UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE des Sciences de la Vie et de la Santé
IGBMC - CNRS UMR 7104 – Inserm U 1258

THÈSE présentée par :

Dana Mariel DIAZ JIMENEZ

soutenue le : **24 Janvier 2022**

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

Discipline/ Spécialité : Biochimie structurale

Structural and functional studies of the BCL7 proteins: novel subunits of the mammalian SWI/SNF complex

---

THÈSE dirigée par :
Mme. BERGAMIN Elisa   Directeur de recherche CNRS, IGBMC, Illkirch
Mme. CHAN Susan   Directeur de recherche CNRS, IGBMC, Illkirch

RAPPORTEURS :
Mme. OCHSENBEIN Françoise   Directeur de recherche I2BC, Gif-sur-Yvette
M. BLAIS Alexandre   Professeur, Université d'Ottawa, Canada

AUTRES MEMBRES DU JURY :
M. WEIXLBAUMER Albert   Directeur de recherche CNRS, IGBMC, Illkirch
Mme.UNGUREANU Daniela   Professeur, Université d'Helsinki, Finlande
                                Directeur de recherche CNRS, IBMC, Strasbourg

MEMBRE INVITE
M. SAUTER Claude

ACKNOWLEDGEMENTS

To my mexican friends at the IGBMC **Mony**, **José**, **Roberto, Manu, Fer** and **Israel.**

You know I am very bad expressing emotions but you are special in my heart.

There are a lot of special persons I met during this years at the IGBMC and I want to say thanks to everyone who shared any experience with me, specially **Kenny** and **Ben**. I know you may want an individual sentence but both mean so much for me and I don't have enough words. To my friends in the first floor crystallographers and yeast friends, and the cineclub.

An special thanks to **Veronique Fischer** for all the help during student paper work.

Valerie Schon for her kindness and support since my arrival.

My friends in Strasbourg specially **Malcolm** for all your kindness, love and help. **Markus**, **Moni**, **Liz**, **Mariant**, **Andrea** for the dancing and brunches before 2020. **Julie**, my dear friend in a short time. **Emi** for you help without knowing me at all. To my nice neighbors, specially **Louis** for sharing his internet.

Special thanks to **Jacqueline Plumbridge** and **Laura Alvarez** for their friendship.

To the **Oddsson** family for the years of support and love (und ein Entlein).

My friends in Mexico specially **Dani**.

And finally thanks to **my family**, **mami** por la fuerza que me das cada día al despertar y por las ganas de poder estar juntas. **Papi** por no dejarme rendir y confiar en mi. Mis **hermanos** por los consejos y apoyo y en esta etapa final por el nuevo bebé que será parte de la familia. A mi **abuelo** por que cada semana tuvo tiempo para hablar a miles de kilometros.

Gracias a la vida

RÉSUMÉ EN FRANÇAIS

Études structurales et fonctionnelles des protéines BCL7 : de nouvelles sous-unités du complexe SWI/SNF.

Introduction.

L'ADN d'une cellule humaine est estimé à environ 2 mètres de long. Cette longueur énorme est emballée dans un noyau de seulement 10 μm de large en s'associant avec des protéines pour former une structure dense et compacte appelée chromatine. L'unité de base de la chromatine est le nucléosome, un complexe protéique en forme de bobine composé de huit sous-unités d'histones et de 147 paires d'ADN enroulées autour [1]. La chromatine a deux états différents : un état ouvert appelé euchromatine et un état fermé appelé hétérochromatine. L'état ouvert est généralement la forme active où les gènes sont transcrits en ARN. Pour cette transcription, et pour la réplication de l'ADN avant la division cellulaire, la structure de la chromatine doit être modifiée pour permettre le passage des grands complexes ARN et ADN polymérase. Les enzymes modificatrices d'histones et les complexes de remodelage de la chromatine ATP-dépendants modifient la structure de la chromatine pour permettre le passage de ces polymérases. Les complexes de remodelage de la chromatine ATP-dépendants utilisent l'énergie de l'hydrolyse de l'ATP pour retirer les nucléosomes de l'ADN ou les faire glisser le long de celui-ci. Ils sont classés en quatre types principaux sur la base des séquences d'acides

aminés de leurs sous-unités catalytiques, INO80, ISWI, NurD et SWI/SNF. Nous étudions une famille spécifique : Les complexes de remodelage de la chromatine Switch/ Sucrose non-fermentable, plus connus sous le nom de SWI/SNF, qui ont été caractérisés pour la première fois chez la levure. Chez les métazoaires, il existe trois types de complexes protéiques SWI/SNF : BAF(BRG1/BRM-associated factor), pBAF (polybromo-associated BAF) et ncBAF (non canonical BAF). Les complexes BAF des mammifères contiennent jusqu'à 29 sous-unités qui, ensemble, ont un poids moléculaire d'environ 2MDa. Alors que les sous-unités catalytiques et régulatrices essentielles ont été étudiées de manière approfondie, ces dernières années, de nouvelles sous-unités dont les fonctions sont moins bien caractérisées ont été décrites. L'importance de ces complexes est relayée par le fait que dans plus de 20% des cancers humains, des mutations dans les gènes codant pour les sous-unités du complexe SWI/SNF sont présentes, certaines sous-unités étant mutées dans des tumeurs malignes spécifiques [2]. Mon étude porte sur les trois nouvelles protéines BCL7, qui sont invariablement présentes dans les complexes SWI/SNF des mammifères. Leurs structures et fonctions sont inconnues, cependant, les gènes qui les codent sont mutés chez les patients atteints de divers types de cancer. Dans le lymphome diffus à grandes cellules B, BCL7A agit comme un suppresseur de tumeur, et cette activité peut être perdue lorsque des mutations qui altèrent la liaison avec le complexe apparaissent [3]. Notre étude porte sur la pertinence fonctionnelle des protéines BCL7 pour le remodelage de la chromatine et leurs interactions avec d'autres sous-unités du complexe. L'objectif de ma recherche était de mieux comprendre la structure et la fonction des protéines BCL7 et de comprendre l'impact des mutations dérivées du cancer. J'ai utilisé des techniques

de cristallographie, de cryo-EM et de biologie moléculaire pour atteindre les objectifs suivants :

1. Déterminer la fonction moléculaire des protéines BCL7.

2. Elucider la structure 3D des protéines BCL7 seules ou en complexe avec des partenaires de liaison.

3. Évaluer l'impact des mutations dérivées du cancer sur la structure et la fonction de BCL7.

Résultats.

Déterminer la fonction moléculaire des protéines BCL7

Afin d'évaluer la liaison des protéines BCL7 à leurs partenaires de liaison, le nucléosome et l'ADN, des expériences EMSA ont été réalisées. Pour cela, de l'ADN marqué Cy5 601 a été amplifié par PCR, purifié par extraction sur gel et utilisé pour assembler le nucléosome par dilution saline. Les protéines BCL7A et BCL7C de pleine longueur se liant à l'ADN Cy5 601 et au NCP ont été testées, montrant que les deux protéines sont capables de se lier à l'ADN et aux nucléosomes. L'alignement des séquences de la famille BCL7 montre que les 50 premiers acides aminés sont conservés au sein des trois protéines et des différentes espèces, ce qui indique que cette région est très probablement le domaine fonctionnel. Il est intéressant de noter que cette région est la seule partie de la protéine où la structure secondaire peut être prédite. Pour réduire la région minimale d'interaction avec le nucléosome, nous avons effectué un alignement de séquences de différentes espèces de BCL7A et j'ai trouvé que les 100 premiers acides aminés sont hautement conservés dans cette protéine. La protéine BCL7A 1-100 purifiée est suffisante pour

se lier aux nucléosomes. Pour quantifier l'affinité de la protéine BCL7A complète avec les nucléosomes, des expériences de thermophorèse à micro-échelle (MST) ont été réalisées. Cy5-NCP à 80nM a été utilisé et différentes concentrations de BCL7A à partir de 30uM ont été utilisées pour titrer les nucléosomes. Un Kd de ~230 nM a été calculé.

Elucider la structure 3D des protéines BCL7 en complexe avec des partenaires de liaison.

J'ai pu purifier jusqu'à homogénéité les protéines BCL7A et BCL7C humaines. La BCL7A complète et les constructions plus courtes ont été surexprimées en fusion avec une étiquette GST clivable par protéase du virus Tobbacco Etch et N-terminale dans des cellules Rosetta (DE3). L'expression de la protéine a été induite par l'ajout d'isopropylthiogalactopyranoside (IPTG, 0,2 mM) pendant 18 heures à 18°C. La protéine a été purifiée par incubation avec des billes de GST; le tag a ensuite été clivé avec la protéase TEV. Après concentration, la protéine sans tag a été purifiée à l'aide d'héparine et de chromatographie par filtration sur gel. Les protéines purifiées ont été utilisées pour effectuer des essais de cristallisation, des études cryo-EM et des expériences de biochimie. Après avoir effectué des prédictions de structure secondaire dans PSIPRED et obtenu des spectres UV de dichroïsme circulaire, nous avons réalisé que les protéines sont pour la plupart non structurées, et que l'ajout d'un partenaire de liaison est nécessaire pour pouvoir obtenir des cristaux. J'ai réalisé un assemblage à grande échelle de nucléosomes recombinants et testé la liaison de BCL7A purifié pour ceux-ci. Les cristaux ont été obteneus avec des techniques de diffusion de vapeur en gouttes assises et suspendues, l'utilisation

de micro-ensemencement a également été réalisée. En fait, le complexe est stable et j'ai pu obtenir des cristaux qui ont diffracté à 6A. La purification d'une version plus courte de BCL7A 1-100 a été effectuée et j'ai pu obtenir une plus grande quantité de cristaux. J'ai essayé différentes stratégies afin d'améliorer la diffraction. Afin d'améliorer la qualité des cristaux, nous avons commencé à modifier les conditions initiales de cryoprotection en passant de 30% de 2-Méthyl-2,4- pentanediol (MPD), à 20% d'éthylène glycol, ou 35% de PEG400, et un mélange de MPD et d'éthylène glycol. En même temps, le complexe a été vérifié en coloration négative et la préparation des échantillons a été optimisée pour CryoEM. La BCL7C complète a été surexprimée et purifiée comme mentionné ci-dessus, à la seule exception que la chromatographie à l'héparine a été remplacée par une chromatographie d'échange de cations. J'ai découvert, après avoir effectué quelques essais de pull down, que cette protéine est capable d'interagir avec BAF47. Le complexe BCL7C-BAF47 a été purifié par double pull down de GST-BCL7C et de SUMO-BAF47 respectivement. Une chromatographie d'exclusion de taille a été réalisée et la présence des deux protéines a été confirmée par immunoprécipitation et spectrométrie de masse. Ce complexe binaire a ensuite été incubé avec des nucléosomes afin d'obtenir un complexe tertiaire que nous avons pu repérer dans les grilles. La préparation des échantillons pour la cryo-EM a été effectuée et j'ai pu collecter deux ensembles de données. Environ 1 700 000 particules ont été automatiquement prélevées et soumises à un traitement d'image. Le raffinement d'environ 500 000 particules a permis d'obtenir la carte 5.5 Å qui montre une densité supplémentaire après deux séries de classification 2D et plusieurs séries de raffinement homogène et hétérogène.

Évaluer l'impact des mutations dérivées du cancer sur la structure et la fonction de BCL7.

Pour évaluer si les mutations cancéreuses rapportées dans la base de données COSMIC affectent la liaison aux nucléosomes, j'ai effectué deux mutations ponctuelles indépendantes sur BCL7A, R11S et P78S, ces mutations ont été introduites à l'aide du kit de mutagenèse dirigée QuickChange II et confirmées par séquençage automatique de l'ADN. J'ai surexprimé et purifié les protéines comme décrit précédemment pour la protéine de type sauvage, et j'ai effectué une analyse SEC-MALS sur les mutants pour confirmer le repliement correct et évaluer la masse moléculaire exacte. L'activité de ces protéines a été testée avec des essais EMSA et j'ai montré que les mutations altèrent la liaison au nucléosome. Enfin, je me suis demandé si les mutations cancéreuses rapportées dans la partie C-terminale de la protéine BCL7 pouvaient également affecter la liaison au nucléosome. J'ai donc cloné et purifié BCL7C L210A comme décrit ci-dessus et j'ai comparé son activité avec celle de la protéine BCL7 wild-type. J'ai pu détecter que cette mutation n'affecte pas la liaison au nucléosome.

Conclusion

Les études structurales difficiles des protéines BCL7 confirment la capacité de ces protéines à se lier non seulement à l'ADN mais aussi aux nucléosomes, même si elles sont dépourvues de domaines canoniques de liaison à l'ADN. J'ai adopté deux approches structurelles: des essais de cristallisation isolant le complexe NCP-BCL7A (1-100) et nous avons mené des études cryo-EM du complexe ternaire NCP-

BCL7C-BAF47 car j'ai découvert que les protéines BCL7 interagissent également avec la sous-unité BAF47. La densité actuelle de la cryo-EM nécessitait une étude plus approfondie. BCL7A et BCL7C nécessite l'interaction avec les partenaires de liaison pour se stabiliser. Plus intéressant encore, des mutations ponctuelles dans les 100 premiers acides aminés de BCL7A affectent la liaison des protéines au nucléosome et ces mutations sont fréquemment trouvées chez les patients atteints de cancer.

Bibliography.

1) 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(6648), 251. 2) Kadoch, C., Hargreaves, D. C., Hodges, C., Elias, L., Ho, L., Ranish, J., & Crabtree, G. R. (2013). Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nature genetics, 45(6), 592 3) Baliñas-Gavira, C., Rodríguez, M. I., Andrades, A., Cuadros, M., Álvarez-Pérez, J. C., ÁlvarezPrado, Á. F., ... & Medina, P. P. (2020). Frequent mutations in the amino-terminal domain of BCL7A impair its tumor suppressor role in DLBCL. Leukemia, 34(10), 2722-2735

TABLE OF CONTENTS

16

FIGURES INDEX

ABBREVIATIONS

2YT: Yeast extract and Tryptone media

ACTB: Actin B

ACTL: Actin-like

ADD: ATRX-DNMT3-DNMT3L

ADP: Adenosine diphosphate

ADP-BeFx: ADP Berillium Fluoride

AML: Acute Myeloid Leukemia

ARID: AT-Rich Interactive Domain-containing protein

ARP: Actin Related Protein

Asf1: Anti-silencing function 1

Asp: Aspartate

ATP: Adenosine Triphosphate

ATRX: ATP-dependent helicase

BAF47: BRG1 Associated Factor 47

BAH: Bromo-Adjacent Homology

BAP: BRM-associated protein

BAZ: Bromodomain Adjacent to Zinc finger domain

BCL7: B-Cell Lymphoma 7

BICRA:BRD4-Interacting Chromatin Remodeling Complex Associated

BICRAL: BICRA Like

BME: Beta Mercaptoethanol

BPTF: Bromodomain PHD finger Transcription Factor

BRD: Bromodomain

BRG1: Brahma-related gene-1

BRK: Brahma and Kismet domain

BRM: Brahma

BS3: BisSulfosuccinimidyl Suberate

Cas9: CRISPR-associated protein 9

C2H2: Acetylene

cBAF: canonical BAF

CBP: CREB Binding Protein

CCRCC: Clear Cell Renal Cell Carcinoma

CECR2: Cat Eye Syndrome Critical Region protein 2 homolog

CENH3: Expressed Centromere Specific Histone 3

CENPA: Centromere protein A

CHD: Chromodomain Helicase DNA-binding

CREB: cAMP-response element binding protein

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

cryo-EM: cryo Electronic Microscopy

CTCL: Cutaneous T-Cell Lymphoma

CTD: C-Terminal Domain

DBD: DNA-binding Domains

DCD: Double Chromodomain

DLS: Dynamic Light Scattering

DNMT3: DNA methyltransferase 3

DNMT3L: DNA methyltransferase 3 like

DPF2: Zinc Finger Protein ubi-d4

DTT: Dithiothreitol

ECL: Enhanced Chemiluminescence

EDTA: Ethylenediamine tetraacetic acid

EMSA: Electrophoretic Mobility Shifts Assays

GH: Gating Helix

GLTSCR: Glioma Tumor Suppressor Candidate Region

GNAT: Gcn5-related N-acetyltransferases

GST: Glutathione S-transferase

H: Histone

HATs: Histone Acetyltransferases

HBL1: Hunchback-like protein

HDACs: Histone Deacetylases

HEK: Embryonic kidney cells

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HG-BCL: High-Grade B-cell non-Hodgkin Lymphoma

HKMT: Histone Lysine Methyltransferases

HP: Heparin

HRP: Horseradish Peroxidase

HSA: Helicase/SANT-Associated

HTH: Helix-Turn-Helix

IDP: Intrinsically Disordered Protein

IDR: Intrinsically Disordered Region

IGH: Immunoglobulin heavy chain

IPTG: Isopropyl β-d-1-thiogalactopyranoside

ISWI: Imitation SWI

ITC: Isothermal Titration Calorimetry

KCl: Potassium Chloride

kDa: kilo Dalton

KS: Lysine Serine

LANA: Latency-Associated Nuclear Antigen

LB: Luria Broth

LLPS: liquid-liquid PS

LMB: Laboratory of Molecular Biology

LSD1: Lysine-specific Demethylase 1

MBT: Malignant Brain Tumor

MDa: Mega Dalton

MPD: 2-Methyl-2,4-Pentanediol

MRT: Malignant Rhabdoid Tumors

MST: Microscale Thermophoresis

MYC: MYC proto-oncogene protein

MYST: MOZ, Ybf2 (Sas3), Sas2, and Tip60

NaCl: Sodium Chloride

NAP1: Nucleosome Assembly Protein 1

ncBAF: non-canonical BAF

NCP: Nucleosome Core Particule

NDR: Nucleosome Depleted region

NFR: Nucleosome-free Regions

NMR: Nuclear Magnetic Resonance

NTT: N-terminal Tail

NuA4: Nucleosome Acetyltransferase of H4

NuRD: Nucleosome Remodeling Deacetylase

NuRF: Nucleosome Remodeling Factor

O2P: Phosphate Oxygen

OD: Optical Density

OSA: Orthologue to Swi1

p53: tumor suppressor protein encoded by the TP53 gene

PAR: Poly ADP-Ribose

pBAP: Polybromo-associated BAP

pBAF: Polybromo-associated BAF

PBRM1: Protein Polybromo 1

PCR: Polymerase Chain Reaction

PEG: Polyethylene Glycol

PHD: Plant Homeodomain

PHF: PHD Finger Protein

PPPS: Polymer-Polymer PS

PRC2: Polycomb Repressive Complexes 2

PS: Phase Separation

PSIPRED: PSI-blast based secondary structure PREDiction

PTM: Post-Translational Modification

PWWP: Pro-Trp-Trp-Pro

QLQ: Glutamine-Leucine-Glutamine

Ran: RAs-related Nuclear protein

RanGTP: Ran Guanosine Triphosphate

RCC1: Regulator of Chromosome Condensation 1

RecA: Recombination Protein A

Req: Requiem

RFX:Regulatory Factor X

RNA: Ribonucleic Acid

RPM: Rotations Per Minute

RPT: Regulatory Protein T-lymphocyte

RSC: Chromatin Structure Remodeling complex

RSF1: Remodeling and Spacing Factor 1

Rtt: Regulator of Ty1 Transposition

RuvB: Recombination ultra violet B

Rvb:RuVB-like

SAM: S-Adenosyl Methionine

SANT: Swi3, Ada2, N-Cor, and TFIIIB

SDS-PAGE: Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

SEC: Size Exclusion Chromatography

SEC-MALS: SEC - Multi-Angle Light Scattering

SET: Su(var)3-9, Enhancer-of-zeste and Trithorax

Sfh1: Snf Five Homolog 1

SHL: Superhelical Location

SLIDE: SANT-like but with several insertions

SMARC: SWI/SNF-related Matrix-associated Actin-dependent Regulator of Chromatin subfamily

SnAc: Snf2 ATP coupling

SNF: Sucrose Non-Fermentable

SOX: Sry-related HMG box

Spt6: Suppressor of Ty 6

SS: Synovial sarcoma

SS18-SSX: Synovial sarcoma with t(X;18)(p11.2;q11.2)

Sth: SNF Two Homolog

SUMO: Small Ubiquitin-like Modifier

SWI/SNF: SWItch/Sucrose Non-Fermentable

TBE: Tris-borate-EDTA

TBS: Tris-Buffered Saline

TBST: TBS and Tween 20

TCEP: Tris(2-Carboxyethyl)Phosphine

TEV: Tobacco Etch Virus

TFBS: Transcription Factor Binding Site

TIP60: Tat-interactive Protein 60

Tris-HCl: Tris-(hydroxymethyl) aminomethane hydrochloride

TTD: Tandem Tumor Domain

Val: Valine

WH: Winged Helix

ZF-CW: Zinc Finger CW

Zn: Zinc

ZNF: Zinc Finger

1 INTRODUCTION

1.1 CHAPTER I CHROMATIN STRUCTURE

In eukaryotes, about 2-meter length of DNA are packed into the nucleus by interacting with a specific type of proteins to form higher-order structures. These small basic proteins called histones, first described by Albrecht Kossel in 1884 (Kossel & 1853-1927, 1928), assemble into octamers. Each histone octamer is composed of two copies of the histone proteins H2A, H2B, H3, and H4; the octamer organizes symmetrically around two central H3 histones that hetero-dimerize with two H4 histones forming a tetramer that interacts with two H2A/H2B heterodimers. About 147 base pairs of DNA wrap around the histone octamer core to form the basic unit of the chromatin, the nucleosome (Luger et al., 1997). DNA interacts mainly with the histone's structured regions, whereas the histone tails are flexible areas where post-translational modifications (PTMs) can occur (Baldi et al., 2020). It is essential to point out that every position where the major groove of the DNA faces the histone octamer is known as "superhelix location" (SHL) (Zhou et al., 2019), being SHL0 at the nucleosomal dyad and at the last region SHL +/- 7(Zhou et al., 2019). Overall the histone octamer interacts in 14 discrete places with the nucleosomal DNA.

Nucleosomes cores are bridged by linker DNA (~50 nucleotides long) and histone H1 that sits on the nucleosome core itself. In the famous model "beads on a string" linker DNA is the string and the nucleosomes cores are the beads (Figure 1B). Nucleosomes assemble into higher order structures, forming the 30nm chromatin fiber. This secondary structure is further compacted until mitotic chromosomes form

(Felsenfeld & Groudine, 2003) (Figure 1A). Nowadays, two different models are proposed to understand the organization of the 30nm fiber structure. The first and more well-known is the "solenoid model" in which nucleosomes are next to each and fold into a one-start helix. The second is referred to as the "zigzag model", and consists of two rows of nucleosomes arranged as zig zag in two rows forming a double helical structure (Tremethick, 2007).

Chromatin exists in two different states: euchromatin and heterochromatin.

1.1.1 Euchromatin

Euchromatin is characterized by an open chromatin structure and contains actively transcribed genes, including their *cis*-regulatory elements (Bannister & Kouzarides, 2011). Acetylation of histones is generally associated with euchromatin. In regions with actively transcribed genes, active enhancers, and promoters, specific histones PTM signatures are found. For instance, active genes are enriched in histone H3 lysine 36 trimethylation (H3K36me3), while active enhancers can be identified with a combination of high levels of histone H3 lysine 4 monomethylation (H3K4me) and histone H3 lysine 27 acetylation (H3K27ac) marks (Bannister & Kouzarides, 2011).

1.1.2 Heterochromatin

Heterochromatin is the condensed inactive form of chromatin. It is found mainly in the nucleus periphery, and it was described the first time when Emil Heitz was performing cytogenetic studies in different species of moss (Fedorova & Zink, 2008).
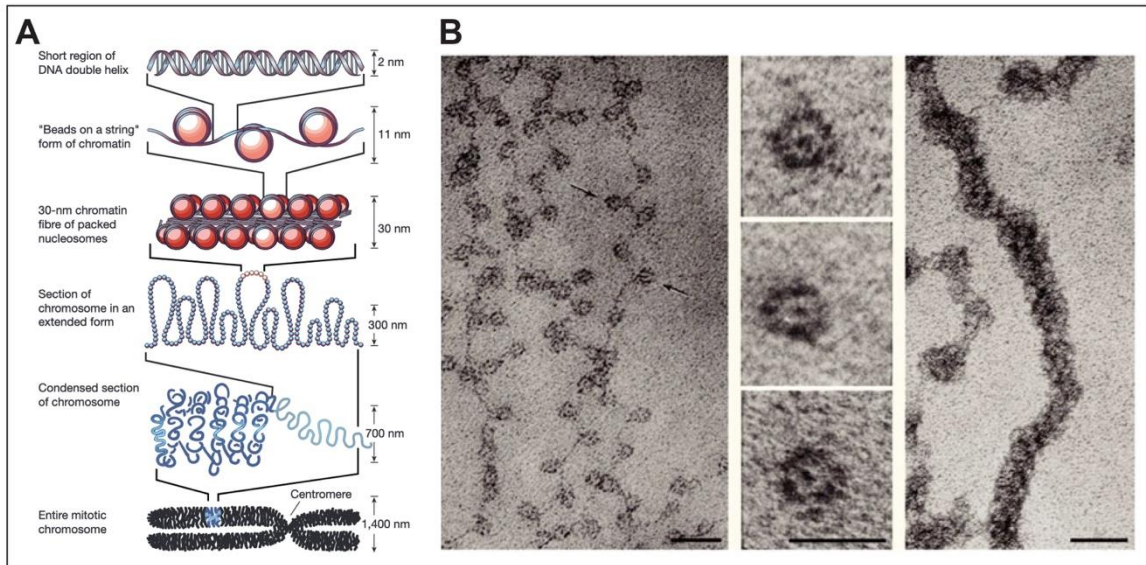
Figure 1. Organization of DNA within the chromatin structure.

A) Modified from (Felsenfeld & Groudine, 2003). Chromatin is a protein-DNA complex that facilitates the storage of DNA in eukaryotic cells. B) (Taken from (Olins & Olins, 2003) historical electron micrographs of chromatin showing beads on a string (left panel), single nucleosomes (center panel) and chromatin spread at moderate ionic strength.

Heterochromatin can be divided into two different classes: facultative and constitutive. Facultative chromatin is related to genes that need to be regulated during morphogenesis or differentiation signals. (J.-H. Lee et al., 2020). One of the best-characterized examples of facultative heterochromatin is the inactive X chromosome of mammals. On the other hand, constitutive heterochromatin is characterized by a strong condensation and is regarded as a mechanism to permanently silence genes such as centromere and telomere regions of repetitive DNA elements in all post-mitotic cells (Janssen et al., 2018). This type of chromatin does not codify for proteins.

1.1.3 Epigenetic Regulation of chromatin

The structure of chromatin is very important as an open conformation allows access to enzymes that can facilitate transcription, DNA repair, cell cycle, and other biological processes. Instead a closed chromatin conformation is refractory to the transcription machinery and other molecules. Variation to the chromatin fiber are introduced in different ways to control gene expression and activity: exchange of histone variants, post translational histone modifications (PTM), ATP dependent chromatin remodeling, DNA methylation, they all contribute to the epigenetic regulation of the chromatin fiber, a way in which the genome is interpreted without variation of the DNA nucleotide sequence. The word epigenetic comes from the Greek prefix "epi" (= over/outside/on the top) to indicate changes introduced "on the top" of the chromatin fiber. All these levels of regulations happen on the basic unit of

chromatin, the nucleosome, which is essentially an octamer of four histones wrapped around by DNA.

1.1.4 Histones and Histone variants

Histones are small basic and highly conserved proteins with a molecular mass of around 10-16 kDa. They have unstructured N- and C-terminal tails and highly conserved histone fold domains (Figure 2B). The histone fold domain is shared between the four histones despite the lack of sequence similarity. The domain is composed of three helices that are separated by two strap loops, and they form heterodimers by interleaving the helices into the "handshake motif" and juxtaposing the strap loops into parallel β bridges (Figure 2A) (Alva et al., 2007). Histones H3 and H4 are positioned in the inner region of the nucleosome, while H2A and H2B are located in the nucleosome surface (Giaimo et al., 2019). These proteins require chaperones that prevent them from making unspecific interactions with other proteins (Y.-J. Park & Luger, 2006); some of these chaperones are NAP1, Asf1, Spt6, etc.

The canonical histones carry N- and C-terminal tails of different lengths; these regions are often modified by covalent modifications, like methylation, acetylation, phosphorylation, and ubiquitination.

Linker histone H1 binds the nucleosomes with linker DNA and stabilizes higher levels of chromatin structure. In its canonical form is composed of a central globular domain, an unstructured N-terminal tail and a larger lysine-rich unstructured C-terminal region (Caterino & Hayes, 2011).
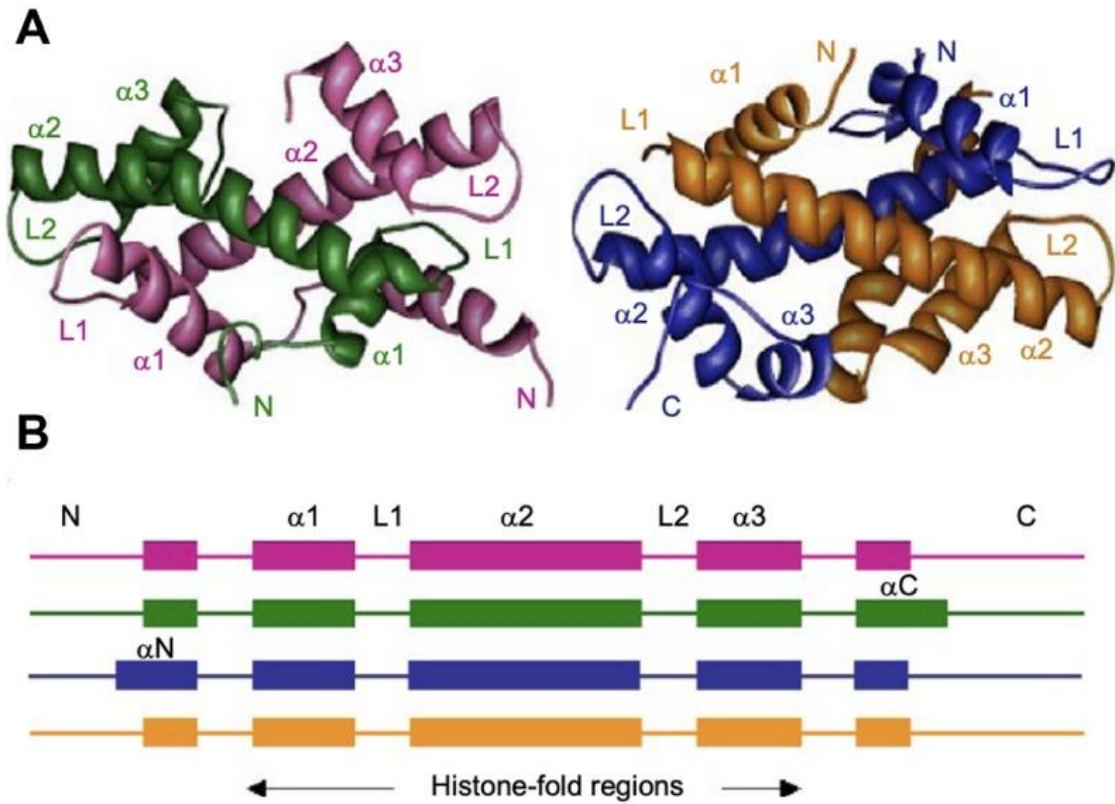
Figure 2. Architecture domain organization of the four canonical histones.

Taken from (Ramaswamy & Ioshikhes, 2013) (A) The secondary structure of handshake motifs; H3–H4 and H2A–H2B formed by the histones H3 (pink), H4 (green), H2A.Z (blue), and H2B (orange) and (B) their schematic representation as colored in (A).

1.1.4.1 Histone variants

In the three domains of life, Bacteria, Archaea, and Eukaryotes, highly basic proteins have evolved to neutralize the negatively charged phosphates of the DNA during the packing of their genome. Eukaryotes have histone variants, that can substitute the canonical histones. The replacement of canonical histones with histone variants is a dynamic process that leads to the regulation of several processes like replication, transcription, and heterochromatin formation (Talbert & Henikoff, 2017). The canonical histones have further differentiated into additional paralogues and variants. Canonical histones are encoded by multi-copy genes that do not contain introns. They are synthesized at S phase for rapid deposition behind the replication fork (Mei et al., 2017), The genes that code for histones variants are biallelic and can be subjected to splicing. Histone variants are deposited independently of replication and are expressed throughout the cell cycle (Giaimo et al., 2019), providing a mode of chromatin differentiation. Changes in chromatin composition by a histone variant is more frequent for histones H2A and H3, while for H4 there are not variants reported to date. The main histones variants are the following:

- **H2A.Z**

The function of this histone variant is still not clear, but it is found in almost all the eukaryotes and it is involved in several processes, such as transcriptional control and DNA repair. Interestingly, the high-resolution structure of the nucleosome containing H2A.Z (Suto et al., 2000) shows that this histone variant presents a larger acidic patch, which is an essential interaction surface for binding partners and

histone H3. Like the other histones, H2A.Z can be post translationally modified and the PTMs in this variant are associated with gene activation, but also with gene repression when the H2A.Z is ubiquitinated. H2A.Z is deposited near or within heterochromatin and it is thought to have a role into controlling aberrant spreading of silencing. This histone variant is inversely correlated with gene expression levels; presumably, this histone variant is evicted during transcription. H2A.Z is exchanged in the chromatin fiber by SWR1 complex, which exchanges a H2A-H2B dimer with a H2A.Z dimer in a ATP-dependent manner (Talbert & Henikoff, 2017). Removal of H2A.Z can happen during the eviction of nucleosomes at promoters or the variant can be specifically removed by INO80, that has been shown to exchange back the canonical H2A-H2B dimer.

- **H2AX**

According to one theory that emerged with some evolutionary analysis, the histone variant H2AX is one of the most ancestral histone forms, meaning that the canonical H2 could instead be the variant (Talbert & Henikoff, 2010, 2017). A four amino acids C-terminal motif of this histone variant can be phosphorylated at serine 139, at site of double-stranded DNA damage and, when phosphorylated, it is referred to as γH2AX. It is thought that this phosphorylation event provides early recruitment of the DNA repair machinery. γH2AX has been widely used as marker for double strand DNA breaks (DSBs) in cell biology.

- **MacroH2A**

The   histone variant macroH2A has a unique structural feature, as it carries an additional larger C-terminal globular domain of about 200 amino acids connected to

the N-terminal histone domain by a linker (Giaimo et al., 2019). It is specific to vertebrates, and it is enriched on the inactive X-chromosome. As with other histone variants, it can be associated with activation or repression of the transcription. It is a crucial player in stabilizing differentiated cell identity, acting as a barrier to somatic cell programming toward pluripotency, and acts as a tumor suppressor in a wide range of cancers (Z. Sun & Bernstein, 2019).

- **H3.3**

Transcriptionally active chromatin is maintained epigenetically by enrichment of the histone variant H3.3. H3.1 and H3.3 are highly conserved and H3.3 differs from the canonical H3.1 by only four-five amino acids. However, H3.3 is deposited in the chromatin fibre in a replication independent manner, whereas H3.1 is assembled into nucleosomes in a replication dependent mode. As a result of this process, actively transcribed regions are enriched with H3.3. These small yet fundamental differences in the histone sequence prevent H3.1 to be deposited by a replication independent pathway, ensuring epigenetic inheritance of active chromatin. Indeed, nucleosomes assembled with H3.3 are co-purified with HIRA and DAXX histone chaperons and nucleosomes assembles with H3.1 co-purify with CAF-1, testifying that they participate in distinct pathways (Ahmad & Henikoff, 2002; Henikoff & Smith, 2015). Incorporation of H3.3 into nucleosomes modulates higher ordered chromatin, resulting into an open chromatin conformation. It is enriched at dynamic regions such us promoter regions, cis-regulatory elements and gene bodies (L. Shi et al., 2017).

- **CenH3**

Also known as CENP-A in humans, this is the H3 variant that is abundant at centromeres and it is assembled into special nucleosomes that form the kinetochore

and it has a major role during mitosis and meiosis; The assembly of CenH3 containing nucleosomes at centromeres does not appear to require centromeric DNA sequences, setting an example for epigenetic inheritance. CENH3 has a wide variability depending on the species. Domains of this variant have evolved differently (Evtushenko et al., 2017). It has an extended N-terminal tail (NTT) and presents a limited number of PTM sites (Sharma et al., 2019)

1.1.4.2 Histone post-translational modifications (PTMs)

Pioneer studies on post-translational modifications (PTMs) were conducted by Vincent Allfrey in the early '60s. He described for the first time the modification of histone structures by the addition of acetyl and methyl groups (Allfrey et al., 1964). The field of post-translational modifications has been blooming in the past decades and now we know that a wide variety of modifications alter chromatin structure, recruit remodelling enzymes as other molecules involved in replication, DNA repair, and recombination.

Nucleosome are subjected to covalent modifications, also known as epigenetic marks. Histones can be modified by several post-translational modifications, that can modify chromatin structure and its dynamics in two ways, interacting directly with nucleosome components altering histone-histone or histone-DNA interactions, or targeting histone-binding domains or PTM's readers (Musselman et al., 2012) (Figure 3). The diversity and biological specificity associated with distinct patterns of PTMs have led to the hypothesis of a ¨histone code¨; a language encoded in the histone's tails (Strahl & Allis, 2000) that allows

other proteins as chromatin remodeling complexes and the transcriptional machinery to be recruited. The histone code is a hypothesis that the transcription of genetic information encoded in the DNA is regulated by chemical modifications to histone proteins, especially on their N-terminal flexible tails. The kind of post translational modification, the exact position at which the modification occurs determine different biological responses. There are enzymes that are responsible for writing the histone modifications (writers), some that read the modifications (readers) and others that remove them (erasers). Common PTMs are methylation, phosphorylation, acetylation and ubiquitination. Epigenetic modifications are made independently of actual DNA sequence changes and lead to the creation of the 'epigenome.

1.1.4.2.1 Histone acetylation

The highly dynamic process of deposition and removal of the acetylation mark is regulated by two prominent families of enzymes, with antagonistic activities: the histone deacetylases (HDACs) and histone acetyltransferases (HATS). These proteins regulate many transcription-mediated biological processes, and their aberrant activities are correlated with several human diseases (C. Y. Lee & Grant, 2019). Acetylation occurs on lysine (K) residues through the addition of an acyl group from an acyl-CoA donor to the ε-amino group of the K side chain on histones, and in some cases also other proteins (Bannister & Kouzarides, 2011). The addition of the acyl group masks the positive charge on the lysine residue, thereby reducing the affinity of the tail for chromatin, leaving the underlying DNA more exposed. The acetylation marks on lysine residues are read by small protein modules called

bromodomains (BrDs), sometimes referred to as "readers". These domains are conserved within many chromatin-associated proteins including some HATs, as well as other posttranslational modification enzymes (referred to as "writers") and ATP-dependent remodeling proteins. HATs are classified into two classes, type A and B. Type A HATs are divided into three subgroups: GNAT, MYST, and CBP/p300 (Hodawadekar & Marmorstein, 2007). These enzymes can modify multiple sites within the terminal histone tails and histone core. Another characteristic of this type of HATs is the association with multiprotein complexes (X.-J. Yang & Seto, 2007), regulating enzyme recruitment activity and substrate specificity (Bannister & Kouzarides, 2011). On the other hand, type B of HATs is mainly found in the cytoplasm, having an essential role in histone deposition. They acetylate residues K5 and K12 of newly synthesized H4 histones, excluding those already part of nucleosomes and some sites of H3 (Parthun, 2007). HDACs are erasers molecules that catalyze the reverse reaction restoring the positive charge of lysine residues, prompting a compact chromatin structure, rendering access by RNA polymerase difficult, and thereby decreasing gene repression. These enzymes are classified into four classes, I-IV, based on their homology to yeast HDACs. Among these groups Class I and II HDACs play a major role into acetylation of N-terminal histone tails. Class I, II, IV are Zn^{2+} dependent enzymes, whereas class III enzymes are NAD-dependent enzymes (S.-Y. Park & Kim, 2020). Imbalances between HATS and HDACs affect local chromatin compaction resulting into improper gene expression and genomic instability. Several anti-histone deacetylase drugs are currently available in the market to treat different types of cancer.
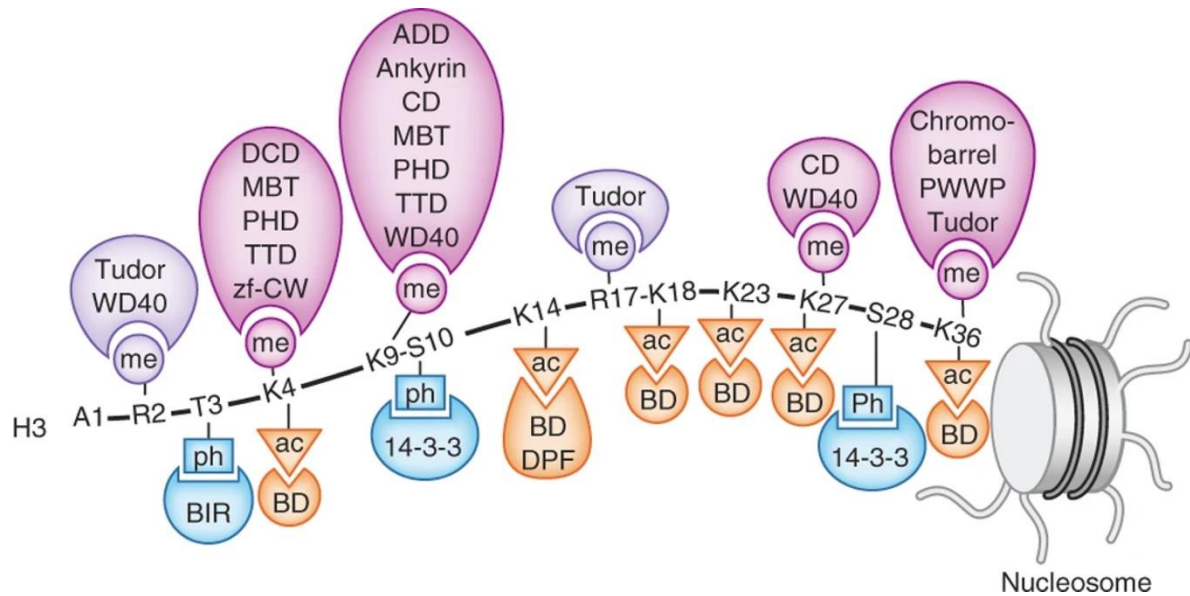
Figure 3. Readers of histone PTMs.

Taken from (Musselman et al., 2012). Cartoon representing an example of post translational modifications on histone H3 tail and common reader domains. The exact position of the residue carrying the mark is indicated.

1.1.4.2.2 Histone Phosphorylation

Histone phosphorylation occurs on serine, threonine and tyrosine residues mainly in the N-terminal histone tails. Phosphatases and kinases are the enzymes that remove or add a phosphate group from ATP to the target amino acid of the side chain. A large number of phosphorylation residues on histones have been described. Histone phosphorylation confers a negative charge to histones resulting in a more open chromatin conformation. Infact the addition of negatively charged phosphate groups to the N-terminal H3 tails may disrupt electrostatic interactions between the basic H3 tails and the negatively charged DNA backbone, thus increasing the accessibility of the genome to nuclear factors. It is thought that acetylation and phosphorylation both cooperate into facilitating gene expression. These two PTMs appear to be coupled on H3 for instance in response to EGF stimualtion: nucleosomes that are acetylated and phosphorylated on the H3 tail (pH3S10, AcH3K9 and AcH3K14) are associated with EGF activated genes. Phosphorylarion of H3S10, H3S28 and H2BS32 are associated with regulation of gene expression of proto-oncogenes such as c-myc and c-fos (Lau et al., 2011). In addition to transcription regualtion, histone phosphorylation is also linked to other cellular processes. For instance, phosphorylation of S139 on H2A.X is an important histone modification that plays a major role in DNA damage response (van Attikum & Gasser, 2005). Phosphorylation of H2A.XS139 is also linked to apoptosis induced signals (Rogakou et al., 2000) and phosphorylation of S10 has long been associated to mitosis and chromsome condensation (Bradbury, 1992). To conclude, histone phosphorylation partecipates in regulating gene expression, cell cycle, DNA damage and other molecular processes that regulate chromatine structure.

1.1.4.2.3 Histone methylation

Methylation occurs at the side chains of lysine (Lys) and arginine (Arg) residues, and this modification does not modify the charge of histones. Lysine residues can be mono-, di-, or tri-methylated, whereas arginine residues can be mono or symmetrically/asymmetrically demethylated (Ng et al., 2009). Histone methyl transferases (HMT) are the enzymes that catalyze the transfer of a methyl group on Lys and Arg residues at specific position on H3 and H4 tails. Histone lysine methyltransferases (HKMT) are highly specific enzymes that catalyze the transfer of a methyl group from S-adenosyl methionine (SAM) to the ε-amino group of lysines (writers). Most of HKMT contain a SET domain (Su(var)3-9, Enhancer of Zeste, Trithorax). Notable examples of HKMTs are SETD1, MLL, PRDM9 (meH3K4), SETDB1, Suv39H1 (meH3K9), SETD2,ASH1L (meH3K36), EZH1/2 (H3K27) (Husmann & Gozani, 2019). Arginine specific methyltransferases transfer a methyl group from SAM to the ω-guanidino group of arginine.

For a long time, methylation was considered an irreversible modification, but lysine-specific demethylase 1 LSD1 was identified (Y. Shi et al., 2004). The demethylation reaction requires a protonated nitrogen, and it is therefore only compatible with mono- and di-methylated lysine substrates (Bannister & Kouzarides, 2011). Methylation is a frequent PTM, arginine, and lysine residues are methylated, changing the hydrophobic character and size of these residues (Musselman et al., 2012). Lysines can be mono, di, or tri-methylated on histones H3, H4, and H1. Six lysines of histone H3 can be methylated: K4, K9, K26, K27, K36, and K79; for histone H4 K20 and finally for histone H1 K26. The readers of these PTMs are the most commonly studied, including ADD (ATRX-DNMT3-DNMT3L), ankyrin, Bromo-

adjacent homology (BAH), chromo barrel, chromodomain, double chromodomain (DCD), MBT (malignant brain tumor), PHD (plant homeodomain), PWWP (Pro-Trp-Trp-Pro), tandem Tudor domain (TTD), Tudor, WD40 and the zinc finger CW (ZF-CW). As mentioned above arginine is mono-methylated or symmetrically/ asymmetrically di-methylated. The most known arginine methylation sites are present in three histones, the first one histone H3: R2, R8, R17, R26; histone H2 R11 and R29 and finally in H4 the R3.

Mutations, genetic translocations and altered gene expression involving HMT enzymes are frequently observed in cancer, developmental disorders and other pathologies (Husmann & Gozani, 2019). Therapeutic compounds targeting specific KMT are currently being tested such as drugs targeting EZH2 the main H3K27 KMT.

1.1.4.2.4 Histone ubiquitylation

The ubiquitination of lysine residues consists on the addition of one (mono-ubiquitination) or more (poly-ubiquitination) ubiquitin monomers (Bannister & Kouzarides, 2011). Three enzymes are involved in ubiquitylation: the activating, conjugating, and ligating enzymes. Poly-ubiquitination creates an irreversible signal for proteosomal-mediated degradation, whereas mono-ubiquitination generates a regulatory signal, which can be reversed by the action of ubiquitin-specific proteases (USPs/UBPs) called deubiquitinating enzymes. In humans, the best-known example of histone ubiquitylation is the histone H2A lysine 119 ubiquitylation (H2AK119ub) implicated in gene silencing and found in facultative heterochromatin. Another example is histone H2B lysine 120 ubiquitylation (H2BK120ub) that is frequently

found in actively transcribed genes; this is a good example of the divergent role that a histone modification can hold, depending on its position (Bannister & Kouzarides, 2011). Histone ubiquitination plays a role into DNA replication, transcription and repair. How the mark at a specific position is read by chromatin enzymes triggers different outcomes.

To conclude, a variety of histone post translational modifications introduce variation into the chromatin fiber, modulating its structure and as a consequence controlling genome expression, genome integrity and other cellular processes.

1.1.4.3 ATP dependent chromatin remodeling

Another mechanism by which nature controls chromatin structure is the ATP-dependent chromatin remodeling. ATP-dependent multiprotein remodeling complexes (remodelers) remodel chromatin using different mechanisms to achieve nucleosome organization, disorganization, ejection or changes in nucleosome composition. But they all have in common an ATP-ase "motor" that translocates DNA from a specific position on the nucleosome, breaking histones-DNA contacts (Clapier et al., 2017) and moving the DNA along the histone surface. Another chapter of this thesis is entirely dedicated to chromatin remodelers.

1.1.5 Nucleosome core particle, basic unit of chromatin

1.1.5.1. 601 DNA sequence

Lowary and Widow made an extensive study of different DNA sequences that led to the discovery of an artificial sequence with a higher affinity for histones than other native sequences; they named it 601 (Lowary & Widom, 1998). Besides the high

affinity that this sequence shows for histones, it does not interfere with the behavior of a native nucleosome. The sequence is 282 base pairs, of which 147 base pairs correspond to the original core sequence. In addition, 135 base pairs are part of the linker regions that flank both sides of the core.

The first nucleosome structure reported containing this sequence showed that 145 bp of the Widom 601 DNA wraps around the histone octamer to form the nucleosome core particle instead of 147 bp of DNA in the canonical human α-satellite nucleosome core particle (Makde et al., 2010, p. 1). We can say that this sequence is the most used in chromatin biology studies in vitro.

1.1.6 Nucleosome structure

The nucleosome is composed of two copies of the histones H2A, H2B, H3, and H4 that form an octamer that is wrapped around by a left-handed superhelix DNA with an approximated length of 147 base pairs. The first high-resolution nucleosome structure showed that the main contacts between histones and DNA are made through the evolutionary well-conserved histone domains (Luger et al., 1997). The globular domains of histones organize 120 base pairs of DNA, while the remaining length of DNA interacts with a unique N-terminal α helix coming exclusively from H3. It is very important to note that every location where the major groove of DNA faces the histone octamer will be designed as a superhelix location (SHL) (Figure 4) if we start counting from the dyad of the nucleosomes being the SHL0 the first position, until SHL7 at the very last region of histone-bound DNA. The histone core is composed by asymmetric hetero-tetramer, defined by hydrophobic interactions between H3 and H4, that interacts at the same time with two H2A-H2B dimers

through several interactions that include a four-helix-bundle structure between H2B and H4 (Zhou et al., 2019) and some other interactions between H2A and the hetero-tetramer H3-H4. Then the side chains of H2A form a remote interface with the L1 loops into the nucleosome core.

Nucleosomes are very dynamic structures in terms of composition and conformation, and they can transition to different states depending on their histone composition and the kind of PTMs they carry. At the same time, nucleosomes have an intrinsic property that causes partial and rapid DNA unwrapping and rewrapping; this is known as DNA breathing (Zhou et al., 2019). In summary, nucleosomes are the highly dynamic unit of chromatin that carries numerous epigenetic marks; therefore, different proteins will interact with them, and each interaction will modify the structure and dynamic.

Another important concept is nucleosome spacing. Nucleosomes usually form vast arrays, and the distance between the neighboring nucleosomes tends to be constant (A. K. Singh & Mueller-Planitz, 2021); this distance is species and tissue-specific (Flaus, 2011).
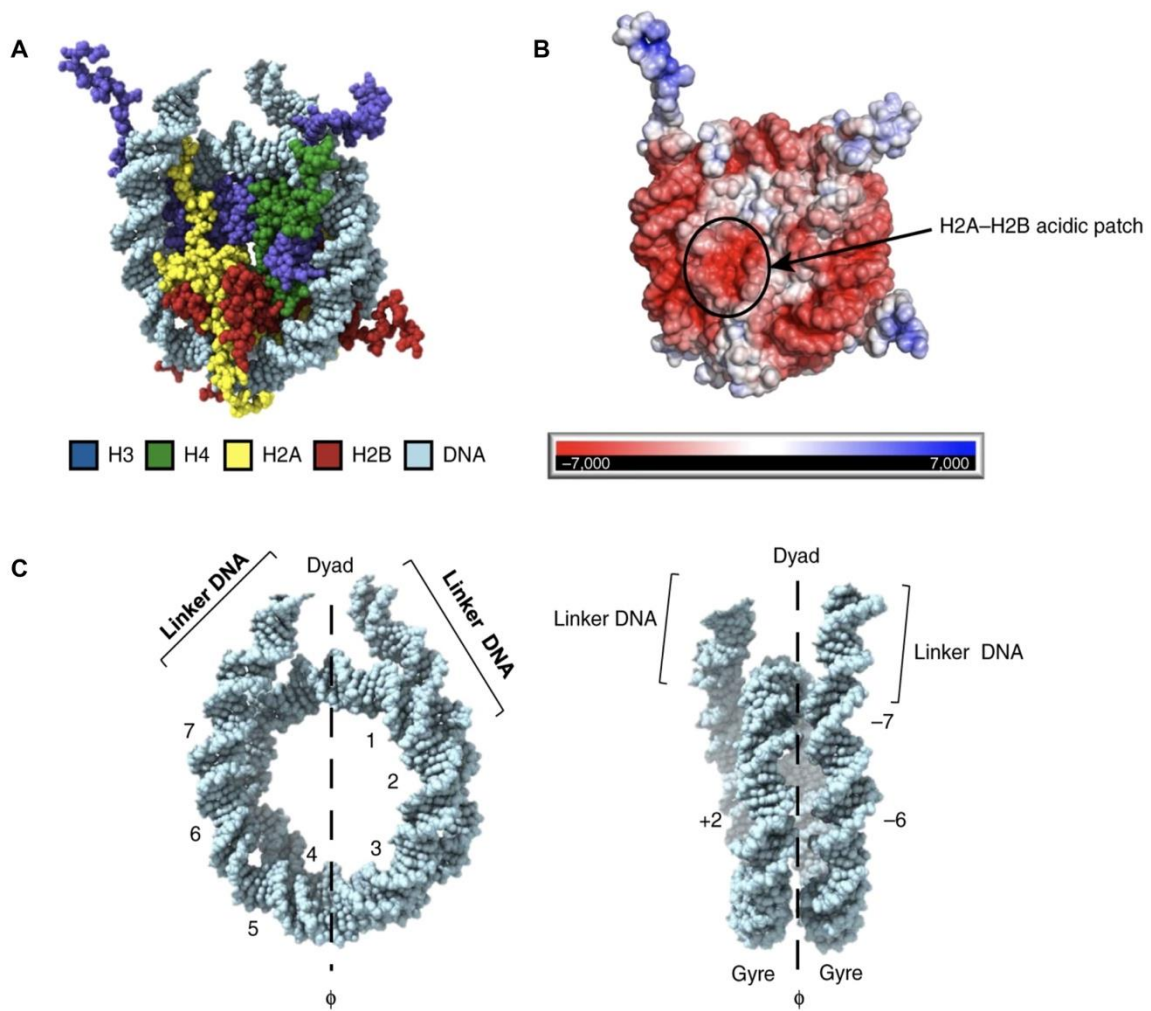
Figure 4. Nucleosome structure.

Taken from (Zhou et al., 2019) A) Nucleosome disc view, model, derived from PDB 1KX5 and PDB 1ZBB79 (DNA). B) Electrostatic potential of the nucleosome surface C) Nucleosomal DNA and linker DNA (from PDB 1ZBB). Along the two-fold axis, nucleosomal DNA (145–147 bp) can be divided into two gyres (approximately 72 bp each). The SHL designation represents the position of each major groove facing inward. The dyad (center of the nucleosomal DNA) is defined as position 0. The numbers 1–7 highlight the SHLs on DNA. Linker DNA is the extranucleosomal DNA, which is located next to the entry-exit site of nucleosomal DNA.

1.1.6.1 Nucleosome depleted regions

Proteins that bind specifically to DNA sequences, such as transcription factors or RNA polymerase II, can encounter restricted access mediated by nucleosomes. Frequently active gene promoters and enhancers are found within so-called nucleosome-depleted regions (NDR), which are flanked by a -1 upstream and a +1 upstream highly positioned nucleosomes (Hughes & Rando, 2014). Nucleosome occupancy at gene promoters is negatively correlated with gene expression levels, and it is well known that highly transcribed genes display a stronger nucleosome depletion than low expressed genes(Hughes & Rando, 2014). Gene promoters can be distinguished based on distinct patterns of nucleosome positioning.

