THE UNIVERSITY *of* EDINBURGH

Investigation into the role of LSH ATPase in chromatin remodelling

Alina Gukova

Doctor of Philosophy – University of Edinburgh – 2020

ABSTRACT

Chromatin remodelling is a crucial nuclear process affecting replication, transcription and repair. Global reduction of DNA methylation is observed in Immunodeficiency-Centromeric Instability-Facial Anomaly (ICF) syndrome. Several proteins were found to be mutated in patients diagnosed with ICF, among them are LSH and CDCA7. LSH is a chromatin remodeller bearing homology to the members of Sf2 remodelling family. A point mutation in its ATPase lobe was identified in ICF. CDCA7 is a zinc finger protein that was recently found to be crucial for nucleosome remodelling activity of LSH. Several point mutations in its zinc finger domain were described in ICF patients.

In vitro and *in vivo* studies have shown that $LSH^{-/-}$ phenotype demonstrates reduction of global DNA methylation, implying that chromatin remodelling LSH functions may be required for the efficient methyltransferase activity, linking this finding to ICF phenotype.

Here, the LSH-mononucleosome interaction was explored *in vitro* using bioinformatics, biochemical, biophysical and structural techniques. LSH purification was further optimised, achieving near 100% purity, which is a useful improvement for any potential structural studies. LSH has been found to interact with the mononucleosome *in vitro* and no DNA linker was required for this interaction, indicating that LSH binds to the nuclesomal core through its ATPase domain. Qualitatively estimated K_d for this interaction was in nanomolar region, which did not translate into complex detection during size exclusion chromatography.

CDCA7 was expressed in insect cell system and semi-purified, however, high nucleic acid presence in the final protein sample precluded any potential studies of CDCA7 interaction with chromatin.

Homology and *ab initio* modelling for LSH and CDCA7, respectively, indicated that LSH is likely to bind the superhelical location 2 (SHL2), however, the exact location of CDCA7 and its interaction with LSH will have to be elucidated in further experimental work.

LAY SUMMARY

Genetic information in living organisms is stored in a molecule called DNA. Being a long molecule, DNA required a special organisation within the cell nucleus. Specialised proteins called histones form a structural basis for DNA storage. Together, DNA and histones form structures called nucleosomes. Nucleus-wide DNA-histone mass is termed chromatin. Dynamic properties of chromatin in the nucleus are ensured by a specific class of proteins termed chromatin remodellers. LSH is a putative chromatin remodeller. A protein called CDCA7 is likely to be crucial in ensuring the remodelling activity of LSH. When mutated or truncated, LSH supposedly loses its remodelling activity and is implicated in a rare genetic disease called Immunodeficiency, centromeric instability, facial anomaly (ICF) syndrome. CDCA7, when mutated, is also implicated in ICF syndrome, indicating that they participate in a common process associated with chromatin remodelling. Here, work has been done to elucidate the mode of LSH-nuleosome interaction. Furthermore, initial work has been carried out to prepare CDCA7 for laboratory studies.

DECLARATION

The work presented in this thesis is the original work of the author. This thesis has been composed by the author and has not been submitted in whole or in part for any other degree.

Alina Gukova

February 2020

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor Dr. Julia Richardson for supervising me over these years and providing useful comments on my thesis drafts and Dr. Irina Stancheva for suggesting the LSH project. Both made me gain crucial insights into scientific work and life in general, into how things should and should not be done.

I would like to express my deep gratitude to my thesis committee: Dr. Philipp Voigt, Dr. Jeyaprakash (JP) Arulanandam and Prof. Bill Earnshaw. Philipp's help was indispensable for carrying out the histone work and making nucleosome. JP provided many important insights into protein biochemistry and structural biology aspects of my project. Bill was the one who encouraged me to explore the CDCA7 topic. Apart from the scientific expertise, all of them provided their personal encouragement and motivation. Without my thesis committee, my work on this project would not be possible. I would like to thank Dr. Chris Jenness and Dr. Hiro Funabiki for their enthusiasm and help with the CDCA7 project, without them all my numerous questions about CDCA7 and the techniques associated with it would take me even longer to work through. The Edinburgh Protein Production Facility (EPPF) team played a cardinal role in my project, providing not only their scientific expertise in the art and skill of protein expression and purification, but also their human support. So, I'm thanking Dr. Liz Blackburn, Dr. Matt Nowicki and Dr. Martin Wear for being so amazing, professionally and personally. Also a big thank you to Dr. Meng Yuan and Dr. Iain McNae for their advice on homology modelling.

I am sending my deepest gratitude to everyone on the 3rd floor of the Swann building, who made it almost pleasant to come to work every day (even including the weekends). Your positive attitude has really offset the negative aspects of a typical PhD existence. Those people include present and former members of Walkinshaw, Cook, Ly and Bramham labs. A special thank you to Dr. Simon Varzandeh, who worked on the LSH project before me and helped me a lot with it, to Dr. Jia Ning and to Dr. Ola Kasprowicz, who witnessed my struggles and supported me through the worst times in this PhD. I am also incredibly grateful to Karen Woodcock and Dr. Caroline Proctor for their emotional and practical support.

I'm very grateful to my viva examiners Prof. Malcolm Walkinshaw and Dr. Helder Ferreira for their advice and useful insights into my project.

Finally, I'd like to thank the University of Edinburgh and the School of Biological Sciences for providing financial support to me and my work and making it possible for me to carry out this work.

ABBREVIATIONS

5mC - 5 methyl-cytosine
A - amper
Å - ångström
aa – amino acid
ABC - ammonium bicarbonate
ACN - acetonitrile
ARPs - actin-related proteins
ATP - adenosine triphosphate
Bp – base pair
BS^3 - bis(sulfosuccinimidyl)suberate
BSA – bovine serum albumin
CC - coiled coil
CDCA7 - Cell division cycle-associated protein 7
CDS - coding DNA sequence
CHD - Chromodomain-Helicase DNA binding
ChIP - chromatin immuno-precipitation
C-NHEJ - non-homologous end joining
CpG – cytosine-phosphate-guanine
Cryo-EM – cryogenic electron microscopy
CV - column volumes
Da - Daltons
DBD - DNA binding domain
DBSs - double-strand breaks
ddH_2O – double-distilled water
DMSO - dimethyl sulfoxide
DNA - deoxyribonucleic acid
DNMT1 - DNA Methyltransferase 1
DNMT3A - DNA Methyltransferase 3A
DNMT3B - DNA Methyltransferase 3B
dNTP - deoxynucleotide triphosphates
dsDNA - double-stranded DNA
DTT - Dithiothreitol
dTTP – deoxythymidine triphosphate
EDC - 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA - Ethylenediaminetetraacetic acid
EMSA - electrophoretic mobility shift assay
ESCs - embryonic stem cells
ESI-MS - electrospray ionisation mass spectrometry
FL -full length
FRET - Fluorescent resonance energy transfer
g - relative centrifugal force

GFP - green fluorescent protein
GraFix - gradient fixation
GSH – reduced glutathione
GST - Glutathione S-transferase
HCL – hydrochloric acid
HDAC1 and HDAC2 – histone deacetylase 1 and 2
HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMM - hidden Markov model
HP1 - heterochromatin protein 1
HSA - helicase/SANT-associated domain
HSCs - hematopoietic stem cells
IAPs - intracisternal A-particles
ICF - Immunodeficiency, centromeric instability, facial anomaly syndrome
IEC - ion exchange chromatography
IMAC - immobilised metal affinity chromatography
IMAC FF - immobilised metal affinity chromatography fast flow
IMAC HP - immobilised metal affinity chromatography high performance
INO80 - INOsitol requiring 80 remodeller
IPTG - isopropyl β-D-1-thiogalactopyranoside
IR800 – infrared 800 nm dye
ISWI - Imitation SWItch remodeller
Kbp – kilo base pair
KD - knockdown
kDa - kiloDaltons
KO – knockout
LB - Lysogeny broth
LCRs - Low complexity regions
LIC - Ligation-independent cloning
LINEs - long interspersed nuclear elements
LSH - Lymphoid-specific helicase
MALDI-ToF - matrix-assisted laser desorption/ionization-time-of-flight
MAO - monoamine oxidase A
Mbp – mega base pair
MDa - megaDaltons
MEFs - murine embryonic fibroblasts
mESCs - the murine embryonic stem cells
MNase - micrococcal nuclease
MSA - Multiple sequence alignment
MW – molecular weight
MW CO – molecular weight cut off
NCP – nucleosomal core particle
Ni-NTA - Nickel-nitrilotriacetic acid
OD – optical density

ORF – open reading frame
PBS – phosphate buffered saline
PCR -polymerase chain reaction
PES - polyethersulfone
pI – isoelectric point
PKB - protein kinase B
PMSF - phenylmethylsulfonyl fluoride
pN – piconewton (force unit)
PTMs – post-translational modifications
RFP - red fluorescent protein
RFUs – relative fluorescent units
RMSD - Root-mean-square deviation of atomic positions
rpm – revolutions per minute
RSC - Remodelling the Structure of Chromatin remodeller
RT – room temperature
SAXS – small angle light scattering
SDS-PAGE – sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC-MALS - Size-exclusion chromatography coupled with multi-angle light scattering
SELEX - Systemic Evolution of Ligands of EXponential enrichment
SFM – serum-free medium
SHL2 – superhelical location 2
SINEs - short interspersed nuclear elements
SOC - Super Optimal Broth
ssDNA – single-stranded DNA
sulfo-NHS - N-hydroxysulfosuccinimide
SWI/SNF - SWItching defective/Sucrose-Non-Fermenting remodeller
TBE – Tris-boric acid-EDTA buffer
TCA - trichostatin A
TE – Tris-EDTA buffer
TEMED - Tetramethylethylenediamine
TFA - trifluoroacetic acid
TGEV - transmissible gastroenteritis virus
U/ml – units per millilitre
V - volt
v/v – volume per volume
w/v – weight per volume
WT – wild type
YFP – yellow fluorescent protein

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

1. INTRODUCTION

1.1 Chromatin organisation and modifications of DNA and chromatin

1.1.1 DNA and chromatin

DNA is a molecule that acts a universal carrier of genetic information across biological species. Structural characteristics of DNA conformation and organisation define the modes of gene expression. At the most basic level DNA exists in three forms: A, B and Z. The most common forms, A-DNA and B-DNA are both right-handed double helices, with B-DNA being the most common native cellular form and A-DNA predominating in dehydrating experimental conditions. Z-DNA is the least common form, having a left-handed orientation and a zigzag geometry that is formed by alternating purines and pyrimidines (Berg et al., 2002).

The helical topology of DNA forms major and minor grooves, which serve as binding interfaces for proteins and small molecules (Fratini et al., 1982; Heinemann et al., 1992). The major groove is the preferred site for transcription factor binding (Pabo and Sauer, 1984), however, certain proteins (e.g. the AT-hook motif-containing proteins) and antibiotics and a DNA fluorochrome Hoechst bind to the minor groove (Pjura et al., 1987).

DNA is organised differently in prokaryotes and eukaryotes. Since prokaryotes do not contain the nucleus, they solve the problem of storing their large DNA molecule (around 0.5-10 Mbp) by compressing it through supercoiling into the nucleoid area via the action of the proteins HU and topoisomerase I (Boubrik et al., 1991; Siedlecki et al., 1983).

Eukaryotic genomes have the advantage of being stored inside the nucleus, where DNA is packaged into nucleosomal subunits. Each nucleosome consists of DNA and the histone octamer - a protein heterocomplex consisting of pairs of four histone proteins H2A, H2B, H3 and H4. The histones form a disk-like structure termed the octamer, which accommodates roughly two gyres of DNA (the total of 146-147 bp) in around 1.7 turn (Luger et al., 1997a). The resulting DNA-protein complex stored within the nucleus is termed chromatin.

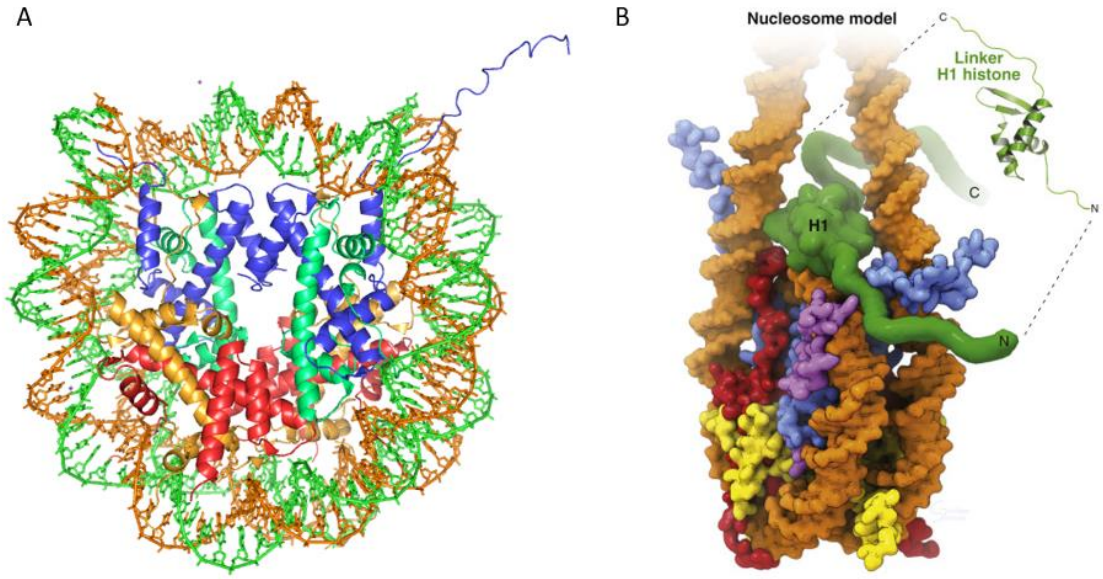


Figure 1.1 Mononucleosome architecture. A) A mononucleosome X-ray structure adapted from Luger et al. 1997 (PBD ID 1AOI). DNA is shown as sticks and ribbons. H2A – orange, H2B – red, H3 – blue, H4 – green. B) A schematic representation of a mononucleosome bound to H1 histone (from Cell Biology, 3rd edition 2017 (Johnson et al., 2017)).

Neighbouring nucleosomes are connected via linker DNA of around 40-50 bp that protrudes from either side of the nucleosome. Cryo-electron microscopy studies showed that linker DNA provide a characteristic zigzag pattern in chromatin compaction (Bednar et al., 1998). Compaction is assisted by the accessory H1 histone that binds roughly at the nucleosome dyad symmetry axis between two linker DNA fragments (***Fig. 1.1***) (Johnson et al., 2017).

1.1.2 Histones and nucleosomes

In vitro experiments helped to elucidate histone octamer formation (Luger et al., 1997a). At high ionic strength (~2 M NaCl) histones form an octamer, whereas at physiological ionic strength (~150 mM) they exist as stable aggregates of H2A-H2B dimers and H3-H4 tetramers in the absence of DNA or as the octamer in the presence of DNA (Luger et al., 1997a). The H3-H4 tetramer is juxtaposed with the dyad axis, whereas the H2A-H2B dimers are situated on the side of the octamer disk opposite the dyad axis (Luger et al., 1997a) (***Fig. 1.1***). The Stability of the octamer is determined by the interaction between the histones. Each histone has a characteristic folding of its three alpha-helices, which are connected by two loops (L1, L2). Each heteromeric L1L2 interaction site serves as a DNA contact location (Luger et al., 1997a).

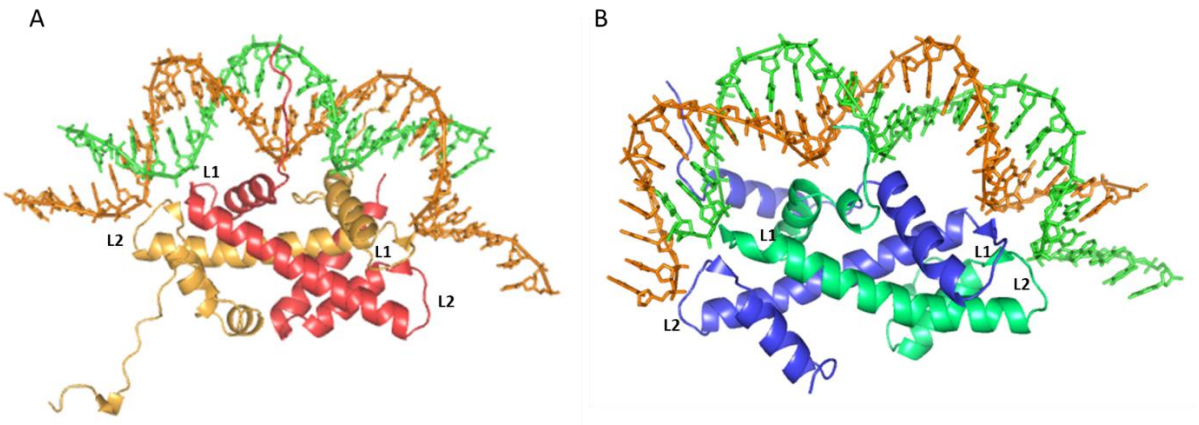


Figure 1.2 L1L2 loops of histones contacting DNA. A) H2A-H2B dimer (orange and red, respectively). B) H3-H4 dimer (blue and green, respectively).

Histones have highly positive net charges with pI around 9-10 allowing them to tightly interact with negatively charged DNA at physiological conditions. DNA-histone octamer affinity measurements depend on multiple factors. Apart from salt concentration, temperature, nucleosome particle concentration and histone variants are other determinants of DNA-octamer affinity (Ausio et al., 1984; Jin and Felsenfeld, 2007; Yager and van Holde, 1984). Energetics studies with 5S ribosomal RNA gene and alpha-satellite DNA sequence involving electrophoretic mobility shift assay (EMSA) and velocity sedimentation studies, demonstrated K_d at picomolar concentrations (Gottesfeld and Luger, 2001).

One of the common DNA sequences used in structural studies of the nucleosome is the 601 Widom sequence (P.T Lowary and Widom, 1998) that was identified by the SELEX (Systemic Evolution of Ligands of EXponential enrichment) technique. The relative free energy (ΔΔG) of DNA binding to the histone octamer compared to 5S RNA gene was -2.9, indicating that the 601 Widom sequence would have an affinity in the low picomolar range. The high stability of such recombinant nucleosomes making them a popular choice for structural studies.

The net charge of the nucleosome depends on the DNA fragment. It was estimated that the octamer, despite positively charged histone N-termini, would also carry a significant portion of negative charge (214 positive charges, 72 negative charges based on the positively and negatively charged amino acids in the histone octamer). This would not completely offset highly negative charge of DNA (310 negative charges for a 155 bp DNA fragment) (Bertin et al., 2007), making the nucleosome overall negatively charged with asymmetrical positively and negatively charged patches on the histone octamer.

One of the important functional regions of the nucleosome is the acidic patch consisting of eight acidic residues located in the histones H2A (E56, E61, E64, D90, E91, E92) and H2B (E102, E110) (Kalashnikova et al., 2013). This cluster interacts with residues 16-24 of the H4 tails of the neighbouring nucleosome (Luger et al., 1997a), therefore participating in chromatin compaction (Fan et al., 2004), as well as with several nucleosome-interacting proteins (Armache et al., 2011; Barbera, 2006; Makde et al., 2010; Roussel et al., 2008).

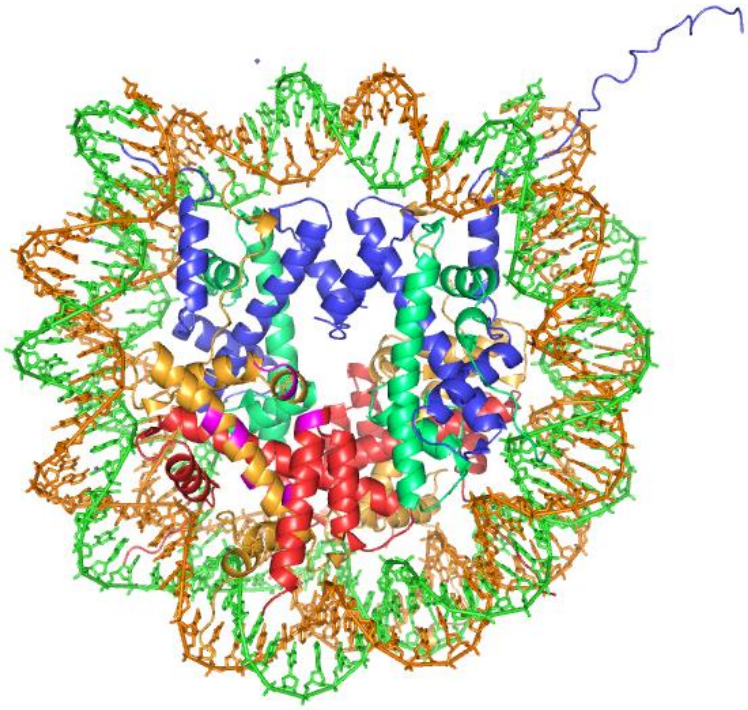


Figure 1.3 A mononucleosome X-ray structure adapted from Luger et al. 1997 (PBD ID 1AOI) showing the location of the acidic patch (H2A (E56, E61, E64, D90, E91, E92) and H2B (E102, E110)) (in magenta). DNA is shown as sticks and ribbons. H2A – orange, H2B – red, H3 – blue, H4 – green.

1.1.3 Histone tails and their modifications

Molecular dynamics simulations demonstrated that histone tails can affect the overall structural stability of the nucleosome, with truncated tails of H2A and H3 introducing structural changes to the histone core (Biswas et al., 2011).

Histone tails modifications is an important mechanism of epigenetic regulation. Histone N-terminal tails carry positive charges due to the abundance of lysines and arginines. This property also makes them a convenient site for numerous covalent post-translational modifications (PTMs). Acetylation is one of the most abundant PTMs, examples include H3K9, H3K14, H3K18, H4K5, H4K8 and H4K12 (Kouzarides, 2007). Those modifications act by

neutralising positive charges and are generally associated with transcriptionally active chromatin, enhancers and gene promoters (Wang et al., 2008). Along with acetylation, phosphorylation of H3S10 is also involved in charge neutralisation of histone tails and is correlated with chromatin condensation in mitosis and meiosis (Wei et al., 1998).

Positively charged residues are also the sites for mono-, di- or tri-methylation. Bulky methyl groups may add local structural variations or act as a binding site for chromatin-interacting proteins. For instance, H3K9 methylation was shown to be the binding site for heterochromatin protein 1 (HP1) (Bannister et al., 2001). Ubiquitylation, due to addition of a large group, may affect nucleosomal conformation and introduce disruptions into intro- and internucleosomal interactions (Kouzarides, 2007).

1.1.4 DNA methylation

DNA is another chromatin site that undergoes methylation, wherein a methyl group is transferred from S-adenyl methionine to the fifth carbon of a cytosine forming 5 methyl-cytosine (5mC) (Moore et al., 2013) predominantly on CpG dinucleotides (Bird, 1980). Nucleosomal DNA is protected from methylation, therefore, linker DNA is the main target or methyl group transfer (Felle et al., 2011). There are two types of DNA methylation: *de novo* and maintenance. *De novo* methylation is catalysed by two methyltranferases DNMT3A and DNMT3B, which modify DNA during embryonic development. These methylation patterns are stable and are crucial for gene imprinting, X chromosome inactivation and silencing of repetitive elements, with disruption of this mechanism implicated in numerous diseases (Gopalakrishnan et al., 2008; Jin et al., 2011).

1.2 Chromatin remodelling

1.2.1 ATP-dependent chromatin remodelling

Enzymes utilising the energy from ATP hydrolysis are one of the most prominent actors in the chromatin metabolism. They confer a variety of outcomes in chromatin remodelling such as DNA unwrapping, nucleosome sliding, histone variant exchange and nucleosome eviction (***Fig. 1.4***) facilitating various nuclear processes.

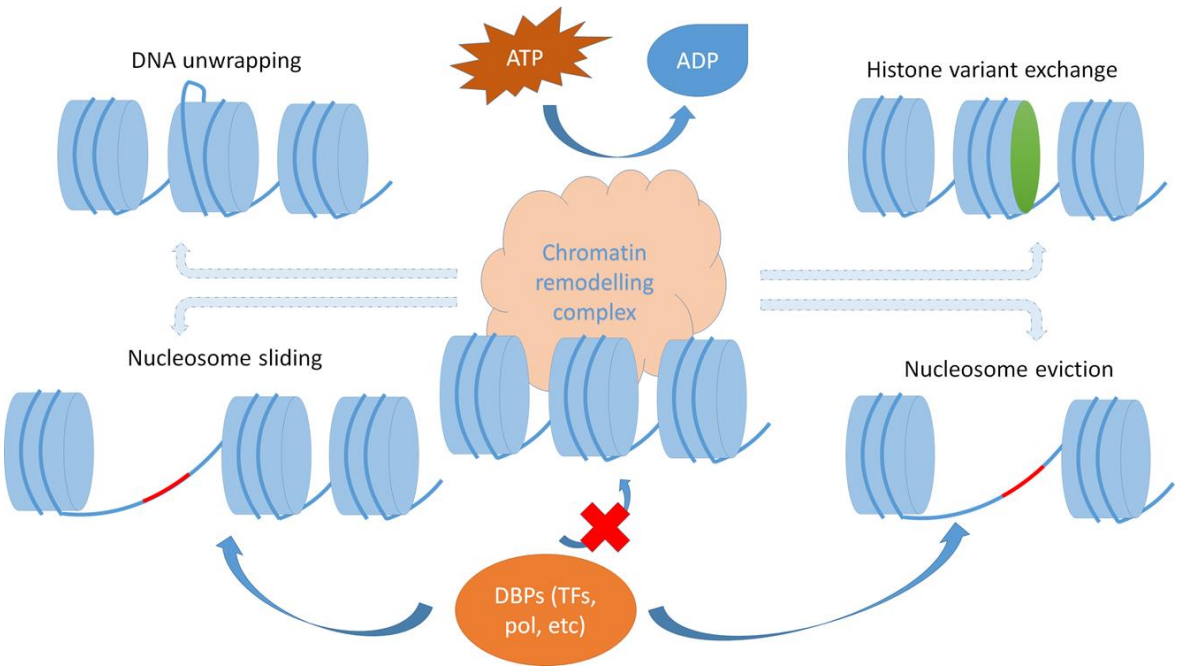
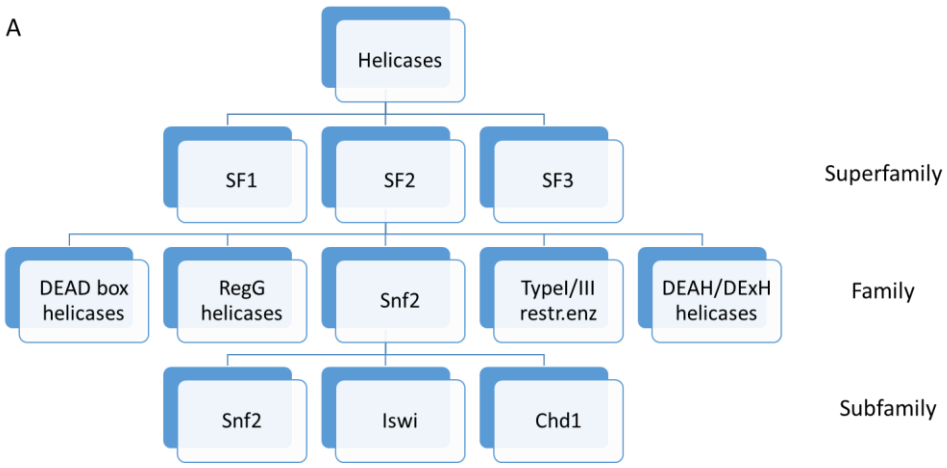


Figure 1.4 Schematic representation of possible functions of ATP energy-utilising chromatin remodelling complex. They include DNA unwrapping, nucleosome slinding, histone variant exchange and nucleosome eviction. All these actions affect accesibility of DNA to various DNA-binding proteins (DBPs), such as transcription factors (TFs), polymerases (pol) etc.

The chromatin remodellers share a common RecA-like domain. The RecA protein is found in *E. coli* and is involved in DNA recombination (Clark and Margulies, 1965). In the centre of the RecA-like domain is a β-sheet with α-helices on each side (Ye et al., 2004). Lobe 1 contains a Walker A motif (or P-loop) containing a highly conserved motif A/GxxxxGKT/S, where x is any amino acid (Ye et al., 2004). The P-loop is responsible for the γ-phosphate coordination during ATP hydrolysis. The Walker B motif is located downstream from Walker A motif and contains the sequence hhhhDE, where h is a hydrophobic residue (Hanson and Whiteheart, 2005). The ATP hydrolysis mechanism is as follows: the γ-phosphate is coordinated by arginines or lysines and the attacking water molecule is polarised by aspartates or glutamates (Ye et al., 2004). The transition state stabilisation also requires a Mg^{2+} (Ye et al., 2004). An ATP molecule is primarily bound to the lobe 1, however, hydrolysis requires participation of the lobe 2, and the interaction between the two lobes requires critically but not exclusively a presence of “arginine fingers” (Ye et al., 2004).

Most commonly, chromatin-remodelling enzymes function as a part of large multisubunit complexes, wherein accessory subunits specify the targeting and functionality of the enzyme.

The most commonly accepted classification separates the chromatin remodelling enzymes into 4 groups according to the similarity of the accessory domains flanking the ATPase core and the remodelling outcome. These are: SWI/SNF (SWItching defective/Sucrose-Non-Fermenting), ISWI (Imitation SWItch), Chromodomain-Helicase DNA binding (CHD) and INOsitol requiring 80 (INO80). Roughly, their functionalities correspond to the following areas: SWI/SNF primarily have a role in chromatin access, as the ATPase faces moderate inhibition from the accessory domains and is able to carry out octamer eviction and nucleosome unwinding; more controlled ATPases of ISWI and CHD provide the basis of forming pre-nucleosomal complexes and regularly spaced nucleosomal arrays; and INO80 is capable of nucleosome editing, wherein histone variants can be evicted and exchanged (Clapier et al., 2017).



B

GROUPINGS	SUBFAMILIES
Snf2-like	ALC1 Iswi
	CHD7 Chd1 Mi-2
	Lsh Snf2
Swr1-like	EP400
	Swr1
	Ino80
SSO1653-like	ERCC6 SSO1653
	Mot1
Rad54-like	Rad54
	ATRX Arip4
	JBP2
	DRD1
Rad5/16-like	Lodestar
	Rad5/16 Ris1
	SHPRH
Distant	SMARCAL1

Figure 1.5 Helicases classification. A) Helicases at superfamily, family and subfamily levels. B) Rooted tree from hidden Markov model (HMM) on the basis on full length alignments of the helicase regions. Adapted from Flaus et al. 2006.

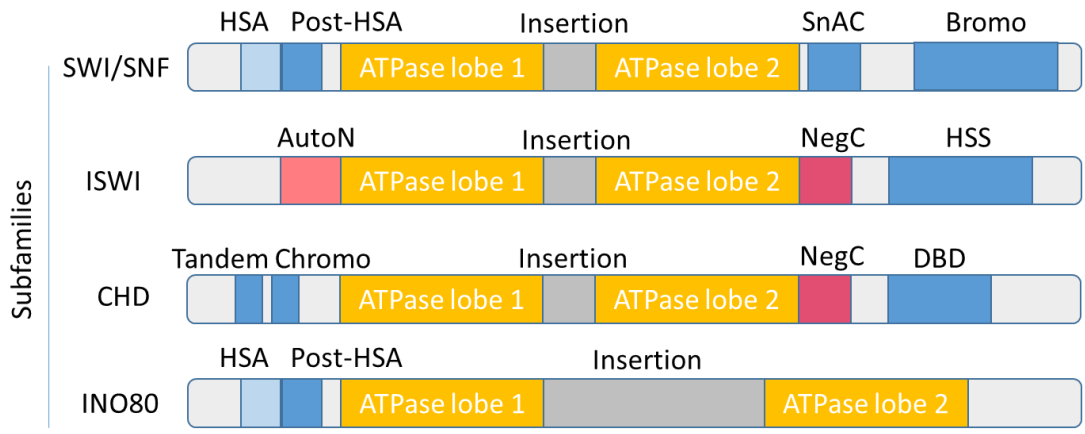


Figure 1.6 Schematic representation of the domains in each subfamily of chromatin remodellers. HSA – helicase-SANT–associated domain; SnAC – Snf2-ATP coupling domain; Bromo - bromodomain; AutoN and NegC – ISWI regulatiory domains ; HSS - HAND–SAND–SLIDE domain; Chromo - chromodomain ; DBD – DNA-binding domain.

1.2.2 Modes of DNA translocation

The RecA-like lobes of the ATPase subunit are at the core of DNA translocation mechanism by chromatin remodellers. The two lobes act like “mittens” sequentially grabbing and releasing the DNA double helix via its backbone atoms, appearing as an inchworm-like processive movement along the DNA double helix. When this action is fixed from the remodellers’ perspective, this inchworm-like movement translates into pumping the DNA helix uni-directionally towards the nucleosome dyad. This movement results in DNA displacement from the-histone octamer and DNA translocation along the histone octamer surface, thus relaxing chromatin or exposing the DNA sites previously blocked by the contact with the nucleosome.

A series of experiments with optical tweezers provided the first evidence of DNA translocation and DNA loop formation by the SWI/SNF subfamily members. Optical tweezers are a well-established biophysical technique, where a DNA molecule is fixed between two polystyrene beads, one of which is attached to a force-measuring optical trap (dynamic) and the other is held by a pipette tip by suction (stationary). The force of DNA extension is translated into the movement of the optical trap using a laser and recorded using a charge-coupled camera. The reduction in distance between the two ends of DNA may indicate formation of a loop. The optical tweezers experiments showed that a DNA loop could only form when a nucleosome, SWI/SNF complex and ATP were present (Zhang et al., 2006), indicating that the loop forms as an extension from the nucleosomal DNA, and effective remodelling requires nucleosomal context. Formation of the loop introduces a DNA torsion with accumulated energy, which subsequently translates into the loop relaxation, with the subsequent DNA rebound and translocation around the histone octamer. This remodelling behaviour was also found to be processive, as the complex could disengage from the nucleosome after a series of translocations (Zhang et al., 2006).

The efficiency of remodelling is ensured by the nucleosome context and the remodeller’s accessory subunits, as demonstrated by comparison of a full RSC complex and its minimal version containing the ATPase core (Sirinakis et al., 2011). Expectedly, the full complex appeared to be more potent and less mobile than the minimal complex, however, the minimal complex was still capable of DNA translocation and even binding of free DNA. The bend already present in a nucleosomal DNA was shown to be an energy-saving factor or a

store of energy that contributes to remodelling once released, as the minimal motor had to dispense its energy to bend the free DNA (Sirinakis et al., 2011). By anchoring to the nucleosome, the complex can exert the remodelling activity through a force sufficient to disrupt DNA-histone contacts. The complex was able to generate up to 30 pN of force, which exceeds the typical forces required to disrupt the DNA (~23 pN) (Sirinakis et al., 2011). This powerful movement was, however, susceptible to stalls when gaps were introduced to the DNA sequence (Harada et al., 2016).

Fluorescent resonance energy transfer (FRET) is a powerful technique to study the behaviour of chromatin remodellers. It allows detection of the distance between two fluorophores conjugated at different locations on a nucleosome, e.g. green fluorescent protein (GFP) on a histone and red fluorescent protein (RFP) on the DNA terminal. A laser then excites GFP, and if RFP is located at a close enough proximity at this point (1-10 nm) and the spectra of the fluorophores overlap, the energy from GFP will be transferred to RFP. The transferred energy is measured using various approaches (sensitised emission, acceptor photobleaching, spectral imaging, anisotropy imaging), allowing nucleosome dynamics to be dissected in great detail, such as 1-2 bp step size per 1 hydrolysed ATP molecule (Saha et al., 2005), which confirmed the earlier findings from optical tweezer experiments (Sirinakis et al., 2011).

The triple-helix strand displacement assay of RSC interaction with the octamer confirmed the processive movement of RSC along the DNA strand. This assay relies on the Hoogsteen base pairing interaction between a double-stranded DNA and a third helix of homopurine-homopidimidine repeats (Kopel, 1996). This interaction is weak enough for the third helix to be displaced by the processive movement of the remodeller. Furthermore, RSC was shown to be capable of interaction with a single-stranded DNA, indicating that this processive movement requires the interaction with only one strand (Saha, 2002). The outcome of the translocation was postulated to be a DNA segment twist that is subsequently propagated as a wave, owing to the accumulation of energy from an undertwisted DNA (Saha, 2002), which is partially dissipated by the introduction of a gap on one of the DNA strands. This model is termed “wave-ratchet-wave” translocation, wherein the remodeller acts as a ratchet, engaging DNA and tracking its translocation unidirectionally from the 3’- to the 5’-end in small increments. This model is similar to the loop diffusion model, however, the energy required for DNA translocation, should come directly from the remodeller’s ATPase activity rather than from the energy accumulated in the loop formation.

This unidirectional movement can occur at either side of the nucleosome dyad axis, allowing DNA translocation on either side, depending of the remodeller binding orientation. The exact location of the engagement between the remodeller and the nucleosome was elucidated from the studies of RSC, where superhelical location 2 (SHL2), found two DNA helical turns away from the nucleosomal dyad axis, was protected from DNAase I digestion, indicating the remodeller binding location (Saha et al., 2005). This location was found to be crucial for other remodellers, such as ISWI (Hwang et al., 2014; Ranjan et al., 2015; Schwanbeck et al., 2004). Chromatin remodellers take advantage of the inherent flexibility of DNA at the SHL2, which would also explain the propensity across the remodellers to bind this region (Edayathumangalam et al., 2005; Luger et al., 1997a).

Several studies point to the loop diffusion model to be more likely. SWI/SNF remodelleres were shown to form ~50 bp loops and ISWI – 9-11 bp loops (Schwanbeck et al., 2004, Strohner et al., 2005; Zhang et al., 2006; Zofall et al., 2006), which demonstrates that SWI/SNF are more disruptive than ISWI. ISW2 was in turn acting through torsional strain at SHL2 (Zofall 2006), indicating that loop formation is more likely to be formed in case of disruptive complexes such as SWI/SNF but not ISWI or Chd1. BRG1 remodeller in the context of a trinucleosome was able to generate DNA loops on the central nucleosome, providing enough force to overcome the neighbouring nucleosome barrier (Fan et al., 2003). A large loop (~50-100 bp) was hypothesised to form on the nucleosomal surface, however, the introduction on small gaps on the tracking strand stalled RSC progression, which is also consistent with a small ~1 bp bulge formation (Lorch et al., 1999; Saha et al., 2005).

Molecular dynamics studies also highlight different possibilities of nucleosome remodelling: rotation-uncoupled mode associated with loop formation and rotation-coupled mode associated with corkscrew motion and 1 bp step size (Niina et al., 2017), however, chromatin remodelling through the loop formation on the surface of the nucleosome still remains a nebulous concept.

Recently DNA torsion model, where DNA is translocated following formation of a small bulge has gained further ground, based on several solved structures, molecular dynamics modelling and biochemical studies of Snf2 family remodellers.

Molecular dynamics simulation of Snf2 binding at SHL2 and formation of twist defects, which subsequently propagated along the nucleosome (Brandani et al., 2018). ATP binding is followed by lobes closure, lobe 1 detachment and movement towards lobe 2 by 1 bp

(Brandani et al., 2018). ATP hydrolysis is concurrent with the weakening of lobe 2 and DNA contacts and movement of lobe 2 towards the dyad by 1 bp. When ADP is released and cycle continues, and, hence, the processive movement of the remodeller occurs.

Two states have been observed: a decreased twist in apo and ADP-bound state and an increased twist with ATP analogues at SHL2 position (Li et al., 2019; Winger et al., 2018). This finding is consistent with recently found distortions in sugar-phosphate backbone and in base stacking (Liu et al., 2017a). Diffusion of this twist is sufficient to disrupt the DNA-histone interactions, presumably without major octamer perturbations, at least in the case of ISW1 (Yan et al., 2019a). Other findings indicate that twist defect is associated with octamer distortions at SHL2, such as histone dimers maintain their contact points with shifted DNA or form new contacts after DNA register shift (Bilokapic et al., 2018). Taken together, there is no one clear picture of nucleosome remodelling, and several modes of DNA translocation can be utilised by different chromatin remodellers to achieve specificity.

1.2.3 SWI/SNF subfamily

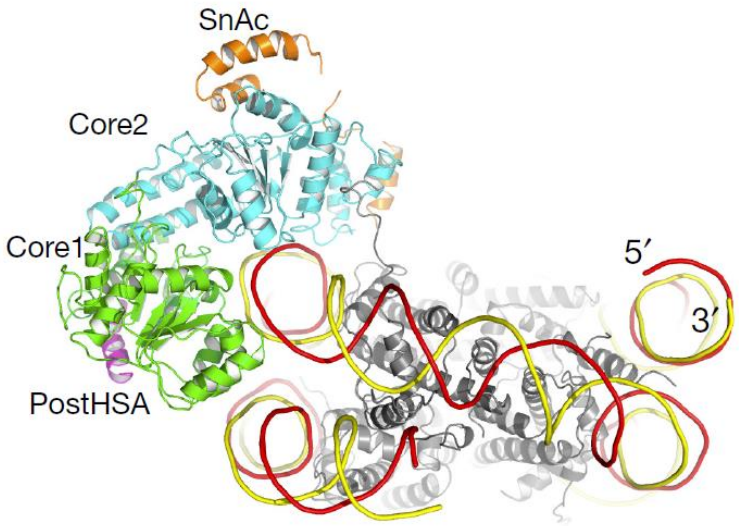


Figure 1.7 Cryo-EM structure of S. cerevisiae Snf2 (ScSnf2) in complex with NCP (PDB ID 5X0Y). Structure is taken from Liu et al. 2017 (Liu et al., 2017b). Histone octamer is shown in dark grey.

The early findings regarding the SWI/SNF subfamily functions in octamer eviction came from the studies of the yeast complex RSC, whose catalytic subunit Sth1 is homologous to

Swi2/Snf2 of the SWI/SNF complex. These are large complexes of approximately 1 MDa with 11-15 subunits, mediating remodelling activity of the ATPase subunit.

Early experiments demonstrated that the RSC complex is capable of completely evicting the histone octamer and forming intermediate double nucleosomes (Lorch et al., 1999). Octamer eviction can play a critical role in gene expression regulation, if the nucleosomal DNA contains various regulatory regions, such as transcription activator binding sites. This was confirmed by using a chimeric transcription factor Gal4-VP16, which enhanced octamer transfer from the nucleosome (Gutiérrez et al., 2007; Owen-Hughes and Workman, 1996). Activation domains present on a transcription factor guide the localisation and stimulation of the remodeller.

SWI/SNF complexes are capable not only of evicting octamers but also of sliding nucleosomes. To be precisely carried out, these distinct functions require accessory subunits and regulatory domains in the remodelling complex. Sth1, the catalytic subunit of the RSC complex, has two key domains: the N-terminal post-HSA domain and protrusion 1, located in the middle of the ATPase domain (Clapier et al., 2016). The Post-HSA domain is located immediately downstream the HSA domain that is known to interact with actin-related proteins (ARPs). ARPs play various roles in cellular processes and are a part of the RSC complex. Removal of ARPs improved nucleosome sliding but impaired ejection (Clapier et al., 2016). Removal of the ARP-binding and ARP-bridging protein Rtt102, another subunit in RSC, reduced sliding, but did not affect ejection (Clapier et al., 2016).

In Sth1, mutations introduced into the post-HSA domain that interacts with ARPs, produced a hyperactive remodeller with excessive ATPase activity, irregularly spaced nucleosomes and a lethal phenotype. By contrast, mutations in the protrusion 1 yielded improved coupling – the amount of translocation relative to the number of hydrolysed ATP molecules – and enhanced translocation and ejection (Clapier et al., 2016). Therefore, post-HSA and protrusion 1 play roles of a “break” and a “clutch”, respectively, specifying the function of RSC.

Further insights into the functions of SWI/SNF remodellers come from the crystal structure of Swi2/Snf2 catalytic subunit in a resting state from the yeast *Myceliophthora thermophila* (Xia et al., 2016). The protein lacks a DNA binding domain (DBD) and, like RSC, Swi2/Snf2 contains an N-terminal post-HSA domain, as well as a SANT-associated domain. The C-terminus harbours a SnAc domain, which may interact with histone tails. Like RSC, Swi2/Snf2

interacts with ARPs through its N-terminus. The essential motifs for ATP hydrolysis are the motifs I (P loop) and VI (arginine fingers, which are disordered when inactive) located in the RecA-like lobes. The motifs Ib, II, V and core 2i were found to be essential for DNA binding (Xia et al., 2016).

The two ATPase RecA-like lobes of *M. thermophila* SWI/SNF stack together in an inactive state, with their ATP hydrolysis-essential motifs distant from each other and the DNA binding domains exposed, rendering the protein in a poised state (Xia et al., 2016). By similarity with an RNA-binding helicase Vasa in a complex with RNA, the authors predicted that, upon substrate binding and activation, the lobe 2 would perpendicularly rotate relative to the lobe 1 (Sengoku et al., 2006; Xia et al., 2016). Like in Sth1, HSA domain downregulates the remodelling activity and its removal yields a hyperactive enzyme (Xia et al., 2016).

In complex with the nucleosome, Swi2/Snf2's motif VI (arginine fingers) becomes ordered and is brought close to the ATP binding element of motif I (P loop) (Liu et al., 2017a). The C-terminal Brace helices, upon activation, protrude from the lobe 2 to contact lobe 1 through interaction with protrusion 1, aiding communication between the two lobes (Liu et al., 2017a). This brace domain is similar to the NegC domain in ISWI remodeller. However, unlike in ISWI, this domain has a positive regulatory function.

The H4 tail is another point of contact in Swi2/Snf2-nucleosome interaction. It contacts the acidic surface on the lobe 2, which is similar to the acidic pocket in ISWI (see below) (Liu et al., 2017a). Again, unlike for ISWI, where H4 plays a critical regulatory role (see below), Swi2/Snf2's remodelling activity is diminished but not blocked by interaction with the H4 tail (Liu et al., 2017a).

Another important residue is W1185, which is also present in CHD and ISWI remodellers. Its bulky side chain is located within the ATP-hydrolysing motif VI and is inserted into the DNA minor groove (Liu et al., 2017a) (***Fig. 1.7***)). Mutation W1185A does not block ATP hydrolysis but diminishes it and abrogates chromatin remodelling activity (Liu et al., 2017a).

1.2.4 ISWI subfamily

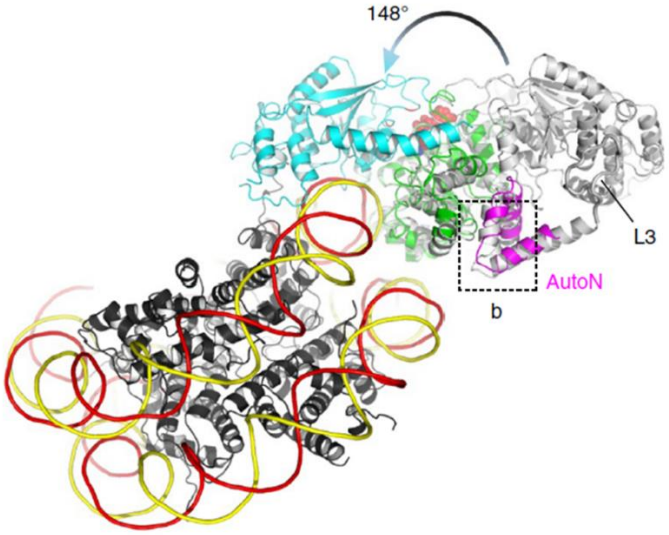


Figure 1.8 Cryo-EM structure of S. cerevisiae ISW1 in active conformation (colour coded) bound to NCP compared with inactive conformation (grey) (PDB ID 6JYL). From Yan et al. 2019 *(Yan et al., 2019a)****. ADP-Bef$_x$ is shown in red.***

Compared with the SWI/SNF subfamily, ISWI chromatin remodellrs are generally smaller (around 0.5 MDa with 2-5 subunits) and are less disruptive to the nucleosome, providing enough force for sliding but not for histone eviction. This conclusion was illustrated by *Drosophila melanogaster* ISWI-containing NURF complex, wherein ISWI was estimated to translocate DNA forming 10 bp bulges (Schwanbeck et al., 2004). By contrast SWI/SNF was capable of producing as large as 50 bp increments (Zofall et al., 2006). ISWI complexes, such as CHRAC, ACF and RSF, are mostly involved in nucleosome spacing and array formation, except for the NURF complex, which degrades spacing regularity (Hamiche et al., 1999).

Unlike SWI/SNF complexes, ISWI subfamily chromatin remodellrs require additional histone and nucleosomal epitopes to exert their action. Upon binding of the catalytic subunit of ISWI, distortions in the linker DNA adjacent to the nucleosome occur, and remodeller-nucleosome complexes form more readily in the presence of a linker DNA (Schwanbeck et al., 2004; Whitehouse et al., 2003).

Interestingly, directionality of the nucleosome repositioning by ISWI was found to be strictly centre-to-end, i.e. the octamer is located at the centre of DNA molecule with a DNA linker on each side, transforming that orientation to asymmetrical positioning with no linker on one side and a long linker on the other. However, when ISWI remodeller is a part of CHRAC

complex, the directionality changes to end-to-centre (Längst et al., 1999; Whitehouse et al., 2003). These data yet again stress the influence of the subunits and accessory domains on chromatin remodelling by the core ATPase. In ISW2 complex, a non-catalytic subunit Itc1 determined the direction and extent of nucleosomal sliding by binding to a distal part of the DNA linker, which may provide a steric block between the nucleosomes ensuring regular spacing (Kagalwala et al., 2004).

To dissect the role of linker DNA in ISWI remodelling, hydroxyl radical footprinting (HRF) was utilised. HRF is a sensitive technique with a principle similar to DNAse I digestion. HRF is based on cleavage by hydroxyl radicals of a hydrogen atom from C4 of the deoxyribose in the minor groove, which only happens when a DNA molecule is exposed. Therefore, the pattern of cleavage allows detection of regions protected by a bound protein. This method demonstrated that the NURF complex protected DNA on the linker region adjacent to the nucleosome, as well as within the nucleosomal core, at SHL2 site (Schwanbeck et al., 2004).

The importance of a linker DNA was confirmed by the higher amount of ATP hydrolysed by ISWI in the presence of a linker compared to a nucleosome core particle only (Whitehouse et al., 2003). Further studies estimated the minimal required linker length for ISWI remodelling to be 20 bp, whilst > 67 bp was an optima length under competitive or limiting conditions (Kagalwala et al., 2004). A triple helix displacement assay hinted at the processivity deterioration of ISWI with a linker length between 0 and 60 bp, where displacement was more efficient on a shorter linker compared to a longer one (Whitehouse et al., 2003).

This dependence on the linker length points to the importance of histone tails for ISWI function, as further movement from histones affects ISWI processivity and, by extension, remodelling activity. A nucleosome with tail-less histones was shown to be more thermally labile than the wild type, indicating lost electrostatic contacts between the histone tails and DNA (Hamiche et al., 2001). In the CHRAC complex, removal of all histone tails reduced affinity between ISWI and the nucleosome but did not disrupt the remodelling activity, whereas removal of H4 tails interrupted remodelling (Clapier et al., 2001). The location of the H4 histone tails appears to be crucial for remodelling, as it is located proximal to SHL2 – the site crucial for remodeller-nucleosome interaction, as well as for sensing the DNA linker length (Hwang et al., 2014). Isolated H4 fused to a GST-tag could neither interact with ISWI in a pull-down assay, nor could it activate ATPase activity of the NURF complex,

demonstrating that a full nucleosome is required as a remodelling substrate (Clapier et al., 2001; Hamiche et al., 2001).

The H4 histone tail harbours an important motif for chromatin remodelling - the basic patch which has the amino acid sequence KRHR at the residues 16-19 (Clapier and Cairns, 2012; Hamiche et al., 2001). The ATPase subunit of ISWI is known to have inhibitory domains AutoN and NegC on its N- and C-terminus, respectively. AutoN interaction with the ATPase domain negatively regulates ATP hydrolysis, whereas NegC does not block ATP hydrolysis but uncouples it from DNA translocation (Clapier and Cairns, 2012). The HSS (HAND-SANT-SLIDE) domain on the C-terminus of the protein binds to the DNA linker (Dang and Bartholomew, 2007). Mutations in or removal of AutoN and NegC promotes nucleosome sliding even without the H4 basic patch, DNA linker or HSS domain (Clapier and Cairns, 2012), demonstrating that the ATPase domain has an intrinsic translocase activity. Therefore, ISWI should be controlled by regulatory domains. In ISWI, AutoN keeps the ATPase in a resting state, and the H4 tail competes with it for the ATPase binding site. This regulation of ISWI by AutoN is achieved due to its motif bearing high similarity to the H4 tail basic patch (Clapier and Cairns, 2012). The DNA linker provides a similar activating mechanism by binding to the HSS domain and triggering conformational changes in the NegC domain, antagonising its inhibitory activity (Clapier and Cairns, 2012).

Further insights into ISWI function came from the high-resolution crystal structure of its catalytic core (AutoN-ATPase-NegC-HSS) in a complex with H4 peptide (Yan et al., 2016). AutoN was found to bind to the acidic surface of RecA-like lobe 2, whereas H4 competes with it for one of the AutoN binding sites. NegC also binds to lobe 2, resulting in allosteric regulation response to DNA linker. C-terminal HSS domain binds the DNA linker, which should be of a sufficient length (25 bp with affinity increasing with increasing DNA length to 146 bp). A shorter linker results in weaker binding to HSS domain, which, in turn, cannot disrupt NegC interaction with lobe 2 (Yan et al., 2016). The integrative approach of cross-linking coupled to mass spectroscopy (XL-MS) and small-angle light scattering (SAXS) of the full length ISWI showed that the ATPase domain exists in a resting open conformation with motifs I and II on lobe 1 and motifs VI on lobe 2 not aligned (Harrer et al., 2018). Nevertheless, the overall protein structure is relatively compact, with the C-terminal SLIDE domain forming extensive cross-links with the N-terminal and ATPse domains,indicating their close proximity) and SLIDE domain (Harrer et al., 2018). Upon activation, the protein adopts a more compact

conformation, as indicated by a reduced radius of gyration as measured by SAXS (Harrer et al., 2018).

1.2.5 CHD subfamily

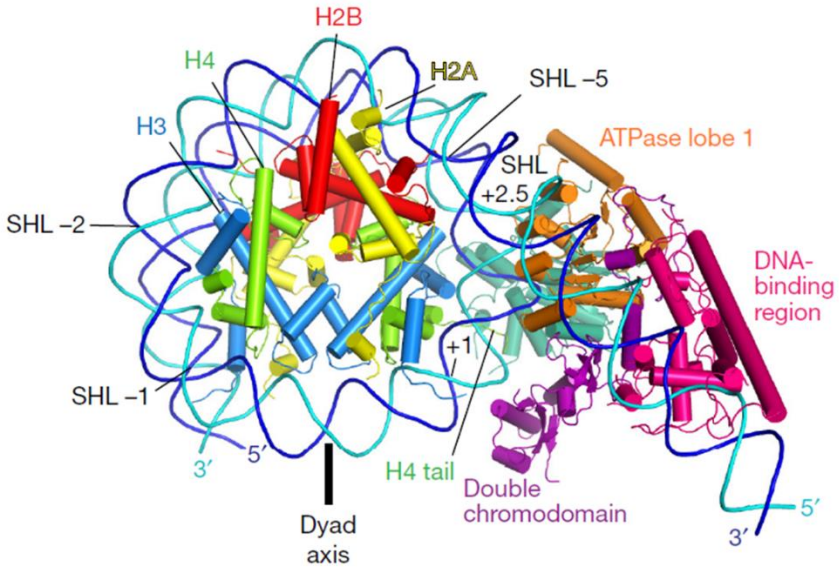


Figure 1.9 Cryo-EM structure of S. cerevisiae Chd1 bound to NCP (PDB ID 5O9G). From Farnung et al. 2017 (Farnung et al., 2017).

The CHD subfamily is closely related to the ISWI subfamily in its mode of action, however, the domains of the ATPase subunit differ. Instead of an N-terminal AutoN domain that negatively regulates ATPase activity by engaging the ATPase domain, CHD proteins have two tandem chromodomains connected by a conserved acidic helix forming a wedge-like structure. This packs against the DNA binding surface of the ATPase core, thus inhibiting its activity by providing a "gating" mechanism. Disruption of the chromodomain-ATPase interface results in full activation of the ATPase domain (Hauk et al., 2010). Two Arg residues, or "arginine fingers", conserved across chromatin remodellers, specifically R579 and R582 in *S. cerevisiae* Chd1, are present in the linker connecting the ATPase lobes, contacting the phosphate tail of ATP when the cleft is in the active, closed conformation (Hauk et al., 2010).

The C-terminus of Chd1 has a DNA binding domain (DBD) that senses the DNA linker length, thus producing regularly spaced nucleosomal arrays. This mode of action was proposed to be guided by the kinetic "release" model: the remodeller initially binds the nucleosome at the position of the highest affinity, gradually shifts it to the position of lower affinity, which favours its disengagement (Rippe et al., 2007). In the inactive state of the CHD1 the N-

terminus interacts with the DBD, as shown by BS3 cross-linking, and a short acidic patch in the N-terminus was shown to mediate this interaction (Sundaramoorthy et al., 2017).

Interdomain communication ensures measured activity of the remodeller. The Chromo-ATPase module provides the platform for DBD binding, as demonstrated by the abrogation of DBD binding upon chromo-ATPase deletion (Nodelman et al., 2017). DBD was shown to bind right at the edge of the nucleosome, interacting with the major groove (Nodelman et al., 2017). In turn, the deletion of the DBD disrupted the catalytic activity of the ATPase domain (Hauk et al., 2010; McKnight et al., 2011). The activity was restored by the addition of a heterologous DBD (McKnight et al., 2011; Patel et al., 2013). Interestingly, the DBD does not seem to be required for catalytic activity *per se*, with the DBD-truncated CHD1 still being able to shift the nucleosome (Patel et al., 2013), indicating that the ATPase has an intrinsic sliding activity. However, without a proper control from DBD, CHD1 could only produce disordered nucleosomal patterns, more similar to SWI/SNF remodellers (Patel et al., 2013), which lack a DBD.

A cryo-EM approach revealed more details about the CHD1-nucleosome complex. It was shown that the ATPase lobes remain in a partially closed state in the absence of a nucleotide, but with addition of a non-hydrolysable ATP analogue ADP-BeF_3 they close completely (Farnung et al., 2017), **Fig. 1.9**). Initial engagement with the nucleosome occurs in the partially closed, pre-translocation state. Complete closure of the lobes occurs upon ATP binding. The lobe 2 of the ATPase domain interacts with the H4 tail (Farnung et al., 2017) – similar interactions with histone tails are found in other remodeller families (Clapier et al., 2001; Ferreira et al., 2007; Hauk et al., 2010; Hwang et al., 2014; Liu et al., 2017a; Yan et al., 2016). The cryo-EM structure of CHD1? showed that the key for DNA recognition by the remodeller is the bend configuration of the DNA wrapped around the histone octamer, as a straight fragment of free DNA would result in a steric clash with the chromodomains (Farnung et al., 2017).

