitose présentent une structure compacte individualisée en X formée par les chromatides sœurs reliées entre elles par un centromère. Cette structure en X permet de condenser de 2 à 3 fois plus le brin d'ADN, par rapport à l'interphase (Paulson et al., 2021). Par la suite, les chromatides sœurs sont séparées lors de l'anaphase.

De nombreux modèles de chromosomes mitotiques ont été proposés. Pour les premiers modèles datant de la première moitié du XX^{e} siècle, la chromatine était considérée comme un collier de perle ou une hélice (Wilson, 1900; Kuwada, 1939). Pendant la fin de la seconde moitié du XX^{e} siècle, deux modèles ont émergé. Le premier modèle, dit modèle à boucles radiales, suggère que les chromosomes mitotiques sont composés d'un squelette protéique sans histone, formant des boucles radiales à des intervalles réguliers (Marsden and Laemmli, 1979). Le second modèle se base sur un repliement hélicoïdal, sans squelette de protéine, aboutissant à la formation d'une fibre de chromatine de 10 à 30 nm enroulée sur elle-même (Sedat and Manuelidis, 1978; Belmont et al., 1987). De nos jours, grâce aux développements des techniques de capture de conformation des chromosomes et de la microscopie électronique, un nouveau modèle domine, selon lequel des boucles de chromatine de 5 à 24 nm se condensent plus fortement en comparaison à l'interphase et adoptent une configuration en escalier hélicoïdal (Ou et al., 2017) (Figure 5A-B). Cette condensation se traduit également par un enrichissement des contacts de longue distance (intervalle entre 2 à 12 Mb), alors que les chromosomes en interphases présentent un enrichissement des contacts inférieurs à 2Mb (Gibcus Johan H. et al., 2018). De plus, des données de Hi-C ont démontré que les chromosomes mitotiques ne présentent pas de niveau d'organisation structurale supérieure aux boucles de chromatine, contrairement aux chromosomes en interphase, avec notamment les TADs (Naumova et al., 2013; Gibcus Johan H. et al., 2018) (Figure 5C).

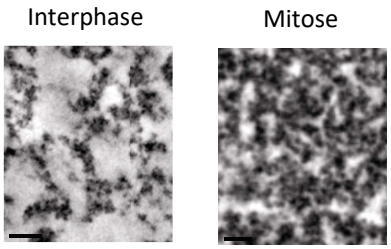
Ainsi, selon les étapes du cycle cellulaire, la chromatine subit des degrés de condensation très variés, en alternant entre une conformation adaptée à l'interphase, et plus particulièrement à la réplication de l'ADN, et une conformation très condensée nécessaire à la mitose et à la ségrégation chromosomique. Cela requiert des mécanismes de régulation finement orchestrés et implique notamment des complexes multiprotéiques, appartenant à la famille des complexes de maintenance structurelle des chromosomes (SMC) conservée

dans tous les domaines de la vie. Ces complexes SMC adoptent une structure en anneau leur permettant de capturer l'ADN. Chez les eucaryotes, la Condensine et la Cohésine, appartenant à cette famille SMC, participent à la condensation et à la cohésion des chromosomes, processus clés pour la division cellulaire.

A



B



C

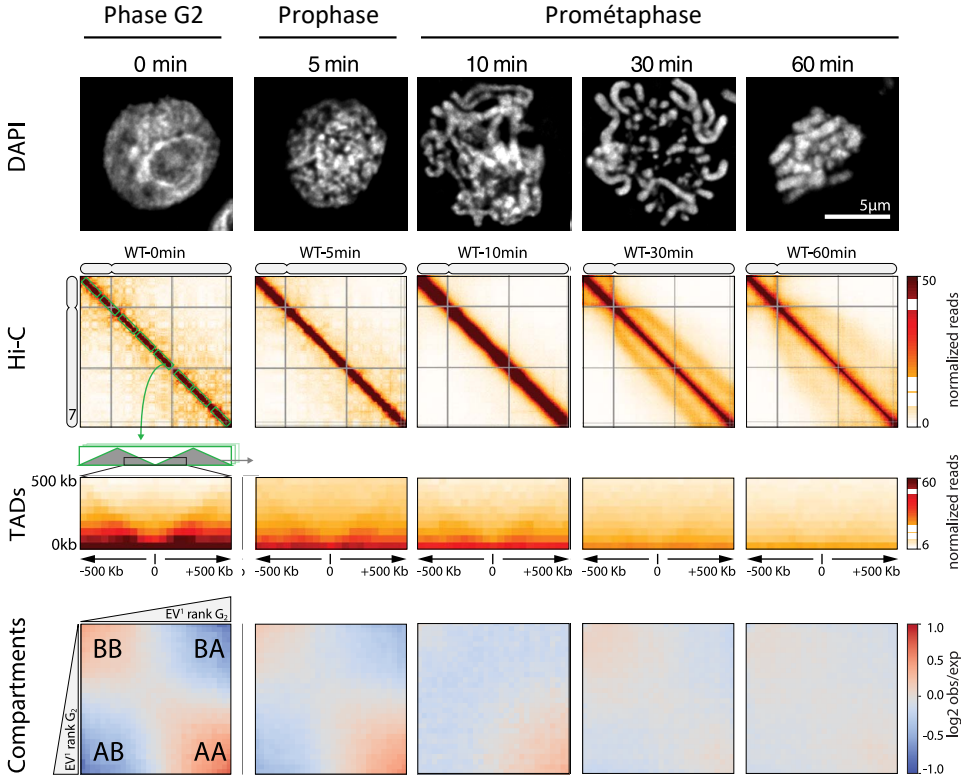


Figure 5 : De la fibre de chromatine au chromosome mitotique.
A. Représentation schématique des différents niveaux de condensation du chromosome mitotique. Adapté de (Ou et al., 2017).
B. Coupe tomographique de cellules épithéliales humaines des petites voies respiratoires pendant l'interphase (gauche) et la mitose (droite), réalisée par la technique de chromEMT. Adapté de (Ou et al., 2017).
C. Suivi de la disparition des TADs et des compartiments lors de la mitose du chromosome 7 dans des cellules DT 40. Les images DAPI permettent de visualiser l'aspect des noyaux et des chromosomes au cours des différentes étapes du cycle cellulaire, à un temps donné. Les cartes de Hi-C ainsi que les cartes d'interaction moyenne centrées sur les TADs montrent des changements conséquents au niveau de la taille des contacts chromatiniens, ainsi que la disparition des TADs lors de la mitose, de même que pour la compartimentation. Adapté de (Gibcus Johan H. et al., 2018).

III. Les complexes SMC : organisateurs de la chromatine

Les complexes SMC sont des complexes multiprotéiques conservés dans l'ensemble du monde du vivant, des bactéries aux archées et aux eucaryotes supérieurs. Ces complexes, qui sont capables de lier l'ADN, sont essentiels à de nombreux processus impliqués dans l'organisation et l'intégrité du génome. Quels que soient les organismes, ils présentent des caractéristiques architecturales communes. Toutefois, ces complexes SMC ont divergé au cours de l'évolution et présentent également des spécificités structurales, régulatrices et fonctionnelles. En règle générale, les archées et les bactéries codent principalement pour un seul complexe SMC, excepté pour les *Pseudomonas*. En revanche, chez les eucaryotes, trois complexes distincts sont retrouvés : la Condensine, la Cohésine et le complexe SMC5/6, qui ont des rôles fonctionnels spécifiques au cours du cycle cellulaire.

3.1. Découverte des protéines SMC

Dans la dernière moitié du XX^{e} siècle, a émergé une théorie selon laquelle des protéines différentes des histones seraient impliquées dans les niveaux d'organisation 3D de la chromatine. Laemmli et al sont parvenus à isoler, par un phénomène de compétition entre les histones et des polyanions sulfatés, des chromosomes mitotiques dépourvus d'histones et dont le repliement n'était pas altéré. Ils ont démontré que le repliement de l'ADN n'est pas altéré par la perte des histones et doit donc passer par le recrutement d'autres protéines, confirmant ainsi cette théorie (Laemmli et al., 1978). Depuis, de nombreux complexes protéiques et protéines impliqués dans l'organisation du génome ont été identifiés, dont les complexes SMC.

L'abréviation SMC, signifiant au départ Stabilité des Mini-Chromosomes, est apparue pour la première fois en 1985, où Larionov et al, ont identifié un gène impliqué dans la stabilité des minichromosomes (plasmides centromériques) chez *S. cerevisiae*, nommé Smc1 (Larionov et al., 1985). Strunnikov et al ont démontré par la suite l'importance de cette protéine pour la viabilité et le maintien de la cohésion des chromatides sœurs lors des processus mitotiques (Strunnikov et al., 1993).

En 1991, Niki et al ont découvert le gène mukB codant pour une protéine de 177 kDa, qui est essentielle à la partition des chromosomes chez *E. coli (Niki et al., 1991)*. En se basant

sur sa séquence en acide aminé, ils ont postulé que cette protéine possède deux domaines globulaires au niveau des extrémités N- et C-terminales et un autre en son centre. Ces domaines sont reliés par des régions en superhélice. Des micrographies électroniques de la protéine MukB ont confirmé cette architecture en trois domaines l'année suivante (Niki et al., 1992).

Des études de comparaison de séquences ont révélé que les protéines Smc1 de *S. cerevisiae* et MukB de *E. coli*, ainsi que d'autres protéines procaryotes et eucaryotes partagent une architecture commune, peuvent lier l'ATP et semblent être indispensables à la dynamique des chromosomes. Il a alors été proposé que ces protéines appartiennent à une famille conservée des procaryotes aux eucaryotes : la famille des protéines SMC, rebaptisée Maintenance Structurelle des Chromosomes. D'autres protéines homologues ont été identifiées, notamment la protéine Smc2 de *S. cerevisiae*, les protéines XCAP-E/SMC2 et XCAP-C/SMC4, qui ont été isolées à partir d'extrait d'œufs de *X. laevis,* la protéine SMC2 de *G. gallus,* ainsi que les protéines SMC1 et SMC2 purifié d'extrait de thymus de veau (Hirano and Mitchison, 1994; Saitoh et al., 1994; Strunnikov et al., 1995; Jessberger et al., 1996; Hirano et al., 1997).

La poursuite des investigations a permis de montrer qu'il n'existe qu'un seul complexe SMC impliqué dans la dynamique des chromosomes chez les procaryotes, alors que plusieurs complexes SMC ayant des fonctions distinctes ont été découverts chez les eucaryotes.

Le premier complexe SMC eucaryote à avoir été identifié est la Condensine, qui participe à la condensation des chromosomes lors de la mitose (Hirano et al., 1997). Ce complexe a été isolé à partir d'extrait de *X. laevis* et est composé de deux sous-unités CAP-E/Smc2 et CAP-C/Smc4, et de trois autres sous-unités auxiliaires : CAP-D/Ycs4, CAP-G/Ycg1 et CAP-H/Brn1 (noms d'usage chez l'homme/*S. cerevisiae*). La condensine a ensuite été trouvée dans différents organismes, comme la levure *S. pombe*, *S. cerevisiae* et l'homme (Sutani et al., 1999; Freeman et al., 2000; Schmiesing et al., 2000).

En 1997, Michaelis et al ont identifié, chez *S. cerevisiae,* le deuxième complexe SMC, la Cohésine, impliqué non pas dans la condensation des chromosomes mitotiques, mais dans la cohésion des chromatides sœurs jusqu'à leur séparation lors de la mitose (Michaelis et al., 1997). Il se compose de deux protéines SMC : SMC1 et SMC3 et d'au moins deux protéines non-SMC : la kleisine Scc1/MCD1/RAD21 et la protéine Scc3/STAG. Ce complexe a été

également été trouvé chez *S. pombe, X. laevis* et chez l'homme (Losada et al., 1998; Tóth et al., 1999; Tomonaga et al., 2000; Sumara et al., 2000).

Enfin, dans les années 2000, le dernier complexe eucaryote SMC, composé des protéines SMC5 et SMC6 et de six protéines auxiliaires Nse1-6, a été identifié comme étant impliqué dans la réplication et la réparation de l'ADN, chez la levure *S. pombe* et chez l'homme (Fousteri and Lehmann, 2000; Taylor et al., 2001).

Ainsi, la famille très conservée des complexes SMC joue un rôle primordial dans de nombreuses transactions avec l'ADN et dans l'organisation des chromosomes. Cela est rendu notamment possible grâce à leur architecture commune, qui leur permet d'adopter une structure en anneau et d'encercler l'ADN. De nombreuses études ont permis de mieux comprendre leur mode d'action concernant la capture de la chromatine, bien qu'il reste de nombreux questionnements. Le tableau 1 récapitule les différentes sous-unités cœur de chaque complexe SMC.

	Procaryotes		Eucaryotes		
	Complexe SMC	MukBEF	Cohésine	Condensine	Complexe SMC5-SMC6
Protéines SMC	SMC	MukB	SMC1A (Smc1)	SMC2 (Smc2)	SMC5 (Smc5)
			SMC3 (Smc3)	SMC4 (Smc4)	SMC6 (Smc6)
Kleisines	ScpA	MukF	RAD21(Scc1)	CAP-H/H2 (Brn1)	NSE4 (Nse4)
Protéines non-SMC	ScpB	MukE	PDS5 (Pds5)	CAP-D2/D3 (Ycs4)	NSE1 (Nse1)
			STAG1/2 (Scc3)	CAP-G/G2 (Ycg1)	NSE2 (Nse2)
			NIPBL (Scc2)	/	NSE3 (Nse3)

Tableau 1 : Composition des différents complexes SMC.
La nomenclature utilisée pour les eucaryotes est celle d'usage pour l'homme, ceux employés chez S. cerevisiae sont indiqués entre parenthèse.

3.2. Architecture générale des complexes SMC

Les complexes SMC procaryotes et eucaryotes partagent une architecture commune, composée de trois protéines cœur. Ils se composent d'un dimère de protéines SMC, auquel se lie une protéine dite kleisine, formant alors une structure en anneau, caractéristique de ces complexes. Il s'agit du complexe SMC canonique. Par la suite, la kleisine permet le recrutement de protéines régulatrices et auxiliaires.

3.2.1. Les protéines SMC

Il existe une seule protéine SMC chez les complexes SMC procaryotes, alors que les eucaryotes en possèdent six (SMC1, SMC2, SMC3, SMC4, SMC5, SMC6). Ce sont des protéines allongées d'environ 1 000 à 1 300 acides aminés et de 120 à 160 kDa, qui se divisent en cinq domaines distincts (Hirano, 2006). Ces protéines possèdent des régions N- et C-terminales globulaires très conservées (d'environ 150 à 160 acides aminés respectivement), ainsi que d'un domaine central organisé, appelé région charnière (d'environ 200 acides aminés). Ces domaines sont reliés les uns aux autres par de longues régions désorganisées (Figure 6A).

La protéine SMC se replie sur elle-même au niveau de la région charnière, entrainant l'arrangement des régions désorganisées en superhélice antiparallèle intramoléculaire d'une longueur d'environ 50 nm et le rapprochement des extrémités N- et C- terminales. La protéine SMC adopte alors une forme de bâtonnet, avec la région charnière à l'une de ses extrémités et le domaine formé des régions N- et C-terminales, appelé tête ATPase, à l'autre extrémité (Figure 6B) (Löwe et al., 2001; Haering et al., 2004).

Les protéines SMC présentent également deux motifs qui interrompent l'organisation en superhélice. Le premier motif, appelé joint, est situé dans les régions en superhélice proches des têtes ATPases et le second motif, appelé coude (¨elbow¨), est situé au milieu des régions en superhélice (Figure 6B). Ces motifs joint et coude semblent avoir, respectivement, une implication dans l'orientation correcte des têtes ATPases des protéines SMC, ainsi que dans différentes conformations adoptées par les complexes SMC (Gligoris et al., 2014; Eeftens et al., 2016; Diebold-Durand et al., 2017a; Bürmann et al., 2019).

Les protéines SMC peuvent dimériser de deux façons différentes et indépendantes. Une dimérisation permanente s'effectue à travers l'interaction de leurs régions charnières. En présence d'ATP, ces protéines peuvent aussi interagir par la dimérisation ATP-dépendante de leurs têtes ATPases (voir partie IV. Le complexe Cohésine).

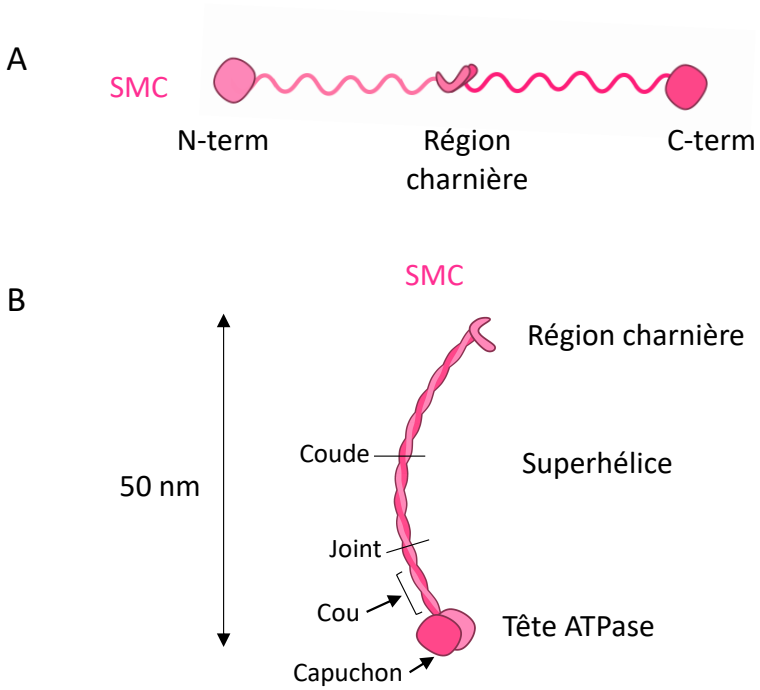


Figure 6 : Composition et architecture des protéines SMC
A. Chaque protéine SMC se compose de trois domaines globulaires : extrémité N-terminale (N-term), la région charnière et l'extrémité C-terminale (C-term).
B. Les protéines SMC se replient sur elles-mêmes permettant la formation de la tête ATPase. Cette tête est séparée de la région charnière par une longue région organisée en superhélice. Les protéines SMC possèdent également deux motifs (joint et coude), qui entrainent une rupture de l'organisation en superhélice et participent à la flexibilité des protéines SMC. Les interfaces avec la kleisine se font avec soit le cou ou le capuchon des protéines SMC.

3.2.2. La protéine kleisine, une sous-unité cœur non-SMC régulatrice

La kleisine est le troisième composant essentiel aux complexes SMC. Elle s'associe au dimère SMC, permettant de fermer la structure en anneau caractéristique de ces complexes. Le terme kleisine vient d'ailleurs du grec kleisimo qui signifie fermeture (Schleiffer et al., 2003).

La famille des kleisines se divise en cinq classes, une pour les procaryotes, avec notamment ScpA de *B. subtilis* et quatre classes pour les eucaryotes : α, β, γ, δ (Nasmyth and Haering, 2005). Chacune de ces classes est associée à un complexe SMC. Les kleisines de la classe α interagissent avec la Cohésine et sont présentes dans tous les organismes eucaryotes. Celles des classes β et γ se lient au dimère SMC de la Condensine et sont retrouvées dans une majorité d'espèces, sauf chez les champignons et *C. elegans*. Quant à celles de la classe δ, elles sont associées au dimère SMC5/SMC6 (Nasmyth and Haering, 2005).

L'ensemble des kleisines ont, dans leurs régions N- et C-terminales globulaires, une forte homologie de séquence (Schleiffer et al., 2003). Ces domaines N- et C-terminaux sont reliés par une région centrale désorganisée, de taille variable selon les espèces et moins conservée. Les structures cristallines des têtes ATPases de SMC en complexe avec les extrémités N- ou C-terminales de la kleisine chez *S. cerevisiae*, *E. coli*, et *B. subtilis* ont permis de définir les zones d'interactions entre ces protéines (Haering et al., 2004; Woo et al., 2009; Bürmann et al., 2013; Gligoris et al., 2014). L'extrémité N-terminale de la kleisine est composée de trois hélices α et interagit avec la région en superhélice située entre la tête ATPase et le motif joint de la protéine ν-SMC, appelée cou (nu pour ¨neck¨) (Haering et al., 2004; Woo et al., 2009). L'extrémité C-terminale de la kleisine adopte une conformation en hélice ailée (¨winged helix¨, WH), qui interagit avec la partie inférieure C-terminale de la tête ATPase de la κ-SMC, appelé capuchon (kappa pour ¨cap¨) (Bürmann et al., 2013; Gligoris et al., 2014). Ainsi, la kleisine interagit de manière asymétrique avec deux protéines SMC (Figure 7A) (Bürmann et al., 2013).

3.2.3. L'assemblage des complexes SMC

L'assemblage des complexes SMC se fait par l'association de la kleisine aux deux protéines SMC, ainsi que par la dimérisation des deux protéines SMC (ν-SMC et κ-SMC) au niveau de leurs régions charnières. Bien que les protéines SMC procaryotes et eucaryotes s'associent respectivement en homo- et en hétérodimères, les régions charnières des protéines SMC présentent une architecture similaire en forme de beignet (Haering et al., 2002; Ku et al., 2010; Kurze et al., 2011). En conséquence, le dimère de protéines SMC entières adopte une structure en V (Higashi et al., 2020; Shi et al., 2020; Vazquez Nunez et al., 2021). Il s'agit de la conformation présentant la plus grande ouverture du complexe SMC (Figure 7A).

Les protéines SMC peuvent également dimériser au niveau de leurs têtes ATPases, qui contribuent chacune aux deux semi-sites de fixation de l'ATP, lors de la liaison de deux molécules d'ATP (Löwe et al., 2001; Lammens et al., 2004). En d'autres termes, les molécules d'ATP se lient au domaine N-terminal de la première tête et au domaine C-terminal de la seconde. Ainsi, la fixation de l'ATP permet la dimérisation (« engagement ») des deux têtes ATPase. L'hydrolyse de l'ATP entraine leur dissociation.

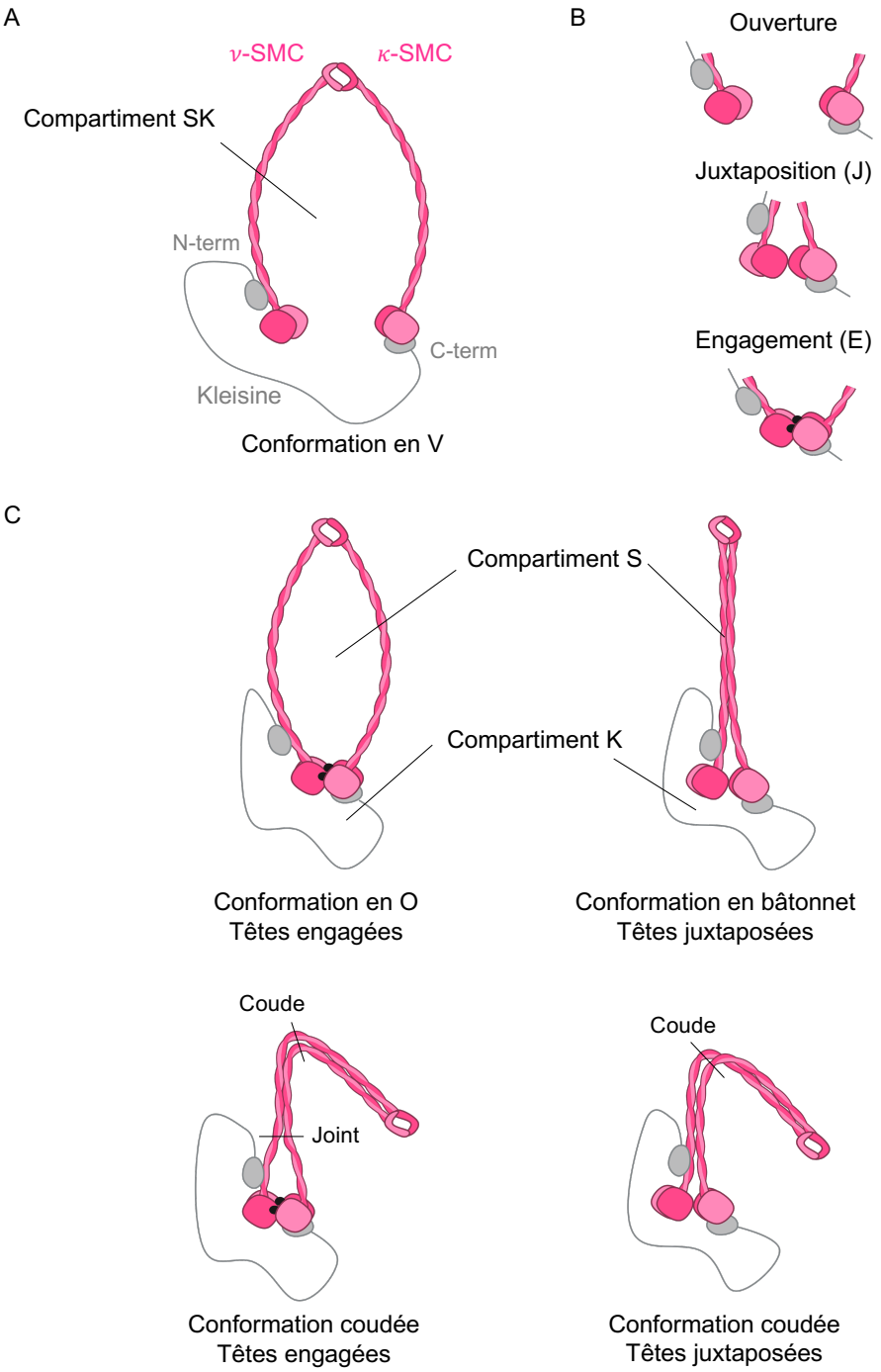


Figure 7 : Composition et architecture du complexe SMC.

A. Le complexe SMC est un complexe tripartite formé par un dimère de protéines SMC et d'une kleisine. La kleisine interagit via ses extrémités N- et C-terminales avec respectivement le cou d'une des protéines SMC (région en superhélice proche de la tête ATPase) et le capuchon de la tête ATPase de l'autre protéine SMC. Les protéines SMC dimérisent au niveau de leurs régions charnières, adoptant une conformation en V. Le complexe SMC peut être divisé en deux compartiments : S et K, respectivement limités par les régions charnières et les têtes ATPases des protéines SMC, et par les têtes ATPases des protéines SMC et la kleisine. Dans la conformation en V, ces deux compartiments sont fusionnés formant le compartiment SK.

B. Les têtes ATPases des protéines SMC peuvent s'associer de plusieurs manières. Elles peuvent ne pas être en contact l'une et l'autre, les têtes sont dites ouvertes. Elles peuvent être proches l'une de l'autre au niveau de leurs extrémités C-terminales et de leurs cous, qui se font face, les têtes sont juxtaposées. Pour finir, les têtes interagissent ensemble, en présence de deux molécules d'ATP (schématisées par un rond noir). Chaque

tête apporte un site de fixation semi-actif, leur interaction formant ainsi le site actif complet permettant la fixation de l'ATP. L'hydrolyse de l'ATP entraine leur dissociation. Cette activité enzymatique est essentielle aux changements conformationnels et aux fonctions des complexes SMC.
C. Le complexe SMC est un complexe dynamique et flexible. Il peut adopter plusieurs conformations. La première est la configuration en O, où nous observons l'engagement des têtes ATPases en présence d'ATP, les régions en superhélices des deux protéines SMC sont éloignées l'une de l'autre, adoptant une forme d'anneau. Les deux compartiments S et K sont représentés. La seconde conformation, appelée bâtonnet, est définie par la juxtaposition des têtes ATPases, ainsi que la proximité des régions en superhélice sur toutes leurs longueurs. A ces conformations, s'ajoutent celles coudées, induites par le repliement des régions en superhélice au niveau du coude, ce qui entraine le rapprochement des régions charnières vers les têtes ATPases. Cette conformation coudée peut être adoptée lors de l'engagement ou la juxtaposition des têtes. En cas d'engagement des têtes, les régions en superhélice jusqu'au motif joint restent à distance permettant l'orientation des têtes engagées, alors que pour les têtes juxtaposées, la totalité des régions en superhélice sont en contact.

Il est intéressant de noter que ces têtes ATPases peuvent rester proches l'une de l'autre (juxtaposées), en absence d'ATP dans certaines conformations du complexe SMC, leurs domaines C-terminaux et leurs cous restant alors à proximité (Diebold-Durand et al., 2017b; Chapard et al., 2019; Lee et al., 2020). Les têtes ATPases des protéines SMC peuvent ainsi adopter trois conformations différentes : elles peuvent être ouvertes, c'est-à-dire éloignées l'une de l'autre, juxtaposées ou engagées en présence d'ATP (Figure 7B). Ce mécanisme d'association/dissociation lié à l'activité ATPase est essentiel pour les changements conformationnels des complexes SMC qui sont eux-mêmes cruciaux à l'ensemble des fonctions de ces complexes (Arumugam et al., 2003).

Il a également été montré par cristallographie et cryo-microscopie électronique et/ou par des expériences de réticulation *in vitro* et *in vivo*, que les régions en superhélice des protéines SMC sont flexibles et peuvent interagir entre elles, le complexe SMC adoptant alors des conformations différentes de celle en V. Nous retrouvons tout d'abord une conformation en O, où les régions en superhélice sont éloignées l'une de l'autre et les têtes sont engagées (Figure 7C). Nous observons également une conformation en bâtonnet, où les régions en superhélice sont à proximité l'une de l'autre et les têtes ATPases sont juxtaposées (Figure 7C). Une autre conformation, appelée coudée, est également observée : les régions en superhélice des deux protéines SMC sont en contact et se replient sur elles-mêmes au niveau de leur coude, entrainant un rapprochement des régions charnières et des têtes ATPases. Cette conformation coudée est à la fois observée avec les têtes juxtaposées et avec les têtes engagées. Dans cette dernière conformation, seules les régions en superhélice entre la tête ATPase et le motif joint ne sont pas en contact, permettant les réarrangements structuraux

nécessaires à l'engagement des têtes (Figure 7C) (Barysz et al., 2015; Soh et al., 2015; Diebold-Durand et al., 2017a; Eeftens et al., 2016; Bürmann et al., 2019).

Les complexes SMC sont donc très dynamiques et peuvent adopter différentes conformations. Les mouvements des têtes ATPases des protéines SMC divisent notamment le grand anneau des complexes SMC en deux compartiments : (i) un compartiment SMC (appelé S), délimité par les interactions entre les régions charnières et entre les têtes ATPases des protéines SMC, (ii) un compartiment kleisine (K), délimité par les interactions entre les têtes ATPases et entre les têtes ATPases et les extrémités N- et C-terminales de la kleisine. Ainsi, l'ouverture des têtes entraine la formation d'un seul et grand compartiment SK, l'engagement et la juxtaposition des têtes conduit à la formation des deux compartiments S et K, avec des degrés d'ouvertures différents (Figure 7C) (Chapard et al., 2019).

Par la suite, une fois le complexe tripartite formé, la kleisine, à travers sa région centrale désorganisée recrute des protéines auxiliaires et régulatrices, nécessaires aux fonctions des complexes SMC.

3.2.4. Les protéines auxiliaires et régulatrices

Il existe deux familles de protéines auxiliaires et régulatrices pour les complexes SMC : les protéines KITEs et HAWKs. Selon le complexe SMC, une seule ou plusieurs de ces protéines peuvent être liées à la kleisine et sont essentielles pour le fonctionnement de ces complexes, en agissant comme régulateurs spécifiques (Palecek and Gruber, 2015). Certaines interagissent également avec le dimère des régions charnières des protéines SMC (Duan et al., 2009; Shi Zhubing et al., 2020).

Les protéines de la famille KITEs (¨kleisin interacting winged-helix tandem elements¨) possèdent une homologie structurale et sont constituées de plusieurs domaines WH en tandem, qui interagissent via plusieurs zones de contact avec la kleisine des complexes SMC procaryotes et celle du complexe SMC5/SMC6 eucaryote (Palecek and Gruber, 2015).

Les protéines de la famille HAWKs (¨HEAT proteins associated with kleisins¨) ont des tailles et structures similaires et se composent d'une répétition de domaines HEAT (¨Huntington/EF3/PP2A/Tor1¨), c'est-à-dire une répétition en tandem de deux hélices α liées par une boucle. Ces protéines se lient aux complexes eucaryotes de la Cohésine et de la Condensine (Wells et al., 2017).

3.3. Les différents complexes SMC : fonctions, compositions, architectures et spécificités

Bien que les complexes SMC soient présents dans l'ensemble des organismes vivants, partagent une architecture commune et jouent un rôle important dans les activités chromosomiques, ils possèdent chacun des spécificités structurales et fonctionnelles.

3.3.1. Les complexes procaryotes

Les complexes SMC procaryotes sont au nombre de trois : SMC-ScpAB, MukBEF et MksBEF. Ces complexes possèdent une architecture similaire et semblent tous impliqués dans la ségrégation chromosomique. Comme mentionné précédemment, il n'existe en très large majorité qu'un seul complexe SMC chez les bactéries. Cependant, seul le complexe SMC-ScpAB est associé à la famille des complexes SMC, les complexes MukBEF et MksBEF sont assimilés à des complexes SMC-like, du fait que, bien qu'ils présentent une architecture commune avec les complexes SMC, ils possèdent une faible homologie de séquences (Soppa et al., 2002; Schleiffer et al., 2003).

3.3.1.1. Le complexe SMC-ScpAB

Le complexe SMC-ScpAB est le complexe SMC le plus répandu chez les bactéries et est le complexe procaryote dont les protéines présentent la plus forte homologie de séquence avec celles de la Condensine(Cobbe et Heck 2004). Ce complexe joue un rôle essentiel dans l'organisation et la ségrégation des chromosomes, en permettant notamment une forte compaction des régions d'ADN répliquées, jusqu'à leur séparation (Britton et al., 1998; Moriya et al., 1998; Wang et al., 2014; Marbouty et al., 2015).

Le complexe SMC-ScpAB se compose d'un homodimère de protéines SMC, de la kleisine ScpA, qui interagit avec le cou d'une des protéines SMC et avec le capuchon de l'autre protéine SMC (Figure 8A) (Bürmann et al., 2013). La protéine KITE, ScpB, s'associe ensuite en dimère à la région centrale de ScpA.

La combinaison des données structurales, de résonance paramagnétique électrique (RPE), de microscopie à force atomique (AFM) et de réticulation *in vivo* a démontré que ce

complexe SMC-ScpAB peut adopter plusieurs conformations (Figure 8B). La première conformation est celle en bâtonnet, où les têtes ATPases, ainsi que les régions en superhélice des protéines SMC sont juxtaposées. Deux conformations en O sont observées, (i) une semi-ouverte se caractérisant par un engagement des têtes en présence d'ATP, ce qui entraine un réarrangement des régions en superhélice proximales, qui s'éloignent l'une de l'autre, le reste de ces régions demeure à proximité (conformation intermédiaire) et (ii) une totalement ouverte, également appelée conformation en anneau, avec l'engagement des têtes associé à un éloignement total des régions en superhélice. Une dernière conformation, en V, a également été déterminée, où les têtes ATPase et les régions en superhélice sont éloignées les unes des autres. Il est intéressant de noter que, pour ce complexe SMC bactérien, la conformation coudée n'a jamais été observée (Fuentes-Perez et al., 2012; Soh et al., 2015; Diebold-Durand et al., 2017b; Kamada et al., 2017; Vazquez Nunez et al., 2021). Les conformations en bâtonnet et en O sont illustrées dans la Figure 8C.

3.3.1.2. Le complexe MukBEF

Le complexe MukBEF est retrouvé chez les entérobactéries, dont font partie les bactéries *E. coli,* ainsi que chez certaines γ-protéobactéries. Ce complexe semble avoir remplacé le complexe SMC-ScpAB, du fait de ses fonctions dans la ségrégation et l'organisation des chromosomes (Danilova et al., 2007; Lioy et al., 2018).

MukBEF se compose d'un homodimère de protéines SMC MukB, de la kleisine MukF, ainsi que d'un dimère de protéines KITE MukE, lié à MukF. La kleisine MukF présente une architecture différente des kleisines des autres complexes SMC, en se divisant en quatre domaines : (i) une extrémité à WH N-terminale, (ii) une région avec un faisceau de quatre hélices (¨helix bundle¨), suivie (iii) d'une région centrale permettant de lier MukE, et (iv) d'une région à WH C-terminale (Figure 9A) (Woo et al., 2009). Cette dernière se lie au cou de la protéine ν-SMC et présente une interface d'interaction similaire à celle des autres complexes SMC. L'interaction avec le capuchon de la seconde SMC , κ-SMC, se fait à travers la région en faisceau d'hélices et non avec la région N-terminale, formant alors un complexe tri-protéique asymétrique (Figure 9B) (Woo et al., 2009).

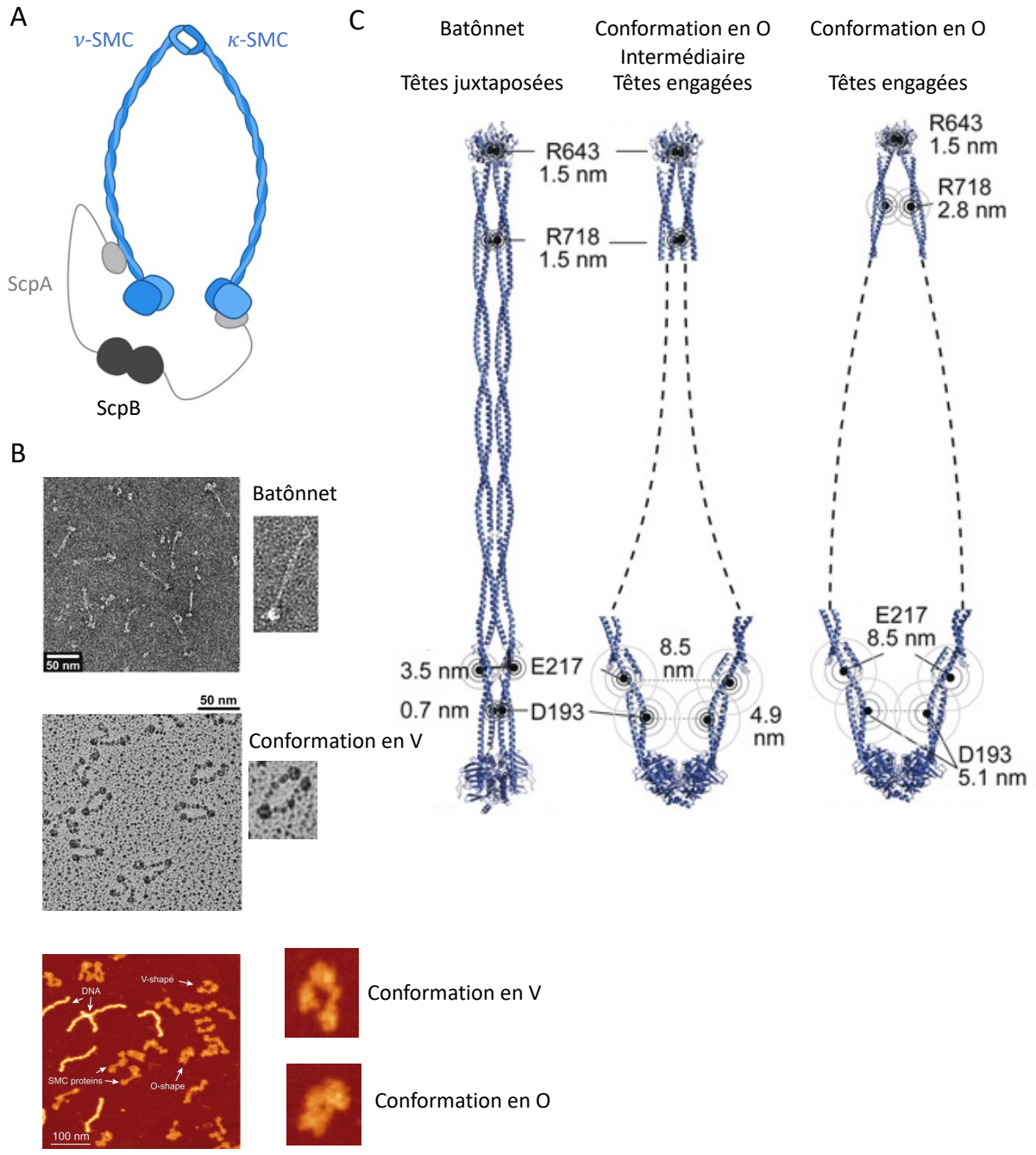


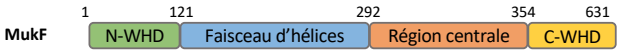
Figure 8 : Le complexe procaryote SMC-ScpAB.

A. Représentation schématique de ce complexe multiprotéique.

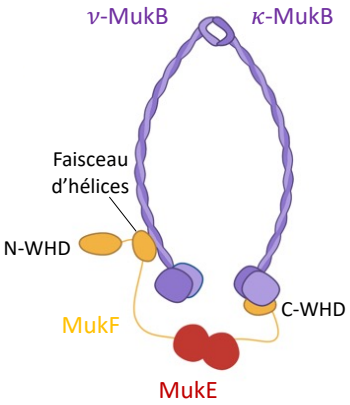
B. Micrographes électroniques (haut, milieu) et image d'AFM (bas) des complexes SMC-ScpAB de B. subtilis et de G. stearothermophilus. Ces complexes peuvent adopter plusieurs conformations : une forme en bâtonnet, en V et en O. Adapté de (Fuentes-Perez et al., 2012; Soh et al., 2015; Kamada et al., 2017).

C. Modèles structuraux des conformations en bâtonnet et en O. La conformation en bâtonnet est associée à la juxtaposition des têtes ATPases et au contact des régions en superhélice des deux protéines SMC dans toutes leurs longueurs. Les deux conformations en O requièrent l'engagement des têtes ATPases et l'ouverture des régions en superhélice, dont le degré varie. La conformation en O intermédiaire présente une plus faible ouverture comparée à la conformation en O totalement ouverte. Il est intéressant de noter que les régions en superhélice entre les têtes ATPases et le motif joint sont éloignées avec la même distance, quelle que soit la conformation adoptée. Adapté de (Vazquez Nunez et al., 2021).

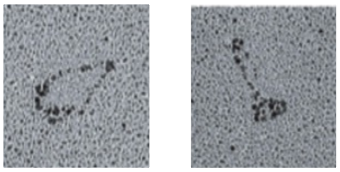
A


MukF
1
121
292
354
631
N-WHD
Faisceau d'hélices
Région centrale
C-WHD

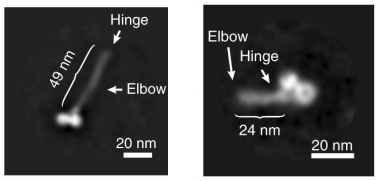
B


ν-MukB
κ-MukB
Faisceau d'hélices
N-WHD
C-WHD
MukF
MukE

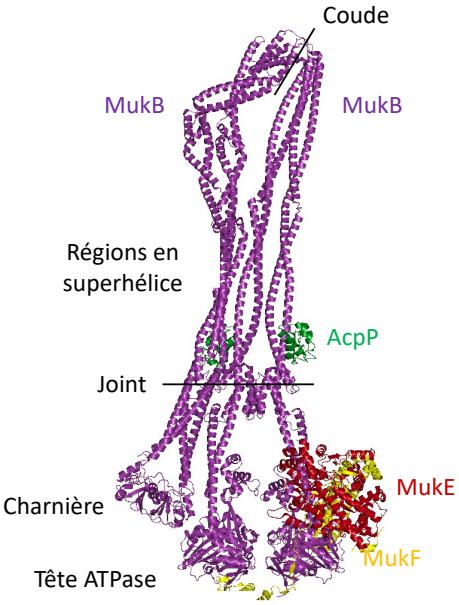
C



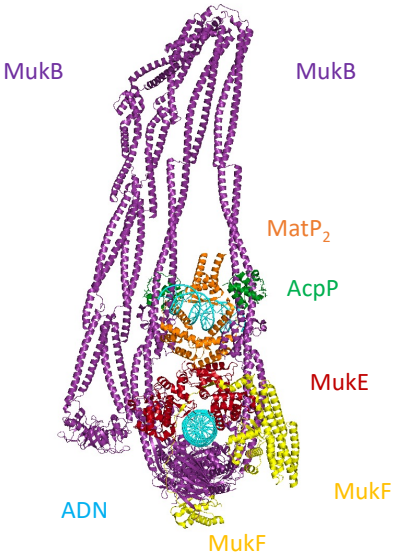
D


Hinge
49 nm
Elbow
20 nm
Elbow
Hinge
24 nm
20 nm

E


Coude
MukB
MukB
Régions en superhélice
AcpP
Joint
MukE
Charnière
MukF
Tête ATPase

F


MukB
MukB
MatP2
AcpP
MukE
MukF
ADN
MukF

G

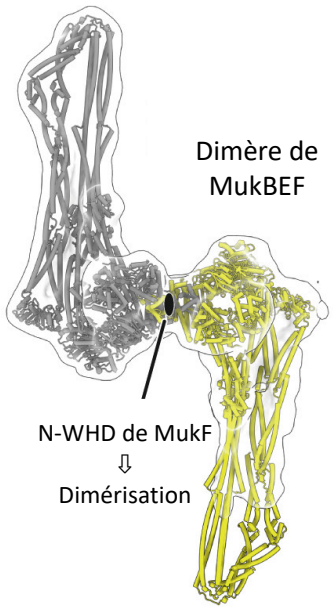

Dimère de MukBEF
N-WHD de MukF
⇩
Dimérisation

Figure 9 : Le complexe MukBEF de E. coli.
A. Représentation en diagramme de la structure primaire de MukF, différente des autres kleisines, et qui se compose de quatre domaines distincts. N-WHD et C-WHD : domaine à hélice ailée N- ou C-terminal. Adapté de (Woo et al., 2009).
B. Représentation schématique de ce complexe.
C. Micrographes électroniques du complexe MukBEF. Différentes conformations, comme celles en anneau (gauche) et en bâtonnet (droite) peuvent être adoptées par ce complexe. Adapté de (Matoba et al., 2005).
D. Moyenne de classes 2D par coloration négative de la conformation en bâtonnet (gauche) et celle coudée (droite) adoptées par le complexe MukBEF. Adapté de (Bürmann et al., 2019).
E. Structure du complexe apo-MukBEF en conformation coudée, associé à la acyl-carrier protéine AcpP PDB : 7NYY). Adapté de (Bürmann et al., 2021).
F. Structure du complexe MukBEF en présence d'ATP, d'ADN, de la protéine AcpP et de la protéine de chargement MatP. MukBEF adopte alors une conformation clampée (PDB : 7NYZ). Adapté de (Bürmann et al., 2021).
G. Structure du dimère apo-$MukB_4E_4F_2$ en tête bêche, superposée dans la carte de cryoEM de 13 Å (PDB : 7NZ4, EMDB : EMD-12664). Adapté de (Bürmann et al., 2021).

Des micrographies électroniques de ce complexe ont permis de déterminer, dès 1992, certaines de ses conformations, notamment celles en V et en bâtonnet (Figure 9C) (Niki et al., 1992; Matoba et al., 2005). Lors d'expériences de microscopie électronique à coloration négative, Bürmann et al ont découvert une nouvelle conformation, où les particules présentent une longueur d'environ 24 nm, soit la moitié de ce qui est habituellement observé pour la forme en bâtonnet (50 nm). Ils ont mis en évidence que cela est dû au repliement au niveau du coude des protéines MukB, entrainant alors le rapprochement des régions charnières et des têtes ATPases (Figure 9D)(Bürmann et al., 2019). Ils sont ensuite parvenus à obtenir une reconstruction cryoEM du complexe MukBEF dans cette conformation coudée sous forme apo, où les têtes de MukB sont juxtaposées et proches de leurs régions charnières, grâce aux repliements de leurs bras en superhélice au niveau de leurs coudes (PDB : 7NYY) (Figure 9E). Une seconde structure de ce complexe en présence d'ATP, d'ADN et de la protéine de chargement MatP a également été résolue (PDB : 7NYZ) (Figure 9F). MukBEF adopte alors une conformation dite clampée, où les différentes sous-unités du complexe s'arrangent dans l'espace de sorte à créer un tunnel permettant la capture de l'ADN. Les têtes ATPases de MukB sont alors engagées (Bürmann et al., 2021). Cela permet ainsi d'observer les réarrangements structuraux opérés lors de la capture de l'ADN au sein de ce complexe. Lors de ces expériences, les complexes MukBEF ont été co-exprimés avec la protéine ¨Acyl carrier¨ AcpP, dont la liaison à MukBEF est essentielle pour l'organisation et la ségrégation chromosomique (Prince et al., 2021).

Il a également été montré que la région N-terminale de MukF est impliquée dans la formation d'un dimère MukF en tête bêche, $MukB_4E_4F_2$ (Figure 9G) (Bürmann et al., 2021). Ce dimère correspond à la conformation fonctionnelle de MukBEF (Fennell-Fezzie et al., 2005; Rajasekar et al., 2019; Woo et al., 2009). Cette dimérisation est caractéristique de MukBEF et a été confirmée *in vitro* par des caractérisations structurales (cristallographie et cryoEM) et chimiques, ainsi que par des expériences *in vivo*, grâce à des études de microscopie à fluorescence de particules uniques non invasive (Woo et al., 2009; Bürmann et al., 2021; Badrinarayanan Anjana et al., 2012).

3.3.1.1. Le complexe MksBEF

Un dernier complexe SMC-like, MksBEF (¨MukBEF like SMC¨) a été découvert en 2011 (Petrushenko et al., 2011). Les études de phylogénie ont montré une homologie concernant la taille et les structures secondaires des protéines SMC et kleisine entre les complexes MksBEF et MukBEF. Les protéines qui constituent ce complexe ont été nommées MksB pour la protéine SMC, MksF pour la kleisine et MksE pour la protéine KITE. Deux différences notoires sont observées entre ces deux complexes. Cela concerne d'une part les protéines MksB qui présentent des régions en superhélice environ deux fois plus petites que ces mêmes régions chez MukB. D'autre part, il n'y a pas d'homologie de séquence entre les sous-unités de ces complexes (Petrushenko et al., 2011).

Ce complexe MksBEF se retrouve dans certaines bactéries, en concomitance avec les complexes SMC-ScpAB ou MukBEF. Ce complexe est majoritairement étudié chez les bactéries du genre *Pseudomonas*, qui codent donc pour deux complexes SMC : SMC-ScpAB et MksBEF. De nombreuses questions restent ouvertes, notamment sur la redondance de fonction d'un tel complexe, alors qu'un autre complexe SMC est déjà présent.

Ce complexe reste très peu caractérisé, toutefois il a été mis en évidence que SMC-ScpAB et MksBEF ont des localisations différentes et participent tous deux à la maintenance et à la ségrégation des chromosomes (Zhao Hang et al., 2020; Lioy et al., 2020). En effet, une mauvaise séparation chromosomique a été observée dans des cellules déficitaires en protéines SMC ou MksB, démontrant donc de la complémentarité fonctionnelle de ces deux complexes (Zhao Hang et al., 2020). La fonction de MksBEF dépend également du système ParABS, système clé pour la partition chromosomique (Lioy et al., 2020). Dans le cas où ce

dernier est fonctionnel, MksBEF est complémentaire du complexe SMC-ScpAB et les deux complexes participent à la séparation des chromosomes. Si le système ParABS est défaillant, le complexe MksBEF semble avoir une importance majeure afin d'aboutir à une ségrégation chromosomique correcte (Lioy et al., 2020).

Il semble donc que le complexe MksBEF soit essentiel pour la division cellulaire chez les *Pseudomonas* et ce malgré la présence d'un autre complexe SMC. Cela démontre l'adaptation des complexes SMC chez les procaryotes au cours de l'évolution, que nous retrouvons également chez les eucaryotes, qui possèdent trois complexes SMC distincts : la Cohésine, la Condensine, et le complexe SMC5/SMC6.

3.3.2. Les complexes eucaryotes : Condensine, Complexe SMC5/SMC6 et Cohésine

Ces trois complexes sont conservés au cours de l'évolution et sont présents dans l'ensemble des eucaryotes, des plus petits organismes, les microsporidies, à l'homme (Nasmyth and Schleiffer, 2004). Ils remplissent des fonctions distinctes dans l'organisation de la chromatine au cours du cycle cellulaire. De plus, pour certains de ces complexes, des fonctions ainsi que l'existence de protéines régulatrices supplémentaires ont été constatées chez les eucaryotes supérieurs.

Le tableau 2 reprend la nomenclature des homologues des protéines SMC, des kleisines ainsi que des protéines auxiliaires associées dans les organismes principalement étudiés. Les noms des protéines utilisées dans ce manuscrit sont les noms d'usage utilisés chez l'homme.

3.3.2.1. La Condensine

La Condensine est impliquée dans la condensation des chromosomes mitotiques, en agissant en concomitance avec la topoisomérase II-alpha, ainsi que dans l'organisation et la ségrégation de l'ADN ribosomal et d'autres séquences d'ADN répétées (Freeman et al., 2000; Machín et al., 2004; Hirano, 2005).

			S. cerevisiae	*S. pombe*	*H. sapiens*
Cohésine					
Sous-unités cœur	κ-Smc		SMC1	Psm1	SMC1A, SMC1B*
	ν-Smc		Smc3	Psm3	SMC3
	α-kleisine		Scc1, Rec8*	Rad21, Rec8*	RAD21, REC8*, RAD21L*
	HEAT-A		Pds5	Pds5	PDS5A, PDS5B
	HEAT-B		Scc3	Psc3, Rec11*	SA1, SA2, SA3*
Protéines régulatrices	Protéines de chargement		Scc2 Scc4	Mis4 Ssl3	NIPBL MAU2
	Acétylase		Eco1	Eso1	ESCO1, ESCO2
	Déacétylase		Hos1	-	HDAC8
	Protéine stabilisatrice		-	-	Sororine
	Protéine déstabilisatrice		Wpl1	Wpl1	WAPL
	Séparase		Esp1	Cut1	ESPL1
	Complexe Shugoshine-phosphatases		Sgo1 PP2A	Sgo1 PP2A	SGOL1 PP2A
Condensine					
Sous-unités cœur	κ-Smc		Smc4	Cut3	SMC4
	ν-Smc		Smc2	Cut14	SMC2
	Condensine −	γ-kleisine	Brn1	Cnd2	CAP-H
		HEAT-A	Ycs4	Cnd1	CAP-D2
		HEAT-B	Ycg1	Cnd3	CAP-G
	Condensine =	β-kleisine	-	-	CAP-H2
		HEAT-A	-	-	CAP-D3
		HEAT-B	-	-	CAP-G2
Protéines régulatrices	Kinase cycline-dépendante		Cdc28	Cdc2	CDK1
	Kinase Aurora B		Ipl1	Akr1	Aurora B
	Kinase polo-like		Cdc5	-	PLK1
Complexes SMC5/SMC6					
Sous-unités cœur	κ-Smc		Smc5	Smc5	SMC5
	ν-Smc		Smc6	Smc6	SMC6
	δ -kleisine		Nse4	Nse4	NSMCE4
	Ligase E3 tandem-WHD		Nse1	Nse1	NSMCE1
	Tandem - WHD		Nse3	Nse3	NSMCE3
	SUMO-ligase		Nse2	Nse2	NSCMCE2
Protéines régulatrices	Protéine de recrutements		Nse5 Nse6	Nse5 Nse6	SLF1 SLF2

Tableau 2 : Nomenclature des différents homologues des complexes SMC.
(variants pour la méiose). Adapté de (Haering and Gruber, 2016; Oravcová and Boddy, 2019).*

Il existe deux types de condensine : la condensine de type I (forme canonique), qui est conservée de la levure à l'homme, et celle de type II, qui se retrouve dans l'ensemble des eucaryotes, excepté chez les *Fungi*, tels que *S. cerevisiae et S. pombe* (Ono et al., 2003; Yeong et al., 2003). Ces deux types de condensines ont des fonctions complémentaires, mais redondantes dans le façonnement des chromosomes, avec cependant des localisations différentes. La Condensine I est située dans le cytoplasme lors de l'interphase et n'accède à l'ADN qu'après la rupture de l'enveloppe nucléaire, alors que la Condensine II se trouve dans le noyau à chaque étape du cycle cellulaire (Hirota et al., 2004; Gerlich et al., 2006a). Lors de la mitose les deux condensines sont situées sur l'axe chromosomique mitotique, mais à des localisations différentes. La Condensine II est plutôt localisée au centre de l'axe, alors que la Condensine I est située en périphérie (Ono et al., 2003; Walther et al., 2018). Des travaux pangénomiques par technique Hi-C ont démontré que les deux types de Condensine induisent la formation de boucles d'ADN de tailles différentes. La Condensine II forme de grandes boucles d'ADN qui sont ensuite réduites en plus petites boucles par la Condensine I (Gibcus Johan H. et al., 2018).

Ces deux Condensines partagent le même hétérodimère SMC avec les sous-unités SMC2/SMC4, mais présentent des sous-unités auxiliaires différentes. La kleisine et les protéines HEAT sont CAP-H (Bnr1), CAP-D2/CAP-G (Ycs4 et Ycg1) pour la Condensine I et CAP-H2, CAP-D3 et CAP-G2 pour la Condensine II (Figure 10A). Il s'agit des noms d'usage des protéines humaines. Les noms des protéines de *S. cerevisiae*, lorsque les variants sont présents, sont indiqués entre parenthèses.

Des expériences de microscopie électronique par coloration négative, par ¨low angle rotary shadowing¨ et d'AFM sèche et liquide révèlent que la Condensine de *S. cerevisiae* peut adopter plusieurs conformations, du fait de ces régions en superhélice très flexibles (Figure 10B) (Anderson et al., 2002; Eeftens et al., 2016). Dans l'ensemble des techniques utilisées, nous retrouvons (i) la conformation en V, (ii) la conformation en bâtonnet, ainsi que (iii) la conformation en O. La conformation en V se caractérise par l'ouverture du complexe. Dans la conformation en bâtonnet, les têtes sont juxtaposées et les régions en superhélice en contact. Lors de la conformation en O, les têtes ATPases de SMC2 et de SMC4 sont engagées et les régions en superhélices sont éloignées (Anderson et al., 2002; Eeftens et al., 2016).

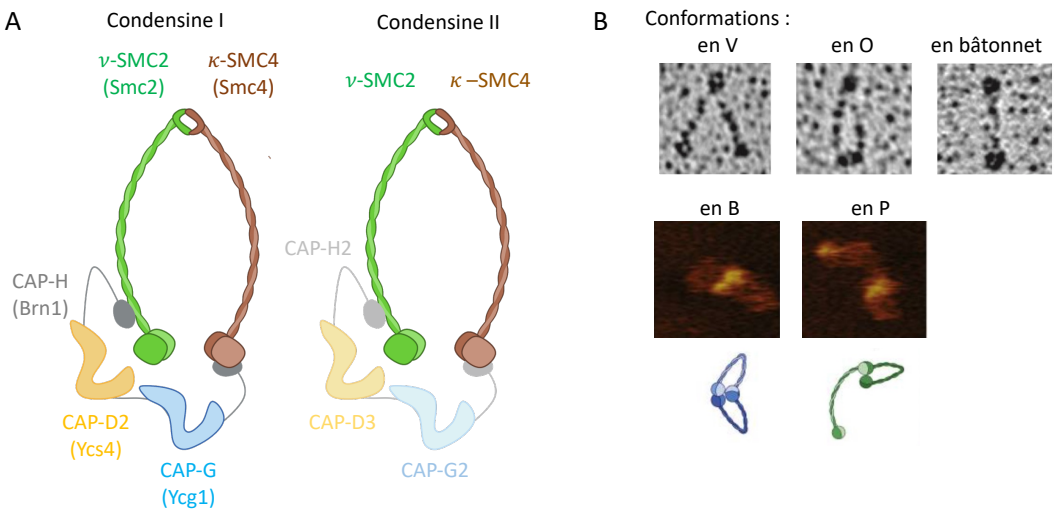


Figure 10 : La Condensine.
A. Représentation schématique des deux variants humains de ce complexe. La nomenclature utilisée chez S. cerevisiae est indiquée entre parenthèses.
B. Micrographes électroniques (haut), et films d'AFM liquide, mettant en évidence les différentes conformations adoptées par la Condensine. Sont retrouvées les conformations en V, en O et en bâtonnet communes à la majorité des complexes SMC. Deux nouvelles conformations ont été identifiées : (i) celle en B, où les têtes ATPases de SMC2 et SMC4 sont engagées et interagissent avec leurs régions charnières, les régions en superhélices sont éloignées les unes des autres et (ii) celle en P, où les têtes ATPases sont éloignées l'une de l'autre et une des têtes interagit avec les régions charnières. Adapté de (Anderson et al., 2002; Eeftens et al., 2016).

Les expériences d'AFM liquide ont mis en évidence deux nouvelles conformations, où les têtes ATPases et les régions charnières de SMC2 et SMC4 interagissent les unes avec les autres : les conformations en B et en P. La première conformation en B correspond à l'interaction entre les têtes ATPases engagées et les régions charnières et à l'éloignement des régions en superhélice. Dans la seconde conformation en P, une seule des têtes interagit avec les régions charnières, grâce au repliement du bras en superhélice (Figure 10B) (Eeftens et al., 2016).

De plus, ces deux dernières années, grâce à l'essor de la cryo-microscopie électronique, plusieurs structures de la Condensine de *S. cerevisiae* sous différentes conformations ont été obtenues, confirmant sa flexibilité et sa dynamique.

En 2020, Lee et al ont résolu une première reconstruction à 8,1 Å de la Condensine de *S. cerevisiae*, composée de Smc2, Smc4, Brn1, Ycs4 et Ycg1 dans un état apo, en absence d'ATP et d'ADN (PDB : 6YVU) (Figure 11A). Ce complexe adopte une conformation coudée, similaire à la conformation apo et coudée de MukBEF. En effet, nous observons un repliement des régions en superhélice, rapprochant ainsi les têtes ATPases et les régions charnières de Smc2 et Smc4, distantes alors d'environ 15 nm. Les régions charnières interagissent avec une partie

des régions en superhélice de Smc2 et Smc4. Cette conformation coudée a également été confirmée par spectrométrie de masse associée à la réticulation. Dans cette structure, les têtes ATPases de Smc2 et Smc4 sont juxtaposées et leurs interfaces avec les extrémités N- et C-terminales de la kleisine Brn1 sont similaires à celles observées dans l'ensemble des complexes SMC. Ycs4 peut interagir avec les têtes de Smc2 et Smc4. Ycg1 est plus mobile et peut adopter plusieurs positions à proximité des têtes ATPases, probablement connecté via son interaction avec la région centrale de Brn1 (Lee et al., 2020).

Lee et al ont également résolu une structure à 7,5 Å de ce complexe sous un état dit ponté, qui n'a jamais été observé auparavant. Cette conformation est caractérisée par la distanciation des têtes ATPases induite par Ycs4. En effet, les têtes ATPases de Smc2 et Smc4 sont séparées l'une de l'autre de 10 nm par Ycs4, qui, agit comme un espaceur en se liant aux deux têtes (PDB : 6YVV) (Figure 11B). Smc2 se courbe alors vers l'extérieur de 50°, avec son motif joint comme point de pivot. Aucune densité n'a été observée pour Ycg1.

Une autre reconstruction à 7,5 Å de la Condensine en présence d'ATP a également été obtenue (PDB : 6YVD) (Figure 11C). Les têtes de Smc2 et de Smc4, ainsi que leurs régions en superhélice se réarrangent, permettant leur engagement. Les régions en superhélice de Smc2 et Smc4 restent à proximité l'une de l'autre et adoptent une conformation coudée. L'extrémité N-terminale de Brn1 n'est plus liée avec le cou de Smc2. Ycs4 se dissocie du corps principal, mais reste à proximité, tout en étant mobile du fait de son interaction avec la région centrale de Bnr1, alors que Ycg1 se lie à la tête de Smc2 (Lee et al., 2020).

Toute comme MukBEF, un état clampé en présence d'ADN et d'ATP ou d'un analogue faiblement hydrolysable de l'ATP, l'ADP.BeF_3^-, a été observé. Cet état correspond à la formation autour du brin d'ADN d'un tunnel composé des différentes sous-unités de la Condensine. Deux structures de la Condensine de *S. cerevisiae*, définissant cet état clampé ont été obtenues. La première structure correspondant au complexe à quatre sous-unités (Smc2, Smc4, Bnr1 et Ycs4) et la seconde au complexe à cinq sous-unités (Smc2, Smc4, Brn1 et Ycs4, Ycg1) (PDB : 7Q2X, 7Q2Y, 7QEN) (Lee Byung-Gil et al., 2022; Shaltiel et al., 2022). Pour ces deux reconstructions à des résolutions comprises entre 3 Å et 3,5 Å, seul le module des têtes a pu être résolu, le complexe entier étant trop flexible. L'ADN se situe entre le sommet des têtes de Smc2 et Smc4 et la partie centrale de Ycs4. Bnr1 est lié au cou de Smc2 et au capuchon de Smc4. Les têtes de Smc2 et Smc4 sont engagées et leurs régions en superhélice s'ouvrent

d’environ 25°, adoptant ainsi une conformation en O (Figure 11D). L’ADN interagit alors avec la tête et le cou de Smc2 et la tête de Smc4 (Lee Byung-Gil et al., 2022; Shaltiel et al., 2022). La seconde structure met également en évidence un module périphérique formé par Ycg1 et la région centrale de Brn1, qui fixe également un brin d’ADN (Figure 11D) (PDB : 7QFW) (Shaltiel et al., 2022).

Ainsi, quatre conformations stables de la Condensine ont été résolues ces dernières années, partageant certaines caractéristiques communes avec les complexes SMC, mais présentant certaines particularités, telles que la conformation pontée, qui n’a encore jamais été observée pour d’autres complexes SMC, ainsi que la fixation de l’ADN par une des protéines HEAT, Ycg1.

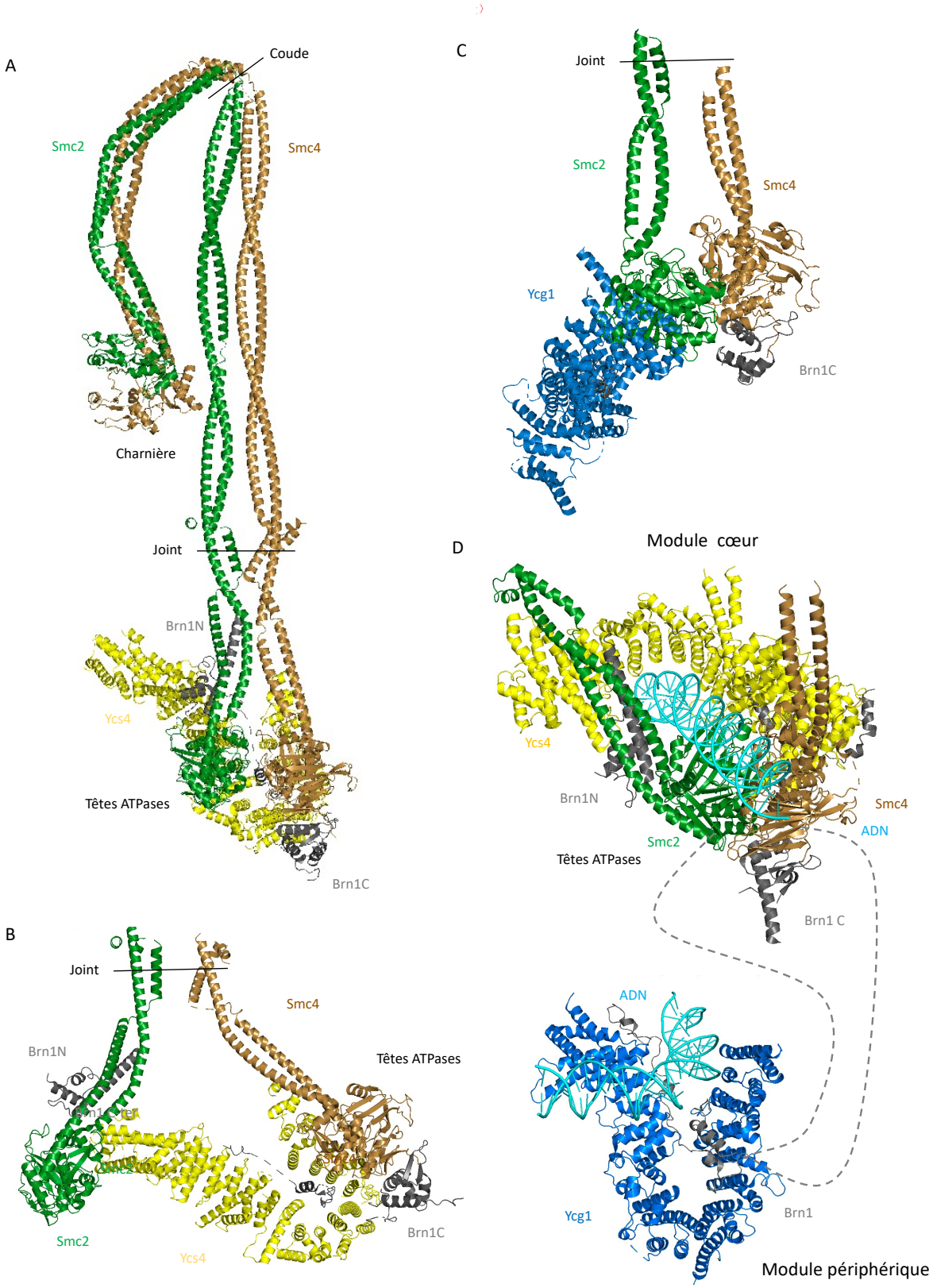


Figure 11 : Reconstruction cryoEM de la Condensine de S. cerevisiae sous différents états.
A. Reconstruction de la forme apo, en absence de nucléotide et d'ADN (PDB : 6YVU). Ce complexe adopte une structure coudée associée à la juxtaposition des têtes ATPases de Smc2 et Smc4, qui interagissent respectivement avec les extrémités N- et C-terminales de Brn1 (Brn1N et Brn1C). La partie centrale de Ycs4 se place à l'arrière des têtes. Adapté de (Lee et al., 2020).

B. Reconstruction de la forme pontée (PDB : 6YVV), où les têtes ATPases de Smc2 et Smc4 sont espacées l'une de l'autre de 10 nm, par Ycs4, qui interagit à travers des interfaces différentes de celles observées sous la forme apo. Adapté de (Lee et al., 2020).
C. Reconstitution de l'engagement des têtes ATPases de Smc2 et Smc4, en présence d'ATP (PDB : 6YVD). L'extrémité N-terminale de Brn1 et la protéine Ycs4 se sont dissociées de ce complexe permettant l'interaction de Smc2 avec la seconde protéine HEAT Ycg1. Adapté de (Lee et al., 2020).
D. Reconstitution de la forme clampée, en présence d'ATP et d'ADN. Le module cœur (PDB : 7QEN), composé des têtes ATPases engagées, de Brn1 et Ycs4, s'arrange de sorte à former un tunnel capturant l'ADN. Ycg1 peut également interagir avec l'ADN formant le module périphérique et est relié au module cœur grâce à son interaction avec Brn1 (PDB : 7Q2Z). Les régions non résolues de RAD21 liant le module cœur au module périphérique sont représentées par des pointillés. Adapté de (Shaltiel et al., 2022).

3.3.2.2. Le complexe SMC5/SMC6

Le complexe SMC5/SMC6 joue un rôle dans les processus de réparation des lésions de l'ADN par recombinaison homologue et en facilitant le recrutement de la Cohésine au niveau des cassures d'ADN. Il est également impliqué dans la réplication de l'ADN, en stabilisant et en redémarrant les fourches de réplication, ainsi que dans la résolution de la tension super-hélicoïdale induite par la réplication (Aragón, 2018; Palecek, 2018).

La majorité des expériences étant faites chez la levure *S. cerevisiae*, la nomenclature de cette espèce sera utilisée pour la description de ce complexe. Ce dernier est composé des protéines cœur Smc5 et Smc6 et de la kleisine Nse4. Les interactions entre ces sous-unités sont similaires à celles observées pour les autres complexes SMC. A ce complexe s'associent 6 protéines ¨Non-SMC Elements¨: Nse1, Nse2, Nse3, Nse5, Nse6 (Figure 12A). Nse1 et Nse3 appartiennent à la famille des protéines KITEs (Palecek and Gruber, 2015). En plus de l'activité ATPase des têtes Smc5 et Smc6, Nse1 et Nse2 confèrent respectivement une activité de liaison à l'ubiquitine et à la protéine SUMO (Taschner et al., 2021). Nse5 et Nse6 participent à la liaison à l'ADN et influencent l'activité ATPase des têtes de Smc5 et Smc6 chez *S. cerevisiae* (Taschner et al., 2021).

Les données de microscopie électronique et de spectrométrie de masse associée à la réticulation mettent en évidence respectivement des conformations en bâtonnet et en anneau (Figure 12B) (Yu et al., 2021; Hallett et al., 2021; Yu et al., 2022). A l'instar du complexe SMC-ScpAB procaryote et au contraire de la Condensine et de la Cohésine, il semble donc que ce complexe n'adopte pas une conformation coudée, où les têtes ATPases et les régions charnières des protéines SMC sont rapprochées.

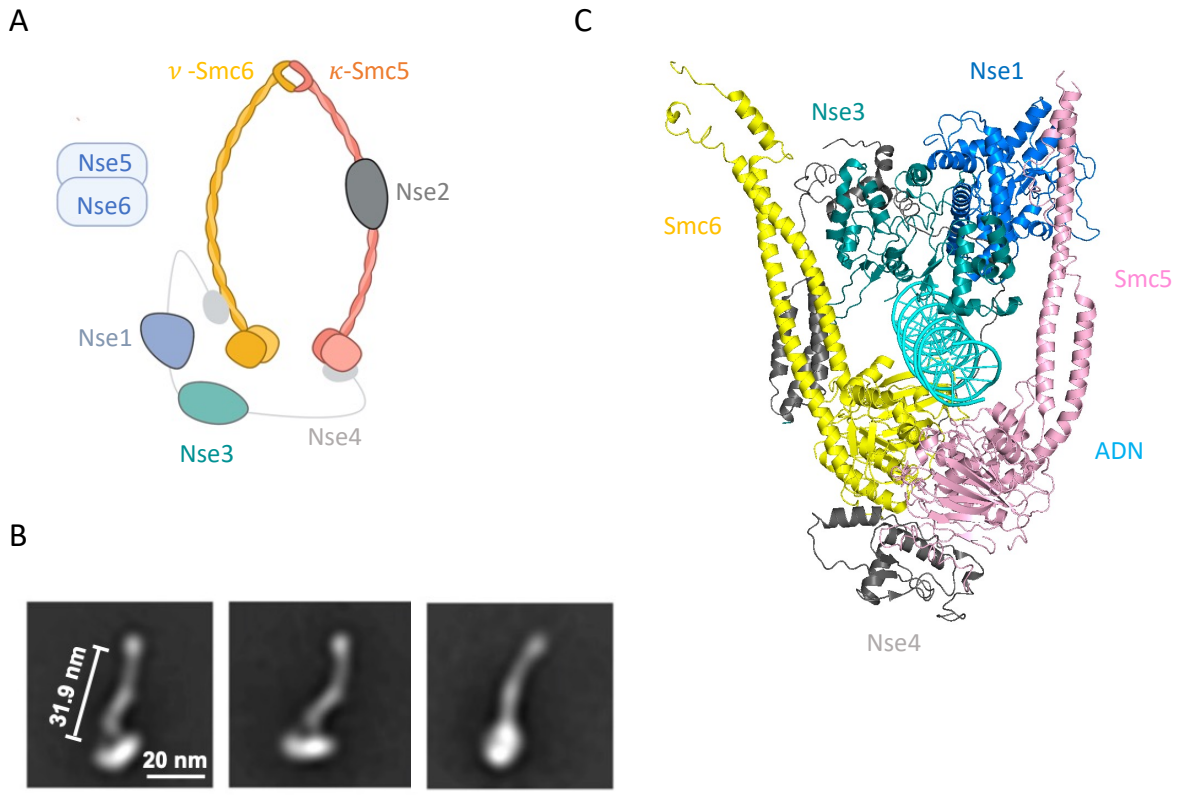


Figure 12 : Le complexe SMC5/SMC6
A. Représentation schématique de ce complexe.
B. Moyenne de classes 2D par coloration négative de la conformation en bâtonnet. Adapté de (Yu et al., 2021). La conformation en anneau a été mise en évidence par spectrométrie de masse associée à la réticulation (Yu et al., 2022).
C. Reconstruction du complexe SMC5/SMC6 de S. cerevisiae, en présence d'ATP et d'ADN (PDB : 7TVE). Il s'agit de la conformation clampée, caractérisée par l'engagement des têtes ATPases de Smc5 et de Smc6, ainsi que de la capture de l'ADN au sein d'une tunnel formé par les protéines cœur du complexe SMC5/SMC6 et de ses protéines auxiliaires. Adapté de (Yu et al., 2022).

La première reconstruction par cryoEM de ce complexe de *S. cerevisiae* (Figure 12C), publiée récemment, révèle comment les différentes sous-unités de ce complexe, Smc5, Smc6, Nse1, Nse3 et Nse4 s'agencent pour piéger l'ADN (Yu et al., 2022). Des mouvements globaux des protéines entières permettant l'encerclement de l'ADN sont observés, avec notamment ceux des protéines Nse1 et Nse4 et l'ouverture des régions en superhélice de Smc5 et Smc6 (Yu et al., 2022). Toutefois, cela n'entraine pas de réarrangements internes majeurs, excepté pour la kleisine NSE4, dont la région centrale, très flexible, se déplace des têtes ATPases vers les régions charnières des protéines Smc5 et Smc6 (Yu et al., 2022). Ces réarrangements entrainent la formation d'un tunnel dans lequel se trouve l'ADN. Bien que cela semble être un mécanisme commun à l'ensemble des complexes SMC, ce complexe SMC5/SMC6 présente quelques caractéristiques spécifiques. Nous retrouvons notamment des différences dans les

interactions entre sa kleisine et l'ADN et entre ses protéines auxiliaires et l'ADN, ainsi que l'absence d'interaction entre le cou de Smc6 et l'ADN, au contraire de la Condensine et de la Cohésine (Yu et al., 2022).

3.3.2.3. La Cohésine

Dans l'ensemble des eucaryotes, la Cohésine intervient tout au long du cycle cellulaire, en participant dès le début de la phase G1 à la réplication de l'ADN, ainsi qu'à sa réparation (Watrin and Peters, 2006). Après la réplication de l'ADN, ce complexe permet la cohésion des chromatides sœurs et prévient leur séparation précoce lors de la mitose (Michaelis et al., 1997).

Chez les eucaryotes supérieurs, ce complexe participe à des missions évolutives supplémentaires. Il est impliqué dans l'organisation 3D du génome, avec notamment la formation de boucles de chromatine et de TADs, ainsi que dans la régulation de l'expression des gènes, grâce au rapprochement d'éléments régulateurs (Fudenberg et al., 2016; Merkenschlager and Nora, 2016; Rao et al., 2017).

Il existe deux types de Cohésine chez les mammifères, l'un est observé lors de la mitose, l'autre lors de la méiose. Le complexe mitotique comprend les sous-unités SMC1A, SMC3, la kleisine RAD21 et les protéines HEAT, STAG1 ou STAG2 et PDS5 ou NIPBL (Figure 13A). Quant au complexe méiotique, il se compose de SMC3 et des variants SMC1B, RAD21L ou REC8, ainsi que de STAG3 (Ward et al., 2016). Les interactions entre ces différentes sous-unités sont communes à l'ensemble des complexes SMC, décrite précédemment.

Par comparaison aux autres complexes eucaryotes, et plus particulièrement à la Condensine, la Cohésine présente la même architecture de base, commune à l'ensemble des complexes SMC. Elle peut également adopter des conformations en V, en O, en bâtonnet, ainsi que celles coudées et celle en B, mais pas la conformation en P, observée uniquement chez la Condensine. Ces différentes conformations ont été observées par microscopie électronique et par AFM (Figure 13B) (Hons et al., 2016; Xiang and Koshland, 2021; Bauer et al., 2021). La conformation en V présente la plus grande ouverture du complexe et est formée par les interactions entre les régions charnières de SMC1A et SMC3, ainsi que les interactions entre leurs têtes ATPases et RAD21. La conformation en O est associée à l'engagement des têtes ATPases de SMC1A et SMC3 en présence d'ATP et s'accompagne de l'ouverture des régions

en superhélices. La conformation en bâtonnet se caractérise par la juxtaposition des têtes ATPases et par la proximité des régions en superhélices. Deux conformations coudées peuvent être observées, les deux présentant un repliement des régions en superhélices au niveau du coude, qui restent à proximité l'une de l'autre et rapprochent les régions charnières des têtes ATPases de SMC1A et SMC3. Toutefois, l'une est associée à la juxtaposition des têtes, alors que l'autre est associée à l'engagement des têtes. Enfin, la conformation en B correspond à l'engagement des têtes et à l'éloignement des régions en superhélices.

Au début de ma thèse, seules les structures résolues par diffraction aux rayons X des deux têtes ATPases isolées de la Cohésine (PDB : 1W1W, 4UX3), ainsi que celles des régions charnières (PDB : 1GXL, 2WD5) étaient disponibles pour le complexe cœur de la Cohésine (Gligoris et al., 2014; Haering et al., 2004, 2002; Kurze et al., 2011). Par la suite, Chapard et al ont mis en évidence les deux états des têtes ATPases de SMC1A et de SMC3, à savoir l'état juxtaposé, en absence d'ATP, et l'état engagé, en présence d'ATP, par des expériences de réticulation *in vitro* (Chapard et al., 2019).

Au cours de ma thèse, à l'instar de la Condensine, plusieurs reconstitutions de ce complexe cœur associé ou non à l'ATP, l'ADN et des protéines auxiliaires telles que NIPBL et STAG ont également été obtenues par cryoEM. Nous retrouvons notamment une forme apo, sans nucléotide ou ADN, ainsi que l'état clampé, où la Cohésine et ses protéines auxiliaires s'arrangent pour former un tunnel capturant l'ADN (PDB : 6QPW, 7OGT, PDB : 6ZZ, 6YUF, 6WGE, 6WG3) (Muir et al., 2020; Petela et al., 2021; Collier et al., 2020; Higashi et al., 2020; Shi Zhubing et al., 2020). Seule sera présentée rapidement dans cette partie une des structures de l'état clampé, afin de discuter des similitudes et différences observées chez les complexes SMC eucaryotes (Shi et al., 2020). L'ensemble des structures seront décrites plus en détail dans la partie IV. Le complexe Cohésine.

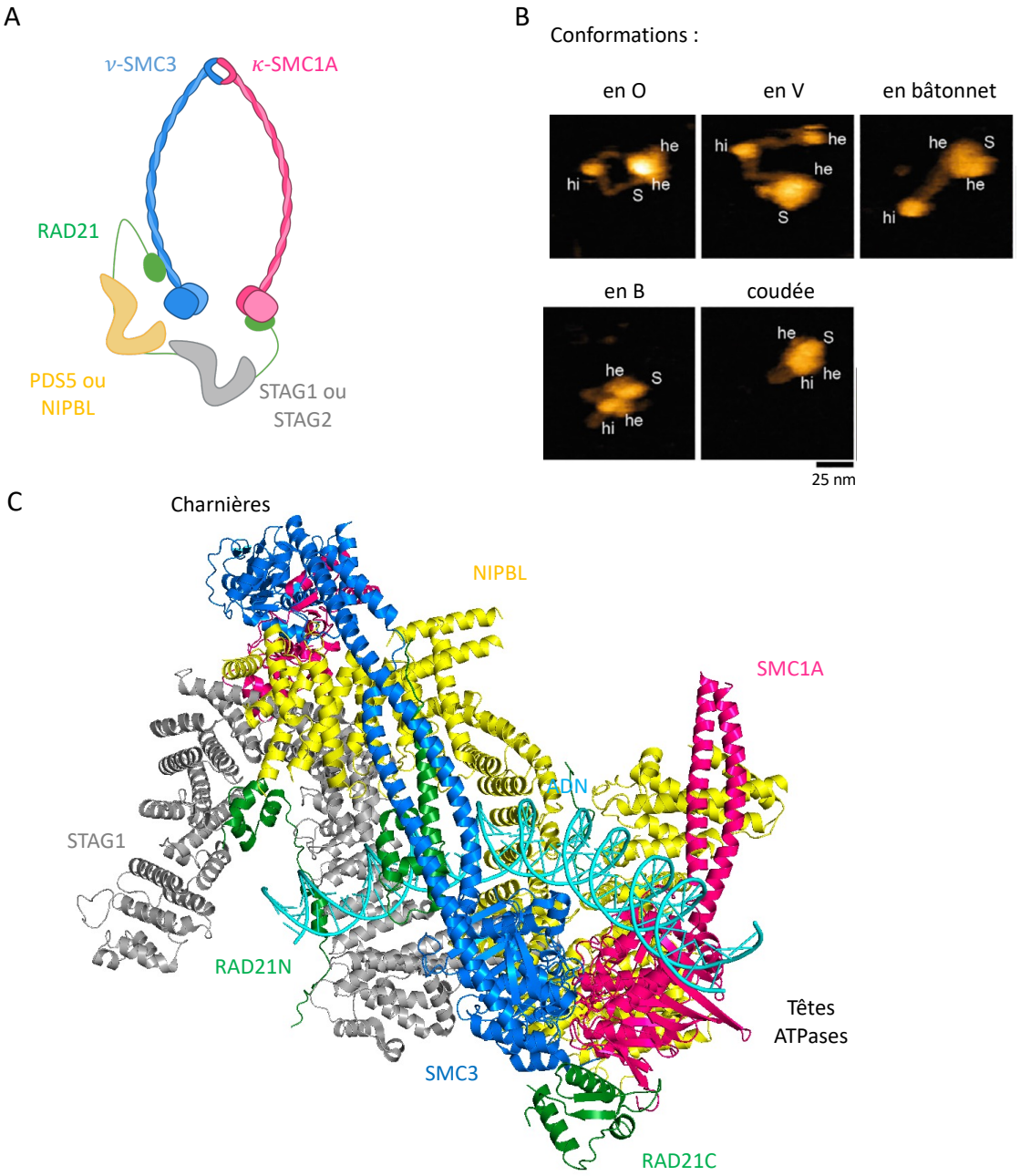


Figure 13 : La Cohésine.

A. Représentation schématique de la Cohésine (variant mitotique).

B. Micrographes d'AFM présentant les différentes conformations que peut adopter la Cohésine : une conformation en O, en V, en bâtonnet, en B et coudée. Les abréviations hi, he et S correspondent respectivement aux régions charnières, aux têtes ATPases de SMC1 et SMC3 et à la sous-unité STAG1. Adapté de (Bauer et al., 2021).

C. Reconstruction du module cœur de la Cohésine (SMC1A, SMC3, RAD21, STAG1), associé à la protéine auxiliaire NIPBL, dans un état clampé. L'ADN est piégé dans un tunnel construit par l'ensemble des protéines composant ce complexe, ainsi que par l'engagement des têtes ATPases, induit par l'AMP.PNP, un analogue non hydrolysable de l'ATP. Les protéines SMC1A et SMC3 adoptent une conformation coudée entrainant le rapprochement des têtes ATPases et des régions charnières. La région charnière de SMC1A interagit également avec STAG1 et NIPBL. RAD21N et RAD21C : extrémités N- et C-terminales de RAD21. Adapté de (Shi et al., 2020).

Shi et al ont obtenu une reconstruction cryoEM à 5,3 Å de l'état clampé de la Cohésine tétramérique humaine, qui est composée de SMC1A, SMC3, RAD21 et STAG1 et associée à la protéine auxiliaire NIPBL, à un analogue de l'ATP non hydrolysable, l'AMP.PNP et à l'ADN (Figure 13C) (Shi et al., 2020). Seul le module des têtes (cœur) a pu être résolu, du fait de la flexibilité des régions en superhélice. Les protéines SMC1A et SMC3 adoptent une conformation coudée, qui conduit au rapprochement des régions charnières vers les têtes ATPases. Les têtes de SMC1A et SMC3 sont engagées et associées respectivement aux extrémités C- et N-terminales de RAD21. NIPBL adopte une conformation en U et est positionné à l'arrière des têtes, avec lesquelles elle interagit directement. NIPBL se lie également aux régions en superhélice proches des têtes, aux régions charnières, ainsi qu'aux protéines RAD21 et STAG1. STAG1 adopte également une conformation en U et s'amarre à l'arrière du complexe, interagissant avec la région charnière de SMC1A. Ces différentes protéines s'assemblent ainsi pour former un tunnel permettant de capturer l'ADN. L'ADN est localisé sur le dessus des têtes ATPases et est en contact avec toutes les protéines composant ce tunnel.

L'état clampé de la Cohésine présente des similarités avec ce même état retrouvé chez la Condensine et le complexe SMC5/SMC6, avec la formation d'un tunnel, induit par l'engagement des têtes ATPases et l'amarrage des protéines auxiliaires. Ces complexes semblent donc partager un mécanisme de base piégeant l'ADN.

En comparant plus spécifiquement les états clampés de la Cohésine humaine et de la Condensine de *S. cerevisiae*, nous observons des similitudes mais également certaines spécificités propres à chaque complexe. Les têtes de SMC1A et SMC3 ou de Smc2 et Smc4 sont engagées et liées aux extrémités de la kleisine RAD21 ou Brn1. Dans cet état, les protéines SMC adoptent une conformation courbée. Toutefois, le repliement des régions en superhélice n'est pas identique, les régions charnières de la Cohésine sont plus proches des têtes ATPases, en comparaison de la distance entre les têtes ATPases et les régions charnières de la Condensine. Les régions charnières de la Cohésine interagissent également avec les deux protéines HEAT, STAG et NIPBL, ce qui n'est pas le cas de la Condensine. De plus, bien que Ycs4 (Condensine) et NIPBL (Cohésine) interagissent avec les protéines SMC, ils ne présentent pas les mêmes interfaces. De même pour la seconde protéine HEAT, STAG1 (Cohésine) et Ycg1 (Condensine), ils n'ont pas la même localisation dans le complexe, le premier interagissant

directement avec le complexe cœur, alors que le second se trouve en périphérie (décrit précédemment). Bien que des études supplémentaires soient nécessaires pour mieux comprendre l'implication et le positionnement des protéines HEAT, ces différences démontrent donc que le mode d'interaction entre l'ADN et les complexes SMC semble être distinct, ce qui est cohérent au vu des différentes fonctions occupées par ces complexes. Cette hypothèse est également renforcée par l'état ponté, ainsi que par la conformation en P, qui n'ont, à ce jour, été observés que chez la Condensine et non chez les autres complexes SMC.

Par ailleurs, la régulation des complexes SMC eucaryotes requiert le recrutement de nombreuses protéines auxiliaires, qui sont propres à chaque complexe et nécessaires à leurs différentes fonctions. La Cohésine est par ailleurs le seul complexe nécessitant l'aide d'un complexe de chargement, comprenant NIPBL (Scc2) et MAU2 (Scc4) (voir Tableau 2, récapitulant l'ensemble des protéines régulatrices) pour la capture de l'ADN.

Ainsi, bien qu'un mode d'action avec des bases communes puisse être défini pour l'ensemble des complexes SMC, chacun de ces complexes présente des spécificités structurales, fonctionnelles et régulatrices. Il est donc important d'étudier chaque complexe SMC séparément, afin de comprendre en termes moléculaires leurs modes d'actions et leurs mécanismes régulateurs sous-jacents, qui leur permettent de remplir leurs fonctions spécifiques.

De plus, de nombreuses différences ont été observées tout au long du cycle cellulaire entre les différentes espèces, notamment dans l'organisation de la chromatine lors de l'interphase et de la mitose. La Condensine et la Cohésine présentent également des caractéristiques évolutives, avec l'implication de différents variants chez les vertébrés, des fonctions additionnelles en plus de leur rôle canonique dans la mitose, ainsi que le recrutement de protéines régulatrices devenant de plus en plus complexe au cours de l'évolution. Cela suggère une mécanistique différente de chaque complexe en fonction des espèces, d'où la nécessité d'étudier chaque complexe de manière distincte et d'une espèce donnée, notamment le complexe de la Cohésine humaine, dont les caractéristiques seront détaillées dans la partie suivante.

IV. Le complexe Cohésine

Comme décrit précédemment, la Cohésine est un complexe essentiel pour l'organisation de la chromatine. Il se retrouve chez tous les eucaryotes, de la levure à l'homme et présente des caractéristiques communes inter-espèces, telles que son architecture, son organisation en anneau, ainsi que ses fonctions dans la réplication et la réparation de l'ADN et dans la cohésion des chromatides sœurs.

Cependant, nous notons des spécificités évolutives chez les eucaryotes supérieurs, avec notamment la complexification du cycle de la Cohésine, qui consiste au chargement de complexe sur l'ADN, à sa stabilisation, puis à sa libération du génome et qui nécessite le recrutement des sous-unités régulatrices spécifiques. La Cohésine des eucaryotes supérieurs possède également des fonctions additionnelles, en participant à l'organisation 3D du génome et à la régulation de l'expression des gènes. Or une majorité des études de la Cohésine se font chez les organismes de levure *S. cerevisiae et S. pombe*. Notre hypothèse de travail est que, bien que la Cohésine humaine possède une architecture et un mécanisme global communs à l'ensemble des Cohésines, elle présente des spécificités fonctionnelles et structurales, qui la distingue de la Cohésine des eucaryotes inférieurs, et qui permettent d'expliquer ses fonctions additionnelles.