1.1.7 Nucleosome recognition

For a very long time, the nucleosome structure was described as a repressive structure that avoids the interactions of transcription factors with the DNA (McGinty & Tan, 2016), but nowadays, we know that most of the DNA on the outer part of the nucleosome is solvent accessible. Some transcription factors can recognize nucleosomes with the DNA grooves aligned, increasing the affinity for nucleosomes over naked DNA.

A large history has been written down since the first nucleosome structure was solved in 1997 (Luger et al., 1997): almost ten years passed till the first co-crystal between nucleosome and LANA peptide was determined (Barbera et al., 2006), and then few years later the first nucleosome structure with RCC1 protein (Makde et al., 2010), etc. After these major efforts, it has been determined that the binding between a nucleosome and a chromatin factor is not dependent on a unique

single component; indeed, proteins recognize multiple elements in the nucleosome (Figure 5), often across multiple and non-contiguous nucleosomal surfaces (McGinty & Tan, 2016).

Around 40 % of the solvent-accessible surface area of the nucleosome is formed by the histone core surface (Zhou et al., 2019), and from this area, the acidic patch formed between H2A-H2B seems to be the most recognizable area for chromatin-binding proteins. In humans, this patch is composed of one aspartic acid, six glutamic acids from H2A, and two more glutamic acids from H2B (McGinty & Tan, 2016).

As mentioned before, the first peptide that was co-crystallized with the nucleosome and found that interact with the acidic patch is the latency-associated nuclear antigen (LANA) peptide; this protein mediates viral genome attachment to the mitotic chromosome (Barbera et al., 2006). LANA peptide has an arginine side chain that was described later as the arginine anchor. Other proteins as RCC1 require additional contacts with nucleosomal DNA facing the histone core through a DNA-binding loop and its N-terminal region (Makde et al., 2010; Zhou et al., 2019).
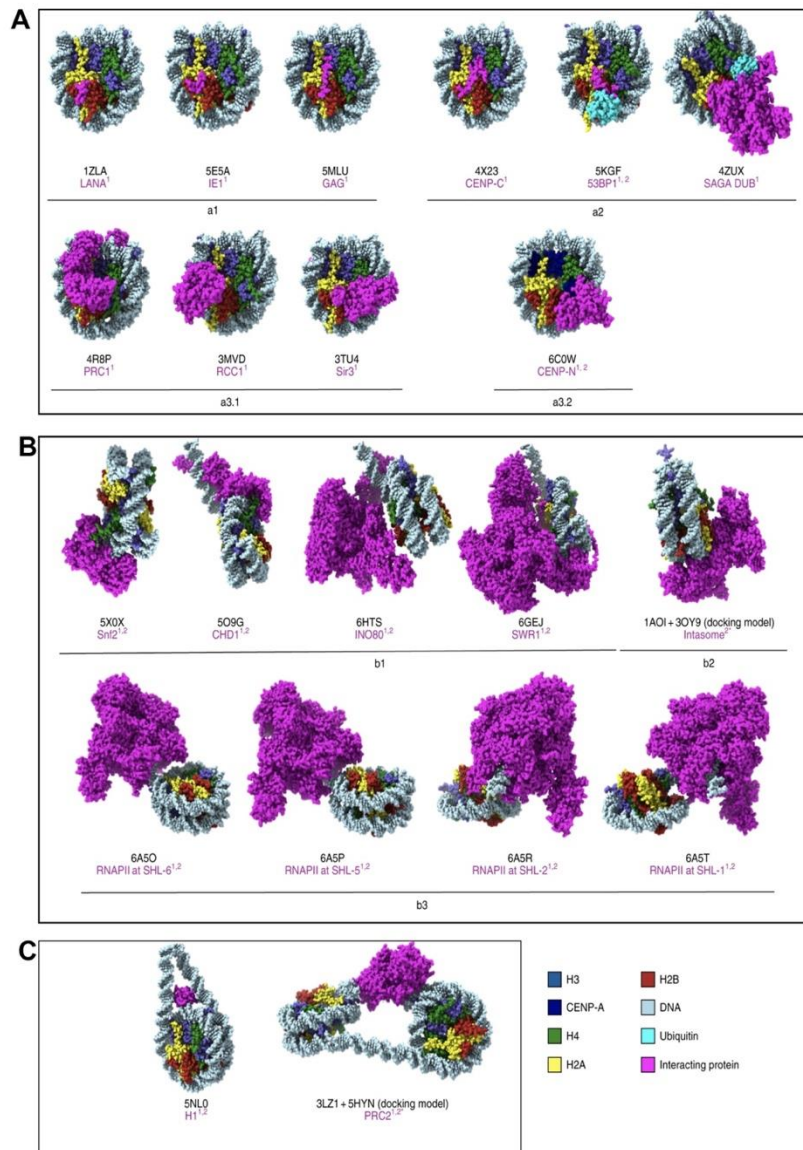
Figure 5. Structures of chromatin-binding factor complexes.

Taken from (Zhou et al., 2019), PDB entries of the structures are shown in the figure. "1" indicates structures obtained by crystallography. "2" indicates structural models built from single-particle cryo-EM maps. "2*" represents docking models generated to interpret cryo-EM maps. A) Proteins targeting the surface of nucleosome: a1, small protein fragments or polypeptides recognizing the acidic patch on the nucleosome surface; a2, proteins recognizing both the acidic patch and epigenetic marks on the nucleosome surface; a3, proteins binding to both histones and nucleosomal DNA on the nucleosome surface (the acidic patch also plays an important role in complex a3.1 but not in a3.2). B) Proteins invading nucleosomal DNA gyres. C) Proteins interacting with linker DNA.

1.1.7.1 NCP recognition by RCC1

The first chromatin protein nucleosome complex structure with 2.9 Å resolution was obtained from crystals of Drosophila RCC1 (regulator of chromosome condensation) bound to the recombinant NCP. The complex size is 300kDa and represented a huge step forward into understanding chromatin recognition. This structure, at the same time, was the first structure of a nucleosome core particle containing the Widom 601 (Makde et al., 2010, p. 1). RCC1 is a protein that generates RanGTP gradients signal that regulates mitotic spindle formation, transport between cytoplasm and nucleus, and nuclear envelope formation. This protein recruits Ran to chromatin and activates Ran's nucleotide exchange activity. RCC1 interacts with the histones and unexpectedly with the DNA.

1.1.7.1.1 RCC1-histone interactions

RCC1 contains a β-propeller domain; this kind of protein usually uses the loops on either face of the β-propeller toroid to interact with other proteins. One of the faces interacts with Ran24, and it was suspected that the opposite face could bind the nucleosome. However, the crystal structure and biochemical data suggested that the β-propeller uses a switchback loop to bind to the acidic patch of the nucleosome. This switchback loop is in the borders of the face that interacts with Ran. Two molecules of RCC1 interact individually with the two depressions formed by residues Glu61, Asp90, and Glu92 of the H2A/H2B acidic patch. Arg 216 and Arg 223 are responsible for these interactions, while a network of hydrogen bonds and Van der Waals interactions mediate the interactions (Makde et al., 2010, p. 1).

Very interestingly, several similarities are seen between RCC1 and LANA peptides. First, both employ two arginine side chains to interact with histone dimer residues, secondly both make van der Waals contacts with H2B Glu102, Leu103, His106, and Val45, and finally, both contain serine side chains that interact with H2A Glu64 (Makde et al., 2010, p. 1).

1.1.7.1.2 RCC1-nucleosome DNA interactions

RCC1 has a DNA binding loop that interacts with nucleosomal DNA via phosphate interaction. The interactions occur between the Lys241 and Arg239 across the major groove of the DNA very close to the SHL6. The NH1 atom of Arg239 interacts in two different with the DNA phosphate group, either through charged interactions or possibly through a hydrogen bond (Arg239 NH1 is 3.5 Å from the Guanine131 O2P atom) (Makde et al., 2010, p. 1). An additional interaction between RCC1 and the nucleosomal DNA is made between the Gln259 located in an adjacent loop. These additional interactions lead to the idea that RCC1 is a non-DNA-sequence-specific chromatin protein (Makde et al., 2010, p. 1).

1.1.7.2 Pioneer transcription factors

Transcription factor binding sites (TFBS) are located within the NDR, but these can be occupied by nucleosomes that prevent binding. However, some transcription factors can bind their cognate DNA sites directly, even in chromatin that is locally compacted by linker histone (Iwafuchi-Doi & Zaret, 2016). They are believed to promote chromatin accessibility in these regions, and that is why they are named

"pioneer transcription factors." A very well know pioneer transcription factor is Sox2; this protein enables the disengagement of terminal nucleosomal DNA from histone proteins and repositioning of the N-terminal tails of histone H4 (Dodonova et al., 2020).

1.1.7.2.1 Pioneer transcription factors SOX2 and SOX11

As mentioned above, pioneer transcription factors bind nucleosomal DNA to allow gene expression to form regions of the genome where chromatin is present in a close conformation. These factors are required for stem-cell pluripotency, cell differentiation, and cell reprogramming (Dodonova et al., 2020; Iwafuchi-Doi & Zaret, 2016; Takahashi & Yamanaka, 2006). Almost two years ago, cryo-EM structures of the human DNA-binding domains (DBD) of SOX2 and SOX11bound to the nucleosome were determined. In this study, a 147 bp DNA sequence named DNA-1 was used for the nucleosome reconstitution; the sequence was selected because it is known to bind the related factor SOX11 (F. Zhu et al., 2018).

For the complex formed between the nucleosome and DBD of SOX2, a 5.5 Å resolution map was obtained and fitted with the crystal structures of the nucleosome and SOX2 DBD; the structure revealed that a single copy of the protein binds to the SHL2. This position, located in the center of the nucleosome, involves specific interactions with the DNA motif TTGT. The resolution of the protein could not be improved; therefore, they pursue the complex made between SOX11 and the nucleosome. In this case, they achieved 3.7 Å resolution, and to facilitate the model

building, they determine the crystal structure of SOX11 DBD with a DNA fragment at 2.5 Å resolution.

1.2 CHAPTER II CHROMATIN REMODELING

DNA replication, transcription, and the different pathways that allow DNA repair are processes that must overcome the physical barriers that chromatin structure provides (Ribeiro-Silva et al., 2019). Several mechanisms act in coordination to give flexibility to the chromatin structure. As described above, the exchange of canonical histones for histone variant, the covalent modifications in histones tails (acetylation, methylation, etc.), and the ATP-dependent chromatin remodeling complexes act on chromatin to regulate its dynamic structure (El Hadidy & Uversky, 2019).

ATP-dependent chromatin remodeling complexes contain an ATPase subunit. This subunit belongs to the DNA helicases family SNF2. Based on the amino acid sequence, this family is further classified into four different types: SWI/SNF (switch/sucrose non-fermentable), INO80, ISWI (imitation SWI), and CHD (chromodomain helicase DNA-binding) (Hota & Bruneau, 2016) (Figures 6,7).
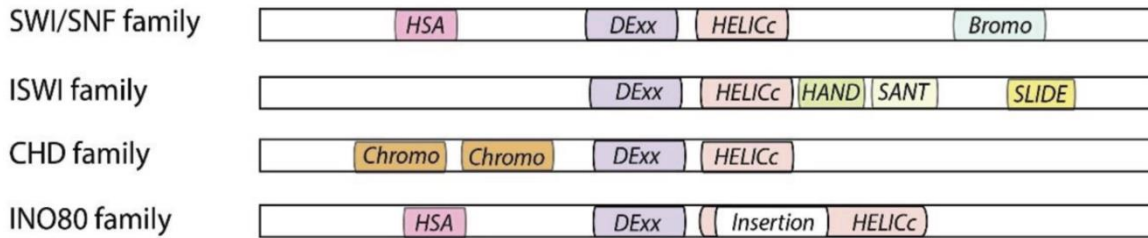
Figure 6. Architecture organization of the ATPase subunits.

Taken from (Längst & Manelyte, 2015) Domain architecture organization of the ATPase subunits found in different ATP-dependent chromatin remodeling complexes.
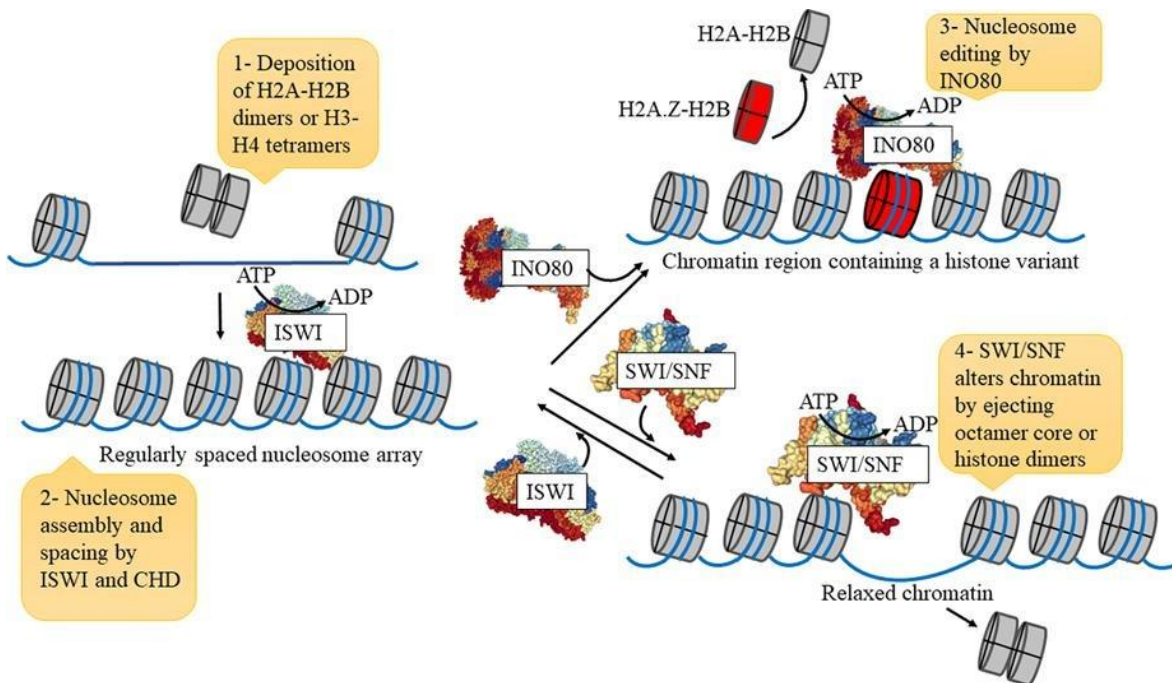
Figure 7. Function of the four families of chromatin remodeling complexes.

Taken from (Sahu et al., 2020). Four families of chromatin remodeling complexes are known. The different actions of these complexes regulate the dynamic changes of chromatin structure. ISWI and CHD can slide nucleosomes, SWI/SNF can eject them.

ATP-dependent chromatin remodelers operate via diverse mechanisms of action and engage different proteins to aid nucleosome recognition; however, they all utilize the "ATP-ase translocase motor" to translocate DNA from a precise position on the nucleosome (Clapier et al., 2017). They all converge to regulate a DNA translocation mechanism which translates into nucleosome sliding, nucleosome eviction, nucleosome spacing, and dimer exchange (Yan & Chen, 2020).

To gain insight on how chromatin remodelers break the histone-DNA interactions it is first important to understand how the catalytic subunit engages the nucleosome. The catalytic subunits Snf2 of yeast SWI/SNF, and ISWI of ISWI family, bind the nucleosome at position SHL2, which is two DNA helical turns away from the nucleosome dyad. Instead INO80 engages the nucleosome at the DNA entry site (Brahma et al., 2017). However, a major difference exists between INO80 and ISWI, being ISWI more efficient, bringing up the hypothesis that SHL2 is a strategic site for chromatin remodeling (Yan & Chen, 2020).

The Snf2 family has been considered a DNA translocase that travels along a DNA duplex tracing only the 3'-5' strand (Singleton et al., 2007; Yan & Chen, 2020). Several non-mutually exclusive mechanisms are contemplated in the literature, and they propose local perturbations of DNA that then propagate throughout the nucleosome. The first model is referred to as twist diffusion model, where a twist distortion is created when a segment of DNA absorbs one base pair, and this movement is propagated to the next DNA segment allowing the movement of the nucleosome of one base pair by one. In this model, the histone-DNA contacts are maintained, but the DNA is rotated by 35°. The second model argues for the formation of a DNA loop of about ten base pairs length generated by the absorption

of DNA into the nucleosome. This large loop disturbs the histone contacts but maintains the rotational phasing (Yan & Chen, 2020). The third model, called the wave-ratchet-wave model, couples the DNA translocation activity to DNA displacement, proposing that a subdomain of the remodeler anchors the enzyme at a specific site on the octamer and the torsion of this pull in and push out the nucleosomal DNA; this model is very similar to the inchworm-like model in helicases.

Despite these three models, how the ATPases induce the DNA twist, loop or wave remains unknown. These questions inspired structural studies that, in the case of Snf2, the ATPase subunit of the yeast SWI/SNF complex (homolog of SMARCA4/BRG1 in humans) led to the proposal of the wave model.

1.2.1 The wave model of Snf2

The ATPase motor domain of Snf2 can slide nucleosomes by itself, and its activity is regulated by the HSA and SnAc domains. In the absence of its substrate, the enzyme adopts an open conformation and stores a 1 bp bulge in its primary DNA-binding cleft at SHL+2 (Yan & Chen, 2020). When ATP is bound to Snf2, it changes to a closed conformation, during which lobe 2 tightly grips the tracking strand and pushes the DNA bulge toward the exit side, completing the 1 bp DNA translocation (Yan & Chen, 2020). After ATP hydrolysis, the phosphate is released, and Snf2 reopens the DNA-binding cleft inducing a 1bp bulge at SHL+2, pulling the DNA from the proximal end where the nucleosomal DNA is distorted and prepared for the next remodeling cycle. In each cycle Snf2 translocates 1 bp of DNA per ATP hydrolysis (Figure 8); this mechanism is consistent in the RSC complex.
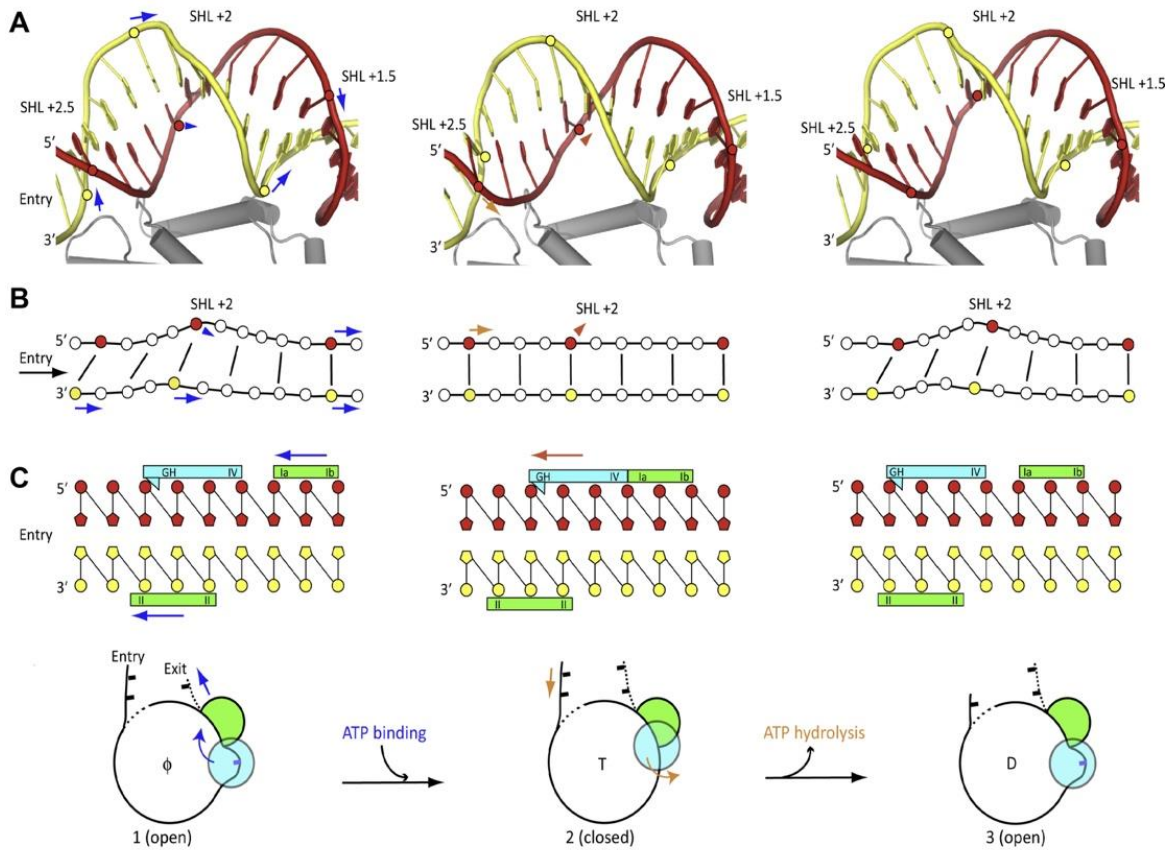
Figure 8. DNA wave model of chromatin remodeling.

Taken from (Yan & Chen, 2020) (A) DNA movements are relative to the histone core in an ATPase cycle (the histone core is the reference point, and the DNA moves). The region around superhelical location (SHL) +2 is shown; the 5′-tracking strand (red) and the 3′-guide strand (yellow) are also depicted. The first, second, and third columns illustrate the apo, ATP-bound, and ADP-bound states, respectively. Arrows indicate the direction of DNA translocation. (B) DNA distortion around SHL +2. (C) Motions of lobes 1 (green) and 2 (cyan) of the motor relative to the DNA Helicase motifs Ia, Ib, II, III, and IV, and the gating helix (GH) are labeled. Arrows indicate the movement of the lobes.

1.2.2 INO80 (Inositol requiring 80)

The INO80 multiprotein complex is highly conserved in all eukaryotes, and it was described and characterized for the first time in *Saccharomyces cerevisiae* (Shen et al., 2000). The complex has a molecular mass of 1.3 MDa, and it is composed of 10-15 subunits mostly found with exact stoichiometry (Shen et al., 2000). It modifies chromatin in different ways, by sliding nucleosomes and exchanging histones. For instance, it can catalyze the exchange of H2A.Z/H2B dimers with free H2A/H2B (Papamichos-Chronakis et al., 2011). The complex performs other activities like DNA repair (specifically at double-strand breaks), represses inappropriate transcription at promoters in the opposite direction to the coding sequence (Willhoft & Wigley, 2020).

The INO80 family of remodeling enzymes is composed of two classes of enzymes, INO80 and SWR1. A universal characteristic of the members of the INO80 family is the hexameric ring of RuvB-like proteins. This structure acts as an axis for the motor subunit, and it is an architectural scaffold upon which other subunits are assembled (Willhoft & Wigley, 2020). The complex also contains actin and actin-related proteins crucial for histone recognition and fundamental for remodeling activity (Shen et al., 2003). The complex has a HSA (helicase-SANT-associated) domain with DNA binding activity. After interaction with the actin and Arps proteins, the HSA domain acquires a more extended α helical conformation, the main characteristic that defines the INO80 family. Currently, there is no structural information on the HSA-Arp module, but evidence shows that it occupies the nucleosomal linker DNA (Figure 9) (Willhoft & Wigley, 2020).

INO80 interacts with the nucleosome in two different positions, the first site of interaction is at SHL6/7, and the second occurs at SHL2/3 (Figure 9). The sliding activity requires the dimerization of two INO80 complexes, and this event can occur via the interaction of the C-terminal domain of the INO80 subunits with DNA. The interaction allows crosstalk between the complexes. However, the ATPase activity is only necessary for one of the complexes that form the dimer for sliding (Willhoft et al., 2017).

1.2.3 ISWI (Imitation switch)

In humans, the ISWI family forms a heterodimer composed of two ATPase subunits SMARCA1 or SMARCA5, and includes one of the six regulator subunits, BAZ1A, BAZ1B, BAZ2A, BPTF, CECR2, or RSF1 (Oppikofer et al., 2017). Some auxiliary subunits like histone binding proteins and histone fold proteins can bind to the heterodimeric core allowing the formation of different family members. For instance, the best-known complex NURF is formed by the association of SMARCA1 and the regulator subunit BPTF. ISWI needs to be activated upon interaction with the tail of H4; therefore, it needs to compete with the substrate of H4: the neighboring nucleosomes. It has been proposed that ISWI piles up nucleosomes, allowing them to compact and lead to gene inactivation (Ludwigsen et al., n.d.). The ATPase-Translocase domain of ISWI has two RecA-like lobes (lobe1 and lobe2) separated by an insertion stretch. The ATPase and coupling regulatory modules of ISWI, Auto-N, and Neg-C flank the translocase motor (Figure 10). The C-terminal region of ISWI is responsible for the sliding activity of the complex. The SANT and SLIDE domains are composed of three helices that resemble a DNA binding domain similar to the DNA binding domain of the C-Myb transcription factor.
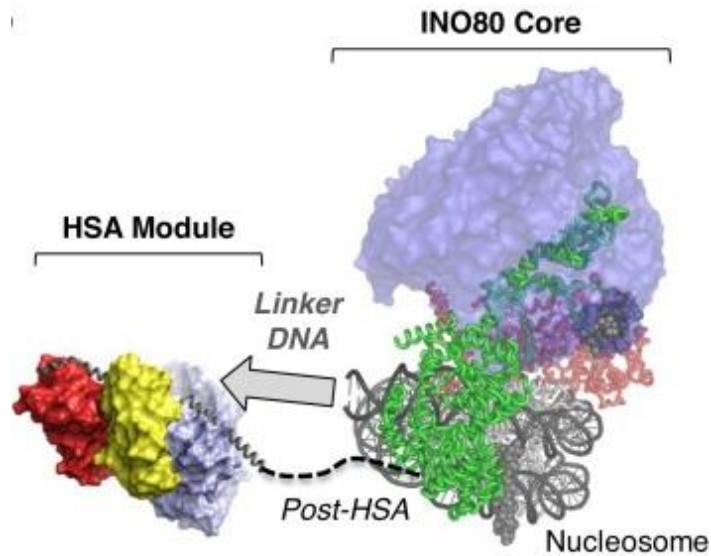
Figure 9. Structure of INO80 bound to the nucleosome.

Modified from (Willhoft & Wigley, 2020). Representation of INO80, the HAS module composed with Arp4 (red), actin (yellow), and Arp8 (light blue), and the proposed location relative to the motor and heterohexamer formed by Rvb1/2.
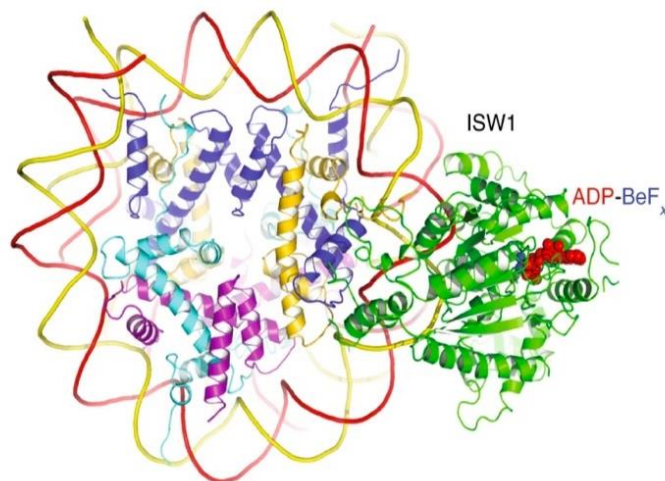
Figure 10. ISWI structure bound to the nucleosome.

Taken from (Yan et al., 2019) A) Overall Structure of the ISW1–nucleosome complex in the ADP-BeFx-bound state.

1.2.4 CHD (Chromodomain Helicase DNA-binding)

The functional role of the family is the least known of all the chromatin remodelers. It has been shown that Chd1 can operate on parallel routes with other chromatin remodelers. Chd1 can shift nucleosomes and induce regular nucleosome spacing. Chd1 contains the DNA-binding domains SANT and SLIDE, a bilobal motor domain that hydrolyses ATP, and a regulatory double chromodomain (Figure 11A). The cryo-EM structure of the yeast Chd1 bound to a nucleosome was reported at a resolution of 4.8 Å (Farnung et al., 2017) (Figure 11B) and shows Chd1 unwrapping two turns of DNA from the histone octamer and binding between the two DNA gyres. The SANT and SLIDE domains contact detached DNA at SHL -7, the ATPase motor binds at SHL +2 and is anchored to the N-terminal tail of histone H4, like observed in Snf2 ATPase structure. The double chromodomain swings toward the nucleosomal DNA at SHL +1, closing on the ATPase subunit. With this movement, the ATP-ase can translocate DNA toward the nucleosome dyad, starting the nucleosome remodeling (Farnung et al., 2017).

These complexes were identified for the first time as a DNA binding factor with a helicase domain. The CHD family consists of nine members, CHD1-9, divided at the same time into three subclasses. The first class corresponding to CHD1 and CHD2 promotes the stabilization of the histone variant H2A.X and the repair by homologous recombination of double-stranded breaks in DNA (Mills, 2017). Subclass II corresponding to CHD3-5 is very similar to class I as it regulates DNA repair and is involved in transcription, but additional functions include potent modulators of cellular proliferation, senescence, and apoptosis (Bagchi et al., 2007; Mills, 2017). Interestingly CHD5 is considered a member of the NURD chromatin

remodeler family (Quan & Yusufzai, 2014). Finally, subclass III is composed of the last three members, involved in transcription and DNA repair and in the development and neurological syndromes (Mills, 2017).
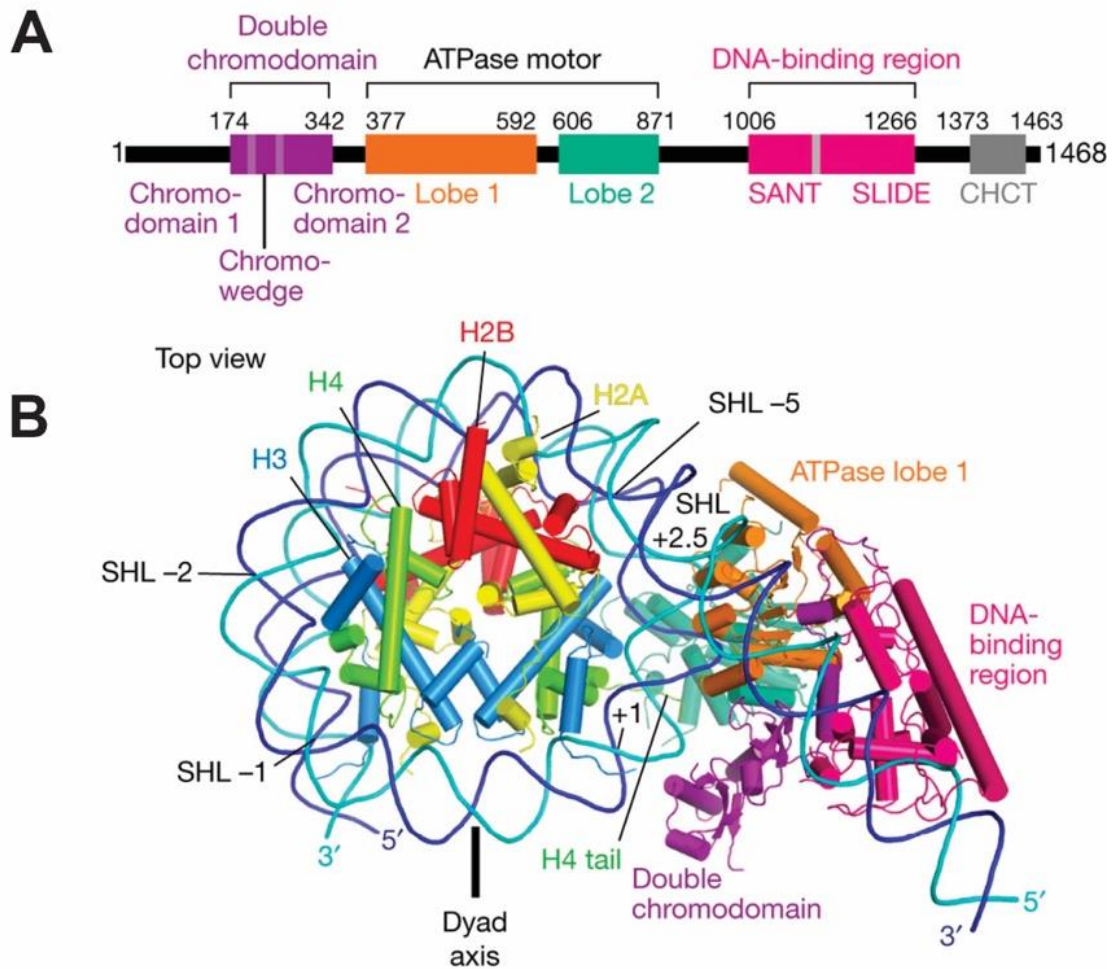
Figure 11. Chd chromatin remodelling complex.

Taken from (Farnung et al., 2017) A) Chd1 domain architecture. B) Structure of nucleosome–Chd1 complex.

1.2.5 SWI/SNF (Switch/Sucrose Non-Fermentable)

The SWI/SNF family was discovered in 1984 by identifying genes that regulate the mating-type switching and sucrose fermentation in yeast (Neigeborn & Carlson, 1984). The early screenings looking for mutations causing defective sucrose fermentation revealed five different Sucrose Non-Fermenting 2-6 (Snf2, Snf3, Snf4, Snf4, Snf5, and Snf6). After extensive proteomic, biochemical, genetic, and structural studies, 29 genes have been identified to codify the subunits of the three different types of human SWI/SNF complexes. The following are the different types of complexes: canonical BAF (cBAF), polybromo-associated BAF (pBAF), and the newest sort identified the non-canonical BAF (ncBAF) complex (Figure12).

The catalytic subunit of these complexes can remodel chromatin via nucleosome eviction, nucleosome exchange and nucleosome sliding with the energy obtained by ATP hydrolysis (Clapier et al., 2017). There are two mutually exclusive ATPase subunits, BRG1 and BRM. They are specific subunits that determine the particular type of complex that may form; for instance, it is well known that the cBAF can have ARID1A or ARID1B forming part of an AT-rich interaction domain, whereas pBAF uses ARID2. Regarding the subunits that contain a PHD domain, cBAF incorporates the DPF2 subunit while pBAF includes PHF10. After proteomic analyses of the synovial sarcoma protein were performed, SS18 protein was identified as a specific member of the cBAF complex, as it was found to precipitate exclusively with ARID1A containing complexes.

Interestingly, the ncBAF does not integrate any ARID or PHD subunits or other subunits being identified before as core subunits like SMARCB1; instead, it

contains BRD9 and GLTSCR1/2. GLTSCR1 replaces ARID subunits, whereas BRD9 is the subunit targeting this non-canonical complex to specific gene locations.
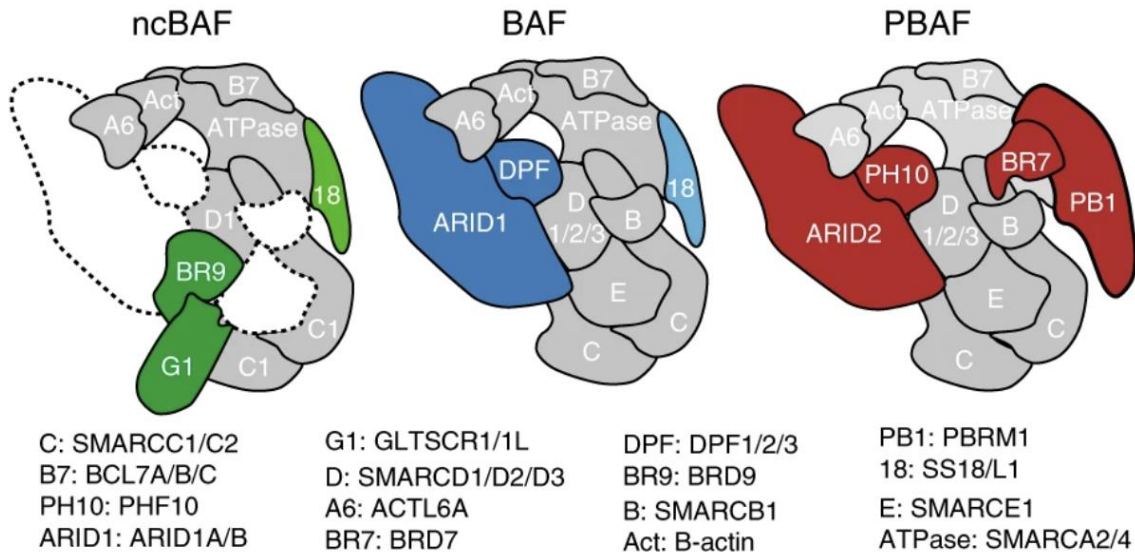
Figure 12. Three types of mammalian SWI/SNF complexes

Modified from (Michel et al., 2018). Schematic depicting biochemical subunit compositions of mammalian ncBAF, cBAF, and PBAF complexes. The blue, red, and green subunits represent BAF-, PBAF- and BRD9/GLTSCR1-specific complexes, respectively. Shared subunits are shown in grey.

1.3 CHAPTER III SWI/SNF COMPLEX

1.3.1 Evolution of SWI/SNF complex

The SWI/SNF complex has been changing for 1,5 billion years extending in composition and widening functions. As mentioned before, the SWI/SNF complex was described for the first time in yeast. In *Saccharomyces cerevisiae, there* are two subfamilies: RSC and SWI/SNF complexes. Both are multiprotein complexes composed of 12-17 subunits with a molecular weight of about 2 MDa. Mutations in these complexes cause defects in sporulation and poor growth (Sudarsanam et al., 2000; C. Wang et al., 2020). In the context of transcription, there is the main difference that distinguishes the two types of complexes; for instance, RSC is responsible for maintaining nucleosome-free regions (NFR) while SWI/SNF plays a role in remodeling nucleosomes during transcription initiation (Klein-Brill et al., 2019; Krietenstein et al., 2016; Nagai et al., 2017; Patel et al., 2019)

The core subunits of the SWI/SNF complex are Swi1, Swi3, Snf5, and Snf6, all of them are necessary for the assembling of a high molecular weight complex in association with Snf2 (the catalytic subunit), the rest of the subunits are Swp29, Swp73, Swp82 and two actin-regulated genes Arp7 and Arp9. The structure has five main lobes: the rigid core, the nucleosome binding module, the pre-HSA stabilization module, the coiled-coil domain, and the ARP module (Figure 13) (C. Wang et al., 2020). To this day, two structures are known, a 4.7 Å structure bound to the nucleosome and a 2.9 Å free state complex (C. Wang et al., 2020) (Wagner et al., 2020).
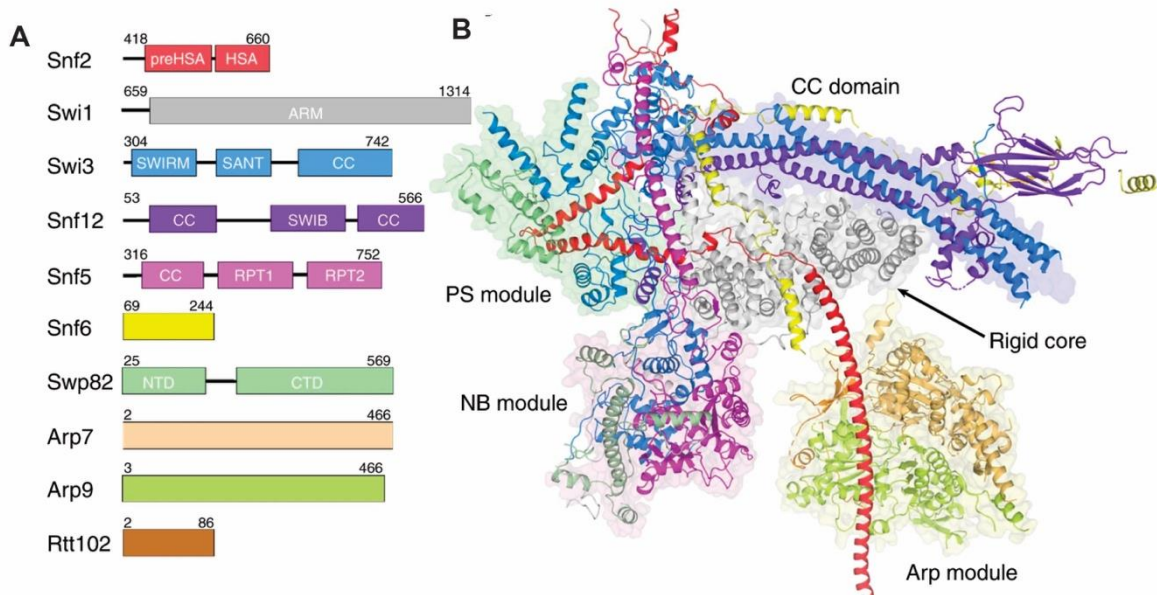
Figure 13. Yeast SWI/SNF complex.

Modified from (C. Wang et al., 2020) A) Domain organization of all subunits included in the final model. B) Cartoon model of the SWI/SNF complex with the individual subunits colored. The rigid core, PS module, NB module, CC domain, and Arp module are colored in gray, green, pink, and orange, respectively.

The second subfamily, RSC, comprises a tripartite core and two flexible lobes(Patel et al., 2019). The core structure is supported by an asymmetric Rsc8 dimer and formed with Sfh1, Rsc6, Rsc9, and Sth1. The ATPase lobe, composed of helicase subunit Sth1, Arp7, Arp9, and Rtt102, is anchored to this core by the N-terminus of Sth1 (Patel et al., 2019).The core forms around 70% of the RSC complex, and the core consists of the subunits Rsc8, Rsc6, Rsc9, Rsc58, Sth1, and Sfh1 (Figure 14). Apart from Rsc58, all the other subunits are evolutionarily conserved in eukaryotes. After some years, a new complex was identified during the screening to detect genes opposing Polycomb mediated repression of homeotic genes in *Drosophila*. The new SWI/SNF or BAP (BRM-associated protein) complex has a homologous Snf2 catalytic subunit named Brm; however, the main target of this complex is the Polycomb complex instead of histones and nucleosomes as it is for the yeast complex (Kadoch & Crabtree, 2015). In Drosophila, two subfamilies are known BAP and pBAP, both containing BRM (Swi2), two copies of BAP155 (also called MOR, the orthologue of Swi3), BAP47 (Arp7), BAP55 (Arp9), BAP60 (Swp73), BAP45 (Snf5) and the new subunits, BAP111 and ACTB. BAP and pBAP differ by their alternative incorporation of either OSA (orthologue to Swi1) alone or BAP180 (Rsc1) together with BAP170 (Rsc9) and SAYP (PHF10), respectively.

A clear transition occurs with the appearance of multicellularity, some subunits are lost from unicellular to multicellular organisms, and new subunits have appeared.
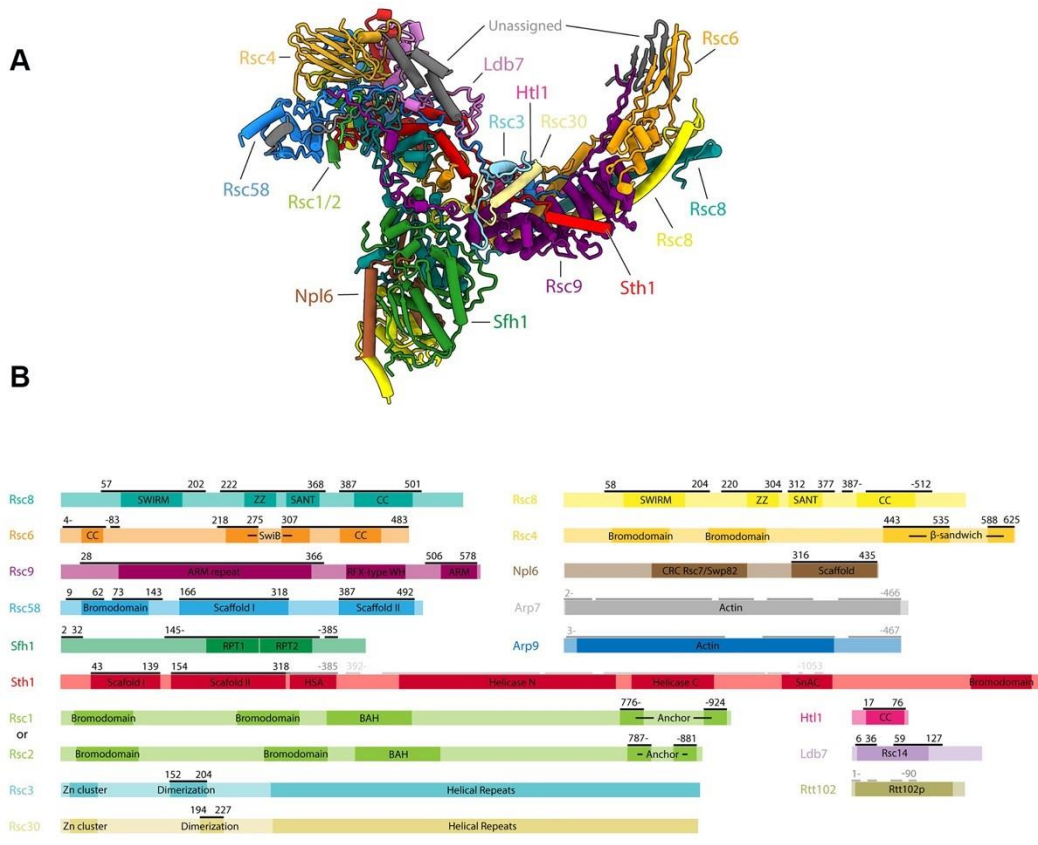
Figure 14. The RSC complex.

Modified from (Patel et al., 2019). A) Model of the RSC core with individual subunits colored and labeled. B) Domain architecture for RSC subunits.

In multicellular organism, during evolution, chromatin underwent several changes, including the appearance of histone H1 (Kadoch & Crabtree, 2015) (Figure 15). A new level of complexity occurs with the appearance of vertebrates and DNA methylation, and new subunits, unique to metazoans, have been discovered to be part of complex. The complexes take on highly specialized functions in the vertebrates thanks to the polymorphic and combinatorial ability to get assembled. Finally, in the late evolution of vertebrates, four neuron-specific subunits appeared. Several subunits have been conserved through millions of years (Figure 15), for instance, the helicase domain of the ATPase subunit, suggesting that fundamental mechanisms are conserved in this group of chromatin remodelers.
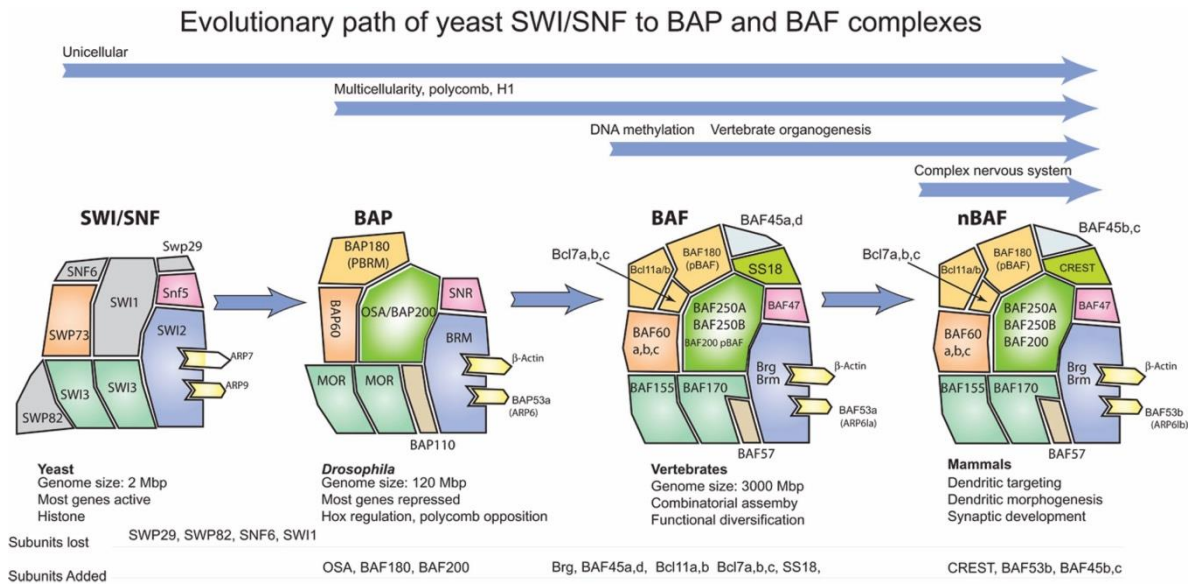
Figure 15. Evolution of SWI/SNF complex.

Taken from (Kadoch & Crabtree, 2015). Evolution of the yeast SWI/SNF complexes to the fly BAP and vertebrate BAF complexes. The figure depicts the subunit structure of these related complexes over the last 500 million years of evolutionary history. Colors are used to indicate homology.

1.3.2 Module organization and assembly of mammalian SWI/SNF

Before the mammalian cryo-EM structure became available (Mashtalir et al., 2020a), the modular organization of the complex was reported (Mashtalir et al., 2018). SWI/SNF complexes were extracted from embryonic kidney cells (HEK293T) by tagging different subunits such as DPF2 and SS18, and purified via glycerol gradients. Intra and inter molecular interactions of the subunits were identified by perfoming combined cross-linkning and mass spectrometry studies (Mashtalir et al., 2018) (Figure 16-17). According to this model SWI/SNF assembles in a combinatorial modular stepwise manner.

In the first module, the ATPase is composed of either SMARCA4 or SMARCA2 and the actin-related proteins ACTB and ACTL6A/B and one of the paralog subunits of DPF2. The second module essential for assembling the entire complex is composed of two copies of SMARCC1 or SMARCC2, one SMARCD paralog, SMARCE1, and one ARID1 or ARID2. The last module was formed with SMARCB1 and one member of the BCL7 family. The same experiments were made in fly and yeast, showing that the modularity of the core module and the ATPase module is highly conserved.

The assembly pathways for the mammalian sub-complexes were characterized by deleting each SWI/SNF component using CRISPR-Cas9 and deleting paralogs of the corresponding subunits to avoid redundancy. The study (Mashtalir et al., 2018) reports that serial immunoprecipitation and density sedimentation followed by mass spectrometry were performed. This study allowed the design of a combinatorial assembly of SWI/SNF (Figure 16-17) (Mashtalir et al., 2018).

The first step of the complex assembly consists of forming homo and heterodimers of SMARCC1 and SMARCC2. Next a trimer is formed with the addition of SMARCD, and the core of the complex is formed. Depending on which subunits interacts with this trimer the canonical BAF, PBAF, or ncBAF form. When SMARCE1 and SMARCB1 are bound, they assembly the core subcomplex that can lead to the formation of cBAF and PBAF complexes

Further interactions with ARID subunits will define the formation of cBAF and PBAF; for instance, ARID1, a member of cBAF, interacts via the C-terminal domain with SMARCC and SMARCD and then another specific subunit DPF2 binds. On the other hand, ARID2 binds and allows BRD7 and PH10 association.

The last step of assembly corresponds to the association of the ATPase subunit and proteins associated with it. The two ATPases, SMARCA4 and SMARCA2, crosslink with BCL7A, BCL7B, BCL7C, SS18, and the actin-related subunits like ACTL6A; ACTL6A has been found to interact at the same time with ARID1.
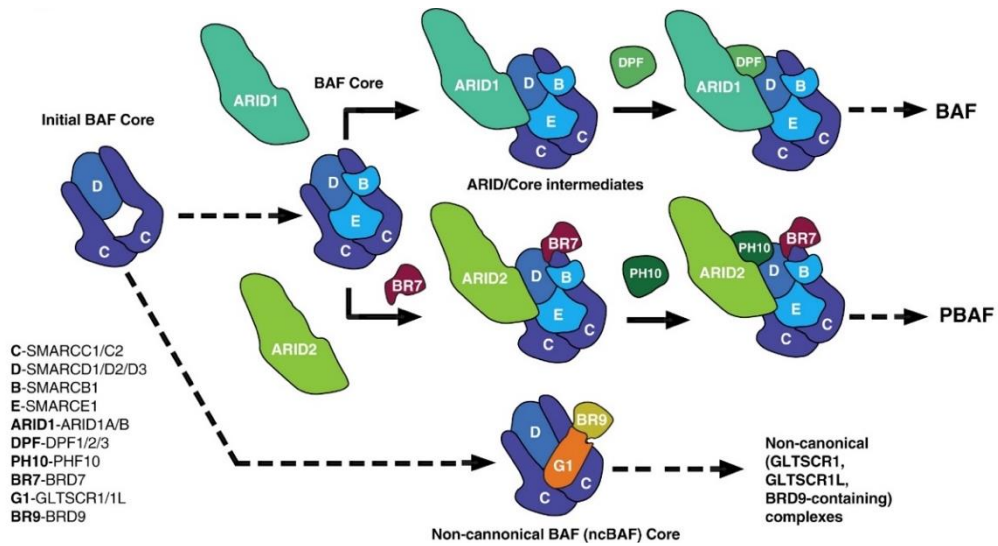
Figure 16. Mammalian SWI/SNF assembly.

Taken and modified (Mashtalir et al., 2018). SWI/SNF assembly branch points are initiated by ARID subunits. Subunit abbreviations are indicated.
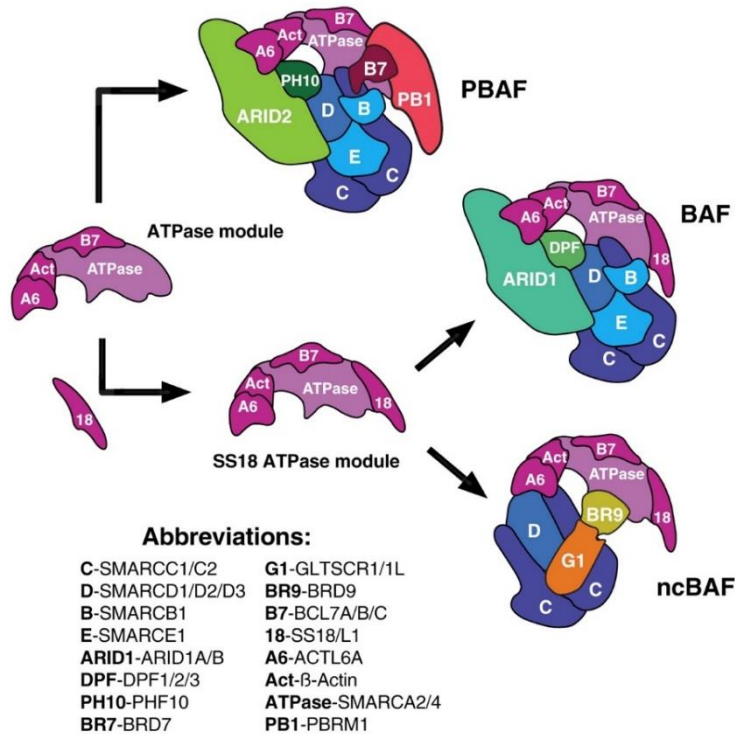
Figure 17. Schematic of the assembly and incorporation of the BAF ATPase module.

Taken and modified from (Mashtalir et al., 2018)

1.3.3 SWI/SNF BAF complex structure

Recently, studies on the structure of SWI/SNF were published. Two studies concerned the yeast SWI/SNF complex (Han et al., 2020; C. Wang et al., 2020) (novel subunits do not exist). One study used a reconstituted mammalian SWI/SNF (novel subunits were omitted) (He et al., 2020); and the fourth study, realized with the endogenous human complex, had overall poor resolution (about 8.5 Å, novel subunits were simply invisible) (Mashtalir et al., 2020). The cryo-EM structure at 3.6 Å (He et al., 2020) is the first high resolution structure of the mammalian SWI/SNF complex, obtained by reconstitution of individually purified subunits. However, some subunits were omitted from the complex and the SWI/SNF reconstitution is not complete. As seen in RSC, the complex is organized into three modules: the ATPase module, the actin-related (ARP), and the base module. (Figure 18). At the overall resolution of 3.6 Å, the group could describe how the ATPase and the base modules "sandwich" the nucleosome in a different manner than previously observed in other chromatin remodelers (He et al., 2020), supporting the idea that the complex can evict nucleosomes.