1.2.6 INO80 subfamily

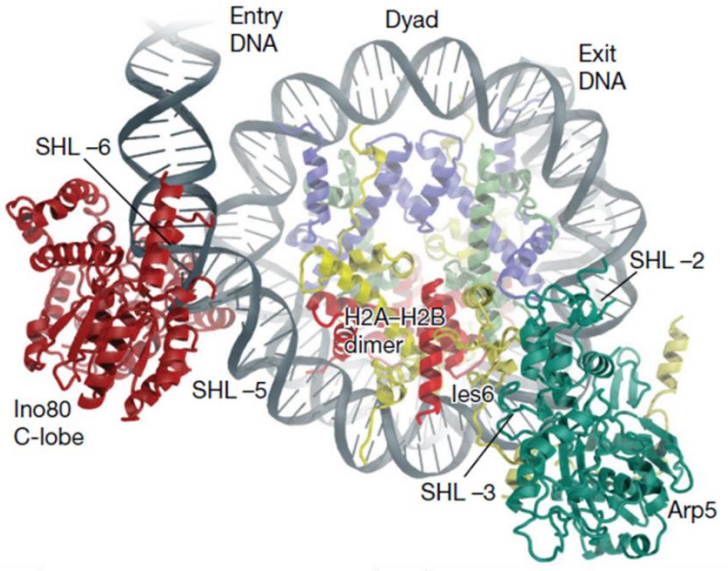


Figure 1.10 Cryo-EM structure of C. thermophilum INO80 bound to NCP (PDB ID 6FML). From Eustermann et al. 2018 (Eustermann et al., 2018).

The proteins from INO80 subfamily perform a function of replication-independent histone variant exchange (Mizuguchi, 2004). Histone variants, such as H2A.Z and H3.3, are expressed throughout the cell cycle and they can be incorporated into chromatin independently of DNA replication. H2A.Z is a highly conserved histone variant of an evolutionary lineage distinct from that of H2A. In yeast genomes, H2A.Z acts as a barrier between transcriptionally active domains and propagating heterochromatin (Meneghini et al., 2003). Even though H2A.Z-containing nucleosomes are similar to the canonical ones, certain structural differences alter their biophysical properties making them a less stable nucleosomal species (Rudnizky et al., 2016). This explains the preference for H2A.Z at promoters, which, in their active form are characterised by the presence of acetylated H2A.Z. In the absence of INO80 proteins, H2A.Z-containing nucleosomes are mislocalised and their presence in promoters does not correspond to transcriptional changes (Papamichos-Chronakis et al., 2011).

Heterogeneous nucleosomes (containing a mixture of H2A dimers: H2A-H2A, H2A.Z-H2A and H2A.Z-H2A.Z) have been found at gene promoters (Luk et al., 2010). The histone replacement reaction by an INO80 subfamily remodeller Swr1 is stepwise and unidirectional: H2A-H2A-containing nucleosomes, but not H2A.Z-H2A.Z, could stimulate Swr1 ATPase, with the highest ATPase activity detected when both the H2A-H2A nucleosome and the H2A.Z-H2B replacement dimer were bound (Luk et al., 2010). Contrary to that mode of action, Ino80

ATPase can incorporate H2A into H2A.Z nucleosome in a concentration-dependent manner, but cannot perform a H2A-to-H2A exchange (Papamichos-Chronakis et al., 2011), highlighting the importance of an H2A.Z-specific epitope. These observations led to a proposed model, wherein Swr1 deposits H2A.Z at a broad range of sites across the genome, with Ino80 performing the reverse reaction, only leaving H2A.Z at metabolically important sites, such as promoters (Papamichos-Chronakis et al., 2011).

The mechanism of H2A.Z exchange was elucidated by comparison of H2A and H2A.Z structures. Whereas the C-terminal domain was found to play an important role in histone-chaperone binding (Hong et al., 2014), the α2 helix and loop 2 were found on the H2A surface and in contact with Swr1 (Ranjan et al., 2015). Similarly to other remodelling families, INO80 proteins interact with the SHL2 site on the nucleosome, and this interaction can be blocked by introduction of gaps at this site (Ranjan et al., 2015).

The mode of action of INO80 family protein has been reinforced by the 3D structure of SWR1 complex obtained by cryo-EM (Nguyen et al., 2013). SWR1 is a 1 MDa complex consisting of 14 subunits. The Swr1 ATPase is in the core of the complex and is surrounded by three functional modules. The N-module includes: the actin-related protein Arp4 which is capable of binding canonical nucleosomes and histones; Swc4 containing a SANT domain that binds unmodified histone tails; and Bdf1 containing bromodomains for binding acetylated histones. The C-module contains Swc3, Swc2, Arp6 and Swc6 and binds free H2A.Z-H2B dimers. Finally, the heterohexameric Rvb1/Rvb2 module consists of the AAA+ ATPases Rvb1 and Rvb2, which form a ring structure. The complex adopts a more elongated conformation upon nucleosome binding, with the Swr1 subunit and the C-module extending away from the Rvb1/Rvb2 ring (Nguyen et al., 2013). SWR1 was observed to make limited contacts with the nucleosomal substrate, unlike the remodellers from other families, which may explain the specialisation of INO80 proteins for histone dimer exchange (Nguyen et al., 2013). The INO80 complex, however, was shown to make extensive contacts with the nucleosome (Tosi et al., 2013), possibly promoting conformational change in the nucleosome and histone dimer dissociation. In the resting state, both complexes adopt a closed conformation (Watanabe et al., 2015).

1.2.7 Mot1 – a remodeller not interacting with the nucleosome

Mot1 (modifier of transcription 1) is an essential SNF2/SWI2 family remodeller in *Saccharomyces cerevisiae*. Like the other members of this remodeller family, Mot1 has conserved ATPase domains comprising the lobes RecA 1 and 2. Unlike the major remodellers described above, it does not interact with a nucleosome to exert its remodelling activity. Its main function is transcription regulation through interaction with TATA binding protein (TBP) and double-stranded DNA. TATA binding protein is a transcription factor binding the TATA box – a DNA sequence found upstream of the transcription start site of promoters of some eukaryotes.

Mot1 can both repress and activate transcription. The predominant role in transcription repression is conferred by ATP-driven dissociation of TBP from dsDNA and activation via preinitiation complex (PIC) reorganisation. These effects were shown to be exerted through direct localisation to Mot1-repressed and Mot1-activated promoters (Auble et al., 1997; Dasgupta et al., 2002).

Mot1 has two major regions: a nonconserved N-terminal part containing HEAT repeats 1-5 (aa 289-1225, Uniprot ID P32333) and a conserved C-terminal part that contains the ATPase domain and HEAT 6 domain (1284-1787, UniProt ID P32333) (Auble et al., 1997). The N-terminus targets Mot1 towards TBP and the C-terminus, while aiding TBP-DNA complex recognition, primarily binds to the DNA upstream of the TATA box and acts in the ATP-driven dissociation reaction, thus removing TBP from dsDNA and exerting transcription repression (Auble et al., 1997). The ATPase domain stabilises the Mot1-TBP-DNA ternary complex and is activated in the presence of TBP (Auble et al., 1997). DNA alone does not activate the ATPase domain and the presence of TBP-DNA complex is required for that (Adamkewicz et al., 2000; Chicca et al., 1998; Wollmann et al., 2011).

Interestingly, the Mot1-TBP interaction was found to play a role in positive regulation of transcription through redistribution of TBP from high-affinity TATA-containing sites to lower-affinity poly(dA-dT) tracts (Zentner and Henikoff, 2013). These tracts alter chromatin architecture and disfavour nucleosome formation, thus making DNA accessible for binding by DNA-interacting proteins (Segal and Widom, 2009). There has been evidence showing that Mot1 assists TBP binding to DNA by clearing the promoters from repressive marks, this way facilitating activation of transcription (Andrau et al., 2002).

Even though Mot1, unlike many other Swi2/Snf2 family proteins, has not been directly linked to nucleosome organisation, MNase experiments demonstrated its involvement in +1 nucleosome positioning – the strongly positioned first nucleosome in the array between the promoter-associated nucleosome-free region and the gene body – through a mechanism that is unclear (True et al., 2016).

Recent structural studies have shed light on Mot1's interaction with TBP and DNA. The crystal structure of Mot1 in resting state showed that Mot1 exists in the auto-inhibited resting state where the N- and C-termini are in close proximity to each other (Butryn et al., 2018). Upon binding to TBP-DNA-NC2 (the transcription regulator negative cofactor 2) complex, Mot1 induces DNA bending, which facilitates its dissociation from TBP (***Fig. 1.11***) (Butryn et al., 2015).

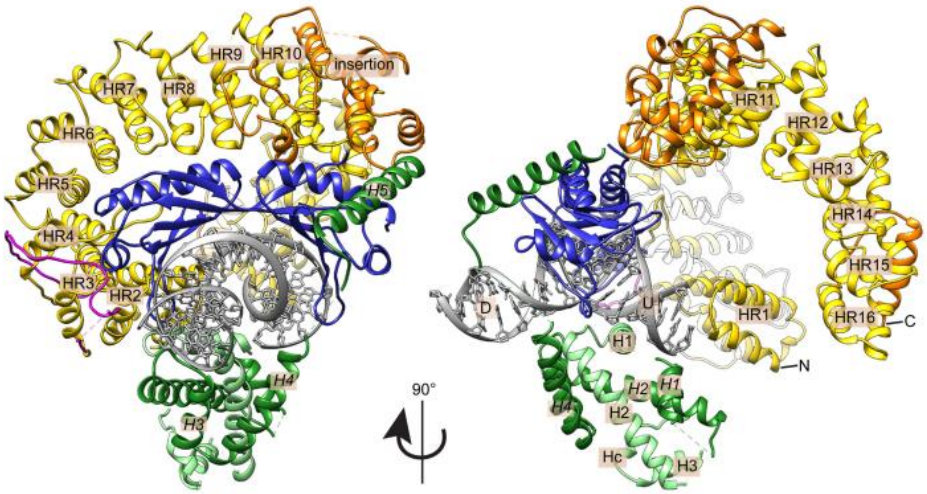


Figure 1.11 Crystal structure of Mot1-TBP-DNA-NC2 complex shown as a cartoon. Yellow – the N-terminal HEAT repeats (HR) of Mot1, orange – insertion domain of Mot1, blue – TBP, green – a TBP-associated protein NC2 (H – NC2 helices), grey – DNA. Mot1 binds predominantly to TBP via contact with helix H5 of NC2. Figure from Butryn 2015.

1.3 Lymphoid-specific helicase (LSH)

1.3.1 Biological functions of LSH

Lymphoid-specific helicase (LSH), also known as HELLS, PASG, SMARCA6 or Nbla10143, is a member of the SNF2 family of chromatin-remodelling ATPases characterised by the presence of two RecA-like ATPase lobes containing seven conserved helicase motifs (Wu and Brosh, 2012). LSH is common to mammals but close homologues are also present in *Arabidopsis*

thaliana (DDM1) (Jeddeloh et al., 1999), yeast (Basenko et al., 2016; Geiman et al., 1998; Laurent et al., 1992), *Xenopus laevis* and *Danio rerio* (Meehan et al., 2015).

Initially the *Lsh* transcripts had been identified by Northern blot in the lymphoid cell lineage exclusively (hence the designation as "lymphoid-specific") (Jarvis, 1996). Initial knockout (KO) studies confirmed its importance for the lineage by showing that LSH is required for T-lymphocyte proliferation (Geiman and Muegge, 2000). Later it was also detected in testis (Geiman et al., 1998) and, consequently, deemed to have a role primarily in recombination, particularly through DNA repair (Brosh, 2013), since these tissues are highly active in Variable, Diverse and Joining (VDJ) and meiotic recombination, respectively. Independently from that, LSH had been identified in MO7e cell line as a decreased transcript following cytokine withdrawal and revealed to belong to SNF2 family chromatin remodellers (Lee et al., 2000).

Further research demonstrated that in adults LSH is not restricted to lymphocytes and testis but is common to actively proliferating cells including the spleen, thymus, bone marrow and small intestine (Raabe et al., 2001). At the embryonic stage, LSH is ubiquitously expressed throughout the organism – a pattern recapitulated in cancer cell lines (Raabe et al., 2001). Thus, LSH expression positively correlates with cell proliferation.

In the cellular context, LSH localisation is restricted to the nucleus, in accordance with the protein's putative functions, and correlates with the S phase (Geiman and Muegge, 2000). Treatment of HEK293 nuclei with micrococcal nuclease (MNase) and subsequent sucrose gradient fractionation confirmed tight association of LSH with chromatin (Meehan et al., 2015). Immunofluorescent staining after LSH overexpression in MEFs showed that LSH predominantly has a diffused nuclear pattern with only occasional pericentric heterochromatic clustering, which increases with co-overexpression of LSH along with a chromo- and chromo-shadow domain active HP1α (Meehan et al., 2015). This pattern is consistent with the result after Triton X-100 wash, where about 50% of LSH could be Triton X-100 extracted whereas the other 50% remained tightly associated with chromatin (Yan, 2003). This pattern was found to be cell cycle-dependent: more LSH was tightly associated with chromatin during S phase than in the rest of the cell cycle (Yan, 2003).

Diffused nuclear localisation of LSH was contrasted to that of its ATPase deficient mutant LSH K237Q, which showed increased chromatin residence time, manifesting in distinct nuclear foci (Lungu et al., 2015). The authors also noticed that this interaction was dependent on

H3K9me3 modification. However, along with the intact ATPase domain, H3K9me3 was found to be important for LSH release from chromatin, rather than recruitment of LSH to it (Lungu et al., 2015).

The ATPase domain was also shown to be crucial for LSH effects on cell cycle progression (Fan, 2005; Geiman and Muegge, 2000; Yan, 2003; Zhou et al., 2009). The Intact ATPase domain blocks cells in S phase upon LSH overexpression, whereas a point mutation K254Q abrogates this effect (Raabe et al., 2001). Thus, LSH effects appear to be highly cell-type and cell-line dependent.

1.3.2 LSH knockouts

The human *Lsh* gene is located on the chromosome 10 and is encoded by 22 exons, whereas its close homologue, the mouse *Lsh*, is located on the chromosome 19 and is encoded by 21 exons. There are two published murine knockout (KO) variants of LSH, both of which were generated by homologous recombination. The first one (KO1) targeted exons 6 and 7 (helicase domains I, II and partly IIa) and the phenotype was characterised by normal embryonic development, but early post-natal lethality, reduced body weight and renal lesions, possibly due to mitochondrial abnormalities (Geiman et al., 2001). The second one (KO2) targeted exons 10-12 (helicase domains II, III and IV) and resulted in low levels of protein, which was also unstable (Sun 2004). This LSH hypomorph resulted in impaired growth, and premature aging manifested in wasting syndrome, bone fragility and early lethality (Sun, 2004). At the cellular level, LSH KO2 mice showed global hypomethylation and replicative senescence (Sun, 2004). It should be noted that the truncated version of mutated LSH was still detectable after immunoprecipitation, albeit only in small amounts, which could account for post-natal survival of LSH KO2 mice compared to KO1.

1.3.3 The role of LSH in DNA methylation

The analysis of the genome-wide methylation profile of the murine LSH KO1 demonstrated the global loss of DNA methylation at both single-copy genes and repeat sequences (Dennis, 2001; Tao et al., 2011). This was later confirmed in the murine embryonic stem cells (mESCs) and murine embryonic fibroblasts (MEFs) (Fan et al., 2003.; Myant et al., 2011; Zhu et al., 2006). Comparison of genic and intergenic regions after LSH deletion, however,

demonstrated a more pronounced effect on methylation levels of intergenic regions (Yu et al., 2014). LSH has also been found to affect the imprinting of and directly associate with a site on the *Cdkn1c* gene (Fan, 2005), however, this was the only imprinted gene linked to LSH.

Treatment of the LSH WT MEFs with a demethylating agent 5'-azacytidine resulted in an increase of centrosome numbers and a phenotype similar to that of the LSH knockouts – an indication that the changes associated with LSH perturbations are dependent on DNA methylation (Fan et al., 2003). The influence of LSH on DNA methylation was found to be independent of the expression of the DNA methyltransferases – DNMT1, DNMT3A and DNMT3B (Dennis, 2001). An experiment in which LSH-deficient MEFs were transfected with an episomal vector demonstrated that LSH is indispensable for *de novo* methylation but is not required for maintenance methylation (Zhu et al., 2006). The effect of LSH on DNA methylation was shown to be more pronounced in highly expressing genes (Tao et al., 2011).

LSH was shown to affect promoter methylation patterns of development-associated genes, such as HOX, *Oct4* and *Nanog* (Myant et al., 2011; Xi et al., 2009; Termanis et al., 2016; Xi et al., 2007; Zhu et al., 2006). In the LSH KO model a subset of Hox genes were hypomethylated and derepressed in mouse embryonic fibroblasts (MEFs), as well as in brain and liver (Xi et al., 2007). By contrast, LSH knockdown (KD) in embryonic stem (ES) cells resulted in partial expression of these genes and maintenance of stemness characteristics, such as anchorage-independent growth and alkaline phosphatase production (Sichuan Xi, Theresa M. Geiman, 2009). Moreover, LSH was found at *Oct4* and *Nanog* promoters by chromatin immuno-precipitation (ChIP). However, *in vivo* experiments showed that *Oct4* subsequently undergoes methylation in later embryonic developmental stages (Xi et al., 2009; Zhu et al., 2006), suggesting an additional or compensatory methylation mechanism, at least in these genes. In LSH rescue experiments, ectopic expression of wild-type LSH in LSH-null mouse embryonic fibroblasts (MEFs) repressed *Rhox2a* and *Rhox6* expression (Termanis et al., 2016), but not *HoxC6* gene, consistent with the previous finding that not all Hox genes are affected by LSH (Xi et al., 2007).

As noted earlier, despite diffuse nuclear localisation, LSH has a certain affinity for major and minor satellite repeats in pericentromeric heterochromatin and co-localises with HP1α (Dunican et al., 2015; Yan, 2003). This pattern was found to be independent of the pre-existing CpG methylation (Yan, 2003). The analysis of the protein-coding genes in LSH KO

MEFs revealed that, even though both exhibited reduced methylation levels, silent genes faced more dramatic loss of methylation than active genes both at transcription start sites (TSS) and gene bodies (Yu et al., 2014). This indicated that methylation in active and silent genes occurs though distinct mechanisms, and LSH has a non-redundant function in methylation of silent genes.

The effect of LSH on DNA methylation was shown to be direct, as indicated by the association of LSH with the sequences that subsequently gained methylation, which was found to be independent of the presence of a functional ATPase domain in LSH (Ren et al., 2015). In contrast, deletion of the DEAH domain (325-342 aa), abolished LSH localisation on the repeat elements (Ren et al., 2015). Mutation of the ATPase domain (K237A), however, altered nucleosome occupancy, with the levels decreasing upon cell differentiation in LSH KO condition (Ren et al., 2015).

The analysis of repeat elements also showed reduction in CG methylation in KO MEFs. However, unlike the protein-coding genes, some repeat sequences, such as intracisternal A-particles (IAPs), and major and minor satellite repeats, demonstrated some de-repression (Dunican et al., 2013; Huang, 2004; Yu et al., 2014). Changes in expression of previously repressed sequences were not solely based on changes in methylation status and in addition were associated with differential histone modifications.

1.3.4 LSH and histone modifications

The interplay between DNA methylation and histone modifications has a crucial influence on gene expression. Hence, it is not surprising that manipulations of LSH have various effects on histone modification status, and *vice versa*. For instance, treatment with histone deacetylase (HDAC) inhibitor trichostatin A (TCA) abolished LSH association with chromatin (Yan et al., 2003), indicating that LSH requires histone PTMs, namely acetylation, for binding. This can be due to local structural changes that affect LSH binding or global rearrangements in chromatin architecture.

Changes in repetitive sequences (major and minor satellite repeats, IAPs and LINEs) upon LSH deletion involve not only loss of DNA methylation, but also gain of active histone marks, such as acetylation and methylation at H3K4 (Dunican et al., 2013; Yan, 2003; Yu et al., 2014), and loss of a repressive mark – methylated H3K9 (Dunican et al., 2013). Promoters of some genes

encoding development-associated transcription factors also underwent loss of methylation on H3K9 (Myant et al., 2011). Contrary to this, one study showed that the regions, which underwent the greatest loss of CG methylation, acquired the repressive H3K27me3 mark, indicating possible redundancy in silencing mechanisms (Yu et al., 2014).

1.3.5 LSH and senescence

The LSH KO MEFs showed impaired centrosome formation and subsequent de-regulation of cell division (Fan et al., 2003). Interestingly, this defect did not lead to any detectable cell transformation, and the LSH-deficient cells, on the contrary, failed to proliferate even after introduction of SV40 large T antigen (Fan et al., 2003), indicating resistance of LSH-deficient MEFs to the cancerous phenotype. This was accompanied by an increase in centrosome numbers compared to the WT and deregulation of spindle formation, resulting in aberrant cell division and subsequent senescence as a response mechanism (Fan et al., 2003). On the contrary, the overexpression of LSH leads to a delay in senescence (Zhou et al., 2009).

The link between LSH deficiency and cellular senescence was further explored in a study of the relationship between LSH and a tumour suppressor protein p16 that stalls progression through cell cycle by inhibiting cyclin dependent kinases (Zhou et al., 2009). In that study, LSH and p16 were shown to have opposite effects on the progression to senescence of a human lung fibroblast cell line 2BS: LSH overexpression rescued the proliferative phenotype and resulted in a decrease in p16, whereas LSH shRNA knock-down resulted in an increase in p16 expression – an effect that was shown to be mediated by HDAC1 and HDAC2. Interestingly, silencing of both LSH and p16 rescued the proliferative phenotype (Zhou et al., 2009), demonstrating that LSH effects in proliferation may be performed by another protein. This study contradicts the earlier finding that LSH overexpression in 2BS cells leads to cell cycle arrest and the intact ATPase domain is crucial for this effect (Raabe et al., 2001).

It was found that LSH promoter activity is under the control of E2F1 - a transcription activator involved in the control of cell cycle and mediating either cell proliferation or apoptosis (Niu et al., 2011).

1.3.6 LSH in DNA integrity maintenance during replication and repair

LSH has been implicated in DNA integrity maintenance during replication and repair (Burrage et al., 2012). Its possible role in replication has been primarily linked to chromosome segregation fidelity. It has been recently found that Irc5 (Increased Recombination Centers 5), the LSH homologue from *Saccharomyces cerevisiae* identified in DNA damage response screen, associates with the Scc1 subunit of the cohesin complex (Litwin et al., 2017). Cohesin is a multisubunit complex that is formed in late G1 as a 3-dimensional cylindrical structure connected to the sister chromatids, tethering them to each other during cell division ensuring proper alignment and segregation (Nasmyth, 2011).

When Irc5 is disrupted, cohesin levels on the centromeres and chromosome arms decrease, causing mild premature separation of sister chromatids and the phenotype similar to that with cohesin loading complex mutants (Hakimi et al., 2002; Huang, 2004; Litwin et al., 2017), indicating that Irc5 may be crucial for cohesion in certain locations or under specific conditions.

The absence of Irc5 resulted in reduction of the cohesin loading complex protein Scc2 levels, which was accompanied by reduction of Scc2 and cohesin subunit Scc1 interaction and compromised cohesin assembly (Litwin et al., 2017). On the other hand, when a cohesin loading complex DAEA box yeast mutant had wild-type Irc5, it could survive in semi-permissive (30^{o}C) conditions (Litwin et al., 2017), but not in the case of an ATPase-deficient Irc5 mutant, which highlights the importance of Irc5 chromatin remodelling activity in cohesin loading and sister chromatid separation. Similar mechanismmay be relevant for LSH functions.

LSH is also linked to the repair of double-strand breaks (DBSs), both in germline and somatic cells. Improper repair of DBSs leads to chromosomal instability and cancer development (Jackson and Bartek, 2009). There are two repair mechanisms in the cell: canonical non-homologous end joining (c-NHEJ) and homologous recombination (HR). HR is based on the presence of a sister chromatid as a repair template, which occurs in S and G2 (San Filippo et al., 2008).

In oocytes, LSH KO leads to incomplete DSB repair (DBSR) and impaired establishment of crossovers in meiosis, which correlated with the loss of methylation at repeat sequences (De La Fuente et al., 2006). Similarly, in spermatocytes, double-strand breaks could not be efficiently repaired, which was indicated by the persistence of asynapsed chromosomes

(Zeng et al., 2011). These findings contradict the initial localisation of LSH outside the areas of active meiotic division (Raabe et al., 2001). In somatic cells, LSH KO leads to a decrease in H2AX histone variant phosphorylation, which is a modification acquired straight after DNA damage to mark the areas that need to be repaired, and reduced survival after ionising radiation exposure with no change detected in the levels of 5meC (Burrage et al., 2012), indicating that there are non-redundant functions of LSH in DNA damage repair in promoting efficient phosphorylation upon DNA damage.

A recently identified interaction of LSH with a DNA repair-associated protein CtIP reinforced the role of LSH in chromatin remodelling. LSH is involved in DSBR during HR in heterochromatin in G2 through recruitment of CtIP at DSB locations (Kollarovic et al., 2018). CtIP is an endonuclease that interacts with the MRE (Mre11/Rad50/Nbs1) DNA damage sensor complex and promotes end resection on DSBs to generate 3'-end single strands (Makharashvili and Paull, 2015), which is a crucial step in HR (Sartori et al., 2007).

DSBR in heterochromatin is complicated by the condensed environment, resulting in less efficient repair of DSB and more frequent mutations in heterochromatin due to the condensed environment and repetitive sequences (Schuster-Böckler and Lehner, 2012; Woodbine et al., 2011). Therefore, chromatin remodelling and decondensation are essential for heterochromatic DSBR. LSH is likely to create a permissive environment in heterochromatin for DSBR machinery recruitment and activity (Kollarovic et al., 2018). This role of LSH in DSBR was, however, redundant, since its deficiency did not result in dramatic loss of HR (Kollarovic et al., 2018). Therefore, it is conceivable that its role in DSBR is restricted to heterochromatin.

LSH activity in DSBR requires the intact ATPase domain, as opposed to the mutated ATP-binding lysine K254R (corresponding to K237 in murine LSH), which failed to recruit CtIP to DSBs (Kollarovic et al., 2018). This interaction was direct as indicated by co-immunoprecipitation and His-pull-down (Kollarovic et al., 2018).

1.3.7 LSH in disease

Genome-wide perturbations in DNA methylation and histone PTMs upon LSH point-mutations or complete deletion have numerous implications in disease, notably in cancer. As

noted previously, LSH is highly expressed in the lymphoid lineage and actively proliferating cells.

Tumour protein p63, whose overexpression results in overcoming senescence, induction of stem-like properties and tumour formation, was found to act by activating LSH through consensus binding sites – LSH was found to be imperative for senescence bypass and able to overcome a Ras-induced senescent phenotype (Keyes et al., 2011). E2F transcription factor has also been found to require LSH for its oncogenic functions, whereas knockdown of LSH ablated E2F-mediated gene expression upon re-entry to the cell cycle (von Eyss et al., 2012).

LSH has also been found to suppress expression of fumarate hydratase by recruitment of G9a – an epigenetic silencing factor (He et al., 2016). Fumarate hydratase is an enzyme present in mitochondria, and it plays a prominent role in the Krebs cycle, facilitating oxidative metabolism and supporting anabolic reactions. Suppression of fumarate hydratase leads to alterations in cellular metabolism and diversion in energy production from oxidative phosphorylation to glycolysis – the phenomenon known as the Warburg effect. By suppressing production of a Krebs cycle intermediate, LSH supports the metabolic switch.

A form of LSH with a C-terminal in-frame deletion of 75 bp was found in 57% of acute myeloid leukaemia and 37% acute lymphoblastic leukaemia samples used in the study (Lee et al., 2000). Intriguingly, this deleted region contains a motif that has been observed to couple ATP hydrolysis to chromatin remodelling (Smith and Peterson, 2005).

Interestingly, not only overexpression of LSH can lead to proliferative abnormalities. A subset of LSH-null mice developed hematopoietic neoplasms as indicated by histopathology of spleen, liver and bone marrow – the regions where LSH is highly expressed (Fan et al., 2008).

An RNA screen in renal cell carcinoma identified LSH as one of the differentially expressed genes compared to normal kidney cells. Statistical analysis demonstrated that lower expression of LSH correlated with longer survival time and could potentially be used as a prognostic marker (Chen et al., 2017). LSH was also found to be a negative predictor of survival in astrocytomas and glyoblastomas – this time LSH expression was dependent on E2F1 transcriptional activation (Xiao et al., 2017).

Cancer, however, is not the only disease affected by LSH. ICF (Immunodeficiency, centromeric instability, facial anomaly) syndrome is an extremely rare autosomal recessive disease, characterised by fatally low immunoglobulin counts with normal or decreased B and

T lymphocyte counts. At the molecular level, the disorder is manifested by reduced methylation of juxta-centromeric heterochromatin repeats of chromosomes 1, 9 and 16 (Ehrlich et al., 2008; Thijssen et al., 2015). The genes with mutations implicated in ICF syndrome are LSH, *de novo* DNA methyltransferase DNMT3B, zinc-finger-containing ZBTB24 and CDCA7 (cell division cycle associated 7). All these genes were found to be involved in centromeric CpG methylation – a manifestation of genetic heterogeneity that converges in a single disease (Thijssen et al., 2015).

1.3.8 LSH homologues

Apart from LSH variants in human, mouse, zebrafish and clawed frog (see Homology Modelling section), there are several LSH homologues in more distant species,for which experimental data have been collected.

Decrease in DNA Methylation 1 (DDdm1) is a chromatin remodeller found in *Arabidopsis thaliana*. Mutations in Ddm1 result in methylation loss mainly in heterochromatin, and to a smaller extent in genes (Jeddeloh et al., 1999). Like LSH, Ddm1 mediates cytosine methylation for repetitive sequence repression in plants, specifically centromeric retrotransposons and centromeric satellite repeats, which is essential for non-coding RNA expression from repetitive sequences (Tan et al., 2018). RNA-directed DNA methylation (RdDM) is a process in plants in which short interfering RNAs (siRNAs) mediate transposon methylation. Ddm1 knockouts in tomato showed severe hypomethylation of heterochromatic transposon regions in CG and CHG, but increased for CHH (C = cytosine, H = A, T or C) (Corem et al., 2018), and the function of Ddm1 was antagonistic to RdDM in rice and tomato (Corem et al., 2018; Tan et al., 2018).

The function of Ddm1 was linked to the presence of the linker histone H1: inactivation of H1 in a Ddm1-deficient phenotype resulted in restored methylation, indicating that DNA methyltransferase require Ddm1 to bypass the H1 barrier for heterochromatic silencing (Lyons and Zilberman, 2017; Zemach et al., 2013). Interestingly, Ddm1 was found to be crucial for methylation of nucleosomal DNA, which was inhibited by H1 in the absence of Ddm1 and restricted primarily to the DNA linker regions (Lyons and Zilberman, 2017).

The homologues from *S. cerevisiae* Irc5 is a Snf2 family member (Flaus and Owen-Hughes, 2011) identified in a genome-wide analysis as gene, deletion of which resulted in an increase

of Rad52 foci in proliferating cells (Alvaro et al., 2007). Irc5 has recently been found to participate in cohesin loading, which requires its intact ATPase domain (Litwin et al., 2017) (see Section 1.3.6). However, this function of Irc5 is not linked to DNA methylation, since it is absent in yeast.

An even more distant homologue from the mould *Neurospora crassa* called Mus30, or Chromatin Remodelling Factor 5, was also found to play a crucial role in preventing genome instability with its levels increased following DNA damage, and this function depended on association of Mus30 with *Neurospora* WDR76 (Basenko et al., 2016). The human homologue of WDR76 is an E3 ubiquitin ligase linker protein and was recently found as a potential tumorigenesis suppressor through degradation of Ras GTPase (Jeong et al., 2019). The lack of Mus30 affected neither double strand break repair (DSBR), HR nor NHEJ (Basenko et al., 2016).

Taken together, the insights into LSH homologues provide additional information about possible functions of LSH in DNA methylation as well as additional pathways not directly related to DNA methylation.

1.3.9 LSH structure and domains

LSH is classified as a chromatin remodelling protein belonging to the Snf2-like subfamily of SF2 superfamily of helicase-like proteins (Mueller-Planitz et al., 2013). An extensive study of Snf2 subfamily members based on sequence similarity placed LSH in close proximity to the Snf2 protein – a catalytic component of SWI/SNF chromatin-remodelling complex from *Saccharomyces cerevisiae* (Flaus, 2006).

There is no crystal structure of LSH available, and all structural insights come from the sequence similarity to characterised protein. The N-terminus of LSH is disordered and contains a coiled coil motif (residues 14-96) that may interact with other proteins. In a pull-down experiment with an LSH fragment, GST-LSH (1-503) resulted in binding of a de novo DNA methyltransferase DNMT3B to the fragment (Myant and Stancheva, 2008). Residues

126-131 represent a nuclear localisation signal; upon its disruption via site-directed mutagenesis the subcellular localisation of LSH changes (Lee et al., 2000) (Fig. 1.12).

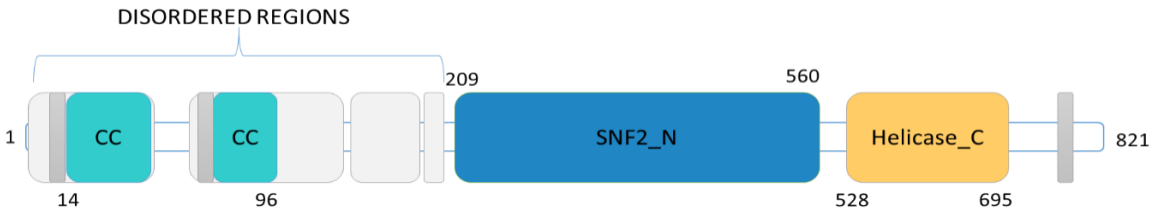


Figure 1.112 Schematic structure of LSH from Mus musculus with coiled coil domains (CC), SNF2 ATPase domain and the C-terminal helicase domain.

One of the main features of the Snf2-like family proteins, including LSH, is the seven helicase-like motifs, spanning the ATPase region. The ATPase region consists of two RecA domains of alpha-beta topology, which are connected by the elongated linker between the helicase motifs III and IV (Flaus, 2006). The region between residues 202 and 394 belongs to the DEAD-like helicase superfamily and includes the ATP-binding Walker A (phosphate-binding loop GILA DEMGLGKT in *M. musculus* LSH) and Walker B motifs (Mg^{2+}-binding aspartate). The ATPase mutant LSH^{K237Q} (part of the Walker A motif) has shown slight reduction in ATPase activity compared to the weakly active ATPase *in vitro* by thin layer chromatography (Kevin Myant's thesis).

A study of Snf2 identified a tryptophan (W1185) residue in the C-terminus upstream of motif VI, which plays an important role in interaction with the nucleosome core (Liu et al., 2017a). The residue inserts itself into the minor groove of DNA and interacts with the 3'-strand. This residue is conserved in mouse LSH (W679) and in Chd1 and ISWI remodellers. Site-directed mutagenesis of this residue to a smaller one or a charged one could reveal its role in LSH interaction with NCP.

DDM1, an LSH homologue in *Arabidopsis thaliana*, has been found to interact with nucleosomal DNA and to remodel it upon ATP addition (Brzeski and Jerzmanowski, 2003). The addition of the protein to a nucleosome resulted in the octamer sliding from the end to the centre of the nucleosome, resulting in distinct nucleosomal species visible on the gel (Brzeski and Jerzmanowski, 2003). A high degree of sequence similarity (49.4% by EMBOSS Needle) between DDM1 and LSH indicates that the latter possesses a similar mode of action.

However, the absence of NCP remodelling activity by LSH *in vitro* shows that additional co-factors or binding partners are likely required.

1.4 Cell division cycle-associated protein 7 (CDCA7) in health and disease

1.4.1 Zinc finger proteins

Zinc finger proteins (ZFPs) are some of the most abundant proteins across species. Their interaction partners include DNA, RNA and proteins and their functions span such processes as transcriptional regulation, protein degradation, DNA repair, signal transduction (Cassandri et al., 2017). ZFPs are implicated in a variety of diseases related to tumorigenesis, diabetes, neurodegeneration and beyond (Cassandri et al., 2017).

The classic zinc finger domain contains C2H2 motif, in which two cysteines and two histidines are coordinated by a zinc ion (***Fig. 1.13 A***) (Zhang et al., 2011). This motif, however, is not applicable to all ZFPs, and other abundant non-classical ZFPs include RING (Really Interesting New Gene) (Linke et al., 2008), PHD (Plant Homeodomain) (Borgel et al., 2016) and LIM (Lin11-Isl1-Mec3) containing proteins (Turner and Miller, 1994) (***Fig. 1.13 A, B***). The classic C2H2 ZF domain has a common conserved motif TGEKP connecting adjacent ZFs (John H Laity et al., 2000) and transitioning from disordered to ordered state upon binding to DNA (J. H. Laity et al., 2000).

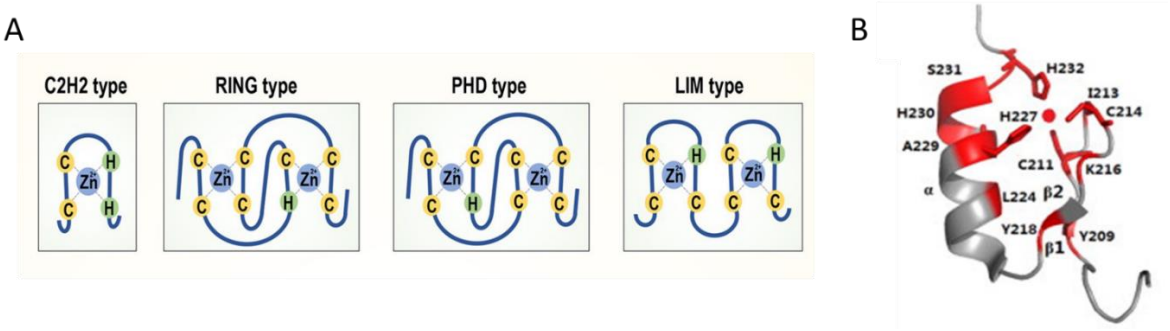


Figure 1.13 Zinc finger domain. A) Schematic representation of the most common types of zinc finger domain (from Cassandri 2017); B) Crystal structure of the Cys2His2-type zinc finger domain of human transcription factor DPF2. The residues chelating the zinc atom (C211, C214, H227, H232) are shown in red sticks (from Zhang et al. 2011).

RING ZFD are distinct from the classical C2H2 ZF-containing motif (Miller et al., 1985). Described RIGN ZFDs contain eight conserved residues (cysteines and histidines), which

coordinate two zinc ions, four residues per zinc ion (Borden and Freemont, 1996). RING ZFD are similar to LIM ZFD and PHD ZFD, however, the central histidine residue is located in distinct positions within the motif (Borden and Freemont, 1996). The initial role for RING-containing proteins was assumed to be protein dimerisation but more research demonstrated this domain to be present abundantly in E3 ubiquitin ligase – a protein transferring a ubiquitin molecule from E2 conjugating enzyme to the substrate (Joazeiro and Weissman, 2000). This function is well characterised in the case of the proto-oncoprotein Mdm2, whose E3 ligase activity is required for the protein p53 ubiquitylation and degradation, resulting in oncogenic phenotype (Lindstrom et al., 2007). Mutations in Mdm2 ZFD results in Mdm2 escaping degradation itself and causing uninhibited degradation of p53, an indication of a cancerous phenotype (Lindstrom et al., 2007).

1.4.2 CDCA7

CDCA7/JPO1 is a ZFP that was identified in a differential expression analysis of cell cycle genes and was shown to have expression patterns similar to cell-division-cycle (CDC) and cyclin genes (Walker, 2001). The CDCA7 gene produces a conserved protein of around 47 kDa, with the orthologues found across different species such a human, chicken and frog (OrthoDB query, https://www.orthodb.org/?query=cdca7). Its expression peaks at the transition from G1 to S phase (Whitfield et al., 2002). CDCA7 is highly expressed in undifferentiated cells and in the colon, thymus and small intestine, with low levels in the spleen, bone marrow and peripheral leukocytes (Prescott et al., 2001).

CDCA7 in transcription

CDCA7 is a responsive gene of a proto-oncogenic transcription factor c-Myc, which plays a prominent role in the cell cycle, apoptosis and cell differentiation (Prescott et al., 2001). CDCA7 does not possess strong transformation activity but it enhances a transformation-defective c-Myc mutant, possibly through the involvement in anchorage-independent growth (Jiménez-P et al., 2018; Prescott et al., 2001). Even though c-Myc was shown to bind directly to the CDCA7 gene, the CDCA7 gene did not show any significant sequence conservation at and around the c-Myc binding site, as opposed to the majority of the c-Myc target genes (Haggerty et al., 2003). CDCA7 interacts with c-Myc in a phosphorylation-

dependent manner: the unphosphorylated form of CDCA7 is present in the nucleus and is associated with c-Myc, but following phosphorylation of threonine 163 by protein kinase B (PKB/AKT), it dissociates from c-Myc, binds to 14-3-3 protein and is sequestered to the cytoplasm (Gill et al., 2013).

c-Myc is not the only transcription factor targeting CDCA7 expression. CDCA7 is also a direct target of E2F1 and E2F4 transcription factors that play a crucial role in cell cycle regulation and DNA replication, with both proteins having strongly correlating expression patterns (Goto et al., 2006). Notch, a transcription factor essential for normal developmental processes, targets CDCA7 in embryonic hematopoietic development: induction of CDCA7 resulted in maintenance of hematopoietic stem cell (HSC) population through activation of the HSC precursors and downregulation of CDCA7 induced hematopoietic differentiation (Goto et al., 2006).

CDCA7 has inverse expression pattern with a transcription factor Pax6 in cortical neuroepithelium progenitors, indicating that Pax6 has a repressive effect on CDCA7 expression and that this process is required for the normal generation of intermediate neural progenitors (Huang et al., 2017).

CDCA7 is not only the target of TFs but itself has been found to possess transcriptional activity in its C-terminal cysteine-rich zinc finger domain (Goto et al., 2006). It was found to be involved in transmissible gastroenteritis virus (TGEV) pathogenicity via microRNA action: microRNA miR-4331 was targeting 3' UTP of CDCA7 mRNA, resulting in transcription inhibition of TGEV gene 7 (Song et al., 2015). Even though in the study by Song et al. CDCA7 has not been directly shown to possess transcriptional activity, the results reinforced that notion that it is a part of a transcriptional pathway. Interestingly, CDCA7 is also a target of long non-conding RNA in promoting colorectal cancer and this process is counterbalanced by CDCA7 downregulation through the action of microRNA miR-302e (Li 2019).

CDCDA7 in cancer

CDCA7 involvement in transcription links it to cancer development and progression. Moreover, its close homologue CDCA7L has been extensively linked to cancer (see below). CDCA7 is often overexpressed in cancers, such as colon, rectum, ovary, lung, uterus and stomach, and particularly in acute and chronic myelogenous leukaemia (Osthus et al., 2005).

Ectopic expression of CDCA7 in transgenic mice correlated with an increased likelihood of development of solid tumours in B and T cells (Osthus et al., 2005). CDCA7 was found to be a predictor of poor prognosis in triple-negative breast cancer through increasing expression of EZH2, a marker of aggressive breast cancer (Ye et al., 2018).

CDCA7 is elevated in lung adenocarcinoma, and its downregulation resulted in arrest in G1 and subsequent apoptosis of cancer cells (Wang et al., 2019). CDCA7 is upregulated in Burkitt's lymphoma and T- and B-lymphoid tumours cell lines and contributes to anchorage-independent growth in lymphomagenesis (Jiménez-P et al., 2018). CDCA7 upregulation is associated with poor prognosis in triple-negative breast cancer and predicts metastatic relapse through upregulating EZH2, which is a marker of aggressive breast cancer (Ye et al., 2018). Consistent with this, CDCA7 silencing was shown to decrease lymphoma tumour cell invasion, which was linked to alterations in cytoskeleton formation through effects on actin and myosin, highlighting CDCA7 important role in metastasis (Martin-Cortazar 2019).

1.4.3 CDCA7 in Immunodeficiency, Centromeric Instability and Facial Anomalies (ICF) syndrome

Like with LSH, CDCA7 mutations are implicated in the recessive autosomal disorder Immunodeficiency, Centromeric Instability and Facial Anomalies (ICF) syndrome. Whereas the LSH mutation-associated form of ICF is termed ICF4, the CDCA7-associated form is termed ICF3 and is linked to several homozygous missense mutations (R274C, R274H, G294V, R304H in human) located in the zinc finger domain (Thijssen et al., 2015). These mutations, like in the case of the other proteins implicated in ICF, are associated in hypomethylation of pericentromeric heterochromatin (Thijssen et al., 2015) but not subtelomeric heterochromatin methylation, as it is the case with DNMT3B (ICF1) (Toubiana et al., 2018). Another ICF-related protein ZBTB24 (ICF2 form) was found to affect transcription of CDCA7 and it was enriched in the CDCA7 promoter, indicating that it is a CDCA7 transcription factor controlling CDCA7 expression through its BTB (BR-C, ttk and bab) domain (Wu et al., 2016a). The converging actions of the ICF-associated proteins were further elucidated by the finding that LSH requires CDCA7 to carry out its chromatin remodelling function (as demonstrated by nucleosome remodelling assay) (***Fig. 1.14***) and ICF-associated CDCA7 mutations abrogate LSH recruitment to chromatin (Jenness et al., 2018).

Apart from LSH, CDCA7 interaction with classic non-homologous end joining (C-NHEJ) proteins Ku70 and Ku80 was demonstrated by co-immunoprecipitation experiments (Unoki et al., 2018). ICF mutations of CDCA7 as well as nuclease treatment abrogated the interaction (Unoki et al., 2018), indicating that CDCA7 interacting with Ku70 and Ku80 is bridged by DNA or a nucleosome, and that CDCA7 mutations may result in the ICF phenotype not only through loss of DNA methylation but also through genome instability. Nevertheless, CDCA7 mutations are still linked to altered DNA methylation, as indicated by the analysis of the global methylome of ICF patients, which revealed differential methylation in CpG-poor regions of genes related to neuronal development in patients with CDCA7 mutations (Velasco et al., 2018).

1.4.4 CDCA7-LSH interaction

Recent findings identified CDCA7 as a direct interaction partner of LSH (Jenness et al., 2018), and this interaction provides molecular conditions for LSH ATPase activation. Mass spectrometry of unmodified or silent chromatin of *Xenopus* egg extracts showed stoichiometric levels of LSH and CDCA7, indicating that they form a complex through the cell cycle, with the exception of M phase and cytokinesis. However, immunodepletion of either protein did not deplete the other, indicating that only a fraction of each protein exists in the complex (Jenness et al., 2018). Importantly, this interaction is conserved, since exogenously expressed CDCA7L also interacted with LSH in egg extract (Jenness et al., 2018)

LSH binding to chromatin was dependent on CDCA7, but CDCA7 did not need LSH for chromatin binding, and CDCA7 ICF mutants R274C/H and R304H blocked its binding to chromatin and failed to recruit LSH (Jenness et al., 2018).

The nucleosome remodelling performed by Jenness et al. (2018) utilised a mononucleosome with symmetric 20 bp DNA linkers on each side and an engineered PstI restriction site 15 bp into the nucleosomal core. Following remodelling, the PstI site is exposed and cleaved by PstI restriction enzyme. The samples are then resolved on a PAGE (native or denaturing). Remodelled nucleosomal species or cleaved DNA (on a native or denaturing gel, respectively) would be detected as a faster migrating band (***Fig. 1.14***).

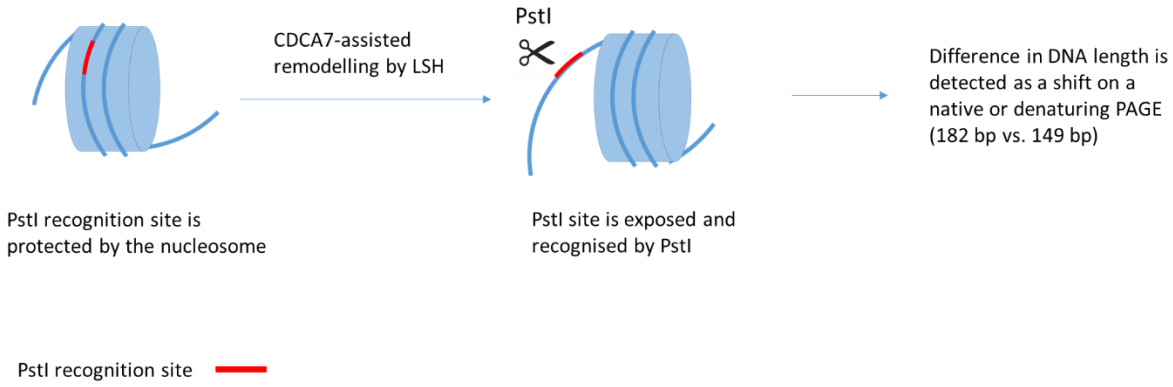


Figure 1.14 Nucleosome remodelling assay used to detect chromatin remodelling by LSH in the presence of CDCA7 as described in Jenness et al. 2018. Nucleosomal DNA contains a PstI restriction site that is protected by the nucleosomal core. After remodelling is performed the site shifts from the core and becomes accessible to PstI restriction enzyme. Enzymatic cleavage results in generation of a nucleosomal species of a smaller size. The size difference is observed on a native or denaturing gel as a shift of a nucleosome (native) or free DNA (denaturing).

1.4.5 Cdca7L - a close homologue of CDCA7

CDCA7 has 45% protein sequence identity and 77% in the C-terminal region with its homologues CDCAL/JPO2/R1. The C-terminus of CDCA7 contains a RING zinc finger domain (Bartholomeeusen et al., 2007). CDCA7L was identified as an inhibitor of monoamine oxidase A (MAO A) – an enzyme that causes degradation of the crucial neurotransmitters serotonin, dopamine and norepinephrine and produces neurotoxic reactive oxygen species – by directly binding to MAO A promoter *in vivo* (Chen et al., 2005). CDCA7L is also a c-Myc responsive gene and protein interacting partner, which rescues transforming activity in a transformation-defective c-Myc mutant (Huang et al., 2005). MAO A inhibiting function of CDCA7L was found to be involved in apoptosis inhibition: c-Myc-mediated expression of CDCA7 inhibited MAO A (Ou et al., 2006). The MAO A inhibition, in turn, was accompanied by the increase in cell-cycle regulators E2F1 or cyclin D1, thus enhancing cell proliferation (Ou et al., 2006).

There is ample evidence of CDCA7L involvement in malignancy formation. CDCA7L is upregulated in common the cancer hepatocellular carcinoma (HCC) and promotes its progression from G1/G0 to S phase through activation of ERK1/2 signalling pathway and cyclin D1 (required for progression through G1) expression (Tian et al., 2013). It is also expressed in human glioblastoma cells, and its high expression was a poor prognostic indicator for glioma patients (Ji et al., 2019; Shen et al., 2019). Downregulation of CDCA7L

was associated with blocking of cell cycle in G1/G0 through reduction of cyclin D1 expression (Ji et al., 2019), indicating that cyclin D1 is a target gene of CDCA7L.

CDCA7L overexpression results in metastatic medulloblastoma through interaction with LEDGF/p75 protein and regulation of protein kinase B/AKT signalling (Chan et al., 2016; Huang et al., 2005). The interaction with p75 is mediated by the N-terminal 61-residue region on CDCA7L (residues 58-119) and allows p75 to tether CDCA7L to chromosome during mitosis (Maertens, 2006). This CDCA7L-p75 interaction was shown to prevent binding of the p75 C-terminus to HIV-1 integrase and overexpression of CDCA7L resulted in a modest inhibition of HIV-1 replication (Bartholomeeusen et al., 2007). Interestingly, CDCA7L overexpression resulted in transactivation of HIV-1 (Bartholomeeusen et al., 2007), indicating that its function as a transcription factor is separate to its function as a competitive binder of p75.

At the epigenetic level, CDCA7L repression in melanoma cells was associated with loss of activating mark H3K4me2 and gain of repressive mark H3K27me3 (Cannuyer et al., 2015).

1.5 Aims of the project

LSH is a putative remodeller and not only does its functions have implications for cancer, DNA repair and *de novo* DNA methylation but it also has an elusive implication in the ICF syndrome. Previously, LSH structure was attempted to be elucidated by high and low resolution structural methods, however, due to a large degree of disorder a detailed LSH structure is hard to solve. Therefore, it was decided to investigate structural aspects of LSH and using homology modelling, as well as to optimise expression of LSH for structural studies.

I attempted to further investigate LSH-nucleosome complex with various DNA linker lengths and histone variants and to form LSH-nucleosome complex to study its biochemical, biophysical and structural properties.

In a parallel study, a newly identified binding partner of LSH termed CDCA7 was studied in more details. Like LSH, CDCA7 has limited structural information and no structural information for its homologues, therefore, I used *ab initio* modelling in attempt to gain further understanding of its interaction with LSH and DNA. Expression, purification and characterisation of CDCA7, as well as investigation of the possible complex formation between it and LSH were attempted.

2. MATERIALS & METHODS

2.1 LSH and CDCA7 cloning, expression and purification

2.1.1 Protein constructs and their characteristics

The following constructs were used in the present study:

Table 2.1 Details of the proteins used the present study.

Protein	Origin organism	Construct (aa)	Tag (C- or N-terminal)	Vector	Expression organism	MW (kDa)	Extinction coefficient (M^{-1} cm^{-1})	pI
LSH	*M. musculus*	1-821	C-6xHis	pFL	*Sf9*	96	77070	8.2
LSH_K237Q*	*M. musculus*	1-821	C-6xHis	pFL	*Sf9*	96	77070	8.2
CDCA7	*M. musculus*	1-382	N-GST-3C	pFL	*Sf9*	44	23920	8.94
CDCA7	*M. musculus*	1-382	N-GST-3C	pEC-KCG	*E. coli*	44	23920	8.94
CDCA7_T1	*M. musculus*	64-382	N-GST-3C	pEC-KCG	*E. coli*	37	22180	8.64
CDCA7_T2	*M. musculus*	151-382	N-GST-3C	pEC-KCG	*E. coli*	27	22180	8.99

Antibiotic resistance: pFL – ampicillin, pEC-KCG – kanamycin

*Previously cloned by Simon Varzandeh

2.1.2 DNA manipulations

2.1.2.1. Primers used for LSH and CDCA7 cloning

The constructs were amplified with the following primers purchased from integrated DNA Technologies:

Table 2.2 Primers used for amplification of the of LSH and CDCA7 constructs used in cloning.

Name	Sequence (5' to 3')
LSH_full_length_F (*Sf9*)	TTTCGCGGATCCATGGCCGAA
LSH_full_length_R (*Sf9*)	CTCGTCGACTCAGTGGTGATGGTTTT
CDCA7_full_length_F_LIC (*Sf9*)	CCAGGGGCCCGACtcgTCCGACTCCGAGGAAGAAGA
CDCA7_full_length_R_LIC (*Sf9*)	GAGTTCGAGATGCAGGCCTAAGCAGTCGGTGGCGGTCTG
CDCA7_full_length_F_LIC (*E. coli*)	CCAGGGGCCCGACtcgATGGAAGCACGTCGTGCACG
CDCA7_T1_F_LIC (*E. coli*)	CCAGGGGCCCGACtcgCGTCATAGCGGTCCGCTGC
CDCA7_T2_F_LIC (*E. coli*)	CCAGGGGCCCGACtcgGGTCGTCATTCACTGCCTGGTCA
CDCA7_R_LIC (*E. coli*)	CAGACCGCCACCGACtgcTTAGGCCTGCATTTCAAATTCTTGTTTCAG

M13_F	CCC AGT CAC GAC GTT GTA AAA CG
M13_R	AGC GGA TAA CAA TTT CAC ACA GG

2.1.2.2. PCR

Templates for PCR amplification were either previously generated in the lab or cDNA was purchased from GeneArt (Thermo Fisher Scientific). Polymerase chain reaction (PCR) (Table 2.3) was carried out on the Biometra TProfessional Gradient 96 Thermocycler. To clone the protein constructs, Q5 polymerase (NEB) was used, according to the manufacturer's protocol.

The thermocycler programme was as follows:

Table 2.3 PCR programme used for construct amplification with Q5 DNA polymerase.

Step	Temperature	Time	Cycles
1	98°C	1 min	
2	95°C	30 s	30x
3	68°C	30 s	
4	72°C	1 min	
5	72°C	3 min	
6	4°C	Hold	

Table 2.4 Reagents with stock concentrations and volumes used in PCR reactions for protein constructs amplification.

Reagent (stock concentration)	Volume
5x Q5 buffer	10 µl
dNTPs (10 µM)	1 µl
DMS0 (100%)	2.5 µl
Primer_F (10 µM)	1 µl
Primer_R (10 µM)	1 µl
DNA template (10 ng/µl)	1 µl
Q5 DNA polymerase	1 µl
ddH_2O	31.5 µl
Total	50 µl

2.1.2.3. Plasmid restriction digest

Plasmids containing the appropriate insert and backbone were incubated with FastDigest restriction enzymes (Thermo Scientific) (see Table 2.4) for 30-60 min at 37°C, and 2-5 µl of each reaction sample was run of 0.5-1% (w/v) agarose gel prepared with SafeView DNA stain

(NBS Biologicals). The appropriate bands were visualised using a UV lamp, cut out, purified from the gel using QIAGEN QIAquick Gel Extraction kit and eluted with 30 µl ddH_2O.

Table 2.5 Reagents and their volumes used in the plasmid restriction digest.

Reagent and stock concentrations	**Volume**
Restriction enzyme (e.g. SalI or BamHI)	1 µl
DNA to 1 µg	X µl
Buffer (10x)	2 µl
ddH_2O	To 20 µl

2.1.2.4 DNA ligation

Ligation reactions involved the gel-purified vector and insert DNA approximately at the molar ratio 1:3, respectively. The reaction (Table 2.5) was incubated at room temperature for 1h or at 16^oC overnight. 1 µl of each resulting reaction was then used to transform chemically competent *E. coli* DH5α cells.

Table 2.6 Reagents and their volumes used for DNA ligation.