Cette partie abordera donc l'ensembles des connaissances de la Cohésine, et plus particulièrement les fonctions, caractéristiques et mode d'action du complexe humain. Je me suis également intéressée à l'implication de la Cohésine dans les maladies. Les noms des protéines utilisées dans ce manuscrit sont ceux d'usage chez l'homme pour la Cohésine mitotique, sauf si explicitement mentionné.

4.1. Les fonctions de la Cohésine

4.1.1. Cohésion des chromatides sœurs (rôle canonique)

La réplication de l'ADN au cours de la phase S et sa condensation en début de mitose conduit à la formation de deux chromatides sœurs. Ces dernières doivent rester unies jusqu'à leur séparation en mitose, afin d'assurer une distribution uniforme de l'information génétique dans chaque cellule fille. Les chromatides sœurs restent ainsi physiquement connectées de la

phase S jusqu'à la métaphase. Il s'agit de la cohésion des chromatides sœurs, qui s'établit dès l'initiation de la réplication de l'ADN. Cette cohésion permet de contrer la force induite par les microtubules du fuseau, jusqu'à la bonne orientation des chromatides sœurs sur la plaque de métaphase, évitant ainsi une ségrégation chromosomique précoce (Tanaka et al., 2000; Eckert et al., 2007).

Chez les eucaryotes, cette cohésion est médiée par la Cohésine, qui grâce à sa structure en anneau peut encercler les deux chromatides sœurs (Gruber et al., 2003; Ivanov and Nasmyth, 2005; Haering et al., 2008). La délétion de la Cohésine entraine une perte précoce de la cohésion des chromatides sœurs, ainsi qu'une bi-orientation anormale de ces chromatides sœurs (Michaelis et al., 1997; Sonoda et al., 2001).

Chez *S. cerevisiae,* la Cohésine se lie aux bras des chromosomes peu avant le début de la réplication de l'ADN, à la fin de la phase G1, et est maintenue jusqu'à l'anaphase, où son clivage par la Séparase permet la séparation des chromatines sœurs et leurs migrations aux pôles opposées de la cellule (Michaelis et al., 1997; Uhlmann et al., 2000). Chez les vertébrés, elle se retrouve associée à l'ADN dès la fin de la télophase du cycle précédent. La majorité des Cohésines sont ensuite dissociées des chromosomes au cours de la prophase, les complexes restants sont enrichis au niveau des centromères et à leur périphérie. La Cohésine centromérique est maintenue jusqu'à la métaphase (Sumara et al., 2000; Waizenegger et al., 2000; Gerlich et al., 2006b).

Il a également été montré que seule une petite fraction de la Cohésine est nécessaire à la cohésion des chromatides sœurs. Heidinger-Pauli et al ont généré des cellules de levure bourgeonnante exprimant seulement 13 % et 30 % de la kleisine Scc1 et ont mis en évidence que la cohésion des chromatides sœurs et la ségrégation des chromosomes ne sont pas affectés par cette réduction, bien que les autres fonctions de la Cohésine, comme la réparation, le soient (Heidinger-Pauli et al., 2010). Cela a également été observé chez des cellules embryonnaires *de D. melanogaster*, où seulement 22 % des cohésines sont nécessaires pour une cohésion correcte (Carvalhal et al., 2018).

Ainsi, la Cohésine eucaryote est essentielle pour la cohésion des chromatides sœurs, participant à la ségrégation et à la distribution uniforme des chromosomes dans chaque cellule fille. Bien que la cohésion des chromatides sœurs soit une fonction canonique de la Cohésine, la faible proportion de ce complexe nécessaire à cette fonction met en évidence

son implication dans d'autres processus, notamment dans la réparation de l'ADN, l'organisation 3D du génome et dans la régulation de l'expression des gènes.

4.1.2. Réplication de l'ADN

En plus de ses fonctions dans la cohésion des chromatides sœurs, la Cohésine de la levure à l'homme participe à de nombreux processus nucléaires impliqués dans le maintien de l'intégrité du génome, notamment la réplication de l'ADN.

D'une part, la Cohésine est enrichie au niveau des origines de réplications et joue un rôle dans l'organisation spatiale des origines de réplication de l'ADN, participant ainsi à la réplication (Guillou et al., 2010; Emerson et al., 2022).

D'autre part, au cours des phases S et G2, la Cohésine promeut la stabilité des fourches de réplication. En effet, elle est essentielle au niveau des fourches de réplications situées dans des régions difficiles à répliquer, comme les télomères. Elle s'accumule également au niveau des fourches de réplications bloquées, en maintenant potentiellement les chromatides sœurs mieux appariées, favorisant ainsi le redémarrage des fourches. Le mécanisme impliqué dans cette fonction présente encore de nombreuses questions, mais il a été montré que la présence de protéines auxiliaires de la Cohésine, telles que STAG1, WAPL et PDS5 et d'autres protéines, dont RAD50, ainsi que l'ubiquitinylation de la Cohésine, sont nécessaires (Tittel-Elmer et al., 2012; Remeseiro et al., 2012; Frattini et al., 2017; Carvajal-Maldonado et al., 2019; Morales et al., 2020).

4.1.3. Réparation des dommages de l'ADN par recombinaison homologue

Les organismes vivants sont exposés de manière continue à des sources environnementales, physiques et chimiques. Il peut s'agir de sources endogènes, telles que les espèces réactives à l'oxygène, ou exogènes, comme les rayons UV et ionisants, les agents alkylants, les amines aromatiques et les hydrocarbures aromatiques polycycliques. Cette exposition peut induire des dommages à l'ADN, notamment des cassures double brin (DSB). Ces dernières peuvent être particulièrement délétères, conduisant à une instabilité génomique (Chatterjee and Walker, 2017). Il est donc primordial que les cellules possèdent

des mécanismes de réparation de l'ADN robustes et fidèles afin d'éliminer les dommages et restaurer l'intégrité de l'ADN.

Chez les eucaryotes, il existe cinq grandes voies de réparation : (i) la réparation par excision de base (BER), (ii) la réparation par excision de nucléotide (NER), (iii) la réparation des mésappariements (MMR), (iv) la jonction des extrémités non homologues (NHEJ) et (v) la recombinaison homologue (HR), les trois premières étant impliquées dans la réparation de cassure simple brin, les deux dernières dans les cassures double-brin.

La BER, active principalement dans la phase G1, corrige les modifications chimiques survenues au niveau d'une base individuelle, induites par des dommages oxydatifs, de désamination et d'alkylation, et qui n'entrainent pas de fortes distorsions de l'hélice d'ADN (Krokan and Bjørås, 2013). La NER permet d'éliminer des lésions plus volumineuses, généralement induites par les rayons UV, les hydrocarbures aromatiques et les agents utilisés en chimiothérapie. Ces dommages entrainent une distorsion de la double hélice d'ADN, bloquant la transcription et la réplication (Costa et al., 2003). La MMR corrige les lésions du nouveau brin synthétisé apparues lors de la réplication (Martín-López and Fishel, 2013).

Les deux dernières voies de réparation, la NHEJ et la HR permettent de réparer les cassures doubles brins induites par des agents chimiques et physiques. La NHEJ est un mécanisme non-conservateur rapide qui repose sur l'alignement et la ligation directe des extrémités de l'ADN sans qu'il y ait une homologie entre les brins. Elle permet ainsi de relier n'importe quel type d'extrémité de l'ADN et est active tout au long du cycle cellulaire, mais plus particulièrement en phase G1. Ce mécanisme se divise en plusieurs étapes : reconnaissance de l'extrémité de l'ADN lésée, recrutement et stabilisation de la machinerie NHEJ au niveau de cette cassure, pontage, traitement et ligation des extrémités de l'ADN (Davis and Chen, 2013; Hustedt and Durocher, 2017). La HR requiert une séquence homologue située sur une des chromatides sœurs et utilisée comme modèle de réparation. De ce fait, cette voie n'est active que pendant les phases S et G2 du cycle cellulaire. Elle se compose de plusieurs étapes : résection de l'ADN au niveau de la lésion, génération d'un brin monocaténaire, invasion de l'ADN double-brin homologue par l'ADN monocaténaire, synthèse par polymérisation des séquences d'ADN manquantes, résolutions des brins d'ADN (Karanam et al., 2012).

Chez les eucaryotes, la Cohésine est un acteur clé dans la détection et la réponse aux dommages de l'ADN, notamment par recombinaison homologue. Cette fonction a été décrite pour la première fois en 1992, avant la découverte de son implication dans la cohésion des chromatides sœurs. Birkenbihl et al ont identifié le gène Rad21 de *S. pombe*, comme étant impliqué dans cette voie de réparation, sa délétion rendant les cellules sensibles aux rayonnements ionisants (Birkenbihl and Subramani, 1992). Son rôle dans la réparation a été initialement associé à sa fonction de cohésion des cohésines sœurs et décrit comme passif : en maintenant à proximité les chromatides sœurs, la Cohésine facilite leurs réparations (Sjögren and Nasmyth, 2001). Par la suite, il a été mis en évidence que des nouvelles molécules de Cohésine sont chargées au niveau des sites de lésion, et que leurs altérations affectent la réparation de l'ADN (Ström et al., 2004). Cela démontre que la Cohésine préchargée pour la cohésion des chromatides sœurs ne confère pas cette fonction de réparation, mais qu'il s'agit de Cohésines *de novo* exclusivement chargées pour cette fonction. De plus, la capacité de la Cohésine à façonner la chromatine est également impliquée dans la réparation de l'ADN, par l'assimilation des TADs, formés par la Cohésine, en tant qu'unités fonctionnelles de la réponse aux dommages de l'ADN. Cette organisation 3D, à travers la formation de boucles d'ADN par extrusion, permet un rapprochement des sites de cassures double-brin et des séquences proximales, ainsi qu'à la propagation de la phosphorylation de l'histone H2Ax, qui est un point clé dans la signalisation de la réparation de l'ADN et qui s'étend principalement à l'intérieur des limites des TADs (Collins et al., 2020; Arnould et al., 2021).

Bien que le mode d'action exact de la Cohésine dans la réparation reste à caractériser précisément, son recrutement au niveau des cassures double-brin semble requérir la présence du complexe de chargement NIPBL-MAU2, celle du complexe de remodelage de la structure de la chromatine (RSC), la phosphorylation de l'histone H2Ax, la phosphorylation et la SUMOylation de RAD21, l'acétylation de SMC3 par ESCO, ainsi que la phosphorylation de SMC1A et SMC3 par le complexe ATM/ATR (Heidinger-Pauli et al., 2010; Kim et al., 2010; Oum Ji-Hyun et al., 2011; McAleenan et al., 2012; Scherzer et al., 2022).

4.1.4. Organisation 3D du génome et de la chromatine

Grâce à l'essor des techniques de conformation chromosomique 3C, 4C, 5C et Hi-C, il a été mis en évidence que la chromatine adopte différents niveaux d'organisation lors de

l'interphase, tels que les boucles de chromatine et les TADs, allant de 40 kb à 1 Mb . Ces niveaux d'organisation, compactant la chromatine, sont décrits dans la partie II. Architecture du génome eucaryote : de la double hélice d'ADN au chromosome.

Un des acteurs clé dans ces organisations d'ordre supérieur est la Cohésine. La dégradation de RAD21, l'une de ses sous-unités cœur, induite par le dégron inductible par l'auxine (AID), élimine les boucles de chromatine. La restauration de RAD21 permet de rétablir leur formation (de Wit et al., 2015; Wutz et al., 2017; Casa et al., 2020). La dégradation, à l'aide de cette même technique, de certaines protéines auxiliaires de la Cohésine, telles que STAG1 et STAG2, ainsi que CTCF, conduit également à une altération de la formation de boucles, démontrant de l'implication de ces protéines dans la formation des boucles de chromatine. Ces protéines sont décrites dans la partie 4.2.2. Les protéines auxiliaires et régulatrices.

Pour ce faire, la Cohésine organise, par un mécanisme appelé extrusion, la chromatine en boucles dynamiques, qui sont ensuite repliées en domaines fonctionnels, les TADs. Ces derniers sont délimités par des sites de liaison de la protéine CTCF asymétriques. Ces sites peuvent avoir deux orientations différentes : une orientation dite convergente, où les sites sont orientés les uns vers les autres, et inversement, une orientation divergente. Le facteur CTCF se fixe à ces sites et sert de point d'ancrage ou de barrage à la Cohésine en fonction de l'orientation du site de liaison. Cela est en accord avec la colocalisation de la Cohésine et de CTCF observées lors d'expériences ChIP-chip (Parelho et al., 2008; Wendt et al., 2008; Li et al., 2020).

Lors de l'extrusion de boucles, la Cohésine va alors rencontrer une protéine CTCF et deux configurations s'ouvrent à elle : soit le site de liaison de CTCF est convergent et la Cohésine s'arrête, la boucles est alors terminée et délimitée, soit le site de liaison est divergent et la Cohésine poursuit l'extrusion de la boucle, cette dernière s'agrandit. Cela permet ainsi de médier la formation et la taille des boucles de chromatine. Les mécanismes impliqués dans la formation de boucles d'ADN nécessitent le recrutement de protéines auxiliaires de la Cohésine, telles que NIPBL/MAU2, WAPL, PDS5, ESCO1, mais ne sont pas précisément compris (Wutz et al., 2017; Davidson et al., 2019; Kim et al., 2019; Wutz et al., 2020).

Chez *S. cerevisiae*, la Cohésine forme également des boucles de chromatine, bien qu'aucun homologue de CTCF n'ait été identifié. La formation de boucles requiert

l'intervention de protéines régulatrices, comme Pds5, Wpl1 et Eco1 (Dauban et al., 2020; Lazar-Stefanita et al., 2017). La Cohésine participe aussi à l'organisation 3D du génome chez *S. pombe* par la formation de globules (Mizuguchi et al., 2014). Toutefois les mécanismes sous-jacents restent à déterminer.

4.1.5. Régulation de la transcription

La Cohésine participe également à la régulation des gènes, notamment du fait de sa capacité à former des boucles d'ADN. Cette fonction est conservée dans l'ensemble des eucaryotes, de la levure à l'homme (Dorsett and Ström, 2012).

Chez les mammifères, la Cohésine associée à CTCF peut favoriser ou empêcher les contacts enhancer–promoteur, initialement distaux, par la formation de boucles de chromatine (Guo et al., 2015). Cela permet ainsi de médier l'accessibilité et l'expression des gènes (Rinzema et al., 2022) . En effet, Thiecke et al ont montré, par Hi-C, que la déplétion par AID de la sous-unité RAD21 de la Cohésine ou de CTCF conduit à l'élimination de la majorité des contacts entre promoteurs, mais qu'un certain nombre persiste, malgré l'absence de Cohésine et CTCF. De plus, d'autres équipes ont mis en évidence que les déplétions de la Cohésine et de CTCF entrainent quelques changements dans la transcription globale, mais bien plus faibles qu'attendus (Busslinger et al., 2017; Rao et al., 2017; Thiecke et al., 2020). Cela suggère qu'il existe probablement des mécanismes alternatifs cohésine-indépendants dans la régulation à longue distance de l'expression des gènes.

De plus, il a été montré que la Cohésine interagit avec d'autres protéines impliquées dans la transcription, telles que les protéines YY1, BRD2, TPO2B, WIZ, ainsi que le complexe Médiateur (Kagey et al., 2010; Uusküla-Reimand et al., 2016; Weintraub et al., 2017; Hsu et al., 2017; Wen et al., 2018) . Ainsi, le rôle de la Cohésine ne se résume pas uniquement aux rapprochements d'éléments régulateurs, mais est plus complexe et est au cœur de nombreuses études.

Chez la levure *S. cerevisiae*, l'inactivation de la Cohésine, ainsi que celle de Scc2 perturbent l'expression des gènes (Skibbens et al., 2010; Bose et al., 2012; Lindgren et al., 2014). La Cohésine de levure exerce donc également une mission de régulation transcriptionnelle, toutefois très peu d'informations sont disponibles sur les mécanismes sous-jacents.

4.1.6. Recombinaison V(D)J et commutation de classes des immunoglobulines.

Chez les vertébrés, il existe une immunité innée, dépourvue de mémoire, et une immunité adaptative, qui permettent de défendre l'hôte contre un agent infectieux. L'immunité adaptative fait intervenir un vaste panel de lymphocytes T et d'immunoglobulines. Plusieurs mécanismes, tels que la recombinaison V(D)J et la commutation de classe, sont à l'origine de cette grande variété de lymphocytes.

La recombinaison V(D)J est un mécanisme de recombinaison somatique de l'ADN, se produisant lors des premiers stades de la maturation des lymphocytes B et T (LB et LT). Elle permet de former un locus VDJ, composé de segments de gènes variables (V), de diversité (D) et de jonction (J) dispersés sur le génome. Ce locus code pour la région variable du récepteur à l'antigène des lymphocytes T et des immunoglobulines. La recombinaison V(D)J permet ainsi de générer une grande diversité de reconnaissance des antigènes par les lymphocytes. Cette recombinaison nécessite le recrutement de l'endonucléase RAG1/RAG2 et la présence de séquences signal de recombinaison (RSS) (Schatz and Ji, 2011).

La commutation de classe (CSR) consiste à modifier, lors de la maturation des lymphocytes B, la chaîne lourde des anticorps permettant d'obtenir des immunoglobulines de différentes classes, mais possédant la même région variable. Ce mécanisme permet de créer une diversité fonctionnelle des immunoglobulines (Xu et al., 2012).

Récemment, la Cohésine, par la formation de boucles d'ADN, a été identifiée comme médiateur dans ces mécanismes de l'immunité adaptative. Ba et al ont mis en évidence que la dégradation de RAD21 dans des lignées cellulaires pro-B de souris arrêtées en G1 conduit à l'élimination de toutes les recombinaisons V(D)J (Ba et al., 2020). La Cohésine est recrutée au niveau des locus des immunoglobulines lors de la CSR et de la recombinaison V(D)J, et permet de rapprocher les régions essentielles pour ces mécanismes par la formation de boucles (Y. Zhang et al., 2019; X. Zhang et al., 2019; Dai et al., 2021). Toutefois, d'autres études sont nécessaires afin de mieux comprendre l'implication de la Cohésine dans ces mécanismes.

4.2. Composition de la Cohésine

4.2.1. L'anneau tripartite, le complexe cœur de la Cohésine

Comme cité précédemment, la Cohésine se compose de trois protéines cœur : les protéines SMC, SMC1A et SMC3, ainsi que la kleisine RAD21. Pour rappel, SMC1A et SMC3 se composent de cinq domaines : deux domaines globulaires correspondant aux régions N- et C-terminales, un domaine central globulaire, dit charnière, et deux régions désorganisées permettant de lier les trois domaines globulaires (Figure 6A). Ces protéines se replient sur elles-mêmes au niveau de leur région charnière, permettant la formation de la tête ATPase, composée des extrémités N- et C-terminales et l'organisation de leurs deux régions désorganisées en superhélice. Ces protéines possèdent également deux motifs discontinus courts au niveau de leurs régions en superhélice, appelés coude et joint (Figure 6B).

SMC1A et SMC3 peuvent former un hétérodimère, à travers l'interaction de leurs régions charnières. Les régions charnières ont une structure en demi-cercle et leur dimérisation, à travers deux interfaces (appelées nord et sud), conduit à la formation d'un anneau de 50 Å de diamètre environ avec un pore central (Figure 14A). L'intérieur de cet anneau est bordé principalement par des résidus chargés positivement (Haering et al., 2002; Kurze et al., 2011; Shi et al., 2020). Des données structurales ont révélé que ce dimère de charnières peut adapter plusieurs conformations : une conformation fermée (PDB : 2WD5) et deux conformations ouvertes en présence d'ADN, où l'ouverture se fait soit au niveau de l'interface nord, soit au niveau de l'interface sud (PDB : 6WG6, 6WG4) (Figure 14A) (Kurze et al., 2011; Shi et al., 2020). Les interfaces de ce dimère des régions charnières sont dynamiques et peuvent alterner les conformations fermées et ouvertes.

La dimérisation de SMC1A et SMC3 peut également se faire par les têtes globulaires de ces protéines qui, en présence d'ATP, forment le module ATPase de la Cohésine (Collier et al., 2020). Ce module sera décrit plus en détail dans la partie 4.6. Le module ATPase de la Cohésine, acteur clé de son mode d'action.

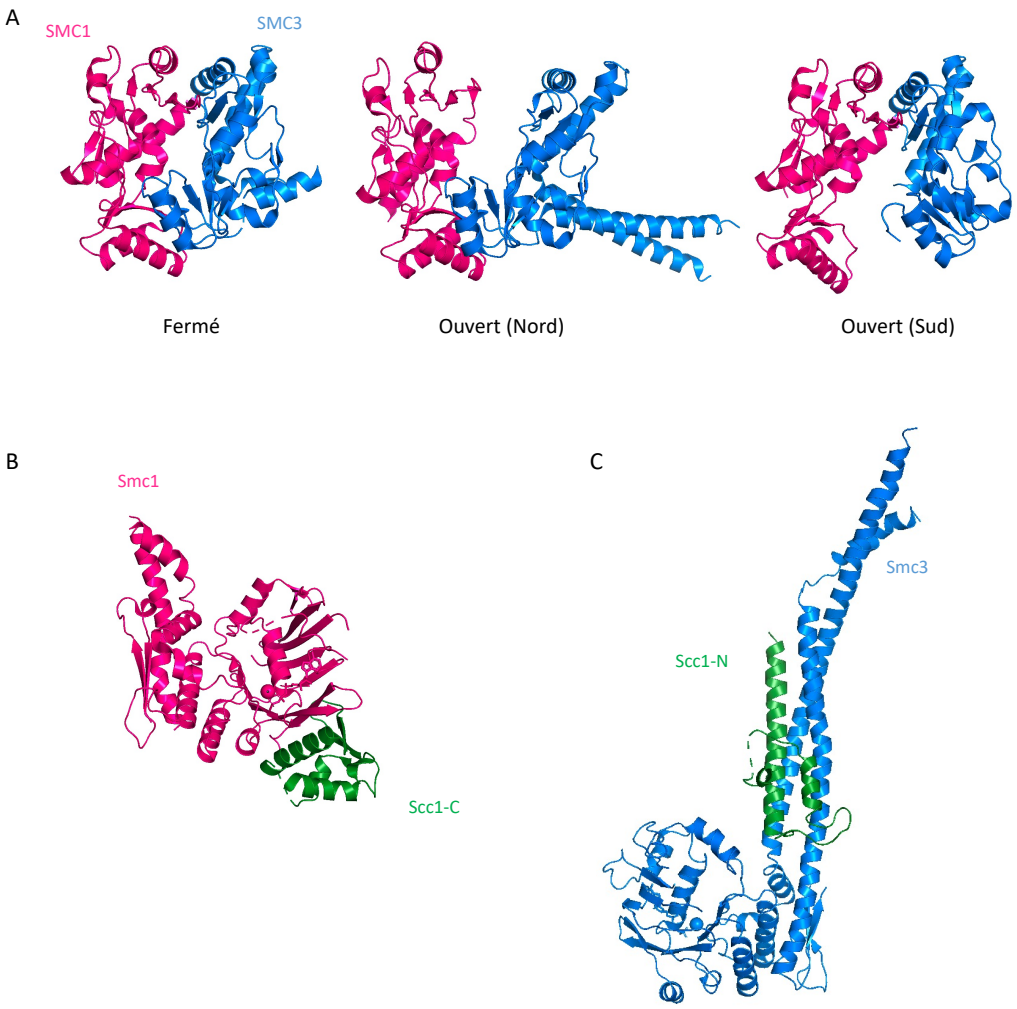


Figure 14 : La Cohésine et ses interfaces SMC1A-SMC3-RAD21.
A. Structure des régions charnières de SMC1A et SMC3 de souris (gauche) et humaines (centre et droite), sous 3 conformations : une conformation fermée, une conformation ouverte au niveau de l'interface nord et une conformation ouverte au niveau de l'interface sud (PDB : 2WD5, 6WG6, 6WG4) (Kurze et al., 2011; Shi et al., 2020).
B. Structure de la tête de Smc1 (SMC1A) de S. cerevisiae. L'extrémité C-terminale de Scc1 (RAD21) interagit avec le capuchon de la tête de Smc1 (PDB : 1W1W) (Haering et al., 2004).
C. Structure de la tête de Smc3 (SMC3) de S. cerevisiae. L'extrémité N-terminale de Scc1 (RAD21) se fixe au niveau du cou de Smc3 (PDB : 4UX3) (Gligoris et al., 2014).

La kleisine RAD21 se lie ensuite aux protéines SMC1A et SMC3. La kleisine RAD21, via son extrémité C-terminale, interagit avec la tête de SMC1A. Plus précisément, cette région C-terminale de RAD21 forme un domaine WH, composé de trois hélices et de deux feuillets β, qui se lie aux deux feuillets β les plus C-terminaux de SMC1A, d'après la structure cristalline à 2,9 Å du complexe Smc1 (SMC1A) / Scc1 (RAD21) de *S. cerevisiae* (Figure 14B) (Haering et al., 2004). La kleisine interagit également avec le cou de SMC3, c'est-à-dire la région de la superhélice proche de la tête ATPase de SMC3, via son extrémité N-terminale. La structure à

3,3 Å de ce complexe chez *S. cerevisiae* résolue en 2014 permet de préciser l'interface entre RAD21 et SMC3 (Gligoris et al., 2014). L'extrémité N-terminale de RAD21 se compose de trois hélices, dont les deux dernières interagissent avec la région en superhélice de SMC3, formant un faisceau de 4 hélices (¨helix bundle¨) (Figure 14C) (Gligoris et al., 2014). La Cohésine, via l'interaction de ses trois protéines cœur, forme ainsi un anneau asymétrique d'environ 40 nm permettant la capture de l'ADN.

4.2.2. Les protéines auxiliaires et régulatrices

La régulation de l'activité de la Cohésine fait intervenir de nombreuses protéines régulatrices (Tableau 2). La majorité d'entre elles se fixent au niveau de la kleisine RAD21. Chez les eucaryotes, nous retrouvons les protéines HAWKs dont font partie les protéines STAG, NIPBL et PDS5, ainsi que la protéine WAPL, l'acétyltransférase ESCO et la désacétylase HDAC8. Les protéines SORORINE et CTCF, spécifiques des vertébrés, se lient également à la Cohésine.

4.2.2.1. La protéine STAG : protéine auxiliaire essentielle pour la Cohésine

La Cohésine s'associe, à travers la région centrale de RAD21, à la protéine HEAT STAG ou SA (¨stromal antigene¨), conservée chez tous les eucaryotes. Cette protéine, toujours liée à l'anneau tripartite et qui est souvent considérée comme la quatrième sous-unité cœur, est indispensable à la stabilité de la Cohésine sur la chromatine, à l'extrusion de boucles d'ADN et sert également d'interface avec certains de ses régulateurs (Haering et al., 2002; Roig et al., 2014; Davidson et al., 2019).

La protéine STAG, composée de dix-sept domaines répétés HEAT, adopte une structure en forme de U, qui est associée à une région N-terminale plus flexible ressemblant à un crochet (Figure 15). Elle interagit avec RAD21 avec l'intérieur de son extrémité en U, via plusieurs interfaces plus ou moins étendues (Hara et al., 2014). Cette protéine peut également lier l'ADN en son centre, ainsi que la région charnière de SMC1A, à travers la base de sa structure en U (Figure 15) (Li et al., 2018; Shi et al., 2020). STAG participe également au recrutement de nombreuses protéines régulatrices, telles que NIPBL, PDS5 et WAPL. Ces interactions sont décrites dans les parties décrivant ces protéines régulatrices.

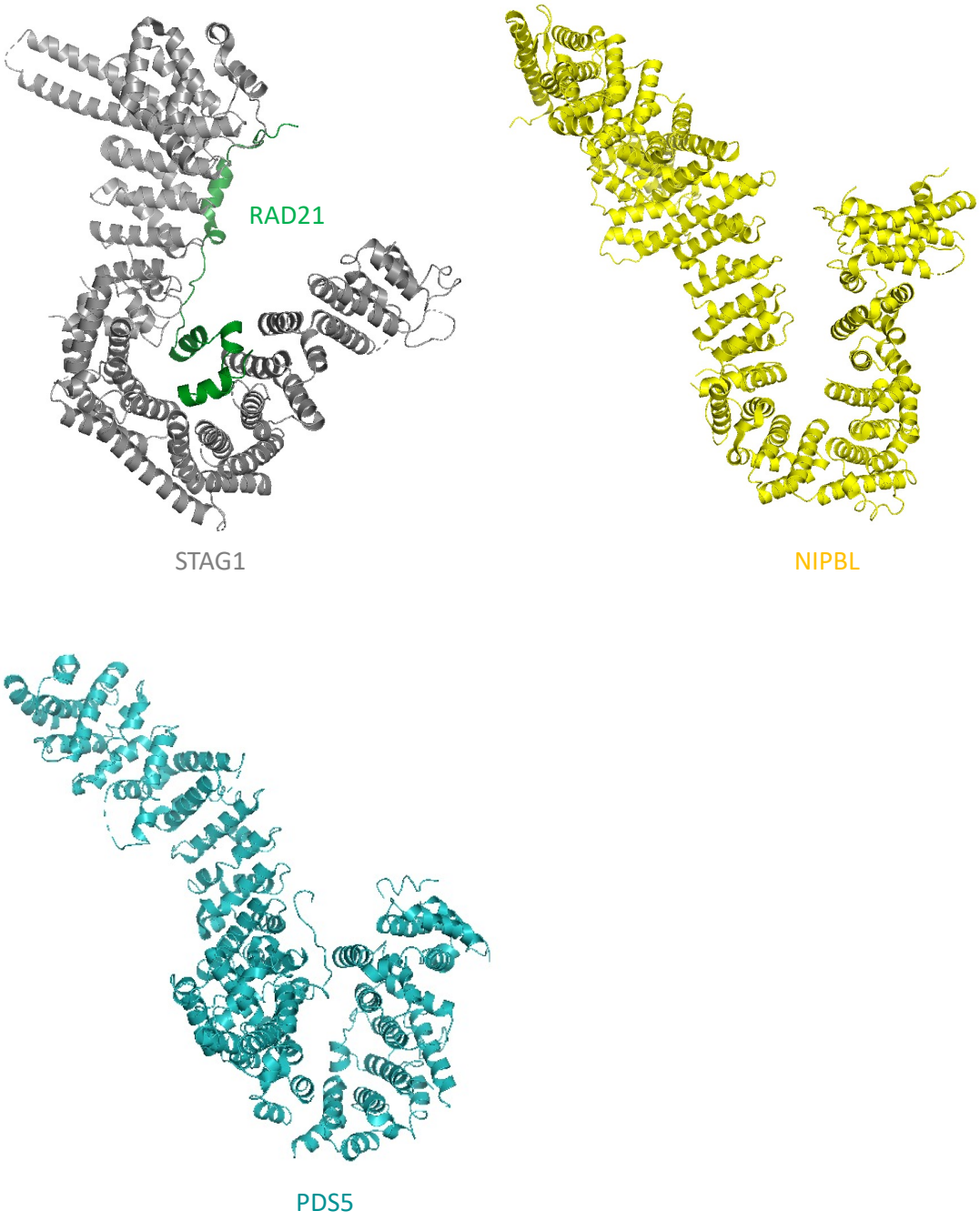


Figure 15 : Structures des protéines auxiliaires et régulatrices HAWKs : STAG1 humain associé à la région centrale de RAD21 (gauche), NIPBL de C. thermophilum (centre) et PDS5 (droite) de L. thermotolerans (PDB : 4PJU, 5T8V, 5F0O).
Ces protéines adoptent toutes une structure en U, avec une extrémité N-terminale en crochet.

Chez l'homme, deux variants mitotiques de STAG, STAG1 et STAG2, sont retrouvés. Ils présentent une homologie de séquence de 70 %, avec une conservation forte au niveau de leurs régions centrales, mais plus faible au niveau des extrémités N- et C-terminales, avec une homologie de 30 - 50 % (Kong Xiangduo et al., 2014).

Des expériences sur des cellules humaines HeLa.2.11 déplétées en STAG1 et STAG2 ont mis en évidence que les sous-unités STAG1 et STAG2 sont toutes les deux impliquées dans la cohésion des chromatides sœurs, avec toutefois des localisations différentes : les complexes Cohésine-STAG1 se retrouvent au niveau des télomères, alors que les complexes Cohésine-STAG2 au niveau des centromères (Canudas and Smith, 2009). STAG2 est majoritairement impliqué dans la réparation des cassures double-brin de l'ADN par recombinaison homologue au cours des phases S et G2 (Kong Xiangduo et al., 2014). Concernant le repliement du génome et la régulation de l'expression de gènes, les protéines STAG1 et STAG2 ont des localisations chevauchantes et différentes. Les complexes Cohésine-STAG1 sont préférentiellement situés dans les grandes boucles, au niveau des sites CTCF et des limites des TADs, alors que les complexes Cohésine-STAG2 sont majoritairement impliqués dans l'organisation plus locales des boucles de chromatine, notamment au niveau des enhancers et des promoteurs, indépendamment de CTCF (Kojic et al., 2018; Cuadrado et al., 2019; Casa et al., 2020). Les cellules peuvent tolérer l'absence d'un des deux variants, mais pas celle des deux, suggérant un mécanisme compensatoire entre STAG1 et STAG2 (Benedetti et al., 2017; van der Lelij et al., 2020). Ainsi, ces deux variants participent à l'ensemble des fonctions de la Cohésine avec des missions à la fois redondantes et distinctes.

4.2.2.2. Les protéines régulatrices

4.2.2.2.1. Le complexe NIPBL/MAU2 : complexe de chargement

Les protéines NIPBL et MAU2 s'associent pour former le complexe de chargement de la Cohésine sur le génome. Ce complexe est nécessaire pour le chargement de la Cohésine sur l'ADN, mais n'est pas requis pour son maintien sur l'ADN (Ciosk et al., 2000). En initiant le chargement, il permet de le contrôler de manière temporelle et spatiale (Kogut et al., 2009).

MAU2 est une protéine composée de 13 motifs de répétions de tétratricopeptides (TRP), adoptant une structure en forme de tube qui englobe l'extrémité N-terminale de NIPBL (Chao et al., 2015; Hinshaw et al., 2015).

L'extrémité C-terminale de NIPBL, composée des motifs HEAT, adopte une structure en forme de crochet et se lie à la fois à l'AD et aux protéines RAD21, STAG, SMC1A et SMC3 (Figure 15). NIPBL interagit avec un court domaine N-terminal de RAD21, via sa région centrale

en U, ainsi qu'avec l'extrémité N-terminale de STAG, via sa région N-terminale. NIPBL peut également se fixer sur les sous-unités SMC1A et SMC3 (Collier et al., 2020; Higashi et al., 2020; Shi Zhubing et al., 2020). Il a également été montré que NIPBL médie l'activation de l'activité ATPase de SMC1A et SMC3 (Murayama and Uhlmann, 2014; Petela et al., 2018).

4.2.2.2.2. Le complexe PDS5 / WAPL / SORORINE : complexe d'association/dissociation

La protéine PDS5 (¨precocious dissociation of sister protein¨) appartient également à la famille des HAWK, avec sa vingtaine de domaines HEAT (Ouyang et al., 2016). Elle joue un rôle important dans l'association de la Cohésine sur l'ADN, notamment pour la cohésion des chromatides sœurs et dans la réparation de l'ADN (Panizza et al., 2000; Wang et al., 2002). Elle adopte une structure en U et se fixe au niveau de la région N-terminale de RAD21, proche de la zone d'interaction entre RAD21 et SMC3 (Figure 15) (Lee et al., 2016). Elle entre en compétition avec NIPBL pour sa liaison avec la Cohésine, les zones d'interaction entre la Cohésine et ses deux protéines auxiliaires étant partiellement chevauchantes (Petela et al., 2018).

Chez les vertébrés, deux variants, PDS5A et PDS5B, sont retrouvés et ont des fonctions redondantes, mais spécifiques dans l'association Cohésine – ADN lors de l'interphase et de la mitose, mais qui restent à définir plus précisément (Al-Jomah et al., 2020; Zhang et al., 2021). Ces variants agissent de manière positive ou négative sur l'association Cohésine-ADN, selon leurs interactions avec les autres protéines régulatrices de la Cohésine, telles que WAPL et la SORORINE.

D'une part, PDS5 permet, à travers son domaine N-terminal, le recrutement de la protéine WAPL (¨wings-apart like¨), conduisant au retrait de la Cohésine sur l'ADN (Kueng et al., 2006). La protéine WAPL, par son extrémité N-terminale composée de motifs FGF, interagit avec la Cohésine, et plus particulièrement avec une courte séquence d'acides aminés centrale de RAD21, ainsi qu'avec la protéine STAG (Kueng et al., 2006; Ouyang et al., 2013).

D'autre part, PDS5, via son domaine N-terminal, peut stabiliser l'interface Cohésine-ADN, par le recrutement de la SORORINE, protéine caractéristique des vertébrés. La SORORINE entre alors en compétition avec WAPL, grâce à son motif EGF, et entraine son déplacement vis-à-vis de PDS5, par un réarrangement conformationnel. Cela conduit ainsi à l'inhibition de la dissociation de la Cohésine de l'ADN induite par WAPL (Nishiyama et al.,

2010a). La SORORINE, à travers son extrémité C-terminale, peut aussi interagir directement avec la Cohésine, via STAG1 ou STAG2 (Zhang and Pati, 2015). La SORORINE est essentielle à la cohésion des chromatides sœurs (Ladurner et al., 2016; Schmitz et al., 2007).

PDS5A et PDS5B peuvent également interagir avec l'acétyltransférase ESCO1, mais pas ESCO2 (Minamino et al., 2015) . De plus, PDS5A lie également la protéine accessoire CTCF (Nora et al., 2020).

4.2.2.2.3. Les protéines ESCO et HDAC8 : stabilisation de l'interface Cohésine-ADN et recyclage

L'association de la Cohésine sur l'ADN est également stabilisée par l'acétyltransférase ESCO, qui est recrutée par PDS5 (Chan et al., 2013). Cette enzyme acétyle les lysines K105 et K106 (chez l'homme), situées sur la tête ATPase de SMC3 et conservées au cours de l'évolution (Ivanov et al., 2002). Cette acétylation assure une liaison stable entre la Cohésine et l'ADN, réduit la mobilité de la Cohésine et participe au recrutement de la SORORINE.

L'homme possède deux orthologues ESCO, ESCO1 et ESCO2. Ces deux variants participent à la cohésion des chromatides sœurs, avec des spécificités mécanistiques, spatiales et temporelles. En effet, la présence ESCO2, et non celle de ESCO1, est essentielle au niveau des centromères (Whelan et al., 2012). L'expression de ESCO1 est stable tout au long du cycle cellulaire, alors que celle d'ESCO2 est fortement augmentée durant la phase S (Hou and Zou, 2005). De plus, seul ESCO 1 présente le site d'interaction de PDS5, dont la liaison est requise pour l'acétylation de SMC3, ESCO2 devant donc faire appel à un autre mécanisme de recrutement (Minamino et al., 2015).

Ces deux variants semblent également présenter d'autres activités que l'acétylation de SMC3, le mimétisme de l'acétylation de SMC3 dans des cellules déplétées en ESCO1 et ESCO2 ne permettant pas une cohésion correcte des chromatides sœurs (Kawasumi et al., 2017). Toutefois, leurs rôles restent à définir.

Par la suite, à la fin du cycle cellulaire, les deux lysines sont désacétylées par l'histone désacétylase 8 (HDAC8), permettant son utilisation dans un nouveau cycle. Chez les vertébrés, HDAC8 possède également un rôle dans la formation de boucles de chromatine, agissant ainsi en opposition à WAPL pendant le cycle cellulaire (van Ruiten et al., 2022).

4.2.2.2.4. Le complexe SHUGOSHIN/PP2A : complexe protecteur

La protéine SHUGOSHINE est une phosphoprotéine qui interagit avec la phosphatase PP2A, formant un complexe se liant aux Cohésines-ADN centromériques. Ce complexe permet de prévenir la dissociation précoce entre la Cohésine et la chromatine au niveau des centromères lors de la mitose (Watanabe and Kitajima, 2005; Xu et al., 2009). SHUGOSHINE se lie à STAG et à RAD21, avec la même interface que WAPL, protégeant ainsi les interactions Cohésine-ADN, de l'action dissociative de WAPL. Ce complexe assure également le maintien de la SORORINE sur la Cohésine et empêche la phosphorylation du complexe de la Cohésine, notamment celle de STAG, évitant ainsi sa dissociation (Kitajima et al., 2006; Nishiyama et al., 2010b; Hara et al., 2014). Les mammifères possèdent deux variants de la SHUGOSHINE, SGO1 et SGO2, qui diffèrent en taille (respectivement 72 kDa et 144 kDa) et présentent des fonctions chevauchantes de protection de la cohésion centromérique lors de la mitose (Yao and Dai, 2012). SGO2 est également impliqué dans la bi-orientation des chromosomes, en participant à l'attachement des microtubules aux kinétochores et en servant de protéine de reconnaissance pour de nombreux facteurs impliqués dans la bi-orientation (Marston, 2015).

4.2.2.2.5. La protéine CTCF : protéine d'ancrage

La protéine CTCF (¨CCCTC-binding factor¨), protéine à 11 doigts de zinc, est conservée chez la majorité des métazoaires et se lie à la séquence d'ADN consensus d'une vingtaine d'acides aminés (Holohan et al., 2007). Cette protéine sert de point d'ancrage à la Cohésine pour la formation de boucles d'ADN selon son orientation, favorise la stabilité de cette dernière sur la chromatine. L'extrémité N-terminale de CTCF interagit directement avec STAG, via la même interface que celle liant WAPL et STAG. CTCF interagit également avec PDS5, via la région permettant la liaison de WAPL et de la Cohésine. CTCF entre alors en compétition avec WAPL, empêchant ainsi le retrait la Cohésine induit par WAPL (Li et al., 2020; Nora et al., 2020).

4.3. Le cycle de la Cohésine

4.3.1. Modes d'interaction entre la Cohésine et l'ADN

Selon ses fonctions, la Cohésine peut interagir avec l'ADN de manière topologique ou non-topologique/pseudo-topologique. Cette dernière correspond à l'entrée de l'ADN sans l'ouverture de l'anneau de la Cohésine. Cette conformation est observée pour le chargement de la Cohésine, ainsi que pour l'extrusion de boucles (Srinivasan et al., 2018; Golov et al., 2021).

La conformation topologique de la Cohésine et de l'ADN correspond à la capture de l'ADN à l'intérieur de la structure annulaire de la Cohésine et requiert l'ouverture transitoire des portes d'entrée/de sortie de la Cohésine. Cette capture topologique est notamment nécessaire pour la cohésion des chromatides sœurs (Haering et al., 2008; Srinivasan et al., 2018). La Cohésine possède deux portes d'entrée/de sortie, une des portes est formée par l'ouverture de l'interface entre les deux régions charnières des protéines SMC1A et SMC3, l'autre correspond à l'ouverture de l'interface entre le cou de SMC3 et l'extrémité N-terminale de RAD21 et est appelé porte de sortie de l'ADN (Kurze et al., 2011; Eichinger et al., 2013; Gligoris et al., 2014; Srinivasan et al., 2018; Higashi et al., 2020).

4.3.2. Mécanisme impliqué dans la cohésion des chromatides sœurs

Pour remplir ses fonctions dans la cohésion des chromatides sœurs, la Cohésine est associée puis dissociée de la chromatine tout au long du cycle cellulaire. Son cycle de régulation est très complexe et fait appel à de nombreux partenaires (Figure 16).

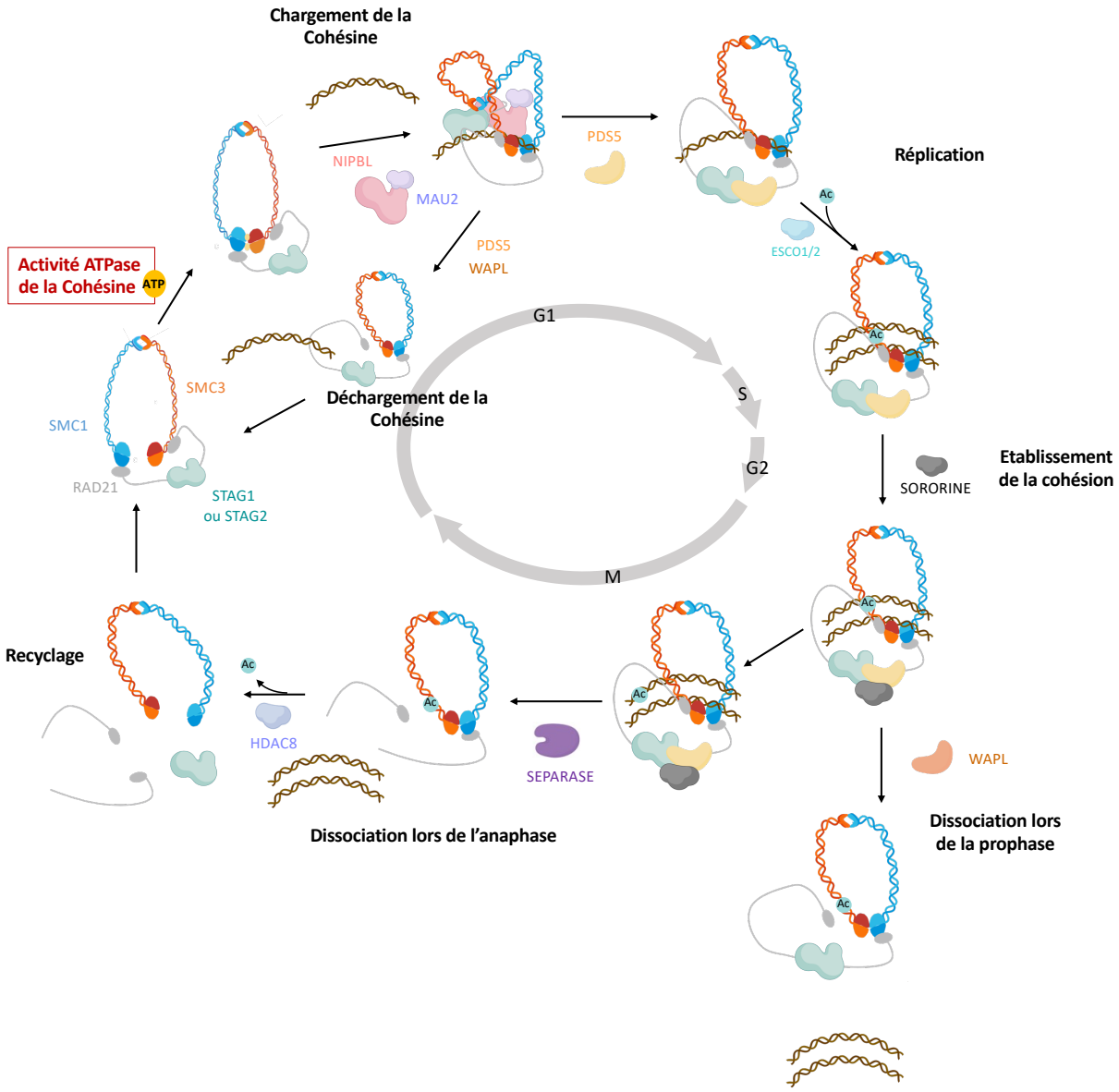


Figure 16 : Schéma du cycle de la cohésine des vertébrés, établissant la cohésion des chromatides sœurs et sa dissolution au cours du cycle cellulaire.

Lors de la fin de la télophase et au cours de la phase G1, la Cohésine est chargée sur la chromatine par le complexe de chargement NIPBL/MAU2. Ce chargement requiert la liaison et l'hydrolyse de l'ATP par les têtes ATPases de SMC1A et SMC3. Il existe deux portes d'entrée envisageables pour la capture de l'ADN au sein de la Cohésine, une à l'interface entre l'extrémité N-terminale de RAD21 et le cou de SMC3, la seconde à l'interface entre les deux régions charnières de SMC1A et SMC3. Lors de la phase G1, la Cohésine peut être déchargée par le complexe WAPL/PDS5. Au cours de la phase S, après la réplication de l'ADN, la cohésion des chromatides sœurs est établie par le chargement de PDS5, de l'acétylation de SMC3, par l'acétyltransférase ESCO1/ESCO2 et par le recrutement de la SORORINE. Cette cohésion est maintenue jusqu'à la prophase, où la majorité des cohésines, excepté celles se situant au niveau des centromères, est dissociée de la chromatine par un mécanicisme dépendant de WAPL et de kinases mitotiques, qui phosphorylent STAG1 ou STAG2 et la SORORINE (non représenté). Au niveau des centromères, la Cohésine est protégée par le complexe SGO1-PP2A (non représenté) et est dissociée lors de l'anaphase, par le clivage de RAD21 par la Séparase, permettant la séparation des chromatides sœurs. SMC3 est ensuite désacétylé par HDAC8 et la Cohésine peut être réutilisée pour le cycle suivant.

4.3.2.1. Chargement de la Cohésine sur l'ADN

La Cohésine, composée de SMC1A, SMC3, RAD21 et STAG1 ou STAG2, est chargée sur les chromosomes grâce au complexe de chargement NIPBL/MAU2 et à l'activité ATPase de SMC1A et SMC3, lors de la télophase et de la phase G2 (Ciosk et al., 2000; Hu et al., 2011). Chez l'homme, la Cohésine est également associée à l'ADN en début de phase S. Ce processus requiert le complexe NIPBL/MAU2, ainsi que l'hélicase réplicative MCM2-7 et la kinase D (Zheng et al., 2018). Chez *S. cerevisiae*, la Cohésine est chargée sur les chromosomes au niveau des centromères, des promoteurs de gènes fortement exprimés, des fourches de réplication et des sites de lésion de l'ADN. Le chargement de la Cohésine implique le remodeleur de la chromatine RSC, qui interagit directement avec le complexe de chargement de la Cohésine Scc2/Scc4 et fournit une matrice exempte de nucléosome (Lopez-Serra et al., 2014; Muñoz et al., 2020). Conservé de la levure à l'homme, le chargement de la Cohésine, notamment au niveau des gènes fortement transcrits, semble également impliqué le complexe Médiateur, qui est un récepteur direct de NIPBL/MAU2, (Kagey et al., 2010; Mattingly et al., 2022).

Avant la réplication de l'ADN, la Cohésine est dynamique, son maintien sur l'ADN nécessite NIPBL et sa libération est induite par le complexe PDS5/WAPL, qui se fixe au niveau de RAD21 et STAG1 ou STAG2 (Ouyang et al., 2013, 2016). Cette libération dépendante de WAPL conduit à la dissociation de la porte de sortie de l'ADN (interface SMC3/RAD21), permettant la sortie de l'ADN (Eichinger et al., 2013; Beckouët et al., 2016). Des molécules de Cohésine peuvent être également chargées *de novo* sur l'ADN naissant au cours de la phase S, au niveau des fourches de réplications (Srinivasan et al., 2020).

4.3.2.2. Établissement de la cohésion des chromatides sœurs

Après la réplication de l'ADN, la cohésion des chromatides sœurs est induite par la stabilisation de l'interaction entre la Cohésine et l'ADN (Gerlich et al., 2006b). Ce processus implique le remplacement de NIPBL par PDS5, ainsi que l'acétylation de deux lysines K105 et K106 de SMC3 par ESCO1 ou ESCO2 chez l'homme (K112 et K113 chez *S. cerevisiae*). Cette acétylation, facilitée et protégée par PDS5, est suffisante pour bloquer l'activité de WAPL chez les levures, mais insuffisantes pour les mammifères (Beckouët et al., 2010; Chan et al., 2013). Chez ces derniers, le recrutement de la SORORINE est requis pour contrer la libération de

l'ADN induite par WAPL et maintenir la cohésion des chromatides sœurs jusqu'à l'anaphase (Zhang et al., 2008; Nishiyama et al., 2010a; Ladurner et al., 2016).

De plus, l'établissement de cette cohésion est lié à de nombreux composants du réplisome, tels que WDHD1, TIMELESS,TIPIN ou DDX11 (Ctf4, Tof1, Csm3, Chl1 chez la levure), dont leurs dérégulations entrainent un défaut de cohésion dans l'ensemble des eucaryotes (Parish et al., 2006; Leman et al., 2010; Samora et al., 2016; Zheng et al., 2018). Deux mécanismes, nécessitant différentes protéines associées aux réplisomes, ont été proposés chez *S. cerevisiae*. La cohésion établie par des Cohésines préalablement chargées lors de la télophase ou la phase G1 fait appel aux protéines Tof1, Csm3, Chl1 et Ctf4, alors que la cohésion établie par des Cohésines nouvellement chargées est dépendante du complexe Ctf18-RFC (Srinivasan et al., 2020).

4.3.2.3. Retrait de la Cohésine et dissolution de la cohésion

Une différence majeure est observée pour cette étape entre la levure et les vertébrés. Chez *S. cerevisiae*, toutes les Cohésines se dissocient de l'ADN lors de la transition entre la métaphase et l'anaphase, alors que chez les vertébrés, la majorité des Cohésines sont libérées de la chromatine lors de la prophase. Seule une petite fraction enrichie au niveau des centromères reste accrochée à la chromatine jusqu'au début de l'anaphase (Waizenegger et al., 2000).

Le processus de dissociation des Cohésines lors de la prophase est médié par un ensemble de phosphorylases. Les kinases mitotiques Plk1, Cdk1 et Aurora B phosphorylent respectivement STAG1 ou STAG2 et la SORORINE, entrainant alors le départ de cette dernière et son remplacement par WAPL. WAPL entraine alors la dissociation de l'interface SMC3 et RAD21, permettant la libération des bras chromosomiques. La Cohésine présente au niveau des centromères est alors protégée par le complexe SGO1–PP2A, qui permet de maintenir la Cohésine et la SORORINE dans un état hypophosphorylé et d'empêcher l'interaction physique entre WAPL et STAG1 ou STAG2 (Liu et al., 2013; Hara et al., 2014). Lors de la transition métaphase-anaphase, la Séparase est activée et permet le clivage de la Cohésine restante au niveau des centromères chez l'homme et de toutes les Cohésines chez la levure, entrainant ainsi la perte de la cohésion des chromatides sœurs. Ces dernières peuvent alors migrer vers les pôles opposés de la cellule. Enfin, une fois la Cohésine dissociée de l'ADN, SMC3 est

désacétylé par l'histone désacétylase HDAC8, permettant le recyclage et la réutilisation de la Cohésine pour le cycle cellulaire suivant (Borges et al., 2010; Deardorff et al., 2012).

4.3.3. Mécanisme impliqué dans l'extrusions des boucles de chromatine

Les fonctions de la Cohésine dans l'organisation du génome, dans la régulation de l'expression des gènes et dans la recombinaison sont médiées par la capacité de la Cohésine à organiser la chromatine en boucle par un mécanisme d'extrusion. Ce processus est symétrique et bidirectionnel et nécessite des interactions non topologiques entre la Cohésine et l'ADN (Srinivasan et al., 2018; Kim et al., 2019; Golov et al., 2021). Ce procédé lui permet d'agrandir des boucles jusqu'à ce qu'elle rencontre un site occupé par CTCF. Ce processus requiert également l'activité ATPase de SMC1A et SMC3 (Vian et al., 2018; Davidson et al., 2019; Kim et al., 2019). L'importance de cette activité ATPase sera discutée dans la partie 4.6. Le module ATPase de la Cohésine, acteur clé de son mode d'action.

De nombreuses questions demeurent concernant ce processus et ses mécanismes sous-jacents, mais plusieurs études ont mis en évidence l'implication de multiples protéines auxiliaires de la Cohésine, telles que CTCF, NIPBL/MAU2, WAPL, PDS5, ESCO1 ainsi que d'autres complexes cellulaires, comme celui de la maintenance de minichromosome (MCM), qui permet également de délimiter les boucles de chromatine (Dequeker et al., 2022).

Le complexe NIPBL/MAU2 participe à la formation et au maintien des boucles *in vitro*, en promouvant notamment l'hydrolyse de l'ATP (Davidson et al., 2019; Kim et al., 2019). PDS5A remplace NIPBL sur la Cohésine, mais semble avoir des fonctions antagonistes dans l'extrusion de boucles, du fait qu'il ne stimule pas l'activité ATPase des têtes de SMC1A et SMC3 (Petela et al., 2018). PDS5B n'a pas de rôle prépondérant dans l'extrusion de boucles de chromatine (van Ruiten et al., 2022). WAPL favorise la dissociation de la Cohésine de la chromatine influençant ainsi la taille des boucles (Haarhuis et al., 2017).

Le facteur CTCF définit les limites des boucles en agissant comme une barrière. Il joue également un rôle protecteur vis-à-vis de la Cohésine, comme la SORORINE le fait dans la cohésion des chromatides sœurs. CTCF empêche la fixation de WAPL sur la Cohésine par compétition, stabilisant ainsi la Cohésine sur la chromatine. CTCF médie donc la longueur et la position des boucles de chromatine (Li et al., 2020). Il a été montré que ESCO, par l'acétylation de SMC3 et combiné à la présence de CTCF, favorise la stabilité de la Cohésine

sur la chromatine. Ce rôle a été attribué du fait de la protection de la Cohésine vis-à-vis de WAPL, selon un mécanisme similaire à celui observé dans la cohésion des chromatides sœurs (Wutz et al., 2020). Toutefois, il a été montré récemment que l'acétylation de SMC3 jouerait également un rôle limitant en association avec PDS5 dans l'extrusion de boucles selon un mécanisme indépendant de WAPL. Au vu de la non stimulation de l'activité ATPase de la Cohésine par PDS5A, Wutz et a ont proposé un modèle selon lequel l'acétylation par ESCO et PDS5 empêcheraient l'élargissement des boucles, mettant en pause l'extrusion de boucles (Wutz et al., 2017; van Ruiten et al., 2022). Cela a également été observé chez la levure (Bastié et al., 2022; Dauban et al., 2020).

Ainsi, l'ensemble des fonctions de la Cohésine requiert une régulation fine et orchestrée et implique de nombreux partenaires. Plusieurs modèles d'extrusion de boucles par la Cohésine ont été proposés, mais d'autres études sont nécessaires afin de mieux comprendre comment la Cohésine piège l'ADN et dans quel compartiment K, S ou S-K se trouve l'ADN tout au long de son cycle (Nichols and Corces, 2018; Higashi et al., 2020; Bauer et al., 2021; van Ruiten et al., 2022).

4.4. État de l'art : les structures de la Cohésine connues à ce jour

Au début de ma thèse, seules les structures résolues par diffraction aux rayons X des deux têtes ATPases isolées de la Cohésine (PDB : 1W1W, 4UX3), ainsi que celles des régions charnière (PDB : 1GXL, 2WD5) étaient disponibles pour le complexe cœur de la Cohésine (Figure 14) (Gligoris et al., 2014; Haering et al., 2004, 2002; Kurze et al., 2011).

Depuis, grâce à l'essor de la cryoEM, plusieurs reconstitutions de la Cohésine sous différentes formes ont été obtenues et permettent de préciser certaines étapes du cycle cellulaire.

D'une part, une structure à une résolution de 3,3 Å des têtes engagées d'un complexe chimérique Smc1, Smc3 et Scc1 de *C. thermophilum* et *S. cerevisiae,* en présence d'ATP-γ-S et de réticulation *in vitro,* a été obtenue et suggère que l'engagement des têtes entraine l'ouverture de l'interface entre Scc1 et Smc3 (Figure 17A) (PDB : 6QPW, EMD-4616) (Muir et al., 2020). Cette structure pourrait également correspondre à l'état qu'adopte la Cohésine après sa dissociation de l'ADN, du fait de l'ouverture de cette porte de sortie.

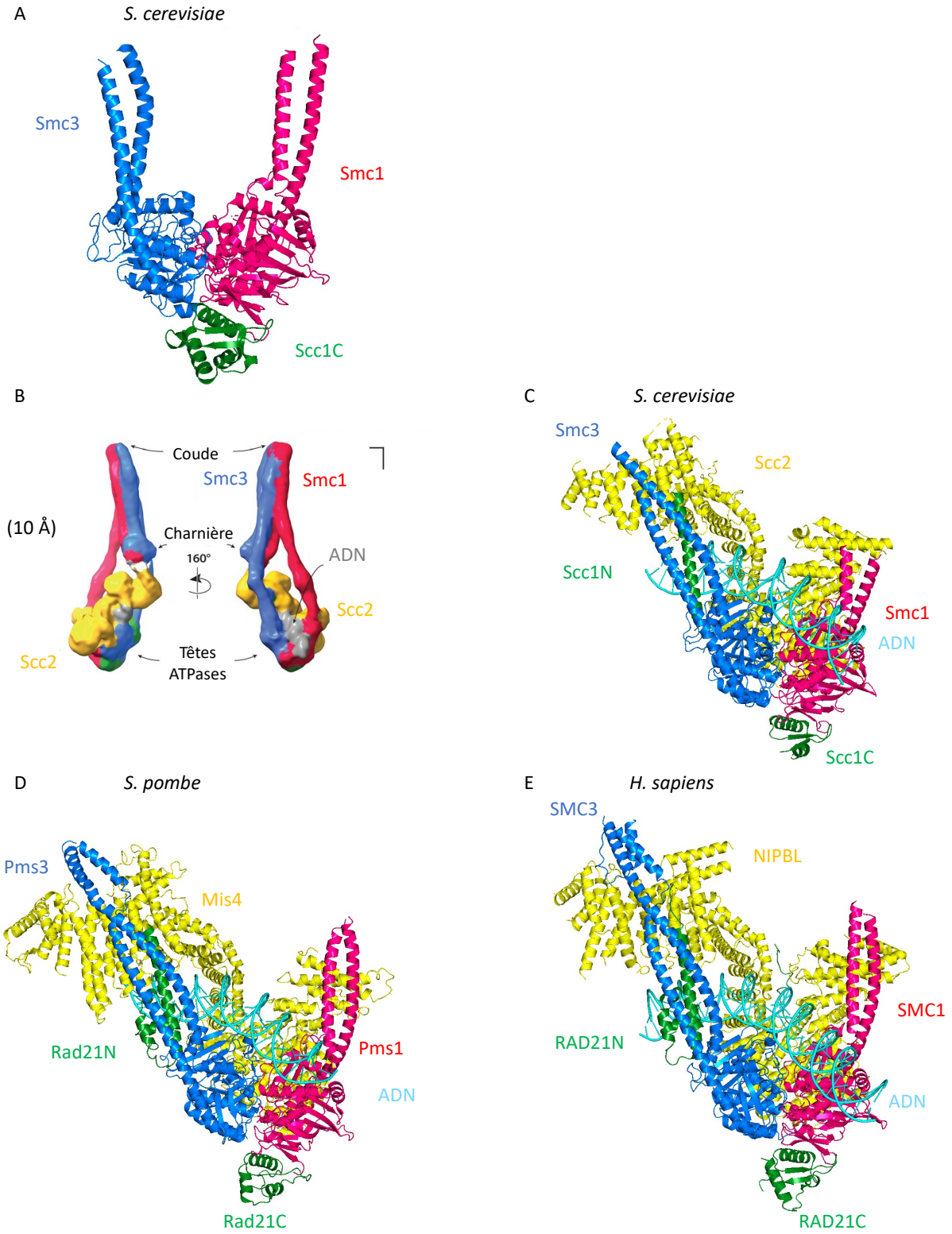


Figure 17 : Les différentes structures connues de la Cohésine.

A. Structure du complexe chimérique Smc1/Smc3/Scc1 de C. thermophilum et S. cerevisiae, en présence d'un analogue de l'ATP faiblement hydrolysable, l'ATP-γ-S. Les têtes ATPases de Smc1 et Smc3 sont engagées. Leur engagement semble induire l'ouverture de l'interface entre l'extrémité N-terminale de RAD21 et le cou de Scm3 (PDB : 6QPW, EMD-4616) (Muir et al., 2020).

B. Carte cryoEM de 10 Å de la Cohésine chez S. cerevisiae, montrant l'ouverture des régions en superhélice au niveau des têtes ATPases en présence d'ADN et d'ATP. La fixation de l'ATP et de l'ADN entraine un

réarrangement de ces régions en superhélice, par comparaison à la structure apo, où ces régions sont rapprochées l'une de l'autre. Adapté de (Collier et al., 2020).
C-D-E. Reconstitution des complexes tripartites de la Cohésine (SMC1A-SMC3-RAD21), associés à NIPBL et à l'ADN, en présence d'ATP ou d'analogues faiblement hydrolysables de l'ATP, chez trois espèces différentes : S. cerevisiae, S. pombe et chez l'homme. Ces complexes adoptent alors une conformation clampée, avec des similarités inter-espèces, mais quelques spécificités propres à chaque espèce (respectivement PDB : 6ZZ6, EMD-11585 – PDB : 6YUF, 3,94 Å, EMD-10930 et PDB : 6WGE, 3.9 Å, EMD-21663) (Collier et al., 2020; Higashi et al., 2020; Shi et al., 2020).

D'autre part, comme observé pour la Condensine et MukBEF, plusieurs reconstructions de l'état clampé de la Cohésine ont été résolues chez *S. cerevisiae* (PDB : 6ZZ6, 3,4 Å, EMD-11585), *S. pombe* (PDB : 6YUF, 3,94 Å, EMD-10930) et chez l'homme (PDB : 6WGE, 3.9 Å, EMD-21663) (Figure 17C-D-E) (Collier et al., 2020; Higashi et al., 2020; Shi et al., 2020). Cette conformation correspond à la seconde étape du cycle de la Cohésine, celle de son chargement sur l'ADN par NIPBL.

Dans cette conformation clampée, le complexe trimérique, SMC1A/Smc1/Pms1, SMC3/Smc3/Pms3 et RAD21/Scc1/Rad21, est associé à NIPBL/Scc2/Mis4, à l'ADN et à l'ATP ou à des analogues de l'ATP non hydrolysables, tels que l'ADP.BeF_3^- et l'AMP.PNP (homme/*S. cerevisiae/S. pombe*). Seule la partie inférieure des complexes a pu être reconstruite en modèle atomique, la partie supérieure correspondant aux régions en superhélice de SMC1A et SMC3 étant trop peu résolutive. Dans cette conformation clampée, les différentes structures présentent une organisation similaire, où l'ADN est piégé dans un tunnel formé par l'ensemble des composants de la Cohésine et de la protéine de chargement NIPBL/Scc2/Mis4. L'ADN interagit avec la majorité des sous-unités, avec toutefois quelques particularités espèces-dépendantes.

Notamment, les têtes ATPases de SMC1A/Smc1/Psm1 et SMC3/Smc3/Psm3 sont engagées et interagissent respectivement avec les extrémités C- et N-terminales de RAD21/Scc1/Rad21. Leurs régions en superhélice proches des têtes sont ouvertes, permettant à l'ADN de se lier à la surface supérieure des têtes ATPases. Le reste des régions en superhélice adoptent une conformation coudée et sont maintenues proche l'une de l'autre (Figure 17B) (Collier et al., 2020; Higashi et al., 2020; Shi et al., 2020).

Dans les trois structures, la région centrale de RAD21 n'a pas pu être reconstruite, du fait que cette protéine très flexible montre très peu ou pas de densité, mais les régions N- et C-terminales liées à SMC1A/Smc1/Psm1 et SMC3/Smc3/Psm3 ont pu l'être. L'engagement des têtes associé à NIPBL/Scc2/Ms4 et à l'ADN n'entraine pas l'ouverture de l'interface entre le

cou de SMC3/Smc3/Psm3 et l'extrémité N-terminale de RAD21/Scc1/Rad21. De plus, l'extrémité N-terminale de RAD21/Rad21 humaine et de *S. pombe*, via une boucle chargée positivement, interagit avec l'ADN. Cette boucle n'est pas présente chez *S. cerevisiae* (Collier et al., 2020; Higashi et al., 2020; Shi et al., 2020).

La protéine NIPBL/Scc2/Mis4 s'ancre, à travers sa région centrale en U au niveau des têtes de SMC1A et SMC3, au niveau de leurs lobes C- et N-terminaux respectivement. NIPBL contacte également le cou de SMC1A et SMC3, via ses régions N- et C-terminale chez *S. pombe* et l'homme. Chez *S. cerevisiae* une interaction entre Scc2 et le cou de Smc3 peut être définie, mais pas pour le cou de Smc1. NIPBL permet ainsi de stabiliser l'engagement des têtes et de fixer l'ADN sur ces têtes ATPases (Collier et al., 2020; Higashi et al., 2020; Shi et al., 2020).

Une reconstruction de l'état clampé a également été obtenue à une résolution de 5,3 Å, mais avec le complexe tétramérique humain (SMC1A, SMC3, RAD21, STAG1) associé à NIPBL et à l'ADN en présence d'un analogue de l'ATP faiblement hydrolysable, l'AMP.PNP (PDB : 6WG3, EMD-21658) (Figure 18) (Shi et al., 2020). Les interactions entre SMC1A, SMC3, RAD21, NIPBL et l'ADN sont similaires à celles décrites dans les paragraphes précédents. Concernant STAG1, il s'amarre à l'arrière du complexe Cohésine-NIPBL-ADN. La partie basse de son U interagit avec les régions charnières de SMC1A-SMC3, principalement par le domaine de SMC1A. STAG1 et NIPBL participeraient donc potentiellement au rapprochement des régions charnières et des têtes ATPases. L'extrémité N-terminale de STAG1 se lie également à l'extrémité N-terminale de NIPBL (Shi et al., 2020).

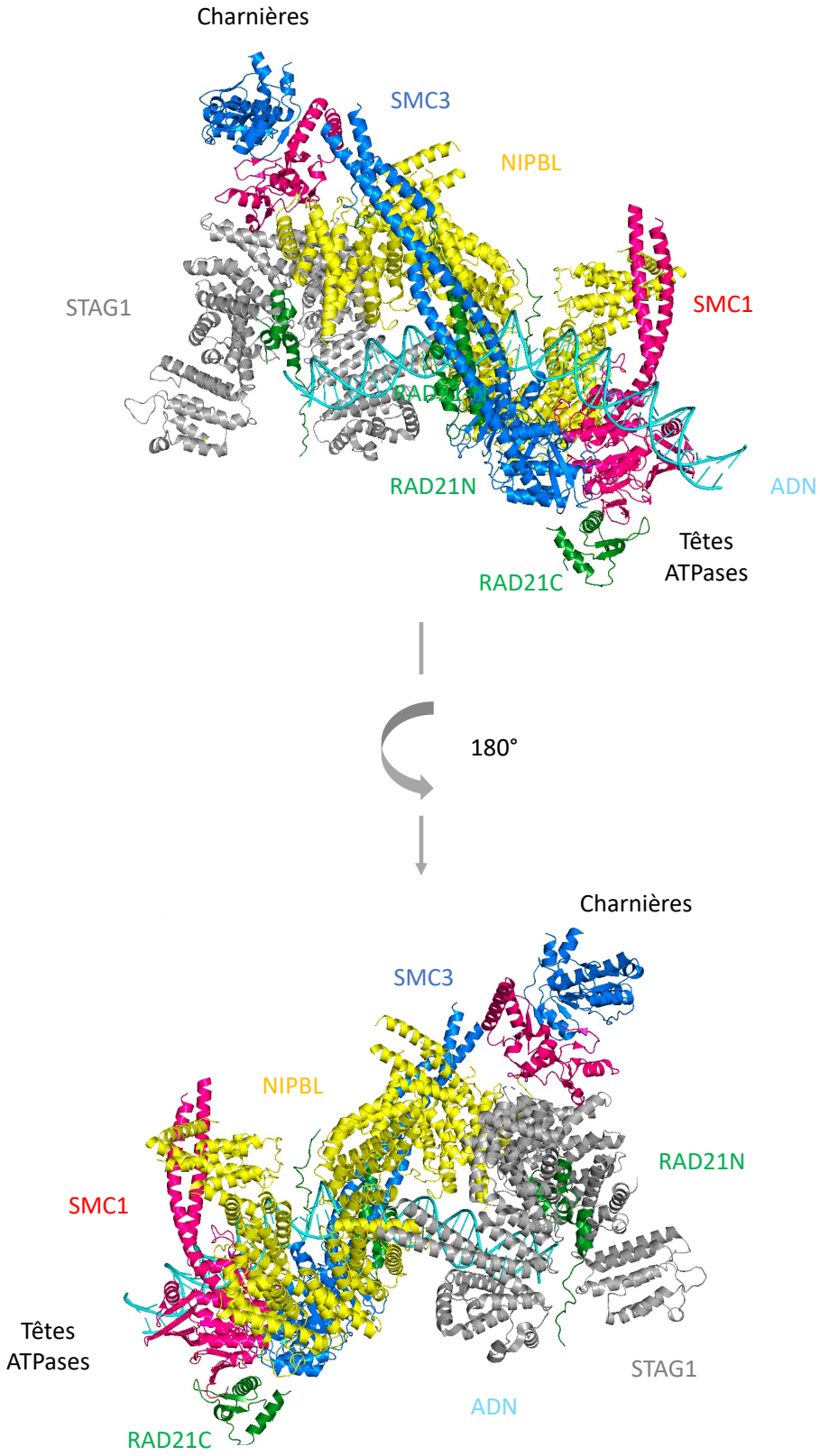


Figure 18 : Reconstruction à 5,3 Å du complexe Cohésine tétramérique humain (SMC1A, SMC3, RAD21 et STAG1), associé à NIPBL et à l'ADN, en présence d'AMP.PNP, sous sa conformation clampée.
Les régions charnières sont rapprochées des têtes ATPases, stabilisées par leurs interactions avec NIPBL et STAG1. L'ADN interagit avec toutes sous-unités présentes, SMC1A, SMC3, RAD21, STAG1 et NIPBL et est piégé dans un tunnel formé par toutes les protéines (PDB : 6WG3, EMD-21658) (Shi et al., 2020).

Par ailleurs, des reconstructions pseudo-atomiques de la conformation apo du complexe trimérique de S. cerevisiae Smc1 (SMC1A), Smc3 (SMC3) et Scc1 (RAD21), associé à Pds5 (PDS5) (EMD-12888) ou Scc2 (NIPBL) (EMD12880), en absence d'ADN et de nucléotide ont été obtenues plus récemment en 2021 (Figure 19) (Petela et al., 2021). Toutefois, l'assimilation de ces structures à une étape du cycle de la Cohésine reste abstraite, du fait des fonctions respectives de Scc2 et de Pds5 dans le chargement de la Cohésine sur l'ADN et sa stabilisation. Dans ces deux structures, le complexe trimérique adopte une structure coudée, où les régions en superhélices de Smc1 et Smc3 sont repliées au niveau de leur coude, rapprochant les régions charnières des têtes ATPases de ces protéines, qui sont séparées d'une distance d'environ 10 nm. Ce repliement est similaire, quelle que soit la protéine HEAT associée au complexe trimérique.

Un modèle pseudo-atomique du complexe Cohésine-Scc2 a pu être résolu, la partie base du complexe lié à Scc2 avec une résolution de 13 Å et les régions en superhélice avec une résolution de 5,5 Å. Scc2 se lie à la tête de SMC1A, mais n'a aucun contact avec Smc3. L'extrémité N-terminale de Scc2 interagit avec la région charnière de Smc1 (Figure 19A-B) (PDB : 7OGT) (Petela et al., 2021). La reconstruction du complexe Cohésine-Pds5 a été obtenue avec une résolution de 13 Å (Figure 19C-D). De manière surprenante, Pds5 ne se lie pas à la tête de Smc1, comme Scc2, mais via sa région C-terminale, interagit avec la tête de Smc3. Cette région de SMC3 comprend les lysines acétylées par Eco1, inhibant la libération de la Cohésion de l'ADN. Cela pourrait ainsi expliquer comment Pds5 peut d'une part promouvoir l'acétylation des lysines de Smc3 en se fixant à la tête de Smc3 et en recrutant la protéine Eco1 et d'autre part la protection de ces lysines de la désacétylation au cours de la phase G2 et de la mitose (Vaur et al., 2012; Chan et al., 2013; Petela et al., 2021). Il est également possible que l'acétylation médiée par Eco1 puisse favoriser la liaison de Pds5 sur la Cohésine, empêchant ainsi la fixation de Scc2, Pds5 et Scc2 étant en compétition pour la liaison à Scc1 (Petela et al., 2018; Dauban et al., 2020).

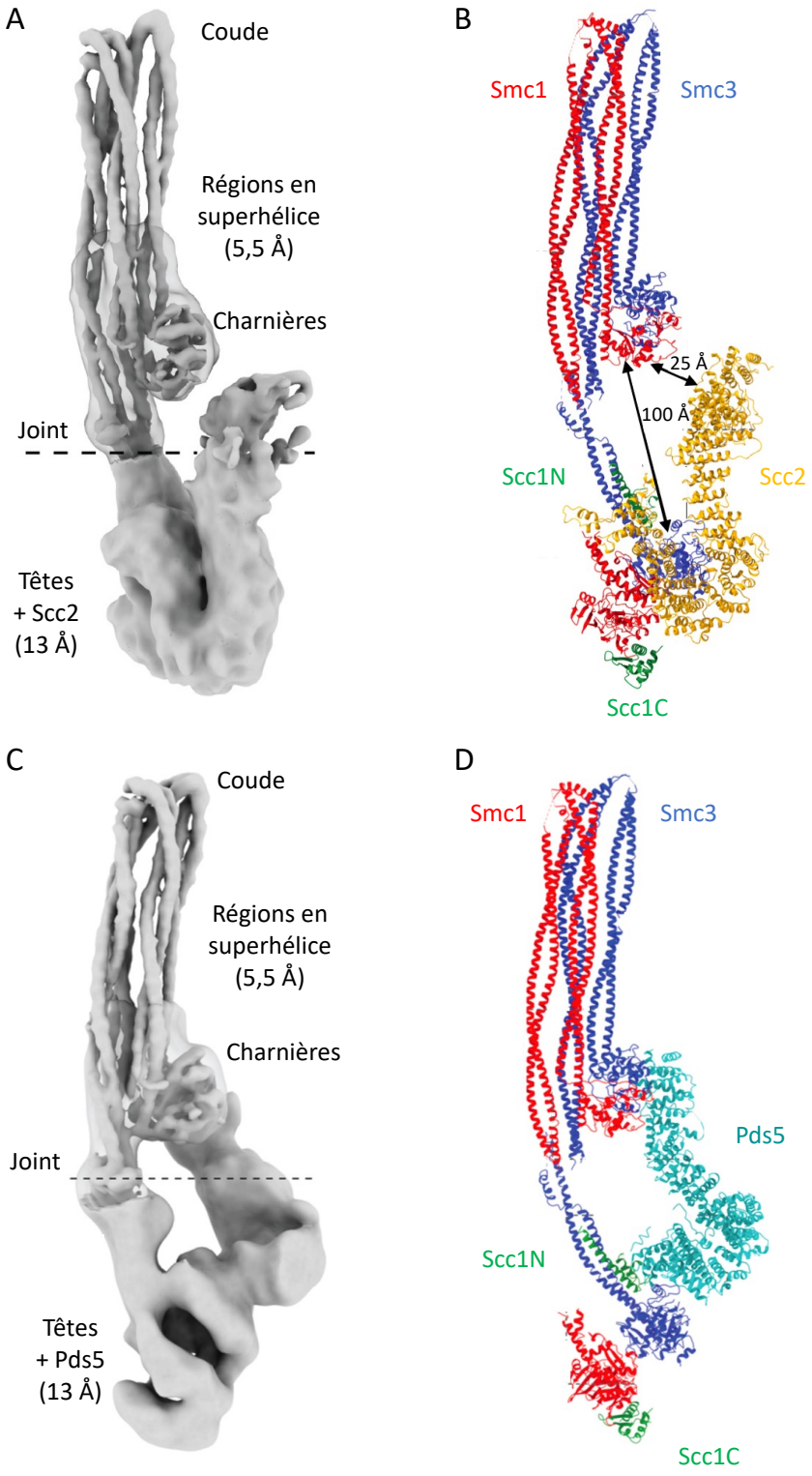


Figure 19 : Le complexe tripartite associé à Scc2 ou Pds5 de S. cerevisiae, sous une forme apo, sans nucléotide et sans ADN.

A-B. Carte et reconstruction pseudo-atomique du complexe Cohésine tripartite-Scc2 (PDB : 7OGT, EMD-12880). Adapté de (Petela et al., 2021).

C-D. Carte et reconstruction pseudo-atomique du complexe Cohésine tripartite-Pds5 (EMD-12888). Adapté de (Petela et al., 2021).

De plus, au cours de ma thèse, mon équipe d'accueil a également obtenu, par cristallographie aux rayons X, les structures des têtes de SMC1A et SMC3 humaines indépendantes, associées respectivement aux extrémités C- et N-terminales de RAD21 sous les formes apo et liées à l'ATP-γ-S et l'ADP, à des résolutions comprises entre 1,9 et 3,11 Å (Figure 20). Cela nous a permis de déterminer les changements conformationnels induits par la liaison et l'hydrolyse de l'ATP par les têtes ATPases de SMC1A et SMC3, au niveau de leurs sites catalytiques et à un niveau plus global des têtes, notamment les mouvements entre leurs domaines globulaires et leurs régions en superhélice. Cela nous a également permis d'identifier pour la première fois, l'état au repos de la porte de sortie de l'ADN, correspondant à l'interface entre l'extrémité N-terminale et le cou de SMC3. Ces changements conformationnels sont détaillés dans la première partie du chapitre 3 - Résultats.

Ainsi, les découvertes structurales de ces dernières années ont permis de mettre en lumière différentes conformations adoptées par la Cohésine au cours de son cycle, notamment lors de certaines des premières étapes, correspondant au chargement non-topologique de la Cohésine sur l'ADN. Mais, de très nombreuses questions demeurent concernant les mécanismes impliqués et le rôle des différentes interfaces et réarrangements structuraux opérés. De nombreuses études sont encore nécessaires pour mieux comprendre le mode d'action de la Cohésine et à plus long terme son implication dans les maladies.

4.5. Implication de la Cohésine dans les pathologies humaines

4.5.1. Cohésinopathies

Les cohésinopathies sont des maladies rares neurodéveloppementales causées majoritairement par des mutations dans les protéines cœur, ainsi que dans les protéines régulatrices et auxiliaires de la Cohésine. Cette famille de pathologies comprend notamment le syndrome de Cornelia de Lange (CdLS), le syndrome de Roberts, le syndrome de cassure chromosomique de Varsovie (¨Warsaw breakage syndrome¨), le syndrome de Juberg-Hayward et le syndrome de Mungan. A ce jour, ces maladies peuvent uniquement être traitées par chirurgie et par l'utilisation de prothèse pour corriger les malformations physiques, par le recours à une éducation spécialisées, avec notamment de l'orthophonie, et par le traitement symptomatique des dysfonctionnements organiques (Mfarej and Skibbens, 2021).

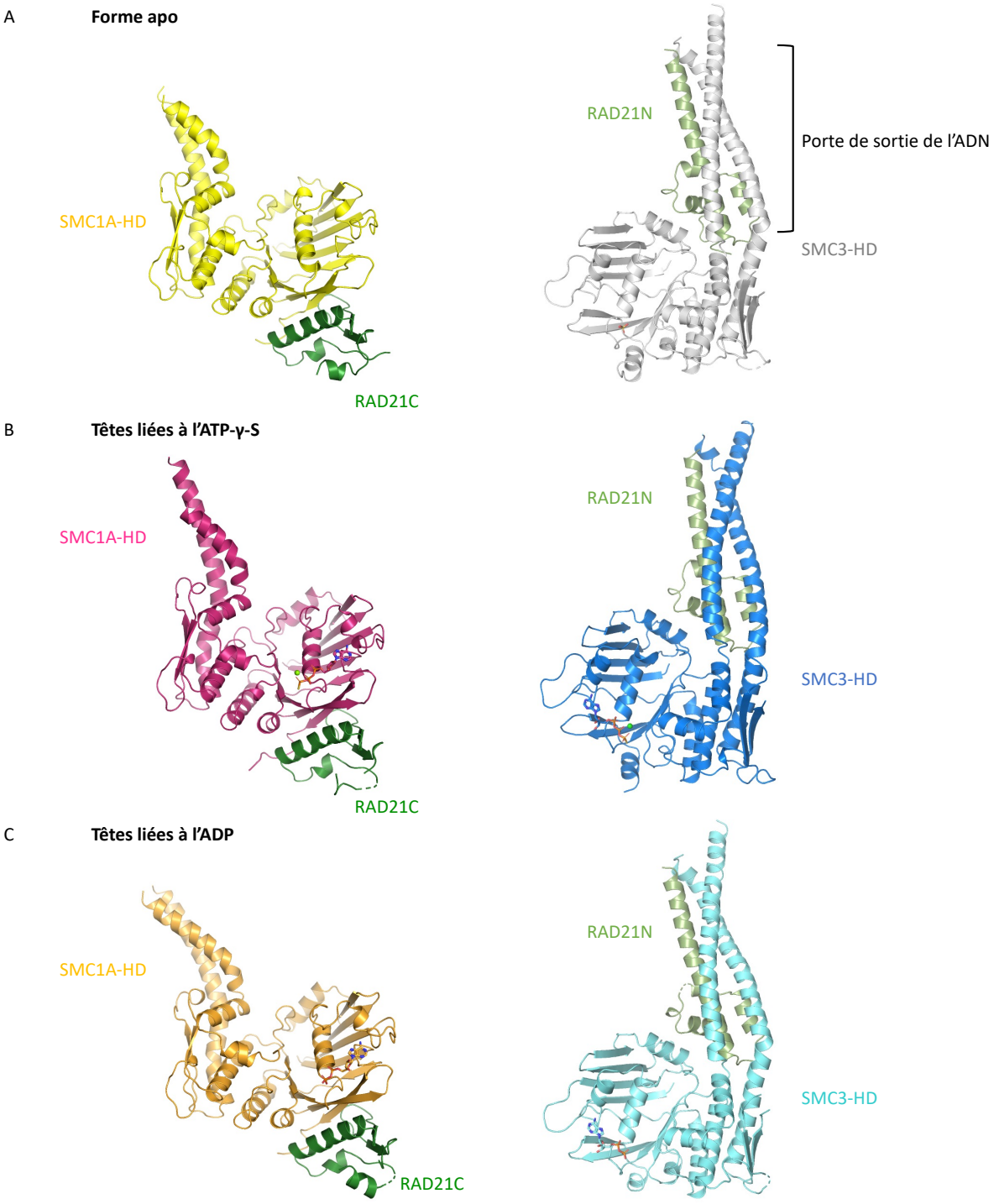


Figure 20 : Les structures des complexes SMC1A-HD/RAD21C et SMC3-HD/RAD21N indépendant, sous forme apo (A), liées à l'ATP-γ-S (B) et liées à l'ADP (C), obtenues au sein de mon équipe d'accueil.
SMC1-HD et SMC3-HD correspondent aux têtes de SMC1A et SMC3, RAD21N et RAD21C aux extrémités N- et C-terminales de RAD21. La porte de sortie de l'ADN correspond à l'interface entre le cou de SMC3 et la région N-terminale de RAD21.

4.5.1.1. Le syndrome de Cornelia de Lange

La cohésinopathie la mieux caractérisée est le syndrome de Cornelia de Lange, maladie rare avec une prévalence de 0,5 - 10 toutes les 100 000 naissances, selon les sources. Ce syndrome se caractérise par une déficience intellectuelle, un trouble de la croissance, des anomalies faciales, une myopie et une malformation des membres supérieurs (Kline et al., 2018). Environ 70% des CdLS impliquent des mutations du gène *Nipbl*, les autres étant liés à des mutations affectant les gènes *Smc1a*, *Smc3*, *Rad21* ou *Hdac8*. D'autres protéines indépendantes de la Cohésine mais ayant un rôle dans la transcription, telles que BRD4 et ANKRD11, se retrouvent également mutées dans le CdLs (Krantz et al., 2004; Deardorff et al., 2007, 2012; Kline et al., 2018). Le phénotype varie en fonction du gène muté. Il est important de noter que ces mutations n'affectent pas l'ensemble des fonctions de la Cohésine, mais que seulement certaines sont ciblées. En effet, les mutations associées au CdLS n'affectent pas la cohésion des chromatides sœurs, mais perturbent la régulation de l'expression des gènes (Remeseiro et al., 2013).

4.5.1.2. Les autres cohésinopathies

Le syndrome de Roberts est un trouble grave du développement associé à un retard de croissance et intellectuel, des malformations des membres mimant la phocomélie et des anomalies crâniofaciales. Cette maladie est associée à des mutations du gène codant pour ESCO2, enzyme acétylant la sous-unité SMC3 de la Cohésine, qui est notamment importante dans la transcription de gènes et dans la réparation des lésions de l'ADN (Vega et al., 2005; Mfarej and Skibbens, 2021).

Le syndrome de cassure chromosomique de Varsovie est associé à une microcéphalie, un retard de croissance et une perte auditive. Ce syndrome est lié à des mutations de l'hélicase DDX11 (Capo-Chichi et al., 2013). Cette dernière favorise la cohésion des chromatides sœurs et la stabilité de la Cohésine sur l'ADN au niveau des réplisomes et pourrait également avoir un rôle dans la formation de boucles de chromatine (Cortone et al., 2018; Pisani, 2019).

Le syndrome de Juberg-Hayward est une maladie rare caractérisée par une microcéphalie, un défaut de croissance, une dysmorphie faciale et des anomalies des

membres supérieurs. Ce syndrome est également associé à des mutations du gène *Esco2* (Kantaputra et al., 2021).

Le syndrome de Mungan est associé à une pseudo-obstruction intestinale et des anomalies cardiaques et est lié à des mutations du gène *Rad21* (Mungan et al., 2003; Bonora et al., 2015).

De nouvelles cohésinopathies impliquant des mutations de STAG1 et STAG2, pouvant notamment conduire à l'holoprosencéphalie, ont été récemment découvertes, mais nécessitent d'être caractérisées plus précisément (Lehalle et al., 2017; Soardi et al., 2017; Mullegama et al., 2019; Kruszka et al., 2019; Cratsenberg et al., 2021; Funato et al., 2022).

Ainsi, la famille des cohésinopathies englobe de nombreux syndromes neurodéveloppementaux rares, qui présentent des caractéristiques propres à chacun et qui sont associés à diverses mutations affectant à la fois les gènes de la Cohésine cœur, de ses protéines associées ou d'autres protéines nucléaires. Il est donc difficile de les étudier du fait de leur hétérogénéité, la complexité étant renforcée par le nombre restreint de patients. Des études futures sont nécessaires pour mieux comprendre les cohésinopathies et l'implication de la Cohésine et sa dynamique dans les pathologies.

4.5.2. Cancers

La dérégulation de la Cohésine et de sa dynamique est associée à un grand nombre de cancers. Cette dérégulation peut être induite (i) par des mutations somatiques sur la Cohésine et sur ses protéines partenaires, (ii) par sa sous-expression ou sur-expression, ainsi que (iii) par des mutations aux sites liants le facteur CTCF et la Cohésine (Katainen et al., 2015; Guo et al., 2018)

En effet, les sous-unités cœurs de la Cohésine sont fréquemment mutées dans des cancers solides, notamment dans 21% des cancers de la vessie, dans 19% des glioblastomes, dans 15% des sarcomes d'Ewing et dans 10% des mélanomes, ainsi que dans certaines leucémies (12% des leucémies myéloïdes aiguës, 46% des leucémies aiguës mégacaryoblastiques associées au syndrome de Down) (Figure 21) (Mintzas and Heuser, 2019). Le gène *Stag2* est le gène de la Cohésine le plus muté et est également l'un des vingt-deux gènes les mutés dans au moins quatre types de cancer, en particulier dans les leucémies myéloïdes, les cancers de la vessie, les cancers du poumon et les glioblastomes (Lawrence et

al., 2014). Les mutations de STAG2 sont généralement des mutations entrainant un décalage du cadre de lecture et des mutations de non-sens, conduisant dans la majorité des cas à des protéines tronquées et non fonctionnelles (Romero-Pérez et al., 2019). Les protéines accessoires de la Cohésine peuvent également se retrouver mutées, avec principalement CTCF, qui est l'un des dix gènes les plus mutés dans trois types de cancers, notamment dans les cancers de l'endomètre et dans les tumeurs malignes myéloïdes et lymphoïdes (Lawrence et al., 2014; Bailey et al., 2021).

La Cohésine peut également être sur-exprimée ou sous-exprimée selon le type de cancer. Sa surexpression est notamment observée dans les cancers du sein, dans les cancers colorectaux, ainsi que dans les cancers du poumon, alors que sa sous-expression a été observée dans les cancers de la bouche et de la gorge (Yamamoto et al., 2006; De Campos Silva et al., 2014; Xu et al., 2011, 2015). L'expression anormale de PDS5 a aussi été observée dans les cancers de la prostate, celui du col de l'utérus, de la tête et du cou et de l'œsophage (Zhang et al., 2021).

La pathogénicité induite par une cohésine mutée, sous- ou sur-exprimée, semble être davantage liée à ses fonctions non canoniques, notamment dans l'organisation du génome, la régulation de l'expression de gène et dans la réparation des lésions de l'ADN, entrainant ainsi une instabilité génomique (Qiu and Huang, 2020; Tothova et al., 2021; Di Nardo et al., 2022).