On the other hand, the group of Walz and Kadoch published the cryo-EM structure of canonical BAF (Mashtalir et al., 2020) at 7.8 Å. After local refinements, they could reach in some regions around 5 Å. However, they could purify the complex from HEK293F cells isolating the complex by precipitating the tagged DPF2 subunit. The complex was bound to the nucleosome and stabilized using the GraFix method. Modeling and mass spectrometry experiments helped to assign density to different subunits (Figure 19). The differences between the recombinant and the

endogenous complexes depend on the subunit composition. A major difference between the two structures is seen in the extra density near the nucleosome acidic patch, which corresponds to the SnAc and post-SnAc domains of SMARCA4, remarking a major difference between humans and yeast SWI/SNF complexes, in yeast the ATPase have more regular contacts with the nucleosome DNA. A second difference is a gap observed between the ATPase and ARP modules formed by the accessory subunits such as BCL7A. In this study, they report for the first time and, thanks to the crosslinking data, that BCL7 proteins are very close to the nucleosome. The density for BCL7 proteins is not visible. For SMARCB1, another difference was observed: the localization of the winged-helix domain in the endogenous complex is close to the exit site of the nucleosomal DNA while in the recombinant complex is far away, the difference can be due to the absence of the rest of the subunits or the multiple conformations that the complex can adopt.

Beside the differences that both structures have, the SWI/SNF complex is composed of three main modules, ATPase, ARP, and base module, and the difficulty of having a high-resolution structure is due to the variability in the composition of the complex and the dynamic and flexibility of each of the components.

Figure 18. Structure of the reconstituted mammalian BAF SWI/SNF complex bound to a nucleosome.

A) Ribbon representation of the 3.7 Å BAF complex organization. Color scheme for BAF subunits is indicated at the bottom (He et al., 2020a).
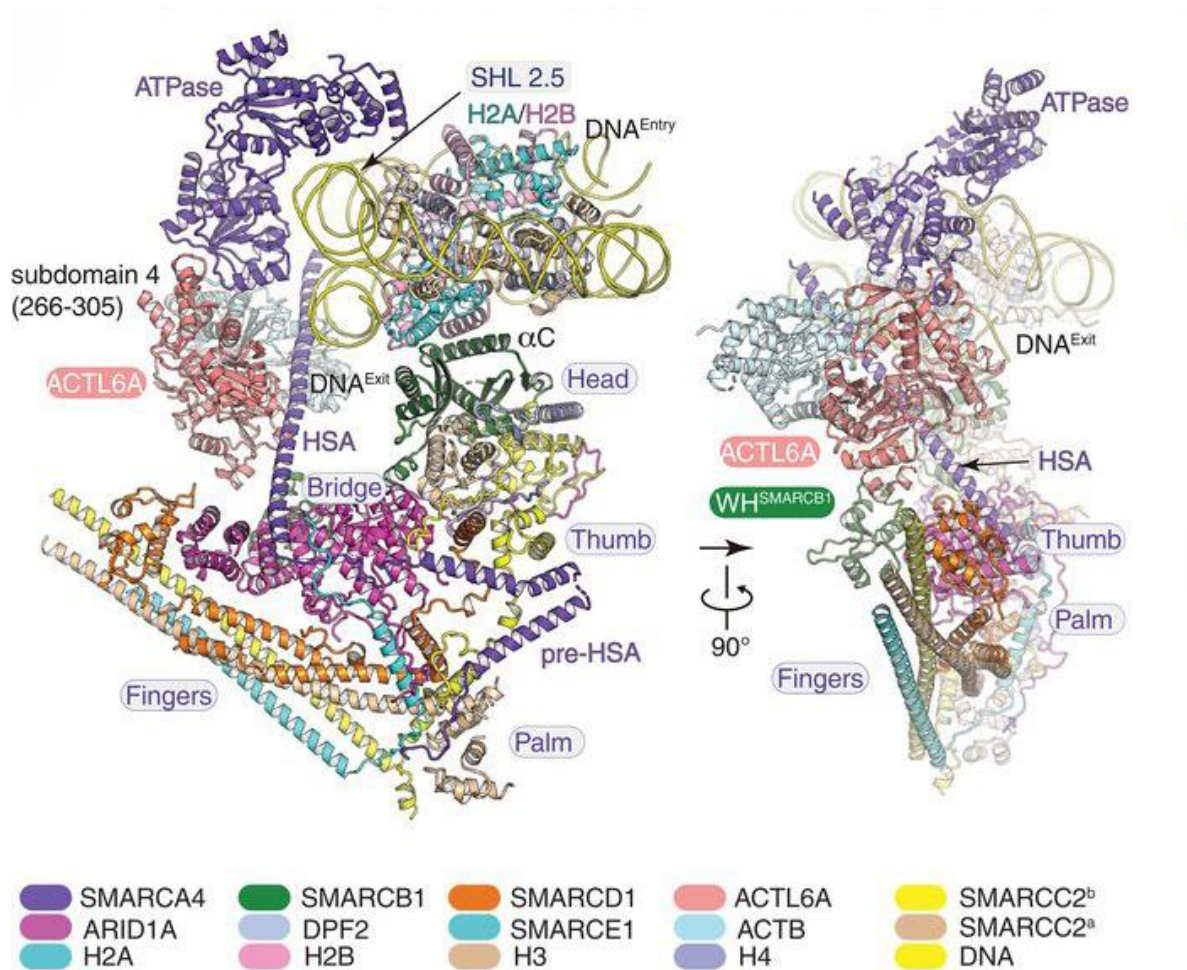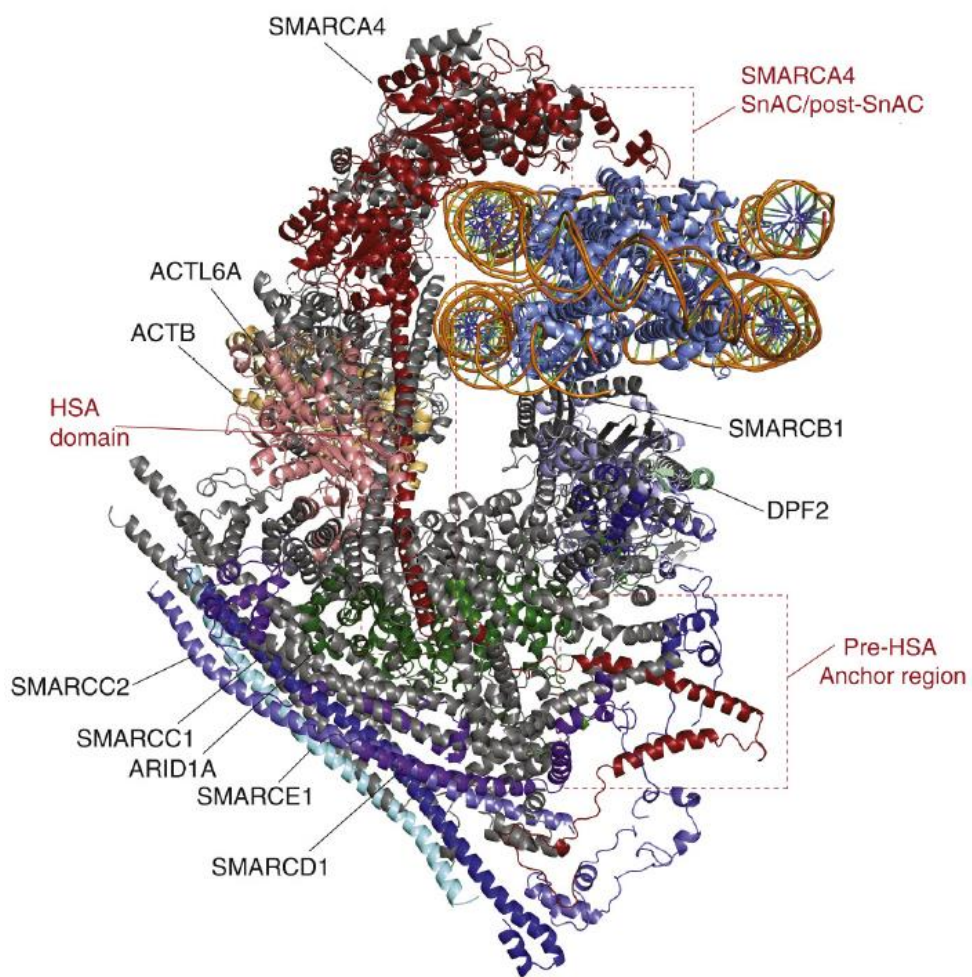
Figure 19. Cryo-EM structure of endogenous human BAF SWI/SNF complex bound to a nucleosome at 7.8 Å (Mashtalir et al., 2020).

1.3.3.1 The ATPase motor

In the structure of the reconstituted complex (He et al., 2020), In absence of ATP, the ATPase module was not solved at high resolution, but they could see that the complex in a pre-engagement state or open conformation. Whereas, when the structure was obtained in the presence of ADP, the ATPases closes from 90 to 70 degrees and contacts the nucleosome at position SHL2, a similar feature shared with the yeast complex, the ATPase motor engages the nucleosomal DNA and uses the energy from ATP hydrolysis generates the translocation of the DNA (He et al., 2020). This module is composed of the catalytic helicase domain of SMARCA4 that interacts with the "top" nucleosome face and with the nucleosomal DNA as superhelical location 2 (SHL2).

1.3.3.1.1 SMARCA4/BRG1

The human SWI/ SNF can have BRG1 or Brahma as a catalytic subunit. These paralogous proteins share around 74% of sequence identity (Trotter & Archer, 2008). The BRG1 homolog Swi2 and the Snf2 protein were identified in yeast when Winston and Carlson searched for proteins involved in the mating-type switching and sucrose non-fermenting pathways (Winston & Carlson, 1992). BRG1 has multiple domains (Figure 20): the ATPase domain, highly conserved through evolution, the C-terminal bromodomain, the AT-hook motif, and towards the N-terminal two domains: the glutamine-leucine-glutamine (QLQ) and the HAS domains.

The C-terminal domain is involved in acetylated lysine recognition, it consists of a left-handed four-helix bundle, and it is well known to recognize N-acetyl lysines

within the tails of H3 and H4, more specifically Lys 9 and Lys 14 of histone three and Lys8 of H4 (Agalioti et al., 2002; M. Singh et al., 2006). The AT-hook motif helps in DNA binding. On the other hand, the N-terminal domain has the QLQ motif corresponding to amino acids 172-208 that could be involved in protein-protein interactions, but the function is not completely clear.

In vitro studies showed that BRG1 on its own can perform nucleosome remodeling, and in the presence of the subunits SMARCC1, SMARCC2, and BAF47, the in vitro remodeling activity is reestablished (Phelan et al., 1999). Tumor suppressor activity has been described when the protein interacts with BRCA1 and p53 (Trotter & Archer, 2008).
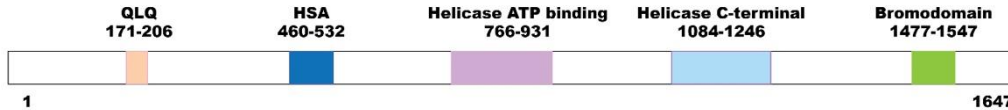
Figure 20. Domain architecture of BRG1.

1.3.3.2 The actin-related module (ARP)

The ARP module is formed by a heterodimer of ACTL6A and ACTB subunits and the α-helix of the HSA domain of SMARCA4 (He et al., 2020a) (Figure 18, Figure 21). This module does not interact with the nucleosome but functions as a bridge and couples the motions between the ATPase and base modules during the chromatin remodeling activity.

1.3.3.2.1 Actin-like 6A (ACTL6A)/ BAF53A

This subunit is crucial for maintaining stem cells during mammalian embryonic development (Bao et al., 2013; Krasteva et al., 2012, p. 53; Lu et al., 2015, p. 6; Y. Xiao et al., 2021). It is found in other chromatin remodelers like INO80 (Brahma et al., 2018) and the NuA4/TIP60 acetyltransferase complex (Lu et al., 2015; Y. Xiao et al., 2021). It is part of the evolutionary conserved actin-related proteins (Arps) classified into two main groups: cytoskeleton or chromatin-associated. ACTL6A, as part of the chromatin-associated group, regulates the structure and function of chromatin. This subunit contains an actin fold domain that allows binding and hydrolysis of ATP (Dion et al., 2010).

ACTL6 has an oncogenic role in tumor progression, promoting epithelial-mesenchymal transition and metastasis or invasion in colon cancer, osteosarcoma, hepatocellular carcinoma, and glioma cells (Meng et al., 2017, p. 53; W. Sun et al., 2017; S. Xiao et al., 2016; Y. Xiao et al., 2021; Zeng et al., 2018).

1.3.3.2.2 Actin cytoplasmic 1, or β-actin (ACTB)

ACTB forms a heterodimer with ACTL6A PBAF and binds to the α helix of the HSA of SMARCA4, forming the ARP module (Han et al., 2020; Mashtalir et al., 2020). This subunit is one of the most abundant proteins in eukaryotes, therefore is found being part of different complexes, and it is related to multiple diseases like developmental malformations–deafness–dystonia syndrome, as well as multiple tumors, such as melanoma, blastoma, and hematologic, lung, urinary, uterine, bladder, kidney, and colorectal cancers (El Hadidy & Uversky, 2019). Human *ACTB* genes are down-regulated in prostate and urinary bladder cancers and up-regulated in breast and kidney cancers (El Hadidy & Uversky, 2019).
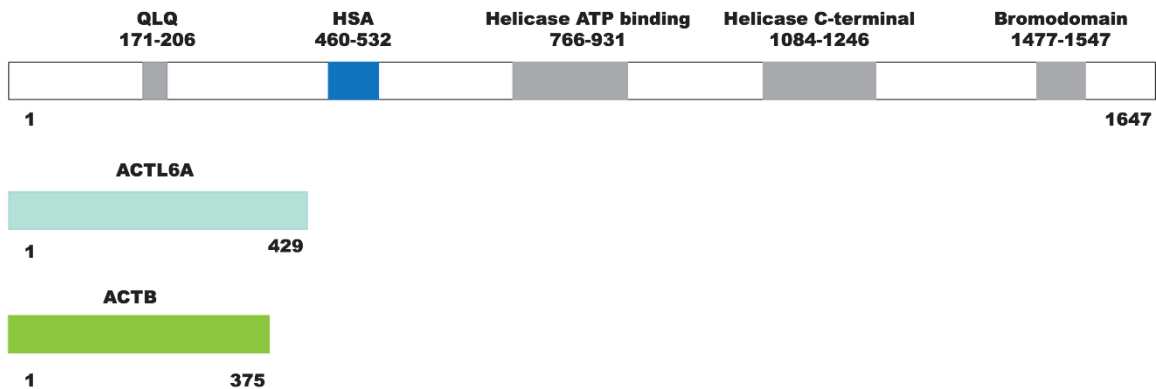
Figure 21. Domain architecture of proteins that form part of the ARP module.

1.3.3.3 Base module organization

The base module accounts for about 80% of the of the entire remodeler complex. Figure 22 provides a summary of the different subunits that compose the base module and the domain architecture is represented for each subunit. It is striking to realize the wide variety of functional domains that the base module contains. Interestingly, the catalytic subunit SMARCA4 is part of the three modules, thus part of it is also located in the base module. The pre-HSA domain of SMARCA4 is anchored to the base, and seven other additional subunits form the base of SWI/SNF: BAF250A, DPF2 (these two exclusive subunits of canonical BAF), SMARCB1, SMARCD1, SMARCE1, and two copies of SMARCC. The base module comprises five submodules: head, thumb, palm, bridge, and fingers.

The head submodule interacts directly with the nucleosome, it is formed by RPT1 and RPT2 domains and C-terminal α helix of SMARCB1, and as mentioned before, this region of SMARCB1 has several positively charged amino acids that bind the acidic patch. It is crucial for the entire remodeling activity of the complex (Valencia, Collings, Dao, St Pierre, et al., 2019). This submodule is formed with the Req domain of DPF2, two SWIRM domains of SMARCC and an insert from ARID1A. Each SWIRM domain binds RPT1 and RPT2, respectively, forming two RPT-SWIRM structures that fold similarly and bind each other; the Req domain of DPF2 and the insert of ARID1A reinforce RPT-SWIRM interactions (He et al., 2020a).

The Head, the Bridge, and the Thumb interact with each other mediated by DPF2, ARID1A, and SMARCCsRPT1 of SMARCB1 is close to the second helix of H2A; this proximity disrupts the octamer and pushes the DNA to the exit site of the

81

nucleosome. On the other hand, the N-terminal winged-helix domain of SMARCB1 can interact with the ARM domain of ARID1A located far away from the nucleosomal DNA but close to the ARP module. This fact suggests a different function for this domain and demonstrates that the presence of this protein is crucial for chromatin remodeling activity.

The largest subunit of the complex is ARID1A, a very flexible protein that is not entirely visible in the structure. However, they could determine that the ARM domain, which consists of seven ARM repeats arranged in superhelical conformation, acts as a "platform" where SMARCA4 and the other subunits of the module can lean. The ARM domain has a zinc finger that helps stabilize and associate ARID1A to the center of the base. One end of ARM binds to the Head and Thumb while the other side interacts with the SWI domain of SMARCD1. The residues of ARM that interact with SMARCA4 and SMARCD1 are highly conserved through evolution, and importantly this is the subunit most frequently found mutated in cancer from the SWI/SNF complex. ARID1A and ARID1B are mutually exclusive subunits for canonical BAF; however, how different is ARID1A from ARID2? They have different domain architecture; however, the N-terminal region is predicted to have seven repeat ARM domains, as seen for the C-terminal domain of ARID1A.

1.3.3.3.1 SMARCB1 (BAF47)

SMARCB1 C-terminal domain (CTD) contains a basic α helix that makes direct contact with the nucleosome acidic patch, and this interaction is crucial for the remodeling activity of the entire SWI/SNF complex (Valencia, Collings, Dao, St

Pierre, et al., 2019). Using techniques like NMR, molecular docking, and photo-crosslinking, frequent mutations found in patients with Coffin-Siris syndrome (a syndrome that causes intellectual disability), localized in the Coiled-coil CTD of SMARCB1 were studied, interestingly single point mutations of these positively charged amino acids disrupt the binding and remodeling activity of the SWI/SNF complex with the nucleosome. BAF47 has an N-terminal winged-helix (WH) domain that interacts with double-stranded DNA and two repeat domains (RPT1 and RPT2). RPT1 interacts with the SWIRM domain of BAF155 (Yan et al., 2017, p. 155). In malignant rhabdoid tumors and epithelioid sarcomas, the loss of SMARCB1 is a genetic driver that leads to tumorigenesis, but very interestingly, the loss is rarely observed, and biallelic loss is only tolerated in specific circumstances (Chun et al., 2016; Versteege et al., 1998).

1.3.3.3.2 SMARCC1/BAF155

This subunit has a very important role, is responsible for maintaining the wholes BAF-complex stoichiometry (Keppler & Archer, 2010). BAF155 stimulates the activity of the ATPase subunit (Phelan et al., 1999). This subunit bound to the BRG1/BRM, BAF47, and BAF170 can form a smaller complex with catalytic activity, demonstrating that these subunits form part of the conserved core/SWI/SNF complex. SMARCC1 has a tumor suppressor activity in colorectal cancer and ovarian carcinoma (DelBove et al., 2011). In prostate cancer, the expression of this gene is completely lost. The human BAF155 has two functional domains, a SANT complex that binds to histone tails and a SWIRM domain requires for the assembly

of the entire complex. The SWIRM domain interacts with the RPT1 of BAF47 (Yan et al., 2017). Besides these domains, the protein has several regions: coiled-coil, poly-Pro, Glu-rich, poly-Ala, and Pro-rich (El Hadidy & Uversky, 2019).

SMARCC2/BAF170

The human BAF170 has a similar domain organization as BAF155 with SWIRM and SANT domains and the same regions with amino acid composition biases and the coiled-coil (El Hadidy & Uversky, 2019). Different from BAF155, this protein has three different isoforms in addition to the canonical isoform. As mentioned before, the RPT1 of BAF47 binds to the SWIRM domain of BAF155, and in BAF170, this is not an exception; the interaction between RPT2 and SMARCC2 has been reported, suspecting that RPT2 folds upon binding to SMARCC2, assembling into a heterotrimer (Chen et al., 2020). The mutations in this gene are associated with the same type of cancer as BAF155, lung, stomach, liver, breast, and uterine. It is part of the conserved core subunits.

1.3.3.3.3 ARID1/BAF250

ARID1 encodes a 250kDa subunit that is evolutionarily conserved, sharing sequence similarity with the yeast Swi1 and the Drosophila gene Osa. This subunit exists in two forms, BAF250A/ARID1A and BAF250B/ARID1B(El Hadidy & Uversky, 2019; X. S. Li et al., 2010). This regulatory subunit enrolls the complex to chromatin either through protein-DNA or protein-protein interactions, allowing the activation of several genes(El Hadidy & Uversky, 2019; Nie et al., 2000). ARID1A is mutated in ovarian

clear-cell carcinoma (Jones et al., 2010) and endometrioid ovarian carcinoma, other mutations in this gene have been linked to Burkitt lymphoma and some subtypes of endometrioid carcinoma. ARID1A is the most frequent mutated gene of the SWI/SNF complex through different types of cancers (Mathur, 2018). The role of ARID1A and ARID1B in tumor progression will depend on the nature of the disease (El Hadidy & Uversky, 2019; Mathur, 2018). These subunits have AT-rich interactions domains (ARID) and several LXXL motifs, including poly-Serine and glutamine-rich regions and a very important C-terminal armadillo (ARM) domain that arranges scaffolds for the core module of the complex.

1.3.3.3.4 SMARCE1/BAF57

This 411 residue-long protein-protein interacts with androgen and estrogen receptors and regulates the nuclear receptor function (Lomelí & Castillo-Robles, 2016), overall BAF57 interacts with a broad number of proteins outside the SWI/SNF complex. This subunit has no identifiable homolog in yeast, and it is specific for higher eukaryotes (Lomelí & Castillo-Robles, 2016). Mutations within this gene are found in meningiomas and sporadic tumors (Savas & Skardasi, 2018). This gene is involved in other cancers, like breast and ovarian, and is a prognosis marker in prostate cancer (El Hadidy & Uversky, 2019; Kerr et al., 2018). BAF57 contains a high-mobility group (HMG) DNA binding domain that belongs to the sequence-nonspecific HMG family (Lomelí & Castillo-Robles, 2016) adjacent to a kinesin-like region.

1.3.3.3.5 PBRM1/BAF180

This protein was identified during a screening looking for tumour supressor genes in breast cancer and it is mutated in around 40% of patients with renal cell carcinoma (Varela et al., 2011). The protein has six bromodomains and two Bromo adjacent domains. Structures of five of the six bromodomains are available, and flexible regions have been identified between them.

1.3.3.3.6 AT-rich interactive domain-containing protein 2 ARID2/BAF200

This protein is part of the pBAF complex, and it is involved in maintaining the stability of the complex; inactivating mutations in this protein have been reported in a variety of human cancers like hepatocellular carcinoma, non-small cell lung carcinoma, and melanomas(M. Li et al., 2011). The protein has an ARID (N-terminal AT-rich DNA interaction domain) domain, two conservative zinc finger motifs, a RFX-type winged helix, proline and glutamine-rich regions. ARID2 has tumor suppressor activity when the AT-rich domain of the protein is mutated (H. Zhao et al., 2011).

1.3.3.3.7 DPF2/BAF45D

This protein binds modified Histones 3 and 4. Mutations in the gene are linked with melanoma, blastoma, and different type of cancers as liver, lung, and uterine.

The protein is a reader subunit containing a requiem (Req) and a PHD domain, and it is can only associate with the fully assembled core module of the SWI/SNF complex. It interacts with RPT2 of BAF47 and interfaces with other subunits as SMARCC1 and ARID1A/B (Mashtalir et al., 2020).
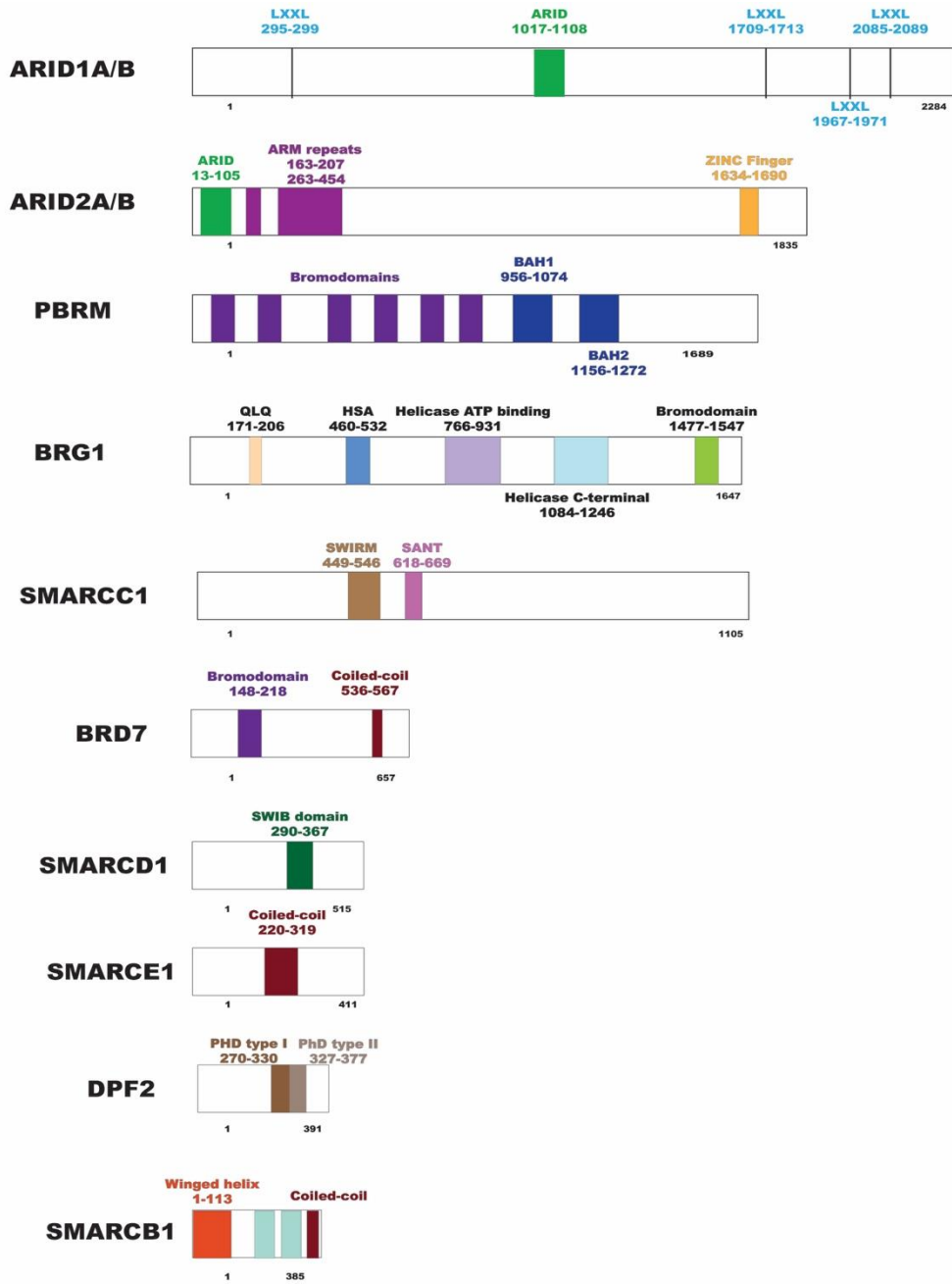
Figure 22. Domain architecture of proteins that form part of the BASE module.

(Only one ortholog is shown in the picture.)

1.3.3.4 Novel subunits of the SWI/SNF complex

Identifying new subunits of the mammalian SWI/SNF came out after extensive proteomic analysis in parallel with biochemical studies of endogenous complexes coming from different cell types, neurons, fibroblasts, neural progenitors, neurons, and mice embryonic stem cells (Kadoch et al., 2013). The newly identified subunits include BCL7A, BCL7B, BCL7C, BCL11A, BCL11B, SS18, and BRD9. The domain architecture for BICRA, BRD9 and SS18 is reported in figure 23. These proteins do not have homologs in the yeast SWI/SNF complex (Kadoch et al., 2013). Different cells types will have different compositions; however, these novel subunits are established as non-exchangeable subunits of the mammalian SWI/SNF complex.

1.3.3.4.1 Synovial sarcoma 18 (SS18)

Synovial sarcoma (SS) is a malignant soft tissue lesion that most of the time affects young adults (Ishii et al., 2018). Around 95% of SS has the SS18 gene located on chromosome 18 translocated to either SSX1, SSX2, or SSX4 gene located on chromosome X. SS18 codifies a 46 kDa protein that interacts with the disordered N-terminal domain of the ATPase subunit. The fusion SS18-SSX oncoprotein binds the nucleosome acidic patch with higher affinity than BAF47, resulting in the disruption of the entire BAF core module, observed in SS (Kadoch & Crabtree, 2013; Mashtalir et al., 2020).

SS18L1: Paralog of SS18 are mutually exclusive in cBAF and ncBAF complexes. It is thought that these subunits can regulate the final step of the complex assembly.

1.3.3.4.2 BICRA (BRD4-interacting chromatin remodeling complex associated) / GLTSCR1 (Glioma tumor suppressor candidate region gene 1)

This protein was recently named BICRA, together with the mutually exclusive paralog BICRAL are part of subunits of the ncBAF complex. The information about these proteins is limited; however, it is known that in the complex, they are incorporated instead of the ARID proteins. This protein's evolutionarily conserved domain, GLTSCR, is needed to interact with the ncSWI/SNF complex.

BICRAL: This is the paralog of BICRA; they share 32% of sequence homology, and they are mutually exclusive, and it can alter the overall SWI/SNF stoichiometry (Barish et al., 2020).

1.3.3.4.3 Bromodomain-containing protein 9 BRD9

This protein is a member of the bromodomain family IV, and together with BICRA, these subunits are specific for the newest described ncBAF complex (Mashtalir et al., 2018). In addition to the bromodomain (BRD), which recognizes acetylated 18lysine residues of histones H3 and H4, for example, diacetylated H4K5acK8ac and dipropionylated H4K5prK8pr (Flynn et al., 2015), this subunit has a DUF3512 domain, essential for assembling the ncBAF complex(X. Wang et al., 2019, p. 9). In cancer, BRD9 plays an oncogenic role in multiple cancer types by regulating tumor cell growth (X. Zhu et al., 2020). It has been linked to acute myeloid leukemia (AML), clear cell renal cell carcinoma (ccRCC) and breast cancer.

Bromodomain inhibitors act as antitumor agents(X. Zhu et al., 2020), and the conserved BRD fold contains a large and deep hydrophobic acetyl lysine binding site, this pocket represents a target for development of small molecule inhibition and

pharmaceutical active molecules (Muller et al., 2011). Therefore, BRD9 features as a target for small molecule inhibition. In cancer therapeutics, two different types of bromodomain inhibitors have been developed, the non-acetylated and acetylated lysine mimetics. In another study the first selective cellular chemical probe for non-BET bromodomain BRD9 was discovered after extensive studies in structure-based design(Theodoulou et al., 2016); the discovery of this molecule provides the ability to select specifically BRD9 and gives new insights for the selection of human bromodomains.
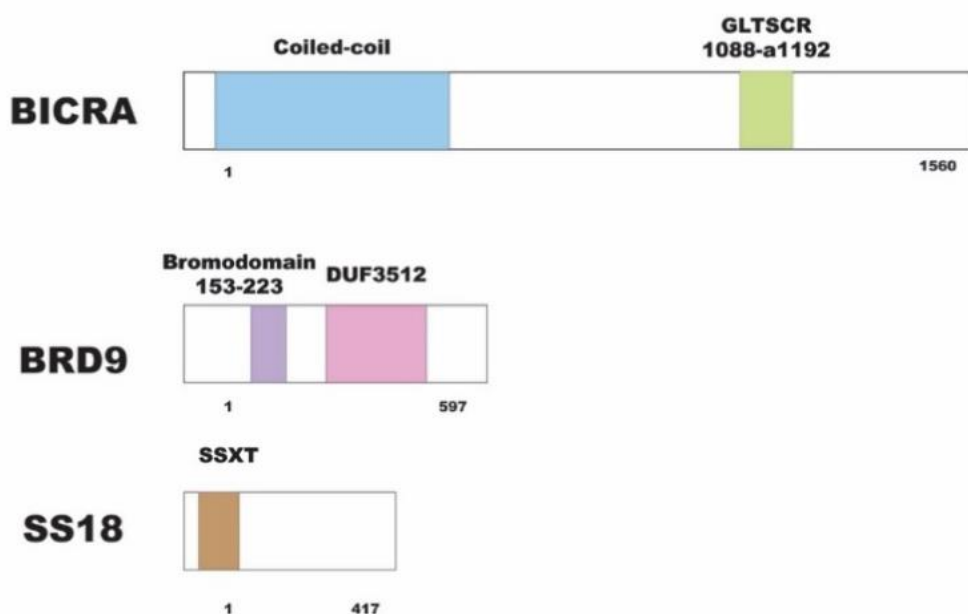
Figure 23. Domain architecture of proteins novel subunits of the SWI-SNF complex.

1.3.4 Structural comparison of SWI/SNF complex

Comparing the structures of the yeast complex with the two human BAF complexes (recombinant and endogenous), it is clear that the base module is highly conserved; however, a major difference is observed between yeast and human complexes nucleosomal DNA-ATPase components interactions. For instance, interactions are looser in the human complexes, but this difference could explain the different functional states captured for the more dynamic human complex (Mashtalir et al., 2020). Secondly, the endogenous human complex has the largest gap between the ARP/ATPase modules and the nucleosome due to the BCL7A subunit. Between the two human complexes, the absence of subunits (as BCL7A) can explain the difference between the localization of the WH domain of BAF47.

In general, all the SWI/SNF family has a characteristic binding to the nucleosomes; they form bilateral acidic-patch interactions. On the top side, the ATPase subunit interacts through the post-SnAc domain, and the bottom face, the C-terminal alpha-helix of BAF47/Snf5, forms a "C clamp" around the nucleosome (Mashtalir et al., 2020)

1.3.5 SWI/SNF complex in human diseases

Mammalian SWI/SNF is interesting from a medical perspective because several of its components is disrupted in human diseases. A recent report showed that almost 20% of human cancers display mutations to mSWI/SNF components (Kadoch et al., 2013a), and several subunits function as bona fide tumour suppressors. Further, deregulation or mutations of mSWI/SNF complex subunits cause neurological

disorders with cognitive dysfunction phenotypes (Ji et al., 2013). Notably, Coffin-Siris syndrome, schizophrenia, and autism spectrum disorders are characterized human diseases caused by defective mSWI/SNF complexes. The most commonly affected SWI/SNF subunits are ARID1A and ARID2. ARID1A mutations are frequent in ovarian cancers, whereas ARID2 mutations are found mostly in melanoma. Mutations in these subunits can lead to loss-of-function, which destabilizes the core module of the SWI/SNF complex (Mashtalir, D'Avino, Michel, Luo, Pan, Otto, Zullow, McKenzie, Kubiak, Pierre, et al., 2018). Another example is SMARCB1 (BAF47): biallelic inactivation of SMARCB1 is found in 95% of malignant rhabdoid tumors (MRT), one of the most aggressive and lethal cancer in early childhood. In synovial sarcoma, a rare type of cancer that found different types of soft tissue, such as muscle or ligaments, a chromosomal translocation produces the fusion of SS18 to SSX (Clark et al., 1994). In the SWI/SNF complex, the fusion protein competes with the wild-type SS18 protein. When the fusion protein incorporates as a subunit of the BAF complex, the larger molecular weight fusion protein evicts SMARCB1(BAF47) from the complex (Kadoch & Crabtree, 2013). SS18-SSX redirects the BAF complexes from enhancers to broad polycomb domains to oppose PRC2-mediated repression and activate bivalent genes; when the fusion protein is suppressed, the reincorporation of BAF47 habilitates the enhancer activation (McBride et al., 2018).

Thank to extensive sequencing studies, many mutations occurring on SWI/SNF subunits are reported. However functional and structural studies are lagging behind and are necessary to translate this knowledge in therapeutic opportunities.

1.4 CHAPTER IV BCL7 FAMILY

1.4.1 B-cell CLL/lymphoma 7 (BCL7) protein family

1.4.1.1 Discovery of the BCL7 family

Chromosomal translocations are a feature of malignancy linked to mental retardation, infertility, and cancer. For years, the molecular characterization of the rearrangement breakpoints and resulting gene products have been of clinical interest, helping the understanding of gene deregulation linked to specific diseases. An example of this is the translocation of the immunoglobulin heavy chain (IGH) locus at chromosome 14q32.33, one consequence of this event is the deregulated expression of oncogenes (Boehm & Rabbitts, 1989; Jadayel et al., 1998). Further studies showed that IGH is involved in a complex three-way translocation common in high-grade B-cell non-Hodgkin lymphoma (HG-BCL). When all the partners of this triple translocation involving IGH, MYC on chromosome 8q24.1 and chromosome 12q24.13 were cloned, a new uncharacterized gene was discovered: B-cell CLL/lymphoma 7 protein family member A (Jadayel et al., 1998; Zani et al., 1996).

The DNA sequence analysis for this new gene exhibited no known protein domains; however, some weak homology was found with caldesmon, an actin-binding protein (Zani et al., 1996). A couple of years later, new sequences released showed similarity to the N-terminal sequence of BCL7A, the two separate transcripts named BCL7B and BCL7C (Jadayel et al., 1998, p. 7). Comparison of the amino acid sequence of the three proteins showed no similarity outside the first 51 amino acids. Interestingly, the conserved region has four potential phosphorylation sites

(Jadayel et al., 1998). Further analysis mapping the position of BCL7B and BCL7C genes located them in chromosome 7q11.23 and 16p11, respectively.

1.4.1.2 BCL7 family in diseases

Extensive genomic studies focused on epigenetic alterations in primary cutaneous T-cell lymphoma (CTCL) demonstrated that patients carrying this disease have highly methylated BCL7A, thus diminished expression of BCL7A represents an unfavorable prognostic sign, suggesting this gene functions as a tumor suppressor in lymphoid cells (van Doorn et al., 2005).

Recurrent point mutations in BCL7A gene have been reported in diffuse large B-cell lymphoma (DLBCL) (Baliñas-Gavira et al., 2020). DLBCL is a common and aggressive type of non-Hodgkin lymphoma that affects B-lymphocytes. It is the most common form of lymphoma, accounting for 30–40% of adult non-Hodgkin lymphoma worldwide. Although most of the patients achieve remission with RCHOP chemotherapy, about 40% of patients relapse and die of the disease (Coiffier et al., 2002). It has been shown that BCL7A undergoes biallelic loss in DLBCL, including a mutational hotspot in the splice donor site of intron one (Baliñas-Gavira et al., 2020). This is a typical feature of tumor suppressor genes. The splice mutations produce a shorter mature mRNA, 81 nucleotides long, that renders a truncated BCL7A protein, lacking 27 amino acids at the N-terminal region of the protein, that cannot bind any longer the SWI/SNF complex. The other reported missense mutation is at the hot spot R11 (Baliñas-Gavira et al., 2020). The fact that the truncated Δ27-BCL7A cannot bind to the SWI/SNF was demonstrated by

immunoprecipitation studies made in HBL1 (cells that lack expression of wild type BCL7A), with transduced Δ27-BCL7A; pull-downs showed no binding between Δ27-BCL7A and the catalytic subunit. When the entire SWI/SNF analysis was carried out with liquid chromatography-tandem mass spectrometry was performed, no complex subunits could be detected. In conclusion, the first 27 amino acids of BCL7A are necessary to bind the SWI/SNF complex. It is clear now that BCL7A has a tumor suppressor activity in DLBCL cells, that is lost by the splice site mutation. Gene ontology enrichment analysis upon wild type BCL7A restoration indicate that BCL7A is involved into B-cell activation processes (Baliñas-Gavira et al., 2020).

BCL7A is expressed at low levels in ovarian cancer tissues and is correlated with survival status. Survival analysis showed that, compared with those who had higher levels of BCL7A expression, patients with ovarian cancer and low levels of BCL7A generally had shorter overall/relapse-free survival times. BCL7A expression could be used as an independent prognostication factor for ovarian cancer patients (Z. Sun et al., 2019).

BCL7C has a role as tumor suppressor in ovarian cancer. BCL7C is downregulated in human ovarian carcinomas, and its underexpression is associated with unfavorable prognosis of ovarian cancer as well as some other types of human cancers. Consistently, depletion of BCL7C reduces apoptosis and promotes cell proliferation and invasion of these cancer cells (Huang et al., 2021). BCL7C suppresses mutant p53-mediated gene transcription by binding to mutant p53, while knockdown of BCL7C enhances the expression of mutant p53 target genes in ovarian cancer cells (Huang et al., 2021).

Glioma is the most common primary brain tumor and represents one of the most aggressive and lethal types of human cancer. In a recent study, the expression level of the BCL7 family and their correlation with prognosis was achieved with bioinformatics and statistical analysis. The expression of BCL7A was low in glioma tissues compared with normal brain tissues (J. Liu et al., 2021). Analysis revealed that patients under chemotherapy that expressed higher levels of BCL7A expression survived longer; therefore, it was determined that BCL7A is a new tumor suppressor gene and can be adopted as a biomarker for independent prognosis in glioma (J. Liu et al., 2021).

BCL7B aberrations have been found in patients affected by the Williams-Beuren syndrome (WBS) (Jadayel et al., 1998; Pérez Jurado et al., 1996). Patients affected by WBS show a variety of phenotypes, including elfin face, mental retardation, reduced spatial reasoning capacity, supravalvular aortic stenosis, and peripheral pulmonic stenosis (Uehara et al., 2015). It has been shown that BCL7B negatively regulates the Wnt-signaling pathway and positively regulates the apoptotic pathway (Uehara et al., 2015).

In pancreatic cancer, BCL7B overexpression is correlated with overall survival (Taniuchi et al., 2018); this was demonstrated with immunohistochemistry experiments that showed BCL7B accumulation in cell protrusions of migrating pancreatic cancer cells. When BCL7B is knockdown, motility, and invasiveness of these cells decreases. Phosphoprotein array analysis was performed to understand if this could be possibly due to a related function of BCL7B with associated intracellular signaling pathways. The analysis demonstrated that BCL7B is involved in the low phosphorylation levels of CREB, a leucine zipper transcription factor

involved in regulating proper cell differentiation (Mayr & Montminy, 2001). Therefore, patients with high levels of BCL7B have the worst overall survival, indicating that BCL7B can be an effective marker to predict the survival of patients with pancreatic cancer (Taniuchi et al., 2018).

Studies conducted in neuronal differentiation and plasticity further underscore the importance of the BCL7 family, showing that BCL7A and BCL7B are highly expressed in the neuron system compared to BCL7C (Wischhof et al., 2017).These studies were performed in mice, and interestingly when Bcl7b is knocked down, the animal survival and behavior are not compromised, while deletion of Bcl7a causes perinatal mortality. It was concluded that in neurons, the expression levels of the three members of the BCL7 family are not the same, and the deletion of one gene is not compensated for the other.

1.4.1.3 BCL7 proteins are part of the SWI/SNF

The BCL7 family counts three members: BCL7A, BCL7B and BCL7C. They are unique to metazoans and are part of the invariable core subunits of the SWI/SNF chromatin remodeling complex (Kadoch et al., 2013a; J. Liu et al., 2021). Cross-linking mass spec data revealed that BCL7 proteins interact with the HSA domain of the catalytic subunit and, with the histone H2B of the nucleosome acidic patch, with BAF47 and DPF2 (Mashtalir et al., 2020). The BCL7 proteins are part of the ATPase module and we think they may act as a bridge or spacer between SMARCA4/BRG1 and SMARCB1/BAF47. The molecular functions and/or role of BCL7 proteins within mSWI/SNF, is essentially unknown. BCL7 proteins are 25 KDa proteins,

characterized by a conserved, ordered N-terminal domain encompassing 51 amino acids and by a non-conserved C- terminal tail (Figure 24). They do not carry any known functional and their C-terminal part is predicted to be disorder.

Our lab has discovered that BCL7 proteins bind DNA and nucleosomes, but interestingly they do not carry any canonical DNA binding domain.
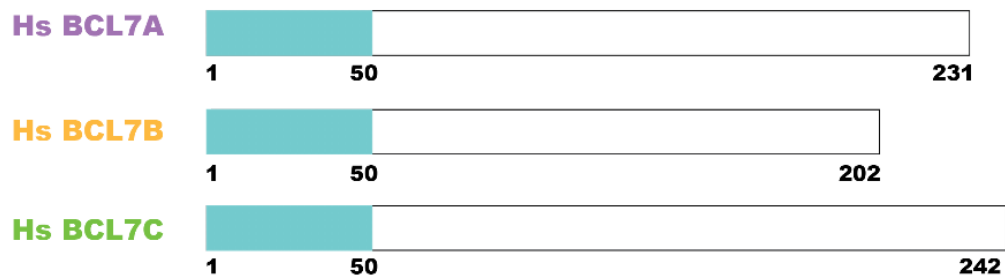
Figure 24. Domain architecture of BCL7 family.

The first 50 amino acids are conserved in the three members. The N-terminal part (aa 1-150) is ordered and the C-terminal tail instead is disordered

1.4.2 Canonical DNA binding domains

1.4.2.1 Helix-turn-helix (HTH) Homeodomain

The Helix-turn-helix domain is the first DNA binding motif discovered, and it was initially described in bacteria. In eukaryotes, a specific class of this motif is widely present from yeast to humans, the "HTH homeodomain". This motif consists of three alpha-helices that are associated with hydrophobic interactions (Figure 25). The helices 2 and 3 resemble the bacterial helix-turn-helix motif. Helix 3 found towards the C-terminal part is known to be the recognition helix, and it is responsible for the interaction with the major groove of the DNA.
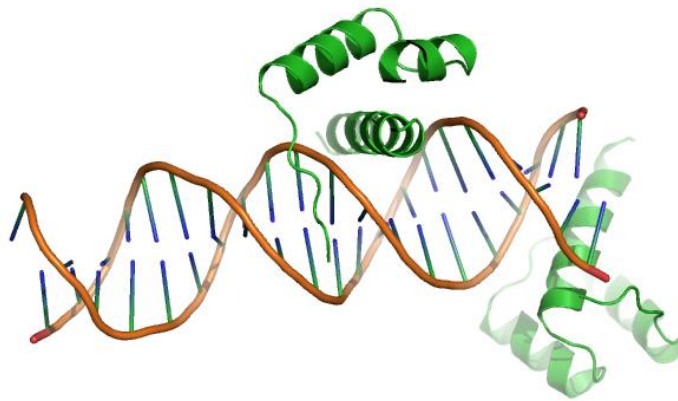
Figure 25. Helix-turn-helix Homeodomain. PDB 1HDD.

1.4.2.2 Zinc Finger domain

The Zinc fingers (ZNF) were described for the first time 40 years ago in the transcription factor IIIa of *Xenopus laevis*. This structure can be defined as any functional and independently folded domain that requires the coordination of one or more zinc ions (Laity et al., 2001). Zinc fingers are very wide in structure and function, and they can bind DNA or RNA and be involved in membrane association and protein-protein interactions. This new class of proteins interacts with three base pairs of DNA and comprises an α-helix and two adjacent β-sheets (Figure 26) (Klug, 2010). An important role has a zinc ion that coordinates the interactions of two cysteines and two histidines between the α-helix and one of the β-sheets. This interaction C2H2 is the most common DNA binding motif in eukaryotes, specifically in their transcription factors.

The first crystal structure of a ZNF protein in complex with a DNA fragment was the mouse transcript factor Zif268, and since then more than 30 types of ZNFs have been described and classified. The Zinc fingers are classified based on the zinc-finger domain structure, but as mentioned above, the most abundant are the C2H2 domain, plant homeodomain PHD and the LIM domains.
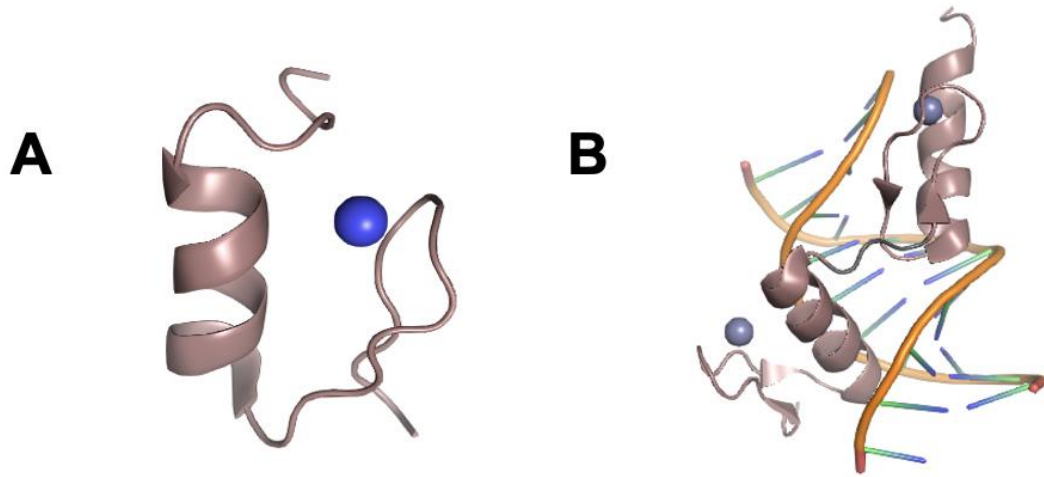
Figure 26. Zinc finger domain structure

A) Single zinc finger from a human enhancer binding protein in solution PDB 3ZNF.
B) Mouse ZFP57 zinc fingers in complex with methylated DNA PDB 4GZN.

1.4.2.3 Leucine zipper

The Leucine zipper domain is composed of two different motifs. The first one rich in leucines and folds as a long bipartite α-helix that is 60 to 80 amino acids length, involved in events of dimerization. The second one is a basic region that can recognize specific sequences of DNA (Vinson et al., 2002). Dimers of leucine zippers are formed by the paired contacts between hydrophobic leucine zippers domains (Figure 27). This conformation allows dimerization in parallel and it bends the helices so that the newly functional dimer forms a flexible fork where the basic domains, at the N-terminal open end, can then interact with DNA. The two leucine zipper are therefore oriented perpendicular to the DNA (O'Shea et al., 1991).
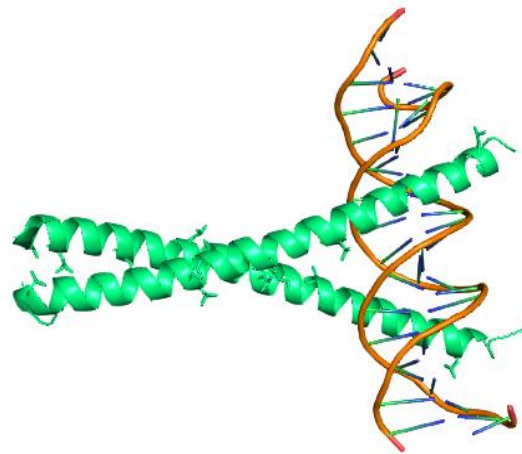
Figure 27. Leucine zipper domain.

Crystal structure of GCN4 basic region leucine zipper bound as a dimer to DNA. PDB 1YSA.

1.4.3 Liquid-liquid phase separation

The mammalian SWI/SNF complex contains several subunits that display disordered regions. The novel subunit, among which the BCL7 proteins, are a good example of molecules carrying intrinsically disordered regions. It is thought that intrinsically disordered regions contribute to the formation of liquid-liquid phase separation but the function of this phenomenon within SWI/SNF is unknown.

In the cells, macromolecules such as proteins, DNA, and RNA are organized based on their functions. This subcellular organization is crucial for a wide range of cellular processes (Alberti, 2017). The nucleus is a highly organized organelle that contains different membrane-less compartments known as nuclear bodies (A & Weber, 2019). The nuclear bodies have a different composition from the nucleoplasm, and they have high concentrations of proteins and acid nucleic. Phase separation is a model that can explain how this nuclear compartmentalization occurs.

Currently, they are two proposed mechanisms of phase separation, the liquid-liquid PS (LLPS) and polymer-polymer PS (PPPS). PPPS is based on active bridging interactions relying on protein binders that compact the chromatin fiber, whereas LLPS is driven by liquid-like multivalent interactions among soluble components with distinct molecular composition and concentration. Both mechanisms can promote the formation of nuclear compartments and are not mutually exclusive,

In the recent years, it has been shown that many proteins that contain intrinsically disordered regions (IDRs) can form liquid-liquid phase separation. The formation of phase separation is driven by intermolecular electrostatic, hydrophobic and cation-π interactions(Pak et al., 2016; Shakya et al., 2020; J. Wang et al., 2018).

1.4.4 Intrinsically disordered proteins

Intrinsically disordered proteins (IDP) proteins do not have a fixed conformation and usually have low sequence complexity. Two different IDPs have been identified, the prion-like composed of polar amino acids like serine, tyrosine, glutamine, asparagine, and glycine, and the second type called low-complexity IDPs containing positive and negative charged amino acids, but they have low content of bulky hydrophobic amino acids (Alberti, 2017).

Proteins that contain intrinsically disordered regions can have different interactions that lead to liquid-liquid phase separation and the formation of condensates (Gallego et al., 2020). In the case of chromatin organization, it has been shown that the linker histone H1 promotes phase separation and acts as a coordinator for condensates formation (Sanulli et al., 2019). In the presence of multi-bromodomain proteins, highly acetylated chromatin forms droplets that simulate the chromatin subdomains present in the nucleus (Gibson et al., 2019).

The first experimental evidence of the existence of phase separation was reported in 2012. Several factors determine phase separation. The first parameter depends on the number of interactive modules a protein has; the bigger the number of modules, the bigger the probability of arranging large complexes as phase-separated droplets. Solubility is another parameter that affects the formation of phase separation; usually, phase-separated proteins have low solubility. (P. Li et al., 2012)

2 METHODS

The BCL7A and BCL7C proteins were subcloned in the vector pGST2 (Figure 28) that has ampicillin resistance in fusion with a tobacco etch virus (TEV) protease cleavable Glutathione S-transferase tag. Recombinant proteins including mutants, and shorter constructs have been overexpressed in *Escherichia coli* by transformation of Lemo21 (DE3) strain which has a chloramphenicol-resistant plasmid and a resistance to phage T1.
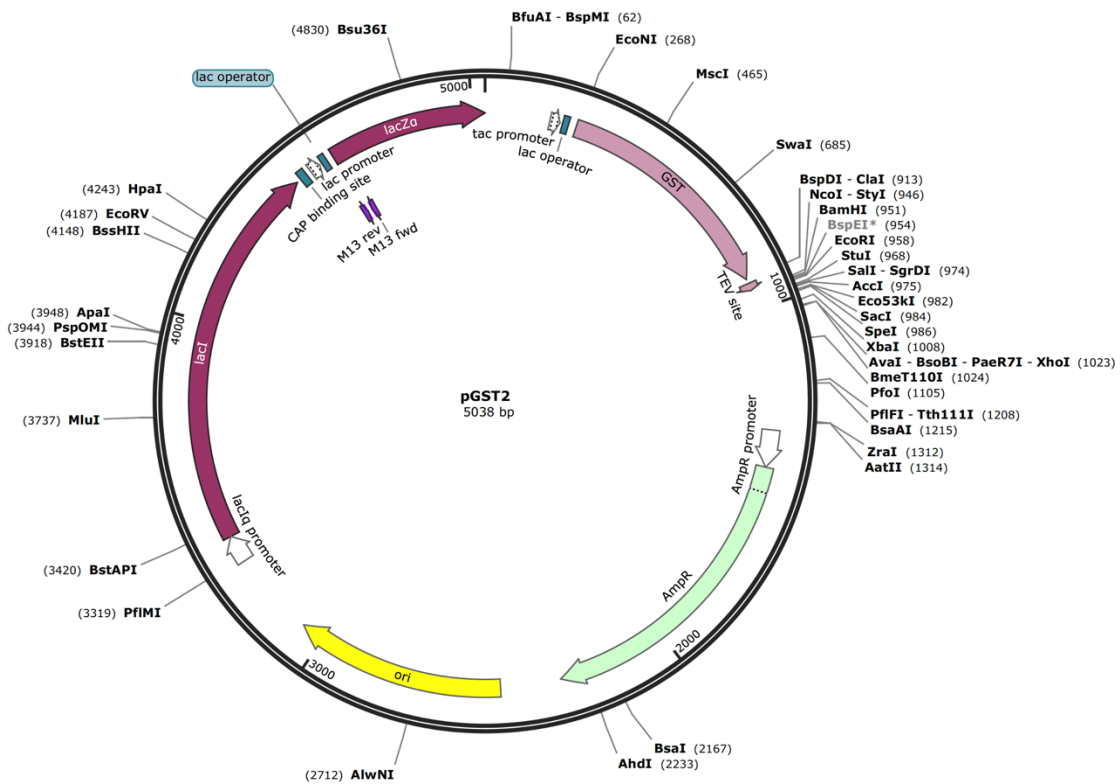
Figure 28. Plasmid map of the pGST2 vector.