Reagent and stock concentrations	**Volume**
T4 DNA ligase	1 µl
T4 DNA ligase buffer (10x)	2 µl
Vector DNA 100 ng	X µl
Insert DNA 200-500 ng	X µl
ddH_2O	To 20 µl

2.1.2.5 Ligation-independent cloning (LIC)

Ligation-independent cloning (LIC) is an efficient method for construct generation. It relies on the production of a vector and an insert with complementarity regions. The insert is then incubated with dATP and T4 DNA polymerase, which possesses a 3'-5'exonuclease activity that creates the complementary overhangs. The vector is incubated with dTTP and T4 DNA polymerase. These reactions generate complementary overhangs, which ligate when the insert and vector are mixed together. Adding only one of the four nucleotides (dATP or dTTP) allows balancing out the exonuclease and polymerase activity of T4 DNA polymerase, stalling of T4 DNA polymerase and creates the overhangs.

CDCA7 coding DNA sequence (CDS), codon-optimised for either insect cell or bacterial expression, was purchased from GeneArt (Thermo Fisher Scientific). The CDS was amplified using the standard PCR protocol with Q5 polymerase (see Section 2.1.2.2 PCR) and primers containing LIC-specific overhangs (see Table 2.2). 2 µg of vector was incubated with 20 U of ZraI in 100 µl reactions to linearise the vector. The insert and vector were gel-extracted using QIAquick Gel Extraction Kit and then, in preparation for LIC, incubated in the presence of dATP and dTTP, respectively, using the reaction mixtures described in Tables 2.6 and 2.7.

Table 2.7 Components and their volumes used for insert preparaton for LIC.

Reagent and stock concentrations	**Volume**
Gel-purified PCR product of the insert DNA (600 ng)	X µl
10x T4 DNA pol. buffer	2 µl
dATP (25 mM)	2 µl
DTT (100 mM)	1 µl
T4 DNA pol. (LIC-qualified from Novagen)	0.4 µl
ddH_2O	X µl to 20 µl

Table 2.8 Components and their volumes used for linearised vector preparaton for LIC.

Reagent and stock concentrations	**Volume**
Linearised vector (450 ng)	X µl
10x T4 DNA pol. buffer	3 µl
dTTP (25 mM)	3 µl
DTT (100 mM)	1.5 µl
T4 DNA pol. (LIC-qualified from Novagen)	0.6 µl
ddH_2O	X µl to 30 µl

Mixes were incubated at RT for 30 min, followed by enzyme inactivation at 75^{o}C for 20 min. Linearised LIC pFL vector for insect cell expression containing an N-terminal GST tag was a kind gift from the Cook lab. 1 µl of linearised vector and 2 µl of insert were mixed and incubated at RT for 10 min, followed by addition of 1 µl of 25 mM EDTA with a further 10 min RT incubation to inactivate any residual enzymes in the mixture. 2 µl of the reaction were used to transform chemically competent *E. coli* DH5α strain cells.

2.1.2.6 Bacterial transformation and colony screening

Chemically competent *E. coli* DH5α cells were thawed on ice and 50-100 ng of plasmid DNA (~0.5-1 µl) was added to them. The reaction was incubated on ice for 20-30 min, heat-shocked at 42°C and was let to recover on ice for 2 min. 200 µl of antibiotic-free SOC medium was added to each reaction. The cells were incubated with shaking at 37°C for 1h, streaked onto an agar plate containing relevant antibiotics (ampicillin for pFL and kanamycin for pEC-KCG) and incubated overnight at 37°C.

The next day the colonies were checked, and plates left at 4°C during the day to avoid colony overgrowth. In the evening the colonies were picked and inoculated in 5 ml or LB medium containing relevant antibiotics and incubated overnight.

The following morning the test cultures were pelleted and processed according to QIAGEN Miniprep kit manufacturer's protocol. The absorbance at 260 nm of purified DNA was measured on a Nanodrop (Thermo Fisher Scientific) to determine the DNA concentration. The purified DNA was digested with relevant restriction enzymes to confirm the clones that had the expected size and intact restriction sites and separated on 0.5-1% (w/v) agarose gel.

2.1.2.7. Sanger sequencing

Plasmids were verified by Sanger sequencing using construct-specific primers. The sequencing mixes were as follows:

Table 2.9 Mixes used in Sanger sequencing reactions.

Reagent and stock concentrations	**Volume**
5x Big Dye buffer	5 µl
5x Big Dye Terminator v3.1 (Thermo Fisher)	5 µl
3.2 µM sequencing primer	1 µl
Plasmid DNA (200-500 ng)	X µl
ddH_2O	X µl to 10 µl

The mixes were run on a PCR thermocycler. The thermocycler programme was as follows:

Table 2.10 Biometra TProfessional Gradient 96 Thermocycler programme for Sanger sequencing with Big Dye terminator.

Step	**Temperature**	**Time**	**Cycles**
1	90°C	30 s	25x
2	50°C	20 s	

3	60°C	4 min	
4	60°C	1 min 15 s	
5	4°C	Hold	

Sanger sequencing was carried out on the isolated clones by Edinburgh Genomics. Sequencing chromatograms were analysed in Benchling software by comparing to the template sequence. Correct clones were used for downstream applications.

2.1.3 Insect cell work

2.1.3.1 Preparation of EMBacY competent cells

Buffers:

TFB1 (100 ml): 30 mM KAc, 100 mM RbCl, 10 mM $CaCl_2$*$2H_2O$, 50 mM $MnCl_2$*$4H_2O$, 15% (v/v) glycerol, pH 5.8

TFB2: 10 mM MOPS, 10 mM RbCl, 10 mM $CaCl_2$*$2H_2O$, 15% glycerol, pH 6.5

A stock of EMBacY was spread on a plate containing 50 µg/ml kanamycin, 10 µg/ml tetracycline and 100 µg/ml carbenicillin and incubated at 37°C overnight. A single colony was inoculated into 5 ml LB with appropriate antibiotics (same as on the plate) and grown overnight. The culture was diluted to 250 ml with LB pre-warmed to 37°C and $MgSO_4$ added to 20 mM. The cells were grown to OD_{600} ~0.6, transferred to autoclaved and chilled centrifuge tubes, left on ice for 10 min and spun down at 4,000 g for 5 min at 4°C. Pellets were resuspended by gentle pipetting in 50 ml of ice-cold TFB1 buffer and left on ice for 5 min. Cells were pelleted at 1,400 g for 10 min at 4°C and resuspended by gentle pipetting in 5 ml (per 125 ml of cells) of ice-cold TFB2 buffer and left on ice for 15 min. Cells were aliquoted in 100 µl into pre-chilled Eppendorf tubes in sterile conditions, snap-frozen in liquid nitrogen and stored at -80°C.

2.1.3.2 Bacmid preparation

Agar plates were prepared containing the antibiotics described in Table 2.11 at the concentrations indicated.

Table 2.11 Antibiotics used for bacmid preparation.

Antibiotic	Final concentration
Gentamycin	10 µg/ml
Kanamycin	50 µg/ml
Tetracyclin	10 µg/ml

Prior to plating the cells, 100 µl of 0.1 M IPTG and 100 µl of 20 mg/ml X-gal for blue-white screening were added to the plates, which were allowed to dry for 30-60 min at 37°C. For bacterial transformation, a vial of EMBacY competent cells were retrieved from -80°C and thawed on ice. 30 ng of pFL plasmid containing the construct of interest were added to the cells, gently flicked and incubated on ice for 45 min, following by a heat shock for 60 s at 42°C and recovery on ice for 3 min. 900 µl of SOC media was added and the cultures were incubated at 37°C with 250 rpm agitation for 4-5 h. Following that, the cultures were centrifuged at 13,000 rpm for 1 min, 750 µl of the supernatant was removed and the cell pellets were resuspended in the remaining supernatant. 40 µl and 200 µl of the culture were streaked onto the prepared agar plates to allow for better colony separation and incubated at 37°C overnight or longer, until the colour develops in the blue-white colony screening. Once the colour developed, white colonies, which contained the construct, were re-streaked on a fresh plate to select for true positive colonies.

To purify bacmids, a single colony was inoculated into 10 ml of LB with the relevant antibiotics (see Table 2.11) at 37°C and 250 rpm overnight. The cells were pelleted and resuspended in 250 µl of P1 resuspension buffer containing 100 µg/ml RNase A of QIAprep Spin Miniprep Kit. 250 µl P2 lysis buffer was added and thoroughly mixed by inversion, followed by addition of 350 µl N3 neutralisation buffer and mixed. The samples were centrifuged at 13,000 rpm for 10 min at 4°C. DNA-containing supernatant was then added to fresh tubes containing 800 µl 100% isopropanol and mixed. The samples were centrifuged at 13,000 rpm for 15 min at 4°C. Supernatants were discarded and the pellets washed in 500 µl 70% ethanol and centrifuged at 13,000 rpm for 5 min at 4°C. The pellets were air dried and resuspended in 50 µl of EB buffer. The sample concentrations were measured at 260 nm, with the 260/280 ratio used to assess DNA purity, diluted to 1 µg/µl, aliquoted in 50 µl for single-use to avoid freeze-thawing and stored at -20°C. Insert-specific and bacmid-specific (M13) primers were used to assess plasmid integration into the bacmid backbone. PCR reactions were visualised following a 1% agarose run and staining with SafeView (NBS Biologicals). The presence of an amplicon indicated correct bacmid. Verified bacmids were used for insect cell transfection.

2.1.3.3 Insect cell culture

Sf9 insect cells isolated from the *Spodoptera frugiperda* cell line IPLB-Sf-21-AE was used for transfection and amplification of recombinant baculovirus, as well as protein expression. Immediately after resuscitation from cryogenic storage, the cells were maintained as an adherent culture in T75 flasks (Corning with vented cap) in approximately 10 ml of Sf-900 II SFM (Thermo Fisher Scientific) at 27^{o}C without antibiotics. The day after resuscitation, the media was changed, and cells were allowed to reach 100% confluence. Following that, the cells were detached by applying horizontal forceful impact on the flask. The suspension cultures were maintained at 27^{o}C with 100 rpm agitation in conical baffled flasks with the culture –to-flask volume ratio in ml not exceeding 1:4.

2.1.3.4 Transfection of *Sf9* cells

All transfection and protein expression work in insect cells was carried out in a sterile cabinet. An aliquot (50 µl at 1 µg/µl) of a relevant bacmid was mixed with 500 µl of pre-warmed Insect cell medium Sf-900 II SFM and gently flicked. 20 µl of X-tremeGENE HP DNA transfection reagent (Sigma-Aldrich) was added to the mix and gently flicked again. The transfection mix was left for 30 min at room temperature (in the cabinet). Meanwhile, healthy (97-100% viability) *Sf9* cells were diluted to 1*10^{6} cell/ml and 2 ml was added to each well in clear 6-well plates. Following the transfection mix incubation, 110 µl of the mix was added to each well, gently mixed and incubated at 27^{o}C. Fluorescence intensity was monitored every 24 h. After 72-96 h, supernatants containing baculovirus were pelleted by centrifugation and filtered in a sterile syringe through a 0.22 µm filter into a sterile 15 ml Falcon tube. The virus was labelled V0 (viral generation 0) and stored at 4^{o}C for no more than 2 months wrapped in foil for light protection.

2.1.3.5 Baculovirus amplification

The V0 stock was used to produce subsequent viral generations of increased viral titres. V1 was generated by adding 2.5 ml of V0 to 50 ml of *Sf9* cells at a cell density 1.5*10^{6} cells/ml (1:20 v/v virus:cell culture). The cells were grown for ~72h or until the total fluorescence of the culture reached around 90%. The cells were pelleted by centrifugation and filtered in a sterile syringe through 0.22 µm filter into a sterile 50 ml Falcon tube. The virus was labelled V1 (viral generation 1) and stored at 4^{o}C for no more than 2 weeks wrapped in foil for light

protection. The pellets were used for a confirmatory pull-down assay (see Section 2.3.4) For V2 generation, 1 ml of V1 was added to 50 ml of *Sf9* at a cell density $1.5*10^6$ cells/ml (1:50 v/v virus:cell culture) and the process was repeated. Either V1 or V2 generation was used for protein expression.

2.1.3.6 Protein expression tests in *Sf9* using baculovirus

1 ml of V1 or 0.75 ml of V2 baculoviral generation was added to 50 ml of *Sf9* at a cell density $1.5*10^6$ cells/ml (1:50 or 1:75 v/v virus:cell culture, respectively). The cells were grown for ~72h or until the total fluorescence of the culture reached around 90%, after which they were pelleted and subjected to a pull-down and/or Western blot to confirm expression of the correct protein construct. The virus:cell culture ratio that showed the most viable cultures or better protein expression levels was selected for the large-scale expression.

2.1.3.7 Large scale protein expression in *Sf9*

Sf9 cells at a density $1.5*10^6$ cells/ml were infected with V1 or V2 at a ratio 1:50 or 1:75, depending of the baculovirus strength. For co-infection, two viruses were added to the same cell culture. The total *Sf9* culture volume was between 0.5 and 2L. Cells were grown for 72h with fluorescence and viability monitored every 24h. Cells were pelleted by centrifugation at 4000 rpm in large centrifuge buckets, the supernatant was discarded, and the pellets were frozen in liquid nitrogen and stored at -80^{o}C until further manipulations.

2.1.3.8 LSH^{WT} purification

Buffers:

Lysis buffer: 50 mM HEPES pH 7, 500 mM NaCl, 5% glycerol, 2 mM MgCl2, 4 mM beta-mercaptoethanol, 0.2 mM PMSF, 1x cOmplete EDTA-free protease inhibitor tablet (Sigma-Aldrich).

Ni-NTA wash buffer: 50 mM HEPES pH 7, 500 mM NaCl, 5% glycerol, 4 mM beta-mercaptoethanol, 0.2 mM PMSF, 20-40 mM imidazole (two-step wash), 1x EDTA-free protease inhibitor tablet

Ni-NTA elution buffer: 50 mM HEPES pH 7, 500 mM NaCl, 5% glycerol, 4 mM beta-mercaptoethanol, 0.2 mM PMSF, 500 mM imidazole, 1x EDTA-free protease inhibitor tablet

LSH low salt buffer: 50 mM HEPES pH 7, 150 mM NaCl, 5% glycerol, 4 mM beta-mercaptoethanol, 0.2 mM PMSF, 1x EDTA-free protease inhibitor tablet

LSH high salt buffer: 50 mM HEPES pH 7, 1 M NaCl, 5% glycerol, 4 mM beta-mercaptoethanol, 0.2 mM PMSF, 1x EDTA-free protease inhibitor tablet

LSH gel filtration buffer: 50 mM HEPES pH 7, 500 mM NaCl, 5% glycerol, 4 mM beta-mercaptoethanol, 0.2 mM PMSF, 1x EDTA-free protease inhibitor tablet

Around 5 to 10 g of LSH^{WT}-containing *Sf9* pellets were resuspended in 5 ml of lysis buffer and sonicated on ice for 30 s on/15 s off for the total of 3 min at 30-40% amplitude. Lysate was incubated with 3 µl of 250 U/µl of Benzonase nuclease (Sigma) or 50 µl of 2000 U/ml of DNAseI (from bovine pancreas, Sigma) on ice for 30-60 min. At this and each following step an aliquot was taken for further visualisation by SDS-PAGE. The insoluble fraction was pelleted by centrifugation at 50,000 g for 45 min at 4°C. The clarified lysate was loaded onto HiTrap IMAC FF 1 ml column on the ÄKTApure system, washed with Ni-NTA buffer and eluted with Ni-NTA elution buffer. Fractions were run on 12% SDS-PAGE and stained with Quick Blue Coomassie stain (Avidity Science). The protein-containing fractions were then applied onto HiPrep 26/10 desalting column and eluted with low salt buffer. The fractions containing the protein were pooled and loaded onto Resource S (GE Lifesciences) ion exchange column and eluted by salt gradient (150 mM - 1 M) with high salt buffer. The fractions containing the protein were again pooled and concentrated to 500 µl for injection onto Superdex200 10/300 24 ml gel filtration column. Purified protein was run on 12% SDS-PAGE and stained with Coomassie stain. Protein concentration calculated from measurement of absorbance at 280 nm on the Nanodrop with MW and extinction coefficient/1000 settings. Protein was concentrated to 0.8-1 mg/ml, aliquoted in 40 µl aliquots, snap-frozen in liquid nitrogen and stored at -80°C.

2.1.3.9 LSH^{K237Q} purification

LSH^{K237Q}-containing *Sf9* pellets were processed the same way as LSH^{WT} but HiPrep 26/10 desalting and Resource S ion exchange columns were omitted since the mutant was meant to be used for biochemical assays and did not require high purity.

2.1.3.10 CDCA7 purification

Buffers:

Lysis buffer: 50 mM HEPES pH 7, 500 mM NaCl, 5% (v/v) glycerol, 2 mM $MgCl_2$, 4 mM β-mercaptoethanol, 0.2 mM PMSF, 2x EDTA-free protease inhibitor tablet

GSH high salt wash buffer: 50 mM HEPES pH 7, 1 M NaCl, 5% (v/v) glycerol, 4 mM β -mercaptoethanol, 0.2 mM PMSF

Around 5 to 10 g of GST-CDCA7-containing *Sf9* pellets were resuspended in 5 ml lysis buffer and sonicated on ice for 30 s on/15 s off for the total of 3 min at 30-40% amplitude. The Lysate was incubated with 3 µl of 250 U/µl of Benzonase nuclease (Sigma) or 50 µl of 2000 U/ml of DNAseI (from bovine pancreas, Sigma) on ice for 30-60 min. At this and each following step an aliquot was taken for further visualisation by SDS-PAGE. The insoluble fraction was pelleted by centrifugation at 50,000 g for 45 min at 4°C.

Meanwhile, GSH resin was prepared. 1 ml of resin (bed volume) was used for every 1 L of cell culture at 1-2 million/ml cell density. The appropriate resin volume was transferred to an Eppendorf tube and spun at 8,000 rpm and the storage buffer was discarded. The resin was washed twice with ddH_20 and twice with CDCA7 lysis buffer for equilibration. The equilibrated resin was then added to the clarified supernatant, and the mixture was incubated for 2-3h at 4°C with rocking. Following that, the sample was micro-centrifuged for 2 min at 5,000 prm and the flow-through transferred to a fresh tube. The resin was washed twice with lysis buffer, then twice with high-salt wash buffer and then twice in lysis buffer again. 1 ml of lysis buffer was added to the resin following the washes and 10 µl of 3C-6xHis protease (Sino Biological) at 1000 U/ml was added for an overnight incubation (16-18h). The samples were checked on an 12% SDS-PAGE gel. Protein concentration was measured by BSA titration and the protein aliquoted in 50 µl aliquots, snap-frozen in liquid nitrogen and stored at -80°C.

2.2 Preparation of nucleosomes

2.2.1 Histone expression and purification

The following histones were prepared:

Table 2.12 Histones from X. laevis expressed from pET3 in E. coli.

Histone Protein	**Construct (aa)**	**MW (kDa)**	**Extinction coefficient (M^{-1} cm^{-1})**
H2A	1-129	14	4470
H2B	1-125	14	7450
H3	1-135	15	4470
H3T	28-135	12	4470

H4	1-102	11	5960

Histones were not tagged.

Acidic patch mutant histones (H2A: E65A E90A E91A; H2B: E105A) were expressed and purified by the Wilson lab (University of Edinburgh) following the standard protocol used in this study.

2.2.1.1 Preparation of inclusion bodies

Buffers:

Wash buffer 1: 50mM Tris pH 7.5, 100mM NaCl, 1mM EDTA, 1mM benzamidine, 1% (v/v) Triton X-100

Wash buffer 2: 50mM Tris pH 7.5, 100mM NaCl, 1mM EDTA, 1mM benzamidine

Unfolding buffer: 7 M guanidine-HCl, 20 mM Tris pH 7.5, 10 mM DTT

Urea buffer A (low ionic strength): 7 M urea, 100 mM NaCl, 10 mM Tris pH 8, 1 mM EDTA, 5 mM β-mercaptoethanol, 0.2 mM PMSF

Xenopus laevis histone constructs (see Table 2.12) were a kind gift from the Voigt lab (University of Edinburgh). Recombinant core histone proteins, along with a truncated version of H3 lacking 1-27 N-terminal residues (Luger et al., 1997a), were expressed in BL21 *E.coli* cells from a pET3 vector containing ampicillin resistance gene. All agar plates contained 100 µg/ml of ampicillin. BL21(DE3)pLysS was used for H2A, H2B, H3 and H3T (with the addition of 34 µg/ml of chloramphenicol), and BL21(DE3) for H4. Additionally, the histone-containing plasmids were separately transformed into XL10 Gold *E. coli* cells for preparation of plasmid stocks. DNA was purified by the Qiagen plasmid Miniprep kit and stored at -20°C for future bacterial transformations. Both transformations were carried out following a standard transformation protocol. In brief, competent cells were allowed to thaw on ice, and around 50 ng of DNA was added to them. Cells were allowed to recover on ice for 20 min, heat-shocked for 35 s at 42°C, further recovered on ice for 2 min. 200 µl of SOC media was added, and the cells were recovered at 37°C for 30 min-1h with shaking. Following this, they were plated onto agar plates containing 100 µg/ml of ampicillin and 34 µg/ml of chloramphenicol, where required, and incubated at 37°C overnight. The following day, colonies were picked from plates and inoculated into 50 ml LB media containing 100 µg/ml of ampicillin and 34 µg/ml of chloramphenicol, where required. The following day, starter cultures were amplified in 2 L LB at 37°C with shaking and allowed to reach to OD_{600} = 0.5-0.7. Protein

expression was induced with 0.2mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 3h, following the cell harvest by centrifugation at 4000 rpm for 20 min at 4°C. Cells then were resuspended in 40 ml of Wash buffer 1, centrifugation repeated, supernatant removed and the pellets snap-frozen in liquid nitrogen or 100% ethanol-dry ice mix and stored at -80°C until further manipulations.

For the protein purification stage, cells were thawed in a warm water bath, and the freeze-thaw cycle repeated. Cells were sonicated for 2 min at 40% amplitude (10 s on/off). Lysates were centrifuged at 23,000 g for 45, and pellets containing the inclusion bodies collected. The pellets were washed twice in Wash buffer 1 and then Wash buffer 2, with centrifugation steps of 10 min at 4°C and 20,000 g performed after each wash. To solubilise inclusion bodies and release the protein, unfolding buffer was added to cell pellets and left at room temperature (RT) for 1-2h with stirring. This was followed by centrifugation at RT for 20min at 23,000 g. Supernatants containing histones were dialysed into urea buffer A (low salt). Following this, precipitates were removed by 20 min centrifugation at 4°C and 23,000g.

2.2.1.2 Tandem ion exchange chromatography

Buffers:

Urea buffer A (low ionic strength): 7 M urea, 100 mM NaCl, 10 mM Tris pH 8, 1 mM EDTA, 5 mM β-mercaptoethanol, 0.2mM PMSF (same as in the previous section)

Urea buffer B (high ionic strength): 7 M urea, 1 M NaCl, 10 mM Tris pH 8, 1 mM EDTA, 5 mM β-mercaptoethanol, 0.2mM PMSF

Histones were purified using HiTrap Q (anion exchange) and SP (cation exchange) columns on an AKTA Pure system. Buffer A (low salt) was used for equilibration and wash, and a linear gradient 0-50% of buffer B (high salt) was used for histone elution. The samples were loaded onto HiTrap Q (top) and SP (bottom) columns connected together and washed extensively. HiTrap Q column collected negatively charged contaminants and was dismantled, and histones eluted from HiTrap SP column using gradient elution with buffer B. Peak fractions were analysed on 15% SDS-PAGE, and relevant fractions pooled. Histone were dialysed into ddH_2O and their concentrations were measured by Bradford assay (Thermo Fisher Scientific) using serial dilutions 1:20, 1:40 and 1:80 and confirmed on 15% SDS-PAGE. Following quantification, histones were dialysed into water containing 5 mM β-mercaptoethanol and lyophilised for long term storage at -80°C.

2.2.1.3 Histone octamer formation

Refolding buffer: 10 mM Tris pH 8, 1 mM EDTA, 2 M NaCl, 5 mM β-mercaptoethanol

Equal amounts of histones (at least 1mg each) in either unfolding or urea buffer B were mixed and dialysed into 500 ml of pre-chilled refolding buffer at 4° C for 2h, overnight and then 2h again the next day. Samples were then centrifuged at 13,000 rpm for 15 min at 4°C to remove precipitations, transferred to new tubes and centrifugation step repeated. Large precipitations indicate wrong histone ratios. Recovered histone octamers underwent gel filtration on a Superdex 200 10/30 column (24 ml). Elution was performed for 1.5 column volumes (CV). Fractions of aggregates and octamers were run on 20% SDS-PAGE for 1.5-2h for good resolution. Fractions were pooled, and octamers were measured by Bradford assay. For long-term storage, octamers were aliquoted for 1-use portions, snap-frozen in liquid nitrogen and stored at -80°C.

2.2.1.4 Primers for 601 Widom DNA sequence amplification

601 Widom DNA sequence for subsequent nucleosome reconstitution was amplified with the following primers to produce DNA fragments of different lengths.

Table 2.13 Primers used for generation of 601 Widom DNA sequence. IR800 – near-infrared fluorescent label covalently attached to the indicated primers.

Name	**Sequence (5' to 3')**
F-601-0_NCP	CTGGAGAATCCCGGTGCCGA
F-601-6_NCP	GATCCCCTGGAGAATCCCGG
F-601-12_NCP	TGATGCGATCCCCTGGAGAATCCCGG
F-601-25_NCP	GCCGCTCTACCATGATGCCGATCCCCTGGAGAATCCCGG
IR800-R-601-0_NCP	IR800-ACAGGATGTATATATGTGACACGTGCC
IR800-F-12mer-non-hairpin	IR800-GTTGCGTGCTTT
R-12mer-non-hairpin	AAAGCACGCAAC

2.2.1.5 PCR amplification of 601 sequence with various DNA **linkers**

Four DNA fragments containing the strong nucleosome positioning Widom (W) 601 sequence were amplified with/without the following DNA linkers: 0W0 – no linker, 6W0 – 6 bp linker, 12W0 – 12 bp linker, 25W0 – 25 bp linker. The reverse primer IR800-R-601-0_NCP was conjugated with IR800 fluorophore (791 nm excitation, 809 nm emission). All primers were purchased from Integrated DNA Technologies. Primers containing a fluorescent tag were

purified by HPLC. The pBlueScript-601 plasmid containing 601 Widom was used as a template. For preparation of 601 Widom DNA, homemade (prepared in the Stancheva lab, University of Edinburgh) Taq polymerase was used.

Reaction mixes were scaled up to 2, 5 or 10 ml to obtain enough DNA for nucleosome reconstitution (tens of micrograms of DNA). The reaction composition is described in table X below.

Table 2.14 Reagents and their stocks for preparation of Widom 601 DNA for nucleosome recostitution.

Reagent and stock concentrations	**Final concentration**	**Volume for 50 µl reaction (µl)**	**Volume for 5 ml reaction (µl)**	**Volume for 10 ml reaction (µl)**
$MgCl_2$ (50 mM)	6 mM	6	600	1200
dNTP (10 mM)	200 µM	1	100	200
DMSO (100%)	7%	3.5	350	700
Taq buffer 10x	1x	5	500	1000
Fwd primer (10 µM)	200 nM	1	100	200
Rev primer (10 µM)	200 nM	1	100	200
Template 10 ng/µl	200 pg/µl	1	80	160
Taq polymerase	2.5 U/ml	1	100	200
H_2O		30.5	3050	7100

Mixes were added in 100 µl aliquots into PCR tubes and PCR performed using the parameters described in Table 2.15 below.

Table 2.15 PCR programme for amplification of Widom 601 DNA

Step		**Temp (** °C)	**Time**	**Cycles**
1	Initial denaturation	98	30 s	
2	Denaturation	98	20 s	30x
3	Annealing	55°C	30 s	
4	Elongation	70°C	30 s	
5	Final elongation	70°C	3 min	
6	Pause	4°C	Hold	

After the PCR reaction, samples were pooled, and purification of the PCR products performed using the Invitrogen PureLink PCR purification kit. Samples were run on 1% agarose gel, stained with Safe View and scanned on a Geldoc scanner and Odyssey scanner (Li-Cor) at 800 nm wavelength to check both DNA content and fluorescence intensities. Absorbance of the purified fragments was measured at 260 nm on Nanodrop. Samples were stored in the PCR purification elution buffer at -20^{o}C.

2.2.1.6 Oligonucleotide annealing

Oiligos were resuspended in ddH_2O to a 1 mM stock before being checked using UV spectroscopy and adjusted to 0.5 mM. Equal volumes of complementary oligonucleotides were mixed with equal volumes of 2x TEN buffer (20 mM Tris pH 7.5, 100 mM NaCl, 2 mM EDTA). The mixture was heated to 90°C for 10 mins in a heat block, with a hot block covering the Eppendorf lids to prevent condensation. The heat block was turned off and blocks were removed with slow cooling to room temperature for ~4 h. The annealed oligonucleotides were analysed using 20% native PAGE. If the DNA was IR800 labelled the gel was scanned on an Odyssey scanner (Li-Cor).

2.2.1.7 Nucleosome reconstitution by salt gradient dialysis

Buffers:
TE buffer: 10 mM Tris pH 8, 1 mM EDTA
6% TBE gel: 5.5 ml ddH_2O, 2.5 ml 2x TBE, 1 ml 30% acrylamide, 25 µl 10% APS, 10 µl TEMED

For a small-scale reconstitution (100 nM DNA), each Widom DNA fragment of interest was mixed with an histone octamer at an appropriate molar ratio (in the range of 0.5-1 moles DNA to 1 moles of histone octamer) in refolding buffer and added to 0.1 ml Slide-a-Lyzer (Thermo Fisher Scientific) dialysis tubes (10,000 MW CO). Samples were placed in 200 ml refolding buffer and 800 ml of TE buffer was added drop-wise by a peristaltic pump at a flow-rate 1 ml/min at 4^{o}C with gentle stirring. Following that, the samples were incubated in 500 ml of fresh TE buffer for further 2 h to remove any residual NaCl.

For large-scale reconstitution (using 500 nM-1 µM DNA), the small scale reconstitution reactions of an appropriate ratio were used as a guideline, the amounts of each reagent were multiplied by 5, mixed and added to 0.5 ml Slide-a-Lyzer dialysis tubes (10,000 MW CO), with the downstream process the same as for the small scale reconstitution.

Following dialysis, samples were transferred to fresh Eppendorf tubes and spun at 13,000 rpm for 3 min at 4°C to remove any precipitates. To assess nucleosome reconstitution, 5 µl of each sample were mixed with 5 µl of 20% (v/v) glycerol and run on a native 6% TBE gel (see Table 2.3.2 for gel recipe) (ran in 0.5x TBE buffer) alongside their respective free DNA fragments at 100 V and 4°C for 1h. Gels were scanned on an Odyssey scanner (Li-Cor) immediately after the run (for the oligos containing a fluorophore) or stained with SafeView nucleic acid stain and visualised on Geldoc scanner (DNA without a fluorophore). To calculate Nucleosome yields, absorbance at 260 nm was measured on Nanodrop, gel images were processed in Image Studio Lite software (https://www.licor.com/bio/image-studio-lite/) to estimate the percentage of free DNA in nucleosome samples.

2.2.1.8 Electrophoretic mobility shift assay (EMSA)

Buffers:

Binding buffer 5x: 100 mM HEPES pH 7, 250 mM NaCl, 50% (v/v) glycerol, 5 mM DTT

LSH storage buffer: 50 mM HEPES pH 7, 500 mM NaCl, 10% (v/v) glycerol, 1 mM DTT

Running buffer 1x: 60 mM Tris, 40 mM CAPS

EMSAs were performed in 20 µl reaction volumes in binding buffer. LSH was diluted in its storage buffer to 10x final concentration and 2 µl added to each respective reaction (titration range 0-250 nM LSH), followed by addition of 18 µl nucleosomal master mix (4 µl of 5x Binding buffer, nucleosome to the final concentration of 10 nM, ddH_2O to the final volume of 20 µl). Reaction mixes were incubated at 4° C for 30 min. Samples were then loaded onto 0.5% (w/v) agarose gel made with Tris-CAPS buffer pH 9.4 and run in 1x Tris-CAPS buffer (pH 9.4) for 1h at 110 V at 4° C. The gels were visualised on an Odyssey scanner (Li-Cor).

2.3 Common biochemical techniques and assays

2.3.1 SDS-PAGE

Denaturing gels were prepared manually, unless indicated otherwise, according to the following recipe:

Table 2.16 Components of the SDS-PAGE gels for different gel percentages.

	Resolving				**Stacking**
Component	**8%**	**12%**	**15%**	**20%**	**4%**
ddH_2O	2.4 ml	1.75 ml	1.2 ml	0.4 ml	1.5 ml
Tris-HCl	1.25 ml	1.25 ml	1.25 ml	1.25 ml	625 µl
10% SDS	50 µl	50 µl	50 µl	50 µl	25 µl
30% acrylamide (29:1)	1.3 ml	2 ml	2.5 µl	3.3 ml	335 µl
10% APS	25 µl	25 µl	25 µl	25 µl	12.5 µl
TEMED	15 µl	15 µl	15 µl	15 µl	15 µl

Loading dye 4x: 200 mM Tris-HCL pH 7, 8% (w/v) SDS, 20% (v/v) Glycerol, a 0.04% (w/v) bromophenol blue, 100 mM DTT

Gels were run in the following buffer:

SDS-PAGE running buffer: 25 mM Tris, 250 mM Glycine, 0.1 % (w/v) SDS

Following the run, gels were rinsed in ddH_2O, and stained with 10-20 ml of either Quick Blue Coomassie dye (Avidity Science) (for mass spectrometry studies) or homemade Blue Silver stain (see below) (for all other applications).

Blue Silver stain (Colloidal Coomassie): 10% H_3PO_4, 10% $AmSO_4$, 20% MeOH, 0.1% G-250

Following the stain, gels were destained in ddH_2O until no background stain remained.

2.3.2 Native PAGE

Native gels were prepared manually according to the following recipe:

Table 2.17 Components and their volumes for preparation of native PAGE of different percentages.

	Gel percentage		
Component	**4.5%**	**6%**	**20%**
ddH_2O	6 ml	5.5 ml	0.8 ml
2x TBE	2.5 ml	2.5 ml	2.5 ml
30% acrylamide (29:1)	1.5 ml	2 ml	6.7 ml
10% APS	50 µl	50 µl	50 µl
TEMED	15 µl	15 µl	15 µl
Total V (ml)	10	10	10

2.3.3 Western Blot

Buffers:

Transfer buffer 1x: 25 mM Tris, 192 mM glycine

Ponceau Red dye: 0.1% (w/v) Ponceau Red, 5% (v/v) acetic acid

Blocking buffer: 5% (w/v) powdered skimmed milk + 0.1% (v/v) Tween-20

Wash buffer 1: PBS + 0.1% (v/v) Tween-20

Wash buffer 2: PBS

Protein resolved on an SDS-PAGE gel were wet-transferred onto a nitrocellulose membrane at 400 mA and 4°C for 1h, visualised with Ponceau Red dye and blocked with blocking buffer for 1h at RT. The relevant primary antibody was diluted in blocking buffer to an appropriate concentration (see Table 2.18 for antibodies) and added to the membrane for 1h at RT or overnight at 4°C. The membrane was washed for 30 min at RT with Wash buffer 1 and incubated with the fluorophore-conjugated secondary antibody (Table 2.19) in blocking buffer for at least 30 min at RT protected from light. The membrane was then washed in Wash buffer 1 for 15-30 min at RT, following with Wash buffer 2 for 15-30 min at RT. Proteins were visualised on an Odyssey scanner (Li-Cor).

Table 2.18 Primary antibodies used for Western blots.

Antibody	Type	Catalogue number	Dilution
α-N-LSH	Mouse monoclonal (recognition of 1-223 aa)	Santa Cruz, sc46665	1:1000
α-CDCA7	Rabbit polyclonal (recognition of 51-128 aa)	Abcam, ab223717	1:1000
α-GST	Rabbit polyclonal	Thermo Fisher, cab4169	1:1000

Table 2.19 Secondary antibodies used for Western blots.

Antibody	Type	Catalogue number	Dilution
IRDye 680	Donkey anti-Rabbit	LI-COR, 926-68023	1:10000
IRDye 800	Donkey anti-Mouse	LI-COR, 926-32212	1:10000

2.3.4 Protein pull-down from crude insect cell lysates

Buffers:
Lysis buffer: 50 mM HEPES pH 7, 150 mM NaCl, 5% (v/v) glycerol, 2 mM $MgCl_2$, 4 mM β-mercaptoethanol, 0.2 mM PMSF, 1x EDTA-free protease inhibitor tablet

Ni-NTA wash buffer (for single protein): 50 mM HEPES pH 7, 500 mM NaCl, 5% glycerol, 4 mM β-mercaptoethanol, 0.2 mM PMSF, 10 mM imidazole, 1x EDTA-free protease inhibitor tablet

GSH wash buffer (for single protein): 50 mM HEPES pH 7, 500 mM NaCl, 5% glycerol, 4 mM β-mercaptoethanol, 0.2 mM PMSF, 1x EDTA-free protease inhibitor tablet

Wash buffer (for protein-protein interaction): 50 mM HEPES pH 7, 150 mM NaCl, 5% glycerol, 4 mM β -mercaptoethanol, 0.2 mM PMSF, 1x EDTA-free protease inhibitor tablet

Pellets of 50 ml V1 or V2 cultures (see Section 2.1.3.5 for v1 v2 production) were resuspended in 2 ml lysis buffer and lysed by mild sonication (Soniprep 150, MSE) on ice at 25% amplitude for 2 min at 10 s on/off. The concentrations of total lysates were measured by Bradford assay, and 100-500 µg of total cell lysate of a single protein was used. To check protein-protein interactions, lysates were mixed and incubated for 1h at 4°C. Meanwhile, 15 µl (bed volume) per sample of the relevant resin was washed in ddH_2O twice and equilibrated in lysis buffer. Each lysate or the lysate mix were added to the resin and incubated with rotation for 1-2h. Following that, the samples were spun in a cooled microfuge at 7,000 rpm. The Flow-through was removed and the samples were washed 3 items with wash buffer. Following that, the samples were run on SDS-PAGE and stained with Coomassie or subjected to Western blot.

2.4 Biophysical techniques

2.4.1 MALDI-ToF and MALDI-FT-ICR mass spectrometry

Protein samples previously resolved on SDS-PAGE gels were processed for mass spectrometry (protein confirmation or identification) as follows:

Buffers:
Buffer 1: 200 mM ammonium bicarbonate (ABC), 50% acetonitrile (ACN)
Buffer 2: 200 mM ABC, 50% ACN, 20 mM DTT
Buffer 3: 200 mM ABC, 50% ACN, 50 mM iodoacetamide
Buffer 4: 20 mM ABC, 50% ACN
Buffer 5: 50 mM ABC, 13 µg/ml trypsin

Proteins were run on an SDS-PAGE gel, stained with Quick Blue Coomassie dye (Avidity Science) and thoroughly destained in ddH_2O. The bands of interest were cut out with a clean scalpel, cut into a few smaller pieces and placed into an Eppendorf tube. The samples were incubated twice in 300 µl Buffer 1 for 30 min at RT, then in Buffer 2 for 1h at RT, then in Buffer 3 for 20 min at RT and protected from light. The samples were washed 3x with 500 µl in Buffer 4, spun at 13,000 for 2 min and dehydrated with ACN for 5 min. ACN was then removed and gel pieces allowed to dry. The samples were then rehydrated in 150 µl Buffer 5, first on ice, then at 32°C overnight, to digest the proteins along with a trypsin blank sample. The next day the samples were sonicated for 5 min in the water bath sonicator.

2.4.2 Matrix preparation for MALDI-ToF

The Matrix mixture can be prepared up to 2 weeks prior to use and stored at 4°C. 10 mg of α-Cyano-4-hydroxycinnamic acid (CHCA) was mixed with 500 µl ddH_2O to saturation, 500 µl acetonitrile and 10 µl 0.01% trifluoroacetic acid (TFA). 1 µl of the sample prepared as in Section 2.4.1 was spotted on a stainless steel 100 sample plate (Applied Biosystems) in triplicates, then an equal volume of matrix was added to it and allowed to air dry. Proteins were confirmed or identified by MALDI-ToF on Bruker UltraflexExtreme instrument using peptide calibration standard (Covered mass range ~1000-3200 Da, Bruker 8206195). The Mascot server (http://www.matrixscience.com/search_form_select.html) was used to identify proteins and the ProteinProspector tool (http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msfitstandard) was used to verify known proteins.

2.4.3 Size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS)

SEC-MALS was carried out using the ÄKTApure system with a Superdex200 10/300 24 ml column coupled with a UV detector, Viscotek SEC-MALS-20 scattering detector and Viscotek VE-3580 refractive index detector. The column was equilibrated with SEC running buffer (50 mM HEPES pH 7, 100 mM NaCl, 5% glycerol, 4 mM β-mercaptoethanol, 0.2 mM PMSF, 1x EDTA-free protease inhibitor tablet) overnight (at 0.5 ml/min flow rate). 100 µl BSA at 1 mg/ml was injected and the monomer peak corresponding to 66 kDa was used to calibrate the mass. Samples were injected in 100 µl volumes at concentrations varying from 0.5 mg/ml

to 5 mg/ml. Analysis of the mass was performed using the OmniSEC software package (Malvern Pananalytical).

2.5 Protein sequence alignment and homology modelling

Sequence alignments were performed using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and visualised using JalView (http://www.jalview.org/). Protein sequences were aligned in full or as separate domains, as indicated in the tables.

Sequences used in homology modelling sections:

Protein	**Species**	**UniProt/PDB ID**
LSH	*M. musculus*	Q60848
LSH	*H. sapiens*	Q9NRZ9
LSH	*B. taurus*	E1BCV0
LSH	*D. rerio*	B7ZD98
LSH	*X. laevis*	Q4V835
CDCA7	*M. musculus*	Q9D0M2
CDCA7L	*M. musculus*	Q922M5
Chd1	*S. cerevisiae*	P32657
Ino80	*S. cerevisiae*	P53115
Ino80	*C. thermophilum*	6FML_G
Snf2	*S. cerevisiae*	P22082
Iswi	*S. cerevisiae*	P38144
Snf2h	*H. sapiens*	O60264
Mot1	*C. thermophilum*	6G7E_B
Mtiswi	*T. thermophilus*	5JXR_B
DDM1	*A. thaliana*	Q9XFH4
Irc5	*S. cerevisiae*	P43610
Mus30	*N. crassa*	Q7SAC4

Homology modelling and one-to-one threading was performed using Phyre2 server and model quality was assessed using Phyre Investigator (Kelley et al., 2015). *Ab initio* modelling was performed using I-TASSER (Yang et al., 2015) and Phyre2 in intensive modelling mode. The PDB files were manipulated using PyMol Molecular Graphics System software (DeLano, 2002).

2.6 Structural methods

2.6.1 Gradient fixation of nucleosomes and LSH-nucleosome complexes (GraFix)

To improve sample quality for potential electron microscopy studies, GraFix (gradient fixation) method was used (Kastner et al., 2008a; Stark, 2010). Briefly, solid sucrose was dissolved in binding buffer containing 20 mM HEPES pH 7 and 50 mM NaCl to concentrations of 10% (w/v) (light solution) and 30% (w/v) (heavy solution). 0.2% (v/v) EM-grade glutaraldehyde was added to the heavy solution (for unfixed controls this step was omitted). Approximately 1 ml of the light solution was added to Beckman ultracentrifugation 2.4ml tubes, followed by careful injection of the heavy solution to the bottom of the tube with a blunt-end needle. Continuous gradients were then formed by BioComp gradient master, model 107 (programme - sucrose 10-30% (w/v)), followed by 1h incubation of the gradients at 4^{o}C for temperature equilibration.

Following that, ~100 µl of a sample were very carefully added on top of the gradient, avoiding bubble formation. The samples were centrifuged in an Optima Ultracentrifuge at 35,000 rpm in a TSS-55 rotor for 20 h at 4^{o}C. Samples were manually fractionated immediately and 100-150 µl fractions were collected. Samples were run on 4.5% TBE native PAGE (see Section 2.3.2) at 100V for 1h at 4^{o}C and visualised using Licor scanner owing to the fluorescent IR800 labelling of nucleosomes. To verify successful cross-linking, the samples were additionally run on 12% SDS-PAGE and visualised on Licor scanner or by colloidal Coomassie staining if only unlabelled protein was used.

2.6.2 LSH-NCP chemical crosslinking

Chemical crosslinking was carried out using either BS^3 (Thermo Fisher Scientific) or EDC (Thermo Fisher Scientific) in the presence of sulfo-NHS (Thermo Fisher Scientific). BS^3, EDC and sulfo-NHS were dissolved in 20 mM HEPES pH 7 to 10 µg/µl stock concentration and then diluted to the required experimental concentrations (Tables 2.20, 2.21). EDC and sulfo-NHS were mixed at the ratio 1:2 (w/w). LSH stock in gel filtration buffer (50 mM HEPES pH 7, 500 mM NaCl, 5% glycerol, 4 mM β-mercaptoethanol, 0.2 mM PMSF, 1x EDTA-free protease inhibitor tablet) at 77.1 µM (7.4 mg/ml) was used at the final concentration of 2.5 µM. Recombinant nucleosome core particle (NCP) stock (final dialysis stem into 50 mM HEPES pH 7 buffer) at 1.7 µM was used at the final concentration of 0.5 µM. Both components were

mixed to their final concentrations with the 10x reaction buffer (200 mM HEPES pH7, 750 mM NaCl) with ddH_2O added to 20 µl making the final NaCl concentration 100 mM and incubated for 30 min on ice. Following that, the samples were incubated with the relevant crosslinker for 2h at 4°C. Reactions were quenched by addition of 1 µl of 1 M Tris pH 7, and the samples were run on a precast SDS-PAGE 4-20% gradient Mini PROTEAN TGX gel (BioRad) and stained with Quick Stain Coomassie.

Table 2.20 Components and their volulmes used in the BS^3 crosslinking test experiment.

	LSH + BS^3				**LSH + NCP + BS^3**			
BS^3 (µg)	0	0	10	20	30	10	20	30
BS^3 (µl)	0	0	1	2	3	1	2	3
LSH (µl)	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
NCP (µl)	-	-	-	-	6	6	6	6
Buffer 10x (µl)	2	2	2	2	2	2	2	2
ddH_2O	17.4	11.4	10.4	9.4	8.4	16.4	15.4	14.4
Final V (µl)	20	20	20	20	20	20	20	20

Table 2.21 Components and their volulmes used in the EDC+sulfo-NHS crosslinking test experiment.

	LSH + EDC+sulfo-NHS				**LSH + NCP + EDC+sulfo-NHS**			
EDC (µg)	0	0	10	20	30	10	20	30
EDC (µl)	0	0	1	2	3	1	2	3
Sulfo-NHS (µg)	0	0	20	40	60	20	40	60
Sulfo-NHS (µl)	0	0	2	4	6	2	4	6
LSH (µl)	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
NCP (µl)	-	-	-	-	6	6	6	6
Buffer 10x (µl)	2	2	2	2	2	2	2	2
ddH_2O	17.4	11.4	10.4	9.4	8.4	16.4	15.4	14.4
Final V (µl)	20	20	20	20	20	20	20	20

3. HOMOLOGY MODELLING OF LSH

3.1 Introduction

Since very limited and low-resolution structural information is only available for LSH (see Simon Varzandeh's thesis, University of Edinburgh), I decided to utilise homology modelling to gain any further insights into the mode of action of LSH in chromatin remodelling and its potential disruption by ICF-associated mutations.

To investigate structural aspects of murine LSH (UniProt Q60848) and to obtain any insights into its complex with the nucleosome homology modelling approach was used. Homology modelling is a useful approach for getting an estimation of a protein structure, when no structural information is available, but a protein has a number of close homologues with known structures. Being a part of SNF2 subfamily, LSH has several close homologues covering the highly conserved ATPase domain (***see Section 1.2***), therefore, it was a good candidate for homology modelling, that here was performed using Phyre2 server.

3.2 LSH homology modelling

A member of the SNF2 family LSH has minimal structural information available, therefore, it is useful to conduct a bioinformatics and homology modelling investigation that may provide further insight regarding the protein's structure and function. The initial step is examination of its secondary structure. The server PsiPred (Jones, 1999) provides the secondary structure prediction based on the protein sequence. It is well established that certain residues have higher propensity to participate in the formation of secondary structure: Methionine, alanine, leucine, glutamate, and lysine ("MALEK") are more likely to be found in α-helices, since their steric and electrostatic properties favour formation of a peptide spiral with stabilising hydrogen bonds. Larger amino acids, such as tryptophan and tyrosine, have disrupting effect on α-helices and are more likely to be found in β-sheets.

Secondary structure prediction by PsiPred based on LSH sequence indicates that a large part of the protein consists of α-helices with intermittent β-sheets (***Fig. 3.1***). This pattern is relevant to the two RecA domain lobes and is compatible with the observed structure of other RecA-contaning proteins, where a β-sheet core of each lobe is surrounded by a bundle of α-helices (Ye et al. 2004). The residues crucial for ATPase activity – the arginine fingers R690 and R693 and lysine K237 that coordinate ATP are all located in the helical regions. This information is further examined below in the homology modelling section.

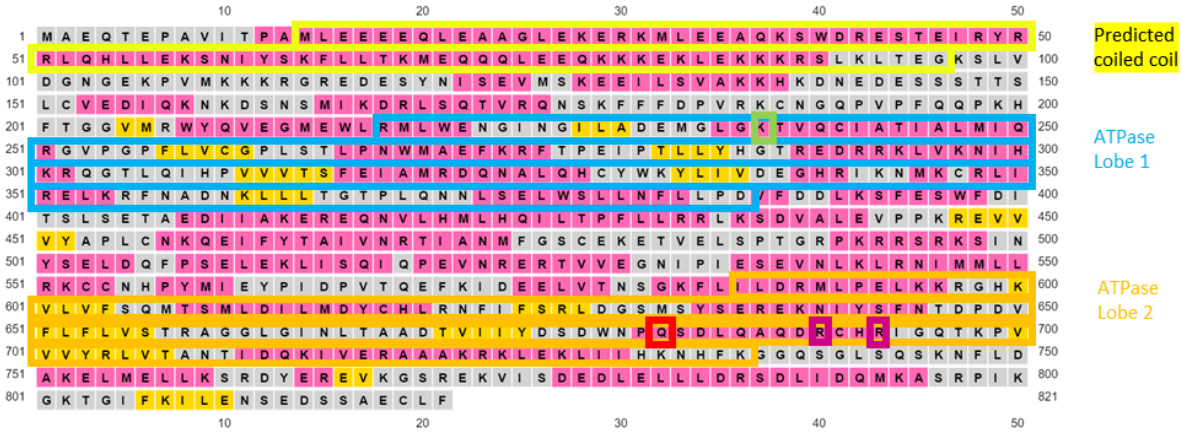


Figure 3.1 Secondary structure prediction of LSH by PsiPred (https://www.ncbi.nlm.nih.gov/pubmed/10493868). Numbers on each side are amino acid positions in the protein. Pink – α-helix, yellow – β-sheet, grey – random coil. LSH domains are labelled as indicated, K237 is in green, Q682 (murine LSH) is in red, arginine fingers R690 and R693 are in purple.

The N-terminal part of LSH (14-94 aa) was predicted by the server COILS to harbour a coiled coil (CC) (***Fig. 3.2***), which is feasible from the secondary structure prediction that indicated abundance of α-helices in that region. When submitted to the server MARCOIL (Delorenzi and Speed, 2002), the N-terminal region of LSH was shown to harbour a heptad repeat pattern (***Fig. 3.3***), reinforcing the notion of a coiled coil located in that region.

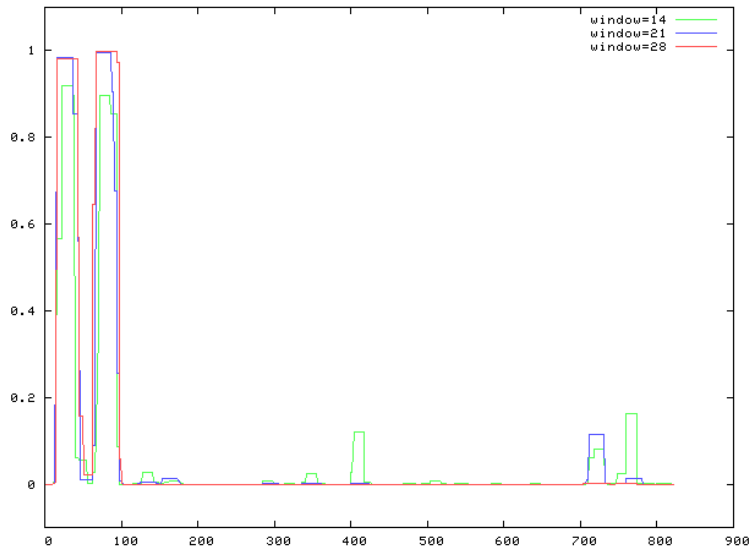


Figure 3.2 Prediction of coiled coil patterns in the sequence of murine LSH by the server COILS. X axis – LSH residues, Y axis – probability for a coiled coil. Windows represent the number of amino acids scanned at a time.

1M[00.3]g	2A[00.6]a	3E[00.9]b	4Q[01.2]e	5T[01.3]f	6E[01.4]g	7P[01.5]a
8A[01.8]b	9V[02.2]c	10I[02.6]d	11T[03.2]e	12P[03.8]f	13A[10.4]g	14M[18.7]a
15L[28.6]g	16E[45.4]a	17E[68.8]b	18E[79.3]c	19E[82.6]d	20Q[86.4]e	21L[87.9]f
22E[91.3]g	23A[92.3]a	24A[93.1]b	25G[93.7]c	26L[95.2]d	27E[95.6]e	28K[95.7]f
29E[95.8]g	30R[95.8]a	31K[95.7]b	32M[95.7]c	33L[95.7]d	34E[95.6]e	35E[95.2]f
36A[94.2]g	37Q[93.4]a	38K[92.7]b	39S[91.3]c	40W[90.0]d	41D[90.5]e	42R[90.8]f
43E[91.3]g	44S[90.0]a	45T[89.8]b	46E[90.1]c	47I[89.1]d	48R[88.2]e	49Y[86.3]f
50R[86.5]g	51R[86.5]a	52L[86.4]b	53Q[86.4]c	54H[86.2]d	55L[86.2]e	56L[86.4]f
57E[86.4]g	58K[86.3]a	59S[86.1]b	60N[85.9]c	61I[85.8]d	62Y[85.6]e	63S[85.8]f
64K[86.5]g	65F[86.7]a	66L[89.3]b	67L[92.4]c	68T[93.8]d	69K[95.9]e	70M[97.1]f
71E[98.2]g	72Q[98.5]a	73Q[98.7]b	74Q[98.8]c	75L[98.8]d	76E[99.0]e	77E[99.1]f
78Q[99.1]g	79K[98.9]a	80K[98.6]b	81K[98.0]c	82E[97.1]d	83K[96.5]e	84L[95.3]a
85E[94.7]b	86K[92.4]c	87K[89.5]d	88K[85.2]e	89R[78.4]f	90S[74.8]g	91L[72.1]a
92K[66.3]b	93L[54.8]c	94T[47.9]d	95E[43.2]e	96G[31.2]d	97K[27.0]e	98S[18.1]f

Figure 3.3 N-terminal region of murine LSH with a heptad repeat pattern as indicated by MARCOIL server. Core positions a and d are in red, surface positions are in blue. Percentage probability to form coiled coil are in brackets.

Canonical coiled coils contain a heptad repeat structure, labelled as *a-g* for each of the seven positions of the heptad. The hydrophobic core of the heptad are in positions *a* and *d* and the rest of the positions are predominantly solvent-exposed hydrophilic residues (Ludwiczak et al., 2019). In the LSH sequence, position *a* contains a mixture of charged and polar uncharged residues, but position d has a few isoleucines and a tryptophan, adding to its hydrophobic qualities (***Fig. 3.3***). Positions *b, e* and *g* are dominated by hydrophilic residues, which is in line with their position on the surface of the coiled coil (***Fig. 3.3***).

Most of the LSH sequences is highly conserved. This predicted coiled coil region is conserved in LSH across species, however, it is unique to LSH and is not present in its homologues or in other member of the chromatin remodelling family (***Fig. 3.4 and 3.5***). The C-terminal region is also conserved and unique to LSH, indicating that the regions flanking the ATPase domain possess specific function related to the LSH translocase activity but dissimilar the functions of the other chromatin remodellers. The residues required for binding and hydrolysing ATP – the arginine fingers and lysine K237 – are all conserved across the species (***Fig. 3.4***). The glutamine Q682 of murine LSH mutated to arginine in ICF syndrome (corresponds to Q699 in human LSH) is not present in yeast Snf2 and Ino80 (***Fig. 3.5***), where it is replaced with histidine and tyrosine, respectively.

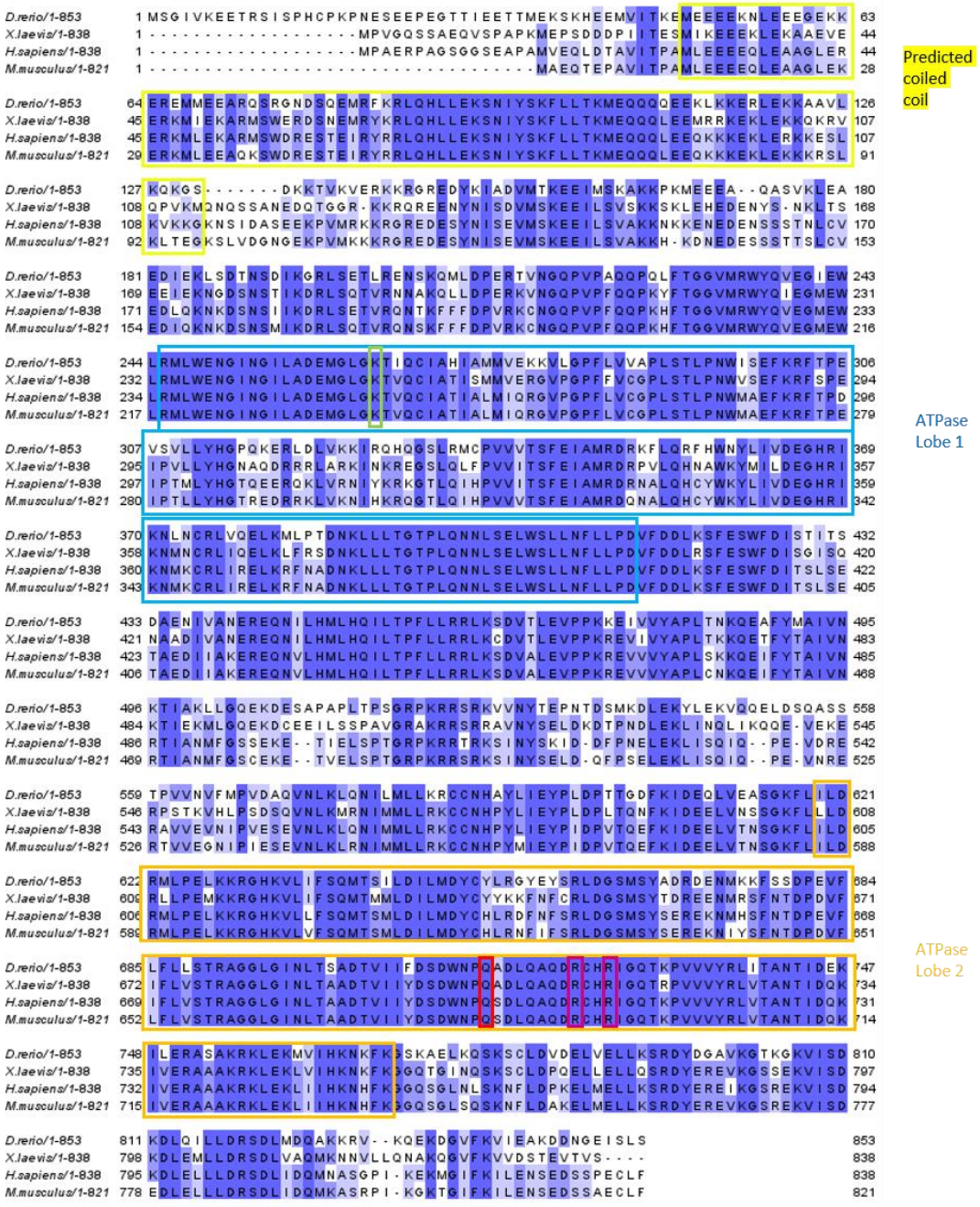


Figure 3.4 LSH homologues across species. Multiple sequence alignment was performed by Clustal Omega and visualised in JalView. LSH domains are labelled as indicated, K237 is in green, basic patch is in brown, Q682 (murine LSH) is in red, arginine fingers R690 and R693 are in purple.