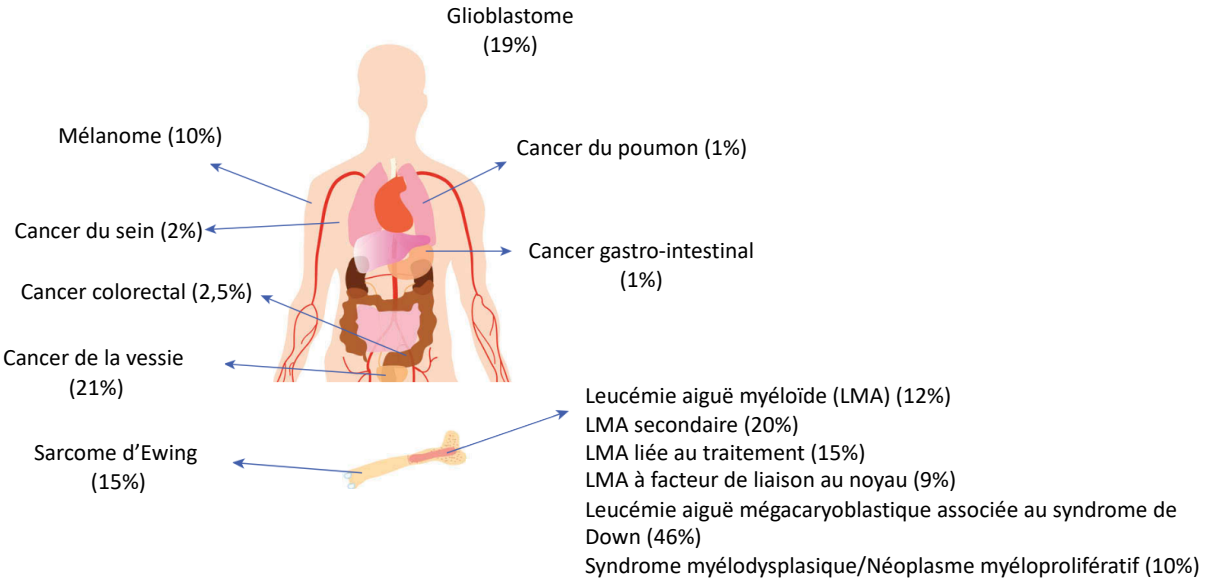


Figure 21 : Fréquences des mutations des sous-unités de la Cohésine, selon les types de cancers.
Les fréquences sont indiquées entre parenthèses. Adapté de (Mintzas and Heuser, 2019).

Ainsi, la Cohésine et ses régulateurs se retrouvent impliqués dans de nombreuses pathologies. Les mutations de la Cohésine se situent sur l'ensemble de la Cohésine, et notamment au niveau des têtes ATPases de SMC1A et SMC3, ainsi que sur la sous-unité STAG. Pourtant la pathogénicité associée à la Cohésine et à ses partenaires protéiques reste peu comprise, du fait de sa complexité structurale, fonctionnelle et de régulation, et ne permet pas encore le développement de thérapies ciblant la Cohésine. C'est pour cela qu'il est essentiel d'étudier le fonctionnement en condition normale de la Cohésine, notamment l'activité au niveau de ses têtes ATPases, afin de mieux comprendre comment sa dynamique est altérée lors des cancers, mais aussi des cohésinopathies.

4.6. Le module ATPase de la Cohésine, acteur clé de son mode d'action

4.6.1. Composition du module ATPase de la Cohésine

Le module ATPase se compose des têtes ATPases de SMC1A et SMC3, formées par l'interaction de leurs régions N- et C-terminales, et des extrémités N- et C-terminale de RAD21. Les têtes de SMC1A et SMC3 sont divisées en deux avec un lobe RecA, correspondant à la partie globulaire de la tête ATPase englobant la poche de liaison à l'ATP, et un lobe hélical qui relie la région supérieure de la partie globulaire à la région en superhélice, grâce à des hélices couplées au motif signature (¨signature-coupling helices¨) (Figures 22A). Ces protéines, tout comme les autres protéines SMC, possèdent deux demi-sites de liaison à l'ATP similaires à ceux observés dans la famille des transporteurs ¨ATP-Binding Casette¨ (transporteur ABC) (Figure 22B). (Hopfner and Tainer, 2003). La dimérisation de deux protéines SMC induite par la liaison de deux molécules d'ATP permet de former des sites catalytiques entiers et actifs. Dans cette conformation, les deux têtes sont positionnées dans une conformation de ying et yang. Chaque site de liaison à l'ATP se compose de motifs partagés par les deux têtes ATPases (Figure 22C).

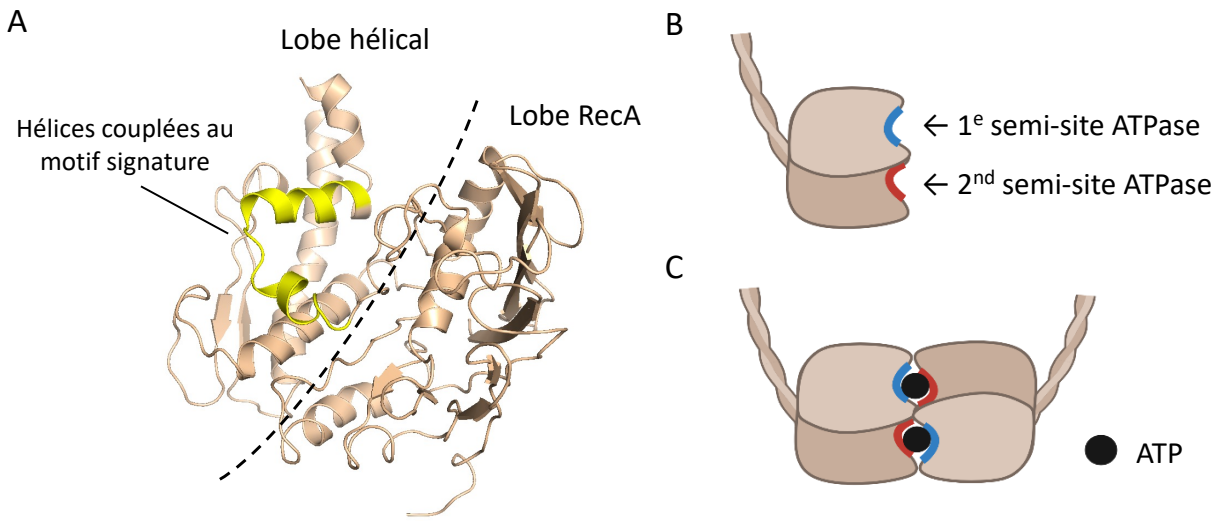


Figure 22 : La tête ATPase des protéines SMC.
A. Chaque tête SMC est divisé en un lobe RecA et un lobe hélical.
B. Chaque tête SMC comprend deux semi-sites de liaison à l'ATP, possédant chacun des motifs différents (représentés en rouge et en bleu).
C. La liaison de deux molécules d'ATP entrainent la dimérisation des têtes de deux protéines SMC. Chaque site de liaison à l'ATP est partagé par des motifs situés sur les deux têtes, permettant la formation d'un site ATPase entier et actif.

Le lobe RecA de chaque tête SMC comprend le motif Walker A ou boucle P, dont la séquence consensus est GXXXXGK(S/T). Ce motif, à travers les groupes amides -NH de son squelette et du résidu lysine consensus, permet de stabiliser la liaison de l'ATP dans le site catalytique, en interagissant avec les β- et γ-phosphates de l'ATP. Cette boucle P est également à proximité du sucre et de la base azotée de l'ATP. Un ion Mg^{2+} est également coordonné par les β- et γ-phosphates de l'ATP et par la sérine/thréonine de la boucle P. L'ion Mg^{2+} est important pour l'hydrolyse de l'ATP (Thomas and Tampé, 2020). La Cohésine présente des motifs non présents chez les transporteurs ABC, tel que la boucle R située dans la vicinité du site ATPase catalytique et présentant une arginine conservée. Cette boucle lie l'α-phosphate de l'ATP et semble stimuler l'hydrolyse de l'ATP (Lammens et al., 2004). Nous retrouvons également le motif Walker B, qui a comme séquence consensus hhhhDE, où h représente un acide aminé hydrophobe. L'aspartate de ce domaine permet la coordination de l'ion Mg^{2+} et le glutamate catalytique est essentiel à l'hydrolyse de l'ATP, en polarisant la molécule d'eau, qui va attaquer le groupement γ-phosphate de l'ATP. Le lobe RecA possède également une boucle D, dont la séquence consensus est SALD, et pour laquelle l'aspartate participe à la dimérisation des deux têtes (Figure 23A) (Thomas and Tampé, 2020)

Quant au lobe hélical, il comprend la boucle Q et le motif signature, dont la séquence consensus est LSGG(Q/E). Le motif signature permet la dimérisation des deux têtes SMC, induite par la liaison de l'ATP. La boucle Q est également située dans le lobe hélical et est caractérisée par la séquence X(X/Q)Q, le X correspondant à n'importe quel résidu, participe à la liaison de l'ATP, en coordonnant l'ion Mg^{2+}, en agissant comme senseur de la présence du γ-phosphates de l'ATP, en connectant les lobes RecA et hélical de chaque tête et en participant à la dimérisation des deux têtes (Figure 23A) (Thomas and Tampé, 2020).

La dimérisation des têtes de SMC1A et SMC3 est induite par la liaison de deux molécules d'ATP et de Mg^{2+}, prises en sandwich entre les deux têtes, formant alors un module bi-ATPase fonctionnel (Arumugam et al., 2003). Chacun des sites catalytiques est formé par les boucles P, R et Q et le motif Walker B d'une des têtes et par le motif signature de l'autre tête. Les têtes de SMC1A et SMC3 sont positionnées de sorte que ces têtes soient en regard l'un de l'autre, le motif signature de la première tête est aligné avec la boucle P de la seconde et inversement, leurs motifs Walker B se font face (Figure 23B) (Arumugam et al., 2003).

4.6.1. Importance du module ATPase dans les fonctions de la Cohésine

L'activité ATPase de la Cohésine est au cœur de l'ensemble des fonctions de la Cohésine. En effet, Arumugam et al ont démontré *in vivo* que les mutants de SMC1A et SMC3 de *S. cerevisiae*, ayant une faible activité d'hydrolyse de l'ATP, ne peuvent pas être chargés sur les chromosomes (Arumugam et al., 2003). Cela a également été confirmé la même année par Weitzer et al, qui ont mis en évidence que l'expression de protéines SMC1A mutées au niveau de leurs motifs Walker A, Walker B ou signature chez *S. cerevisiae* entraine une perte de la cohésion des chromatides sœurs (Weitzer et al., 2003).

Par ailleurs, l'activité ATPase est également essentielle pour les autres mécanismes de la Cohésine, notamment dans la formation de boucles d'ADN. En effet, l'utilisation d'un autre nucléotide que l'ATP, comme le cytidine triphosphate ou celle de mutants SMC1A et SMC3 ne pouvant pas lier ou hydrolyser l'ATP, ne permet pas de constituer des boucles *in vitro* (Davidson et al., 2019; Kim et al., 2019). De plus, les Cohésines associées à PDS5 présentent une faible activité ATPase et ne sont pas capables d'extruder des boucles d'ADN, confirmant d'autant plus l'importance de l'activité ATPase dans ce processus (Petela et al., 2018; Davidson et al., 2019).

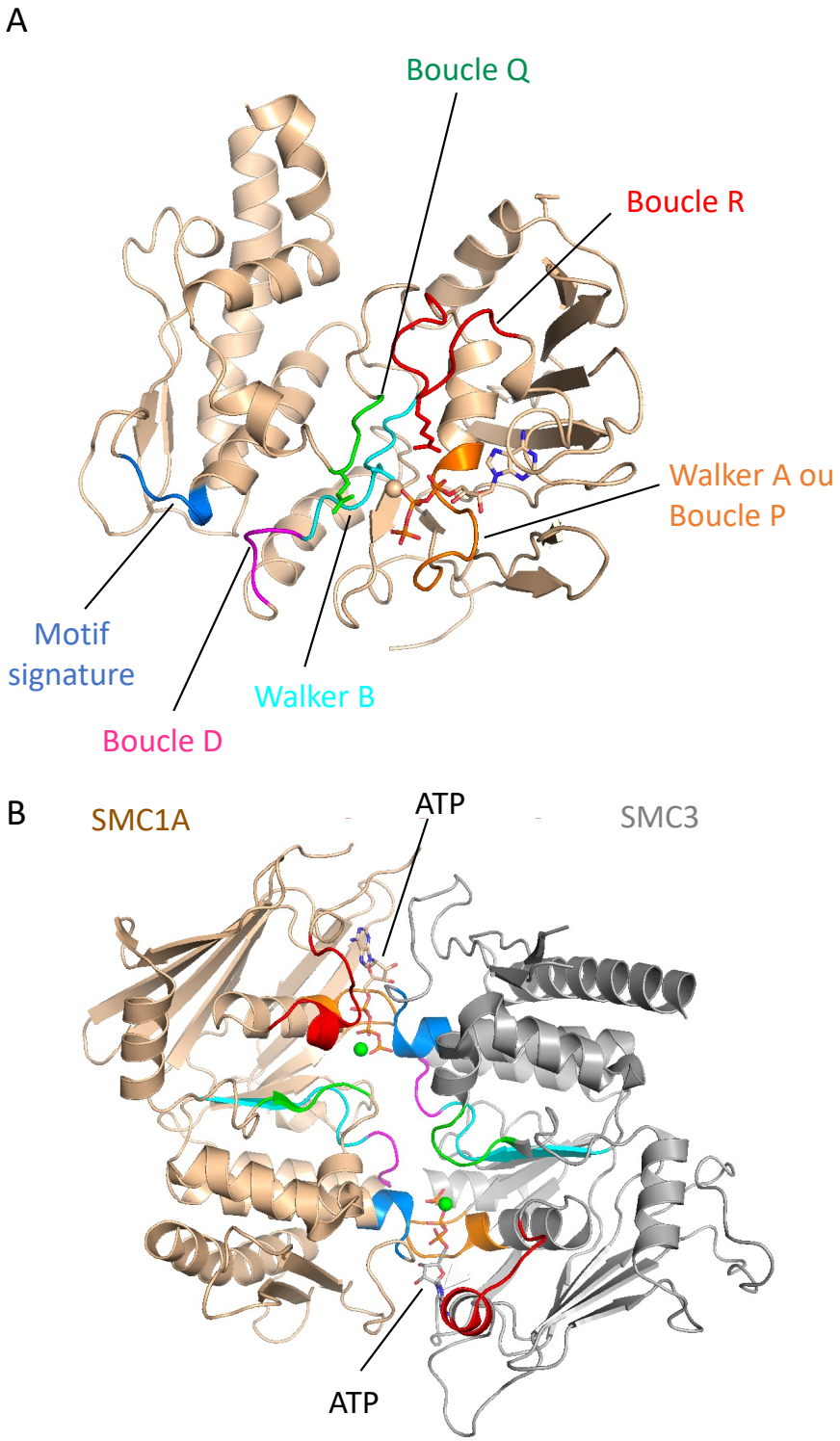


Figure 23 : Les têtes ATPases de la Cohésine partagent des motifs participant à la liaison et à l'hydrolyse de l'ATP, similaires à ceux des transporteurs ABC.

A. Chaque tête présente les motifs nécessaires à son activité ATPase suivants : les Walkers A (ou boucle P, orange) et B (cyan), les boucles Q (vert), R (rouge) et D (magenta), ainsi que le motif signature (bleu).

B. La liaison d'ATP entraine la dimérisation des têtes de SMC1A et SMC3. Chaque site est composé des boucles P, Q et R et du motif Walker B d'une des têtes et du motif signature de l'autre tête et inversement. Les têtes ATPases peuvent ainsi lier deux molécules d'ATP et former un module bi-ATPase fonctionnel.

Plus particulièrement, l'activité ATPase, associée au complexe de chargement NIPBL/MAU2, est requise pour le chargement de la Cohésine sur l'ADN. L'utilisation d'analogues de l'ATP faiblement ou non-hydrolysables ne permet pas le chargement de la Cohésine sur l'ADN, seule l'utilisation d'ATP le permet (Hu et al., 2011; Murayama and Uhlmann, 2014; Davidson et al., 2019; Kim et al., 2019). De plus, il a été montré, *in vitro*, que les Cohésines, s'étant liées à l'ADN de manière stable, maintiennent une activité ATPase et que cette activité est cruciale pour le maintien de cette interface (Murayama and Uhlmann, 2014; Camdere et al., 2015). L'activité ATPase est aussi requise pour l'acétylation de SMC3. L'utilisation de mutants des têtes ATPases affectant la liaison ou l'hydrolyse de l'ATP réduit de manière drastique cette acétylation, par comparaison aux protéines sauvages (Ladurner et al., 2014). Enfin, l'activité ATPase de la Cohésine participe également à la libération de la Cohésine, mais les mécanismes impliqués restent méconnus (Murayama and Uhlmann, 2015; Elbatsh et al., 2016).

Ainsi, la liaison et l'hydrolyse de l'ATP par les têtes ATPases de la Cohésine sont essentielles à l'ensemble des fonctions de ce complexe et ce tout au long du cycle de la Cohésine (chargement, stabilisation et dissociation de la Cohésine sur l'ADN). Toutefois, nos connaissances sur ce module ATPase sont fragmentaires et restent à caractériser précisément, bien que certaines structures soient déjà disponibles, détaillées dans la partie 4.4. État de l'art : les structures de la Cohésine connues à ce jour du chapitre 1 - Introduction.

Objectifs du travail de thèse

Mon laboratoire d'accueil a comme thématique principale l'étude des bases moléculaires de la Régulation de la Chromatine et de la Transcription et concentre ses efforts sur la caractérisation de la Cohésine, complexe clé dans la stabilité du génome et la division cellulaire. La dérégulation de son activité se retrouve impliquée dans de nombreuses pathologies, telles que dans certaines maladies neurodéveloppementales, dites cohésinopathies, et dans de nombreux cancers tous types confondus.

Ces dernières années, des avancées majeures, rendues possibles grâce à l'essor des techniques de capture de la conformation des chromosomes, ainsi que celle de la cryoEM, ont permis de mieux définir les fonctions de la Cohésine au sein du noyau, ainsi que sa régulation au cours du cycle cellulaire. Le cycle de fonctionnement de la Cohésine est un processus complexe qui fait appel à un grand nombre de régulateurs. Ce cycle se décompose en plusieurs étapes, du chargement à la stabilisation et au déchargement de la Cohésine sur l'ADN. De plus, de nombreuses études ont montré l'importance de l'activité ATPase de la Cohésine dans toutes ses fonctions, allant de la cohésion des chromatides sœurs à l'organisation 3D du génome. Il existe une étroite collaboration entre cette activité ATPase et les protéines accessoires de la Cohésine. Toutefois, les connaissances des bases moléculaires et structurales liées à cette activité ATPase et au cycle ATPase de la Cohésine restent incomplètes. Il est donc essentiel de déterminer la dynamique de la Cohésine humaine, induite par la liaison et l'hydrolyse de l'ATP par ses protéines SMC1A et SMC3, afin de mieux comprendre son mode d'action physiologique et de mieux appréhender ses dysfonctionnements observés dans les cohésinopathies et les cancers.

Au début de ma thèse, très peu d'informations étaient disponibles sur la Cohésine humaine dans la littérature. Ce n'est qu'en 2020, que les premières structures du module ATPase humain, couplé à STAG1, à NIPBL, et à l'ADN ont été publiées. Cela a permis de poser les prémices de la capture de l'ADN par la Cohésine, mais a laissé de nombreuses questions ouvertes concernant la mécanistique impliquée dans l'activité ATPase et la dynamique de la Cohésine au cours du cycle ATPase. Notamment, comment les sous-unités de la Cohésine interagissent-elles les unes avec les autres lors du cycle de la Cohésine, en particulier lors de

la liaison/hydrolyse de l'ATP. Mais aussi, quel est le rôle STAG1 ou STAG2, protéine auxiliaire associée en permanence à l'anneau tripartite ?

Il est donc essentiel d'étudier précisément le module ATPase et cette question a été la base de mon projet de thèse. Lors de ma thèse, j'ai été amenée, par une approche de biologie structurale intégrative, à (i) caractériser la dynamique moléculaire de la Cohésine, induite par la liaison de l'ATP et qui entraine l'engagement des têtes de SMC1A et SMC3, et à (ii) comprendre comment cette dynamique est potentiellement impactée par STAG1 et STAG2.

A mon arrivée dans le laboratoire d'accueil, mon équipe avait déjà initié les premières recherches sur les têtes ATPases de la Cohésine humaine. Les protocoles initiaux de l'expression et de la purification des têtes ATPases de SMC1A et SMC3 humaines (SMC1A-HD, SMC3-HD), associées respectivement aux extrémités C- et N-terminales de RAD21, étaient déjà établis. Une structure cristalline de SMC1A dans sa conformation apo avait également été résolue. Lors de ma thèse, mon équipe d'accueil a également obtenu les structures de SMC1A-HD et de SMC3-HD, associées respectivement aux extrémités de RAD21 correspondantes, sous formes apo pour SMC3-HD et liées à l'ATP-γ-S et à l'ADP pour SMC1A-HD et SMC3-HD. Ces structures sont décrites dans la partie 4.4. État de l'art : les structures de la Cohésine connues à ce jour du chapitre 1 - Introduction et dans le manuscrit présenté en partie I. Etude du module ATPase, acteur clé dans le cycle ATPase de la Cohésine dans le chapitre 3 – Résultats.

Ces travaux préliminaires ont permis de poser des bases solides afin d'adresser des questions plus complexes sur la dynamique de la Cohésine et d'initier l'étude de la Cohésine associée à un de ses partenaires, STAG1 ou STAG2. Mes objectifs de thèse ont ainsi porté sur :

- la reconstitution du module ATPase engagé, composé des complexes SMC1A-HD/RAD21C et SMC3/RAD21N en présence d'ATP ou d'analogues faiblement hydrolysables de l'ATP, en utilisant la chromatographie d'exclusion de taille et l'ultracentrifugation analytique.

- la caractérisation structurale de ce module engagé, en essayant trois techniques : la cristallographie aux rayons, la diffusion des rayons X aux petits angles et la cryoEM.

- la reconstitution du module ATPase associée à STAG1 et STAG2, afin de pouvoir définir le rôle de STAG1 ou STAG2 sur le module ATPase.

Chapitre 2 - Matériels et méthodes

La méthodologie que j'ai utilisée au cours de ma thèse pour la caractérisation de mes protéines et complexes protéiques d'intérêt a été basée sur une approche de biologie structurale intégrative combinant des méthodes moléculaires, biochimiques, biophysiques et structurales. Cette approche permet une meilleure compréhension du mode d'action des protéines et complexes d'intérêt, étape préalable clé à des études sur leur rôle dans la cellule.

Cette méthodologie a été initiée par une première étape de biologie moléculaire pour générer les vecteurs d'expression dans lesquels sont insérés les gènes codant pour mes protéines/complexes protéiques d'intérêt. Des tests d'expression et de purification à petite échelle à moyen débit ont ensuite été réalisés, afin d'identifier les constructions protéiques et conditions expérimentales de purification les plus prometteuses. Ces tests ont permis de poser les bases pour la prochaine étape, l'expression et la purification à grande échelle des protéines et complexes protéiques. Cette dernière étape a été clé et a nécessité de nombreuses optimisations, afin d'obtenir des protéines et complexes protéiques suffisamment purs, homogènes et stables de manière à poursuivre leur caractérisation. Cette dernière a alors pu débuter, en combinant plusieurs techniques biochimiques (chromatographie par exclusion de taille, tests d'activité enzymatique, …), biophysiques (titration calorimétrique isotherme, ultracentrifugation analytique, …) et structurales (cristallographie aux rayons X, cryo-microscopie électronique, …).

Les différentes parties de ce chapitre décrivent l'ensemble des techniques que j'ai utilisées au cours de ma thèse pour caractériser le module ATPase de la Cohésine humaine.

I. Biologie moléculaire - Clonages

1.1. Vecteurs d'expression bactériens de la série pET-MCN

Les vecteurs de la série pET-MCN (¨pET multi-cloning and expression¨) dérivent respectivement des plasmides pET15b, pCDF-DUet et pACYC11b et ont été développés au sein de mon équipe d'accueil (Romier et al., 2006; Perrakis and Romier, 2008; Diebold et al., 2011; Marek et al., 2013; Haffke et al., 2015). Il existe trois familles de plasmides : les vecteurs pnEA

(pmA), les vecteurs pnCS (pmS) et les vecteurs pnYC (pmC). Ils présentent tous une architecture commune et contiennent notamment un promoteur T7, qui est compatible avec les cellules bactériennes ayant un système d'expression basé sur la polymérase T7, telles que les cellules BL21 (DE3) et BL21 (DE3) pRARE2 pLysS, cellules principalement utilisées lors de ma thèse (Figure 24). Ces vecteurs possèdent notamment le gène codant pour l'inhibiteur LacI de l'opéron lactose et son site de fixation lacO en aval du promoteur T7. Cela permet d'induire l'expression des protéines avec un analogue non-hydrolysable du lactose, l'isopropyl β-D-1-thiogalactopyranoside (IPTG), tout en minimisant les fuites d'expression.

Les trois familles de vecteurs pmA, pmS et pmC possèdent des résistances aux antibiotiques différentes (respectivement à l'ampicilline, à la spectinomycine et au chloramphénicol) et des origines de réplications différentes (respectivement pBR322, CDF13 et p15A). Ils peuvent donc être utilisés simultanément pour la co-expression de plusieurs protéines d'intérêt au sein d'une même cellule pour la reconstitution de complexes protéiques. Les différents membres de chaque famille peuvent coder pour différentes étiquettes de purification par affinité (par ex. étiquette histidine, Flag, Strep) et/ou protéines de fusion (par ex. thiorédoxine, GST, MBP), en position N- et/ou C-terminales, ou donner des protéines complètement natives (Figure 24). Ce large panel de vecteurs développé au sein de mon laboratoire d'accueil est utilisé afin de tester une multitude de combinaisons, en jouant sur l'étiquette de fusion et sa position, ainsi que sur la co-expression de plusieurs protéines et de leurs différentes constructions. Cela permet ainsi de déterminer les conditions idéales pour l'expression des complexes protéiques.

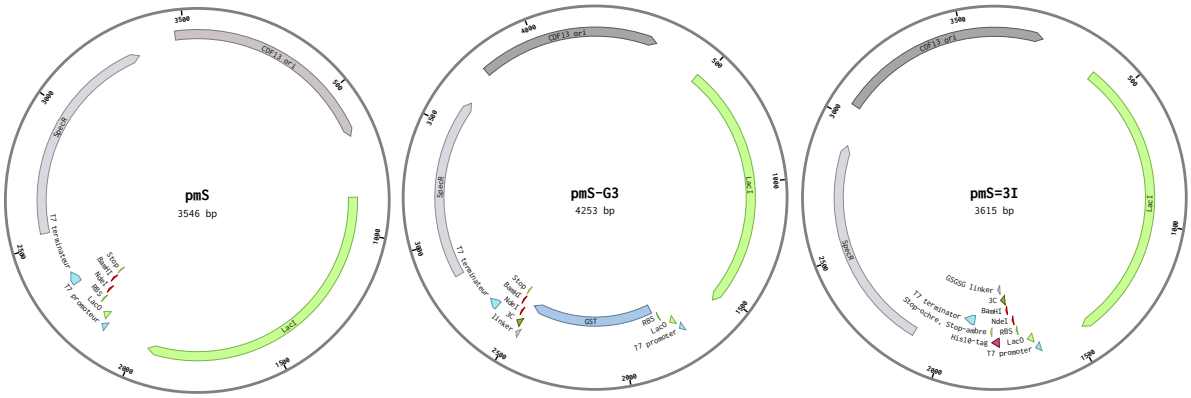


Figure 24: Exemple de plasmides pmS utilisés pour l'obtention d'une protéine d'intérêt native (pmS ; gauche), d'une protéine couplée à une étiquette GST et un site de coupure à la protéase 3C à son extrémité N-terminale (pmS-G3 ; centre) et à un site de coupure à la protéase 3C suivi d'une étiquette His_{10} à son extrémité C-terminale (pmS=3I ; droite).

Ces plasmides possèdent un promoteur T7, un site de liaison aux ribosomes (RBS), un site de clonage multiple (MCS) contenant les sites de coupure des enzymes de NdeI et BamHI, qui sont utilisés pour l'insertion du gène d'intérêt, et d'un terminateur T7.

1.2. Méthodes de clonages

1.2.1. Clonage par assemblage de Gibson

Les gènes d'intérêt ont été amplifiés par réaction en chaîne par polymérase (PCR). La réaction de PCR se fait par le mélange de la matrice ADN ou du plasmide contenant le gène d'intérêt (10 ng), de 0,5 µL de chaque amorce (100 µM), de 5 µL de dNTP (2 mM) (New England BioLabs, NEB), de 10 µL de tampon de PCR (NEB) et 0,5 µL de Phusion ADN polymérase 2 U/µL (NEB), complémenté avec de l'eau pour un volume final de 50 µL. Les amorces ont été créées de sorte à posséder des régions homologues aux gènes d'intérêt mais également au vecteur d'expression utilisé.

La PCR est réalisée à l'aide du thermocycleur Mastercycler Nexus GX2 (Eppendorf), d'après le programme suivant :

- un cycle initial à 98 °C pendant 5 min,

- trente cycles composés chacun d'une étape de dénaturation (95 °C pendant 30 sec), d'une étape d'hybridation (58 °C pendant 30 sec) et d'une étape d'élongation (72 °C pendant 30 sec par kilo-paire de base),

- un cycle final à 72 °C pendant 5 min.

Les fragments amplifiés sont purifiés sur colonne (NucleoSpin ®, Macherey Nagel), selon les recommandations du fabriquant.

En parallèle de la préparation des inserts, les vecteurs utilisés sont digérés par les enzymes de restriction NdeI et BamHI (NEB), selon les recommandations du fabriquant. Les vecteurs ouverts sont ensuite traités par la phosphatase alcaline de crevette (NEB) pendant 1h à 37°C, pour éviter qu'ils ne se referment sur eux-mêmes. Les vecteurs sont ensuite purifiés sur gel d'agarose 1% et sur colonne (NucleoSpin®, Macherey Nagel). Ils sont prêts à être utilisés pour les réactions de clonage. Les fragments amplifiés peuvent alors être insérés entre les sites NdeI/BamHI des vecteurs respectifs, par la méthode de Gibson.

La méthode de Gibson permet d'assembler le vecteur d'expression linéarisé et le fragment d'intérêt amplifié. Cette méthode est basée sur l'utilisation de trois enzymes : la T5

exonucléase, la Phusion ADN polymérase et la Taq ADN ligase. La T5 exonucléase dégrade les extrémités 5' des fragments du vecteur linéarisé et de l'insert produit par PCR. Les extrémités dénudées complémentaires peuvent alors s'hybrider et les extrémités 3' libres sont allongées par la Phusion ADN polymérase. La Taq ADN ligase lie les différents fragments entre eux et permet de reconstituer le plasmide contenant le gène d'intérêt (Figure 25) (Gibson et al., 2009, 2010).

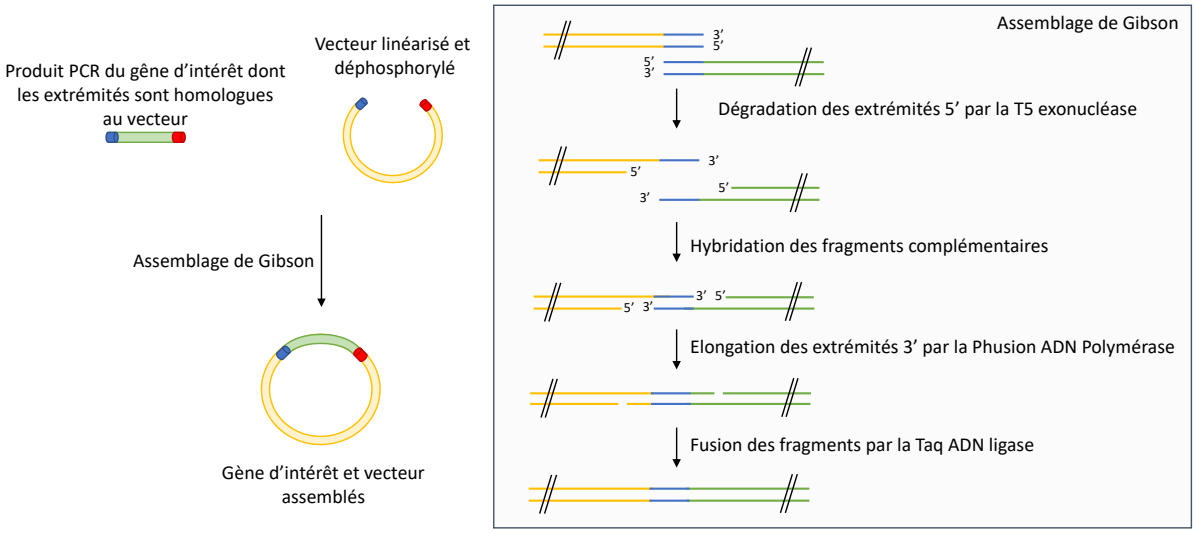


Figure 25 : Principe de la réaction d'assemblage de Gibson.

Cette méthode consiste à ajouter 15µL de mélange de Gibson, contenant les enzymes T5 exonucléase (NEB), Phusion ADN polymérase (NEB) et Taq ligase (NEB), à l'insert (150 ng) et au plasmide correspondant (50 ng), dans un volume final de 20 µL. Le mélange est chauffé à 50 °C pendant 1 h. 5 µL du produit de la réaction sont ensuite utilisés pour transformer des cellules *E. coli* DH5α chimio-compétentes et les plasmides obtenus sont vérifiés par séquençage (Eurofins).

1.2.2. Assemblage des constructions des têtes ATPases des protéines SMC1A et SMC3 et de RAD21 par PCR nichée

La majorité des clonages ont été réalisés à l'aide des techniques décrites précédemment. Cependant, la reconstitution des constructions des têtes ATPases des protéines SMC1A et SMC3 humaines (SMC1A-HD, SMC3-HD), ainsi que celles des constructions de RAD21 a nécessité de faire appel à une approche plus complexe : celle de la PCR nichée.

1.2.2.1. Les têtes ATPases de SMC1A et SMC3

SMC1A-HD et SMC3-HD sont composées respectivement de leurs extrémités N- et C-terminales, ainsi que d'une petite portion de leurs régions en superhélice, qui sont reliées entre elles par un linker peptidique suivi d'un site de clivage de la thrombine (ESSKHPASLVPRGS) (Figure 26A).

Lorsque j'ai commencé ma thèse, mon équipe de recherche avait déjà réalisé des études préliminaires sur les têtes ATPases SMC1A-HD et SMC3-HD humaines. De ce fait, des premières constructions de SMC1A-HD et de SMC3, indiquée sur la figure 26A et présentant une certaine longueur des régions en superhélice, étaient déjà disponibles (SMC1ACC et SMC3J). Afin d'étudier l'impact de ces régions en superhélices sur les têtes ATPases de SMC1A et SMC3, des constructions plus courtes et plus longues, avec un linker différent (GSGSLVPRGSGS), ont été réalisées par moi-même et par Marina Victoria Gomez, une étudiante en thèse de l'équipe (SMC1AJ, SMC1ACCsh, SMC3CC) (Figure 26A).

Une approche similaire a été utilisée pour toutes les constructions ayant été produites par PCR nichée. Les parties codant pour les domaines N- et C-terminaux (NTD et CTD) de SMC1A et de SMC3 ont été amplifiées séparément par PCR. L'amorce sens du NTD comprend le site de restriction de NdeI et l'amorce antisens de CTD le site de restriction de BamHI. L'amorce antisens du NTD possède une séquence chevauchante d'une vingtaine de nucléotides correspondant au début de l'amorce sens du CTD. Les séquences NTD et CTD amplifiées séparément sont purifiées sur gel d'agarose 1% et sur colonne (Kit NucleoSpin ®, Macherey Nagel), selon les recommandations du fabriquant. Une seconde PCR est réalisée avec les fragments obtenus et avec les amorces bornant les extrémités 5' et 3' du fragment entier. Le produit de la PCR est purifié sur colonne (Kit Nucleospin ®, Macherey Nagel) et utilisé pour le clonage par la méthode de Gibson décrit précédemment. Le produit obtenu est transformé dans les cellules DH5α de *E. coli*. Les plasmides purifiés sont séquencés dans la partie du gène d'intérêt pour vérifier la présence de la construction génique (Eurofins).

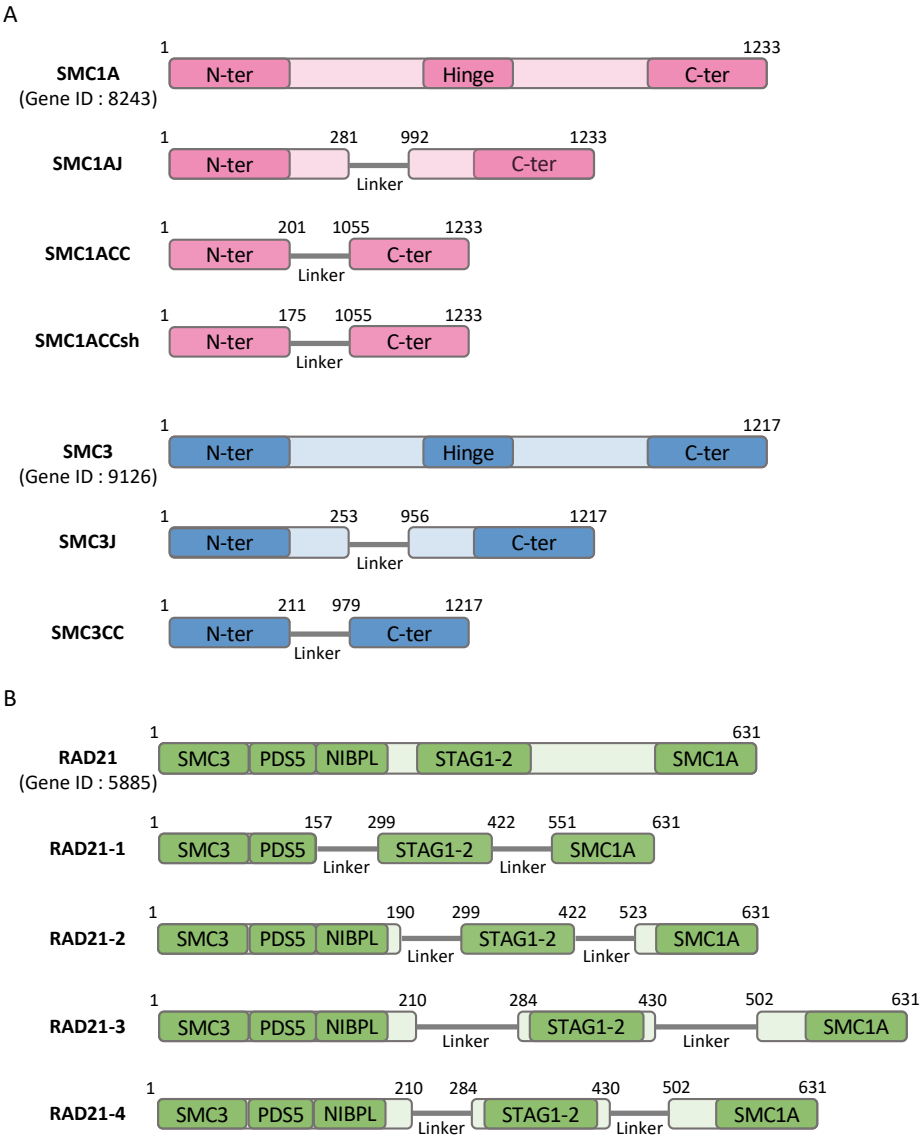


Figure 26 : Représentation des différentes constructions de SMC1A et SMC3 (A) et de RAD21 (B) réalisées et utilisées pendant ma thèse.
Les annotations sur les constructions correspondent à des régions spécifiques de chacune des protéines. Notamment dans le cas de RAD21, celles-ci indiquent les zones d'interaction de cette protéine avec les autres sous-unités de la Cohésine.

1.2.2.2. Les constructions de RAD21

J'ai réalisé cinq constructions de RAD21, qui comprennent trois domaines d'interaction. Le premier correspond à la région N-terminale interagissant avec les protéines SMC3 et PDS5, le second domaine à la région centrale liant les protéines STAG1 et STAG2 et le dernier domaine à la région C-terminale interagissant avec la protéine SMC1A. Les constructions 2, 3 et 4 comprennent également un domaine de liaison à NIPBL. Ces domaines sont espacés par différents linkers courts (Figure 26B).

Pour obtenir ces constructions, j'ai utilisé un protocole semblable à celui décrit pour la reconstitution des têtes de SMC1A et de SMC3. Les parties codant pour les trois domaines de RAD21 ont été amplifiées séparément. L'amorce sens du premier domaine comprend le site de restriction de NdeI et l'amorce antisens du dernier domaine le site de restriction de BamHI. L'amorce antisens du premier domaine possède une séquence chevauchante d'une vingtaine de nucléotides correspondant au début de l'amorce sens du second domaine et ainsi de suite pour le reste des amorces. Les trois fragments amplifiés séparément des différents domaines sont purifiés sur gel d'agarose 1% et sur colonne (Kit NucleoSpin®, Macherey Nagel). Une dernière PCR est ensuite réalisée avec les trois fragments obtenus et avec les amorces bornant les extrémités 5' et 3' du fragment entier. Le produit de la PCR est purifié sur colonne (Kit Nucleospin, Macherey Nagel) et utilisé pour le clonage par la méthode de Gibson décrit précédemment. Le produit obtenu est transformé dans les cellules DH5α de *E. coli* et les plasmides obtenus sont vérifiés par séquençage (Eurofins).

1.2.3. Concaténations des vecteurs pour l'expression de complexes protéiques

La co-expression bactérienne par co-transformation permet l'expression de deux à trois protéines, avec l'utilisation concomitante maximale de trois vecteurs. Au-delà, des problèmes de croissance cellulaire apparaissent, du fait de l'utilisation d'un trop grand nombre d'antibiotiques différents. De plus, j'ai utilisé dans certains cas le vecteur pRARE2 (décrit dans le chapitre 2 - Matériels et méthodes), qui porte une résistance au chloramphénicol et une origine de réplication p15A, ce qui empêche l'utilisation du vecteur pmC dans le cas des cellules BL21 pRARE2 pLysS. C'est pourquoi, pour la reconstitution de plus gros complexes en bactérie, la concaténation de plusieurs gènes sur un même vecteur d'expression est préférable voire nécessaire et permet de s'affranchir de cette limite. Il existe plusieurs types de concaténation dont la concaténation polycistronique et celle multi-promoteur (Busso et al., 2011; Diebold et al., 2011; Haffke et al., 2015).

La concaténation polycistronique permet de mettre de manière simultanée plusieurs gènes sous le contrôle d'un seul et même promoteur T7. Elle est basée sur la présence du site de restriction de l'endonucléase I-CeuI (NEB) en aval du site RBS et du premier gène d'intérêt. C'est à ce niveau qu'est inséré le second gène d'intérêt. Le principe de cette réaction est

présenté dans la figure 27A. Cette technique a été utilisée pour concaténer sur un même vecteur SMC1A-HD et SMC3-HD.

La concaténation multi-promoteur permet quant-à-elle d'exprimer de manière simultanée plusieurs protéines, chaque gène étant sous contrôle de son propre promoteur et terminateur T7. Cette technique est envisagée lorsque les protéines à transcrire ont une taille conséquente, ce qui peut poser des problèmes pour l'expression polycistronique bactérienne. Cette technique se base sur la présence du site de restriction de l'endonucléase PI-PspI (NEB), qui est en amont du promoteur, du site RBS et du premier gène d'intérêt. C'est à ce niveau que le fragment, comprenant la séquence du promoteur T7, celle du second gène d'intérêt, ainsi que celle du terminateur T7, est inséré. Le principe de cette réaction est présenté dans la figure 27B.

Cette technique a été utilisée pour insérer les différentes constructions de RAD21 dans un vecteur contenant les gènes de SMC1A-HD et SMC3-HD, afin d'obtenir un plasmide qui code pour ces trois protéines. Cela a été fait dans le but de les co-exprimer avec les protéines STAG1 ou STAG2, ainsi qu'avec les autres protéines auxiliaires de la Cohésine.

Le protocole est similaire quelle que soit la technique de concaténation utilisée, seules l'enzyme et les amorces diffèrent. 1 µg du plasmide contenant le premier gène d'intérêt est digéré par l'enzyme I-CeuI (NEB) ou PI-PspI (NEB), puis déphosphorylé par la CIP (¨Calf-intestinal alkaline phosphatase¨, NEB) en suivant les recommandations du fabricant. Le vecteur linéarisé est purifié sur gel d'agarose 1% et grâce au kit NucleoSpin®. Le plasmide contenant le second gène d'intérêt est amplifié par PCR, comme décrit précédemment, en utilisant les amorces de concaténation qui présente une homologie de séquence avec les sites de coupure de I-CeuI ou PI-PsPI. Le produit de la PCR est purifié sur colonne (Kit NucleoSpin®, Macherey Nagel), selon les recommandations du fabriquant. Le fragment amplifié et le vecteur linéarisé sont assemblés par la méthode de Gibson détaillée précédemment et transformés dans les cellules DH5α de *E. coli.*

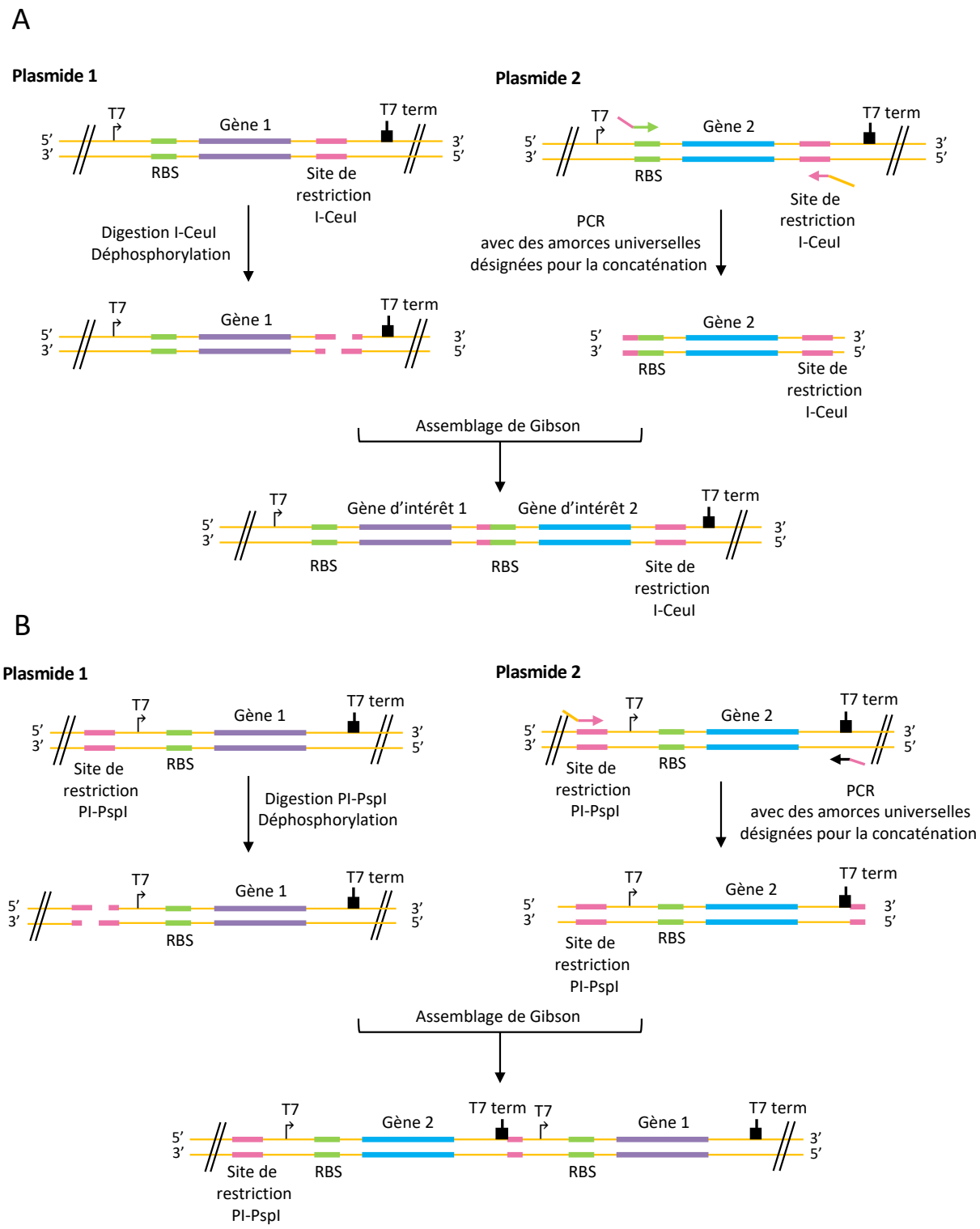


Figure 27 : Principe de la concaténation des vecteurs pET-MCN basée sur le site de restriction I-CeuI (A), aboutissant à une concaténation polycistronique, ou sur le site de PI-PspI, aboutissant à une concaténation multi-promoteur (B).

1.2.4. Mutagénèse dirigée par réplication circulaire du plasmide ou « Rolling circle »

La méthode de réplication circulaire du plasmide permet d'introduire des mutations ponctuelles dans les gènes codant pour les protéines d'intérêt. Cette méthode consiste à amplifier en totalité par PCR le plasmide d'intérêt avec des amorces contenant la ou les mutation(s) à introduire. Après amplification, le produit de la PCR est incubé pendant 1h à 37°C avec l'enzyme DpnI (NEB), qui permet la digestion du plasmide natif méthylé (parent) issu de son amplification chez *E. coli*, le plasmide synthétisé par PCR étant non méthylé. Le produit de la réaction obtenu est ensuite purifié sur colonne (Kit Nucleospin®, Macherey Nagel) selon les recommandations du fabriquant et transformé dans les cellules DH5α de *E. coli*. Les plasmides purifiés sont séquencés dans la partie du gène d'intérêt pour vérifier la présence de la mutation voulue (Eurofins.)

1.2.5. Récapitulatif des constructions des protéines utilisées pendant ma thèse

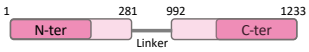
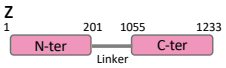
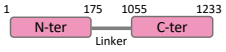
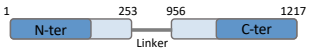
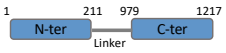
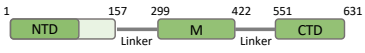
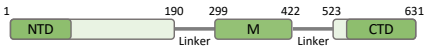
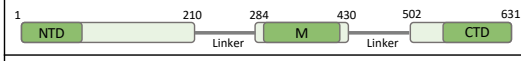
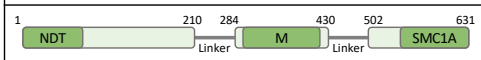
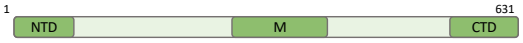
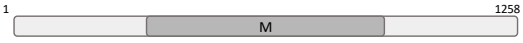
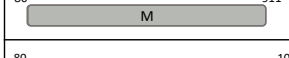
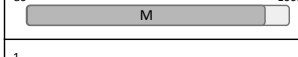
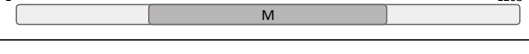
Nom	Constructions	Mutants			
SMC1A-HD					
SMC1AJ		E1157Q			
SMC1ACC		E1157Q	L1128V	D1163E	C1073S, C1115S, E1157Q
SMC1ACCsh		E1157Q			
SMC3-HD					
SMC3J		E1144Q	L1115V	D1150E	
SMC3CC		E1144Q	D92C, C1134S, E1144A	D120C, C1134S, E1144Q	
RAD21					
RAD21N		K25C C35S C78S	K26C, C35S, C78S		
RAD21C		C585S			
RA21-1					
RAD21-2					
RAD21-3					
RAD21-4					
RAD21-FL					
STAG1					
STAG1-1					
STAG1-2					
STAG1-FL					
STAG2					
STAG2-1					
STAG2-2					
STAG2-FL					

Tableau 3 : Liste des constructions et mutants utilisés pendant ma thèse.
NTD : domaine N-terminal, CTD : domaine C-terminal, M : région centrale.

II. Expression et purification de protéines

2.1. Test d'expression et de purification à petite échelle

Les tests d'expression et de purification à petite échelle ont été développés au sein de mon équipe d'accueil et permettent de tester un panel de conditions expérimentales dans un laps de temps restreint (Vincentelli and Romier, 2013, 2016). Cette technique permet de définir les constructions et les étiquettes de purification idéales, les conditions de culture (souche bactérienne, milieu de culture, température, temps d'induction, …), ainsi que les conditions de purification (tampon de lyse, résine d'affinité, …) les plus optimales afin d'obtenir une protéine ou un complexe protéique en grande quantité et soluble.

En plus de permettre l'identification des conditions les plus prometteuses pour l'expression et la purification de protéine, cette méthode permet en cas de co-expression protéique d'identifier les régions d'interaction nécessaires à la formation d'un complexe, en testant notamment différentes régions de chacune des protéines. La co-expression présente également de nombreux avantages. En effet, l'expression de protéines interagissant l'une avec l'autre modifie le comportement de chaque partenaire et peut entrainer une amélioration de leur solubilité. Pour la reconstitution d'un complexe, la place de l'étiquette de purification est un paramètre important à prendre en compte. En effet, selon son positionnement et sa taille, l'étiquette peut empêcher l'interaction entre les deux partenaires, par gêne stérique, d'où l'importance de tester différentes combinaisons avec une étiquette positionnée d'une part sur le premier partenaire, d'autre part sur le second. Ces tests sont donc une technique robuste pour l'identification et la reconstitution de complexes.

Cette technique comprend deux étapes : une étape d'expression en *E. coli* et une étape de purification par affinité, qui repose sur la fixation spécifique et réversible d'une protéine à une étiquette qui peut se lier à une matrice, la résine d'affinité.

2.1.1. Expression

Les tests à petite échelle permettent de tester l'expression et la co-expression des protéines d'intérêt. Plusieurs souches bactériennes ont été testées, avec en majorité les cellules BL21 (DE3) et les cellules BL21 (DE3) pRARE2 pLysS. Cette dernière souche bactérienne

fournit d'une part 7 ARNs de transfert rarement utilisés chez *E. coli* et donc faiblement produits, mais présents en abondance chez les eucaryotes. D'autre part, elle permet l'expression du lysozyme T7 qui, en se fixant sur la polymérase T7, réduit l'expression basale de cette dernière avant l'induction par addition de lactose ou d'IPTG. Cela permet à la fois de faciliter la production de protéines eucaryotes et de s'affranchir de problèmes de croissance ou de toxicité cellulaire.

Les cellules *E. coli* sont (co-)transformées avec les vecteurs de (co-)expression d'intérêt, par choc thermique pendant 1 min, à 42 °C, puis mises sur glace pendant 5 min et régénérées par ajout de milieu 2xLB pendant 1 h, à 37 °C. Les cellules sont ensuite ensemencées dans des boîtes de Pétri, contenant du milieu LB/agar supplémenté avec l'antibiotique ou les antibiotiques (en cas de co-transformation) adéquats (100 µg/ml ampicilline, 34 µg/ml chloramphénicol et/ou 50 µg/ml spectinomycine). Ces boîtes sont incubées à 37 °C pendant la nuit.

Pour la culture, plusieurs milieux de culture peuvent être testés : (1) le milieu Luria-Bertani doublement concentré (2xLB), (2) le milieu Graffinity ® (Graffinity Pharmaceuticals GmbH), qui se compose d'un milieu A supplémenté en glucose (0,5%) et d'un milieu B supplémenté en lactose (0,6%), en isopropyl-β-1-thiogalactopyranoside (IPTG, Euromedex, 0,5 mM) et en acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique (HEPES, Fluka, 20 mM) pH 7,4 et (3) le milieu d'auto-induction Terrific Broth (TB, Formedium). Leurs compositions sont détaillées dans le tableau 4. 3-4 colonies sont utilisées pour inoculer des mini-cultures de 4 mL de milieu 2xLB ou auto-induit TB et de 2mL de milieu A du Graffinity ®, en plaques de 24 puits (GE Healthcare). Les plaques sont incubées pendant 6 h à 37 °C, puis amenées à une température de 25 °C. L'induction de l'expression des protéines est réalisée par l'ajout de 0,7 à 1mM d'IPTG pour le milieu 2xLB ou par l'ajout de 2 mL du milieu B pour le milieu Graffinity®.

Pour le milieu d'induction AI.TB, qui contient du glucose et du lactose, les bactéries se développent et consomment en premier le glucose. Une fois ce dernier entièrement consommé, le lactose est ensuite internalisé et utilisé. L'ajout de lactose ou d'IPTG induit un changement de conformation du répresseur lacI, qui ne peut plus se lier au promoteur LacUV5 entrainant alors la production de la polymérase T7 et la synthèse des protéines d'intérêt. Il n'est donc pas nécessaire d'ajouter de l'IPTG pour le milieu auto-induit, seule la température est diminuée pour faciliter, comme avec les autres milieux, le repliement des protéines. A

noter que le milieu Graffinity reproduit ce type de condition, mais est supplémenté en IPTG pour une induction plus rapide qui sera ensuite soutenue par le lactose. Les milieux TB et Graffinity, du fait de leur richesse en nutriments, permettent d'atteindre des densités optiques supérieures à celles obtenues avec du milieu 2xLB, facilitant l'utilisation des mini-cultures.

Les cellules sont incubées à 25 °C pendant la nuit. Les cellules sont ensuite centrifugées pendant 15 min à 3500 rpm et le milieu de culture est éliminé. Les culots cellulaires sont alors utilisés directement ou conservés à -20 °C.

Milieu	Composition	Induction
2xLB	20 g/L NaCl 20 g/L Tryptone 10 g/L Extrait de levure	700 mM IPTG
Graffinity ®	2xLB 0,5 % Glucose	2xLB 0,6 % Lactose 20 mM Hepes pH7
Milieu d'auto-induction TB	12 g/L Tryptone 24 g/L Extrait de levure 3,3 g/L $(NH_4)_2SO_4$ 6,8 g/L KH_2PO_4 7,1 g/L Na_2HPO_4 0,5 g/L Glucose 2 g/L Lactose 0,15 g/L $MgSO_4$ 0,003 g/L Oligoélément	/

Tableau 4 : Composition des milieux de culture utilisés.

2.1.2. Purification

Les culots cellulaires sont re-suspendus dans 1,2 mL de tampon de lyse. Plusieurs tampons avec notamment différentes salinités, pH, … peuvent être testés. Les cellules sont lysées par sonication avec un sonicateur à quatre têtes (Vibracell 75043, Bioblock Scientific), pendant 1 min avec une amplitude d'onde de 40% et une alternance de 2 s de sonication et 2 s de pause. Un échantillon, l'extrait total, est alors prélevé et contient l'ensemble des

protéines présentes dans les cellules bactériennes. Les cellules lysées sont ensuite centrifugées pendant 20 min, à 4000 rpm, permettant d'éliminer les débris cellulaires et les protéines insolubles. Le surnageant, correspondant à la fraction soluble, est prélevé.

L'étape de purification par affinité débute par le chargement de la fraction soluble sur une résine d'affinité, préalablement équilibrée avec le tampon de lyse correspondant et incubé pendant 1 à 2 h à 4 °C. Les résines fréquemment utilisées sont (i) la résine Glutathion Sepharose ® 4B (GE Healthcare), (ii) la résine TALON ® Superflow Metal Affinity Resin (Takara), et (iii) la StrepTactin® Sepharose High Performance (GE Healthcare), qui permettent respectivement la rétention de protéines couplées à l'étiquette GST, l'étiquette poly-histidines et l'étiquette Twin-Strep. Leurs caractéristiques sont détaillées dans le tableau 5.

Le surnageant est enlevé et la résine, liant les protéines d'intérêt couplées à l'étiquette de purification, est lavée deux fois avec le tampon de lyse, afin d'éliminer les protéines non retenues par la résine ou par leur interaction avec la protéine étiquetée. La résine est re-suspendue dans 30 µL de tampon Laemmli (100 mM chlorhydrate de tris(hydroxyméthyl)méthylamine (Tris-HCl) pH 6,8, 4 % dodécylsulfate de sodium (SDS), 20 % glycérol, 0,2 % bleu de bromophénol, 5 % β-mercaptophénol), entrainant le détachement des protéines de la résine. Les échantillons finaux sont analysés par électrophorèse sur gel d'acrylamide 10-15 % en présence de SDS (SDS-PAGE) et révélés par coloration au bleu de Coomassie. Nous obtenons ainsi 3 échantillons : (i) l'extrait total, qui contient toutes protéines exprimées, (ii) la fraction soluble, qui convient toutes les protéines solubles, (iii) l'échantillon final qui contient toutes les protéines fixées sur la résine d'affinité.

Les tests d'expression et de purification permettent ainsi de déterminer les conditions optimales permettant la bonne expression et solubilité d'une protéine, en jouant sur les différents paramètres discutés ci-dessus. A l'issue de ces tests, il est décidé de poursuivre l'optimisation de ces tests à petite échelle ou de passer à la prochaine étape, qui est de produire la(les) protéine(s) d'intérêt à grande échelle.

2.2. Expression et purification de protéines recombinantes à grande échelle

2.2.1. Production

Pour la production à grande échelle, il s'agit d'une mise à l'échelle du protocole décrit précédemment. Les cellules BL21(DE3) ou BL21(DE3) pRARE2 pLysS de *E. coli* sont (co-) transformées avec les vecteurs de (co-)expression d'intérêt. Une dizaine de colonies, obtenues des petites boîtes de pétri, sont ensuite utilisées pour un ré-étalement sur des grandes boîtes de Pétri (milieu LB/agar supplémenté avec les antibiotiques correspondants), et incubées à 37 °C, pendant la nuit, afin d'être réamplifiées et d'éviter les contaminations. Les bactéries sont ensuite re-suspendues dans 20 mL de milieu 2x LB et leur densité optique à 600 nm est mesurée afin de déterminer le volume à utiliser pour l'inoculation, la DO_{600} finale recherchée étant d'environ 0,1. Cette solution bactérienne sert ensuite à inoculer les bactéries dans des fioles en verre de 5 L, contenant 1 L de milieu culture et les antibiotiques correspondants. Les milieux majoritairement utilisés sont le milieu 2xLB ou le milieu d'auto-induction TB. Ces fioles sont ensuite incubées à 37 °C sous agitation à 200 rpm jusqu'à ce que la densité optique à 600 nm soit égale à 1,5, après la phase exponentielle de croissance. L'utilisation d'une DO élevée a été déterminée lors des tests à petite échelle et montre un meilleur ratio biomasse/expression protéique pour les protéines étudiées. Les fioles sont ensuite refroidies à une température de 18 - 25°C. L'induction de l'expression des protéines est réalisée par l'ajout d'IPTG (0,7 mM) pour le milieu 2xLB. Pour le milieu d'auto-induction, seul le changement de température est nécessaire. Les cultures sont incubées sous agitation à 25 °C pendant 12 à 15 h. Les cellules sont ensuite centrifugées pendant 25 min, à 4000 rpm à 4°C et le surnageant est éliminé. Les cellules sont re-suspendues dans le tampon de lyse adéquat (40 ml de tampon / 1 L de culture initiale), dont la composition a été déterminée à partir des tests d'expression et de purification à petite échelle. Les cellules sont conservées à - 20 °C, jusqu'à leur prochaine utilisation.

2.2.2. Purification

2.2.2.1. Lyse des cellules

Pour 3 L de culture initiale, les cellules décongelées subissent une sonication sur glace, trois cycles de 1 min 30, avec une alternance d'impulsion/pause de 5 s et une amplitude de 30%), grâce au sonicateur Labsonic Sartorius®. Les cellules sont centrifugées, pendant 1 h à

18 000 rpm et à 4 °C pour clarifier le lysat et précipiter les débris cellulaires et les protéines insolubles.

2.2.2.2. Chromatographie d'affinité

La chromatographie d'affinité est une technique basée sur la fixation spécifique et réversible d'une étiquette de purification fusionnée de manière covalente avec la protéine d'intérêt à une matrice. Les trois techniques d'affinités majoritairement utilisées pendant ma thèse reposent sur l'interaction entre l'étiquette et un cation divalent ou sur l'interaction entre un enzyme et un substrat. Les caractéristiques des étiquettes de fusion, ainsi que leurs résines associées sont présentées dans le tableau 5. Cette technique de chromatographie comprend plusieurs étapes. La première étape consiste à fixer la protéine d'intérêt fusionnée avec l'étiquette, alors que les autres protéines présentes dans le lysat clarifié ne sont pas retenues. Une succession de lavages permet ensuite d'éliminer de potentiels contaminants non spécifiques. Enfin, la dernière étape consiste à soit éluer par compétition la protéine d'intérêt associée à l'étiquette de fusion de la résine, en utilisant le compétiteur adéquat en fonction de la résine, soit à la dissocier de la résine par coupure de la séquence spécifique située entre l'étiquette de fusion et la protéine d'intérêt, à l'aide d'une protéase.

Concernant le protocole utilisé, la fraction soluble obtenue de la lyse est incubée pendant 1 à 2 h à 4 °C avec la résine d'affinité choisie, qui a été préalablement équilibrée avec le tampon de lyse. La résine est ensuite centrifugée pendant 10 min à 1000 rpm et à 4 °C. Un échantillon du lysat, appelé fraction non retenue et correspondant aux protéines non retenues, est prélevé et le surnageant est enlevé. La résine est lavée trois fois avec un large excès de tampon de lyse, en utilisant une étape de centrifugation entre chaque lavage. L'échantillon est alors soit élué par compétition, soit coupé à l'étiquette de fusion par une protéase, permettant de libérer la protéine d'intérêt. Pour la première technique, l'échantillon est élué, en étant incubé pendant 15min avec le tampon d'élution adapté, en deux temps. Le volume de tampon utilisé est adapté en fonction du volume de culture initial, de 0,5 ml pour 1L de culture jusqu'à 5 ml pour 6L de culture. Concernant le clivage par une enzyme, la résine est re-suspendue dans un volume de 5 mL de tampon de lyse et digéré pendant la nuit par la protéase 3C (1 mg/ml), produite au sein de mon laboratoire d'accueil, à raison de 1 U de protéase pour 25 μg de protéine à cliver. Le site de coupure de cette enzyme

se situe entre la protéine d'intérêt et son étiquette. La protéine clivée est récupérée et sa purification se poursuit par une purification par héparine ou par chromatographie par échangeuse d'ion, afin d'éliminer notamment l'ADN contaminant.

Étiquette d'affinité	Glutathion-S-Transférase (GST)	Poly-histidine $(His)_6$ ou $(His)_{10}$	Twin-Strep
Taille de l'étiquette	26 kDa	1,3 kDa	2,9 kDa
Résine	Glutathion Sepharose ® 4B (GE Healthcare)	TALON ® Superflow Metal Affinity Resin (Takara)	StrepTactin ® Sepharose High Performance (GE Healthcare)
Molécule liée à la matrice	Glutathion	Cobalt	Strep-Tactin
Mode d'élution	Glutathion (Sigma) ou clivage par une protéase	Imidazole (Sigma) ou clivage par une protéase	Desthiobiotine (Sigma) ou clivage par une protéase
Avantages	- Purification rapide, non dénaturante, très sélective - Augmentation potentielle de la solubilité de la protéine d'intérêt	- Purification rapide, non dénaturante, sélective - Étiquette de très petite taille	- Purification rapide, non dénaturante, très sélective -Étiquette de très petite taille
Inconvénients	- Dimérisation de l'étiquette GST -Gêne stérique, du fait de la taille de l'étiquette	- Liaisons aspécifiques avec d'autres protéines (présentant naturellement des régions riches en poly-histidines ou liant les métaux	- Possible diminution du rendement lors d'une purification en ¨batch¨

Tableau 5 : Principales étiquettes d'affinité utilisées

2.2.2.3. Purification par héparine

La chromatographie d'affinité à l'héparine permet de purifier une large gamme de protéines, comprenant des enzymes, des inhibiteurs de protéase à sérine, des facteurs de

croissance, des protéines de matrice extracellulaire, des protéines liant les acides nucléiques, des récepteurs hormonaux et des lipoprotéines selon deux modes principaux d'interaction. En effet, l'héparine peut être utilisée en tant que ligand d'affinité et retenir notamment des facteurs de croissance et l'antithrombine III. L'héparine peut également, grâce à son action d'échangeur de cations, fixer des protéines liant les acides nucléiques, en mimant la structure polyanionique de ces derniers. L'élution des protéines se fait ensuite par une augmentation de la force ionique grâce à un gradient de NaCl.

Lors de ma thèse, j'ai utilisé la colonne HiTrap® Heparin HP de 5 ml (Cytiva), branchée sur un système de purification AKTA Pure (GE Healthcare). L'échantillon contenant la protéine d'intérêt est dilué afin d'obtenir une concentration de 50 mM de NaCl (tampon de liaison). L'échantillon est ensuite chargé sur cette colonne, préalablement équilibrée avec 10 volumes colonne (CV) avec le tampon de liaison. La colonne est lavée avec 5 CV avec ce même tampon, puis l'échantillon est élué par un gradient continu de 10 CV de NaCl allant de 50 mM à 1 M. Cette étape permet ainsi de séparer les protéines d'intérêt de l'ADN contaminant, co-purifié lors de la première étape de purification par affinité.

2.2.2.4. Purification par échangeuse d'ions

La chromatographie par échangeuse d'ions permet la séparation des protéines et molécules selon leur charge. Cette technique se base sur l'interaction entre les protéines chargées et la matrice stationnaire de la colonne de charge opposée. Le choix de l'échangeuse d'ions dépend alors du point isoélectrique de la protéine, ainsi que du pH du tampon considéré.

J'ai utilisé les colonnes échangeuses d'anion fort HiTrap ® Q HP de 1 ou 5 ml (Cytiva), branchée sur un système de purification AKTA Pure (GE Healthcare). L'échantillon contenant la protéine d'intérêt est dilué afin d'obtenir une concentration de 50 mM de NaCl. L'échantillon est ensuite chargé sur cette colonne, préalablement équilibrée avec 10 volumes colonne (CV) avec le tampon de purification contenant 50 mM NaCl. La colonne est ensuite lavée avec 5 CV avec ce même tampon, puis l'échantillon est élué par un gradient continu de 10 CV allant de 50 mM à 1 M de NaCl. Tout comme la purification par héparine, cette étape permet dans notre cas de séparer les protéines d'intérêt de l'ADN contaminant, qui a été co-purifié lors de la première étape de purification par affinité.

2.2.2.5. Chromatographie d'exclusion stérique

La chromatographie d'exclusion stérique (SEC) permet de séparer les molécules selon leur taille moléculaire et leur forme, en fonction de leur volume d'élution dans des colonnes remplies d'un gel poreux. Les molécules de grande taille s'éluent en premier, celles de plus petites tailles pénètrent dans les pores à des degrés différents selon leur taille et sont éluées en dernier. Il s'agit généralement de la dernière étape de purification, qui permet notamment de séparer les agrégats ou les impuretés résultants de la dégradation des protéines d'intérêt, ainsi que d'évaluer l'homogénéité de l'échantillon (formation de monomère, dimères, …) et de s'assurer que l'échantillon est dans un tampon bien défini.

Durant ma thèse, j'ai utilisé plusieurs colonnes de SEC branchées sur le système de chromatographie AKTA Pure (GE Healthcare). La première est la colonne analytique Superdex® 200 10/300 GL (GE Healthcare). Cette colonne présente une gamme de séparation de 10 à 600 kDA et a un volume de résine de 24 mL, avec une injection maximale de 0,5 - 1mL d'échantillon. Cette colonne est majoritairement utilisée lors des premiers essais à grande échelle, où le volume de culture est plutôt faible (1 – 3 L de culture) ou lorsque de faibles quantités injectées sont nécessaires.

La seconde colonne est la colonne préparative Superdex® 200 16/200. Cette colonne présente la même gamme de séparation que celle décrite précédemment, avec toutefois un volume de résine (124 mL) ainsi qu'un volume d'injection (5 mL) supérieurs. Elle permet donc de purifier de plus grandes quantités de protéines nécessaires à la caractérisation biophysique, biochimique et structurale. Ces deux colonnes Superdex ont été utilisées pour la purification des complexes SMC1A-HD/RAD21C et SMC3-HD/RAD21N.

La dernière colonne utilisée est la Superose® 6 increase 10/300 GL, qui possède une gamme de séparation allant de 5 à 5000 kDa, un volume de résine de 24 mL et une injection maximale de 0,5 – 1 mL. Cette colonne a notamment été utilisée pour la purification du complexe quaternaire de la Cohésine SMC1A-HD/SMC3-HD/RAD21 /STAG1 ou STAG2.

Concernant le protocole, après l'étape de purification par l'héparine ou par échangeuse d'anion, l'échantillon est concentré à l'aide d'un concentrateur (Amicon, Ultracel 30 kDa – 100 kDa, selon les protéines concentrées) jusqu'à obtenir le volume désiré de 1 ou 5 mL. L'échantillon est chargé sur la colonne de chromatographie d'exclusion de taille choisie, préalablement équilibrée par 1,2 CV du le tampon de lyse supplémenté avec de

l'hydrochlorate de (tris(2-carboxyéthyl)phosphine (1 mM TCEP) ou du dithiothréitol (2mM) et du chlorure de magnésium (5 mM $MgCl_2$). Ensuite, 30 µL de chaque fraction des pics d'intérêt sont prélevés et mélangés à 20 µL de tampon Laemmli et ces fractions sont analysées par migration sur gel SDS-PAGE à 10 ou 15 %. Les protéines sont révélées par coloration au bleu de Coomassie et les fractions analysées. Les fractions contenant les protéines d'intérêt sont sélectionnées et mélangées, puis concentrées à l'aide d'un concentrateur (Amicon, Ultracel 30 ou 100 kDa) jusqu'à obtenir une concentration protéique de l'ordre du mg/mL (concentration différente selon la protéine purifiée et le type d'expérience utilisée par la suite). Cette concentration est estimée par la mesure de son absorbance à 280nm ou par l'utilisation du réactif Bradford.

III. Caractérisation physico-chimique

Cette caractérisation permet de déterminer les propriétés physiques et chimiques des protéines d'intérêt, ainsi que d'évaluer leur stabilité, leur homogénéité et leur activité enzymatique.

3.1. Diffusion dynamique de la lumière

La diffusion dynamique de la lumière (DLS) est une technique estimant la taille des particules en solutions, ainsi que leur distribution. Elle est basée sur la théorie des mouvements browniens des particules (mouvements thermiques aléatoires), selon laquelle les petites particules se déplacent plus rapidement, tandis que les grandes particules se déplacent plus lentement en solution. Ces mouvements sont modélisés par la relation de Stokes-Einstein.

Au cours d'une expérience de DLS, un laser traverse l'échantillon et l'intensité de la lumière laser diffusée par les particules est mesurée en fonction du temps. Des variations de ces intensités sont induites par les mouvements browniens, qui eux-mêmes diffèrent selon la taille des particules. Le traitement informatique de ces données renseigne sur le coefficient de diffusion transversal, permettant alors de déterminer le diamètre hydrodynamique des particules. Ainsi, la taille de l'échantillon et le pourcentage de polydispersité de l'échantillon, indicateur de son homogénéité ou de la formation d'agrégat, peuvent être déterminés.

Avant toute mesure de DLS, les échantillons sont centrifugés pendant 10 min à 14 000 rpm, afin d'éliminer les gros agrégats potentiels. Les mesures ont été prises avec l'appareil DLS DynaPro NanoStar (Wyatt), à 4°C et en utilisant 50 µL d'échantillon.

3.2. SEC-MALS

Le SEC-MALS est le couplage de la chromatographie d'exclusion de taille (SEC) et la diffusion de lumière multi-angles (MALS). Cette technique permet d'évaluer la masse molaire exacte des protéines en solution, ainsi que leur état oligomérique, grâce à la mesure de l'indice de réfaction ,tout en s'affranchissant du volume d'élution des protéines (Some et al., 2019).

Cette technique a été utilisée pour caractériser la protéine STAG1-1. Cette expérience a été menée par la plateforme de biologie structurale de l'IGBMC, à l'aide d'un système de chromatographie Ettan microLC (GE Healthcare) couplé à la colonne Superdex® 200 10/300, d'un détecteur de diffusion de lumière multi-angles miniDawn™ TREOS™, d'un module QELS pour la mesure du rayon hydrodynamique et d'un réfractomètre Optilab® T-rEX (Wyatt Technology).

3.3. Spectrométrie de masse

La spectrométrie de masse est une technique permettant l'identification et la quantification des protéines, par l'ionisation de peptides provenant de la digestion de ces protéines (Menet, 2011). Cette technique est utilisée afin de vérifier que les protéines d'intérêt sont bien celles obtenues et est réalisée par la plateforme de protéomique de l'IGBMC.

Pour ce faire, les protéines d'intérêt sont migrées sur gel SDS-PAGE, leurs bandes correspondantes sont coupées du gel et digérées par la trypsine, afin d'extraire les peptides propres à chaque protéine. Ces peptides sont ensuite ionisés et analysés grâce à l'appareil Orbitrap Elite™.

3.4. Titration calorimétrique isotherme

La titration calorimétrique isotherme (ITC) est une technique analytique permettant de mesurer l'affinité (constante d'affinité K_D) et d'autres paramètres thermodynamiques (enthalpie ΔH, entropie ΔS, enthalpie libre ΔG, stœchiométrie) d'une liaison entre deux partenaires. Il s'agit d'une technique de référence pour les études quantitatives d'interactions moléculaires. C'est la seule technique capable de détecter tous les paramètres de l'interaction simultanément en une seule expérience. De plus, il n'est pas nécessaire d'immobiliser ou de marquer de manière fluorescente l'un des partenaires. Toutefois, cette méthode nécessite une grande quantité de protéines et des tampons strictement identiques pour les échantillons, ce qui peut être un frein pour son utilisation.

Le calorimètre est composé de deux cellules : la cellule de référence, qui contient de l'eau et la cellule d'échantillon, qui contient le premier partenaire en solution dans un tampon adéquat. Une seringue complète ce dispositif et contient le deuxième partenaire (ligand ou protéine). Lors d'injection répétée du ligand dans la cellule échantillon, une chaleur est dégagée ou absorbée lors de l'interaction entre les deux partenaires. Le calorimètre mesure alors cette variation de chaleur, en établissant la puissance nécessaire pour maintenir une différence de température nulle entre la cellule de référence et la cellule d'échantillon (Figure 28).

Les mesures ont été réalisées en collaboration avec le Dr. Eric Ennifar et Karl Brillet (IBMC, Strasbourg), avec l'appareil Microcal PEAQ-ITC (Malvern). Nos mesures ont consisté à titrer les complexes SMC1A-CC/RAD21C ou SMC3-CC/RAD21N WT et mutants EQ (SMC1ACC-E1157, SMC3CC-E1144Q) avec de l'ATP (Sigma), son analogue non-hydrolysable, l'ATPγS (Jenabioscience) ou l'ADP (Sigma). Ces ligands ont été dissous dans le même tampon que les protéines à 4 mM. Les mesures ont été effectuées à 5°C. 250 µL du complexe d'intérêt (concentration comprise entre 200 et 300 µM) sont placés dans la cellule d'échantillon. L'ATP, l'ATPγS ou l'ADP est chargé dans une seringue et injecté toutes les 120 secondes dans la cellule échantillon, jusqu'à ce que la réaction de liaison atteigne l'équilibre (plateau). Chaque injection de ligand produit une variation de chaleur, qui est représentée graphiquement en fonction du temps. Chaque pic est intégré et normalisé en fonction du ratio molaire (ligand/échantillon), grâce au logiciel d'analyse de Microcal PEAQ-ITC. L'isotherme final permet de déduire l'affinité (K_D), la stœchiométrie (n), l'enthalpie ΔH, l'enthalpie libre ΔG et

l'entropie ΔS de l'interaction entre l'ATP, l'ATPγS ou l'ADP et les complexes d'intérêt (Figure 28).

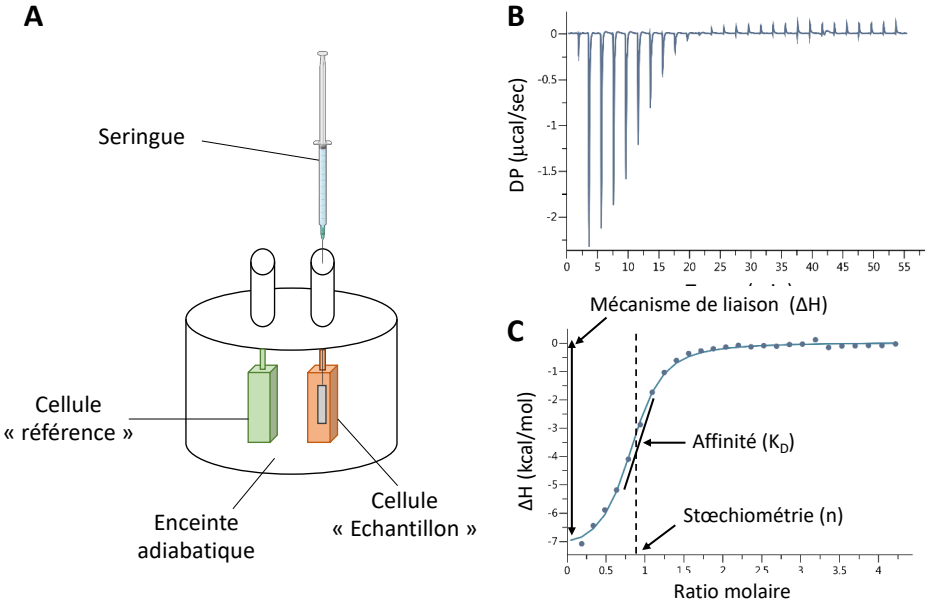


Figure 28 : Principe de la titration calorimétrique isotherme.
A) Représentation schématique des différents composants du calorimètre. La protéine d'intérêt est déposée dans la cellule échantillon et la deuxième molécule (ligand ou protéine) est injectée dans cette cellule à l'aide de la seringue. La chaleur libérée (réaction exothermique) ou absorbée (réaction endothermique) par l'interaction entre les deux composés est mesurée par rapport à la cellule de référence.
B) Exemple de données brutes obtenues lors d'une expérience d'ITC, correspondant à la libération de la chaleur mesurée lors de chaque injection du ligand. Un phénomène de saturation est observé au cours du temps.
C) L'analyse des données, par intégration et normalisation, permet de déterminer le changement d'enthalpie (ΔH), l'affinité (K_D), et la stœchiométrie.

3.5. Expériences de réticulation *in vitro* (crosslinking)

La réticulation désigne le processus de jonction chimique intermoléculaire de deux ou plusieurs molécules par la formation d'une liaison covalente. Cette technique peut être utilisée pour étudier les interactions entre protéines, en facilitant notamment l'identification de domaines protéiques proches l'un de l'autre ou en stabilisant leur interaction.

Les liaisons covalentes sont formées par des réactifs de réticulation, qui agissent sur différents groupements chimiques et selon différentes longueurs d'espacement entre les molécules. Ces réactifs présentent plusieurs groupements chimiques, qui vont réagir et connecter les groupes fonctionnels, tels que les amines primaires, les sulfhydryles, les carbonyles, …

Il existe 3 classes d'agents de réticulation : (i) les réactifs homobifonctionnels, (ii) les réactifs hétérobifontionnels et (iii) les photoréactifs. Les premiers possèdent des groupements réactifs identiques à chaque extrémité et lient généralement des groupements fonctionnels similaires. Les seconds possèdent deux groupements réactifs différents et peuvent lier des groupements fonctionnels distincts. Les troisièmes sont des agents hétérofonctionnels qui s'activent uniquement sous lumière ultraviolette ou visible (Scientific, 2012).

Lors de ma thèse, cette technique a été utilisée pour identifier les différentes conformations adoptées par la superhélice proche de la tête ATPase de SMC3 par la Dr. Marie-Laure Diebold-Durand. Pour ce faire, un réactif homobifonctionnel, le bis-maléimidoéthane (BMOE, Sigma) a été utilisé. Il possède deux groupements maléimides à ses extrémités, qui sont séparés par un espaceur de 8 Å et interagissent avec les groupes sulfhydryles (-SH), présents sur les cystéines, formant alors une liaison thioéther (Figure 29). Ainsi, le BMOE permet de réticuler deux cystéines étant dans un voisinage proche de 8 Å.

Pour ces expériences, toutes les cystéines endogènes des constructions SMC1ACC, SMC3CC, RAD21N et RAD21C ont été mutées en sérines pour éviter les réticulations non spécifiques initialement observées entre le BMOE et les cystéines natives. Les résidus D92 et D120 de SMC3CC, situés sur la tête, et les résidus K25 ou K26 de RAD21, situés dans la région interagissant avec la superhélice de SMC3, sont mutés en cystéines. Ces résidus ont été choisis en fonction de leur distance. Les complexes SMC3CC-D120C/RAD21N-K25C et SMC3CC-D92C/RAD21N-K26C ont été purifiés selon le protocole standard décrit ci-dessus. Les protéines SMC3CC/RAD21N mutées ont été mélangées à une concentration finale de 6 µM, seules ou en présence de SMC1ACC/RAD21 (7 µM), ATP (20 µM), ADN (10 µM) et MgCl2 (40 µM). Les réactions ont été incubées pendant 5 min sur la glace ou à température ambiante avant l'addition de 0,5 mM de BMOE ou de DMSO (contrôle). La réaction a été bloquée par l'ajout de tampon Laemmli après 2, 5 et 10 min et chauffée à 70°C pendant 5 min. Les échantillons ont été chargés sur gel d'acrylamide et l'intensité des bandes a été mesurée à l'aide du logiciel Fiji (Schindelin et al., 2012). Les expériences ont été réalisées en triplicata.

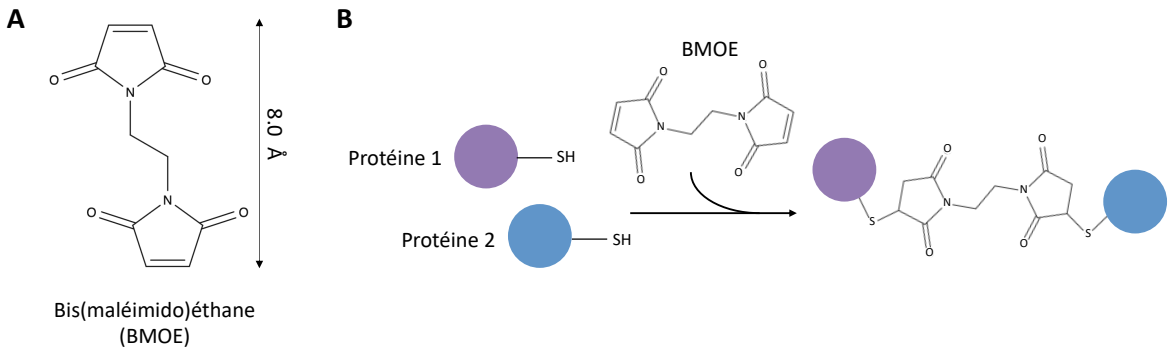


Figure 29 : Principe de la réticulation par le BMOE.
A) Structure du BMOE.
B) Réaction entre le BMOE et les groupements sulfhydryles des cystéines présentes sur les protéines 1 et 2, qui sont dans un voisinage proche de 8 Å. Le groupement maléimide du BMOE réagit avec le groupement sulfhydryle pour former une liaison thioéther stable.

3.6. Test enzymatique

La mesure de l'activité ATPase des complexes SMC1A-HD/RAD21C et SMC3-HD/RAD21N sauvages et mutants a été effectuée à l'aide du kit EnzChek Phosphate Assay (Thermo Fischer Scientific). Il s'agit d'un test spectrophotométrique en microplaque, qui permet de quantifier le phosphate inorganique en solution, correspondant au phosphate libéré lors de l'hydrolyse de l'ATP par les têtes de SM1A et de SMC3. Ce kit se compose du substrat 2-amino-6-mercapto-7-méthylpurine riboside (MESG), de l'enzyme purine nucléoside phosphorylase (PNP) et de phosphate, afin d'établir la gamme phosphate utilisée lors du calcul de l'activité. L'enzyme PNP convertit le MSG, grâce à la présence de phosphate libéré lors de l'hydrolyse, en ribose 1-phosphate et en 2-amino-6-mercapto-7-methyl-purine (Figure 30). Le MESG possède une absorbance à 330nm, alors que le produit final présente une absorbance maximale à 360 nm. Ainsi, la mesure de l'absorbance à 360 nm permet de déterminer l'activité ATPase de SMC1A et SMC3.

Ces essais nécessitent 20 µM des complexes d'intérêt (testés séparément) ou 10 µM de chacun des complexes protéiques préalablement mélangés. Ces complexes ont été incubés avec 1 mM d'ATP. L'activité d'hydrolyse de l'ATP a été évaluée à 30 °C, en mesurant l'absorbance à 360 nm toutes les 42 secondes pendant 2 heures, à l'aide d'un lecteur de plaques spectrophotométrique (TECAN). Les expériences ont été réalisées en duplicata et triplicata.

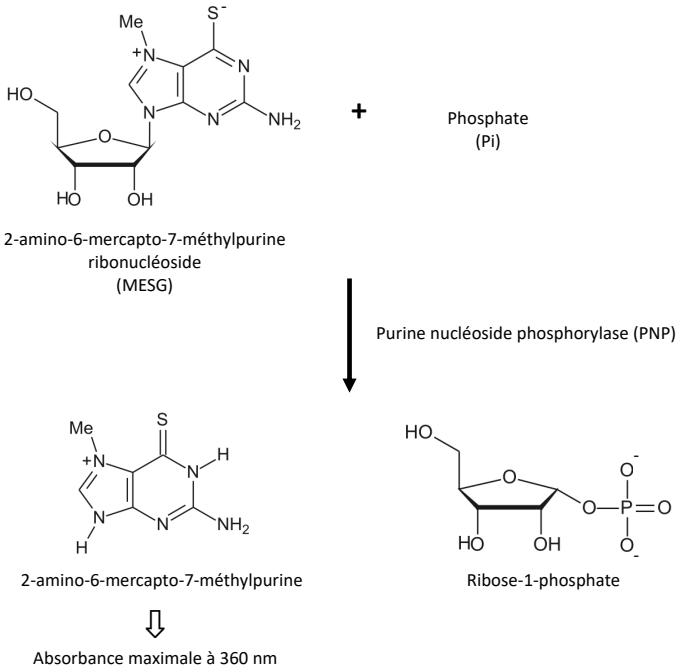


Figure 30 : Conversion enzymatique du MESG en ribose-1-phosphate et 2-amino-6-mercapto-7-méthylpurine par la purine nucléoside phosphorylase (PNP).
Le changement d'absorption à 360 nm qui accompagne cette réaction permet de quantifier le phosphate inorganique (Pi) consommé lors de la réaction. Adapté selon le protocole du fabriquant (Thermofisher).

3.7. Reconstitution du module ATPase

3.7.1. Par méthode biochimique : SEC

En plus de son importance dans les dernières étapes de purification, la SEC permet également d'évaluer les interactions protéines-protéines. En effet, du fait de la discrimination par la taille et de la forme des protéines d'intérêt, il est possible de suivre la formation ou non d'un complexe entre plusieurs protéines. En effet, un décalage du volume d'élution vers les hauts poids moléculaires est observé lors de la formation d'un complexe par rapport aux volumes d'élution des protéines seules (Figure 31). La SEC présente certains avantages en comparaison à d'autres techniques, comme l'ITC. Tout d'abord, elle ne nécessite pas de préparations particulières pour la réalisation de l'expérience et est assez rapide à mettre en place. De plus, la persistance d'un complexe lors de son passage à travers une colonne d'exclusion de taille et soumis à un débit constant et démontre que le complexe est stable dans le temps et non transient. Les composants du complexe interagissent les uns avec les autres de manière forte au cours du temps. C'est ainsi une technique de choix pour confirmer la formation du module ATPase (SMC1A-HD/RAD21C/SMC3-HD/RAD21N/ATP).

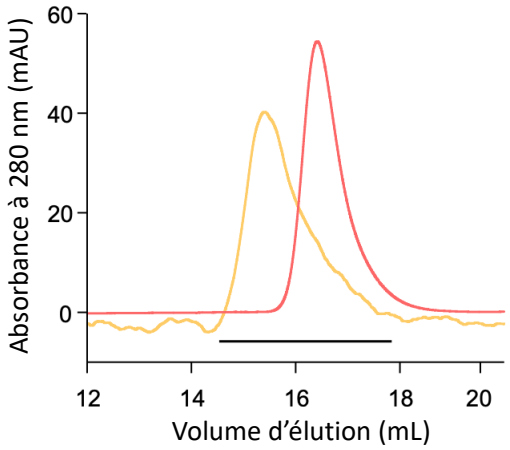


Figure 31 : Résultats typiques d'une expérience de reconstitution d'un complexe par SEC, où est observé un décalage vers les plus hauts poids moléculaires lors de la formation d'un complexe (courbe jaune) par rapport au profil de SEC des protéines seules (courbe orange).