A single colony was picked and grown ON into 200mL of Luria Broth (LB) supplemented with chloramphenicol and ampicillin. This culture was used to inoculate 8L of Terrific Broth. The cells were grown to an OD600 of 0.6 at 37°C, and protein expression was induced by the addition of isopropylthiogalactopyranoside (IPTG) 0.2 mM for 18 hr at 19°C. The protein was purified to homogeneity using a 3-step procedure.

2.1 BCL7A proteins purification

Cells were harvested by centrifugation suspended in lysis buffer (50 mM Tris HCl pH 8.0, NaCl 300mM, 0.1% Triton, 5mM BME), and lysed by sonication with 40% amplitude at intervals of one minute on/ one minute off for a total of four minutes. The lysate was clarified by centrifugation at 18,000rpm for 1 hour.

Although these are highly insoluble proteins, a soluble fraction present in the clarified lysate was recovered by incubating two hours with Glutathione S-transferase (GST) sepharose beads. After incubation beads were extensively - washed with wash buffer (50 mM Tris HCL pH 8.0, 300mM NaCl, 1mM DTT) followed by overnight cleavage with Tobacco Etch Virus (TEV) protease, deoxyribonuclease DNAse, and 2 mM $MgCl_2$ at 4°C. Then the protein was concentrated and further purified with a 5 mL Heparin HP column (GE Healthcare) pre-equilibrated in Buffer A (50mM NaCl, 50mM Tris HCl 7.5). The protein was eluted in 2mL fractions during a 20-column volume gradient ranging from 50mM to 2M NaCl. The protein eluted approximately at a concentration of 300mM NaCl. The fractions containing the eluted protein were pulled and concentrated to 1 ml and further purified by size exclusion

chromatography Superdex 200 (GE Healthcare) column equilibrated in size exclusion buffer (150mM NaCl, 20 mM Tris HCl pH 8.0, 1mM DTT) (Figure 29). Peak fractions were pulled, and protein concentration was estimated using the extinction coefficient of the different constructs at 280. After reaching the desired concentration, the proteins were frozen in liquid nitrogen in small aliquots and stored at -80°C. The single point mutants of BCL7A FL R11S, P78S, and L210A were obtained by performing quick-change mutagenesis with the QuikChange II system Agilent kit and confirmed by sequencing.

A
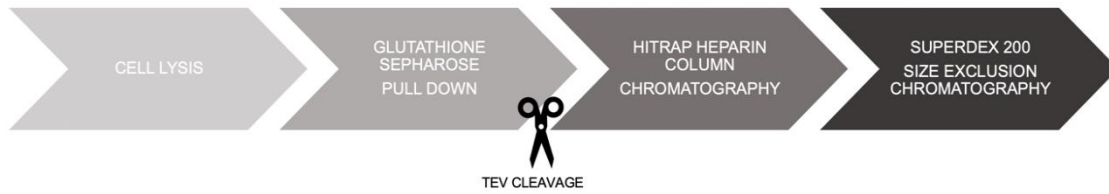
B
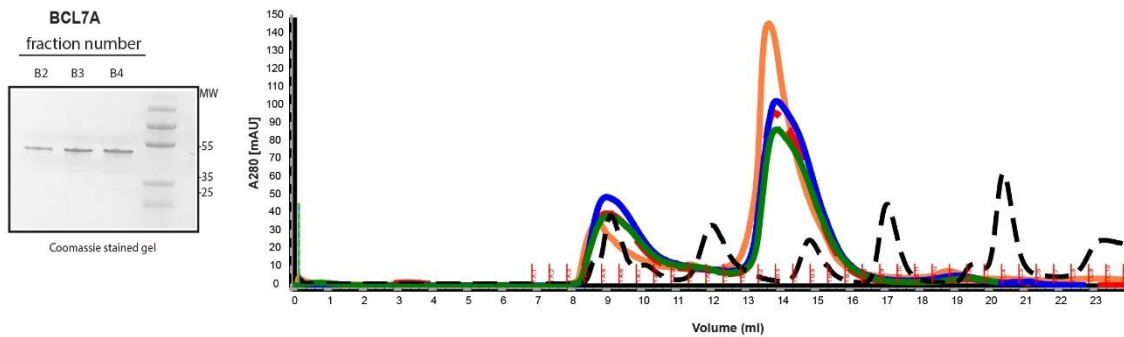
Figure 29. Purification of BCL7A proteins

A) Summary diagram of the optimized purification steps for BCL7A proteins.
B) Comparison of size exclusion chromatograms of BCL7A WT (orange), BCL7A R11S (blue), BCL7A P78S (green), BCL7A L210A(red) overlapped with Biorad Standards (black dot line). The left panel shows BCL7A WT SDS-PAGE gel Coomassie-stained.

2.2 BCL7C proteins purification

Cells were harvested by centrifugation and resuspended in lysis buffer (50 mM Tris pH 8.0, 300mM NaCl, 0.1% Triton, 5mM BME), and lysed by sonication with 40% amplitude at intervals of one minute on/ one minute off for a total of four minutes. The lysate was clarified by centrifugation at 18,000rpm for 1 hour.

Although these are highly insoluble proteins, a soluble fraction present in the clarified lysate was recovered by incubating two hours with GST beads, and beads were extensively washed with wash buffer (50 mM Tris HCL pH 8.0, 300mM NaCl, 1mM DTT) followed by overnight cleavage with TEV protease, deoxyribonuclease DNAse, and 2 mM $MgCl_2$ at 4°C. Then the protein was concentrated and further purified with a 1 mL Resource S column (GE Healthcare) pre-equilibrated in Buffer A (50mM NaCl, 50mM Hepes 8.0). The protein was eluted in 2mL fractions during a 20-column volume gradient ranging from 50mM to 2M NaCl. The protein eluted approximately at 300mM NaCl. The fractions containing the eluted protein were pulled and concentrated to 1 ml and further purified by size exclusion chromatography Superdex 200 (GE Healthcare) column equilibrated in size exclusion buffer (150mM NaCl, 20 mM Tris HCl pH 8.0, 1mM DTT) (Figure 30). Peak fractions were pulled, and protein concentration was estimated using the extinction coefficient of the different constructs at 280. After reaching the desired concentration, the proteins were frozen in liquid nitrogen in small aliquots and stored at -80°C.

A
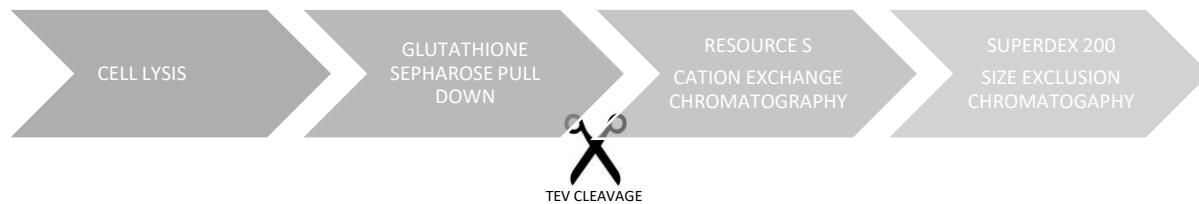
B
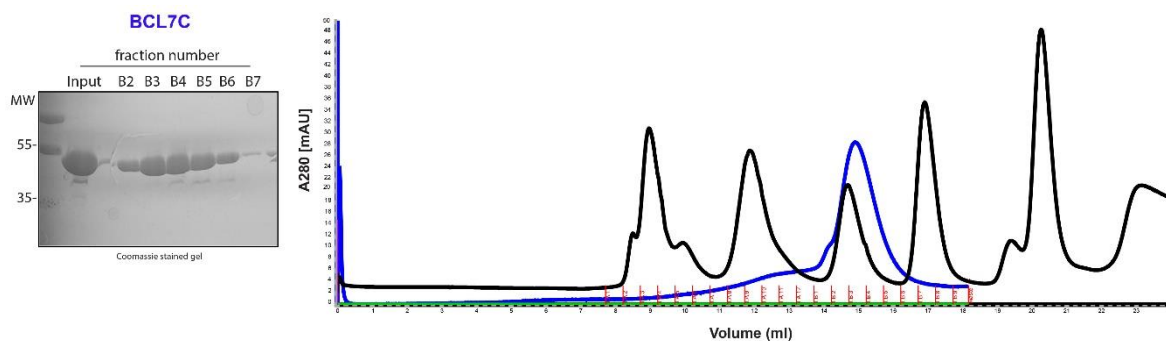
Figure 30. BCL7C purification.

A) Summary diagram of the optimized purification steps for BCL7C proteins.
B) Size exclusion chromatogram of BCL7C WT (blue) overlapped with Biorad Standards (black line). A Coomassie-stained gel of BCL7C WT protein is shown.

2.3 BAF47 Purification

Human BAF47 has been subcloned in a pHis parallel vector in fusion with a 6xHis-SUMO-tag and overexpressed in Rosetta cells. A single colony was picked and inoculated overnight into 200mL of LB with the corresponding antibiotics. This culture was used to inoculate 8L of Terrific Broth media. The cells were grown to an OD600 of 0.6 at 37°C, and protein expression was induced by adding isopropylthiogalactopyranoside (IPTG) 0.2mM for 18 hr at 19°C. The protein was purified to homogeneity using a 2-step procedure. Cells were harvested by centrifugation and resuspended in Lysis Buffer (50 mM Tris HCl pH 8.0, NaCl 300mM, 0.1% Triton, 5mM BME), and sonicated with 40% amplitude at intervals of one minute on/ one minute off for a total of four minutes. The total lysate was clarified by centrifugation at 18,000rpm for 1 hour. The clarified lysate was mixed with 2ml of Ni^{2+}-NTA agarose beads and incubated for 45 minutes. The beads were packed into an Econo column and washed with ten volumes of a column with wash buffer (300mM NaCl, 20mM Tris pH 8.0, 1mM DTT). Serial dilution elutions were performed with ten different elution buffers with an initial imidazole concentration of 50mM to a final of 500mM. Fractions were checked by SDS-PAGE electrophoresis and, fractions containing BAF47 were pulled together, and 50µl of 1mg/ml ULP1 protease capable of cleaving at the C-terminus end of SUMO protein was added and 10uL of DNAse. The last step of purification was performed by size exclusion chromatography. The fractions corresponding to the protein were verified on a 12% SDS PAGE electrophoresis (Figure 31), fractions with the pure protein were pulled together concentrated with AMICON 10 kDa to 7 mg/ml and stored at -80 °C.

A

SUMO CLEAVAGE

B

BAF47
fraction number
B1 B2 B3 B4 B5

MW
55-
35-
25-

Coomasie stained gel

A280 [mAU]

Volume
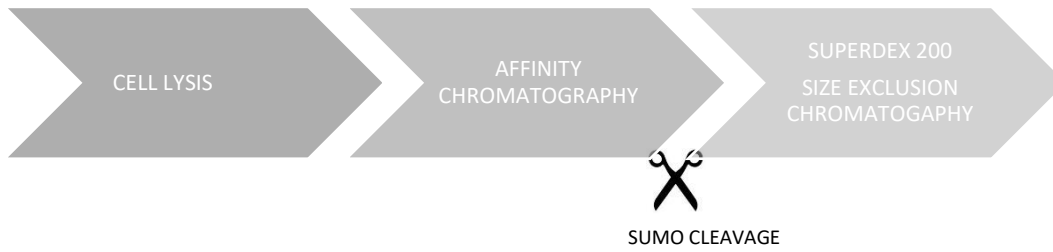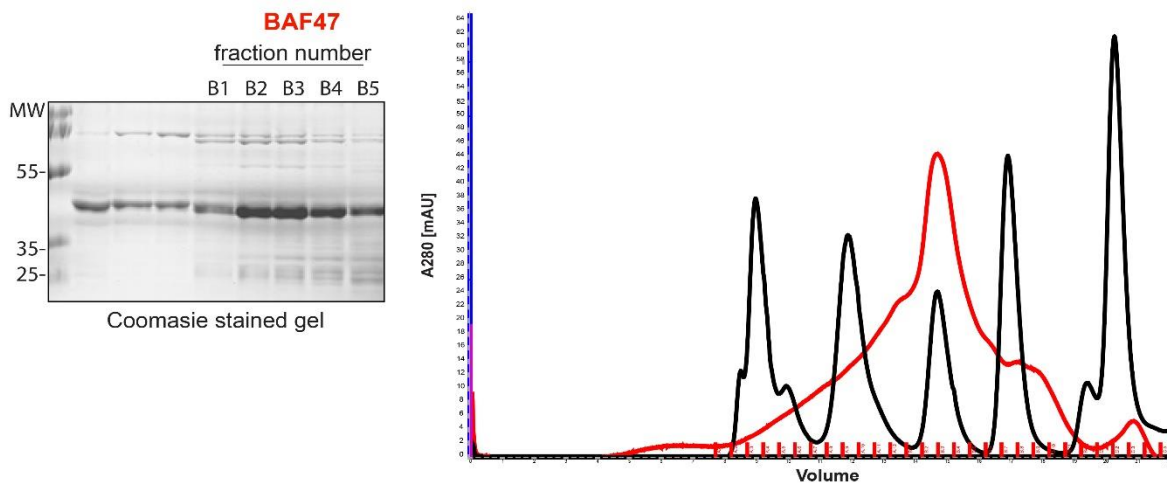
Figure 31. Purification of BAF47 protein.

A) Summary diagram of the optimized purification steps for BAF47.
B) Size exclusion chromatogram of BAF47 (red) overlapped with Biorad Standards (black line).
The left panel shows BCL7C WT SDS-PAGE gel Coomassie-stained.

2.4 Complex preparation

BAF47/BCL7C complex

The purification of the complex was performed with a double pulldown, followed by size exclusion chromatography. As described previously for the individual protein purifications, the proteins were expressed and growth individually in Rosetta cells. One pellet coming from 8L growth was used for each protein.

BAF47-6xHis-SUMO-tag and BCL7C-GST tagged pellets were resuspended individually in 100mL of Lysis Buffer (300mM NaCl, 20mM Tris pH 8.0, 0.1% Triton and 5mM BME), sonicated (6 cycles of 1 min separate by 1 min of incubation at four °C with an amplitude of 40%), the total lysate was clarified centrifuging 50min at 18000RPM at four °C. The soluble fractions were mixed and incubated with 2ml Nickel beads previously equilibrated in Wash buffer. After 45 minutes of incubation, the complex was eluted using serial elution buffers from the Nickel beads previously washed through an Econo column (as described above for BAF47). The presence of protein was verified with Bradford, and fractions containing protein were pulled together and incubated for an hour with GST beads. The GST beads containing the two proteins were washed two times at 1000RPM and incubated with 20uL of SUMO protease and 50uL of TEV protease overnight at 4C.

The beads were spun 1000RPM for 10 minutes, and the flow-through was collected after pouring the beads in an Econo column. The approximate volume is 10mL, and this was concentrated to a final volume of 1mL to inject in the size exclusion column. The fractions were collected, and the presence of both proteins and purity of the complex was evaluated with a 12% SDS PAGE electrophoresis gel. To confirm the

presence of the desired proteins, the bands were sent to mass spectrometry, and western blots were performed.

2.5 Western blot

The same fractions from the size exclusion chromatogram were selected and run in SDS-PAGE 12% gel. The gel ran for 45min at 200V. The next step was to transfer the protein from the gel to a membrane of nitrocellulose. The transfer starts with the cassette assembly, cathode, two blotting papers (with the same size as the gel), gel membrane, two blotting paper, and anode; all the components were soaked with transfer buffer (250mM Tris, 1.9M Glycine). The transfer is done with 100V, one hour at 4C, and needs to be done in a cold environment, and the buffer must be pre-chilled.

The membrane was then stained with Ponceau colorant for 10min then washed to see the protein transfer. The membrane was blocked for 1 hour with 5% milk prepared in TBS. Primary antibodies were used independently for BCL7C and BAF47. The membrane was blocked with TBST containing 5% milk and incubated first with the primary antibody (anti-BCL7C mouse, Qiagen) diluted at 1/10000 in TBST/5% milk for 1 hour at room temperature and then with the secondary anti-mouse antibody HRP (horseradish peroxidase, GE Healthcare) diluted at 1/5000 in TBST/5% milk for 1 hour at room temperature. Finally, the HRP signal is detected (Clarity Western ECL substrate, BioRad). Between each step, the membrane was washed twice for 10 minutes in TBST and visualized on the Chemidoc Touch Imaging System (BioRad).

2.6 Nucleosome core particle (NCP) assembly

In the lab we perform nucleosome assembly via salt gradient to obtain large quantities of nucleosomes for biochemical and biophysical assays as well for crystallization and Cryo-EM studies, we follow a protocol based in (Dyer et al., 2004; Luger et al., 1999) With this method we obtain a high yield of homogeneous mononucleosomes. We use *Xenopus laevis* core histones, H2A, H2B, H3 and H4, and 147-bp dsDNA with the 'widom' 601 sequence.

The purification that we use in the lab to purify histone octamer is a faster method that does not involve denaturalizing the proteins. The protocol is taken from (Shim et al., 2012) "Polycistronic co-expression and nondenaturing purification of histone octamers." We purify the 601 DNA from the protocol.

2.7 Octamer purification

The polycistronic plasmid pET29a-rbs-6xHis-thrombin-H2A-rbs-H2B-rbs-H3-rbs-H4-thrombin-6xHis was transformed and expressed in BL21(DE3) pLysS cells. One colony was taken and grown overnight in 200mL of LB with 200uL of chloramphenicol and kanamycin and left overnight at 37°C. From the overnight culture, 8L of bacteria were grown in 2YT media. Induction was done with 0.4mM IPTG when the OD of the culture reached 0.4 and left the induction overnight at 30 °C. The cultures were harvested at 4000RPM for 20 min, and the pellets were stored at -80 °C.

The pellet was resuspended in 200ml of lysis buffer (2M NaCl, 20mM Tris pH 8.0, 0.1 mM TCEP). The pLysS cells were sonicated (1min on, 1min off, Amplitude 40%)

(Vibra Cell, Bioblock scientific) and clarified by spin down at 18000 RPM for 45min.

The clarified supernatant was transferred in a 200ml bottle, and 5mL of Talon beads previously equilibrated with wash buffer (2M NaCl, 20mM Tris pH 8.0) were added. Talon beads have more specificity than Nickel beads, and both can be used.

The incubation with Talon bead is for 45 minutes; no longer times are recommended. Beads are spun down 10 minutes at 1000 RPM, and two washes are made, changing buffer and spinning down again each time.

2.8 SEC-MALS of BCL7A and BCL7A mutants

BCL7A wild type and mutants were analyzed by size exclusion chromatography-multi-angle light scattering (SEC-MALS) with a Superdex 200 (GE Healthcare) size exclusion column. The column was extensively equilibrated in buffer composed 20 mM Tris HCl pH 8.0, 150 mM NaCl, 1mm DTT) at room temperature before applying 100 µL BCL7A through a capillary loop. The system flow was maintained at 0.5 mL/min. Light scattering and UV absorbance data at 280 nm were collected across a 25 mL volume post-injection using a miniDAWN TREOS light scattering detector (Wyatt).

2.9 Electrophoretic mobility shift assay

Mobility shifts assay were performed with fluorescently labeled nucleosomes. The nucleosomes are assembled using the salt dilution with recombinant histones and Cy5-labelled 601 DNA produced by PCR. The labelled probes are incubated at 10 nM in 30 µl binding reactions containing 50 mM Tris HCl pH 8.5, 20 mM KCl, 10 mM

b-mercaptoethanol, 10% (v/v) glycerol. Purified proteins are added to the binding reactions at varying concentrations, and the reactions are incubated on ice for 30 minutes. Samples are then loaded on non-denaturing 4% acrylamide gels (60:1 ratio of acrylamide to bis-acrylamide) containing 10% glycerol, using 0.5 X TBE (40 mM Tris HCl, 45 mM boric acid and 1 mM EDTA) as running buffer. After electrophoresis, gels are immediately scanned using a Typhoon phosphoimager.

2.10 Binding studies of BCL7A and nucleosome by microscale thermophoresis (MST)

We determined the binding constant with measurements for equilibrium binding. We measured BCL7A-NCP complex interactions on a temperature equilibrated Monolith NT.115 (NanoTemper) using 40 nM fluorescently labeled Cy5 nucleosomes. For each set of measurements, a serial dilution of BCL7A protein was prepared at 2X final concentration starting with 30 µM using the following buffer: 20 mM HEPES pH 8.0, 150 mM NaCl, and mixed equal volume of nucleosome stock solution (prepared in the same buffer) to a final nucleosome concentration of 40nM. Samples were given 10 minutes to equilibrate and were spun down in centrifuge 3 min a 10,000rpm before loading in the Monolith TM NT.115 Premium Coated capillaries.

Experiments were conducted at 25°C in the MST Buffer (20 mM HEPES pH 8.0, 150 mM NaCl, 0.05% Tween-20) with LED power at 80% and MST power set at 20%. Binding curves were fitted to three sets of replicates. Data were analyzed in the Nano Temper Analysis software using a *Kd* fitting analysis on temperature jump data. Data points at the extremes of the range were excluded from the analysis. Straight lines

were fitted to the unsaturated and saturated portions of the data in triplicate using GraphPad Prism

2.11 Crystallization of BCL7A-NCP complex

The complexes BCL7A-NCP and BCL7A (1-100)-NCP were isolated after purification by size exclusion chromatography (Superose 6 Increase) at 2mg/ml and initial co-crystallization trials were performed by sitting drop vapor diffusion, using the mosquito robot (SPT Labtech).

2.12 Labeling of BCL7A (1-100)

We labelled the protein through the N-α amino group using (5-(and-6)-Carboxyrhodamine 6G, succinimidyl. The BCL7A 1-100 protein was purified as described before, with some exceptions. We did buffer exchange chromatography after the heparin purification, with the following buffer: 20mM Hepes pH 7.5, 300mM NaCl, instead of using a Tris HCl base buffer. We performed this extra step to remove reactive nucleophiles that may interfere with the labeling reaction in buffer with Tris HCl.

We labelled 5 mg of chromatin protein at 2-4 mg/ml.

We dissolved the (5-(and-6)-Carboxyrhodamine 6G, succinimidyl ester in dimethylformamide (DMF) at 10 mM and follow the same mixture as in (McGinty et al., 2016).

250 μl 235 μM BCL7A1-100

738 μl labeling buffer (10 mM HEPES pH 7.5, 300 mM NaCl)

12 μl 10 mM carboxyrhodamine succinimidyl ester

1000 μl total volume

We left the reaction at room temperature for 1 hour.

After the reaction finished, we performed a desalting column purification with the optimized buffer for the protein: 20mM Tris pH 8.0, 150mM NaCl, 1mM DTT. We concentrated and quantified the protein as in the method (McGinty et al., 2016). The fluorophore absorbs at 280 nm, meaning that the total absorbance at 280 is the sum of BCL7A 1-100 and the absorbance of the fluorophore. To calculate the absorbance coming from the fluorophore, we measured the absorbance at 532nm. Thus, by measuring the absorbance of the labeled protein at 280 nm and 532 nm, we can calculate the concentration of the protein and the fluorophore in our labeled protein samples. The BCL7A 1-100 complex was purified in the following buffer: 20mM Tris pH 8.0, 100mM NaCl, 1mM DTT.

2.13 Cryo-EM Data processing

Images were recorded using Serial EM on Titan Krios microscope at 300 keV of acceleration voltage and each equipped with a K3 camera (Gatan, Inc., Pleasanton, CA, USA). The target defocus ranges were set to -0.6 to -1.6 μm. The movies contained 40 frames, collected in super-resolution counting mode with a pixel size of 0.862 Å/pixel, and exposures ranging from 5 to 6.4 e-/A2/s, which corresponds to a total dose of ~ 50.93 e-/Å2. The pre-processing was made with warp software that enables evaluation and correction the cryo-EM data during data acquisition; it provides an interface between data acquisition software and 3D refinement of pre-processed data software like cryoSPARC and RELION.

3 RESULTS

3.1 Aim 1. Determine the molecular function of BCL7 proteins

Electrophoretic Mobility Shifts Assays (EMSA) were performed; labeled Cy5 601 DNA was amplified by PCR, purified, and used for NCP small scale assembly. Full length BCL7A and BCL7C proteins binding to Cy5 601 DNA and NCP was tested, showing that both proteins bind DNA and nucleosomes (Figure 32).
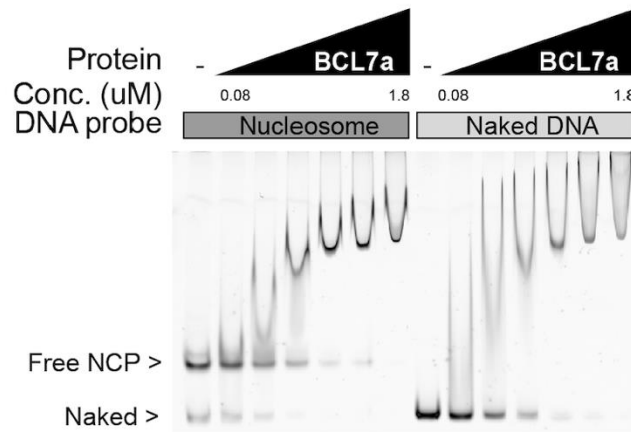

Figure 32. BCL7A binds the nucleosome core particle and free DNA.

Electrophoretic mobility shift assay (EMSA) showing the binding of BCL7A with nucleosome and naked 601 DNA.

```
                                  1                  11                 21                 31                 41
Consensus                         ----msgRSv         RAETRsRakD         dikrVmaaie         KVRkWEKKWV         TvgDTslRIY
Conservation                              .          .       :         ::    :  .:.          ***  *         :  **.::**
Homo_sapiens                      ----MSGRSV         RAETRSRAKD         DIKRVMAAIE         KVRKWEKKWV         TVGDTSLRIY
Mus_musculus                      ----MSGRSV         RAETRSRAKD         DIKRVMAAIE         KVRKWEKKWV         TVGDTSLRIY
Pan_troglodytes                   ----MSGRSV         RAETRSRAKD         DIKRVMAAIE         KVRKWEKKWV         TVGDTSLRIY
Canis_lupus                       MAEPGARNRD         PGAQTAWMEF         QLAFCVIPLS         LDIVREKKWV         TVGDTSLRIY
Danio_rerio                       ----MSGRSV         RAETRSRAKD         DIKRVMAAIE         KVRKWEKKWV         TVGDTSLRIY
Salmo_salar                       ----MSGRSV         RAETRSRAKD         DIKRVMAAVE         KVRNWEKKWV         TVGDTSLRIY
Drosophila_melanogaster           -----MSRSV         RAETRSRAKD         DIKRVMQAVD         KVRHWEKKWV         TISDTTMKIY
Caenorhabditis_elegans            --MYSANRSH         RAETRNRSKD         ELRKVINSLE         KVRRWEKKLV         LIKDTNIRIY

                                  51                 61                 71                 81                 91
Consensus                         KWVPVtepkv         ddknKnKkKg         kdekcgseVt         tPENSSspgm         mdmhddnsnq
Conservation                      ****::                 :  *         :  .   .  :         .   ...             .
Homo_sapiens                      KWVPVTEPKV         DDKNKNKKKG         KDEKCGSEVT         TPENSSSPGM         MDMHDDNSNQ
Mus_musculus                      KWVPVTEPKV         DDKNKNKKKG         KDEKCGSEVT         TPENSSSPGM         MDMHDDNSNQ
Pan_troglodytes                   KWVPVTEPKV         DDKNKNKKKG         KDEKCGSEVT         TPENSSSPGM         MDMHDDNSNQ
Canis_lupus                       KWVPVTEPKV         DDKNKNKKKG         KDEKCGSEVT         TPENSSSPGM         MDMHDDTSNQ
Danio_rerio                       KWVPVTEPKS         DD-NKNKKKG         KDDKYGSEVT         TPENSSSPGM         MDMHDDNSNQ
Salmo_salar                       KWVPVTEPKS         NDKSKNKKKA         NNDKYGSEIT         TPENSSSPGM         MDMHDDNSNQ
Drosophila_melanogaster           KWVPIAS---         --ASEKKAKL         ESSPGSAAVR         RPPPGSG---         ----------
Caenorhabditis_elegans            KWVPVSAQNI         MAPPKIKEVK         EVDEESNQVP         SAENSQD---         ---------S

                                  101                111                121                131                141
Consensus                         sSiadaSPik         qENSsnssPa         pepnsavpsd         gteakadeaq         adgkehpgae
Conservation                      :   .   .          .  ....  .:             .                 .                .     .
Homo_sapiens                      SSIADASPIK         QENSSNSSPA         PEPNSAVPSD         GTEAKVDEAQ         ADGKEHPGAE
Mus_musculus                      SSIADASPIK         QENSSNSSPA         PETNPPVPSD         GTEAKADEAQ         ADGKEHPGAE
Pan_troglodytes                   SSIADASPIK         QENSSNSSPA         PEPNSAVPSD         GTEAKVDEAQ         ADGKEHPGAE
Canis_lupus                       SSIADASPIK         QENSSNSSPA         PEPSSAVPGE         GTDAKADETQ         ADGKELPGAE
Danio_rerio                       SSIADSSPLK         QETSNNTSPA         PEPMAASQND         SSDLKNDQ--         YPSKQPSSGQ
Salmo_salar                       SSIADCSPMK         QENSNSASPA         QEAMATSQSD         GIEAKSDQSQ         SPGKEQLKST
Drosophila_melanogaster           VTPVGGSKSD         KENSQKGTPT         PPQITPSYQG         LTAEDSNTCF         SVVSDSQGAD
Caenorhabditis_elegans            TSVTQPPPQF         DINEDSNFST         GDHFDSDSNQ         TFEPQNYQGG         ATGSTDFSS-

                                  151                161                171                181                191
Consensus                         dasseqnsqs         sme-smnsse         kverQpsges         glaaetsais         qd--------
Conservation                                                 .
Homo_sapiens                      DASDEQNSQS         SMEHSMNSSE         KVDRQPSGDS         GLAAETSAIS         QD--------
Mus_musculus                      DASEEQNSQS         SMENSVNSSE         KAERQPSAES         GLAAETSAVS         QD--------
Pan_troglodytes                   DASDEQNSQS         SMEHSMNSSE         KVDRQPSGDS         GLAAETSAIS         QVPRSRSQRG
Canis_lupus                       DASDEQNSQS         SMENSMNSSE         KVERQPSGDA         GLAAEMSAIS         QD--------
Danio_rerio                       DHKSEQNHCS         SES--MTSKR         DSKSQGDSES         FLDSSKSAQE         LE--------
Salmo_salar                       DSKSGNVHPS         AFSNTSTPKR         DRRSTGRNEP         SSDTPQDSQE         SE--------
Drosophila_melanogaster           FVSS------         -----MPFSE         DSNSQGSDG-         ----------         ----------
Caenorhabditis_elegans            ----------         -----MRDAE         MTSKQP----         ----------         ----------

                                  201                211                221                231
Consensus                         ----------         ----egvpPs         KkmKleasqq         nseem
Conservation
Homo_sapiens                      ----------         ---LEGVPPS         KKMKLEASQQ         NSEEM
Mus_musculus                      ----------         ---LEGVPPS         KKMKLEASQQ         NSEEM
Pan_troglodytes                   SQIGREPIGL         SGDLEGVPPS         KKMKLEASQQ         NSEEM
Canis_lupus                       ----------         ---LEGVPPS         KKMKLEASQQ         NSEEM
Danio_rerio                       ----------         ----DGAPPS         KKGKIDSSSE         ES---
Salmo_salar                       ----------         ----DGTPPN         KKGKLETPSD         FLKTS
Drosophila_melanogaster           ----------         --------PV         KRLKTSD---         -----
Caenorhabditis_elegans            ----------         ----------         ----------         -----
```

Figure 33. Sequence Alignment of BCL7A.

Sequence alignment of BCL7A among different species made in CLUSTALW. The first 100 amino acids are conserved.

Sequence alignment of the BCL7 family shows that the first 50 amino acids are conserved within the three proteins, remarking that this region must be the functional domain. Interestingly this region is the only part of the protein where we can predict a secondary structure. To verify if this is the domain that allows binding, we tried the purification of the first 50 amino acids of BCL7A, and this protein was not possible to obtain. After sequence alignment of different BCL7A species (Figure 33), we found that the first 100 amino acids are highly conserved within this protein. The purified BCL7A 1-100 protein is sufficient to bind to nucleosomes but not with the same affinity as the full-length protein does. We could see the same effect with BCL7C 1-100 protein (Figure 34).
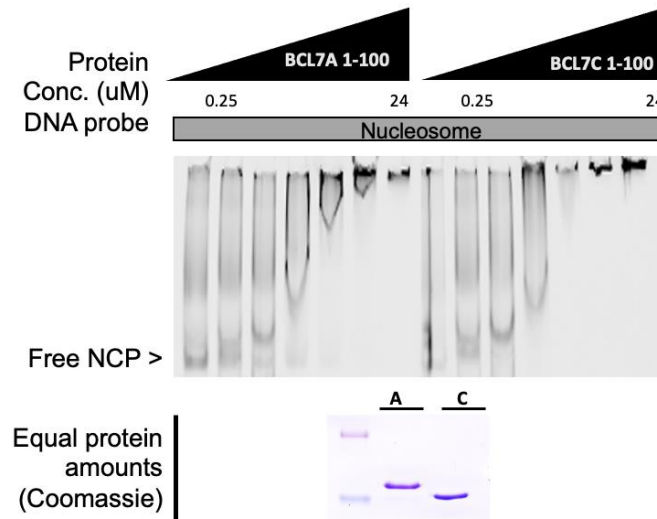
Figure 34. The first 100 amino acids of BCL7A and BCL7C are sufficient for nucleosome binding.

EMSA showing the binding of BCL7A 1-100 and BCL7C1-100 with nucleosome; equal amounts of protein were used as shown in the coomassie stained gel.

Unexpectedly, when we expressed the rest of the protein and tested the ability for binding to the nucleosome, we found that the C-terminal part can interact with the nucleosome (Figure 35).
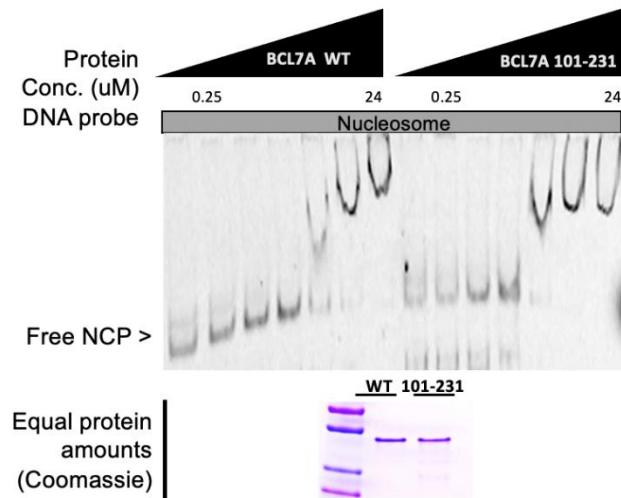
Figure 35. The C-terminal part of BCL7A contributes to nucleosome binding.

EMSA showing the binding of BCL7A WT and BCL7A 101-231 with nucleosome, equal amounts of protein were used.

3.1.1 BCL7A binds the NCP with high affinity

We showed above that the complex between BCL7A, and NCP is stable. To determine the affinity of BCL7A for NCP, we measured the Kd of the complex using microscale thermophoresis (MST). To perform this experiment, we assembled NCP with Cy5 601 DNA. In a Nanotemper, we completed several rounds of titrations. We normalized the fluorescence intensity changes corresponding to different concentrations of BCL7A and calculated that the Kd of the protein for NCP is around 230 nM (Figure 36), indicating that the protein has a very high affinity for NCP in the absence of other SWI/SNF components.

One site -- Total	
Best-fit values	
Bmax	0.9204
Kd	227.0
NS	1.478e-005
Background	0.02406
95% CI (profile likelihood)	
Bmax	0.7848 to 1.089
Kd	130.1 to 392.4
NS	-1.327e-005 to 3.957e-005
Background	-0.03008 to 0.07539
Goodness of Fit	
Degrees of Freedom	8
R squared	0.9913
Sum of Squares	0.01494
Sy.x	0.04322
Number of points	
# of X values	12
# Y values analyzed	12
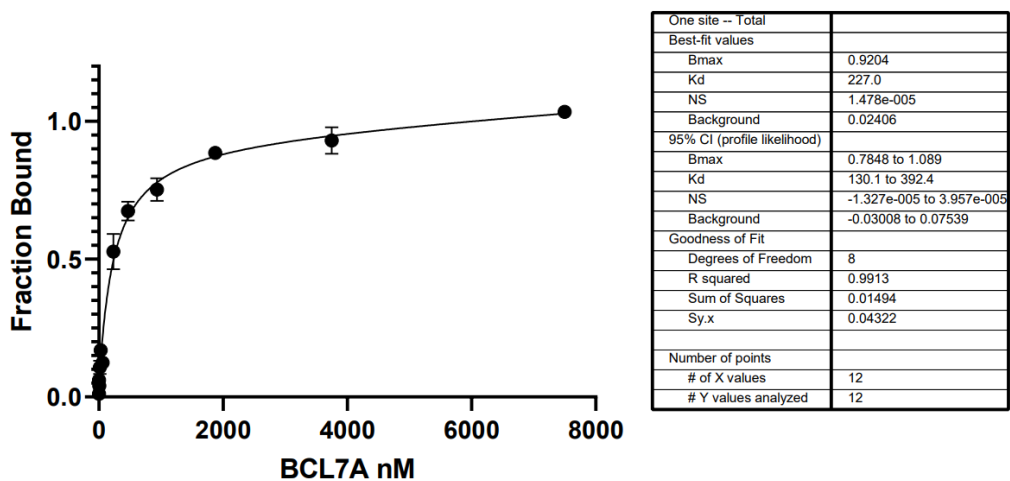
Figure 36. BCL7A binding affinity for the nucleosome by microscale thermophoresis (MST)

Microscale thermophoresis of BCL7A with labeled Cy5 nucleosomes shows a Kd in a nanomolar range. Cy5 601 DNA sequence was used for the assembly of NCP. A solution of 150nM nucleosomes was titrated against increasing concentrations of BCL7A from 50µM to nanomolar amounts. The Kd of 230nM was calculated from the normalized fluorescence intensity plotted against the BCL7A concentration.

3.1.2 Secondary structure prediction

The secondary prediction softwares are a wonderful tool to predict the local secondary structure of proteins and identify domains; as we were interested into knowing the function of the BCL7 proteins, and more specifically BCL7A and BCL7C, we performed protein secondary prediction using PSIPRED to identify a functional domain (Figure 37), however no canonical domains were identified. As mentioned before both proteins share homology in the first 50 amino acids from the N-terminal region. The conserved region is predicted to have two alpha helices and two beta strands; however, the rest of the molecule is characterized by low complexity and it is disordered. For instance, BCL7A predictions show five more helices at the end of the sequence.

According to secondary structure prediction, BCL7C protein carries an alpha helix and a beta strand within the first ninety amino acids, and the rest of the molecule does not carry any secondary structure feature and it is disordered. The presence of such an extended disordered part of the molecule triggers questions about its role. If the conserved domain is responsible for binding to the SWI/SNF complex and to the nucleosome what is the rest of the molecule doing and which characteristic brings to the proteins.

A

Legend:

■ Strand (yellow)	**Conf:** -▁▂▃▄▅█+ Confidence of prediction
■ Helix (pink)	**Cart:** 3-state assignment cartoon
— Coil	**Pred:** 3-state prediction
	AA: Target Sequence
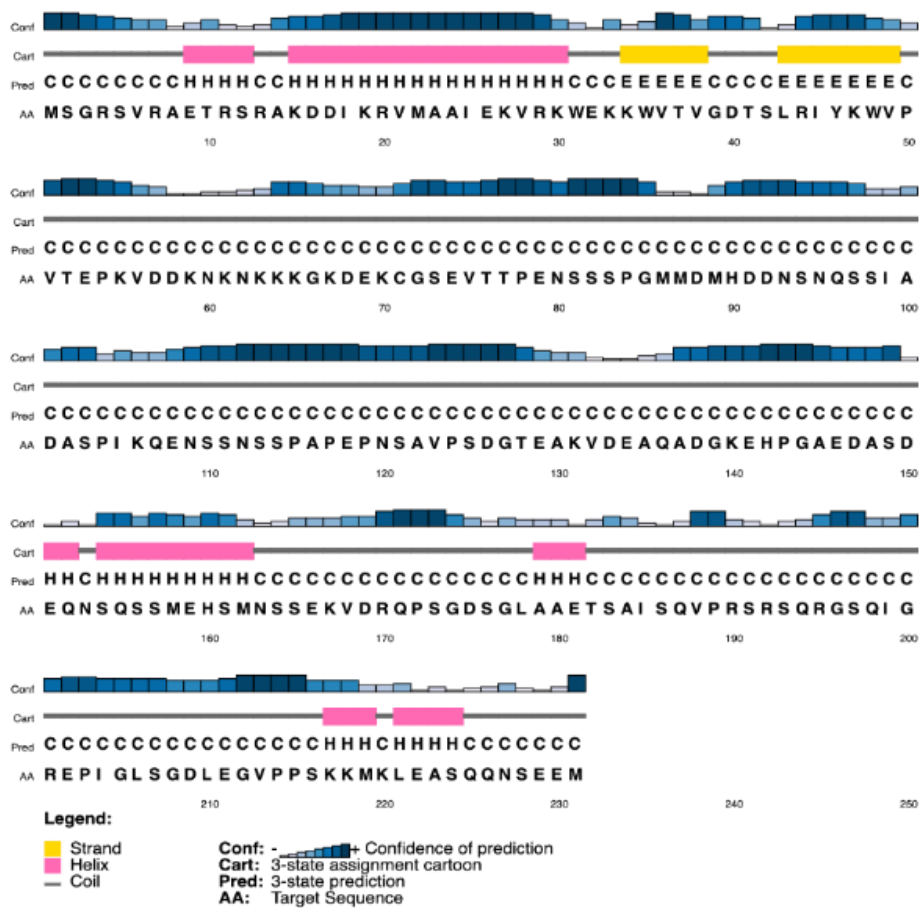
Figure continues on next page.

B

Figure 37. PSIPRED sequence alignment of BCL7A and BCL7C.

3.1.3 Spectroscopic properties of BCL7A

Circular dichroism is an extraordinary technique that allows to determine quickly the secondary structure of proteins. Proteins containing α-helixes are characterized by negative bands at 222 nm and 208 nm and a positive band at 193 nm, whereas proteins with well-defined antiparallel β-sheets display negative bands at 218 nm and positive bands at 195 nm. BCL7A spectra were obtained in buffers of different compositions or at different protein concentrations. To test the effect of different buffers BCL7A was prepared at a concentration of 30 µM in one of the following buffers: 1) 50mM Tris pH 8.0, 150mM NaCl; 1mM DTT; 2) 20 mM Sodium Phosphate pH 7.5, 150 mM NaCl; 3) 20 mM Sodium Phosphate; 150 mM NaCl; 1mM DTT (Figure 38A). The spectrum of each buffer was subtracted from each protein measurement. Spectra of BCL7A proteins were obtained at different concentrations as well (Figure 38B).
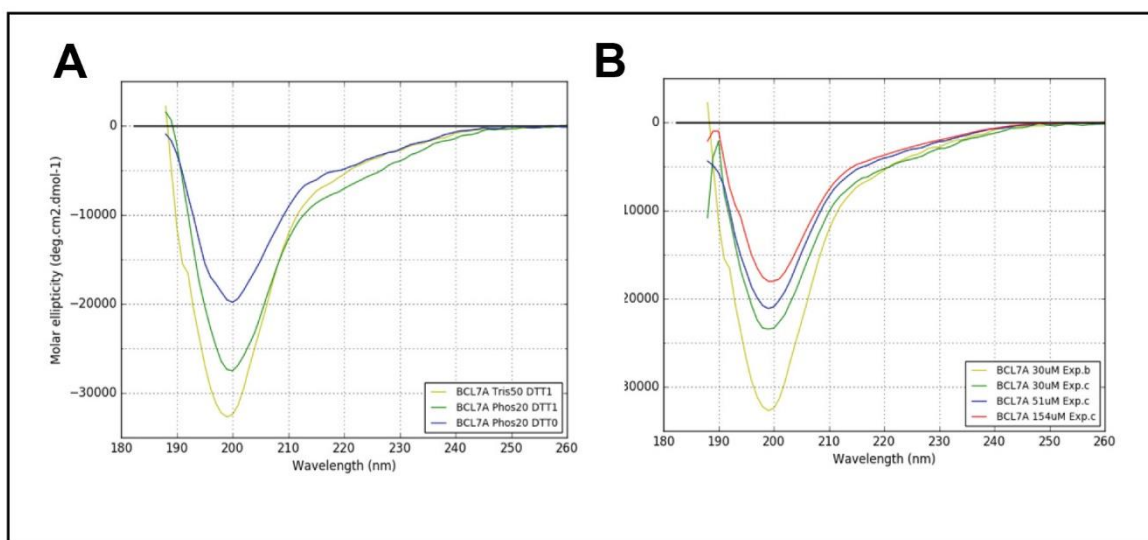
Figure 38. CD far UV Spectrum of BCL7A

A) Spectrum of BCL7A with different buffers, B) and with different protein concentrations.

The spectra of BCL7A are shown in figure 38. Deconvolution analysis was made with the Contin program on the SDP48 database. The CD analysis shows that full-length BCL7A is at least 75% disordered independently than protein concentration and buffer composition. BCL7A CD spectrum correspond to a spectrum typical of IDPs with a minimum centered at 200nm and the absence of strong negative signals characteristic of α-helixes and/or β-sheets secondary structures; For a long time, the spectra of IDPs were considered as the spectra of a 'random coil' polypeptide.

3.1.4 BCL7A is involved in phase separation.

Proteins with intrinsically disordered regions can drive phase separation above a critical concentration. Because the interactions that drive the event of phase separation are weak, salt concentration and temperature are critical parameters to be considered as triggers for phase separation formation as well. During BCL7A purifications and in vitro manipulations, we have never observed a behavior typical of IDPs or noticed droplet formation. As parallel in cell studies with BCL7A were conducted in the lab by other lab members, we discovered that the recruitment of BCL7A at sites of DNA damage is PARP dependent. PARP is an enzyme that deposits the PTM poly(ADP-ribose) (PAR). Thus, we decided to investigate if poly(ADP-ribose) could have an impact on phase separation formation. We also keep into consideration RNA as RNA is known to be able to mediate phase separation formation and in vitro studies performed by other lab members showed that BCL7A binds RNA as well. We tested the formation of phase separation with 150 μM BCL7A protein in presence or absence of 15 μM of NCP, with the addition

of 10 µM of poly(ADP-ribose) or 5µM of dsRNA. The experiments were performed in buffer consisting of 20mM Hepes pH 8.0 and 150mM KCl.

On a glass cover slide we mixed 1 µL of BCL7A (full length or the truncated version encompassing amino acids 1-100) with or without 1 µL of NCP, with or without 1µL of poly(ADP-ribose) or RNA and observed the drop through a stereoscopic microscope for several minutes. We realized that, when the reaction mixture contained BCL7A, NCP and poly(ADP-ribose) or RNA, phase separation occurs. Phase separation was not observed in absence of NCP or poly(ADP-ribose) or RNA. When phase separation occurs in the drops, an initial precipitation event occurs within the first 5 minutes, and after 10 minutes, this precipitation turns into phase separation; however, it takes seconds for the droplets to arrange all together into a bigger unique drop and then disappear. Very interestingly, when the reaction mixture contained BCL7A (1-100) instead of BCL7A full length, the appearance of droplets was even more prominent (Figures 39, 40). More experiments need to be performed to investigate this phenomenon.
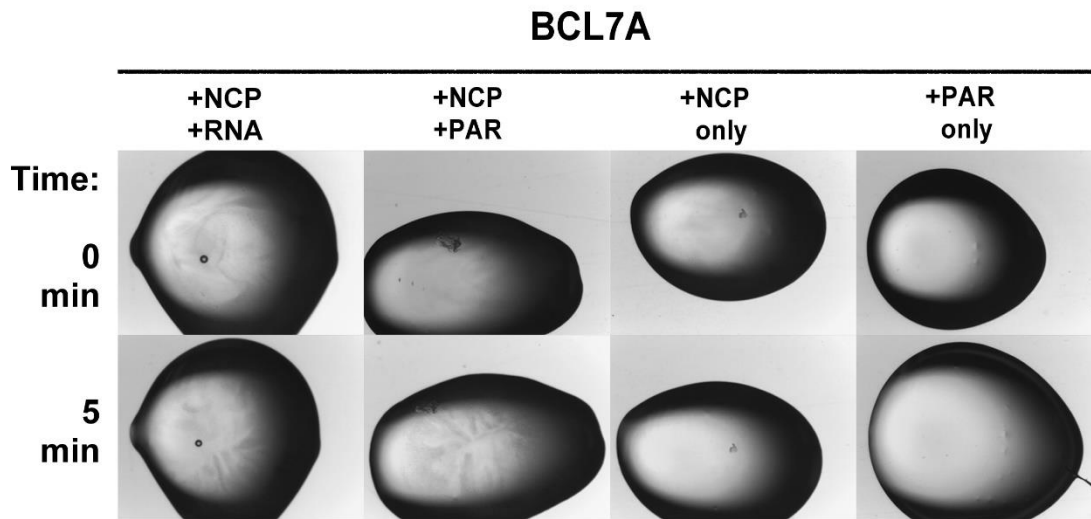
BCL7A

Figure 39. Phase separation observed with BCL7A.
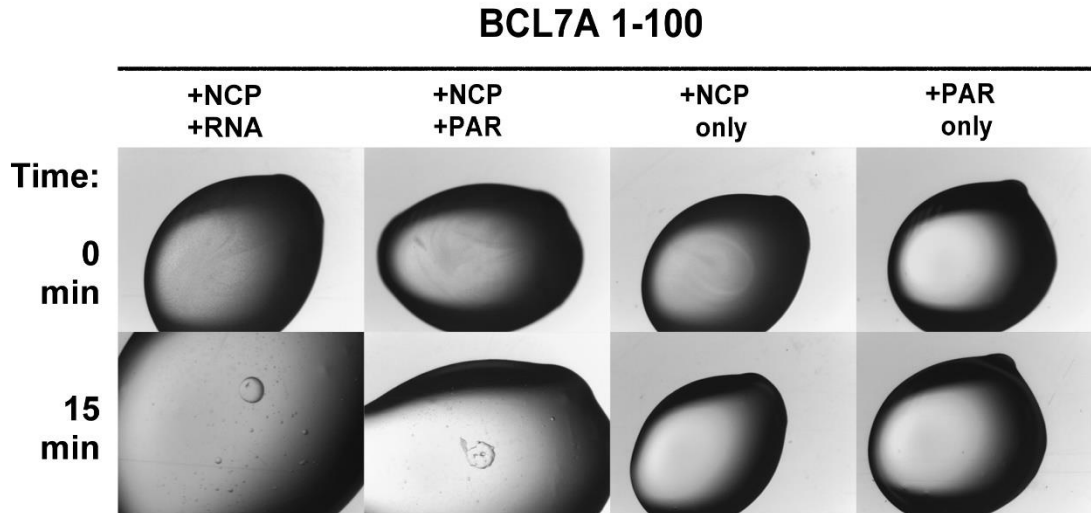
BCL7A 1-100

Figure 40. Phase separation of BCL7A 1-100.

3.2 Aim 2 Elucidate the 3D structure of BCL7 proteins in complex with binding partners

Before starting the co-crystallization project, we tried to crystallize the recombinant BCL7A, BCL7C, and shorter constructs of both proteins individually. After extensive crystallization trials, we were able to get crystals for BCL7A 1-100 construct (Figure 41) in the following conditions: 10% PEG 20K; 2% 2,4 Dioxane; 0.1M Bicine pH 9.0 at 6mg/mL. The crystals were very fragile and disappear after a couple of hours; several strategies employed to reproduce crystals; for instance, microbatch under oil trials were performed with no success.
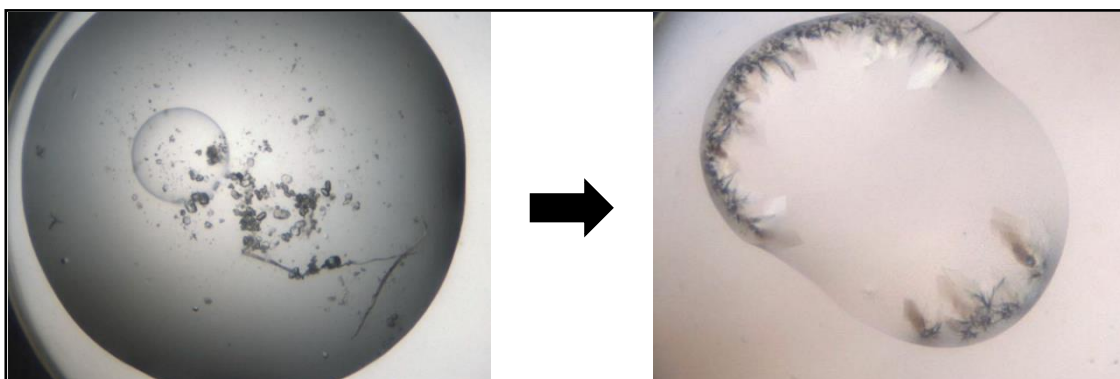
Figure 41. Initial Crystallization trials of BCL7A (1-100).

Crystals of BCL7A 1-100 in the original conditions and how crystals were obtained after the additive screening with 30% 1,8 diaminooctane at 6mg/ml, the protein concentration was crucial; however, these crystals disappeared after a couple of hours.

In order to elucidate the 3D structure of the BCL7 proteins in complex with binding partners, we used two different methods, X-ray crystallography and Cryo-EM. We were able to purify different complexes, being the first one BCL7A full-length protein bound to the nucleosome, the molecular mass of the complex is 223 kDa that falls within a range suitable to be explored both with X-ray crystallography and Cryo-EM.

EMSA experiments confirmed that BCL7 proteins directly interact with NCP and naked DNA. Because this protein is a core subunit of the SWI/SNF complex, it makes sense to use the nucleosome as the binding partner that will allow us to determine the structure of the BCL7 proteins. We prepared large scale assembly of NCP producing a batch of nucleosomes at a concentration of ~ 30μM and incubated it with a 3-fold molar excess of full-length BCL7A protein. The molar excess of BCL7A was used to ensure that the nucleosome was saturated with the protein, avoiding a heterogeneous sample. The complex was isolated by size exclusion chromatography (Superose 6 Increase) (Figure 42). As control NCP alone was purified by size exclusion chromatography on the same column. When comparing the elution time of NCP alone versus NCP-BCL7A complex, we observed that the complex NCP-BCL7A elutes in an earlier peak corresponding to higher molecular weight species. Further characterization of the complex was performed by coomassie stained SDS-PAGE gel showing a pure complex with BCL7A and nucleosomes. EMSA confirmed complex integrity, as we can see that the NCP is shifted in the peak fractions.
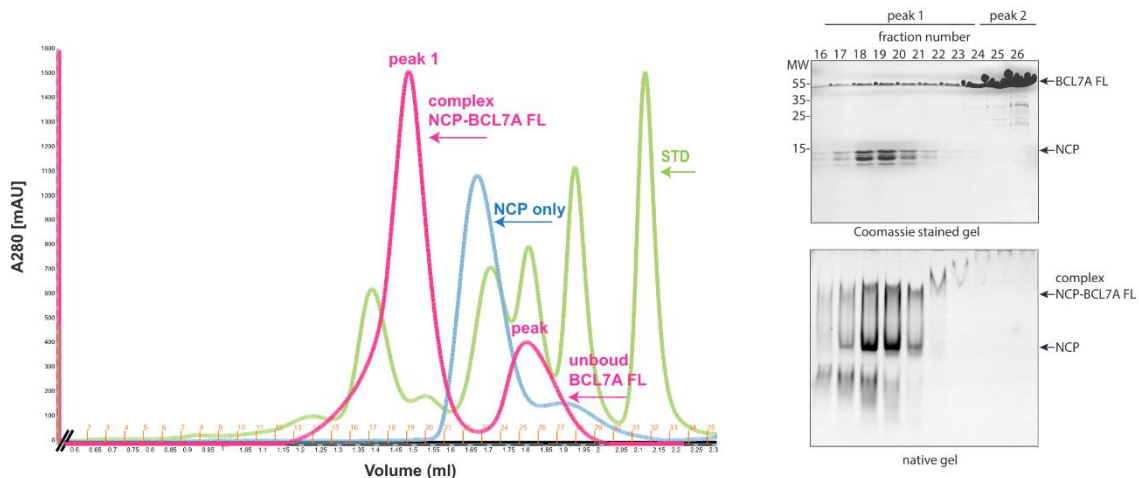
Figure 42. In vitro characterization of the NCP-BCL7A interaction.