To investigate structural aspects of murine LSH (UniProt Q60848) and to obtain any insights into its complex with the nucleosome, a homology modelling approach was used, utilising Phyre2 server for model generation. Being a part of SNF2 subfamily, LSH has several close homologues, and their alignment covered the highly conserved ATPase domain (***Fig. 3.5 and see Section 1.2 of Introduction***). High percentage identity of the overall sequences, as well as local alignment of each ATPase lobe (***Table 3.1***), allowed reliably to consider the remodellers from different subfamilies as modelling templates.

No homology modelling was available for the terminal regions of LSH. Previously, *ab initio* modelling of the N-terminus LSH^{1-176} was performed by Simon Varzandeh (Simon Varzandeh, 2016), however, two structures generated by I-TASSER and Phyre were dissimilar and were shown as a helical bundle or an elongated helical structure, respectively. In this study no *ab initio* modelling was performed. The basic patch (K855, R880 and K885) present in Snf2 that plays a role in the interaction with SHL-6 site on the nucleosome by acting as an electrostatic spring to direct the direction of translocation was not conserved in the other remodellers (***Fig. 3.5***). In LSH it is replaced by arginine, glutamine and histidine (R293, Q323, H328), therefore, there may still be a basic patch in LSH, despite the presence of uncharged glutamine (***Fig. 3.10***). In Chd1 it is unlikely to be present, with aspartate, arginine and serine in the respective positions. Ino80 and ISWI are still likely to have a partial basic patch, with lysine, alanine and lysine in Ino80 and asparagine, lysine and lysine in ISWI on the relevant positions (***Fig. 3.5***). Therefore, this basic patch may still be relevant for LSH functions.

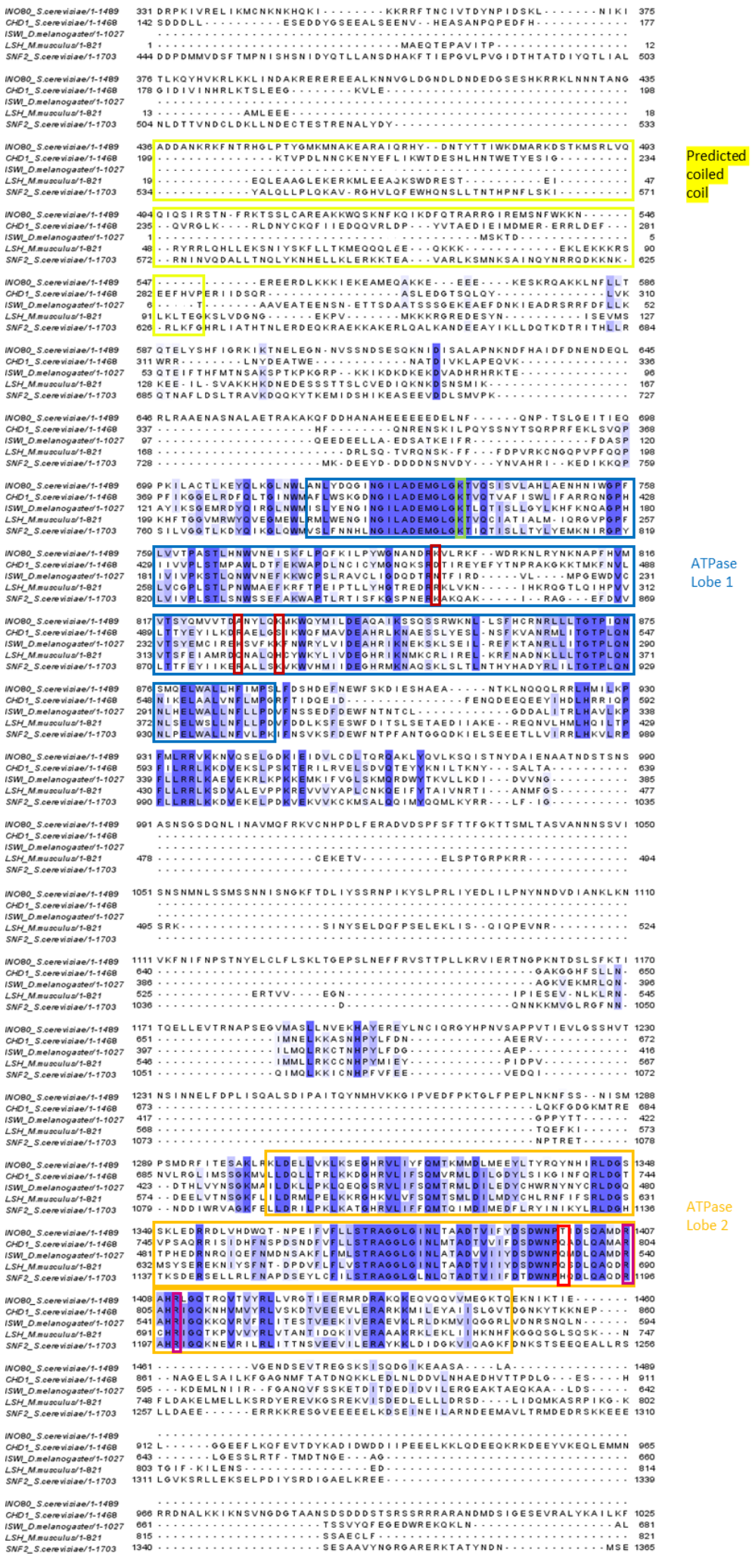

INO80_S.cerevisiae/1-1489
CHD1_S.cerevisiae/1-1468
ISWI_D.melanogaster/1-1027
LSH_M.musculus/1-821
SNF2_S.cerevisiae/1-1703
Predicted coiled coil
ATPase Lobe 1
ATPase Lobe 2

Figure 3.5 Multiple sequence alignment of LSH and its homologues yeast Chd1, yeast Ino80, fly ISWI and yeast Snf2 from the Sf2 remodellers family. Alignment was performed with Clustal Omega and visualised in JalView. LSH domains are labelled as indicated, K237 is in green, basic patch (R293, Q323, H328) is in brown, Q682 (murine LSH) is in red, arginine fingers R690 and R693 are in purple.

Table 3.1 Chromatin remodellers used as the templates for homology modelling with their sequence identity percentages generated by BLAST Protein. Total query cover indicates the cover of LSH, % identity indicates the percentage of identical amino acids in the aligned sequences.

Protein	**Species**	**Total query cover**	**Total % identity**	**ATPase Lobe 1 query cover**	**ATPase Lobe 1 % identity**	**ATPase Lobe 2 query cover**	**ATPase Lobe 2 % identity**
Chd1	*S. cerevisiae*	73%	47.6%	98%	42.5%	87%	56.2%
Ino80	*C. thermophilum*	80%	34.7%	98%	46.2%	87%	55.6%
Snf2	*S. cerevisiae*	69%	42%	98%	46.4%	90%	49%
Iswi	*S. cerevisiae*	61%	42.5%	98%	50.6%	83%	54.7%

Chd1 and LSH have 47.6% total identity and 42.5% and 56.2% identity for lobe 1 and 2, respectively (***Table 3.1***), which is a high enough similarity between the two proteins for homology modelling. Pairwise sequence alignment demonstrated extensive regions of identity in the ATPase region (***Fig. 3.6***). Interestingly, certain regions of the Chd1 fragment aligned to the LSH N-terminal coiled coil were identical (Fig. below). I then submitted this region of Chd1 to the COILS server, however, no coiled coils were predicted (not shown).

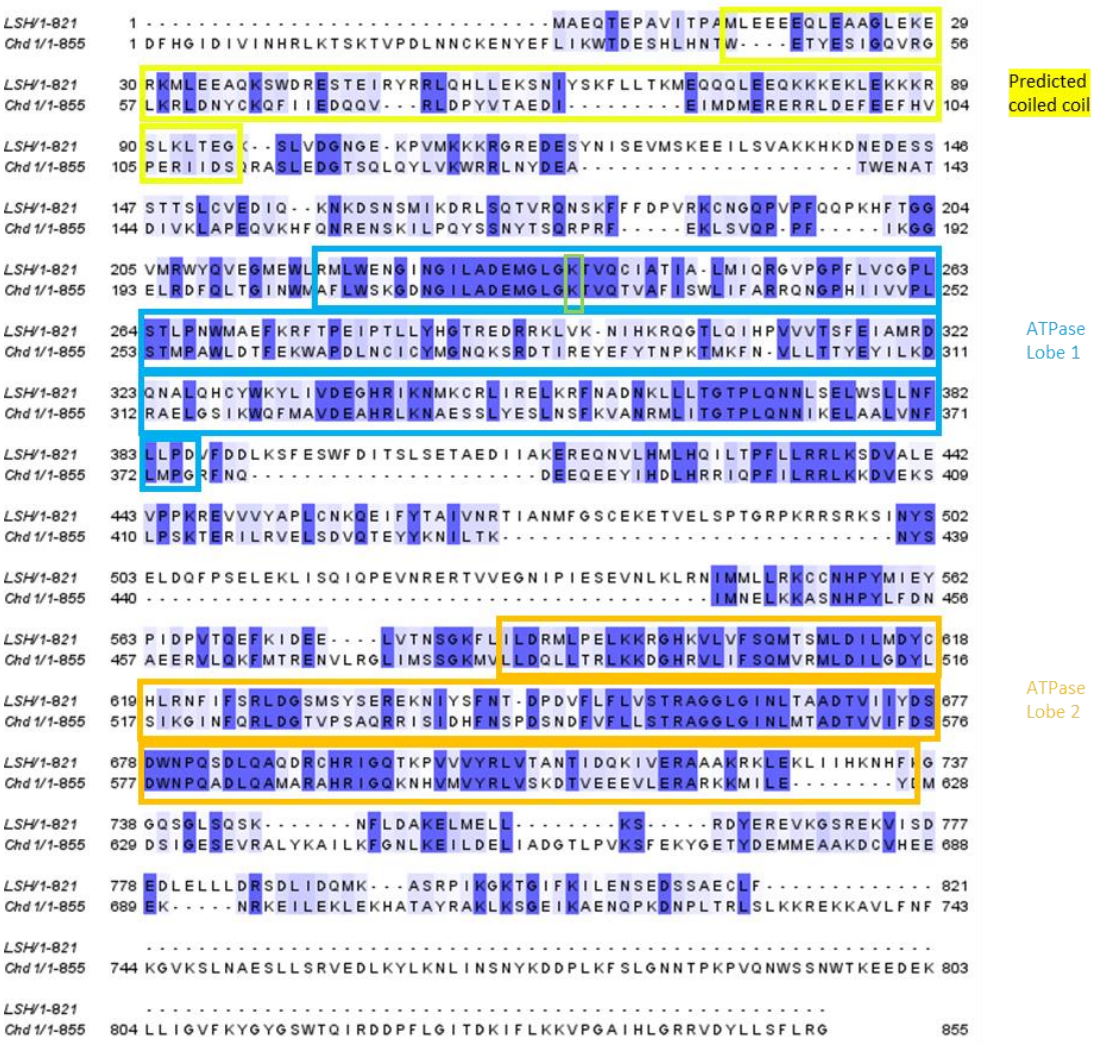


Figure 3.6 Pairwise sequence alignment of murine LSH and yeast Chd1. N-terminal coiled coil predicted by COILS server in yellow; LSH domains as indicated by ProSite: ATPase lobe 1 in blue, ATP-binding lysine K237 in green, ATPase lobe 2 in orange. Alignment was performed by EMBOSS Needle and visualised in JalView.

The LSH model based on yeast Chd1 bound to the nucleosome and ADP-BeF and poised for catalysis is based on a cryo-EM structure with 4.8 Å resolution (***Table 3.2***), which is on the lower end of the resolution spectrum but sufficient for a rough estimation of the LSH model (***Fig. 3.7***). The model generated by Phyre2 has 100% confidence and 45% identity (***Table 3.2***), which indicates a good quality model.

The superimposition of LSH on Chd1 demonstrated LSH bound to the nucleosome core particle (NCP) at SHL2 region, about 20 bp away from the dyad, which is a common binding site for chromatin remodellers (***Fig. 3.7 A and B***). The NCP used in the Chd1 structural study is an NCP containing 146 bp core with a single DNA linker of 14 bp on the entry side of the nucleosome.

In the template structure, Chd1 binds at this location on the tracking stand to move DNA in the 3' to 5' direction (Farnung et al., 2017). The cryo-EM structure of Chd1 captures the detachment of DNA from the nucleosomal surface and rotation of the linker by about 60^{o} from the nucleosomal trajectory about the dyad axis (Farnung et al., 2017). LSH binds the NCP core and since no ATPase-flanking domains were modelled, it is impossible to evaluate the location of the LSH termini. There is the histone H4 tail in proximity to the ATPase lobes. In Chd1 H4 N-terminus acts as an allosteric activator (Hauk et al., 2010) that interacts with the region outside of the chromodomains, presumably in the acidic pocket located in the lobe 2 through modifications of H4K16 and H4K20 (Farnung et al., 2017). Mutations in the chromodomains can affect the interaction between the N-terminus of H4 and nucleosome interface. Chd1 triple mutants KAK and AAA at positions E265, D266, E268 (an acidic region EDIE265-268) had a much higher affinity for a 16 bp dsDNA compared to the wild type in the absence of the H4 tail (Hauk et al., 2010).

This region was not modelled for LSH (the model only covers the region corresponding to the residues 192-731 of LSH and 366-1011 of Chd1). LSH sequence inspection revealed a region EKLE98-101 in LSH that has a similar residue context to EDIE265-268 in Chd1. This sequence in LSH, however, contains a positively charged residue K99 and cannot be characterised as an acidic pocket. Besides that, this region is surrounded by lysine triplets on each side, making the region more likely to have an overall positive charge (***Fig. 3.5***).

Checking the region proximal to the H4 N-terminus in the LSH model does not indicate the presence of any alternative acidic pockets (***Fig. 3.7***). The loop YMIEY at LSH$^{558-562}$ corresponds to FMTREN in Chd1 and the α-helical portion at LSH$^{609-617}$ (SMLDILMDYC) is similar to Chd1$^{725-730}$ (RMLDILGDYL). Even though these regions are similar, it is not clear what role they play in the NCP interaction and remodelling. These regions do not seem to have acidic pockets based on the residue sequence that would be expected to interact with a basic patch KRHR on H4 N-terminus. However, the electrostatic model of LSH (***Fig. 3.7 C***) demonstrates this region as acidic (shown in red). Therefore, it is not clear if H4 N-terminus plays the same role in LSH-NCP interaction as with the other chromatin remodellers.

When submitted to Phyre Investigator and assessed by ProQ2 server, the quality of the β-sheet core of the ATPase lobes as well as the helical regions proximal to them were good, however, no further quantitative information was given in Phyre Investigator regarding the model quality (***Fig. 3.7 E***). When compared with PsiPred secondary structure prediction, the

modelled β-sheet from Phyre2 were in accordance with PsiPred data with a few exceptions: residues ILA and TLLY in lobe 1 and FSRL and FLFLVS in lobe 2 were modelled as loops (***Fig. 3.7 D***).

The Ramachandran analysis from Phyre Investigator (not shown) demonstrated that most residues in the model lied in favourable and good regions, including those containing important functional residues (***Fig. 3.7 F***).

The Q682 residue in murine LSH relevant to ICF syndrome was shown to locate on the loop of a good-to-average quality (yellow), as indicated by ProQ2 (***Fig. 3.7 E and F***). When mutated to arginine, this residue has several disruptive possibilities: it may disrupt the interaction between lobe 2 and the loop in lobe 1 connecting two α-helices, may disrupt the location of the proximal α-helix in lobe 1 or may interfere with DNA gyre sliding (***Fig. 3.7 D and 3.14***). This lobe 1 loop also a good-to-average quality (yellow) (***Fig. 3.7 E***), which may indicate that is has a secondary structure other than a loop, without invalidating its location. Therefore, the ICF mutation may prevent the lobes from closing, thus compromising ATP binding are/or hydrolysis. The proximal α-helix in lobe 1 has low modelling quality according to the ProQ2 result.

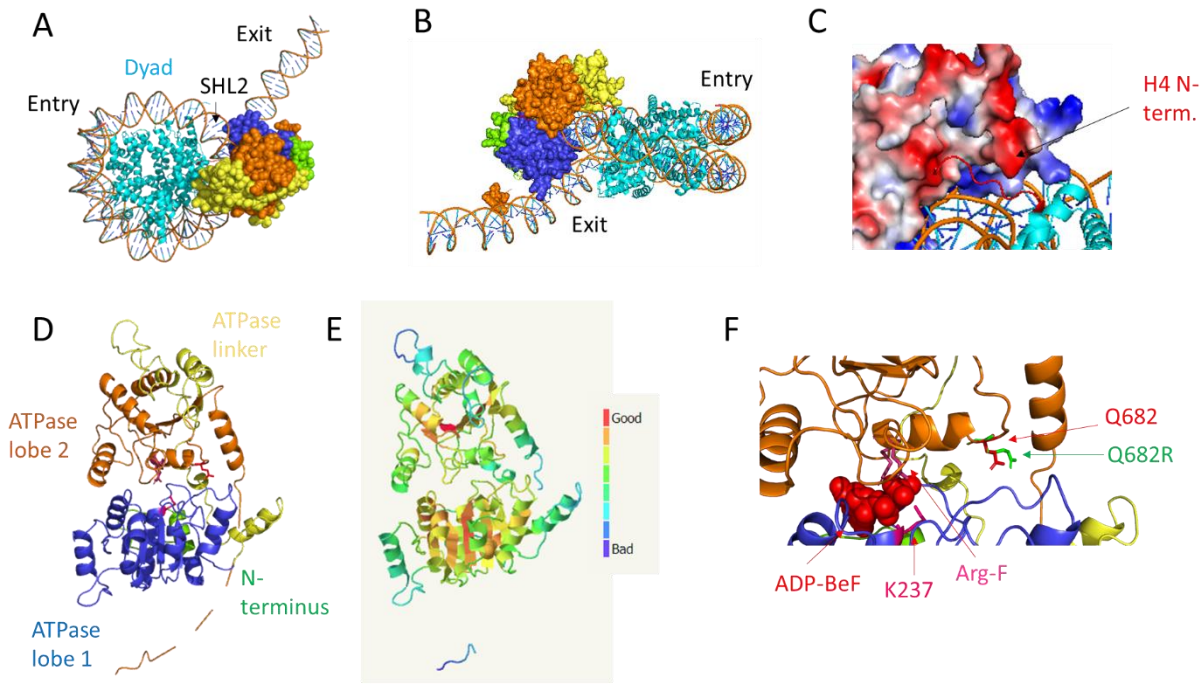


Figure 3.7 Phyre2 homology modelling of murine LSH ATPase domain based on yeast Chd1 template (PDB c5o9Gw). A) LSH ATPase (represented as spheres) bound to the nucleosome (represented as cartoon in cyan) at SHL2 site – top view; B) LSH ATPase bound to the nucleosome at SHL2 site – side view; C) Electrostatic surface representation of LSH model with histone H4 N-terminus located proximal to LSH ATPase lobe; basic residues- blue, acidic residues - red; D) Colour coding of the LSH model represented as cartoon; E) ProQ2 model of the LSH ATPase model quality assessment generated in Phyre Investigator with the colour coding scale; F) close up of the LSH catalytic region.

Table 3.2 LSH model information based on Chd1.

Protein	Template RDB ID & publication	Resolution (Å) and method	Confidence %	Id%	Alignment coverage % (LSH aa)
Chd1	5o9g, chain W (Farnung et al., 2017)	4.8, cryo-EM	100	45	65 (192-733)

Ino80 remodeller poised for catalysis in complex with the NCP was the next template for LSH modelling. Ino80 and LSH have high sequence identity overall (over 30%) and in the ATPase region (over 40% in lobe 1 and over 50% in lobe 2) (***Table 3.2 and Fig. 3.8***). Unlike LSH, Ino80 has a long linker between the two lobes, however, its function is predominantly histone exchange, which LSH is unlikely to perform, since Ino80 is a part of a large complex, and LSH not being a part of a stable complex, may not produce enough force for such disruptive actions, therefore, this structural variation is expected.

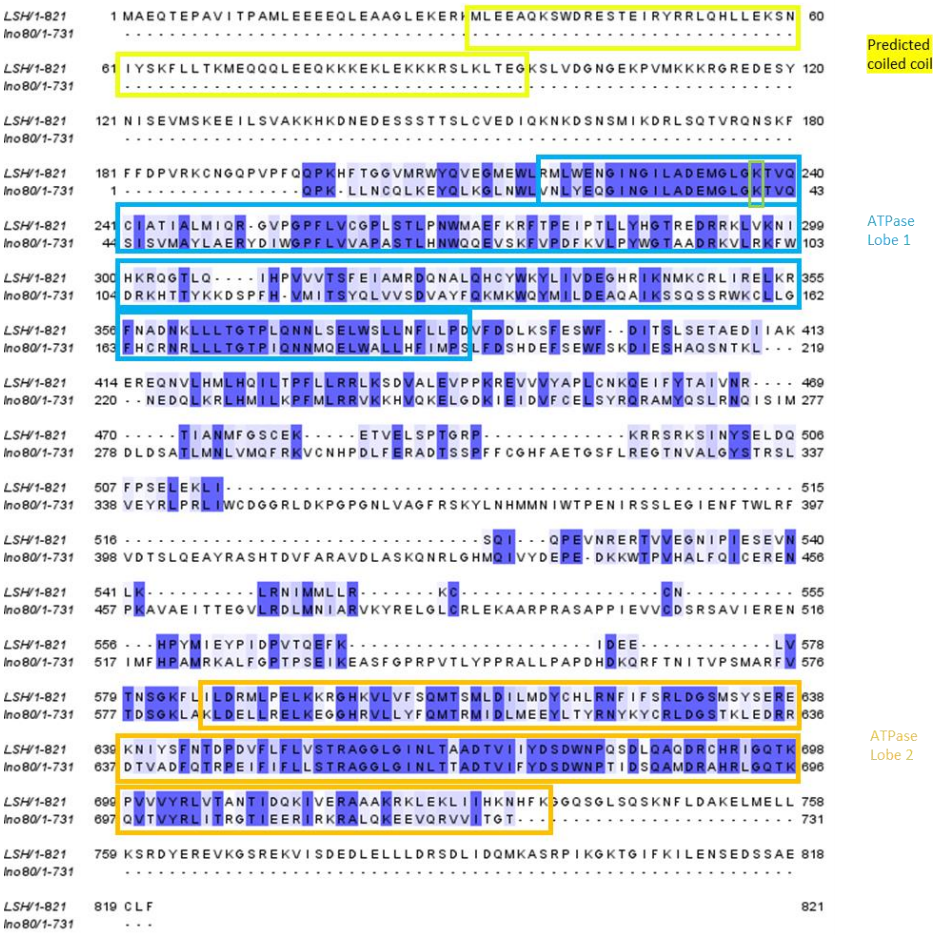


Figure 3.8 Pairwise sequence alignment of murine LSH and Ino80 from C. thermophilum. N-terminal coiled coil predicted by COILS server in yellow; LSH domains as indicated by ProSite: ATPase lobe 1 in blue, ATP-binding lysine K237 in green, ATPase lobe 2 in orange. Alignment was performed by EMBOSS Needle and visualised in JalView.

In the model, based on a cryo-EM structure (***Table 3.3***) LSH is shown bound at SHL-6, with the ATPase lobe clamping DNA gyre (***Fig. 3.9 A and B***). The NCP superimposed from the Ino80 structure demonstrates unwrapping of the DNA linker away from the nucleosomal core by 15 bp, illustrating disruptive effect of Ino80 on the NCP (Eustermann et al., 2018).

The quality of the model was assessed in Phyre Investigator. ProQ2 analysis agreed with the model quality based on Chd1: the core β-sheets in both lobes had the highest confidence with surrounding α-helices demonstrating good-to-average quality (***Fig. 3.9 D***). β-sheet location corresponded with PsiPred prediction (***Fig. 3.1 and Fig 3.9 C***). Ramachandran analysis was mostly favourable, with only a few surface loop regions indicated as disallowed (not shown).

The examination of the catalytic cleft of LSH demonstrated that Q682 faces the cleft and is located in a helical region, which contradicts the location in the Chd1-based model (***Fig. 3.8 F and Fig. 3.9 E***). However, the disruptive role of the Q682R mutation is still relevant in this model. The loop in lobe 1 is in proximity to this residue, and the mutation may prevent efficient closure of the lobes. Examination of the helix sequence supports this hypothesis: LSH has an alanine stretch and two positively charged residues (K722, R723), which may cause electrostatic disruption when glutamine is mutated to arginine. However, it should be noted that the helix harbouring those residues was not modelled with high quality (Fig. below D), therefore, its location and secondary structure should be taken with caution. On the other hand, it may also disrupt the lobe 2 itself, which, in turn, prevents the translocase closure. Similarly, to the Chd1-based model, the linker connecting two ATPase lobes had low model quality due to low level of similarity of this region between the two proteins.

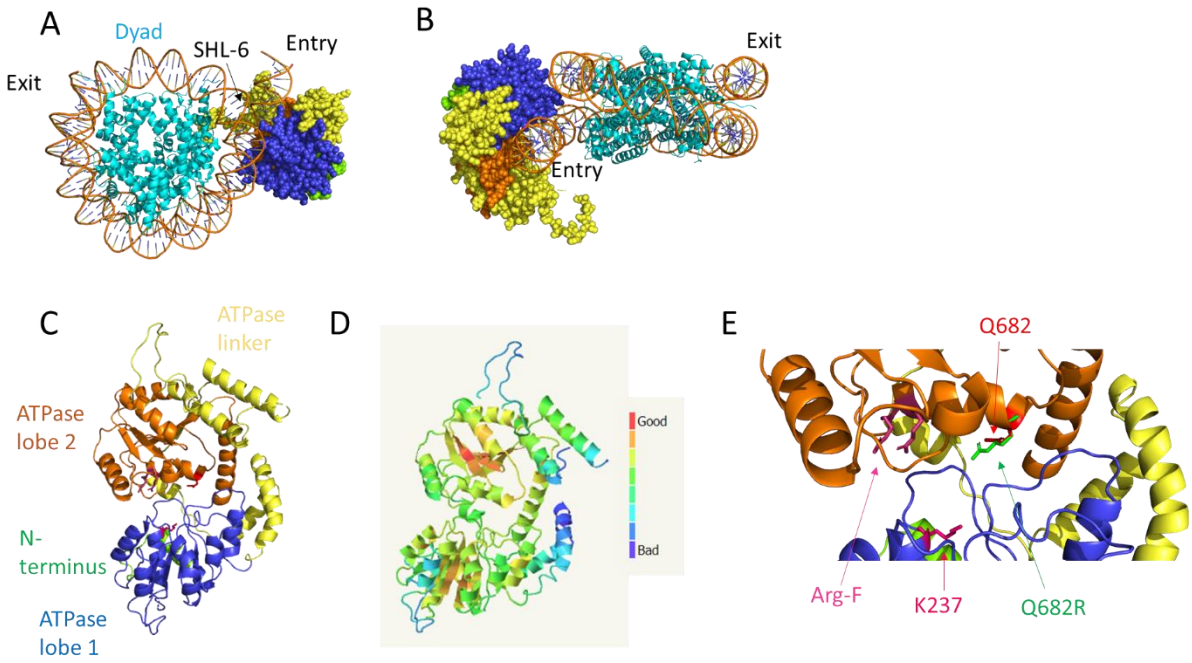


Figure 3.9 Phyre2 homology modelling of murine LSH ATPase domain based on Ino80 from C. thermophilum template (PDB c6fml_G). A) LSH ATPase (represented as spheres) bound to the nucleosome (represented as cartoon in cyan) at SHL2 site – top view; B) LSH ATPase bound to the nucleosome at SHL2 site – side view; C) Colour coding of the LSH model represented as cartoon; D) ProQ2 model of the LSH ATPase model quality assessment generated in Phyre Investigator with the colour coding scale; E) close up of the LSH catalytic region.

Table 3.3 LSH model information based on Ino80.

Protein	Template RDB ID & publication	Resolution (Å) and method	Confidence %	Id%	Alignment coverage % (LSH aa)
Ino80	6fml, chain G (Eustermann et al., 2018)	4.3, cryo-EM	100	49	64 (199-732)

Yeast Snf2 bound to the NCP (***Table 3.4***) was another template used for LSH homology modelling due to high sequence identity: 42% overall, 46.4% in lobe 1 and 49% lobe 2 (***Table 3.3***). The C-terminal region of LSH bears some similarity to the Snf2 region flanking lobe 2. Pairwise alignment of the C-terminal regions of LSH and Snf2 (Fig. below) in BLAST Protein reveal identity of 28%, which was too low to consider homology. This region of Snf2 is located on the surface of the remodeller (Ye et al. 2019) away from the NCP, therefore it is not clear what higher similarity could implicate for the function of the relevant C-terminal region of LSH.

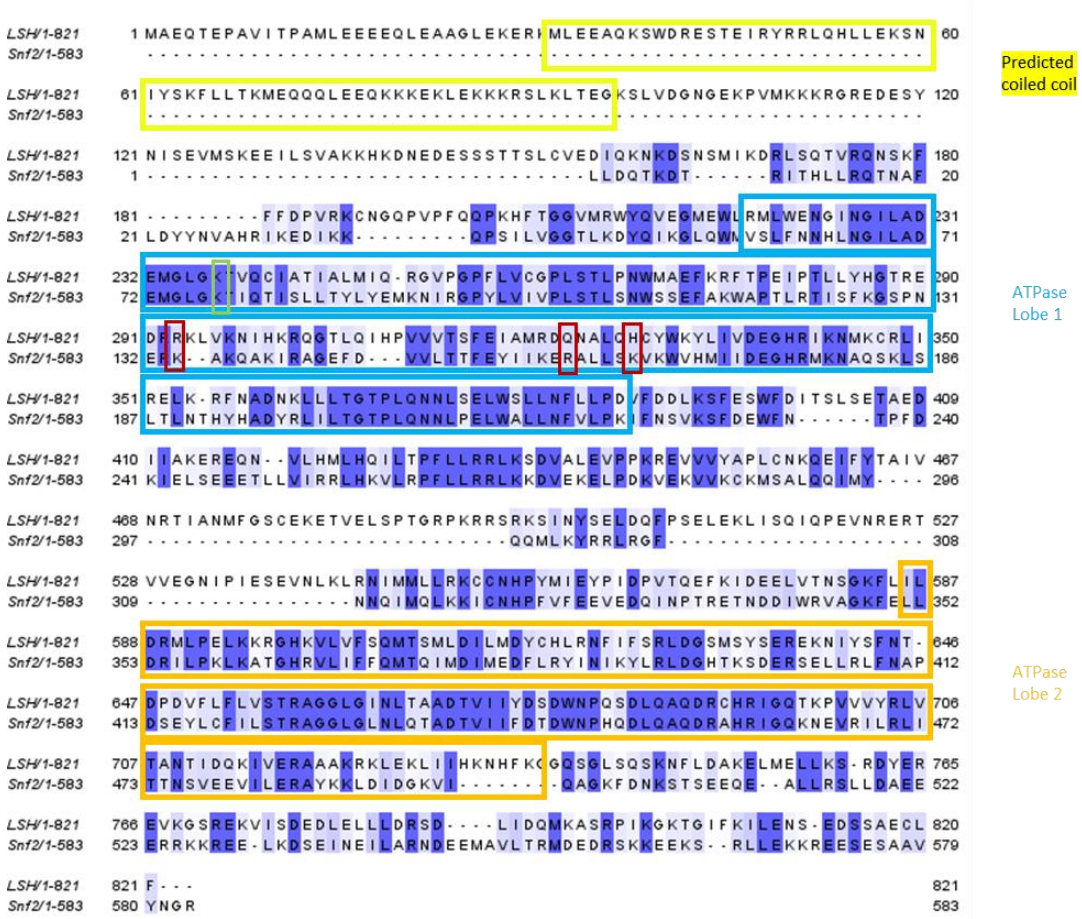


Figure 3.10 Pairwise sequence alignment of murine LSH and yeast Snf2. N-terminal coiled coil predicted by COILS server in yellow; LSH domains as indicated by ProSite: ATPase lobe 1 in blue, ATP-binding lysine K237 in green, Snf2 basic patch (R293, Q323, H328) is in brown, ATPase lobe 2 in orange. Alignment was performed by EMBOSS Needle and visualised in JalView.

The initial examination of the model based on the Snf2 structure in active state with the two core domains realigned showed 100% and 41% identity, as indicated by Phyre2 (***Table 3.4***). As in the case of the Chd1-based model of LSH (***Fig. 3.7 C***), the N-terminus of the histone H4 is protruding towards the ATPase domain, however, its role in LSH interaction is unclear due to the absence of any acidic patches in that region of LSH.

ProQ2 assessment, as in the models above, demonstrated high quality of the β-sheet cores of each ATPase lobe, and the α-helices surrounding them had lower but nevertheless relatively high quality, making this part of the model reliable for analysis (***Fig. 3.11 D and E***). As in the models above, the helices on the surface of the ATPase domain are of low quality, as indicated by ProQ2.

Ramachandran analysis as a part of Phyre Investigator (not shown) revealed that the majority of the residues were located in favourable regions, including those harbouring the Q682R mutation site. As in the previous models, ICF-relevant glutamine is located at the border of the helix facing the NCP-binding cleft (***Fig. 3.11 F***). The proximal lobe 2 helices had relatively low model quality, therefore, it is hard to determine whether Q682R mutation would disrupt the integrity of lobe 2.

The basic patch that plays a role of an electrostatic spring in Snf2 and induces twist defects (Brandani et al., 2018) is located proximal to SHL-6 (not shown). In LSH these residues are R293, Q323, H328 (***Fig. 3.10***), and not all being positively charged they may not play the same role in LSH. Structural information would be required to test that.

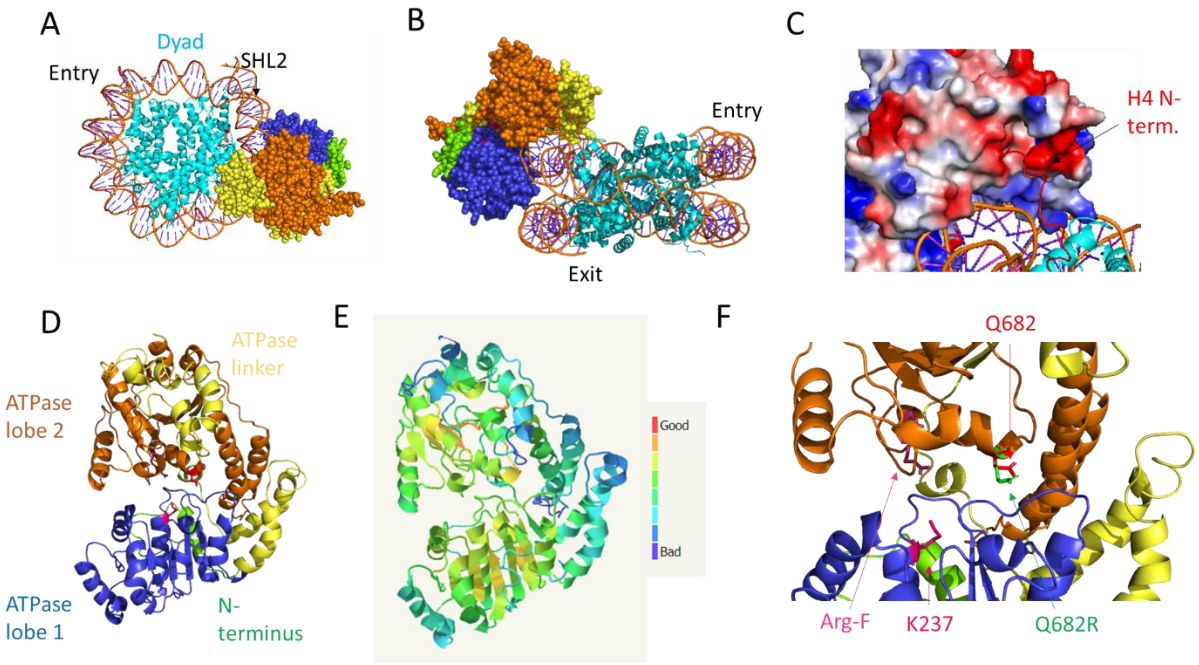


Figure 3.11 Phyre2 homology modelling of murine LSH ATPase domain based on yeast Snf2 template (PDB c5x0y_O). A) LSH ATPase (represented as spheres) bound to the nucleosome (represented as cartoon in cyan) at SHL2 site – top view; B) LSH ATPase bound to the nucleosome at SHL2 site – side view; C) Electrostatic surface representation of LSH model with histone H4 N-terminus located proximal to LSH ATPase lobes; basic residues- blue, acidic residues - red; D) Colour coding of the LSH model represented as cartoon; E) ProQ2 model of the LSH ATPase model quality assessment generated in Phyre Investigator with the colour coding scale; F) close up of the LSH catalytic region.

Table 3.4 LSH model information based on Snf2.

Protein	Template RDB ID & publication	Resolution (Å) and method	Confidence %	Id%	Alignment coverage % (LSH aa)
Snf2	5x0y, chain O (Liu et al., 2017a)	4.7, cryo-EM	100	41	74 (188-796)

Finally, the yeast ISWI cryo-EM structure was used as a template due to high overall sequence homology (when two full sequences were aligned) and ATPase-specific homology (***Table 3.4 and Fig. 3.12***). Similarly to Chd1, a region of ISWI upstream lobe 1 showed some similarity to the coiled coil predicted in the LSH N-terminus (***Fig. 3.12***), however, COILS server did not predict any coiled coils in that region of ISWI (not shown).

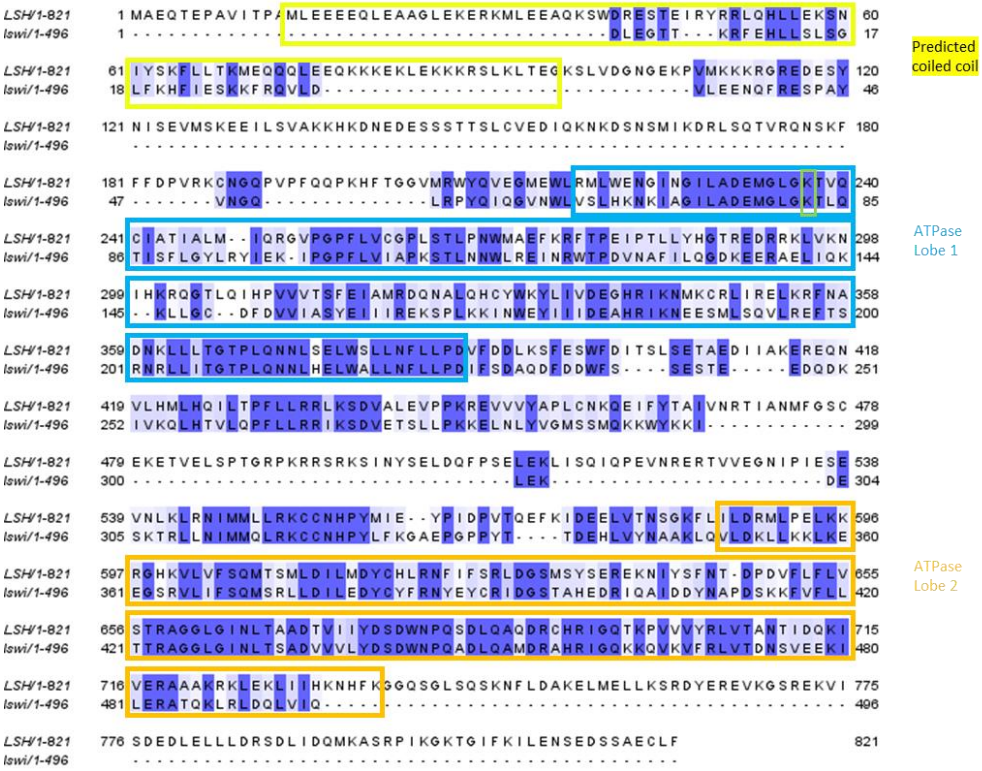


Figure 3.12 Pairwise sequence alignment of murine LSH and yeast ISWI. N-terminal coiled coil predicted by COILS server in yellow; LSH domains as indicated by ProSite: ATPase lobe 1 in blue, ATP-binding lysine K237 in green, ATPase lobe 2 in orange. Alignment was performed by EMBOSS Needle and visualised in JalView.

ISWI-based model of LSH showed 100% confidence with 49% identity. The template is ISWI in activated, nucleosome-bound and ADP-BeF-bound state. ProQ2 assessment demonstrated good quality of the β-sheets at the core of ATPase lobes and slightly lower quality of the helices surrounding them, as it was in the models described above (***Fig. 3.13 E***). The Ramachandran analysis performed by Phyre Investigator indicated predominantly favourable locations of the residues in the model (not shown). The location of Q682 was in agreement with the other LSH models (based on Chd1, Ino80 and Snf2) and was shown to face the cleft between the lobes (***Fig. 3.13 F***). As in the models described above, Q682R mutation did not

appear to directly disrupt ATP binding, as it is located away from the arginine fingers and the ATP-binding lysine K237, however, it has the potential to interfere with ATP hydrolysis by preventing efficient closure of the lobes.

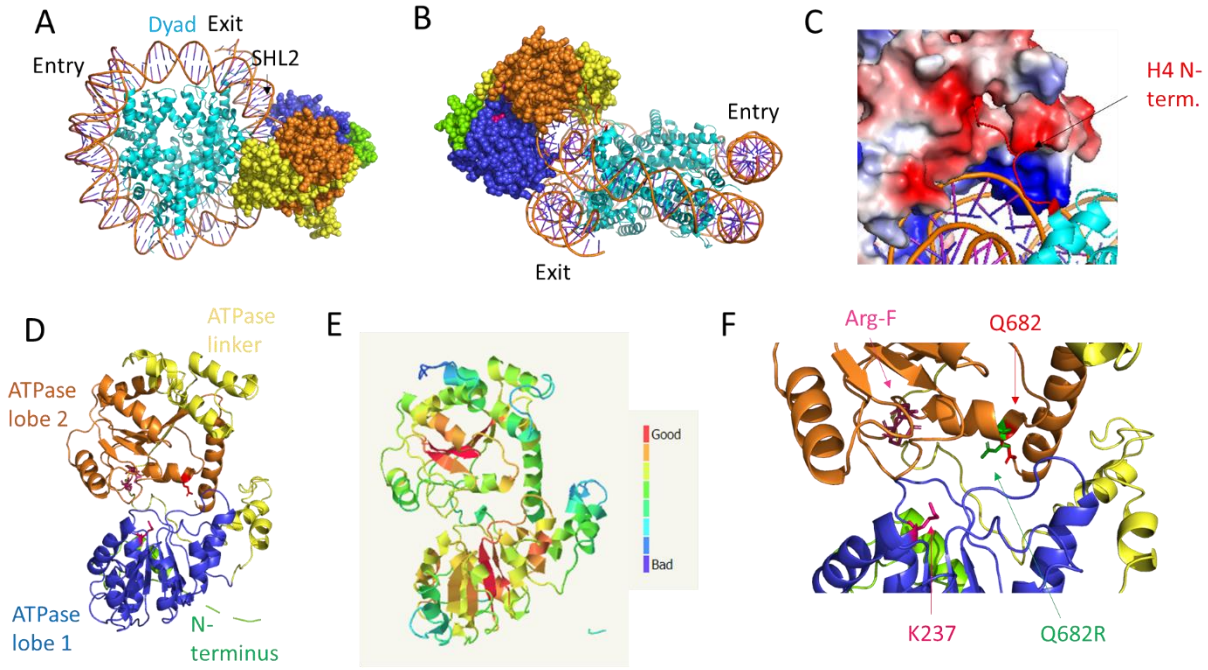


Figure 3.13 Phyre2 homology modelling of murine LSH ATPase domain based on yeast ISWI template (PDB c6iro_L). A) LSH ATPase (represented as spheres) bound to the nucleosome (represented as cartoon in cyan) at SHL2 site – top view; B) LSH ATPase bound to the nucleosome at SHL2 site – side view; C) Electrostatic surface representation of LSH model with histone H4 N-terminus located proximal to LSH ATPase lobes; basic residues- blue, acidic residues - red; D) Colour coding of the LSH model represented as cartoon; E) ProQ2 model of the LSH ATPase model quality assessment generated in Phyre Investigator with the colour coding scale; F) close up of the LSH catalytic region.

Table 3.5 LSH model information based on ISWI.

Protein	Template RDB ID & publication	Resolution (Å) and method	Confidence %	Id%	Alignment coverage % (LSH aa)
ISWI	6iro, chain L (Yan et al., 2019a)	3.4, cryo-EM	100	49	65 (192-731)

The ICF-related mutation located in lobe 2 may interfere with DNA strand electrostatic properties. To investigate that, I checked the position of Q682R mutation relative to the proximal DNA gyre (***Fig. 3.14***). Even though the models did not allow to estimate the position of the residue rotamers, it is possible to hypothesise regarding the disruptive potential of this mutation located in proximity to DNA. The Chd1-based model demonstrated proximal positioning of the arginine mutant with respect to DNA (***Fig. 3.14 A***). In the Ino80-based LSH model Q682 is located in proximity to the DNA strand, but does not seem to be in direct

contact with it (***Fig. 3.14 B***). The Snf2- and ISWI-based models appear to support the hypothesis of the disruptive effect Q682R mutation on DNA interaction (***Fig. 3.14 C and D***), even though rotamer positioning of a real structure may disprove the electrostatic effect of the mutation on LSH function.

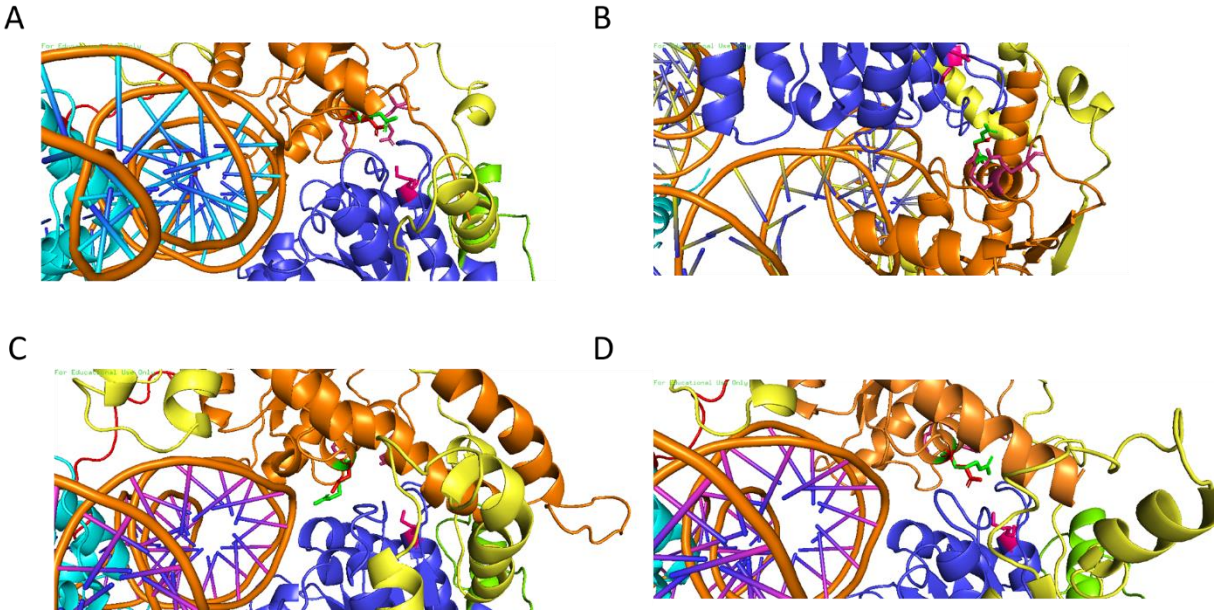


Figure 3.14 LSH models showing positioning of the Q682R mutation with respect to nucleosome DNA. A) Chd1-based model of LSH (in cartoon) colour-coded as follows: N-terminus – green, ATPase lobe 1 – blue, ATPase lobe 2 – orange, ATPase linker – yellow; residues shown as sticks: K237 - magenta, Q682 – red, Q682R – green; nucleosomal DNA strands shown as cartoon; Chd1 PDB ID 5o9g; B) Ino80-based model of LSH (in cartoon) colour-coded as in A; residues shown as sticks as in A; nucleosomal DNA strands shown as cartoon; Ino80 PDB ID 6fml; C) Snf2-based model of LSH (in cartoon) colour-coded as in A; residues shown as sticks as in A; nucleosomal DNA strands shown as cartoon; Snf2 PDB ID 5x0y; D) ISWI-based model of LSH (in cartoon) colour-coded as in A; residues shown as sticks as in A; nucleosomal DNA strands shown as cartoon; ISWI PDB ID 6iro.

Recently solved structures of Snf2 bound to the nucleosome bound at SHL2 site in complex with NCP and in different transition states (has not been used by Phyre2 as a template, therefore I used them as templates for one-to-one threading. The cryo-EM structures have a medium resolution of 4.31 Å, 3.62 Å and 4.22 Å for the complex in apo, ADP-bound and ADP-BeF-bound states, respectively (Li et al., 2019). The models show the similarities between the apo- and ADP-bound states, where they adopt an open conformation (***Fig. 3.15 A, B and D, E***) and the ATP analogue ADP-BeF_3 locks the lobes in open conformation (***Fig. 3.15 C and F***). The models demonstrate the change in the location of Q682 residue relative to the helix in lobe 2 and the loop in lobe 1. Closure of the ATPase brings ICF-affected glutamine in close

contact with those regions of the ATPase. Therefore, it is feasible to assume that Q682R mutation (corresponding to Q699R in human) affects *in vitro* ATP hydrolysis but not ATP binding, since no disruption appears to occur in the region directly contacting ATP molecule – the arginine fingers and lysine K237 (***Fig. 3.15 D, E, F***).

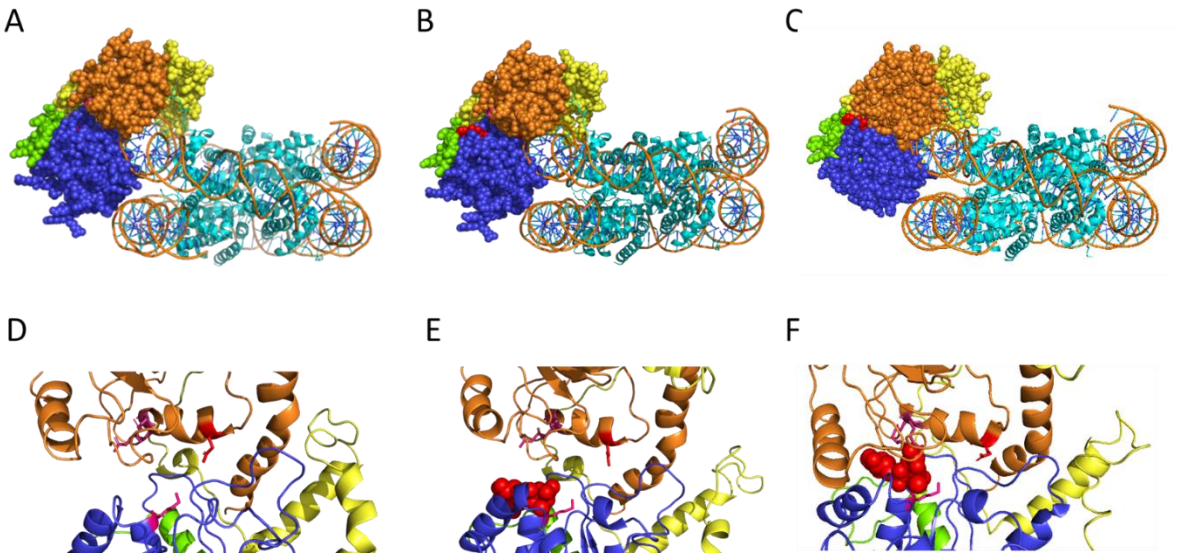


Figure 3.15 LSH homology modelling with yeast Snf2 bound to a nucleosome (Li et al. 2019) as a template. A) LSH model in complex with a nucleosome bound at SHL2 in apo state (template PDB ID 5z3l), LSH domains are colour-coded as follows: N-terminus – green, ATPase lobe 1 – blue, ATPase lobe 2 – orange, ATPase linker – yellow; residues shown as sticks: K237 - magenta, Q682 – red, Q682R – green; B) LSH model in complex with a nucleosome bound at SHL2 in ADP-bound state (template PDB ID 5z3o), LSH domains are colour-coded as in A; C) LSH model in complex with a nucleosome bound at SHL2 in ADP-BeF-bound state (template PDB ID 5z3v), LSH domains are colour-coded as in A; D) Close-up of the LSH nucleosome binding cleft (as a cartoon) as a part of a model of LSH bound to a nucleosome in the apo state: as sticks - arginine fingers R690 and R693 in purple, lysine 237 in hot pink, glutamine Q682 mutated in ICF syndrome in red; E) Close-up of the LSH nucleosome binding cleft (as a cartoon) as a part of a model of LSH bound to a nucleosome in the ADP-bound state, catalytic residues are colour-coded as in D; F) Close-up of the LSH nucleosome binding cleft (as a cartoon) as a part of a model of LSH bound to a nucleosome in the ADP-BeF-bound state, catalytic residues are colour-coded as in D.

LSH has a few poorly characterised homologues, including yeast Irc5, bread mould Mus30 and a relatively well studied remodeller from *A. thaliana* DDM1 (***Table 3.6 and Fig. 3.16***). LSH and its homologues have relatively high percentage identity of the full sequences (30-40%) and even higher identity when only the ATPase regions were compared (50-60%) (***Table 3.6 and Fig. 3.16***). Since DDM1 is a relatively well studied homologue, compared to the other two, to further verify modelling quality, I looked at the models of DDM1 to compare them with the ones of LSH. Like with the LSH models, all the models of DDM1 produced by Phyre2 had 100% confidence and above 40% identity, making them high quality on initial examination.

Table 3.6 LSH homologues with their sequence identity percentages generated by BLAST Protein. Total query cover indicates the cover of LSH, % identity indicates the percentage of identical amino acids in the aligned sequences.

Protein	**Species**	**Total query cover**	**Total % identity**	**ATPase Lobe 1 query cover**	**ATPase Lobe 1 % identity**	**ATPase Lobe 2 query cover**	**Lobe 2 % identity**
DDM1	*A. thaliana*	97%	36.8%	98%	58.3%	90%	55%
Irc5	*S. cerevisiae*	70%	44.5%	98%	59.8%	90%	53%
Mus30	*N. crassa*	76%	48.9%	95%	53.8%	100%	57.8%

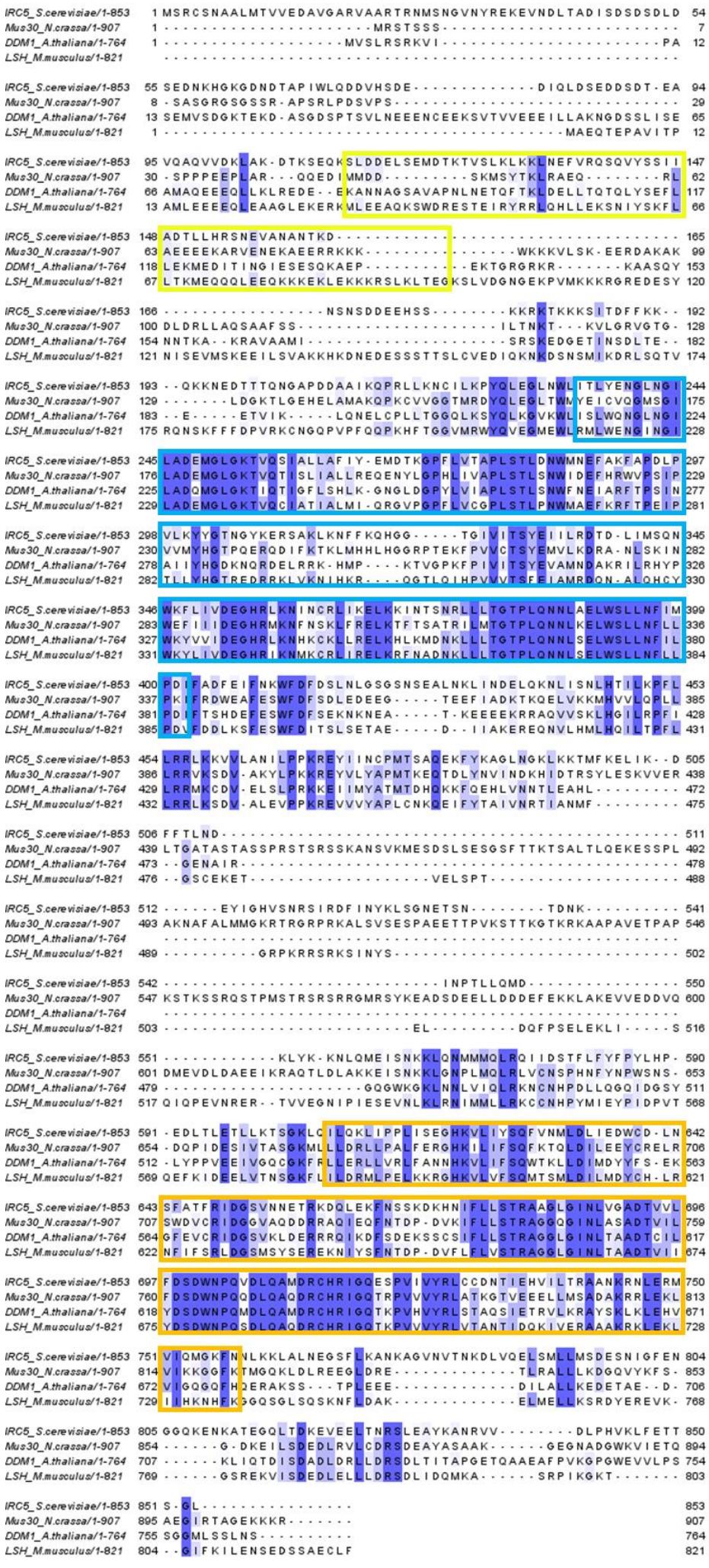

IRC5_S.cerevisiae/1-853
Mus30_N.crassa/1-907
DDM1_A.thaliana/1-764
LSH_M.musculus/1-821
Predicted coiled coil
ATPase Lobe 1
ATPase Lobe 2

Figure 3.16 Multiple sequence alignment of LSH and its homologues Irc5, Mus30 and DDM1. Alignment was performed with Clustal Omega and visualised in JalView. LSH domains are labelled as indicated, K237 is in green, Q682 (murine LSH) is in red, arginine fingers R690 and R693 are in purple.