Pour ce faire, j'ai utilisé les complexes SMC1ACC/RAD21C WT, E1157Q, D1163E ou L1128V et les complexes SMC3CC/RAD21N WT et E1144Q ou SMC3J/RAD21N E1144Q, L115V ou D1150E. Ces mutations affectent le site catalytique des têtes ATPases de SMC1A et de SMC3, réduisant ainsi leur activité d'hydrolyse résiduelle. J'ai également testé la formation du module ATPase en présence d'ATP (Sigma), en présence de deux analogues de l'ATP faiblement hydrolysables, l'ATP-γ-S (Jenabioscience) et l'AMP-PNP (Jenabioscience), ainsi qu'en présence d'ADP (Sigma). Ces complexes ont été purifiés indépendamment et dilués à une concentration finale de 50 µM. Ils ont ensuite été chargés sur une colonne Superdex S200 10/300 (GE Healthcare), préalablement équilibrée avec le tampon utilisé pour la chromatographie d'exclusion de taille, c'est-à-dire soit sans nucléotide, soit supplémentés avec le nucléotide choisi (1 mM ADP, 1 mM ATP, 0,5 mM ATP-γ-S ou 0,5 mM AMP-PNP). La liste des combinaisons testées est décrite dans le chapitre 3 - Résultats.

3.7.2. Par méthode biophysique : ultracentrifugation analytique

L'ultracentrifugation analytique (AUC) est une technique puissante et polyvalente pour caractériser les protéines et les complexes protéiques en solution. Elle est basée sur les propriétés de sédimentation des protéines en solution, lorsque ces dernières sont soumises à une force centrifuge extrême. Le comportement des protéines est suivi en temps réel par un système optique (absorbance, interférence de Rayleigh ou fluorescence), en mesurant le mouvement de ces protéines lors de la centrifugation en fonction de la distance à l'axe de

rotation. L'évolution de la répartition des protéines en fonction du temps est ainsi déterminée. Il existe deux approches d'AUC : l'AUC à vitesse de sédimentation (AUC-SV) et l'AUC à l'équilibre de sédimentation (AUC-SE) (Schuck et al., 2016; Ebel and Birck, 2021).

La première approche, l'AUC-SV, détermine la vitesse de sédimentation, d'après l'équation de Svedberg, qui est propre à chaque protéine/complexe et qui dépend de leur taille, de leur forme et de leur densité, ainsi que des propriétés du tampon et de la centrifugeuse. Cette technique permet d'évaluer à la fois l'homogénéité/hétérogénéité d'un échantillon et ses propriétés de liaison (Figure 32). Elle présente certains avantages pour l'étude des complexes protéiques, du fait d'une résolution hydrodynamique élevée en fonction de la taille permettant la séparation des espèces libres et liées (Schuck et al., 2016; Ebel and Birck, 2021).

Quant à l'AUC-SE, elle permet d'analyser l'état d'équilibre final des protéines, où la vitesse de migration de ces dernières est exactement compensée par leur diffusion. Cela permet alors de déterminer de manière plus précise la masse moléculaire des protéines (Schuck et al., 2016).

Lors de ma thèse, j'ai utilisé la technique d'AUC-SV pour caractériser la formation du module ATPase. Les expériences ont été réalisées avec la centrifugeuse ProteomeLab® XL (Beckman Coulter). Les complexes SMC1ACC-EQ/RAD21C and SMC3CC-EQ/RAD21N ont été purifiés séparément et dilués à une concentration finale de 1 mg/mL pour tester la présence d'homodimères ou à un ratio de 1 : 0.9 avec une concentration finale de 1 mg/mL pour tester la formation du module ATPase. Les nucléotides testés ADP, ATP, AGS, AMP-PNP ont été ajoutés aux échantillons à une concentration finale de respectivement 1 mM, 1 mM, 0,5 mM et 0,5 mM. Les échantillons ont ensuite été centrifugés à 4 °C et à une vitesse de 42,000 rpm. Les mesures d'absorbance à 280 nm et d'interférence ont été prises pendant 24 h, avec un intervalle de 9 min entre chaque prise. Concernant le traitement des données, le volume spécifique partiel de la protéine (υ), la densité (ρ), et la viscosité (η) des échantillons ont été estimés grâce au logiciel SEDNTERP (Laue, 1992). Les données ont été analysées dans le logiciel SEDFIT (Schuck, 2000), en utilisant l'analyse de la distribution c(s) continue. Les graphiques des superpositions des courbes d'AUC ont été réalisées dans le logiciel GUSSI (Brautigam, 2015).

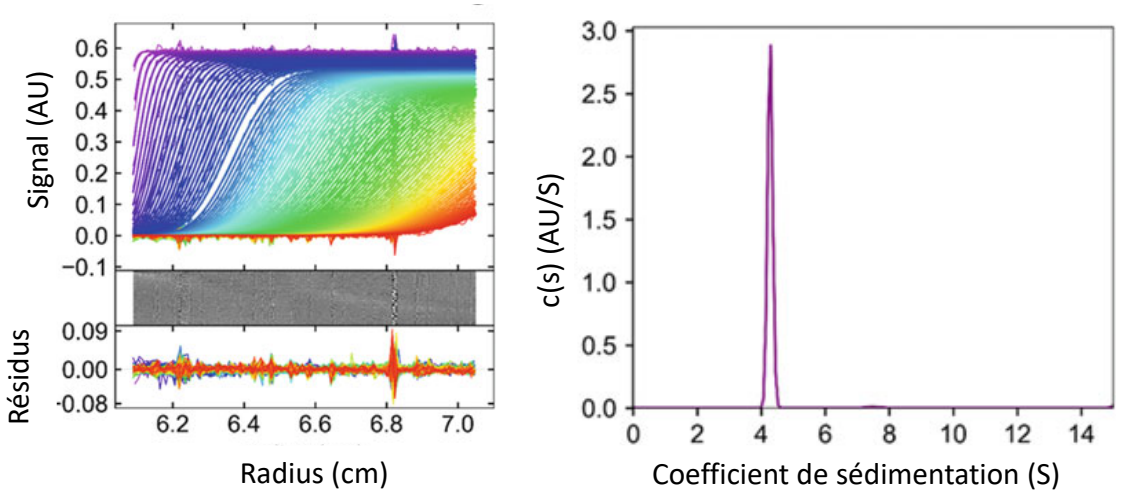


Figure 32 : Résultats typiques d'une expérience de AUC-SV (ici avec du BSA), avec le profil des vitesses de sédimentations expérimentales (gauche) et le profil obtenu après traitement informatique (droite), permettant de déterminer un coefficient de sédimentation caractéristique de chaque protéine. Adapté de (Ebel and Birck, 2021).

IV. Caractérisation structurale

La résolution de la structure d'une protéine a pour but de mieux comprendre à l'échelle moléculaire son organisation 3D et son mode d'action. Il existe plusieurs méthodes de biologie structurale, qui permettent l'obtention de structure de basse à haute résolution. La technique de diffusion des rayons X aux petits angles (SAXS) permet d'obtenir une enveloppe de la protéine d'intérêt en solution, alors que les techniques de cristallographie par diffraction aux rayons X et de cryo-microscopie électronique (cryoEM), ainsi que la résonance magnétique nucléaire aboutissent à des structures de haute résolution. Lors de ma thèse, je me suis tout d'abord concentrée sur la technique de cristallographie pour obtenir une structure du module ATPase de la Cohésine. Cependant, du fait de la flexibilité intrinsèque du module ATPase, cela ne s'est pas avéré concluant. Je me suis alors tournée vers une autre technique, la cryoEM, pour résoudre la structure de ce complexe.

4.1. Essai de cristallisation pour la cristallographie par diffraction aux rayons X

La cristallographie par diffraction aux rayons X est fondée sur la diffraction des rayons X par la matière sous forme cristalline (Hammond, 2015; Hoffmann, 2020). Cette technique nécessite une première étape clé et indispensable : la cristallisation de l'échantillon protéique. Pendant ce processus, les molécules des protéines passent d'un état désordonné en solution

à un état ordonné dans un cristal. Cet état est induit par l'ajout d'agents précipitant, tels que les sels (sulfate d'ammonium, phosphate de potassium, …), les polymères (polyéthylène glycérol, PEG), par l'ajout d'additifs, ainsi que par les paramètres expérimentaux (concentration et pureté de la protéine, pH, force ionique, …). Il existe donc une multitude de conditions à tester afin de trouver les paramètres favorisant la cristallisation. De nombreux screens commerciaux sont disponibles à la plateforme de biologie structurale intégrative de l'IGBMC.

J'ai mené des essais de cristallisation sur le module ATPase en présence d'ATP en utilisant plusieurs combinaisons des constructions de SMC1A-HD et SMC3-HD (SMC1ACC/RAD21C/SMC3J/RAD21N et SMC1ACCsh/RAD21C/SMC3CC/RAD21N) à différentes concentrations (5, 7,5 et 10 mg/mL).

Ces essais ont été réalisés dans des plaques de cristallisation MRC2 96 puits à 3 gouttes (Lens Crystallization Plate, Swissci) et de nombreux screens commerciaux ont été testés : Classics (Qiagen), PACT (Molecular Dimension), PEG (Qiagen), JCSG (Qiagen) et Wizard (Rigaku), SALT RX (Hampton Research) et TOP 96 (Anatrace). Les gouttes de cristallisation ont été préparées selon la méthode de goutte assise, à l'aide du robot Mosquito Crystal (TTP Labtech ®), avec un réservoir de 40 µL. Pour chaque condition de cristallisation, un mélange 1 :1 de 200 nL de solution protéique et 200 nL de solution de cristallisation a été réalisé à une température de 4 °C et ou de 20 °C, sous atmosphère humide (80%). Les gouttes de cristallisation ont été observées au microscope après 1, 3, 7 jours et plus. Quelques conditions parmi les screens commerciaux ont donné des cristaux prometteurs. J'ai alors réalisé de nombreuses optimisations, en jouant plus précisément sur le pH, la température, la concentration saline, etc, dans le but d'améliorer la qualité des cristaux. Les quelques cristaux de taille suffisante obtenus ont été congelés dans leurs solutions de cristallisation supplémentées avec 20 % de glycérol ou 20 % de PEG200 (cryo-protectants) et congelés dans de l'azote liquide. Les cristaux ont ensuite été testés au Swiss Ligh Source Synchrotron (SLS, Switzerland) ou sur le système de diffraction des rayons X de l'IGBMC. Certains cristaux ont présenté un pouvoir diffractant, mais insuffisant pour la résolution de structure.

4.2. Diffusion des rayons X aux petits angles

La diffusion des rayons X aux petits angles permet de déterminer à basse résolution la forme, la conformation et l'état d'assemblage des macromolécules en solution. Cette technique mesure l'intensité des rayons X diffusés par un échantillon en fonction de l'angle de diffusion (Kikhney and Svergun, 2015). Les expériences de SAXS ont été menées sur le module ATPase engagé en présence d'ATP, à une concentration de 2 mg/ml et réalisées par la plateforme de biologie structurale de l'IGBMC. L'appareil de diffusion des rayons X aux petits angles BioSAXS-1000 (Rigaku) a été utilisé. Le traitement des données a été fait grâce au programme DAMMIN, qui, à partir de la diffusion de rayons X aux petits angles des protéines, permet de déterminer la forme *ab initio* à basse résolution de macromolécules orientées aléatoirement en solution (Svergun, 1999).

4.3. Cryo-microscopie électronique

La cryoEM est une technique permettant d'étudier la structure moléculaire de complexes protéiques dans des conditions proches de leur état natif. Elle consiste à observer, à l'aide d'un microscope électronique à transmission (TEM), ces complexes macromoléculaires, qui ont été inclus dans de la glace vitreuse, ce qui permet à terme d'élucider leur structure 3D à partir des images de projections 2D collectées. Les différentes étapes de cette méthode sont illustrées dans la figure 33.

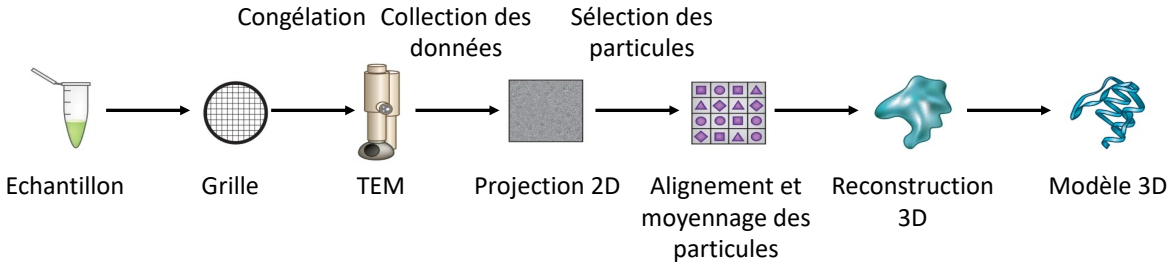


Figure 33 : Les différentes étapes de cryoEM, de l'échantillon protéique au modèle 3D.
Adapté de (Doerr, 2016).

La cryoEM a connu une forte évolution depuis la création du premier microscope à transmission électronique dans les années 1930 par le physicien allemand Ernst Ruska. Un autre tournant a été le développement d'une méthode pour la préparation cryogénique des échantillons pour les particules uniques par Jacques Dubochet dans les années 1980. Cette technique prend un nouvel essor depuis une dizaine d'années, du fait des innovations techniques majeures concernant l'instrumentation, la préparation des grilles, l'acquisition des données, ainsi que les algorithmes et ressources informatiques utilisés pour le traitement des données. Il est maintenant possible par cette technique de résoudre à l'échelle atomique des structures protéiques et notamment de complexes protéiques (Cheng Yifan, 2018; Lyumkis, 2019).

La cryoEM présente également de nombreux avantages par rapport aux autres techniques de biologie structurale, comme la cristallographie. Notamment, elle nécessite l'utilisation de moins de matériel protéique et s'affranchit des étapes limitantes de cristallisation et de traitement des cristaux. Elle permet également d'étudier les protéines sous forme native en solution et, dans une certaine mesure, elle peut être utilisée avec des échantillons hétérogènes, permettant d'obtenir des informations sur la dynamique des protéines. Toutefois, cette technique n'est pas parfaite, la taille de l'objet étudié pouvant notamment être un paramètre limitant (Cheng Yifan, 2018). Dans mon cas, le module ATPase est un petit objet d'environ 120 kDa, ce qui peut être limitant pour cette méthode et m'a obligé à jouer sur de nombreux paramètres, décrits dans le dernier paragraphe de cette section, afin d'obtenir des données de qualité.

4.3.1. Coloration négative

La coloration négative est utilisée avant tout essai de cryo-microscopie électronique. Il s'agit d'une technique de fixation basée sur l'adsorption de l'échantillon sur une grille de carbone, qui est ensuite recouvert par un agent de contraste dense aux électrons, l'acétate d'uranyle. Ce dernier augmente le contraste des images, en se fixant préférentiellement autour des particules adsorbées, l'échantillon apparaissant plus clair que ce qui l'entoure. Cette technique permet de vérifier la qualité de l'échantillon, son homogénéité et sa concentration, et d'évaluer le comportement de l'échantillon sur la grille (morphologie, forme et taille des particules) (Gonen, 2021).

Le protocole utilisé est le suivant : les grilles sont traitées par décharge de plasma (80% d'argon, 20% d'oxygène), avec l'appareil Model 1070 NanoClean (Fischione instruments), pendant 1 min 30. Les grilles sont alors plus hydrophiles, facilitant l'adsorption de l'échantillon sur les grilles. L'échantillon est fixé avec du glutaraldéhyde 1% pendant 5 min, à température ambiante, puis déposé sur grille. Après 1 min d'incubation, l'excédent est retiré à l'aide d'un papier filtre appliqué sur la tranche de la grille. 10 µL de solution d'acétate d'uranyle 2 % sont alors déposés sur la grille et incubés pendant 1 min, l'excédent étant ensuite également retiré à l'aide d'un papier filtre. Les échantillons ainsi colorés sont ensuite analysés directement par microscopie électronique en transmission avec le microscope Tecnai TF-20 (FEI).

4.3.2. Cryo-EM et analyse des particules isolées

4.3.2.1. Préparation de l'échantillon protéique

L'échantillon utilisé pour la préparation des grilles doit être stable et ne pas présenter d'agrégats. C'est pourquoi il est recommandé d'utiliser un échantillon fraîchement purifié et non concentré, provenant de la dernière étape de SEC. L'échantillon est ensuite centrifugé pendant 20 min à 14 000 rpm et à 4 °C.

Les échantillons du module ATPase SMC1ACC-EQ/RAD21C/SMC3CC-EQ/RAD21N/ATP, présenté dans ce manuscrit ont une concentration comprise entre 0,1 - 0,5 mg/ml.

4.3.2.2. Préparation des grilles et congélation

Il existe une multitude de grilles dans le commerce, ainsi que des grilles préparées en interne par la plateforme de biologie structurale à l'IGBMC. Ces grilles diffèrent par le matériel utilisé (cuivre, or, …), par la taille de leurs trous (1.2/1.3, 2/2, …), par la présence d'un support (carbone, or fin, …). Le choix du type de grilles est important et impacte la qualité des échantillons et des données, en influençant notamment la répartition des particules, la formation d'agrégats, ou encore l'épaisseur de glace (Weissenberger et al., 2021). A l'instar de la coloration négative, les grilles sont traitées par une décharge de plasma, afin de créer une surface négative hydrophile. Le tableau 6 récapitule les différents types de grilles testées ainsi que leurs caractéristiques.

Grilles	Caractéristiques	Avantages	Référence
Quantifoil® (Cuivre)	Couche de carbone (épaisseur 10 – 12 nm) Maillage de 200 ou 400 carrés par pouce	- Faible bruit de fond - Beaucoup de grilles différentes disponibles (formes, tailles, espacement des trous, tailles de mailles, …)	(Ermantraut et al., 1998)
C-flat®	Couche de carbone (épaisseur 20 ou 40 nm) Maillage de 200 ou 400 carrés par pouce	- Distribution plus uniforme des particules au niveau des trous - Épaisseur de glace plus uniforme	(Quispe et al., 2007)
UltraAuFoil® (Or)	Une fine feuille d'or (épaisseur de 500 Å) recouvre le support Maillage entre 200 et 400 carrés par pouce	- Meilleure répartition des particules - Amélioration de la stabilité de l'imagerie - Protection de l'échantillon vis à vis des électrons	(Russo and Passmore, 2014)
Graphène (Préparation par la plateforme de biologie structurale de l'IGBMC)	Fine couche de graphène sur un support carbone	- Meilleure répartition des particules - Concentration en protéine plus faible - Amélioration du signal/bruit - Amélioration de la stabilité de l'imagerie	(Pantelic et al., 2011)
Or pégylé (Préparation par la plateforme de biologie structurale de l'IGBMC)	Monocouche hydrophile (PEG) sur un support or	- Décharge de plasma non nécessaire - Meilleure hydrophilie - Renforcement de l'intégrité structurelle du support troué - Diminution de l'affinité des protéines pour le support et ainsi des orientations préférentielles - Amélioration de la stabilité de l'imagerie	(Meyerson et al., 2014)

Tableau 6 : Récapitulatif des grilles utilisées et leurs caractéristiques.

Pour minimiser les dommages induits par le faisceau d'électron et préserver la structure native de l'échantillon, ce dernier doit être incorporé dans une fine couche de glace vitreuse, évitant ainsi la formation de cristaux de glace (Weissenberger et al., 2021). Pour ce faire, la congélation doit être très rapide et se fait grâce au Vitrobot ® Mark IV (FEI). Cet appareil permet de déposer l'échantillon sur la grille dans une enceinte à température et atmosphère contrôlées, d'éliminer l'excès avec du papier filtre (buvardage) et de le plonger rapidement dans un bain d'éthane liquide. Les paramètres de chaque étape peuvent être ajustés par l'opérateur, afin d'obtenir une épaisseur de glace idéale (Figure 34).

Les grilles utilisées pour le module ATPase ont été soumises à une décharge de plasma pendant 2 min (excepté pour les grilles en or pégylé) à l'aide de l'appareil Model 1070 NanoClean (Fischione instrument) avec une puissance de 30 - 35%. 3 µL d'échantillon ont été déposés sur chaque grille dans l'enceinte du Vitrobot® Mark VI sous une atmosphère saturée en eau (humidité 100%) et à une température de 4 ou 6 °C et l'excès de l'échantillon enlevé par buvardage avec différentes forces de pression (4-8) pendant des temps différents (3-5 s). L'échantillon est ensuite immergé dans un bain d'éthane liquide. La grille contenant l'échantillon est ensuite immergée et stockée dans l'azote liquide.

4.3.2.3. Criblage des grilles et collecte des données

Les microscopes à transmission électronique se composent d'une source d'émission d'électron accélérés dans une colonne sous vide et sous haut voltage, appelée canon à électron, de lentilles électromagnétiques, d'un emplacement pour positionner l'échantillon à imager, de lentilles de projection pour agrandir l'image et d'un système d'acquisition d'image. Un canon produit un faisceau d'électrons, qui passe alors par des lentilles condenseur. Cela permet de régler l'angle et la taille d'incidence du faisceau. Le faisceau atteint et traverse ensuite l'échantillon. Une première image est formée par la lentille objectif, est ensuite agrandie par des lentilles intermédiaires et protectrices et est projetée sur le détecteur (Fioramonte et al., 2017).

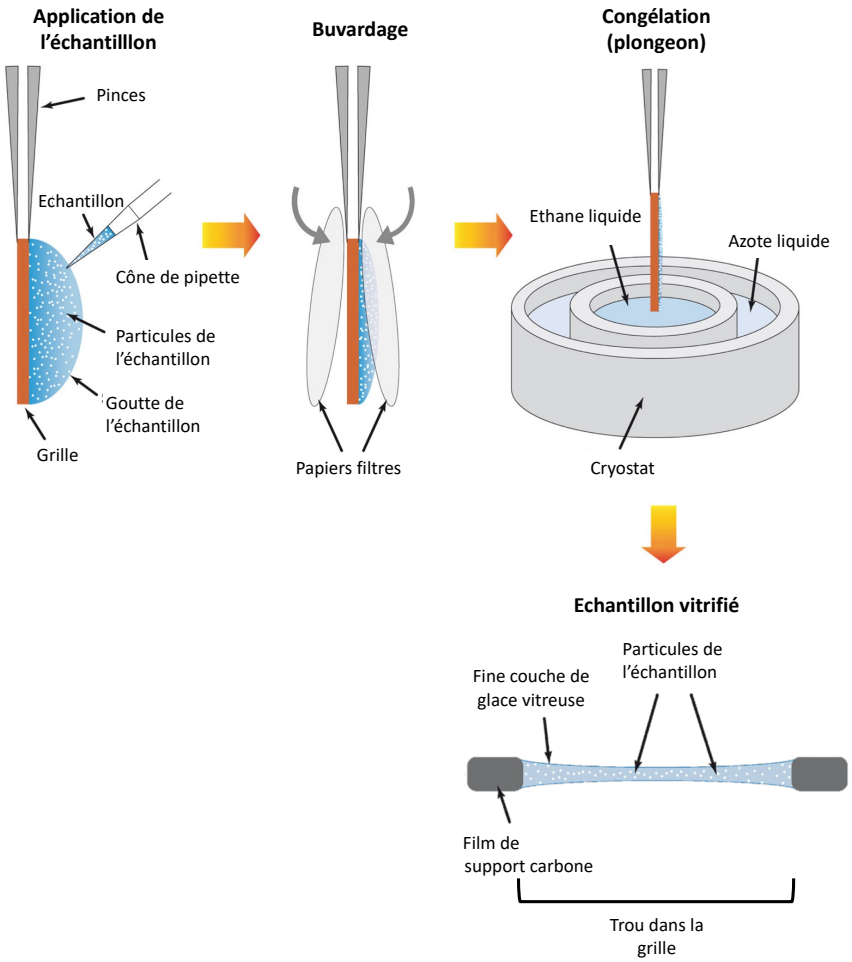


Figure 34 : Préparation et congélation des grilles de microscopie électronique. Adapté de (Sgro and Costa, 2018).

Les progrès techniques ont permis de doter les microscopes d'une meilleure optique, avec notamment un éclairage parallèle ainsi que le développement de correcteurs permettant de corriger l'aberration sphérique des lentilles. Les microscopes les plus récents sont également dotés d'un transfert d'échantillon intégré contrôlé par ordinateur, d'un porte-échantillon, ainsi que de détecteurs d'électrons directs plus performants, qui permettent d'améliorer le rapport signal/bruit des images de cryoEM. Ces avancées, en complément des développements de pointe en informatique pour le traitement de données, ont participé à la révolution de la résolution de la cryoEM en permettant d'atteindre des hautes résolutions impensables il y a encore quelques années (Kuei et al., 2020).

Au cours de ma thèse, j'ai utilisé deux microscopes, qui sont disponibles sur la plateforme de biologie structurale de l'IGBMC : le Titan Krios™ et le cryoTEM Glacios™ (Glacios) (FEI). Ce dernier fonctionne à 200 keV avec une caméra par comptage direct

d'électron K2 Summit™. Quant au Titan Krios™, il fonctionne à 300 keV et possède un correcteur d'aberration sphérique, de détecteurs directs d'électron Falcon 3 et K3 et d'un filtre à énergie BioQuantum. Le criblage des grilles permettant l'évaluation des conditions de préparation des grilles et l'optimisation de vitrification a été réalisé sur le Glacios™.

Les collectes ont été faites sur le Titan Krios™ ou sur le Glacios™. Ce dernier peut présenter un avantage par rapport au Titan Krios ™ pour la collecte de petits objets, tel que le module ATPase, du fait de son voltage à 200 kV, améliorant le rapport signal/bruit (Peet et al., 2019). Les films ont été alors enregistrés respectivement à une accélération de 200 keV ou 300 keV avec la caméra Falcon III ou la K2 en mode comptage et avec une taille de pixel de 0,901 ou 1,4 Å, à un grossissement de 45.000 et une dose totale d'environ 45 e^-. $Å^{-2}$. L'acquisition automatisée des données a été réalisée à l'aide du logiciel SerialEM (Mastronarde, 2005). La plage de défocalisation a été réglée de -0,8 à -2,0. Un total de 4000 – 4500 micrographes par collecte a été obtenu.

4.3.2.4. Traitement et affinement des données

Chaque micrographe contient de quelques particules à quelques centaines de particules. Ces particules sont observées sous différentes orientations, chaque particule correspondant à une représentation bidimensionnelle du complexe étudié. La combinaison de toutes les orientations des particules permet la reconstruction 3D finale. Les différentes étapes du traitement des données sont illustrées dans la figure 35.

La première étape du traitement de données de cryoEM consiste à corriger le mouvement des particules induit lors de la traversée du faisceau d'électrons. Ce phénomène est appelé mouvement induit par le faisceau et a été corrigé à l'aide du logiciel WARP (Tegunov and Cramer, 2019). Cette correction est essentielle pour récupérer le contenu à haute résolution.

La seconde étape est de corriger la fonction de transfert de contraste (CTF). Cette fonction permet de traduire la fonction d'onde de sortie du microscope en une image finale grâce à la transformée de Fourier. Ce transfert est représenté sous la forme d'une onde sinusoïdale. Elle est impactée par l'aberration sphérique du microscope et par la défocalisation de l'image, qui est différente pour chaque micrographe collecté. Cette étape a également été réalisée avec le logiciel WARP (Tegunov and Cramer, 2019).

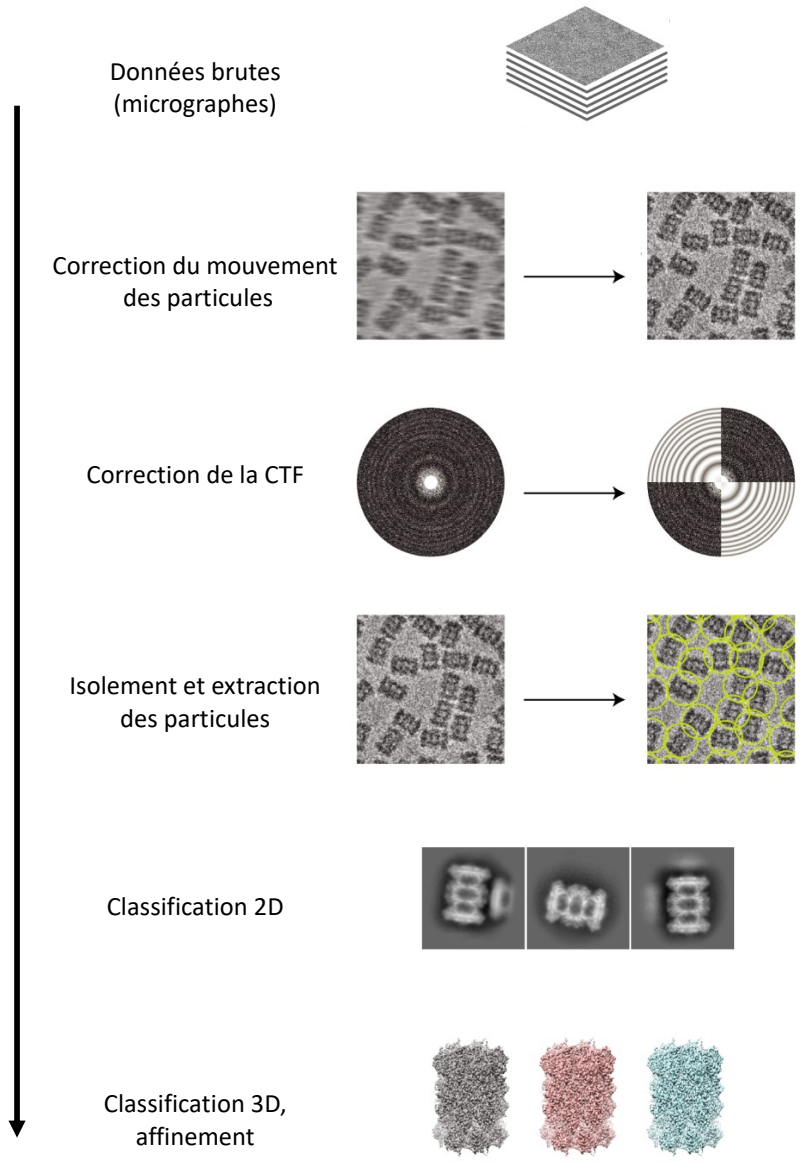


Figure 35 : Les étapes nécessaires à l'analyse de particules isolées en cryoEM.
Adapté de (Tegunov and Cramer, 2019).

Les particules sont ensuite isolées et extraites, selon des boîtes de taille définies en fonction de la taille des particules. Cette étape peut être réalisée de manière semi-automatique ou totalement automatique par les logiciels WARP (Tegunov and Cramer, 2019), crYOLO (Wagner et al., 2019), RELION-3 (Scheres, 2012; Zivanov et al., 2018), ou cryoSPARC (Punjani et al., 2017, 2020).

Les particules extraites sont ensuite classées en groupe, selon leur similarité et leurs orientations dans la glace. En moyenne, les particules sont séparées en 50 ou 100 classes en fonction du nombre de particules isolées initialement. Cette étape permet d'une part de faire un premier tri dans les particules. Celles qui sont dans des classes avec une mauvaise

résolution ou avec un bruit trop conséquent sont rejetées, les autres sont gardées pour la suite du traitement des données. D'autre part, cette étape apporte des informations concernant le comportement des protéines d'intérêt (orientation préférentielle, hétérogénéité, …). La classification 2D des particules, ainsi que le reste du traitement des données, peuvent se faire à l'aide des logiciels RELION-3 (Scheres, 2012; Zivanov et al., 2018) ou cryoSPARC (Punjani et al., 2017, 2020). Toutefois, pendant ma thèse, j'ai majoritairement utilisé le logiciel cryoSPARC, ses algorithmes de calcul s'étant montrés plus performant pour la reconstruction du module ATPase.

Les classes 2D sélectionnées sont ensuite utilisées pour la reconstruction 3D du complexe. Il s'agit d'un processus itératif, où chaque reconstruction sert de modèle de référence pour le cycle suivant. La résolution s'améliore au fur et à mesure des cycles et lorsqu'elle a atteint un plateau, la reconstruction finale est générée.

Des premiers modèles *ab-initio* à basse résolution sont produits. Ils permettent de déterminer l'hétérogénéité des protéines d'intérêt et d'éliminer les particules, qui ne contribuent pas à l'obtention d'une reconstruction 3D de bonne qualité. Ces premiers modèles sont ensuite affinés par affinement non-uniforme. Ce dernier peut améliorer de manière significative la résolution de la carte finale, notamment pour les protéines/complexes de petite taille, par rapport à un affinement homogène. Il tient compte de la variabilité structurale, en conservant tout signal important pour le traitement des données, mais en éliminant le bruit des régions désordonnées, permettant ainsi d'améliorer la qualité et la résolution de la reconstruction 3D finale.

Le jeu de données, qui a permis d'aboutir à la reconstruction du module ATPase, décrit dans la partie I. Etude du module ATPase, acteur clé dans le cycle ATPase de la Cohésine du chapitre 3 - Résultats, a été traité selon le protocole suivant. Les corrections du mouvement induit par le faisceau et de la CTF ont été faites sur les 4146 micrographes collectés et ont été réalisées à l'aide du logiciel WARP (Tegunov and Cramer, 2019) et son réseau neuronal BoxNet2. Ce même programme a aussi permis l'isolement et l'extraction des 399 773 particules. Le reste du traitement de données a été fait avec le logiciel cryoSPARC (Punjani et al., 2017, 2020). La classification 2D de ces particules a permis de sélectionner 204 379 particules contribuant aux classes 2D de meilleure qualité. Ces particules ont été soumises à une reconstruction *ab initio* en deux classes. Cela permet de reconstruire une carte sans avoir

à soumettre une carte de référence au préalable, minimisant ainsi les risques de biais (Punjani et al., 2017). Cette étape est ensuite suivie d'un affinement non-uniforme. Cet affinement est particulièrement adapté pour le traitement de petites molécules, permettant d'obtenir une carte finale de meilleure qualité et de plus haute résolution (Punjani et al., 2020). Cela a conduit à l'obtention de deux cartes pour le module ATPase de la Cohésine humaine : une carte à 4.04 Å (avec 110 099 particules) et une carte à 4.50 Å (94 280 particules).

4.3.2.5. Construction et validation du modèle du module ATPase de la Cohésine

La carte du module ATPase à 4.04 Å a été utilisée pour construire un modèle de ce complexe ATPase. Un premier modèle a été construit avec le logiciel COOT (Emsley et al., 2010), en utilisant les structures de SMC1ACC/RAD21C, SMC1ACCsh/RAD21C et SMC3CC/RAD21N obtenues par cristallographie dans mon laboratoire d'accueil. Ce modèle a ensuite été optimisé avec le programme d'affinement en espace réel de Phenix (Liebschner et al., 2019), en utilisant les paramètres suivants : (i) des contraintes de structures secondaires, (ii) une minimisation globale, qui permet de régulariser la géométrie du modèle, (iii) un morphing et un affinement par recuit simulé (Afonine et al., 2018). Le morphing permet de déformer un modèle afin de le faire correspondre à une carte de densité électronique (Terwilliger et al., 2013). La méthode par recuit stimulé permet des mouvements d'atomes plus importants qu'avec une méthode de minimisation, afin de mieux explorer la surface de la fonction énergétique à la recherche des minima énergétiques les plus bas (Cavarelli, 2009).

Le modèle a ensuite été amélioré par plusieurs cycles de construction manuelle dans le logiciel COOT (Emsley et al., 2010) et d'affinement en espace réel par minimisation énergétique dans Phenix. La validation du modèle a été obtenue avec les programmes Coot et MolProbity. MolProbity permet d'obtenir une validation globale et locale des constantes entre les atomes, ainsi que la géométrie covalente du modèle structurale (Prisant et al., 2020).

4.3.2.6. Optimisations nécessaires à l'amélioration de l'échantillon et à la qualité des données

L'étude structurale du module ATPase par cryoEM a nécessité un travail très conséquent d'optimisation, afin d'améliorer la qualité de l'échantillon et d'obtenir des grilles de qualité

suffisante. La première étape a été de tester différentes concentrations de l'échantillon, pour éviter de saturer les grilles de particules, tout en ayant suffisamment de particules pour la collection de données. Plusieurs optimisations ont également été nécessaires pour affiner l'épaisseur de la glace, (i) en utilisant plusieurs types de grilles (détaillées dans le tableau 6) et (ii) en modifiant les paramètres de la décharge du plasma (temps et puissance) ou du buvardage (temps et pression de buvardage, température et humidité de la chambre).

De plus, les premiers criblages et collectes ont montré que le module ATPase adopte une orientation préférentielle, se caractérisant par une distribution non uniforme des vues de projections des particules et entrainant une baisse globale de la résolution de la reconstruction. L'adsorption des particules à l'interface air/eau lors de la préparation des grilles est l'une des causes principales de ces orientations. Pour éviter ces dernières, d'autres grilles ont été testées (UltraAuFoil®, graphène, or pégylé), ainsi que certains détergents, qui permettent de modifier les propriétés de l'interface air/eau. J'ai notamment utilisé le tween 20 (0,05 %, Sigma), l'octyl-β-glucoside (OG, 0,0025 % - 0,005 %, Sigma), le n-dodecyl-β-D-maltoside (DDM, 0,2%, Sigma).

Après de nombreuses optimisations, j'ai ainsi pu améliorer la qualité de l'échantillon et des grilles, ce qui m'a permis d'obtenir après traitement des données une reconstruction du module ATPase à 4,04 Å.

Chapitre 3 - Résultats

I. Etude du module ATPase, acteur clé dans le cycle ATPase de la Cohésine

Publication en cours de soumission : ¨Specific conformational dynamics of the ATPase head domains and DNA exit gate mediate the Cohesin ATPase cycle.¨

1.1. Introduction & description des travaux

La compréhension des mécanismes moléculaires qui gouvernent les différentes étapes du cycle de la Cohésine humaine, et plus particulièrement celui de son module ATPase, est essentielle afin de mieux appréhender le mode d'action physiologique de la Cohésine, ainsi que sa dérégulation dans les cohésinopathies et les cancers. Cela permettra à plus long terme le développement de stratégies thérapeutiques.

Le module ATPase de la Cohésine se compose des têtes ATPases de SMC1A et SMC3, associées respectivement aux extrémités C- et N-terminales de RAD21. Mon équipe d'accueil s'est intéressée à la caractérisation des différentes étapes du cycle ATPase de la Cohésine humaine, allant de la liaison de l'ATP aux têtes de SMC1A et SMC3 à son hydrolyse, en passant par l'engagement des têtes. Pour ce faire, nous utilisons des approches *in vitro*, telles que des techniques biophysiques, biochimiques et structurales, et *in vivo*, en utilisant comme animal modèle le poisson zèbre. Le projet *in vitro* a été divisé en deux axes, d'une part la caractérisation des têtes ATPases de SMC1A et SMC3 seules, et d'autre part celle de l'engagement des têtes.

Au cours de ma thèse, j'ai participé, d'une part, à l'initiation des étapes du premier projet, en collaboration avec une étudiante en thèse de l'équipe, Marina Vitoria Gomes. Nous avons optimisé le protocole initial de purification, afin d'éliminer les acides nucléiques contaminants et améliorer ainsi l'homogénéité des complexes SMC1A-HD/RAD21C et SMC3-HD/RAD21N. Nous avons également réalisé plusieurs constructions pour les têtes de SMC1A et SMC3, qui présentent des régions en superhélice plus ou moins longues, afin de déterminer si ces régions participent ou non à la stabilisation et à la solubilisation des têtes ATPases de

SMC1A et SMC3. Les différentes constructions réalisées sont décrites dans le tableau 3, chapitre 2 – Matériels et méthodes. J'ai également participé à la mise au point des conditions expérimentales pour les expériences de titration calorimétrique et au développement de tests enzymatiques. Par la suite, la caractérisation des têtes seules a été poursuivie par Marina Vitoria Gomes.

Pour ma part, j'ai concentré mes efforts sur le second projet, celui de l'engagement des têtes ATPases. Il s'agit de mon projet de thèse principal. La reconstitution d'un module ATPase engagé requiert la présence d'ATP, qui entraine, par sa liaison aux têtes de SMC1A et SMC3, l'engagement de ces têtes. Il s'agit d'un complexe transitoire et instable et de ce fait, très difficile à étudier. Lorsque j'ai débuté ma thèse, aucune structure d'un tel complexe n'était disponible.

Pour reconstituer ce complexe engagé, j'ai tout d'abord déterminé les conditions expérimentales optimales, afin d'obtenir un complexe suffisamment stable pour poursuivre sa caractérisation. J'ai notamment testé différents mutants de SMC1A et SMC3 avec une activité d'hydrolyse plus faible, tels que les mutants EQ, LV et DE, qui affectent l'hydrolyse de l'ATP, mais permettent encore sa fixation. J'ai également utilisé différentes concentrations d'ATP, ainsi que des analogues faiblement hydrolysables de l'ATP, comme l'ATP-γ-S et l'AMP.PNP. J'ai également joué sur les conditions de purification (tampons, pH, …) et optimisé les ratios de protéines afin d'obtenir une stœchiométrie idéale. L'ensemble de ces tests m'ont permis de déterminer les conditions optimales pour la formation de ce module engagé et mettent en évidence l'importance de l'utilisation d'ATP et des mutants EQ des têtes ATPases humaines de SMC1A et SMC3. La formation stable de ce complexe engagé a été confirmée par chromatographie d'exclusion de taille et par AUC-SV.

J'ai ensuite initié la caractérisation structurale de ce module ATPase engagé par cristallographie aux rayons X. Toutefois, bien qu'ayant réussi à obtenir des cristaux après plusieurs mois d'optimisation, ces derniers n'ont pas présenté une diffraction suffisante pour la résolution de structure. Je me suis alors tournée vers deux autres techniques structurales : la cryoEM et le SAXS. J'ai mené ces études avec l'aide du personnel de la plateforme de biologie structurale de l'IGBMC. Grâce à la technique de SAXS, j'ai pu obtenir une première enveloppe *ab-initio* révélant une première organisation de ce complexe engagé (données non présentées). Concernant la cryoEM, après de multiples optimisations des grilles, de la

réalisation de plusieurs collectes et de traitements de données, j'ai pu obtenir une carte de ce module ATPase à une résolution de 4.0 Å et construire un modèle de ce complexe engagé.

L'ensemble des structures obtenues par l'équipe nous permettent de reconstituer le cycle ATPase de la Cohésine. En comparant nos résultats avec la structure de l'état clampé de la Cohésine en présence de NIPBL et d'ADN (Shi et al., 2020), nous pouvons ainsi définir les réarrangements structuraux opérés lors des différentes étapes du cycle de la Cohésine. Ces premières étapes comprennent la liaison de l'ATP au niveau des têtes de SMC1A et SMC3, l'engagement de ces dernières, l'interaction de l'ADN et NIPBL avec la Cohésine et enfin l'hydrolyse de l'ATP. Nos résultats mettent en évidence que la liaison de l'ATP aux têtes ATPases les maintient dans un état inactif. Bien que l'engagement des têtes entraine des changements conformationnels, ces derniers ne sont pas suffisants pour activer les têtes, qui sont dans un état semi-actif. C'est la liaison de l'ADN et de NIPBL, qui permettent d'activer pleinement les têtes ATPases et ainsi obtenir une activité ATPase maximale. De plus, nos travaux révèlent une certaine flexibilité de la porte de sortie de l'ADN formée par l'interface RAD21/SMC3, confirmée également par des expériences de réticulation *in vitro*. En complément, nos travaux sur le poisson zèbre mettent en évidence l'importance de cette interface dans les fonctions de la Cohésine.

1.2. Publication

Specific conformational dynamics of the ATPase head domains and DNA exit gate mediate the Cohesin ATPase cycle

Vitoria Gomes, M.[1,2,3,4,§], Landwerlin, P.[1,2,3,4,§], Diebold-Durand, M.-L[1,2,3,4,§], Shaik, T.B.[1,2,3,4], Troesch, E.[1,2,3,4], Weber, C.[1,2,3,5], Durand, A.[1,2,3,4], Brillet, K.[6], Dulac, L.[1,2,3,4,5], Antony, P.[1,2,3,4], Watrin, E.[7], Ennifar, E.[6], Golzio, C.[1,2,3,5] and Romier, C.[1,2,3,4,8,*]

[1] Université de Strasbourg, IGBMC UMR 7104- UMR-S 1258, F-67400 Illkirch, France

[2] CNRS, UMR 7104, F-67400 Illkirch, France

[3] Inserm, UMR-S 1258, F-67400 Illkirch, France

[4] IGBMC, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Department of Integrated Structural Biology, F-67400 Illkirch, France

[5] IGBMC, Institut de Génétique et de Biologie Moléculaire et Cellulaire, Department of Translational Medicine and Neurogenetics, F-67400 Illkirch, France

[6] 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.

[7] CNRS, Université de Rennes, Institut de Génétique et Développement de Rennes, UMR 6290, Rennes, France.

[8] Lead contact. romier@igbmc.fr.

[§]These authors have contributed equally to this work.

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

Summary

The Cohesin complex is key to eukaryotic genomes organization. ATP binding and hydrolysis are essential to Cohesin functions, but the molecular mechanisms underlying Cohesin ATPase activity remain poorly understood. We characterized distinct steps of the human Cohesin ATPase cycle. Non-engaged SMC1A and SMC3 ATPase domains adopt specific inactive conformations caused by a SMC1A relaxed state and a SMC3/RAD21 DNA exit gate resting state. ATP-dependent ATPase domains engagement leads to a semi-productive ATPase module with a flexible DNA exit gate, and formation of the fully productive ATPase module requires a SMC1A/SMC3 interface reorganization through NIPBL and DNA binding. Functional analyses in zebrafish demonstrate the importance of the full DNA exit gate and its dynamic. The distinct states and concerted movements of Cohesin ATPase domains and DNA exit gate thus mediate the Cohesin ATPase cycle and can be targeted by Cohesin regulators.

Keywords: Genome organization; Chromatin; SMC complexes; ATPase activity; DNA exit gate; Active sites asymmetry; Coiled coils; Conformational changes; Loop extrusion; Resting state.

Introduction

Structural Maintenance of Chromosomes (SMC) complexes play key roles in genome organization in all three kingdoms of life. Accumulating evidence shows that SMC complexes use an ATP-dependent loop extrusion mechanism to perform these roles[1-6]. The various prokaryotic and eukaryotic SMC complexes have, however, evolved different functions that are supported by specific sets of core and regulatory proteins[3,6,7]. How shared and evolutionary-divergent mechanisms of SMC complexes contribute to their specific functions remains poorly understood.

Among the eukaryotic SMC complexes, Cohesin acts in the cohesion of sister chromatids, the segregation of chromosomes, in DNA repair, in V(D)J recombination, in transcription regulation, and in the organization of the genome by forming and maintaining chromatin loops and Topological Associated Domains (TADs)[3,6]. 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). SMC1A and SMC3 subunits harbor at each of their extremities a hinge domain and an ABC-like ATPase head domain (HD), both separated by a long intramolecular coiled coil, while RAD21^{Scc1} only contains a few folded domains and conserved motifs that are used for interaction with the other Cohesin core and regulatory subunits[6,8,9].

Formation of the Cohesin core complex involves SMC1A and SMC3 heterodimerization through their hinge domains and the asymmetric bridging by the RAD21^{Scc1} kleisin subunit of SMC1A and SMC3 HDs (Fig. 1A)[10-14]. Specifically, the RAD21^{Scc1} N-terminal region (RAD21N) is formed of a small helical domain (HelD) followed by a long α-helix. This latter helix interacts with the proximal coiled coil (CC) emerging from the globular domain (GD) of the SMC3 ATPase HD. In contrast, the RAD21^{Scc1} C-terminal region (RAD21C) folds into a Winged-Helix Domain that directly binds to the GD of the SMC1A HD, thus closing the tripartite ring characteristic of SMC core complexes[10,12].

The DNA-binding STAG1 or STAG2 (SA^{Scc3}) subunit, which is stably associated with the core Cohesin complex and is often considered as its fourth core subunit, plays an important role in the different Cohesin functions, but its precise mode of action still remains elusive[14-28].

The SMC1A and SMC3 HDs also heterodimerize upon binding two ATP molecules in a composite manner to form the Cohesin ATPase module. This latter interaction, known as engagement, is transient and is lost upon ATP hydrolysis. An asymmetry of the Cohesin active sites in the mechanisms underlying genomic cohesion has been reported but remains poorly understood[15,24,29-32].

Cohesin carries out its various functions using ATP-dependent mechanisms, either stably entrapping DNA in a topological manner, or performing loop extrusion to generate chromatin loops by a mechanism that requires pseudo-topological DNA binding[29,30,33-38]. These two activities participate to

the Cohesin cycle throughout the cell cycle, synergistically contributing to Cohesin functions and being dependent on the its regulators. Notably, stable association of core Cohesin with DNA initially requires its recruitment to the genome by the MAU2^{Scc4}/NIPBL^{Scc2} complex followed by ATP binding and hydrolysis, leading to topological entrapment[23,32,33,39-42].

Stabilization of Cohesin on the genome further requires acetylation of SMC3 by ESCO1 or ESCO2 (ESCO^{Eco1}) on its K105 and K106 lysine residues (human numbering used throughout), a process requiring ATP-dependent engagement and hydrolysis, and the replacement of NIPBL^{Scc2} by PDS5[27,32,43-51]. While acetylation is sufficient for genomic stabilization of Cohesin in yeast, vertebrate Cohesin additionally requires the action of SORORIN[32,48,52-54]. Unloading of Cohesin from the genome also requires PDS5 when complexed with WAPL and is equally engagement-dependent[15,16,20,24,25,43,48,55-58].

The mechanism of topological DNA entrapment occurs through the SMC1A-hinge/SMC3-hinge interface (DNA entry gate) (Fig. 1A)[38,59,60], although the SMC3 HD/RAD21N interface has also been suggested as an alternative entry gate[24,61]. This latter SMC3 HD/RAD21N interface is generally recognized as the DNA exit gate during G1 and during the prophase[15,24,43,56,59,62], although this interface has also been questioned[63]. In contrast, DNA exit from the Cohesin ring during the anaphase occurs through the proteolytic cleavage of RAD21^{Scc1} by Separase[64-66].

NIPBL^{Scc2} also plays a central role in the Cohesin loop extrusion mechanism, notably through the enhancement of the Cohesin ATPase activity, which is at the core of this mechanism[23,34,37,67]. This stimulation is higher in the presence of DNA although DNA alone barely shows any stimulatory activity. The recent structures of *S. cerevisiae*, *S. pombe* and human Cohesins bound to DNA and to NIPBL^{Scc2} C-terminal region have provided unprecedented information on the intimate interactions of NIPBL^{Scc2}, DNA and the Cohesin ATPase module[14,61,68]. Notably, NIPBL^{Scc2} is favoring HDs engagement, interacting with SMC1A and SMC3 GDs and CCs, these three proteins firmly clamping the DNA between themselves. These structures, which have been obtained in the presence of SMC1A and SMC3 mutants having a low ATP hydrolysis rate, have been described as being in a pre-ATP hydrolysis state that we hereafter term productive state[14,61,68].

Formation of chromatin loops by Cohesin is a highly dynamic and regulated process. Notably, the length of the loops depends not only on NIPBL^{Scc2} and ATP binding and hydrolysis, but also on the action of PDS5, WAPL and SMC3 acetylation and, in vertebrates, of the transcription factor/chromatin insulator CTCF[16,20,22,28,55,57,58,69-78].

Cycles of ATP binding and hydrolysis are crucial, as for other SMC complexes, to the required conformational dynamics of Cohesin[14,18,61,62,68,79-90]. Notably, Cohesin can adopt either an open ring or a rod conformation depending on whether the SMC1A and SMC3 coiled coils are dissociated or interact

with each other by co-alignment. In this latter conformation, both SMC1A and SMC3 HDs are in close vicinity, being juxtaposed[81,87].

In contrast, ATP-dependent engagement reorganizes the HDs interface and prevents the SMC1A and SMC3 coiled coils to interact with each other at least in the vicinity of the HDs up to their joint element (Fig. 1A)[14,61,68]. Cohesin coiled coils can also bend at the SMC1A and SMC3 elbow elements, enabling the hinges to come in close vicinity to the HDs[14,61,68,80,87]. How these various conformational movements participate to Cohesin functions is still unclear. A swing-and-clamp mechanism for loop extrusion mediated by the conformational flexibility of Cohesin has however recently been proposed[91]. ATP binding and hydrolysis by Cohesin is therefore key to all Cohesin functions. How this ATPase activity drives the conformational dynamics underlying these functions and how this process is regulated are fundamental questions that require further investigations. Notably, while deciphering the molecular basis of the Cohesin ATPase activity is crucial to understand the structure/function relationships of this complex, our current knowledge on the Cohesin ATPase module remains fragmented. Especially, the precise conformational changes undergone by the SMC1A and SMC3 HDs upon ATP binding, HDs heterodimeric engagement, and ATP hydrolysis are not fully understood.

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, and upon their ATP-dependent engagement. Our study provides the most complete view of the Cohesin ATPase cycle, thus contributing to advance our knowledge of the central mechanisms that underlie Cohesin functions. We notably show that the human SMC1A HD adopts an inactive relaxed conformation and displays a specific conformational dynamic in its different nucleotide binding states. In contrast, human SMC3 is locked in an ATP-hydrolysis inactive state due to a so far uncharacterized resting state adopted by the DNA exit gate. Moreover, a productive conformation of the SMC1A and SMC3 HDs is not fully induced upon HDs engagement but traps the ATPase module in a specific semi-productive state. The fully productive ATPase module formation requires additional structural changes mediated and stabilized by NIPBL^{Scc2} and DNA binding. Our functional analyses in zebrafish further reveal the importance of the full Cohesin DNA exit gate and its dynamic in vertebrates.

Results

Human SMC1A and SMC3 ATPase domains display distinct nucleotide binding properties

To analyze the human Cohesin ATPase module, different constructs of the human SMC1A and SMC3 ATPase head domains (HDs) have been created by varying the length of the proximal coiled coil (CC) emerging from their globular domains (GD). Those included short (SMC1ACCsh), up to the joint

element (SMC1ACC and SMC3CC), and including the joint element (SMC1AJ and SMC3J) constructs (Fig. S1A,B). The RAD21N and RAD21C constructs were designed according to previous work (Fig. S1C)[10,12]. Independent expression of the SMC1A, SMC3, RAD21N and RAD21C constructs led to insoluble proteins. However, co-expression of these proteins yielded soluble SMC1A HD/RAD21C and SMC3 HD/RAD21N complexes. The SMC1ACC/RAD21C and SMC3CC/RAD21N complexes, due to their highest stability, were chosen for our comparative investigations, unless stated.

We first assessed the ATPase activity of the SMC1ACC/RAD21C and SMC3CC/RAD21N complexes, either alone or when mixed together. The independent complexes displayed a very low ATPase activity, but a robust increase in activity was observed upon mixing of the two complexes (Fig. 1B). We next measured the Kd of ADP and ATP for the individual SMC1ACC/RAD21C and SMC3CC/RAD21N complexes using isothermal titration calorimetry (ITC). For ATP, due to the residual ATPase activity of both complexes, Kd measurements were carried out with the wild-type (WT) and with the low ATPase activity EQ mutants (SMC1A-E1157Q and SMC3-E1144Q) as well as with ATP and the low hydrolysable ATPγS analogue.

In the case of ADP, while both complexes displayed a single binding event, the Kd of ADP for SMC1ACC/RAD21C (27.6 µM) was 15 times higher than that for SMC3CC/RAD21N (1.8 µM) (Fig. 1C and S2A-B; Table S1). In contrast, in the presence of ATP or ATPγS, while the SMC1ACC/RAD21C complex showed a single binding event, the SMC3CC/RAD21N complex unambiguously showed two events (Fig. 1D and S2A-B). Since ADP binding led to a single event, we reasoned that the second binding event could be due to a homodimerization at high concentration of the SMC3CC/RAD21N complex in the presence of ATP and ATPγS. We therefore processed the ITC data for SMC3 with a model including two consecutive binding events: an initial ATP-binding event followed by a homodimerization event. This model gave a good fitting of the experimental data and yielded one Kd in the µM range and another in the low nM range for the two events (Fig. S2B; Table S1).

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 also corresponded to the higher measured Kd. Comparative analysis of the results obtained showed that both ATP and ATPγS have Kd in a similar range when binding to SMC1ACC/RAD21C and to SMC3CC/RAD21N (Table S1). Nevertheless, major differences were observed when comparing the ITC-measured thermodynamic signatures of the SMC1ACC/RAD21C and SMC3CC/RAD21N complexes binding to ADP, ATP and ATPγS. While the binding reaction to SMC1ACC/RAD21C was entropy-driven, suggesting an increase in solvent entropy, both binding events of SMC3CC/RAD21N were enthalpy-driven (Fig. S2C-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. Crystallization attempts for both SMC1ACC/RAD21C and SMC1ACCsh/RAD21C complexes yielded crystals in the three states (apo, ADP-bound and ATPγS-bound) diffracting between 1.36 and 2.50 Å and belonging to different space groups (Tables S2 and S3). In contrast, crystallization attempts with the ATP and ATP-hydrolysis mimics ADP-BeF_3, ADP-AlF_x and ADP-VO_4 remained unsuccessful. Structure determination was carried out by molecular replacement using the *S. cerevisiae* Smc1 HD/ATPγS structure (PDB entry 1W1W) as model.

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 (P-loop/Walker A motif) 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 (Fig. 2A). In the loop structures, both conformations were found, which were most likely caused by the low occupancy binding into the nucleotide-binding pocket of an ion, a cleaved phosphate or a bound ADP or ATPγS molecule.

A closed conformation has also been observed in the *Chaetomium thermophilum* Condensin Smc2 HD, which hampers ATP binding[83]. Our different structures and our ITC results show, however, that the closed conformation in SMC1A cannot prevent nucleotide binding but can potentially 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 neighboring residues (Fig. 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 (Fig. S2C). Comparison with the apo structure of *C. thermophilum* Smc1 HD shows that, in this organism, the Smc1 P-loop displays an open conformation[92].

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

In its closed conformation, the human SMC1A P-loop is further stabilized by the formation of a larger hydrogen bond network formed between the carbonyl oxygen of S36, the hydroxyl of S39, and the side chains of K13, S14, N40, D43 and R57 (Fig. 2C). The involvement of this latter residue is unexpected since in the fungi Smc1 apo and ATPγS-bound structures, the R-loop, including its arginine residue equivalent to R57, is not observed in density[12,92]. Specifically, ordering of the R-loop only occurs upon the formation of the productive complex in the presence of $NIPBL^{Scc2}$ and DNA, suggesting that the R-loop helps sensing the formation of this complex and transmits this information to the ATPase site[68].

In contrast to what is observed in fungi, 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 is conserved in these structures. In the ADP-bound and ATPγS-bound SMC1A HD structures, an additional interaction is formed between the N40 side chain and the sugar of these nucleotides (Fig. 2D). Furthermore, the specific structural rearrangements induced upon ATPγS binding bring the R57 side chain closer to the nucleotide, and the R57 guanidino group also binds to the α phosphate of the nucleotide, still maintaining its interactions with the N40 and D43 side chains (Fig. 2D). These interactions are also observed in the human productive complex, where the R-loop displays a similar conformation to that observed in our structures[14].

To further characterize the role of R57 in the human SMC1A HD, we have either mutated this residue into alanine (R57A) or used the Cornelia de Lange Syndrome (CdLS)-characterized 58-62 deletion mutant (Δ58-62)[93], and analyzed the ability of these mutants to hydrolyze ATP. When investigated on their own, the SMC1ACC-R57A/RAD21C and SMC1ACC-Δ58-62/RAD21C mutants showed no ATP hydrolysis (Fig. 2E). However, the loss of enzymatic activity compared to the WT complex was only ~10% when SMC3CC/RAD21N was added to the reaction. Therefore, in human SMC1A, R57 plays an earlier role in organizing the nucleotide binding site prior to HDs engagement, both by participating to the stabilization of the closed P-loop conformation in the apo state and subsequently by stabilizing ATP binding prior to engagement. This contrasts with the case of fungi Smc1, where the R-loop appears only to play a role once the DNA is bound to the engaged ATPase HDs.

SMC1A HD undergoes specific rotational movements upon binding of nucleotides

We next used our different structures of the SMC1ACC/RAD21C complex to analyze the global structural changes occurring upon ADP and ATP binding. 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 (Fig. 3A). Notably, a rotational movement of the region of the RecA-lobe interacting with the adenine moiety of the ADP molecule is observed toward the RecA-lobe/Helical-lobe interface. In turn, the Helical-lobe is pushed and rotates away from the RecA-lobe (Fig. 3A).

Upon ATPγS-Mg binding, the rotational movement of the RecA-lobe region binding to the adenine moiety toward the RecA-lobe/Helical-lobe interface is conserved and even amplified, notably with the P-loop moving toward this interface compared to what is observed in the ADP-bound structure (Fig. 3B-C). However, in contrast to ADP binding, ATPγS-Mg binding causes the Helical-lobe to move toward the RecA-lobe. This is due to the movement of the Q-loop toward the SMC1A nucleotide binding site, which positions the amide group of the Q137 side chain in productive contact to the Mg ion (1.9 Å) and to the γ group of the ATPγS nucleotide (2.6 Å), a feature that is neither observed in our apo nor in

our ADP-bound structures (Fig. 3B-C). These opposite movements of the RecA-lobe and of the Helical-lobe upon ATPγS-Mg binding cause a shrinkage of the SMC1A HD and induce a tightening of the RecA-lobe/Helical-lobe interface. This tightening only induces a few changes in side chain conformations rather than a major reorganization of this interface, and enables the formation of a crystallographic homodimer.

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 essentially as a rigid domain upon ADP and ATPγS binding. Specifically, we observe that the SMC1A CC rotates in opposite directions upon the respective binding of ADP and ATPγS, when compared to a central position of this CC observed in our apo structures, the amplitude of this rotation being larger in the presence of ATPγS (Fig. 3D; Video 1). Importantly, 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 from it upon ADP binding (Fig. 3C,E).

Comparative analysis of the *C. thermophilum* apo Smc1 HD with the *S. cerevisiae* Smc1 HD in its ATPγS-bound form, upon keeping Scc1C fixed for superposition, shows similar rotational movements from the apo form to a shrinked ATPγS-bound form, albeit of smaller amplitude (Fig. S3A-B). This is also induced by the movement of the P-loop and of the Q-loop, including a productive interaction of glutamine Q151 with the Mg ion (2.1 Å) and the γ group of the nucleotide (2.9 Å) (Fig. S3C).

SMC1A HD adopts a relaxed conformation regardless of its nucleotide-binding state

We next compared our SMC1ACC/RAD21C structures with that of the SMC1A HD in the productive complex[14], further keeping RAD21C as reference for superposition. This comparison reveals a major movement of the Helical-lobe with respect to the RecA-lobe upon formation of the productive complex (Fig. 4A). This movement, which has been termed lever effect in the bacterial/archeal case[94], does not cause major structural changes within the RecA- and Helical-lobes themselves, but induces a sliding of the Helical-lobe along the RecA-lobe. The amplitude of this movement is very large, as demonstrated by the displacement by ~7-8 Å of the second signature-coupling helix of SMC1A that precedes the N-terminal coil of this HD (Fig. 4A). This large movement occurring upon leverage induces a significant reorganization of the interface between the RecA-lobe and the Helical-lobe of the SMC1A HD.

This sliding movement is almost perpendicular to the aforementioned planar rotational movement observed upon ADP and ATPγS binding to the SMC1A HD, showing that these conformational changes are not related (Fig. 3E and 4A; Video 1). The leverage also repositions the SMC1A CC in a central position, as observed in our apo structure (Fig. 3D and 4B). Despite this central repositioning of the Helical-lobe, the lever effect keeps Q137 at a productive distance to the ATP binding site[14]. While the Helical-lobe overall structure does not change significantly between both structures, we nevertheless

observe that the signature motif and the preceding F-loop undergo significant conformational and positional changes upon leverage (Fig. 4C).

Importantly, since neither ADP nor ATPγS binding can induce the productive reorganization of the SMC1A RecA-lobe/Helical-lobe interface in the SMC1ACC/RAD21C complex, we conclude that the SMC1A HD adopts on its own a stable relaxed conformation, regardless of its nucleotide-binding state, which is different from that of the productive state. In the fungi case, a similar relaxed conformation is observed that is not significantly modified upon ATPγS binding. Notably, formation of the fungi productive complex also requires a similar lever effect, even if its amplitude is smaller (~4-5Å) (Fig. S3D-E).

Comparison of our SMC1ACC/RAD21C and SMC1ACCsh/RAD21C structures with that of the productive SMC1A HD reveals that the SMC1ACCsh conformation is closer to that of SMC1A HD in the productive state than in the SMC1ACC/RAD21C complex (Fig. 4A-B). Accordingly, only a small lever effect (1-2 Å) is observed between our SMC1ACCsh/RAD21C structures and the productive SMC1A-HD/RAD21C structure. In agreement, these structures display a similar RecA-lobe/Helical-lobe interface and we also observe a central repositioning of the SMC1ACCsh CC compared to the ATPγS-bound SMC1ACC/RAD21C structure (Fig. 4B).

Inspection of the packing in the crystals of the SMC1ACCsh/RAD21C complex, which is almost identical whatever the nucleotide-binding state, shows however that two symmetry-related complexes interact in a head-to-tail manner in the crystals (Fig. S4). Specifically, the artificial linker built to bridge the two coils of one complex inserts itself into the groove formed between the RecA-lobe and Helical-lobe of the second complex. These artificial interactions favor and stabilize the sliding of both lobes with respect to each other.

Despite their artificial character, the analysis of the SMC1ACCsh/RAD21C structures provide significant information on the SMC1A HD dynamic. Indeed, in the ATPγS-bound SMC1ACCsh/RAD21C structure, Q137 is positioned close but at a non-productive distance to the Mg ion (2.8 Å) and to the γ phosphate of the nucleotide (3.8 Å) (Fig. 4D). Therefore, only the full lever effect can bring Q137 in a productive position. In addition, the F-loop and signature motif are not significantly involved in crystal packing, and no ATP-driven homodimerization of this complex occurs in the SMC1ACCsh/RAD21C crystals. Nevertheless, we observe a conformational and positional changes of the F-loop and signature motif in the SMC1ACCsh/RAD21C structures that are similar to those observed in the productive SMC1A HD/RAD21C structure (Fig. 4C). Thus, the structural and conformational changes undergone by these two functional elements are coupled and are intimately linked to the rearrangements occurring at the RecA-lobe/Helical-lobe interface.

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, ATPγS as well as ADP-BeF_3, ADP-AlF_x 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 (Table S4). Crystals were also obtained with ADP-BeF_3 but diffracted only to 4 Å. Structure determination was carried out by molecular replacement using the yeast Smc3-HD/ATPγS structure (PDB entry 4UX3) as a model. Despite different space groups as well as heterodimerization in the presence of ATPγS and ADP-BeF_3, all structures obtained showed a strong conformational similarity in contrast to what had been observed for SMC1A (Fig. 5A; Video 2). In our structures, the SMC3 R-loop is well defined in density and adopts, as previously observed[10,14,61,68], a different conformation than that of the SMC1A R-loop. Specifically, the SMC3 R57 side chain is exposed at the surface of the SMC3 HD RecA-lobe, being thus poised for interaction with an incoming DNA molecule or any other macromolecule. Similarly, the two lysines K105 and K106, which are targets of the acetylase $ESCO^{Eco1}$, are also solvent exposed (Fig. 5A). Analysis of the SMC3 nucleotide binding site showed that the P-loop adopts an open conformation in all our structures (Fig. 5B), which may result in the apo structure from the binding of a sulphate ion stemming from the crystallization buffer.

Most importantly, in all our structures the Q-loop is positioned far away (~9 Å) from the nucleotide binding site and the Q141 side chain is turned away from this binding site (Fig. 5B). Comparison of our structures with the structure of the SMC3 HD in the productive complex[14] shows that formation of the productive complex also induces a lever effect that moves the SMC3 second signature-coupling helix by ~4-5 Å (Fig. 5C). The associated conformational changes strongly affect the positioning of the Q-loop in the productive complex, notably with Q141 which is repositioned toward the ATP-binding site (Fig. 5C-D; Video 2).

Moreover, in contrast to our observations on the SMC1A HD, specific additional changes are observed in the SMC3 HD upon formation of the productive complex that are similar to those described upon bacterial/archeal Smc HD homodimerization and leverage[94]. These notably include the conjoint conformational changes in the wedge element and in the R-loop (Fig. 5D). A repositioning but not a conformational rearrangement of the signature motif is also observed between the two structures.

Comparison of the *S. cerevisiae* published Smc3 structures[10,68] shows that a lever effect (6-7 Å) is also required to form the productive state in *S. cerevisiae* (Fig. S5A). The larger lever effect observed for the *S. cerevisiae* Smc3 HD than that for the human SMC3 HD results from a more relaxed conformation of the yeast HD, which also requires a specific conformational rearrangement of the signature motif loop to reach the productive state (Fig. S5B). In addition, a repositioning of the Q-loop is observed

between the ATPγS-bound and productive structures of the *S. cerevisiae* Smc3 HD, while it already occurs between the apo and ATPγS-bound states of the fungi Smc1 HD (Fig. S3C and S5A).

The human DNA exit gate in the SMC3CC/RAD21N complex adopts a resting state

Surprisingly, a major specific and unexpected structural feature observed in the human SMC3CC/RAD21N complex is a ~45° kink of the CC emerging from the SMC3 GD compared to the straight conformation displayed by the SMC3 CC in the productive complex (Fig. 6A). This kink is due to a simultaneous bending of both α-helices forming the human SMC3 CC that occurs right above the region where these helices leave the SMC3 GD, and just below the region where they interact with RAD21N (Fig. 6A-B; Video 2).

The flexibility of the short helical stretches of the human SMC3 CC that enables their bending is demonstrated by their two slightly different conformations observed in our different structures, including between non-crystallographic symmetry mates in a single crystal form. These small changes also slightly influence the positioning of the wedge element and, in turn, of the R-loop. This kink is observed in all our SMC3CC/RAD21N structures, regardless of the nucleotide-binding state, the dimerization state and the crystal packing.

Although the folding of RAD21N and its interaction with the CC are not changed by the kink, the new path of the CC positions differently the RAD21N small helical domain (HelD) in the groove at the interface between the SMC3 GD RecA-lobe and Helical-lobe (Fig. 6B). The resulting complex thus stabilizes the DNA exit gate on the SMC3 GD, and we have named resting state this specific conformation of the human SMC3 HD/RAD21N DNA exit gate.

In comparison, in the productive state, the RAD21N HelD is positioned above the SMC3 GD, both domains only forming a single contact through the interaction established between the RAD21N R54 side chain and the SMC3 N123 main chain carbonyl (Fig. 6C-D)[14]. In addition, the productive RAD21N HelD interacts extensively with the bound DNA, both through electrostatic interactions and through direct contacts, such as those that can be formed by RAD21N K25 and K26 with the DNA phosphate backbone (Fig. 6D).

In the resting state, RAD21N residues 5 to 26 pack against the SMC3 GD forming hydrophobic and hydrogen bonding interactions. Notably, the first small α-helix of the RAD21N HelD, composed of residues 13 to 23, packs against the SMC3 RecA-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 (Fig. 6B,E).

While residues 13 to 17 of this small helix form main chain/main chain interactions with the SMC3 GD, residues 18 to 23 are involved in main chain/side chain and side chain/side chain interactions with this GD (Fig. 6E). Specifically, tryptophan W18 indole nitrogen makes a pivotal hydrogen bond with the

main chain carbonyl of SMC3 S127. Following this helix, RAD21N K25 and K26 side chains interact respectively with the SMC3 GD D120 and D92 side chains. In addition, the R54 guanidino group makes several interactions with the SMC3 GD, notably with the acidic patch formed by the side chains of S131, S133 and D1135 (Fig. 6E).

Strikingly, the position of RAD21N in the resting state is not compatible with the forward movement of the SMC3 second signature-coupling helix occurring upon leverage (Fig. 6F). Therefore, the resting state keeps SMC3 into an inactive conformation by stabilizing the RAD21N HelD on the SMC3 GD and by hampering the structural rearrangements required to reach the productive state. Importantly, our results show that neither ADP nor ATPγS binding are sufficient to release the inactive state of SMC3 imposed by the resting conformation of the DNA exit gate.

Stable engagement of human SMC1A and SMC3 HDs in the presence of ATP

To understand the molecular mechanisms leading to the release of the SMC1A and SMC3 HDs inactive conformations, we have investigated the conformational dynamics of these HDs upon their ATP-dependent engagement. We first looked at conditions favoring stable HDs interaction by using size exclusion chromatography since weakly formed complexes cannot resist the dilution conditions of this technique. Analytical ultracentrifugation velocity measurements were used to confirm the results obtained.

Initial analyses with the WT and EQ constructs revealed that only the SMC1ACC-EQ/RAD21C and the SMC3CC-EQ/RAD21N mutant complexes can stably associate when ATP is present (Fig. 7A-B and S6A). In contrast, combination of WT/WT, WT/EQ and EQ/WT complexes failed to provide a stably engaged ATPase module in the presence of ATP. The fact that RAD21N was not lost in these latter experiments shows that the DNA exit gate remains shut upon engagement and subsequent ATP hydrolysis. Moreover, injection of the two EQ mutants on the ATP-loaded chromatographic column without preincubation with ATP was sufficient to form the complex, demonstrating of its rapid assembly and stability.

In contrast, both size exclusion chromatography and analytical ultracentrifugation showed that while ATP causes engagement, ADP, ATPγS and AMP-PNP cannot (Fig. S6B). In addition, we did not observe a homodimerization of the independent EQ mutants in the presence of ATP, demonstrating of the specificity of the heterodimeric engagement (Fig. S6C). Several signature motif (SMC1A L1128V and SMC3 L1115V) and D-loop (SMC1A D1163E and SMC3 D1150E) mutants have previously been shown, like the EQ mutants, to have a low ATPase activity and to enable engagement[15,29,31]. However, no combination of these mutants yielded stable engagement (Fig. S6D).

Molecular basis of SMC1A and SMC3 HDs engagement

We used our stable engaged ATPase module for structural analyses. Crystallization attempts of the complex only yielded poorly diffracting crystals and cryo-electron microscopy was used instead. Slight crosslinking conditions caused sample aggregation and the purified samples obtained directly from the size exclusion chromatography were therefore used for cryo-EM data collection. Processing of the collected data yielded two cryo-EM maps at 4.0 and 4.5 Å, both representing unambiguously the engaged complex (Fig. S7 and Table S5). While the first map, based on a majority of the particles, was well resolved and used to model the engaged ATPase module, the second map showed more heterogeneity with specific regions not very well defined. This was particularly the case for the SMC1A HD, suggesting that SMC1A can retain some conformational flexibility upon engagement, reminiscent of the higher SMC1A flexibility observed in our crystallographic analyses.

The model of the engaged human Cohesin ATPase module was obtained upon fitting our high-resolution crystallographic structures into the 4.0 Å resolution cryo-EM map, followed by cycles of manual building and automatic real space refinement (Fig. 7C and Table S5). The structure of the engaged human Cohesin ATPase module shows similarities to that bound to NIPBL^{Scc2} and DNA in the productive complex (Fig. 7D). Notably, upon engagement both SMC1A and SMC3 HDs undergo lever effects that release them from their inactive states (Fig. 7E-F). Engagement also repositions both the SMC3 wedge element and the SMC3 R-loop, as seen in the productive complex, showing that these changes are not caused by NIPBL^{Scc2} and DNA binding. As expected, the SMC1A R-loop retains its conformation as observed in the productive but also in the independent SMC1A structures.

Despite these similarities, the structures of the engaged and productive ATPase module show major differences. Indeed, while both SMC1A and SMC3 GDs, when considered independently, exhibit a strong structural similarity in the engaged and productive complexes (Fig. 7E-F), their complex is tighter in the engaged than in the productive state. Notably, when comparing both complexes by superposing one of their HDs, we observe a displacement of the second HD (Fig. 7D). Refinement of the ATPase module taken from the productive complex against our cryo-EM map also resulted in the same tightening, showing that the starting model for structure refinement is not the cause of the difference observed. In addition, this difference was observed when comparing our structure of the engaged module with those of the human, but also of the two yeasts productive complexes[14,61,68].

This difference in the SMC1A/SMC3 GDs interface is mainly caused by the binding of NIPBL^{Scc2} to the engaged complex in the vicinity of the SMC1A F-loop. Indeed, while the F-loop itself is not significantly affected by NIPBL^{Scc2} binding, the insertion of NIPBL^{Scc2} at this SMC1A/SMC3 interface forces both HDs two move apart (Fig. 7D). This movement of both GDs with respect to each other reaches up to the other side of the complex, causing the observed modification of the interface between the SMC1A and

SMC3 HDs. This modification does not prevent engagement but alters the interactions between the SMC1A and SMC3 GDs. Notably, the different SMC1A and SMC3 motifs that form both composite ATP binding sites are slightly displaced with respect to each other. These changes are not sufficient to prevent ATP binding but are expected to affect the productive character of both sites.

Notably, in the engaged ATPase module, while the SMC3 site appears poised for hydrolysis, with Q141 now turned toward the active site, the SMC1A active site adopts a non-productive conformation, with SMC1A Q137 which is turned toward the SMC3 signature motif rather than toward the ATP γ-phosphate (Fig. 7G-H). In contrast, the glutamines of the SMC1A and SMC3 Q-loops are both turned toward their respective ATP binding site within the yeasts and human productive complexes. Therefore, engagement can occur even if the ATP binding sites are not adopting a full productive organization. Consequently, the ATPase activity at both sites can be regulated, potentially in an independent manner, by slight modifications of the organization of these composite sites.

The other major difference between the engaged and productive ATPase module concerns both SMC1A and SMC3 CCs. In the case of SMC1A, the binding of $NIPBL^{Scc2}$ is responsible for the additional stabilization of 3-5 helical turns within each coil, which corresponds to the interaction region of these coils with $NIPBL^{Scc2}$ hook domain (Fig. 7E). Overall, however, the CC emerging from the SMC1A GD is structurally very similar in both structures, demonstrating of its stability.

In contrast, no density is observed in our cryo-EM map for the DNA exit gate (Fig. 7C). Specifically, the density for the SMC3 CC is lost exactly where this CC is kinked in our crystallographic structures, and no density is observed for RAD21N. This absence of density of the DNA exit gate is already observed at the 2D classification level (Fig. S7) and agrees with the fact that the lever effect is incompatible with the resting state of the DNA exit gate (Fig. 6F). Thus, engagement releases the inactive resting conformation of the DNA exit gate without, however, causing a straightening of the SMC3 CC that would enable the DNA exit gate to adopt a productive conformation (Fig. 7F).

The DNA exit gate becomes flexible upon engagement

To confirm the flexibility of the DNA exit gate observed in our cryo-EM structure, we have performed crosslinking experiments in solution using our SMC1ACC and SMC3CC EQ mutants. Based on our structures and the structure of the productive complex, we identified two pairs of residues in the SMC3 HD and RAD21N (SMC3CC-D120/RAD21N-K25 and SMC3-D92/RAD21N-K26), involving the aforementioned RAD21N K25 and K26 residues (Fig. 6E), that could be crosslinked in the resting state but not in the productive state, when mutated in cysteines. The residues of the first pair were ideally positioned for cross-linking with Bis-maleimidoethane (BMOE) in the resting state (8.5 Å) but not in the productive state (14.0 Å). Those of the second pair were a bit more distant in the resting state (9.5 Å) and very far apart in the productive state (25.0 Å).

Crosslinking of the SMC3CC-D120C/RAD21N-K25C complex gave a robust crosslinking (> 40-50%) at 4°C and room temperature (Fig. 7I and S8A). The amount of crosslink was not significantly modified in the presence of SMC1A, alone or supplemented with either ATP or ATP and DNA. These results therefore confirmed the capacity of the RAD21N to come in close proximity to the SMC3 GD and showed that RAD21N does not dissociate from the SMC3 HD upon ATP-dependent engagement, in agreement with our reconstitution experiments.

The use of the SMC3-D92C/RAD21N-K26C complex also showed a similar crosslinking pattern upon addition of SMC1A. However, additional supplementation of ATP or ATP and DNA caused significant changes in this pattern in a time-dependent manner. This difference is explained by the faster increase of the distance between the SMC3 GD and RAD21N upon movement of the DNA exit gate in the case of this pair. Thus, upon ATP-dependent engagement, the DNA exit gate resting state is released and the gate becomes flexible but neither adopts a stable productive conformation nor opens (Fig. 7J and S8B). This flexibility does not affect the conformation of the SMC3 GD since the position of the SMC3 Helical-lobe is fixed by the HDs engagement and the conformational change of the SMC3 wedge region uncouples the movements of the SMC3 CC to those of the RecA-lobe.

The DNA exit gate and its RAD21N HelD component are functionally important in vertebrates

Our biochemical and structural results suggest a functional role for the human RAD21N HelD. Sequence comparison reveals that this domain is almost fully conserved (sequence identity > 99%) in vertebrates (Fig. S1C), indicating that the role of the RAD21N HelD should be conserved in these organisms. To assess the functional importance of the RAD21N HelD region in vertebrates, we obtained a characterized *rad21a* zebrafish mutant line[95] which we utilized to evaluate the effect of truncating mutations in these regions. Homozygous larvae exhibited large pericardial edema, microcephaly and curved bodies at 3 days post-fertilization (dpf), which mirrored clinical features of CdLS (Fig. 8A and S9)[96]. These phenotypes were specific and could be rescued efficiently upon injection of 200 pg of WT zebrafish full-length *rad21a* mRNA (Fig. 8A-C and S9).

To test the requirement of the RAD21 HelD domain in Cohesin function, and notably of its N-terminal helix that enables the resting state conformation of the DNA exit gate, we injected a mutant mRNA that did not code for the first 21 amino acids of RAD21 (RAD21-Δ21) into homozygous mutant eggs. The RAD21-Δ21 mRNA failed to rescue the RAD21-/- phenotypes compared to the wild-type mRNA (Fig. 8B-C and S9). Specifically, body curvature, eye diameter (as assessment of the microcephaly) and pericardial edema of RAD21-Δ21-injected larvae were affected indistinguishably from non-injected *rad21a* homozygous mutant larvae. The loss of rescue observed with the RAD21-Δ21 mRNA indicated that the absence of the first 21 residues induces a loss of function of *rad21a*, thus confirming the critical role of the RAD21 HelD domain for Cohesin function in vertebrates.

We made a comparative analysis with a *rad21a* mRNA lacking the full RAD21N region (RAD21-Δ87). This mutant, which should prevent the formation of the Cohesin complex by destroying the DNA exit gate, led to an aggravated phenotype rather than a phenotypic rescue as shown by the presence of necrotic tissues in the head of homozygous mutant embryos at 1 dpf. Similarly, injection of the zebrafish RAD21-Δ87 mRNA into wild-type eggs led to the same severe phenotype suggesting that this truncating mutation acts as a dominant or gain of function allele. Accordingly, acridine orange staining of RAD21-Δ87 mRNA-injected embryos revealed a significant increase in the number of apoptotic cells in the head compared to control embryos or embryos injected with WT *rad21a* mRNA (Fig. 8D). This demonstrates the essential function of RAD21N in vertebrates, including its N-terminal HelD domain.

Discussion

SMC1A and SMC3 HDs undergo specific conformational dynamics during the Cohesin ATPase cycle

The Cohesin ATPase cycle is poorly understood despite its key involvement in the Cohesin functions. Notably, the specific roles of the SMC1A and SMC3 head domains and their dynamic prior, during and after their ATP-dependent engagement have remained elusive. Here we provide the most complete molecular view of the human Cohesin ATPase cycle that reveals the specific dynamics of the SMC1A and SMC3 HDs along the ATPase cycle in support of Cohesin functions (Video 3).

The role and importance of the non-engaged SMC1A and SMC3 HDs are poorly characterized. Inter-species and inter-SMC complexes comparisons did not clearly identify specific conformations that could be mechanistically important. We show here that, on their own, both SMC1A and SMC3 HDs adopt specific inactive conformations whose conformational changes can be regulated. Specifically, ATP binding to the SMC1A HD, which is regulated but not prevented by an initial closed conformation of the SMC1A P-loop, does not release this HD from its relaxed inactive state but induces specific conformational movements that prepare it for engagement. In contrast, ATP binding to the SMC3 HD does not modify its stable resting inactive conformation, which is poised for engagement.

In an open ring conformation of the Cohesin complex, when non-engaged, both HDs are likely to adopt such inactive conformations that can serve as specific docking platforms for Cohesin regulators and partners. Specifically, the observed high flexibility of the regions linking the SMC1A and SMC3 HDs to the Cohesin arms[14,61,68,87] is expected to contribute to the conformational independence of these HDs. Whether HDs juxtaposition upon coiled coils co-alignment can alter SMC1A and SMC3 HDs inactive states remains unknown. We were however unable to stabilize their juxtaposed complex in the absence of ATP in our reconstitution experiments, which suggests a loose association of the HDs upon juxtaposition.

Although both HDs can be relieved from their inactive states through specific lever movements of their Helical-lobes with respect to their RecA-lobes, such conformational changes need to be induced and stabilized to prevent the SMC1A and SMC3 HDs to stay or return to their inactive states. ATP-dependent HDs engagement is sufficient to cause such changes, but the resulting engaged ATPase module only adopts a semi-productive conformation. Additional specific conformational changes of both HDs, caused by NIPBL^{Scc2} and DNA binding, are required to bring the ATPase module in its full productive conformation. Following ATP hydrolysis and disengagement, both HDs return to their inactive states. The higher affinity of the SMC3 HD for ADP than for ATP suggests a potential regulatory step for the next round of engagement.

The semi-productive structure of the engaged ATPase module supports an asymmetry of the Cohesin ATPase active sites

The semi-productive conformation of the engaged ATPase module is due to the specific interface of both SMC1A and SMC3 HDs upon engagement that induces slight conformational changes and repositioning, compared to the productive state, of the motifs (WalkerA/P-loop, Walker B, Q-loop, signature motif, D-loop) that contribute to the composite SMC1A and SMC3 ATPase active sites. Although these small modifications are not sufficient to prevent ATP binding and engagement, they alter the active sites, affecting notably the productive character of the SMC1A active site, but less that of the SMC3 active site. Similar observations on the asymmetry of the Cohesin ATPase active sites have been previously reported by several *in vitro* and *in vivo* studies on yeast and human Cohesins[15,24,29-32].

The characterization of the asymmetry of the Cohesin ATPase active sites stems from the use of several mutants, notably in the Walker B (EQ), signature motif (LV), and D-loop motif (DE) of the SMC1A and SMC3 active sites. These mutants have been shown to display a reduced ATPase activity but to retain engagement. These effects can be explained for the EQ mutant that directly affects the ATPase activity due to the mutation of a major catalytic residue that, however, does not significantly contribute to ATP binding and engagement. This is not the case of the LV and DE mutants whose effect is most likely indirect due to the small structural rearrangements they will cause and that will induce a limited reorganization of the ATP binding site they contribute to. This should perturb the productive character of the active site affected without preventing ATP binding and engagement, similarly to what is observed for the engaged ATPase module.

This fundamental difference between the EQ and LV/DE mutants should however cause the latter mutants to retain a higher residual ATPase activity than the former. Indeed, the intrinsic flexibility that we observe of the engaged ATPase module can induce some remodeling of the active sites over time to lead to metastable productive conformations of these sites. This is assessed by our reconstitution experiments, where the LV and DE mutants cannot stabilize the engaged ATPase module, and by the

fact that the LV/DE mutants display milder effects than the EQ mutants *in vivo*[15,29,31]. The same reasoning applies for the requirement of both EQ mutants for the stabilization of the engaged ATPase module in our reconstitution experiments. It has been suggested that, due to this asymmetry, ATP hydrolysis at the SMC3 active site should be sufficient for HDs disengagement[32]. Our structure of the engaged ATPase module supports such a hypothesis, even though a longer residency time of ADP could require further events for disengagement[30].

Molecular basis of Cohesin ATPase activity stimulation by NIPBL^{Scc2} and DNA

The comparative analysis of the structures of the ATPase module in the engaged and productive states reveals the changes induced by NIPBL^{Scc2} and DNA binding to reach the full productive conformation. Those changes go beyond a mere stabilization of the interaction between the SMC1A and SMC3 HDs. Indeed, while engagement causes a lever effect in both HDs, inducing the proper positioning of the SMC3 R-loop for DNA docking and of the SMC1A F-loop for NIPBL^{Scc2} binding, this latter interaction is key in transforming the semi-productive conformation of the engaged ATPase module into its fully productive conformation by reorganizing the interface between the SMC1A and SMC3 GDs.

In addition, NIPBL^{Scc2} binding to SMC1A occurs not only at the F-loop but also at the SMC1A CC, stabilizing additional helical turns of both SMC1A coils upon interaction (Fig. 7E). This double interaction of NIPBL^{Scc2} with the SMC1A F-loop and CC can only occur when the SMC1A Helical-lobe is in a productive conformation since, in the SMC1A relaxed conformation, the F-loop is positioned differently with respect to the CC and adopts a different conformation (Fig. 4C). This implies either that NIPBL^{Scc2} binds to the SMC1A HD when its Helical-lobe is already in a productive conformation, or that NIPBL^{Scc2} causes this lobe to adopt such a productive conformation. Since the productive conformation of the F-loop is linked to the productive interface between the SMC1A RecA-lobe and Helical-lobe, these specific interactions of NIPBL^{Scc2} with the SMC1A Helical-lobe will stabilize the productive conformation of the full SMC1A HD, counteracting the intrinsic flexibility of this HD that we observe in the engaged complex.

The binding of NIPBL^{Scc2} to the engaged ATPase module also contributes to the stabilization of this module, notably through the interaction of the NIPBL^{Scc2} E-loop with both SMC1A and SMC3 GDs (Fig. 7D)[14]. The DNA plays a similar role by docking onto both GDs. These interactions do not cause significant structural changes but certainly favor the full productive conformation of the ATPase module over its semi-productive conformation. In addition, the interaction of the NIPBL^{Scc2} nose domain with the SMC3 joint element further participates to the anchoring of NIPBL^{Scc2} onto the ATPase module but also plays a specific role in stabilizing the DNA exit gate in its productive conformation, which thus also contributes to the stabilization of the rearranged SMC1A GD/SMC3 GD interface. Therefore, NIPBL^{Scc2} and DNA binding to the engaged ATPase module not only favor engagement but

induce and stabilize specific conformational changes that transform the semi-productive engaged ATPase module into a full productive complex (Fig. 9).

The DNA exit gate resting state and engagement-induced flexibility can support the swing-and-clamp mechanism for loop extrusion

While the comparative analysis of the engaged and productive ATPase modules reveals the changes occurring upon NIPBL^{Scc2} binding to this complex, they do not inform on how and when this macromolecule docks onto the ATPase module to form the productive complex. Specifically, this docking can initially occur on any of the non-engaged HDs or on the already engaged ATPase module. Interestingly, a swing-and-clamp mechanism as recently been proposed for the formation of the productive complex during loop extrusion by human Cohesin[91]. This mechanism implies an initial spontaneous swinging of the Cohesin arms, upon their bending at their elbow elements, that brings and docks NIPBL^{Scc2}, the SMC1A and SMC3 hinges and the DNA onto the non-engaged SMC3 HD/RAD21N complex. Subsequent engagement/clamping of the HDs then induces conformational changes that release the hinges, while NIPBL^{Scc2} and DNA are repositioned to form the productive complex.

Our analysis of the human Cohesin ATPase cycle shows that the swing-and-clamp mechanism can build on the DNA exit gate resting state and its subsequent engagement-dependent flexibility. Notably, the stable resting conformation of the non-engaged SMC3 HD/RAD21N complex, which is independent of the SMC3 HD nucleotide binding state, can provide a unique stable docking surface for the incoming NIPBL^{Scc2}/hinges/DNA complex. ATP-dependent HDs engagement can then cause the formation of the semi-productive engaged ATPase module, which then releases the resting conformation of the DNA exit gate. The subsequent movement of the DNA exit gate can then enable NIPBL^{Scc2} to bind productively to the engaged ATPase module, bringing in turn this module in its full productive conformation (Fig. 9).

Considering the interaction between the SMC3 joint element and the NIPBL^{Scc2} nose in the productive complex, the NIPBL^{Scc2} nose is likely to interact with the joint element upon docking onto the resting SMC3 HD/RAD21N complex. This could cause the release of the hinges, which have been shown to interact with the NIPBL^{Scc2} nose during the swinging step[91], for a new round of loop extrusion. In addition, the movement of the DNA exit gate can help the Cohesin to alternate between a strong grip on the DNA molecule, in the productive state, and a looser grip, when the DNA exit gate is still oscillating, enabling the DNA to follow the movement of the exit gate and of NIPBL^{Scc2} (Fig. S10A-D).

Specifically, in the productive complex, the DNA is laying on the highly positively charged electrostatic potential of the pit formed by the engaged SMC1A and SMC3 GDs (Fig. S10A). In this complex, the SMC1A CC shows no strong positively or negatively electrostatic potential and appears mainly to act

as a rigid wall to maintain the DNA in its binding chamber. In contrast, the DNA exit gate is strongly positively charged, and contributes to the binding of the DNA and to its curvature through electrostatic interactions and direct contacts such as those made by the RAD21N HelD K25 and K26 lysine residues (Fig. S10A-B). In contrast, in a non-productive conformation of the DNA exit gate, the RAD21 HelD is displaced, which widens the DNA binding chamber, easing initial DNA movements and binding (Fig. S10C-D).

How the DNA remains bound to the HDs once the ATP is hydrolyzed and the HDs are disengaged is unclear. In the resting state, however, the RAD21N HelD forms a specific and highly positively charged patch together with the SMC3 GD that is different from that observed when the DNA exit gate is in a productive conformation (Fig. S10E-F). This patch could contribute to maintain DNA binding after disengagement. Juxtaposition of the SMC1A HD with the SMC3 HD and STAG^{Scc3} could also participate to this stabilization of the DNA after disengagement.