BCL7A forms a stable complex with the nucleosome core particle (NCP). A) Size exclusion chromatography of BCL7A bound to NCP performed on a Superose 6 Increase column showing the shorter elution time of the BCL7A–NCP complex. The molecular ratio of BCL7A: NCP applied to the column was 3:1. Pink, the chromatogram of the BCL7A–NCP complex; blue line, the chromatogram of NCP in the absence of BCL7A; green line, the chromatogram of standards run on the same column. B) SDS-PAGE gel of the fractions from the chromatogram in part A showing the presence of BCL7A and histones. C) Native gel of the same fractions stained with RedSafe shows the nucleosomes' integrity; the mobility shift observed in fractions 21-–28 demonstrates the formation of the complex.

3.2.1 Characterization of the purified complex using Dynamic Light Scattering (DLS)

We performed measurements for free NCP, BCL7A, and the NCP-BCL7A complex (Figure 43). The estimated molecular weight for the NCP was 208 KDa, whereas for BCL7A proteins, in the absence of ligand, the measurements threw aberrant reading (Figure 43 B). Measurements of the BCL7A-NCP complex confirmed a good quality of the sample and the analysis yielded a molecular weight around 300kDa.
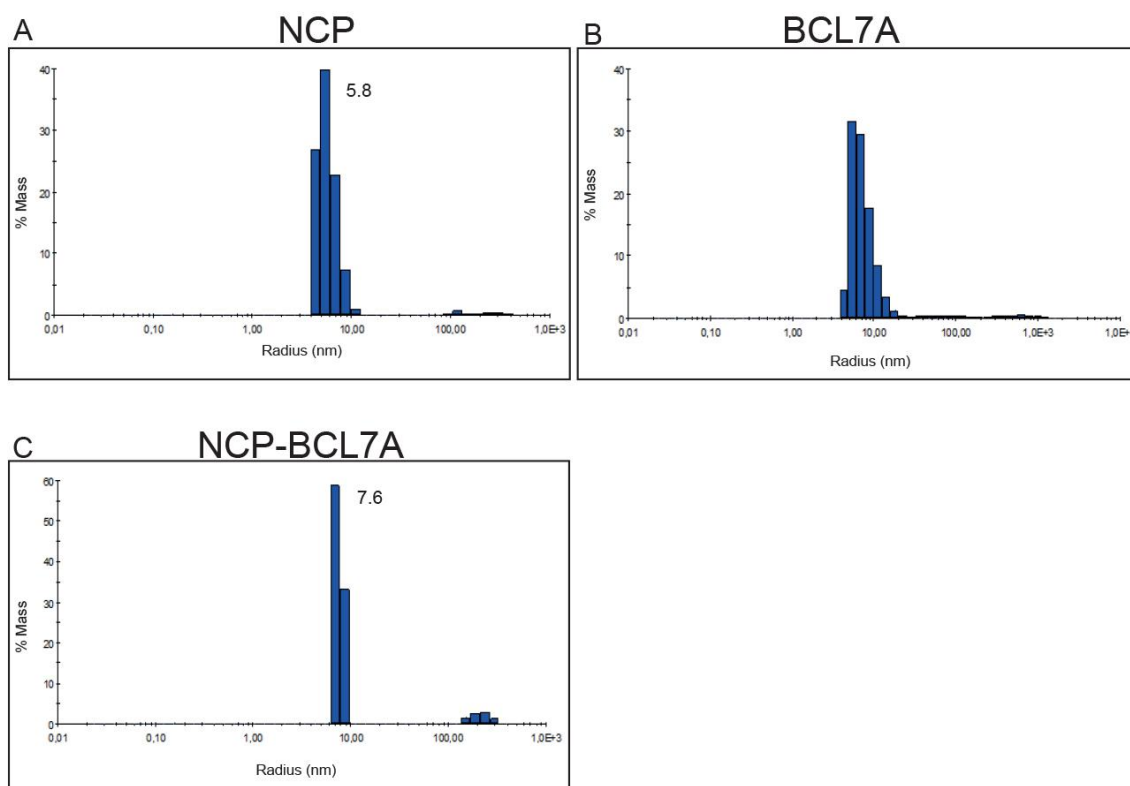
Figure 43. DLS profile of BCL7A and NCP.

A) NCP DLS profile shows a major peak with a radius of 5.8nm and an expected Molecular weight of 208 kDa.
B) A polydisperse peak with inconsistent molecular weight is seen for BCL7A protein.   C) Upon binding to NCP the DLS profile shows a major peak of 7.6 nm and an expected molecular weight of 300kDa.

3.2.2 Crystallization of BCL7A in complex with NCP.

The crystallization process of a chromatin complex is highly variable and depends on the individual complex properties. By using several commercial crystallization screens, we tested different ranges of pH, salts, and precipitants.

After the complex purification performed at a 3:1 ratio (BCL7A:NCP) in a SEC buffer consisting of 20mM Tris pH 8.0, NaCl 100mM, 1mM DTT, we performed initial screening of the complex at 2mg/mL in 96 wells plates using the Mosquito robot using a drop size of 100nL + 100 nL drops. We obtained crystals (Figure 44) in the following condition: 0.2M Sodium thiocyanate, 0.1M HEPES pH7, 15% v/v Pentaerythritol propoxylate (5/4 PO/OH).
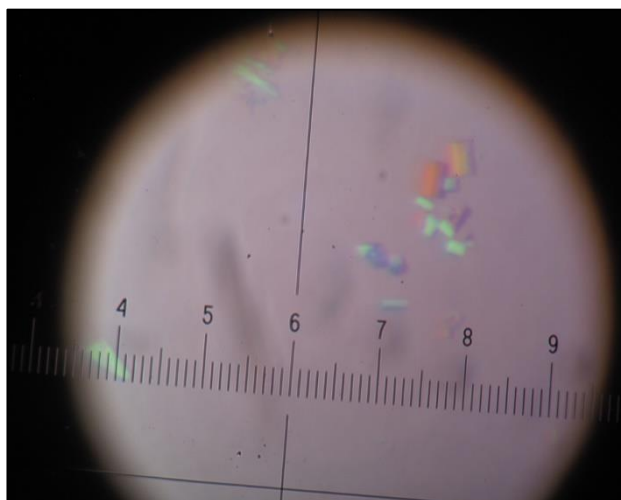
Figure 44. BCL7A-NCP crystals.

Original crystals obtained in 100+100 nL drops of the full-length BCL7A bound to the nucleosomes.

To obtain bigger crystals, we set 1μL+1μL bigger sitting drops. The bigger crystals screened in house diffracted to 20 Å resolution, and they were further sent to the synchrotron Soleil; the resolution did not improve (Figure 45).

Different cryo-protectant solutions were tested like PEG400, different Pentaerythritol propoxylate (5/4 PO/OH) concentrations (up to 55%), and different MPD concentrations. The best way to freeze the crystals was to soak them in a step wise manner in a cryo-protectant solution containing increasing amounts of MPD, from 10 to 30% MPD. As reported in the literature (Hanson et al., 2003), we decided to mount the crystals directly into the N_2 gas stream adjusted to 152 K. Then we performed annealing on the beam by lowering the gas stream temperature to 100K for data collection. The diffraction was poor and indexing was not possible. However, by measuring the distance between the weak diffraction spots, we could guess that the space group could be P61. This space group is the tetracyclic hexagonal.
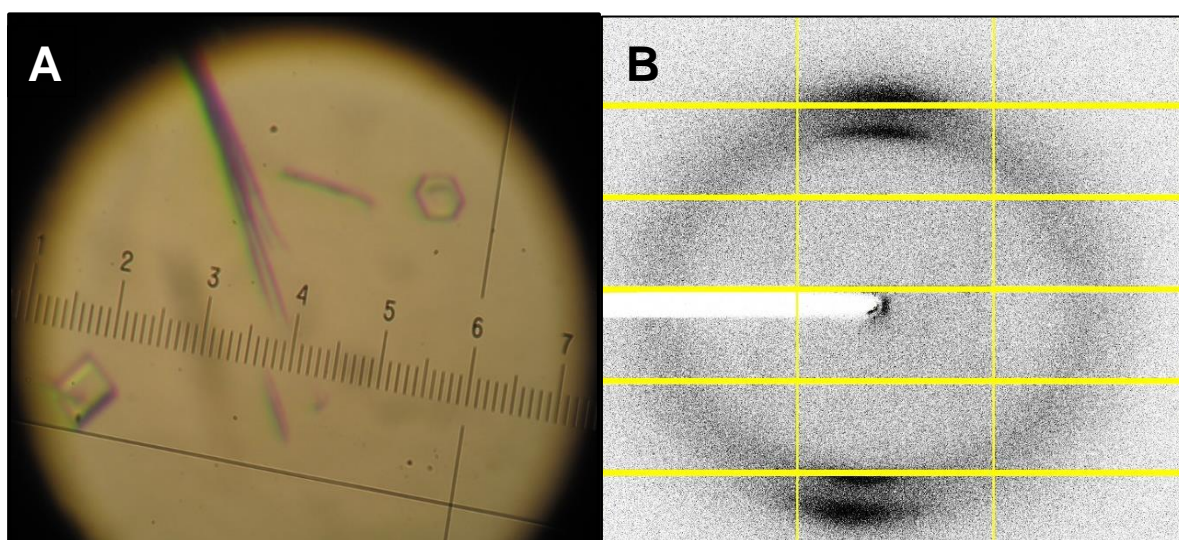
Figure 45. BCL7A-NCP complex crystals.

A) Full-length BCL7A-NCP crystals grown in 1μL drops.
B) Diffraction pattern BCL7A-NCP crystal

The crystals appearing in this condition had the shape of hexagonal bars (Figure 44, 45A); despite several attempts, crystals were difficult to reproduce. We further optimized this condition by setting 1 + 1 μL hanging drops in 24 well plates and adjusting the condition slightly. To induce re-appearance of the crystals in drops set up with freshly purified NCP-BCL7A complexes, we tried streak seeding and beads seeding, two techniques that prompt nucleation of crystals by introducing pre-formed nuclei into the crystallization drop. Reproducibility of the crystals was very challenging. We investigated if the crystals were co-crystals of NCP and BCL7A and contained indeed both proteins, some crystals were harvested, rinsed three times in mother liquor, then dissolved in water and run on a SDS-PAGE gel. Given the small amount of protein, proteins were detected by silver staining (Figure 46). We were able to confirm the presence of the BCL7 protein and the histones of the nucleosome. However we noticed the appearenace of a band corresponding to BCL7A running at a lower molecular weight (Figure 46) than expected, possibly due to some degradation events. When we run a drop where almost no crystals are found, the expected band is present, indicating that the event of this degradation may occur to form a complex that can crystallize. We thought that the crystals were obtained when BCL7A spontaneusly degrades to a truncated version of itself.

The corresponding band was sent to mass spectrometry, but silver staining interfered with the reading of the results, and we were not able to identify the region of the protein that corresponds to this lower molecular mass band.

Based on our biochemistry experiments, we know that the first 100 amino acids of BCL7A are sufficient for binding to the nucleosome. So we decided to isolate a stable complex with the nucleosome and a shorter construct of BCL7A (BCL7A 1-100).
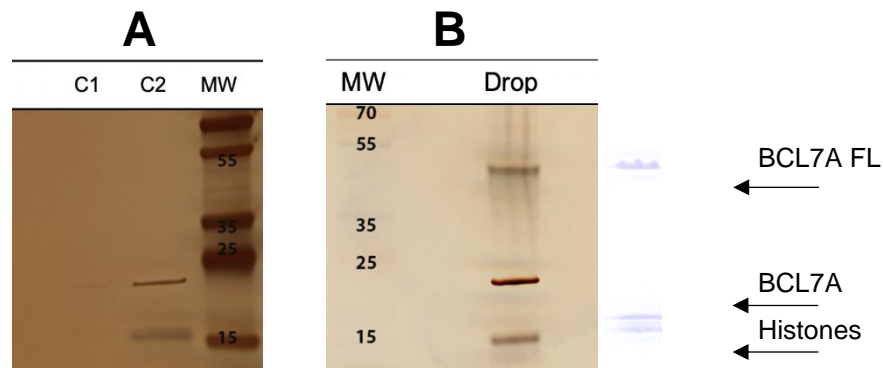
Figure 46. Silver stained SDS PAGE gels of BCL7A+NCP crystals.

A) C1 corresponding to crystal number one and C2 crystals number two in both cases, several bands can be observed, the highest band below the 25kDa band corresponding presumably to BCL7A and around the 15 kDa marker at least three bands corresponding to the histones can be observed. B) A second silver staining gel was performed. One drop coming from the crystallization trials that failed to produce crystals was resolved on SDS-PAGE and silver stained. Besides the histones two higher bands were detected. The band around 55 KDa corresponds to full length BCL7A, and the band at 35 KDa could corresponds to a degradation product coming from the full-length protein.

3.2.2.1 Molecular replacement

The crystal diffracted to a resolution of around 6.5 Å. (Figure 47). Data processing (Table-1) has been done using the software XDS. The crystal belonged to space group P2₁, the screw axis along b has been confirmed by looking at the systematic absences in the *0h0* plane, where all the *h* odd reflections are absent. Matthews coefficient predicts the presence of two molecules of nucleosomes in the asymmetric unit.
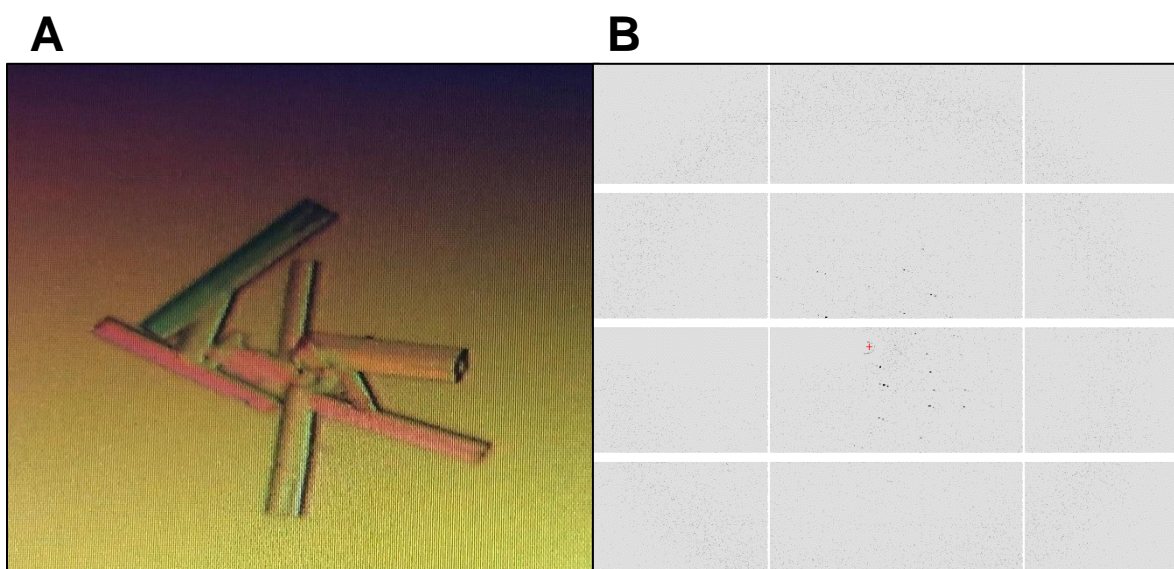
Figure 47. Analysis of BCL7A-NCP complex crystals.

A) Crystals of BCL7A-NCP complex sent to the synchrotron for data collection.
B) Diffraction pattern from the data collection

Table 1. Crystallography table

Resolution	47.71- 6.50 (7.27-6.50) Å
Space group	$P\,2_1$
Unit cell parameters	a=105.460 Å ; b=218.170 Å ; c=106.180 Å α=90° ; β=118.22° ; γ=90°
R_{meas}	0.14 (1.56)
1/ σ	5.0 (1.8)
Number of reflections	43045 (11204)
Completeness	99.5(99.2)
Redundancy	5.1
CC (1/2)	37%
R_{work} (%)	31
R_{free} (%)	34

90°
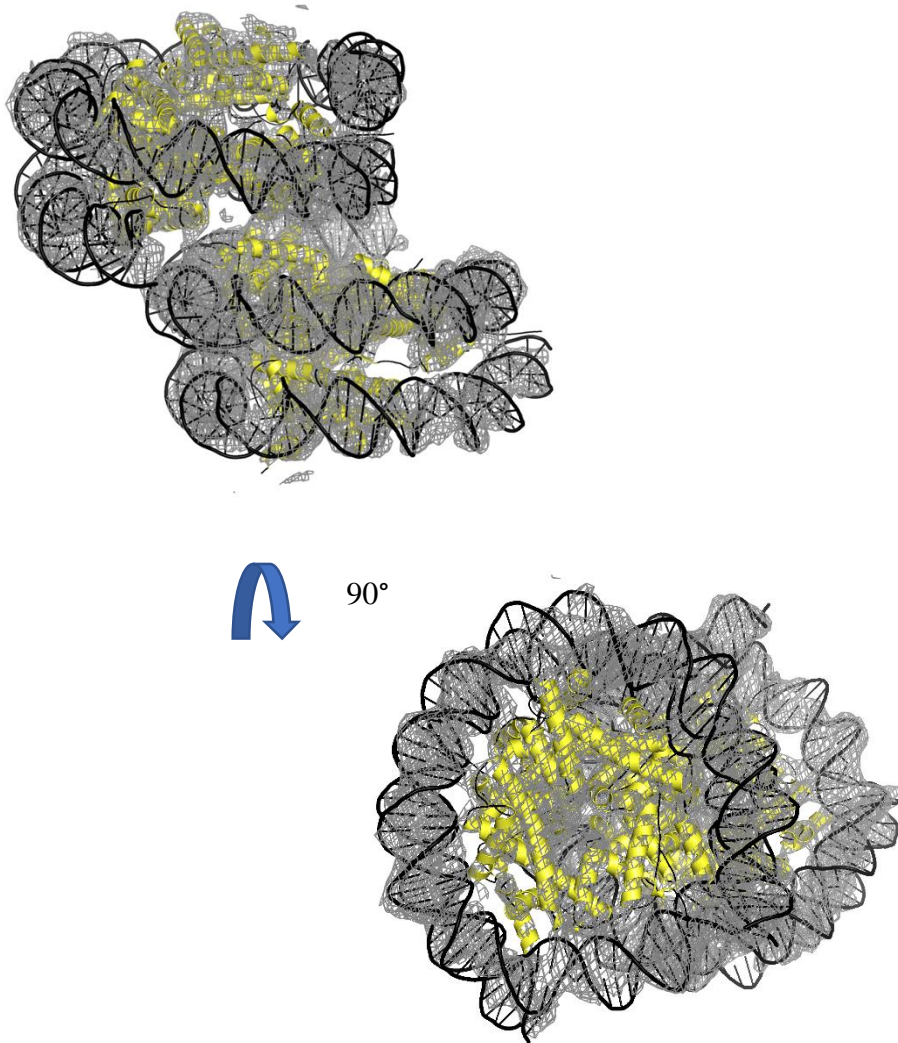
Figure 48. Asymmetric unit showing stack of 2 nucleosomes.
Stacking between the nucleosomes occurs via the octamer region. Octamer shown in yellow,
2Fo-Fc map shown in grey and DNA shown in black. B) Fig A rotated 90°, electron density
map fits well with the DNA as well as with the octamer.

Although crystals diffracted poorly, it was possible to index the data set and proceed with molecular replacement to solve the structure. Initial phases are obtained by molecular replacement using 3MGQ.pdb as search model using the program PHASER in CCP4. We have single solution from Phaser with a LLG score of 778. With the obtained solution, we performed rigid body refinement for 20 cycles using REFMAC5 in PHENIX and that yielded us an R_{work} of 31 % and R_{free} of 34%. Experimental map (2Fo-Fc) at contour level 1 and Fo-Fc map with contour level 3 is opened in coot to look for extra density, we could see the presence of two nucleosomes in the asymmetric unit stacked against one top of other through octamer-interface interactions (Figure 48) also we see that the density is fitting very well with the refined model in both the DNA as well as the octamer region, but we could not find any extra density which could correspond to approximately 25 kDa of BCL7A protein.

We decided to apply symmetry mates, showing that no extra density for the BCL7A protein is found: we could see the nice packing of the nucleosomes (Figure 49). Hence, we concluded that these crystals are empty nucleosomes. We have checked the region close to the octamer to see whether there is some extra density for the BCL7A and we could not any find extra density. One plausible explanation for the lack of BCL7A density in this crystal form is the tight interaction between the nucleosomes via the octamer, which could occlude the binding site for the BCL7A protein. The two nucleosomes repeat when the symmetry operator is turned on, and hence the crystal packing could also block the binding site for BCL7A protein.
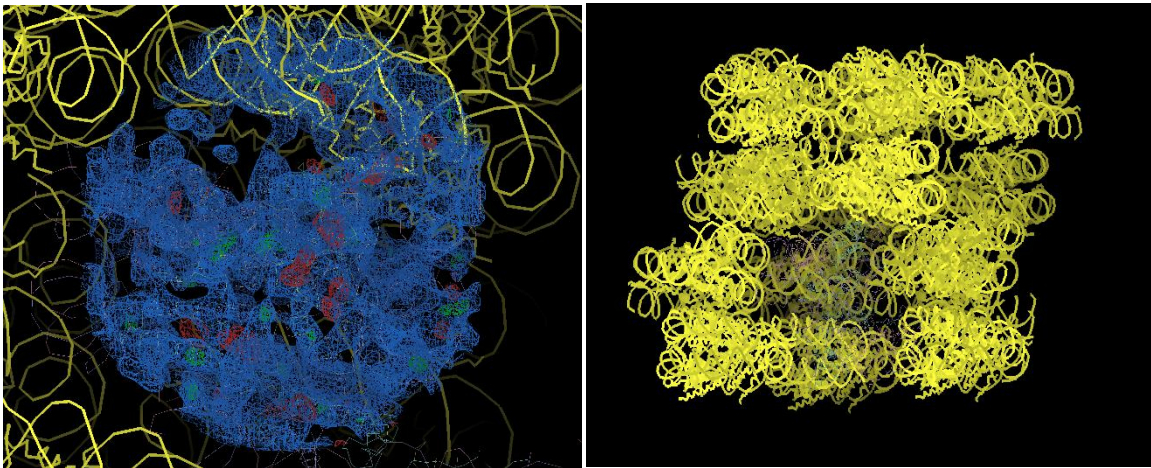
Figure 49. Analysis after molecular replacement.

A) Map resulting from molecular replacement, octamer region view and analysis, B) Symmetry mates analysis.

Crystals also suffered of radiation decay and when, we collected new data sets on other crystals (Figure 50) several diffraction patterns were of poor quality and we had to discard them leaving us with a small number of frames and a data set that could not be indexed and on which MR did not work.
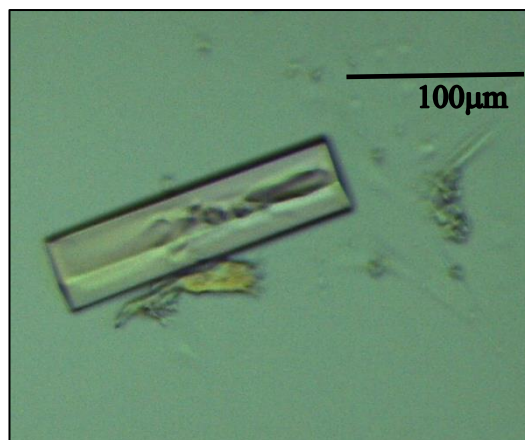
Figure 50. BCL7A-NCP crystal.

BCL7A-NCP crystal obtained in 0.2M Sodium thiocyanate, 0.1M HEPES pH7, 15% v/v Pentaerythritol propoxylate (5/4 PO/OH).

3.2.3 Crystallization of BCL7A 1-100 in complex with NCP

Considering that we detected a degraded form of BCL7A in the crystallization trials with NCP and BCL7A full-length, we decided to work with a shorter version of BCL7A encompassing the first 100 amino acids and shown to be sufficient to nucleosome binding according to our EMSA experiments (BCL7A 1-100).

A challenging problem when working with chromatin complexes is that nucleosomes can crystallize by themselves without the binding partner bound to it. As the production of large scale nucleosomes is an extremely laborious process, this is a problem of particular concern. To confirm the presence of BCL7A 1-100 bound to the nucleosome in the crystals, we followed and modified the method used in Son Tang laboratory (McGinty et al., 2016) that consists of the labelling of the chromatin protein with a fluorophore to be able to see the protein during the co-crystallization trials. The complex NCP-BCL7A(1-100) was isolated by size exclusion chromatography in a buffer consisting of: 20 mM Tris pH 8.0, 100mM NaCl; 1mM DTT. The same condition was used to purify the NCP- BCL7A full length complex. The protein was labelled by cross-linking the amino groups of the protein amino acids with 5-(and-6)-Carboxyrhodamine 6G, succinimidyl ester; the labelling did not interfere with the binding of the protein to the nucleosome as I could isolate a nice complex (Figure 51).
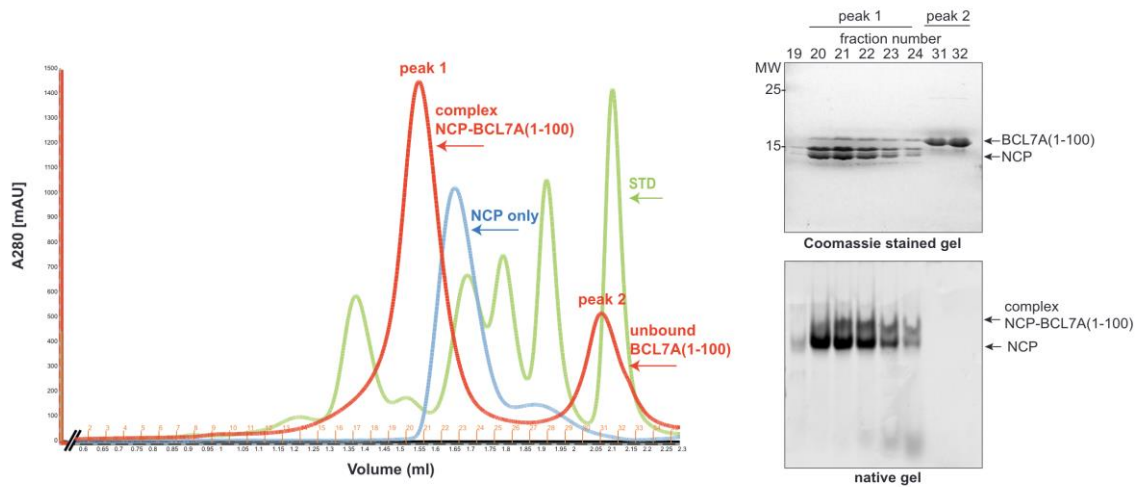
Figure 51. In vitro characterization of the NCP-BCL7A (1-100) interaction.

BCL7A 1-100 labeled protein forms a stable complex with the nucleosome core particle (NCP). A) SDS-PAGE of the fractions from the chromatogram in part A showing the presence of BCL7A 1-100 and histones. B) Native gel of the same fractions showing the integrity of the nucleosomes; the mobility shift observed in fractions 20-–24 demonstrates the formation of the complex. C) Size exclusion chromatography of BCL7A 1-100 bound to NCP performed on a Superose 6 Increase. Redline, the chromatogram of the BCL7A–NCP complex; blue line, the chromatogram of NCP alone; green line, the chromatogram of standards run on the same column.

We performed two complex purifications simultaneously, with labeled and unlabeled BCL7A 1-100. We set up crystallization trials at 20 °C in 96 wells sitting drops MRC2 plates using the Mosquito robot and different commercial crystallization kits. Crystals were obtained in several conditions in the following kits: Nucleix (Qiagen), LMB (Molecular Dimensions), Midas (Molecular Dimensions). Specific composition of crystallization conditions is listed in Table 2. Crystals grew after two days and stopped growing on the fifth day. The labelling did not interfere with crystal growth as we could obtain crystals with both the unlabelled and labelled protein. To confirm the presence of BCL7A 1-100 in the crystals, we observed the labelled and unlabelled crystals under a microscope equipped with proper filter to detect the fluorophore. We used a Nikon SM7 1500 stereoscope equipped with a Nikon Intensilight C-HGFI light source, a DsRed filter (Ex 545/30 nm, DM 570 nm, BA 620/60 nm), and a CoolSnapEZ Turbo 1394 camera. The crystals obtained with unlabeled protein were used as a negative control.

In figures 52 and 53, the crystals of the two most successful conditions can be observed. We got crystals in several other conditions (Table 2); however, we followed the best two conditions where crystals were bigger and easier to fish. The morphology of the crystals changed depending on the condition; for instance, a more conical shape is observed in conditions coming from Nucleix kit.
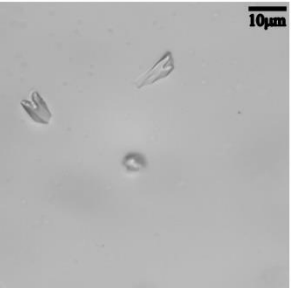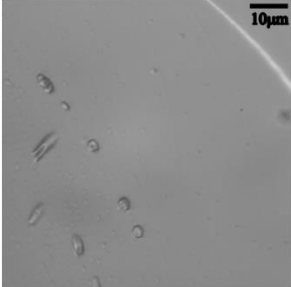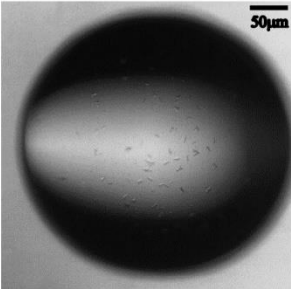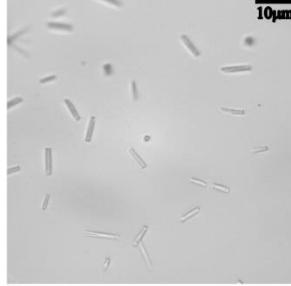
Kit	Condition	Crystals
Nucleix E7	Buffer: 50mM Sodium succinate 5.5 Additive: 1mM Cobalt chloride; 10mM Magnesium chloride Precipitant: 10% Isopropanol	
Nucleix E8	Buffer: 50mM Sodium cacodylate 6.0 Additive: 1mM Spermine; 20mM Magnesium chloride Precipitant: 15% ethanol	
Nucleix E9	Buffer: 50mM Sodium cacodylate 7.0 Additive: 1mM Spermine; 20mM Magnesium chloride; 1mM Cobalt chloride Precipitant: 15% Ethanol	
Midas E9	Buffer: 100mM Tris 8.0 Salt: 100mM Lithium sulfate Precipitant: 25% Jeffamine ED-2003	

Table 2 Crystallization conditions where crystals of the complex BCL7A (1-100) NCP were obtained, and the presence of the chromatin protein was confirmed.

CONDITION: 100mM Tris 8.4, Salt: 200mM Lithium sulfate, Precipitant: 40% PEG400
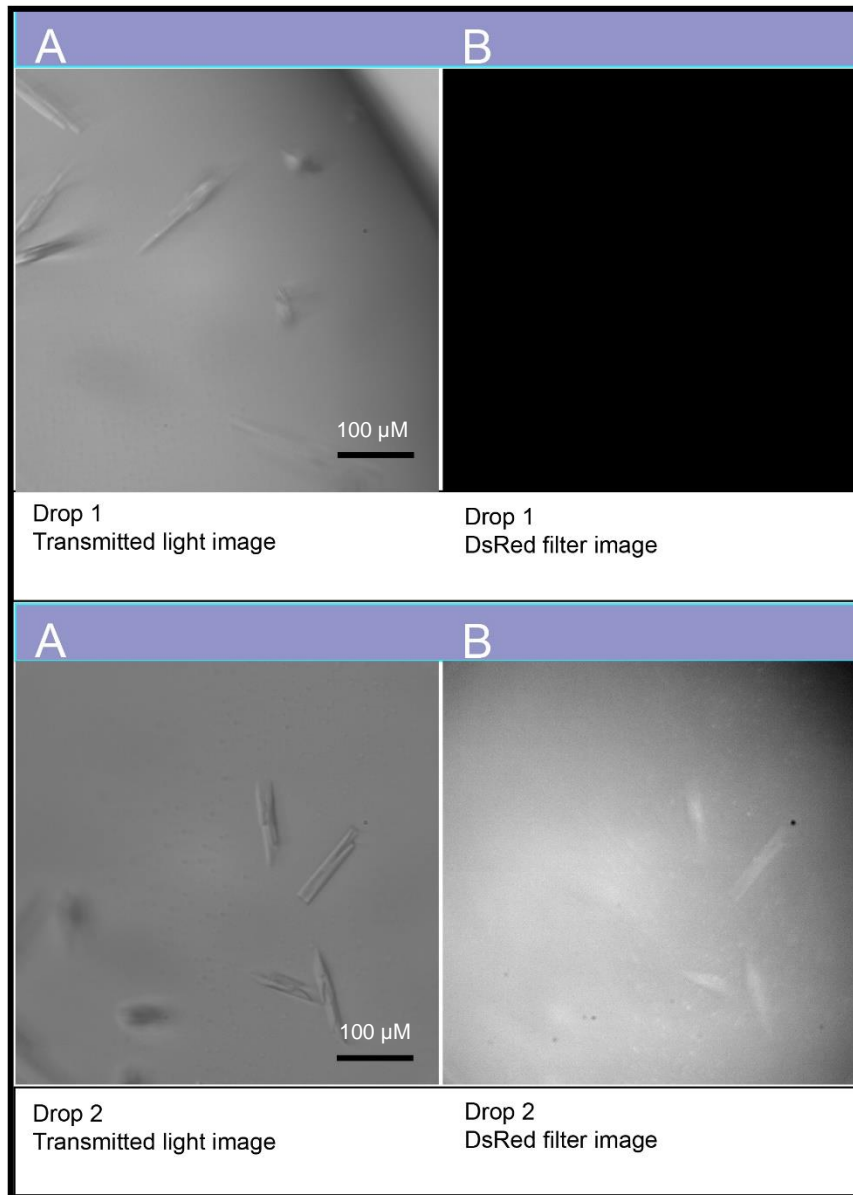
Figure 52. Crystals of BCL7A1-100_NCP complex.

A, B) Crystals obtained from unlabeled BCL7A (1-100)-NCP complex, left image taken with transmitted light, right image taken after using a DsRed filter image without any fluorescent signal detected. B, D) Crystals obtained from labeled BCL7A 1-100_NCP complex, left image taken with transmitted using a DsRed filter image where a fluorescent signal can be detected, indicating the presence of BCL7A(1-100).

CONDITION: 100mM Bis-Tris propane 7.1 Precipitant: 18% Ethanol
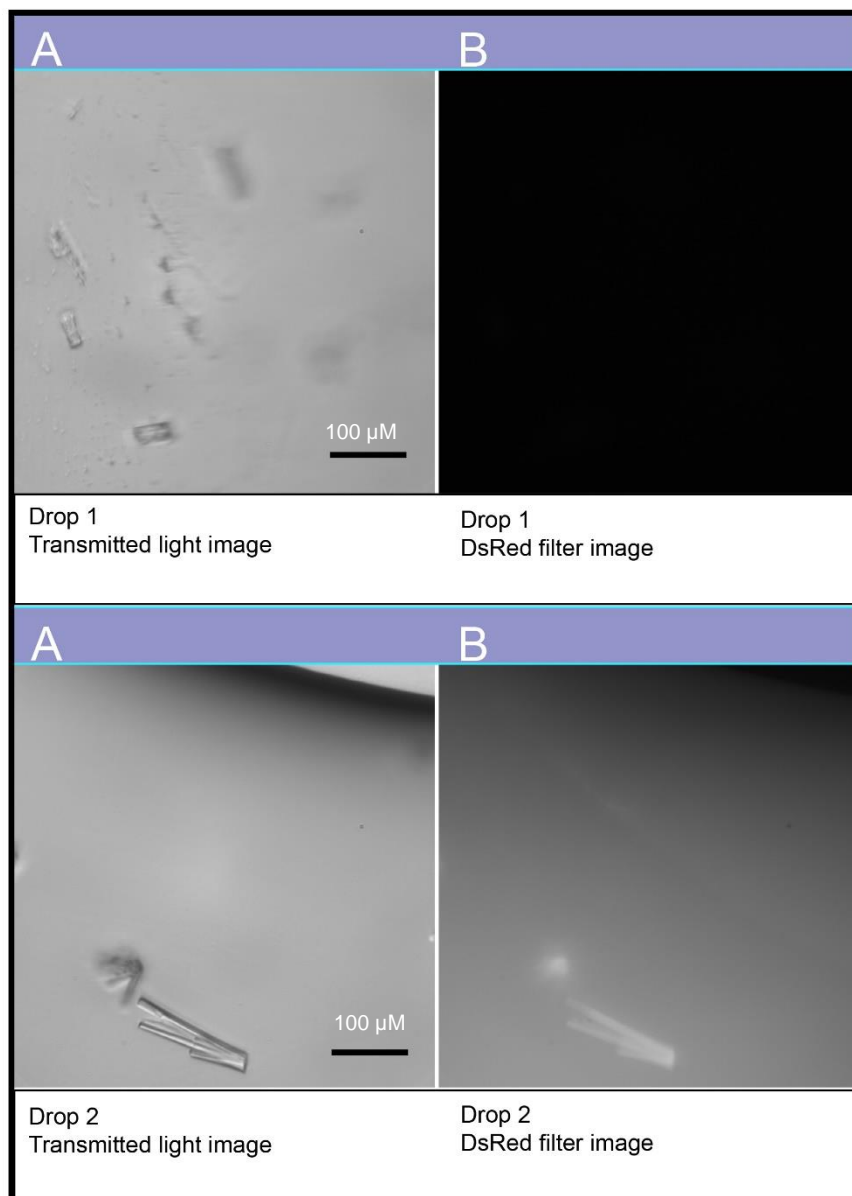
Figure 53. Crystals of BCL7A1-100_NCP complex.

A,B) Crystals obtained from unlabeled BCL7A (1-100)-NCP complex, left image taken with transmitted light, right image taken after using a DsRed filter image without any fluorescent signal detected. B,C) Crystals obtained from labeled BCL7A(1-100)-NCP complex, left image taken with transmitted light, right image taken after using a DsRed filter image where a fluorescent signal can be detected, indicating the presence of BCL7A using a DsRed filter image where a fluorescent signal can be detected, indicating the presence of BCL7A 1-100.

The crystals were very unstable and fragile; the smaller crystals disappeared very rapidly after opening the drop, and the bigger crystals, around 100 microns, broke during the attempts to fish them. However, some pieces of the crystals were sent These crystals were sent to Synchrotron Soleil, and we could collect two data sets that revealed two main issues: mosaicity and twinning. For the third data set, our molecular replacement failed. In the attempt to obtain better crystals, we decided then to perform additive screening in these two conditions, and the following are the list of most successful compounds (Figure 54):

Condition: 100mM Bis-Tris propane 7.1 Precipitant: 18% Ethanol
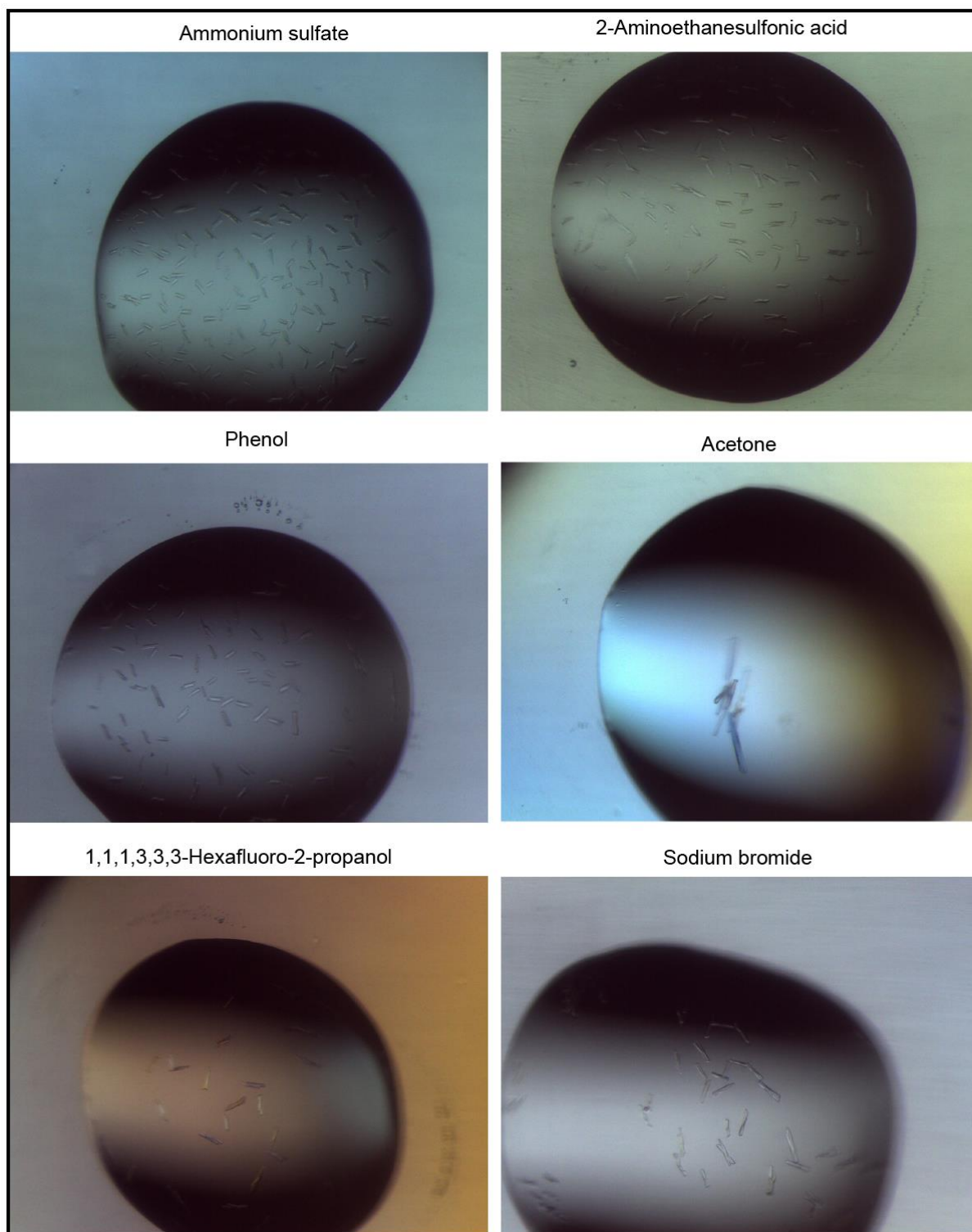
Figure 54. Crystals of BCL7A (1-100)-NCP complex obtained after additive screening.

The crystals presented several issues: they were multiple, fragile, and hollow. However, we were able to fish several crystals and screen them at the X-ray. The diffraction improved from 20 to 6 Å (Figure 55). Post-crystallization soak was particularly important. Transferring crystals from 20 to 4°C degrees gave us time to fish them and to test different cryo-protectants, like ethylene glycol, MPD, alcohols, and PEGs.
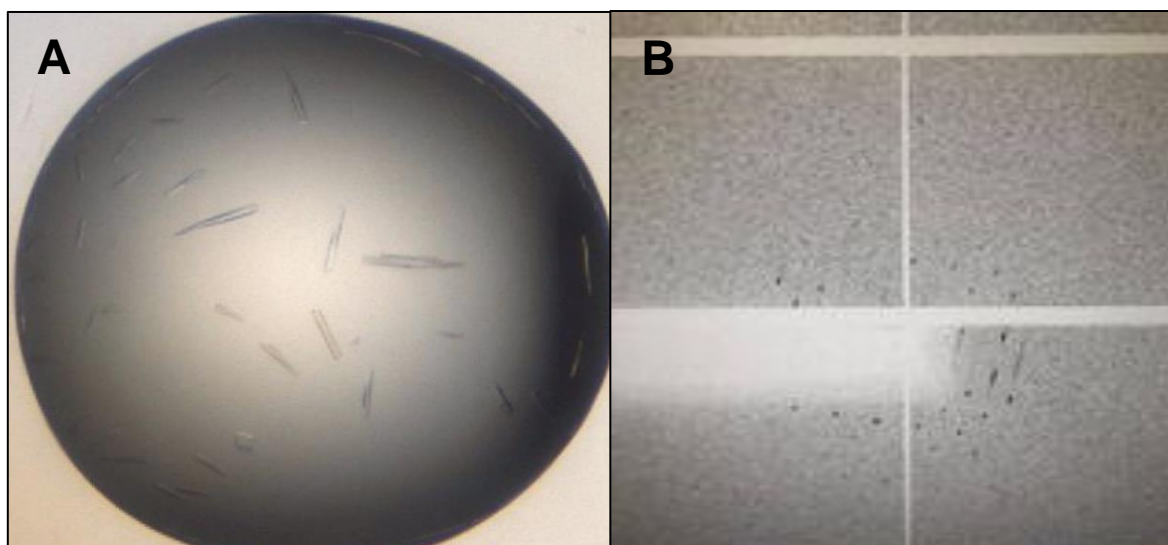
Figure 55. Crystals and diffraction pattern of BCL7A (1-100)-NCP complex.

A) Crystals obtained in 100mM Bis-Tris propane 7.1; 18% Ethanol and 50mM Sodium fluoride. B) Diffraction pattern of one of the crystals.

We obtained a data set but it had a serious twinning problem. Twinning occurs when two separate crystals share some of the same crystal lattice points in a symmetrical manner. Twinning does not produce simple diffraction patterns and molecular replacement failed to find a solution. Canonical de-twinning tricks did not work. Further crystallization attempts are conducted currently in the lab to obtain

better diffracting crystals, optimizing the cryo-protectant conditions, and screening crystals obtained with other species such as drosophila BCL7A.

3.2.4 Cryo-EM studies of BCL7A in complex with NCP.

For Cryo-EM studies, we reconstituted the BCL7A-NCP complex trying different buffers that could allow us to use different cross-linkers. The most common cross-linker used are glutaraldehyde and BS3. Another modification in the conditions was the salt. For crystallization trials, we used NaCl; however, for Cryo-EM, we decided to use KCl because it has been shown to help with the contrast of obtained images, and it is a more physiological salt.

3.2.4.1 Negative staining

We have successfully performed negative staining of the sample; this technique is frequently used to verify the concentration of the sample and helps to achieve the optimization of sample preparation for cryo-EM experiments. We used uranyl acetate as a colorant reagent, blotting the excess and seeing the integrity of the complex in a T20 electron microscope (Figure 56). We proceeded to start the optimization of the cryo-grids; for this, we tested different blotting times, blotting force, types of grids, use of detergents, etc., when freezing with the Vitrobot to improve the quality of ice thickness. Another test was made on the usage of cross-linkers like glutaraldehyde and BS3. Usually, for SWI/SNF complexes, the most common cross-linker used is BS3, but in our case, we got better results using 0.1% glutaraldehyde.
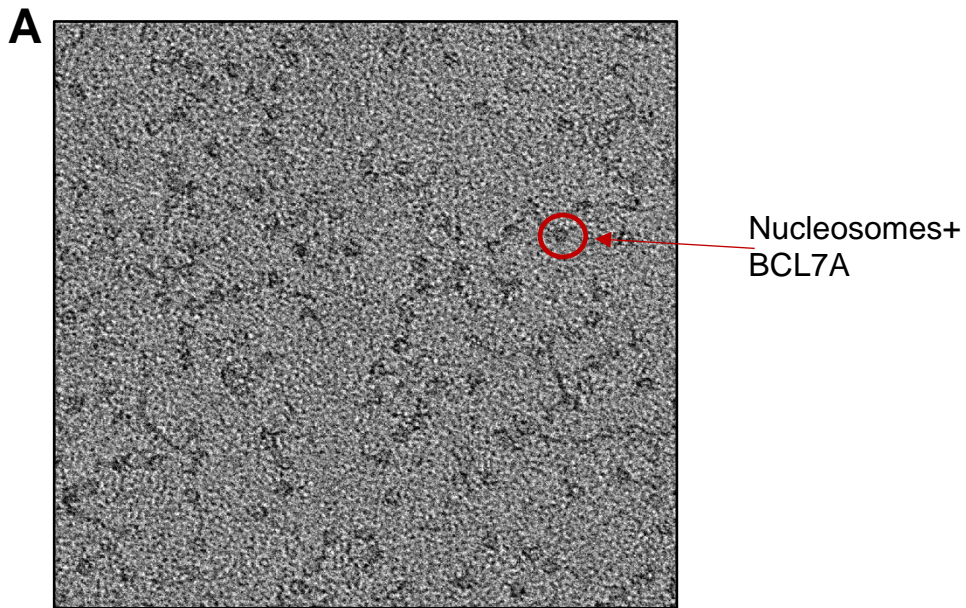
Figure 56. Negative staining of BCL7A-NCP complex.

nucleosomes are marked on the image, the concentration of the complex is 2mg/mL.

A brief summary of the conditions tried for this complex is in table 3. Different blotting times, force, the different cross-linkers (Figure 57) and detergents tested individually and in combination are noted.

At this time, we had collected images of two grids on the Titan Krios cryo-electron microscope, the complex as it is (Figure 58A), and a sample with 0.1% of Glutaraldehyde (Figure 58B). These grids were made with the following settings: blotting time four and blotting force 5 using Quantifoil R 2/2 grids.
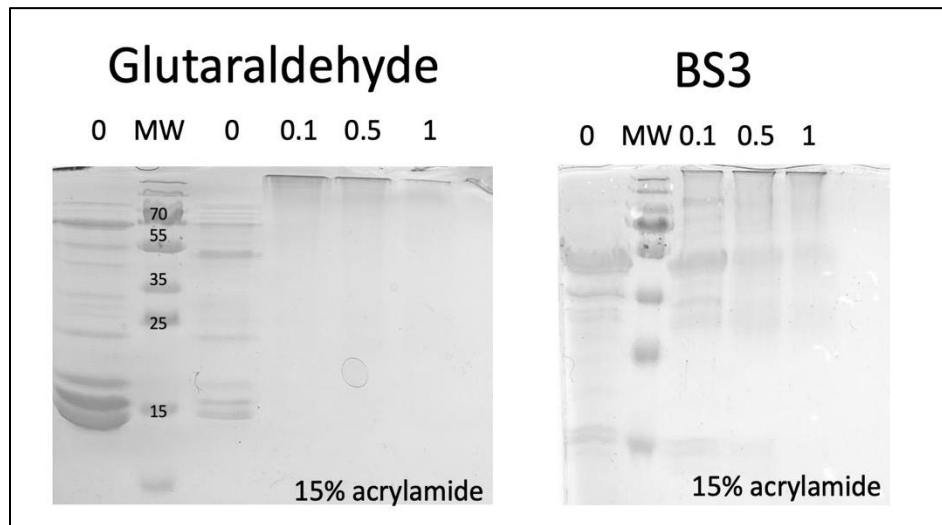
Figure 57. BCL7A-NCP complex crosslinked test.

Coomassie staining SDS PAGE gels showing the crosslinked BCL7A-NCP in the presence of Glutaraldehyde and BS3.

The particle distribution for the sample without a cross-linker was better than the one with 0.1% of glutaraldehyde; however, after data processing, we solved a high-resolution nucleosome (3.7 Å) but not density for BCL7A was visible. The sample with cross-linker looks crowded, but we could collect a data set at the Titan Krios cryo-electron microscope.
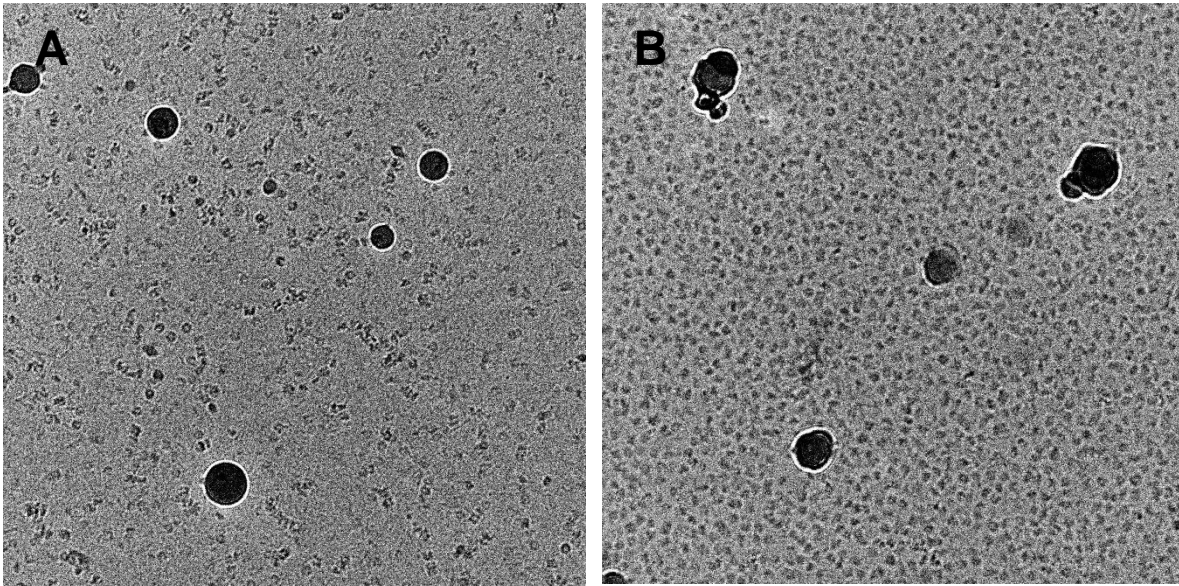
Figure 58. Cryo-EM grids BCL7A-NCP complex

Micrographs of BCL7A-NCP without crosslinker (A) and in the presence of 0.1% of crosslinker (B)

#Grid	Blot force	Blot time	Condition
1	5	3	
2	5	4	
3	5	3	0.1% Glut
4	5	4	0.1% Glut
5	5	4	0.01% NP40
6	5	3	0.01% NP40
7	5	3	0.1% Glut, 0.01% NP40
8	5	4	0.1% Glut, 0.01% NP40

Table 3. Different conditions tested for the elaboration of Cryo-EM grids

After data processing in Relion and Cryosparc (Figure 60), we obtained a cryo-EM density at 4.41 Å (Figure 59), and this density appeared to be only the nucleosome, as we could not detect the extra density corresponding to BCL7A.
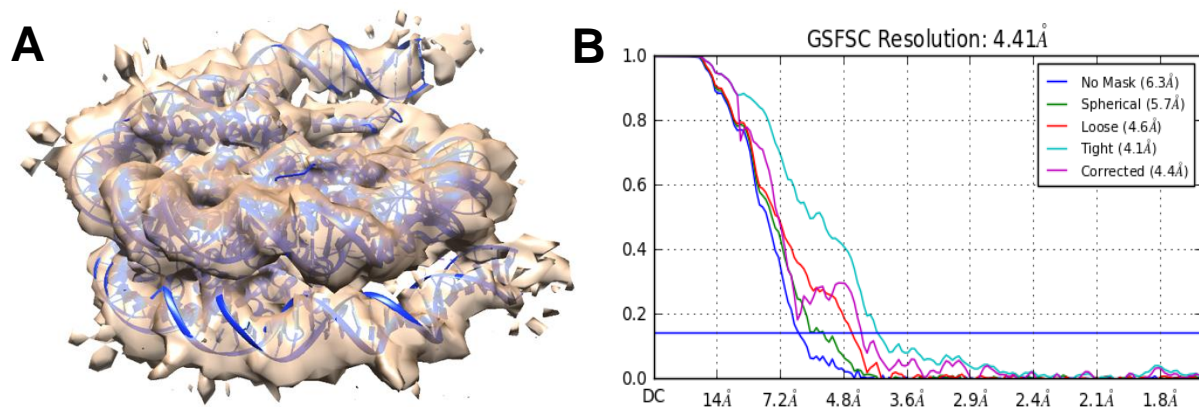
Figure 59. Structure obtained of BCL7A-NCP.