The model of DDM1 based on Chd1 bound to the nucleosome demonstrates high similarity to the LSH model and the RMSD value generated by PyMol after superimposition of the models was 0.000 Å with 390 Cα after outlier rejection (***Table 3.7***). The β-sheet core of the ATPase domain in both lobes again was the best modelled portion of the protein (***Fig. 3.17 C***). The α-helical bundle surrounding the β-sheets were of average-to-high and average quality according to ProQ2 gradation (***Fig. 3.17 C***). The region connecting two ATPase lobes is shown as a mixture of α-helices and loops and is not present in other remodellers, however, the low model quality of this part does not allow to make conclusions about this region (***Fig. 3.17 C***). The Q682 residue located at the centre of the ATPase lobe 2 and facing the DNA-binding cleft is located in a loop attached to an α-helix reinforcing the hypothesis that Q682R mutation would prevent effective closure of the lobe and, hence, may disrupt ATP hydrolysis.

The outer α-helices of the ATPase domain have average model quality (***Fig. 3.17 C***), however, their exact role in the protein function is unclear, since they may also interact with the LSH terminal regions that have not been modelled. The biggest discrepancy between the models is the flexible loop belonging to the ATPase linker and located at the outermost part of lobe 2, which has some helical elements in DDM1 but is only a loop in LSH, however, the model quality is the lowest in the whole model, therefore no conclusions can be drawn from this region (***Fig. 3.17 A, B, C***). The same is relevant for the N-terminus: in LSH it is absent but modelled as a loop in DDM1, however, the bad modelling quality makes this region unreliable for analysis.

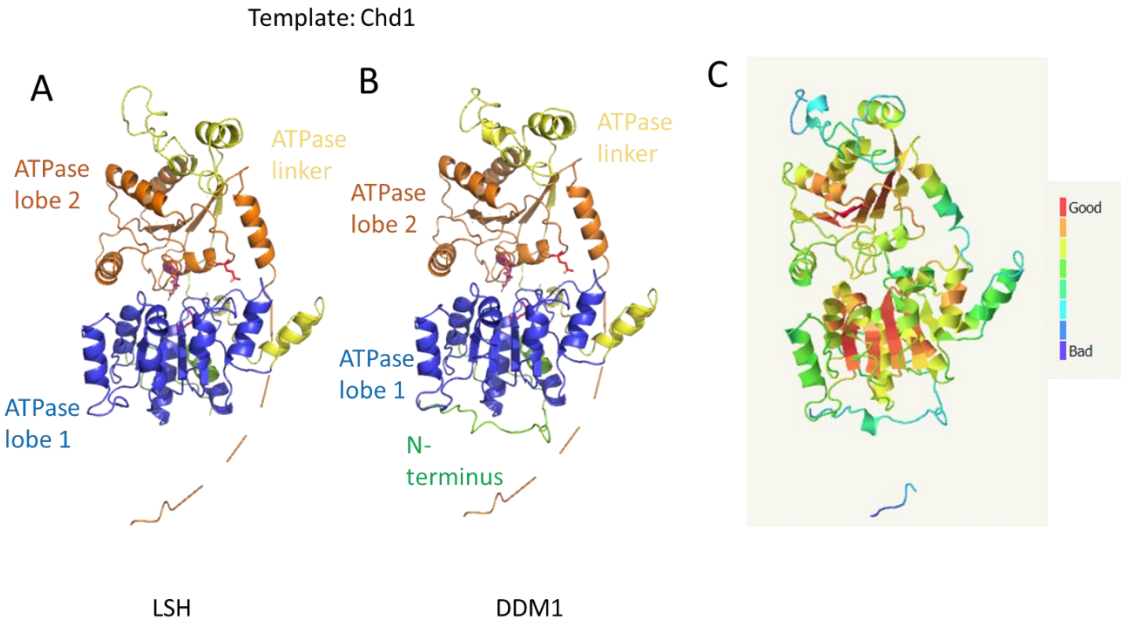


Figure 3.17 Phyre2 homology modelling of murine LSH and A. thaliana DDM1 ATPase domain based on yeast Chd1 template (PDB ID 5o0g_W). A) LSH ATPase shown as cartoon with the labelled domains and Q682 residue shown in pink sticks; B) DDM1 ATPase shown as cartoon with the labelled domains and Q682 residue shown in sticks; C) ProQ2 model quality analysis of DDM1 generated in Phyre Investigator with the colour coding scale.

Table 3.7 LSH and DDM1 model information based on Chd1-NCP complex and model alignment information.

Protein	**Protein template RDB ID**	**Initial Cα atoms aligned**	**Final Cα atoms aligned**	**Cα RMSD (Å)**
Chd1-NCP	5o9g, chain W	426	390	0.000

Ino80 in a complex with the nucleosome was the next template used for DDM1 modelling. The characteristics of the model bear similarities to the one based on Chd1 described above. The RMSD value between LSH and DMM1 was 0.001 Å with 397 Cα (***Table 3.8***). As in the previously described models, the β-sheet cores of both ATPase lobes have the highest modelling quality as estimated by ProQ2 (***Fig. 3.18 C***). The ATPase loop due to low homology has the lowest modelling quality. The Q682 residue is facing the DNA-binding cleft and similarly to the other models indicates possible ATP hydrolysis disruption or DNA-interaction disruption (see Ino80-based modelling of LSH above).

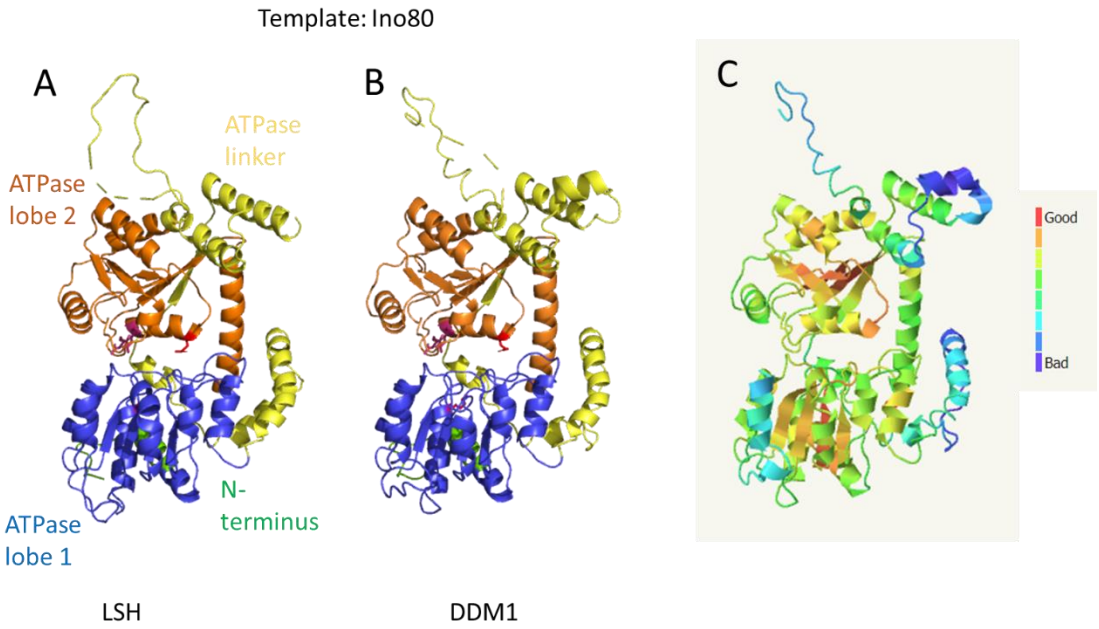


Figure 3.18 Phyre2 homology modelling of murine LSH and A. thaliana DDM1 ATPase domain based on Ino80 template (PDB ID 6fml_G) from C. thermophilum. A) LSH ATPase shown as cartoon with the labelled domains and Q682 residue shown in sticks; B) DDM1 ATPase shown as cartoon with the labelled domains and Q682 residue shown in sticks; C) ProQ2 model quality analysis of DDM1 generated in Phyre Investigator with the colour coding scale.

Table 3.8 LSH and DDM1 model information based on Ino80-NCP complex and model alignment information.

Protein	**Protein template RDB ID**	**Initial Cα atoms aligned**	**Final Cα atoms aligned**	**Cα RMSD (Å)**
Ino80-NCP	6fml, chain G	452	397	0.001

Snf2-based model of DDM1 shares high similarity with the LSH model. The RMSD value after outlier rejection was 0.001 Å with 412 Cα (***Table 3.9***). The highest quality of the model is again in the β-sheet region of the ATPase domain, however, this modelling quality is lower compared with the other models, according to ProQ2. The surface of the protein has low modelling quality, therefore, the exact location of the surface helices and loops are debatable. Since no terminal regions of LSH were modelled, the function implications of the protein surface and its interaction with the protein termini are only open to speculation, so far. Similarly, the lobe 1 helices harbouring the basic patch in Snf2 (Brandani et al., 2018) (not shown) have very low modelling quality, therefore, it is difficult to make assumptions about the relevant region in LSH.

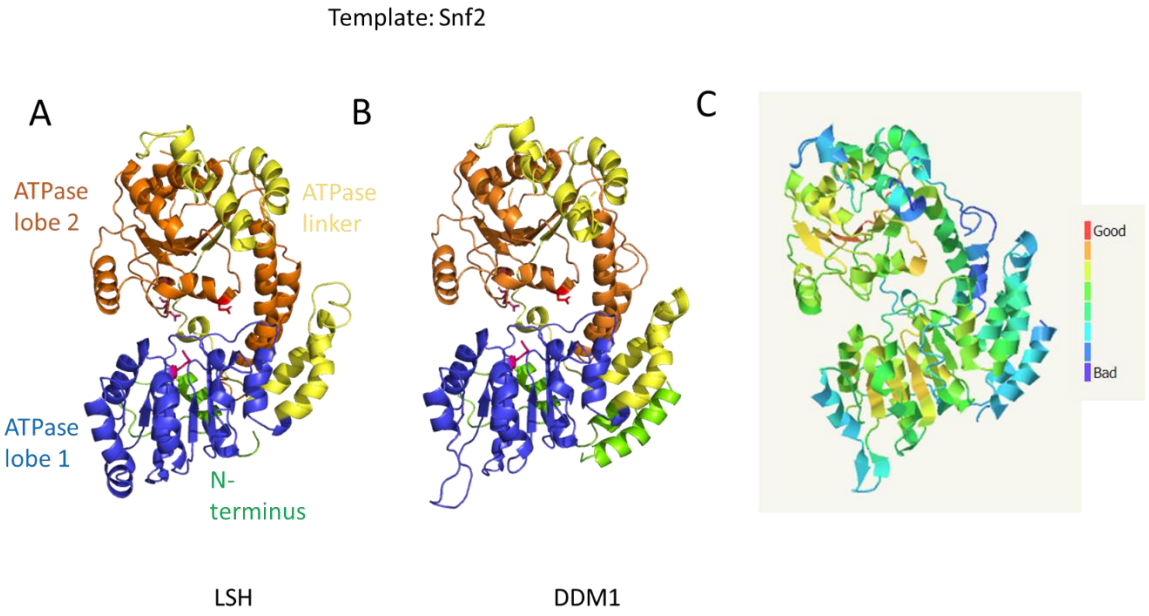


Figure 3.19 Phyre2 homology modelling of murine LSH and A. thaliana DDM1 ATPase domain based on yeast Snf2 template (PDB ID 5x0y_O). A) LSH ATPase shown as cartoon with the labelled domains and Q682 residue shown in sticks; B) DDM1 ATPase shown as cartoon with the labelled domains and Q682 residue shown in sticks; C) ProQ2 model quality analysis of DDM1 generated in Phyre Investigator with the colour coding scale.

Table 3.9 LSH and DDM1 model information based on Snf2-NCP complex and model alignment information.

Protein	Protein template RDB ID	Initial Cα atoms aligned	Final Cα atoms aligned	Cα RMSD (Å)
Snf2-NCP	5x0y, chain O	498	412	0.001

The ISWI-based model of DDM1 is, again, similar to the LSH one with RMSD of 0.001 Å with 385 Cα after outlier rejection (***Table 3.10***). The β-sheet cores and some helical regions surrounding them have high modelling quality, supporting the notion that the ATPase lobe core of LSH consists of parallel β-sheet, similarly to other Snf2-family remodellers. The modelling of Q682 region has average-to-high quality, along with the proximal loop on lobe 1, however, the proximal perpendicular helix of lobe 2 has low modelling quality, making the mechanism of the culprit unclear. Again, the modelling does not rule out different possibilities of how the mutation may compromise LSH function.

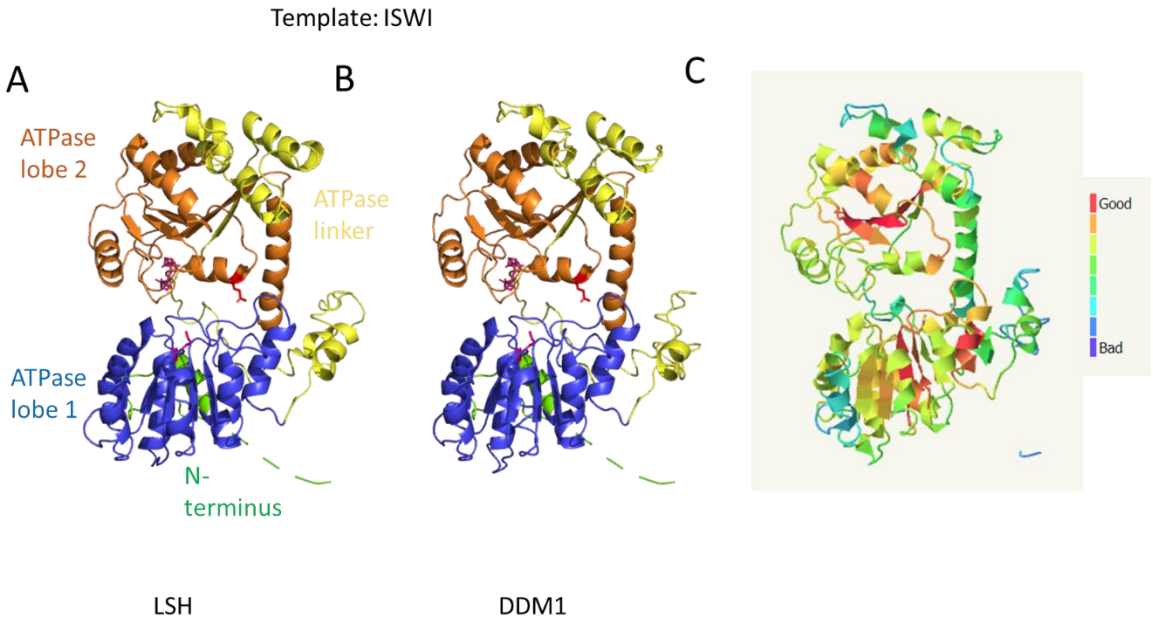


Figure 3.20 Phyre2 homology modelling of murine LSH and A. thaliana DDM1 ATPase domain based on yeast ISWI template (PDB ID 6iro_L). A) LSH ATPase shown as cartoon with the labelled domains and Q682 residue shown in sticks; B) DDM1 ATPase shown as cartoon with the labelled domains and Q682 residue shown in sticks; C) ProQ2 model quality analysis of DDM1 generated in Phyre Investigator with the colour coding scale.

Table 3.10 LSH and DDM1 model information based on ISWI-NCP complex and model alignment information.

Protein	**Protein template RDB ID**	**Initial Cα atoms aligned**	**Final Cα atoms aligned**	**Cα RMSD (Å)**
ISWI-NCP	6iro, chain L	451	385	0.001

3.3 Discussion

Due to the lack of high-resolution structural data on LSH, homology modelling using the structures of chromatin remodellers from Snf2 family (Chd1 - PDB ID 5o9g, Ino80 - PDB ID 6fml, Snf2 - PDB ID 5x0y, ISWI – PDB ID 6iro) was performed by Phyre2 (Kelley et al., 2015) to gain further insights into LSH possible structure and the functional aspects on the wild type and the ICF syndrome-associated mutant.

PsiPred (Jones, 1999) prediction of the secondary structure revealed that full-length LSH has extensive α-helical regions with intermittent β-sheet regions and random coils (***Fig. 3.1***). The N-terminal part of LSH is predicted to harbour a coiled coil by the servers COILS and MARCOIL, implying either dimerisation or protein-protein interaction propensity of LSH (Truebestein and Leonard, 2016). It has previously been shown by SEC-MALS experiments (Simon

Varzandeh, 2016) that LSH is present in solution as a monomer, therefore, this coiled coil may be a site of protein-protein interaction rather than dimerisation. The binding partner of LSH CDCA7 (Jenness et al., 2018), through a tight DNA interaction via its zinc finger domain, may provide a docking site for LSH binding to the nucleosome, and the interaction between the two proteins is likely to occur through their coiled coil regions.

Previously, the N-terminus of LSH was suspected to interact with the *de novo* DNA methyltransferase DNMT3B, presumably through the coiled coil (Myant and Stancheva, 2008). This experiment utilised recombinant proteins, and an N-terminal GST-tag was fused to LSH N-terminus. S. Varzandeh could not detect LSH and DNMT3B interaction in vitro in the absence of the nucleosome (Simon Varzandeh, 2016). As observed in my experiments (see discussion on CDCA7), LSH has certain non-specific affinity for the GST-tag, therefore, this finding would require verification in a more reliable system. It should be noted that even though LSH may be functionally related to DNMT3B, this interaction was not detected in a mass spectrometry study of *Xenopus laevis* eggs (Jenness et al., 2018), indicating a possible transient nature of the interaction. Therefore, future experiments will require further optimisation of LSH truncations to elucidate the role of its N-terminus.

Apart from the predicted coiled coil no further information could be deduced from the LSH N-terminal due to lack of homology with the other Snf2-family remodellers (***Fig. 3.5***). This is hardly surprising, since remodellers are unified by the highly homologous ATPase domain, however, other domains are accessory rendering each remodeller subfamily its unique characteristics. This can be the case for LSH, since it appears to have non-redundant functions in silencing heterochromatic repetitive elements (Yu et al., 2014). Unlike other remodellers that have N-terminal regulatory domain such as AutoN, HSA, post-HSA and chromodomain (Clapier and Cairns, 2012, Clapier et al., 2016, Yan et al., 2016, Hauk et al., 2010) with well-studied functions, the N-terminus of LSH does not have such well-characterised domains. Recently identified conserved motif ppHSA present in human LSH$^{64-85}$ (mouse LSH$^{48-69}$) and other remodellers was shown to have an *in vivo* function and to stabilise the N-terminus by clamping it to ATPase lobe 2 in ISWI (Ludwigsen et al., 2017), which provides a room for speculation for this domain in LSH.

Another Snf2/Swi2 related protein, Mot1 also has an extensive non-conserved but functionally important N-terminal region (Auble et al., 1997). Unlike other Snf2/Swi2 remodellers, its primary interaction partner is not the nucleosome but TATA-binding protein

(TBP) and duplex DNA (Wollmann et al., 2011). In this context the N-terminal domain of Mot1 binds to TBP and, while binding to DNA with Snf2/Swi2 ATPase domain, weakens the TBP-DNA interaction (Wollmann et al., 2011). Rad54 utilises its unstructured N-terminus to bind Rad51 (Raschle et al., 2004). Similarly, LSH may utilise its N-terminus in a protein-protein interaction with a DNA-binding protein such as CDCA7, to compensate for the lack of a DNA-binding domain in other regions apart from its ATPase domain. Another interacting partner of LSH N-terminus is the transcription factor E2F3, which has been found to specify LSH positioning at the transcription start site (von Eyss et al., 2012).

Another important feature of N-terminus is its role in LSH stability. Previous attempts to express the ATPase core with various N-terminal truncations of LSH resulted in high level of protein degradation and chemical crosslinking demonstrated proximity of the N-terminus to the ATPase core (Simon Varzandeh, 2016). Therefore, the N-terminus appears to have an important role in LSH both in apo state and in complex with the nucleosome, however, its structural characteristics remain vague. Previously, *ab initio* modelling of the N-terminus LSH^{1-176} was performed by Simon Varzandeh (Simon Varzandeh, 2016). The two structures generated by I-TASSER and Phyre were a helical bundle and an elongated helical structure linked by loops, respectively. Optimisation of expression conditions would be required to biochemically elucidate the exact role of the LSH N-terminus.

Homology modelling based on various Snf2 family remodellers (Chd1, Ino80, Snf2, ISWI) demonstrated that LSH is likely to have an ATPase domain structurally similar to the other Snf2 family ATPase (***see Section 1.2***), with the central parallel β-sheets surrounded by α-helices harbouring the Walker A and B motifs (Ye et al. 2004). The orientation of the lobes and motif similarity to the well-studied remodellers allow to assume a similar remodelling mechanism in LSH. The lysine K237 residue of murine LSH has been shown to play an important role in ATP binding (Burrage et al., 2012; Ren et al., 2015; Termanis et al., 2016) is located in lobe 1 and is likely to interact with γ-phosphate on ATP (Ye et al. 2004). Lysine K237 forms an ATP-binding structure along with the arginine fingers located on lobe 2 (***Fig. 3.7 F, 3.9 E, 3.11 F, 3.13 F***). Consistently with mutagenesis experiments (Burrage et al., 2012; Ren et al., 2015; Termanis et al., 2016), mutation of K237 abolishes cellular LSH activity, presumably via the lack of ATP binding.

Homology modelling provided further insights into the role of the ICF-relevant mutation. Q682R mutation in murine LSH, which corresponds to Q699R in human LSH, is located away

from the ATP binding site (***Fig. 3.7 F, 3.9 E, 3.11 F, 3.13 F***), therefore, solely from the modelling it was not clear whether the ATP binding itself was affected by this mutation. The mutated residue is located in ATPase lobe 2 opposite the modelled loop in lobe 1 in LSH models based on different chromatin remodellers. Introduction of a positively charged and more protruding residue may disrupt conformation of lobe 2 and thus interfere with lobes closure upon ATP binding and chromatin remodelling. Alternatively, the mutation may disrupt lobes closure without introducing any deformation to lobe 2. Experiments in embryonic stem cells demonstrated that K237A mutant and deletion of DEAH box mutant could not restore methylation levels in repeat elements after introduction of those mutations to LSH-null cells, presumably via abrogated binding of *de novo* DNA methyltransferase DNMT3B (Ren et al., 2015). This indicates that LSH forms nucleosome-free areas that can then be accessed by DNMT3B, but not necessarily via direct protein-protein interaction and DNMT3B recruitment. Interestingly, LSH knock out was associated with reduced nucleosome occupancy at repeat elements (Ren et al., 2015), indicating that ATPase function of LSH is important for nucleosome positioning and not only forming nucleosome-free regions. These findings, however, do not specify whether ATP molecule was able to bind LSH. Thin layer chromatography (TLC) performed by Kevin Myant (Kevin Bryan Myant, 2008) demonstrated that LSH, albeit with low efficiency, could perform weak ATP hydrolysis both in the presence of free DNA and a nucleosome, compared to the lack of stimulus. This, however, still does not elucidate the requirements for LSH-ATP binding and the effects of Q682R mutation on it.

Since it appeared that Q682 residue is located proximally the DNA gyre, it is feasible to suggest that an introduction of a positively charged residue could strengthen the electrostatic potential between LSH and DNA and prevent LSH from effective sliding along the strand. Furthermore, Q682R mutation is located two residues away from a conserved tryptophan (W679 in murine LSH), which inserts itself into the DNA duplex minor groove as shown in the Snf2 cryo-EM structure (Liu et al., 2017a). Mutation W1185A in Snf2 only reduces ATP hydrolysis but almost block remodelling activity (Liu et al., 2017a). It is not clear from the structural data how this mutation causes this effect but it is possible to assume that this tryptophan may provide a lever effect or an additional contact point necessary for the remodeller's sliding across the DNA strand or a similarly critical contact point. In a similar manner, Q682R may not disrupt ATP binding and hydrolysis, but may prevent efficient sliding due to stabilisation of LSH on the phosphate backbone. Mutagenesis experiments with LSH would be required to determine that. For that the restriction enzyme accessibility

remodelling assay should be initially optimised for murine LSH and CDCA7, with LSH Q682 and W679 residues subsequently mutated. Various residues, such as smaller and uncharged alanine or glycine, another positively charged residue such as lysine, or a bulky residue such as tryptophan, could be introduced at the position 682 to see the effects. Fluorescence resonance energy transfer (FRET) assay could be another project to dissect the roles of these LSH residues. However, this can only be achieved with a robust expression system and abundant protein yield.

Homology modelling highlighted several similarities between the well-studied remodellers and LSH. One of them is remodeller binding location on a nucleosome. LSH models based on Chd1, Snf2 and ISWI were shown to bind at superhelical location 2 (SHL2) (***Fig. 3.7 A and B, 3.9 A and B, 3.11 A and B, 3.13 A and B***) due to intrinsic instability at this site favouring translocation initiation (Li et al., 2019; Winger et al., 2018). SHL2 location around 20 bp away from the nucleosome dyad on the entry DNA gyre is a standard binding location for Snf2 family remodellers (McKnight et al., 2011; Saha et al., 2005; Schwanbeck et al., 2004; Zofall et al., 2006). Recent studies showed that remodeller binding at SHL2 induces a DNA bulge, which then propagates across the histone octamer in a corkscrew-like manner resulting in translocation (Li et al., 2019; Winger et al., 2018; Yan et al., 2019b). This inchworm mechanism was contrasted to translocation via loop-formation (Zhang et al., 2006; Zofall et al., 2006). The LSH models based on transition states of Snf2 demonstrated how LSH fits into the structure, which adopts open conformation in apo or ADP-bound state, mimicking resting state before and after ATP hydrolysis and closed conformation when bound to an ATP analogue ADP-BeF (***Fig. 3.15***). This indicates a lack of major conformational change before ADP binding and after its release. Similarly to the recently identified mechanism, LSH is likely to be capable of pumping DNA towards the dyad in a two-step manner: first the DNA downstream SHL2 is pulled towards the remodeller once it is bound, this action creates under-twisting, which is then resolved at the step two following ATP hydrolysis (Li et al., 2019; Winger et al., 2018). Energy consumption of a twist defect formation is still unknown (Li et al., 2019) and it is unclear whether LSH being a relatively small protein compared to the known remodellers can induce this nucleosomal state, however, high similarity of the ATPase domain indicates that this mechanism can also be relevant to LSH.

The LSH model based on Ino80 was bound to SHL6 (***Fig. 3.9 A and B***). Even though LSH and Ino80 share high identity of the ATPase domain (***Table 3.1***) and Phyre2 had good quality

parameters (***Table 3.3***), this localisation of LSH should be taken with caution. Ino80 is a part of a large megadalton multi-subunit complex (Tosi et al., 2013), whereas LSH has not been found in a complex (Myant and Stancheva, 2008). Ino80 ATPase was found to bind SHL6, whereas SHL2 was bound by Arp5-Ies6 – the proteins also present in Ino80 complex (Brahma et al., 2017). SHL6 is located closer to the entry site of the nucleosome and it is a good position to introduce significant disruptions into the nucleosome core, unlike a more subtle bulge formation induce at SHL2 due to its closer location to the edge of the nucleosome. Ino80 is known to perform histone variant exchange, form centrally positioned nucleosomes and was shown to induce major disruptions of ~15 bp, with auxiliary subunits were shown to act on SHL2 as a counter grip (Ayala et al., 2018; Eustermann et al., 2018). Even though the exact mechanism of LSH remodelling has not yet been elucidated, it is unlikely that is acts similarly to Ino80 due to its physical capacities. Interestingly, Liu et al. and Li et al. also found that Snf2 was able to bind the 601-containing nucleosome at SHL6 position in a subset of samples, however, biological relevance of this interaction was not investigated in those studies (Li et al., 2019; Liu et al., 2017a).

Remodelling potential of Snf2 family proteins is ensured not only by the ATPase domain, but additional interaction with the nucleosome. This was reflected in LSH homology modelling, which shows histone H4 tails protruding towards the ATPase lobes (***Fig. 3.7 C, 3.11 C, 3.13 C***). In the relevant remodellers the H4 tail contains an important nucleosomal epitope – the basic patch – which plays a role in remodelling, remodeller auto-inhibition, recruitment and orientation, linker-length sensitivity and nucleosomal array formation (Clapier, 2002; Clapier et al., 2001; Clapier and Cairns, 2012; Dang et al., 2006; Hwang et al., 2014). LSH was not found to contain any similar acidic pockets, nevertheless, its N-terminus is abundant in glutamates, which may provide contact points with the H4 basic patch, however, their potential functions can only be determined experimentally.

Histone H3 tail is another significant binding partner for Snf2 family remodellers, predominantly via its modifications (Mansfield et al., 2011; Santos-Rosa et al., 2003). H3 tail acetylation was shown to increase affinity of SWI/SNF and RSC for the nucleosome through their bromodomains (Chatterjee et al., 2011). SANT domain present in subunits of those remodelling was also found to play a role in H3 tail binding (Boyer et al., 2004). Chromodomains of Chd1 interact with methylated H3 tail and their deletion impairs binding and remodelling of the nucleosome (Flanagan et al., 2005). In LSH homology modelling H3

tail is located away from the modeller ATPase domain and no obvious contacts were present in any of the models (***Fig. 3.7 A, 3.9 A, 3.11 A, 3.13 A***). However, the domains lacking in the LSH models may be relevant for this interaction, even though no bromodomains and chromodomains have been identified in LSH. Histone interaction may be an important feature of an LSH binding partner, such as CDCA7, however, this hypothesis is not yet substantiated by evidence.

Additional nucleosomal epitope relevant to remodelling was recently described in molecular modelling simulation of Snf2. It has been found that the basic patch (K855, R880 and K885) in lobe 1 of Snf2 directs the orientation of the remodeller by acting as an electrostatic spring (Brandani et al., 2018). This portion of lobe 1 is located in the similar position in LSH and can potentially have a similar function. On examination on the LSH sequence, the relevant site was represented by R293, Q323, H328, which does not make it a *bona fide* basic patch. However, this site can be examined experimentally by mutating RQH to KRK to mimic the Snf2 basic patch. Along with this, mutating all three residues to alanine would provide additional insights and a potential negative control.

Additional information about LSH can be gained from its distant homologues DDM1, Irc5 and Mus30 (Alvaro et al., 2007; Basenko et al., 2016; Jeddeloh et al., 1999). Irc5 was recently been found to facilitate cohesin loading on the centromeres via its ATPase domain activity (Litwin et al., 2017), indicating functions independent of DNA methylation. It would be interesting to evaluate this function in LSH profile and to explore the mechanisms alternative or linked to DNA methylation. Recent findings demonstrated the role of chromatin remodelling in cohesin loading: the remodelling complex RSC was found to not only to generate nucleosome-free regions but also independently from its remodelling function act as a docking site for Scc2-Scc4 cohesin loading complex (Muñoz et al., 2019).

Mus30 has been linked to ubiquitylation via its association with a E3 ubiquitin ligase homologue WDR76 (Basenko et al., 2016). These functions were, however, outside the scope of this homology modelling.

Irc5 and Mus30 do not have extensive experimental information, therefore, DDM1 was chosen for homology modelling. The resulting models based on the same templates as for LSH were highly similar to the LSH models (***Fig. 3.17, 3.18, 3.19, 3.20***). Even though LSH and DDM1 are present in very different organisms, they both participate in DNA methylation of heterochromatic regions (Tan et al., 2018). Activity of DDM1 was shown as a requirement to

bypass linker histone H1 as a barrier to silencing in heterochromatin as well as providing nucleosome-free regions for establishment of DNA methylation (Lyons and Zilberman, 2017; Zemach et al., 2013). Both LSH- and DDM1-containing cells exhibit 10 bp methylation periodicity (Jimenez-Useche et al., 2013), which is lost in their respective mutants (Lyons and Zilberman, 2017), which could arise from DNA loop formation on the surface of the nucleosome or methylation on the linker DNA, which is then shifted to the nucleosomal core. These findings provide further insight into the ability of LSH to methylate heterochromatic regions, which, as in the absence of DDM1, can also be performed by other remodellers (Lyons and Zilberman, 2017). This propensity of LSH towards heterochromatin may be related to the more compact size of LSH compared to other remodellers, which are also often a part of large multi-subunit complexes.

Homology modelling, although providing some information in the absence of high resolution structural data, has obvious limitations. As mentioned before, the terminal regions of LSH could not be modelled due to low or lacking homology (***Fig. 3.5***). More fine details, such as DNA bulge formation observed in other remodellers (Li et al., 2019; Winger et al., 2018) could be speculated but not reliable extrapolated to LSH, since it is not clear what force is required for this DNA translocation mode and if it requires additional force from interactions with nucleosomal epitopes and auxiliary domains and subunits. Additionally, only experimental and structural data would allow to test the common features of LSH and other remodellers, such as α13 gating helix, which stabilises the tracking DNA strand and prevents it from backward motion (Li et al., 2019). Mutations of this region impair remodelling, without compromising ATP hydrolysis (Li et al., 2019). This portion of the LSH model had low quality, as indicated by ProQ2 (***Fig. 3.11***), therefore, functions of this LSH region are not clear.

To conclude, homology modelling of LSH not only provided some insights into LSH structure and functions but also highlighted limitations of this approach and reinforced the importance of optimising LSH expression for more fundamental structural and biochemical studies.

4. EXPRESSION AND PURIFICATION OF LSH

4.1 Introduction

LSH is a poorly studied member of the chromatin remodelling family and its function is implicated in development of ICF syndrome. To gain further insights into LSH structure and functions by carrying out biochemical and structural studies I needed to produce the protein in the amount and purity adequate for structural and biochemical studies. Previously, LSH has been attempted to be expressed in *E. coli*, however, this expression system was not adequate for the murine LSH (Kevin Bryan Myant, 2008). Therefore, all subsequent purification attempts have been carried out in the insect cell system, which allows to generate proteins with the current post-translational modifications (PTMs).

Previously, LSH has been expressed by Simon Varzandeh (Simon Varzandeh, 2016), however, the closer examination of the remaining stocks produced by him, revealed that to carry out high-resolution structural studies (e.g. cryo-EM) the protein purity could be increased. Therefore, I attempted to further optimise the purification process of LSH.

4.2 Cloning and expression of LSH and LSH^{K237Q}

As per previous expression attempts, murine LSH could not be expressed in bacteria. Hence, an insect cell expression system using baculoviral infection was utilised. The system differs from *E.coli* in that it takes advantage of the insect eukaryotic protein expression machinery, ensuring proper folding and post-translational modification (PTM) allocation of the expressed protein (Sørensen, 2010).

Wild-type murine LSH and LSH^{K237Q} had been previously cloned by Simon Varzandeh into a pFL plasmid, with a C-terminal 6xHis tag without a TEV cleavage site between the tag and the protein's N-terminus. The LSH^{K237Q} has a mutation in its ATPase domain, which results in the DNA methylation abolished as revealed by cell studies (Termanis et al., 2016). Since I was planning to carry out a nucleosome remodelling assay, using WT LSH and CDCA7, LSH^{K237Q} would act as a useful negative control. The WT LSH construct also contained the gene ORF of a *de novo* DNA methyltransferase DNMT3B. Since my experiments did not include DNMT3B, I needed to reclone LSH-6xHis into another pFL plasmid. Full-length LSH^{WT} was amplified by PCR from the pFL-LSH-6xHis-DNMT3B construct using Q5 polymerase and primers containing SalI and BamHI restriction sites (see Section 2.1.2.1 of Materials and methods). I digested both an empty pFL plasmid and LSH PCR product using SalI and BamHI enzymes and ligated

using a standard protocol (see Section 2.1.2.4 of Materials and methods), so that the LSH gene was inserted downstream of a polH promoter (***MSC2, Fig. 4.1***). Correct clones were confirmed by Sanger sequencing (see Section 2.1.2.7 of Materials and methods).

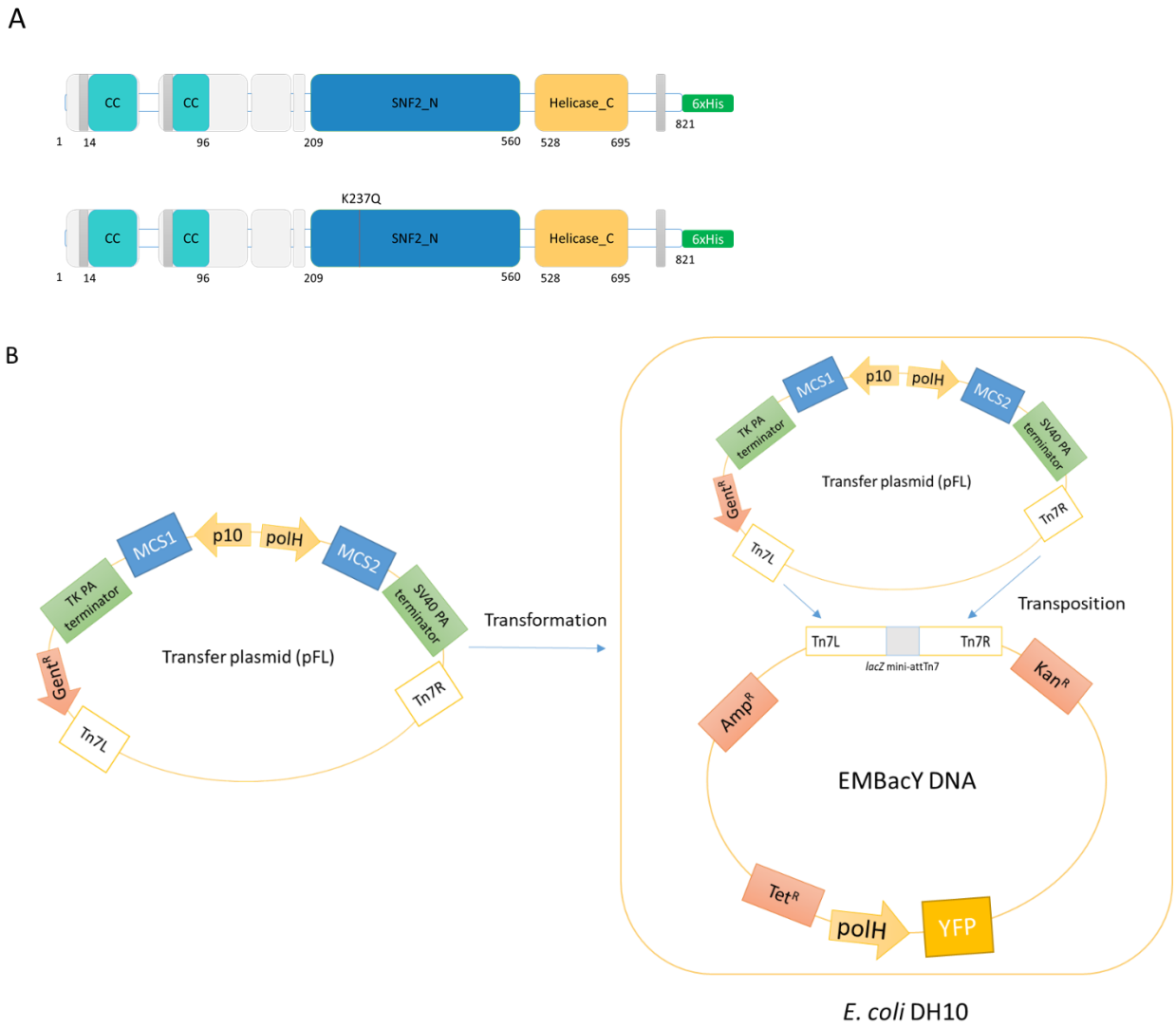


Figure 4.1 Cloning of the LSH construct. A) LSH^{WT} and LSH^{K237Q} construct schematics. B) Transformation and transposition processes in preparation of a bacmid.

For transfection into Sf9 insect cells, bacmids were generated using the MultiBac expression system (Fitzgerald et al., 2006). The system requires a two-step process. First, a transfer plasmid (in this case, pFL) was generated by cloning (***see Section 2.1.2 of Materials and methods and Fig. 4.1***). It contains two promoters – polyhedron (polH) and p10 – allowing for cloning of two gene ORFs. LSH was inserted downstream of the polH promoter. Second, competent *E. coli* DH10 strain containing the EMBacY baculoviral genome with transposition

sites and *lacZ* gene that allows to utilise blue-white screening. *LacZ* encodes β-galactosidase, which is induced by a lactose analogue IPTG (isopropyl β-D-1-thiogalactopyranoside). β-galactosidase can then break down a dye-linked substrate X-gal present in the agar, producing blue pigment. In the event of successful transposition, *lacZ* gene is disrupted and X-gal is not broken down, resulting in white (positive) colonies. EMBacY competent cells are transformed with a transfer plasmid, and Tn7 transposition takes place, which should disrupt *lacZ* gene.

Unlike the MultiBac backbone, the EMBacY backbone also encodes yellow fluorescent protein (YFP) (***Fig. 4.1***) under the control of the polH promoter. YFP expression conveniently allows detection of the expressed protein using a microscope with a UV-lamp with an emission filter of 520 to 550 nm transmission range. The caveat here is that the protein of interest and YFP are expressed under separate polH promoters, and it is difficult to estimate the levels of protein expression by YFP intensity, however, in this study, it reliably correlated with the presence of the protein of interest.

Successful production of the resulting bacmid constructs was verified by PCR using a pair of primers, one of which is complementary to the inserted gene ORF, and the other is complementary to the EMBacY backbone (see Section 2.1.2.1 of Materials and methods). In the case of LSH cloning, the forward primer was complementary to the N-terminus of LSH and the reverse primer was the universal M13 reverse primer (M13 R) complementary to the *lacZ* gene (***Fig. 4.2***). The expected sizes of LSH^{WT} and LSH^{K237Q} were 2.5 kb and 3 kb, respectively (note, the difference in sizes was due to different forward primers used to amplify the constructs).

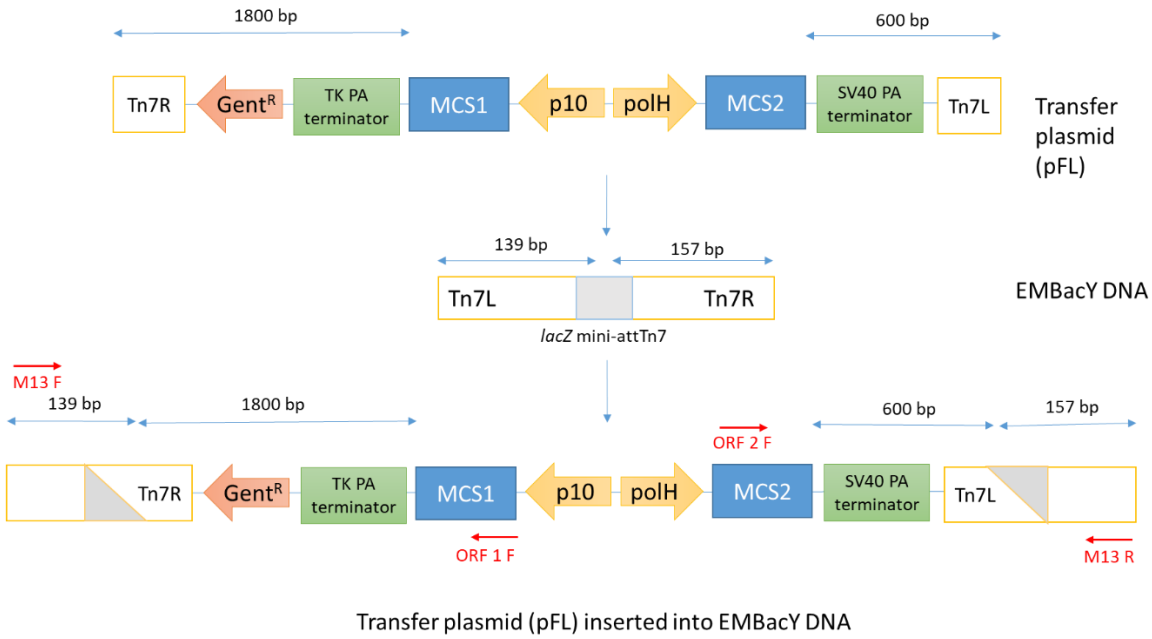


Figure 4.2 Schematic representation of bacmid with primers used for veryfication PCR after tranfer plasmid insertion into EMBacY DNA. Top - The part of the transfer plasmid that is transposed into the EMBacY DNA. Middle – Transposition site of the EMBacY DNA. Bottom – resulting construct of the transfer plasmid inserted into the EMBacY DNA.

It should be noted that the quality of the PCR product (smears present in the amplicons, ***Fig. 4.3 A,B***) did not correlate with the subsequent levels of protein expression (see below), and even though the PCR product of LSH shows poor amplification compared to that of LSH^{K237Q}, it was not reflected at the expression stage (***Fig. 4.4***).

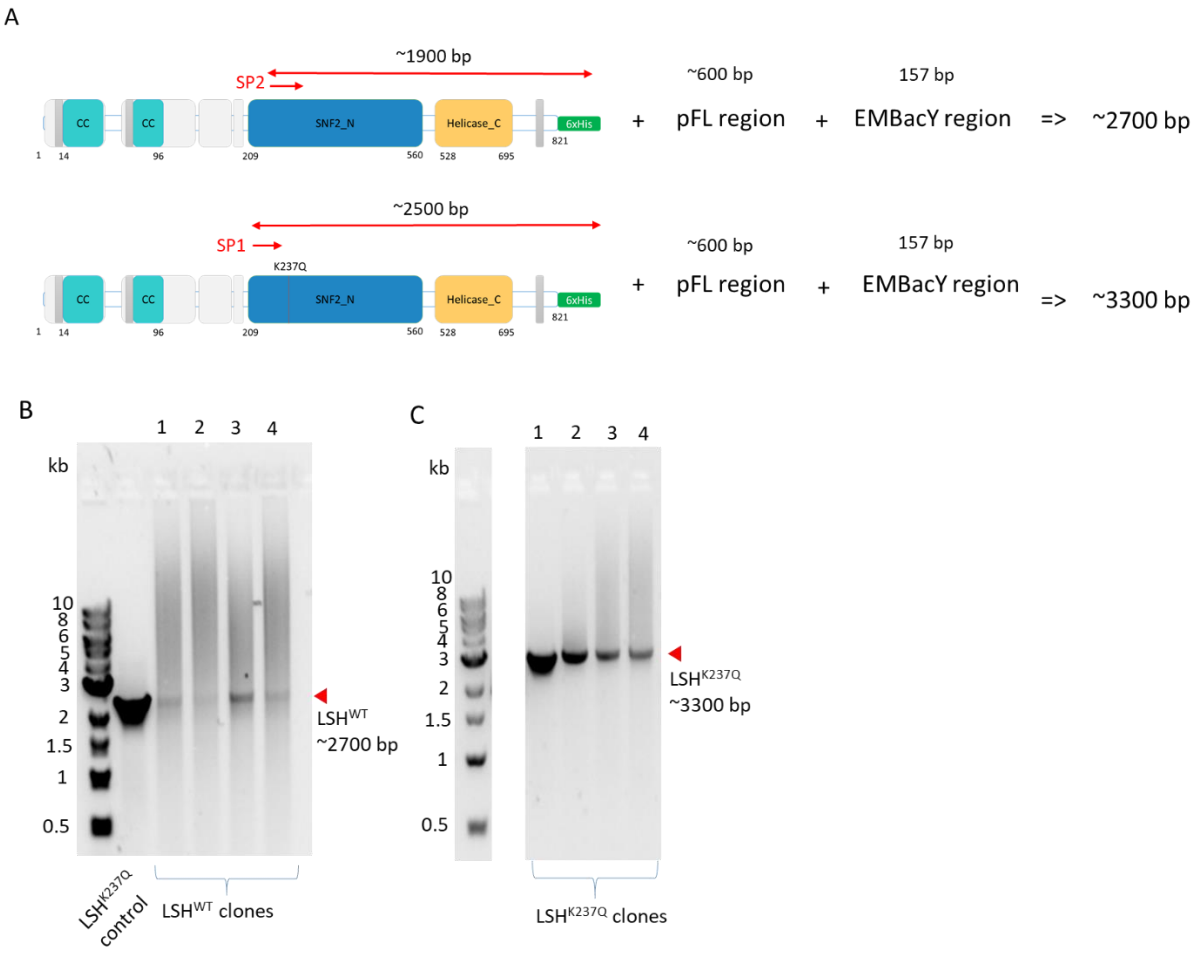


Figure 4.3 PCR verification of LSH bacmid. A) Schematic representation of the bacmid PCR approach used to verify LSH wide-type and mutant. Amplified regions of LSH, pFL plasmid and EMBacY DNA are indicated. B) LSH^{Q237K}-containing transfer plasmid previously made by Simon Varzandeh was used as a positive control for the LSH^{WT} clones; C) LSH^{Q237K} bacmids verified by PCR. The sizes of the amplicons correspond to the portions or the genes of interest, transfer plasmids and the transposition-related portions of EMBacY backbone. (CDCA7 data not relevant to this section cut out these lanes and leave a gap between the marker and the LSH mutant clones).

For a good expression of the protein of interest it is important to avoid freeze-thawing of bacmids, therefore they should be aliquoted for a single usage and stored at -20°C (***see Section 2.1.3.2***). Even though DNA is a stable molecule, bacmids are large DNA fragments, and freeze-thawing and above 0°C temperatures may have a negative effect on its integrity. Another important point is the viability of insect cells. It is not advised to perform transfections and infections if cell viability is below 95%. Passage number is also an important factor, and in the present experiments I tended to only use Sf9 below passage 20 for transfections and below passage 30 for infection and large scale expressions. It is also important to bear in mind that the cell viability does not guarantee high expression levels, at least in the case of Sf21 cells.

Bacmid #3 LSH^{WT} and #1 LSH^{K237Q} (***Fig. 4.3 A, B***) were used for transfection into Sf9 cells using a 6-well plate (see Section 2.1.3.3 of Materials and methods). After 48h, a YFP signal could be detected, peaking at 72h, at which point supernatant containing baculoviral particles was pooled from the wells as the V_0 viral generation. This was followed by infection of 50 ml of Sf9 at 1-1.5 million/ml at 1:20 (v:v) V_0 to Sf9 ratio to produce V_1. Again, at 72h fluorescent signal reached ~90-95%, supernatant was collected and the pellet (10-50 ml) was used for a pull-down and Western blot. At this point, V_1 was used directly for expression tests (in this case, the volume of the prior viral infection should be 100 ml) or for further generation of V_2. In both cases, a ratio of 1:50 (v:v) V_1 to Sf9 was used. Expression tests with V2 required 1:75 or 1:100 ratio (v:v) V_2 to Sf9. For expression tests fluorescence and viability were monitored after 24h. Generally, 72h was the time point of cell harvesting, with the highest YFP signal and a reasonable viability (85-95%). Generally, Sf9 cells are used for virus generation and High Five are used for expression, as the latter produce higher amounts of protein. However, in this case no High Five cells were available, therefore, I used Sf9 for all steps of protein production.

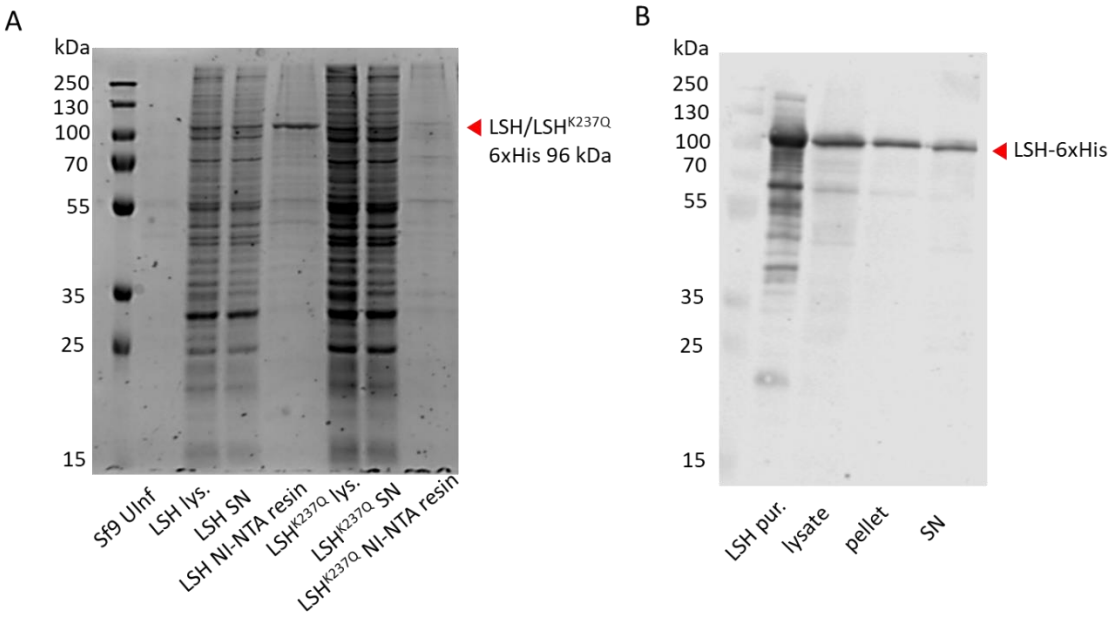


Figure 4.4 Evalution of LSH expression by pull-down and Western blot.
A) Coomassie stained 12% SDS-PAGE gel of Ni-NTA pull-down confirming LSH expression. The samples shown are uninfected Sf9 (lane 2), LSH total lysate (lane 3) , LSH supernatant (soluble fraction) (etc) and the fraction bound to the resin.
B) Western blot of LSH after wet transfer onto neutrocellulose membrane, with anti-LSH 1ry antibody and visualisation of Licor IR800-conjugated 2ry anti-mouse antibody. SN – supernatant. State what is in each lane.

Expression of LSH was verified by Protein expression was verified by a Ni-NTA pull-down (to ensure that the expressed proteins can be visualised by Coomassie staining) and by Western blot (***Fig. 4.4 B***). It should be noted that LSH^{K237Q} did not express well in this instance, however, subsequent tranformation (not shown) solved that problem.

4.3 Purification of LSH^{WT}

The presence of 6xHis tag on the N-terminus of LSH allows to utilise a robust approach of immobilised metal affinity chromatography (IMAC) as a first step of LSH purification, in which a spherical bead is attached to a chelating ligand nitrilotriacetic acid (NTA), charged with Ni^{2+} that, in turn, binds histidines in the 6xHis tag. Ni-NTA resin can be used as a slurry or as a packed column. Initially, as a test purification, LSH was incubated with the resin for 2-3h and then the resin was washed and the protein eluted through centrifugation steps, as detailed in Materials and Methods (***see Section 2.1.3.8***). However, that resulted in a lot of residual contaminants (not shown). Therefore, after incubation, the resin was packed into a gravity column for washes and elution (***Fig. 4.5 A***). The eluted protein was still heavily contaminated, due to insufficient number of washes and low concentration of imidazole in the wash buffer (10 mM). As the second purification step, LSH was applied on a HiTrap Heparin HP 1 ml column (***Fig. 4.5 B***). Heparin columns are routinely used for purification of DNA-binding proteins, through anionic sulphated glucosaminoglycans, which mimic negatively charged DNA. The electrostatic interaction is disrupted by an increase in ionic strength, making the method similar to ion exchange.

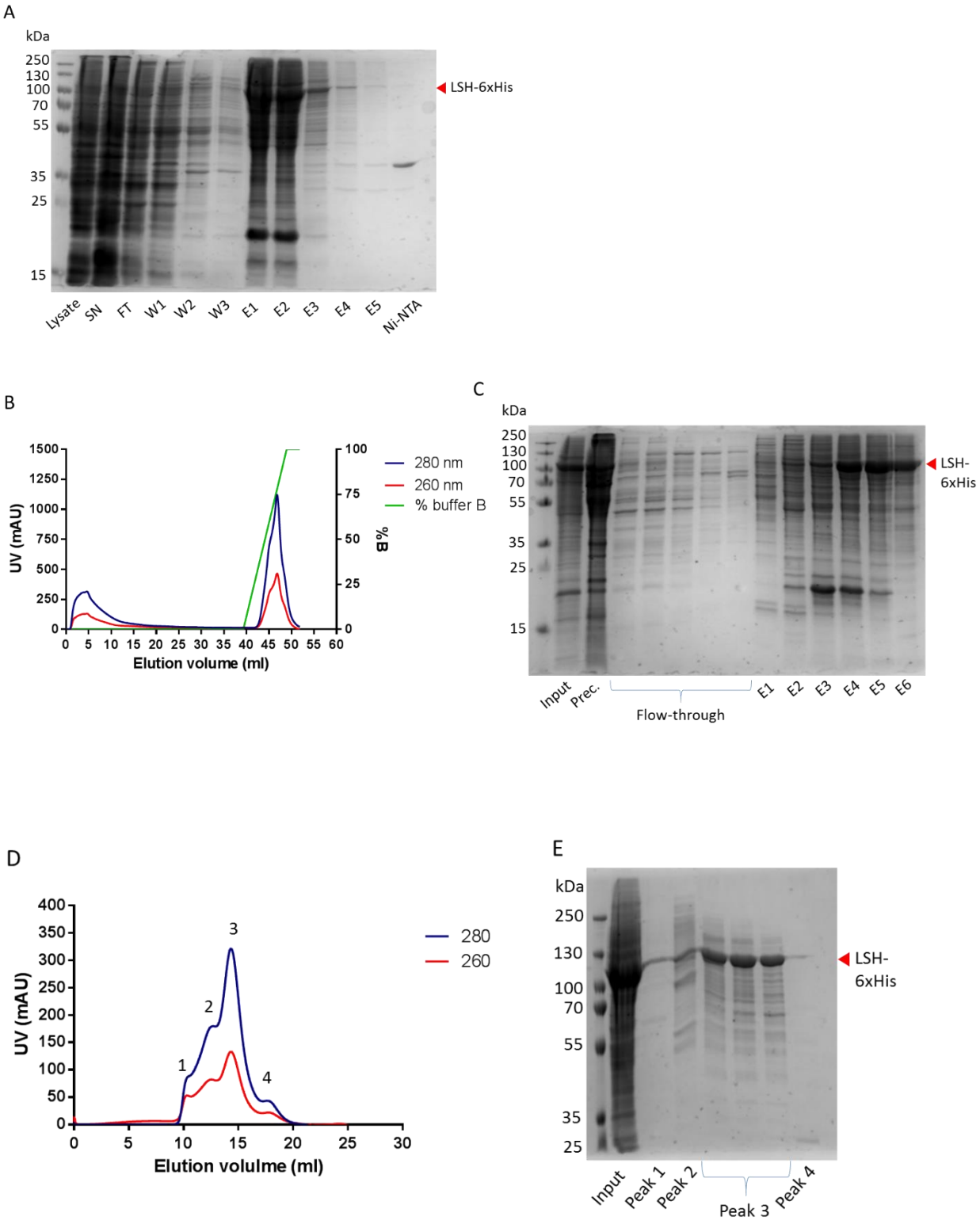


Figure 4.5 Crude LSH purification from Sf9 cells with Ni-NTA resin as the first step, followed by the heparin column and SEC.