Alternative docking modes of NIPBL^{Scc2}

Considering the large interface between NIPBL^{Scc2} and the SMC1A HD in the productive complex, it cannot be excluded that these two macromolecules initially interact together to form this complex. Such a binding could occur prior or after HDs engagement. In the former case, NIPBL^{Scc2} binding can favor the productive conformation of the SMC1A HD, potentially facilitating subsequent ATP binding and engagement with SMC3, while in the latter case, the prior formation of the engaged ATPase module provides a full docking platform for NIPBL^{Scc2}. In these two mechanisms, however, the non-productive conformation of the DNA exit gate appears to hamper its interaction with the NIPBL^{Scc2} nose domain in its productive conformation (Fig. S10G-H).

The formation of the productive complex, when NIPBL^{Scc2} initially binds to the SMC1A HD, would therefore require that both the NIPBL^{Scc2} nose domain and the DNA exit gate move conjointly from initial non-productive conformations to be able to interact as observed in the productive complex. These latter movements could be eased by the reported flexibility of NIPBL^{Scc2} and that we observe for the DNA exit gate[14,19,61,68,97]. The bound DNA could favor this movement by attracting electrostatically the DNA exit gate, notably the HelD. Our crosslinking experiments show however that the interaction of the DNA exit gate with the DNA is not sufficient to stabilize the productive conformation of the gate, implying that the conjoint positioning of NIPBL^{Scc2} onto the DNA is required to stably form the productive complex (Video 3).

Alternatively, it cannot be excluded that initial binding of NIPBL^{Scc2} to the SMC1A HD prevents the interaction between the NIPBL^{Scc2} nose domain and the DNA exit gate. This could yield other interactions between NIPBL^{Scc2} and the ATPase module. Specifically, NIPBL^{Scc2} needs to switch between a loop extrusion mechanism, where the DNA is bound non-topologically, and a genomic loading

mechanism, where the DNA is bound topologically. A Cohesin complex where the DNA exit gate remains mobile, as observed in our structure of the engaged ATPase module, could participate to a topological entrapment mechanism.

Potential opening of the DNA exit gate based on the engagement-dependent flexibility of this gate

In agreement, the studies reporting an asymmetry of the Cohesin active sites showed that this difference between the active sites is observed for the NIPBL^{Scc2}-dependent DNA topological entrapment by Cohesin, but also for the ESCO^{Eco1}-mediated SMC3 K105/K106 acetylation and for the PDS5/WAPL-dependent Cohesin genomic unloading[15,24,29-32]. Thus, these different functions, which all require HDs engagement at some stage, could make use of the resting SMC3 HD/RAD21N complex and of the semi-productive ATPase module where the DNA exit gate shows engagement-dependent flexibility (Fig. 9).

Notably, in the case of Cohesin unloading, an engagement-dependent opening of the DNA exit gate is required. Although ATP binding and hydrolysis have been proposed to be sufficient for this opening in the fungi case[92], we do not observe such an opening in our various experiments on the human engaged ATPase module, even with wild-type SMC1A and SMC3 HDs and in conditions where ATP hydrolysis occurs. Interestingly, the region that links the C-terminal extremity of RAD21N to the PDS5 and NIPBL^{Scc2} binding motifs is shorter for the former (~10 residues) than for the latter (~60 residues) (Fig. S1C). This could suggest a potential opening mechanism in the presence of PDS5 and WAPL where the engagement-dependent induced flexibility of the DNA exit gate would exert pulling forces on RAD21N and cause the opening of the DNA exit gate.

The suggested sequestration of the RAD21N HelD by PDS5[48] could further facilitate the exit of the DNA. Interestingly, a role for the RAD21N HelD has also been suggested in the studies reporting DNA topological entrapment through the DNA exit gate[24,61]. In addition, our functional experiments in zebrafish demonstrate of the RAD21N HelD importance in vertebrates. Furthermore, the vicinity of the SMC3 K105 and K106 to the DNA exit gate suggests a modulation of the DNA exit gate dynamics upon recruitment of specific Cohesin regulators by these lysines either acetylated or non-acetylated.

Evolutionary differences of the ATPase module and of the DNA exit gate

The difference in gate opening upon ATP binding/hydrolysis between the fungi and human Cohesin adds to the significant differences highlighted by our study on the HDs and ATPase module from these different organisms. In addition, the RAD21N HelD domain is less conserved at the sequence and structural levels in *S. cerevisiae* Scc1 and several other fungi (Fig. S1C and S11A-D)[10,68], suggesting a specific evolution of this domain in *S. cerevisiae* that could be involved in specific mechanisms.

However, whether the *S. cerevisiae* Smc3 CC displays a conformational flexibility, as that we observed for the human SMC3 CC, and that would be mechanistically important, remains to be investigated.

In the case of the *S. pombe* Cohesin, the Rad21N HelD domain structure and position in the productive complex are similar to those of the human RAD21N HelD (Fig. S11E)[61], suggesting that the *S. pombe* SMC3^{Psm3} HD/Rad21N complex could adopt a resting conformation that would potentially be released by ATP-dependent engagement. In line with this observation, the specific crosslinks observed between the *S. pombe* SMC3 ^{Psm3} HD, Rad21N, NIPBL^{Mis4}, and DNA in the so-called initial binding state compared to the productive/DNA gripping state[61] could correspond to specific interactions occurring upon a docking of NIPBL^{Mis4} and DNA onto a resting SMC3^{Psm3} HD/Rad21N complex, prior to HDs engagement and formation of the DNA gripping complex.

Finaly, the flexibility of the SMC3 CC, which enables the resting state of the DNA exit gate, is observed at the same position in Condensin Smc2, at least in the absence of Brn1N[83,85]. The Brn1N HelD also displays some sequence and structural similarity with that of human RAD21N, including the conservation of a N-terminal helix and of K26 (Brn1 K42; Fig. S11F-G), thus raising the question whether the Condensin DNA exit gate could also adopt a resting state that would be released upon HDs engagement. Specifically, the recent structures of the yeast Condensin ATPase module bound to DNA and to the regulatory subunit Ycs4 show important similarities to the published Cohesin productive complexes[14,61,68,86,88]. Similarly, the reported asymmetry of the Condensin ATPase active sites[98] could also indicate the formation of a semi-productive engaged Condensin ATPase module that would be transformed into a full productive complex upon binding of Ycs4 and DNA.

Conclusions

Collectively, our results reveal the highly dynamic nature of the human Cohesin ATPase cycle that requires specific and concerted conformational movements of the Cohesin ATPase head domains and DNA exit gate. Notably, the Cohesin dynamics cause the HDs to adopt diverse conformations that can modulate the Cohesin ATPase activity and can be the target of the various Cohesin regulators to support the different Cohesin functions (Fig. 9). Understanding how these different conformations and interactions occur throughout the Cohesin cycle will be crucial to further characterize the structure/function relationships of this complex.

Methods

Cloning

The SMC1A and SMC3 ATPase head domain (HD) constructs were generated by PCR using the full length human *smc1a* and *smc3* genes as templates. The sequences coding for the hinges and for the coiled-coils up to the joint elements were replaced by sequences coding for a short protein linker (either ESSKHPASLVPRGS or GSGSLVPRGSGS), as previously reported[10,12]. Point mutations were introduced into the constructions using rolling circle or nested PCR strategies. The constructs were cloned by Gibson assembly into bacterial co-expression vectors[99-102] to code either for native proteins (SMC1A-HD and SMC3-HD constructs) or for proteins followed by a 3C protease cleavage site and a 10xhistidine fusion tag (RAD21N and RAD21C constructs).

Protein complexes production and purification

The same large-scale overproduction and purification methods have been used for all protein complexes, unless stated. The SMC1A or SMC3 HDs-coding plasmids were respectively co-transformed with the RAD21C- or RAD21N-coding plasmids into chemically competent *Escherichia coli* BL21(DE3) cells (Novagen). 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 from 3 L of culture were resuspended with 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, lysate clarification was performed by centrifugation (1 hour at 17,000 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. Removal of nucleic acid contaminants was performed using 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). 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

Measurement of the ATPase activity of wildtype and mutant SMC1A-HD/RAD21C and SMC3-HD/RAD21N complex 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 assessed at 30 °C, by measuring the absorbance at 360 nm every 42 seconds for 2 hours, using a spectrophotometer plate reader (TECAN). The ATPase activities, expressed in Pi molecules released per dimer and per minute shown in the figures were calculated in the linear range of the curves. The experiments have been performed in either duplicates or 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). Each nucleotide was injected into 189 to 330 µM of either the wildtype or the EQ-mutant SMC1ACC/RAD21C or SMC3CC/RAD21N complexes, respectively. ITC measurements were performed at 5 °C using a PEAQ-ITC microcalorimeter (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 then fitted with either one binding event or two binding events models, using the AFFINImeter analysis software.

Cross-linking experiments

For crosslinking experiments, all endogenous cysteines of the SMC1ACC, SMC3CC, RAD21N and RAD21C constructs were mutated into serines to avoid initially observed unspecific crosslinking. The specific cysteines were then introduced in the SMC3CC (D92C and D120C) and RAD21N (K25C or K26C) constructs. All mutated complexes were purified following standard protocol described above. Mutated SMC3CC/RAD21N proteins were mixed to a final concentration of 6 µM alone or in the presence of SMC1ACC/RAD21 (7 µM), ATP (20 µM), DNA (10 µM) and $MgCl_2$ (40 µM). Reactions were incubated for 5 min on ice or at room temperature before addition of 0.5 mM BMOE or DMSO. The reaction was quenched by addition of Laemmli buffer containing β-mercaptoethanol after 2, 5 and 10 minutes, and was further heated at 70°C for 5 min. Samples were loaded on acrylamide gel and band intensity was measured using Fiji[103]. All experiments were performed in triplicates.

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 Tables S2, S3 and S4.

Crystallographic structure determination, model building and refinement

X-ray diffraction data were collected at the SOLEIL and SLS synchrotrons. High resolution diffraction data (ranging from 1.4 to 3.0 Å) were obtained and processed by indexation, integration, and scaling within the XDS program[104]. Data were merged using Aimless from the CCP4 software suite[105]. The various SMC1A HD/RAD21C and SMC3 HD/RAD21N structures were solved by molecular replacement using PhaserMR[106] using respectively the yeast Smc1 HD/Scc1C (PDB entry 1W1W) and Smc3 HD/Scc1N (PDB entry 4UX3) structures as models. The initial models were subsequently iteratively built and refined using the Coot and Phenix refinement programs[107,108]. All refined models were verified with Molprobity[109] and showed good refinement statistics (Tables S2-S4).

Reconstitution of the ATP-engaged ATPase module by size exclusion chromatography

For the characterization of the stably engaged ATPase module by size exclusion chromatography, the WT or EQ independently purified SMC1ACC/RAD21C and SMC3CC/RAD21N complexes were supplemented or not with 0.5/1 mM of the different nucleotides (ADP, ATP, ATPγS, AMP-PNP), diluted to a concentration of 50 µM, and loaded onto a Superdex S200 10/300 column (GE Healthcare) equilibrated with a buffer containing 200 mM NaCl, 10 mM Tris-HCl pH 8.5 mM $MgCl_2$ and 1 mM TCEP, either without nucleotide or supplemented with 0.5/1 mM of the chosen nucleotide. The experiments performed with the WT and EQ, LV and DE mutants were performed initially using the same conditions but incorporating the SMC3J/RAD21N complex instead of the SMC3CC/RAD21N complex.

Reconstitution of the ATP-engaged ATPase module by analytical ultracentrifugation (AUC) experiments

Analytical ultracentrifugation sedimentation velocity experiments were performed with a ProteomeLab® XL (Beckman Coulter) at 4°C and 42,000 rpm with absorbance detection at 280 nm. Independently purified SMC1ACC-EQ/RAD21C and SMC3CC-EQ/RAD21N complexes, stored in a buffer composed of 200 mM NaCl, 10 mM Tris-HCl pH 8.5 mM $MgCl_2$ and 1 mM TCEP, were diluted to a final concentration of 1 mg/mL for experiments with the homodimers or mixed with a 1:0.9 ratio to a final concentration of 1 mg/mL for the heterodimers. ADP, ATP and low-hydrolysable ATP analogs (ATPγS,

AMP-PNP) were added to the diluted samples at a concentration of 0.5/1 mM prior to measurements. The A280 nm scans data were acquired at 9 min intervals for 24h. The SEDNTERP software[110] was used to estimate the partial specific volume of the protein (υ), the density (ρ), and the viscosity (η) of the samples. Data were analyzed in the SEDFIT software[111] using the continuous c(s) distribution analysis. AUC graphs were rendered in the GUSSI software[112].

Cryo-electron microscopy sample preparation and data acquisition

For the preparation of the cryo-electron microscopy (cryo-EM) samples, the independently purified SMC1ACC-EQ/RAD21C and SMC3CC-EQ/RAD21N samples were mixed with a 1:0.9 ratio and loaded onto a Superdex 200 16/600 column (GE Healthcare) equilibrated with the purification buffer (200 mM NaCl, 10 mM Tris-HCl pH 8.2 mM $MgCl_2$ and 1 mM TCEP) supplemented with 1 mM ATP. The engaged complex obtained from the most concentrated chromatographic fraction was diluted to 0.3 mg/mL. A 3 µl aliquot of the sample was applied onto an in-house PEGylated gold 1.2/1.3 grid[113]. The grid was blotted for 4s (blot force 5) and flash-frozen in liquid ethane using Vitrobot Mark IV (FEI) at 4°C and 100% humidity. Micrographs were acquired on a Glacios Cryo-TEM operated at 200kV on a K2 Summit camera in counting mode. Automated data acquisition was carried out using the SerialEM software[114] at a 45,000 magnification in a nanoprobe TEM mode, which yielded a pixel size of 0.901 Å. The defocus range was set from -0.8 to -2.0. 40 movies frames were recorded at a dose rate of 7.4 electrons per $Å^2$ per second to a total dose of 44.42 e/$Å^2$. A total of 4146 micrographs were collected.

Cryo-electron microscopy data processing, model building and refinement

The detailed image processing and its related statistics are summarized in Fig. S7. All micrographs were CTF estimated and motion-corrected using Warp[115]. Particles were picked using the neural network BoxNet2 of Warp, yielding 399,773 particles positions. Subsequently, particles were extracted for processing in cryoSPARC[116] to perform 2D classification. 204,379 particles belonging to the best 2D classes were subjected to *ab initio* reconstruction with 2 classes, followed by a non-uniform refinement using default parameters. This refinement on the two classes obtained led to a 4.04 Å map and a 4.50 Å map of the engaged complex. Only the first map, obtained with a majority of the particles, was of sufficient quality for model building.

Model building was performed using our crystal structures of the SMC1ACC/RAD21C, SMC1ACCsh/RAD21C and SMC3CC/RAD21N for docking into the cryo-EM map using UCSF chimera[117]. After few manual adjustments, real-space refinement of the model in Phenix[118] with secondary structure restraints, global minimization, morphing, and simulated annealing was performed. The model was then further improved by several cycles of manual building/adjustment in Coot[108] and real space refinement in Phenix. Model validation was performed within Coot and with MolProbity[119].

Zebrafish rescue experiments and immunostaining

Zebrafish (*Danio rerio*) were raised and maintained as described[120]. The mutant line rad21a$^{hi2529Tg/+}$ (AB genetic background) was obtained from the International Resource Center for Zebrafish (ZIRC#ZDB-ALT-041006-8). The zebrafish line reproduces normally. The *rad21a* heterozygote mutants exhibit no phenotype. The *rad21a* homozygote mutants died at 5 days post-fertilization (dpf) from severe pericardial edema. For this study, viable heterozygous mutant adults were crossed and obtained larvae were individually scored and genotyped following the procedure previously described[95]. All the experiments were at least duplicated, all phenotypes were scored blind to the genotypes, and all statistical analysis were performed using GraphPad Prism v8.0.2.263 (GraphPad Software, San Diego, CA) and R (R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/).

For wild-type (WT) and mutant rescue experiments, *Danio rerio* WT and truncated *rad21a* messenger RNAs were Sanger sequenced, cloned into pCS2 vector and transcribed using the SP6 mMessage Machine kit (Invitrogen). We injected 200 pg RNA (WT or mutants) into mutant or WT zebrafish embryos at the 1- to 2-cell stage. Eye diameter, body curvature and pericardial edema were scored on injected larvae at 3 dpf fixed in 4% PFA and washed in PBS-Tween 0.1%. Body curvature (Fig. S9A) and eye diameter (Fig. S9B) were measured using Fiji and statistical differences were assessed using ANOVA followed by Turkey's test for post-hoc analysis. For edema scoring, embryos at 3 dpf were classified into three groups: none, mild, and severe based on the size of the edema compared with an age-matched control group from the same clutch (Fig. S9C) and a Fisher's exact test was performed to determine significance.

Acridine orange experiment was performed on embryos at 1dpf. The embryos were dechorionated and incubated at 28°C for 30 min in E3 embryo medium supplemented with 2 µg/mL acridine orange. After extensive washing, embryos were anesthetized with tricaine and imaged as Z-stacks with GFP green light excitation. Cell counting was performed with Fiji using the ITCN plugin. Statistical differences were assessed with a *t*-test. All animal experiments were carried out according to the guidelines of the Ethics Committee of IGBMC and ethical approval was obtained from the French Ministry of Higher Education and Research under the number APAFIS#15025-2018041616344504.

Generation of the graphical supports

The graphics in the figures have been generated with Pymol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC), Coot[108], Chimera[117], GraphPad Prism (GraphPad Software, San Diego, CA) and BioRender.com.

PDB deposition

The structures described in this manuscript have been deposited under the following PDB codes: 8A6K, 8A6Y, 8A72, 8A75, 8A77, 8A79, 8A7M, 8A7W, 8AZY, 8A80, 8A81, 8A87, 8A88, 8A89, 8A8A, 8A8B, 8A8P.

Acknowledgments

This work and the authors of this manuscript have been supported by funds from the Fondation ARC (Association pour la Recherche sur le Cancer; ARCPJA20181208268 (CR), ARCPJA2021060003715 (CR) and DOC20180507150 (PL)), from the FRM (Fondation pour la Recherche Médicale; FDT202106012973 (PL)), from the ANR (Agence Nationale de la Recherche; JCJC ANR-17-CE12-0006 (CG)), and by institutional funds from the Centre National de la Recherche Scientifique (CNRS), the Institut National de la Santé et de la Recherche Médicale (INSERM) and the Université de Strasbourg. This work has been published under the framework of the IGBMC INRT Labex (ANR-10-LABX-030-INRT (MVG)), the IdEx Unistra (ANR-10-IDEX-0002 (MLDD, CR)) and the EUR IMCBio (ANR-17-EURE-0023 (MVG)) of the French Investments for the Future Program. The authors acknowledge the support and the use of resources of the French Infrastructure for Integrated Structural Biology (FRISBI; ANR-10-INBS-0005-01) and Instruct-ERIC. The cryo-electron microscope Glacios was co-financed by the European Union, with the European Regional Development Fund (ERDF), the Strasbourg Eurometropole and FRISBI. The cryo-electron microscope Titan was co-financed by FRISBI, the Alsace Region, the General Council of Bas-Rhin and the Ministry of Higher Education and Research. We thank Catherine Birck, Alastair Mc Ewen, Nils Marechal and Pierre Poussin-Courmontagne of the IGBMC Structural Genomics platform for their help during biochemical, biophysical and structural data collection and analysis. We thank Gabor Papai, Nils Marechal and Albert Weixlbaumer for advice on cryo-EM structure refinement. We thank Piotr Sosnowski and Helgo Schmidt for advice with the ATPase assays. We thank the Zebrafish International Resource Center (ZIRC, Oregon, US) for providing the zebrafish line used in this study, and the IGBMC Zebrafish Facility, in particular Sandrine Geschier, for maintenance and care of the zebrafish line. We thank Marianne Lemée for help with Acridine orange experiments on zebrafish and Gaëlle Hayot for help with statistical analyses. We thank members of the SOLEIL and SLS synchrotrons for the use of their beamline facilities and for help during data collection.

Author contributions

MVG, PL, TBS, ET and MLDD made the clones and purified the proteins. MVG and PL performed the ATPase assays. MLDD and MVG performed the crosslinking experiments. KB and EE collected the isothermal titration calorimetry data. MVG, TBS and CR performed the crystallographic analyses. PL, AD, CR and MLDD performed the cryo-EM analyzes. MLDD, CW, LD and CG performed the zebrafish

analyzes. MVG, PL, MLDD, TBS, AD, EE, CG and CR planned the experiments. All authors analyzed and discussed the data. CR wrote the manuscript and all authors commented on the manuscript.

Competing financial interests

The authors declare no competing interests.

Figure 1

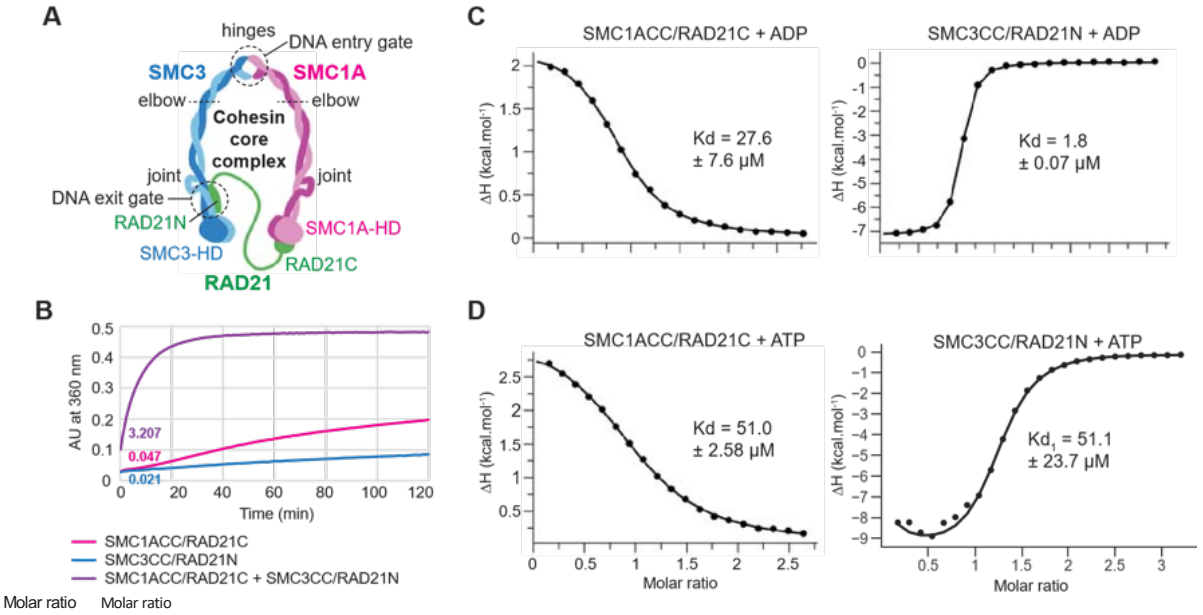


Figure 1. Distinct ATP hydrolysis and ADP/ATP binding properties of the human SMC1A and SMC3 ATPase head domains.

A. Schematic representation of the human Cohesin core complex. Specific elements are labeled.

B. ATPase activity of the independent and mixed SMC1ACC/RAD21C and SMC3CC/RAD21N complexes. Mixing of both complexes leads to a robust increase in ATPase activity, whereas the independent complexes display very little activity. The ATPase activity is given in Pi molecules released per dimer and per minute.

C. Measurements of the Kd of ADP for the SMC1ACC/RAD21C and SMC3CC/RAD21N complexes. SMC3CC/RAD21N shows a higher affinity for ADP than SMC1ACC/RAD21C.

D. Same as in (C) for ATP. In contrast to ADP, both complexes show a similar Kd for ATP.

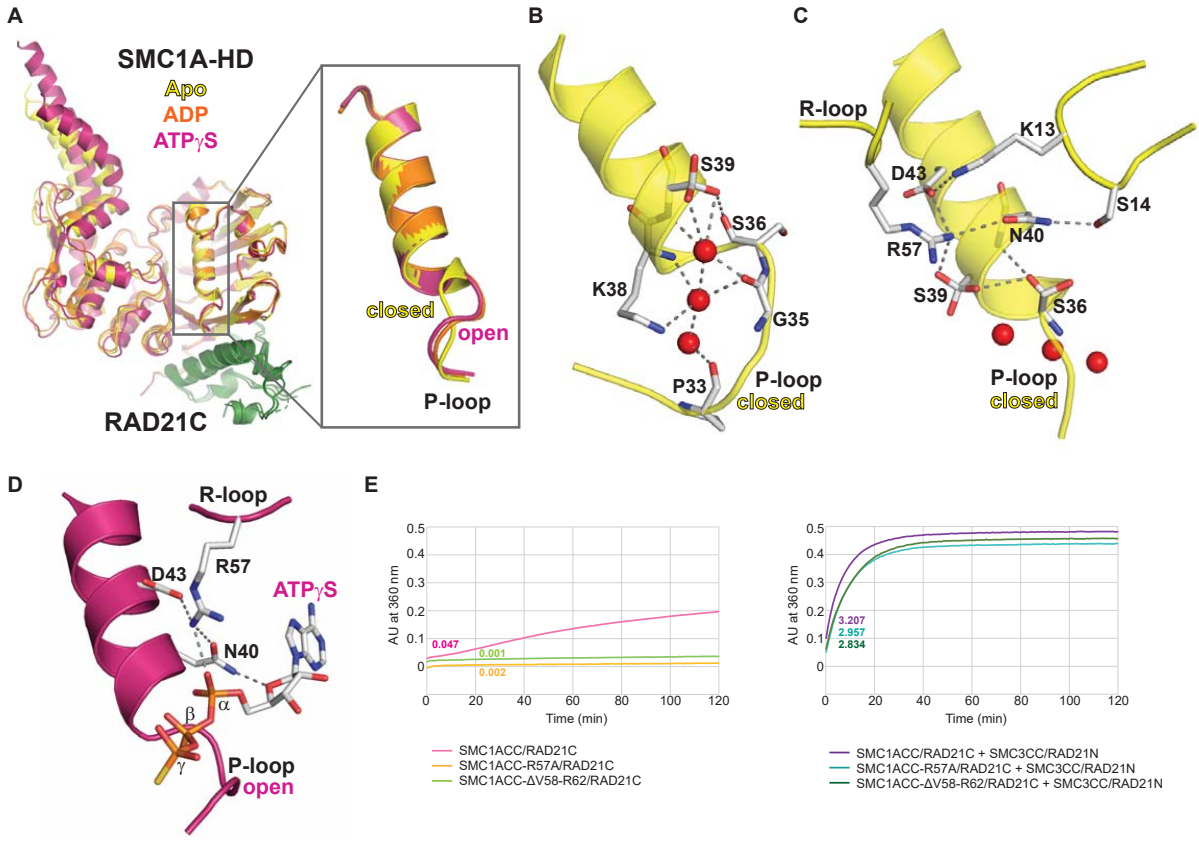


Figure 2. Organization of the SMC1A HD ATP binding site by the P-loop and R-loop.

A. Ribbon representation of the SMC1ACC/RAD21C complex in apo, ADP-bound and ATPγS-bound forms upon superposition of the α-helix following the P-loop. In the apo form, the P-loop adopts a closed conformation, whereas it has an open conformation in the nucleotide-bound forms.

B. Interaction network, including water molecules (red spheres), that stabilize the apo closed conformation of the P-loop.

C. Interaction network that stabilizes R57 from the SMC1A R-loop and enables this residue to participate to the organization of the SMC1A ATP binding site already prior to nucleotide binding.

D. Interaction of SMC1A R57 with a bound ATPγS molecule that enable this residue to stabilize nucleotides in the ATP binding site.

E. ATPase activity of the independent (left panel) and SMC3-bound (right panel) SMC1A R57A and CdLS ΔV58-R62 mutants. While the mutations prevent ATP hydrolysis by the independent SMC1A HD, they only slightly affect the ATPase activity of the engaged complex. The ATPase activity is given in Pi molecules released per dimer and per minute.

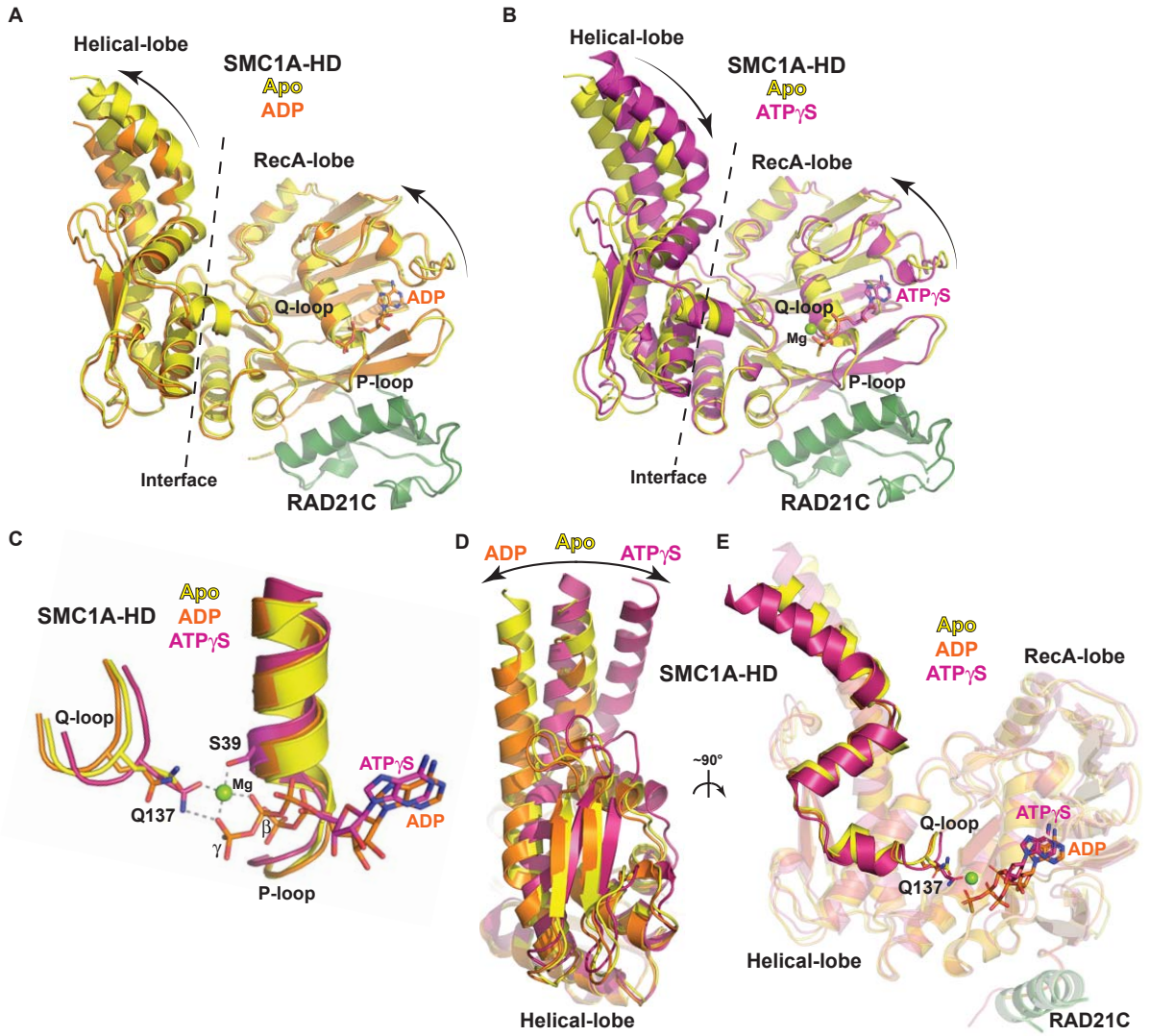


Figure 3. Specific conformational movements of the SMC1A HD upon nucleotide binding.

A. Conformational movements of the SMC1A HD, keeping RAD21C fixed, upon ADP binding. ADP binding causes rotational movements in the same direction for the RecA-lobe and Helical-lobe.

B. Same as in (A) upon ATPγS and Mg binding. In contrast to ADP binding, the RecA-lobe and Helical lobe of the SMC1A HD rotate in opposite directions upon ATPγS binding, causing a shrinkage of the HD.

C. Specific movements of the SMC1A Q-loop and Q137 depending on the nucleotide binding state. Only ATPγS binding and its associated changes bring Q137 in productive distance to the Mg ion and γ group of the ATPγS molecule.

D. Opposite rotational movements of the SMC1A Helical-lobe upon ADP and ATPγS binding.

E. The movements observed in (D) are in a same plane that passes through Q137, explaining the movements of this catalytic residue in the various nucleotide-binding states.

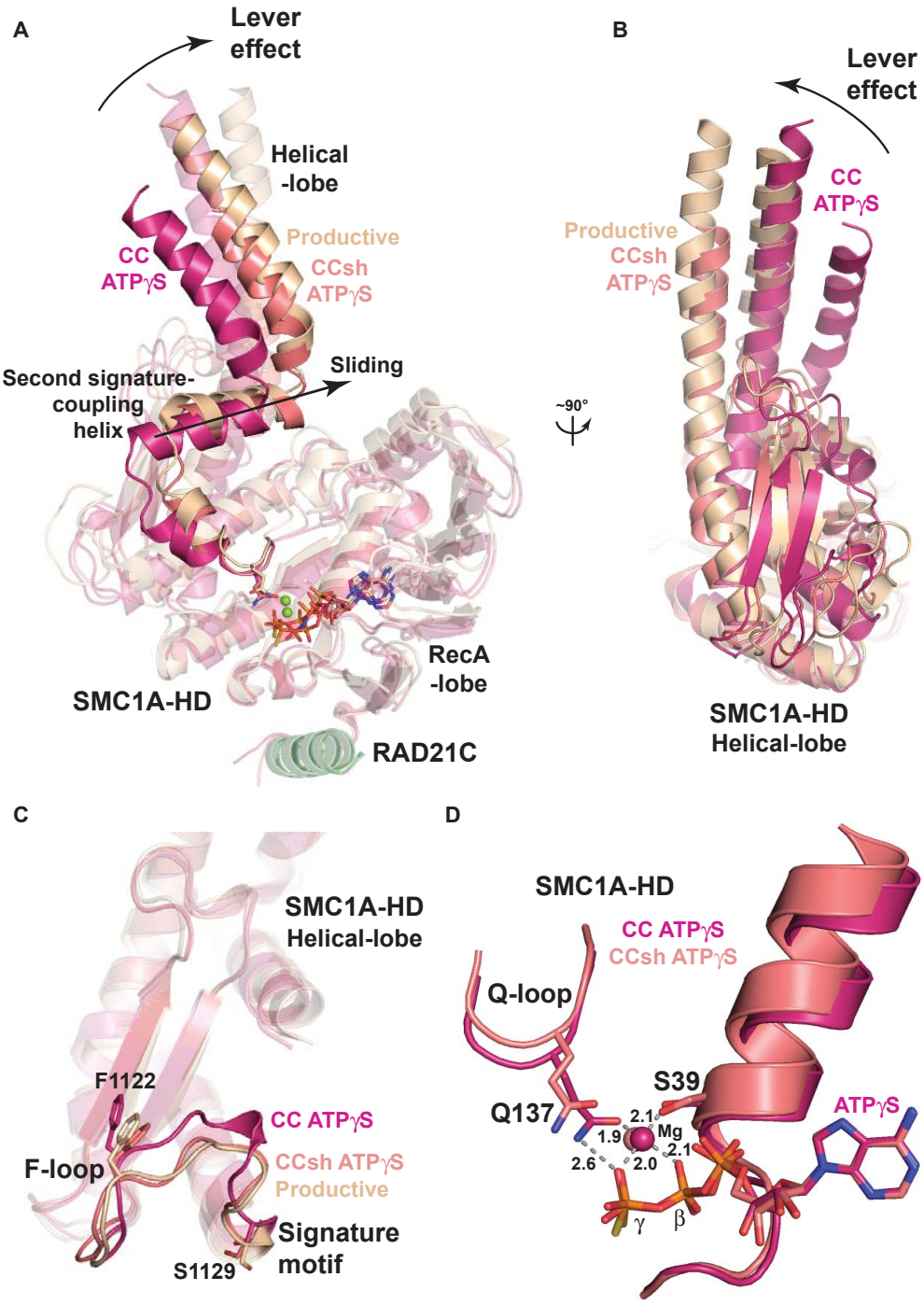


Figure 4. SMC1A adopts a relaxed conformation different from that in the productive complex.

A. Movements of the Helical-lobe with respect to the RecA-lobe of SMC1A upon formation of the productive complex. These movements are caused by a lever effect which induces a displacement of the CC and a sliding of the second signature-coupling helix of SMC1A. These movements are similar in the SMC1ACCsh which undergoes an artificial lever effect. Therefore, the SMC1A HD adopts a relaxed conformation, regardless of its nucleotide-binding state, which is released by engagement.

B. Same as in (A) in a 90° view focusing on the SMC1A Helical-lobe. A repositioning of the SMC1A CC, like in the apo form, is observed upon the lever effect caused by the formation of the productive complex.

C. Conformational and positional changes occurring in the SMC1A F-loop and signature motif upon formation of the productive complex. These are the only significant changes occurring in the helical lobe and are intimately linked to the reorganization of the RecA-Lobe/Helical-lobe interface as shown by the similar changes observed in the SMC1ACCsh structure that do not homodimerize upon crystallization.
D. The Q-loop and Q137 do not adopt a fully productive position in the SMC1ACCsh/RAD21C structures due to the intermediate lever effect artificially induced on this HD.

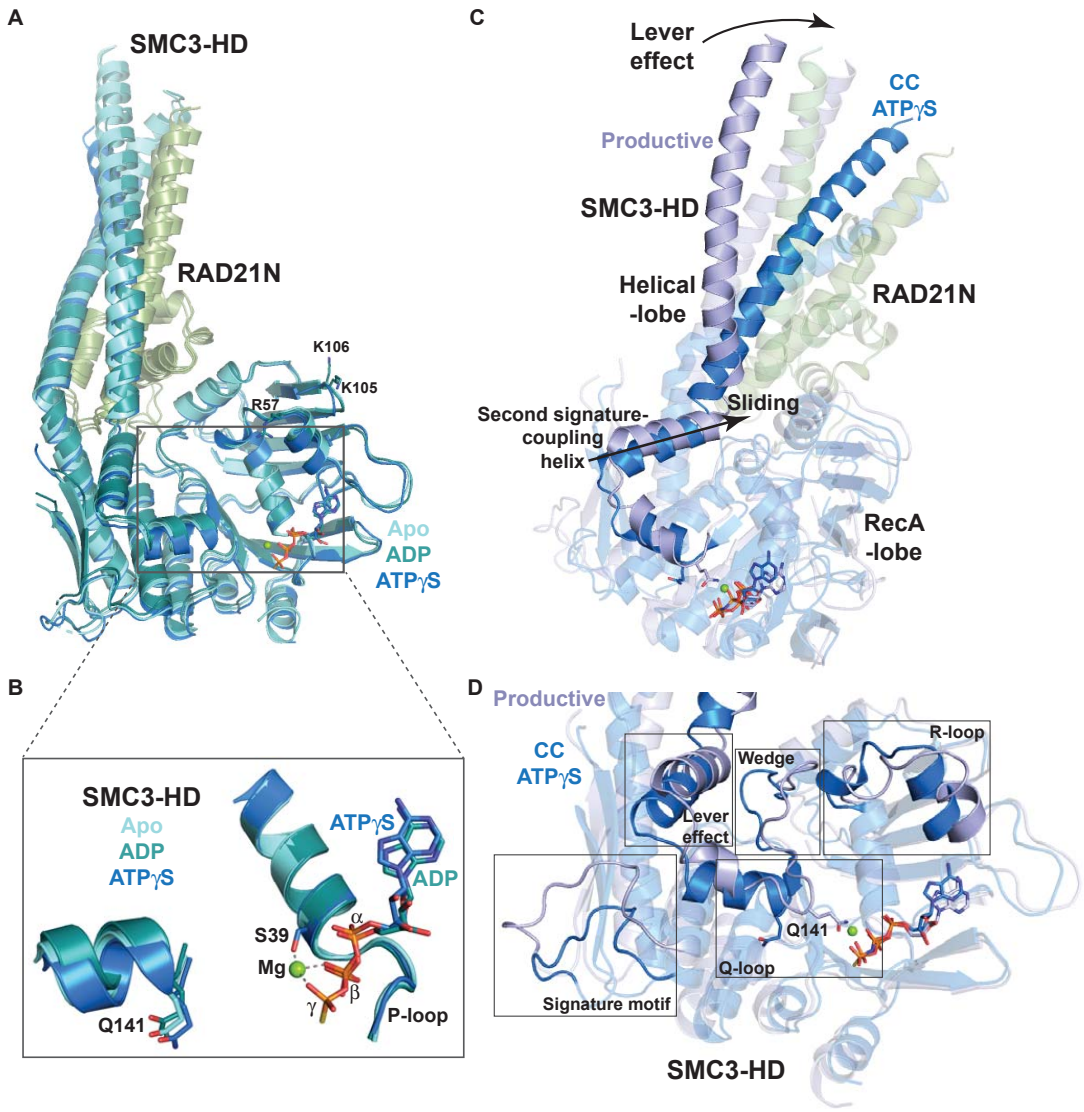


Figure 5. SMC3 adopts a stable inactive conformation regardless of its nucleotide-binding state.

A. Superposition of the SMC3CC/RAD21N structures in various nucleotide-binding states (apo, ADP-bound and ATPγS-bound) showing that, in contrast to SMC1A, the SMC3 HD adopts a stable conformation regardless of its nucleotide-binding state.

B. Close-up on the ATP binding site of the structures shown in A. Q141 remains turned away from the active site showing that the SMC3 HD adopts an inactive conformation whatever its nucleotide-binding state.

C. Conformational changes of the SMC3 HD upon formation of the productive complex, including, like for SMC1A, a lever effect that slides the second signature-coupling helix of this HD.

D. Multiple specific changes in the SMC3 HD induced by the formation of the productive complex.

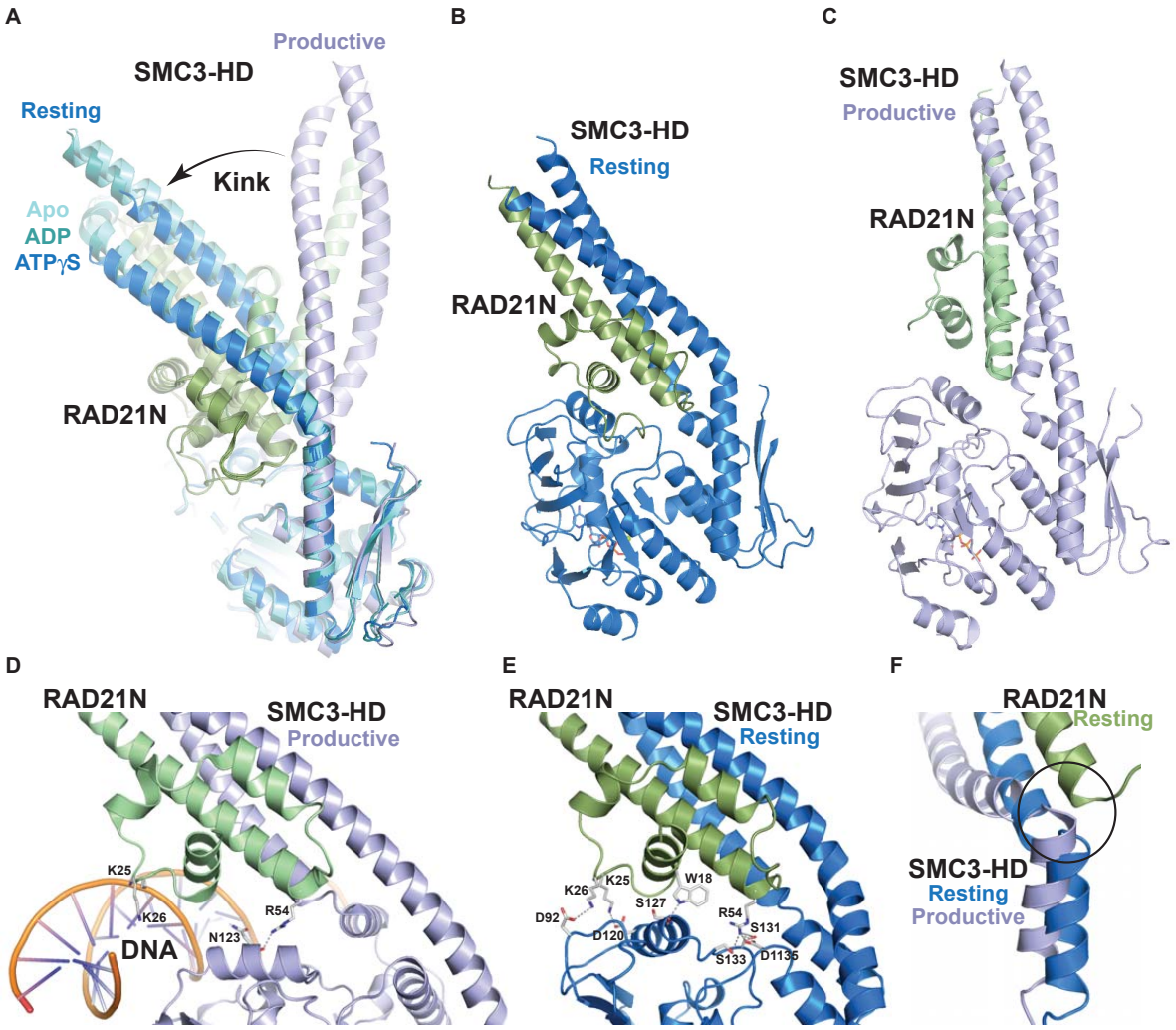


Figure 6. The Cohesin DNA exit gate adopts a stable resting state when the SMC3 HD is non-engaged.

A. Comparison of the independent SMC3 HD structures with that of this HD in the productive complex. A major kink of the SMC3 CC is observed in the non-engaged structures that repositions the RAD21N HelD without perturbing the structure and integrity of the DNA exit gate. We have termed resting state this novel conformation of the DNA exit gate.

B. Structure of the SMC3 HD/RAD21N complex in its resting conformation showing the positioning of the RAD21N HelD domain at the interface between the RecA-lobe and Helical-lobe of the SMC3 HD.

C. Structure of the SMC3 HD/RAD21N complex in its productive conformation showing that the DNA exit gate is further away and hardly interacts with the SMC3 GD.

D. Interactions made by the RAD21N HelD domain with the SMC3 GD and notably the DNA in the productive complex.

E. Interactions made by the RAD21N HelD domain with the SMC3 GD in the resting state.

F. Steric clashes occurring (circle) between the SMC3 second signature-coupling helix in the productive conformation and the RAD21N long α-helix paralleling the SMC3 CC in the resting state. The resting conformation of the DNA exit gate is therefore incompatible with the lever effect in the SMC3 HD.

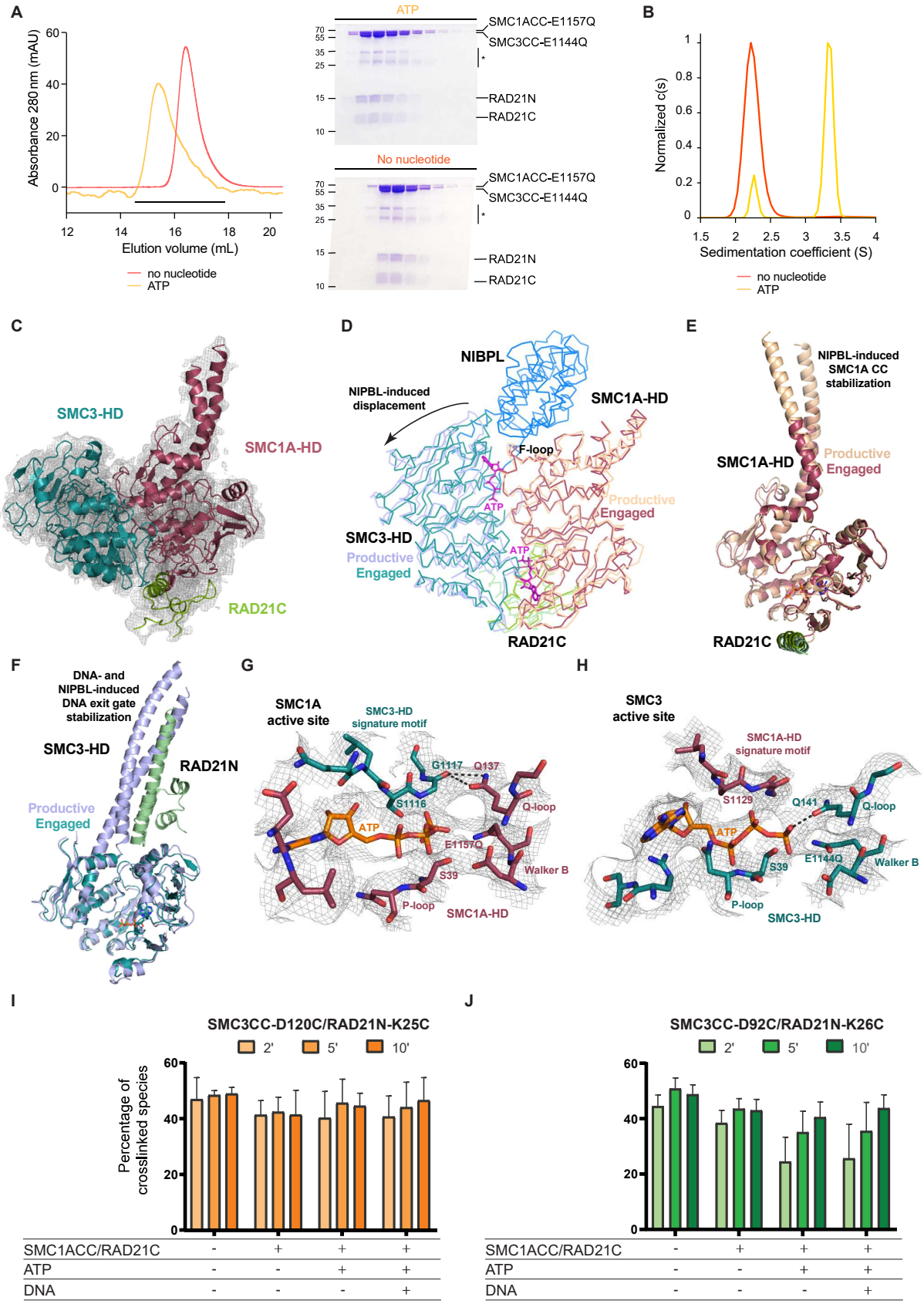


Figure 7. ATP-dependent engagement of the SMC1A-EQ and SMC3-EQ HDs yields a semi-productive engaged ATPase module with a flexible DNA exit gate.

A. Size exclusion chromatographic profiles and associated SDS-PAGE analyses of the mixed SMC1ACC-EQ/RAD21C and SMC3CC-EQ/RAD21N mutants in the absence (red) and the presence (orange) of ATP. In the presence of ATP, a shift of the peak toward the high molecular weight fractions demonstrates of the stable ATP-dependent engagement of the mutants to form the engaged ATPase module. The fractions analyzed by SDS-PAGE are indicated with a black bar in the chromatographic profiles.
B. Analytical ultracentrifugation velocity experiments confirming that the mutants used in (A) can form the engaged ATPase module in the presence of ATP.
C. Structural model of the human engaged ATPase module displayed within the 4.0 Å resolution cryo-EM map. The map unambiguously shows the absence of a conformationally stable DNA exit gate, implying that, upon engagement, the DNA exit gate is released from its resting state and becomes flexible.
D. Superposition of the engaged and productive ATPase modules using the SMC1A HD as reference for superposition. A displacement of the SMC3 HD with respect to the SMC1A HD is observed between both structures that is due to the binding of NIPBL (blue) to the SMC1A F-loop at the SMC1A/SMC3 interface. For clarity, only the region of NIPBL binding to the F-loop and the NIPBL E-loop are displayed.
E. Superposition of the SMC1A HD in the engaged and productive complexes. Both structures display a very high similarity, showing that engagement already induces most of the conformational changes observed for this HD in the productive complex. The major difference concerns the stabilization of the region of each coil that interacts with the C-terminal part of the hook domain of NIPBL.
F. Superposition of the SMC3 HD in the engaged and productive complexes. Both structures also display a very high similarity, showing that engagement also induces most conformational changes observed in the productive complex. The DNA exit gate composed of the SMC3 CC and RAD21N is however not seen in the engaged ATPase module due to its inherent flexibility and is only stabilized in its productive conformation by the binding of NIPBL and DNA. For simplicity, the SMC3 joint element is not shown.
G. Representation of the motifs forming the SMC1A composite ATPase active site and of the bound ATP molecule embedded in the 4.0 Å cryo-EM map contoured at 5 σ. The active site is in a non-productive mode, notably with SMC1A Q137 which is turned toward and interacts with the SMC3 signature motif.
H. Same as in (G) for the SMC3 composite ATPase active site where the positioning of the different motifs and of Q141 are compatible with productivity.
I. Quantification of crosslinked species for the SMC3CC-D120C/RAD21N-K25C pair in experiments performed at room temperature. The supplementation of the SMC1ACC/RAD21C, ATP and DNA is indicated underneath the graph. Crosslinking of the SMC3 GD and RAD21N is retained even upon ATP-dependent engagement and DNA binding, showing that the exit gate remains shut. All experiments were done in triplicates.
J. Same as in (I) for the SMC3CC-D192C/RAD21N-K26C pair. The longer distance between the SMC3 GD and RAD21N in an extended state of the SMC3 CC enables the observation of the DNA exit gate resting state release and dynamic upon engagement. DNA binding does not alter significantly this dynamic.

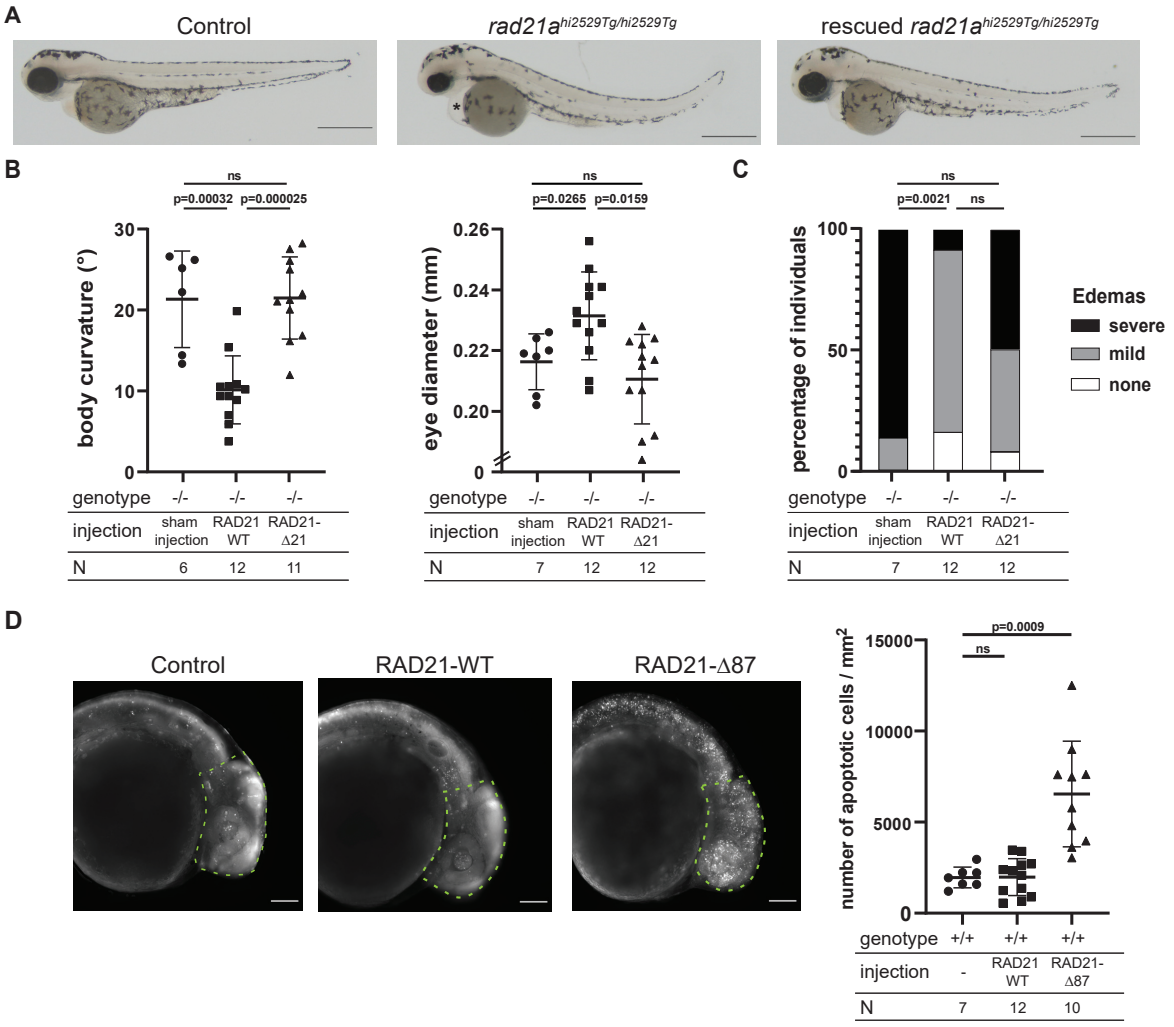


Figure 8. Functional importance of the RAD21N HelD domain and of the DNA exit gate in the zebrafish.

A. From left to right, representative lateral images of zebrafish rad21$^{+/+}$ (wildtype; WT), homozygous mutant rad21$^{hi2529Tg/hi2529Tg}$, and homozygous mutant rad21$^{hi2529Tg/hi2529Tg}$ larvae rescued with 200 pg of mRNA encoding WT full-length *Danio rerio rad21* at 3 days post-fertilization (dpf). Scale bar, 0.5 mm. * indicates the presence of a pericardial edema.

B. Dot plots measurements of body curvature (°) and eye diameter (mm) for homozygous mutant rad21$^{hi2529Tg/hi2529Tg}$ larvae injected with sham or full-length WT *Danio rerio rad21a* (RAD21 WT) or with the *rad21a* mRNA lacking the first 21 residues (RAD21-Δ21). Statistical significance was assessed by ANOVA followed by a Tukey's test for post-hoc analysis. All experiments have been performed in biological duplicates. The p-values are indicated on the graphs. N corresponds to the number of embryos per condition.

C. Bar graph showing the presence of pericardial edema for homozygous mutant rad21$^{hi2529Tg/hi2529Tg}$ larvae injected with sham or full-length WT *Danio rerio rad21a* (RAD21 WT) or with the *rad21a* mRNA lacking the first 21 residues (RAD21-Δ21). Larvae were binned into three categories: severe, mild or absent (none). Statistical significance was assessed by a Fisher's exact test. All experiments have been performed in biological duplicates. The p-values are indicated on the graphs. N corresponds to the number of embryos per condition.

D. From left to right, representative lateral images of zebrafish rad21$^{+/+}$ (wildtype; WT) larvae and WT larvae injected with full length WT *Danio rerio rad21* (RAD21 WT) or with the *rad21* mRNA lacking the first 87 residues (RAD21-Δ87) at 1 day post-fertilization and stained with acridine orange to visualize apoptotic cells. The number of apoptotic cells were counted in the highlighted brain area (dotted line on the images). The dot plot represents the number of apoptotic cells/mm². A *T*-test was conducted between pairs of conditions to determine significance. The p-values are indicated on the graphs. Acridine orange staining has been performed two times. N corresponds to the number of embryos per condition. Scale bar, 0.1 mm.

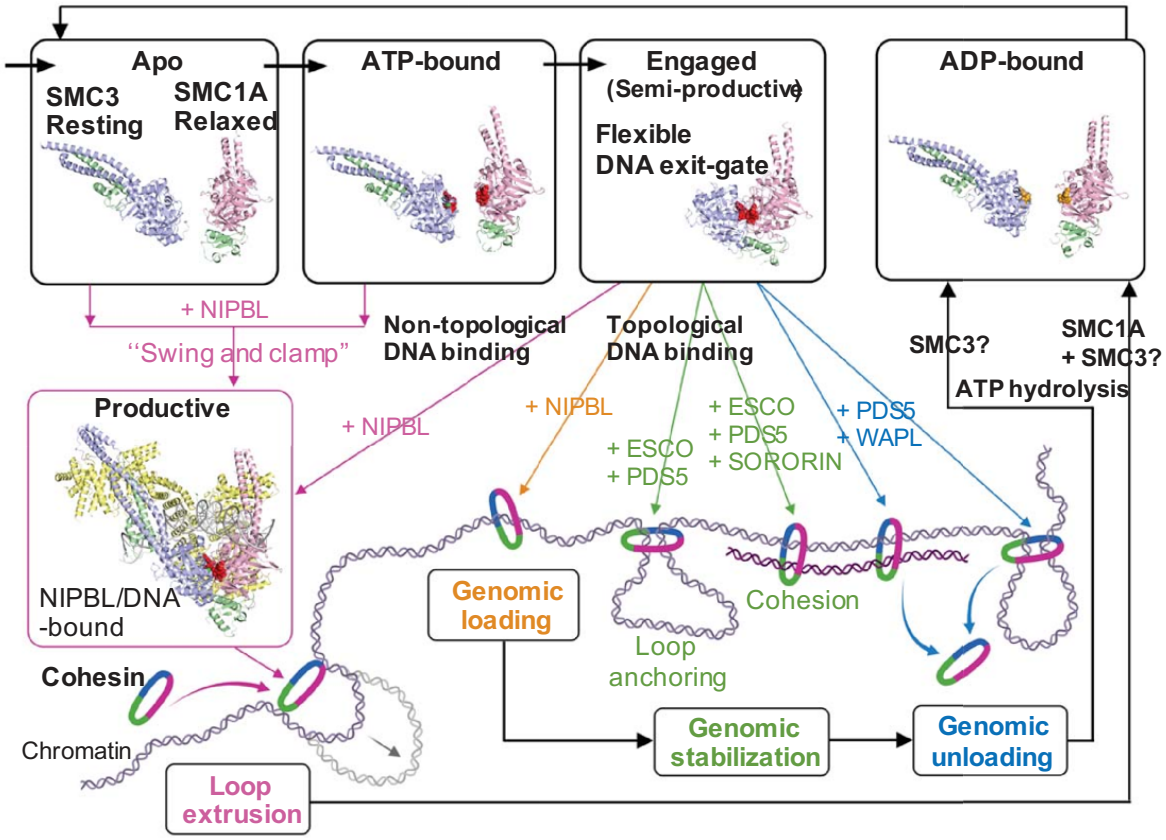


Figure 9. Model of ATP-dependent mechanistic pathways underlying the Cohesin functions

Model of the implication of the Cohesin ATPase cycle in the mechanisms supporting the Cohesin functions. The SMC1A and SMC3 HDs are in inactive relaxed and resting conformations prior and after their ATP-dependent engagement. Engagement induces major structural changes in both HDs but leads to a semi-productive conformation of the ATPase module. This semi-productive conformation, which is due to an asymmetry of the SMC1A and SMC3 active sites, could be used by regulators to drive the genomic loading, stabilization and unloading steps of Cohesin. In contrast, this semi-productive conformation is modified by the binding of $NIPBL^{Scc2}$ and DNA to lead to the productive complex involved in loop extrusion. Dissociation of the ATPase module results from ATP hydrolysis at one or both active sites, depending on the conformation of this module. All these mechanisms act synergistically to support the various Cohesin functions.

Video 1. Conformational changes of the SMC1A HD during the ATPase cycle.

Conformational changes undergone by the SMC1A HD upon switching from its apo to its ATP-bound, productive, and ADP-bound states.

Video 2. Conformational changes of the SMC3 HD during the ATPase cycle.

Conformational changes undergone by the SMC3 HD upon switching from its apo to its ATP-bound, productive, and ADP-bound states.

Video 3. Model of the Cohesin ATPase cycle.

Model for the formation of the productive ATPase module of human Cohesin upon docking of NIPBL and DNA onto the semi-productive engaged ATPase module, including a conjoint movement of the NIPBL nose and DNA exit gate. Alternative mechanisms for the formation of the productive complex and the implication of the ATPase module in topological DNA loading and unloading are provided in the discussion.

References

1. Cutts, E.E. & Vannini, A. Condensin complexes: understanding loop extrusion one conformational change at a time. *Biochem Soc Trans* **48**, 2089-2100 (2020).
2. 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).
3. Davidson, I.F. & Peters, J.M. Genome folding through loop extrusion by SMC complexes. *Nat Rev Mol Cell Biol* **22**, 445-464 (2021).
4. Higashi, T.L. & Uhlmann, F. SMC complexes: Lifting the lid on loop extrusion. *Curr Opin Cell Biol* **74**, 13-22 (2022).
5. van Ruiten, M.S. & Rowland, B.D. On the choreography of genome folding: A grand pas de deux of cohesin and CTCF. *Curr Opin Cell Biol* **70**, 84-90 (2021).
6. Yatskevich, S., Rhodes, J. & Nasmyth, K. Organization of Chromosomal DNA by SMC Complexes. *Annu Rev Genet* **53**, 445-482 (2019).
7. Haering, C.H. & Gruber, S. Snapshot: SMC protein complexes Part I. *Cell* **164**, 326 (2016).
8. Gligoris, T. & Lowe, J. Structural Insights into Ring Formation of Cohesin and Related Smc Complexes. *Trends Cell Biol* **26**, 680-693 (2016).
9. Hopfner, K.P. & Tainer, J.A. Rad50/SMC proteins and ABC transporters: unifying concepts from high-resolution structures. *Curr Opin Struct Biol* **13**, 249-55 (2003).
10. Gligoris, T.G. et al. Closing the cohesin ring: structure and function of its Smc3-kleisin interface. *Science* **346**, 963-7 (2014).
11. 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).
12. Haering, C.H. et al. Structure and stability of cohesin's Smc1-kleisin interaction. *Mol Cell* **15**, 951-64 (2004).
13. Kurze, A. et al. A positively charged channel within the Smc1/Smc3 hinge required for sister chromatid cohesion. *EMBO J* **30**, 364-78 (2011).
14. Shi, Z., Gao, H., Bai, X.C. & Yu, H. Cryo-EM structure of the human cohesin-NIPBL-DNA complex. *Science* **368**, 1454-1459 (2020).
15. Beckouet, F. et al. Releasing Activity Disengages Cohesin's Smc3/Scc1 Interface in a Process Blocked by Acetylation. *Mol Cell* **61**, 563-574 (2016).
16. 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).
17. Hara, K. et al. Structure of cohesin subcomplex pinpoints direct shugoshin-Wapl antagonism in centromeric cohesion. *Nat Struct Mol Biol* **21**, 864-70 (2014).
18. Hons, M.T. et al. Topology and structure of an engineered human cohesin complex bound to Pds5B. *Nat Commun* **7**, 12523 (2016).
19. Kikuchi, S., Borek, D.M., Otwinowski, Z., Tomchick, D.R. & Yu, H. Crystal structure of the cohesin loader Scc2 and insight into cohesinopathy. *Proc Natl Acad Sci U S A* **113**, 12444-12449 (2016).
20. Kueng, S. et al. Wapl controls the dynamic association of cohesin with chromatin. *Cell* **127**, 955-67 (2006).
21. Li, Y. et al. Structural basis for Scc3-dependent cohesin recruitment to chromatin. *Elife* **7**(2018).
22. Li, Y. et al. The structural basis for cohesin-CTCF-anchored loops. *Nature* **578**, 472-476 (2020).
23. Murayama, Y. & Uhlmann, F. Biochemical reconstitution of topological DNA binding by the cohesin ring. *Nature* **505**, 367-71 (2014).
24. Murayama, Y. & Uhlmann, F. DNA Entry into and Exit out of the Cohesin Ring by an Interlocking Gate Mechanism. *Cell* **163**, 1628-40 (2015).
25. Ouyang, Z. et al. Structure of the human cohesin inhibitor Wapl. *Proc Natl Acad Sci U S A* **110**, 11355-60 (2013).

26. Roig, M.B. et al. Structure and function of cohesin's Scc3/SA regulatory subunit. *FEBS Lett* **588**, 3692-702 (2014).
27. Rowland, B.D. et al. Building sister chromatid cohesion: smc3 acetylation counteracts an antiestablishment activity. *Mol Cell* **33**, 763-74 (2009).
28. Wutz, G. et al. ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesin(STAG1) from WAPL. *Elife* **9**(2020).
29. Camdere, G., Guacci, V., Stricklin, J. & Koshland, D. The ATPases of cohesin interface with regulators to modulate cohesin-mediated DNA tethering. *Elife* **4**(2015).
30. 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).
31. Elbatsh, A.M.O. et al. Cohesin Releases DNA through Asymmetric ATPase-Driven Ring Opening. *Mol Cell* **61**, 575-588 (2016).
32. Ladurner, R. et al. Cohesin's ATPase activity couples cohesin loading onto DNA with Smc3 acetylation. *Curr Biol* **24**, 2228-37 (2014).
33. Arumugam, P. et al. ATP hydrolysis is required for cohesin's association with chromosomes. *Curr Biol* **13**, 1941-53 (2003).
34. Davidson, I.F. et al. DNA loop extrusion by human cohesin. *Science* **366**, 1338-1345 (2019).
35. Golfier, S., Quail, T., Kimura, H. & Brugues, J. Cohesin and condensin extrude DNA loops in a cell cycle-dependent manner. *Elife* **9**(2020).
36. Haering, C.H., Farcas, A.M., Arumugam, P., Metson, J. & Nasmyth, K. The cohesin ring concatenates sister DNA molecules. *Nature* **454**, 297-301 (2008).
37. Kim, Y., Shi, Z., Zhang, H., Finkelstein, I.J. & Yu, H. Human cohesin compacts DNA by loop extrusion. *Science* **366**, 1345-1349 (2019).
38. 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).
39. 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).
40. Hinshaw, S.M., Makrantoni, V., Harrison, S.C. & Marston, A.L. The Kinetochore Receptor for the Cohesin Loading Complex. *Cell* **171**, 72-84 e13 (2017).
41. 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).
42. Weitzer, S., Lehane, C. & Uhlmann, F. A model for ATP hydrolysis-dependent binding of cohesin to DNA. *Curr Biol* **13**, 1930-40 (2003).
43. 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).
44. Chan, K.L. et al. Pds5 promotes and protects cohesin acetylation. *Proc Natl Acad Sci U S A* **110**, 13020-5 (2013).
45. 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).
46. Lee, B.G. et al. Crystal Structure of the Cohesin Gatekeeper Pds5 and in Complex with Kleisin Scc1. *Cell Rep* **14**, 2108-2115 (2016).
47. Muir, K.W. et al. Structure of the Pds5-Scc1 Complex and Implications for Cohesin Function. *Cell Rep* **14**, 2116-2126 (2016).
48. 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).
49. Rolef Ben-Shahar, T. et al. Eco1-dependent cohesin acetylation during establishment of sister chromatid cohesion. *Science* **321**, 563-6 (2008).
50. Unal, E. et al. A molecular determinant for the establishment of sister chromatid cohesion. *Science* **321**, 566-9 (2008).

51. 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).
52. Ladurner, R. et al. Sororin actively maintains sister chromatid cohesion. *EMBO J* **35**, 635-53 (2016).
53. 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).
54. Nishiyama, T. et al. Sororin mediates sister chromatid cohesion by antagonizing Wapl. *Cell* **143**, 737-49 (2010).
55. Bernard, P. et al. Cell-cycle regulation of cohesin stability along fission yeast chromosomes. *EMBO J* **27**, 111-21 (2008).
56. 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).
57. 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).
58. Tedeschi, A. et al. Wapl is an essential regulator of chromatin structure and chromosome segregation. *Nature* **501**, 564-8 (2013).
59. 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).
60. Gruber, S. et al. Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. *Cell* **127**, 523-37 (2006).
61. Higashi, T.L. et al. A Structure-Based Mechanism for DNA Entry into the Cohesin Ring. *Mol Cell* **79**, 917-933 e9 (2020).
62. Huis in 't Veld, P.J. et al. Characterization of a DNA exit gate in the human cohesin ring. *Science* **346**, 968-72 (2014).
63. 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).
64. 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).
65. Lin, Z., Luo, X. & Yu, H. Structural basis of cohesin cleavage by separase. *Nature* **532**, 131-4 (2016).
66. 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).
67. 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).
68. 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).
69. Dauban, L. et al. Regulation of Cohesin-Mediated Chromosome Folding by Eco1 and Other Partners. *Mol Cell* **77**, 1279-1293 e4 (2020).
70. Feytout, A., Vaur, S., Genier, S., Vazquez, S. & Javerzat, J.P. Psm3 acetylation on conserved lysine residues is dispensable for viability in fission yeast but contributes to Eso1-mediated sister chromatid cohesion by antagonizing Wpl1. *Mol Cell Biol* **31**, 1771-86 (2011).
71. Haarhuis, J.H.I. et al. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* **169**, 693-707 e14 (2017).
72. Kanke, M., Tahara, E., Huis In't Veld, P.J. & Nishiyama, T. Cohesin acetylation and Wapl-Pds5 oppositely regulate translocation of cohesin along DNA. *EMBO J* **35**, 2686-2698 (2016).
73. Nora, E.P. et al. Molecular basis of CTCF binding polarity in genome folding. *Nat Commun* **11**, 5612 (2020).
74. Rao, S.S.P. et al. Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e24 (2017).

75. Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51-56 (2017).
76. Srinivasan, M. et al. Scc2 counteracts a Wapl-independent mechanism that releases cohesin from chromosomes during G1. *Elife* **8**(2019).
77. Vian, L. et al. The Energetics and Physiological Impact of Cohesin Extrusion. *Cell* **175**, 292-294 (2018).
78. 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).
79. 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).
80. Burmann, F. et al. A folded conformation of MukBEF and cohesin. *Nat Struct Mol Biol* **26**, 227-236 (2019).
81. 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).
82. Diebold-Durand, M.L. et al. Structure of Full-Length SMC and Rearrangements Required for Chromosome Organization. *Mol Cell* **67**, 334-347 e5 (2017).
83. Hassler, M. et al. Structural Basis of an Asymmetric Condensin ATPase Cycle. *Mol Cell* **74**, 1175-1188 e9 (2019).
84. 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).
85. 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).
86. Lee, B.G., Rhodes, J. & Lowe, J. Clamping of DNA shuts the condensin neck gate. *Proc Natl Acad Sci U S A* **119**, e2120006119 (2022).
87. Petela, N.J. et al. Folding of cohesin's coiled coil is important for Scc2/4-induced association with chromosomes. *Elife* **10**(2021).
88. Shaltiel, I.A. et al. A hold-and-feed mechanism drives directional DNA loop extrusion by condensin. *Science* **376**, 1087-1094 (2022).
89. Soh, Y.M. et al. Molecular basis for SMC rod formation and its dissolution upon DNA binding. *Mol Cell* **57**, 290-303 (2015).
90. 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).
91. Bauer, B.W. et al. Cohesin mediates DNA loop extrusion by a "swing and clamp" mechanism. *Cell* **184**, 5448-5464 e22 (2021).
92. 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).
93. Rohatgi, S. et al. Facial diagnosis of mild and variant CdLS: Insights from a dysmorphologist survey. *Am J Med Genet A* **152A**, 1641-53 (2010).
94. 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).
95. Xu, B., Sowa, N., Cardenas, M.E. & Gerton, J.L. L-leucine partially rescues translational and developmental defects associated with zebrafish models of Cornelia de Lange syndrome. *Hum Mol Genet* **24**, 1540-55 (2015).
96. Kline, A.D. et al. Diagnosis and management of Cornelia de Lange syndrome: first international consensus statement. *Nat Rev Genet* (2018).
97. Chao, W.C. et al. Structure of the cohesin loader Scc2. *Nat Commun* **8**, 13952 (2017).
98. Elbatsh, A.M.O. et al. Distinct Roles for Condensin's Two ATPase Sites in Chromosome Condensation. *Mol Cell* **76**, 724-737 e5 (2019).

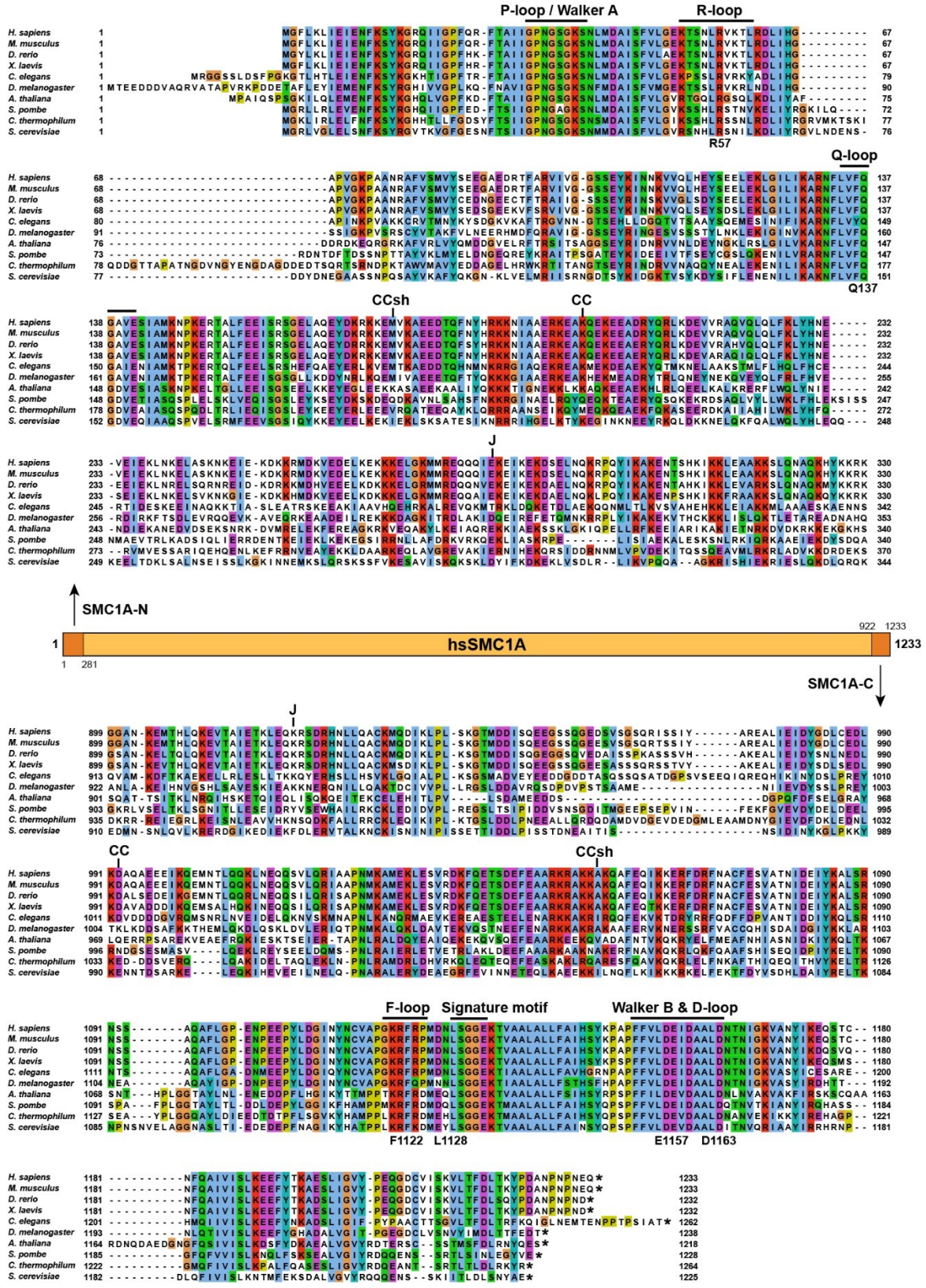
99. Diebold, M.L., Fribourg, S., Koch, M., Metzger, T. & Romier, C. Deciphering correct strategies for multiprotein complex assembly by co-expression: application to complexes as large as the histone octamer. *J Struct Biol* **175**, 178-88 (2011).
100. Fribourg, S. et al. Dissecting the interaction network of multiprotein complexes by pairwise coexpression of subunits in E. coli. *J Mol Biol* **306**, 363-73 (2001).
101. 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).
102. 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).
103. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676-682 (2012).
104. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-32 (2010).
105. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760-3 (1994).
106. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
107. 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).
108. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography* **66**, 486-501 (2010).
109. Williams, C.J. et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293-315 (2018).
110. Laue, T.M., Shah, B.D., Ridgeway, T.M., Pelletier, S.L. *Computer-aided Interpretation of Sedimentation Data for Proteins*, (Royal Society of Chemistry, Cambridge [England], 1992).
111. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J* **78**, 1606-19 (2000).
112. Brautigam, C.A. Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. *Methods Enzymol* **562**, 109-33 (2015).
113. Meyerson, J.R. et al. Self-assembled monolayers improve protein distribution on holey carbon cryo-EM supports. *Sci Rep* **4**, 7084 (2014).
114. Mastronarde, D.N. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51 (2005).
115. Tegunov, D. & Cramer, P. Real-time cryo-electron microscopy data preprocessing with Warp. *Nat Methods* **16**, 1146-1152 (2019).
116. Punjani, A., Zhang, H. & Fleet, D.J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat Methods* **17**, 1214-1221 (2020).
117. Pettersen, E.F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (2004).
118. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877 (2019).
119. Prisant, M.G., Williams, C.J., Chen, V.B., Richardson, J.S. & Richardson, D.C. New tools in MolProbity validation: CaBLAM for CryoEM backbone, UnDowser to rethink "waters," and NGL Viewer to recapture online 3D graphics. *Protein Sci* **29**, 315-329 (2020).
120. Westerfield, M. *The zebrafish book 5th edition - A guide for the laboratory use of zebrafish (Danio rerio)*, (University of Oregon Press, Eugene, 2007).

Specific conformational dynamics of the ATPase head domains and DNA exit gate mediate the Cohesin ATPase cycle

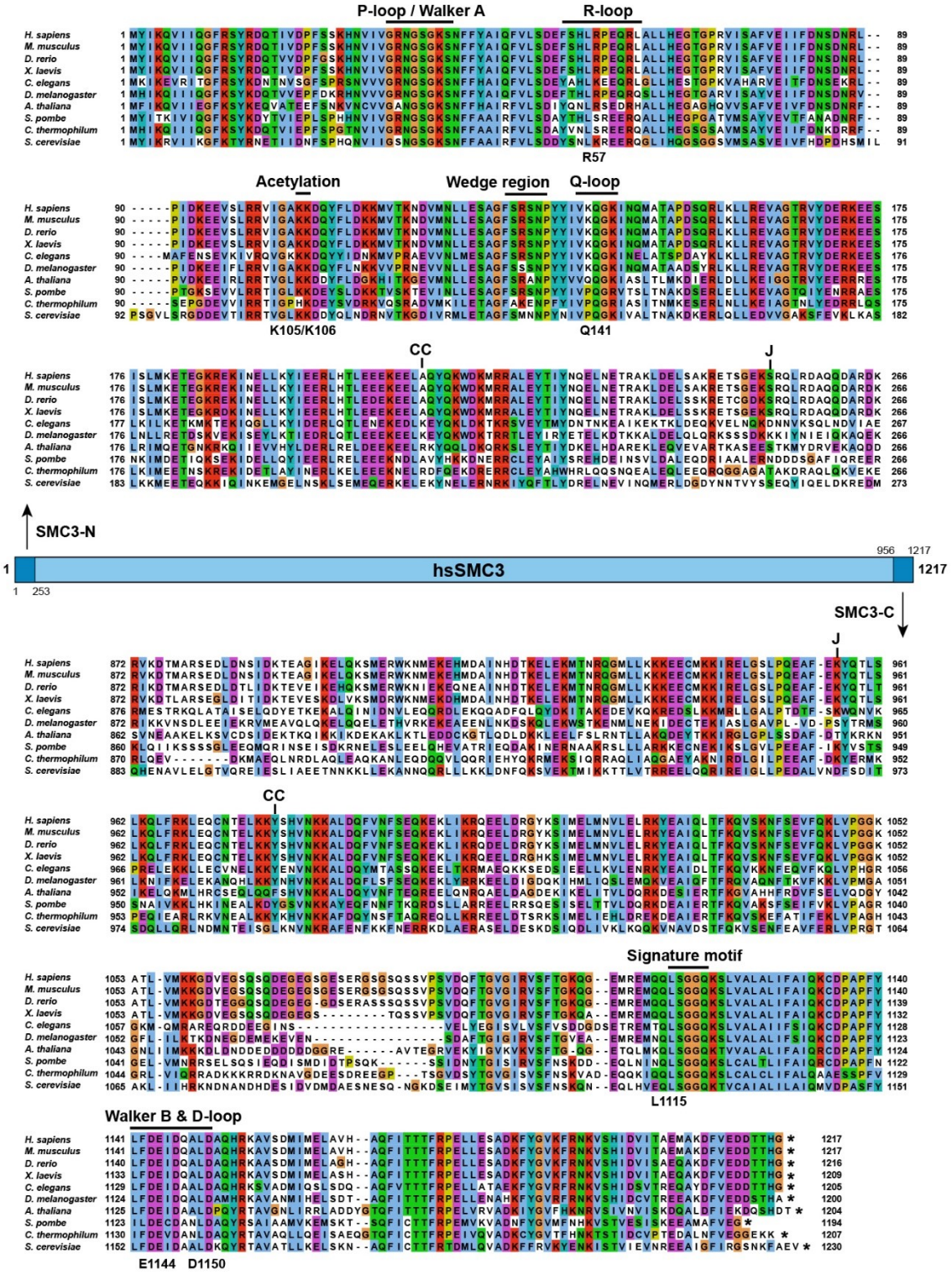
Vitoria Gomes, M., Landwerlin, P., Diebold-Durand, M.-L, Shaik, T.B., Troesch, E., Weber, C., Durand, A., Brillet, K., Dulac, L., Antony, P., Watrin, E., Ennifar, E., Golzio, C. and Romier, C.

Supplementary Information

A


P-loop / Walker A
R-loop
R57
Q-loop
Q137
CCsh
CC
J
SMC1A-N
1
922
1233
hsSMC1A
1233
1
281
SMC1A-C
J
CC
CCsh
F-loop
Signature motif
Walker B & D-loop
F1122
L1128
E1157
D1163
H. sapiens
M. musculus
D. rerio
X. laevis
C. elegans
D. melanogaster
A. thaliana
S. pombe
C. thermophilum
S. cerevisiae

B


P-loop / Walker A
R-loop
R57
Acetylation
Wedge region
Q-loop
K105/K106
Q141
CC
J
SMC3-N
hsSMC3
1
253
956
1217
SMC3-C
Signature motif
L1115
Walker B & D-loop
E1144
D1150
H. sapiens
M. musculus
D. rerio
X. laevis
C. elegans
D. melanogaster
A. thaliana
S. pombe
C. thermophilum
S. cerevisiae

C

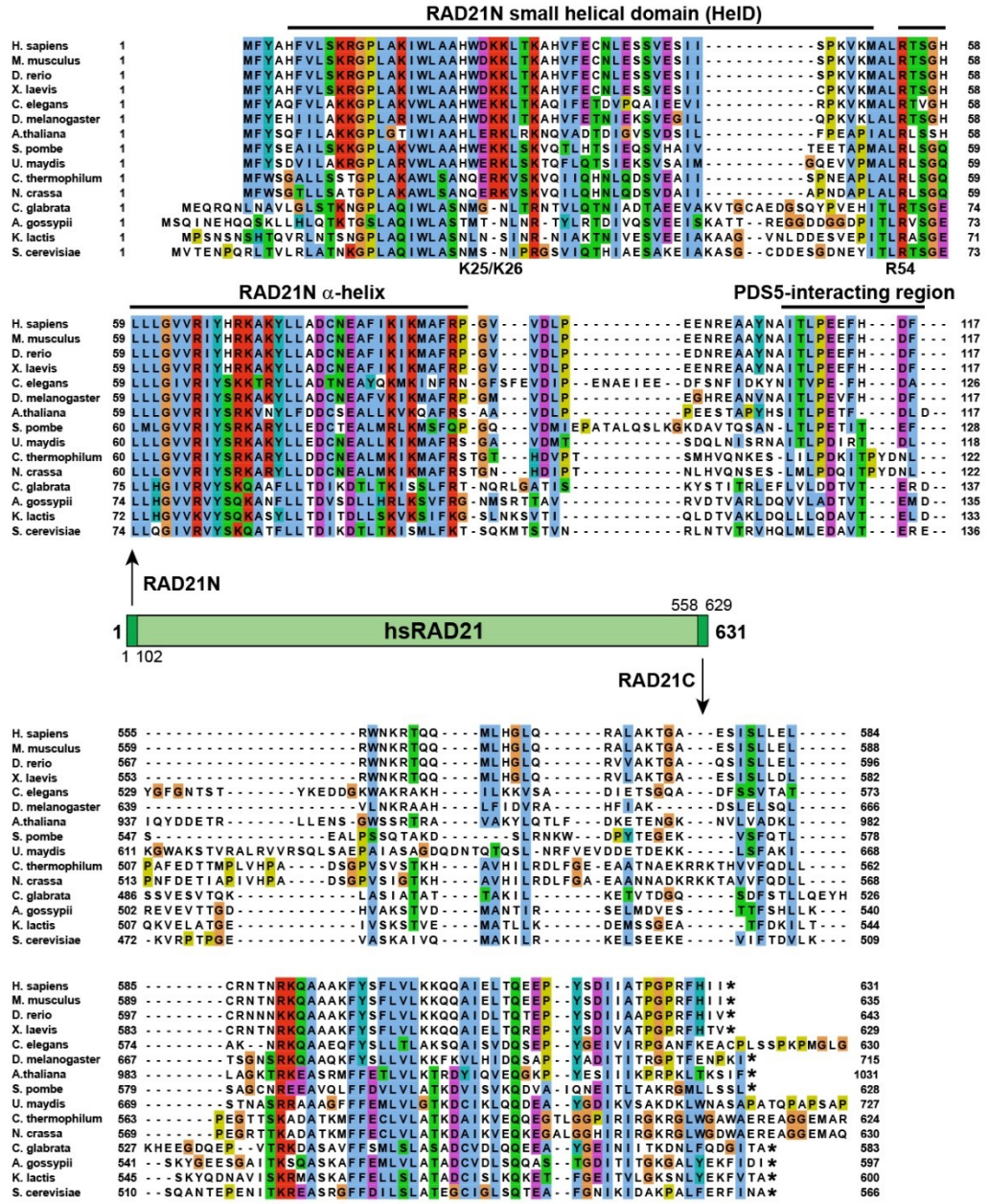
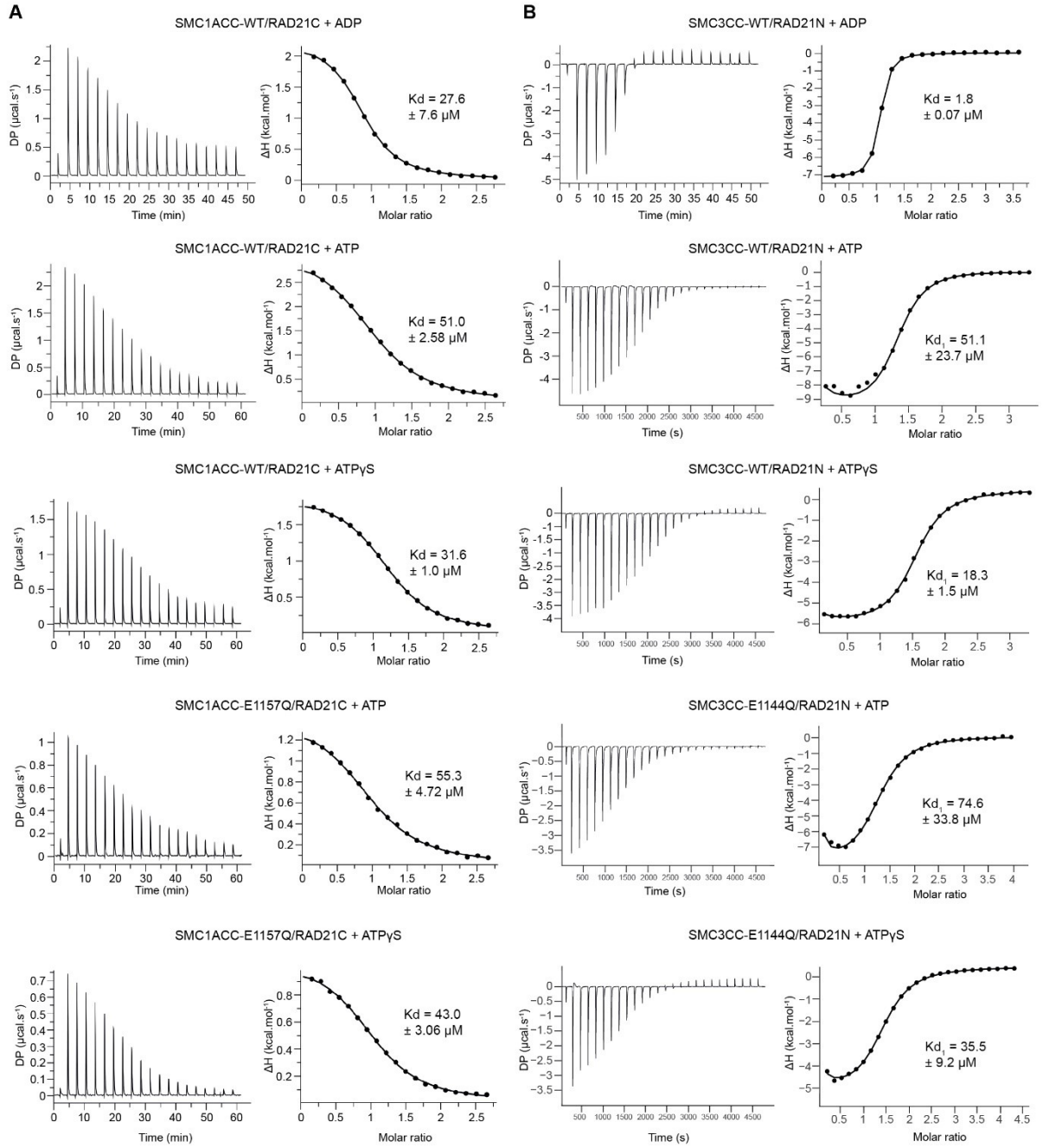


Figure S1. Sequence alignments of SMC1A HD, SMC3 HD, RAD21N and RAD21C.