(A) The estimated resolution obtained in Cryosparc (B)
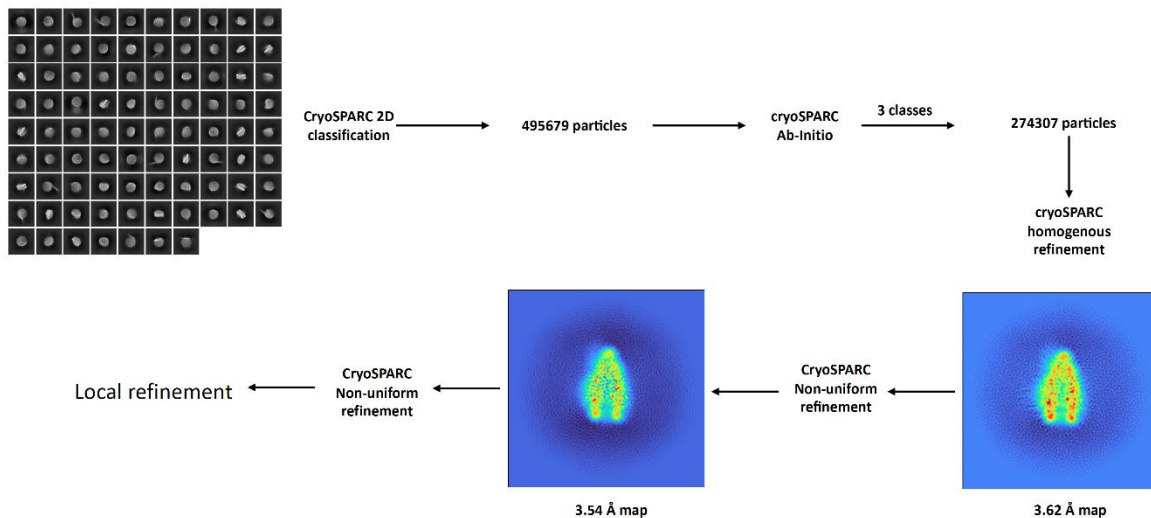
Figure 60. Flow chart describing the data processing in cryoSPARC.

3.2.5 Cryo-EM studies of BCL7C-BAF47 in complex with NCP.

Pull-down assays were used to define binary interactions between BCL7A, BCL7C, and binding partners such as BAF47. We were able to obtain a stable complex of full-length BCL7C and BAF47 by using proteins overexpressing BCL7C as GST tagged and BAF47 as sumo tagged protein, respectively. The complex was purified by double pull-down, followed by size exclusion chromatography (Figure 61).

Further because it is known that both proteins bind the nucleosome, we tested the complex ability to bind the nucleosome and isolate the ternary complex by size exclusion chromatography (Figure 62).
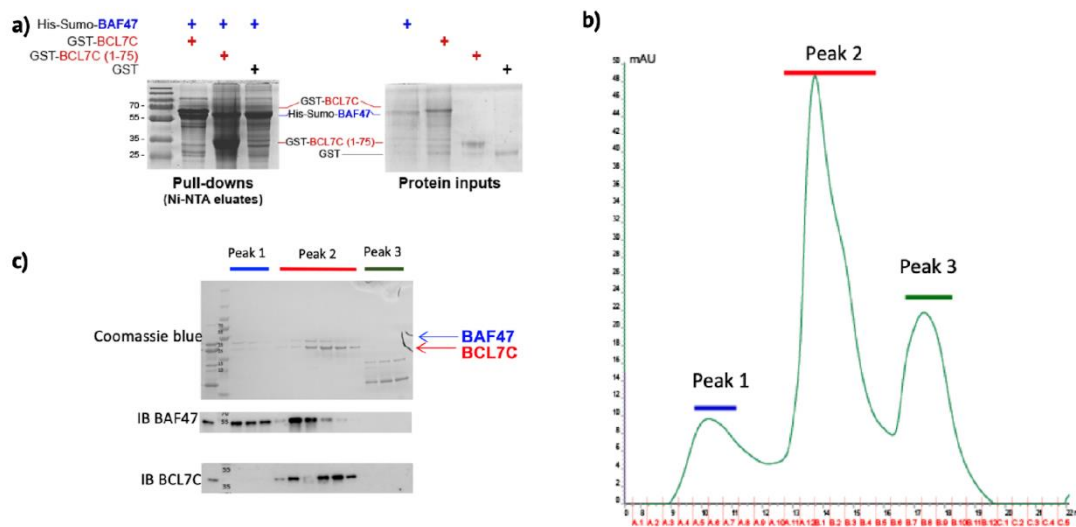
Figure 61. The identification of the BAF47-BCL7C complex.

A) **Pull-down assay that shows the binary interactions between BCL7C and BAF47.**
B) **Size-exclusion profile of the complex after it was purified by double pull down.**
C) **Western blot confirming the presence of both proteins.**
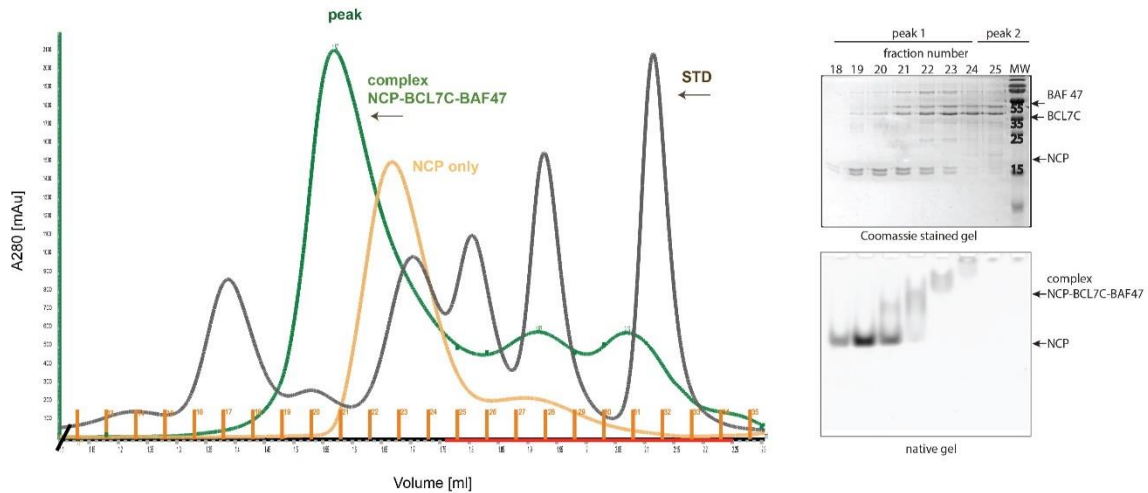
Figure 62. BCL7C-BAF47 binary complex binds to the nucleosome core particle (NCP).

A) Size exclusion chromatography of the ternary complex on a Superose 6 Increase. The green line, the chromatogram of the BCL7C–BAF47-NCP complex; orange line, the chromatogram of NCP alone; black line, the chromatogram of standards run on the same column.
B) SDS-PAGE of the fractions from the chromatogram in part A showing the presence of BCL7C, BAF47, and histones.
C) Native gel of the same fractions shows the nucleosomes' integrity; the mobility shift observed in fractions 22-–24 demonstrates the formation of the complex.

As shown in the figures 61 and 62, we were able to isolate BCL7C and BAF47 complex but also a stable BCL7C-BAF47-NCP complex, feasible to conduct cryo-EM studies. We prepared grids and collected two different data sets from this complex.
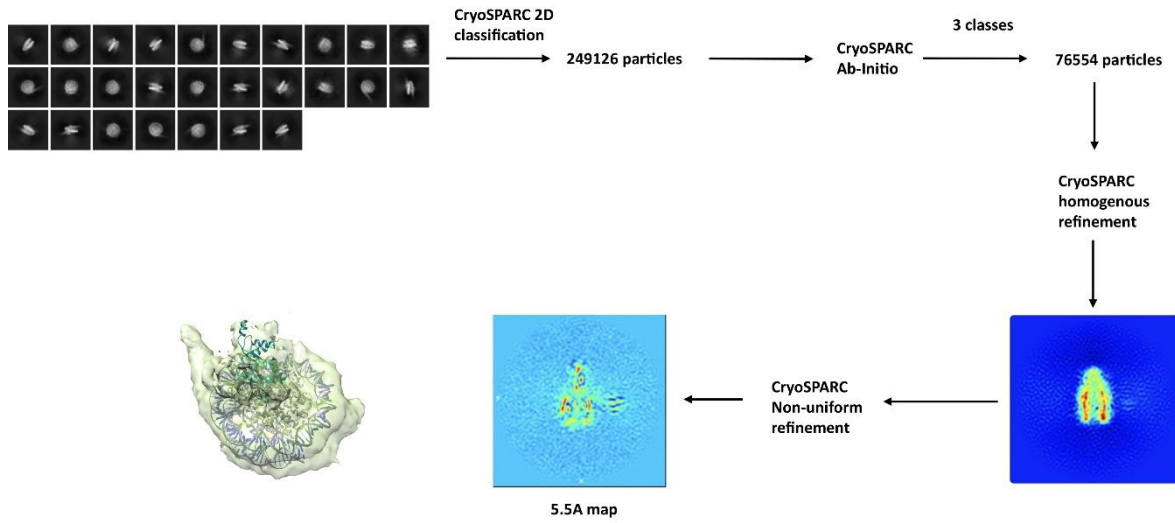
Figure 63. Flow chart data processing BCL7C-BAF47-NCP
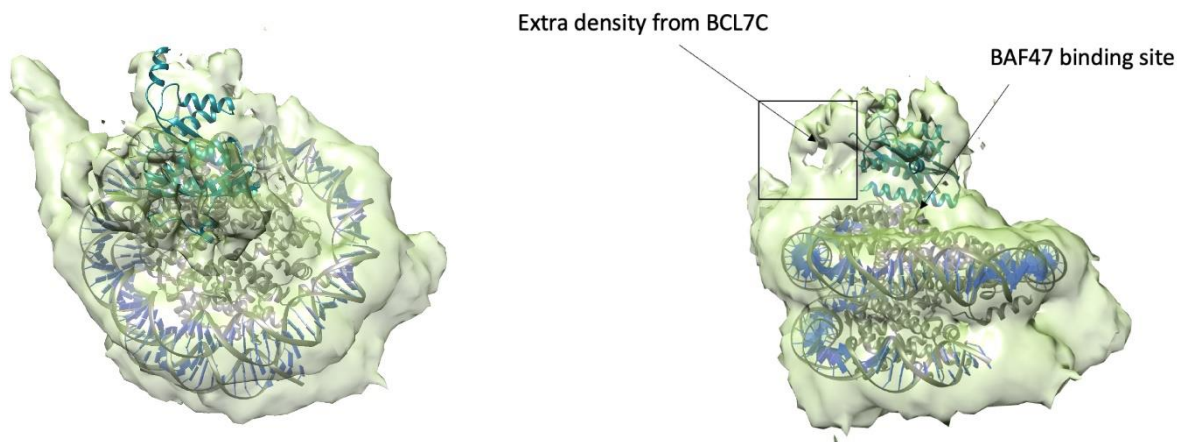
Figure 64. Density map BCL7C-BAF47-NCP.

After pre-processing steps were performed followed by several rounds of 2D classification, 3D reconstruction and local refinements (Figure 63) we obtained a density map of the ternary complex at 5.5 Å resolution (Figure 64). We were able to map BAF47, and we can see extra density corresponding to BCL7C, however, the resolution needs to be improved to be able to perform any building of the BCL7 molecule.

3.3 Aim 3 Evaluate the impact of cancer-derived mutations on BCL7 structure and function

To evaluate if BCL7A cancer mutations reported in the COSMIC database affect the binding of the molecule to the nucleosomes, we over-expressed and purified three different mutant proteins: BCL7A R11S, P78S and L210A. To make sure that the mutant proteins behaved like wild type proteins and were healthy proteins, we performed SEC-MALS experiments. SEC-MALS analysis of the mutants confirmed the correct folding of the protein, mono dispersity and provided the exact mass of the proteins (Figure 65).

The activity of these proteins was tested with EMSA assays performed with Cy5 labelled nucleosome. Figure 66 shows that the mutations R11S and P78S impair binding to the nucleosome, whereas L210A seems to have a smaller impact. In order to be more precise in the impact of cancer mutations, the images need to quantified to obtain apparent binding affinities.
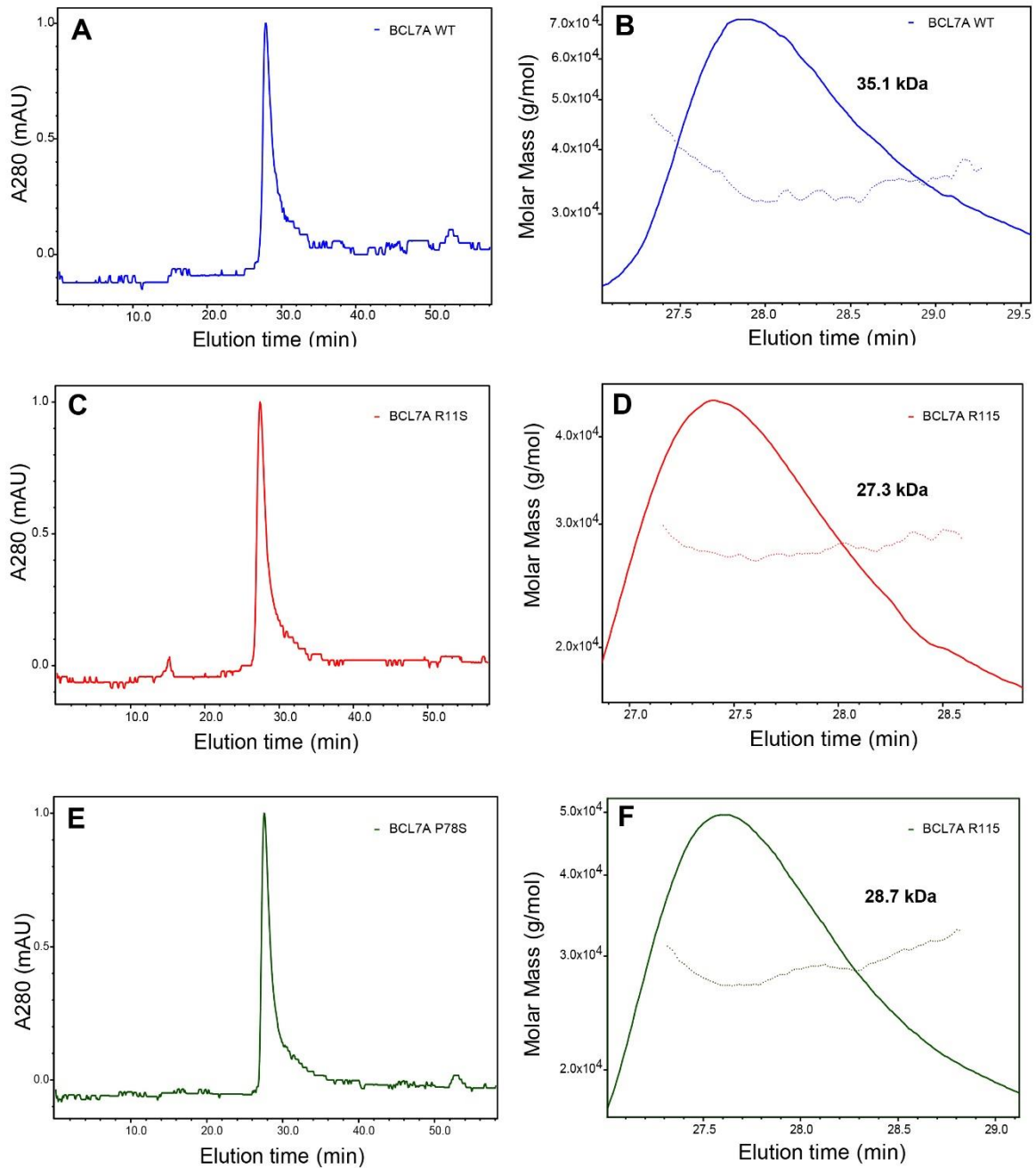
Figure 65. SEC-MALS analysis of BCL7A proteins using Superdex 200 column.

A),C) and E) Chromatograms of BCL7A WT, BCL7A R11S , and BCL7A P78S , respectively. B),D) ,and E) The three chromatograms exhibit only monomers. The monomer Molecular Weight values exhibit high homogeneity.
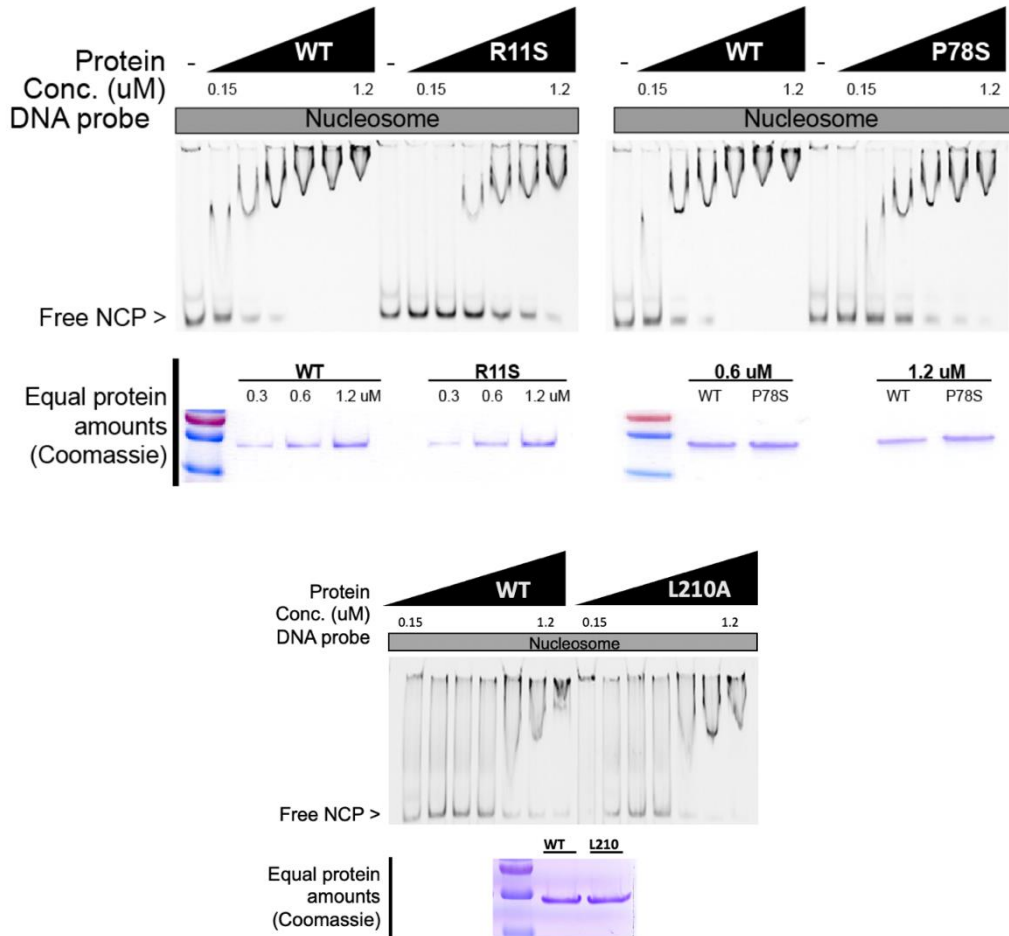
Figure 66. Mutations reported in cancer patients impair the binding of BCL7A to the nucleosome.

4 DISCUSSION

4.1 Determine the molecular function of BCL7 proteins

During the course of my PhD, I have studied human chromatin proteins that belong to the family of BCL7 proteins. I have worked with both BCLA and BCL7C, but I have focused mainly on BCL7A. Although BCL7 proteins have been discovered many years ago, it is only recently that they have been associated to the mammalian SWI/SNF(Kadoch et al., 2013a). They are part of the invariable core subunits of the complex, but their function in the context of the SWI/SNF complex and more generally in chromatin remodelling is not known.

When I started the project, the recombinant proteins were already cloned and basic overexpression and preliminary purification protocols were established. During the purification, large amounts of DNA were co-purifying with BCL7 proteins, suggesting that the proteins could bind DNA. Because we wanted to solve the three-dimensional structure of BCL7 proteins by X-ray crystallography and perform some biochemical studies to investigate the activity of the proteins, I have optimized the protein purification protocol to eliminate traces of DNA that could interfere with the homogeneity of the sample, affecting crystallization trials and the quality of in vitro experiments. An important step that I introduced in the purification protocol of BCL7 proteins was the ion exchange chromatography. BCL7A has a theoretical pI of 4.6, meanwhile BCL7C has a basic pI of 10. Thus, BCL7A and BCL7C were purified using anion exchange (Resource Q) and cation exchange (Resource S) chromatography respectively; With the same purpose, I have also tried heparin

affinity chromatography: this method was the most efficient at removing nucleic acids traces and allowed recovery of higher amounts of proteins.

Because the proteins co-eluted with significant amounts of DNA and are part of a chromatin remodeling complex, we tested the ability of the proteins to bind to the nucleosome and free DNA. Interestingly the proteins bind to both the nucleosome and free DNA. Using the analysis coming from the secondary structure prediction and sequence alignments, we designed different constructs that will help us identify the part of the protein responsible for the binding. We knew that the first 50 amino acids were conserved, so we suspected that this region could contain the functional domain; however, this part of the protein could not be expressed and purified. For BCL7A, we tested the first 100 amino acids that were sufficient to bind to the nucleosome. On the other hand, for BCL7C, we found that the first 100 and 75 amino acids were enough for nucleosome binding; however, the affinity of truncated versions of BCL7 proteins for the nucleosome decreases considerably.

For BCL7A, we went a step further, and we explored the function of the rest of the protein. It is important to remember that the C-terminal part of BCL7 proteins does not carry any secondary structure feature and it is predicted to be disordered. We speculate that the C-terminal part confers a peculiar behaviour to the protein; BCL7 proteins elute at higher predicted MW on size exclusion chromatography but their monomeric nature in solution was confirmed by SEC-MALS analysis. They also run at higher MW on SDS-PAGE gels. This behaviour is lost with the truncated versions of BCL7 proteins like BCL7A (1-100) or BCL7C (1-75). We discovered that the C-terminal part of the protein can bind the nucleosomes as well.

In order to better characterize the interaction BCL7A-NCP, we wondered what is the affinity of the protein for the nucleosome. We used different techniques to answer this question: microscale thermophoresis (MST) and gel shift assays (EMSA). Different trials were also made using isothermal titration calorimetry (ITC), but due to the large amount of sample required for the ITC experiments, we could not finalize the standardization of the ITC experiments. MST is a technique that requires smaller amounts of samples and it is based on the detection of temperature-induced changes in fluorescence of a target as function of the concentration of a non-fluorescent ligand.

We decided to set up the MST assay and we adopted two different strategies. In the first strategy the, BCL7 proteins were labelled with a fluorescent dye. However, we found out that the protein and its initial fluorescence were not stable enough. In the second strategy, we used fluorescent nucleosomes and unlabelled protein instead, which gave a stable fluorescent signal. Therefore, we followed this second strategy to perform the experiments. We then tested different buffers in order to optimize the assay. We have chosen a buffer composed to 20 mM Hepes pH 7.5, 100 mM KCl and 0.05% Tween. The addition of tween was critical to achieve good experimental conditions and stable signal. Once we established the conditions where the nucleosome and the protein were stable, we investigated at which concentrations of BCL7 and NCP it was best to perform the experiments. Generally lower nucleosome and protein concentrations worked better. The estimated Kd of BCL7A for NCP is ~230nM; this value is comparable with the binding affinities reported for other chromatin proteins that bind the nucleosome like the LANA peptide (Barbera et al., 2006). However, it is important to

remember that BCL7A interacts with other subunits of the SWI/SNF, and it is possible that interactions with other subunits alter the affinity of the protein for the nucleosome.

For a long time, IDPs' spectrums were considered random, with a minimum cantered at 200nm and the absence of strong negative signals characteristic to alpha-helix and/or beta-sheet structures; this was the case for BCL7A that has some secondary structure elements but is mostly unstructured. Proteins with these features are implicated in multiple macromolecular interactions, and they undergo structural transitions upon binding to the binding partner; however, these transitions can be challenging to study.

Further experiments can be performed with the different regions of the protein. It would be interesting to determine if the structured part is independent from the disordered region of if they are interacting, and more important it would be interesting to determine if the binding with the nucleosome or DNA make changes in the secondary structure composition.

4.2 Elucidate the 3D structure of BCL7 proteins in complex with binding partners

Crystallization of proteins containing IDR is a very a big challenge in structural biology. As mentioned before, they usually need to be stabilized by a binding partner to ensure proper folding; however, before initiating the co-crystallization trials we tried to crystallize the structured part of the proteins. In the case of BCL7A the first 100 amino acids gave hexagonal crystals that appeared two hours after setting the

drops, however these crystals disappeared only after a couple of hours. We then obtained co-crystal of BCL7A and the nucleosome but they diffracted poorly. In the attempt to try to improve crystal diffraction we explored several crystallization conditions, additives like monovalent and divalent salts, different precipitants, variation of pH and crystallization geometries, sitting and hanging drops. We tried optimization also of cryo-protectants and crystal dehydration and annealing techniques. A famous crystals structure of an enzyme bound to the nucleosome is the structure of RCC1-NCP complex (McGinty & Tan, 2016). The diffraction of crystals of this complex dramatically improved by soaking the crystals in different dehydrating solutions. The use of orthologs species of the protein may help to improve the quality of the crystal as well, as it has been shown for the NCP-RCC1 complex.

To determine the three-dimensional structure of the BCL7A-NCP complex we conducted at the same time cryo-EM studies. Sample preparation was optimized for grids preparation. We investigated in which conditions the complex maintain its stability by testing different buffers and salt conditions, at the same time we tested if the usage of cross-linkers and detergent was needed. At the same time, we tested different blotting times and blotting forces while freezing the grids with the Vitrobot system to determine the conditions that allow to obtain a layer of ice of a good thickness ideal for data collection. The grids were screened at the Titan Kryos or Glacios microscope and we could find the best condition that showed a good integrity of the sample and the nucleosomes. We were able to perform three different complex preparation, NCP-BCL7A, NCP-BAF47-BCL7C and NCP-BAF47-BCL7A-BRG1 and perform four different rounds of data collection. Data processing

for the NCP-BCL7A complex lead to the 3D reconstruction of empty nucleosomes. We could not see the density of the BCL7 proteins. We don't know if it is due to the fact that a large part of the BCL7 protein is disordered and moves even when the molecule is bound to the nucleosome or if it due to the fact that the freezing of the sample breaks the complex resulting into a low occupancy of BCL7 on the nucleosome. In this case larger data sets would need to be acquired to be able to have enough particles containing the complex. Data processing of the NCP-BCL7C-BAF47 complex lead to the 3D reconstruction of the nucleosome with visible extra density bound to it at a resolution of ~5 Å. We were able to fit into the cryo-EM density the 3D structures of BAF47 and the nucleosome; we can see an extra density corresponding to BCL7C. However, the resolution is not high enough to allow building of BCL7 into the density. More particles need to be averaged in to improve the resolution, thus more data acquisition is needed. We can see that BCL7 interacts extensively with the nucleosome and BAF47, but the molecular details of interactions are not visible yet. We also followed up by creating a larger sub complex containing also BRG1 with the hope that the ATPase subunit may stabilize BCL7 even further. Data processing of data obtained from these sample is being continued in the lab.

4.3 Aim 3 Evaluate the impact of cancer-derived mutations on BCL7 structure and function

We investigated the impact of mutations reported in cancer patients on BLC7 ability to interact with the nucleosome and discovered that these cancer mutations indeed impair binding to the nucleosome. As the structure of BCL7 proteins is not available, it is impossible to map the mutations on the structure to rationalize them. When the

structure will become available, we will have a better insight on the mechanism of action of these proteins. It would be also interesting to test the impact of mutations present in the disordered part of the molecule.

5 CONCLUSION

BCL7 proteins contain a small ordered N-terminal motif and a larger C-terminal tail that is disordered. They are part of the mSWI/SNF complex but little is known about their function and their contribution to chromatin remodelling. In this work, we demonstrate that the three members of this family, BCL7A, BCL7B and BCL7C, interact with free DNA and with the nucleosome with high affinity. We identified that the conserved N-terminal region of the proteins, that is highly conserved, is sufficient for this nucleosome binding function. Interestingly, we show that BCL7A C-terminal region also contributes to the binding of the nucleosome.

We studied three different point mutations reported in the COSMIC data base for BCL7A, R11, P78 and L210. The first two mutations located in the conserved N-terminal region showed impaired binding to the nucleosome, however the mutation in the C-terminal appear to have a milder effect. It is possible that BCL7A binds the nucleosome though the conserved N-terminal domain and that the rest of the molecule helps stabilizing the interaction. This idea is supported by our DLS experiments. BCL7A DLS spectra shows heterogeneity and a high molecular weight due to aggregation or the lack of secondary structure. Instead the complex BCL7A-NCP displays a more ordered spectrum.

We were able to identify BAF47 as binding partner of BCL7C. BAF47 also binds the nucleosome and BAF47-NCP-BCL7C form a stable ternary complex. We speculate that the binding of both proteins stabilizes and enhances the interactions with the nucleosome. Further studies await to be conducted to obtain the 3D structure of the BCL7 proteins and to further characterize their mode of action.

Bibliography

A, P., & Weber, S. C. (2019). Evidence for and against Liquid-Liquid Phase Separation in the Nucleus. *Non-Coding RNA*, *5*(4), 50. https://doi.org/10.3390/ncrna5040050

Agalioti, T., Chen, G., & Thanos, D. (2002). Deciphering the transcriptional histone acetylation code for a human gene. *Cell*, *111*(3), 381–392. https://doi.org/10.1016/s0092-8674(02)01077-2

Ahmad, K., & Henikoff, S. (2002). Histone H3 variants specify modes of chromatin assembly. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(Suppl 4), 16477–16484. https://doi.org/10.1073/pnas.172403699

Alberti, S. (2017). Phase separation in biology. *Current Biology*, *27*(20), R1097–R1102.

Allfrey, V. G., Faulkner, R., & Mirsky, A. E. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *51*(5), 786.

Alva, V., Ammelburg, M., Söding, J., & Lupas, A. N. (2007). On the origin of the histone fold. *BMC Structural Biology*, *7*(1), 17. https://doi.org/10.1186/1472-6807-7-17

Bagchi, A., Papazoglu, C., Wu, Y., Capurso, D., Brodt, M., Francis, D., Bredel, M., Vogel, H., & Mills, A. A. (2007). CHD5 is a tumor suppressor at human 1p36. *Cell*, *128*(3), 459–475. https://doi.org/10.1016/j.cell.2006.11.052

Baldi, S., Korber, P., & Becker, P. B. (2020). Beads on a string—Nucleosome array arrangements and folding of the chromatin fiber. *Nature Structural & Molecular Biology*, *27*(2), 109–118. https://doi.org/10.1038/s41594-019-0368-x

Baliñas-Gavira, C., Rodríguez, M. I., Andrades, A., Cuadros, M., Álvarez-Pérez, J. C., Álvarez-Prado, Á. F., de Yébenes, V. G., Sánchez-Hernández, S., Fernández-Vigo, E., Muñoz, J., Martín, F., Ramiro, A. R., Martínez-Climent, J. A., & Medina, P. P. (2020). Frequent mutations in the amino-terminal domain of BCL7A impair its tumor suppressor role in DLBCL. *Leukemia*, *34*(10), 2722–2735. https://doi.org/10.1038/s41375-020-0919-5

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, *21*(3), 381–395. https://doi.org/10.1038/cr.2011.22

Bao, X., Tang, J., Lopez-Pajares, V., Tao, S., Qu, K., Crabtree, G. R., & Khavari, P. A. (2013). ACTL6a enforces the epidermal progenitor state by suppressing SWI/SNF-dependent induction of KLF4. *Cell Stem Cell*, *12*(2), 193–203. https://doi.org/10.1016/j.stem.2012.12.014

Barbera, A. J., Chodaparambil, J. V., Kelley-Clarke, B., Joukov, V., Walter, J. C., Luger, K., & Kaye, K. M. (2006). The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. *Science (New York, N.Y.)*, *311*(5762), 856–861. https://doi.org/10.1126/science.1120541

Barish, S., Barakat, T. S., Michel, B. C., Mashtalir, N., Phillips, J. B., Valencia, A. M., Ugur, B., Wegner, J., Scott, T. M., Bostwick, B., Undiagnosed Diseases Network, Murdock, D. R., Dai, H., Perenthaler, E., Nikoncuk, A., van Slegtenhorst, M., Brooks, A. S., Keren, B., Nava, C., … Bellen, H. J. (2020). BICRA, a SWI/SNF Complex Member, Is Associated with BAF-Disorder Related Phenotypes

in Humans and Model Organisms. *American Journal of Human Genetics*, *107*(6), 1096–1112. https://doi.org/10.1016/j.ajhg.2020.11.003

Boehm, T., & Rabbitts, T. H. (1989). A chromosomal basis of lymphoid malignancy in man. *European Journal of Biochemistry*, *185*(1), 1–17. https://doi.org/10.1111/j.1432-1033.1989.tb15074.x

Bradbury, E. M. (1989). K. E. Van Holde. Chromatin. Series in molecular biology. Springer-Verlag, New York. 1988. 530 pp. $98.00. *Journal of Molecular Recognition*, *2*(3), i–i. https://doi.org/10.1002/jmr.300020308

Bradbury, E. M. (1992). Reversible histone modifications and the chromosome cell cycle. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *14*(1), 9–16. https://doi.org/10.1002/bies.950140103

Brahma, S., Ngubo, M., Paul, S., Udugama, M., & Bartholomew, B. (2018). The Arp8 and Arp4 module acts as a DNA sensor controlling INO80 chromatin remodeling. *Nature Communications*, *9*(1), 3309. https://doi.org/10.1038/s41467-018-05710-7

Brahma, S., Udugama, M. I., Kim, J., Hada, A., Bhardwaj, S. K., Hailu, S. G., Lee, T.-H., & Bartholomew, B. (2017). INO80 exchanges H2A.Z for H2A by translocating on DNA proximal to histone dimers. *Nature Communications*, *8*, 15616. https://doi.org/10.1038/ncomms15616

Caterino, T. L., & Hayes, J. J. (2011). Structure of the H1 C-terminal domain and function in chromatin condensationThis paper is one of a selection of papers published in a Special Issue entitled 31st Annual International Asilomar Chromatin and Chromosomes Conference, and has undergone the Journal's usual peer review process. *Biochemistry and Cell Biology*, *89*(1), 35–44. https://doi.org/10.1139/O10-024

Chen, G., Zhou, H., Liu, B., Wang, Y., Zhao, J., Giancotti, F. G., & Long, J. (2020). A heterotrimeric SMARCB1-SMARCC2 subcomplex is required for the assembly and tumor suppression function of the BAF chromatin-remodeling complex. *Cell Discovery*, *6*, 66. https://doi.org/10.1038/s41421-020-00196-4

Chun, H.-J. E., Lim, E. L., Heravi-Moussavi, A., Saberi, S., Mungall, K. L., Bilenky, M., Carles, A., Tse, K., Shlafman, I., Zhu, K., Qian, J. Q., Palmquist, D. L., He, A., Long, W., Goya, R., Ng, M., LeBlanc, V. G., Pleasance, E., Thiessen, N., … Marra, M. A. (2016). Genome-Wide Profiles of Extra-cranial Malignant Rhabdoid Tumors Reveal Heterogeneity and Dysregulated Developmental Pathways. *Cancer Cell*, *29*(3), 394–406. https://doi.org/10.1016/j.ccell.2016.02.009

Clapier, C. R., & Cairns, B. R. (2009). The biology of chromatin remodeling complexes. *Annual Review of Biochemistry*, *78*, 273–304. https://doi.org/10.1146/annurev.biochem.77.062706.153223

Clapier, C. R., Iwasa, J., Cairns, B. R., & Peterson, C. L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Reviews Molecular Cell Biology*, *18*(7), 407–422. https://doi.org/10.1038/nrm.2017.26

Clark, J., Rocques, P. J., Crew, A. J., Gill, S., Shipley, J., Chan, A. M., Gusterson, B. A., & Cooper, C. S. (1994). Identification of novel genes, SYT and SSX, involved in the t(X;18)(p11.2;q11.2) translocation found in human synovial sarcoma. *Nature Genetics*, *7*(4), 502–508. https://doi.org/10.1038/ng0894-502

Coiffier, B., Lepage, E., Briere, J., Herbrecht, R., Tilly, H., Bouabdallah, R., Morel, P., Van Den Neste, E., Salles, G., Gaulard, P., Reyes, F., Lederlin, P., & Gisselbrecht, C. (2002). CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *The New England Journal of Medicine*, *346*(4), 235–242. https://doi.org/10.1056/NEJMoa011795

Corbeski, I., Horn, V., van der Valk, R. A., le Paige, U. B., Dame, R. T., & van Ingen, H. (2018). Microscale Thermophoresis Analysis of Chromatin Interactions. In R. T. Dame (Ed.), *Bacterial Chromatin* (Vol. 1837, pp. 177–197). Springer New York. https://doi.org/10.1007/978-1-4939-8675-0_11

DelBove, J., Rosson, G., Strobeck, M., Chen, J., Archer, T. K., Wang, W., Knudsen, E. S., & Weissman, B. E. (2011). Identification of a core member of the SWI/SNF complex, BAF155/SMARCC1, as a human tumor suppressor gene. *Epigenetics*, *6*(12), 1444–1453.

Dion, V., Shimada, K., & Gasser, S. M. (2010). Actin-related proteins in the nucleus: Life beyond chromatin remodelers. *Current Opinion in Cell Biology*, *22*(3), 383–391. https://doi.org/10.1016/j.ceb.2010.02.006

Dodonova, S. O., Zhu, F., Dienemann, C., Taipale, J., & Cramer, P. (2020). Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. *Nature*, *580*(7805), 669–672. https://doi.org/10.1038/s41586-020-2195-y

Dyer, P. N., Edayathumangalam, R. S., White, C. L., Bao, Y., Chakravarthy, S., Muthurajan, U. M., & Luger, K. (2003). Reconstitution of Nucleosome Core Particles from Recombinant Histones and DNA. In *Methods in Enzymology* (Vol. 375, pp. 23–44). Elsevier. https://doi.org/10.1016/S0076-6879(03)75002-2

Dyer, P. N., Edayathumangalam, R. S., White, C. L., Bao, Y., Chakravarthy, S., Muthurajan, U. M., & Luger, K. (2004). Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods in Enzymology*, *375*, 23–44. https://doi.org/10.1016/s0076-6879(03)75002-2

Eberharter, A., Längst, G., & Becker, P. B. (2003). A Nucleosome Sliding Assay for Chromatin Remodeling Factors. In *Methods in Enzymology* (Vol. 377, pp. 344–353). Elsevier. https://doi.org/10.1016/S0076-6879(03)77021-9

El Hadidy, N., & Uversky, V. N. (2019). Intrinsic Disorder of the BAF Complex: Roles in Chromatin Remodeling and Disease Development. *International Journal of Molecular Sciences*, *20*(21), 5260.

Evtushenko, E. V., Elisafenko, E. A., Gatzkaya, S. S., Lipikhina, Y. A., Houben, A., & Vershinin, A. V. (2017). Conserved molecular structure of the centromeric histone CENH3 in Secale and its phylogenetic relationships. *Scientific Reports*, *7*(1), 17628. https://doi.org/10.1038/s41598-017-17932-8

Farnung, L., Vos, S. M., Wigge, C., & Cramer, P. (2017). Nucleosome-Chd1 structure and implications for chromatin remodelling. *Nature*, *550*(7677), 539–542. https://doi.org/10.1038/nature24046

Fedorova, E., & Zink, D. (2008). Nuclear architecture and gene regulation. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1783*(11), 2174–2184. https://doi.org/10.1016/j.bbamcr.2008.07.018

Felsenfeld, G., & Groudine, M. (2003). Controlling the double helix. *Nature*, *421*(6921), 448–453.

Flaus, A. (2011). Principles and practice of nucleosome positioning in vitro. *Frontiers in Life Science*, *5*(1–2), 5–27. https://doi.org/10.1080/21553769.2012.702667

Flynn, E. M., Huang, O. W., Poy, F., Oppikofer, M., Bellon, S. F., Tang, Y., & Cochran, A. G. (2015). A Subset of Human Bromodomains Recognizes Butyryllysine and Crotonyllysine Histone Peptide Modifications. *Structure (London, England: 1993)*, *23*(10), 1801–1814. https://doi.org/10.1016/j.str.2015.08.004

Forbes, S. A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., Kok, C. Y., Jia, M., De, T., Teague, J. W., Stratton, M. R., McDermott, U., & Campbell, P. J. (2015). COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Research*, *43*(D1), D805–D811. https://doi.org/10.1093/nar/gku1075

Gallego, L. D., Schneider, M., Mittal, C., Romanauska, A., Carrillo, R. M. G., Schubert, T., Pugh, B. F., & Köhler, A. (2020). Phase separation directs ubiquitination of gene-body nucleosomes. *Nature*, *579*(7800), 592–597.

Giaimo, B. D., Ferrante, F., Herchenröther, A., Hake, S. B., & Borggrefe, T. (2019). The histone variant H2A.Z in gene regulation. *Epigenetics & Chromatin*, *12*(1), 37. https://doi.org/10.1186/s13072-019-0274-9

Gibson, B. A., Doolittle, L. K., Schneider, M. W., Jensen, L. E., Gamarra, N., Henry, L., Gerlich, D. W., Redding, S., & Rosen, M. K. (2019). Organization of chromatin by intrinsic and regulated phase separation. *Cell*, *179*(2), 470–484.

Gibson, B. A., Zhang, Y., Jiang, H., Hussey, K. M., Shrimp, J. H., Lin, H., Schwede, F., Yu, Y., & Kraus, W. L. (2016). Chemical genetic discovery of PARP targets reveals a role for PARP-1 in transcription elongation. *Science (New York, N.Y.)*, *353*(6294), 45–50. https://doi.org/10.1126/science.aaf7865

Gunawardana, J., Chan, F. C., Telenius, A., Woolcock, B., Kridel, R., Tan, K. L., Ben-Neriah, S., Mottok, A., Lim, R. S., Boyle, M., Rogic, S., Rimsza, L. M., Guiter, C., Leroy, K., Gaulard, P., Haioun, C., Marra, M. A., Savage, K. J., Connors, J. M., … Steidl, C. (2014). Recurrent somatic mutations of PTPN1 in primary mediastinal B cell lymphoma and Hodgkin lymphoma. *Nature Genetics*, *46*(4), 329–335. https://doi.org/10.1038/ng.2900

Han, Y., Reyes, A. A., Malik, S., & He, Y. (2020a). Cryo-EM structure of SWI/SNF complex bound to a nucleosome. *Nature*, *579*(7799), 452–455. https://doi.org/10.1038/s41586-020-2087-1

Han, Y., Reyes, A. A., Malik, S., & He, Y. (2020b). Cryo-EM structure of SWI/SNF complex bound to a nucleosome. *Nature*, *579*(7799), 452–455. https://doi.org/10.1038/s41586-020-2087-1

Hanson, B. L., Harp, J. M., & Bunick, G. J. (2003). The well-tempered protein crystal: Annealing macromolecular crystals. *Methods in Enzymology*, *368*, 217–235. https://doi.org/10.1016/s0076-6879(03)68012-2

He, S., Wu, Z., Tian, Y., Yu, Z., Yu, J., Wang, X., Li, J., Liu, B., & Xu, Y. (2020a). Structure of nucleosome-bound human BAF complex. *Science*, *367*(6480), 875–881.

He, S., Wu, Z., Tian, Y., Yu, Z., Yu, J., Wang, X., Li, J., Liu, B., & Xu, Y. (2020b). Structure of nucleosome-bound human BAF complex. *Science (New York, N.Y.)*, *367*(6480), 875–881. https://doi.org/10.1126/science.aaz9761

Henikoff, S., & Smith, M. M. (2015). Histone variants and epigenetics. *Cold Spring Harbor Perspectives in Biology*, *7*(1), a019364. https://doi.org/10.1101/cshperspect.a019364

Hodawadekar, S. C., & Marmorstein, R. (2007). Chemistry of acetyl transfer by histone modifying enzymes: Structure, mechanism and implications for effector design. *Oncogene*, *26*(37), 5528–5540. https://doi.org/10.1038/sj.onc.1210619

Hota, S. K., & Bruneau, B. G. (2016). ATP-dependent chromatin remodeling during mammalian development. *Development*, *143*(16), 2882–2897. https://doi.org/10.1242/dev.128892

Huang, C., Hao, Q., Shi, G., Zhou, X., & Zhang, Y. (2021). BCL7C suppresses ovarian cancer growth by inactivating mutant p53. *Journal of Molecular Cell Biology*, *13*(2), 141–150. https://doi.org/10.1093/jmcb/mjaa065

Hughes, A. L., & Rando, O. J. (2014). Mechanisms underlying nucleosome positioning in vivo. *Annual Review of Biophysics*, *43*, 41–63. https://doi.org/10.1146/annurev-biophys-051013-023114

Husmann, D., & Gozani, O. (2019). Histone lysine methyltransferases in biology and disease. *Nature Structural & Molecular Biology*, *26*(10), 880–889. https://doi.org/10.1038/s41594-019-0298-7

Huttlin, E. L., Ting, L., Bruckner, R. J., Gebreab, F., Gygi, M. P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., Dong, R., Guarani, V., Vaites, L. P., Ordureau, A., Rad, R., Erickson, B. K., Wühr, M., Chick, J., Zhai, B., … Gygi, S. P. (2015). The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell*, *162*(2), 425–440. https://doi.org/10.1016/j.cell.2015.06.043

Ishii, M., Suehara, Y., Sano, K., Kohsaka, S., Hayashi, T., Kazuno, S., Akaike, K., Mukaihara, K., Kim, Y., Okubo, T., Takamochi, K., Takahashi, F., Kaneko, K., & Saito, T. (2018). Proteomic signatures corresponding to the SS18/SSX fusion gene in synovial sarcoma. *Oncotarget*, *9*(101), 37509–37519. https://doi.org/10.18632/oncotarget.26493

Iwafuchi-Doi, M., & Zaret, K. S. (2016). Cell fate control by pioneer transcription factors. *Development (Cambridge, England)*, *143*(11), 1833–1837. https://doi.org/10.1242/dev.133900

Izhar, L., Adamson, B., Ciccia, A., Lewis, J., Pontano-Vaites, L., Leng, Y., Liang, A. C., Westbrook, T. F., Harper, J. W., & Elledge, S. J. (2015). A Systematic Analysis of Factors Localized to Damaged Chromatin Reveals PARP-Dependent Recruitment of Transcription Factors. *Cell Reports*, *11*(9), 1486–1500. https://doi.org/10.1016/j.celrep.2015.04.053

Jadayel, D. M., Osborne, L. R., Coignet, L. J., Zani, V. J., Tsui, L.-C., Scherer, S. W., & Dyer, M. J. (1998). The BCL7 gene family: Deletion of BCL7B in Williams syndrome. *Gene*, *224*(1–2), 35–44.

Janssen, A., Colmenares, S. U., & Karpen, G. H. (2018). Heterochromatin: Guardian of the Genome. *Annual Review of Cell and Developmental Biology*, *34*, 265–288. https://doi.org/10.1146/annurev-cellbio-100617-062653

Jl, R., W, W., & Gr, C. (2013). From neural development to cognition: Unexpected roles for chromatin. *Nature Reviews. Genetics*, *14*(5). https://doi.org/10.1038/nrg3413

Jones, S., Wang, T.-L., Shih, I.-M., Mao, T.-L., Nakayama, K., Roden, R., Glas, R., Slamon, D., Diaz, L. A., Vogelstein, B., Kinzler, K. W., Velculescu, V. E., & Papadopoulos, N. (2010). Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science (New York, N.Y.)*, *330*(6001), 228–231. https://doi.org/10.1126/science.1196333

Kadoch, C., & Crabtree, G. R. (2013). Reversible disruption of mSWI/SNF (BAF) complexes by the SS18-SSX oncogenic fusion in synovial sarcoma. *Cell*, *153*(1), 71–85. https://doi.org/10.1016/j.cell.2013.02.036

Kadoch, C., & Crabtree, G. R. (2015). Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Science Advances*, *1*(5), e1500447. https://doi.org/10.1126/sciadv.1500447

Kadoch, C., Hargreaves, D. C., Hodges, C., Elias, L., Ho, L., Ranish, J., & Crabtree, G. R. (2013a). Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nature Genetics*, *45*(6), 592–601. https://doi.org/10.1038/ng.2628

Kadoch, C., Hargreaves, D. C., Hodges, C., Elias, L., Ho, L., Ranish, J., & Crabtree, G. R. (2013b). Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nature Genetics*, *45*(6), 592–601. https://doi.org/10.1038/ng.2628

Kadoch, C., Hargreaves, D. C., Hodges, C., Elias, L., Ho, L., Ranish, J., & Crabtree, G. R. (2013c). Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nature Genetics*, *45*(6), 592–601. https://doi.org/10.1038/ng.2628

Keppler, B. R., & Archer, T. K. (2010). Ubiquitin-dependent and ubiquitin-independent control of subunit stoichiometry in the SWI/SNF complex. *Journal of Biological Chemistry*, *285*(46), 35665–35674.

Kerr, K., Qualmann, K., Esquenazi, Y., Hagan, J., & Kim, D. H. (2018). Familial Syndromes Involving Meningiomas Provide Mechanistic Insight Into Sporadic Disease. *Neurosurgery*, *83*(6), 1107–1118. https://doi.org/10.1093/neuros/nyy121

Klein-Brill, A., Joseph-Strauss, D., Appleboim, A., & Friedman, N. (2019). Dynamics of Chromatin and Transcription during Transient Depletion of the RSC Chromatin Remodeling Complex. *Cell Reports*, *26*(1), 279-292.e5. https://doi.org/10.1016/j.celrep.2018.12.020

Klug, A. (2010). The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annual Review of Biochemistry*, *79*, 213–231.

Kossel, A., & 1853-1927. (1928). *Protamines and histones*. Longmans, Green and co. https://scholar.google.com/scholar_lookup?title=protamines+and+histones&author=Kossel%2C+A.+%28Albrecht%29&publication_year=1928

Krasteva, V., Buscarlet, M., Diaz-Tellez, A., Bernard, M.-A., Crabtree, G. R., & Lessard, J. A. (2012). The BAF53a subunit of SWI/SNF-like BAF complexes is essential for hemopoietic stem cell function. *Blood*, *120*(24), 4720–4732. https://doi.org/10.1182/blood-2012-04-427047

Krietenstein, N., Wal, M., Watanabe, S., Park, B., Peterson, C. L., Pugh, B. F., & Korber, P. (2016). Genomic Nucleosome Organization Reconstituted with Pure Proteins. *Cell*, *167*(3), 709-721.e12. https://doi.org/10.1016/j.cell.2016.09.045

Lacy, S. E., Barrans, S. L., Beer, P. A., Painter, D., Smith, A. G., Roman, E., Cooke, S. L., Ruiz, C., Glover, P., Van Hoppe, S. J. L., Webster, N., Campbell, P. J., Tooze, R. M., Patmore, R., Burton, C., Crouch, S., & Hodson, D. J. (2020). Targeted sequencing in DLBCL, molecular subtypes, and outcomes: A Haematological Malignancy Research Network report. *Blood*, *135*(20), 1759–1771. https://doi.org/10.1182/blood.2019003535

Laity, J. H., Lee, B. M., & Wright, P. E. (2001). Zinc finger proteins: New insights into structural and functional diversity. *Current Opinion in Structural Biology*, *11*(1), 39–46. https://doi.org/10.1016/s0959-440x(00)00167-6

Längst, G., & Manelyte, L. (2015). Chromatin Remodelers: From Function to Dysfunction. *Genes*, *6*(2), 299–324. https://doi.org/10.3390/genes6020299

Lau, A. T. Y., Lee, S.-Y., Xu, Y.-M., Zheng, D., Cho, Y.-Y., Zhu, F., Kim, H.-G., Li, S.-Q., Zhang, Z., Bode, A. M., & Dong, Z. (2011). Phosphorylation of histone H2B serine 32 is linked to cell transformation. *The Journal of Biological Chemistry*, *286*(30), 26628–26637. https://doi.org/10.1074/jbc.M110.215590

Lee, C. Y., & Grant, P. A. (2019). Role of Histone Acetylation and Acetyltransferases in Gene Regulation. In *Toxicoepigenetics* (pp. 3–30). Elsevier. https://doi.org/10.1016/B978-0-12-812433-8.00001-0

Lee, J.-H., Kim, E. W., Croteau, D. L., & Bohr, V. A. (2020). Heterochromatin: An epigenetic point of view in aging. *Experimental & Molecular Medicine*, *52*(9), 1466–1474. https://doi.org/10.1038/s12276-020-00497-4

Li, M., Zhao, H., Zhang, X., Wood, L. D., Anders, R. A., Choti, M. A., Pawlik, T. M., Daniel, H. D., Kannangai, R., Offerhaus, G. J. A., Velculescu, V. E., Wang, L., Zhou, S., Vogelstein, B., Hruban, R. H., Papadopoulos, N., Cai, J., Torbenson, M. S., & Kinzler, K. W. (2011). Inactivating mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma. *Nature Genetics*, *43*(9), 828–829. https://doi.org/10.1038/ng.903

Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J. V., King, D. S., & Banani, S. F. (2012). Phase transitions in the assembly of multivalent signalling proteins. *Nature*, *483*(7389), 336–340.