A) Coomassie stained 12% SDS-PAGE gel of LSH that partially isolated from the total lysate. Three washes with 10 mM imidazole were performed, followed by elution with 500 mM imidazole. Wash and elution buffers contained 500 mM NaCl. 2% of each sample volume loaded per well. SN –

supernatant, FT – flow-through, W1-3 – washes, E1-5 – imidazole elutions, Ni-NTA – Ni-NTA resin after elution.
B) Chromatogram of LSH from HiTrap Heparin HP 1 ml column with UV detection (280 and 260 nm) and % of high salt elution buffer B (containing 1M NaCl, for details see Section 2.1.3.8).
C) Coomassie stained 12% SDS-PAGE gel of LSH after HiTrap Heparin HP 1 ml column. 2% of each sample volume loaded per well.
D) Chromatogram of LSH from Superdex200 10/30 24 ml column with UV detection (280 and 260 nm).
E) Coomassie stained 12% SDS-PAGE gel of LSH after Superdex200 10/30 24 ml column. 2% of each sample volume loaded per well.

Ni-NTA purification step was performed in a buffer containing 500 mM NaCl and the protein was eluted using 500 mM imidazole, making this ionic strength incompatible with binding to the heparin column. Therefore, an overnight dialysis step into a lower ionic strength buffer (150 mM NaCl) was required. That, however, resulted in significant precipitation of LSH, possibly caused by the combination of the protein concentration in the sample and the presence of impurities. The sample was spun down to pellet the precipitate, and the supernatant was applied on the heparin column.

150 mM NaCl allowed protein binding to the column, and separation of a portion of impurities, released in a flow-through (***Fig. 4.5 B***). The elution step with the top concentration of 1 M NaCl allowed release of LSH from the column at 78.3% of buffer B concentration (780 mM NaCl, 39.8 mS/cm conductivity), however, the gradient was only 10 CV, or 10 ml. Extending the gradient to 20 CV would allow better separation of LSH and other impurities.

Finally, the sample was applied to a Superdex200 10/300 24 ml column. It allowed further separation from impurities, with LSH eluting as a main peak at 14.3 ml, but the final protein was around 80% pure, which is sufficient for functional assays but not pure enough for structural studies (***Fig. 4.5 D and E***). Around 0.6 mg of LSH at 1.3 mg/ml was purified from a 600 ml culture, with a large fraction of LSH lost at the dialysis step.

In the attempt to increase LSH purity, I used a HiTrap IMAC HP (high performance) 1 ml on the ÄKTApure system, with subsequent imidazole washes of 20 mM for 30 CV and 40 mM for 15 CV (***Fig. 4.6 A***). That resulted in LSH of around 80% (***Fig. 4.6 B***). However, I had to switch to a HiTrap IMAC FF (fast flow) 1 ml column, since the HP columns have a smaller bead size, which allows for improved purification at the cost of loading efficiency. A 1 ml HP column was not sufficient for loading 150 ml lysate and resulted in column clogging and an increase in column pressure. Therefore, the rest of the sample was loaded on a HiTrap IMAC FF 1 ml column (***Fig. 4.6 C***). The eluted protein had more impurities compared with a HiTrap IMAC

HP 1 ml column (***Fig. 4.6 D***). Since I was planning to do subsequent purification steps, this was not considered a problem.

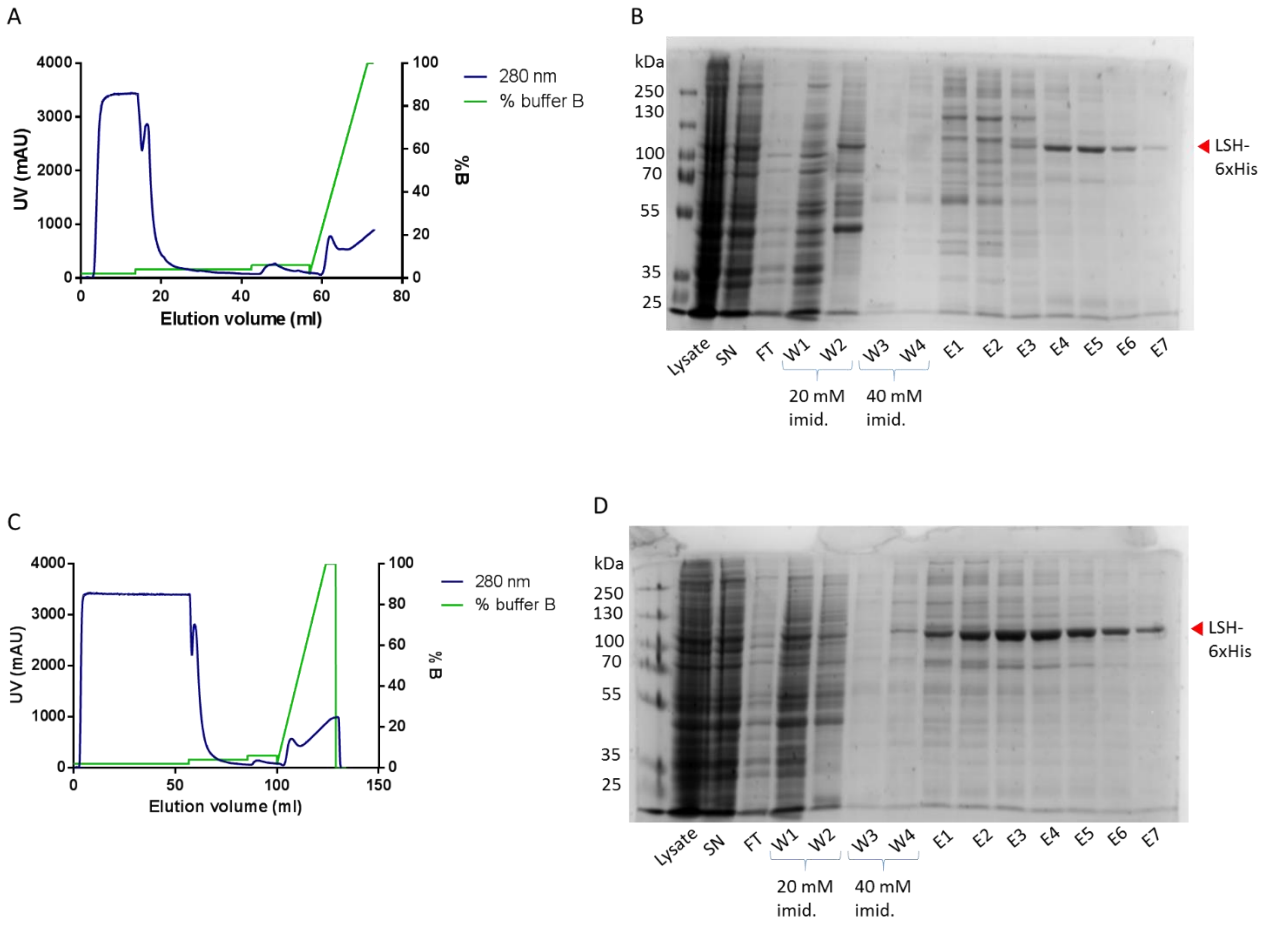


Figure 4.6 LSH purification using HiTrap IMAC HP and FF 1 ml columns.
A) Chromatogram of LSH applied on HiTrap IMAC HP 1 ml with UV detection (280 nm) and % of elution buffer B (see Section 2.1.3.8).
B) Coomassie stained 12% SDS-PAGE gel of LSH after HiTrap IMAC HP 1 ml. 2% of each sample volume loaded per well.
C) Chromatogram of LSH applied on HiTrap IMAC FF 1 ml with UV detection (280 nm) and % of elution buffer B (see Section 2.1.3.8).
D) Coomassie stained 12% SDS-PAGE gel of LSH after HiTrap IMAC FF 1 ml. 2% of each sample volume loaded per well.

To avoid the time-consuming and potentially detrimental dialysis step, I instead used the HiPrep 26/10 desalting column packed with Sephadex G-25 fine resin for a quick (10-15 min) buffer exchange. A Desalting column is a type of gel filtration column with pores too small for protein and peptide partitioning, thus inhibiting only salt and other small molecules. The Elution buffer contains the desirable salt concentration. Since previously dialysis resulted in

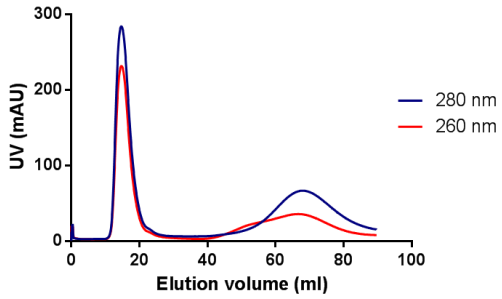
a lot of protein precipitation, I attempted to avoid a drastic decrease in ionic strength and chose 250 mM NaCl as the final salt concentration. LSH did not seem to precipitate and appeared as a single peak in the void volume immediately at the start of elution (***Fig. 4.7 A***).

After desalting, instead of a second step heparin column, the sample was applied on a Resource S column, which acts as a strong cation exchanger, binding positively charged proteins. It has superior properties compared to standard ion exchange resin due to homogeneous size and distribution of resin particles.

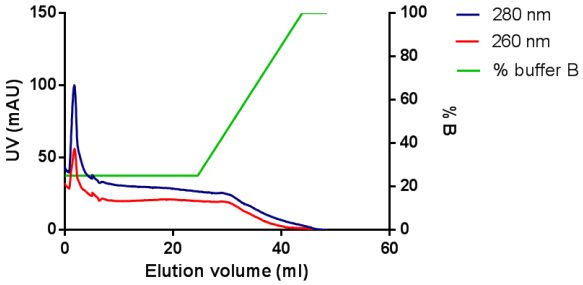
The theoretical pI of LSH is 8.2, indicating an overall positive charge in a buffer close to physiological pH (7-7.5). After LSH was applied on the Resource S column, it appeared in the flow-through (***Fig. 4.7 B***). It is, however, not uncommon that some proteins predicted to be positively charged, in fact, bind to anion exchange resin, instead of cation exchange, possibly through some negatively charged pockets or surfaces. However, a more plausible explanation for the lack of binding is the excessive ionic strength. Therefore, I diluted the rest of the sample to bring the NaCl concentration to 150 mM, which allowed LSH binding as indicated by a test run (***Fig. 4.7 C***). Then the rest of the sample was loaded on the column allowing separation of the main peak from minor impurities (***Fig. 4.7 D, E***).

As a final step, LSH was applied on the Superdex200 10/300 24 ml column to perform size exclusion chromatography (***Fig. 4.7 F, G***), which allowed separation of LSH (peak 2) from minor impurities (peak 1). Around 0.2 mg of protein that was close to 100% purity was obtained. That was an improvement from previous LSH purification approaches, and the sample would be adequate for structural studies, such as by cryo-electron microscopy (cryo-EM).

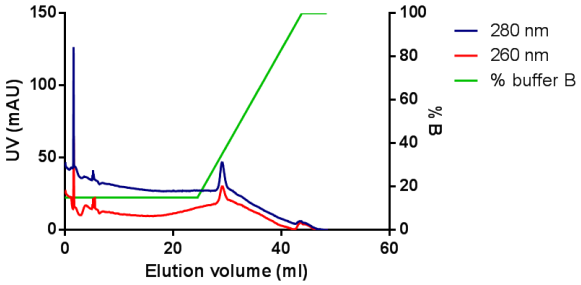
A


UV (mAU)
300
200
100
0
0
20
40
60
80
100
Elution volume (ml)
280 nm
260 nm

B


UV (mAU)
150
100
50
0
0
20
40
60
Elution volume (ml)
100
80
60
40
20
0
% B
280 nm
260 nm
% buffer B

C


UV (mAU)
150
100
50
0
0
20
40
60
Elution volume (ml)
100
80
60
40
20
0
% B
280 nm
260 nm
% buffer B

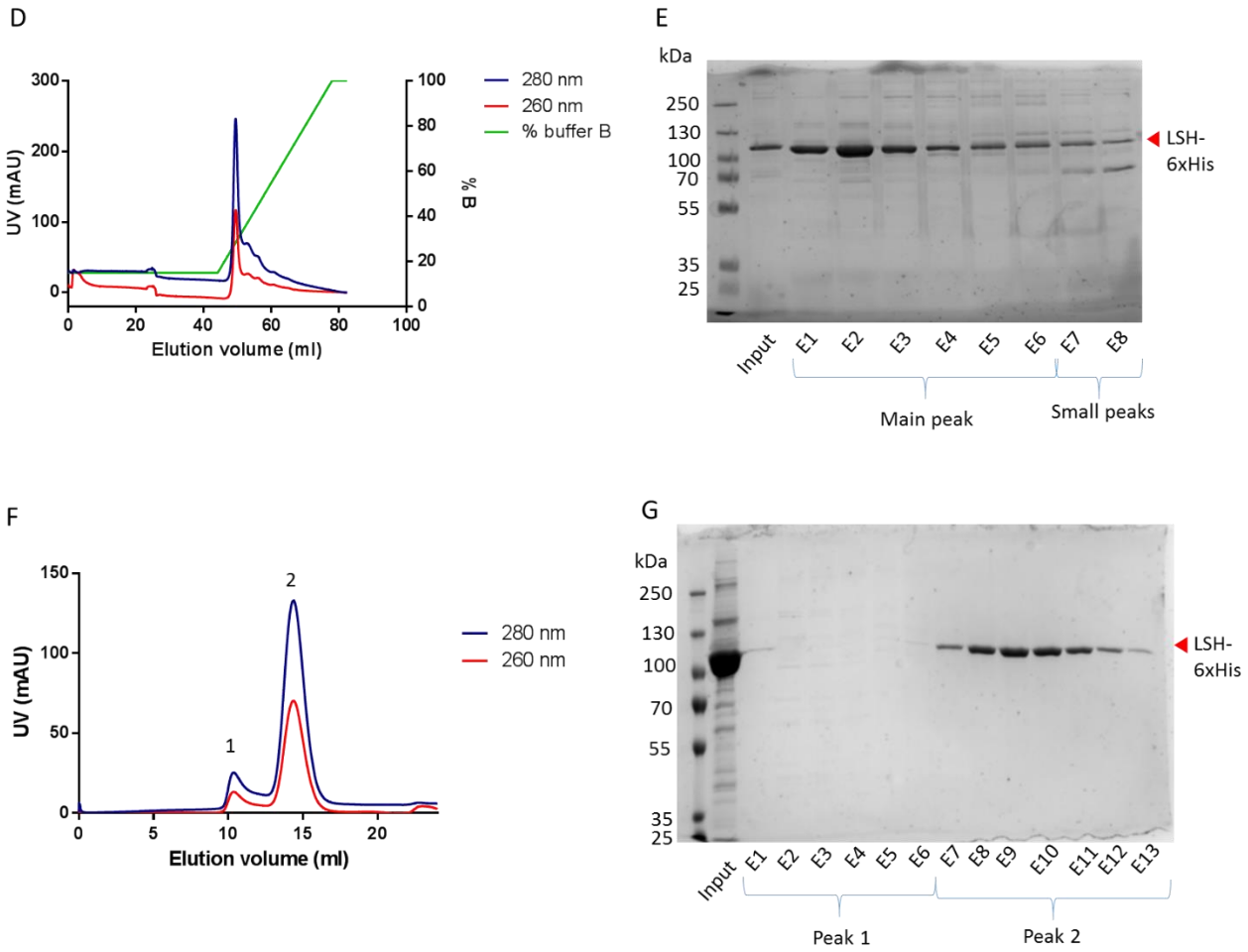


Figure 4.7 LSH purification using HiPrep 26/10 desalting, Resource S ion exchange and Superdex200 10/300 gel filtration columns.
A) Chromatogram of LSH applied on HiPrep 26/10 desalting column with UV detection (280 and 260 nm) and eluted with 250 mM NaCl-containing buffer.
B) Chromatogram of test application of LSH in 250 mM NaCl-containing buffer on Resource S column with UV detection (280 and 260 nm) and % of elution buffer B (1 M NaCl).
C) Chromatogram of test application of LSH in 150 mM NaCl-containing buffer applied on Resource S column with UV detection (280 and 260 nm) and % of elution buffer B (1 M NaCl).
D) Chromatogram of large-scale application of LSH in 150 mM NaCl-containing buffer applied on Resource S column with UV detection (280 and 260 nm) and % of elution buffer B (1 M NaCl).
E) Coomassie stained 12% SDS-PAGE gel of LSH after Resource S column. 2% of each sample volume loaded per well.
F) Chromatogram of LSH gel filtration on Superdex200 10/300 24 ml column with UV detection (280 and 260 nm) showing separation of impurities (peak 1) and LSH (peak 2).
G) Coomassie stained 12% SDS-PAGE gel of LSH after Superdex200 10/300 24 ml column. 2% of each sample volume loaded per well.

4.4 Purification of LSH^{K237Q}

LSH^{K237Q} is an ATPase-inactive mutant with a disrupted ATP binding site. The construct was previously cloned by Simon Varzandeh and was meant to be used in the nucleosome

remodelling assay replicated from Jenness et al. (Jenness et al., 2018). Since the mutant would not be used in any structural studies, I did not aim for the highest purity of final material and followed the same approach as for LSH WT but with minor alterations.

As the first step, I used the HiTrap IMAC FF column with imidazole washes as high as 40 mM (***Fig. 4.8 A***). That allowed to obtain a semi-pure protein upon elution (***Fig. 4.8 B***). That was followed by the HiPrep 26/10 desalting column to remove imidazole (***Fig. 4.8 C***) and, finally, Superdex200 10/300 (***Fig. 4.8 D, E***). The final protein was over 90% pure (***Fig. 4.8 E***) with a yield of 0.2 mg at 0.8 mg/ml.

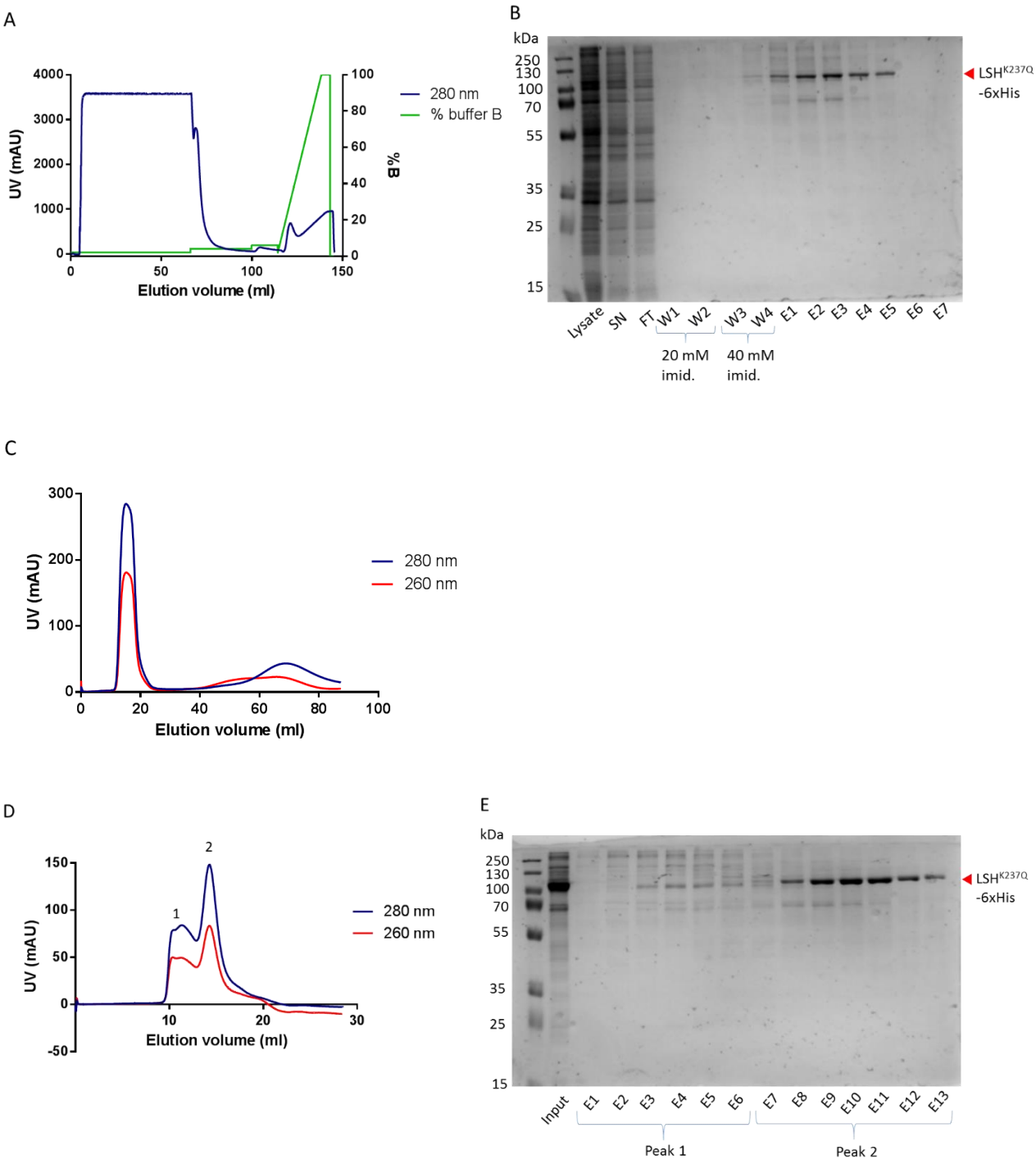


Figure 4.8 LSH^{K237Q} purification using HiTrap IMAC FF 1 ml, HiPrep 26/10 desalting and Superdex200 10/300 gel filtration columns.
A) Chromatogram of LSH^{K237Q} applied on HiTrap IMAC FF 1 ml column with UV detection (280 nm) and % of elution buffer B (500 mM imidazole).
B) Coomassie stained 12% SDS-PAGE gel of LSH after HiTrap IMAC FF 1 ml column. 2% of each sample volume loaded per well.
C) Chromatogram of LSH^{K237Q} applied on HiPrep 26/10 desalting column with UV detection (280 and 260 nm) and eluted with 500 mM NaCl-containing buffer.

D) Chromatogram of LSH^{K237Q} gel filtration on Superdex200 10/300 24 ml column with UV detection (280 and 260 nm) showing separation of impurities (peak 1) and LSH (peak 2).
E) Coomassie stained 12% SDS-PAGE gel of LSH^{K237Q} after Superdex200 10/300 24 ml column. 2% of each sample volume loaded per well.

4.5 Conclusions

Previously LSH has been successfully expressed using baculovirus expression system, however, purification process could be further optimised for potential high-resolution structural studies. LSH^{WT} and LSH^{K237Q} were successfully recloned into pFL vectors for insect cells expression and expressed in *Sf9* cells. I utilised chromatography columns for each purification step instead of incubation with resin (in this case, Ni-NTA), since it provided purity superior to resin. I also added the ion exchange step using Resource S column. Resource columns have smaller matrix beads with homogenous size, which allows to efficiently separate impurities. The final step of gel filtration resulted in LSH of over 95% purity, albeit at a low yield (200 µg), which still would be sufficient for a cryo-EM study. The yield could be improved by expressing LSH in HighFive insect cells, however, during the present expression work the cells were not available to me. Overall, murine LSH expression and purification has been optimised to a level adequate for structural work.

The ATPase inactive mutant LSH^{K237Q} was expressed in the similar fashion to LSH^{WT}, however, the Resource S step has been omitted, since the mutant was going to be used to recapitulate the remodelling assay performed by Jenness et al. (Jenness et al. 2018), and resulting 85-90% purity of the protein was sufficient for biochemical assays.

5. INTERACTION BETWEEN LSH AND THE MONONUCLEOSOME

5.1 Introduction

Chromatin remodellers exert their enzymatic functions by interacting with the chromatin units, nucleosomes, and free DNA and proteins, as in the case of Mot1. LSH was previously shown to interact with free DNA and a mononucleosome containing a DNA linker protruding from of the sides of the nucleosomal core particle (NCP), but not a mononucleosome without any linkers (see Simon Varzandeh's thesis), impying that LSH contains a domain that binds extranucleosomal DNA. Here, I wanted to test whether LSH binding to the nucleosome is dependent on the presence of DNA linkers or a type of a histone octamer. I used recombinant mononucleosomes made up of *X. laevis* histones, which were formed into histone octamers, and a synthetic DNA sequence Widom 601 (see Section 2.2.1.4 of Materials and methods). These recombinant nucleosomes are a standard system for investigating chromatin units and proteins that interact with them. Here, I used 601 DNA with linkers of various length to estimate what length is optimal for LSH binding. I also used different octamers – wild type, an octamer containing H3 with 27 aa truncation on the N-terminus and an octamer containing mutations in the acidic patch (see below) – to check if LSH interacts with the nucleosomal core.

5.2 Expression and purification of histones from *Xenopus laevis*

To make mononucleosomes, recombinant histones derived from frog *Xenopus laevis* were expressed in *E. coli,* isolated as part of inclusion bodies and purified by ion exchange chromatography (IEC) (***Fig. 5.1***). As histones are small proteins completely lacking tryptophan residues and having only a few tyrosine residues, it is challenging to accurately determine their concentration. This, however, is a critical step, since the success of octamer formation depends of the equimolarity of histones mixed. Hence, a range of methods, such as Bradford assay, measurement of absorbance at 280 nm and densitometry, was used. A truncated version of H3 histone (H3T lacking first 27 residues from the N-terminus that constitute the histone tail protruding from the nucleosomal core) was added to the histone set to be potentially used in experiments with DNMT3B – a putative LSH interaction partner that has been shown to bind the N-terminus of H3 (Rondelet et al., 2016). In the initial experimental plan, LSH-DNMT3B complex would have been assessed with and without the mononucleosome to check the complex stability.

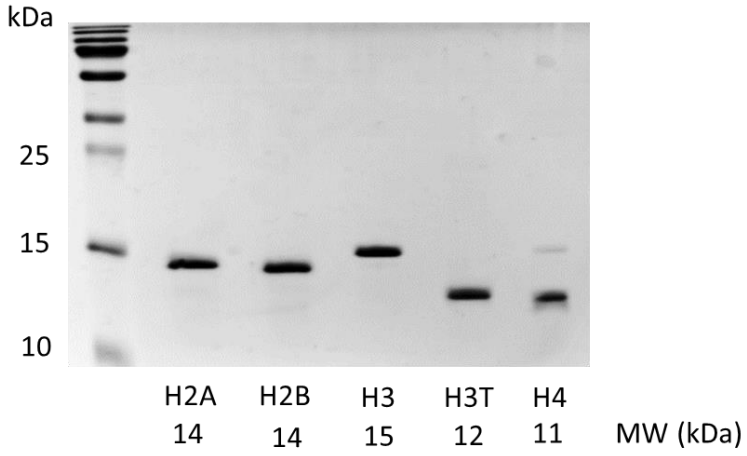


Figure 5.1 Purified histones form Xenopus laevis. **20% SDS-PAGE.**

5.3 Octamer formation and purification by gel filtration

To prepare octamers, histones were mixed at equimolar ratios, each at the concentration around1 mg/ml, with a slight excess of H2A and H2B to ensure successful reconstitution and dialysed into octamer refolding buffer (10 mM Tris pH 8, 1 mM EDTA, 2 M NaCl, 5 mM β-mercaptoethanol). During dialysis, histones transform from disordered aggregates to H2A-H2B dimers and H3-H4 tetramers and subsequently to histone octamers, which are stable at 2 M NaCl. To separate octamers from any aggregates or any smaller species that might have formed during reconstitution, gel filtration was used. The principle of gel filtration is based on separation of the species in a sample based on their hydrodynamic radius, which in turn affects their partitioning into the solid phase of the column's resin. Smaller and globular particles will partition into the pores more readily, compared with larger and more elongated particles, therefore, the former will be eluted later, i.e. at a higher elution volume. Hence, the elution volume depends on the size and shape of a molecule. A calibration curve based on globular proteins of known sizes is used to determine the elution profile for each column. From the previous experiments in the Voigt lab, it was known that octamers containing these particular histones have expected molecular mass of around 108 kDa and elute at 12.4 ml on Superdex200 24 ml column. In the present run, the main peak, supposedly containing histone octamers eluted at 12.8 ml (***Fig. 5.2A***), which indicates smaller molecular size compared to the run in the Voigt lab. However, the columns were physically different, and the present run

was done on a slightly compressed column, therefore, the total column volume was less than the manufacturer's specifications. Hence, it is possible to assume that the main peak is, indeed, the wild type histone octamer, which was subsequently confirmed by 20% SDS-PAGE, where each histone could be seen and all histones have comparable band intensity (***Fig. 5.3, lanes labelled oct.-WT and oct.-H3T***). An octamer previously purified in the lab was used as a control, however, as it turned out to contain a mixture of frog and human histones, the pattern of histone migration is slightly different (***Fig. 5.3 lane labelled oct.-H.s./X.l.***).

In the present run, a peak eluted in the void volume at around 8ml. The presence of aggregates indicated that either the dialysis conditions were suboptimal, e.g. transition into refolding buffer was not gradual enough, or, more likely, that the ratio of mixed histones was not equimolar, which is plausible, considering the difficulty of their accurate quantification. A smaller peak at ~15.2ml indicated the presence of smaller species, such as histone tetramers or dimers (***Fig. 5.2A***). The octamer containing the truncated version of H3 (predictably eluted later than the wild type octamer, at 13.1 ml (***Fig. 5.2 B***), reflecting the difference of approximately 5 kDa between the octamer types (108.8 kDa – wild type, 103.1 kDa – truncated version).

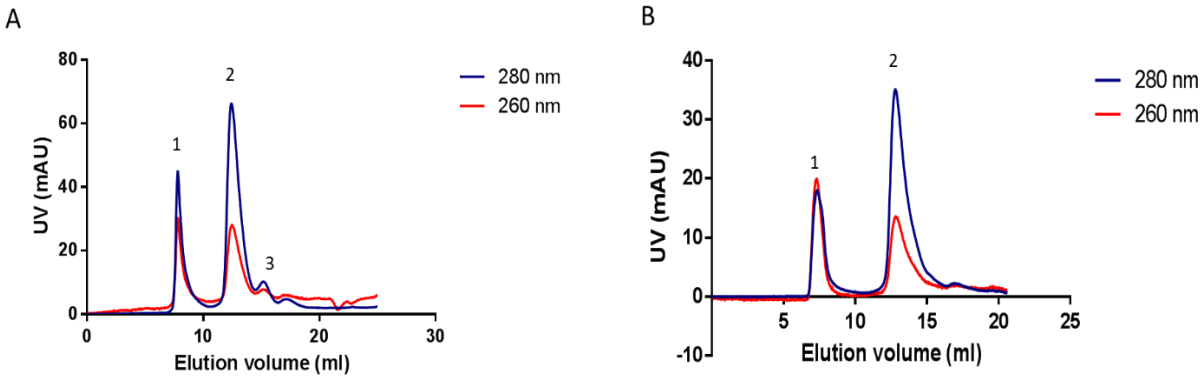


Figure 5.2 Size exclusion chromatography of histone octamers. Octamers containing wild type histones A) and N-terminally truncated H3 B) from Xenopus laevis were purified by isocratic elution with octamer refolding buffer. Peak 1 represents histone aggregates and peak 2 represents octamers, peak 3 represents smaller species, such as tetramers or dimers.

Table 5.1 Elution profile following SEC of octamers.

Peak #	Peak description	WT oct. elution V (ml)	H3T oct. elution V (ml)	Molecular mass (kDa)
1	Aggregates	8.2	7.9	-
2	Octamers	12.8	13.1	108/102
3	Tetramers or dimers	15.2	15	~60

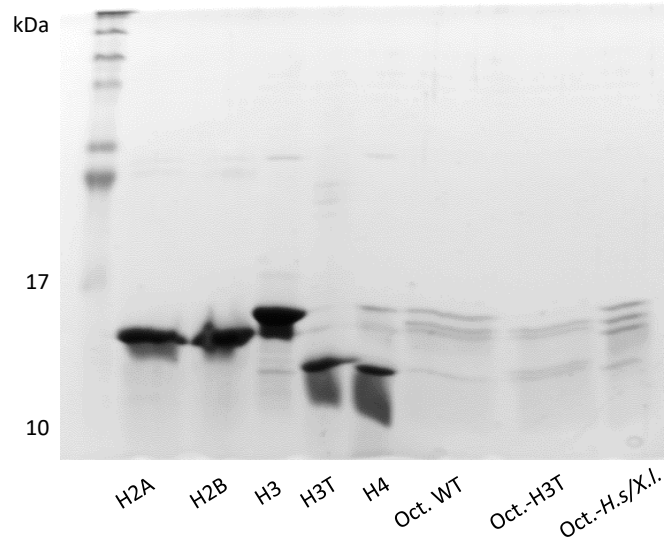


Figure 5.3 Purified histones from Xenopus laevis and reconstituted octamers containing wild type and N-terminally truncated H3. Samples were separated on 20% SDS-PAGE at 250 V for 120 min for enhanced resolution. Individual histones are labelled H2A, H2B, H3 and H4 for the wild types and H3T for the H3 truncation; Oct. WT – octamer with Xenopus laevis wild type histones; Oct.-H3T - octamer with Xenopus laevis truncated histone 3. Oct.-H.s/X.l. - octamer with a mixture of Homo Sapience and Xenopus laevis histones (made by Simon Varzandeh).

5.4 Nucleosome reconstitution and quantification

Being basic units of chromatin, nucleosomes are complexes of histone octamers and DNA wrapped around them. They are naturally present in the form of arrays, but to study a chromatin remodeller's mode of action, mononucleosomes are sufficient, since a remodeller-mononucleosome complex allows to capture the details of the interaction. To form a mononocleosome, a synthetically derived strong positioning DNA sequence termed 601 (Lowary and Widom, 1998) was amplified by PCR (***see Section 2.2.1.4***). Chromatin

remodellers' activity can be dependent on the presence of a DNA linker extending from the nucleosome core particle (NCP). LSH, the protein investigated in this study, has a poorly characterised mode of interaction with a mononucleosome, and it cannot be ruled out that it also requires a linker for efficient engagement with NCP. Previous experiments performed by Simon Varzandeh suggested that LSH requires at least 26 bp DNA linker for binding and it could not bind a nucleosome without any linkers.

Hence, 601 sequence was amplified with a range of shorter DNA linkers: 0W0 (601 Widom sequence without any linkers), 6W0 (601 with a 6bp linker on the 5'-end of the 601 sequence), 12W0 (601 with a 12bp linker on the 5'-end of the 601 sequence) and 25W0 (601 with a 25bp linker on the 5'-end of the 601 sequence) (see Section 2.2.1.3 of Materials and methods for oligo sequences). The reverse primer amplifying 3' of the 601 sequence without a DNA linker was conjugated to IR800 fluorophore. It was ensured that the linkers did not contain any palindromic restriction sites in their sequences, as it can result in secondary structure formation in the linker sequence and, therefore, artefactual results.

Successful nucleosome reconstitution results in formation of mononucleosomes with minimal amounts of free DNA or octamers. The optimal ratios were determined empirically for each new DNA and octamer preparation, for example, the 25W0 fragment was titrated with the fixed amount of the octamer to determine the optimal reconstitution ratio (***Fig. 5.4 A***). Reconstituted mononucleosomes were run alongside the free DNA fragment to verify successful reconstitution by the shift on a native TBE gel (***Fig. 5.4 B***). Generally, in my octamer reconstitution set-ups the DNA-octamer ratios of 0.5:1 or 0.7:1 resulted in minimal amounts of free DNA or aggregated octamers.

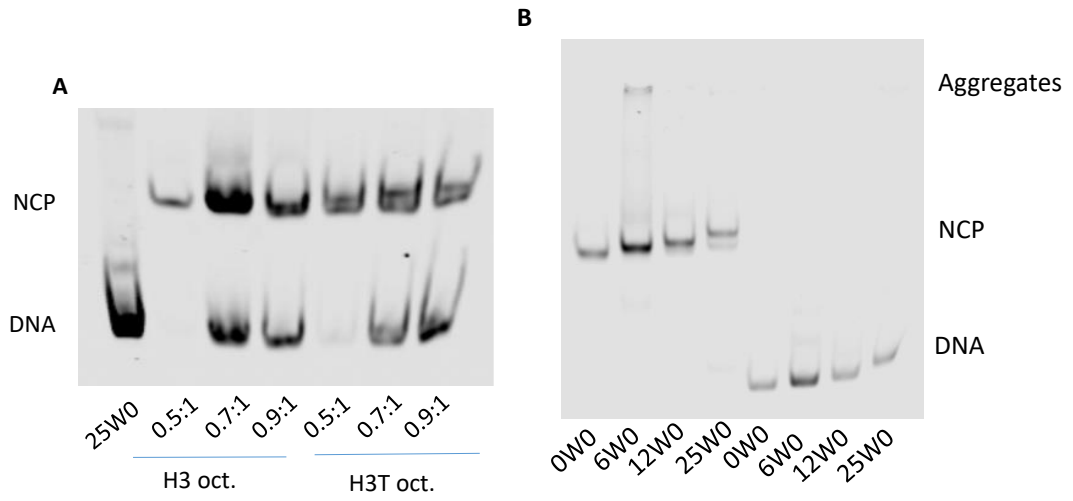


Figure 5.4 Nucleosome reconstituted with X.l. octamers and the 601 sequence with various linkers resolved on 6% TBE native PAGE and visualised as an IR800 signal on Licor Osyssey scanner. A) Titration of 25W0 DNA fragment with the wild type and H3T octamers to determine the optimal reconstitution ratio; molar ratios are shown. B) DNA fragments with various linker lengths mixed with the wild type octamers at 0.5:1 molar ratio. NCP – nucleosome core particle.

It is worth noting that the 601 sequence is known for its uniform positioning of the octamer (Lowary and Widom, 1998). Other sequences, such as alpha-satellite DNA or 5S RNA, may wrap around the octamer in a variety of ways, which results in different nucleosomal species visible on a gel (Luger et al., 1999). When different nucleosomal species are present due to non-uniformal positioning, a heating step helps to obtain molecules with uniform DNA positioning. This problem should not occur with the 601 sequence, since it is a synthetically generated fragment with enhanced affinity for the octamer, compared with naturally occurring sequences. However, in the present study, nucleosome reconstitution resulted in two bands present when H3T-containing octamer was used (***Fig. 5.4 A, lanes H3T oct***.). Similarly, an extra band was present after 25W0 NCP WT octamer reconstitution (***Fig. 5.4 B, lane 25W0 NCP***), however, this pattern was not observed in subsequent reconstitutions, even though the preparation method was the same. Taq polymerase, which was used to make Widom 601 DNA, has relatively low proofreading activity may introduce errors, however, it is unlikely for small fragments, such as the ones used in this study. Nevertheless, potential mutations may weaken the affinity of the DNA fragment for the octamer, resulting in non-uniform positioning.

Nucleosome quantification is complicated by the fact that the molecules contain both DNA and proteins, therefore, it was rationalised that A_{260} will give a mixture of DNA and protein absorbance and will not be accurate, therefore, I used three approaches to NCP

quantification: measurement at A260 by Nanodrop, scanning of fluorescence signal and DNA staining with SYBR Safe. Since all the DNA fragments used here (0W0, 6W0, 12W0 and 25W0) contained the same 3' fluorescent IR800 label, it was expected that the concentration of NCPs would be directly proportional to the fluorescence intensity. To check that, all four DNA fragments of the same concentration determined by Nanodrop were separated on a 6% TBE native PAGE alongside a standard curve based on 6W0 samples of known concentrations and NCPs of unknown concentrations and visualised either using IR800 fluorescent signal and SYBR Safe (***Fig. 5.5 A and 5B***, respectively) and the standard curve based on 6W0 SYBR Safe staining signal (***Fig. 5.5 C***).

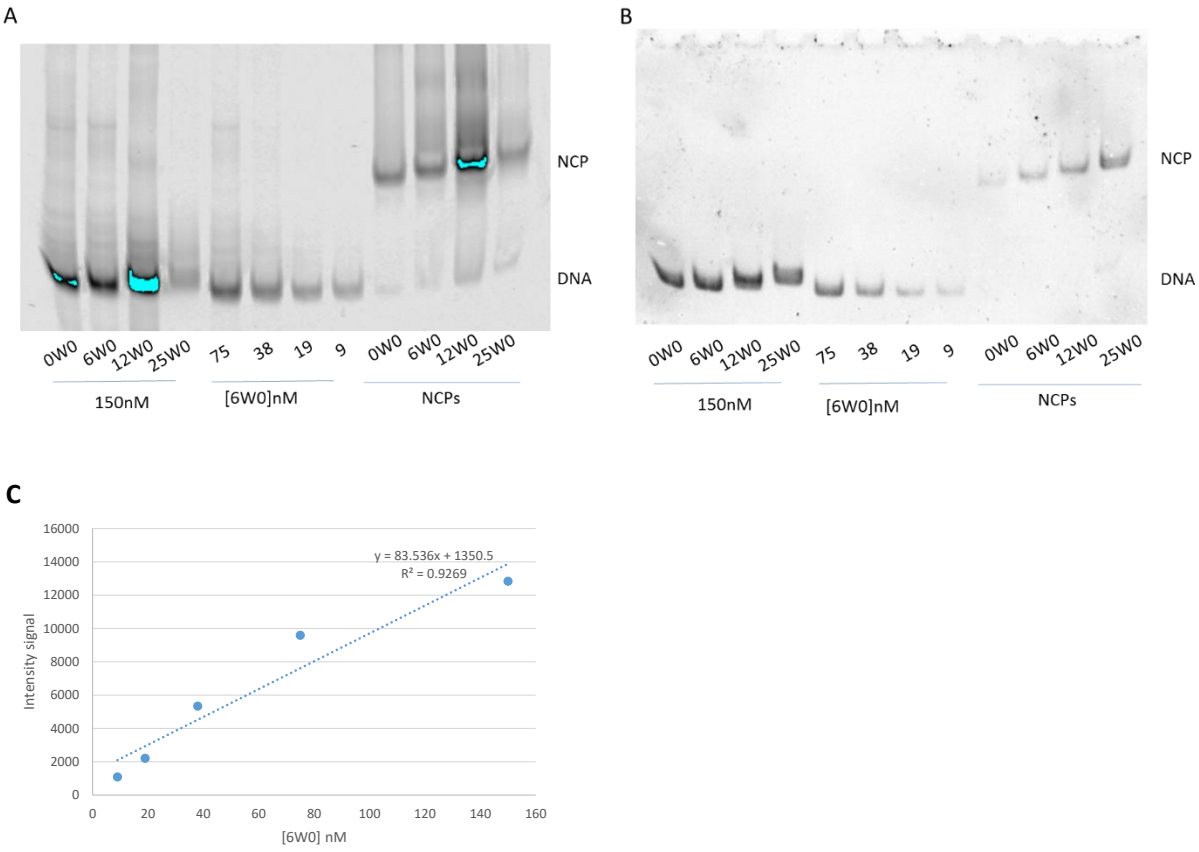


Figure 5.5 Evaluation of NCP reconstitution. NCPs were visualised on a 6% TBE native PAGE using A) IR800 fluorescence intensity or B) SYBR Safe staining. C) Standard curve was generated from 6W0 titration visualised by SYBR Safe.

The comparison of the visualisation and quantification methods demonstrated that fluorescence intensity was not directly proportional to the nucleosome concentration but SYBR Safe staining was more consistent with Nanodrop measurements (***Fig. 5.5 A and B***, respectively). Predictably, dilution of the 6W0 DNA fragment (***Fig. 5.5***, ***from 75 down to 9***

nM) and staining with SYBR Safe generated a nearly linear standard curve (***Fig. 5.5 C***), indicating correct dilution, which did not translate into the IR800 signal (***Fig. 5.5 A***). This result demonstrated that fluorescence intensity should be evaluated cautiously, and Nanodrop quantification was selected as more accurate and quick.

Measurement of the nucleosomes at 260nm on a Nanodrop spectrophotometer demonstrated $A_{260/280}$ ratio of 1.8, which corresponds either to a DNA sample free of protein contaminants or to a DNA sample containing proteins lacking aromatic groups, which are the main source of absorbance at 280 nm. Histones have a few tyrosine residues that give a weaker absorbance at 280nm and a few phenylalanines that absorb very weakly at 260nm. Hence, histones were not anticipated to interfere with DNA absorbance, and its estimated concentration by Nanodrop divided by the MW of the DNA fragment was used to estimate molarity of each nucleosome.

5.5 LSH binding to the NCP was not affected by various DNA linker present on the NCP

It has previously been established that LSH can bind both free DNA and a nucleosome (Simon Varzandeh, 2016). However, from the previous experiments, it was not clear whether LSH requires a linker extending from the nucleosome core particle (NCP). Previous electrophoretic mobility shift assays (EMSA) demonstrated conflicting findings, wherein an NCP lacking a linker (0W0) could not be shifted by increasing concentrations of LSH in some experiments (Kevin Bryan Myant, 2008; Simon Varzandeh, 2016) but was shifted in other set of experiments (Laura FitzPatrick's MSc report, University of Edinburgh). It should be noted that S. Varzandeh used an octamer that contained a mixture of human and frog histones, however, the human histones are highly homologous to the frog histones and the differences in the octamers are unlikely to account for the difference in the shift. In attempt to resolve this discrepancy and to establish whether LSH required linker DNA to bind an NCP, EMSA was performed in triplicate with fixed amount of the nucleosome containing four different DNA linkers used in this study and increasing amounts of LSH.

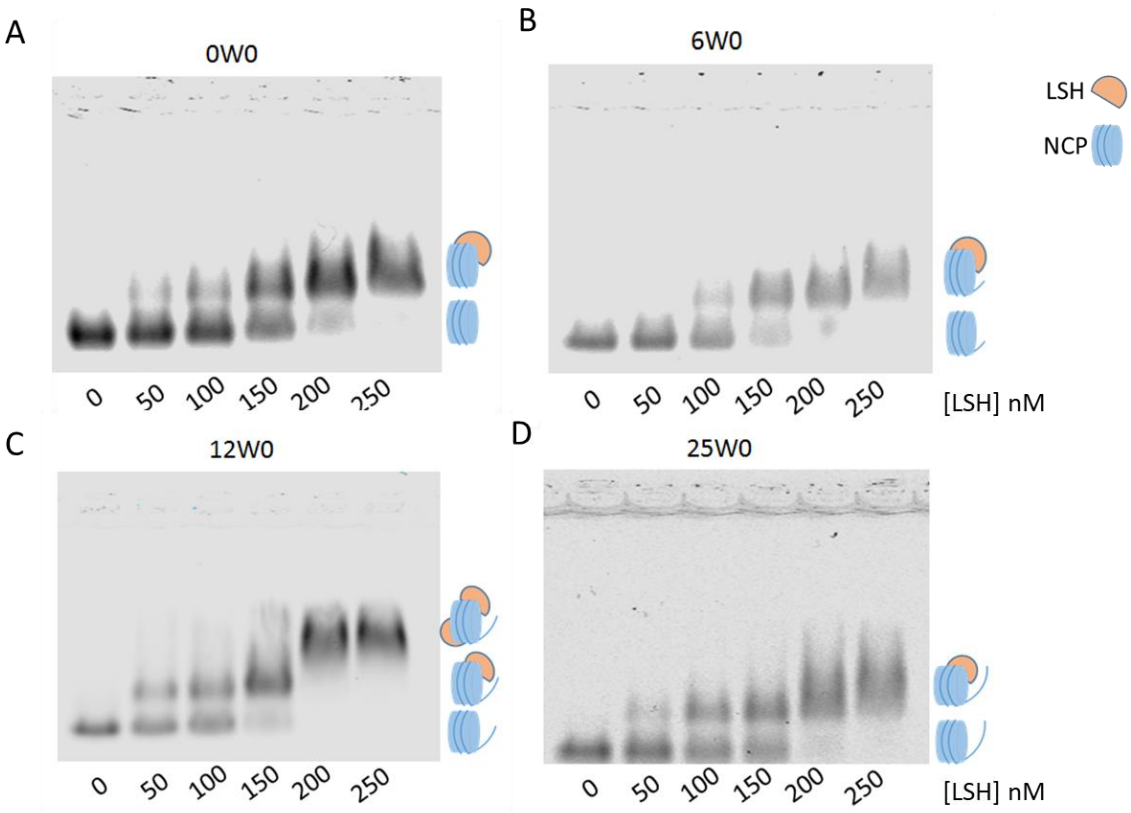


Figure 5.6 EMSA with 10nM of NCP with various linkers and increasing concentrations of LSH. Shift assay on 0.5% agarose gel visualised by IR800 fluorescence intensity for the nucleosomes containing DNA fragments A) 0W0; B) 6W0; C) 12W0; D) 25W0.

From the EMSA experiments in this study, it appears that LSH does not require a DNA linker to bind to a nucleosome, as indicated by the shift of the 0W0 NCP. A shift occurs with all fragments used in this study, however, the 12W0 NCP demonstrated a clear double shift (***Fig. 5.6 C***), which may indicate either oligomerisation of LSH or a more likely possibility that it binds NCP at the two locations – super-helical location 2 (SHL2) and super-helical location -2 (SHL-2) (***see Section 2.1***). There may be a slight double shift with the other NCP fragments, however, it was hard to distinguish a clear double shift from a smear (***Fig. 5.6 A, B, D***). However, it is unlikely that LSH binds 12W0 in a way that differs from the other NCPs in this experiment. If LSH, like the other chromatin remodellers, binds through its ATPase lobes at the SHL2 and SHL-2, this should be done irrespective of the linker DNA.

This is not entirely unpredictable as chromatin remodellers have a propensity to bind an NCP at SHL2, which ensures an adequate placement of mechanical force for DNA translocation along the octamer. This mechanism was observed in the mode of action of Snf2, the closest

LSH homologue from yeast (Liu et al., 2017b). Snf2 was also shown to interact with an NCP at SHL6, located roughly opposite SHL2 (Liu et al., 2017b). That would explain the super-shift clearly observed in the case of 12W0 NCP-LSH interaction (***Fig. 5.6 A***). Other remodellers, such a CHD1, which has two DNA-binding domains, have a higher affinity for an NCP in the presence of DNA linker (Nodelman et al., 2017). It cannot be ruled out that LSH interacts both with a nucleosomal DNA and a linker DNA, perhaps, recognising the linker DNA through the same mechanism as free DNA. Nevertheless, the exact mode of interaction between LSH and an NCP cannot be elucidated solely based on EMSA experiments.

5.6 LSH-NCP complex with free DNA and nucleosomal mutants

Next, I wanted to check if the octamer itself influences LSH interaction with the NCP. To do that, various types of free DNA (a short double-stranded 12-mer fragment, a double-stranded 159-mer fragment) and the nucleosome (containing 0W0 or 12W0 DNA in a complex with the wild type octamer (WT oct.), H3 N-terminal 27 aa truncation (H3T oct.) and an acidic patch mutant (AP-M) octamers) were used. The acidic patch mutant contained the following alanine substitutions: H2A (E65A E90A E91A) and H2B (E105A). Even though LSH is not expected to interact with H3 N-terminal tail, as seen from homology modelling (see Section 3.1 of Results). However, I decided to test this octamer variant to gain experimental data regarding the effects of H3 N-terminus truncation on LSH-NCP interaction and to substantiate LSH-NCP homology modelling.

Comparing the short DNA fragment with a long one would show if LSH binds DNA in a particular location or if it just recognises double-stranded DNA nonspecifically. The octamer variants would elucidate whether LSH interacts with the octamer core during its interaction with the NCP. As indicated from the previous experiments performed by Simon Varzandeh (Simon Varzandeh, 2016), LSH binds free DNA, therefore, it is not clear whether its shift of the NCP is due to the recognition of the nucleosomal fold or simply the recognition of the double-stranded DNA.

Preparation of Nucleosome core particles containing variant histones proceeded as previously: I performed initial octamer titration into 100 nM of 601 DNA to obtain the appropriate reconstitution ratios (***Fig. 5.7 A, B***). For all of the variants, After almost no free DNA was present at 1:2 DNA:octamer molar ratio (***Fig. 5.7 A, B***). Then, using 1:2 ratio obtained from the test reconstitution, I performed a scale up to 500 nM 601 DNA (***Fig. 5.7***

C). Double-stranded oligonucleotide 12-mer was obtained by annealing two ss12-mer fragments (IR800-F-12mer-non-hairpin: GTTGCGTGCTTT, R-12mer-non-hairpin: AAAGCACGCAAC) (***Fig. 5.7 D***). All samples were resolved on a native PAGE.

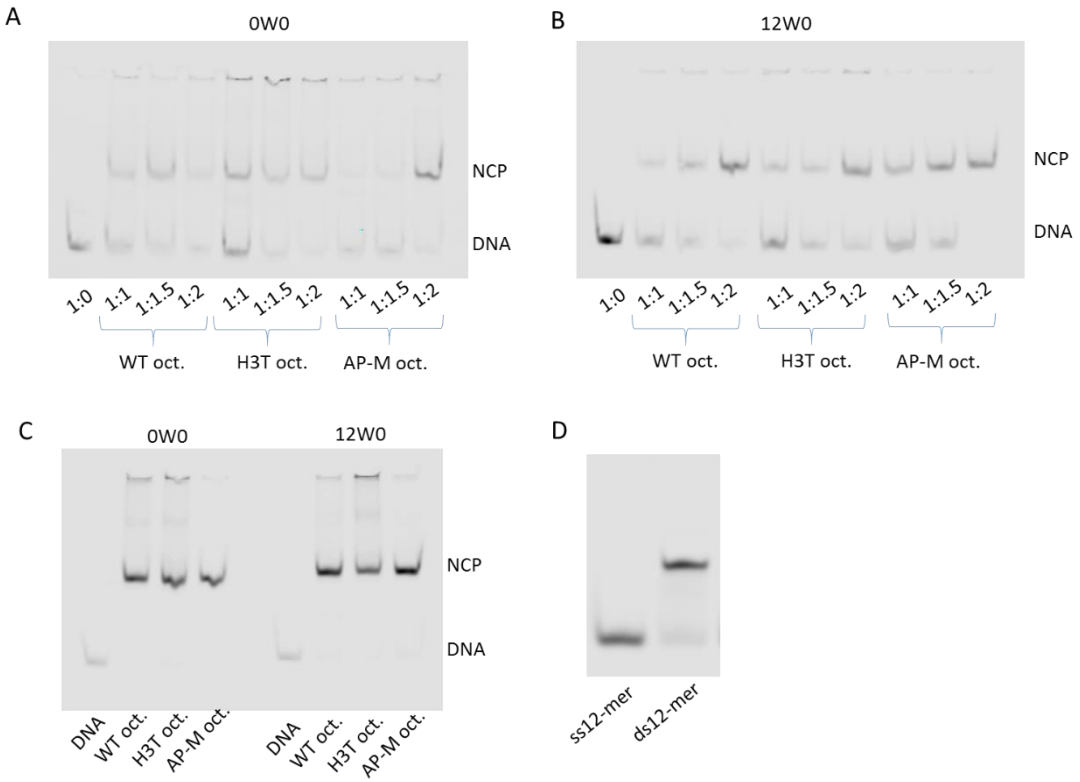


Figure 5.7 Reconstitution of various nucleosome core particles (NCPs) containing the wild type (WT) histone octamer, the octamer with a truncated H3 variant (H3T) and an acidic patch mutant octamer (AP-M) used for EMSAs. A) Test reconstitution of 0W0 601 DNA fragment with different octamers, ratios are DNA:oct.; B) Test reconstitution of 12W0 601 DNA fragment with different octamers, ratios are DNA:oct.; C) Scale up (100 µl) of successful (1:2 DNA:oct. in all octamer variants) reconstitution reactions. D) Annealing of 12-mer.

Following reconstitutions, validation and quantification of the DNA fragments (ds12-mer and ds159-mer) and NCPs (0W0 and 12W0 with WT, H3T and AP-M octamers), they were incubated with increasing amounts of LSH and binding was analysed by EMSA (***Fig. 5.8***). Incubation of LSH with the ds12-mer resulted in a shift, confirming that LSH is capable of recognising and binding short DNA fragments (***Fig. 5.8 top***). The shifts at 100 and 150 nM of LSH may represent a 1:1 stoichiometry and a relatively small complex, however, at 200 and 250 nM LSH the complex appeared larger and some of it was inhibited in wells (***Fig. 5.8 top***), which may be due to LSH aggregation or oligomerisation.

Comparison of ds12-mer oligo to the ds159-mer fragment demonstrated a similar pattern, however, it appeared that LSH was more efficient in shifting the longer DNA fragment (**Fig. *5.8 top***). This can be due to the presence of some ssDNA in the ds12-mer sample. LSH has not been shown to bind ssDNA, therefore, some residual ssDNA may result in the smear on the left hand side gel but not on the middle gel (***Fig. 5.8 top***). The shift pattern of ds159-mer and 12W0-NCP-WT appeared similar (***Fig. 5.8 top, middle, right***), indicating that the presence of dsDNA rather than NCP folds is the key for LSH binding.