A. Sequence alignment of the N- and C-terminal semi-ATPase domains of SMC1A^{Smc1} from various organisms. The clustalX coloring scheme is used for representing sequence conservations. The major sequence/structural elements and residues discussed in the text are indicated. Asterisks indicate the C-terminal end of the sequences. Boundaries of the clones used in the study (J, CC, CCsh) are marked. In the central representation of the full-length SMC1A, the clones used are represented at both extremities and the numbering of the longer clone is indicated.

B. Same as in (A) for the N- and Cterminal semi-ATPase domains of SMC3$^{Smc3.}$

C. Same as in (A) for the N- and C-terminal domains of RAD21^{Scc1}.


A
SMC1ACC-WT/RAD21C + ADP
Kd = 27.6
± 7.6 µM
SMC1ACC-WT/RAD21C + ATP
Kd = 51.0
± 2.58 µM
SMC1ACC-WT/RAD21C + ATPγS
Kd = 31.6
± 1.0 µM
SMC1ACC-E1157Q/RAD21C + ATP
Kd = 55.3
± 4.72 µM
SMC1ACC-E1157Q/RAD21C + ATPγS
Kd = 43.0
± 3.06 µM
B
SMC3CC-WT/RAD21N + ADP
Kd = 1.8
± 0.07 µM
SMC3CC-WT/RAD21N + ATP
Kd1 = 51.1
± 23.7 µM
SMC3CC-WT/RAD21N + ATPγS
Kd1 = 18.3
± 1.5 µM
SMC3CC-E1144Q/RAD21N + ATP
Kd1 = 74.6
± 33.8 µM
SMC3CC-E1144Q/RAD21N + ATPγS
Kd1 = 35.5
± 9.2 µM
DP (µcal.s-1)
ΔH (kcal.mol-1)
Time (min)
Time (s)
Molar ratio

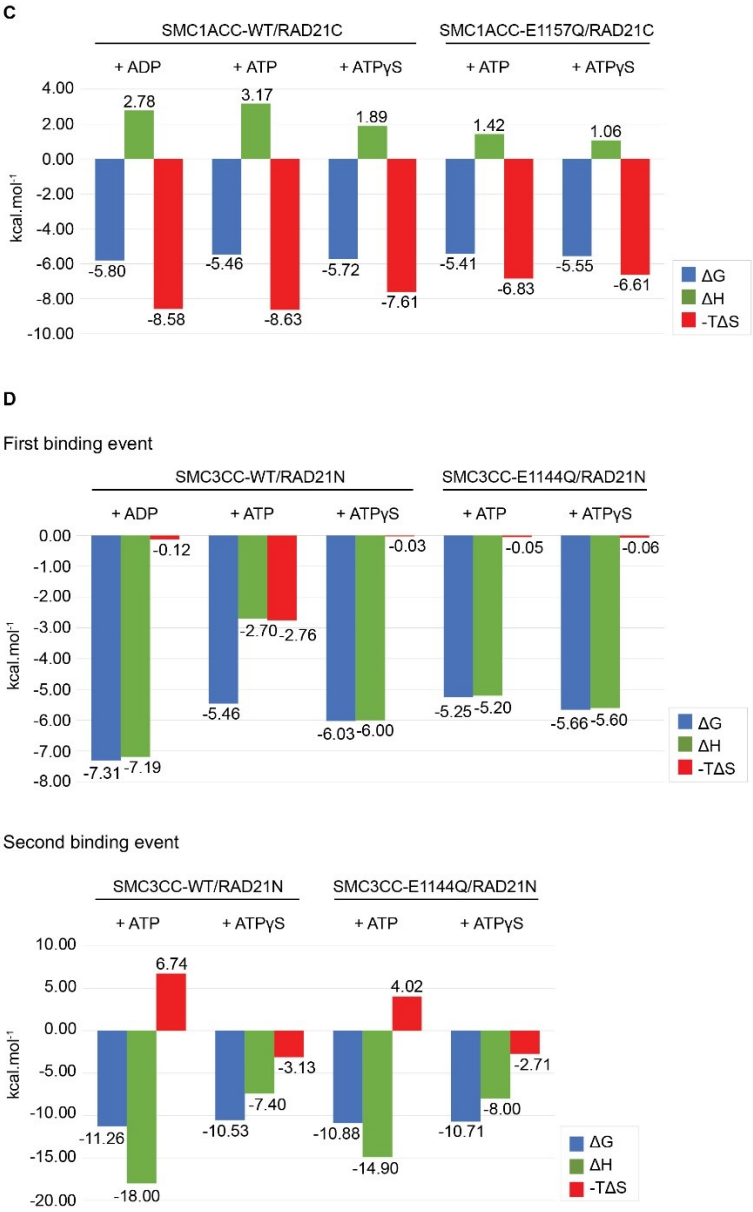


Figure S2. Analysis by isothermal titration calorimetry (ITC) of ADP, ATP and ATPγS binding to the SMC1A HD and SMC3 HD.

A. Measurements of the Kd of ADP, ATP and ATPγS for the SMC1ACC/RAD21C complex using the SMC1ACC WT and E1157Q mutant constructs. The measurements profiles are provided on the left panels, the interpolations on the right panels. The Kd values are given above the interpolation curves. The complete thermodynamic parameters obtained from these experiments are given in Table S1.
B. Same as in (A) for the SMC3CC/RAD21N complex with the WT and E1144Q mutant constructs. Only the higher Kd value is given for the experiments with ATP and ATPγS that include two events. The complete thermodynamic parameters obtained from these experiments are given in Table S1.
C. Thermodynamic signatures obtained from the ITC experiments performed with the SMC1ACC/RAD21C complex.
D. Same as in (C) for the SMC3CC/RAD21N complex.

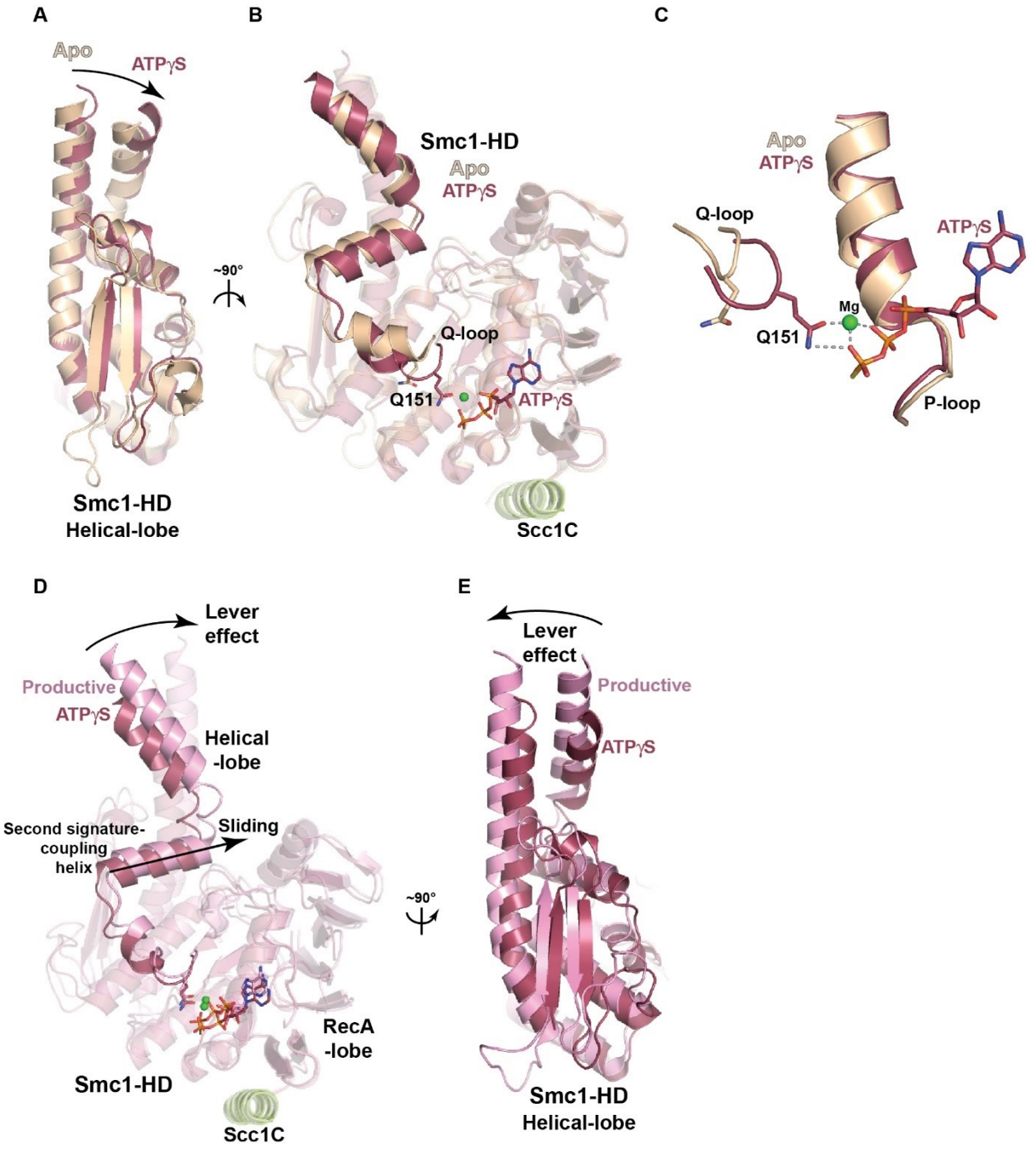


Figure S3. Conformational movements of the fungi Smc1 HD.

A. Rotational movement of the Helical-lobe of fungi Smc1 HD, from the *Chaetomium themophilum* apo structure to the *S. cerevisiae* ATPγS-bound structure.
B. Same as in (A) in a ~90° view. The rotational movement occurs in a plan that goes through Q151.

C. Same as in (B) with a closeup on the Q-loop and Q151.

D. Lever effect observed on the *S. cerevisiae* Smc1 HD upon formation of the productive complex with a forward movement of the CC and a sliding of the second signature-coupling helix.
E. Same as in (D) in a ~90° view focusing on the Helical-lobe. The lever effect moves the Smc1 CC back to a central apo position.

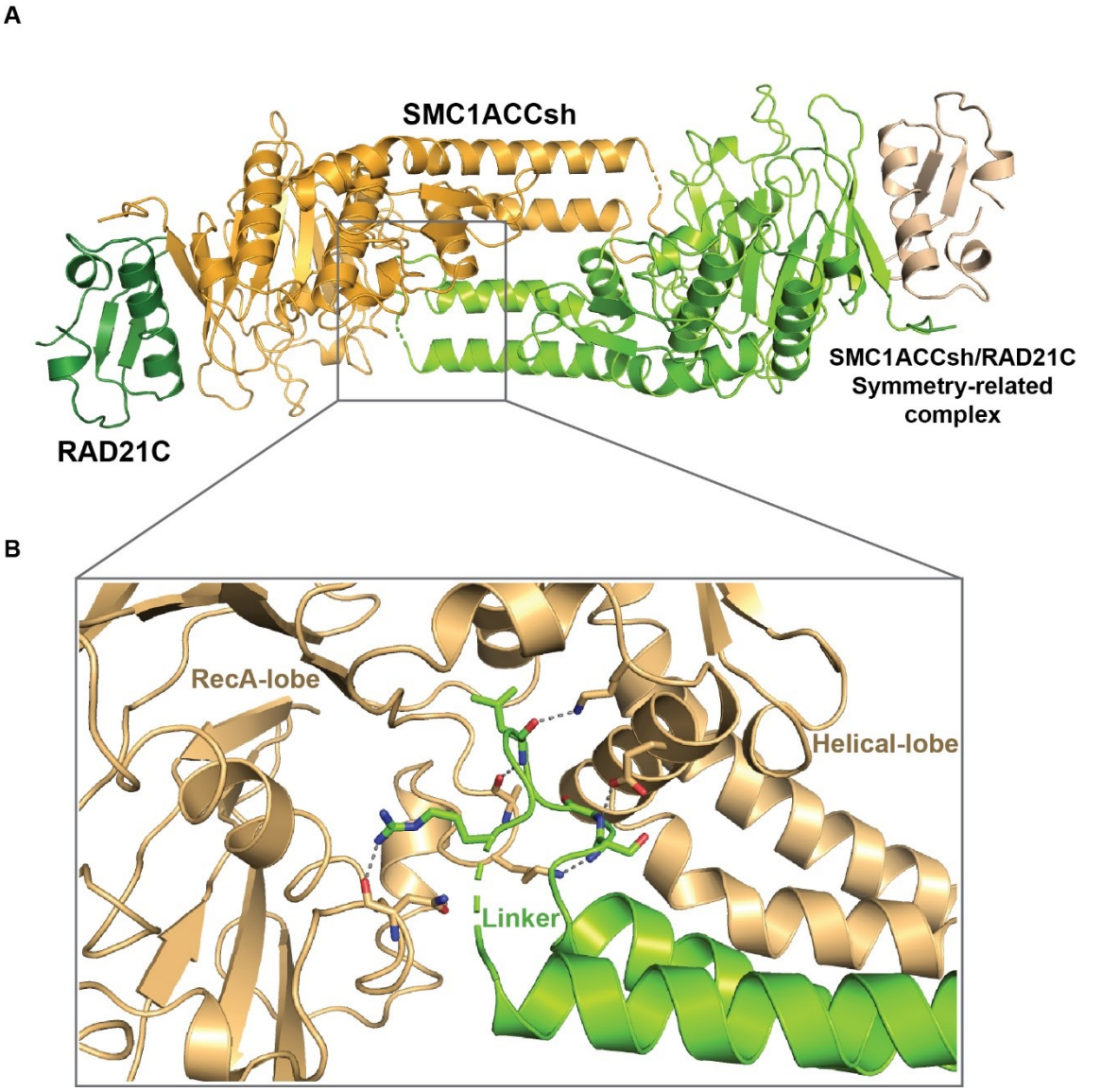


Figure S4. Artificial lever effect induced by the crystal packing of the SMC1ACCsh/RAD21C complex.

A. Crystal packing in the crystals of the SMC1ACCsh/RAD21C complex that occurs regardless of the nucleotide binding state.

B. Details of the binding of the artificial linker, which bridges the SMC1ACCsh N- and C-terminal coils, at the SMC1ACCsh RecA-lobe/Helical-lobe interface and that induces an artificial lever effect-like sliding of the two lobes with respect to each other.

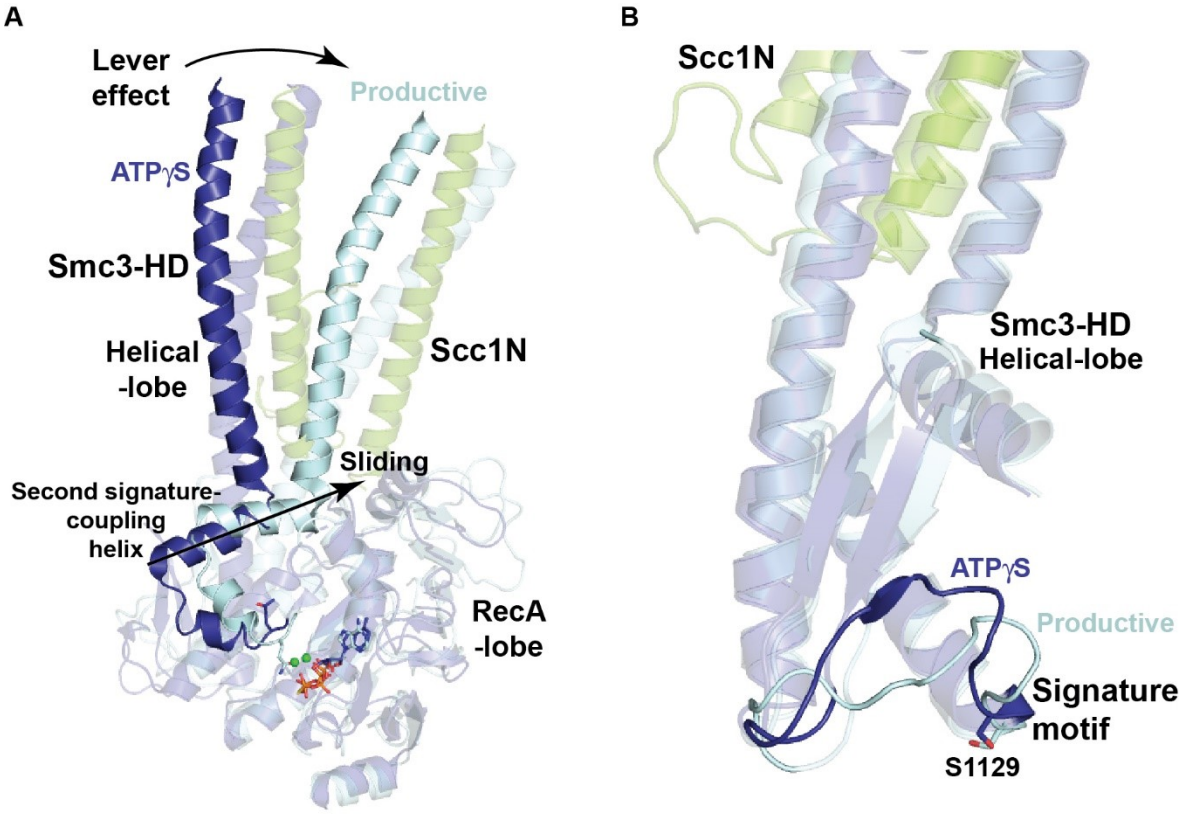


Figure S5. Lever effect on the *S. cerevisiae* SMC3 HD upon formation of the productive complex.

A. Lever effect on the SMC3 HD upon formation of the productive complex that induces a forward movement of the CCs and a sliding movement of the second signature-coupling helix.
B. Conformational changes and repositioning of the loop harboring the Smc3 HD signature motif upon formation of the productive complex.

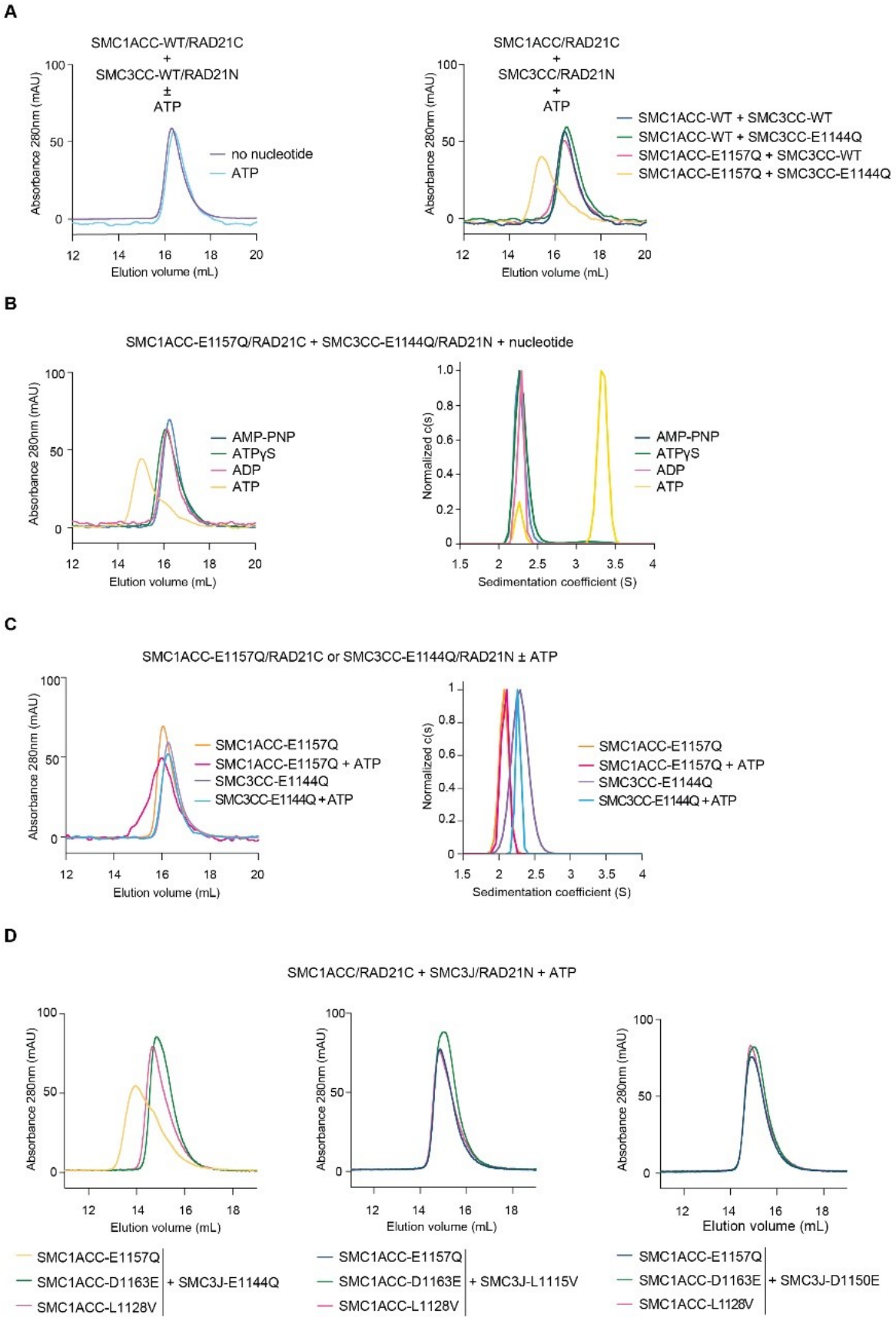

A
SMC1ACC-WT/RAD21C
+
SMC3CC-WT/RAD21N
±
ATP
no nucleotide
ATP
SMC1ACC/RAD21C
+
SMC3CC/RAD21N
+
ATP
SMC1ACC-WT + SMC3CC-WT
SMC1ACC-WT + SMC3CC-E1144Q
SMC1ACC-E1157Q + SMC3CC-WT
SMC1ACC-E1157Q + SMC3CC-E1144Q
Absorbance 280nm (mAU)
Elution volume (mL)
B
SMC1ACC-E1157Q/RAD21C + SMC3CC-E1144Q/RAD21N + nucleotide
AMP-PNP
ATPγS
ADP
ATP
Normalized c(s)
Sedimentation coefficient (S)
C
SMC1ACC-E1157Q/RAD21C or SMC3CC-E1144Q/RAD21N ± ATP
SMC1ACC-E1157Q
SMC1ACC-E1157Q + ATP
SMC3CC-E1144Q
SMC3CC-E1144Q + ATP
D
SMC1ACC/RAD21C + SMC3J/RAD21N + ATP
SMC1ACC-E1157Q
SMC1ACC-D1163E
SMC1ACC-L1128V
+ SMC3J-E1144Q
+ SMC3J-L1115V
+ SMC3J-D1150E

Figure S6. Reconstitution of the human engaged ATPase module in the presence of WT and mutant SMC1A and SMC3 HDs and various nucleotides.

A. Size-exclusion chromatography profiles of reconstitution experiments of the human engaged ATPase module using the WT and EQ mutant SMC1ACC/RAD21C and SMC3CC/RAD21N complexes in the absence and the presence of ATP. Stable engagement requires both EQ mutants and ATP.
B. Same as in (A) (left panel) with the EQ mutants and various nucleotides, and confirmation by analytical ultracentrifugation experiments (right panel). Only ATP is able to stably maintain engagement.
C. Same as in (B) for the characterization of SMC1ACC/RAD21C or SMC3CC/RAD21N homodimers in the absence or the presence of ATP. No stable homodimerization is observed.
D. Same as in (A) using, beside the EQ mutants, the LV and DE mutants that have been used to demonstrate the asymmetry of the Cohesin ATPase active sites. None of the combinations, except with the two EQ mutants, can stabilize the engaged ATPase module. Of note, these initial experiments have been performed with the SMC3J/RAD21N complex, showing that stable engagement also occurs in the presence of the SMC3 joint element.

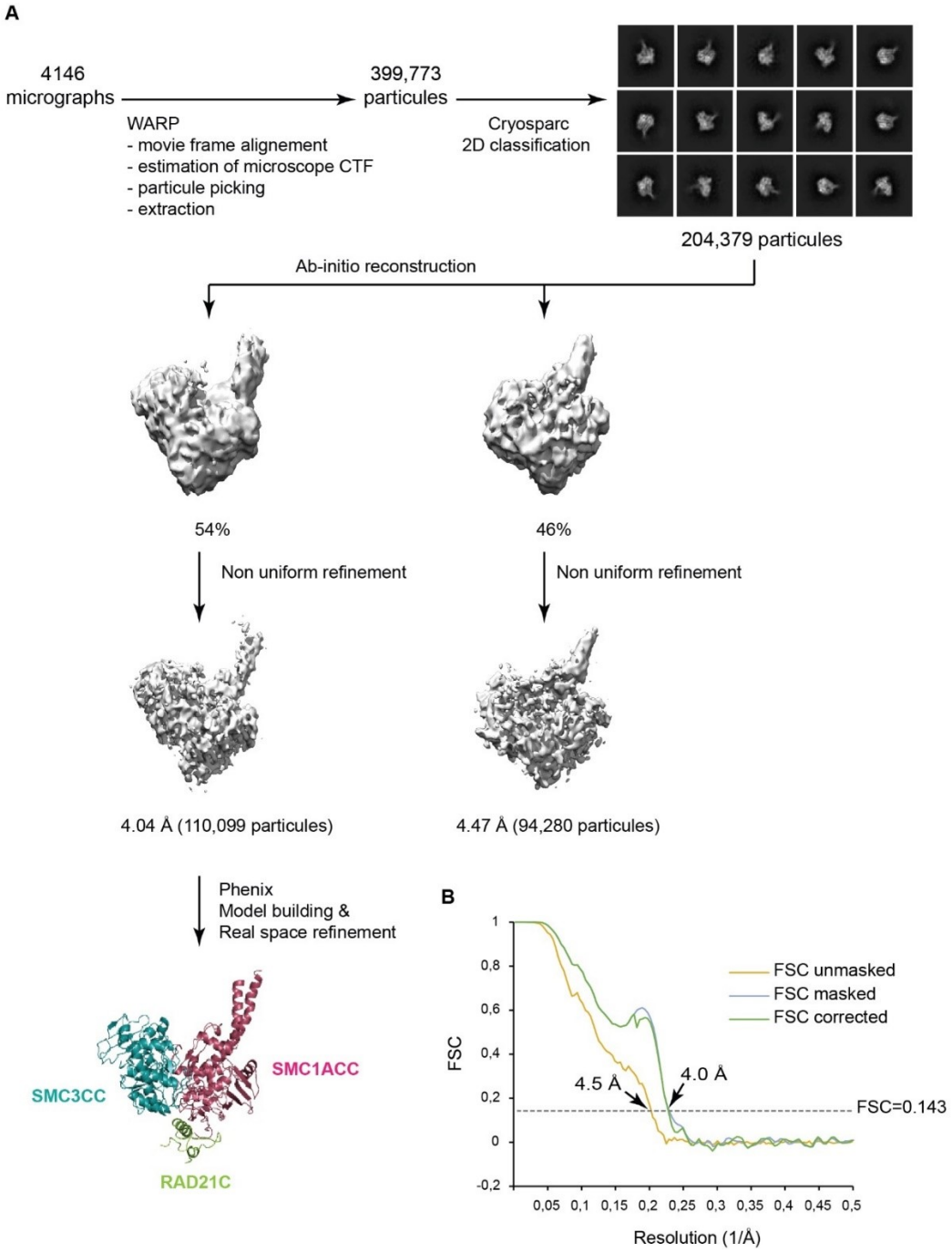


Figure S7. Cryo-EM data processing for the human engaged ATPase module.

A. Flow-chart of the processing of the cryo-EM data obtained on the human engaged ATPase module and final cryo-EM maps obtained. The programs and the number of particles used in each step are indicated. The 2D class averages are shown as good classes representatives. The processing led to two reconstructions with an average resolution of 4.04 Å and 4.47 Å. The 4.04 Å map, obtained from a majority of the particles, was used for model building and real-space refinement. The lower resolution cryo-EM map could not be used for model building due to inherent heterogeneity.

B. FSC plot of the final density 4.04 Å map of the human engaged ATPase module. The plot shows the unmasked (yellow), masked (blue) and masking-effect-corrected (green) FSC curves. The resolution at which the gold-standard FSC curve drops below the 0.143 threshold is indicated.

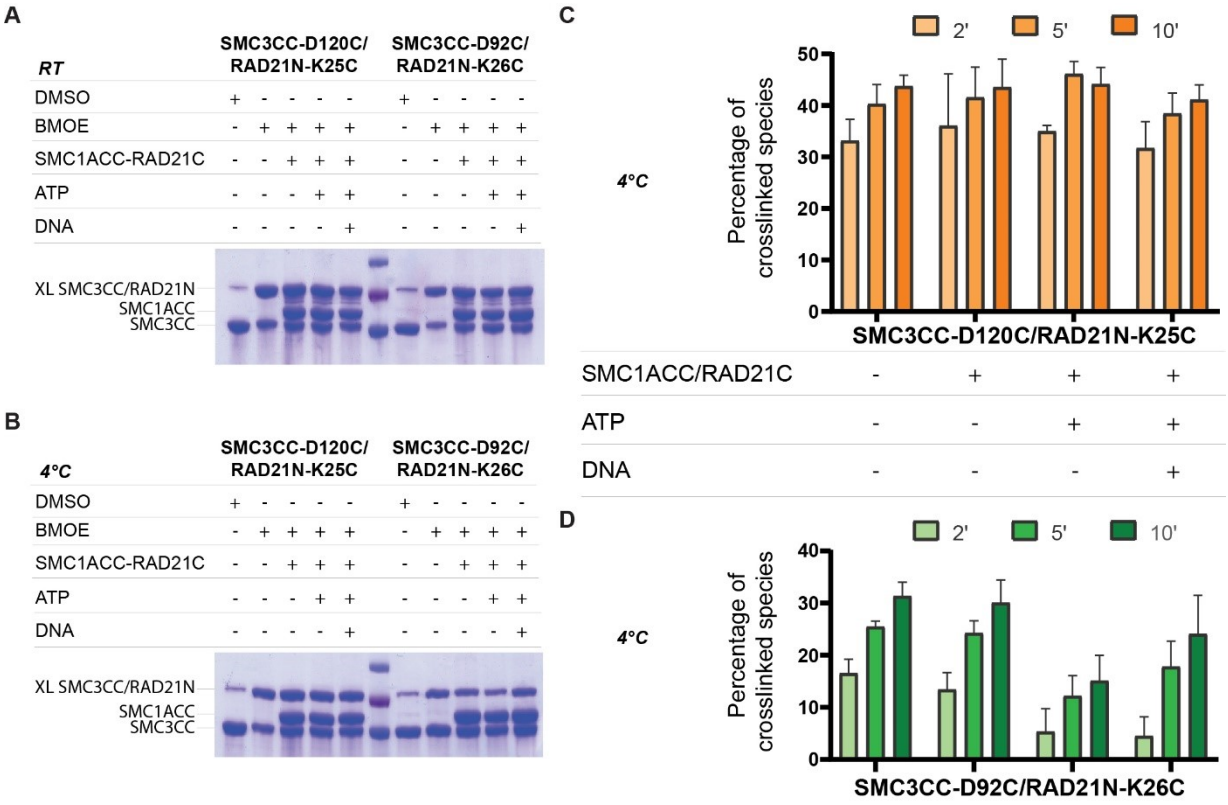


Figure S8. Crosslinking experiments to assess the flexibility of the DNA exit gate upon engagement.

A. SDS-PAGE analysis of the 5-minutes crosslinking experiments performed at room temperature (RT) for both crosslinking pairs. XL, crosslinked species. B. Same as in (A) with experiments performed at 4°C.

B. Quantification of crosslinked species for the SMC3CC-D120C/RAD21N-K25C pair in experiments performed at 4°C. The supplementation of the SMC1ACC/RAD21C, ATP and DNA is indicated underneath the graph. All experiments were done in triplicates.

C. Same as in (C) for the SMC3CC-D92C/RAD21N-K26C pair. The supplementation of the SMC1ACC/RAD21C, ATP and DNA is indicated above the graph in (C). All experiments were done in triplicates.

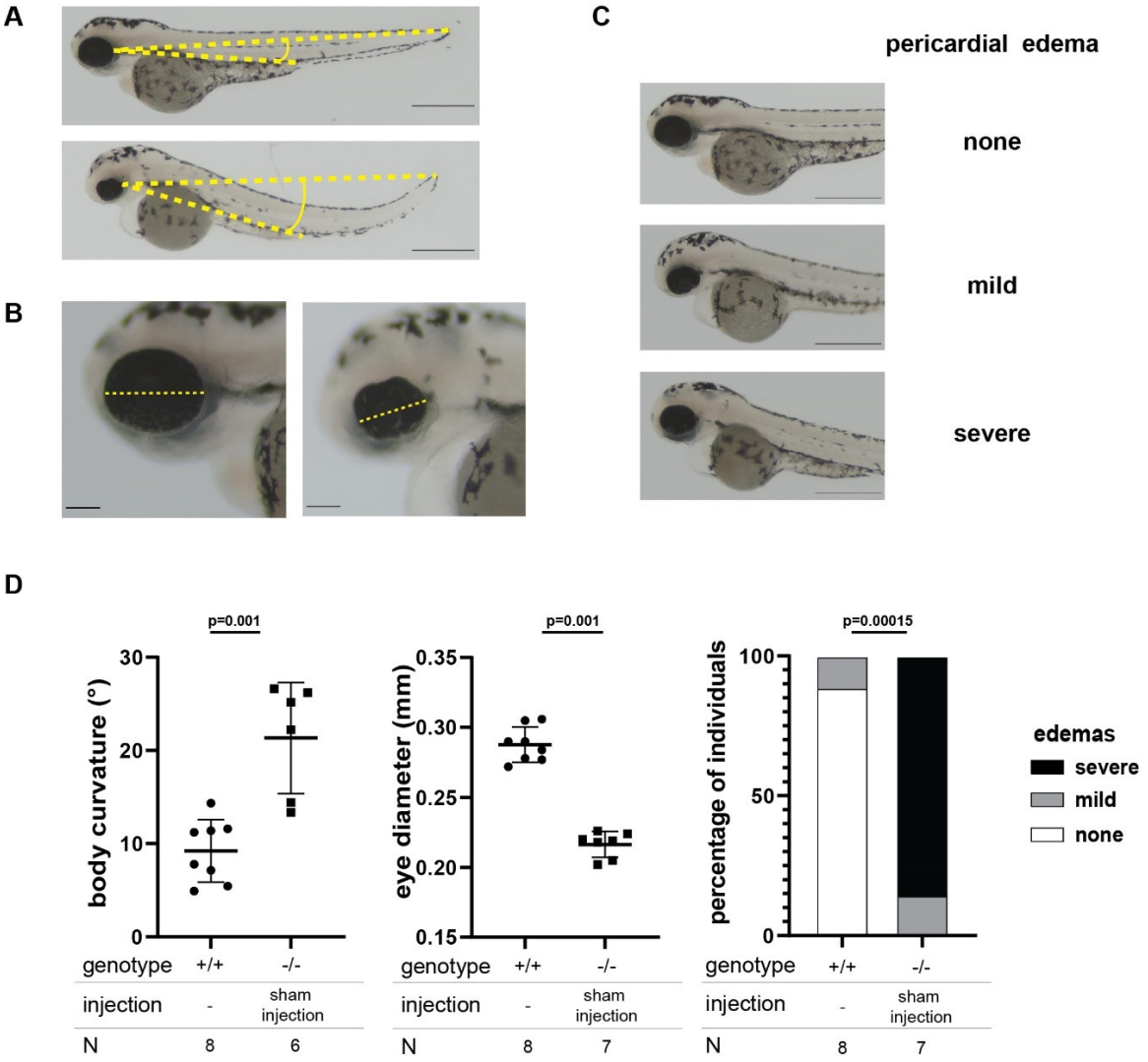


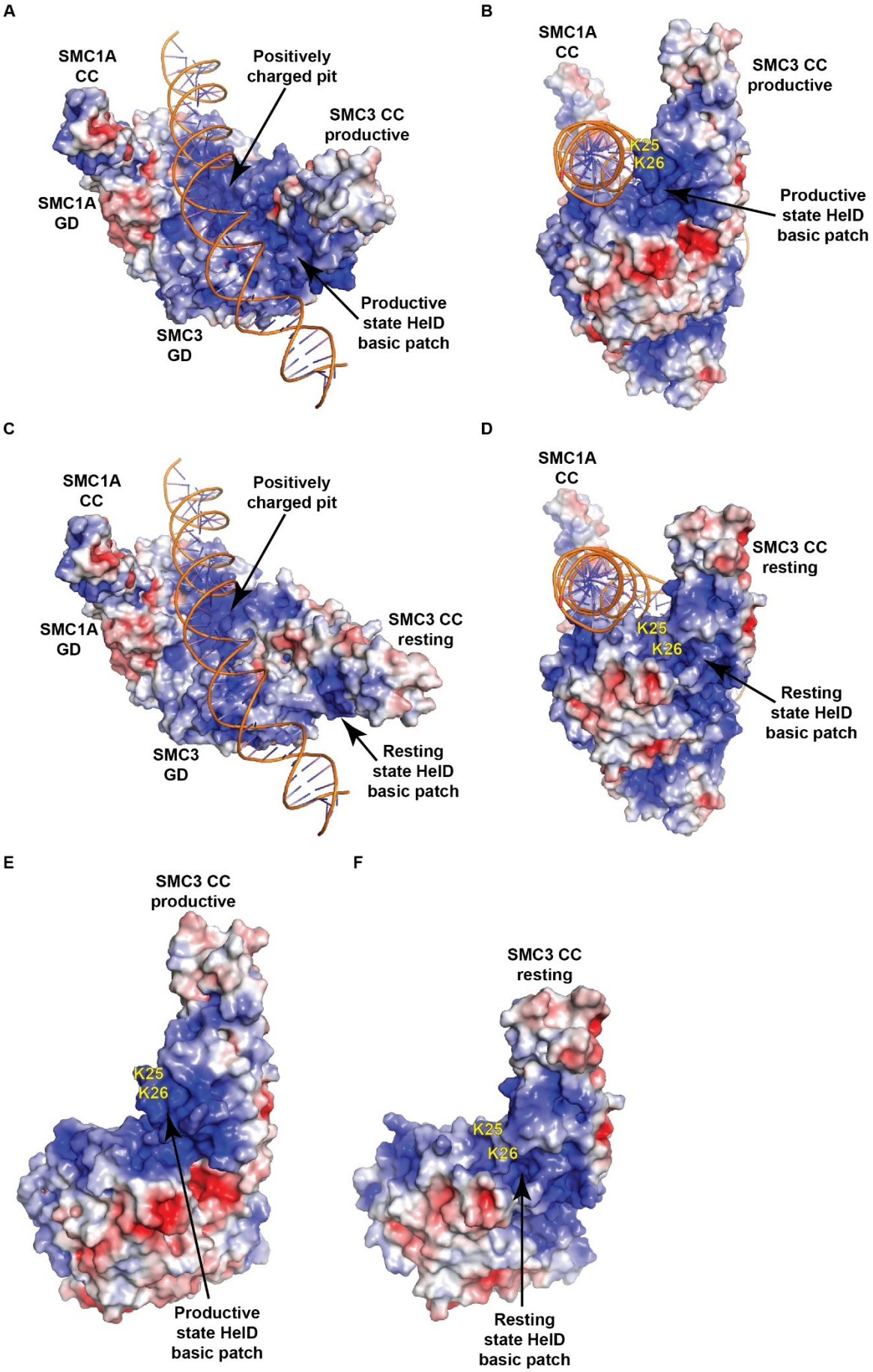
Figure S9. Assessment of the zebrafish body curvature, eye diameter and pericardial edema.

A. Assessment of the body curvature of the zebrafish embryos by measurement of the angle formed between a line going from the tip of the tail to the pigment line that emerges from the lateral side of the eye and a line starting from the same point to the anal disc. Zebrafish larvae with a curved body (lower panel) have a wider angle than WT larvae (upper panel).

B. Assessment of the zebrafish eye diameter (as assessment of microcephaly) measured from the external side of the eye, where the pigment line emerges, to the opposite side.

C. Scoring of the zebrafish pericardial edema. From top to bottom, representative images of zebrafish larvae at 3 dpf with no, mild, or severe pericardial edema.

D. Dot plots showing body curvature (°) and eye diameter (mm) and bar graph showing the presence of pericardial edema for WT and homozygous mutant rad21$^{hi2529Tg/hi2529Tg}$ larvae. For pericardial edema, larvae were binned into three categories: severe, mild or absent (none) as shown in (C). Statistical significance was assessed by ANOVA followed by a Tukey's test for post-hoc analysis for the body curvature and eye diameter, and a Fisher's exact test for the pericardial edema. All experiments have been performed two times. The p-values are indicated on the graphs. N corresponds to the number of larvae per condition.


A
SMC1A
CC
Positively
charged pit
SMC3 CC
productive
SMC1A
GD
Productive
state HelD
basic patch
SMC3
GD
B
SMC1A
CC
SMC3 CC
productive
K25
K26
Productive
state HelD
basic patch
C
SMC1A
CC
Positively
charged pit
SMC1A
GD
SMC3 CC
resting
SMC3
GD
Resting
state HelD
basic patch
D
SMC1A
CC
SMC3 CC
resting
K25
K26
Resting
state HelD
basic patch
E
SMC3 CC
productive
K25
K26
Productive
state HelD
basic patch
F
SMC3 CC
resting
K25
K26
Resting
state HelD
basic patch

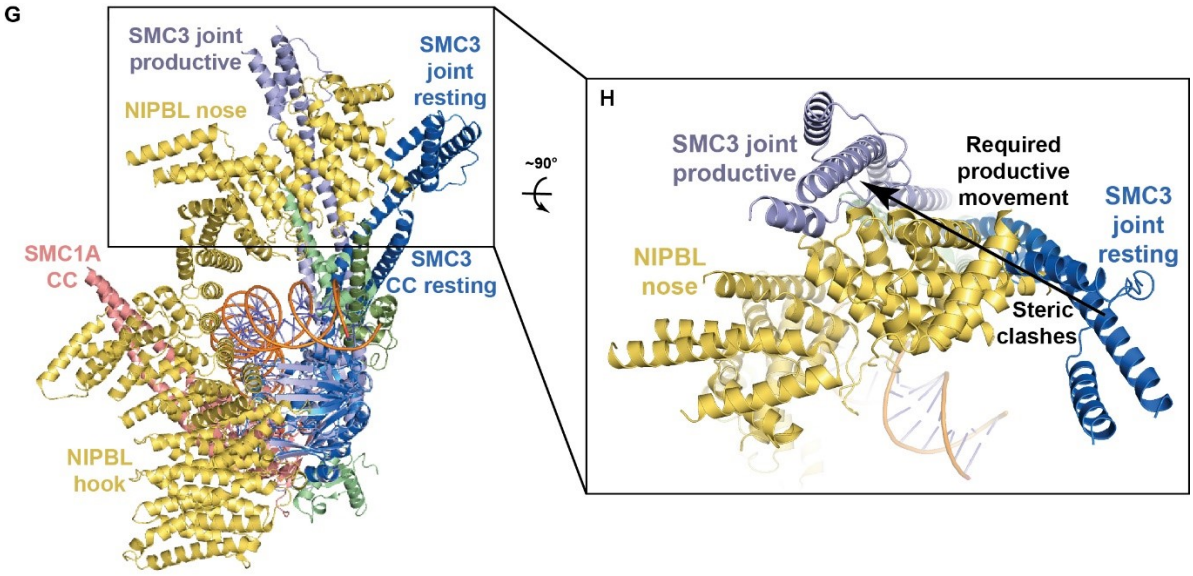


Figure S10. Different properties of the SMC3C HD and DNA exit gate in the resting and productive conformations.

A,B. Two ~90° views of DNA bound to the productive ATPase module represented by its electrostatic surface (red, negatively charged; blue, positively charged). The RAD21N HelD domain, which harbors the K25 and K26 lysines, is highly positively charged and contributes to the gripping of the DNA in its binding chamber and to its curvature.

C,D. Same as in (A,B) with the SMC3 HD in a resting position. The displacement of the SMC3 CC, and therefore of the HelD, widens the DNA binding chamber, facilitating DNA entry and movement.

E. Electrostatic surface representation of the SMC3CC/RAD21N complex in its productive conformation (calculations made without the SMC1A HD).
F. Electrostatic surface representation of the resting SMC3CC/RAD21N complex. This stable inactive complex displays a different surface and electrostatic properties than in the productive conformation shown in (E). These features can be used as specific docking platform for the incoming NIPBL/SMC1A and SMC3 hinges/DNA complex in the swing-and-clamp mechanism for human Cohesin loop extrusion. G. Ribbon representation of the human productive complex (SMC1A HD, pink; SMC3 HD, lavender (productive)/blue (resting); RAD21, green; NIPBL, yellow; DNA,orange) with the SMC3 CC shown both in the productive and resting conformations.
H. Closeup of (G) on the position of the NIPBL nose with respect to the SMC3 joint element. Only the productive positioning of the SMC3 joint is compatible with the productive positioning of the NIPBL nose. The resting positioning of the SMC3 joint as well as any intermediary position clash with the NIPBL nose, showing that, when NIPBL initially docks onto the SMC1A HD, both domains either have to move conjointly in order to form their productive interaction, or are prevented to interact. These alternative mechanisms could underlie different Cohesin mechanistic pathways and functions.

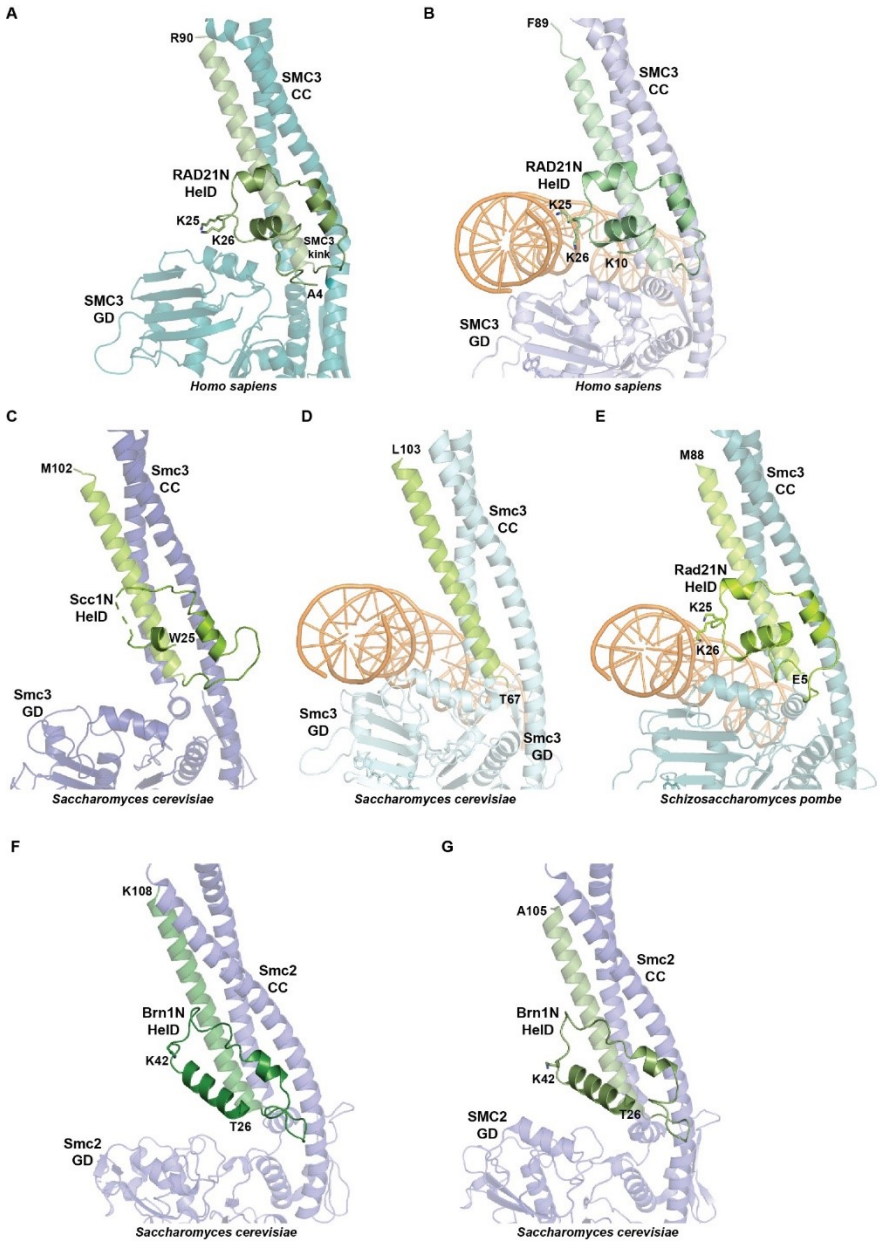


Figure S11. DNA exit gate conformations in Cohesin and Condensin.

Ribbon representations of the DNA exit gate in Cohesin and Condensin. Specific residues, including the N-terminal residue are labeled. DNA is colored orange.

A. Human Cohesin DNA exit gate in the resting conformation (this study).

B. Human Cohesin DNA exit gate in the productive conformation (PDB entry 6WGE).

C. *S. cerevisiae* Cohesin DNA exit gate in the ATPγS-bound conformation (PDB entry 4UX3).

D. *S. cerevisiae* Cohesin DNA exit gate in the productive conformation (PDB entry 6ZZ6).

E. *S. pombe* Cohesin DNA exit gate in the productive conformation (PDB entry 6YUF).

F. *S. cerevisiae* Condensin DNA exit gate in the apo non-engaged conformation (PDB entry 6YVU).

G. *S. cerevisiae* Condensin DNA exit gate in the apo bridged conformation (PDB entry 6YVV).

Table S1. Thermodynamic parameters of ADP, ATP and ATPγS binding to SMC1ACC/RAD21C and SMC3CC/RAD21N measured by Isothermal Titration Calorimetry (ITC)

Cell		Syringe	Fitting model: HD + NucI → HD-NucI			Fitting model: HD + NucI → HD-NucI; HD-NucI + HD-NucI → HD-NucI homodimer			
			K_{d1} (μM)	ΔH_1 (kcal.mol^{-1})	N_1	K_{d1} (μM)	ΔH_1 (kcal.mol^{-1})	K_{d2} (nM)	ΔH_2 (kcal.mol^{-1})
SMC1ACC/RAD21C	wt	ADP	27.6 ± 7.6	2.78 ± 0.19	0.85 ± 0.03	-	-	-	-
SMC1ACC/RAD21C	wt	ATP	51.0 ± 2.58	3.17 ± 0.048	1.00 ± 0.01	-	-	-	-
SMC1ACC/RAD21C	wt	ATPγS	31.6 ± 1.00	1.89 ± 0.014	1.21 ± 0.01	-	-	-	-
SMC1ACC/RAD21C	E1157Q	ATP	55.3 ± 4.72	1.42 ± 0.038	1.05 ± 0.01	-	-	-	-
SMC1ACC/RAD21C	E1157Q	ATPγS	43.0 ± 3.06	1.06 ± 0.021	1.01 ± 0.01	-	-	-	-
SMC3CC/RAD21N	wt	ADP	1.8 ± 0.07	-7.19 ± 0.02	0.98 ± 0.02	-	-	-	-
SMC3CC/RAD21N	wt	ATP	-	-	-	51.1 ± 23.7	-2.7 ± 1.2	1.4 ± 0.7	-18 ± 0.8
SMC3CC/RAD21N	wt	ATPγS	-	-	-	18.3 ± 1.5	-6.0 ± 1.9	5.3 ± 0.6	-7.4 ± 1.8
SMC3CC/RAD21N	E1144Q	ATP	-	-	-	74.6 ± 33.8	-5.2 ± 1.0	2.8 ± 1.8	-14.9 ± 0.6
SMC3CC/RAD21N	E1144Q	ATPγS	-	-	-	35.5 ± 9.2	-5.6 ± 0.8	3.8 ± 1.0	-8.0 ± 0.7

Table S2. Crystallization conditions and crystallographic statistics for the SMC1ACC/RAD21C complexes.

SMC1ACC/ RAD21C	WT apo	EQ mutant apo	WT loop
PDB entry	**8A77**	**8A72**	**8A79**
Crystallization conditions	0.1 M Hepes pH 7.0; 10% w/v PEG 6000	0.2 M sodium malonate dibasic monohydrate; 0.1 M Bis-Tris propane pH 6.5; 20% w/v PEG 3350	0.1 M Hepes pH 7.0; 10% w/v PEG 6000
Data collection*			
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 (%)	98.6 (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.*

Table S2. Crystallization conditions and crystallographic statistics for the SMC1ACC/RAD21C complexes (continued).

SMC1ACC/ RAD21C	EQ mutant Loop	EQ mutant ADP	EQ mutant ATPγS-Mg
PDB entry	**8A75**	**8A6K**	**8A6Y**
Crystallization conditions	0.2 M sodium bromide; 0.1 M Bis-Tris propane pH 6.5; 20% w/v PEG 3350	0.2 M ammonium formate; 20% w/v PEG 3350	0.1 M MMT pH 7.0; 25% w/v PEG 1500
Data collection*			
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	6.8 (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	6469	6578	3312
Ligand/ion	19	54	32
Water	389	64	32
B-factors ($Å^2$)			
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 (°)	0.886	0.965	1.024

** Values in parentheses are for the highest-resolution shell.*

Table S3. Crystallization conditions and crystallographic statistics for the SMC1ACCsh/RAD21C complexes.

SMC1ACCsh/ RAD21C	EQ mutant Apo	EQ mutant Loop	EQ mutant ADP
PDB entry	**8A80**	**8A81**	**8A7W**
Crystallization conditions	0.2 M ammonium nitrate pH 6.3; 20% w/v PEG 3350	0.2 M sodium sulfate; 20 % w/v PEG 3350	0.2 M sodium malonate; 0.1 M Bis Tris propane pH 6.5; 20 % w/v PEG 3350
Data collection*			
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 (°)	0.809	0.831	0.833

** Values in parentheses are for the highest-resolution shell.*

Table S3. Crystallization conditions and crystallographic statistics for the SMC1ACCsh/RAD21C complexes (continued).

SMC1ACCsh/ RAD21C	EQ mutant ADP-Mg	EQ mutant AGS-Mg
PDB entry	**8A7M**	**8A7Y**
Crystallization conditions	0.2 M magnesium chloride; 0.1 M MES pH 6.0; 20 % w/v PEG 6000	0.2 M sodium formate; 0.1 M Bis Tris propane pH 6.5; 20 % w/v PEG 3350
Data collection*		
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 (°)	0.829	0.881

** Values in parentheses are for the highest-resolution shell.*

Table S4. Crystallization conditions and crystallographic statistics for the SMC3CC/RAD21N complexes.

SMC3CC/ RAD21N	EQ mutant Apo	EQ mutant ADP	WT ADP-Mg
PDB entry	**8A87**	**8A88**	**8A8B**
Crystallization conditions	0.21M $NaSO_4$; 0.1M Bis-tris propane pH7; 16% w/v PEG3350	0.1 M Na/K phosphate pH 7.5; 0.1 M HEPES pH 7.5; 15% v/v PEG Smear High; 10% v/v ethylene glycol	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
Data collection*			
Space group	$P4_1\,2_1\,2$	$P4_1\,2_1\,2$	$P4_1\,2_1\,2$
Cell dimensions			
a, b, c (Å)	90.79, 90.79, 236.87	89.95, 89.95, 233.91	89.71, 89.71, 234.65
α, β, γ (°)	90.00, 90.00, 90.000	90.00, 90.00, 90.000	90.00, 90.00, 90.000
Resolution (Å)	50.00 – 2.60 (2.76 – 2.60)	50.00 – 2.45 (2.60 – 2.45)	50.00 – 3.00 (3.18 – 3.00)
Rsym or Rmerge	0.280 (3.842)	0.154 (2.621)	0.247 (3.684)
I / σI	14.93 (1.02)	22.60 (1.50)	16.53 (1.11)
Completeness (%)	99.9 (99.7)	99.9 (99.7)	99.9 (99.6)
Redundancy	26.2 (25.7)	26.1 (26.4)	26.0 (26.9)
CC(1/2)	99.9 (36.6)	99.9 (58.1)	99.9 (43.1)
Refinement			
Resolution (Å)	49.60 – 2.60	49.03 - 2.45	49.27 – 3.00
No. reflections	31408	36209	19996
Rwork / Rfree	0.216 / 0.255	0.195 / 0.239	0.196 / 0.264
Number of atoms			
Protein	4051	3918	3931
Ligand/ion	21	27	28
Water	9	110	4
B-factors ($Å^2$)			
Protein	71.85	74.80	96.65
Ligand/ion	93.20	55.69	79.86
Water	54.47	62.44	71.39
R.m.s. deviations			
Bond lengths (Å)	0.008	0.008	0.009
Bond angles (°)	0.899	0.940	1.080

** Values in parentheses are for the highest-resolution shell.*

Table S4. Crystallization conditions and crystallographic statistics for the SMC3CC/RAD21N complexes (continued).

SMC3CC/ RAD21N	EQ mutant ATPγS-Mg (form 1)	EQ mutant ATPγS-Mg (form 2)
PDB entry	**8A89**	**8A8A**
Crystallization conditions	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	0.1 M Sodium chloride; 0.1 M Bicine pH 9.0; 20% w/v PEG 550 MME
Data collection*		
Space group	C2 2 2_1	$P6_4$ 2 2
Cell dimensions		
a, b, c (Å)	118.24, 154.46, 136.54	201.07, 201.07, 92.38
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	50.00 – 2.25 (2.39 - 2.25)	50.00 – 3.11 (3.30 – 3.11)
Rsym or Rmerge	0.098 (2.233)	0.256 (4.814)
I / σI	20.32 (1.34)	17.00 (1.03)
Completeness (%)	98.0 (94.2)	99.5 (98.1)
Redundancy	13.6 (13.1)	39.1 (39.1)
CC(1/2)	100.0 (81.6)	99.9 (63.7)
Refinement		
Resolution (Å)	47.21 – 2.25	44.64 – 3.11
No. reflections	58103	20146
Rwork / Rfree	0.202 / 0.253	0.187 / 0.238
Number of atoms		
Protein	7529	3955
Ligand/ion	64	32
Water	90	-
B-factors ($Å^2$)		
Protein	75.14	112.95
Ligand/ion	47.68	93.80
Water	56.96	-
R.m.s. deviations		
Bond lengths (Å)	0.008	0.008
Bond angles (°)	0.927	0.997

** Values in parentheses are for the highest-resolution shell.*

Table S5. Cryo-EM data collection, refinement and validation statistics. PDB entry: 8A8P.

Data collection and processing	
Microscope and Camera	Glacios™ cryo-TEM - K2 Summit
Magnification	45,000x
Voltage (kV)	200
No. of micrographs	4146
Electron exposure ($e^-/Å^2$)	44.42
Defocus range	-0.8 to -2.0
Pixel size (Å)	0.901
Symmetry imposed	C1
Initial particle images (no.)	399,773
Final particle images (no.)	110,099
Map resolution (Ă)	4.04
FSC threshold	0.143
Map resolution range (Å)	3.0 - 10.4
Refinement	
Initial model used	SMC1ACC/RAD21C and SMC3CC/RAD21N crystal structures
Model resolution range (Å)	4.0
FSC threshold	0.143
Map sharpening *B* factor ($Ă^2$)	-208.6
Model composition	
Non-hydrogen atoms	5830
Protein residues	727
Ligands	2 ATP
B factors ($Å^2$)	
Protein	89.78
Ligand	67.61
R.m.s. deviations	
Bond lengths	0.003
Bond angles	0.582
Validation	
MolProbity score	2.23
Clashscore	16.16
Poor rotamers (%)	0
Ramachandran plot	
Favored (%)	91.05
Allowed (%)	8.95
Outlier (%)	0

1.2. Discussion

1.2.1. Caractéristique et rôle de l'engagement des têtes ATPases humaines de la Cohésine

1.2.1.1. L'engagement des têtes humaines nécessite la contribution des deux sites ATPases et est dépendant de l'ATP

L'engagement des têtes ATPases de SMC1A et de SMC3 est induit par la fixation de l'ATP. L'hydrolyse entraine la dissociation des têtes. Il est donc nécessaire d'utiliser des protéines SMC1A et SMC3 mutées au niveau de leur site catalytique afin d'obtenir le module ATPase sous une forme engagée stable. Ces mutations doivent affecter l'hydrolyse de l'ATP, mais ne pas perturber la liaison de l'ATP aux têtes ou l'engagement des têtes. Plusieurs mutants SMC1A et SMC3 de *S. cerevisiae* présentant ces caractéristiques ont été décrits dans la littérature (Elbatsh et al., 2016). Nous retrouvons notamment la mutation de l'acide glutamique catalytique en glutamine du Walker B (EQ), la mutation de la leucine en valine du motif signature (LV) et la mutation de l'acide aspartique en acide glutamique dans la boucle D (DE). Plus précisément, le résidu E du Walker B est un résidu catalytique participant à l'hydrolyse de l'ATP. Le résidu L du motif signature est proche du sucre de l'ATP et est important pour l'engagement des têtes. Le résidu D de la boucle est proche du γ-phosphate de l'ATP et participe à l'alignement de la molécule d'eau nécessaire à l'hydrolyse de l'ATP. Ces trois mutations chez la levure permettent la fixation de l'ATP, l'engagement des têtes et présentent une activité ATPase réduite similaire à celle obtenue par les mutations EQ (Elbatsh et al., 2016).

J'ai donc reproduit ces mutations pour les protéines humaines SMC1A (WT, E1157Q, D1163E, L1128V) et SMC3 (WT, E1144Q, L1115V, D1150) et testé l'engagement des têtes en présence d'ATP par chromatographie d'exclusion de taille, selon plusieurs combinaisons décrites dans les données supplémentaires du manuscrit (Figure S6A et S6D du manuscrit, p. 207).

Seul l'engagement des mutants EQ de SMC1A et SMC3 a été observé, mettant en évidence qu'il s'agit de la seule combinaison possible afin d'obtenir un complexe engagé stable. La mutation du résidu E catalytique des têtes ATPases et son impact direct sur l'hydrolyse de l'ATP sont donc suffisants pour former et stabiliser un complexe engagé. Bien

que l'activité d'hydrolyse des mutants LV et DE soient similaires à celle des mutants EQ, les mutants LV ou DE n'engendrent pas un complexe des têtes engagées suffisamment stable. Cela peut notamment s'expliquer par le fait que ces mutants LV et DE, lors de la fixation de l'ATP, entrainent des réarrangements structuraux induisant un engagement des têtes plus instable que celui observé par les mutants EQ. Il serait également envisageable que les mutants LV et DE, au moment de la fixation de l'ATP, miment un état proche de celui observé lors de l'hydrolyse de l'ATP, favorisant le désengagement des têtes. Cela corroborerait les données *in vitro* obtenues chez *S. cerevisiae*, qui montrent que ces mutants LV et DE affectent de manière moindre la cohésion des chromatides sœurs et la viabilité de *S. cerevisiae*, comparativement aux mutants EQ (Elbatsh et al., 2016).

De plus, les expériences menées sur SMC1A et SMC3 WT et EQ démontrent que la prévention de l'hydrolyse sur uniquement un des deux sites catalytiques n'est pas suffisante pour maintenir de manière stable la dimérisation des têtes ATPases (Figure 7 et S6A du manuscrit, p. 187 et 207). La mutation EQ sur les deux têtes ainsi que la liaison de deux molécules d'ATP sont nécessaires pour l'engagement stable des têtes de SMC1A et de SMC3.

Par ailleurs, afin d'améliorer la stabilité, des expériences de reconstitution du module ATPase SMC1A EQ - SMC3 EQ engagé ont été menées en présence d'ATP et de deux analogues faiblement hydrolysables de l'ATP, l'ATP-γ-S et l'AMP.PNP (Figure S6B du manuscrit, p. 207). Dans l'ATP-γ-S, un des atomes d'oxygène du γ-phosphate a été remplacé par un atome de soufre et dans l'AMP.PNP, l'atome d'oxygène liant les β- et γ- phosphates a été remplacé par un atome d'azote. Les résultats de SEC et d'AUC-SV démontrent que seule la liaison d'ATP aux têtes ATPases supporte l'engagement des têtes. Les analogues faiblement hydrolysables de l'ATP ne permettent donc pas la dimérisation des têtes, bien qu'ils puissent se fixer aux têtes isolées de SMC1A et SMC3 (données confirmées par ITC pour l'ATP-γ-S). Toutefois, le remplacement d'un atome d'oxygène par un azote pour l'AMP.PNP et par un soufre pour l'ATP-γ-S entraine la formation d'un angle du triphosphate différent de celui d'un ATP. Il pourrait alors se former une gêne stérique entre l'analogue et les têtes ATPases de SMC1A et SMC3, empêchant leur dimérisation. Cela concorde avec la littérature, qui révèle qu'en cas d'utilisation d'ATP-γ-S ou d'AMP.PNP, il a été nécessaire de recourir à la réticulation *in vitro* aspécifique ou spécifique, afin de stabiliser les têtes ATPases dans une conformation engagée (Muir et al., 2020; Shi et al., 2020). Alors qu'au contraire, la stabilisation des têtes engagées

par la présence d'ATP et l'utilisation des mutants EQ a été suffisante pour stabiliser cet état engagé (Collier et al., 2020).

1.2.1.2. L'engagement des têtes ATPases induit une flexibilité de la porte de sortie SMC3/RAD21 chez l'homme

La porte de sortie de l'ADN, formée par l'interface SMC3/RAD21, est maintenue dans une conformation de repos avant et pendant la liaison de l'ATP à la tête de SMC3 (observations basées sur les structures cristallines de la tête de SMC3 sous formes apo et liée à l'ATP-γ-S obtenues au sein de mon équipe d'accueil). A ce moment, RAD21, à travers son petit domaine hélical N-terminal, interagit de manière extensive avec le domaine globulaire de SMC3, en plus de son interaction avec le cou de SMC3. Lors de l'engagement des têtes de SMC1A et SMC3 en présence d'ATP, nous observons un effet de levier de la deuxième hélice couplée au motif signature de SMC3, également observable dans la conformation productive de la Cohésine. La densité de notre carte cryoEM s'arrêtant à ce niveau-là, nous n'avons pas pu reconstruire au-delà de ce domaine, la région en superhélice de SMC3 et de RAD21 (Figure 7C du manuscrit, p. 187). Cette absence de densité suggère une flexibilité de cette porte de sortie de l'ADN.

Pour confirmer cette flexibilité, nous avons réalisé des expériences de réticulation *in vitro,* en identifiant des résidus de SMC3 et de RAD21, qui une fois mutés en cystéine sont suffisamment proches lors de l'état de repos pour être réticulés par du BMOE. Ce dernier réticule les cystéines éloignées d'environ 8 Å. Au contraire, lorsque le complexe est sous sa forme productive, les résidus sont trop éloignés l'un de l'autre pour permettre une réticulation. Il s'agit des résidus des paires SMC3CC-D120/RAD21N-K25 et SMC3-D92/RAD21N-K26, qui sont séparés respectivement de 8,4 Å et de 9,5 Å à l'état de repos et séparés de 14 Å et 25 Å à l'état productif. La réticulation *in vitro* de la première paire présente des résultats similaires dans l'ensemble des conditions testées en absence ou présence de SMC1A, d'ATP et d'ADN, et ne permet pas de discriminer la présence des différents états (Figure 7I du manuscrit, p. 187). Cela peut provenir du fait que la distance de cette paire observée pour le complexe productif est trop proche de l'intervalle de distance entre les cystéines que peut lier le BMOE, qui est compris entre 6 Å et 11 Å (Green et al., 2001). Toutefois, la réticulation de ces résidus met en évidence que l'interface SMC3/RAD21 n'est

pas perdue lors de l'engagement des têtes en présence d'ATP et d'ADN. Ce qui corrobore nos données de reconstitution du module engagé en présence d'ATP par SEC, où les quatre protéines (SMC1A-HD, SMC3-HD, RAD21C et RAD21N) sont présentes (Figure 7A du manuscrit, p. 187). La deuxième paire identifiée SMC3-D92/RAD21N-K26 semble donc être mieux adaptée pour l'étude de cette flexibilité. En effet, nos résultats montrent que l'engagement des têtes et la fixation de l'ADN entrainent une diminution de la position de repos, correspondant à la libération de l'état de repos de cette région (Figure 7J du manuscrit, p. 187). Ainsi, nos essais de réticulation ont mis en évidence que lors de l'engagement des têtes, l'état de repos de la porte de sortie de l'ADN est délivré et devient flexible, mais ne parvient pas à adopter une conformation productive stable. De plus la liaison entre SMC3 et RAD21 est maintenue, n'entrainant pas l'ouverture de cette porte de sortie.

Nos travaux mettent ainsi en lumière la flexibilité de la porte de sortie, jamais décrite à ce jour et qui permet de préparer la Cohésine à son interaction avec l'ADN, en libérant cette région de son état de repos.

1.2.1.3. La porte de sortie de l'ADN dans l'interface SMC3/RAD21 semble avoir un comportement différent selon les espèces

Nos travaux sur la caractérisation structurale et de réticulation *in vitro* mettent en évidence l'implication de l'engagement des têtes ATPases de SMC1A et de SMC3 humaine dans la flexibilité de la porte de sortie de l'ADN chez l'homme. La Cohésine semble donc adopter un comportement différent par rapport aux autres espèces. En effet, Muir et al ont obtenu la reconstruction d'un module ATPase chimérique en présence d'ATP-γ-S et à l'aide de réticulation *in vitro* à une résolution de 3,3 Å. Ce complexe est composé de la tête ATPase de Smc1 de *C. thermophilum*, des extrémités N- et C-terminales de Scc1 de *S. cerevisiae* et de la tête ATPase de Smc3 de *S. cerevisiae*. Les domaines N- et C-terminaux de Smc1 ne sont pas reliés entre eux, comme dans nos constructions : le domaine N-terminal de Smc1 (Smc1N) est exprimé seul, alors que le domaine C-terminal (Smc1C) est relié à l'extrémité N-terminale de Scc1 (Scc1N).

Contrairement à notre structure cryoEM du module ATPase humain, la région en superhélice proche de la tête de Smc3 a pu être construite, la densité à ce niveau étant un peu mieux définie. Toutefois, aucune densité n'a été observée pour Scc1-N, qui est reliée à

Smc1C. Afin de confirmer l'ouverture de l'interface Smc3/Scc1, cette équipe a tout d'abord vérifié par chromatographie d'exclusion de taille la présence de Scc1 lors de la purification, en amont de l'étape permettant d'obtenir le complexe engagé et donc en absence de nucléotide (ATP ou ATP-γ-S). Muir et al ont mené des essais de rétention et ont montré qu'après clivage de Smc1 et Scc1, lors de l'engagement des têtes induit par l'ajout d'ATP ou d'ATP-γ-S, une diminution de Scc1 est observée et peut correspondre à l'ouverture de l'interface Smc3 et Scc1, par comparaison à la condition sans nucléotide. Toutefois, les résultats sont à prendre avec précaution, du fait qu'il s'agisse de la reconstruction d'un complexe chimérique. De plus, il semble que Muir et al ont utilisé des protéines SMC1A et SMC3 WT. Or si le complexe présente potentiellement des caractéristiques similaires à celles décrites pour le complexe humain, il se peut que l'engagement des têtes WT induit par liaison d'ATP ou d'ATP-γ-S ne soit pas suffisamment stable. La perte de Scc1N serait donc associée à la dissociation des têtes de Smc1 et Smc3 lors de ces essais de rétention. Cette flexibilité reste donc à confirmer sur un complexe de levure non chimérique.

Cette différence potentielle inter-espèce sur la porte de sortie de l'ADN ouvre de nombreuses questions et plusieurs hypothèses peuvent être proposées. La première étant que cette porte adopte un comportement différent à la suite de l'engagement des têtes, sa flexibilité pour l'homme et son ouverture chez la levure, renforçant notre hypothèse de travail selon laquelle la Cohésine présente un mode d'action et une dynamique différente selon les espèces. Les deux structures du module ATPase engagé pourraient également correspondre à des étapes différentes du cycle de régulation de la Cohésine. Le module ATPase humaine semble correspondre à l'étape préliminaire à la fixation de NIPBL et de l'ADN, alors que le module ATPase engagé chimérique pourrait correspondre à l'étape finale de dissociation du complexe de l'ADN induite par WAPL et entrainant l'ouverture de cette porte de sortie.

Ainsi, plusieurs hypothèses concernant cette porte de sortie de l'ADN sont possibles et restent à être caractérisées précisément. Il serait également intéressant d'étudier cette flexibilité/ouverture chez d'autres espèces eucaryotes, telles que S. *pombe* et la drosophile.

1.2.2. Proposition d'un modèle décrivant les premières étapes du cycle de la Cohésine humaine, allant de la liaison de l'ATP à l'interaction de l'ADN et NIPBL

L'ensemble de nos structures obtenues au sein mon équipe d'accueil (structures cristallographiques des complexes SMC1A/RAD21-C et SMC3/RAD21-C sous plusieurs formes apo et associées à l'ATP-γ-S et à l'ADP, structure cryoEM du module ATPase engagé), ainsi que les structures du complexe Cohésine-NIPBL-ADN obtenues par Shi et al, nous permettent de caractériser en termes moléculaires les mouvements opérés lors du cycle ATPase de la Cohésine humaine. Pour plus de clarté, la conformation adoptée par le complexe Cohésine-NIPBL-ADN est évoquée comme étant la conformation productive.

1.2.2.1. La liaison de l'ATP maintient les têtes ATPases de SMC1A et SMC3 dans un état de repos, mais propre à chaque tête

Les structures cristallographiques des complexes humains SMC1A/RAD21-C et SMC3/RAD21-C, sous une forme apo et associées à l'ATP-γ-S, obtenues dans mon équipe d'accueil, nous permettent de définir les mouvements opérés lors de la liaison de l'ATP.

La liaison de l'ATP-γ-S à la tête de SMC1A entraine un réarrangement de certains motifs catalytiques, ainsi qu'une modification globale rapprochant le lobe RecA du lobe hélical (Figure 36A).

En effet, nous observons l'ouverture du Walker A (ou boucle P) induite par la libération de molécules d'eau et permettant à cette boucle d'interagir avec le nucléotide. Cela a également été confirmé par nos données d'ITC, où la liaison de l'ATP-γ-S implique une contribution entropique, provenant de la libération des molécules d'eau. La liaison de l'ATP-γ-S à la tête de SMC1A entraine également des réarrangements structuraux de la boucle R, lui permettant d'interagir avec le sucre, ainsi qu'avec l'α-phosphate du nucléotide. Cette boucle R permet alors de stabiliser le nucléotide au niveau du site catalytique de la tête ATPase.

Par ailleurs, le rapprochement du lobe hélical et du lobe Rec-A, suite à la liaison de l'ATP-γ-S, est induit par l'ouverture du Walker A, ainsi que par le repositionnement de la boucle Q, qui interagit avec un ion Mg^{2+} et avec le γ-phosphate du nucléotide. Ces

mouvements permettent de préparer et d'adopter une conformation compatible avec l'engagement des têtes.

Ainsi la liaison de l'ATP-γ-S à la tête de SMC1A entraine majoritairement un réarrangement global de sa tête, avec quelques réarrangements structuraux locaux permettant la stabilisation de l'ATP jusqu'à l'engagement des têtes. Cette liaison ne permet pas d'activer la tête de SMC1A, mais la maintient sous un état inactif, dit relaxé.

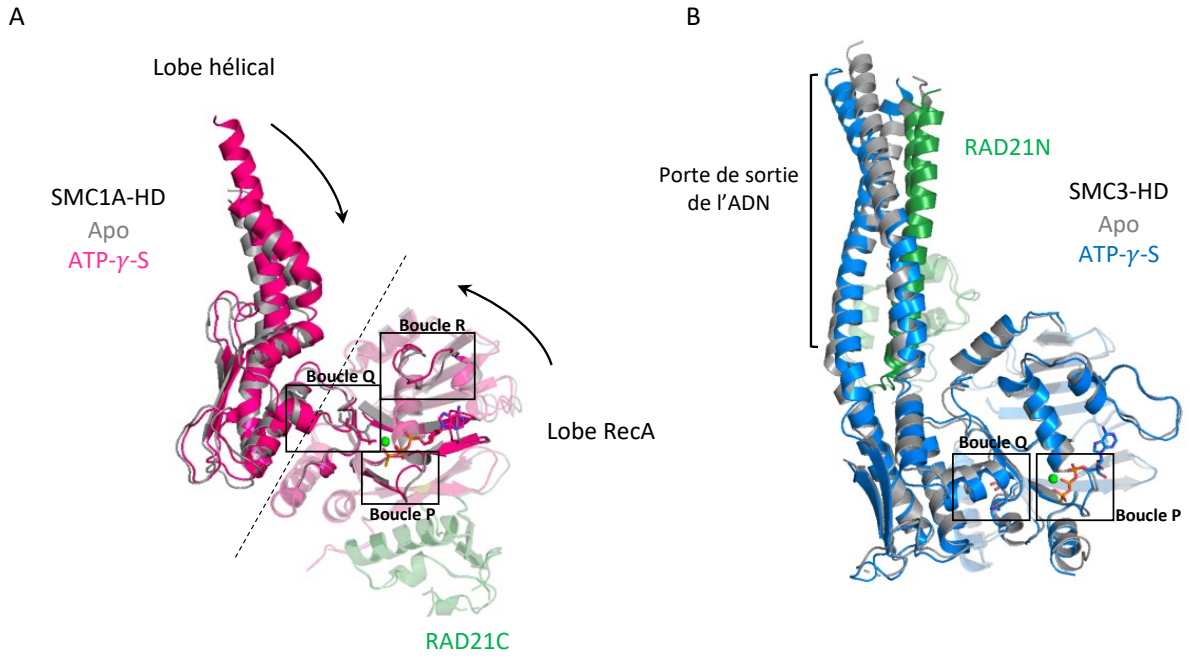


Figure 36 : Changements conformationnels opérés lors de la liaison de l'ATP-γ-S au niveau des têtes de SMC1A (A) et SMC3 (B).
A. La liaison de l'ATP-γ-S entraine un rapprochement des lobes hélical et RecA, ainsi que l'ouverture de la boucle P et le réarrangement des boucles R et Q. Toutefois, la tête de SMC1A est maintenue dans un état inactif.
B. Très peu de mouvements au niveau des motifs catalytiques et à un niveau plus global sont observés lors de la liaison de l'ATP-γ-S au niveau de la tête de SMC3. Cette dernière est déjà préformée pour sa liaison avec l'ATP et est maintenue sous une forme inactive. La porte de sortie de l'ADN adopte également un état de repos.

Concernant la tête ATPase humaine de SMC3, les formes apo et liée à l'ATP-γ-S, présentent une forte similarité conformationnelle (Figure 36B). Contrairement à SMC1A, très peu de mouvements sont observés dans le site catalytique lors de la liaison du nucléotide. Le Walker A est déjà ouvert avant la fixation du nucléotide. La boucle Q reste éloignée du site de liaison, même après la fixation du nucléotide. Il semble donc que SMC3 soit déjà préformé pour son interaction avec l'ATP-γ-S et est maintenu donc une conformation inactive. De plus, dans les formes apo et liée à l'ATP-γ-S, la porte de sortie de l'ADN, décrite précédemment, est

stabilisée dans un état de repos, RAD21 interagissant de manière extensive avec le domaine globulaire de SMC3.

La liaison de l'ATP aux têtes humaines de SMC1A et SMC3 induit très peu de changements conformationnels. Les têtes sont alors gardées sous une conformation inactive avec toutefois certaines caractéristiques propres à chaque tête. Quelques réarrangements structuraux sont nécessaires à SMC1 pour interagir avec le nucléotide et le préparer à l'engagement, alors que SMC3 semble adopter une conformation permettant directement la fixation de l'ATP et son engagement.

1.2.2.2. L'engagement des têtes entraine la semi-activation des têtes et la flexibilité de l'interface SMC3/RAD21

La comparaison de nos structures SMC1A/RAD21-C et SMC3/RAD21-N indépendantes, ainsi que celle du module ATPase engagé obtenue par cryoEM nous permet de mieux comprendre la dynamique structurale induite par l'engagement des têtes humaines. L'engagement des têtes n'entraine pas de changements structuraux majeurs au sein des têtes, mais des réarrangements plus globaux (Figure 37)

Pour la tête humaine de SMC1A, cet engagement entraine un effet de levier du lobe hélical par rapport au lobe RecA, induit par le glissement de la seconde hélice couplée au motif signature de SMC1A. Cet effet levier conduit à (i) la réorganisation de l'interface entre les deux lobes, ainsi qu'à (ii) un mouvement de la région en superhélice de SMC1A et (iii) au réarrangement du motif signature et de la boucle F, qui semble être essentielle pour la liaison de NIPBL. La boucle Q de SMC1A est maintenue à distance du nucléotide, avec notamment le résidu Q137 non orienté vers l'ATP, mais tourné vers le motif signature de SMC3. L'engagement des têtes ne permet donc pas l'activation de SMC1A.

Concernant la tête humaine de SMC3, l'engagement des têtes induit également un effet de levier au niveau de la seconde hélice couplée au motif signature, qui est le dernier domaine observable dans la densité de la carte du module ATPase engagé et qui a pu être reconstruit (Figure 37B). Toutefois, cet effet de levier est également retrouvé dans la structure du complexe productif, confirmant l'induction de ce levier lors de l'engagement des têtes. Cet effet entraine également le repositionnement de la boucle Q, avec notamment le résidu Q141, qui est alors orienté vers le site de liaison de l'ATP au contraire de SMC1A. Nous observons

également le réarrangement de la boucle R, ainsi que celui de la région ¨wedge¨, située à l'interface supérieure du lobe RecA et du lobe hélical (Figure 37B). L'engagement des têtes permet ainsi d'activer la tête ATPase SMC3, mais pas celle de SMC1A, jouant un rôle de régulateur/contrôleur dans l'activité ATPase de la Cohésine.

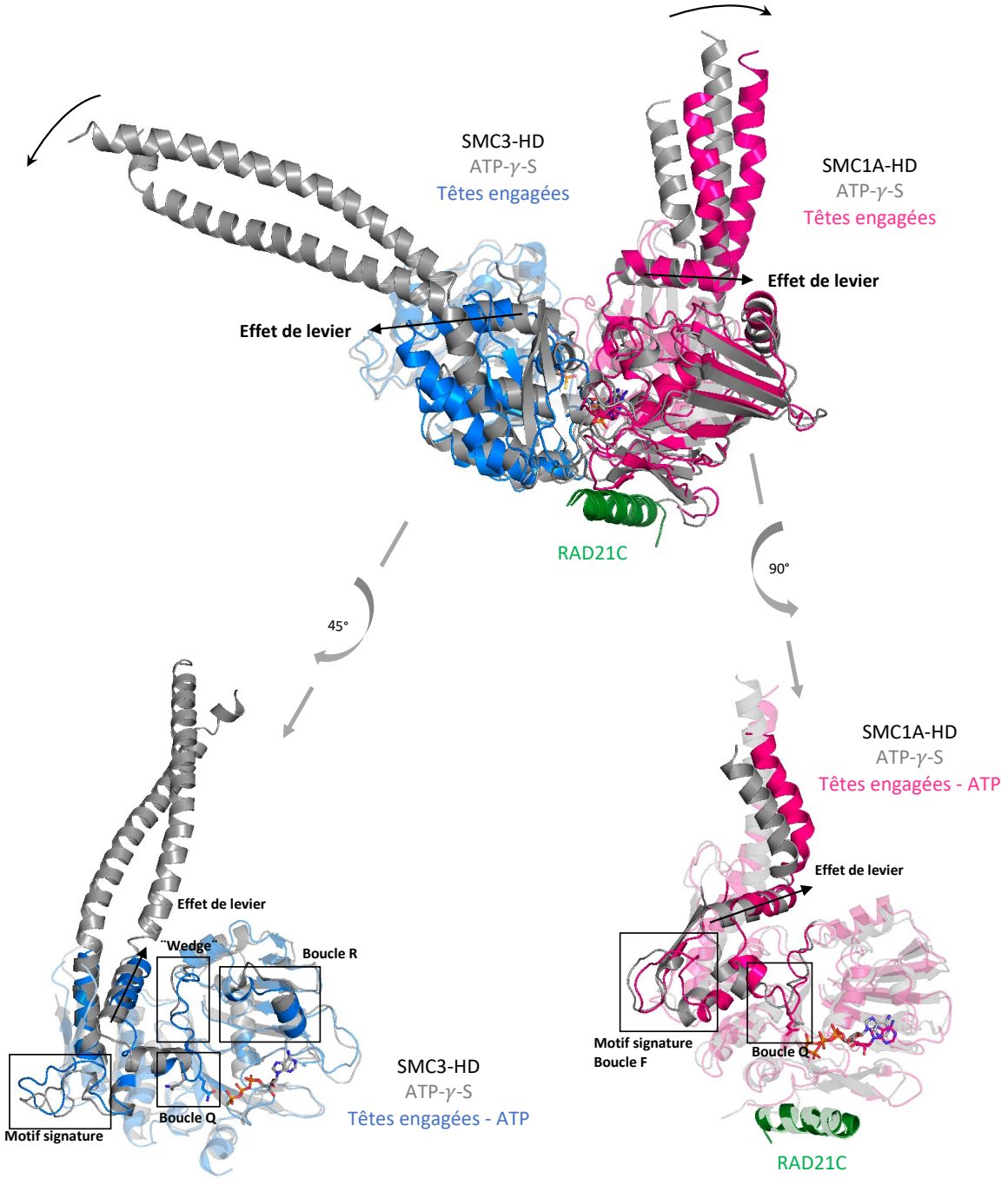


Figure 37 : Changements conformationnels induits lors de l'engagement des têtes de SMC1A et SMC3.
L'engagement des têtes entraine quelques réarrangements structuraux induits par l'effet de levier observés au niveau des hélices couplés aux motif signature de chaque tête. Cela conduit aux repositionnements de certains motifs catalytiques (encadrées en noir) et à la flexibilité de la porte de sortie de l'ADN. L'engagement des têtes permet l'activation de la tête ATPase de SMC3, mais maintient la tête de SMC1A dans un état inactif.

L'engagement des têtes entraine également la libération de la porte de sortie de l'ADN SMC3/RAD21 de son état de repos. Cette région devient flexible et peut alors interagir avec l'ADN, sans toutefois adopter la conformation productive (données confirmées par des expériences de réticulation *in vitro*). Ainsi, l'engagement des têtes participe à la préparation de l'interaction entre le module ATPase, l'ADN et NIPBL.

1.2.2.3. La fixation de l'ADN et de NIPBL entraine l'activation complète des têtes ATPases

La comparaison de notre structure du module ATPase engagé à la conformation de l'état productif humain (complexe Cohésine-ADN-NIPBL), nous permet de définir les changements conformationnels induits par la liaison de NIPBL et de l'ADN au niveau des têtes ATPases engagées (Figure 38 et Figure 7 du manuscrit, p. 187).

En superposant ces deux structures, nous observons que la liaison de l'ADN et de NIPBL favorise l'éloignement des deux têtes ATPases, ce qui conduit à une modification de leurs interfaces. En effet, NIPBL interagit à proximité de la boucle F de SMC1A et entraine l'écartement des deux lobes RecA de SMC1A et SMC3 et un alignement correct des sites enzymatiques. Cet éloignement permet également de modifier l'orientation de certains résidus. Notamment le résidu Q137 de la boucle Q de SMC1A est alors tourné vers le nucléotide, lorsque l'ADN et NIPBL sont associés aux têtes. La liaison de ces deux composants permet donc d'obtenir les deux sites catalytiques sous forme active, capables d'hydrolyser l'ATP.

De plus, NIPBL, stabilise le module ATPase, en interagissant avec les régions en superhélice de SMC1A et plus particulièrement avec celles de SMC3, permettant ainsi de stabiliser le module ATPase de la Cohésine dans une conformation productive. Cette conformation productive permet de stabiliser l'interaction entre l'ADN et la porte de sortie de l'ADN (interface SMC3/RA21).

Ainsi les changements conformationnels induits par l'engagement des têtes doivent être stabilisés par l'ADN et NIPBL, afin d'activer pleinement les têtes ATPases humaines. NIPBL et l'ADN n'induisent pas de modifications conformationnelles majeures, mais conduisent à un alignement correct des motifs ATPases. Ces résultats permettent ainsi de mieux comprendre comment NIPBL et l'ADN favorisent l'hydrolyse de l'ATP et l'importance de ces deux

composants pour cette activité. Nos résultats permettent ainsi d'expliquer pourquoi les têtes seules ont une faible activité ATPase, de même lors l'ajout de l'ADN seul ou de NIPBL seul, mais qu'une augmentation significative de l'hydrolyse de l'ATP est observée en présence d'ADN et de NIPBL (Davidson et al., 2019).

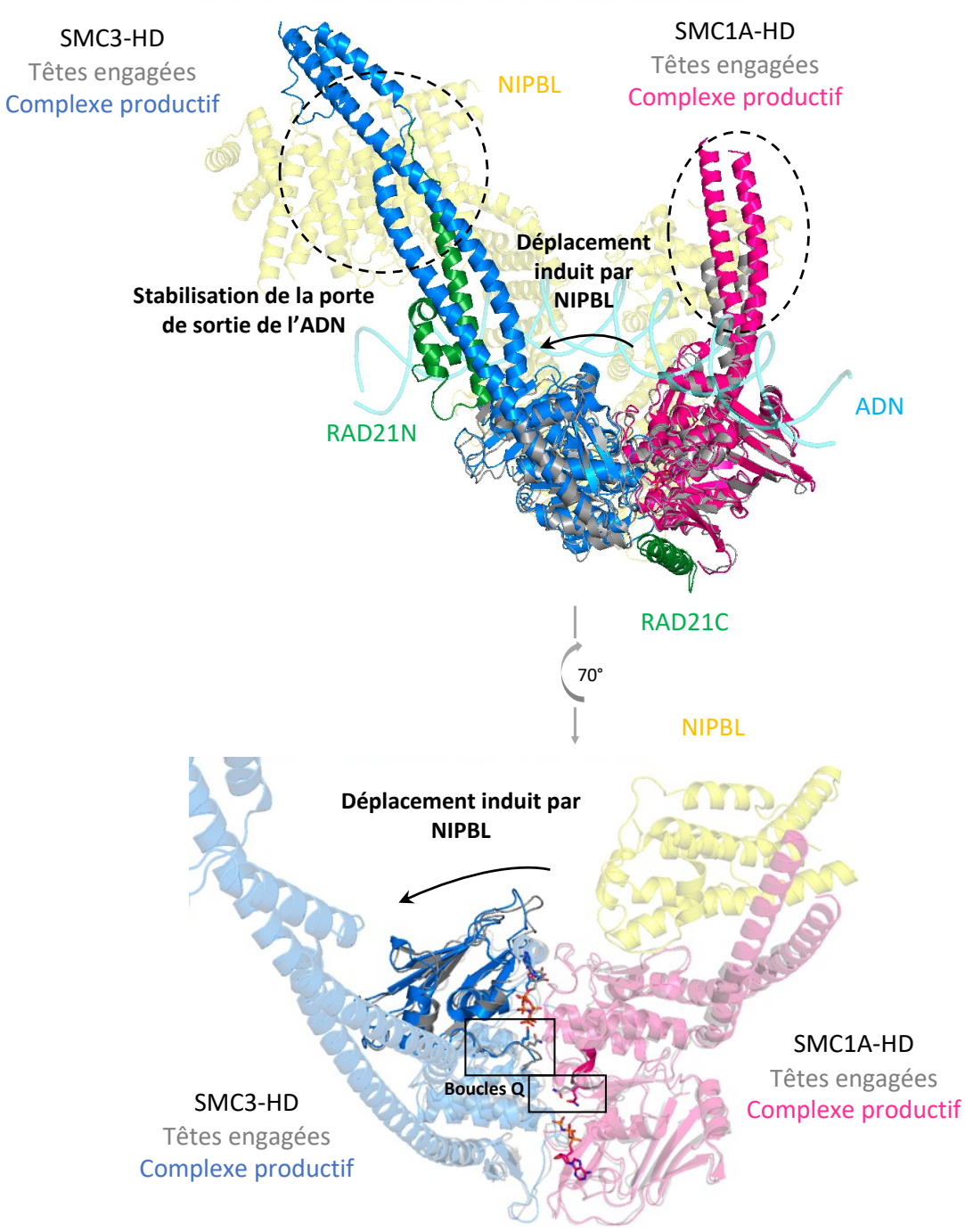


Figure 38 : Changements conformationnels induits par la liaison de NIPBL et de l'ADN au niveau des têtes ATPases de SMC1A et SMC3 engagées.