Li, X. S., Trojer, P., Matsumura, T., Treisman, J. E., & Tanese, N. (2010). Mammalian SWI/SNF--a subunit BAF250/ARID1 is an E3 ubiquitin ligase that targets histone H2B. *Molecular and Cellular Biology*, *30*(7), 1673–1688. https://doi.org/10.1128/MCB.00540-09

Liu, J., Gao, L., Ji, B., Geng, R., Chen, J., Tao, X., Cai, Q., & Chen, Z. (2021). BCL7A as a novel prognostic biomarker for glioma patients. *Journal of Translational Medicine*, *19*(1), 335. https://doi.org/10.1186/s12967-021-03003-0

Liu, X., Li, M., Xia, X., Li, X., & Chen, Z. (2017a). Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure. *Nature*, *544*(7651), 440–445. https://doi.org/10.1038/nature22036

Liu, X., Li, M., Xia, X., Li, X., & Chen, Z. (2017b). Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure. *Nature*, *544*(7651), 440–445. https://doi.org/10.1038/nature22036

Lomelí, H., & Castillo-Robles, J. (2016). The developmental and pathogenic roles of BAF57, a special subunit of the BAF chromatin-remodeling complex. *FEBS Letters*, *590*(11), 1555–1569. https://doi.org/10.1002/1873-3468.12201

Lowary, P. T., & Widom, J. (1998a). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of Molecular Biology*, *276*(1), 19–42. https://doi.org/10.1006/jmbi.1997.1494

Lowary, P. T., & Widom, J. (1998b). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of Molecular Biology*, *276*(1), 19–42. https://doi.org/10.1006/jmbi.1997.1494

Lu, W., Fang, L., Ouyang, B., Zhang, X., Zhan, S., Feng, X., Bai, Y., Han, X., Kim, H., He, Q., Wan, M., Shi, F.-T., Feng, X.-H., Liu, D., Huang, J., & Songyang, Z. (2015). Actl6a protects embryonic stem cells from differentiating into primitive endoderm. *Stem Cells (Dayton, Ohio)*, *33*(6), 1782–1793. https://doi.org/10.1002/stem.2000

Ludwigsen, J., Pfennig, S., Singh, A. K., Schindler, C., Harrer, N., Forné, I., Zacharias, M., & Mueller-Planitz, F. (n.d.). Concerted regulation of ISWI by an autoinhibitory domain and the H4 N-terminal tail. *ELife*, *6*, e21477. https://doi.org/10.7554/eLife.21477

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, *389*(6648), 251–260. https://doi.org/10.1038/38444

Luger, K., Rechsteiner, T. J., & Richmond, T. J. (1999). Preparation of nucleosome core particle from recombinant histones. In *Methods in Enzymology* (Vol. 304, pp. 3–19). Academic Press. https://doi.org/10.1016/S0076-6879(99)04003-3

Makde, R. D., England, J. R., Yennawar, H. P., & Tan, S. (2010). Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature*, *467*(7315), 562–566. https://doi.org/10.1038/nature09321

Marcon, E., Ni, Z., Pu, S., Turinsky, A. L., Trimble, S. S., Olsen, J. B., Silverman-Gavrila, R., Silverman-Gavrila, L., Phanse, S., Guo, H., Zhong, G., Guo, X., Young, P., Bailey, S., Roudeva, D., Zhao, D., Hewel, J., Li, J., Gräslund, S., … Greenblatt, J. (2014). Human-Chromatin-Related Protein Interactions Identify a Demethylase Complex Required for Chromosome Segregation. *Cell Reports*, *8*(1), 297–310. https://doi.org/10.1016/j.celrep.2014.05.050

Mashtalir, N., D'Avino, A. R., Michel, B. C., Luo, J., Pan, J., Otto, J. E., Zullow, H. J., McKenzie, Z. M., Kubiak, R. L., Pierre, R. S., Valencia, A. M., Poynter, S. J., Cassel, S. H., Ranish, J. A., & Kadoch, C. (2018). Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. *Cell*, *175*(5), 1272-1288.e20. https://doi.org/10.1016/j.cell.2018.09.032

Mashtalir, N., D'Avino, A. R., Michel, B. C., Luo, J., Pan, J., Otto, J. E., Zullow, H. J., McKenzie, Z. M., Kubiak, R. L., St. Pierre, R., Valencia, A. M., Poynter, S. J., Cassel, S. H., Ranish, J. A., & Kadoch, C. (2018). Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. *Cell*, *175*(5), 1272-1288.e20. https://doi.org/10.1016/j.cell.2018.09.032

Mashtalir, N., D'Avino, A. R., Michel, B. C., Luo, J., Pan, J., Otto, J. E., Zullow, H. J., McKenzie, Z. M., Kubiak, R. L., St Pierre, R., Valencia, A. M., Poynter, S. J., Cassel, S. H., Ranish, J. A., & Kadoch, C. (2018). Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. *Cell*, *175*(5), 1272-1288.e20. https://doi.org/10.1016/j.cell.2018.09.032

Mashtalir, N., Suzuki, H., Farrell, D. P., Sankar, A., Luo, J., Filipovski, M., D'Avino, A. R., St Pierre, R., Valencia, A. M., Onikubo, T., Roeder, R. G., Han, Y., He, Y., Ranish, J. A., DiMaio, F., Walz, T., & Kadoch, C. (2020a). A Structural Model of the Endogenous Human BAF Complex Informs Disease Mechanisms. *Cell*, *183*(3), 802-817.e24. https://doi.org/10.1016/j.cell.2020.09.051

Mashtalir, N., Suzuki, H., Farrell, D. P., Sankar, A., Luo, J., Filipovski, M., D'Avino, A. R., St Pierre, R., Valencia, A. M., Onikubo, T., Roeder, R. G., Han, Y., He, Y., Ranish, J. A., DiMaio, F., Walz, T., & Kadoch, C. (2020b). A Structural Model of the Endogenous Human BAF Complex Informs Disease Mechanisms. *Cell*, *183*(3), 802-817.e24. https://doi.org/10.1016/j.cell.2020.09.051

Mathur, R. (2018). ARID1A loss in cancer: Towards a mechanistic understanding. *Pharmacology & Therapeutics*, *190*, 15–23. https://doi.org/10.1016/j.pharmthera.2018.05.001

Mayr, B., & Montminy, M. (2001). Transcriptional regulation by the phosphorylation-dependent factor CREB. *Nature Reviews. Molecular Cell Biology*, *2*(8), 599–609. https://doi.org/10.1038/35085068

McBride, M. J., Pulice, J. L., Beird, H. C., Ingram, D. R., D'Avino, A. R., Shern, J. F., Charville, G. W., Hornick, J. L., Nakayama, R. T., Garcia-Rivera, E. M., Araujo, D. M., Wang, W.-L., Tsai, J.-W., Yeagley, M., Wagner, A. J., Futreal, P. A., Khan, J., Lazar, A. J., & Kadoch, C. (2018). The SS18-SSX Fusion Oncoprotein Hijacks BAF Complex Targeting and Function to Drive Synovial Sarcoma. *Cancer Cell*, *33*(6), 1128-1141.e7. https://doi.org/10.1016/j.ccell.2018.05.002

McGinty, R. K., Henrici, R. C., & Tan, S. (2014). Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature*, *514*(7524), 591–596. https://doi.org/10.1038/nature13890

McGinty, R. K., Makde, R. D., & Tan, S. (2016). Preparation, crystallization and structure determination of chromatin enzyme/nucleosome complexes. *Methods in Enzymology*, *573*, 43–65. https://doi.org/10.1016/bs.mie.2016.01.003

McGinty, R. K., & Tan, S. (2016). Recognition of the nucleosome by chromatin factors and enzymes. *Current Opinion in Structural Biology*, *37*, 54–61. https://doi.org/10.1016/j.sbi.2015.11.014

McGinty, R. K., & Tan, S. (2021). Principles of nucleosome recognition by chromatin factors and enzymes. *Current Opinion in Structural Biology*, *71*, 16–26. https://doi.org/10.1016/j.sbi.2021.05.006

Mei, Q., Huang, J., Chen, W., Tang, J., Xu, C., Yu, Q., Cheng, Y., Ma, L., Yu, X., & Li, S. (2017). Regulation of DNA replication-coupled histone gene expression. *Oncotarget*, *8*(55), 95005–95022. https://doi.org/10.18632/oncotarget.21887

Meng, L., Wang, X., Liao, W., Liu, J., Liao, Y., & He, Q. (2017). BAF53a is a potential prognostic biomarker and promotes invasion and epithelial-mesenchymal transition of glioma cells. *Oncology Reports*, *38*(6), 3327–3334. https://doi.org/10.3892/or.2017.6019

Michel, B. C., D'Avino, A. R., Cassel, S. H., Mashtalir, N., McKenzie, Z. M., McBride, M. J., Valencia, A. M., Zhou, Q., Bocker, M., Soares, L. M. M., Pan, J., Remillard, D. I., Lareau, C. A., Zullow, H. J., Fortoul, N., Gray, N. S., Bradner, J. E., Chan, H. M., & Kadoch, C. (2018). A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. *Nature Cell Biology*, *20*(12), 1410–1420. https://doi.org/10.1038/s41556-018-0221-1

Middeljans, E., Wan, X., Jansen, P. W., Sharma, V., Stunnenberg, H. G., & Logie, C. (2012). SS18 Together with Animal-Specific Factors Defines Human BAF-Type SWI/SNF Complexes. *PLoS ONE*, *7*(3), e33834. https://doi.org/10.1371/journal.pone.0033834

Mills, A. A. (2017). The Chromodomain Helicase DNA-Binding Chromatin Remodelers: Family Traits that Protect from and Promote Cancer. *Cold Spring Harbor Perspectives in Medicine*, *7*(4), a026450. https://doi.org/10.1101/cshperspect.a026450

Muller, S., Filippakopoulos, P., & Knapp, S. (2011). Bromodomains as therapeutic targets. *Expert Reviews in Molecular Medicine*, *13*, e29. https://doi.org/10.1017/S1462399411001992

Musselman, C. A., Lalonde, M.-E., Côté, J., & Kutateladze, T. G. (2012). Perceiving the epigenetic landscape through histone readers. *Nature Structural & Molecular Biology*, *19*(12), 1218–1227. https://doi.org/10.1038/nsmb.2436

Nagai, S., Davis, R. E., Mattei, P. J., Eagen, K. P., & Kornberg, R. D. (2017). Chromatin potentiates transcription. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(7), 1536–1541. https://doi.org/10.1073/pnas.1620312114

Narlikar, G. J., Sundaramoorthy, R., & Owen-Hughes, T. (2013). Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell*, *154*(3), 490–503. https://doi.org/10.1016/j.cell.2013.07.011

Neigeborn, L., & Carlson, M. (1984). Genes Affecting the Regulation of SUC2 Gene Expression by Glucose Repression in SACCHAROMYCES CEREVISIAE. *Genetics*, *108*(4), 845–858.

Ng, S. S., Yue, W. W., Oppermann, U., & Klose, R. J. (2009). Dynamic protein methylation in chromatin biology. *Cellular and Molecular Life Sciences*, *66*(3), 407–422. https://doi.org/10.1007/s00018-008-8303-z

Nie, Z., Xue, Y., Yang, D., Zhou, S., Deroo, B. J., Archer, T. K., & Wang, W. (2000). A specificity and targeting subunit of a human SWI/SNF family-related chromatin-remodeling complex. *Molecular and Cellular Biology*, *20*(23), 8879–8888. https://doi.org/10.1128/MCB.20.23.8879-8888.2000

Olins, D. E., & Olins, A. L. (2003). Chromatin history: Our view from the bridge. *Nature Reviews. Molecular Cell Biology*, *4*(10), 809–814. https://doi.org/10.1038/nrm1225

Oppikofer, M., Bai, T., Gan, Y., Haley, B., Liu, P., Sandoval, W., Ciferri, C., & Cochran, A. G. (2017). Expansion of the ISWI chromatin remodeler family with new active complexes. *EMBO Reports*, *18*(10), 1697–1706. https://doi.org/10.15252/embr.201744011

O'Shea, E. K., Klemm, J. D., Kim, P. S., & Alber, T. (1991). X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science (New York, N.Y.)*, *254*(5031), 539–544. https://doi.org/10.1126/science.1948029

Pak, C. W., Kosno, M., Holehouse, A. S., Padrick, S. B., Mittal, A., Ali, R., Yunus, A. A., Liu, D. R., Pappu, R. V., & Rosen, M. K. (2016). Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Molecular Cell*, *63*(1), 72–85. https://doi.org/10.1016/j.molcel.2016.05.042

Papamichos-Chronakis, M., Watanabe, S., Rando, O. J., & Peterson, C. L. (2011). Global Regulation of H2A.Z Localization by the INO80 Chromatin-Remodeling Enzyme Is Essential for Genome Integrity. *Cell*, *144*(2), 200–213. https://doi.org/10.1016/j.cell.2010.12.021

Park, S.-Y., & Kim, J.-S. (2020). A short guide to histone deacetylases including recent progress on class II enzymes. *Experimental & Molecular Medicine*, *52*(2), 204–212. https://doi.org/10.1038/s12276-020-0382-4

Park, Y.-J., & Luger, K. (2006). The structure of nucleosome assembly protein 1. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(5), 1248–1253. https://doi.org/10.1073/pnas.0508002103

Parthun, M. R. (2007). Hat1: The emerging cellular roles of a type B histone acetyltransferase. *Oncogene*, *26*(37), 5319–5328. https://doi.org/10.1038/sj.onc.1210602

Patel, A. B., Moore, C. M., Greber, B. J., Luo, J., Zukin, S. A., Ranish, J., & Nogales, E. (2019). Architecture of the chromatin remodeler RSC and insights into its nucleosome engagement. *ELife*, *8*, e54449. https://doi.org/10.7554/eLife.54449

Pathare, G. R., Decout, A., Glück, S., Cavadini, S., Makasheva, K., Hovius, R., Kempf, G., Weiss, J., Kozicka, Z., Guey, B., Melenec, P., Fierz, B., Thomä, N. H., & Ablasser, A. (2020). Structural mechanism of cGAS inhibition by the nucleosome. *Nature*, *587*(7835), 668–672. https://doi.org/10.1038/s41586-020-2750-6

Pérez Jurado, L. A., Peoples, R., Kaplan, P., Hamel, B. C., & Francke, U. (1996). Molecular definition of the chromosome 7 deletion in Williams syndrome and parent-of-origin effects on growth. *American Journal of Human Genetics*, *59*(4), 781–792.

Phelan, M. L., Sif, S., Narlikar, G. J., & Kingston, R. E. (1999a). Reconstitution of a core chromatin remodeling complex from SWI/SNF subunits. *Molecular Cell*, *3*(2), 247–253.

Phelan, M. L., Sif, S., Narlikar, G. J., & Kingston, R. E. (1999b). Reconstitution of a Core Chromatin Remodeling Complex from SWI/SNF Subunits. *Molecular Cell*, *3*(2), 247–253. https://doi.org/10.1016/S1097-2765(00)80315-9

Pillonel, V., Juskevicius, D., Ng, C. K. Y., Bodmer, A., Zettl, A., Jucker, D., Dirnhofer, S., & Tzankov, A. (2018). High-throughput sequencing of nodal marginal zone lymphomas identifies recurrent BRAF mutations. *Leukemia*, *32*(11), 2412–2426. https://doi.org/10.1038/s41375-018-0082-4

Quan, J., & Yusufzai, T. (2014). The tumor suppressor chromodomain helicase DNA-binding protein 5 (CHD5) remodels nucleosomes by unwrapping. *The Journal of Biological Chemistry*, *289*(30), 20717–20726. https://doi.org/10.1074/jbc.M114.568568

Ramaswamy, A., & Ioshikhes, I. (2013). Chapter Four—Dynamics of Modeled Oligonucleosomes and the Role of Histone Variant Proteins in Nucleosome Organization. In R. Donev (Ed.), *Advances in Protein Chemistry and Structural Biology* (Vol. 90, pp. 119–149). Academic Press. https://doi.org/10.1016/B978-0-12-410523-2.00004-3

Reddy, A., Zhang, J., Davis, N. S., Moffitt, A. B., Love, C. L., Waldrop, A., Leppa, S., Pasanen, A., Meriranta, L., Karjalainen-Lindsberg, M.-L., Nørgaard, P., Pedersen, M., Gang, A. O., Høgdall, E., Heavican, T. B., Lone, W., Iqbal, J., Qin, Q., Li, G., … Dave, S. S. (2017). Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell*, *171*(2), 481-494.e15. https://doi.org/10.1016/j.cell.2017.09.027

Ribeiro-Silva, C., Vermeulen, W., & Lans, H. (2019). SWI/SNF: complex complexes in genome stability and cancer. *DNA Repair*.

Rogakou, E. P., Nieves-Neira, W., Boon, C., Pommier, Y., & Bonner, W. M. (2000). Initiation of DNA fragmentation during apoptosis induces phosphorylation of H2AX histone at serine 139. *The Journal of Biological Chemistry*, *275*(13), 9390–9395. https://doi.org/10.1074/jbc.275.13.9390

Sahu, R. K., Singh, S., & Tomar, R. S. (2020). The mechanisms of action of chromatin remodelers and implications in development and disease. *Biochemical Pharmacology*, *180*, 114200. https://doi.org/10.1016/j.bcp.2020.114200

Sanulli, S., Trnka, M., Dharmarajan, V., Tibble, R., Pascal, B., Burlingame, A., Griffin, P., Gross, J., & Narlikar, G. (2019). HP1 reshapes nucleosome core to promote phase separation of heterochromatin. *Nature*, *575*(7782), 390–394.

Savas, S., & Skardasi, G. (2018). The SWI/SNF complex subunit genes: Their functions, variations, and links to risk and survival outcomes in human cancers. *Critical Reviews in Oncology/Hematology*, *123*, 114–131. https://doi.org/10.1016/j.critrevonc.2018.01.009

Schuck, P. (2000). Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophysical Journal*, *78*(3), 1606–1619. https://doi.org/10.1016/S0006-3495(00)76713-0

Shakya, A., Park, S., Rana, N., & King, J. T. (2020). Liquid-Liquid Phase Separation of Histone Proteins in Cells: Role in Chromatin Organization. *Biophysical Journal*, *118*(3), 753–764. https://doi.org/10.1016/j.bpj.2019.12.022

Sharma, A. B., Dimitrov, S., Hamiche, A., & Van Dyck, E. (2019). Centromeric and ectopic assembly of CENP-A chromatin in health and cancer: Old marks and new tracks. *Nucleic Acids Research*, *47*(3), 1051–1069. https://doi.org/10.1093/nar/gky1298

Shen, X., Mizuguchi, G., Hamiche, A., & Wu, C. (2000). A chromatin remodelling complex involved in transcription and DNA processing. *Nature*, *406*(6795), 541–544. https://doi.org/10.1038/35020123

Shen, X., Ranallo, R., Choi, E., & Wu, C. (2003). Involvement of actin-related proteins in ATP-dependent chromatin remodeling. *Molecular Cell*, *12*(1), 147–155. https://doi.org/10.1016/s1097-2765(03)00264-8

Shi, J., Hua, X., Zhu, B., Ravichandran, S., Wang, M., Nguyen, C., Brodie, S. A., Palleschi, A., Alloisio, M., Pariscenti, G., Jones, K., Zhou, W., Bouk, A. J., Boland, J., Hicks, B., Risch, A., Bennett, H., Luke, B. T., Song, L., … Landi, M. T. (2016). Somatic Genomics and Clinical Features of Lung Adenocarcinoma: A Retrospective Study. *PLoS Medicine*, *13*(12), e1002162. https://doi.org/10.1371/journal.pmed.1002162

Shi, L., Wen, H., & Shi, X. (2017). The Histone Variant H3.3 in Transcriptional Regulation and Human Disease. *Journal of Molecular Biology*, *429*(13), 1934–1945. https://doi.org/10.1016/j.jmb.2016.11.019

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A., & Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell*, *119*(7), 941–953. https://doi.org/10.1016/j.cell.2004.12.012

Shim, Y., Duan, M.-R., Chen, X., Smerdon, M. J., & Min, J.-H. (2012). Polycistronic co-expression and non-denaturing purification of histone octamers. *Analytical Biochemistry*, *427*(2), 190–192. https://doi.org/10.1016/j.ab.2012.05.006

Singh, A. K., & Mueller-Planitz, F. (2021). Nucleosome Positioning and Spacing: From Mechanism to Function. *Journal of Molecular Biology*, *433*(6), 166847. https://doi.org/10.1016/j.jmb.2021.166847

Singh, M., D'Silva, L., & Holak, T. A. (2006). DNA-binding properties of the recombinant high-mobility-group-like AT-hook-containing region from human BRG1 protein. *Biological Chemistry*, *387*(10–11), 1469–1478. https://doi.org/10.1515/BC.2006.184

Singleton, M. R., Dillingham, M. S., & Wigley, D. B. (2007). Structure and mechanism of helicases and nucleic acid translocases. *Annual Review of Biochemistry*, *76*, 23–50. https://doi.org/10.1146/annurev.biochem.76.052305.115300

Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, *403*(6765), 41–45. https://doi.org/10.1038/47412

Sudarsanam, P., Iyer, V. R., Brown, P. O., & Winston, F. (2000). Whole-genome expression analysis of snf/swi mutants of Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, *97*(7), 3364–3369. https://doi.org/10.1073/pnas.97.7.3364

Sun, W., Wang, W., Lei, J., Li, H., & Wu, Y. (2017). Actin-like protein 6A is a novel prognostic indicator promoting invasion and metastasis in osteosarcoma. *Oncology Reports*, *37*(4), 2405–2417. https://doi.org/10.3892/or.2017.5473

Sun, Z., & Bernstein, E. (2019). Histone variant macroH2A: From chromatin deposition to molecular function. *Essays in Biochemistry*, *63*(1), 59–74. https://doi.org/10.1042/EBC20180062

Sun, Z., Sun, L., He, M., Pang, Y., Yang, Z., & Wang, J. (2019). Low BCL7A expression predicts poor prognosis in ovarian cancer. *Journal of Ovarian Research*, *12*(1), 41. https://doi.org/10.1186/s13048-019-0518-0

Suto, R. K., Clarkson, M. J., Tremethick, D. J., & Luger, K. (2000). Crystal structure of a nucleosome core particle containing the variant histone H2A.Z. *Nature Structural Biology*, *7*(12), 1121–1124. https://doi.org/10.1038/81971

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, *126*(4), 663–676. https://doi.org/10.1016/j.cell.2006.07.024

Talbert, P. B., & Henikoff, S. (2010). Histone variants—Ancient wrap artists of the epigenome. *Nature Reviews. Molecular Cell Biology*, *11*(4), 264–275. https://doi.org/10.1038/nrm2861

Talbert, P. B., & Henikoff, S. (2017). Histone variants on the move: Substrates for chromatin dynamics. *Nature Reviews Molecular Cell Biology*, *18*(2), 115–126. https://doi.org/10.1038/nrm.2016.148

Taniuchi, K., Furihata, M., Naganuma, S., Dabanaka, K., Hanazaki, K., & Saibara, T. (2018). BCL7B, a predictor of poor prognosis of pancreatic cancers, promotes cell motility and invasion by influencing CREB signaling. *American Journal of Cancer Research*, *8*(3), 387–404.

Theodoulou, N. H., Bamborough, P., Bannister, A. J., Becher, I., Bit, R. A., Che, K. H., Chung, C., Dittmann, A., Drewes, G., Drewry, D. H., Gordon, L., Grandi, P., Leveridge, M., Lindon, M., Michon, A.-M., Molnar, J., Robson, S. C., Tomkinson, N. C. O., Kouzarides, T., … Humphreys, P. G. (2016). Discovery of I-BRD9, a Selective Cell Active Chemical Probe for Bromodomain Containing Protein 9 Inhibition. *Journal of Medicinal Chemistry*, *59*(4), 1425–1439. https://doi.org/10.1021/acs.jmedchem.5b00256

Tremethick, D. J. (2007). Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber. *Cell*, *128*(4), 651–654. https://doi.org/10.1016/j.cell.2007.02.008

Trotter, K. W., & Archer, T. K. (2008). The BRG1 transcriptional coregulator. *Nuclear Receptor Signaling*, *6*, e004. https://doi.org/10.1621/nrs.06004

Uehara, T., Kage-Nakadai, E., Yoshina, S., Imae, R., & Mitani, S. (2015). The Tumor Suppressor BCL7B Functions in the Wnt Signaling Pathway. *PLoS Genetics*, *11*(1), e1004921. https://doi.org/10.1371/journal.pgen.1004921

Valencia, A. M., Collings, C. K., Dao, H. T., St. Pierre, R., Cheng, Y.-C., Huang, J., Sun, Z.-Y., Seo, H.-S., Mashtalir, N., Comstock, D. E., Bolonduro, O., Vangos, N. E., Yeoh, Z. C., Dornon, M. K., Hermawan, C., Barrett, L., Dhe-Paganon, S., Woolf, C. J., Muir, T. W., & Kadoch, C. (2019). Recurrent SMARCB1 Mutations Reveal a Nucleosome Acidic Patch Interaction Site That Potentiates mSWI/SNF Complex Chromatin Remodeling. *Cell*, *179*(6), 1342-1356.e23. https://doi.org/10.1016/j.cell.2019.10.044

Valencia, A. M., Collings, C. K., Dao, H. T., St Pierre, R., Cheng, Y.-C., Huang, J., Sun, Z.-Y., Seo, H.-S., Mashtalir, N., Comstock, D. E., Bolonduro, O., Vangos, N. E., Yeoh, Z. C., Dornon, M. K., Hermawan, C., Barrett, L., Dhe-Paganon, S., Woolf, C. J., Muir, T. W., & Kadoch, C. (2019). Recurrent SMARCB1 Mutations Reveal a Nucleosome Acidic Patch Interaction Site That Potentiates mSWI/SNF Complex Chromatin Remodeling. *Cell*, *179*(6), 1342-1356.e23. https://doi.org/10.1016/j.cell.2019.10.044

van Attikum, H., & Gasser, S. M. (2005). The histone code at DNA breaks: A guide to repair? *Nature Reviews. Molecular Cell Biology*, *6*(10), 757–765. https://doi.org/10.1038/nrm1737

van Doorn, R., Zoutman, W. H., Dijkman, R., de Menezes, R. X., Commandeur, S., Mulder, A. A., van der Velden, P. A., Vermeer, M. H., Willemze, R., Yan, P. S., Huang, T. H., & Tensen, C. P. (2005). Epigenetic Profiling of Cutaneous T-Cell Lymphoma: Promoter Hypermethylation of Multiple Tumor Suppressor Genes Including BCL7a, PTPRG, and p73. *Journal of Clinical Oncology*, *23*(17), 3886–3896. https://doi.org/10.1200/JCO.2005.11.353

Varela, I., Tarpey, P., Raine, K., Huang, D., Ong, C. K., Stephens, P., Davies, H., Jones, D., Lin, M.-L., Teague, J., Bignell, G., Butler, A., Cho, J., Dalgliesh, G. L., Galappaththige, D., Greenman, C., Hardy, C., Jia, M., Latimer, C., … Futreal, P. A. (2011). Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature*, *469*(7331), 539–542. https://doi.org/10.1038/nature09639

Versteege, I., Sévenet, N., Lange, J., Rousseau-Merck, M. F., Ambros, P., Handgretinger, R., Aurias, A., & Delattre, O. (1998). Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature*, *394*(6689), 203–206. https://doi.org/10.1038/28212

Vinson, C., Myakishev, M., Acharya, A., Mir, A. A., Moll, J. R., & Bonovich, M. (2002). Classification of human B-ZIP proteins based on dimerization properties. *Molecular and Cellular Biology*, *22*(18), 6321–6335. https://doi.org/10.1128/MCB.22.18.6321-6335.2002

Wagner, F. R., Dienemann, C., Wang, H., Stützer, A., Tegunov, D., Urlaub, H., & Cramer, P. (2020a). Structure of SWI/SNF chromatin remodeller RSC bound to a nucleosome. *Nature*, *579*(7799), 448–451. https://doi.org/10.1038/s41586-020-2088-0

Wagner, F. R., Dienemann, C., Wang, H., Stützer, A., Tegunov, D., Urlaub, H., & Cramer, P. (2020b). Structure of SWI/SNF chromatin remodeller RSC bound to a nucleosome. *Nature*, *579*(7799), 448–451. https://doi.org/10.1038/s41586-020-2088-0

Wan, C., Borgeson, B., Phanse, S., Tu, F., Drew, K., Clark, G., Xiong, X., Kagan, O., Kwan, J., Bezginov, A., Chessman, K., Pal, S., Cromar, G., Papoulas, O., Ni, Z., Boutz, D. R., Stoilova, S.,

Havugimana, P. C., Guo, X., … Emili, A. (2015). Panorama of ancient metazoan macromolecular complexes. *Nature*, *525*(7569), 339–344. https://doi.org/10.1038/nature14877

Wang, C., Guo, Z., Zhan, X., Yang, F., Wu, M., & Zhang, X. (2020). Structure of the yeast Swi/Snf complex in a nucleosome free state. *Nature Communications*, *11*(1), 3398. https://doi.org/10.1038/s41467-020-17229-x

Wang, J., Choi, J.-M., Holehouse, A. S., Lee, H. O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A., Drechsel, D., Poser, I., Pappu, R. V., Alberti, S., & Hyman, A. A. (2018). A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell*, *174*(3), 688-699.e16. https://doi.org/10.1016/j.cell.2018.06.006

Wang, X., Wang, S., Troisi, E. C., Howard, T. P., Haswell, J. R., Wolf, B. K., Hawk, W. H., Ramos, P., Oberlick, E. M., Tzvetkov, E. P., Ross, A., Vazquez, F., Hahn, W. C., Park, P. J., & Roberts, C. W. M. (2019). BRD9 defines a SWI/SNF sub-complex and constitutes a specific vulnerability in malignant rhabdoid tumors. *Nature Communications*, *10*(1), 1881. https://doi.org/10.1038/s41467-019-09891-7

Willhoft, O., McCormack, E. A., Aramayo, R. J., Bythell-Douglas, R., Ocloo, L., Zhang, X., & Wigley, D. B. (2017). Crosstalk within a functional INO80 complex dimer regulates nucleosome sliding. *ELife*, *6*, e25782. https://doi.org/10.7554/eLife.25782

Willhoft, O., & Wigley, D. B. (2020). INO80 and SWR1 complexes: The non-identical twins of chromatin remodelling. *Current Opinion in Structural Biology*, *61*, 50–58. https://doi.org/10.1016/j.sbi.2019.09.002

Winston, F., & Carlson, M. (1992). Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. *Trends in Genetics: TIG*, *8*(11), 387–391. https://doi.org/10.1016/0168-9525(92)90300-s

Wischhof, L., Maida, S., Piazzesi, A., Gioran, A., Barragan Sanz, K., Irsen, S., Beyer, M., Schultze, J. L., Dyer, M. J., Salomoni, P., Ehninger, D., Nicotera, P., & Bano, D. (2017). The SWI/SNF subunit Bcl7a contributes to motor coordination and Purkinje cell function. *Scientific Reports*, *7*, 17055. https://doi.org/10.1038/s41598-017-17284-3

Worden, E. J., Zhang, X., & Wolberger, C. (2020). Structural basis for COMPASS recognition of an H2B-ubiquitinated nucleosome. *ELife*, *9*, e53199. https://doi.org/10.7554/eLife.53199

Xiao, S., Chang, R.-M., Yang, M.-Y., Lei, X., Liu, X., Gao, W.-B., Xiao, J.-L., & Yang, L.-Y. (2016). Actin-like 6A predicts poor prognosis of hepatocellular carcinoma and promotes metastasis and epithelial-mesenchymal transition. *Hepatology (Baltimore, Md.)*, *63*(4), 1256–1271. https://doi.org/10.1002/hep.28417

Xiao, Y., Lin, F.-T., & Lin, W.-C. (2021). ACTL6A promotes repair of cisplatin-induced DNA damage, a new mechanism of platinum resistance in cancer. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(3), e2015808118. https://doi.org/10.1073/pnas.2015808118

Yan, L., & Chen, Z. (2020). A Unifying Mechanism of DNA Translocation Underlying Chromatin Remodeling. *Trends in Biochemical Sciences*, *45*(3), 217–227. https://doi.org/10.1016/j.tibs.2019.09.002

Yan, L., Wu, H., Li, X., Gao, N., & Chen, Z. (2019). Structures of the ISWI–nucleosome complex reveal a conserved mechanism of chromatin remodeling. *Nature Structural & Molecular Biology*, *26*(4), 258–266. https://doi.org/10.1038/s41594-019-0199-9

Yan, L., Xie, S., Du, Y., & Qian, C. (2017). Structural Insights into BAF47 and BAF155 Complex Formation. *Journal of Molecular Biology*, *429*(11), 1650–1660. https://doi.org/10.1016/j.jmb.2017.04.008

Yang, C., van der Woerd, M. J., Muthurajan, U. M., Hansen, J. C., & Luger, K. (2011). Biophysical analysis and small-angle X-ray scattering-derived structures of MeCP2-nucleosome complexes. *Nucleic Acids Research*, *39*(10), 4122–4135. https://doi.org/10.1093/nar/gkr005

Yang, X.-J., & Seto, E. (2007). HATs and HDACs: From structure, function and regulation to novel strategies for therapy and prevention. *Oncogene*, *26*(37), 5310–5310.

Zani, V., Asou, N., Jadayel, D., Heward, J., Shipley, J., Nacheva, E., Takasuki, K., Catovsky, D., & Dyer, M. (1996). Molecular cloning of complex chromosomal translocation t(8;14;12)(q24.1;q32.3;q24.1) in a Burkitt lymphoma cell line defines a new gene (BCL7A) with homology to caldesmon. *Blood*, *87*(8), 3124–3134. https://doi.org/10.1182/blood.V87.8.3124.bloodjournal8783124

Zeng, Z., Yang, H., & Xiao, S. (2018). ACTL6A expression promotes invasion, metastasis and epithelial mesenchymal transition of colon cancer. *BMC Cancer*, *18*(1), 1020. https://doi.org/10.1186/s12885-018-4931-3

Zhao, B., Xu, P., Rowlett, C. M., Jing, T., Shinde, O., Lei, Y., West, A. P., Liu, W. R., & Li, P. (2020). The molecular basis of tight nuclear tethering and inactivation of cGAS. *Nature*, *587*(7835), 673–677. https://doi.org/10.1038/s41586-020-2749-z

Zhao, H., Wang, J., Han, Y., Huang, Z., Ying, J., Bi, X., Zhao, J., Fang, Y., Zhou, H., Zhou, J., Li, Z., Zhang, Y., Yang, X., Yan, T., Wang, L., Torbenson, M. S., & Cai, J. (2011). ARID2: A new tumor suppressor gene in hepatocellular carcinoma. *Oncotarget*, *2*(11), 886–891. https://doi.org/10.18632/oncotarget.355

Zhou, K., Gaullier, G., & Luger, K. (2019). Nucleosome structure and dynamics are coming of age. *Nature Structural & Molecular Biology*, *26*(1), 3–13. https://doi.org/10.1038/s41594-018-0166-x

Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., Dodonova, S. O., Nitta, K. R., Morgunova, E., Taipale, M., Cramer, P., & Taipale, J. (2018). The interaction landscape between transcription factors and the nucleosome. *Nature*, *562*(7725), 76–81. https://doi.org/10.1038/s41586-018-0549-5

Zhu, X., Liao, Y., & Tang, L. (2020). Targeting BRD9 for Cancer Treatment: A New Strategy. *OncoTargets and Therapy*, *13*, 13191–13200. https://doi.org/10.2147/OTT.S286867

ARTICLE IN PREPARATION

Running title:

BCL7 proteins, novel subunits of the mammalian SWI/SNF complex, bind the nucleosome core particle

Dana Diaz[1], Stephanie Siebert[1], Franck Martin[1], Julie Lafouge[1], Elisa Bergamin[1#]

[1] Institut de Génetique et de Biologie Moléculaire et Cellularie. 1 Rue Laurent Fries, 67404 Illkirch Cedex, Strasbourg, France

[#] To whom the correspondence should be addressed:

email: bergamie@igbmc.fr

tel: +33 368763575

ABSTRACT

BCL7 proteins are recently identified subunits of the mammalian SWI/SNF (mSWI/SNF) chromatin remodeler complex. Mutations in BCL7 proteins are associated with different kind of cancers including blood malignancies. The information on the molecular function and on the structure of BCL7 proteins is to date very limited. Here we report that BCL7 proteins bind directly the nucleosome core particle (NCP) and free DNA with high affinity. We demonstrate that BCL7 proteins form defined complexes with the NCP and we identify the conserved N-terminal part of BCL7 proteins as sufficient to nucleosome binding. We further characterize the impact of BCL7 protein mutations reported in cancer patients on NCP binding.

INTRODUCTION

At the most fundamental level, eukaryotic DNA is packaged into chromatin by making nearly two turns around an octamer of histone proteins, forming nucleosomes. Several multi-protein, chromatin-modifying complexes contain at their core an ATPase subunit that remodels chromatin: replacement of core histones with histone variants that influence DNA accessibility; sliding, addition or eviction of nucleosomes from the chromatin fibre to modulate the compaction of DNA or expose certain regions[1,2]. Deregulation of this process can severely impact gene expression patterns and genome integrity. Their importance is such that chromatin remodelling protein mutations are strongly associated to several diseases, including cancer. Unfortunately, we currently do not understand well enough the molecular details of their mode of action to be able to translate this into improved cancer treatment.

All eukaryotes have multiple ATP-dependent chromatin remodelling complexes, which tend to be highly conserved, from simple yeasts to metazoans, testifying to their importance. Their ATPases all have homology to the S. cerevisiae enzyme SNF2 and are classified into sub- families based on the amino acid sequence of the ATPase: SNF2, ISWI, CHD, and INO80. The complexes accomplish their remodelling function by antagonizing the strong interaction between the histone octamer and DNA, to alter chromatin structure. In addition to their ATPase subunit, these complexes contain more associated proteins that contribute to their function. The mammalian switch/sucrose non-fermentable (mSWI/SNF) complex is composed of at least 11 subunits, coming from the product of 29 genes and multiple paralogs generating wide diversity in its composition. It appears as a ~ 2MDa

complex. It comprises a set of evolutionary conserved 'core and enzymatic' subunits, but also 'auxiliary' subunits present only in animals thought to reflect increasing biological complexity. All complexes contain a conserved ATPase subunit, either SMARCA4 (BRG1) or SMARCA2 (BRM), that catalyses the hydrolysis of ATP. The advent of sensitive and large-scale proteomics methods has shed new light on the composition of the human SWI/SNF. Studies focused on identifying mSWI/SNF components[3,4], as well as others determining the composition of most nuclear multiprotein complexes[5,6,7] have revealed the identity of previously unsuspected mSWI/SNF subunits such us BCL7A, B and C, BRD7, BRD9, SS18, SS18L1, GLTSCR1. Without exception, these proteins do not exist in yeast, and only participate in the metazoan complexes. The roles of most accessory subunits and their function within SWI/SNF remain poorly defined.

We have undertaken pioneering work to address this knowledge gap and we investigated the role of the BCL7 proteins. They are labelled as "Cancer Genes" by the COSMIC database as they are often mutated in a variety of cancers and established tumor suppressors [8,9]. Despite this, the structure, the molecular details of interaction and the function of these auxiliary yet important subunits are poorly defined at best. The topological information is limited to the observation that BCL7 proteins make contact with BAF47, BRG1 and DPF2[10].

It has been reported that these subunits may have a role in DNA damage repair (DDR)[11], but our knowledge of their actual function in this process is very limited. Further, recent studies have reported the structure of the SWI/SNF complex. Unfortunately, they do not inform us on BCL7 proteins: two studies concerned the yeast SWI/SNF complex and BCL7 are absent in this species[12,13]; one study used a

reconstituted mSWI/SNF and BCL7 were omitted[14]; and the fourth study with the endogenous human complex had overall poor resolution and BCL7 proteins were not visible[15].

Interestingly, although mSWI/SNF is an NCP-displacing complex of 11 proteins, only two of those, SMARCB1 and the ATPase BRG1[16,17], have clearly mapped NCP-binding domains. Thus, much remains to be discovered about the interactions between mSWI/SNF and its substrate.

Here, we used biochemical approaches to provide evidence that BCL7 proteins bind directly the nucleosome core particle to mediate mSWI/SNF complex nucleosome remodelling activity *in vitro* helping the regulation of chromatin architecture. Understanding how BCL7 proteins interact with the nucleosome, the fundamental building block of eukaryotic chromatin, is essential to understand the function of this enigmatic protein in health and disease. Here we present a rigorous biochemical analysis of the complexes formed between BCL7 protein and nucleosomes and we characterize a truncated version of BCL7 proteins (encompassing amino acids 1-100) and show that the first 100 amino acids of BCL7 proteins are sufficient for nucleosome binding. We find that single amino acids mutations within the conserved N-terminal part of BCL7 proteins, which are frequent in cancer patients, impair nucleosome binding.

RESULTS

BCL7 proteins are chromatin associated proteins that are part of the human SWI/SNF complex. The BCL7 family is formed by three proteins encoded by three different genes: BCL7A, B and C. They lack any domain of known molecular function or shared with other proteins. According to sequence alignment among the BCL7 family members and among different species the first 50 amino acids of the BCL7 protein sequence are conserved (Fig. 1).

Human BCL7A isoform 1 is composed of 231 amino acids. Circular dichroism (CD) data of full length human recombinant BCL7A indicate that it is about 75% unstructured consistent with theoretical predictions (Fig. 2) and that the structured region corresponds to the N-terminus of the protein (first 50 conserved amino acids).

BCL7 proteins form complexes with nucleosomes and free DNA

In initial experiments, we noticed that the three BCL7 proteins tend to co-purify with significant amounts of DNA, suggesting a possible DNA-binding function. Electrophoretic mobility shift assays (EMSA) confirmed this intuition. EMSA performed with recombinant human BCL7 proteins and free DNA or nucleosomal probes revealed that they can bind both types of substrates (Fig. 3), further indicating that they may indeed participate in chromatin remodelling. Nucleosomes reconstituted with Cy5 labelled 147 bp DNA fragment derived from the strong positioning '601' DNA sequence were incubated for 30 minutes on ice with increasing amounts of full length human recombinant BCL7 proteins and the complexes were resolved on a 5% native gel polyacrylamide gel. The addition of increasing amounts of BCL7 protein resulted in a progressive reduction of free

nucleosomes and the appearance of shifted bands indicating complex formation. The same behavior was observed with BCL7A (1-100), a truncated version of BCL7A protein that contains the conserved N-terminus, providing evidence that the first 100 amino acids of BCL7A are sufficient for nucleosome binding (Fig. 3). The binding affinity of BCL7A (1-100) to the nucleosome seems however reduced when compared to the full-length protein. Further, EMSA experiments showed that BCL7 proteins bind equally well free Cy5 labelled 601 DNA (Fig. 3), indicating that they do not discriminate against histone-bound nucleosomal DNA and free DNA. To confirm that BCL7 proteins do not exhibit preference for free DNA or nucleosomal DNA, a competition assay was performed between nucleosomes or free 601 DNA and a non-specific competing DNA fragment (salmon sperm DNA). Figure 4 demonstrates that addition of non-specific DNA to the reactions of BCL7A with nucleosome or 601 DNA results in a decrease of the nucleosome-BCL7A or 601 DNA-BCL7A complexes respectively. Together, our results demonstrate that BCL7 proteins bind well both nucleosomes and free DNA.

BCL7A does not have any binding preference for different DNA substrates

To investigate whether BCL7A exhibits any binding preference toward different DNA substrates, similar EMSA assays were performed with BCL7A full length and short Cy5 labelled double stranded DNA probes (12 nucleotides long). Three types of substrates were tested to screen for the impact of base composition (the substrates were enriched for adenosine(A) and thymidine(T) or guanosine(G) and cytosine(C)), of cytosine methylation, and of ends shape (blunt or overhang ends) (Fig. 5). BCL7A recognised with no preference methylated DNA and non-methylated DNA, GC and

AT rich DNA fragments. BCL7A didn't show any preference toward different DNA ends either. This observation was further confirmed by performing EMSA experiments incubating increasing amounts of BCL7A with a plasmid treated with enzymes that generate different DNA ends: 3' overhang, 5' overhang and blunt (Fig. 6). These experiments show that BCL7 proteins bind free DNA with no sequence or length preference.

BCL7A contributes to nucleosome remodeling *in vitro*

To examine the activity of BCL7 proteins on their substrate the nucleosome, we performed a chromatin remodeling assay with the recombinant ATPase subunit BRG1 in presence and absence of BCL7A. Nucleosome were reconstituted with a Cy5 labelled DNA sequence corresponding to the 601 DNA positioning sequence carrying additional 50 and 54 extra nucleotides on each side (50N54 601 DNA) to allow nucleosome sliding. As shown in figure 7 BCL7A was able to enhance nucleosome remodeling by BRG1 as smaller amounts of BRG1 are required to remodel chromatin in presence of BCL7A. This experiment provides insight into the role of BCL7 proteins within the SWI/SNF complex and into their role of stimulating the ATPase activity and aiding chromatin remodeling.

BCL7A forms stable complexes with nucleosomes

We conducted a rigorous biochemical analysis of the complexes formed between BCL7 proteins and the nucleosome. Our studies show that BCL7 proteins form defined complexes with the nucleosome.

SEC

The complexes were analyzed by size exclusion chromatography (SEC). Nucleosomes were incubated with a 3:1 molar excess of full length BCL7A or truncated BCL7A(1-100). The samples were injected individually on a Superose-6 Increase 3.2/300 size exclusion column. Each chromatogram displays a single symmetric peak, indicating the homogeneity of the preparations. Samples containing either full length BCL7A or BCL7A(1-100) or NCP alone were used as a control. The integrity and the composition of the complexes were assessed by running the peak fractions of the SEC purifications on native gels and SDS-PAGE gels (Fig. 8). Interestingly, full length BCL7 proteins elute in a peak corresponding to a higher molecular weight than a regular 25 kDa protein, which is probably due to lack of globular folding of BCL7 proteins and the contribution of the disordered part of the molecule.

MALS

The stoichiometry of the nucleosome-BCL7A complex was investigated using SEC-MALS. This method provides information on the sample purity and monodispersity (SEC) and exact molecular mass (MALS) of the molecules. The particle molecular weight is the basis for stoichiometry determination of complexes in solution. The NCP alone or the complex NCP-BCL7A were injected on a Superose-6 Increase 10/300 size exclusion column connected to a light scattering detector. The analysis of the molar mass at the center part of the peak reveals mono-dispersity for all samples. The NCP alone eluted in a single peak, with a measured molecular mass of 200.4 kDa (198.8 kDa calculated). A mixture of NCP and BCL7A, with the former

in molar excess, eluted in two peaks. The first peak contained both the NCP and BCL7A, and the second peak contained BCL7A only. MALS analysis shows that the second peak corresponding to BCL7A alone has a molecular mass of 29 kDa (25.0 kDa calculated), indicative of a monomeric species. The first peak yields a molecular mass of 208.7 KDa (223.8 kDa calculated) which hints to the molecular mass of a 1:1 complex. While for the NCP and BCL7A alone the measured molar mass agreed with the calculated molecular weight, for the complex NCP-BCL7A it is slightly smaller than the predicted MW corresponding to a 1:1 molar ratio complex (Table 1). This may be due to an effect of dilution of the complex in the SEC column leading to an equilibrium between complex and free NCP (Fig. 9).

AUC

To better understand the structure of the BCL7A-nucleosome complex identified by EMSA experiments, we employed sedimentation velocity in the analytical ultracentrifuge (SV-AUC). Full length BCL7A is a monomer in solution with a sedimentation coefficient of 1.84 S, which corresponds exactly to the MW of 25 KDa. The diffusion-corrected sedimentation coefficient distributions of NCP alone and in complex with BCL7A are shown in Fig. 10. Under our experimental conditions NCP alone sediments as a homogeneous 11.08 S species, consistent with earlier studies of isolated nucleosome particles[18,19] and the complex NCP-BCL7A sediments as 11.25 S species, suggesting the formation of a 1:1 molar ratio complex Fig. 10. Interestingly, no abnormality was detected in the frictional coefficient of the NCP-BCL7A complex which could be anomalously high due to the intrinsically disordered nature of BCL7 proteins. The observation fosters the idea that BCL7 proteins lose

their disordered properties upon binding the nucleosome and folding on their binding partner. Dynamic light scattering (DLS) experiments support this hypothesis. DLS is an appropriate technique to monitor expansion or compaction of protein molecules and it is particularly relevant for intrinsically disordered proteins (IDP) as it can help distinguishing from compactly folded and unfolded states. The DLS profile of BCL7A alone is extended and disordered in its free state, but it assumes a less extended conformation upon interacting with nucleosomes (data not shown).

BCL7 proteins bind the nucleosome with high affinity

Our biochemistry studies have revealed a robust interaction of BCL7 proteins with the nucleosome. In order to better characterize this interaction, we sought to determine the equilibrium dissociation constant (Kd) and determine the binding affinity of BCL7A for the nucleosome. Microscale thermophoresis (MST) experiments were performed with Cy5-labelled nucleosomes and increasing amounts of full length BCL7A. MST measures temperature induced changes in fluorescence. Using this method, we obtained a binding affinity of ~230 nM (Fig. 11), a fairly typical value for molecules that bind nucleosomes [20]. We also calculated the binding affinity for full length BCL7A by quantifying EMSA experiments performed with Cy5 labelled nucleosomes and increasing amounts of BCL7 protein and obtained a Kd of ~ 280 nM for BCL7A and ~ 300 nM for BCL7C FL (Fig. 12). The discrepancy in Kd values is imputable to the different experimental methods employed. Nevertheless, these results confirm that BCL7 proteins bind the nucleosome with high affinity. We also determined the affinity of the first 100 amino acids of BCL7A for the nucleosomes and shown that the truncated BCL7A (1-100)

has a lower affinity for the nucleosomes (Kd of ~900 nM, Fig. 12), underlying the possible important engagement of the non-conserved part of BCL7 proteins into NCP recognition.

Effect of BCL7 point mutations found in cancer patients on nucleosome binding

We investigated if BCL7 protein point mutations, detected in cancer patients and reported in the cosmic database (https://cancer.sanger.ac.uk/cosmic), would affect directly nucleosome binding. Mutations were selected according to the highest reported occurrence frequency and exact position in the protein sequence: residues in the conserved N-terminal part and in the region in close proximity were considered (Fig. 13). We tested the ability of two BCL7A mutants to bind the nucleosome: R11S, P78S. BCL7A R11S mutant is particularly interesting as R11, located in the conserved N-terminus, could be involved in binding the acid patch of the nucleosome. All the mutants were purified as the wild type proteins and correct folding and monodispersity of the proteins were assessed by SEC-MALS. The mutants eluted in a single peak, MALS analysis yielded a molecular mass of 27 kDa for BCL7A R11S and 28 kDa for BCL7A P78S (25.0 kDa calculated) (Fig. 14). We performed EMSA experiments with Cy5 labelled nucleosomes and different amounts of BCL7A WT or BCL7A R11S or BCL7A P78S mutants as mentioned above. Figure 12 shows that the binding of BCL7A R11S and P78S mutants to the nucleosome is impaired, with a calculated apparent Kd of 662 nM and 550 nM respectively, a 50% reduction compared to the WT protein.

EXPERIMENTAL PROCEDURES

Protein expression and purification.

The cDNA encoding full length human BCL7A and BCL7A (1-100), full length BCL7B and BCL7C, were subcloned into the parallel vector pGST2 that includes a TEV-cleavable, N-terminal glutathione sulfotransferase (GST) tag. Each vector encoding BCL7 proteins was transformed into the E.*coli* strain Rosetta(DE3), and cultures were grown in Terrific Broth media at 37°C to an OD_{600} of 0.6. Protein expression was induced by the addition of IPTG (0.2 mM) for 16 hr at 19°C. Cells were harvested, resuspended in lysis buffer (50 mM Tris [pH 8.0], 300 mM NaCl, 0.1% [v/v] Triton X-100, 5 mM 2-mercaptoethanol), lysed by sonication and clarified by centrifugation. The supernatant was collected and applied onto glutathione sepharose beads for 2 hours at 4°C and washed extensively with lysis buffer. GST-BCL7 was TEV cleaved on beads for 16 hours at 4°C. The untagged protein was further purified by heparin affinity chromatography followed by size exclusion chromatography (Superdex 200) preequilibrated in 20mM Tris [pH 8.0], 150mM NaCl and 1 mM DTT. BRG1 was overexpressed in Sf9 cells and purified as previously reported [21].

Nucleosomal DNA and large-scale nucleosome reconstitution.

147-bp palindromic DNA fragment derived from the strong positioning 601 DNA sequence[22] was used. To produce sufficient quantity of nucleosomes for structural studies, a large-scale protocol was employed. The DNA was prepared in milligram quantities using a plasmid with multiple concatenated copies of the 601 sequence,

followed by restriction digestion of the individual units and size selection, as reported previously[23]. Core histones were produced by co-expression in E.*coli* and were purified in native form following an established protocol[24]. The two components were assembled into NCPs by over-night dialysis to reduce NaCl from 2M to 0.25M. NCPs are heat-shifted at 37 $^{\circ}$C for 120 minutes. Proper assembly was assessed on non-denaturing polyacrylamide gels (5% acrylamide 37.5:1 acrylamide:bis-acrylamide ratio, 0.25X TGE, 0.03% v/v NP-40) pre-run for 1 hour at 4 $^{\circ}$C and run for 1 hour at 120 volts on Mini-Protean 3 gels (Bio-Rad).

Chromatin remodelling assay.

The remodelling assay used was the nucleosome sliding method[25]. A piece of 601 DNA sequence[22] was flanked by additional DNA sequences of about 50 nucleotides on each side. The DNA was made fluorescent by PCR amplification using primers with Cy5 fluorophores at their 5' ends. Fluorescent DNA was assembled into nucleosomes using the stepwise dilution method in the presence of recombinant histones. Remodeling reactions were assembled by adding the fluorescent NCP, purified recombinant BCL7A and buffer. The reactions were started upon addition of 2 mM Mg-ATP and incubation at 30ºC and they were stopped with addition of gamma-thio-ATP and glycerol. Samples were analysed on a Tris-Glycine-EDTA non-denaturing acrylamide gel. After electrophoresis, gels were immediately scanned using the phosphorimager typhoon. Remodeling was evidenced by the appearance, in ATP-containing reactions, of faster migrating species.

Electrophoretic mobility shift assays.

Fluorescent nucleosomes were assembled using the salt dilution method with recombinant histones and Cy5-labelled 601 DNA produced by PCR. The labelled probes were incubated at 10 nM in 30 µl binding reactions containing 50 mM Tris-HCl pH 8.5, 20 mM KCl, 10 mM b-mercaptoethanol, 10% (v/v) glycerol. Purified proteins were added to the binding reactions at varying concentrations, and the reactions are incubated on ice for 30 minutes. Samples were then loaded on non-denaturing acrylamide gels (5% acrylamide 37.5:1 acrylamide:bis-acrylamide ratio, 0.25X TGE, 0.03% v/v NP-40) containing 10% glycerol, using 0.5X TGE (40 mM Tris-HCl, 45 mM boric acid and 1 mM EDTA) as running buffer.

Calculation of apparent Kd:

After electrophoresis, gels were immediately scanned using the Typhoon phosphorimager, at 10 micron resolution, adjusting the PMT value so as to obtain the strongest signal without saturating any pixel. Image files were saved in 16-bit.gel format and imported into Fiji. The "Analyze Gels" function was employed to quantitate the signal in three zones representing naked DNA, free NCP and shifted signal (above NCP). The free NCP values in each lane were used, expressed as fraction of free NCP in the absence of protein, and fitted to Hill's equation:

$$1 - Fp \sim \frac{1}{\left(\frac{Kd}{[BCL7]}\right)^n + 1}$$

where Fp is the fraction of free probe (labelled NCP), Kd is the apparent dissociation constant, [BCL7] is the micromolar concentration of ligand (BCL7 protein), and n is

Hill's coefficient. Fitting was performed using the nlsLM function from the 'minpack.lm' package in R. Initial values for Kd and n were set to 0.25 and 1, respectively.

Microscale Thermophoresis experiments.

For the study of the interactions between BCL7A wild type and nucleosome, microscale thermophoresis experiments were performed using a Monolith NT.115 (NanoTemper, LLC, Munich, Germany). For all the thermophoretic experiments, premium coated capillaries were used. All data sets were collected at 25° C. Reactions were performed in buffer consisting of 20 mM HEPES pH 8.0, 150 mM NaCl, 0.05% Tween-20. BCL7A was prepared using 1:1 serial dilution of the protein into the buffer with the highest concentration of 30 μM. sixteen 10-μL samples were thus prepared ranging in concentration from 30 μM to ~1 nM. The association was initiated by the addition of an equal volume of 40 nM Cy5 labelled NCP. The mixtures were incubated in the dark at room temperature for 10 minutes before being loaded into capillary tubes and inserted into the apparatus for data acquisition. The LED power was 80% and the MST power was 20%. Data analysis were performed in the Nano Temper Analysis software using a *Kd* fitting analysis on temperature jump data. Triplicates were used for data fitting. Data points at the extremes of the range were excluded from the analysis. Straight lines were fitted to the unsaturated and saturated portions of the data in triplicate using GraphPad Prism.

Nucleosome competition assay.

Cy5 labelled nucleosomes or Cy5 labelled 601 DNA were pre-incubated with different amounts of BCL7 proteins and a constant amount of salmon sperm DNA. Reaction mixtures were kept on ice 30 minutes before resolving on a 5% native PAGE gel. The gels were visualized by fluorescence staining.

Analytical ultracentrifugation.

Sedimentation velocity (SV) experiments were performed at 20°C and 32,000 RPM in a Beckman Coulter ProteomeLab XL-I instrument (CBI-IGBMC, Integrated Structural Biology Platform). 400 µl of samples were loaded into analytical cell assemblies with 12 mm charcoal-filled Epon double-sector centerpieces and sapphire windows. Absorbance data at 280 nm and interference data were recorded. Sedimentation data were analyzed with SEDFIT software (version 16.1c) using the continuous sedimentation coefficient distribution model c(s)[26].

Size exclusion chromatography / multi-angle light scattering.

Size exclusion chromatography combined with multi-angle light scattering (SEC-MALS) was performed with a Superpose 6 (10/300 GL) column at 0.5 ml/min, 25ºC, preequilibrated in 20 mM Tris pH 8.0, 150 mM NaCl, 1 mM DTT, to separate the sample before performing the MALS measurements. BCL7A wild type and mutants R11S and P78S, nucleosome alone or in presence of a 1:3 molar excess of BCL7A were individually injected in a volume of 50µl of sample at 5 mg/ml. The molar mass for each molecule was determined with the ASTRA software (Wyatt Technologies) SEC-MALS is an accurate tool for determining the molar mass of proteins in solution.