When LSH was incubated with the NCPs containing 0W0 DNA and different octamers (WT, H3T or AP-M), all the nucleosomal variants were shifted (***Fig. 5.8 B***), demonstrating that difference in octamers does not affect LSH binding to different 0W0 NCPs. There is a double shift similar to the one that was clearly observed after incubation with 12W0 NCP (***Fig. 5.6 C***), which reinforces the notion that LSH preferentially binds at two nucleosomal locations (most likely SHL-2 and SHL2), regardless of the presence of DNA linker. The fact that in the previous EMSA experiment it was not clearly observed may be due to the quality of the gel in the previous experiment (***Fig. 5.6 A***). The smeary pattern may also indicate that there is a variety of locations along the nucoleosome where LSH can bind. There is an apparent slight reduction in affinity with acidic patch mutations (***Fig. 5.8 B***), with the shift occurring at slightly higher LSH concentrations (250 nM LSH was enough to completely shift 0W0 NCP-WT whereas some NCP remained unshifted at 250 nM LSH with 0W0 NCP-AP-M, ***Fig. 5.8 B, C***), which may indicate that LSH uses the regions flanking the ATPase domain to interact with the octamer. This pattern was also observed with the 12W0 NCP fragments (***Fig. 5.8. C***), further reinforcing the hypothesis that LSH does not rely on any DNA linkers for binding the nucleosome and interacts only with the nucleosomal core.

The nucleosomes containing 12W0 DNA and different octamers (WT, H3T and AP-M) were also incubated with LSH, which resulted in a similar result to that of 0W0-containing nucleosomes. As previously, a step-wise shift was observed (***Fig. 5.8 C***), hinting at LSH binding at two different locations. Similarly to the 0W0-contaning NCPs, acidic patch mutant appeared to reduce LSH affinity for the NCP, with the WT NCP being completely shifted by 200 nM LSH (***Fig. 5.8C left***), whereas the AP-M NCP was shifted by 250 nM LSH (***Fig. 5.8C right***).

H3T octamer did not appear to make a drastic change to LSH binding for the NCP containing either 0W0 or 12W0 DNA (***Fig. 5.8 B, C***), but there may be a slight increase in affinity, since

the bands of the shifted NCPs at 200 and 250 nM of LSH appear more compact compared to the WT NCPs, however, like with the wild type, the shift occurred at 200 nM of LSH, indicating that the difference in band appearance may be due to the resulting shape of the nucleosome contributing to the migration pattern, rather than due to the difference in affinity (***Fig. 5.8 B left and middle, 5.8 C left and middle***). The nucleosomes containing acidic patch mutated octamers appeared distinct from the wild type and H3 histone truncated octamers (***Fig. 5.8 B, C***), with the shifted bands appearing more smeary than the wild type. Additionally, the complete shift did not seem to occur even at the highest concentration of LSH, 250 nM, indicating that the patch may play a role in LSH interaction with the nucleosome. One of the terminal regions of the protein may form interactions with this nucleosomal area, however, their exact orientation could not be deduced from homology modelling (see Section 3.2) and chemical crosslinking experiment did not capture any obvious intermolecular protein-protein interactions (see below).

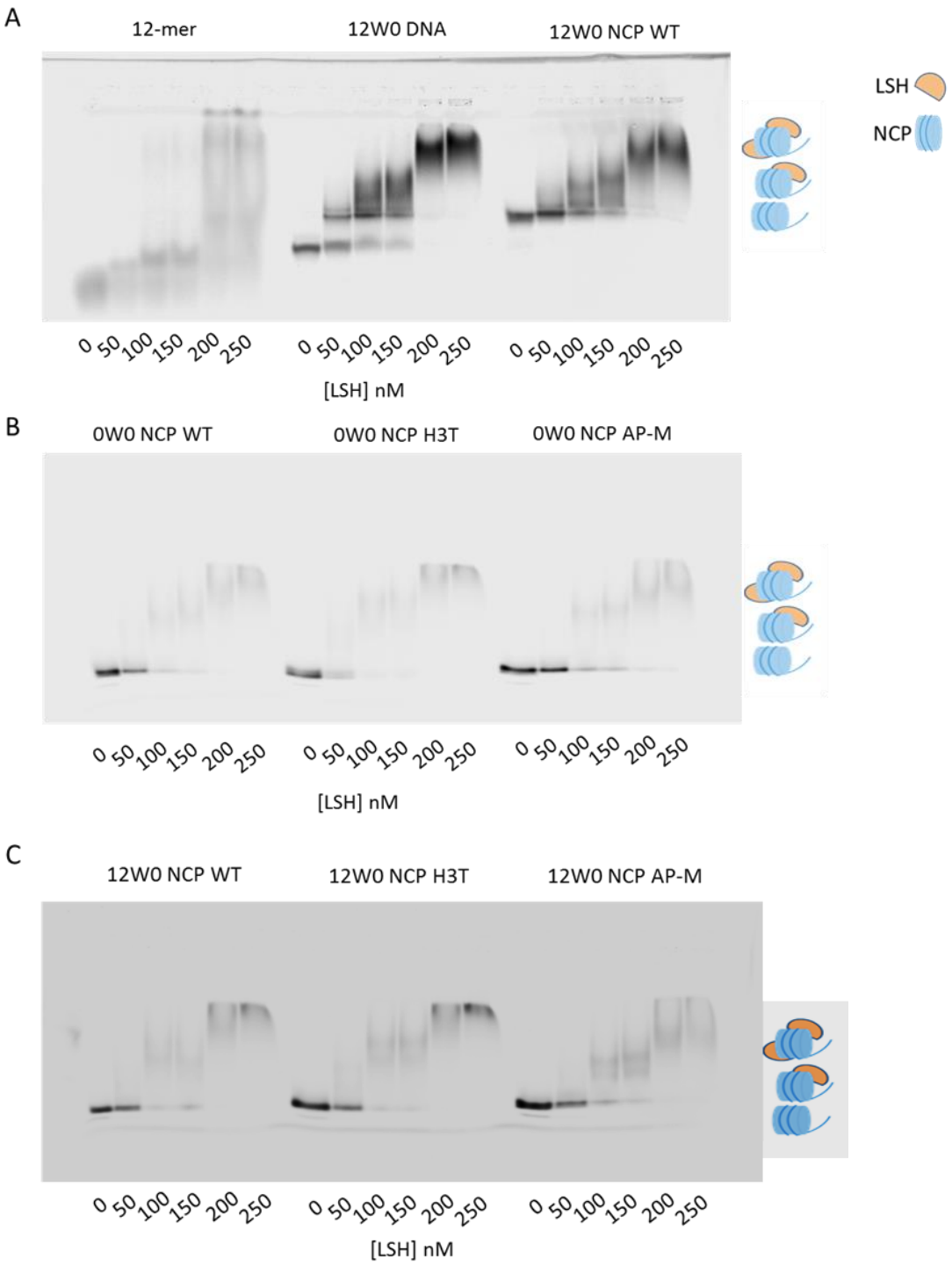


Figure 5.8 EMSA with various DNA fragments and NCPs and LSH titration. A) LSH titrated with 10 nM 12-mer DNA fragment, 159-mer (12W0) DNA fragment and 10nM 12W0 NCP. B) 10 nM of 0W0 NCP with indicated octamers (WT - wild type, H3T – N-terminally truncated H3, AP-M - acidic patch mutant) with LSH titration; C) 10 nM of 12W0 NCP with indicated octamers (WT - wild type, H3T – N-terminally truncated H3, AP-M - acidic patch mutant) with LSH titration. 0.5% agarose gel, Tris-CAPS buffer pH 9.4 run at 4°C and visualised through IR800 fluorescent signal.

It is known that LSH recognises acetylated chromatin and histone deacetylation through inhibition of HDAC by trichostatin A (TCA) abolishes LSH association with it (Yan et al., 2003), indicating that acetylation is a requirement for LSH binding to chromatin in cells. Recombinant NCPs used in the present experiments do not have post-translational modifications (PTMs) and LSH is still capable of binding them indicating that here LSH relies only on recognition of NCP itself rather than any histone tail modifications. To check whether LSH has an altered affinity to NCPs containing recombinant octamers compared to the native ones, I reconstituted NCPs with native octamers isolated from HeLa cells (native octamers were isolated and purified by Dr. Alba Abad, Arulanandam lab). These NCPs were used to perform an EMSA experiment to compare the shift patterns between the NCPs containing recombinant and native histone to determine whether PTMs influence LSH binding *in vitro*. and performed EMSA.

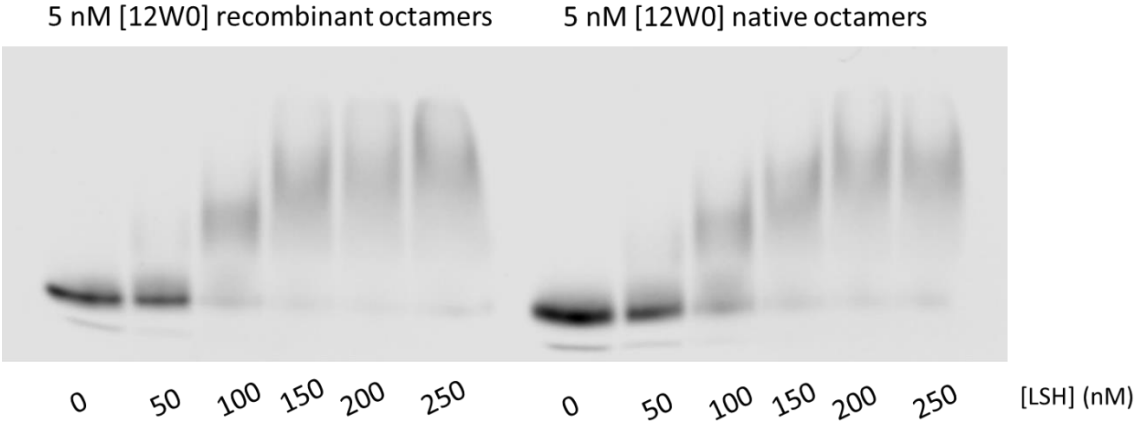


Figure 5.9 EMSA experiment of LSH binding to NCPs containing 12W0 601 DNA fragment and recombinant or native octamers. 0.5% agarose gel, Tris-CAPS buffer pH 9.4 run at 4°C.

The presence of any PTMs on HeLa cell histones did not appear to improve LSH affinity to the NCPs *in vitro* and even make the affinity slightly lower (***Fig. 5.9, recombinant vs. native octamers***). This may be due to the presence of different PTMs on the histone tails since they were isolated from asynchronous cells. A clearer comparison would be possible by chemically adding specific PTMs, namely acetylation, on the tails of recombinant histones, which might have improved affinity between NCPs and LSH, however, this assumption was not tested in the present study. The H3T-containing octamer could be used to add an N-terminal tail with defined modifications.

5.7 NCP-LSH complex was not detected by SEC-MALS at room temperature

To characterise the LSH-NCP complex detected by EMSA, size-exclusion chromatography-multiple angle light-scattering (SEC-MALS) was then used. SEC-MALS is a technique that allows separation of molecules based on its hydrodynamic radius and measurement of their molecular masses. It is more reliable for protein and complex characterisation than standard gel filtration, since the results from a gel filtration run rely on comparison of the protein's elution profile with a standard set of that consists of mostly globular proteins, such as aprotinin, ovalbumin and conalbumin (https://www.gelifesciences.co.jp/catalog/pdf/28407384.pdf). Gel filtration relying on a globular protein standard can provide some insight but may be misleading since the protein migration through the matrix also depends on the protein shape or conformation. Elongated proteins are less likely to partition into the pores, hence, they elute earlier than would a globular protein, and appear larger. SEC coupled with MALS allows to overcome this caveat and to obtain the absolute molecular weight of a molecule, without relying on calibration, by passing the protein of interest through a chamber that is illuminated by a laser (Meier and Heinzmann, 2017) (***Fig. 5.10***). Protein concentration is determined by the refractive index (RI) detection, which allows determining molecular weight across the sample peak (Wyatt, 1993). Once the laser has passed through the sample, the protein absorbs the energy from the incident beam. The resulting intensity of light scattered by the protein is then captured by detectors surrounding the chamber. The light scattering allows to determine the dimensions of the protein (or any other macromolecule) through the mean square radius of gyration – the average squared distance to a point in the object from the centre of its mass.

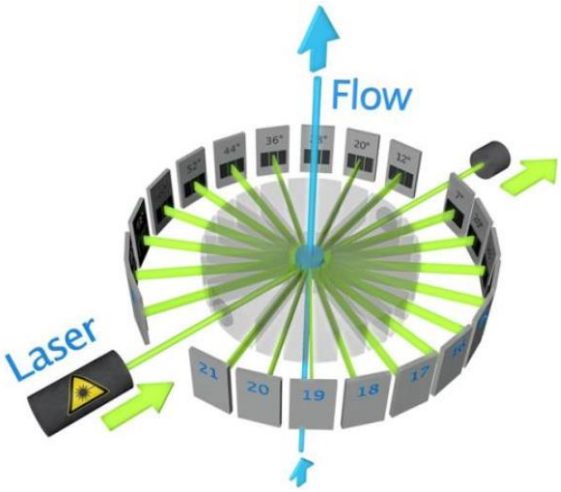


Figure 5.10 Schematic representation of a MALS detector. Grey rectangles represent 21 detectors located at the angles according to the number on each detector. Flow represent the sample flow following separation by gel filtration. Laser passes through the sample resulting in energy scattering, this energy is then captured by the detectors. Figure from Meier and Heinzmann 2017.

All these advantages make SEC-MALS a useful technique for determining not only the size of the components of a potential complex, but also allow determining the absolute mass of the complex, thus, elucidating its stoichiometric characteristics.

A few things should be taken into consideration when comparing analysis of LSH-NCP complex formation by EMSA and by SEC-MALS. First, even though both experiments were performed at 150 mM NaCl, the EMSA experiments (***Fig. 5.6 and 5.8***) were performed at 4^{o}C but the SEC-MALS instrument set up was only possible at the room temperature. Increase in temperature is a factor affecting complex stability. To test if the complex can survive those altered conditions, I repeated the EMSA analysis of LSH-NCP complex formation at room temperature and at two NaCl concentrations – 150 mM and 250 mM (***Fig. 5.11***). It appeared that the increase in salt slightly changed the affinity of the components and resulted in incomplete shift of NCP as well as the increase in temperature affected affinity of LSH for the NCP (***compare with Fig. 5.8C***), however, the differences between the experiments at 4^{o}C and room temperature may also be due to experimental variability.

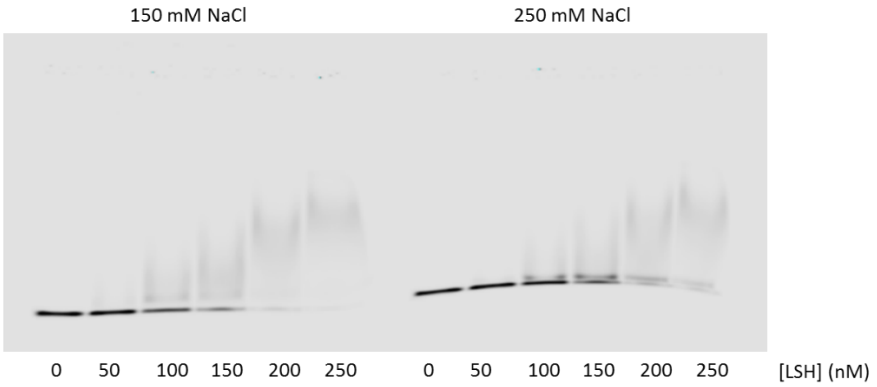


Figure 5.11 EMSA experiment with 10 nM 12W0 NCP and increasing concentrations of LSH. 0.5% agarose gel, Tris-CAPS buffer pH 9.4 run at room temperature.

Second, the amount of material used in the SEC-MALS study differed from that in EMSAs. In the case of EMSAs, the amount of NCPs was 10 nM and the highest LSH concentration was 250 nM, making the ratio to shift the NCP 1:25 NCP to LSH. As was visually estimated from the EMSA experiments, the qualitative K_d for the LSH-NCP interaction is around 100 nM LSH (***Fig. 5.11***) and the highest amount of the complex in the EMSA set up would be 10 nM, however, this amount would not be detected by the instrument. The purified LSH stock (prepared by Simon Varzandeh) was 10 mg/ml, or 104 µM. This high concentration would allow to dilute LSH to a desired concentration, however, scaling up NCP proved to be more problematic, and typically the highest NCP concentration did not exceed 10 µM. Therefore, the highest NCP concentration that could be injected was around 5 µM. Since the samples undergo further dilution of around 14-fold from the initially injected concentrations (***Table 5.2***) while migrating through the column, the complex that might have been formed could undergo partial or complete disintegration. The limiting component of the complex is the NCP concentration, therefore, the highest amount of the complex would be 5 µM at the present NCP molarity. If the complex has a 1:1 or 1:2 stoichiometry of NCP to LSH and the estimated K_d is around 100 nM, injecting 5 µM NCP and 26 µM of LSH would allow to detect the complex. This amount of LSH was chosen due to an attempt to inject as much LSH as possible within the 100 µl of injected volume to capture the complex formation.

When the components (5 µM 12W0 NCP and 26 µM LSH) were applied on Superdex200 24 ml column separately, their molecular masses measured by the instrument were close to their expected molecular masses (observed 205 kDa for NCP and 99 kDa for LSH, Table 4.2)

(***Fig. 5.12 A, B***). NCP 12W0 eluted as a single peak at 11.3 ml and with 80% recovery (MALS data, not shown), indicating that 150 mM NaCl was sufficient to disrupt any potential non-specific interaction with the column matrix (**Fig. 5.12 A**). LSH (***Fig. 5.12 B***), on the other hand, displayed a poor recovery of only around 28% (MALS data), which was unexpected since previously this non-specific interaction had not been observed. This may be due to the lower salt concentration in this experiment compared to previous LSH runs on Superdex 200, since previously all LSH purification steps were performed at 500 mM NaCl. There was an additional peak that eluted after LSH (***Fig. 5.12 B***) but it was outside the column volume indicating that the particle was very small. It was not possible to investigate what it was, since fractions could not be collected on this instrument.

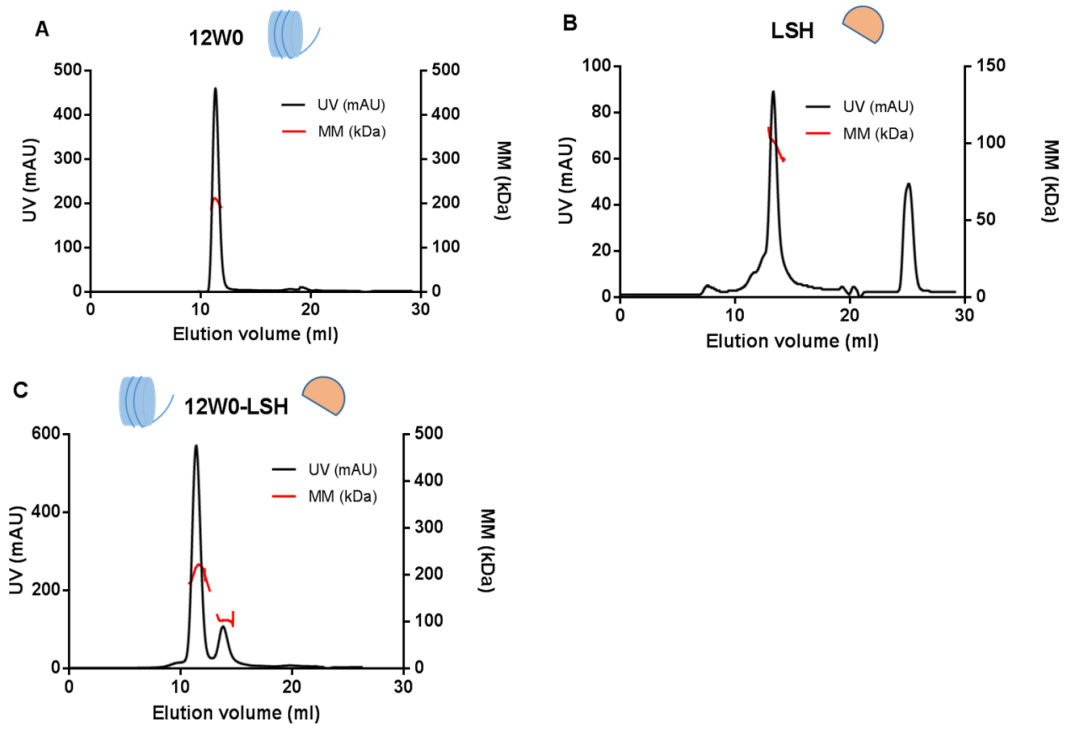


Figure 5.12 12W0 NCP-LSH SEC-MALS chromatograms. A) Elution of 12W0 NCP with absorbance at 280 nm and molecular mass indicated on the left and right hand Y axes respectively; B) Elution of LSH with absorbance and molecular mass indicated on the left and right hand Y axes respectively; C) 12W0 NCP and LSH following their mixture and a 30 min incubation at 4°C with absorbance and molecular masses.

Table 5.2 Characteristics and parameters of the 12W0 NCP-LSH SEC-MALS run.

	12W0 NCP	LSH	12W0 NCP and LSH together
Injected amount	0.5 mg/ml (5 µM)	2.5 mg/ml (26 µM)	0.5 mg/ml (5 µM), 2.5 mg/ml (26 µM)
Elution volume (ml)	11.3	13.3	11.4, 13.3
UV 260/280	1.8	0.5	1.8, 0.5
Expected MM (kDa)	205	96	
Observed MM (kDa)	205	99	211, 102

When the NCP and LSH were mixed at 5 µM and 26 µM, respectively, incubated at 4°C for 30 min and injected on the column in a total of 100 µl, the components eluted separately and no shifts were observed, indicating that the complex did not form or that it formed but did not survive the run (***Fig. 5.12C***). The elution volumes of NCP and LSH were 11.3 ml and 13.3 ml, respectively, when injected separately, and these values did not change after injection of the mixture, indicating the lack of a shift due to the absence of the complex. The increase in the temperature compared to the original EMSA performed at 4°C (***Fig. 5.6***) should not have affected the complex as indicated by the EMSA performed at room temperature (***Fig. 5.11***). It should also be noted that the incubation volume was 100 µl, whereas the samples undergo roughly a 14-fold dilution as they pass through the column. This dilution may interfere with the kinetics of the complex and cause complex dissociation, even though the nanomolar affinity predicted by EMSAs should be sufficient to overcome this dilution. This discrepancy may be explained by the overestimation of affinity by EMSA. EMSA is notorious for its limitations, creating an environment where the components behave differently compared to the solution. Once the complex is applied on the gel it is stabilised by the caging effect of the gel, hence, showing an overestimated affinity (Fried and Bromberg, 1997; Hellman and Fried, 2007). Therefore, the SEC-MALS method was not adequate to detect NCP-LSH interaction.

5.8 Gradient fixation trials of LSH-NCP

Since the ultimate step in the LSH-NCP complex investigation would be examination via electron microscopy (EM), optimisation of the complex preparation was explored. Gradient

fixation (GraFix) is a common step in preparation for single-particle EM. The sample is ultracentrifuged into an increasing gradient of glycerol or sucrose and a fixative, such as formaldehyde or glutaraldehyde (Kastner et al., 2008a; Stark, 2010). Sucrose provides a cushioning effect for the sample that migrates through the gradient and is simultaneously being fixed in increasing concentration of glutaraldehyde. The centrifugal forces disrupt any aggregates, whereas the fixative at a low concentration provides intramolecular cross-links. This results in a stabilised complex that is structurally more homogeneous and can endure longer adsorption times on the EM grid, resulting in improved resolution of the sample and reduced noise. Labile samples, such as spliceosomes and 70S ribosome, have been successfully visualised using cryo-EM combined with GraFix. The structure of a nucleosome, both free and in a complex with a protein, have also been solved by combining these methods (Liu et al., 2017a; Nguyen et al., 2013). Therefore, GraFix was selected as a promising approach to stabilise the complex and to improve the sample quality.

Following methods reported in the literature, 10-30% (w/v) sucrose and 0.2% (v/v) glutaraldehyde gradients were chosen to crosslink LSH to an NCP (***see Section 2.6.1***). Glutaraldehyde is an amine-reactive homobifunctional crosslinker (has identical reactive groups on each side of the molecule) and the preferred reagent in crosslinking for electron microscopy. It is a larger molecule compared to another fixation reagent formaldehyde, therefore, it is less prone to introducing artefacts. The initial approach was to check how these conditions affect the fixation of 0W0 NCP, to make sure that no intermolecular cross-linking occurs. A 100 µl of 0W0 NCP sample at 100 nM was added to 10-30% sucrose gradients with or without 0.2% glutaraldehyde. Following an 18h centrifugation at 35,000 rpm 0W0 with and without glutaraldehyde were manually fractionated and analysed on a native 6% TBE PAGE (***Fig. 5.13 A, B***).

The comparison of fixed and unfixed NCPs demonstrated two bands at the bottom of the gradient in the unfixed sample and one band species in the fixed sample, indicating that centrifugation introduced some aggregation to the NCPs (***Fig. 5.13 A***). This extra band was not present in the fixed sample, indicating that fixation might have crosslinked the NPCs and they were pulled to the bottom of the gradient leaving the mononucleosomes in the middle of the gradient (***Fig. 5.13 B***). The weaker signal of the fixed sample may indicate that some fraction of the NCP has aggregated due to crosslinking or that glutaraldehyde caused some IR800 quenching (***Fig. 5.13 B***). The comparison of input (NCP before loading on the gradient),

fixed and unfixed NCPs demonstrated that the bands present in the fixed and unfixed samples, indeed, corresponded to the input NCP (***Fig. 5.13 C***).

To check for successful fixation, the samples were assessed by 15% SDS-PAGE (***Fig. 5.13 D***). The input and unfixed samples have a similar migration pattern, whereas the fixed sample was not visible on the gel. This may be due to two possibilities: either the sample was cross-linked, therefore, being too big to enter the gel of this percentage, or the denaturing conditions further affected the fluorophore, making the sample even less detectable. The presence of faint low molecular weight traces of unknown origin in the fixed sample lane may indicate that the sample could not enter the gel. Future experiments will require optimisation of each GraFix step and include introduction of molecular weight marker mix run on a gradient, gradient preparation and fractionation and different approach to the final product detection.

The analysis of the gels has a few caveats. First, the fractionation was done manually, which is likely to introduce error, and the sedimentation pattern cannot be used to estimate molecular weight of the sample. Second, manual fractionation is more time consuming, and the gradients were likely to undergo some dissipation following centrifugation. These two factors are likely to explain the broad sedimentation profile of the samples.

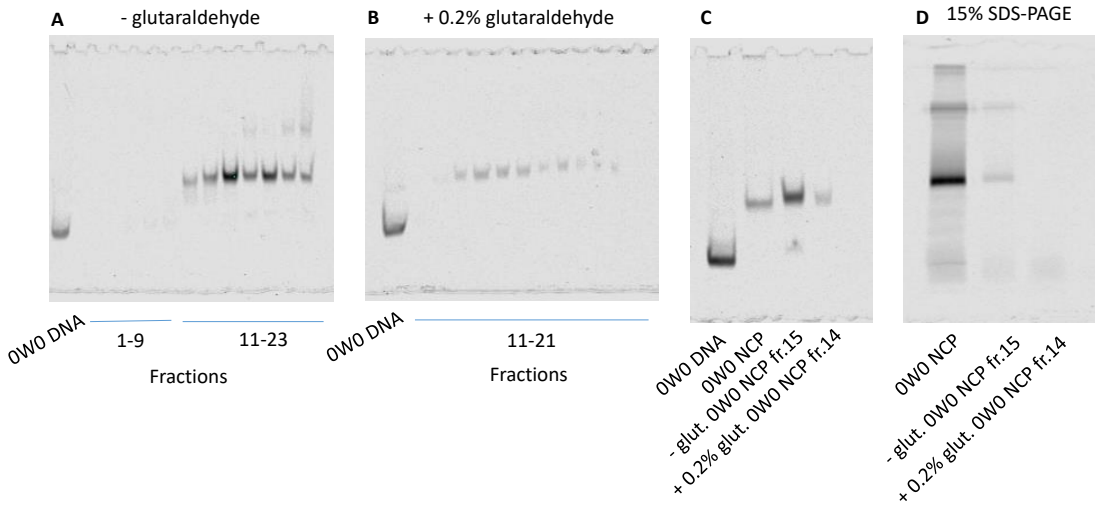


Figure 5.13 PAGE analysis of GraFix tests on 0W0 NCP. Native PAGE of ulatracentrifuged A) Unfixed NCPs. B) Fixed NCPs. C) Comparison of 0W0 DNA (lane 1), input, fixed and unfixed NCPs, Lanes 2,3 and 4, respectively D) Comparison of input, fixed and unfixed NCP migration patterns under denaturing conditions. All samples were run on 4.5% TBE native PAGE, unless stated otherwise. All

samples were visualised by fluorescence intensity of IR800. Fr. – fraction. Note: only every other fraction was run on the gels.

5.9 LSH-NCP complex was not detected by chemical crosslinking

To further explore the ways to capture LSH-NCP complex, I tried chemical crosslinking with BS3 and EDC/sulfo-NHS (***see Section 2.6.2***). Crosslinking is a process where a chemical with a specific reactive group interacts with a functional group on a protein or other molecule, such as DNA, thus forming a covalent bond. BS3 (bis(sulfosuccinimidyl)suberate) is a homobifunctional amine-to-amine crosslinker at each end of its 11.4 Å spacer arm (***Fig. 5.14***). It reacts with primary amines on N-termini of the protein and on lysine residues.

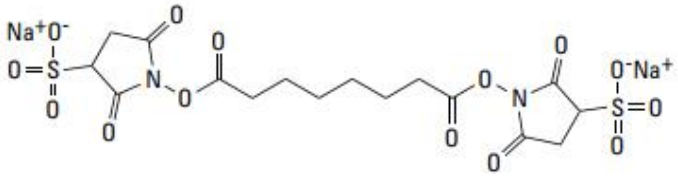


Figure 5.14 Chemical structure of BS3 (bis(sulfosuccinimidyl)suberate) crosslinker (from Thermo Scientific Crosslinking Technical Handbook).

EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride) is a carbodiimide crosslinker that conjugates carboxyl groups to primary amines. Sulfo-NHS (N-hydroxysulfosuccinimide) is added to the EDC reaction to activate carboxylic group (-COOH) for crosslinking with primary amines (***Fig. 5.15***).

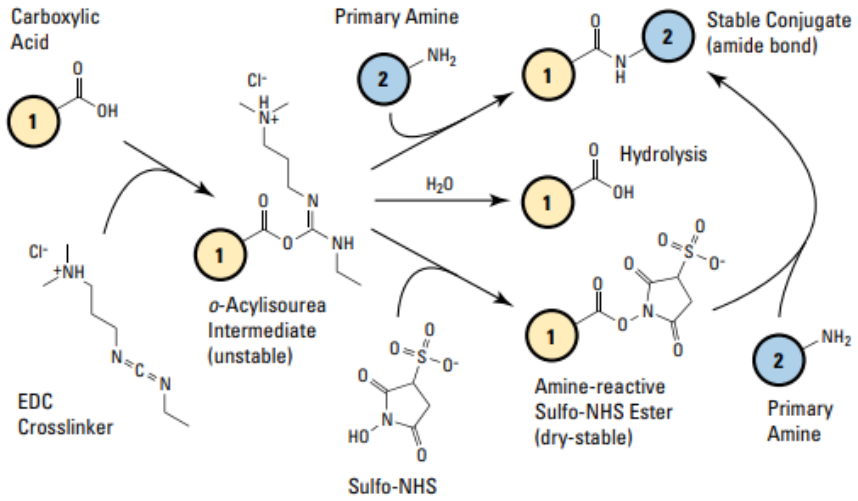


Figure 5.15 Crosslinking reaction with EDC (1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride) and sulfo-NHS (N-hydroxysulfosuccinimide). Carboxylic group reacts with EDC to form an unstable O-acylisourea conjugate, which then reacts with sulfo-NHS forming stable and amine-reactive intermediate. This intermediate is later used to form an intramolecular or intermolecular crosslinking product with a primary amine. Diagram from Thermo Scientific Crosslinking Technical Handbook.

Homology modelling and EMSAs of LSH in a complex with NCP indicated that interaction with nucleosomal DNA would be the most likely contact between two molecules. However, the N- and C-termini of LSH could not be reliably modelled due to low levels of homology with the other chromatin remodellers. The EMSA experiment with acidic patch-mutated octamer showed a slight decrease in affinity (***Fig. 5.8 B, C***). Hence, even though LSH may not form functional interaction with NCP, its flexible termini may be in close proximity to histone tail and the octamer core. Therefore, I tried using BS3 and EDC-sulfo-NHS crosslinking to test this hypothesis.

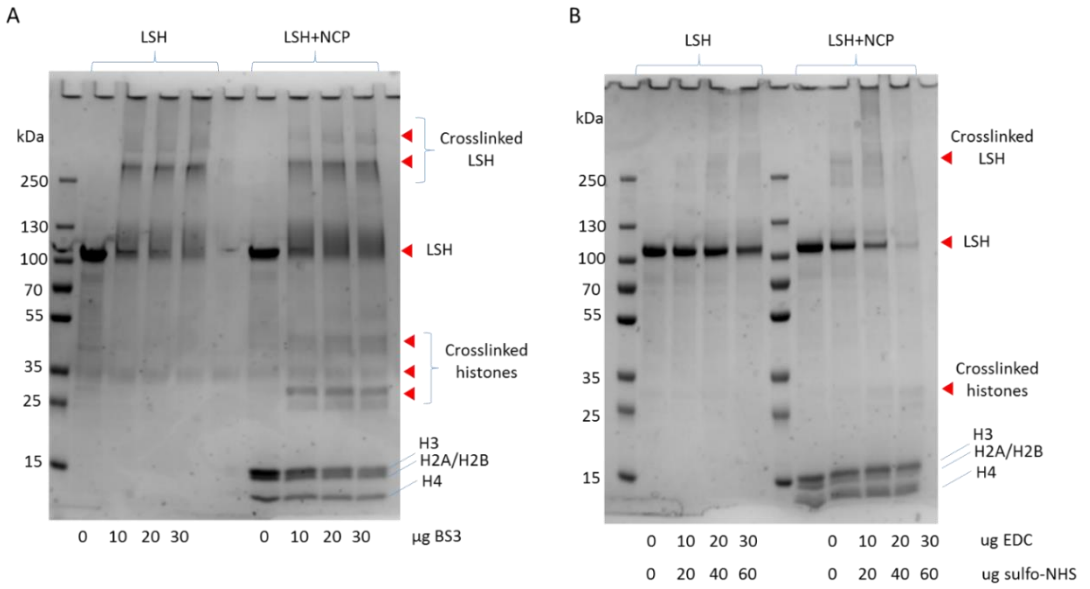


Figure 5.16 Crosslinking reaction between LSH and NCP. A) Crosslinking reactions of 4.8 µg of LSH (2.5 µM) with and without 1 µg of NCP (0.5 µM) and the indicated amounts of BS3; B) Crosslinking reactions of 4.8 µg of LSH (2.5 µM) with and without 1 µg of NCP (0.5 µM) and the indicated amounts of EDC and sulfo-NHS.

4.8 µg of LSH (2.5 µM) was incubated with and without 1 µg of NCP (0.5 µM) for 30 min on ice, and a crosslinker was added to the mix and further incubated for 2h at 4°C. The crosslinked products were resolved on 4-20% precast gradient gels. BS3 crosslinking appeared to be more efficient than EDC-sulfo-NHS, since a larger portion of LSH and NCP was shifted to the crosslinked position on the gel. However, the patterns of LSH and LSH-NCP appeared almost identical indicating that no new crosslinked species were formed following the addition of NCP. Hence, the crosslinking experiment was not pursued further.

5.10 Discussion

To further investigate the LSH-nucleosome interaction I used a common system of a recombinant mononucleosome made of *Xenopus* histones and 601 Widom DNA sequence. The aim was to evaluate if LSH required a DNA linker for binding an NCP or if it binds directly to nucleosomal core. Along with this question I wanted to check if the addition of octamer variants, such as truncated H3 or mutated acidic patch on the octamer, affects LSH affinity for the NCP.

Compared with the other remodellers, LSH does not appear to have a DNA-binding domain, apart from its ATPase domain, however, this role may be carried out by its N- or C-termini.

As mentioned earlier, the N-terminus of LSH harbours a ppHSA motif, common to other chromatin remodellers and found to interact with lobe 2 in ISWI to stabilise the protein, without contributing to its catalytic activity (Ludwigsen et al., 2017). However, this motif is not expected to interact with DNA.

Previous EMSA experiments performed by Simon Varzandeh (Simon Varzandeh, 2016) indicated that LSH required a 25-bp DNA linker to bind the nucleosome and binding was not detected when only a nucleosomal core containing 147 bp was present. This finding was not confirmed in the present study.

To determine which DNA linker minimal length is required for LSH binding, I reconstituted nucleosomes containing no linker DNA or a 6-bp, 12-bp or 25-bp linker on one side (***Fig. 5.6***). From the EMSA experiments performed with nucleosomes with variable DNA linkers it appeared that LSH does not require a DNA linker for binding to the NCP, with similar shifts present in all samples. This reinforced the hypothesis that LSH predominantly binds nucleosomal DNA and does not require linker DNA for binding *in vitro*. This is consistent with the homology modelling data showing LSH binds the nucleosome at SHL2, however, future work elucidating the positions of LSH termini on the nucleosome will be required to verify any potential contacts with the linker DNA.

Different chromatin remodellers have various requirements for DNA linkers. For example, CHD2 required ~40 bp (Liu et al., 2015), ISW2 was shown to require at least 20 bp for binding and remodelling (Brehm, 2000; Kagalwala et al., 2004; Zofall et al., 2004) and around 75-80 bp for efficient association and remodelling (Dang et al., 2006; Deindl et al., 2013). On the other hand, there was evidence that ISWI remodeller can bind nucleosomal core without requiring a linker DNA (Whitehouse et al., 2003), however, linker DNA was indeed required for ISWI ATPase activity, with progressively longer linker resulting in more ATP hydrolysis (Whitehouse et al., 2003). Sth1 subunit of the RSC complex could bind dsDNA of at least 15 bp but required 15-85 bp for increasing and optimal ATPase activity (Saha, 2002). Since LSH remodelling activity has been detected (Jenness et al., 2018), but its ATPase activity has not been dissected, it is still unclear whether LSH requires a linker for efficient ATP hydrolysis, even though it can bind nucleosomal core. It should be noted that Jenness et al. used symmetrical 20-bp linkers, and linker-free nucleosome has not been probed for a remodelling assay. Since the remodelling experiment that they used relied on a PstI restriction site that was initially protected by nucleosomal core and did not necessarily require linker DNA on

both sides, it would be interesting to check if this remodelling could be performed on no-linker or a single-side linker DNA to see their effect on ATPase activity of LSH.

Next, LSH binding to the nucleosome resulted in a double shift, which was clearly observable in case of 12-bp linker DNA but is not very clear with other linkers or the no-linker core (***Fig. 5.6***). This can be due to several reasons. It is established that LSH can bind free dsDNA (see below), but the minimal DNA length is not known. It is possible that in the case of linker-free nucleosome LSH binds the core at the superhelical location 2 (***see Section 3.2***), which is a preferred site for chromatin remodellers (Hauk et al., 2010; Li et al., 2019; Liu et al., 2017a; McKnight et al., 2011; Schwanbeck et al., 2004; Zofall et al., 2006). However, even at that location LSH can bind SHL2 on two surfaces on the nucleosomal disc, which requires two binding events, however, it is not clear how discrete they are and if in the present study the first shift corresponds to the single or double SHL2 binding. Another possible binding location is SHL6 (Eustermann et al., 2018), however, Ino80 is the only remodeller that preferentially binds this location with known biological functions. Snf2 was observed to bind SHL6 with unknown implications (Li et al., 2019; Liu et al., 2017a). Therefore, it is feasible to assume that LSH can bind all four of those locations (two SHL2, two SHL6) *in vitro* via its ATPase domain, as well as the linker DNA, when it is present. Structural studies would be required to elucidate LSH binding pattern to the nucleosome.

This non-specific pattern of LSH binding was observed when free and nucleosomal DNA were compared. The initial experiment comparing the binding to 12-mer, 159-mer or a 12W0 NCP did not reveal any obvious differences, indicating that LSH can bind short fragments of free DNA, as well as longer fragments. Therefore, LSH does not rely on nucleosomal curvature for binding and can recognise free duplex DNA. Moreover, the EMSA shift pattern between 159-mer and 12W0 NCP, which contain the same DNA fragments, but the latter is present in the mononuclesomal context, appeared very similar indicating that LSH is predominantly interacting with the nucleosomal DNA and not the octamer, however, EMSA experiments should not be overinterpreted due to the inherent limitations of this technique.

Previous experiments with an HDAC inhibitor trichostatin A abolished LSH binding to chromatin (Zhou et al., 2009), indicating that LSH may be dependent on histone post-translational modifications. However, this was not the case *in vitro*, as indicated by comparison of LSH binding to nucleosomes containing native and recombinant octamer. Therefore, the main driver of LSH-NCP interaction is still more likely to be solely DNA itself.

Nevertheless, when LSH was incubated with a nucleosome lacking N-terminal H3 tails, the migration pattern appeared distinct from the wild-type, and the shifted bands appear more compact, even though there is no difference in nucleosomal band disappearing upon LSH concentration increase (***Fig. 5.8 B and C***). Whereas predominant point of interaction of H3 with chromatin remodellers is through its post-translational modifications (Chatterjee et al., 2011; Flanagan et al., 2005; Mansfield et al., 2011; Santos-Rosa et al., 2003), LSH indeed may form distinct contacts with the nucleosome upon H3 tail deletion. However, the biological implication for LSH interaction with unmodified H3 tail is unclear and would require further investigation. In would be interesting to explore the effects of histone modifications such as acetylation, methylation and ubiquitylation (Krajewski, 2018; Levendosky et al., 2016) on LSH interaction with the nucleosome. As it appears in other chromatin remodellers, histone modifications are important for orientation and specificity but not for binding *per se*, hence, would not necessarily be detected by EMSA.

Furthermore, histones play an important role in a formation of a nucleosomal epitope termed the acidic patch (Kalashnikova et al., 2013). Symmetrical acidic patches and SHL2 site symmetry on each side of the nucleosomal disc and provide a basis for distinct chromatin remodelling patterns for different remodellers, which use DNA linker as the nucleosomal cue for end-to-centre or centre-to-end positioning (McKnight et al., 2011; Ilana M. Nodelman et al., 2017; Patel et al., 2013; Schwanbeck et al., 2004; Whitehouse et al., 2003). The acidic patch has been shown to act as a remodelling activator in Ino80 and ISWI, in the latter by counteracting the inhibitory AutoN an NegC motifs (Gamarra et al., 2018). Mutations of the acidic patch on just one side of the nucleosome resulted in disruption of DNA linker sensing by Snf2h and shifting the histone octamer towards the shorter DNA linker, as opposed to the longer one (Dann et al., 2017; Levendosky et al., 2016; Levendosky and Bowman, 2019), indicating that the acidic patch is an important factor in insuring remodelling directionality. *In vivo* acidic patch mutations disrupting nucleosome regularity by ISWI have been associated with cancer (Dao et al., 2019).

Here the nucleosomes were reconstituted with mutated H2A (E65A E90A E91A) and H2B (E105A), and, therefore, contained symmetrical acidic patches on each side, whereas Snf2h studies relied on H2A residue mutations (E61A/E64A/D90A/E92A) (Levendosky and Bowman, 2019), highlighting variations in potential experimental set-ups. In the EMSA experiments performed here the nucleosomes appeared completely shifted, as indicated by the

disappearance of the IR800-labelled nucleosome band in a manner compared to the wild-type and the nucleosome containing truncated H3 tail histones (***Fig. 5.8 B and C***). The difference was again, however, in the compaction of the band in the acidic patch mutants, which produced a smeary migration pattern. Whereas this may indicate the difference in LSH binding to the nucleosome, this result should not be overinterpreted, as the different appearance can be due to difference in electrostatic potential explained by the reduction of the negative charge after alanine mutations. Therefore, rigorous structural data, along with biochemical remodelling assays exploring acidic patch variant in LSH remodelling potential should be explored.

Since the LSH-nucleosome complex was detected by EMSA with qualitatively estimated K_d of ~100 nM, even though proper quantification was not successful due to smeary appearance of the bands (***Fig. 5.6***). This affinity was comparable to some other remodellers to 20-40 bp dsDNA with calculated K_d, for example, ISWI had K_d of ~18 nM (Al-Ani et al., 2014) and Chd2 had K_d of ~160 nM (Liu et al., 2015). However, these results cannot be directly compared to the affinity for the nucleosome and a rigorous kinetic study would be required for LSH-nucleosome affinity. However, LSH appeared to bind the nucleosome with nanomolar affinity, which should produce a stable complex. Assuming that the complex formed during the 30 minute incubation of LSH and NCP mix prior to SEC-MALS run, it appears that the complex disintegrated during migration through the column. The injected amounts of LSH and NCP (26 µM and 5 µM, respectively) should have been sufficient for complex formation, if the K_d estimated from EMSA was reliable. Nevertheless, EMSA experiments may demonstrate overestimated affinity, possibly due to caging effect of the gel. It is likely that LSH binding to NCP is transient and is stabilised by interaction with additional proteins, such as CDCA7.

As previously demonstrated by K. Myant and S. Varzandeh who used sucrose gradient sedimentation and SEC-MALS, respectively (Kevin Bryan Myant, 2008; Simon Varzandeh, 2016), LSH is not present in a protein complex or as a dimer. This was confirmed in my SEC-MALS experiment (***Fig. 5.12, Table 5.2***), where LSH migrated through the column as a monomer. Dimerisation of chromatin remodellers, even though not relevant to all of them, plays an important part in chromatin architecture. For example, Snf2h (part of the ACF complex) dimerises on the nucleosome, occupying two SHL2 sites and, by alternating its conformations (Armache et al., 2019; Leonard and Narlikar, 2015), produces alternating

motions that result in central positioning of the nucleosome (Racki et al., 2009), however, from that EM structure it was not apparent if two Snf2h molecules physically interact with each other and would form a dimer without the bridging nucleosome. Nevertheless, when Snf2h molecules were covalently linked, remodelling was accelerated, indicating that physical interaction between two monomers are functionally important (Leonard and Narlikar, 2015). This can also be the case for LSH, a hint of which was observed by EMSA in the form of a double shift (***Fig. 5.6***), however, it is still unclear whether these multiple binding events have functional relevance. On the other hand, Chd1 was shown to perform bidirectional sliding as a monomer (Qiu et al., 2017).

The attempts to stabilise the complex using crosslinking with glutaraldehyde and ultracentrifugation (GraFix) (Kastner et al., 2008b) or chemical crosslinking using BS^3 and EDC+sulfo-NHS have not been successful. In the first instance, NCP on its own appeared to be crosslinked, as indicated by its lack of migration through denaturing gel compared with uncrosslinked NCP. However, LSH on its own was not observed on the gel, indicating either that it aggregated and could not migrate through the gel or that the migrated amount was not sufficient to be detected by Coomassie staining. GraFix, however, is extensively utilised for cryo-EM studies for chromatin remodeller-nucleosome complexes (Li et al., 2019; Machida et al., 2018; Winger et al., 2018).

In the second instance, neither of the chemical crosslinking approaches resulted in an obvious band distinct from the LSH-only control. It is hardly surprising, since LSH is more likely to interact with DNA rather than the octamer core. Nevertheless, LSH terminal regions should be close enough to certain histone tails, for example H4 and H3 that are positioned close to SHL2 site (Luger et al., 1997b), and this proximity was likely to be captured by the long range BS^3 crosslinking. It should be noted that recently identified motif ppHSA in the N-terminus of LSH is not expected to interact with H4 (Ludwigsen et al., 2017). Considering that the histone tails, as well as LSH termini, are flexible and may not form enough of crosslinks of the same entity to be observed on the gel. Chemical crosslinking may require further optimisation, a crucial part of which would be scaling up LSH and especially NCP. Optimised crosslinking experiment with LSH and octamer wild-type would be a good starting point for the experiments involving truncated histone tails.

Another possibility to capture LSH-nucleosome complex would be site-specific UV crosslinking of DNA and protein, which involves modification of DNA with 5-Iodouracil (Flett

et al., 2018). This sophisticated approach, however, requires lengthy optimisation and would require initial labelling of a simpler system with a short dsDNA fragment.

Overall, the LSH-NCP complex appeared to form, predominantly due to LSH-DNA interaction through the LSH ATPase domain, however, the affinity could be in µM rather than nM range, which would require more material to detect it. Additionally, CDCA7 could be a stabilising agent, however, further experiments are required to test that.

6. BIOINFORMATICS ANALYSIS, CLONING, EXPRESSION AND PURIFICATION OF CDCA7

6.1 Introduction

CDCA7 has been recently identified to be a binding partner of LSH, however, information available on CDCA7 is very limited. With the ultimate goal of expressing and purifying murine CDCA7 for biochemical and structural studies and to evaluate its binding to LSH, I initially utilised a host of bioinformatic tools to evaluate the sequence characteristics and associated biochemical properties of CDCA7. One of the aims of this part of work was to produce mg amounts of CDCA7 to carry out biochemical and biophysical characterisation studies of CDCA7. For that, I attempted to express the protein in two different systems – insect and bacterial cells. Murine CDCA7 has proved to be difficult to purify, as high amount of nucleic acid contamination was present following various purification protocols.

6.2 CDCA7 bioinformatics analysis

Being a DNA-binding protein, murine CDCA7 has a high isoelectric point (theoretical pI 8.94 as indicated in ProtParam), which indicates it has extensive positively charged regions participating in protein-DNA or protein-protein interactions.

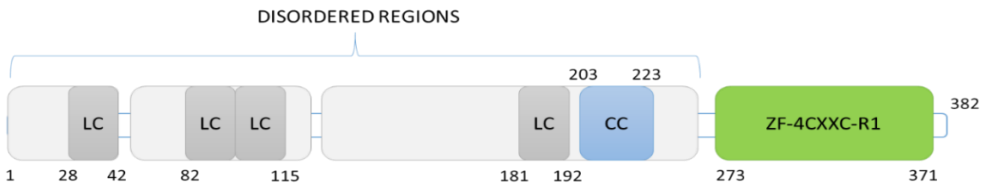


Figure 6.1 Schematic representation of CDCA7 domains, adapted from Pfam server. Grey boxes represent disordered regions, LC – low complexity, CC – coiled coil, ZF-4CXXC-R1 – zinc finger domain. Amino acid numbers are indicated below the schematic.

Table 6.1 CDCA7 amino acid regions and their charachteristics as predicted by Pfam (visualised on Fig. 6.1).

Region (amino acids)	Characteristics
1-28	Disorder
29-42	Low complexity
32-127	Disorder
82-115	Low complexity
131-253	Disorder
181-192	Low complexity
203-223	Coiled coil
203-213	Low complexity
273-371	ZF-4CXXC-R1

As can be seen from the Pfam domain prediction (***Fig. 6.1, Table 6.1***), CDCA7 has extensive regions of structural disorder and can be characterised as an intrinsically disordered protein (IDP). Those disordered regions include highly conserved parts of the protein, indicating they likely have a biological function. There are several proteins that transition from disordered to folded state upon binding their biological target. Alternatively, disordered domains can function as flexible linkers crucial for proper folding and functioning of their juxtaposed regions (Dyson and Wright, 2005). Biological rationalisation for such regions is their increased specificity, since the target is required for proper folding, and that they can be an easier target for degradation, which is an additional fine-tuning mechanism for their interactions and biological functions, such as in signalling and transcription regulation, where fast binding-unbinding mode is utilised (Dunker et al., 2000; Oldfield et al., 2005). Additionally, disordered proteins may bind in a “fly-casting mechanism”, in which initial weak contacts are formed away from the biding site, and the protein gradually folds upon approaching its bona fide bidning site (Shoemaker et al., 2000). The main driving force for protein folding is hydrophobic interactions, which are weak intermolecular forces, but their cumulative action ensures protein stability. Propensity for hydrophobic interaction formation can be deduced from the protein sequence. IDPs can be identified by their amino acid composition, which is biased towards polar charged residues that provide repulsion. Due to the lack of stable structural features, IDPs are notoriously hard to crystallise and to conduct structural studies on. IDPs are predominantly studied by computational approaches, such as molecular dynamics simulations, and by biophysical techniques, such as surface plasmon resonance and electron paramagnetic resonance (Na et al., 2018; Stanley et al., 2015).

CDCA7 has four (two of them located close to each other and can be counted as one) predicted regions of low complexity ($CDCA7^{29-42, 82-115, 181-192, 203-213}$) (***Fig. 6.1***). Low complexity regions (LCRs) are those that have simple repetitive patterns of amino acid composition (Coletta et al., 2010). They can contain a few amino acid types or just one amino acid. Coletta et al. found that LCRs located in the central part of a protein tend to be predominantly related to transcription processes based on Gene Ontology terms (Coletta et al., 2010). This correlates with the central LCR of CDCA7 that flanks the c-Myc interacting domain. It was also found that LCR-containing proteins tend to be a part of highly connected interaction networks (Coletta et al., 2010). Interestingly, CDCA7 has two runs of glutamates – one (E_{10}) in the second predicted low complexity region ($CDCA7^{106-115}$) and another one (E_{11}) downstream the last low complexity region ($CDCA7^{204-215}$), although E_{11} is predicted to be a part of the coiled coil ($CDCA7^{203-223}$) (see below). Glutamate stretches, or runs, were more frequently found in human proteins, however, in general such stretches are more common in fly (Karlin et al., 2002). Human proteins containing such stretches are often implicated in diseases(Karlin et al., 2002). It should be noted that these glutamate stretches are not well conserved, as indicated by the multiple sequence alignment (MSA) (***Fig. 6.4***).

CDCA7 has predicted sites for post-translational modifications (PTMs). PTMcode2 is a database of known and predicted PTMs of proteins (Minguez et al., 2015). It showed that murine CDCA7 has three predicted phosphorylation sites – T170, T203 and S225. T170 corresponds to T163 in human, which is a site of 14-3-3 protein interaction (Gill et al., 2013) (see Section 1.4 of Introduction). UniProt also lists S197 as an additional PTM. These sites were predicted by similarity with human CDCA7, in which the PTMs were discovered during a proteomics study of HeLa cells (Zhou et al., 2013), or, as in the case of S225, during phosphoproteomics of mouse tissues (Huttlin et al., 2010).

UniProt predicts a coiled coil (CC) at positions 203-223. CCs are formed by supercoiling of two or more α-helices through close packing of their side chains (Truebestein and Leonard, 2016) and have highly versatile biological functions, including self-oligomerisation and protein-protein interaction (Burkhard et al., 2001). Even though there are rules guiding which residues are to form the heptad repeat of a CC, sequences can be highly variable. Pfam predicted one $CC^{203-223}$ flanking the zinc finger domain (***Fig. 6.1***). However, the ExPASy server COILS predicted two CCs – one consistent with the Pfam prediction and another one that partly overlaps with the middle LCR (residues 100-150 and 200-225) (***Fig. 6.2***). It should be

noted that a simple visual analysis of CDCA7 sequence does not allow to observe any obvious heptad repeats, which is expected, since CCs can be highly discontinuous (Lupas and Gruber, 2005). The COILS server used for identification of putative CCs is based on Parry matrices of residue frequencies that analyse asymmetric distribution of the heptad segments within defined windows (Parry, 1982). Both CCs predicted by COILS include the glutamate stretches. These CCs are potential regions of interaction with LSH (which has a predicted CC in its N-terminus [14-96]), if such an interaction exists.

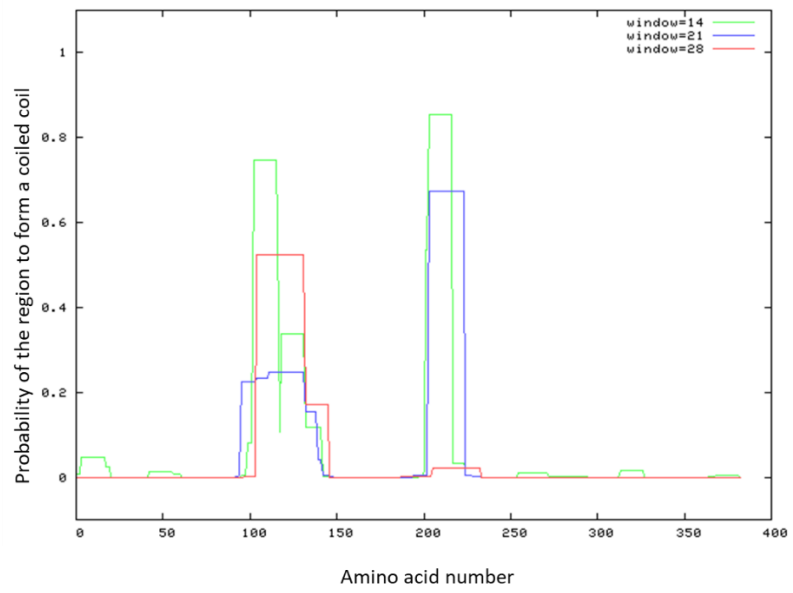


Figure 6.2 Prediction of coiled coil patterns in the sequence of murine CDCA7 by the server COILS. Windows represent the number of amino acids scanned at a time (14,21, 28).

The server MARCOIL (Delorenzi and Speed, 2002) predicts coiled coils in the same regions as COILS, adding more confidence to this prediction. In the first predicted coiled coil, the core position at *a* include a mixture of charged and hydrophobic residues, whereas position *d* accommodates negatively charged residues followed by small hydrophobic isoleucine and leucines (***Fig. 6.3 A***). The solvent-facing positions (*b, c, e, f, g*) predominantly contain polar and charged residues (***Fig. 6.3 A***). The second coiled coil is also dominated by hydrophobic and charged residues in the core *a* and *d* positions and the solvent-facing positions are occupied mostly by charged and polar residues (***Fig. 6.3 B***). Therefore, predicted coiled coils may indeed be present in CDCA7.

A

99N[01.7]c	100D[01.9]d	101S[02.8]e	102H[03.9]f	103S[06.2]g	104D[09.7]a	105S[37.9]b
106E[66.3]c	107E[75.9]d	108E[90.8]e	109E[94.5]f	110E[95.9]g	111E[96.4]a	112E[97.4]b
113E[97.7]c	114E[97.7]d	115E[97.9]e	116D[97.8]f	117G[97.8]g	118M[97.9]a	119N[98.0]b
120F[98.0]c	121L[98.2]d	122E[98.2]e	123K[98.2]f	124R[98.2]g	125A[98.1]a	126L[98.1]b
127N[98.1]c	128I[98.1]d	129K[98.0]e	130Q[98.0]f	131N[97.8]g	132K[97.6]a	133A[97.5]b
134M[97.3]c	135L[97.1]d	136A[96.6]e	137K[96.3]f	138L[95.7]g	139M[95.2]a	140S[94.4]b
141E[93.6]c	142L[91.3]d	143E[81.9]e	144S[44.1]f	145F[03.6]g	146P[00.7]a	147G[00.3]b

B

197S[03.3]g	198L[03.7]a	199G[04.0]b	200A[04.6]c	201L[05.2]d	202P[06.0]e	203T[13.0]a
204E[25.4]b	205E[29.8]c	206E[32.2]d	207E[33.8]e	208E[34.4]f	209E[34.5]g	210E[34.6]a
211E[34.7]b	212E[34.5]c	213E[34.0]d	214E[32.7]e	215D[30.1]f	216K[27.8]g	217Y[24.3]a
218M[22.2]b	219L[20.8]c	220V[19.9]d	221R[19.4]e	222Q[18.5]f	223R[16.8]g	224K[14.1]a
225S[11.0]b	226M[07.9]c	227D[05.9]b	228S[04.7]c	229Y[03.6]d	230M[03.1]e	231N[02.6]f

Figure 6.3 N-terminal region (residues 99-147 and 197-231) of murine CDCA7 with two heptad repeat patterns as indicated by MARCOIL server. Core positions a and d are in red, surface positions are in blue. Percentage probability to form coiled coil is in brackets. A) First and B) second coiled coil predicted.