La fixation de NIPBL et de l'ADN au niveau des têtes entraine un éloignement des têtes permettant l'alignement correct des deux sites enzymatiques partagés par SMC1A et SMC3. NIPBL stabilise également les régions en superhélice de SMC1A et la porte de sortie de l'ADN. Cela permet ainsi de stabiliser le module ATPase de la Cohésine sous un état productif et d'activer pleinement les têtes ATPases.

Il est également intéressant de noter que la Cohésine humaine sous sa forme productive a été obtenue en présence d'AMP.PNP et a été réticulé *in vitro* de manière aspécifique. Or nos expériences sur le module ATPase mettent en évidence la nécessité d'utiliser de l'ATP pour obtenir un engagement stable des têtes, l'AMP.PNP induisant probablement une gêne pour l'interaction des phosphates avec les résidus des têtes, confirmée par nos expériences de reconstruction du module ATPase. De plus, Lacabanne et al ont montré que les analogues faiblement hydrolysables de l'ATP peuvent induire des conformations différentes selon une étude sur des transporteurs ABC, bien que ces analogues correspondent au même stade du cycle de l'ATP (pré-, post-hydrolyse) (Lacabanne et al., 2020). Il serait ainsi intéressant d'obtenir ce même complexe productif sous une forme liée à l'ATP, afin de voir si les mêmes changements ou non sont opérés en fonction du nucléotide lié.

Nos résultats permettent donc de mieux comprendre le rôle de NIPBL et de l'ADN dans l'activation des têtes ATPases de SMC1A et de SMC3 et de préciser les mécanismes moléculaires concernant les premières étapes du cycle de la Cohésine. Nous avons ainsi une meilleure compréhension de l'activité ATPase de la Cohésine, qui est essentielle pour l'ensemble de ses fonctions. Toutefois, il reste encore de nombreuses questions concernant cette activité ATPase, son rôle précis et sa médiation. Comment est-elle régulée précisément au cours du cycle de la Cohésine ? Est-ce que les protéines auxiliaires, autre que NIPBL, sont-elles impliquées dans cette régulation ? Notamment, est-ce que STAG1 ou STAG2, la protéine auxiliaire considérée comme la quatrième sous-unité de la Cohésine, influe-t-elle sur cette activité ATPase ? En effet STAG1 et STAG2 jouent un rôle primordial dans les fonctions de la Cohésine, toutefois leur mode d'action n'a pas été caractérisé précisément. De plus, au vu de la structure du complexe productif associé à STAG1, qui démontre que STAG1 participe également à la liaison de l'ADN et est très proche du module ATPase, il se pose la question de savoir si ce dernier a un impact sur ce module (Shi et al., 2020). Cela m'amène donc à mon prochain objectif qui a été la reconstitution du module ATPase associé à STAG1 ou à STAG2, afin de caractériser ce complexe.

II. Reconstitution du module ATPase associé à STAG1/STAG2

Afin de déterminer l'influence de STAG1 et STAG2 sur le module ATPase, j'ai mené des essais de reconstitution de ce complexe à quatre partenaires (SMC1A-HD, SMC3-HD, RAD21, STAG1 ou STAG2). Pour ce faire, j'ai utilisé les mêmes constructions décrites dans la partie précédente pour les têtes ATPases, à savoir les constructions SMC1ACC-EQ et SMC3CC-EQ. J'ai ensuite défini des constructions de différentes tailles pour RAD21, comprenant le domaine de reconnaissance pour STAG1 et STAG2, ainsi que les domaines de liaison essentiels aux deux autres protéines à motifs HEAT, PDS5 et NIPBL, afin de déterminer les constructions les plus adéquates en termes d'interaction et de stabilité.

J'ai ensuite mené des essais de reconstitution de ce complexe quadripartite, par deux approches. La première a consisté à exprimer et purifier de manière indépendante d'une part le complexe SMC1ACC-EQ/SMC3-EQ/RAD21 et, d'autre part la protéine STAG1 ou STAG2, pour ensuite reconstituer le complexe à quatre partenaires, en les mélangeant. La seconde approche a été de co-exprimer et de co-purifier les quatre partenaires en même temps.

1. Choix des constructions de RAD21 et de STAG1 et STAG2

1.2. Constructions de RAD21

Pour reconstituer le complexe quadripartite, en complément de la protéine entière, j'ai réalisé quatre nouvelles constructions de RAD21, comprenant les domaines qui interagissent avec les têtes de SMC1A, de SMC3 et de STAG1/STAG2. Il s'agit des régions N- et C-terminales, ainsi que de la région centrale de RAD21. Les constructions possèdent également le site de liaison de PDS5 et certaines celui de NIPBL.

RAD21 est une protéine prédite avec des domaines désorganisés, prédiction faite avec l'aide du programme PrDOS (Ishida and Kinoshita, 2007), excepté au niveau de ses régions N- et C-terminales, ainsi que de sa région centrale. Nous avons donc décidé de raccourcir certaines parties flexibles liant les régions d'interactions avec les protéines cœur et auxiliaires de la Cohésine et de les remplacer par des linkers différents selon les constructions. Le découpage de ces constructions a été basé sur les alignements de séquences de RAD21 de différents eucaryotes et selon la littérature, nous permettant d'identifier les zones

d'interactions de RAD21 avec les autres partenaires de la Cohésine (Figure 39). Le fait d'enlever les régions désorganisées permet d'améliorer la stabilité et la solubilité de la protéine en comparaison à la protéine entière.

Nous avons choisi de réaliser des constructions de tailles différentes, la plus petite de 371 acides aminés (RAD21-1) ne possède que les domaines de liaison de SMC3, de PDS5, de STAG1/STAG2 et de SMC1A et a donc le moins de régions désorganisées et donc de flexibilité. Les autres constructions RAD21-2, -3 et -4 possèdent tous les domaines de liaison à SMC1A, SMC3, STAG1/STAG2, PDS5 et à NIPBL, ainsi que des régions désorganisées plus ou moins longues selon la construction (Figure 40A). Les différentes constructions sont également schématisées dans le chapitre 2 - Matériels et méthodes.

1.3. Constructions de STAG1 et STAG2

J'ai également réalisé différentes constructions des protéines STAG1 et STAG2 humaines, dans le but d'obtenir une protéine plus stable et soluble par rapport aux protéines entières. Pour ce faire, je me suis basée sur les structures cristallines connues de ses protéines. Deux structures pour STAG1 sont disponibles, l'une correspondant à la région N-terminale et centrale allant du résidu 86 au résidu 420, et l'autre à un domaine plus central allant du résidu 459 au résidu 915. Une seule structure cristallographique de STAG2 est connue et décrit le domaine N-terminal et central allant du résidu 80 au résidu 1060. J'ai alors fait deux constructions de STAG1 et STAG2, prenant comme limites les résidus décrits ci-dessus. La figure 40B récapitule les constructions de STAG1 et STAG2 utilisées, également schématisées dans le chapitre 2 - Matériels et méthodes. Ces constructions ont ensuite été clonées dans des vecteurs d'expression natifs ou avec des étiquettes différentes en position N- ou C-terminale, telles que l'étiquette His_6 et GST en position N-terminale, et His_{10} en position C-terminale.

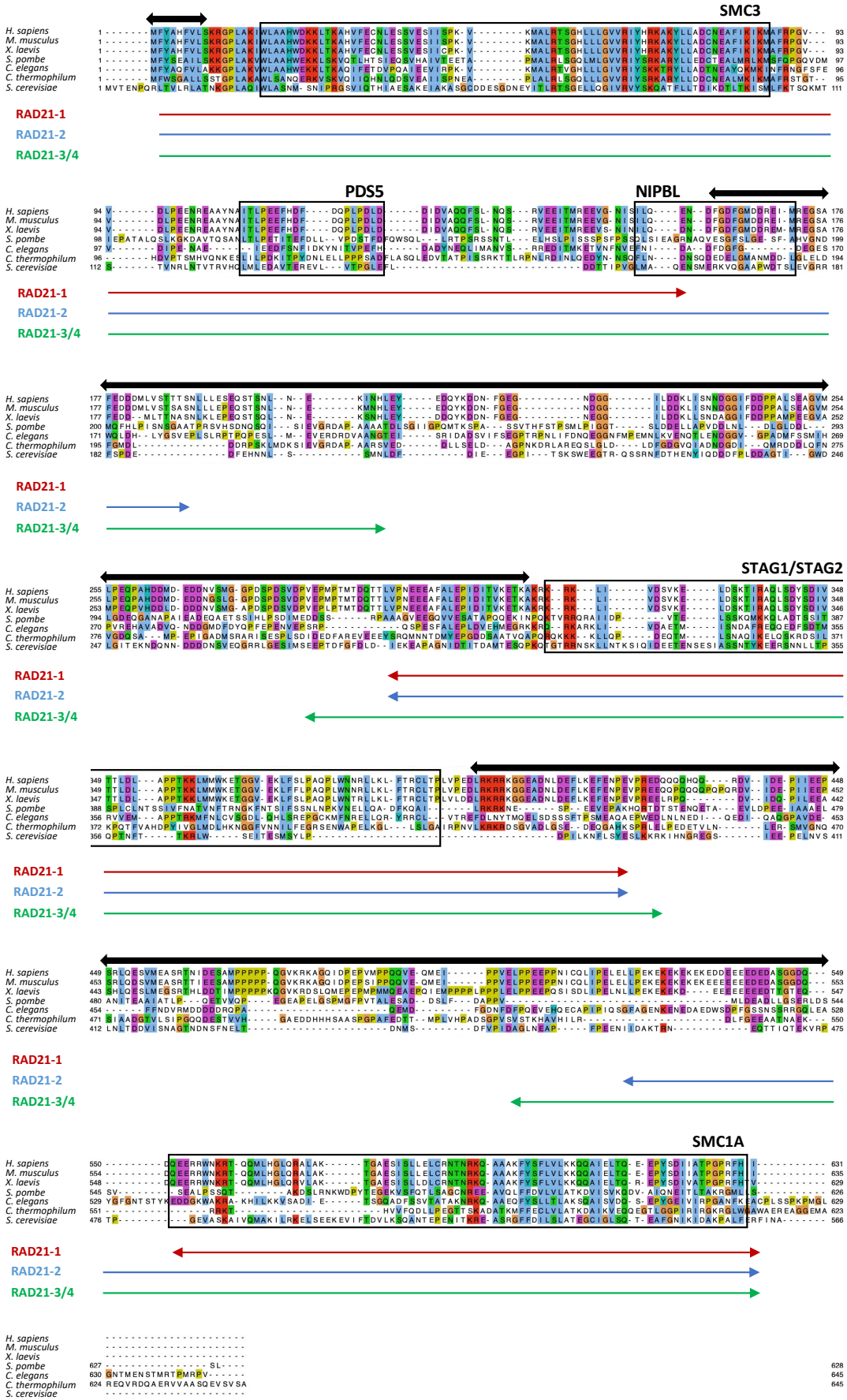

SMC3
PDS5
NIPBL
STAG1/STAG2
SMC1A
RAD21-1
RAD21-2
RAD21-3/4
H. sapiens
M. musculus
X. laevis
S. pombe
C. elegans
C. thermophilum
S. cerevisiae

Figure 39 : Alignement des séquences de RAD21 de divers organismes eucaryotes.
La conservation des séquences est représentée selon la coloration clustalX. Les régions désorganisées de RAD21 sont indiquées par les flèches noires à double sens. Les domaines de RAD21 interagissant avec les autres protéines cœur de la Cohésine, ainsi qu'avec ses protéines auxiliaires, sont indiqués par un encadré. La découpe des constructions de RAD21 a été basée sur cet alignement, ainsi que sur les données disponibles dans la littérature. Les limites des domaines des différentes constructions de RAD21 sont indiquées par les flèches en rouge pour RAD21-1, la construction la plus courte, en bleu pour RAD21-2 et en vert pour RAD21-3 et -4. Ces domaines sont reliés les uns autres par des linkers. Les constructions 3 et 4 présentent la même fragmentation des domaines d'interaction de RAD21, seuls les linkers utilisés divergent. La figure 40A récapitule les constructions de RAD21 utilisées, également schématisées dans le chapitre 2 - Matériels et méthodes. Ces constructions ont ensuite été clonées dans des vecteurs d'expression natifs ou avec des étiquettes différentes en C-terminal, comme l'étiquette His_{10}.

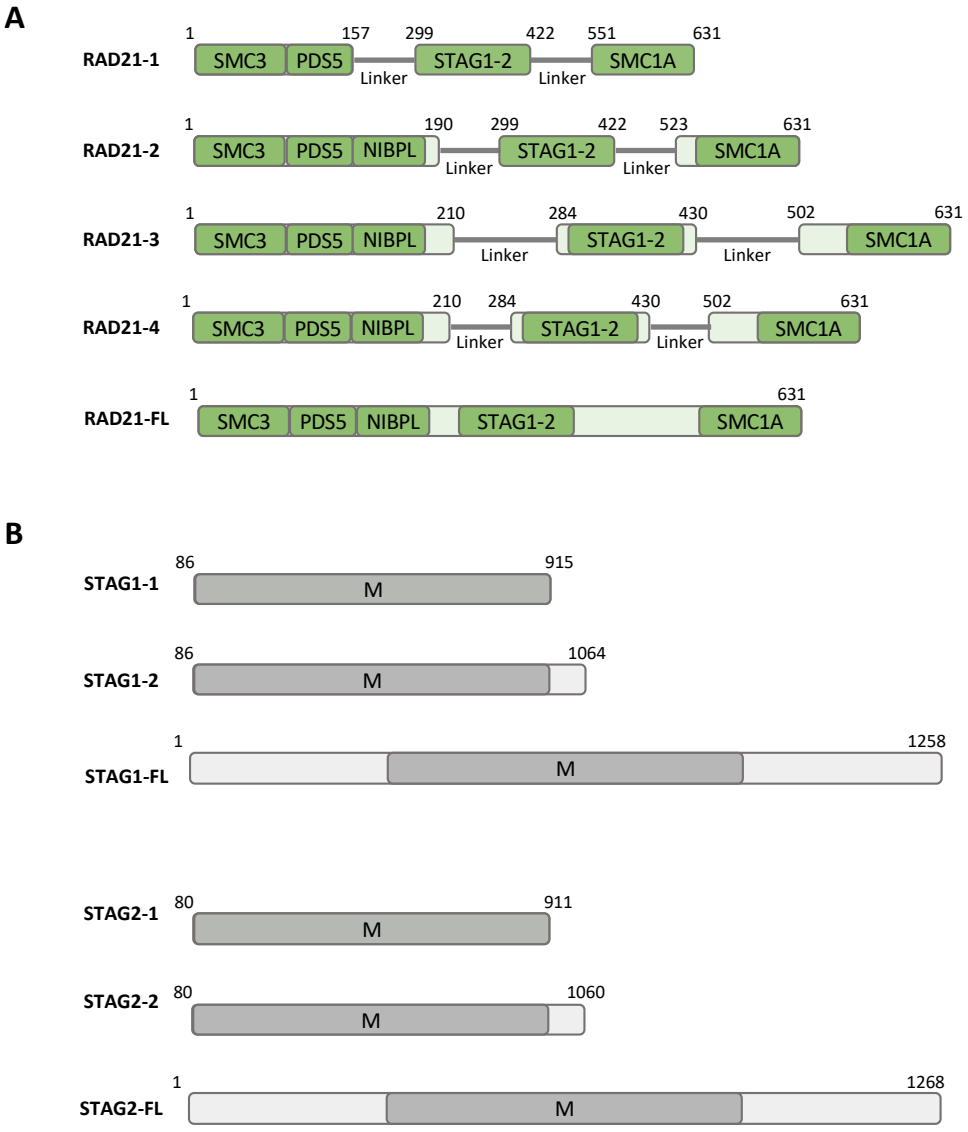


Figure 40 : Récapitulatif des constructions de RAD21 (A), ainsi que celles de STAG1 et STAG2 (B) utilisées pour la reconstruction du complexe tripartite ou quadripartite de la Cohésine.

2. Reconstitution des complexes quadripartites de la Cohésine

L'expression et la purification de protéines font appel à plusieurs étapes. La première est de déterminer les constructions les plus adéquates en termes de production et de stabilité de la protéine ou du complexe d'intérêt, ainsi que les conditions d'expression et de purification les plus prometteuses à petite échelle. Le passage à plus grande échelle nécessite plusieurs étapes d'optimisation, en combinant notamment plusieurs étapes de purifications par affinité et par SEC. Il est ensuite nécessaire d'étudier le comportement de la protéine obtenue en solution, afin de déterminer si la protéine ou le complexe est apte à être utilisé pour sa caractérisation.

Basé sur les données de la littérature, j'ai tout d'abord envisagé la reconstitution du complexe quadripartite en plusieurs étapes, la première étant d'exprimer et de purifier les protéines STAG1 et STAG2 d'une part et de co-exprimer et co-purifier les mutants EQ des têtes de SMC1A et de SMC3 et d'une des constructions de RAD21 d'autre part. Enfin, une dernière étape permet l'assemblage du complexe quadripartite par le mélange des quatre protéines.

2.1. Première approche : purifications indépendantes du module ATPase et de STAG1/STAG2, puis assemblage du complexe quadripartite

2.1.1. Expression et purification de STAG1

J'ai commencé ces essais pour la protéine STAG1. J'ai ainsi mis en place un protocole pour l'expression et la purification à grande échelle de STAG1. La qualité de la protéine purifiée obtenue m'a permis de caractériser son comportement en solution par la technique de DLS. De manière succincte, les trois constructions de STAG1 (STAG1-FL, STAG1-1 et STAG1-2), associées à trois étiquettes de fusion différentes (His_6 ou GST à l'extrémité N-terminale ou His_{10} à l'extrémité C-terminale) ont été exprimées à petite échelle à l'aide de cellules BL21 pRARE2 pLysS dans un milieu de culture AI.TB ou 2xLB. Deux tampons de purification ont été testés, le premier se composant de 40mM de Tris-HCl, pH 7,5 et de 150 mM de NaCl, le second de 40mM Tris-HCl, pH 7,5 et 400mM de NaCl, soit une matrice de 36 combinaisons différentes.

Plusieurs constructions étaient prometteuses, avec notamment His_6-STAG1-1, et ont été utilisées par la suite pour un test de purification à grande échelle effectué à partir d'un

litre de culture et en utilisant une purification en deux étapes, par affinité et par exclusion de taille. Ces tests ont mis en évidence que la construction His_6-STAG1-1 purifiée avec un tampon de 400 mM de NaCl permet un rendement de 0,5 mg/L, avec un échantillon stable avec une bonne polydispersité (21,2%) mesurée par DLS. La protéine présente toutefois une légère dégradation observable sur gel SDS-PAGE. Une contamination aux acides nucléiques est également visible sur le chromatographe de la SEC. De ce fait, des optimisations concernant l'expression et la purification ont ensuite été réalisées afin d'obtenir un échantillon le plus pur et homogène possible et ont permis d'aboutir au protocole final illustré dans la figure 41. La purification comprend trois étapes, une purification par affinité, une purification par échangeuse d'anion, qui permet de séparer la protéine d'intérêt de l'ADN contaminant, et une purification par SEC. A la fin de cette dernière étape, quelques dégradations mineures sont toujours présentes, malgré les différentes étapes de purification et co-éluent avec la protéine d'intérêt lors de la SEC.

De plus, nous observons deux pics sur le profil de chromatographie : un premier très faible, correspondant probablement à une forme multimérique et un second majoritaire, attribué au monomère. Afin de confirmer cette hypothèse et de confirmer les fractions d'intérêt, une expérience de SEC-MALS a été réalisée sur les fractions correspondantes à la forme monomérique, par la plateforme de biologie de l'IGBMC. Cette expérience démontre que l'échantillon est très homogène avec très peu d'agrégats et presque exclusivement composé de monomères de 95,9 kDa (96 %), valeur proche du poids théorique de STAG1-1 (98,3 kDa). Les fractions correspondantes à la protéine STAG1-1 monomérique ont été prélevées, concentrées et congelées afin d'être utilisées pour la reconstitution du complexe quadripartite de la Cohésine . Le rendement de cette purification et de 0,5 mg/L de culture.

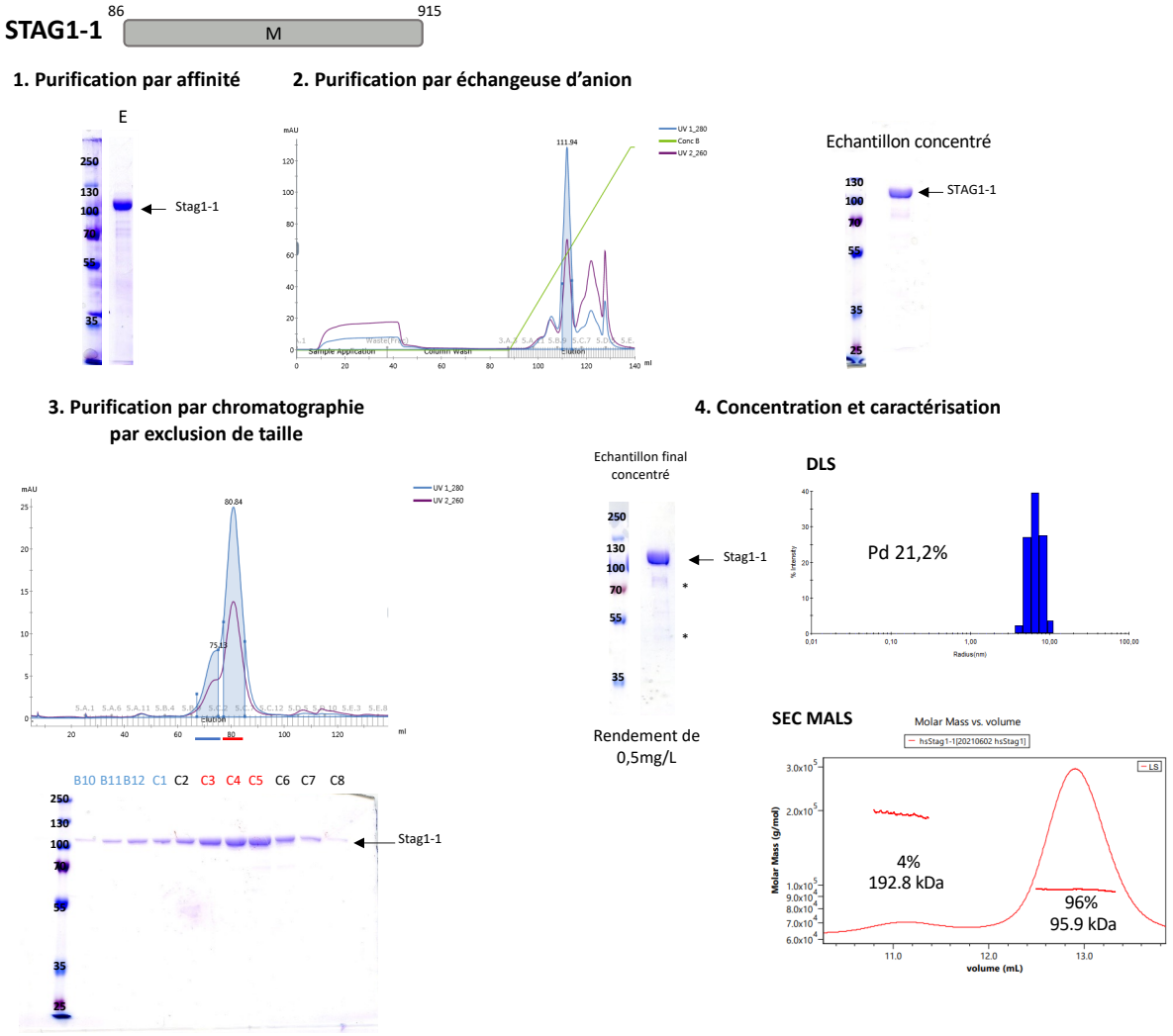


Figure 41 : Purification et caractérisation de la protéine STAG1-1.
Après une purification par affinité (1), l'éluât (E) contient la protéine d'intérêt STAG1-1 dont l'étiquette a été clivée et correspond à l'échantillon utilisé pour la purification par échangeuse d'anion (2). Cette dernière permet d'éliminer les contaminants d'ADN et certaines dégradations/contaminations protéiques. S'en suit une étape de purification par une chromatographie par exclusion de taille (3). Sur le profil de chromatographie d'exclusion de taille, deux pics ont été observés, un pic correspondant à une forme multimérique et un pic majoritaire correspondant au monomère de STAG1-1. Les fractions du pic majoritaire ont été concentrées, déposées sur gel et analysées par DLS et SEC-MALS. Les dégradations sont indiquées par un astérisque. L'expérience de SEC MALS montre un échantillon très homogène presque exclusivement composé de monomères (96%). Ces fractions de STAG1-1 ont été utilisées pour les expériences décrites dans les prochaines parties.

2.1.2. Expression et purification de STAG2

La même approche que celle décrite précédemment a été réalisée pour STAG2. Les tests à petite échelle ont été réalisés avec les trois constructions de STAG2 avec trois étiquettes de fusion distinctes, ainsi qu'avec les mêmes conditions de culture et de purification. Ces tests ont permis d'identifier plusieurs constructions et conditions prometteuses. Je n'ai pu réaliser

que les premiers tests d'expression et de purification à grande échelle sur 1L de culture et comprenant une purification par affinité et par exclusion de taille, pour chaque construction et condition de culture. Parmi toutes les conditions testées, deux conditions semblent encourageantes. Il s'agit de la construction STAG2-2 associée à étiquette His_6 en position N-terminale exprimée dans un milieu de culture AI.TB ou 2xLB et respectivement purifiée dans un tampon contenant 400 ou 150 mM de sel. Leurs profils de la chromatographie d'exclusion de taille sont illustrés à la figure 42. Des optimisations sont encore nécessaires pour obtenir un échantillon le plus homogène et pur possible.

Au vu de l'avancée de mes travaux et du temps nécessaire au développement et à l'optimisation d'une nouvelle procédure de purification, j'ai choisi d'utiliser uniquement la construction de STAG1-1 pour mener les essais de reconstitution du complexe quadripartite.

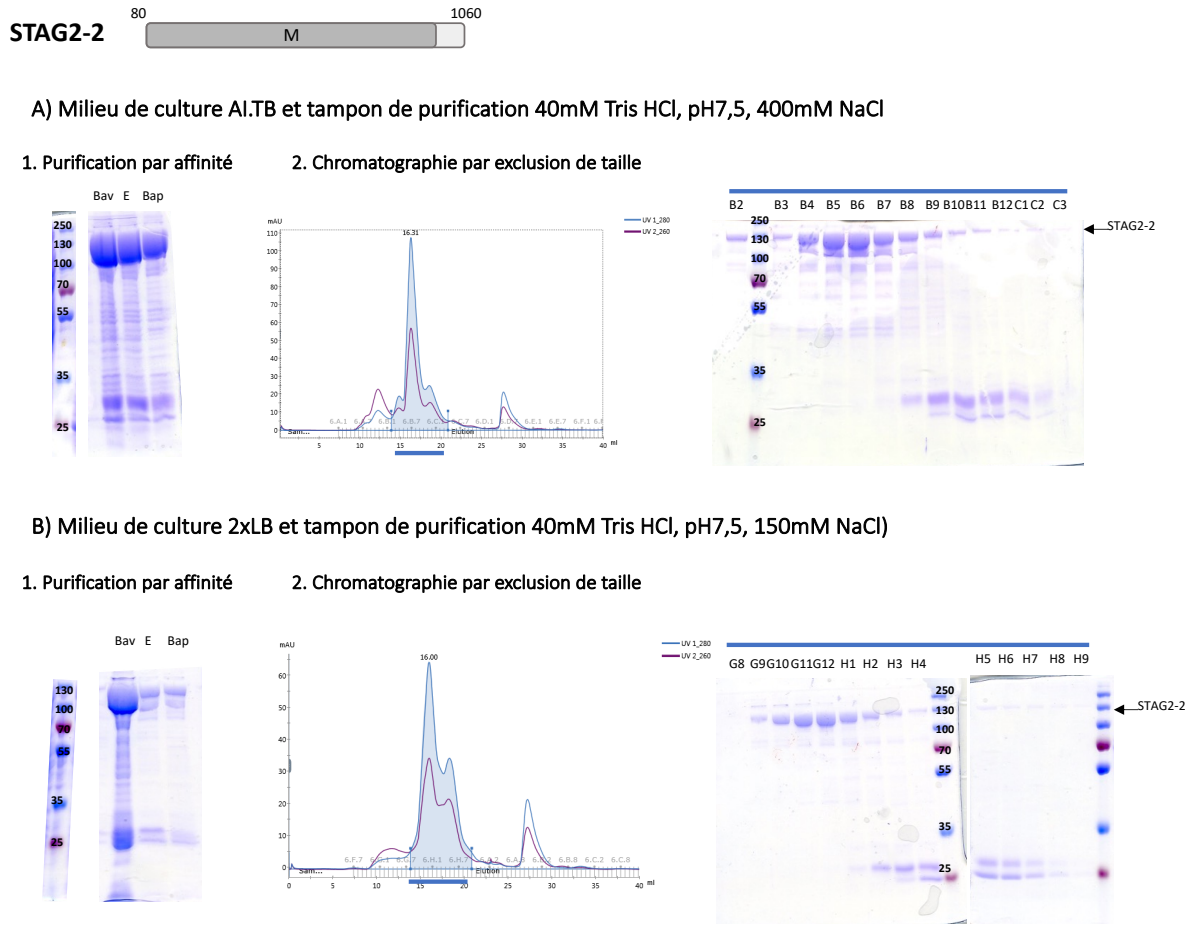


Figure 42 : Premiers essais de la purification de la protéine STAG2-2, construction la plus prometteuse. *Cette expérience se divise en deux étapes : une purification par affinité (1), suivie d'une chromatographie par exclusion de taille (2), représentées par des gels SDS-PAGE et chromatogrammes. Ces essais permettent de poser les bases pour l'établissement d'un protocole d'expression et de purification de STAG2-2. De nombreuses optimisations sont toutefois nécessaires pour obtenir une protéine la plus pure et homogène possible. Bav : Billes avant élution, E : éluât avec de l'imidazole 250 mM, Bap : billes après élution.*

2.1.3. Expression et purification du complexe SMC1A-HD/SMC3-HD/RAD21

Pour cette expérience, j'ai utilisé les constructions SMC1ACCsh-EQ, SMC3CC-EQ, ainsi que les cinq constructions de RAD21 (RAD21-FL, -1, -2, -3 et -4). Afin de co-exprimer et de co-purifier les trois protéines, plusieurs expériences de clonage basées sur la concaténation ont été nécessaires et m'ont permis d'incorporer les trois gènes d'intérêt sur un même vecteur. Cela m'a permis de diminuer le nombre d'antibiotiques nécessaires à la co-expression de plusieurs protéines et d'améliorer les conditions de croissance.

J'ai ensuite réalisé des expériences de co-expression et de co-purification à petite échelle des têtes de SMC1A et de SMC3 associées aux cinq constructions de RAD21. L'étiquette His_{10} de purification utilisée pour la purification par affinité est située à l'extrémité C-terminale de RAD21. Lors de ces tests à petite échelle, j'ai examiné différents paramètres, tels que le choix de la souche bactérienne (BL21 ou BL21 pRARE2 pLysS), du milieu de culture (AI.TB ou 2xLB), ainsi que du tampon de purification avec un pouvoir tampon différent (20 mM ou 40 mM de Tris-HCl pH 8) et une concentration saline différente (50, 200 ou 500 mM de NaCl). Seuls les résultats les plus probants, bien que limités, des fractions solubles fixées sur les billes d'affinité des échantillons sont illustrés à la figure 43. L'analyse des gels SDS-PAGE des fractions solubles a été complexe, d'une part du fait du bruit lié aux fixations aspécifiques des protéines aux billes TALON®. D'autre part, RAD21 possède très peu d'arginines, de lysines et d'histidines, principaux résidus permettant la coloration des protéines par le bleu de Coomassie et certaines de ses constructions présentent des tailles similaires aux têtes de SMC1A et SMC3. La bande correspondante à RAD21 sur les gels est alors très faiblement colorée et difficilement identifiable. La présence de RAD21 et de ses différentes constructions a été supposée par la présence des bandes correspondant aux têtes de SMC1A et de SMC3, qui sans RAD21 deviennent insolubles et ne seraient pas co-purifiées, et donc non présentes sur les gels SDS-PAGE des fractions solubles. Les bandes de SMC1A et de SMC3 étant plus intenses, pour les constructions RAD21-1, -2 et -4, nous émettons l'hypothèse que ces constructions de RAD21 soient les plus adaptées pour les tests à grande échelle.

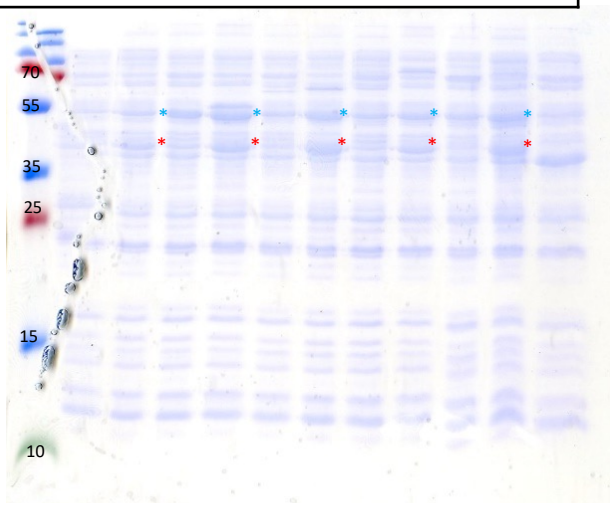
Malgré ces difficultés, j'ai pu trouver quelques conditions positives, bien que limitées. Il s'agit notamment de l'utilisation des bactéries pRARE2 pLysS, celle du milieu de culture 2xLB, ainsi que l'emploi de tampon avec une concentration saline de 200 mM. Concernant les constructions de RARD21 utilisées, étant donné qu'il a été difficile d'extrapoler à partir de ces

résultats à petite échelle, j'ai décidé de tester la co-expression et la co-purification de ce complexe avec les cinq constructions de RAD21.

Tests à petit échelle (purification par affinité)

Co-expression et co-purification de SMC1ACCsh-EQ/SMC3CC-EQ/RAD21
Milieu de culture 2xLB - Tampon de purification 20mM Tris-HCl pH8, 200 mM NaCl

		1		2		3		4		FL	
Rad21	Ø	✓		✓		✓		✓		✓	
	- His_{10}		✓		✓		✓		✓		✓
SMC1ACC-EQ		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SMC3CC-EQ		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓



Protéines		Poids moléculaire
SMC1ACCsh-EQ		40,6 kDa
SMC3CC-EQ		52,3 kDa
Rad21FL	Ø	71,7 kDa
	His_{10}-	74,4 kDa
RAD21-1	Ø	42,3 kDa
	His_{10}-	45,0 kDa
RAD21-2	Ø	49,2 kDa
	His_{10}-	52,0 kDa
RAD21-3	Ø	58,4 kDa
	His_{10}-	61,2 kDa
RAD21-4	Ø	56,5 kDa
	His_{10}-	59,2 kDa

Figure 43 : Gels SDS-PAGE des résultats les plus prometteurs des tests à petite échelle de la co-expression et co-purification du complexe humain SMC1A-HD/SMC3-HD/RAD21.
L'identification de la bande correspondant à RAD21 est très difficile, du fait du bruit aspécifique des billes TALON et de la faible fixation du bleu sur RAD21. La présence de RAD21 est alors confirmée par des bandes plus intenses pour les têtes SMC1A et SMC3, indiquées par des astérisques rouge et bleu respectivement et qui sont retenues grâce à la construction RAD21 fusionnée à l'étiquette His_{10} à son extrémité C-terminale. RAD21 Ø indique que cette protéine est sous forme native, sans étiquette.

Les tests préliminaires de purification à grande échelle se font à partir d'un litre de culture et sont composés de deux étapes chromatographiques, une purification par affinité suivie d'une purification par chromatographie par exclusion de taille, grâce à la colonne analytique Superdex® 200 10/300 GL. Les protéines ont été éluées des billes TALON® par ajout d'imidazole, l'étiquette restant fusionnée à la construction de RAD21. Ces expériences ont mis en évidence que la formation du complexe avec la construction la plus courte de RAD21 (RAD21-R1) semble être la plus optimale avec une assez bonne stœchiométrie, une dissociation du complexe semble être observée pour les autres constructions de RAD21, ainsi

que de nombreuses dégradations. De plus, il reste également compliqué de déterminer la bande correspondant aux constructions de RAD21, excepté pour RAD21-1, au vu des nombreuses bandes présentes sur les gels SDS-PAGE. Il est également possible que les constructions de RAD21 co-migrent avec une des têtes, notamment pour RAD21-2 dont le poids est très proche de celui de la tête de SMC3. Un complexe tripartite, composé des constructions RAD21-2 ou 4, pourrait être formé en faible quantité, mais la bande de RAD21 reste à identifier précisément (Figure 44). Au vu des chromatogrammes et des rendements protéiques évalués sur les gels SDS-PAGE, les résultats de ces expressions et purifications à grande échelle sont mitigés pour l'ensemble des conditions testées, bien que RAD21-1 semble donner les résultats les plus encourageants. C'est pour cela que j'ai réalisé un dernier essai pour tenter d'optimiser le protocole d'expression et de purification du complexe SMC1A/SMC3/RAD21-1.

Lors de l'étape d'optimisation de l'expression et de purification du complexe à trois, j'ai utilisé la construction SMC1ACC-EQ, à la place de SMC1ACCsh-EQ, construction majoritairement utilisée pour les travaux de la première partie des résultats sur le module ATPase. J'ai également joué sur les paramètres d'expressions, notamment sur la densité des bactéries au moment d'induire l'expression avec de l'IPTG, en testant une densité optique, dite classique, autour des 0,8 et d'une densité bien plus élevée proche de 2. J'ai également modifié quelques paramètres de purification, en clivant cette fois l'étiquette de fusion de RAD21-1, en ajoutant un mélange d'inhibiteur de protéines (cOmplete™ Protease Inhibitor, PIC, Roche®) et en utilisant une autre colonne, la Superpose® 6 increase 10/300GL qui permet une meilleure séparation pour les protéines de poids moléculaires proches du complexe quadripartite (SMC1A-HD, SMC3-HD, RAD21, STAG1) et que j'ai utilisée par la suite pour la reconstitution de complexe à quatre partenaires. Cette étape m'a ainsi permis de comparer les complexes à trois partenaires et à quatre partenaires.

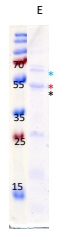
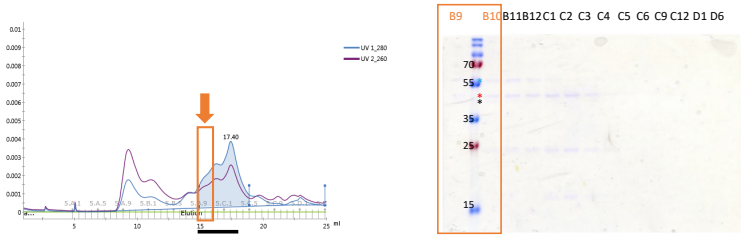
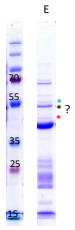
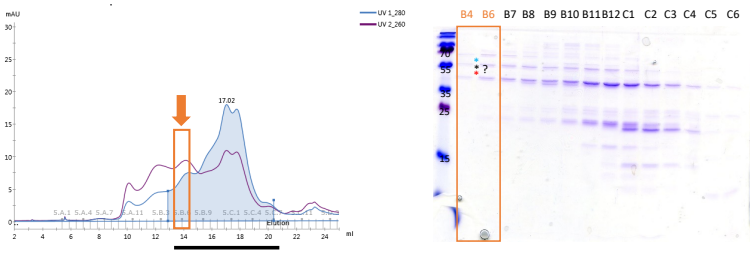
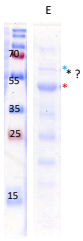
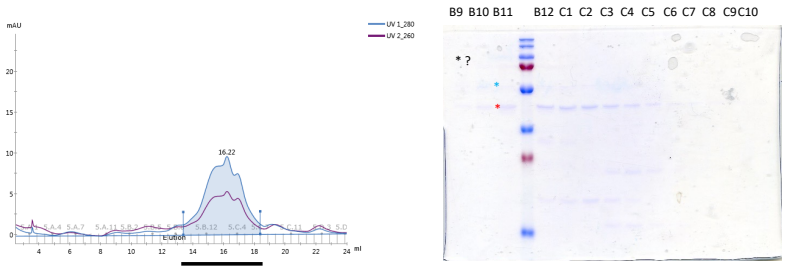
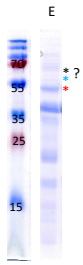
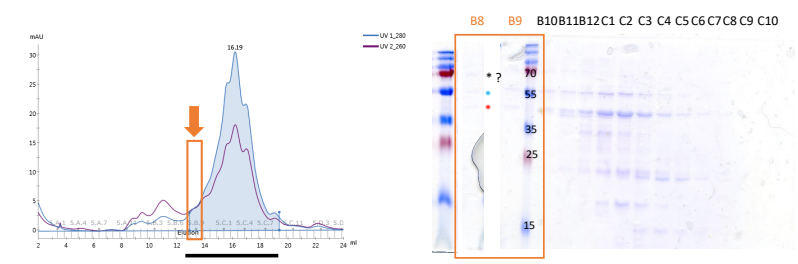
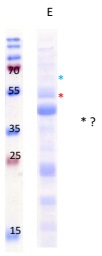
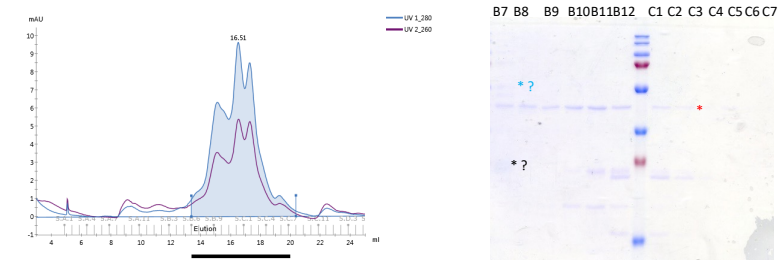
	Purification par affinité	**Purification par chromatographie d'exclusion de taille** (colonne Superdex® 200 10/300 GL)
SMC1CCsh-EQ (40,6 kDa) SMC3CC-EQ (52,3 kDa) RAD21-1 (45 kDa)		
SMC1CCsh-EQ (40,6 kDa) SMC3CC-EQ (52,3 kDa) RAD21-2 (52 kDa)		
SMC1CCsh-EQ (40,6 kDa) SMC3CC-EQ (52,3 kDa) RAD21-3 (61,2 kDa)		
SMC1CCsh-EQ (40,6 kDa) SMC3CC-EQ (52,32 kDa) RAD21-4 (59,2 kDa)		
SMC1CCsh-EQ (40,6 kDa) SMC3CC-EQ (52,32 kDa) RAD21-FL (74,4 kDa)		

Figure 44 : Résultats des premiers essais à grande échelle du complexe formé par les têtes de SMC1A et SMC3 et par une des constructions de RAD21.

Gels SDS-PAGE et chromatogrammes de SEC des différentes étapes de purification. E : éluât des protéines d'intérêt après la purification d'affinité par l'ajout de 250 mM d'imidazole dans le tampon de purification.

Les résultats suivant la purification par affinité montrent un profil similaire, quelle que soit la densité optique au moment de l'induction, les têtes ATPases étant visibles sur gels SDS-PAGE, mais RAD21 restant toujours difficile à identifier (Figure 45). Toutefois, lors de la chromatographie par exclusion de taille nous n'observons pas de complexe, les têtes de SMC1A et SMC3 sont éluées indépendamment l'une après l'autre en fonction de leur poids moléculaire. Cela indique donc que le clivage de l'étiquette semble entrainer une dégradation de RAD21, ce qui conduit à la dissociation du complexe (Figure 45). Suite à ces résultats négatifs, j'ai abandonné cette approche de reconstitution du complexe en plusieurs étapes et j'ai donc commencé une nouvelle approche basée sur la reconstitution de ce complexe par la co-expression et la co-purification des quatre protéines de la Cohésine en même temps.

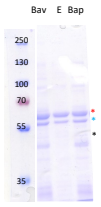
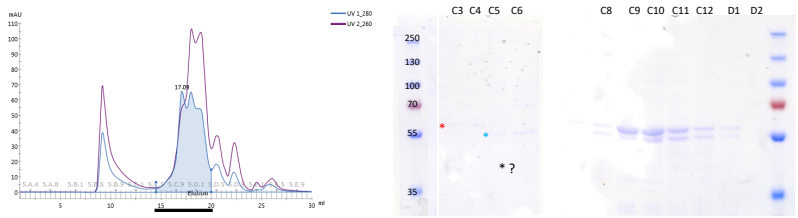
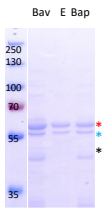
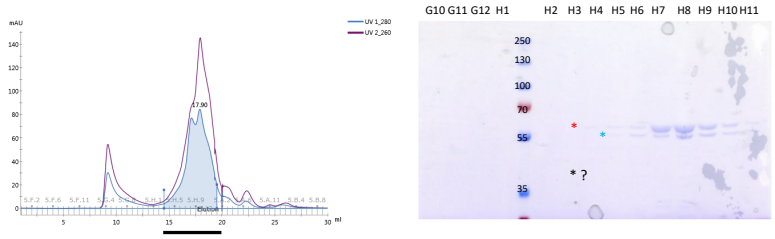
	Purification par affinité	**Purification par chromatographie d'exclusion de taille** (colonne Superose® 6 10/300)
DO d'induction : 2 SMC1CC-EQ (51,3 kDa) SMC3CC-EQ (52,3 kDa) RAD21-1 (45 kDa)		
DO d'induction : 0,9 SMC1CC-EQ (51,3 kDa) SMC3CC-EQ (52,3 kDa) RAD21-2 (52 kDa)		

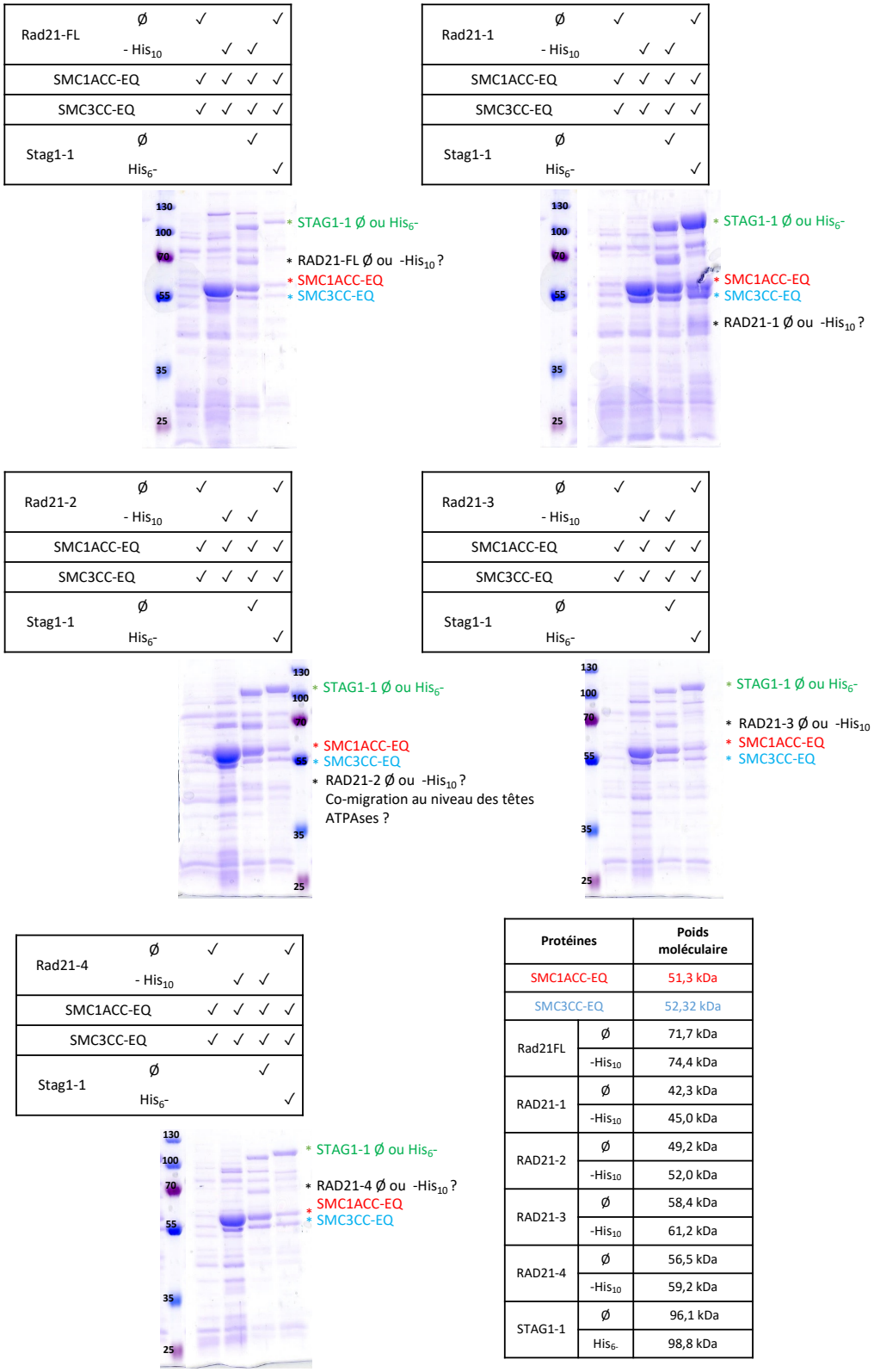
Figure 45 : Essai d'optimisation pour la co-expression et la co-purification du complexe quadripartite, en jouant sur la densité optique d'induction lors de l'expression et en clivant l'étiquette sur RAD21-1 par la protéase 3C.

Gels SDS-PAGE et chromatogrammes de SEC des différentes étapes de purification. E : éluât des protéines d'intérêt après la purification d'affinité par coupure 3C, Bav : billes avant la coupure, Bap : billes après la coupure.

2.1.4. Deuxième approche : expression et purification des quatre partenaires de la Cohésine

Cette approche a également nécessité des expériences de clonage de concaténation de vecteurs dont les techniques ont été décrites dans le chapitre 2 - Matériels et méthodes, et qui sont similaires à celles décrites dans la partie précédente et qui ne seront pas détaillées ici. J'ai ensuite réalisé des essais à petite échelle afin de déterminer les conditions les plus adéquates pour le passage à grande échelle. Ces tests ont également été optimisés par l'ajout d'imidazole (5 mM) lors des lavages permettant de réduire les fixations aspécifiques de protéines aux billes TALON®. Les tests d'affinité ont été faits grâce soit à la fusion de l'étiquette His_{10} à l'extrémité N-terminale de RAD21, soit à la fusion de l'étiquette His_6 à l'extrémité N-terminale de STAG1-1. Les cinq constructions de RAD21 (RAD21-1, -2, -3, -4, -FL) ont été testées afin de déterminer si STAG1 stabilise RAD21 par son interaction. Les paramètres d'expression et de purification sont similaires à ceux testés précédemment et seuls les gels SDS-PAGE des fractions solubles des expériences les plus prometteuses sont détaillés à la Figure 46. L'optimisation du protocole de tests à petite échelle permet de réduire le bruit et d'identifier les bandes pouvant correspondre aux différentes protéines de la Cohésine. J'ai également pu déterminer les bandes pouvant correspondre à RAD21, bien que cela reste à confirmer à grande échelle.

STAG1-1 semble avoir un impact positif sur la stabilité de RAD21 et évite sa dégradation, la stœchiométrie des têtes étant meilleure en présence de STAG1-1. La purification par affinité grâce à l'étiquette présente sur RAD21 semble donner de meilleurs rendements, les bandes correspondant à SMC1A et SMC3 étant plus intenses que lors de la purification par affinité via l'étiquette de STAG1-1, excepté pour le complexe associé à RAD21-1 qui présente un profil similaire quelle que soit la localisation de l'étiquette de purification. Ainsi, les résultats à petite échelle de l'expression et de la purification du complexe quadripartite ont été plus encourageants, en comparaison à ceux obtenus pour le complexe formé par les têtes de SMC1A et SMC3, ainsi que par une des constructions de RAD21. Cela m'a permis de passer à la prochaine échelle, notamment pour le complexe incluant la construction RAD21-1, qui semble être la plus stable et qui présente le rendement le plus prometteur, par comparaison de l'intensité des bandes des différentes conditions.



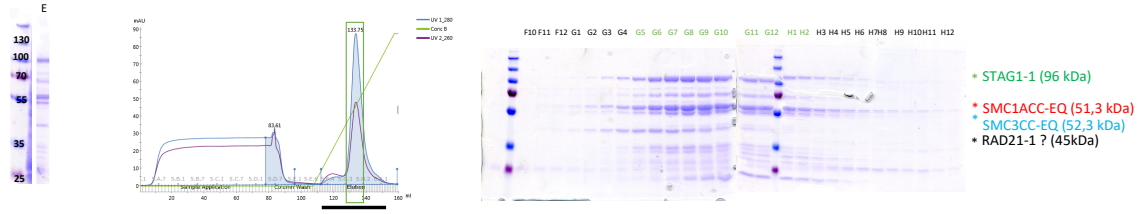
Protéines		Poids moléculaire
SMC1ACC-EQ		51,3 kDa
SMC3CC-EQ		52,32 kDa
Rad21FL	Ø	71,7 kDa
	-His10	74,4 kDa
RAD21-1	Ø	42,3 kDa
	-His10	45,0 kDa
RAD21-2	Ø	49,2 kDa
	-His10	52,0 kDa
RAD21-3	Ø	58,4 kDa
	-His10	61,2 kDa
RAD21-4	Ø	56,5 kDa
	-His10	59,2 kDa
STAG1-1	Ø	96,1 kDa
	His6-	98,8 kDa

Figure 46 : Résultats des tests à petite échelle des complexes quadripartites.
Gels SDS-PAGE des fractions solubles après l'étape purification par affinité. Cette étape d'affinité se fait soit grâce à l'étiquette His_{10} en position C-terminale de RAD21 ou l'étiquette His_6 en position N-terminale de STAG-1. Le symbole Ø indique la protéine native, $-His_6$ et $-His_{10}$ la protéine couplée à une étiquette. Un astérisque est présent lorsque la bande correspondant à une des protéines est visible à la taille attendue. Un point d'interrogation indique qu'une incertitude concernant l'identification de la protéine persiste.

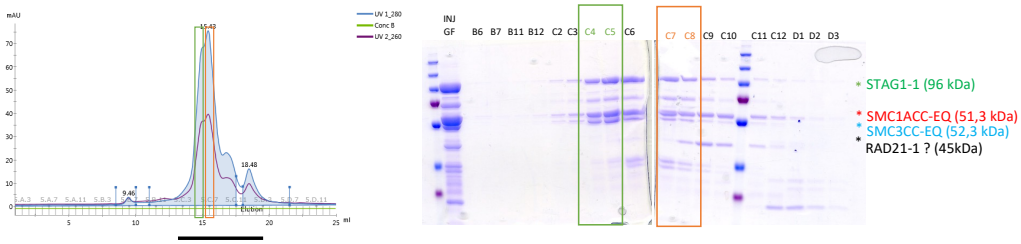
Après ces tests à petite échelle, j'ai donc débuté un essai pilote à grande échelle à partir de 3L de culture, puis réalisé des optimisations. Du fait que le clivage de l'étiquette de RAD21-1 entraine sa dégradation (résultats décrits dans la partie précédente sur la purification du complexe à trois partenaires) et par manque de temps, je me suis focalisée sur le complexe SMC1ACC-EQ/SMC3CC/RAD21-R1/STAG1-1 avec l'étiquette portée par la protéine RAD21-1. J'ai tout d'abord testé différents tampons avec un pouvoir tampon différent (20 ou 40 mM Tris-HCl) et une concentration saline de 200 mM. J'ai également testé l'utilisation de glycérol, ainsi que l'utilisation d'un mélange PIC (Roche®), afin d'éviter la dégradation des protéines. De plus, j'ai ajouté une étape de purification par héparine, afin d'éliminer l'ADN contaminant. La Figure 47 illustre les différentes étapes de purification pour obtenir le profil de chromatographie d'exclusion stérique à partir d'une colonne Superpose® 6 increase 10/300GL. Ce dernier met en évidence deux pics, un premier correspondant à un épaulement, suivi d'un second pic. Concernant le volume d'élution du premier pic, il correspondrait au volume attendu pour l'élution d'un complexe quadripartite, celui du second pic à celui d'une possible dissociation du complexe. Nous pouvons déterminer précisément sur les gels SDS-PAGE les bandes correspondantes aux protéines STAG1-1, SMC1ACC-EQ et SMC3CC-EQ, celle supposée de RAD21 est plus difficile à définir, surtout pour les fractions du pic 1 (Figure 47). Les gels SDS-PAGE montrent que les fractions du pic 1, et cela même après leur concentration, présentent une plutôt bonne stœchiométrie des têtes de SMC1A et SMC3. Au contraire des fractions du second pic, où nous observons une surreprésentation de la tête de SMC1A, en comparaison de celle de SMC3. De nombreuses dégradations sont présentes dans les deux pics, bien qu'elles soient plus intenses dans le second pic. Après concentrations, le rendement pour le premier pic est inférieur à de 0,01 mg/L de culture, le rendement du second est de 0,0069 mg/L de culture. Ces rendements sont très faibles.

Afin de confirmer la présence des quatre partenaires de la Cohésine, les gels SDS-PAGE des échantillons des deux pics ont été envoyés à la plateforme de protéomique de l'IGBMC, pour qu'ils soient analysés par spectrométrie de masse, technique permettant d'obtenir la cartographie peptidique des protéines. Les bandes des gels caractérisées sont détaillées à la Figure 47.

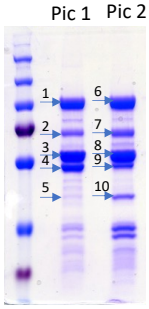
1. Purification par affinité **2. Purification par héparine**



3. Purification par chromatographie par exclusion stérique (Superpose ® 6 increase 10/300GL)



4. Concentration et gel analysé par spectrométrie de masse



Pic	Bandes	Protéines correspondantes (analyses par spectrométrie de masse)	Pic	Bandes	Protéines correspondantes (analyse par spectrométrie de masse)
1	1	STAG1-1	2	6	STAG1-1
	2	Dégradation de STAG1-1		7	Dégradation de STAG1-1
	3	SMC1ACC-EQ RAD21-1 Dégradation de STAG1-1		8	SMC1ACC-EQ RAD21-1 Dégradation de STAG1-1
	4	SMC3CC-EQ		9	SMC3CC-EQ
	5	Dégradation de STAG1-1 RAD21-1 Répresseur de l'opéron lactose (E. coli)		10	Répresseur de l'opéron lactose (E. coli) Dégradation de STAG1-1 RAD21-1

Figure 47 : Résultats de la purification à grande échelle du complexe quadripartite SMC1ACC-EQ/SMC3CC-EQ/RAD21-1/STAG1-1.

Gels SDS-PAGE et chromatogrammes des différentes étapes de purification, ainsi que les résultats des analyses de spectrométrie de masse des différents pics pouvant correspond au complexe à quatre partenaires. E : éluât. Les bandes correspondant aux protéines étudiées sont indiquées par un astérisque. Lorsque plusieurs protéines co-migrent sur une même bande, les protéines sont indiquées dans un ordre décroissant quantitatif pour le tableau des données de spectrométrie de masse.

Concernant les échantillons concentrés des deux pics, les quatre bandes identifiées correspondant bien aux quatre protéines d'intérêt, bien que STAG1-1 se dégrade et une de ses dégradations co-migre notamment avec RAD21-1 et la tête de SMC1, expliquant potentiellement l'intensité légèrement plus forte de la bande de SMC1, STAG1-1 fixant très bien le bleu de Coomassie. Au vu de la présence de nombreuses dégradations, STAG1-1 ne semble pas stable dans les conditions de purification. Pour confirmer cela, j'ai réalisé une expérience de DLS sur STAG1-1 purifié seul dans le même tampon de purification, qui indique un signal multimodal caractéristique d'un mauvais comportement protéique.

Concernant RAD21, la bande identifiée comme étant celle de RAD21-1 des deux pics correspond bien à cette protéine, mais n'est pas la bande majoritaire de RAD21-1. Cette bande

correspond à la co-migration de RAD21-1, de dégradations de STAG1-1, ainsi que d'un inhibiteur de l'opéron lactose de *E. coli*. La bande majoritaire de RAD21-1 co-migre également avec la tête de SMC1A dans les deux pics, participant également à l'intensité de la bande correspondant à SMC1ACC. Il est possible que RAD21-1 adopte plusieurs conformations qui peuvent entrainer une migration différente, expliquant ainsi la présence des deux bandes correspondant à RAD21.

Les quatre sous-unités du complexe sont également observées dans les deux pics et présentent le même profil de couverture, bien que les quantités soient différentes. Il se peut donc que soit le premier pic corresponde au complexe à quatre et le second à une dissociation, soit le complexe à quatre partenaires est présent dans les deux pics, mais adopte des conformations différentes, entrainant une élution à des volumes différentes.

Des études supplémentaires sont encore nécessaires, afin de compléter et améliorer ce protocole initial et poursuivre la caractérisation biochimique et structurale de ce complexe quadripartite de la Cohésine. Cette caractérisation sera discutée dans la partie discussion.

2.2. Discussion

2.2.1. Optimisation du protocole d'expression et de purification du complexe quadripartite SMC1ACC/SMC3CC/RAD21/STAG1-1

La reconstitution d'un complexe quadripartite de la Cohésine a été difficile à mettre en place. La première approche basée sur la littérature est divisée en trois étapes : (i) la co-expression et la co-purification des têtes de SMC1A et de SMC3, ainsi que d'une construction de RAD21, (ii) l'expression et la purification de STAG1 et (iii) la reconstitution du complexe par mélange, s'est avérée infructueuse. J'ai alors envisagé une seconde approche, qui implique la co-expression et la co-purification des quatre partenaires. Les tests à petite échelle de cette approche ont semblé prometteurs et encourageants. Ils mettent en évidence et confirment la formation du complexe quadripartite à cette échelle.

Cependant, les tests à grande échelle n'ont pas apporté des résultats aussi probants, le rendement étant inférieur à 0,01 mg/L de culture et de nombreuses dégradations sont présentes, notamment pour STAG1. Au vu des résultats, il serait souhaitable d'utiliser un autre système d'expression que celui de *E. coli*, comme le système bacculovirus/cellules d'insecte.

En effet, ce système est utilisé par les équipes concurrentes travaillant sur la Cohésine humaine et présente des rendements protéiques suffisants pour sa caractérisation et des échantillons très homogènes avec peu de dégradations (Shi et al., 2020). Ce système d'expression, comparé au système *E. coli*, permet l'ajout de modifications post-traductionnelles qui peuvent être cruciales pour la stabilité d'une protéine. RAD21 et STAG1 subissent notamment de nombreuses phosphorylations, ce qui pourrait potentiellement améliorer leur stabilité et de ce fait le rendement protéique final (Mehta et al., 2012; Cheng et al., 2020). Ce système est plus onéreux que le système *E. coli* et nécessite de repartir sur des nouvelles expériences de clonage, de tests d'expression et de purification.

Concernant la purification, l'utilisation de détergents, tels que le Tween 20, le (3-((3-cholamidopropyl) diméthylammonio)-1-propanesulfonate) (CHAPS), ainsi que NP-40 pourraient également favoriser la stabilité de ce complexe à quatre partenaires. J'ai pu réaliser des expériences préliminaires utilisant du glycérol, comme stabilisateur, mais les résultats n'ont pas été convaincants et montrent une dissociation du complexe quadripartite. Des expériences préliminaires avec de l'ATP dans le tampon de purification ont montré que l'engagement des têtes n'a pas d'influence sur la purification du complexe, mettant en évidence qu'il s'agit bien d'un manque de stabilité induit par les protéines STAG1-1 et RAD21-1 et non par les têtes ATPases de SMC1A et SMC3. L'ensemble de ces paramètres devraient permettre d'obtenir un complexe suffisamment stable afin de poursuivre la caractérisation biochimique et structurale de ce complexe.

2.2.2. Caractérisation future du complexe quadripartite de la Cohésine

De nombreuses optimisations sont encore nécessaires pour l'obtention d'un complexe tripartite ou quadripartite de la Cohésine et n'ont pas pu être réalisées au cours de ma thèse. Toutefois ces optimisations nous auraient certainement permis d'obtenir des complexes suffisamment stables pour réaliser leur caractérisation. L'obtention de ces deux complexes nous permettrait de déterminer l'influence potentielle de STAG1 sur le module ATPase, en comparant notamment l'activité ATPase de ces deux complexes par des tests enzymatiques. Des approches de cryoEM et de réticulation *in vitro* pourront également être menées, afin d'étudier le comportement de STAG1 vis-à-vis du module ATPase et plus particulièrement au niveau de la porte de sortie de l'ADN.

De plus, des études caractérisant l'affinité et la cinétique de la liaison de ce module ATPase et de l'ADN, en absence et associé à STAG1, apporteraient également des informations novatrices sur la dynamique de la Cohésine. Pour ce faire, l'ITC, technique utilisée pour la caractérisation les têtes ATPases, ne semble pas être la technique la plus adéquate, au vu de la quantité protéique nécessaire et la haute sensibilité de cette technique, qui n'est pas adaptée à l'étude de complexes à trois ou quatre partenaires. Toutefois, l'utilisation de technique, comme l'interférométrie à bicouche (¨BioLayer Interferometry¨, BLI) ou la résonance plasmonique de surface (SPR), pourrait être envisagée. La BLI est une méthode optique permettant de mesurer les interactions moléculaires. Elle analyse les variations du motif d'interférence à la surface d'un biocapteur, sur lequel est immobilisée une des protéines et dont l'association/dissociation d'une molécule ou d'une partenaire protéique entraine un changement dans le modèle d'interférence (Sultana and Lee, 2015). La SPR est également une méthode optique, permettant de mesurer les interactions protéines-protéines. Une des protéines est immobilisée sur une puce et la liaison de son/ses partenaires entrainent une modification de l'angle de la réflexion, qui est mesurée et analysée (Douzi, 2017). Ces deux techniques permettraient donc de mieux caractériser les interactions entre les différents constituants de la Cohésine et son association avec l'ADN et ainsi de mieux comprendre le rôle de STAG1 dans le mode d'action de la Cohésine.

2.2.3. Le système *E. coli* et les tests à petite échelle sont des outils efficaces pour tester les interactions entre les complexes tripartite/quadripartite de la Cohésine et ses autres protéines auxiliaires

Bien que le système d'expression de *E. coli* ne semble pas être complètement adapté pour l'expression des complexes tripartite et quadripartite de la Cohésine à grande échelle, son utilisation à petite échelle est suffisante pour déterminer la formation de ces complexes. Ces tests à petite échelle avec un système d'expression *E. coli* présentent de nombreux avantages, ils sont rapides à mettre en place, demandent peu de ressources et sont optimisés de sorte à pouvoir tester un panel de conditions en une seule expérience. Ces tests sont donc adaptés pour étudier les interactions entre la Cohésine et ses autres protéines auxiliaires, telles que NIPBL, PDS5, etc, et cela en absence ou en présence d'ATP et/ou d'ADN. Il est également envisageable de mener des tests de compétition entre ces différentes protéines.

Cela nous permettrait ainsi de mieux caractériser et mieux comprendre les structures récemment publiées, où les trois composants de la Cohésine de *S. cerevisiae* (Smc1, Smc3, Scc1) forment un complexe avec Scc2 et Pds5 en absence d'ADN et d'ATP (Petela et al., 2021). Petela et al ont notamment montré que Scc2 et Pds5 peuvent interagir avec une des têtes ATPases de Smc1 et Smc3 en absence d'ATP et d'ADN, ce qui paraît surprenant au vu des fonctions de ses protéines qui participent au chargement et la stabilisation de la Cohésine sur l'ADN. Ces tests de co-expression nous permettraient ainsi de confirmer les interactions entre ces différents constituants observés par cryoEM et qu'il ne s'agit pas d'artefacts structuraux (Petela et al., 2021).

Il serait également intéressant de caractériser les interactions entre les régions charnières de SMC1A et de SMC3, avec STAG1, observées pour la première fois lors de la publication de la structure du complexe productif, et d'étudier la possible interaction entre PDS5 et ces mêmes régions charnières observée chez la levure (Shi et al., 2020; Petela et al., 2021). Ainsi, les tests d'interaction nous permettraient de mieux définir les interactions entre les composants de la Cohésine cœur et ses protéines accessoires, clé pour la régulation de la Cohésine tout au long de son cycle.

2.2.4. STAG1-STAG2 : vers une recherche appliquée

La Cohésine se retrouve dérégulée dans de nombreuses pathologies, incluant des maladies neurodéveloppementales et des cancers. La Cohésine est donc une cible de choix pour le développement de nouveaux traitements thérapeutique.

Le gène *Stag2* est notamment l'un des vingt-deux gènes les plus mutés dans au moins quatre types de cancer, conduisant majoritairement à une protéine tronquée et non fonctionnelle (Lawrence et al., 2014). De nombreuses études ont montré que l'absence de STAG2 est compensée au moins en partie par STAG1. Toutefois, l'absence simultanée des deux protéines entraine une diminution de la croissance cellulaire (Benedetti et al., 2017; van der Lelij et al., 2020; Bailey et al., 2021). Les cellules cancéreuses déficientes en STAG2 sont donc sensibles à l'inhibition de STAG1. Une approche de létalité synthétique peut alors être envisagée. Le principe de cette approche repose sur le fait que l'inhibition de deux gènes est létale pour la cellule, alors que ce dernier reste viable dans le cas de l'inhibition d'un seul des gènes. L'inhibition de STAG1 serait donc une bonne approche thérapeutique pour cibler la

cohésine, en développant notamment des petites molécules inhibant STAG1 et empêchant son incorporation au sein de la Cohésine. Cela peut se faire par des études de ¨drug design¨ basées sur la structure, qui nécessitent toutefois l'obtention de la structure de STAG1 à haute résolution. Actuellement, trois structures de STAG1 sont disponibles, les deux premières étant des structures cristallographiques des domaines allant du résidu 86 à 406 et du résidu 457 au résidu 911, avec des zones non définies et ayant respectivement une résolution de 2,37 Å et 3,17 Å (PDB : 6RRK et 6RRC) (van der Lelij et al., 2020). La dernière structure est une structure cryoEM de STAG1 provenant du complexe productif de la Cohésine (Cohésine associée à NIPBL et à l'ADN) à une résolution de 5,3 Å (PDB : 6WG3) (Shi et al., 2020). La première étape serait donc d'obtenir une structure de l'ensemble de STAG1 à haute résolution. Cela permettrait, par comparaison avec la structure de STAG2 disponible (PDB : 4PJU), d'identifier les zones spécifiques de STAG1. Ces zones doivent être différentes de STAG2, afin de cibler spécifiquement STAG1, de le déstabiliser et d'empêcher son incorporation au sein de la Cohésine. Des premiers essais de cristallisation sont en cours et sont les prémices d'un projet à plus long terme, qui est le développement de petites molécules ciblant la Cohésine. De nombreuses études sont ainsi nécessaires pour approfondir nos connaissances sur la Cohésine, sa dynamique, son implication dans les maladies et son ciblage thérapeutique.

Chapitre 4 - Conclusion et discussion générale

I. Conclusion

Au cours de ce travail de thèse, j'ai étudié la dynamique de la Cohésine humaine, en particulier celle liée à la fixation de l'ATP et son hydrolyse au niveau des têtes ATPases de SMC1A et SMC3, sous-unités cœur de la Cohésine. Cette activité ATPase est cruciale pour l'ensemble des fonctions de la Cohésine, notamment au niveau de son interaction avec l'ADN et pour la formation de boucles de chromatine par extrusion. Pour ce faire, j'ai analysé les interactions entre ses sous-unités par des approches biochimiques, biophysiques et structurales.

Mon premier objectif a été de reconstituer le module ATPase, composé des complexes SMC1A-HD/RAD21C et SMC3-HD/RAD21N sous une forme engagée en présence d'ATP, et de le caractériser. Ce complexe est par nature très instable, et j'ai été amenée à essayer de nombreuses approches afin d'identifier les conditions expérimentales nécessaires à la formation d'un complexe suffisamment stable pour sa caractérisation. J'ai ainsi montré que l'utilisation d'ATP, et pas d'analogues faiblement hydrolysables de l'ATP, ainsi que l'utilisation concomitante des mutants EQ des têtes ATPases de SMC1A et SMC3, sont nécessaires à l'obtention d'un tel complexe. La reconstitution du module ATPase engagé a été confirmée par chromatographie d'exclusion de taille, ainsi que par ultracentrifugation analytique. J'ai ensuite poursuivi par sa caractérisation structurale, en utilisant plusieurs techniques : la cristallographie aux rayons X, le SAXS et la cryoEM.

Cette dernière m'a permis d'obtenir une reconstruction du module ATPase à une résolution de 4 Å. Cette structure, en complément des structures des têtes ATPases SMC1A-HD/RAD21C et SMC3-HD/RAD21N indépendantes, sous formes apo, liée à l'ATP-γ-S ou à l'ADP obtenues au sein de mon équipe d'accueil, ainsi que des structures disponibles dans la littérature, a permis de caractériser le cycle ATPase de la Cohésine et de mettre en lumière la dynamique de la Cohésine lors des premières étapes de son cycle, à savoir avant et lors de son chargement sur l'ADN. Mes données mettent en évidence l'importance de l'engagement des

têtes pour les préparer à leur liaison à l'ADN, notamment en conduisant à la flexibilité de la porte de sortie de l'ADN, correspondant à l'interface SMC3/RAD21.

L'engagement des têtes permet également de réguler de manière asymétrique l'activité ATPase des têtes, la tête de SMC3 étant positionnée en conformation active, alors que celle de SMC1A est maintenue en position inactive. L'activation totale des deux têtes ATPases est induite par la fixation de l'ADN et de NIPBL.

De plus, mon équipe d'accueil a montré *in vivo*, chez le poisson zèbre, que cette porte de sortie de l'ADN, notamment l'interaction de RAD21 et la tête globulaire de SMC3, qui régule la fixation de l'ADN sur la tête de SMC3 est essentielle pour le fonctionnement physiologique de la Cohésine. Nos travaux mettent ainsi en lumière la dynamique du cycle ATPase de la Cohésine humaine, en identifiant notamment les changements conformationnels opérés lors de la liaison de l'ATP, de l'engagement des têtes ATPases de SMC1A et de SMC3, ainsi que lors de la fixation de l'ADN et de NIPBL sur ces têtes. A noter que j'ai pu récemment améliorer la résolution à 3,7 Å de la carte cryoEM du module, qui était auparavant de 4 Å. Cela va permettre de caractériser d'autant plus les interactions entre les différentes sous-unités de la Cohésine. Il reste toutefois à comprendre comment les différentes protéines auxiliaires de la Cohésine impactent cette dynamique. Ce qui m'a amené à mon second objectif de thèse, qui a été de comprendre l'influence de STAG1/STAG2 sur le module ATPase humain.

Pour ce faire, j'ai tenté de reconstituer un module ATPase plus complet permettant l'association de STAG1/STAG2 dans le système d'expression bactérien *E. coli*, en essayant deux approches. La première approche a consisté à co-exprimer et co-purifier d'une part le complexe tripartite SMC1-HD/SMC3-HD/RAD21, d'exprimer et de purifier d'autre part STAG1 et STAG2 indépendamment, et de mélanger les quatre constituants lors d'une dernière étape de reconstitution. Au vu des résultats peu fructueux obtenus lors du passage à plus grande échelle, j'ai utilisé une autre approche, qui a impliqué la co-expression et la co-purification des quatre partenaires en même temps. Cette approche reste à optimiser, car je n'ai pas pu effectuer toutes les optimisations nécessaires lors de ma thèse, en testant notamment un autre système d'expression, ainsi que d'autres conditions de purifications. Cela devrait permettre à mon équipe de pouvoir caractériser ce complexe quadripartite de la Cohésine, afin de définir l'influence de STAG1/STAG2 sur le module ATPase et plus particulièrement lors de l'engagement des têtes et de la liaison de l'ADN. Ce projet s'inscrit également dans une

étude de plus grande envergure et à plus long terme de recherche appliquée, permettant le développement de petites molécules ciblant STAG1 et empêchant son incorporation au sein de la Cohésine.

II. Perspectives

Nos travaux apportent ainsi des informations novatrices majeures concernant la dynamique de la Cohésine, en précisant les mécanismes moléculaires du module ATPase de la Cohésine avant et pendant la liaison de l'ATP, lors de l'engagement des têtes et du chargement de ce module sur l'ADN par NIPBL. Toutefois, de nombreuses questions restent en suspens et nécessitent des études supplémentaires. Cela concerne notamment la capture de l'ADN au sein du complexe de la Cohésine. Comment l'ADN et NIPBL s'arriment-ils sur la Cohésine ? S'y fixent-ils simultanément ou successivement ? Cela implique-t-il d'avoir les têtes ATPases indépendantes ou des têtes préalablement engagés induites par la liaison de molécules d'ATP ?

Récemment un modèle de ¨balancement et de clampage¨ a été proposé pour la formation de boucles par extrusion (Bauer et al., 2021). Les auteurs ont défini par FRET à partir de molécule seule, les interactions entre la Cohésine et NIPBL, en absence et en présence d'ATP et d'ADN. D'après ce modèle, avant la liaison de l'ATP, NIPBL, à travers son extrémité N-terminale, interagit avec les régions charnières de SMC1A et de SMC3 et avec l'ADN. Le basculement des régions en superhélice de SMC1A et de SMC3, induit par la flexion au niveau de leur coude, entraine le rapprochement de NIPBL, de l'ADN et des régions charnières vers la tête de SMC3 non engagée. La fixation des molécules d'ATP entraine ensuite l'engagement des têtes de SMC1A et SMC3, qui conduit à des réarrangements structuraux permettant la libération des régions charnières de SMC1A et de SMC3 vis-à-vis de NIPBL, le repositionnement de NIPBL et de l'ADN, ainsi que l'ouverture des régions en superhélice de SMC1A et SMC3. L'ADN, NIPBL et le module ATPase adoptent alors la conformation productive, où l'ADN est pris au piège par l'ensemble des protéines présentes. L'hydrolyse de l'ATP conduit au désengagement des têtes ATPases et à la dissociation de NIPBL. Le cycle peut alors rependre (Figure 48) (Bauer et al., 2021).

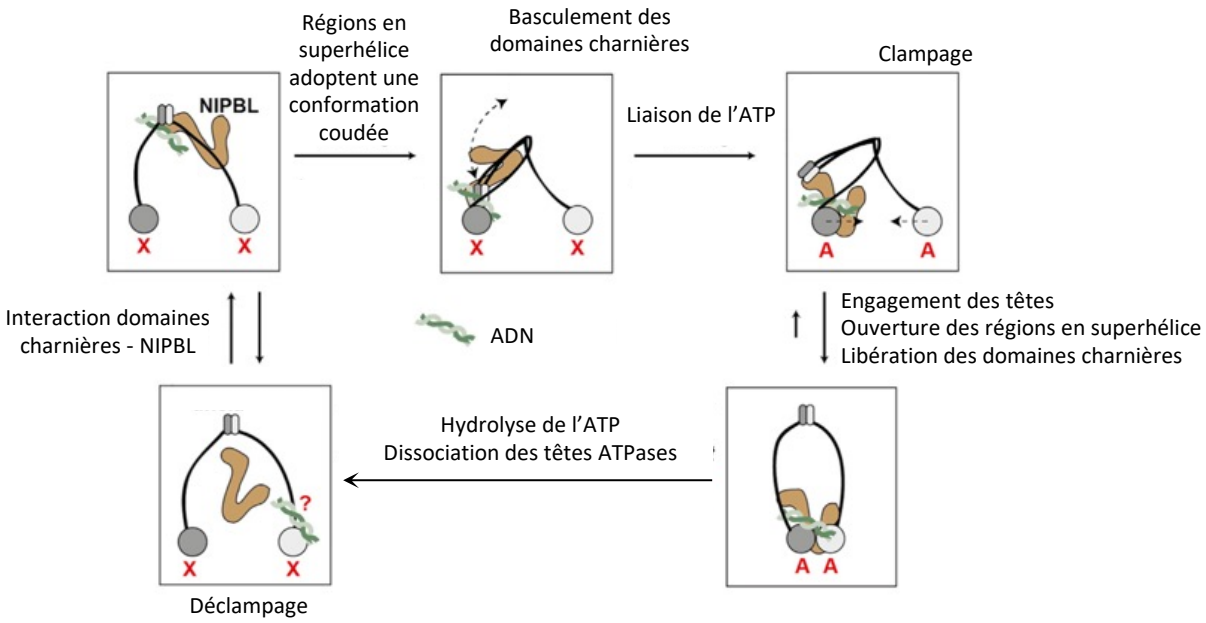


Figure 48 : Modèle du basculement – clampage lors de l'extrusion de boucles par la Cohésine.
X désigne l'état apo des têtes ATPases de SMC1A et SMC3, sans ATP. A correspond aux têtes de SMC1A et SMC3 liées à l'ATP. Adapté de (Bauer et al., 2021).

Nos travaux concordent avec ce modèle. En effet, le complexe SMC3-HD/RAD21N dans sa conformation au repos, sans nucléotide, pourrait servir de point d'ancrage pour NIPBL, à la suite de son rapprochement induit par le basculement des régions en superhélice de SMC1A et SMC3. Lors de cette étape, il reste toutefois à déterminer comment s'agencent les régions charnières, la tête de SMC3, NIPBL et l'ADN. Notamment, l'ADN ne peut pas interagir avec SMC3 en position de repos, comme cela est observé dans le complexe productif, cette interaction étant bloquée par l'interface extensive du lobe globulaire de SMC3 et de RAD21N. Nos travaux sur l'engagement des têtes sont également en accord avec le modèle ¨balancement et clampage¨, qui montrent que l'engagement des têtes conduit à des réarrangements structuraux associés notamment à la flexibilité de la porte de sortie de l'ADN et qui permettent alors à NIPBL et à l'ADN de s'ancrer et de stabiliser les têtes de SMC1A et de SMC3 engagées, adoptant alors un état productif.

Toutefois, plusieurs questions restent ouvertes. Ce modèle n'aborde pas l'ensemble des différentes conformations visualisées par AFM, telle que les conformations en bâtonnet avec les têtes juxtaposées. Les têtes ATPases adoptent-elles les trois conformations (séparés, juxtaposées, engagées) lors de la formation de boucles ? Leurs études de FRET à partir d'une seule molécule sur la forme apo de la Cohésine associée à NIPBL ont été faites en présence

d'ADN, mais y-a-t-il des interactions des régions charnières et des têtes ATPases avec NIPBL en absence d'ADN ? Se forment-elles dans ses conditions ou la présence concomitante de NIPBL et l'ADN est-elle nécessaire ? Qu'en est-il également de STAG1/STAG2 considéré comme la quatrième sous-unité de la Cohésine. Les données actuelles sont controversées : la structure cryoEM du complexe productif humaine de Shi et al (2020) indique que STAG1 pourrait se situer proche des têtes ATPases de SMC1A et SMC3 et interagir avec NIPBL et le domaine charnière de SMC1A.

Au contraire, Bauer et al (2021), par des expériences de FRET à partir de molécule seule, ont démontré que STAG interagit de manière transiente et non stable avec les régions charnières dans le cas où la Cohésine est soit en présence de NIPBL et d'ATP, soit en présence de NIPBL, d'ATP et d'ADN. Cela pourrait s'expliquer par le fait que, d'après ces expériences, la proximité des domaines charnières aux têtes ATPases n'est possible qu'en absence d'ATP. La fixation d'ATP et l'engagement des têtes entrainent l'éloignement des domaines charnières. Il aurait été intéressant de vérifier la position de STAG1 vis-à-vis des têtes SMC1A et SMC3, ainsi que vis-à-vis de NIPBL afin de pouvoir mieux comprendre comment STAG1 se positionne autour du module ATPase.

De plus, la structure récente de la Cohésine de *S. cerevisiae* associé à Scc2 (NIPBL), dans une conformation apo (sans ADN et sans ATP) coudée, montre que Scc2 interagit avec la tête de Smc1 et les domaines charnières de Smc1 et Smc3 (Petela et al., 2021). Associer ces structures à une étape du cycle de régulation de la Cohésine est complexe, du fait que normalement Scc2 participe au chargement de la Cohésine sur l'ADN. Cette structure pourrait correspondre à l'étape de post-hydrolyse de l'ATP, représentant le désengagement des têtes et la libération de l'ADN. Il serait intéressant de faire des expériences de FRET à molécule unique sur un état de post-hydrolyse de la Cohésine, avec l'utilisation d'analogues de l'ATP mimant cet état, comme l'ADP conjugué à l'orthovanadate ($ADP.VO_4^{3-}$), pour confirmer l'interaction entre NIPBL et SMC1A chez l'homme. Il se peut également que cette structure soit spécifique à *S. cerevisiae* et ne s'observe pas pour la Cohésine humaine.

Ainsi mon travail de thèse associé aux travaux menés par mon équipe d'accueil, par comparaison aux structures et données de la littérature, a permis de mettre en lumière les mécanismes moléculaires impliqués de la dynamique de la Cohésine, plus particulièrement lors de la liaison de l'ATP, de l'engagement des têtes de SMC1A et de SMC3, ainsi que lors de

l'interaction de NIPBL et de l'ADN avec la Cohésine. Ces travaux posent les bases pour d'une part mieux comprendre la dynamique de la Cohésine, cruciale pour l'ensemble de ses fonctions, de la cohésion des chromatides sœurs à l'organisation 3D du génome et en passant par la régulation des gènes. D'autre part, une meilleure compréhension du mode d'action physiologique de la Cohésine permet de fournir les bases moléculaires nécessaires aux développements de stratégies thérapeutiques ciblant la Cohésine, dont l'activité est dérégulée dans de nombreuses maladies neurodéveloppementales et dans un grand nombre de cancers. Notamment, une de nos approches serait de développer des petites molécules ciblant STAG1, afin d'empêcher son incorporation au sein de la Cohésine.

Bibliographie

Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A., Adams, P.D., 2018. Real-space refinement in PHENIX for cryo-EM and crystallography. Acta Crystallogr D Struct Biol 74, 531–544. https://doi.org/10.1107/S2059798318006551

Al-Jomah, N., Mukololo, L., Anjum, A., Al Madadha, M., Patel, R., 2020. Pds5A and Pds5B display non-redundant functions in mitosis and their loss triggers Chk1 activation. Frontiers in cell and developmental biology 8, 531.

Allan, J., Harborne, N., Rau, D.C., Gould, H., 1982. Participation of core histone "tails" in the stabilization of the chromatin solenoid. Journal of Cell Biology 93, 285–297. https://doi.org/10.1083/jcb.93.2.285

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. Annual review of genetics 52, 89–107.

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

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

Avery, O.T., MacLeod, C.M., McCarty, M., 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types III : induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. Journal of Experimental Medicine 79, 137–158. https://doi.org/10.1084/jem.79.2.137

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

Badrinarayanan Anjana, Reyes-Lamothe Rodrigo, Uphoff Stephan, Leake Mark C., Sherratt David J., 2012. In Vivo Architecture and Action of Bacterial Structural Maintenance of Chromosome Proteins. Science 338, 528–531. https://doi.org/10.1126/science.1227126

Bailey, C.G., Gupta, S., Metierre, C., Amarasekera, P.M.S., O'Young, P., Kyaw, W., Laletin, T., Francis, H., Semaan, C., Sharifi Tabar, M., Singh, K.P., Mullighan, C.G., Wolkenhauer, O., Schmitz, U., Rasko, J.E.J., 2021. Structure-function relationships explain CTCF zinc finger mutation phenotypes in cancer. Cell Mol Life Sci 78, 7519–7536. https://doi.org/10.1007/s00018-021-03946-z

Barysz, H., Kim, J.H., Chen, Z.A., Hudson, D.F., Rappsilber, J., Gerloff, D.L., Earnshaw, W.C., 2015. Three-dimensional topology of the SMC2/SMC4 subcomplex from chicken condensin I revealed by cross-linking and molecular modelling. Open biology 5, 150005.

Bastié, N., Chapard, C., Dauban, L., Gadal, O., Beckouët, F., Koszul, R., 2022. Smc3 acetylation, Pds5 and Scc2 control the translocase activity that establishes cohesin-dependent chromatin loops. Nature Structural & Molecular Biology 29, 575–585. https://doi.org/10.1038/s41594-022-00780-0

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.

Beckouët, F., Hu, B., Roig, M.B., Sutani, T., Komata, M., Uluocak, P., Katis, V.L., Shirahige, K., Nasmyth, K., 2010. An Smc3 acetylation cycle is essential for establishment of sister chromatid cohesion. Molecular cell 39, 689–699.

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

Belmont, A.S., Sedat, J.W., Agard, D.A., 1987. A three-dimensional approach to mitotic chromosome structure: evidence for a complex hierarchical organization. Journal of Cell Biology 105, 77–92. https://doi.org/10.1083/jcb.105.1.77

Benedetti, L., Cereda, M., Monteverde, L., Desai, N., Ciccarelli, F.D., 2017. Synthetic lethal interaction between the tumour suppressor STAG2 and its paralog STAG1. Oncotarget 8, 37619–37632. https://doi.org/10.18632/oncotarget.16838

Bintu, B., Mateo, L.J., Su, J.-H., Sinnott-Armstrong, N.A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A.N., Zhuang, X., 2018. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science 362, eaau1783. https://doi.org/10.1126/science.aau1783

Birkenbihl, R.P., Subramani, S., 1992. Cloning and characterization of rad21 an essential gene of Schizosaccharomyces pombe involved in DNA double-strand-break repair. Nucleic acids research 20, 6605–6611.

Boettcher, B., Barral, Y., 2013. The cell biology of open and closed mitosis. Nucleus 4, 160–165.

Bolzer, A., Kreth, G., Solovei, I., Koehler, D., Saracoglu, K., Fauth, C., Müller, S., Eils, R., Cremer, C., Speicher, M.R., Cremer, T., 2005. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. PLoS Biol 3, e157–e157. https://doi.org/10.1371/journal.pbio.0030157

Bonora, E., Bianco, F., Cordeddu, L., Bamshad, M., Francescatto, L., Dowless, D., Stanghellini, V., Cogliandro, R.F., Lindberg, G., Mungan, Z., Cefle, K., Ozcelik, T., Palanduz, S., Ozturk, S., Gedikbasi, A., Gori, A., Pippucci, T., Graziano, C., Volta, U., Caio, G., Barbara, G., D’Amato, M., Seri, M., Katsanis, N., Romeo, G., De Giorgio, R., 2015. Mutations in RAD21 disrupt regulation of APOB in patients with chronic intestinal pseudo-obstruction. Gastroenterology 148, 771-782.e11. https://doi.org/10.1053/j.gastro.2014.12.034

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.

Bose, T., Lee, K.K., Lu, S., Xu, B., Harris, B., Slaughter, B., Unruh, J., Garrett, A., McDowell, W., Box, A., Li, H., Peak, A., Ramachandran, S., Seidel, C., Gerton, J.L., 2012. Cohesin Proteins Promote Ribosomal RNA Production and Protein Translation in Yeast and Human Cells. PLOS Genetics 8, e1002749. https://doi.org/10.1371/journal.pgen.1002749

Brautigam, C.A., 2015. Chapter Five - Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data, in: Cole, J.L. (Ed.), Methods in Enzymology. Academic Press, pp. 109–133. https://doi.org/10.1016/bs.mie.2015.05.001

Britton, R.A., Lin, D.C.-H., Grossman, A.D., 1998. Characterization of a prokaryotic SMC protein involved in chromosome partitioning. Genes & development 12, 1254–1259.

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

Bürmann, F., Funke, L.F.H., Chin, J.W., Löwe, J., 2021. Cryo-EM structure of MukBEF reveals DNA loop entrapment at chromosomal unloading sites. Molecular Cell 81, 4891-4906.e8. https://doi.org/10.1016/j.molcel.2021.10.011

Bürmann, F., Lee, B.-G., Than, T., Sinn, L., O'Reilly, F.J., Yatskevich, S., Rappsilber, J., Hu, B., Nasmyth, K., Löwe, J., 2019. A folded conformation of MukBEF and cohesin. Nature Structural & Molecular Biology 26, 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. Nature structural & molecular biology 20, 371–379.

Busslinger, G.A., Stocsits, R.R., van der Lelij, P., Axelsson, E., Tedeschi, A., Galjart, N., Peters, J.-M., 2017. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. Nature 544, 503–507. https://doi.org/10.1038/nature22063

Busso, D., Peleg, Y., Heidebrecht, T., Romier, C., Jacobovitch, Y., Dantes, A., Salim, L., Troesch, E., Schuetz, A., Heinemann, U., Folkers, G.E., Geerlof, A., Wilmanns, M., Polewacz, A., Quedenau, C., Büssow, K., Adamson, R., Blagova, E., Walton, J., Cartwright, J.L., Bird, L.E., Owens, R.J., Berrow, N.S., Wilson, K.S., Sussman, J.L., Perrakis, A., Celie, P.H.N., 2011. Expression of protein complexes using multiple Escherichia coli protein co-expression systems: A benchmarking study. Journal of Structural Biology 175, 159–170. https://doi.org/10.1016/j.jsb.2011.03.004

Camdere, G., Guacci, V., Stricklin, J., Koshland, D., 2015. The ATPases of cohesin interface with regulators to modulate cohesin-mediated DNA tethering. Elife 4, e11315.