DISCUSSION

BCL7A is characterized by an N-terminal motif of about 50 amino acids that is ordered and conserved and a C-terminal tail that is disordered and non-conserved. To date, no structural information exists on BCL7 proteins and their function is poorly characterized as well. Interestingly the ordered N-terminus is well conserved but does not resemble any domain of known molecular function.

BCL7 proteins have been recently identified as part of the human SWI/SNF, which is a multiprotein complex that remodels chromatin. So far only the ATPase subunit and BAF47 have been reported to interact directly with the nucleosome and it is unclear whether and where other subunits interact directly with the nucleosome. Here we report evidence that the BCL7 proteins bind the nucleosomes and free DNA with high affinity.

Nucleosome reconstituted with 147 bp 601 DNA sequence have highly ordered and tightly associated DNA ends. Tightly folded nucleosomes are a good substrate for BCL7A proteins, thus we did not investigate further the binding to nucleosomes with linker DNA or nucleosomes with histone variants that give rise to less-tightly organized DNA ends.

We present biochemical characterization of BCL7 proteins and in particular of BCL7A. We demonstrate that one molecule of BCL7A forms a well-defined complex with nucleosomes, without displacing any of the core histones. We also confirmed that the N-terminal part of BCL7A (BCL7A (1-100)) is sufficient to nucleosome binding and we isolated a stable complex by SEC. However, the binding affinity of the truncated BCL7A for the nucleosome is reduced, indicating that the

rest of the molecule has a role in nucleosome recognition or into stabilizing the complex.

In our experimental conditions, we characterized that BCL7A does not display preferential binding for any specific DNA sequence or different DNA ends and it also binds DNA without ends (circular plasmid). As consequence, we postulate that binding specificity to the chromatin fibre is driven by putative contacts with the histones, or through other SWI/SNF subunits.

We also investigated how BCL7A may mediate nucleosome remodeling activity *in vitro* by performing remodeling assays with the recombinant ATPase subunit BRG1 in presence and absence of BCL7A. We noticed that BCL7A aids chromatin remodeling probably by unwrapping the DNA that protects the histones. As BCL7A does not seem to bind specifically to DNA ends, it is unlikely that the molecule unwraps the entry/exit points of the DNA, but probably by bulging DNA away from the histone octamer or by translocating the DNA sequence around the octamer surface.

Keeping in consideration the SEC-MALS profiles of the BCL7A-NCP complex, the Kd calculated by the EMSA experiments, and the DLS spectra, we argue that BCL7A C-terminus contributes to NCP binding and undergoes induced folding upon recognition of its partner, becoming ordered. Further studies await to be conducted to investigate the role of the non-ordered region of BCL7 proteins in chromatin remodeling. The high conformational flexibility of BCL7 C-terminus probably allows molecular mechanisms that are unlikely for ordered proteins. BCL7 proteins functionality most likely requires at least some degree of conformational flexibility and structural dynamics.

The mSWI/SNF is interesting from a medical perspective because one or more of its components is disrupted in human diseases. A recent report showed that 19% of human cancer exomes display mutations to mSWI/SNF components[27] , and several subunits function as bona fide tumour suppressors; their combined mutation rate approaches that of the most frequently mutated tumour suppressor, p53. The newly identified subunits contribute significantly to this phenomenon; for instance, BCL7A is the target of driver mutations in diffuse large B-cell lymphoma[9] . However, the impact of all these mutations on mSWI/SNF function and on the cancerous phenotype is poorly characterized. Oncology geneticists have accumulated a formidable amount of data linking mSWI/SNF mutations to cancer cases, yet comparatively little has been done to translate this knowledge into better patient care. Thus, there is strong evidence that mutations to the core and the novel mSWI/SNF components are linked to cancer. However, biochemical insight on the molecular function, structure and importance of these proteins, and impact of mutations, is currently incomplete, blocking translational efforts.

We biochemically evaluated the impact on nucleosome recognition of two point mutations reported in patients affected by different cancers: R11S and P78S, both located in the N-terminal part of BCL7A. The zygosity of these mutations is unknown. The protein mutants were well-behaved according to SEC-MALS analysis, thus, the mutations don't affect the structural integrity of BCL7 proteins. Using EMSA experiments we calculated the binding affinity of these protein mutants and report that the mutations decrease binding affinity for the nucleosome by about 50% (Fig. 12). The number of reported patients affected by haematological malignancy (diffuse large B cell lymphoma [DLBCL]) carrying the BCL7A R11 mutation[28,29,30] increased

significantly in the recent years and it is now considered a hotspot. R11 is particularly interesting as it is located in the conserved, structured region of BCL7A, is part of a predicted α-helix, and we speculate that R11 could be a canonical arginine (arginine anchor[31]) that inserts into the narrow cavity defined by α2 and α3 helices of H2A and the C-terminal helix of H2B, that form the nucleosome acid patch. To date, the specific contribution of the arginine to nucleosome binding and enzymatic activity have only been reported for a few nucleosome complexes [31,32,33,34]. It is interesting to consider that BCL7 interacts directly with BAF47[10] (and data not shown) and that the C-terminal α helix of BAF47 (CTD domain) binds the acid patch of the nucleosome[17]. Competition experiments performed with the LANA peptide[17], known to bind the canonical acidic patch of the nucleosome, do not abrogate binding of BAF47 CTD domain to the nucleosome, hinting to the fact that BAF47 may bind to other exposed regions of the nucleosome. Therefore, BCL7A could bind the canonical acidic patch on the nucleosome via R11 and simultaneously BAF47. BCL7A may also modulate the binding of BAF47 to the acidic patch. It would be informative to perform nucleosome competition assays with the LANA peptide and BCL7 proteins to confirm binding to the canonical acidic patch and to test whether BAF47 binding to the NCP is modulated by the presence of BCL7A. It has been reported that R11 in BCL7B is a site of the post translational modification (PTM) poly-ADP-ribosylation (PARylation)[35] and R11 is highly conserved across the BCL7 family. This is especially interesting because PTMs are thought to regulate interaction of chromatin proteins with the nucleosome[36]. There is the possibility that PARylation or other PTMs may somehow regulate association/dissociation of BCL7 and the nucleosome.

BCL7A P78 mutant was selected for its reported frequency in patients affected by lung adenocarcinoma[37] and vicinity to the N-terminal ordered part of the molecule. Besides the acidic patch, the two nucleosomes regions that are critical for chromatin proteins binding are the histone elbow, formed by H3 histone α1 helix and L1 loop and the histone H2B C-terminal helix. It has been reported that SWI/SNF ATPase subunit (Snf2) recognize the histone elbow[38]. As BCL7 proteins are direct binding partners of the ATPase BRG1, it is possible that BCL7A region encompassing P78 could be interacting with H2B C-terminal helix, due to the close vicinity to the acidic patch. In addition to binding to exposed histone surfaces of the nucleosome, many proteins interact with DNA that wraps around the histone octamer. It is also possible that P78 and surrounding residues interact with the nucleosomal DNA and it would be important to test how these mutants behave on naked 601 DNA. It would be interesting as well to test other disease associated mutations located in the C-terminus of the protein, such as L210 and see if the mutation affects nucleosome or DNA binding.

Our biochemical studies provide a molecular basis for understanding the deleterious effects of some missense mutations occurring in BCL7 proteins on binding to chromatin. Structural information on BCL7 proteins would allow mapping of these mutation on the structure and provide further mechanistic insight on the role of the mutations on BCL7 function.

While in recent years the increasing number of cryo-EM structures has provided a better structural insight of chromatin proteins and nucleosome binding, their functional characterization is still lagging behind. Mechanistic investigation of chromatin proteins is essential to be able to contribute to biomedical innovations.

REFERENCES

1.   Clapier, C. R. & Cairns, B. R. The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* **78**, 273–304 (2009).

2.   Narlikar, G. J., Sundaramoorthy, R. & Owen-Hughes, T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* **154**, 490–503 (2013).

3.   Middeljans, E. *et al.* SS18 Together with Animal-Specific Factors Defines Human BAF-Type SWI/SNF Complexes. *PLoS ONE* **7**, e33834 (2012).

4.   Kadoch, C. *et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* **45**, 592–601 (2013).

5.   Marcon, E. *et al.* Human-Chromatin-Related Protein Interactions Identify a Demethylase Complex Required for Chromosome Segregation. *Cell Rep.* **8**, 297–310 (2014).

6.   Huttlin, E. L. *et al.* The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **162**, 425–440 (2015).

7.   Wan, C. *et al.* Panorama of ancient metazoan macromolecular complexes. *Nature* **525**, 339–344 (2015).

8.   Forbes, S. A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43**, D805–D811 (2015).

9.   Reddy, A. *et al.* Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell* **171**, 481-494.e15 (2017).

10.   Mashtalir, N. *et al.* Modular Organization and Assembly of SWI/SNF Family Chromatin Remodeling Complexes. *Cell* **175**, 1272-1288.e20 (2018).

11.   Izhar, L. *et al.* A Systematic Analysis of Factors Localized to Damaged Chromatin Reveals PARP-Dependent Recruitment of Transcription Factors. *Cell Rep.* **11**, 1486–1500 (2015).

12.   He, S. *et al.* Structure of nucleosome-bound human BAF complex. *Science* **367**, 875–881 (2020).

13.   Wagner, F. R. *et al.* Structure of SWI/SNF chromatin remodeller RSC bound to a nucleosome. *Nature* **579**, 448–451 (2020).

14.   Han, Y., Reyes, A. A., Malik, S. & He, Y. Cryo-EM structure of SWI/SNF complex bound to a nucleosome. *Nature* **579**, 452–455 (2020).

15.   Mashtalir, N. *et al.* A Structural Model of the Endogenous Human BAF Complex Informs Disease Mechanisms. *Cell* **183**, 802-817.e24 (2020).

16.   Liu, X., Li, M., Xia, X., Li, X. & Chen, Z. Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure. *Nature* **544**, 440–445 (2017).

17.   Valencia, A. M. *et al.* Recurrent SMARCB1 Mutations Reveal a Nucleosome Acidic Patch Interaction Site That Potentiates mSWI/SNF Complex Chromatin Remodeling. *Cell* **179**, 1342-1356.e23 (2019).

18.   Bradbury, E. M. K. E. Van Holde. Chromatin. Series in molecular biology. Springer-Verlag, New York. 1988. 530 pp. $98.00. *J. Mol. Recognit.* **2**, i–i (1989).

19.   Yang, C., van der Woerd, M. J., Muthurajan, U. M., Hansen, J. C. & Luger, K. Biophysical analysis and small-angle X-ray scattering-derived structures of MeCP2-nucleosome complexes. *Nucleic Acids Res.* **39**, 4122–4135 (2011).

20.   Corbeski, I. *et al.* Microscale Thermophoresis Analysis of Chromatin Interactions. in *Bacterial Chromatin* (ed. Dame, R. T.) vol. 1837 177–197 (Springer New York, 2018).

21.   Phelan, M. L., Sif, S., Narlikar, G. J. & Kingston, R. E. Reconstitution of a Core Chromatin Remodeling Complex from SWI/SNF Subunits. *Mol. Cell* **3**, 247–253 (1999).

22.   Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).

23.   Dyer, P. N. *et al.* Reconstitution of Nucleosome Core Particles from Recombinant Histones and DNA. in *Methods in Enzymology* vol. 375 23–44 (Elsevier, 2003).

24.   Shim, Y., Duan, M.-R., Chen, X., Smerdon, M. J. & Min, J.-H. Polycistronic coexpression and nondenaturing purification of histone octamers. *Anal. Biochem.* **427**, 190–192 (2012).

25.   Eberharter, A., Längst, G. & Becker, P. B. A Nucleosome Sliding Assay for Chromatin Remodeling Factors. in *Methods in Enzymology* vol. 377 344–353 (Elsevier, 2003).

26.   Schuck, P. Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophys. J.* **78**, 1606–1619 (2000).

27.   Kadoch, C. *et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* **45**, 592–601 (2013).

28.   Lacy, S. E. *et al.* Targeted sequencing in DLBCL, molecular subtypes, and outcomes: a Haematological Malignancy Research Network report. *Blood* **135**, 1759–1771 (2020).

29.   Gunawardana, J. *et al.* Recurrent somatic mutations of PTPN1 in primary mediastinal B cell lymphoma and Hodgkin lymphoma. *Nat. Genet.* **46**, 329–335 (2014).

30.   Pillonel, V. *et al.* High-throughput sequencing of nodal marginal zone lymphomas identifies recurrent BRAF mutations. *Leukemia* **32**, 2412–2426 (2018).

31.   McGinty, R. K., Henrici, R. C. & Tan, S. Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature* **514**, 591–596 (2014).

32.   Pathare, G. R. *et al.* Structural mechanism of cGAS inhibition by the nucleosome. *Nature* **587**, 668–672 (2020).

33.   Zhao, B. *et al.* The molecular basis of tight nuclear tethering and inactivation of cGAS. *Nature* **587**, 673–677 (2020).

34.   Worden, E. J., Zhang, X. & Wolberger, C. Structural basis for COMPASS recognition of an H2B-ubiquitinated nucleosome. *eLife* **9**, e53199 (2020).

35.   Gibson, B. A. *et al.* Chemical genetic discovery of PARP targets reveals a role for PARP-1 in transcription elongation. *Science* **353**, 45–50 (2016).

36.   McGinty, R. K. & Tan, S. Principles of nucleosome recognition by chromatin factors and enzymes. *Curr. Opin. Struct. Biol.* **71**, 16–26 (2021).

37.   Shi, J. *et al.* Somatic Genomics and Clinical Features of Lung Adenocarcinoma: A Retrospective Study. *PLoS Med.* **13**, e1002162 (2016).

38.   Liu, X., Li, M., Xia, X., Li, X. & Chen, Z. Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure. *Nature* **544**, 440–445 (2017).

FIGURE 1

```
BCL7A    MSGRSVRAETRSRAKDDIKRVMAAIEKVRKWEKKWVTVGDTSLRIYKWVPVTEPKVDDKN        60
BCL7B    MSGRSVRAETRSRAKDDIKKVMAAIEKVRKWEKKWVTVGDTSLRIFKWVPVTDSKEKEKS        60
BCL7C    MAGRTVRAETRSRAKDDIKKVMATIEKVRRWEKRWVTVGDTSLRIFKWVPVVDPQEEERR        60

BCL7A    KNKKKGKDEKC---GSEVTTPENSSSPGMMDMHDDNSNQSSIADASPIKQENSSNSSPAP       117
BCL7B    KSNSSAARE----PNGFP-SDASANSSLLLEFQDENSNQSSVSDVYQLKVDSSTNSSPSP       115
BCL7C    RAGGGAERSRGRERRGRGASPRGGGPLILLDLNDENSNQSFHSEGSL-QKGTEPSPGGTP       119

BCL7A    EPNSAVPSDGTEAKV-DEAQADG-----KEHPGAEDAS--DEQNSQSS----------ME       159
BCL7B    QQSESLSPAHTSDFRTDDSQPPTLGQEILEEPSLPSSEVADEPPTLTKEEPVPLETQVVE       175
BCL7C    QPSRPVSPAGPPEGVPEEAQPPRLGQERD--PGGITAGSTDEPPMLTKEEPVPELLEA--       175

BCL7A    HSMNSSEKVDRQPSGDSGLAAETSA---------------ISQVPRSRSQRGS-QIGREP       203
BCL7B    EEEDSGAP--------------PLKRFCVDQPT----------VPQTASES---------       202
BCL7C    --EDSGVRMTRRALHEKGLKTEPLRRLLPRRGLRTNVRPSSMAVPDTRAPGGGSKAPRAP       233

BCL7A    IGLSGDLEGVPPSKKMKLEASQQNSEEM             231
BCL7B    ---------------------------             202
BCL7C    RTI-------PQGKGR------------             242
```

Figure 1. Multiple sequence alignment of BCL7 protein family.
Sequence alignment of human BCL7 proteins was performed with Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and shows that the first 50 amino acids of the BCL7 proteins are highly conserved (green shade). The violet shade shows another patch of interestingly conserved amino acids.
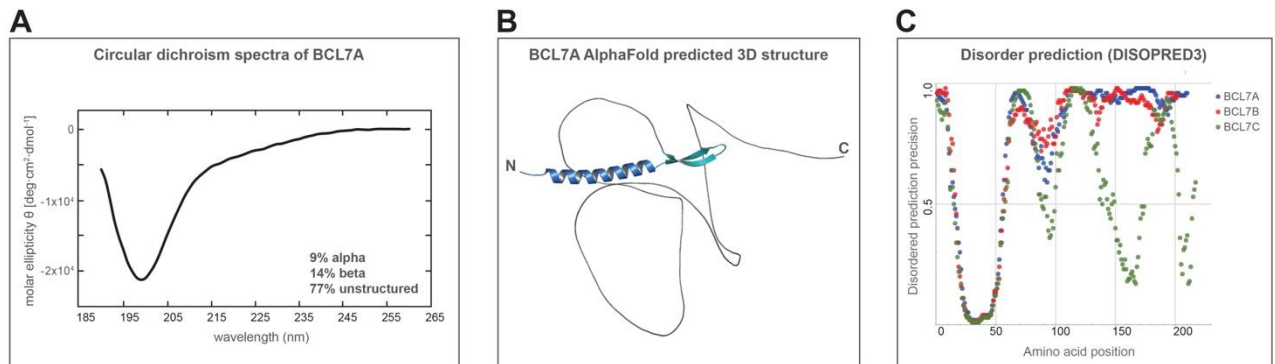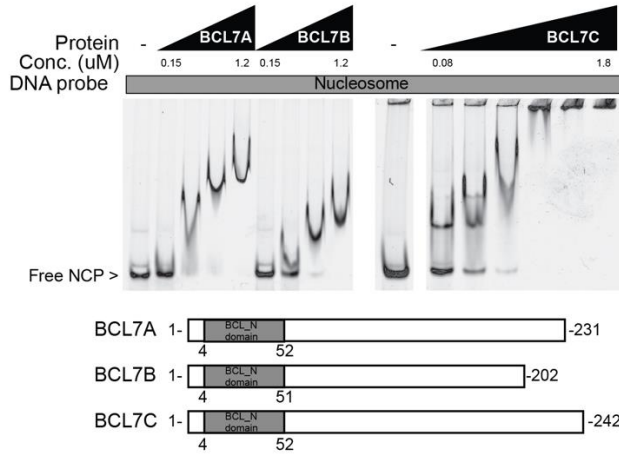
FIGURE 2

Figure 2. Prediction of BCL7A ordered and disordered regions.
A) Circular dichroism spectra of human BCL7A; **B)** AlphaFold computation of BCL7A three-dimensional structure; **C)** DISOPRED3 predictions of BCL7A disordered nature. The three approaches converge into confirming that the N-terminus (first 50 amino acids) of BCL7A are ordered and the rest of the molecule is disordered.
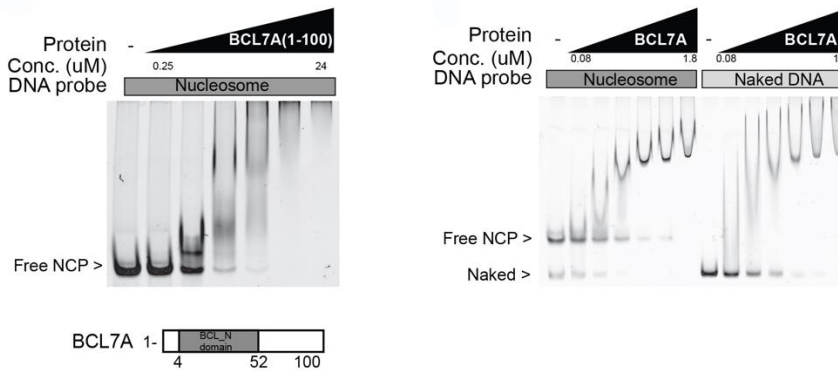
FIGURE 3

A

B

Figure 3. BCL7 proteins bind the nucleosome core particle and free DNA.
A) Electrophoretic mobility assay (EMSA) confirmed that all three BCL7 proteins bind the NCP. **B)** The first 100 amino acids of BCL7A proteins are sufficient for NCP binding. **C)** BCL7 proteins bind well both the 601-DNA and the NCP. Domain architecture of BCL7 proteins is shown at the bottom of the insets.
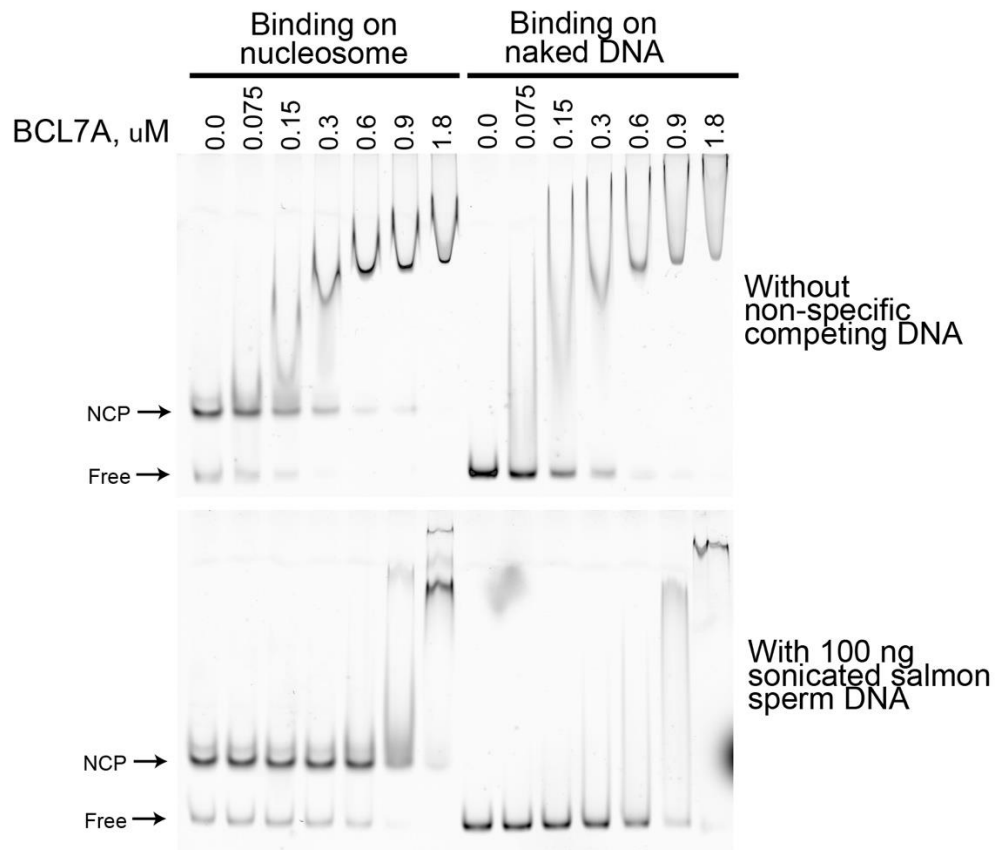
FIGURE 4

Figure 4. BCL7 proteins bind free DNA. EMSA assays were performed with increasing amount of BCL7A and Cy5 labelled nucleosomes or free 601 DNA in presence of salmon sperm DNA added as competitor. The competitor is able to abrogate binding of BCL7A to nucleosomes and 601 DNA, demonstrating that free DNA competes binding of BCL7 proteins to nucleosomes and free 601 DNA.
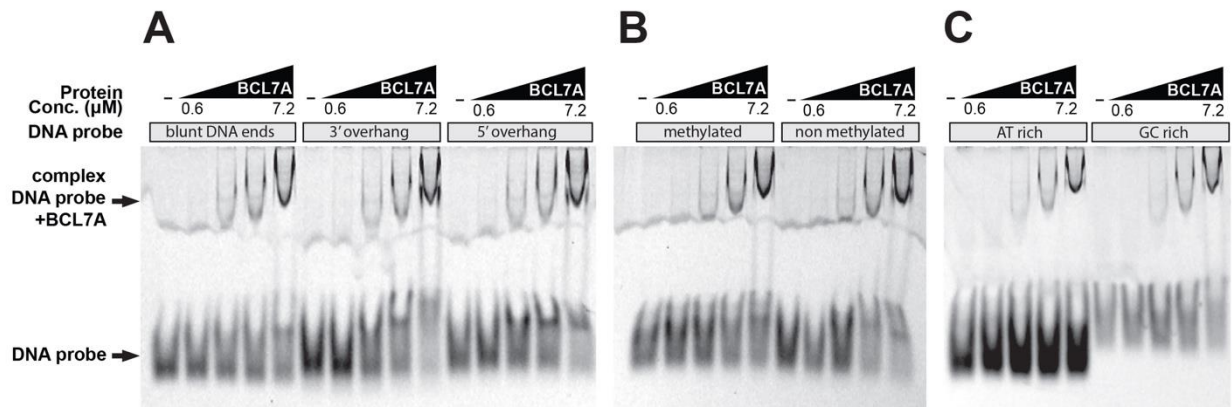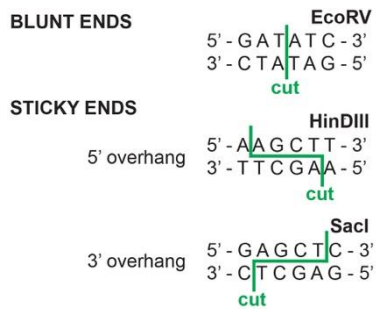
FIGURE 5

Figure 5. BCL7A binding preference over different DNA probes. EMSA experiments were performed with increasing amounts of BCL7A WT and a variety of Cy5-labelled DNA probes to test binding preference over different DNA ends **(A)**, DNA methylation **(B)** and DNA base composition **(C)**. BCL7A does not show any preferential binding specificity.
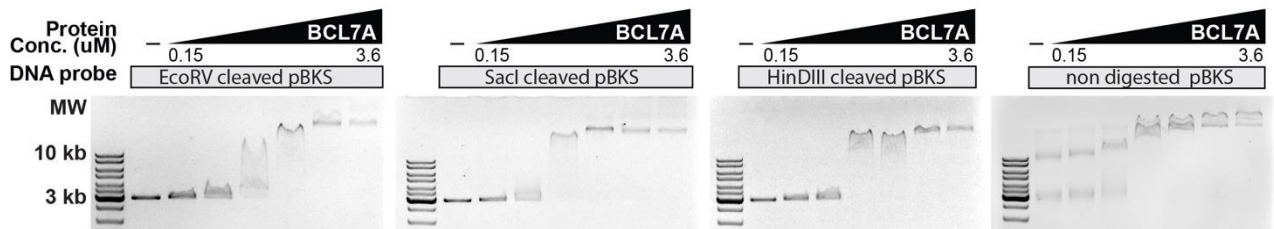
FIGURE 6

A

B

Figure 6. BCL7A binding specificity A) Restriction enzyme cutting and generation of different DNA ends. EcoRV cut produces blunt DNA ends. HinDIII cut produces 5' single-stranded DNA and SacI cut produces 3' single stranded DNA. **B)** Binding preference of BCL7A toward different DNA ends. The plasmid pBKS was digested with EcoRV or SacI or HinDIII enzymes to obtain DNA substrates respectively with blunt, 3' single-standed DNA or 5' single-stranded DNA ends. Increasing amounts of BCL7A were incubated with equal amounts of the different DNA substrates and the reactions were resolved by agarose gel electrophoresis. BCL7A does not show any binding preference toward different DNA ends. BCL7 interacts with non-digested DNA as well, showing that BCL7 does not specifically interacts with DNA ends.
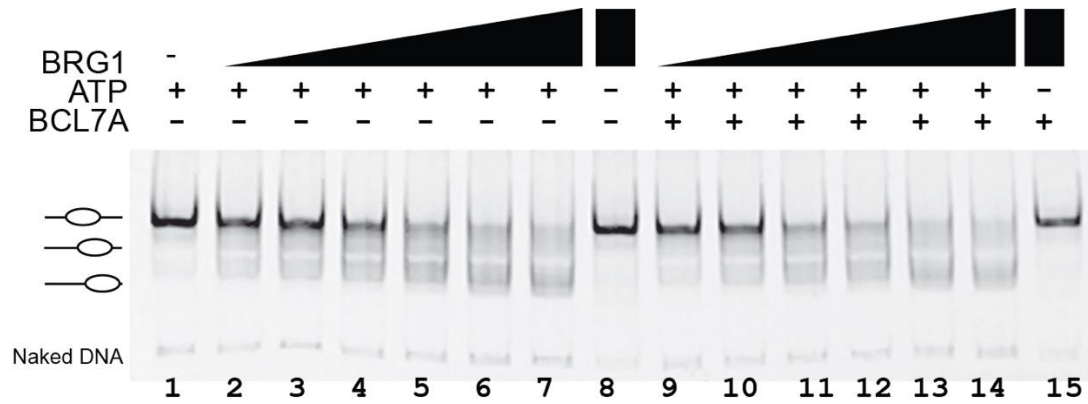
FIGURE 7

Figure 7. BCL7A aids remodeling activity of BRG1 *in vitro*.

Nucleosomes were assembled with Cy5 labelled 50N54 601 DNA and the remodelling assay was performed with increasing amounts of recombinant BRG1 alone or with a fixed amount of BCL7A. BCL7A increases the intensity of the remodelled band (e.g. comparing line 4 and line 11). No remodelling occurs without ATP.
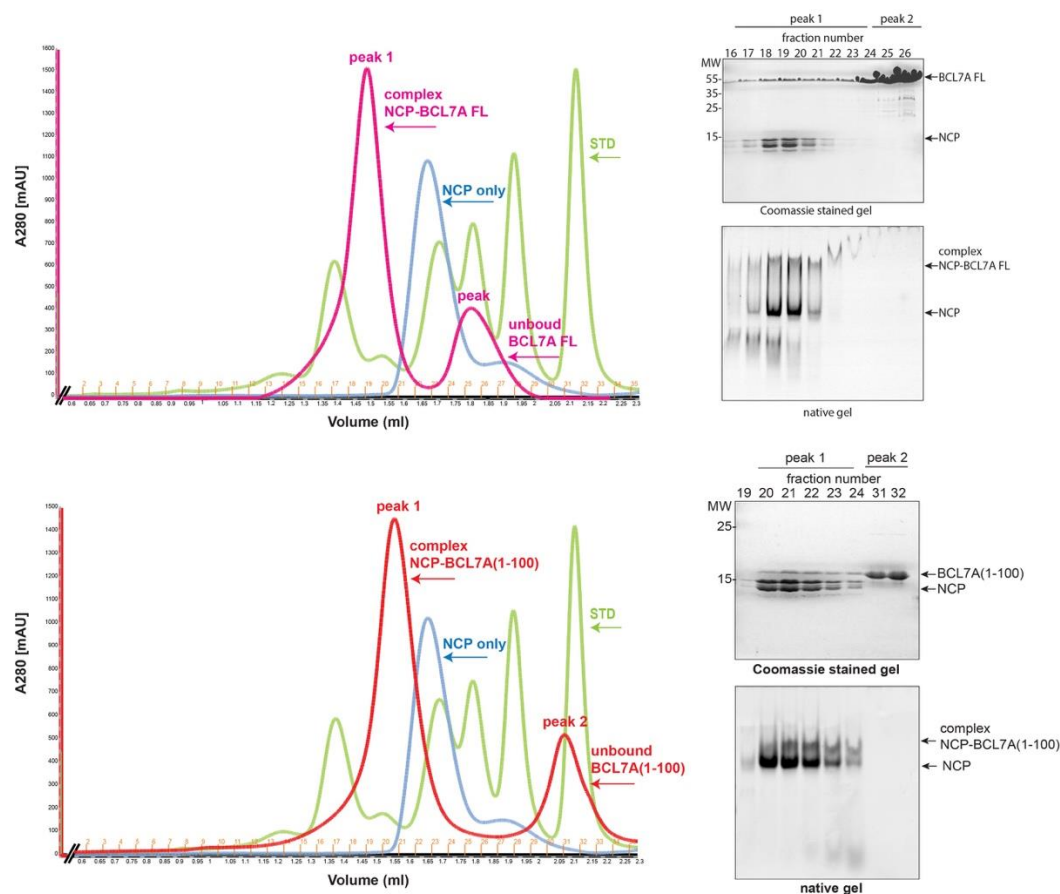
FIGURE 8

Figure 8. In vitro characterization of the NCP-BCL7A interaction.

Size exclusion chromatography profile (Superose 6 Increase 3.2/300) of the NCP-BCL7A FL complex (fig. **8A**) and NCP-BCL7A (1-100) complex (fig. **8B**) overlaid on the profile of NCP alone and molecular weight standards, plotted on the same relative scale. A_{280} traces from the SEC runs are shown as coloured solid lines; the complex NCP-BCL7A FL is coloured magenta (fig. **8A**), the complex NCP-BCL7A (1-100) is coloured red (fig. **8B**), the NCP alone is coloured blue, the molecular weight standards are coloured green. Proteins and complexes are indicated in corresponding colours. Coomassie stained SDS-PAGE gel of the peak fractions is shown on the top right panel confirming the presence of the subunits and complex formation (fig. **8A**: lanes 17-22 complex NCP-BCL7A FL, lanes 23-25 unbound BCL7A; fig. **8B**: lanes 19-24 complex NCP-BCL7A(1-100), lanes 31-32 unbound BCL7A(1-100). The EMSA of the respective fractions is shown on the bottom right panel and confirms the complex formation and nucleosome integrity. Nucleic acids (601 DNA) were detected on the native gel by staining with RedSafe.
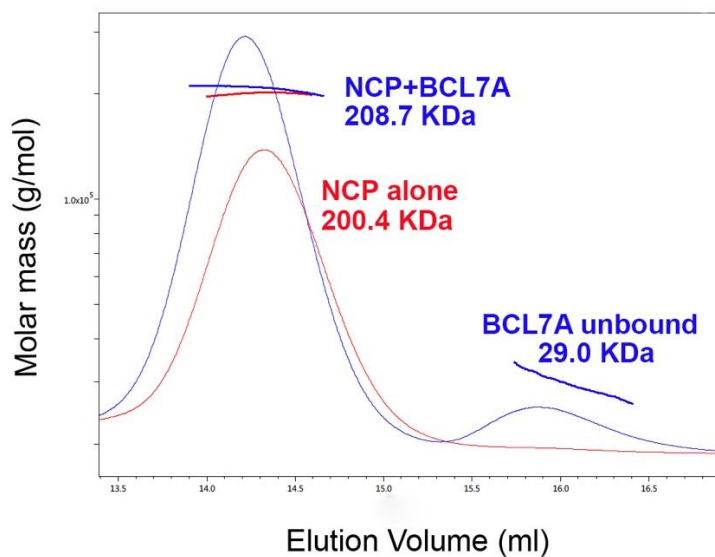
FIGURE 9

Figure 9. SEC-MALS analysis of nucleosomes and nucleosomes-BCL7A complex

The MALS-derived molecular mass distributions are plotted as individual points in the colors corresponding to the A_{280} traces, with the scale shown on the left-hand side. NCP only is colored red, NCP-BCL7A complex is colored blue.
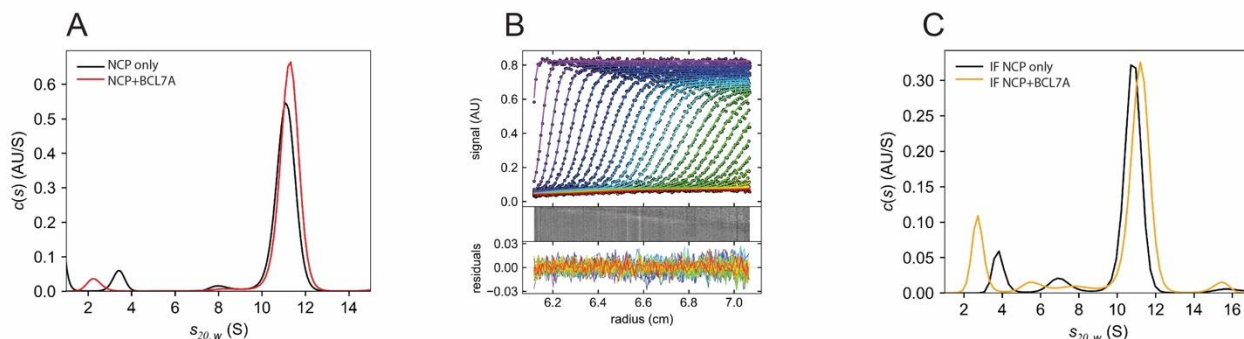
FIGURE 10

Figure 10. Analytical ultracentrifugation of BCL7A-nucleosome complex.

A) Sedimentation coefficient distribution c(s) based on absorbance data at 280 nm of NCP alone and NCP with BCL7A. Integration of the main peak gives a s-value of 11.08 S for NCP and 11.25 S for NCP+BCL7A. **B)** Superposition of the experimental and fitted sedimentation velocity profiles for NCP+BCL7A.100 scans were loaded in SEDFIT, covering the complete sedimentation process in equal time-intervals. Residuals of the fit are displayed in gray or absorbance scales. Sedimentation velocity profiles were obtained at 280 nm, at 20°C, and 32,000 rpm using double-sector centerpieces with 12 mm optical path length. Analysis was performed with the first 100 profiles, collected during 300 min. The c(s) analysis was obtained considering 200 particles with s-values in the 0–25 S range and a partial specific volume of 0.66 mL/g solvent density of 1.003 g/mL and a viscosity of 1.017 cp. A confidence level of 0.68 was used for the regularization procedure. The rmsd was 0.0073. The fitted frictional ratio was 1.50. **C)** Interference sedimentation coefficient c(s) distribution profiles of NCP (black) and NCP+BLCL7A (orange) samples. A major peak is observed at 10.8 S and 11.15 S for NCP and NCP+BCL7A respectively. SV data were analyzed into SEDFIT using the sedimentation coefficient distributions c(s) model (Schuck, 2000) and then loaded into the software GUSSI (Brautigam, 2015) for superimposed plotting and peak integration. The corresponding estimated MW are 207 KDa for nucleosome only and 223 KDa for the nuclesome-BCL7A complex.

FIGURE 11

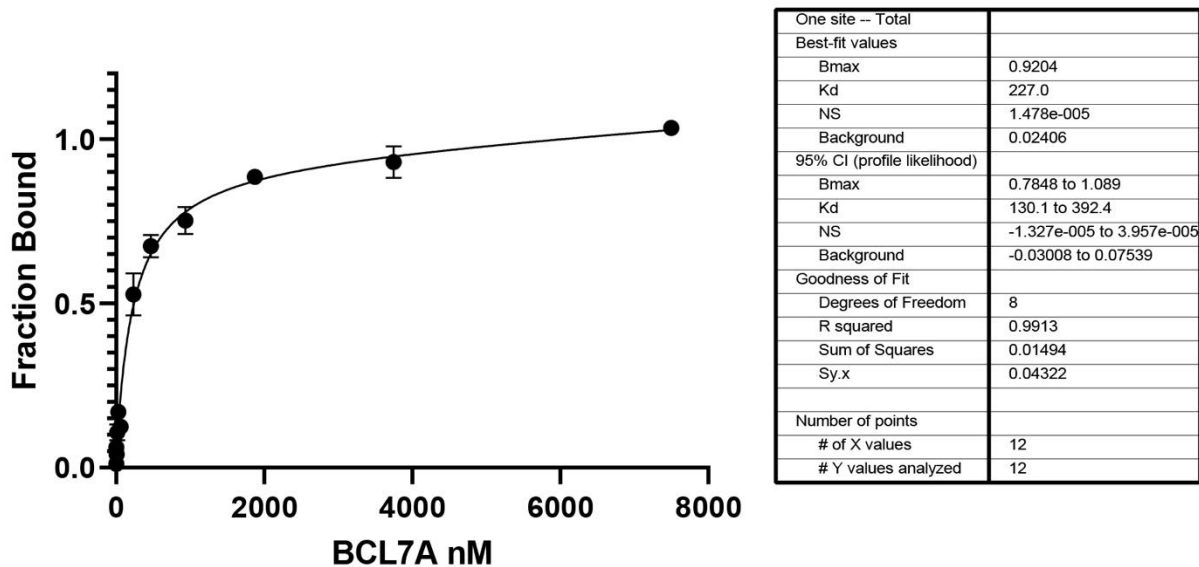| One site -- Total	
Best-fit values	
Bmax	0.9204
Kd	227.0
NS	1.478e-005
Background	0.02406
95% CI (profile likelihood)	
Bmax	0.7848 to 1.089
Kd	130.1 to 392.4
NS	-1.327e-005 to 3.957e-005
Background	-0.03008 to 0.07539
Goodness of Fit	
Degrees of Freedom	8
R squared	0.9913
Sum of Squares	0.01494
Sy.x	0.04322
Number of points	
# of X values	12
# Y values analyzed	12

Figure 11. BCL7A binding affinity for the nucleosome by microscale thermophoresis (MST). Binding affinity of BCL7A wild type to the nucleosome was calculated with MST experiments performed in triplicates. BCL7A binds the nucleosome with high affinity (Kd=230 nM).
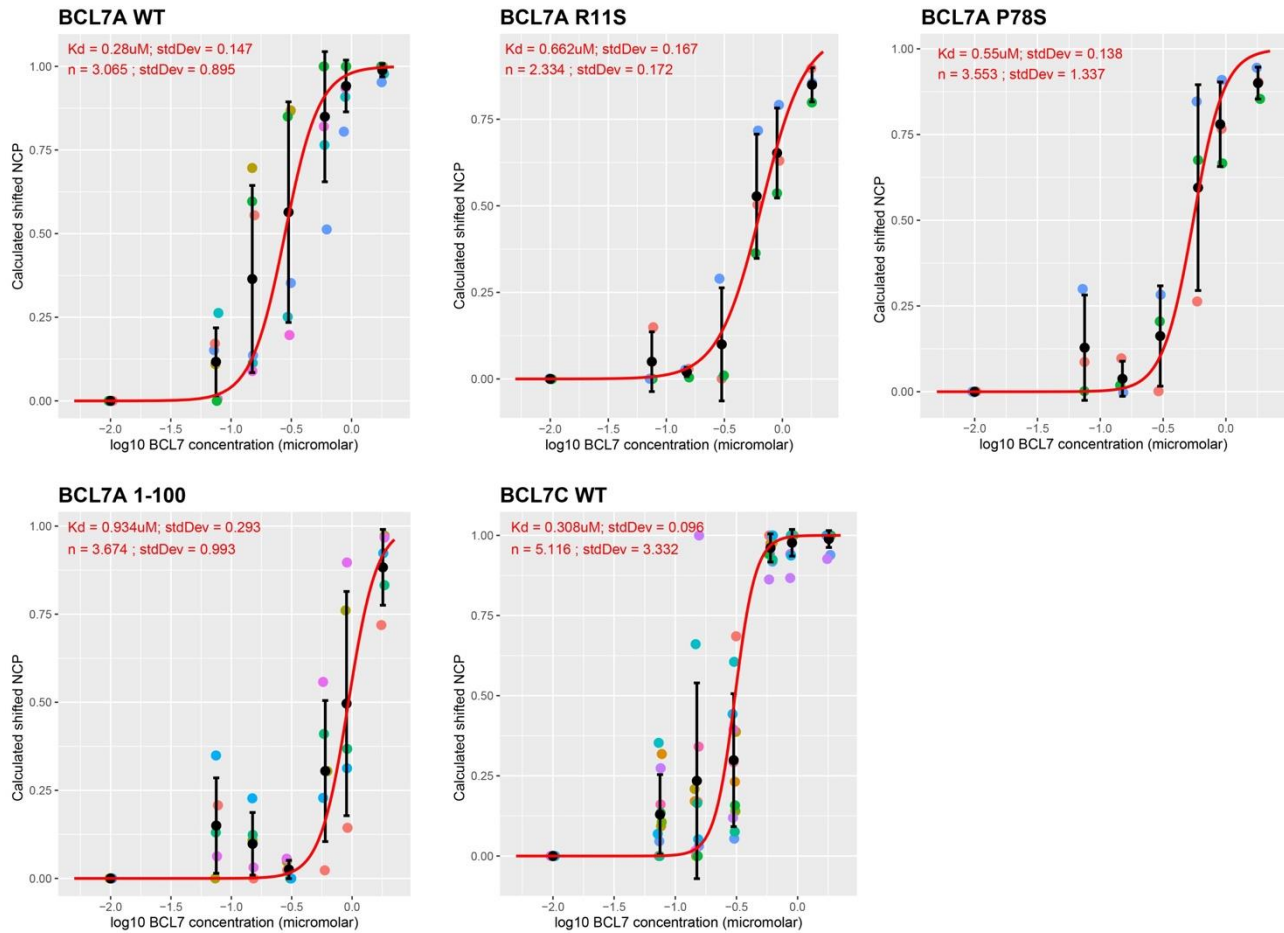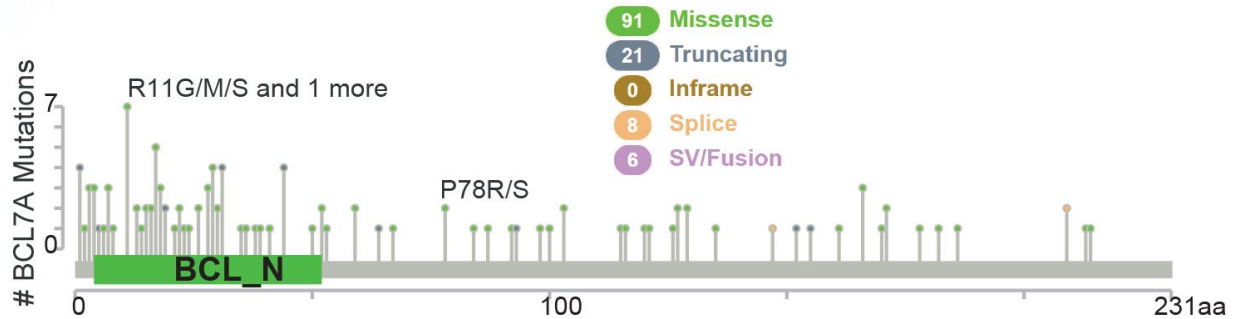
FIGURE 12

Figure 12. BCL7A binding affinity for the nucleosome calculated by electrophoresis mobility shift assay (EMSA). Several EMSA experiments were performed with Cy5 labelled nucleosomes and increasing amounts of BCL7A protein wild type, truncated BCL7A (1-100), mutants R11S and P78S and BCL7C wild type. Calculated apparent Kd values for each experiment are indicated. Each dot color represents data for one individual replicate. The black dots represent the mean of the replicates and black bars represents standard deviation. The red curves show predicted values based on the calculated Kd and Hill coefficient (n). BCL7 concentrations are shown in log10 after adding 0.01 uM.
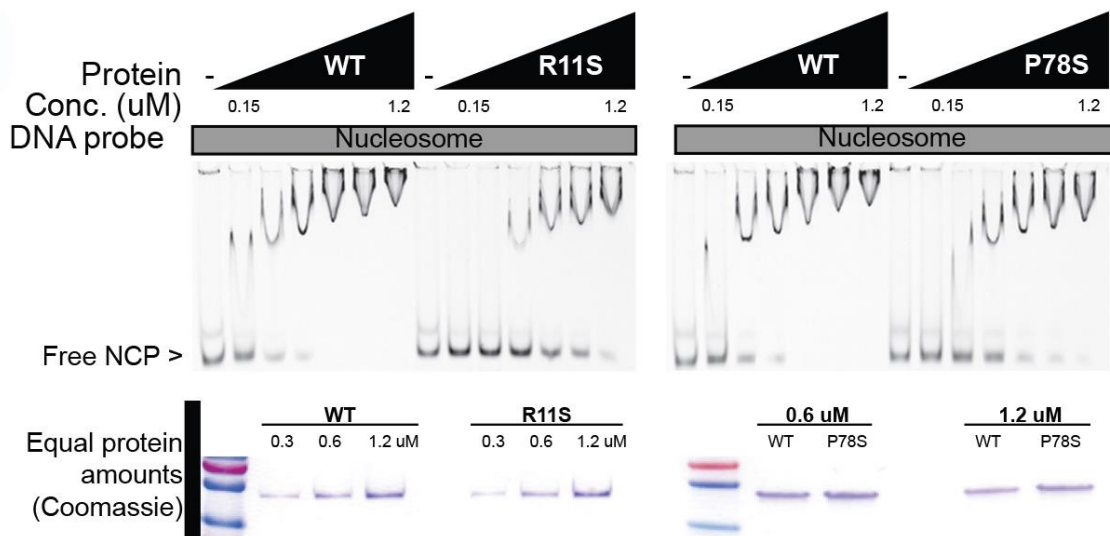
FIGURE 13

A

B

Figure 13. Mutations reported in cancer patients impair the binding of BCL7A to the nucleosome. A) The most frequent mutations found in cancer are indicated on the domain architecture representation of human BCL7A. R11S and P78S are labelled on the respective amino acids. **B)** EMSA experiments showing that BCL7A R11S and P78S mutants reduce BCL7A affinity to the NCP. Coomassie stained SDS-PAGE gels are reported at the bottom showing equal amounts of wild type and mutant proteins.
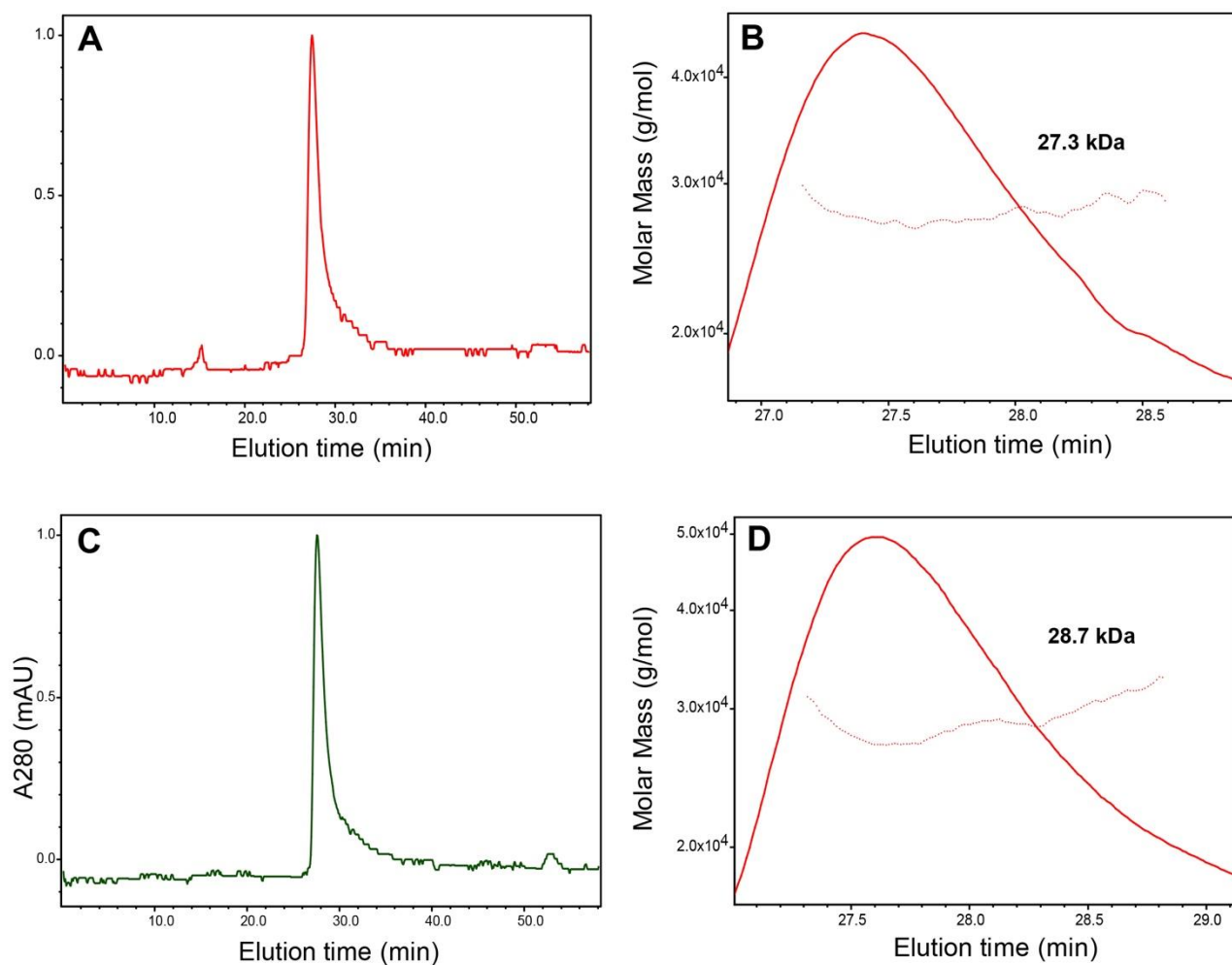
FIGURE 14

Figure 14. BLC7A mutants fold and behave like wild type protein. SEC-MALS experiments were performed with BCL7A R11S and P78S mutants and the SEC profile indicates proper elution of the mutants (left panels) and MALS confirms the exact molecular mass of monomeric well-behaved proteins (right panels). Experimental molecular mass of each mutant is reported in the respective insets.

TABLE 1

| Molecule or complex | SEC MALS Molecular weight (kDa) | |
	Observed	Calculated
NCP (147 bp DNA)	200.4 ± 0.1%	198.8
NCP-BCL7A	208.7 ± 0.2%	223.8
BCL7A WT	29.0 ± 1.8%	25.0
BCL7A R11S	27.7 ± 9.4%	25.0
BCL7A P78S	28.7 ± 10.1%	25.0

Table 1. Summary of the exact molecular masses of BCL7 proteins wild type and mutants and BCL7A-NCP complex of the experiments reported in figure 9 and figure 14.

Dana Mariel Díaz Jiménez

Structural and functional studies of the BCL7 proteins : novel subunits of the mammalian SWI/SNF complex

Résumé

La chromatine est une structure dynamique régulée par différents mécanismes épigénétiques, parmi lesquels le remodelage de la chromatine dépendant de l'ATP. La dérégulation de ce processus peut avoir de graves répercussions sur les schémas d'expression des gènes et l'intégrité du génome. Leur importance est telle que les mutations des protéines de remodelage de la chromatine sont fortement associées à plusieurs maladies, dont le cancer. Les protéines BCL7 sont de nouvelles sous-unités centrales récemment identifiées du complexe de remodelage de la chromatine ATP-dépendant SWI/SNF des mammifères. Les mutations des protéines BCL7 sont associées à différents types de cancers, y compris les hémopathies malignes. Les informations sur la fonction moléculaire et sur la structure des protéines BCL7 sont à ce jour très limitées. En utilisant des approches biochimiques et structurelles, ce projet visait à mieux comprendre la structure et la fonction de ces sous-unités auxiliaires. Nous rapportons ici que les protéines BCL7 se lient directement à la particule centrale du nucléosome (NCP) et à l'ADN libre avec une haute affinité. Nous démontrons que les protéines BCL7 forment des complexes définis avec la NCP et nous identifions la partie N-terminale conservée des protéines BCL7 comme suffisante pour la liaison au nucléosome. Nous caractérisons en outre l'impact des mutations des protéines BCL7 rapportées chez les patients cancéreux sur la liaison au NCP.

Mots-clés : remodelage de la chromatine, épigénétique, SWI/SNF, BCL7, nucléosome, cancer.

Résumé en anglais

Chromatin is a dynamic structure regulated by different epigenetic mechanisms, among which ATP-dependent chromatin remodeling. Deregulation of this process can severely impact gene expression patterns and genome integrity. Their importance is such that chromatin remodeling protein mutations are strongly associated to several diseases, including cancer. BCL7 proteins are recently identified novel core subunits of the mammalian SWI/SNF ATP-dependent chromatin remodeler complex. Mutations in BCL7 proteins are associated with different kind of cancers including blood malignancies. The information on the molecular function and on the structure of BCL7 proteins is to date very limited. Using biochemical and structural approaches, this project aimed to gain insight into the structure and function of these auxiliary subunits. Here we report that BCL7 proteins bind directly the nucleosome core particle (NCP) and free DNA with high affinity. We demonstrate that BCL7 proteins form defined complexes with the NCP and we identify the conserved N-terminal part of BCL7 proteins as sufficient to nucleosome binding. We further characterize the impact of BCL7 protein mutations reported in cancer patients on NCP binding.

Key words: chromatin remodeling, epigenetics, SWI/SNF, BCL7, nucleosome, cancer.