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.

Capo-Chichi, J.-M., Bharti, S.K., Sommers, J.A., Yammine, T., Chouery, E., Patry, L., Rouleau, G.A., Samuels, M.E., Hamdan, F.F., Michaud, J.L., Brosh Jr, R.M., Mégarbane, A., Kibar, Z., 2013. Identification and Biochemical Characterization of a Novel Mutation in DDX11 Causing Warsaw Breakage Syndrome. Human Mutation 34, 103–107. https://doi.org/10.1002/humu.22226

Carvajal-Maldonado, D., Byrum, A.K., Jackson, J., Wessel, S., Lemaçon, D., Guitton-Sert, L., Quinet, A., Tirman, S., Graziano, S., Masson, J.-Y., Cortez, D., Gonzalo, S., Mosammaparast, N., Vindigni, A., 2019. Perturbing cohesin dynamics drives MRE11 nuclease-dependent replication fork slowing. Nucleic Acids Research 47, 1294–1310. https://doi.org/10.1093/nar/gky519

Carvalhal, S., Tavares, A., Santos, M.B., Mirkovic, M., Oliveira, R.A., 2018. A quantitative analysis of cohesin decay in mitotic fidelity. Journal of Cell Biology 217, 3343–3353.

Casa, V., Gines, M.M., Gusmao, E.G., Slotman, J.A., Zirkel, A., Josipovic, N., Oole, E., van IJcken, W.F., Houtsmuller, A.B., Papantonis, A., 2020. Redundant and specific roles of cohesin STAG subunits in chromatin looping and transcriptional control. Genome research 30, 515–527.

Castagnetti, S., Oliferenko, S., Nurse, P., 2010. Fission yeast cells undergo nuclear division in the absence of spindle microtubules. PLoS biology 8, e1000512.

Cavarelli, J., 2009. Détermination des structures 3D des macromolécules biologiques par diffraction X. Partie 2. Techniques de l'ingénieur. Analyse et caractérisation.

Chan, K.-L., Gligoris, T., Upcher, W., Kato, Y., Shirahige, K., Nasmyth, K., Beckouët, F., 2013. Pds5 promotes and protects cohesin acetylation. Proc Natl Acad Sci U S A 110, 13020–13025. https://doi.org/10.1073/pnas.1306900110

Chao, W.C.H., Murayama, Y., Muñoz, S., Costa, A., Uhlmann, F., Singleton, M.R.,

2015. Structural Studies Reveal the Functional Modularity of the Scc2-Scc4 Cohesin Loader. Cell Reports 12, 719–725. https://doi.org/10.1016/j.celrep.2015.06.071

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

Chargaff, E., 1950. Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. Experientia 6, 201–209. https://doi.org/10.1007/BF02173653

Chatterjee, N., Walker, G.C., 2017. Mechanisms of DNA damage, repair, and mutagenesis. Environ Mol Mutagen 58, 235–263. https://doi.org/10.1002/em.22087

Cheng, H., Zhang, N., Pati, D., 2020. Cohesin subunit RAD21: From biology to disease. Gene 758, 144966.

Cheng Yifan, 2018. Single-particle cryo-EM—How did it get here and where will it go. Science 361, 876–880. https://doi.org/10.1126/science.aat4346

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

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

Collins, P.L., Purman, C., Porter, S.I., Nganga, V., Saini, A., Hayer, K.E., Gurewitz, G.L., Sleckman, B.P., Bednarski, J.J., Bassing, C.H., Oltz, E.M., 2020. DNA double-strand breaks induce H2Ax phosphorylation domains in a contact-dependent manner. Nature Communications 11, 3158. https://doi.org/10.1038/s41467-020-16926-x

Cooper, G.M., Hausman, R., 2000. A molecular approach. The Cell. 2nd ed. Sunderland, MA: Sinauer Associates.

Cortone, G., Zheng, G., Pensieri, P., Chiappetta, V., Tatè, R., Malacaria, E., Pichierri, P., Yu, H., Pisani, F.M., 2018. Interaction of the Warsaw breakage syndrome DNA helicase DDX11 with the replication fork-protection factor Timeless promotes sister chromatid cohesion. PLOS Genetics 14, e1007622. https://doi.org/10.1371/journal.pgen.1007622

Costa, R.M.A., Chigançcas, V., da Silva Galhardo, R., Carvalho, H., Menck, C.F.M., 2003. The eukaryotic nucleotide excision repair pathway. Biochimie 85, 1083–1099. https://doi.org/10.1016/j.biochi.2003.10.017

Cratsenberg, D.M., Winningham, P.J., Starr, L.J., 2021. Second reported individual with a partial STAG2 deletion: middle interhemispheric variant holoprosencephaly in STAG2-related cohesinopathy. Clinical Dysmorphology 30.

Cremer, M., von Hase, J., Volm, T., Brero, A., Kreth, G., Walter, J., Fischer, C., Solovei, I., Cremer, C., Cremer, T., 2001. Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. Chromosome Research 9, 541–567. https://doi.org/10.1023/A:1012495201697

Cremer, T., Cremer, M., 2010. Chromosome territories. Cold Spring Harb Perspect Biol 2, a003889–a003889. https://doi.org/10.1101/cshperspect.a003889

Crick, F., 1962. The genetic code. Sci Am 207, 66–74. https://doi.org/10.1038/scientificamerican1062-66

Croft, J.A., Bridger, J.M., Boyle, S., Perry, P., Teague, P., Bickmore, W.A., 1999. Differences in the Localization and Morphology of Chromosomes in the Human Nucleus. Journal of Cell Biology 145, 1119–1131. https://doi.org/10.1083/jcb.145.6.1119

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

Dahm, R., 2008. Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. Human Genetics 122, 565–581. https://doi.org/10.1007/s00439-007-0433-0

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

Danilova, O., Reyes-Lamothe, R., Pinskaya, M., Sherratt, D., Possoz, C., 2007. MukB colocalizes with the oriC region and is required for organization of the two Escherichia coli chromosome arms into separate cell halves. Molecular microbiology 65, 1485–1492.

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

Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., Richmond, T.J., 2002. Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution††We dedicate this paper to the memory of Max Perutz who was particularly inspirational and supportive to T.J.R. in the early stages of this study. Journal of Molecular Biology 319, 1097–1113. https://doi.org/10.1016/S0022-2836(02)00386-8

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

Davis, A.J., Chen, D.J., 2013. DNA double strand break repair via non-homologous end-joining. Transl Cancer Res 2, 130–143. https://doi.org/10.3978/j.issn.2218-676X.2013.04.02

De Campos Silva, T., Deb, S., Xu, H., Yan, Y., Fox, S.B., Mortensen, N., McKay, M., 2014. Effect of RAD21 overexpression on prognosis in KRAS-mutant colorectal carcinomas. JCO 32, 453–453. https://doi.org/10.1200/jco.2014.32.3_suppl.453

De Souza, C.P., Osmani, S.A., 2007. Mitosis, not just open or closed. Eukaryotic cell 6, 1521–1527.

Deardorff, M.A., Bando, M., Nakato, R., Watrin, E., Itoh, T., Minamino, M., Saitoh, K., Komata, M., Katou, Y., Clark, D., Cole, K.E., De Baere, E., Decroos, C., Di Donato, N., Ernst, S., Francey, L.J., Gyftodimou, Y., Hirashima, K., Hullings, M., Ishikawa, Y., Jaulin, C., Kaur, M., Kiyono, T., Lombardi, P.M., Magnaghi-Jaulin, L., Mortier, G.R., Nozaki, N., Petersen, M.B., Seimiya, H., Siu, V.M., Suzuki, Y., Takagaki, K., Wilde, J.J., Willems, P.J., Prigent, C., Gillessen-Kaesbach, G., Christianson, D.W., Kaiser, F.J., Jackson, L.G., Hirota, T., Krantz, I.D., Shirahige, K., 2012. HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. Nature 489, 313.

Deardorff, M.A., Kaur, M., Yaeger, D., Rampuria, A., Korolev, S., Pie, J., Gil-Rodríguez, C., Arnedo, M., Loeys, B., Kline, A.D., 2007. Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation. The American Journal of Human Genetics 80, 485–494.

Dekker, J., Misteli, T., 2015. Long-range chromatin interactions. Cold Spring Harbor perspectives in biology 7, a019356.

Dequeker, B.J., Scherr, M.J., Brandão, H.B., Gassler, J., Powell, S., Gaspar, I., Flyamer, I.M., Lalic, A., Tang, W., Stocsits, R., 2022. MCM complexes are barriers that restrict cohesin-mediated loop extrusion. Nature 606, 197–203.

de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Verstegen, M.J.A.M.,

Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H.L., de Laat, W., 2015. CTCF Binding Polarity Determines Chromatin Looping. Molecular Cell 60, 676–684. https://doi.org/10.1016/j.molcel.2015.09.023

Di Nardo, M., Pallotta, M.M., Musio, A., 2022. The multifaceted roles of cohesin in cancer. Journal of Experimental & Clinical Cancer Research 41, 1–11.

Diebold, M.-L., Fribourg, S., Koch, M., Metzger, T., Romier, C., 2011. Deciphering correct strategies for multiprotein complex assembly by co-expression: Application to complexes as large as the histone octamer. Journal of Structural Biology 175, 178–188. https://doi.org/10.1016/j.jsb.2011.02.001

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., 2017a. 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

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., 2017b. 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

Doerr, A., 2016. Single-particle cryo-electron microscopy. Nature Methods 13, 23–23. https://doi.org/10.1038/nmeth.3700

Dorsett, D., Ström, L., 2012. The ancient and evolving roles of cohesin in gene expression and DNA repair. Current Biology 22, R240–R250.

Douzi, B., 2017. Protein–protein interactions: surface plasmon resonance, in: Bacterial Protein Secretion Systems. Springer, pp. 257–275.

Duan, X., Yang, Y., Chen, Y.-H., Arenz, J., Rangi, G.K., Zhao, X., Ye, H., 2009. Architecture of the Smc5/6 Complex of Saccharomyces cerevisiae Reveals a Unique Interaction between the Nse5-6 Subcomplex and the Hinge Regions of Smc5 and Smc6 *. Journal of Biological Chemistry 284, 8507–8515. https://doi.org/10.1074/jbc.M809139200

Durbin, R.M., Altshuler, D., Abecasis, G.R., Bentley, D.R., Chakravarti, A., De La Vega, F.M., Donnelly, P., Egholm, M., Gabriel, S.B., Lander, E.S., Lehrach, H., Mardis, E.R., McVean, G.A., Nickerson, D.A., Schafer, A.J., Sherry, S.T., Wang, Jun, Gibbs, R.A., Deiros, D., Metzker, M., Reid, J., Wheeler, D., Wang, Jun, Li, J., Jian, M., Li, G., Liang, H., Tian, G., Wang, B., Wang, Jian, Wang, W., Yang, H., Zheng, Huisong, Lander, E.S., Altshuler, D., Ambrogio, L., Bloom, T., Cibulskis, K., Fennell, T.J., Gabriel, S.B., Jaffe, D.B., Shefler, E., Sougnez, C.L., Bentley, D.R., Gormley, N., Humphray, S., Kokko-Gonzales, P., Stone, J., McKernan, K.J., Costa, G.L., Ichikawa, J.K., Lee, C.C., Sudbrak, R., Lehrach, H., Borodina, T.A., Dahl, A., Davydov, A.N., Marquardt, P., Mertes, F., Rosenstiel, P., Schreiber, S., Soldatov, A.V., Timmermann, B., Tolzmann, M., Egholm, M., Affourtit, J., Ashworth, D., Attiya, S., Bachorski, M., Buglione, E., Burke, A., Caprio, A., Celone, C., Clark, S., Desany, B., Gu, L., Guccione, L., Kao, K., Kebbel, A., Knowlton, J., Labrecque, M., McDade, L., Mealmaker, C., Minderman, M., Nawrocki, A., Niazi, F., Pareja, K., Ramenani, R., Riches, D., Song, W., Turcotte, C., Wang, S., Mardis, E.R., Wilson, R.K., Dooling, D., Fulton, L., Fulton, R., Weinstock, G., Durbin, R.M., Burton, J., Carter, D.M., Churcher, C., Coffey, A., Cox, A., Palotie, A., Quail, M., Skelly, T., Stalker, J., Swerdlow, H.P., Turner, D., De Witte, A., Giles, S., Gibbs, R.A., Wheeler, D., Bainbridge, M., Challis, D., Sabo, A., Yu, F., Yu, J., Wang, Jun, Fang, X., Guo, X., Li, R., Li, Y., Luo, R., Tai, S., Wu, H., Zheng, Hancheng, Zheng, X., Zhou, Y., Li, G., Wang, Jian, Yang, H., Marth, G.T., Garrison, E.P., Huang, W., Indap, A., Kural, D., Lee, W.-P., Fung Leong, W., Quinlan, A.R., Stewart, C., Stromberg, M.P., Ward, A.N., Wu, J.,

Lee, C., Mills, R.E., Shi, X., Daly, M.J., DePristo, M.A., Altshuler, D., Ball, A.D., Banks, E., Bloom, T., Browning, B.L., Cibulskis, K., Fennell, T.J., Garimella, K.V., Grossman, S.R., Handsaker, R.E., Hanna, M., Hartl, C., Jaffe, D.B., Kernytsky, A.M., Korn, J.M., Li, H., Maguire, J.R., McCarroll, S.A., McKenna, A., Nemesh, J.C., Philippakis, A.A., Poplin, R.E., Price, A., Rivas, M.A., Sabeti, P.C., Schaffner, S.F., Shefler, E., Shlyakhter, I.A., Cooper, D.N., Ball, E.V., Mort, M., Phillips, A.D., Stenson, P.D., Sebat, J., Makarov, V., Ye, Kenny, Yoon, S.C., Bustamante, C.D., Clarke, L., Flicek, P., Cunningham, F., Herrero, J., Keenen, S., Kulesha, E., Leinonen, R., McLaren, W.M., Radhakrishnan, R., Smith, R.E., Zalunin, V., Zheng-Bradley, X., Korbel, J.O., Stütz, A.M., Humphray, S., Bauer, M., Keira Cheetham, R., Cox, T., Eberle, M., James, T., Kahn, S., Murray, L., Chakravarti, A., Ye, Kai, De La Vega, F.M., Fu, Y., Hyland, F.C.L., Manning, J.M., McLaughlin, S.F., Peckham, H.E., Sakarya, O., Sun, Y.A., Tsung, E.F., Batzer, M.A., Konkel, M.K., Walker, J.A., Sudbrak, R., Albrecht, M.W., Amstislavskiy, V.S., Herwig, R., Parkhomchuk, D.V., Sherry, S.T., Agarwala, R., Khouri, H.M., Morgulis, A.O., Paschall, J.E., Phan, L.D., Rotmistrovsky, K.E., Sanders, R.D., Shumway, M.F., Xiao, C., McVean, G.A., Auton, A., Iqbal, Z., Lunter, G., Marchini, J.L., Moutsianas, L., Myers, S., Tumian, A., Desany, B., Knight, J., Winer, R., Craig, D.W., Beckstrom-Sternberg, S.M., Christoforides, A., The 1000 Genomes Project Consortium, Corresponding author, Steering committee, Production group: Baylor College of Medicine, BGI-Shenzhen, Broad Institute of MIT and Harvard, Illumina, Life Technologies, Max Planck Institute for Molecular Genetics, Roche Applied Science, Washington University in St Louis, Wellcome Trust Sanger Institute, Analysis group: Agilent Technologies, Baylor College of Medicine, Boston College, Brigham and Women's Hospital, Cardiff University, T.H.G.M.D., Cold Spring Harbor Laboratory, Cornell and Stanford Universities, European Bioinformatics Institute, European Molecular Biology Laboratory, Johns Hopkins University, Leiden University Medical Center, Louisiana State University, US National Institutes of Health, Oxford University, The Translational Genomics Research Institute, 2010. A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073. https://doi.org/10.1038/nature09534

Ebel, C., Birck, C., 2021. Sedimentation Velocity Methods for the Characterization of Protein Heterogeneity and Protein Affinity Interactions, in: Poterszman, A. (Ed.), Multiprotein Complexes: Methods and Protocols. Springer US, New York, NY, pp. 155–171. https://doi.org/10.1007/978-1-0716-1126-5_9

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

Eeftens, J.M., Katan, A.J., Kschonsak, M., Hassler, M., de Wilde, L., Dief, E.M., Haering, C.H., Dekker, C., 2016. Condensin Smc2-Smc4 Dimers Are Flexible and Dynamic. Cell Reports 14, 1813–1818. https://doi.org/10.1016/j.celrep.2016.01.063

Eichinger, C.S., Kurze, A., Oliveira, R.A., Nasmyth, K., 2013. Disengaging the Smc3/kleisin interface releases cohesin from Drosophila chromosomes during interphase and mitosis. The EMBO Journal 32, 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

Emerson, D.J., Zhao, P.A., Cook, A.L., Barnett, R.J., Klein, K.N., Saulebekova, D., Ge, C., Zhou, L., Simandi, Z., Minsk, M.K., 2022. Cohesin-mediated loop anchors confine the locations of human replication origins. Nature 1–8.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66, 486–501.

https://doi.org/10.1107/S0907444910007493

Ermantraut, E., Wohlfart, K., Tichelaar, W., 1998. Perforated support foils with pre-defined hole size, shape and arrangement. Ultramicroscopy 74, 75–81. https://doi.org/10.1016/S0304-3991(98)00025-4

Fakhro, K.A., Staudt, M.R., Ramstetter, M.D., Robay, A., Malek, J.A., Badii, R., Al-Marri, A.A.-N., Khalil, C.A., Al-Shakaki, A., Chidiac, O., Stadler, D., Zirie, M., Jayyousi, A., Salit, J., Mezey, J.G., Crystal, R.G., Rodriguez-Flores, J.L., 2016. The Qatar genome: a population-specific tool for precision medicine in the Middle East. Human Genome Variation 3, 16016. https://doi.org/10.1038/hgv.2016.16

Fennell-Fezzie, R., Gradia, S.D., Akey, D., Berger, J.M., 2005. The MukF subunit of Escherichia coli condensin: architecture and functional relationship to kleisins. The EMBO Journal 24, 1921–1930. https://doi.org/10.1038/sj.emboj.7600680

Finch J T, Klug A, 1976. Solenoidal model for superstructure in chromatin. Proceedings of the National Academy of Sciences 73, 1897–1901. https://doi.org/10.1073/pnas.73.6.1897

Fioramonte, M., Gozzo, F.C., de Oliveira, C.L.P., Portugal, R.V., van Heel, M., 2017. Advance Techniques in Biophysics, in: Misra, G. (Ed.), Introduction to Biomolecular Structure and Biophysics: Basics of Biophysics. Springer Singapore, Singapore, pp. 245–274. https://doi.org/10.1007/978-981-10-4968-2_9

Flemming, W., 1882. Zellsubstanz, kern und zelltheilung. Vogel.

Fousteri, M.I., Lehmann, A.R., 2000. A novel SMC protein complex in Schizosaccharomyces pombe contains the Rad18 DNA repair protein. The EMBO Journal 19, 1691–1702. https://doi.org/10.1093/emboj/19.7.1691

Frattini, C., Villa-Hernández, S., Pellicanò, G., Jossen, R., Katou, Y., Shirahige, K., Bermejo, R., 2017. Cohesin Ubiquitylation and Mobilization Facilitate Stalled Replication Fork Dynamics. Molecular Cell 68, 758-772.e4. https://doi.org/10.1016/j.molcel.2017.10.012

Freeman, L., Aragon-Alcaide, L., Strunnikov, A., 2000. The condensin complex governs chromosome condensation and mitotic transmission of rDNA. The Journal of cell biology 149, 811–824.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., Mirny, L.A., 2016. Formation of Chromosomal Domains by Loop Extrusion. Cell Reports 15, 2038–2049. https://doi.org/10.1016/j.celrep.2016.04.085

Fuentes-Perez, M., Gwynn, E., Dillingham, M., Moreno-Herrero, F., 2012. Using DNA as a fiducial marker to study SMC complex interactions with the atomic force microscope. Biophysical journal 102, 839–848.

Funato, M., Uehara, T., Okada, Y., Kaneko, H., Kosaki, K., 2022. Cohesinopathy presenting with microtia, facial palsy, and hearing loss caused by STAG1 pathogenic variant. Congenital Anomalies 62, 82–83. https://doi.org/10.1111/cga.12454

Gerlich, D., Hirota, T., Koch, B., Peters, J.-M., Ellenberg, J., 2006a. Condensin I Stabilizes Chromosomes Mechanically through a Dynamic Interaction in Live Cells. Current Biology 16, 333–344. https://doi.org/10.1016/j.cub.2005.12.040

Gerlich, D., Koch, B., Dupeux, F., Peters, J.-M., Ellenberg, J., 2006b. Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication. Current Biology 16, 1571–1578.

Gibcus Johan H., Samejima Kumiko, Goloborodko Anton, Samejima Itaru, Naumova Natalia, Nuebler Johannes, Kanemaki Masato T., Xie Linfeng, Paulson James R., Earnshaw William C., Mirny Leonid A., Dekker Job, 2018. A pathway for mitotic chromosome formation. Science 359, eaao6135. https://doi.org/10.1126/science.aao6135

Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.-Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E.A., Young, L., Qi,

Z.-Q., Segall-Shapiro, T.H., Calvey, C.H., Parmar, P.P., Hutchison, C.A. 3rd, Smith, H.O., Venter, J.C., 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329, 52–56. https://doi.org/10.1126/science.1190719

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A. 3rd, Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6, 343–345. https://doi.org/10.1038/nmeth.1318

Gligoris, T.G., Scheinost, J.C., Bürmann, F., Petela, N., Chan, K.-L., Uluocak, P., Beckouët, F., Gruber, S., Nasmyth, K., Löwe, J., 2014. Closing the cohesin ring: structure and function of its Smc3-kleisin interface. Science (New York, N.Y.) 346, 963–967. https://doi.org/10.1126/science.1256917

Golov, A.K., Golova, A.V., Gavrilov, A.A., Razin, S.V., 2021. Sensitivity of cohesin–chromatin association to high-salt treatment corroborates non-topological mode of loop extrusion. Epigenetics & Chromatin 14, 36. https://doi.org/10.1186/s13072-021-00411-w

Gonen, S., 2021. Progress Towards CryoEM: Negative-StainNegative stain Procedures for Biological Samples, in: Gonen, T., Nannenga, B.L. (Eds.), CryoEM: Methods and Protocols. Springer US, New York, NY, pp. 115–123. https://doi.org/10.1007/978-1-0716-0966-8_5

Green, N.S., Reisler, E., Houk, K., 2001. Quantitative evaluation of the lengths of homobifunctional protein cross-linking reagents used as molecular rulers. Protein science 10, 1293–1304.

Greenwald, W.W., Li, H., Benaglio, P., Jakubosky, D., Matsui, H., Schmitt, A., Selvaraj, S., D'Antonio, M., D'Antonio-Chronowska, A., Smith, E.N., Frazer, K.A., 2019. Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression. Nat Commun 10, 1054–1054. https://doi.org/10.1038/s41467-019-08940-5

Gruber, S., Haering, C.H., Nasmyth, K., 2003. Chromosomal cohesin forms a ring. Cell 112, 765–777.

Guillou, E., Ibarra, A., Coulon, V., Casado-Vela, J., Rico, D., Casal, I., Schwob, E., Losada, A., Méndez, J., 2010. Cohesin organizes chromatin loops at DNA replication factories. Genes & development 24, 2812–2822.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., 2015. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. Cell 162, 900–910.

Guo, Y.A., Chang, M.M., Huang, W., Ooi, W.F., Xing, M., Tan, P., Skanderup, A.J., 2018. Mutation hotspots at CTCF binding sites coupled to chromosomal instability in gastrointestinal cancers. Nature communications 9, 1–14.

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

Haarhuis, J.H.I., van der Weide, R.H., Blomen, V.A., Yáñez-Cuna, J.O., Amendola, M., van Ruiten, M.S., Krijger, P.H.L., Teunissen, H., Medema, R.H., van Steensel, B., Brummelkamp, T.R., de Wit, E., Rowland, B.D., 2017. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. Cell 169, 693-707.e14. https://doi.org/10.1016/j.cell.2017.04.013

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

Haering, C.H., Gruber, S., 2016. SnapShot: SMC protein complexes part I. Cell 164, 326–326.

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

Haffke, M., Marek, M., Pelosse, M., Diebold, M.-L., Schlattner, U., Berger, I., Romier, C., 2015. Characterization and production of protein complexes by co-expression in Escherichia coli. Methods Mol Biol 1261, 63–89. https://doi.org/10.1007/978-1-4939-2230-7_4

Hallett, S.T., Schellenberger, P., Zhou, L., Beuron, F., Morris, E., Murray, J.M., Oliver, A.W., 2021. Nse5/6 is a negative regulator of the ATPase activity of the Smc5/6 complex. Nucleic Acids Research 49, 4534–4549. https://doi.org/10.1093/nar/gkab234

Hammond, C., 2015. The basics of crystallography and diffraction. International Union of Crystal.

Hara, K., Zheng, G., Qu, Q., Liu, H., Ouyang, Z., Chen, Z., Tomchick, D.R., Yu, H., 2014. Structure of cohesin subcomplex pinpoints direct shugoshin-Wapl antagonism in centromeric cohesion. Nat Struct Mol Biol 21, 864–870. https://doi.org/10.1038/nsmb.2880

Harashima, H., Dissmeyer, N., Schnittger, A., 2013. Cell cycle control across the eukaryotic kingdom. Trends in Cell Biology 23, 345–356. https://doi.org/10.1016/j.tcb.2013.03.002

Heidinger-Pauli, J.M., Mert, O., Davenport, C., Guacci, V., Koshland, D., 2010. Systematic Reduction of Cohesin Differentially Affects Chromosome Segregation, Condensation, and DNA Repair. Current Biology 20, 957–963. https://doi.org/10.1016/j.cub.2010.04.018

Heitz, E., 1928. Das heterochromatin der moose. Bornträger.

Hershey, A.D., Chase, M., 1952. INDEPENDENT FUNCTIONS OF VIRAL PROTEIN AND NUCLEIC ACID IN GROWTH OF BACTERIOPHAGE. Journal of General Physiology 36, 39–56. https://doi.org/10.1085/jgp.36.1.39

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. Mol Cell 79, 917-933.e9. https://doi.org/10.1016/j.molcel.2020.07.013

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

Hirano, T., 2006. At the heart of the chromosome: SMC proteins in action. Nature Reviews Molecular Cell Biology 7, 311–322. https://doi.org/10.1038/nrm1909

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., Kobayashi, R., Hirano, M., 1997. Condensins, Chromosome Condensation Protein Complexes Containing XCAP-C, XCAP-E and a Xenopus Homolog of the Drosophila Barren Protein. Cell 89, 511–521. https://doi.org/10.1016/S0092-8674(00)80233-0

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

Hirota, T., Gerlich, D., Koch, B., Ellenberg, J., Peters, J.-M., 2004. Distinct functions of condensin I and II in mitotic chromosome assembly. Journal of Cell Science 117, 6435–6445. https://doi.org/10.1242/jcs.01604

Hoffmann, F., 2020. Introduction to crystallography. Springer Nature.

Holohan, E.E., Kwong, C., Adryan, B., Bartkuhn, M., Herold, M., Renkawitz, R., Russell, S., White, R., 2007. CTCF Genomic Binding Sites in Drosophila and the Organisation

of the Bithorax Complex. PLOS Genetics 3, e112. https://doi.org/10.1371/journal.pgen.0030112

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. Nature Communications 7, 12523. https://doi.org/10.1038/ncomms12523

Hopfner, K.-P., Tainer, J.A., 2003. Rad50/SMC proteins and ABC transporters: unifying concepts from high-resolution structures. Current Opinion in Structural Biology 13, 249–255. https://doi.org/10.1016/S0959-440X(03)00037-X

Hou, F., Zou, H., 2005. Two human orthologues of Eco1/Ctf7 acetyltransferases are both required for proper sister-chromatid cohesion. Molecular biology of the cell 16, 3908–3918.

Howard, A., 1953. Synthesis of deoxyribonucleic acid in normal and irradiated ceils and its relation to chromosome breakage. Heredity suppl 6, 261–273.

Hsieh, T.-H.S., Cattoglio, C., Slobodyanyuk, E., Hansen, A.S., Rando, O.J., Tjian, R., Darzacq, X., 2020. Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. Molecular Cell. https://doi.org/10.1016/j.molcel.2020.03.002

Hsieh, T.-H.S., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., Rando, O.J., 2015. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. Cell 162, 108–119. https://doi.org/10.1016/j.cell.2015.05.048

Hsu, S.C., Gilgenast, T.G., Bartman, C.R., Edwards, C.R., Stonestrom, A.J., Huang, P., Emerson, D.J., Evans, P., Werner, M.T., Keller, C.A., 2017. The BET protein BRD2 cooperates with CTCF to enforce transcriptional and architectural boundaries. Molecular cell 66, 102–116.

Hu, B., Itoh, T., Mishra, A., Katoh, Y., Chan, K.-L., Upcher, W., Godlee, C., Roig, M.B., Shirahige, K., Nasmyth, K., 2011. ATP Hydrolysis Is Required for Relocating Cohesin from Sites Occupied by Its Scc2/4 Loading Complex. Current biology : CB 21, 12–24. https://doi.org/10.1016/j.cub.2010.12.004

Hustedt, N., Durocher, D., 2017. The control of DNA repair by the cell cycle. Nature cell biology 19, 1–9.

Iashina, E.G., Velichko, E.V., Filatov, M.V., Bouwman, W.G., Duif, C.P., Brulet, A., Grigoriev, S.V., 2017. Additive scaling law for structural organization of chromatin in chicken erythrocyte nuclei. Phys. Rev. E 96, 012411. https://doi.org/10.1103/PhysRevE.96.012411

International Human Genome Sequencing Consortium, 2004. Finishing the euchromatic sequence of the human genome. Nature 431, 931–945. https://doi.org/10.1038/nature03001

Ishida, T., Kinoshita, K., 2007. PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic acids research 35, W460–W464.

Ivanov, D., Nasmyth, K., 2005. A topological interaction between cohesin rings and a circular minichromosome. Cell 122, 849–860.

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

Jerkovic´, I., Cavalli, G., 2021. Understanding 3D genome organization by multidisciplinary methods. Nature Reviews Molecular Cell Biology 22, 511–528. https://doi.org/10.1038/s41580-021-00362-w

Jessberger, R., Riwar, B., Baechtold, H., Akhmedov, A.T., 1996. SMC proteins constitute two subunits of the mammalian recombination complex RC-1. The EMBO journal 15, 4061–4068.

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

Kamada, K., Su'etsugu, M., Takada, H., Miyata, M., Hirano, T., 2017. Overall Shapes of the SMC-ScpAB Complex Are Determined by Balance between Constraint and Relaxation of Its Structural Parts. Structure 25, 603-616.e4. https://doi.org/10.1016/j.str.2017.02.008

Kantaputra, P.N., Dejkhamron, P., Intachai, W., Ngamphiw, C., Kawasaki, K., Ohazama, A., Krisanaprakornkit, S., Olsen, B., Tongsima, S., Ketudat Cairns, J.R., 2021. Juberg-Hayward syndrome is a cohesinopathy, caused by mutation in ESCO2. European Journal of Orthodontics 43, 45–50. https://doi.org/10.1093/ejo/cjaa023

Karanam, K., Kafri, R., Loewer, A., Lahav, G., 2012. Quantitative Live Cell Imaging Reveals a Gradual Shift between DNA Repair Mechanisms and a Maximal Use of HR in Mid S Phase. Molecular Cell 47, 320–329. https://doi.org/10.1016/j.molcel.2012.05.052

Katainen, R., Dave, K., Pitkänen, E., Palin, K., Kivioja, T., Välimäki, N., Gylfe, A.E., Ristolainen, H., Hänninen, U.A., Cajuso, T., 2015. CTCF/cohesin-binding sites are frequently mutated in cancer. Nature genetics 47, 818–821.

Kawasumi, R., Abe, T., Arakawa, H., Garre, M., Hirota, K., Branzei, D., 2017. ESCO1/2's roles in chromosome structure and interphase chromatin organization. Genes & development 31, 2136–2150.

Khorana, H.G., 1968. Synthetic Nucleic Acids and the Genetic Code. JAMA 206, 1978–1982. https://doi.org/10.1001/jama.1968.03150090054013

Kikhney, A.G., Svergun, D.I., 2015. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. FEBS letters 589, 2570–2577.

Kim, B.-J., Li, Yehua, Zhang, J., Xi, Y., Li, Yumei, Yang, T., Jung, S.Y., Pan, X., Chen, R., Li, W., Wang, Y., Qin, J., 2010. Genome-wide Reinforcement of Cohesin Binding at Pre-existing Cohesin Sites in Response to Ionizing Radiation in Human Cells *. Journal of Biological Chemistry 285, 22784–22792. https://doi.org/10.1074/jbc.M110.134577

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

Kline, A.D., Moss, J.F., Selicorni, A., Bisgaard, A.-M., Deardorff, M.A., Gillett, P.M., Ishman, S.L., Kerr, L.M., Levin, A.V., Mulder, P.A., Ramos, F.J., Wierzba, J., Ajmone, P.F., Axtell, D., Blagowidow, N., Cereda, A., Costantino, A., Cormier-Daire, V., FitzPatrick, D., Grados, M., Groves, L., Guthrie, W., Huisman, S., Kaiser, F.J., Koekkoek, G., Levis, M., Mariani, M., McCleery, J.P., Menke, L.A., Metrena, A., O'Connor, J., Oliver, C., Pie, J., Piening, S., Potter, C.J., Quaglio, A.L., Redeker, E., Richman, D., Rigamonti, C., Shi, A., Tümer, Z., Van Balkom, I.D.C., Hennekam, R.C., 2018. Diagnosis and management of Cornelia de Lange syndrome: first international consensus statement. Nature Reviews Genetics 19, 649–666. https://doi.org/10.1038/s41576-018-0031-0

Kogut, I., Wang, J., Guacci, V., Mistry, R.K., Megee, P.C., 2009. The Scc2/Scc4 cohesin loader determines the distribution of cohesin on budding yeast chromosomes. Genes & development 23, 2345–2357.

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

Kong Xiangduo, Ball Alexander R., Pham Hoang Xuan, Zeng Weihua, Chen Hsiao-Yuan, Schmiesing John A., Kim Jong-Soo, Berns Michael, Yokomori Kyoko, 2014. Distinct Functions of Human Cohesin-SA1 and Cohesin-SA2 in Double-Strand Break Repair. Molecular and Cellular Biology 34, 685–698. https://doi.org/10.1128/MCB.01503-13

Kornberg Roger D., 1974. Chromatin Structure: A Repeating Unit of Histones and

DNA. Science 184, 868–871. https://doi.org/10.1126/science.184.4139.868

Krantz, I.D., McCallum, J., DeScipio, C., Kaur, M., Gillis, L.A., Yaeger, D., Jukofsky, L., Wasserman, N., Bottani, A., Morris, C.A., 2004. Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. Nature genetics 36, 631–635.

Krokan, H.E., Bjørås, M., 2013. Base excision repair. Cold Spring Harbor perspectives in biology 5, a012583.

Kruszka, P., Berger, S.I., Casa, V., Dekker, M.R., Gaesser, J., Weiss, K., Martinez, A.F., Murdock, D.R., Louie, R.J., Prijoles, E.J., 2019. Cohesin complex-associated holoprosencephaly. Brain 142, 2631–2643.

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

Kuei, B., Aplan, M.P., Litofsky, J.H., Gomez, E.D., 2020. New opportunities in transmission electron microscopy of polymers. Materials Science and Engineering: R: Reports 139, 100516.

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

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

Kuwada, Y., 1939. Chromosome Structure A critical review. Cytologia 10, 213–256.

Lacabanne, D., Wiegand, T., Wili, N., Kozlova, M.I., Cadalbert, R., Klose, D., Mulkidjanian, A.Y., Meier, B.H., Böckmann, A., 2020. ATP analogues for structural investigations: case studies of a DnaB helicase and an ABC transporter. Molecules 25, 5268.

Ladurner, R., Bhaskara, V., Huis, 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.

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. The EMBO journal 35, 635–653.

Laemmli, U., Cheng, S., Adolph, K., Paulson, J., Brown, J., Baumbach, W., 1978. Metaphase chromosome structure: the role of nonhistone proteins. Presented at the Cold Spring Harbor symposia on quantitative biology, Cold Spring Harbor Laboratory Press, pp. 351–360.

Lammens, A., Schele, A., Hopfner, K.-P., 2004. Structural biochemistry of ATP-driven dimerization and DNA-stimulated activation of SMC ATPases. Current Biology 14, 1778–1782.

Larionov, V.L., Karpova, T.S., Kouprina, N.Y., Jouravleva, G.A., 1985. A mutant of Saccharomyces cerevisiae with impaired maintenance of centromeric plasmids. Current Genetics 10, 15–20. https://doi.org/10.1007/BF00418488

Laue, T.M., 1992. Computer-aided interpretation of analytical sedimentation data for proteins. Analytical ultracentrifugation in biochemistry and polymer science. 90–125.

Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., Getz, G., 2014. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 505, 495.

Lazar-Stefanita, L., Scolari, V.F., Mercy, G., Muller, H., Guérin, T.M., Thierry, A., Mozziconacci, J., Koszul, R., 2017. Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle. The EMBO Journal 36, 2684–2697.

https://doi.org/10.15252/embj.201797342

Lee, B.-G., Merkel, F., Allegretti, M., Hassler, M., Cawood, C., Lecomte, L., O'Reilly, F.J., Sinn, L.R., Gutierrez-Escribano, P., Kschonsak, M., 2020. Cryo-EM structures of holo condensin reveal a subunit flip-flop mechanism. Nature structural & molecular biology 27, 743–751.

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

Lee Byung-Gil, Rhodes James, Löwe Jan, 2022. Clamping of DNA shuts the condensin neck gate. Proceedings of the National Academy of Sciences 119, e2120006119. https://doi.org/10.1073/pnas.2120006119

Lehalle, D., Mosca-Boidron, A.-L., Begtrup, A., Boute-Benejean, O., Charles, P., Cho, M.T., Clarkson, A., Devinsky, O., Duffourd, Y., Duplomb-Jego, L., 2017. STAG1 mutations cause a novel cohesinopathy characterised by unspecific syndromic intellectual disability. Journal of medical genetics 54, 479–488.

Leman, A.R., Noguchi, C., Lee, C.Y., Noguchi, E., 2010. Human Timeless and Tipin stabilize replication forks and facilitate sister-chromatid cohesion. Journal of cell science 123, 660–670.

Levene, P., 1919. The structure of yeast nucleic acid. Studies from the Rockefeller Institute for Medical Research 30, 221.

Li, D., He, M., Tang, Q., Tian, S., Zhang, J., Li, Y., Wang, D., Jin, L., Ning, C., Zhu, W., Hu, S., Long, K., Ma, J., Liu, J., Zhang, Z., Li, M., 2022. Comparative 3D genome architecture in vertebrates. BMC Biol 20, 99–99. https://doi.org/10.1186/s12915-022-01301-7

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

Lian, A.T.Y., Chircop, M., 2016. Mitosis in Animal Cells, in: Bradshaw, R.A., Stahl, P.D. (Eds.), Encyclopedia of Cell Biology. Academic Press, Waltham, pp. 478–493. https://doi.org/10.1016/B978-0-12-394447-4.30064-5

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

Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J., Moriarty, N.W., Oeffner, R.D., Poon, B.K., Prisant, M.G., Read, R.J., Richardson, J.S., Richardson, D.C., Sammito, M.D., Sobolev, O.V., Stockwell, D.H., Terwilliger, T.C., Urzhumtsev, A.G., Videau, L.L., Williams, C.J., Adams, P.D., 2019. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D Struct Biol 75, 861–877. https://doi.org/10.1107/S2059798319011471

Lindgren, E., Hägg, S., Giordano, F., Björkegren, J., Ström, L., 2014. Inactivation of the budding yeast cohesin loader Scc2 alters gene expression both globally and in response to a single DNA double strand break. null 13, 3645–3658. https://doi.org/10.4161/15384101.2014.964108

Lioy, V.S., Cournac, A., Marbouty, M., Duigou, S., Mozziconacci, J., Espéli, O.,

Boccard, F., Koszul, R., 2018. Multiscale structuring of the E. coli chromosome by nucleoid-associated and condensin proteins. Cell 172, 771–783.

Lioy, V.S., Junier, I., Lagage, V., Vallet, I., Boccard, F., 2020. Distinct activities of bacterial condensins for chromosome management in Pseudomonas aeruginosa. Cell reports 33, 108344.

Liu, H., Rankin, S., Yu, H., 2013. Phosphorylation-enabled binding of SGO1–PP2A to cohesin protects sororin and centromeric cohesion during mitosis. Nature cell biology 15, 40–49.

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. Nature genetics 46, 1147–1151.

Losada, A., Hirano, M., Hirano, T., 1998. Identification of Xenopus SMC protein complexes required for sister chromatid cohesion. Genes & development 12, 1986–1997.

Löwe, J., Cordell, S.C., van den Ent, F., 2001. Crystal structure of the SMC head domain: an ABC ATPase with 900 residues antiparallel coiled-coil inserted. Journal of molecular biology 306, 25–35.

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

Lyumkis, D., 2019. Challenges and opportunities in cryo-EM single-particle analysis. Journal of Biological Chemistry 294, 5181–5197. https://doi.org/10.1074/jbc.REV118.005602

Machín, F., Paschos, K., Jarmuz, A., Torres-Rosell, J., Pade, C., Aragón, L., 2004. Condensin regulates rDNA silencing by modulating nucleolar Sir2p. Current biology 14, 125–130.

Marbouty, M., Le Gall, A., Cattoni, D.I., Cournac, A., Koh, A., Fiche, J.-B., Mozziconacci, J., Murray, H., Koszul, R., Nollmann, M., 2015. Condensin- and Replication-Mediated Bacterial Chromosome Folding and Origin Condensation Revealed by Hi-C and Super-resolution Imaging. Molecular Cell 59, 588–602. https://doi.org/10.1016/j.molcel.2015.07.020

Marek, M., Kannan, S., Hauser, A.-T., Moraes Mourão, M., Caby, S., Cura, V., Stolfa, D.A., Schmidtkunz, K., Lancelot, J., Andrade, L., Renaud, J.-P., Oliveira, G., Sippl, W., Jung, M., Cavarelli, J., Pierce, R.J., Romier, C., 2013. Structural Basis for the Inhibition of Histone Deacetylase 8 (HDAC8), a Key Epigenetic Player in the Blood Fluke Schistosoma mansoni. PLOS Pathogens 9, e1003645. https://doi.org/10.1371/journal.ppat.1003645

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

Marston, A.L., 2015. Shugoshins: tension-sensitive pericentromeric adaptors safeguarding chromosome segregation. Molecular and cellular biology 35, 634–648.

Martín-López, J.V., Fishel, R., 2013. The mechanism of mismatch repair and the functional analysis of mismatch repair defects in Lynch syndrome. Familial cancer 12, 159–168.

Mastronarde, D.N., 2005. Automated electron microscope tomography using robust prediction of specimen movements. Journal of Structural Biology 152, 36–51. https://doi.org/10.1016/j.jsb.2005.07.007

Matoba, K., Yamazoe, M., Mayanagi, K., Morikawa, K., Hiraga, S., 2005. Comparison of MukB homodimer versus MukBEF complex molecular architectures by electron microscopy reveals a higher-order multimerization. Biochemical and Biophysical Research Communications 333, 694–702. https://doi.org/10.1016/j.bbrc.2005.05.163

Matthews, H.K., Bertoli, C., de Bruin, R.A.M., 2022. Cell cycle control in cancer. Nature Reviews Molecular Cell Biology 23, 74–88. https://doi.org/10.1038/s41580-021-

00404-3

Mattingly, M., Seidel, C., Muñoz, S., Hao, Y., Zhang, Y., Wen, Z., Florens, L., Uhlmann, F., Gerton, J.L., 2022. Mediator recruits the cohesin loader Scc2 to RNA Pol II-transcribed genes and promotes sister chromatid cohesion. Current Biology 32, 2884-2896.e6. https://doi.org/10.1016/j.cub.2022.05.019

McAleenan, A., Cordon-Preciado, V., Clemente-Blanco, A., Liu, I.-C., Sen, N., Leonard, J., Jarmuz, A., Aragón, L., 2012. SUMOylation of the α-Kleisin Subunit of Cohesin Is Required for DNA Damage-Induced Cohesion. Current Biology 22, 1564–1575. https://doi.org/10.1016/j.cub.2012.06.045

Mehta, G.D., Rizvi, S.M.A., Ghosh, S.K., 2012. Cohesin: A guardian of genome integrity. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1823, 1324–1342. https://doi.org/10.1016/j.bbamcr.2012.05.027

Menet, M.-C., 2011. Principes de la spectrométrie de masse. Revue Francophone des Laboratoires 2011, 41–53. https://doi.org/10.1016/S1773-035X(11)71211-4

Merkenschlager, M., Nora, E.P., 2016. CTCF and cohesin in genome folding and transcriptional gene regulation. Annual review of genomics and human genetics 17, 17–43.

Meyerson, J.R., Rao, P., Kumar, J., Chittori, S., Banerjee, S., Pierson, J., Mayer, M.L., Subramaniam, S., 2014. Self-assembled monolayers improve protein distribution on holey carbon cryo-EM supports. Sci Rep 4, 7084–7084. https://doi.org/10.1038/srep07084

Mfarej, M.G., Skibbens, R.V., 2021. An ever-changing landscape in Roberts syndrome biology: Implications for macromolecular damage. PLOS Genetics 16, e1009219. https://doi.org/10.1371/journal.pgen.1009219

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. J Biol Chem 295, 146–157. https://doi.org/10.1074/jbc.RA119.011099

Moriya, S., Tsujikawa, E., Hassan, A.K., Asai, K., Kodama, T., Ogasawara, N., 1998. A Bacillus subtilis gene-encoding protein homologous to eukaryotic SMC motor protein is necessary for chromosome partition. Molecular microbiology 29, 179–187.

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. Molecular Genetics & Genomic Medicine 7, e00501.

https://doi.org/10.1002/mgg3.501

Mungan, Z., Akyüz, F., Bugra, Z., Yönall, O., Oztürk, S., Acar, A., Çevikbaş, U., 2003. Familial Visceral Myopathy with Pseudo-Obstruction, Megaduodenum, Barrett's Esophagus, and Cardiac Abnormalities. The American journal of gastroenterology 98, 2556–60. https://doi.org/10.1111/j.1572-0241.2003.08707.x

Muñoz, S., Passarelli, F., Uhlmann, F., 2020. Conserved roles of chromatin remodellers in cohesin loading onto chromatin. Current Genetics 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.

Murayama, Y., Uhlmann, F., 2014. Biochemical reconstitution of topological DNA binding by the cohesin ring. Nature 505, 367–371. https://doi.org/10.1038/nature12867

Nasmyth, K., Haering, C.H., 2005. The structure and function of SMC and kleisin complexes. Annual review of biochemistry 74, 595.

Nasmyth, K., Schleiffer, A., 2004. From a single double helix to paired double helices and back. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 359, 99–108.

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A., Dekker, J., 2013. Organization of the mitotic chromosome. Science 342, 948–953. https://doi.org/10.1126/science.1236083

Nichols, M.H., Corces, V.G., 2021. Principles of 3D compartmentalization of the human genome. Cell reports 35, 109330.

Nichols, M.H., Corces, V.G., 2018. A tethered-inchworm model of SMC DNA translocation. Nature structural & molecular biology 25, 906–910.

Niki, H., Imamura, R., Kitaoka, M., Yamanaka, K., Ogura, T., Hiraga, S., 1992. E.coli MukB protein involved in chromosome partition forms a homodimer with a rod-and-hinge structure having DNA binding and ATP/GTP binding activities. The EMBO Journal 11, 5101–5109. https://doi.org/10.1002/j.1460-2075.1992.tb05617.x

Niki, H., Jaffe, 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.

Nirenberg, M.W., 1963. The genetic code. II. Scientific American 208, 80–94.

Nishiyama, T., Ladurner, R., Schmitz, J., Kreidl, E., Schleiffer, A., Bhaskara, V., Bando, M., Shirahige, K., Hyman, A.A., Mechtler, K., 2010a. Sororin mediates sister chromatid cohesion by antagonizing Wapl. Cell 143, 737–749.

Nishiyama, T., Ladurner, R., Schmitz, J., Kreidl, E., Schleiffer, A., Bhaskara, V., Bando, M., Shirahige, K., Hyman, A.A., Mechtler, K., Peters, J.-M., 2010b. Sororin Mediates Sister Chromatid Cohesion by Antagonizing Wapl. Cell 143, 737–749. https://doi.org/10.1016/j.cell.2010.10.031

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, 5612–5612. https://doi.org/10.1038/s41467-020-19283-x

Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., Bruneau, B.G., 2017. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell 169, 930-944.e22. https://doi.org/10.1016/j.cell.2017.05.004

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 Ada L., Olins Donald E., 1974. Spheroid Chromatin Units (ν Bodies). Science 183, 330–332. https://doi.org/10.1126/science.183.4122.330

Oliva, A., Rosebrock, A., Ferrezuelo, F., Pyne, S., Chen, H., Skiena, S., Futcher, B., Leatherwood, J., 2005. The cell cycle–regulated genes of Schizosaccharomyces pombe. PLoS biology 3, e225.

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

Oravcová, M., Boddy, M.N., 2019. Recruitment, loading, and activation of the Smc5–Smc6 SUMO ligase. Current genetics 65, 669–676.

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

Ouimette, J.-F., Rougeulle, C., Veitia, R.A., 2019. Three-dimensional genome architecture in health and disease. Clinical Genetics 95, 189–198. https://doi.org/10.1111/cge.13219

Oum Ji-Hyun, Seong Changhyun, Kwon Youngho, Ji Jae-Hoon, Sid Amy, Ramakrishnan Sreejith, Ira Grzegorz, Malkova Anna, Sung Patrick, Lee Sang Eun, Shim Eun Yong, 2011. RSC Facilitates Rad59-Dependent Homologous Recombination between Sister Chromatids by Promoting Cohesin Loading at DNA Double-Strand Breaks. Molecular and Cellular Biology 31, 3924–3937. https://doi.org/10.1128/MCB.01269-10

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. Proc Natl Acad Sci U S A 110, 11355–11360. https://doi.org/10.1073/pnas.1304594110

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

Palecek, J.J., 2018. SMC5/6: Multifunctional player in replication. Genes 10, 7.

Palecek, J.J., Gruber, S., 2015. Kite proteins: a superfamily of SMC/Kleisin partners conserved across bacteria, archaea, and eukaryotes. Structure 23, 2183–2190.

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

Pantelic, R.S., Suk, J.W., Magnuson, C.W., Meyer, J.C., Wachsmuth, P., Kaiser, U., Ruoff, R.S., Stahlberg, H., 2011. Graphene: Substrate preparation and introduction. Journal of Structural Biology 174, 234–238. https://doi.org/10.1016/j.jsb.2010.10.002

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., 2008. Cohesins functionally associate with CTCF on mammalian chromosome arms. Cell 132, 422–433.

Parish, J.L., Rosa, J., Wang, X., Lahti, J.M., Doxsey, S.J., Androphy, E.J., 2006. The DNA helicase ChlR1 is required for sister chromatid cohesion in mammalian cells. Journal of Cell Science 119, 4857–4865.

Paulson, J.R., Hudson, D.F., Cisneros-Soberanis, F., Earnshaw, W.C., 2021. Mitotic chromosomes. Semin Cell Dev Biol 117, 7–29. https://doi.org/10.1016/j.semcdb.2021.03.014

Peet, M.J., Henderson, R., Russo, C.J., 2019. The energy dependence of contrast and damage in electron cryomicroscopy of biological molecules. Ultramicroscopy 203, 125–131. https://doi.org/10.1016/j.ultramic.2019.02.007

Perrakis, A., Romier, C., 2008. Assembly of protein complexes by coexpression in

prokaryotic and eukaryotic hosts: an overview. Methods Mol Biol 426, 247–256. https://doi.org/10.1007/978-1-60327-058-8_15

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

Petela, N.J., Gonzalez Llamazares, A., Dixon, S., Hu, B., Lee, B.-G., Metson, J., Seo, H., Ferrer-Harding, A., Voulgaris, M., Gligoris, T., Collier, J., Oh, B.-H., Löwe, J., Nasmyth, K.A., 2021. Folding of cohesin's coiled coil is important for Scc2/4-induced association with chromosomes. eLife 10, e67268. https://doi.org/10.7554/eLife.67268

Petrushenko, Z.M., She, W., Rybenkov, V.V., 2011. A new family of bacterial condensins. Mol Microbiol 81, 881–896. https://doi.org/10.1111/j.1365-2958.2011.07763.x

Pisani, F.M., 2019. Spotlight on Warsaw Breakage Syndrome. Appl Clin Genet 12, 239–248. https://doi.org/10.2147/TACG.S186476

Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., Thurman, R.E., Cheng, Y., Gülsoy, G., Dennis, J.H., Snyder, M.P., Stamatoyannopoulos, J.A., Taylor, J., Hardison, R.C., Kahveci, T., Ren, B., Gilbert, D.M., 2014. Topologically associating domains are stable units of replication-timing regulation. Nature 515, 402–405. https://doi.org/10.1038/nature13986

Portillo-Ledesma, S., Tsao, L.H., Wagley, M., Lakadamyali, M., Cosma, M.P., Schlick, T., 2021. Nucleosome Clutches are Regulated by Chromatin Internal Parameters. J Mol Biol 433, 166701–166701. https://doi.org/10.1016/j.jmb.2020.11.001

Prince, J.P., Bolla, J.R., Fisher, G.L.M., Mäkelä, J., Fournier, M., Robinson, C.V., Arciszewska, L.K., Sherratt, D.J., 2021. Acyl carrier protein promotes MukBEF action in Escherichia coli chromosome organization-segregation. Nature Communications 12, 6721. https://doi.org/10.1038/s41467-021-27107-9

Prisant, M.G., Williams, C.J., Chen, V.B., Richardson, J.S., Richardson, D.C., 2020. New tools in MolProbity validation: CaBLAM for CryoEM backbone, UnDowser to rethink "waters," and NGL Viewer to recapture online 3D graphics. Protein Sci 29, 315–329. https://doi.org/10.1002/pro.3786

Punjani, A., Rubinstein, J.L., Fleet, D.J., Brubaker, M.A., 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nature Methods 14, 290–296. https://doi.org/10.1038/nmeth.4169

Punjani, A., Zhang, H., Fleet, D.J., 2020. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. Nature Methods 17, 1214–1221. https://doi.org/10.1038/s41592-020-00990-8

Qiu, Y., Huang, S., 2020. CTCF-mediated genome organization and leukemogenesis. Leukemia 34, 2295–2304. https://doi.org/10.1038/s41375-020-0906-x

Quispe, J., Damiano, J., Mick, S.E., Nackashi, D.P., Fellmann, D., Ajero, T.G., Carragher, B., Potter, C.S., 2007. An Improved Holey Carbon Film for Cryo-Electron Microscopy. Microscopy and Microanalysis 13, 365–371. https://doi.org/10.1017/S1431927607070791

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

Rao, S.S.P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., Huang, X., Shamim, M.S., Shin, J., Turner, D., Ye, Z., Omer, A.D., Robinson, J.T., Schlick, T., Bernstein,

B.E., Casellas, R., Lander, E.S., Aiden, E.L., 2017. Cohesin Loss Eliminates All Loop Domains. Cell 171, 305-320.e24. https://doi.org/10.1016/j.cell.2017.09.026

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., Aiden, E.L., 2014. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell 159, 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021

Remeseiro, S., Cuadrado, A., Carretero, M., Martínez, P., Drosopoulos, W.C., Cañamero, M., Schildkraut, C.L., Blasco, M.A., Losada, A., 2012. Cohesin-SA1 deficiency drives aneuploidy and tumourigenesis in mice due to impaired replication of telomeres. The EMBO Journal 31, 2076–2089. https://doi.org/10.1038/emboj.2012.11

Remeseiro, S., Cuadrado, A., Kawauchi, S., Calof, A.L., Lander, A.D., Losada, A., 2013. Reduction of Nipbl impairs cohesin loading locally and affects transcription but not cohesion-dependent functions in a mouse model of Cornelia de Lange Syndrome. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1832, 2097–2102. https://doi.org/10.1016/j.bbadis.2013.07.020

Ricci, M.A., Manzo, C., García-Parajo, M.F., Lakadamyali, M., Cosma, M.P., 2015. Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo. Cell 160, 1145–1158. https://doi.org/10.1016/j.cell.2015.01.054

Rinzema, N.J., Sofiadis, K., Tjalsma, S.J.D., Verstegen, M.J.A.M., Oz, Y., Valdes-Quezada, C., Felder, A.-K., Filipovska, T., van der Elst, S., de Andrade dos Ramos, Z., Han, R., Krijger, P.H.L., de Laat, W., 2022. Building regulatory landscapes reveals that an enhancer can recruit cohesin to create contact domains, engage CTCF sites and activate distant genes. Nature Structural & Molecular Biology 29, 563–574. https://doi.org/10.1038/s41594-022-00787-7

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

Romero-Pérez, L., Surdez, D., Brunet, E., Delattre, O., Grünewald, T.G.P., 2019. STAG Mutations in Cancer. Trends in Cancer 5, 506–520. https://doi.org/10.1016/j.trecan.2019.07.001

Romier, C., Ben Jelloul, M., Albeck, S., Buchwald, G., Busso, D., Celie, P.H.N., Christodoulou, E., De Marco, V., van Gerwen, S., Knipscheer, P., Lebbink, J.H., Notenboom, V., Poterszman, A., Rochel, N., Cohen, S.X., Unger, T., Sussman, J.L., Moras, D., Sixma, T.K., Perrakis, A., 2006. Co-expression of protein complexes in prokaryotic and eukaryotic hosts: experimental procedures, database tracking and case studies. Acta Crystallogr D Biol Crystallogr 62, 1232–1242. https://doi.org/10.1107/S0907444906031003

Rowley, M.J., Nichols, M.H., Lyu, X., Ando-Kuri, M., Rivera, I.S.M., Hermetz, K., Wang, P., Ruan, Y., Corces, V.G., 2017. Evolutionarily conserved principles predict 3D chromatin organization. Molecular cell 67, 837–852.

Russo, C.J., Passmore, L.A., 2014. Electron microscopy: Ultrastable gold substrates for electron cryomicroscopy. Science 346, 1377–1380. https://doi.org/10.1126/science.1259530

Saitoh, N., Goldberg, I.G., Wood, E.R., Earnshaw, W.C., 1994. ScII: an abundant chromosome scaffold protein is a member of a family of putative ATPases with an unusual predicted tertiary structure. The Journal of cell biology 127, 303–318.

Samora, C.P., Saksouk, J., Goswami, P., Wade, B.O., Singleton, M.R., Bates, P.A., Lengronne, A., Costa, A., Uhlmann, F., 2016. Ctf4 links DNA replication with sister chromatid cohesion establishment by recruiting the Chl1 helicase to the replisome. Molecular cell 63, 371–384.

Sauerwald, N., Kingsford, C., 2018. Quantifying the similarity of topological domains across normal and cancer human cell types. Bioinformatics 34, i475–i483.

https://doi.org/10.1093/bioinformatics/bty265

Schatz, D.G., Ji, Y., 2011. Recombination centres and the orchestration of V(D)J recombination. Nature Reviews Immunology 11, 251–263. https://doi.org/10.1038/nri2941

Scheres, S.H.W., 2012. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. Journal of Structural Biology 180, 519–530. https://doi.org/10.1016/j.jsb.2012.09.006

Scherzer, M., Giordano, F., Ferran, M.S., Ström, L., 2022. Recruitment of Scc2/4 to double-strand breaks depends on γH2A and DNA end resection. Life Sci. Alliance 5, e202101244. https://doi.org/10.26508/lsa.202101244

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. Nature Methods 9, 676–682. https://doi.org/10.1038/nmeth.2019

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.

Schmiesing, J.A., Gregson, H.C., Zhou, S., Yokomori, K., 2000. A human condensin complex containing hCAP-C–hCAP-E and CNAP1, a homolog of Xenopus XCAP-D2, colocalizes with phosphorylated histone H3 during the early stage of mitotic chromosome condensation. Molecular and cellular biology 20, 6996–7006.

Schmitz, J., Watrin, E., Lénárt, P., Mechtler, K., Peters, J.-M., 2007. Sororin is required for stable binding of cohesin to chromatin and for sister chromatid cohesion in interphase. Current Biology 17, 630–636.

Schoenfelder, S., Fraser, P., 2019. Long-range enhancer–promoter contacts in gene expression control. Nature Reviews Genetics 20, 437–455. https://doi.org/10.1038/s41576-019-0128-0

Scholey, J.M., Brust-Mascher, I., Mogilner, A., 2003. Cell division. Nature 422, 746–752. https://doi.org/10.1038/nature01599

Schuck, P., 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J 78, 1606–1619. https://doi.org/10.1016/S0006-3495(00)76713-0

Schuck, P., Zhao, H., Brautigam, C.A., Ghirlando, R., 2016. Basic principles of analytical ultracentrifugation. CRC Press Boca Raton.

Scientific, T., 2012. Thermo scientific crosslinking technical handbook. Waltham (USA): Thermo Scientific.

Sedat, J., Manuelidis, L., 1978. A direct approach to the structure of eukaryotic chromosomes. Presented at the Cold Spring Harbor symposia on quantitative biology, Cold Spring Harbor Laboratory Press, pp. 331–350.

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

Sgro, G.G., Costa, T.R.D., 2018. Cryo-EM Grid Preparation of Membrane Protein Samples for Single Particle Analysis. Front Mol Biosci 5, 74–74. https://doi.org/10.3389/fmolb.2018.00074

Shaltiel, I.A., Datta, S., Lecomte, L., Hassler, M., Kschonsak, M., Bravo, S., Stober, C., Ormanns, J., Eustermann, S., Haering, C.H., 2022. A hold-and-feed mechanism drives directional DNA loop extrusion by condensin. Science 376, 1087–1094.

Shi, Z., Gao, H., Bai, X., Yu, H., 2020. Cryo-EM structure of the human cohesin-NIPBL-DNA complex. Science 368, 1454. https://doi.org/10.1126/science.abb0981

Shi Zhubing, Gao Haishan, Bai Xiao-chen, Yu Hongtao, 2020. Cryo-EM structure of the human cohesin-NIPBL-DNA complex. Science 368, 1454–1459. https://doi.org/10.1126/science.abb0981

Sjögren, C., Nasmyth, K., 2001. Sister chromatid cohesion is required for postreplicative double-strand break repair in Saccharomyces cerevisiae. Current Biology 11, 991–995.

Skibbens, R.V., Marzillier, J., Eastman, L., 2010. Cohesins coordinate gene transcriptions of related function within Saccharomyces cerevisiae. Cell Cycle 9, 1601–1606.

Soardi, F.C., Machado-Silva, A., Linhares, N.D., Zheng, G., Qu, Q., Pena, H.B., Martins, T.M., Vieira, H.G., Pereira, N.B., Melo-Minardi, R.C., 2017. Familial STAG2 germline mutation defines a new human cohesinopathy. NPJ genomic medicine 2, 1–11.

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., 2015. Molecular basis for SMC rod formation and its dissolution upon DNA binding. Molecular cell 57, 290–303.

Some, D., Amartely, H., Tsadok, A., Lebendiker, M., 2019. Characterization of proteins by size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS). JoVE (Journal of Visualized Experiments) e59615.

Song Feng, Chen Ping, Sun Dapeng, Wang Mingzhu, Dong Liping, Liang Dan, Xu Rui-Ming, Zhu Ping, Li Guohong, 2014. Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units. Science 344, 376–380. https://doi.org/10.1126/science.1251413

Sonoda, E., Matsusaka, T., Morrison, C., Vagnarelli, P., Hoshi, O., Ushiki, T., Nojima, K., Fukagawa, T., Waizenegger, I.C., Peters, J.-M., Earnshaw, W.C., Takeda, S., 2001. Scc1/Rad21/Mcd1 Is Required for Sister Chromatid Cohesion and Kinetochore Function in Vertebrate Cells. Developmental Cell 1, 759–770. https://doi.org/10.1016/S1534-5807(01)00088-0

Soppa, J., Kobayashi, K., Noirot-Gros, M., Oesterhelt, D., Ehrlich, S.D., Dervyn, E., Ogasawara, N., Moriya, S., 2002. Discovery of two novel families of proteins that are proposed to interact with prokaryotic SMC proteins, and characterization of the Bacillus subtilis family members ScpA and ScpB. Molecular microbiology 45, 59–71.

Srinivasan, M., Fumasoni, M., Petela, N.J., Murray, A., Nasmyth, K.A., 2020. Cohesion is established during DNA replication utilising chromosome associated cohesin rings as well as those loaded de novo onto nascent DNAs. Elife 9, e56611.

Srinivasan, M., Scheinost, J.C., Petela, N.J., Gligoris, T.G., Wissler, M., Ogushi, S., Collier, J.E., Voulgaris, M., Kurze, A., Chan, K.-L., 2018. The cohesin ring uses its hinge to organize DNA using non-topological as well as topological mechanisms. Cell 173, 1508–1519.

Strahl, B.D., Allis, C.D., 2000. The language of covalent histone modifications. Nature 403, 41–45. https://doi.org/10.1038/47412

Ström, L., Lindroos, H.B., Shirahige, K., Sjögren, C., 2004. Postreplicative Recruitment of Cohesin to Double-Strand Breaks Is Required for DNA Repair. Molecular Cell 16, 1003–1015. https://doi.org/10.1016/j.molcel.2004.11.026

Struhl, K., Segal, E., 2013. Determinants of nucleosome positioning. Nature Structural & Molecular Biology 20, 267–273. https://doi.org/10.1038/nsmb.2506

Strunnikov, A.V., Hogan, E., Koshland, D., 1995. SMC2, a Saccharomyces cerevisiae gene essential for chromosome segregation and condensation, defines a subgroup within the SMC family. Genes & development 9, 587–599.

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. Journal of Cell Biology 123, 1635–1648. https://doi.org/10.1083/jcb.123.6.1635

Sultana, A., Lee, J.E., 2015. Measuring protein-protein and protein-nucleic acid interactions by biolayer interferometry. Current protocols in protein science 79, 19–25.

Sumara, I., Vorlaufer, E., Gieffers, C., Peters, B.H., Peters, J.-M., 2000. Characterization of vertebrate cohesin complexes and their regulation in prophase. The Journal of cell biology 151, 749–762.

Sutani, T., Yuasa, T., Tomonaga, T., Dohmae, N., Takio, K., Yanagida, M., 1999. Fission yeast condensin complex: essential roles of non-SMC subunits for condensation and Cdc2 phosphorylation of Cut3/SMC4. Genes & development 13, 2271–2283.

Svergun, D.I., 1999. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophysical journal 76, 2879–2886.

Szabo, Q., Donjon, A., Jerković, I., Papadopoulos, G.L., Cheutin, T., Bonev, B., Nora, E.P., Bruneau, B.G., Bantignies, F., Cavalli, G., 2020. Regulation of single-cell genome organization into TADs and chromatin nanodomains. Nat Genet 52, 1151–1157. https://doi.org/10.1038/s41588-020-00716-8

Tanaka, T., Fuchs, J., Loidl, J., Nasmyth, K., 2000. Cohesin ensures bipolar attachment of microtubules to sister centromeres and resists their precocious separation. Nature cell biology 2, 492–499.

Taschner, M., Basquin, J., Steigenberger, B., Schäfer, I.B., Soh, Y., Basquin, C., Lorentzen, E., Räschle, M., Scheltema, R.A., Gruber, S., 2021. Nse5/6 inhibits the Smc5/6 ATPase and modulates DNA substrate binding. The EMBO journal 40, e107807.

Taylor, E.M., Moghraby, J.S., Lees, J.H., Smit, B., Moens, P.B., Lehmann, A.R., 2001. Characterization of a Novel Human SMC Heterodimer Homologous to the Schizosaccharomyces pombe Rad18/Spr18 Complex. MBoC 12, 1583–1594. https://doi.org/10.1091/mbc.12.6.1583

Tegunov, D., Cramer, P., 2019. Real-time cryo-electron microscopy data preprocessing with Warp. Nature Methods 16, 1146–1152. https://doi.org/10.1038/s41592-019-0580-y

Terwilliger, T.C., Read, R.J., Adams, P.D., Brunger, A.T., Afonine, P.V., Hung, L.-W., 2013. Model morphing and sequence assignment after molecular replacement. Acta Crystallographica Section D: Biological Crystallography 69, 2244–2250.

Thiecke, M.J., Wutz, G., Muhar, M., Tang, W., Bevan, S., Malysheva, V., Stocsits, R., Neumann, T., Zuber, J., Fraser, P., 2020. Cohesin-dependent and-independent mechanisms mediate chromosomal contacts between promoters and enhancers. Cell reports 32, 107929.

Thomas, C., Tampé, R., 2020. Structural and mechanistic principles of ABC transporters. Annual review of biochemistry 89, 605–636.

Tittel-Elmer, M., Lengronne, A., Davidson, M.B., Bacal, J., François, P., Hohl, M., Petrini, J.H.J., Pasero, P., Cobb, J.A., 2012. Cohesin Association to Replication Sites Depends on Rad50 and Promotes Fork Restart. Molecular Cell 48, 98–108. https://doi.org/10.1016/j.molcel.2012.07.004

Tomonaga, T., Nagao, K., Kawasaki, Y., Furuya, K., Murakami, A., Morishita, J., Yuasa, T., Sutani, T., Kearsey, S.E., Uhlmann, F., 2000. Characterization of fission yeast cohesin: essential anaphase proteolysis of Rad21 phosphorylated in the S phase. Genes & Development 14, 2757–2770.

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

Tothova, Z., Valton, A.-L., Gorelov, R.A., Vallurupalli, M., Krill-Burger, J.M., Holmes, A., Landers, C.C., Haydu, J.E., Malolepsza, E., Hartigan, C., Donahue, M., Popova, K.D., Koochaki, S., Venev, S.V., Rivera, J., Chen, E., Lage, K., Schenone, M., D'Andrea, A.D., Carr, S.A., Morgan, E.A., Dekker, J., Ebert, B.L., 2021. Cohesin mutations alter DNA damage repair and chromatin structure and create therapeutic vulnerabilities in MDS/AML. JCI Insight 6.

https://doi.org/10.1172/jci.insight.142149

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

Uhlmann, F., Nasmyth, K., 1998. Cohesion between sister chromatids must be established during DNA replication. Current Biology 8, 1095–1102. https://doi.org/10.1016/S0960-9822(98)70463-4

Uhlmann, F., Wernic, D., Poupart, M.-A., Koonin, E.V., Nasmyth, K., 2000. Cleavage of Cohesin by the CD Clan Protease Separin Triggers Anaphase in Yeast. Cell 103, 375–386. https://doi.org/10.1016/S0092-8674(00)00130-6

Uusküla-Reimand, L., Hou, H., Samavarchi-Tehrani, P., Rudan, M.V., Liang, M., Medina-Rivera, A., Mohammed, H., Schmidt, D., Schwalie, P., Young, E.J., 2016. Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. Genome biology 17, 1–22.

van der Lelij, P., Newman, J.A., Lieb, S., Jude, J., Katis, V., Hoffmann, T., Hinterndorfer, M., Bader, G., Kraut, N., Pearson, M.A., Peters, J.-M., Zuber, J., Gileadi, O., Petronczki, M., 2020. STAG1 vulnerabilities for exploiting cohesin synthetic lethality in STAG2-deficient cancers. Life Sci Alliance 3, e202000725. https://doi.org/10.26508/lsa.202000725

van Ruiten, M.S., van Gent, D., Sedeño Cacciatore, Á., Fauster, A., Willems, L., Hekkelman, M.L., Hoekman, L., Altelaar, M., Haarhuis, J.H., Brummelkamp, T.R., 2022. The cohesin acetylation cycle controls chromatin loop length through a PDS5A brake mechanism. Nature structural & molecular biology 1–6.

Vaur, S., Feytout, A., Vazquez, S., Javerzat, J.-P., 2012. Pds5 promotes cohesin acetylation and stable cohesin–chromosome interaction. EMBO reports 13, 645–652. https://doi.org/10.1038/embor.2012.72

Vazquez Nunez, R., Polyhach, Y., Soh, Y.-M., Jeschke, G., Gruber, S., 2021. Gradual opening of Smc arms in prokaryotic condensin. Cell Reports 35. https://doi.org/10.1016/j.celrep.2021.109051

Vega, H., Waisfisz, Q., Gordillo, M., Sakai, N., Yanagihara, I., Yamada, M., Van Gosliga, D., Kayserili, H., Xu, C., Ozono, K., 2005. Roberts syndrome is caused by mutations in ESCO2, a human homolog of yeast ECO1 that is essential for the establishment of sister chromatid cohesion. Nature genetics 37, 468–470.

Verhest, A., Heimann, P., 2008. Basic Cytogenetics and the Role of Genetics in Cancer Development. Comprehensive Cytopathology 23–45.

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. Adv Exp Med Biol 896, 43–58. https://doi.org/10.1007/978-3-319-27216-0_4

Vincentelli, R., Romier, C., 2013. Expression in Escherichia coli: becoming faster and more complex. Curr Opin Struct Biol 23, 326–334. https://doi.org/10.1016/j.sbi.2013.01.006

Virchow, R.L.K., 1859. Die Cellularpathologie in ihrer Begründung auf physiologische und pathologische Gewebelehre. A. Hirshcwald.

Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., Raisch, T., Prumbaum, D., Quentin, D., Roderer, D., Tacke, S., Siebolds, B.,

Schubert, E., Shaikh, T.R., Lill, P., Gatsogiannis, C., Raunser, S., 2019. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. Communications Biology 2, 218. https://doi.org/10.1038/s42003-019-0437-z

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

Waldeyer, W., 1888. Über Karyokinese und ihre Beziehungen zu den Befruchtungsvorgängen. Archiv für mikroskopische Anatomie 32, 1–122.

Wall, J.D., Stawiski, E.W., Ratan, A., Kim, H.L., Kim, C., Gupta, R., Suryamohan, K., Gusareva, E.S., Purbojati, R.W., Bhangale, T., Stepanov, V., Kharkov, V., Schröder, M.S., Ramprasad, V., Tom, J., Durinck, S., Bei, Q., Li, J., Guillory, J., Phalke, S., Basu, A., Stinson, J., Nair, S., Malaichamy, S., Biswas, N.K., Chambers, J.C., Cheng, K.C., George, J.T., Khor, S.S., Kim, J.-I., Cho, B., Menon, R., Sattibabu, T., Bassi, A., Deshmukh, M., Verma, A., Gopalan, V., Shin, J.-Y., Pratapneni, M., Santhosh, S., Tokunaga, K., Md-Zain, B.M., Chan, K.G., Parani, M., Natarajan, P., Hauser, M., Allingham, R.R., Santiago-Turla, C., Ghosh, A., Gadde, S.G.K., Fuchsberger, C., Forer, L., Schoenherr, S., Sudoyo, H., Lansing, J.S., Friedlaender, J., Koki, G., Cox, M.P., Hammer, M., Karafet, T., Ang, K.C., Mehdi, S.Q., Radha, V., Mohan, V., Majumder, P.P., Seshagiri, S., Seo, J.-S., Schuster, S.C., Peterson, A.S., GenomeAsia100K Consortium, 2019. The GenomeAsia 100K Project enables genetic discoveries across Asia. Nature 576, 106–111. https://doi.org/10.1038/s41586-019-1793-z

Walther, N., Hossain, M.J., Politi, A.Z., Koch, B., Kueblbeck, M., Ødegård-Fougner, Ø., Lampe, M., Ellenberg, J., 2018. A quantitative map of human Condensins provides new insights into mitotic chromosome architecture. Journal of Cell Biology 217, 2309–2328.

Wang, S.-W., Read, R.L., Norbury, C.J., 2002. Fission yeast Pds5 is required for accurate chromosome segregation and for survival after DNA damage or metaphase arrest. Journal of Cell Science 115, 587–598.

Wang, X., Tang, O.W., Riley, E.P., Rudner, D.Z., 2014. The SMC Condensin Complex Is Required for Origin Segregation in Bacillus subtilis. Current Biology 24, 287–292. https://doi.org/10.1016/j.cub.2013.11.050

Ward, A.L., Hopkins, J., Mckay, M., Murray, S., Jordan, P.W., 2016. Genetic Interactions Between the Meiosis-Specific Cohesin Components, STAG3, REC8, and RAD21L. G3: Genes|Genomes|Genetics 6, 1713–1724.

Watanabe, Y., Kitajima, T.S., 2005. Shugoshin protects cohesin complexes at centromeres. Philosophical Transactions of the Royal Society B: Biological Sciences 360, 515–521.

Watrin, E., Peters, J.-M., 2006. Cohesin and DNA damage repair. Experimental Cell Research 312, 2687–2693. https://doi.org/10.1016/j.yexcr.2006.06.024

Watson, J.D., Crick, F.H.C., 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature 171, 737–738. https://doi.org/10.1038/171737a0

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., 2017. YY1 is a structural regulator of enhancer-promoter loops. Cell 171, 1573–1588.

Weissenberger, G., Henderikx, R.J.M., Peters, P.J., 2021. Understanding the invisible hands of sample preparation for cryo-EM. Nature Methods 18, 463–471. https://doi.org/10.1038/s41592-021-01130-6

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.

Wen, Z., Huang, Z.-T., Zhang, R., Peng, C., 2018. ZNF143 is a regulator of chromatin loop. Cell Biology and Toxicology 34, 471–478.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto, N., Maeshima, K., Shirahige, K., Peters, J.-M., 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature 451, 796.

Whelan, G., Kreidl, E., Wutz, G., Egner, A., Peters, J., Eichele, G., 2012. Cohesin acetyltransferase Esco2 is a cell viability factor and is required for cohesion in pericentric heterochromatin. The EMBO journal 31, 71–82.

Wilson, E.B., 1900. The cell in development and inheritance. Macmillan.

Woo, J.-S., Lim, J.-H., Shin, H.-C., Suh, M.-K., Ku, B., Lee, K.-H., Joo, K., Robinson, H., Lee, J., Park, S.-Y., Ha, N.-C., Oh, B.-H., 2009. Structural Studies of a Bacterial Condensin Complex Reveal ATP-Dependent Disruption of Intersubunit Interactions. Cell 136, 85–96. https://doi.org/10.1016/j.cell.2008.10.050

Woodcock, C.L., Ghosh, R.P., 2010. Chromatin higher-order structure and dynamics. Cold Spring Harb Perspect Biol 2, a000596–a000596. https://doi.org/10.1101/cshperspect.a000596

Wutz, G., Ladurner, R., St Hilaire, B.G., Stocsits, R.R., Nagasaka, K., Pignard, B., Sanborn, A., Tang, W., Varnai, C., Ivanov, M.P., 2020. ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesinSTAG1 from WAPL. Elife 9.

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

Xiang, S., Koshland, D., 2021. Cohesin architecture and clustering in vivo. Elife 10, e62243.

Xu, H., Yan, M., Patra, J., Natrajan, R., Yan, Y., Swagemakers, S., Tomaszewski, J.M., Verschoor, S., Millar, E.K., van der Spek, P., Reis-Filho, J.S., Ramsay, R.G., O'Toole, S.A., McNeil, C.M., Sutherland, R.L., McKay, M.J., Fox, S.B., 2011. Enhanced RAD21 cohesin expression confers poor prognosis and resistance to chemotherapy in high grade luminal, basal and HER2 breast cancers. Breast Cancer Research 13, R9. https://doi.org/10.1186/bcr2814

Xu, W., Ying, Y., Shan, L., Feng, J., Zhang, S., Gao, Y., Xu, X., Yao, Y., Zhu, C., Mao, W., 2015. Enhanced expression of cohesin loading factor NIPBL confers poor prognosis and chemotherapy resistance in non-small cell lung cancer. J Transl Med 13, 153–153. https://doi.org/10.1186/s12967-015-0503-3

Xu, Z., Cetin, B., Anger, M., Cho, U.S., Helmhart, W., Nasmyth, K., Xu, W., 2009. Structure and Function of the PP2A-Shugoshin Interaction. Molecular Cell 35, 426–441. https://doi.org/10.1016/j.molcel.2009.06.031

Xu, Z., Zan, H., Pone, E.J., Mai, T., Casali, P., 2012. Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. Nature Reviews Immunology 12, 517–531. https://doi.org/10.1038/nri3216

Yamamoto, G., Irie, T., Aida, T., Nagoshi, Y., Tsuchiya, R., Tachikawa, T., 2006. Correlation of invasion and metastasis of cancer cells, and expression of the RAD21 gene in oral squamous cell carcinoma. Virchows Archiv 448, 435–441. https://doi.org/10.1007/s00428-005-0132-y

Yang, N., Sheridan, A.M., 2014. Cell Cycle, in: Wexler, P. (Ed.), Encyclopedia of Toxicology (Third Edition). Academic Press, Oxford, pp. 753–758. https://doi.org/10.1016/B978-0-12-386454-3.00273-6

Yao, Y., Dai, W., 2012. Shugoshins function as a guardian for chromosomal stability in nuclear division. Cell Cycle 11, 2631–2642.

Yeong, F.M., Hombauer, H., Wendt, K.S., Hirota, T., Mudrak, I., Mechtler, K., Loregger, T., Marchler-Bauer, A., Tanaka, K., Peters, J.-M., Ogris, E., 2003. Identification of a Subunit of a Novel Kleisin-β/SMC Complex as a Potential Substrate of Protein Phosphatase 2A. Current Biology 13, 2058–2064. https://doi.org/10.1016/j.cub.2003.10.032

Yu, Y., Li, S., Ser, Z., Kuang, H., Than, T., Guan, D., Zhao, X., Patel, D.J., 2022. Cryo-EM structure of DNA-bound Smc5/6 reveals DNA clamping enabled by multi-subunit conformational changes. Proceedings of the National Academy of Sciences 119, e2202799119.

Yu, Y., Li, S., Ser, Z., Sanyal, T., Choi, K., Wan, B., Kuang, H., Sali, A., Kentsis, A., Patel, D.J., 2021. Integrative analysis reveals unique structural and functional features of the Smc5/6 complex. Proceedings of the National Academy of Sciences 118, e2026844118.

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. International Journal of Molecular Sciences 22, 5868.

Zhang, N., Pati, D., 2015. C-terminus of Sororin interacts with SA2 and regulates sister chromatid cohesion. Cell Cycle 14, 820–826. https://doi.org/10.1080/15384101.2014.1000206

Zhang, X., Zhang, Y., Ba, Z., Kyritsis, N., Casellas, R., Alt, F.W., 2019. Fundamental roles of chromatin loop extrusion in antibody class switching. Nature 575, 385–389.

Zhang, Y., Zhang, X., Ba, Z., Liang, Z., Dring, E.W., Hu, H., Lou, J., Kyritsis, N., Zurita, J., Shamim, M.S., 2019. The fundamental role of chromatin loop extrusion in physiological V (D) J recombination. Nature 573, 600–604.

Zhao Hang, Bhowmik Bijit K., Petrushenko Zoya M., Rybenkov Valentin V., Bowman Grant R., 2020. Alternating Dynamics of oriC, SMC, and MksBEF in Segregation of Pseudomonas aeruginosa Chromosome. mSphere 5, e00238-20. https://doi.org/10.1128/mSphere.00238-20

Zheng, G., Kanchwala, M., Xing, C., Yu, H., 2018. MCM2–7-dependent cohesin loading during S phase promotes sister-chromatid cohesion. eLife 7, e33920. https://doi.org/10.7554/eLife.33920

Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., Scheres, S.H., 2018. New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife 7, e42166. https://doi.org/10.7554/eLife.42166

Liste des figures et tableaux

I. Liste des figures

Figure 1 : Representation schematique des differents étapes du cycle cellulaire eucaryote. 22

Figure 2 : Les differents niveaux d'organisation de la chromatine dans les cellules eucaryotes en interphase. 26

Figure 3 : Nucleosome et fibre de chromatine. 28

Figure 4 : Representation des TADs (A) et des territoires chromosomiques (B) 32

Figure 5 : De la fibre de chromatine au chromosome mitotique. 35

Figure 6 : Composition et architecture des proteines SMC 40

Figure 7 : Composition et architecture du complexe SMC. 42

Figure 8 : Le complexe procaryote SMC-ScpAB. 47

Figure 9 : Le complexe MukBEF de E. coli. 49

Figure 10 : La Condensine. 54

Figure 11 : Reconstruction cryoEM de la Condensine de S. cerevisiae sous differents etats. 57

Figure 12 : Le complexe SMC5/SMC6 59

Figure 13 : La Cohesine. 62

Figure 14 : La Cohesine et ses interfaces SMC1A-SMC3-RAD21. 74

Figure 15 : Structures des proteines auxiliaires et regulatrices HAWKs : STAG1 humain associe a la region centrale de RAD21 (gauche), NIPBL de C. thermophilum (centre) et PDS5 (droite) de L. thermotolerans (PDB : 4PJU, 5T8V, 5F0O). 76

Figure 16 : Schema du cycle de la cohesine des vertebres, etablissant la cohesion des chromatides sœurs et sa dissolution au cours du cycle cellulaire 82

Figure 17 : Les differentes structures connues de la Cohesine. 87

Figure 18 : Reconstruction a 5,3 Å du complexe Cohesine tetramerique humain (SMC1A, SMC3, RAD21 et STAG1), associe a NIPBL et a l'ADN, en presence d'AMP.PNP, sous sa conformation clampee. 90

Figure 19 : Le complexe tripartite associe a Scc2 ou Pds5 de S. cerevisiae, sous une forme apo, sans nucleotide et sans ADN. 92

Figure 20 : Les structures des complexes SMC1A-HD/RAD21C et SMC3-HD/RAD21N independant, sous forme apo (A), liees a l'ATP-γ-S (B) et liees a l'ADP (C), obtenues au sein de mon equipe d'accueil. 94

Figure 21 : Frequences des mutations des sous-unites de la Cohesine, selon les types de cancers. 97

Figure 22 : La tete ATPase des proteines SMC. 99

Figure 23 : Les tetes ATPases de la Cohesine partagent des motifs participant a la liaison et a l'hydrolyse de l'ATP, similaires a ceux des transporteurs ABC. 101

Figure 24: Exemple de plasmides pmS utilises pour l'obtention d'une proteine d'interet native (pmS ; gauche), d'une proteine couplee a une etiquette GST et un site de coupure a la protease 3C a son extremite N-terminale (pmS-

G3 ; centre) et a un site de coupure a la protease 3C suivi d'une etiquette His_{10} a son extremite C-terminale (pmS=3I ; droite). 106
Figure 25 : Principe de la reaction d'assemblage de Gibson. 108
Figure 26 : Representation des differentes constructions de SMC1A et SMC3 (A) et de RAD21 (B) realisees et utilisees pendant ma these. 110
Figure 27 : Principe de la concatenation des vecteurs pET-MCN basee sur le site de restriction I-CeuI (A), aboutissant a une concatenation polycistronique, ou sur le site de PI-PspI, aboutissant a une concatenation multi-promoteur (B). 113
Figure 28 : Principe de la titration calorimetrique isotherme. 128
Figure 29 : Principe de la reticulation par le BMOE. 130
Figure 30 : Conversion enzymatique du MESG en ribose-1-phosphate et 2-amino-6-mercapto-7-methylpurine par la purine nucleoside phosphorylase (PNP). 131
Figure 31 : Resultats typiques d'une experience de reconstitution d'un complexe par SEC, ou est observe un decalage vers les plus hauts poids moleculaires lors de la formation d'un complexe (courbe jaune) par rapport au profil de SEC des proteines seules (courbe orange). 132
Figure 32 : Resultats typiques d'une experience de AUC-SV (ici avec du BSA), avec le profil des vitesses de sedimentations experimentales (gauche) et le profil obtenu apres traitement informatique (droite), permettant de determiner un coefficient de sedimentation caracteristique de chaque proteine. Adapte de (Ebel and Birck, 2021). 134
Figure 33 : Les differentes etapes de cryoEM, de l'echantillon proteique au modele 3D. 136
Figure 34 : Preparation et congelation des grilles de microscopie electronique. Adapte de (Sgro and Costa, 2018). 141
Figure 35 : Les etapes necessaires a l'analyse de particules isolees en cryoEM. 143
Figure 36 : Changements conformationnels operes lors de la liaison de l'ATP-γ-S au niveau des tetes de SMC1A (A) et SMC3 (B). 229
Figure 37 : Changements conformationnels induits lors de l'engagement des tetes de SMC1A et SMC3. 231
Figure 38 : Changements conformationnels induits par la liaison de NIPBL et de l'ADN au niveau des tetes ATPases de SMC1A et SMC3 engagees. 233
Figure 39 : Alignement des sequences de RAD21 de divers organismes eucaryotes. 238
Figure 40 : Recapitulatif des constructions de RAD21 (A), ainsi que celles de STAG1 et STAG2 (B) utilisees pour la reconstruction du complexe tripartite ou quadripartite de la Cohesine. 238
Figure 41 : Purification et caracterisation de la proteine STAG1-1. 241
Figure 42 : Premiers essais de la purification de la proteine STAG2-2, construction la plus prometteuse. 242
Figure 43 : Gels SDS-PAGE des resultats les plus prometteurs des tests a petite echelle de la co-expression et co-purification du complexe humain SMC1A-HD/SMC3-HD/RAD21. 244
Figure 44 : Resultats des premiers essais a grande echelle du complexe forme par les tetes de SMC1A et SMC3 et par une des constructions de RAD21. 246

Figure 45 : Essai d'optimisation pour la co-expression et la co-purification du complexe quadripartite, en jouant sur la densite optique d'induction lors de l'expression et en clivant l'etiquette sur RAD21-1 par la protease 3C. 247
Figure 46 : Resultats des tests a petite echelle des complexes quadripartites. 249
Figure 47 : Resultats de la purification a grande echelle du complexe quadripartite SMC1ACC-EQ/SMC3CC-EQ/RAD21-1/STAG1-1. 251
Figure 48 : Modele du basculement – clampage lors de l'extrusion de boucles par la Cohesine. 260

II. Liste des tableaux

Tableau 1 : Composition des differents complexes SMC. 38
Tableau 2 : Nomenclature des differents homologues des complexes SMC. 52
Tableau 3 : Liste des constructions et mutants utilises pendant ma these. 115
Tableau 4 : Composition des milieux de culture utilises. 118
Tableau 5 : Principales etiquettes d'affinite utilisees 122
Tableau 6 : Recapitulatif des grilles utilisees et leurs caracteristiques. 139

Université de Strasbourg

Pauline LANDWERLIN

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

Dynamique conformationnelle du module ATPase de la Cohésine humaine

La Cohésine, complexe à 4 sous-unités (SMC1A, SMC3, RAD21, STAG1/2), est un effecteur majeur de l'organisation 3D du génome eucaryote. Il participe à de nombreuses transactions avec l'ADN (cohésion des chromatides sœurs, réparation de l'ADN, régulation de la transcription et organisation 3D du génome). La Cohésine possède une activité ATPase essentielle à sa dynamique et à ses fonctions et qui est portée par SMC1A et SMC3. Ces protéines forment un hétérodimère en présence d'ATP (module ATPase). Le rôle de l'activité ATPase de la Cohésine sur sa dynamique et ses mécanismes moléculaires est peu compris. Au cours de ma thèse j'ai caractérisé les interactions entre SMC1A, SMC3 et RAD21 par des méthodes biophysiques, biochimiques et structurales. J'ai obtenu une structure cryo-EM à 4 Å du module ATPase. Je me suis également intéressée à la 4^{e} sous-unité de la Cohésine STAG1/2 et son influence sur le module ATPase. Mes travaux de thèse révèlent des informations novatrices sur les mécanismes impliqués dans l'activité ATPase de la Cohésine, notamment au niveau de la porte de sortie de l'ADN, permettant ainsi une meilleure compréhension de sa dynamique et de ses fonctions.

Mots clés : organisation du génome ; chromatine ; complexes SMC ; cohésine ; activité ATPase ; porte de sortie de l'ADN ; réarrangements structuraux

Cohesin, a four-subunit complex (SMC1A, SMC3, RAD21, STAG1/2), is a key effector of the 3D organization of the eukaryotic genome. It is involved in many transactions with DNA (sister chromatid cohesion, DNA repair, regulation of transcription and 3D genome organization). Cohesin has an ATPase activity which is essential to its dynamics and to all its functions and is carried by SMC1A and SMC3. These proteins form a heterodimer in the presence of ATP (ATPase module). The role of the ATPase activity of Cohesin on its dynamics and molecular mechanisms is poorly understood. During my thesis, I characterized the interactions between SMC1A, SMC3, and RAD21 by biophysical, biochemical, and structural methods. I obtained a 4 Å cryo-EM structure of the ATPase module. I also focused on the 4th subunit of the STAG1/2 cohesin and its influence on the ATPase module. My thesis work reveals novel information on the mechanisms involved in the ATPase activity of Cohesin, especially at the DNA exit gate, allowing a better understanding of its dynamics and functions.

Mots clés : Genome organization ; chromatine; SMC complexes ; cohesin ; ATPase activity ; DNA exit gate; conformational changes