CDCA7 has several orthologues in mammals and more distant species such as clawed frog and zebrafish (***Fig. 6.4***). UniProt entries of the orthologues, although mostly unreviewed, include such classes as *Actinoprerygii* (ray-finned fishes) and *Avis* (birds). The most distant orthologues have around 50% identity to the murine CDCA7, as indicated by a BLAST search on UniProt. The ICF syndrome-associated mutations (R274C/H, G294V, R304H in human corresponding to R285, G304, R315 in mouse) are all located in the zinc finger domain and are all highly conserved (***Fig. 6.4***). The species present in the alignment were selected based on their curated status in UniProt (mammalian species). The non-mammalian species (*X. laevis, D. rerio, X. tropicalis*), although unreviewed, were selected to check the conservation across more distant species.

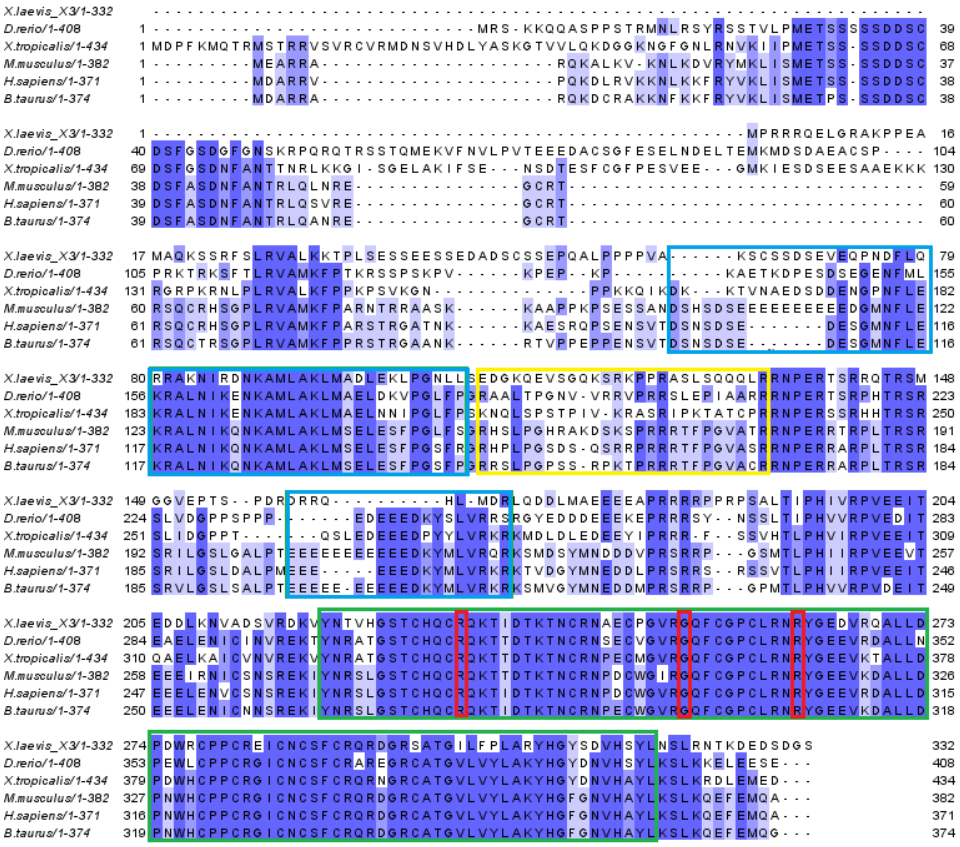


Figure 6.4 Multiple sequence alignment of CDCA7 orthologues performed in ClustalOmega and visualised in JalView. Domains are labelled as follows: C-Myc interacting domain (yellow) and zinc finger domain (green) are indicated on UniPro; coiled coiled regions (blue) were predicted by the COILS server. Relevant murine residues (R285, G304, R315) mutated in ICF syndrome in human LSH are in red.

MSA in ClustalOmega revealed extensive conserved regions in CDCA7. Server COILS predicted two coiled coils, that are 50-60% conserved across the species used in the present alignment (***Fig. 6.2***). The c-Myc interaction domain does not appear to be highly conserved across the species but about a half of it is conserved in the mammalian species present in the alignment (human, mouse, bull), namely the region PRRRTFPGVATR (***Fig. 6.4***). The C-terminal zinc finger domain is highly conserved. None of the CDCA7 orthologues have structural information in the Protein Data Bank, therefore studying CDCA7 by homology modelling would not give reliable information. Nevertheless, sequence analysis can provide some insight into the protein profile.

The server PONDR (Garner et al., 1999; Peng et al., 2006; Romero et al., 1997) indicates the overall percentage disorder to be 49%, which echoes the result from Pfam, which predicted extensive regions of disorder in CDCA7. The ordered region corresponds with the ZF-4CXXC domain (***Fig. 6.1 and 6.5***). Interestingly, there are two disordered regions predicted with high

confidence in the middle part of the protein that overlap with the predicted CCs (***Figs. 6.1, 6.2 and 6.5***). The black lines covering the threshold region indicate possible transition from disordered to ordered state (residues 70-123 and 157-219) in the presence of binding and/or stabilising partners. This analysis is confirmed by the disorder prediction tool IUPred2A (Dosztányi, 2018; Mészáros et al., 2018) and the predictor of binding regions in disordered proteins ANCHOR2 (Dosztanyi et al., 2009; Mészáros et al., 2009). There are several described zinc finger proteins (e.g. the ones with the Cys_2His_2 motif) that undergo conformational changes in their 3D folds in the presence of DNA and adopt structural characteristics required for their biological functions (John H Laity et al., 2000). Examples include the AT-hook containing HMG (high mobility group) and ETS (E26 transformation specific) transcription factor proteins (Levy et al., 2007; Ozturk et al., 2014). CDCA7 may follow a similar mechanism of folding and activity.

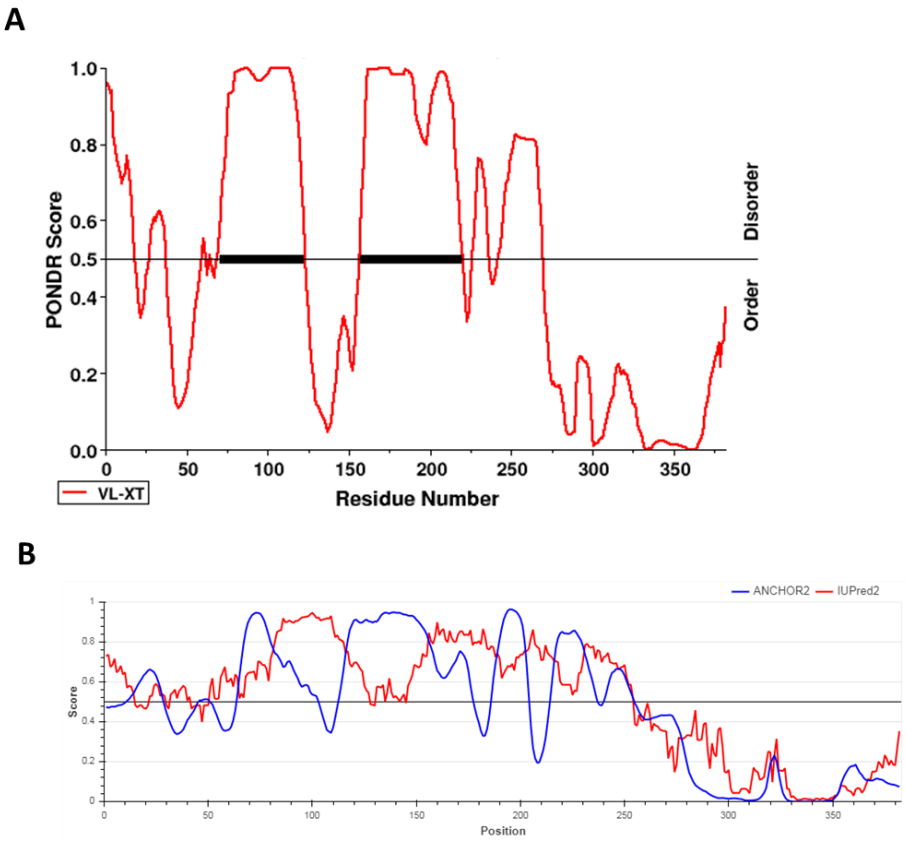


Figure 6.5 CDCA7 disorder profile as indicated by A) PONDR and B) IUPred2A. In A, the black lines covering the threshold region indicate possible transition from disordered to ordered state. In B, ANCHOR2 represents putative protein binding regions.

PsiPred (Jones, 1999) prediction of secondary structure demonstrated extensive regions of random coil (***Fig. 6.6***), which corresponds with the results from PONDR and IUPred2A (***Fig. 6.5***). Both predicted coiled coils correspond to the mostly helical regions, with both glutamate stretches located in those regions, and only a few short β-sheets (***Fig. 6.6***). The c-myc interaction domain is predicted to be a random coil, which agrees with the disorder prediction (***Fig. 6.5***). The ICF syndrome-associated mutations (Thijssen et al., 2015) are located in ZF domain, either in the random coil or the end of an α-helix (***Fig. 6.6***).

Figure 6.6 Secondary structure prediction by PsiPred. Numbers on each side are amino acid positions in the protein. Pink – helix, yellow – strand, grey – random coil. CDCA7 domains are labelled as follows: coiled coils predicted by COILS server – blue, c-myc interaction domain - yellow box, zinc finger domain - green, ICF-associated mutations for the murine sequence: R285, G304, R315 are in red boxes.

A close homologue of CDCA7, CDCA7L (CDCA7-like), is a protein found to have a monoamine oxidase inhibitor function (Chen et al., 2005). Pairwise sequence alignment in EMBOSS Needle demonstrated that the two proteins have high percentage sequence identity in the coiled coil 1 and zinc finger domain (ZFD), as was confirmed by the alignment in Blast Protein. However, the second coiled coil present in CDCA7 is not present in CDCA7L (***Fig. 6.7, Table 6.2***). The N-terminal part of CDCA7L not present in CDCA7 harbours a specific interaction domain (SID) with LEDGF (lens epithelium-derived growth factor)/p75 – a binding target for HIV integrase (Bartholomeeusen et al., 2007). The ZFD of CDCA7L was found to harbour a RING motif, as indicated by SMART database search (Bartholomeeusen et al., 2007).

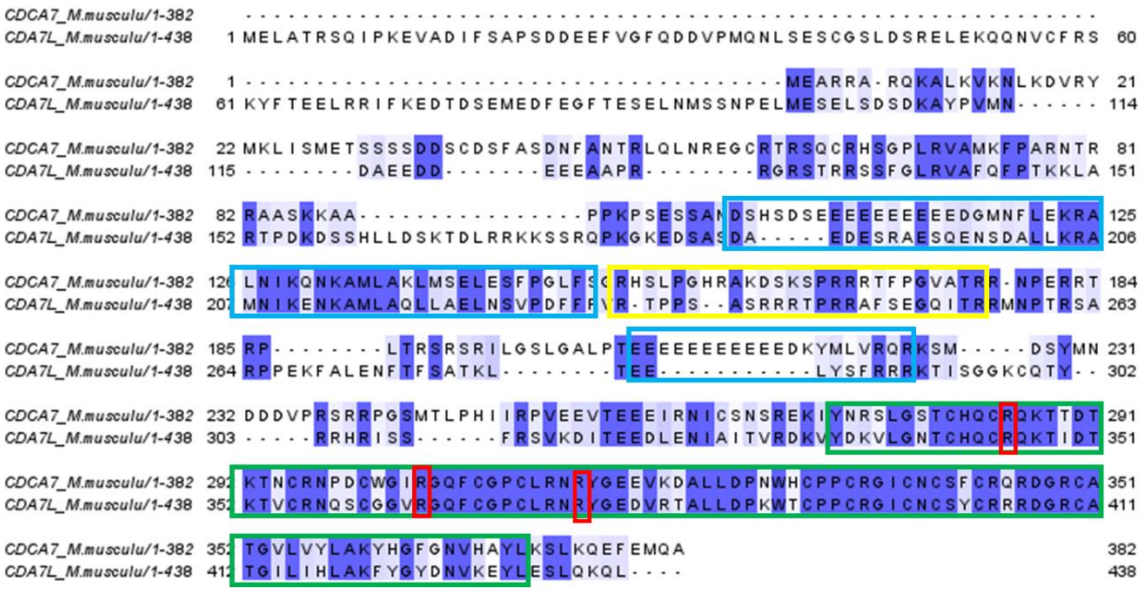


Figure 6.7 Pairwise sequence alignment of murine CDCA7 and CDCA7L performed in EMBOSS Needle and visualied in JalView. Domains are labelled as follows: C-Myc interacting domain (yellow) and zinc finger domain (green) are indicated on UniPro; coiled coiled regions (blue) were predicted by the COILS server. Residues mutated in ICF syndrome are in red.

Table 6.2 Percentage identity (%Id.) in amino acid sequences of murine CDCA7 and its homologue murine CDCA7L as indicated by BLAST Protein.

	Species	Total % id.	Coiled coil 1 % id.	Coiled coil 2 % id.	ZF % id.
CDCA7L	*M. musculus*	44.3	52.2	No similarity	73.7

In vivo yeast one-hybrid assay demonstrated that CDCA7L had a putative DNA-binding activity by directly binding the sequence 5'-ccg gac gcg cag ccc cgc ccg ccc gcc tac gcg cag-3' found in monoamine oxidase A core promoter, which contains Sp1 transcription factor binding sites (Chen et al., 2005). This is in line with the propensity of many zinc finger containing proteins to bind CG-rich sequences (Paillard, 2004), however, RING ZFD are more likely to bind proteins than nucleic acids (Borden and Freemont, 1996).

Currently there is no high resolution structural information, for either CDCA7 or CDCA7L. Therefore, the possibility of homology modelling was tested in Phyre2, but only demonstrated the lack of templates with >30% identity. Therefore, *ab initio* modelling in Phyre2 and I-TASSER (Kelley et al., 2015; Yang et al., 2015) were attempted.

Ab initio structure prediction is utilised if a protein does not have known homologues or none of its homologues have structural data. Small proteins (<120 residues) are better candidates

for this task rather than large proteins (Lee and Skolnick, 2007). I-TASSER models are constructed by iterative fragment assembly simulations which removes steric clashes and constructs models of correct topology even if clashes are present in the threading templates (Lee and Skolnick, 2007; Wu et al., 2007). Phyre2 uses an extended list of templates, which are then used for generating models and the missing fragments are constructed by simplified *ab initio* modelling (Kelley et al., 2015).

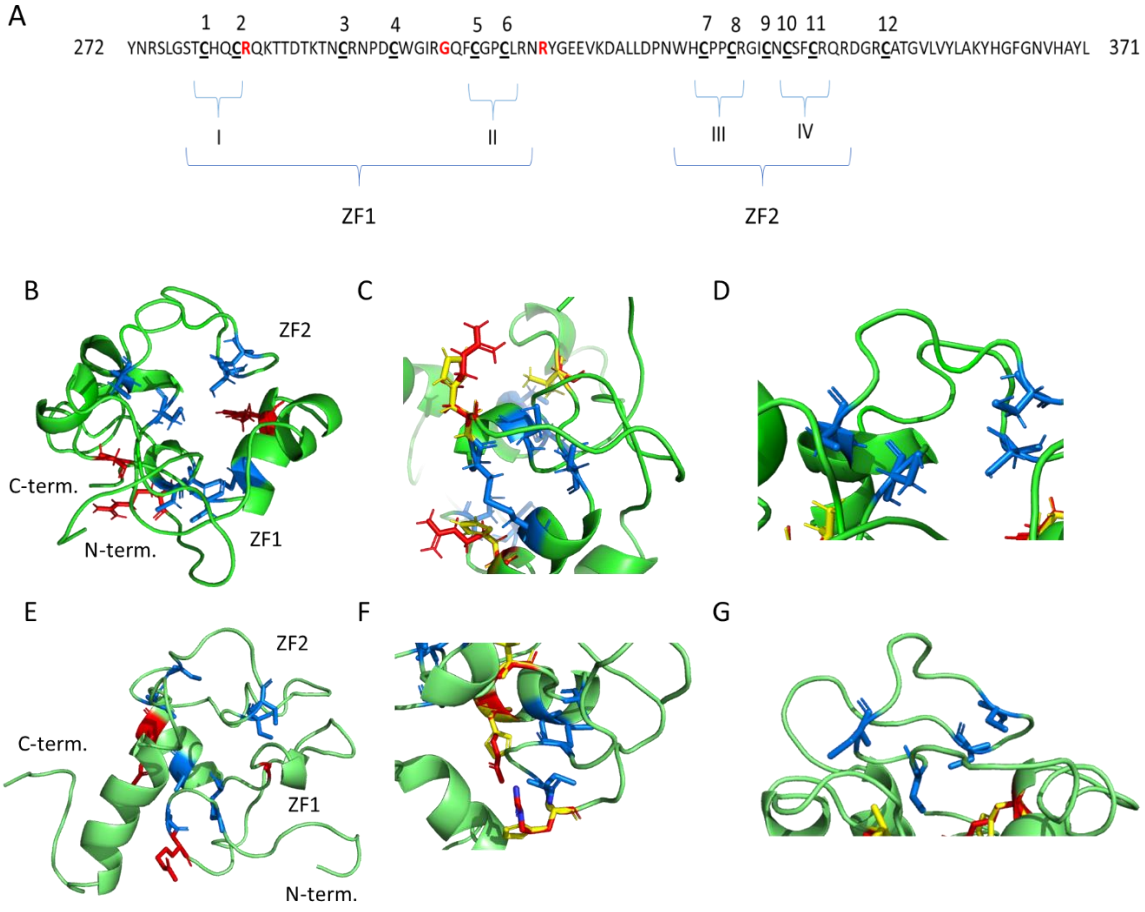


Figure 6.8. Ab initio modelling of the zinc finger domain of murine CDCA7. A) CDCA7 zinc finger domain with 4 cysteine pairs. All cysteines present in ZF are labelled in bold and underlined. ICF-associated mutations are labelled in red. Zinc finger labelling adapted from Chen et al. 2005. B) CDCA7 cartoon model predicted by I-TASSER: cysteines participating in the zinc finger domain are in blue sticks, residues R285, G304, R315 mutated in ICF syndrome are in red sticks; C) Close-up on the zinc finger 1 by I-TASSER: ICF mutations R285C/H, G304V, R315H are in yellow sticks superimposed on the non-mutated residues in red sticks; D) Close-up on the zinc finger 2 by I-TASSER: ICF mutations R285C/H, G304V, R315H are in yellow sticks; E) CDCA7 cartoon model predicted by Phyre2: cysteines participating in the zinc finger domain are in blue sticks, residues R285, G304, R315 mutated in ICF syndrome are in red sticks; F) Close-up on the zinc finger 1 by Phyre2: ICF mutations R285C/H, G304V, R315H are in yellow sticks superimposed on the non-mutated residues in red sticks;; G) Close-up on the zinc finger 2 by Phyre2: ICF mutations R285C/H, G304V, R315H are in yellow sticks.

To simplify modelling only the zinc finger domain was submitted covering CDCA7$^{272-371}$ wild type murine sequence. Both modelling programs produced structures consisting of α-helices connected by loops (***Fig. 6.8 B and E***). The I-TASSER model has a C-score of -1.99, which indicates it is a model of an average confidence with -5 being the lowest C-score and 2 the highest (Yang et al., 2015). The TM-score of the I-TASSER model is 0.48, which is close but below the cut-off of >0.5 for a model of correct topology (Zhang and Skolnick, 2004). Phyre2 generated a model with confidence value of >90% and 77% of the submitted sequence modelled. Phyre2 does not provide further information on model quality and the Phyre Investigator option is not available for intensive modelling.

The two models do not appear to have similar shapes (***Fig. 6.8 B and E***), however, due to the abundance of disordered region, the absence of a bound ligand and bound zinc atoms a distinct shape would not be expected. Even though there are few short β-sheets fragments predicted by PsiPred (***Fig. 6.6***), no such secondary structure was predicted in either model (***Fig. 6.8 B and E***). In the I-TASSER model only the sequence GPCLRNR was predicted to be an α-helix, the rest were either partially or completely loops (***Fig. 6.6 and 6.8 B***). The Phyre2 model showed the sequences GPCLRNR and VLVYLAKY as an α-helix, the rest were loops (***Fig. 6.6 and 6.8 E***).

Initial inspection of both models showed correct orientation of the zinc finger-participating residues, which fold into pockets that can potentially coordinate one zinc ion per pocket (***Fig. 6.8 B and E***). However, the orientation of the cysteines folds the domain in a way more resembling LIM configuration, rather than RING or PHD (***Fig. 6.8 B and E***), since LIM domain fold relies on continuous cysteine orientation, whereas RING and PHD folds have a central criss-cross orientation of the central cysteines , which participate in neighbouring zinc coordinating pockets (Cassandri et al., 2017). Moreover, examination of the ZFD sequence does not show any histidines located away from the cysteine residues (***Fig. 6.8 A***). There are histidine residues immediately neighbouring the cysteines, therefore, it is not clear whether or how they participate in Zn^{2+} coordination. It is also possible to assume that the cysteines form two neighbouring classic C2H2 ZFD, since the sequence does not appear to provide a clear answer.

The residues mutated in ICF (R285C/H, G304V, R315H) are located right next to the Zn^{2+}-coordinating cysteines in ZF1, therefore, it is viable to suggest they interfere with ion binding

through imidazole ring introduction (***Fig. 6.8 A, C, F***). However, this was not determined in this present ab initio modelling due to inability to superimpose the models onto known structures, so that the zinc atoms locate to the relevant areas within the zinc binding pocket. The ZF2 is located away from the ICF mutations, however, in the native conformation it could also be affected by these mutations through altered folding of the domain (***Fig. 6.8 A, D, G***). Finally, it is still not clear whether the main interacting partner of CDCA7 ZFD is DNA or protein, and if the ICF mutations introduce conformational changes without disrupting zinc ion binding but interfering with CDCA7's rigidity or topological parameters, which causes disruption of binding to its target molecule.

I-TASSER also uses the TM-align algorithm, which performs structural alignment based on structural similarity by scanning all structures in the PDB library (Zhang and Skolnick, 2005). The top two hits for structural similarity with the constructed model were DNA methyltransferase-like protein DNMT3L in complex with H3 peptide (Ooi et al., 2007 PDB ID 2PVC) and the ADD domain of *de novo* DNA methyltransferase DNMT3A in complex with H3 peptide (Otani et al., 2009 PDB ID 3A1B). Both proteins are predicted to have GATA-1-like and PHD ZFDs (UniProt O88508 and Q9CWR8 for DNMT3A and DNMT3L, respectively).

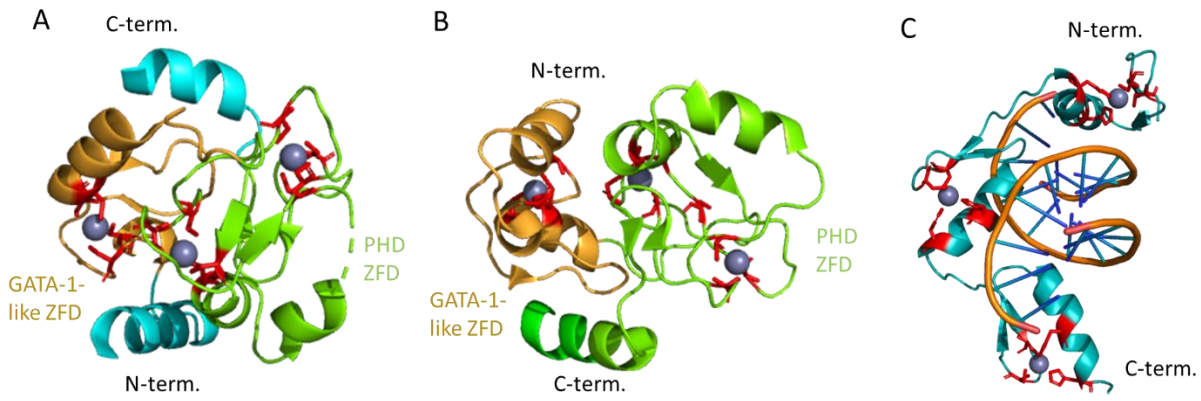


Figure 6.9 Crystal structures of DNMT3A, DNMT3L and EGF1 zinc finger domains. A) DNMT3A ZFD (PDB ID 3A1B) with the domains labelled is indicated. Structure adapted from Otani et al. 2009; B) DNMT3L ZFD (PDB ID 2PVC) with the domains labelled is indicated. Structure adapted from Ooi et al. 2007; C) EGF1 (PDB ID 1ZAA). Structure adapted from Pavletich and Pabo 1991. In all structures, zinc-coordinating cysteines are in red sticks, zinc atoms are shown as grey spheres.

Even though both DNMT3A and DNMT3L are known to have a PHD ZFD, they are distinct from classic PHD domain containing histidine, and instead their three zinc atoms are

coordinated by six pairs of cysteines (***Fig. 6.9 A and B***), of which one zinc atom is coordinated by the GATA-1-like domain with cysteine residues located continuous along the peptide backbone, and two are coordinated by the PHD domain and have the classic criss-cross orientation of cysteines in zinc coordination (***Fig. 6.9 A and B***). A classical C2H2 zinc finger protein EGF1 bound to DNA (***Fig. 6.9 C***) is shown as a comparison to demonstrate DNA-binding mode of a zinc finger domain. Here, pairs of cysteines and histidines coordinate each zinc atom, and the α-helix-β-sheet tandems insert themselves into the DNA major groove (Pavletich and Pabo, 1991). EGF1 has 28% identity with CDCA7 in their ZFD, which was too low to be used as a homology modelling template.

Taken together, *ab initio* modelling of CDCA7 did not show its ZFD as a classic RING domain due to the orientation of cysteine residues. Therefore, the modelling confirms the propensity of the C-terminal region to coordinate bivalent cations, such as zinc, but does not elucidate the type of molecule CDCA7 ZFD is likely to bind.

6.3 Insect cell expression of CDCA7

6.3.1 CDCA7 cloning into pFL vector

Initially, I planned to clone full-length insect-cells codon optimised CDCA7 with an N-terminal MBP (Maltose Binding Protein) tag to keep the construct close to the one used in the Jenness paper (Jenness et al., 2018). MBP tag is a large tag of approximately 42 kDa that can improve protein stability and solubility. To do that I used a plasmid that contained an MBP ORF and tried to create a construct for a fusion protein using overlapping PCR and then Gibson assembly. In the initial approach with overhanging PCR and traditional ligation-based cloning, I tried using BamHI and SalI and then NheI and SphI restriction sites, however, the attempts to insert the resulting amplicon into the pFL plasmid were not successful with any of the approaches. Even though the plates contained a large number of evenly spread colonies, all checked clones turned out to be incorrect, with restriction digest showing fragment of wrong sizes (not shown). One of the reasons for that may have been non-specific recombination due to the lack of recA1 gene in DH5α strain.

In parallel with MBP tag cloning, I tried using a vector for ligation independent cloning (LIC) containing an N-terminal GST-tag. Insertion of murine CDCA7 with LIC-compatible overhangs into pFL-GST vector with a 3C cleavage site resulted in production of a correct clone, which was verified by Sanger sequencing (clone pFL-GST-CDCA7). The GST tag has an additional

caveat of forming dimers, which can be problematic in stoichiometry and binding studies, however, I was planning to cleave the tag after the first step of purification with GSH resin, so the tag dimerization would not affect downstream processes.

6.3.2 CDCA7 expression and purification in Sf9 cells

After generating pFL-GST-CDCA7 plasmid, inserting it into EMBacY cells and verifying the resulting bacmid by PCR (***see Section 2.1.2.2***), I transfected Sf9 cells with GST-CDCA7 construct, according to the protocol described in Materials and Methods (***see Section 2.1.3.4***). 50 ml culture infected for 96h using V2 were lysed and incubated with GSH resin for 3h and eluted with the buffer containing 20 mM GSH. Even though a fraction of GST-CDCA7 was removed from the resin, the majority was not eluted (***Fig. 6.10***), indicating non-specific interaction. Therefore, I decided to displace the protein from the resin by cleaving off GST tag with 3C.

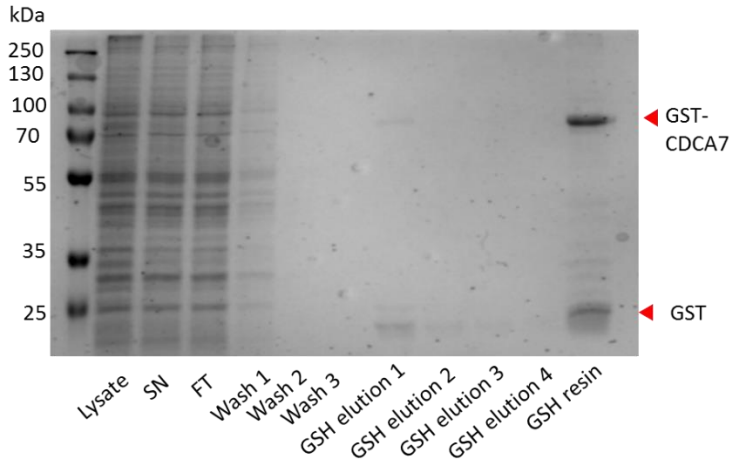


Figure 6.10 GST pull-down of GST-CDCA7 from Sf9. Unsuccessful elution from the resin was attempted using 20 mM GSH. Coomassie stained 12% SDS-PAGE, 0.5% of the total sample loaded.

It is a common practice to use 3C fused to a GST tag, allowing the protease to be inhibited on the resin through binging to GSH to avoid downstream contamination of the sample. However, available 3C-GST has a MW close to that of CDCA7 (~44 kDa). Therefore, I chose to use 3C fused to a His tag, resulting in a 22 kDa construct that can subsequently be separated by gel filtration. After an overnight incubation the resin-bound protein with 3C, CDCA7 appeared in the flow-through indicating successful cleavage (***Fig. 6.11 A, lane FT***). Along with

the full-length CDCA7, two extra bands also appeared in the flow-through (***Fig. 6.11 A***). Initially, it was assumed that the predicted high degree of disorder inherent to the protein renders it unstable and results in degradation products of around 29-30 kDa in size. However, a Western blot against CDCA7 epitopes revealed that the fragments are either not CDCA7 degradation products or they are but are lacking the relevant epitopes (***Fig. 6.11 B***).

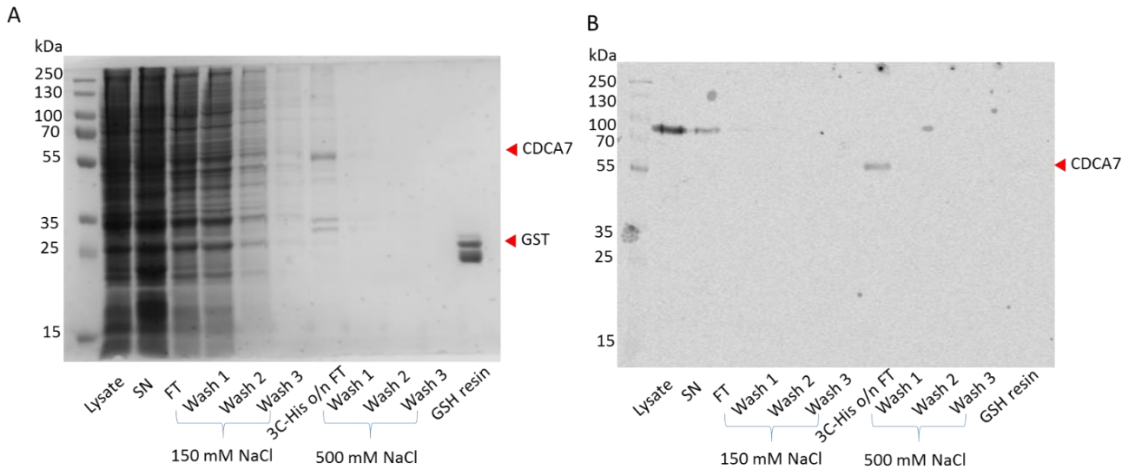


Figure 6.11 Test purification of murine CDCA7 expressed in Sf9 cells: A) GST pull-down of GST-CDCA7 with subsequent cleavage by 3C. 0.5% of total sample volume was loaded and stained with Coomassie. B) Western blot of GST pull-down: same samples as for the Coomassie stain were used. The membrane was probed with anti-CDCA7 primary polyclonal antibody raised in rabbit and anti-rabbit secondary antibody conjugated to Licor IR800 fluorophore.

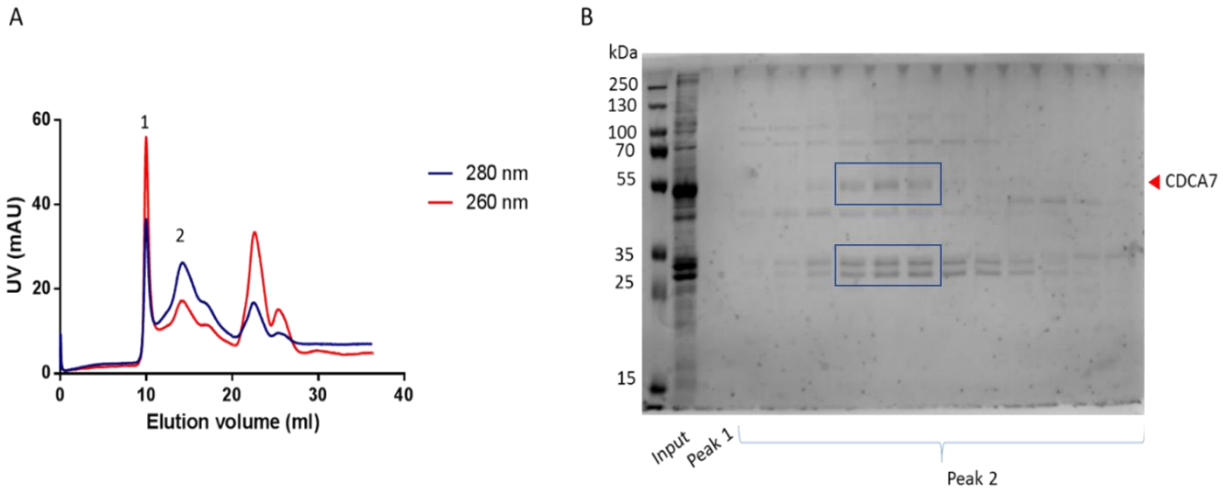


Figure 6.12 Gel filtration of CDCA7. A) Chromatogram of gel filtration on Superdex200 10/300 24 ml column. B) Coomassie stained 12% SDS-PAGE of CDCA7 GF, 0.5% of the total sample loaded; the bands in blue frames were excised from the gel and processed for MALDI-TOF mass spectrometry.

After GSH resin, the sample was applied on a Superdex200 10/30 24 ml column to further separate CDCA7 from contaminants by size exclusion chromatography (SEC). Even though prior to gel filtration the sample appeared ~80% pure, the chromatography run revealed a more complex sample with multiple contaminating bands (***Fig. 6.12 A***). An SDS-PAGE showed that the second peak that consisted of a purified protein ($A_{260/280}$ was 0.6) was CDCA7 with the same smaller bands present in the input, as well as extra contaminants (***Fig. 6.12 B***). The bands were present in the sample before SEC, however, as can be seen in the input, CDCA7 and the smaller bands were the dominating proteins in the sample, but after SEC the portion of CDCA7 was reduced compared to the contaminants (***Fig. 6.12 B***). This may indicate non-specific interaction of CDCA7 with the dextran-agarose matrix. Another observation from SEC was the elution volume of CDCA7, which in the present run was 14.4 ml. 96 kDa LSH that was previously purified on this column and eluted at 14.2 ml, whereas 44 kDa CDCA7 eluted at 14.4 ml, indicating a negligible difference between the protein sizes. This indicates that either CDCA7 formed dimers of around 88 kDa or it co-eluted as a complex with the smaller MW contaminants. If those contaminants were the smaller bands of around 27-29 kDa present on SDS-PAGE, a complex with CDCA7 would be of around 100 kDa. However, this number is an approximation, since SEC elution volume does not directly depend on molecular mass but

rather on the hydrodynamic radius of a protein, nevertheless, an estimation can be made that CDCA7 eluted in a complex with smaller MW proteins.

Therefore, it was decided to confirm the presence of CDCA7 and to reveal the identity of the bands by mass spectrometry (***Fig. 6.12***). Tryptic digest of the CDCA7 and the smaller bands was followed by MALDI-TOF (***see Section 2.4.1***). Peptide spectrum of the CDCA7 band was, indeed, matching that of murine CDCA7 as revealed by the ProteinProspector database (not shown). However, that smaller bands turned out not to match that spectrum. Therefore, Mascot software was used for identification of those proteins by a probabilistic scoring algorithm (Pappin et al., 1993). It revealed that the smaller bands were 14-3-3-ε (epsilon) and 14-3-3-ζ (zeta). Since CDCA7 is predicted to be mostly disordered and *ab initio* modelling did not demonstrate a distinct globular shape, its elution profile on SEC (***Fig. 6.12***) followed the parameters of a globular protein, therefore, it can be hypothesised that in complex with the 14-3-3 proteins CDCA7 adopts a more compact globular structure or formation of extensive contacts, which may also explain its resistance to extraction even with higher salt (***Fig. 6.11***) and to elution with 20 mM GSH (***Fig. 6.10***).

This result is consistent with the earlier finding by Gill et al., who identified 14-3-3 bound to CDCA7 when the latter was phosphorylated at T163 (Gill et al., 2013). This interaction sequesters CDCA7 into cytoplasm (Gill et al., 2013). 14-3-3 is a family of adapter proteins with a multitude of binding partners and a range of signalling functions. The fact that CDCA7 is tightly bound to 14-3-3 proteins may indicate that it is phosphorylated, as would be expected following insect cell expression, and that is may have a specific conformation or state relevant for cytoplasmic but not necessarily nuclear environment. However, this hypothesis would require empirical evidence.

In an attempt to remove contaminating bands, I tried to purify CDCA7 on the ÄKTApure system, which would allow a more stringent purification process with less protein loss. Prior to loading the sample onto a GSTrap column, the cell pellet was lysed at 300 mM NaCl concentration and incubated with 25 U/ml benzonase for 1.5h to digest DNA and RNA present in the sample. Following that, the sample was slowly loaded on a GSTrap FF 1 ml column at 100 µl/min and then incubated overnight with 3C for on-column cleavage. Following elution of the protein from the column, it was washed with 20 ml of 20 mM GSH-containing buffer. The resulting trace revealed extensive contamination with nucleic acids (***Fig. 6.13 A, B***).

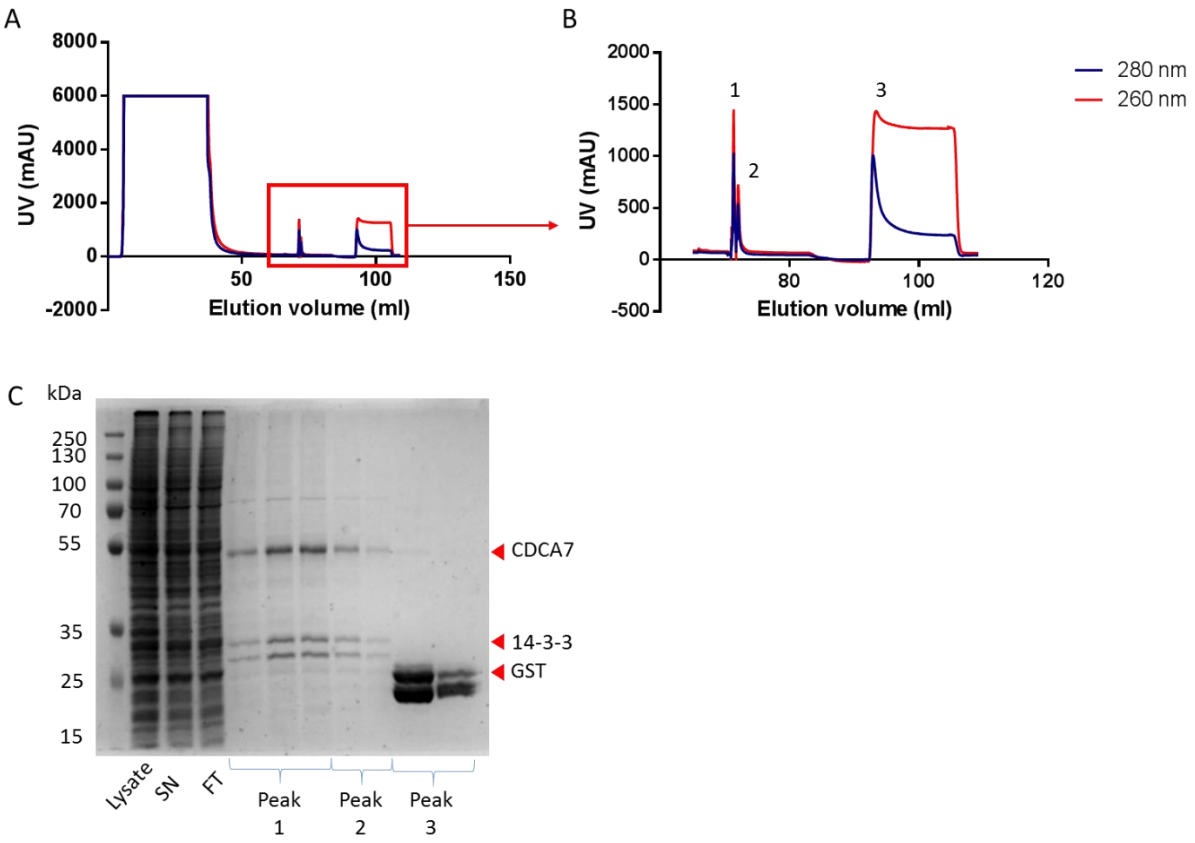


Figure 6.13 Purification of CDCA7 on GSTrap column. A) Chromatogram of a GSTrap run showing application of 40 ml lysate with subsequent column wash, incubation with 3C overnight and 20 ml wash with 20 mM GSH.B) Magnified section of the chromatogram shown on A. Peaks 1 and 2 correspond to CDCA7, as revealed by Coomassie stain, peak 3 corresponds to GST eluted from the column. C) Coomassie stained 12% SDS-PAGE of CDCA7 GSTrap run, 0.5% of the total sample loaded.

Since the previous purification attempts demonstrated nucleic acid contamination, it was decided to increase the NaCl concentration at the lysis stage to disrupt interactions with the contaminants. I chose 500 mM and 1 M NaCl. As can be seen from GST pull-downs, CDCA7 was present in the soluble fraction and bound to the resin, indicating that high salt did not cause protein precipitation at either concentration or regardless of glycerol presence (***Fig. 6.14 A, B***). Since it was not possible to measure the extent of nucleic acid contamination from a pull-down, I decided to use 1 M NaCl in the lysis buffer in the large-scale prep to increase the chances to disrupt any interactions with contaminants. I also decided to keep the GST tag and to attempt elution with 20 mM glutathione following a rationale that high salt may release the protein from contaminants that provided protein stability, therefore, it would be beneficial to keep the tag.

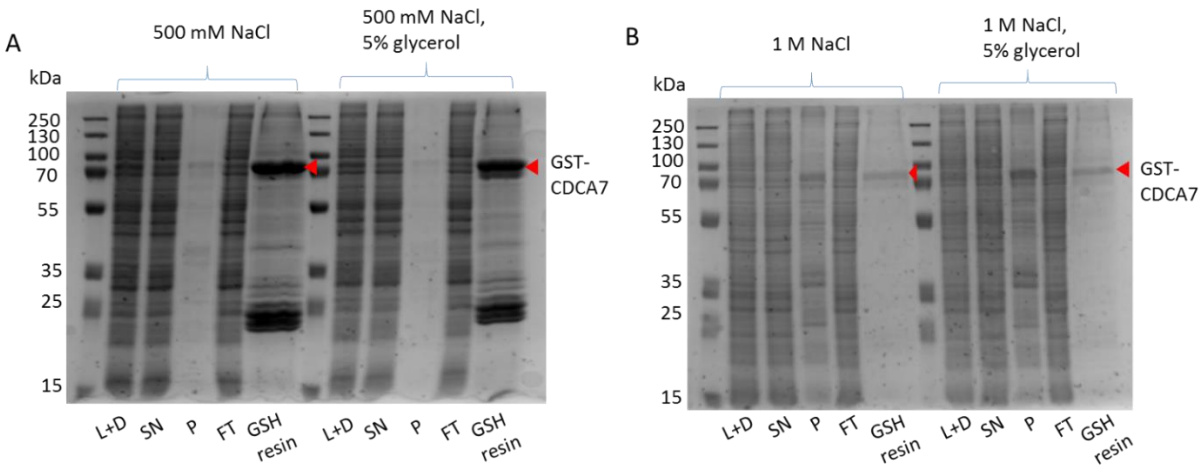


Figure 6.14 12% SDS-PAGE of fractions from GST pull-down of CDCA7 at NaCl concentration indicated above each gel. 15 ml of Sf9 culture following a 72h test expression with V1. L+D – lysate incubated with DNAseI for 30', SN – supernatant, P – pellet, FT – flow-through.

As in the previous attempt to elute GST-CDCA7 using 20 mM GSH (***Fig. 6.10***), the protein could be eluted only partially (***Fig. 6.15 A***). Nevertheless, the elution fractions were pooled (~10 ml in total) and concentrated using PES concentrators to 1 ml. However, I came across two problems: either due to high salt or due to protein's electrostatics, it got stuck to the concentrator membrane, and then the rest got stuck to the column (elution was in 1 M NaCl), so the protein peak on the chromatogram represents only GST (***Fig. 6.15 C***). This may be due to other contaminants being present on the column from previous runs or the way protein behaves on this resin.

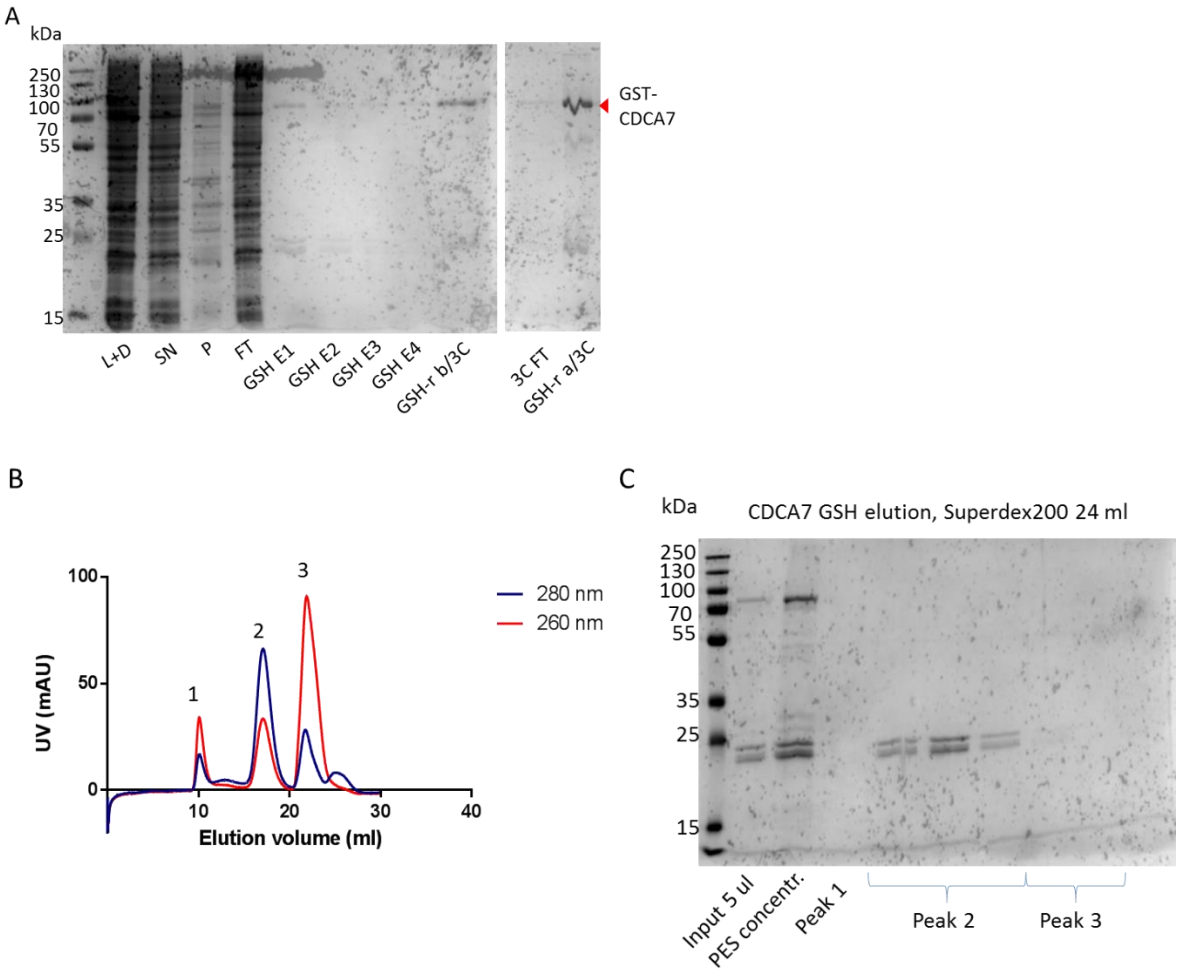


Figure 6.15 GST resin purification step and gel filtration of CDCA7. A) Sf9 lysate was processed and incubated with GSH resin for 2h and subsequently with 3C overnight. L+D – lysate incubated with DNAseI for 30', SN – supernatant, P – pellet, FT – flow-through, GSH E1-4 – GSH elutions, GSH-r b/3C – GSH resin before 3C cleavage, 3C FT – flow-through after 3C cleavage, GSH-r a/3C – GSH resin after 3C cleavage. B) Chromatogram of GST-CDCA7 gel filtration on Superdex200 10/300 24 ml column. B) Coomassie stained 12% SDS-PAGE of CDCA7 GF, 0.5% of the total sample loaded.

Since there was a concern that 1M salt may cause protein precipitation and the usefulness of such high salt for removing contaminating DNA was not established form the previous purification attempt, I used 500 mM NaCl in the next purification and this time to repeat GST tag cleavage (***Fig. 6.16***). Following an overnight incubation with 3C, CDCA7 was present in the flow-though and on the resin. However, this time pooling the fractions was avoided, since previously PES resin caused protein precipitation.

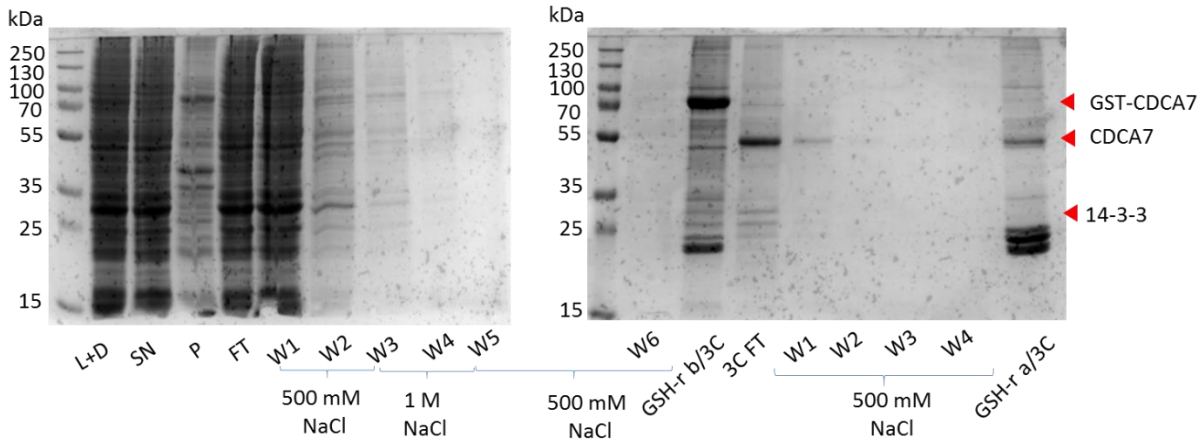


Figure 6.16 CDCA7 purification with GSH resin and 3C cleavage of GST tag at 500 mM NaCl concentration. L+D – lysate incubated with DNAseI for 30', SN – supernatant, P – pellet, FT – flow-through, W – wash, GSH-r b/3C – GST resin before 3C incubation, GSH-r a/3C – GST resin after 3C incubation.

Measurement of the UV absorption of the sample at 260 nm and 280 nm (using the Nanodrop) again indicated nucleic acid contamination ($A_{260/280}$ ~2), therefore, I decided to repeat incubation with DNAseI, rationalising that an additional incubation may be more efficient due to removal of extra contaminants that might have prevented DNAseI from efficient digest. Therefore, a fraction of semi-pure CDCA7 sample was incubated with DNAseI for 30' on ice (***Fig. 6.17***), however, that did not affect $A_{260/280}$, indicating that DNA may either be an important stabilising component of CDCA7 or it binds non-specifically. There was a band of around 25 kDa that initially appeared as a small contaminant, but with additional DNAseI incubation its proportion in the sample increased (***Fig. 6.17***). DNAseI has MW of around 30 kDa and is more likely to be the band the one running at 35 kDa just above 14-3-3 proteins. Therefore, that smaller band could be a degradation product of either CDCA7 or DNAseI, which could be identified by mass spectrometry, however, I did not pursue that question.

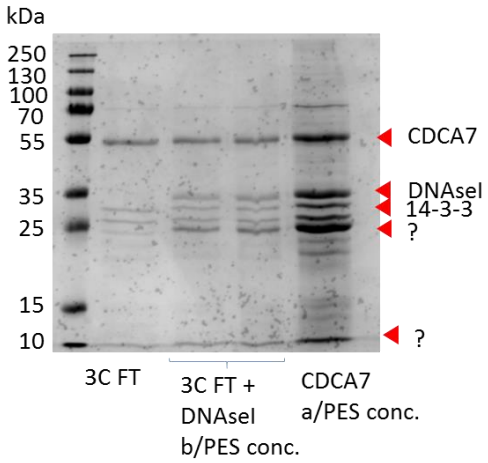


Figure 6.17 Repeated incubation of semi-pure CDCA7 with DNAseI.

Another attempt to remove contaminating nucleic acid was to use a heparin column. After the initial GSH resin step (not shown) the sample was applied on a HiPrep 26/10 desalting column (not shown) to reduce NaCl concentration from 500 mM to 150 mM and then on a HiTrap Heparin HP 1 ml column (***Fig. 6.18 A***). Reduction in ionic strength allowed to remove a contaminant, which on a Coomassie-stained SDS-PAGE gel turned out to be 14-3-3 (***Fig. 6.18 B***). However, $A_{260/280}$ was still around 1, indicating that nucleic acids were not efficiently removed. Comparison of the input and eluted CDCA7 indicated that the protein was not inhibited on the column (***Fig. 6.18 B***).

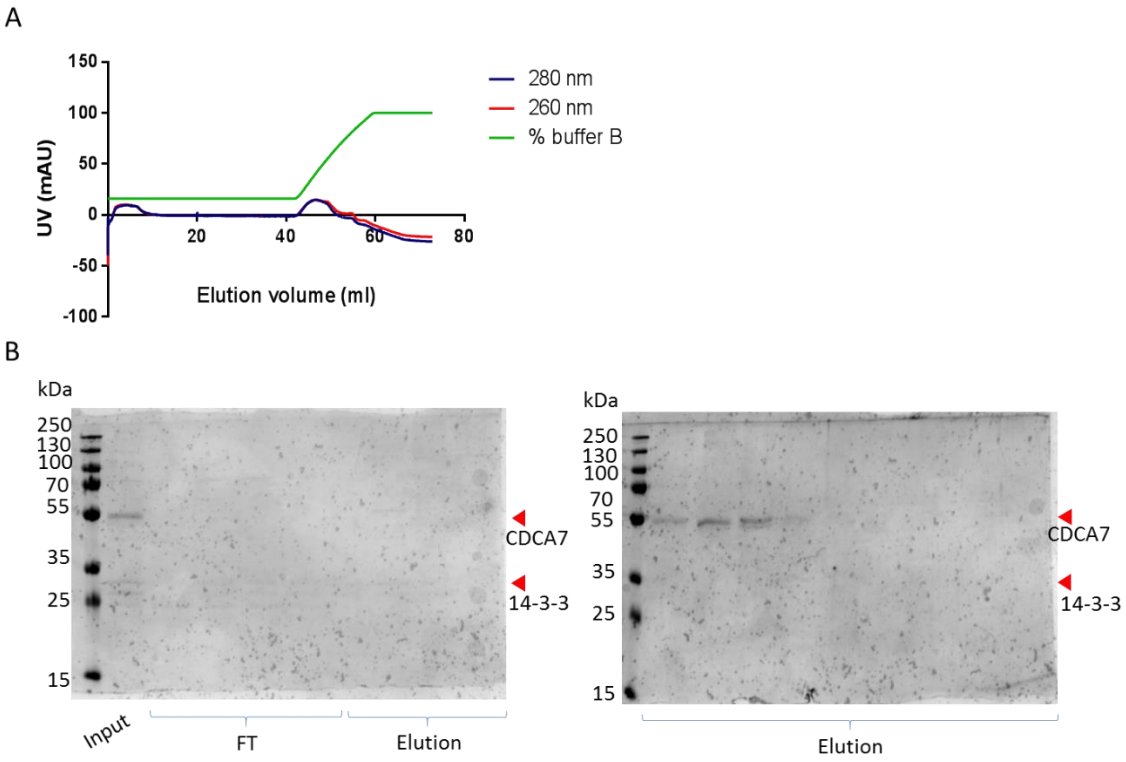


Figure 6.18 CDCA7 run on a HiTrap heparin HP 1 ml column. A) Chromatogram of CDCA7 eluted with gradient 0.15-2M NaCl. B) Coomassie stained 12% SDS-PAGE following the run on HiTrap heparin HP 1 ml column.

Since GSH purification step resulted in a protein that was around 90% pure, I decided to only use that step to avoid any potential problems with downstream purification steps. I repeated preparation of Sf9 lysate, incubation of the soluble fraction with GSH resin for 2h and then with 3C-His overnight at 4°C. As previously, 3C cleavage was efficient and resulted in CDCA7 with some 14-3-3 contamination (***Fig. 6.19 A***).

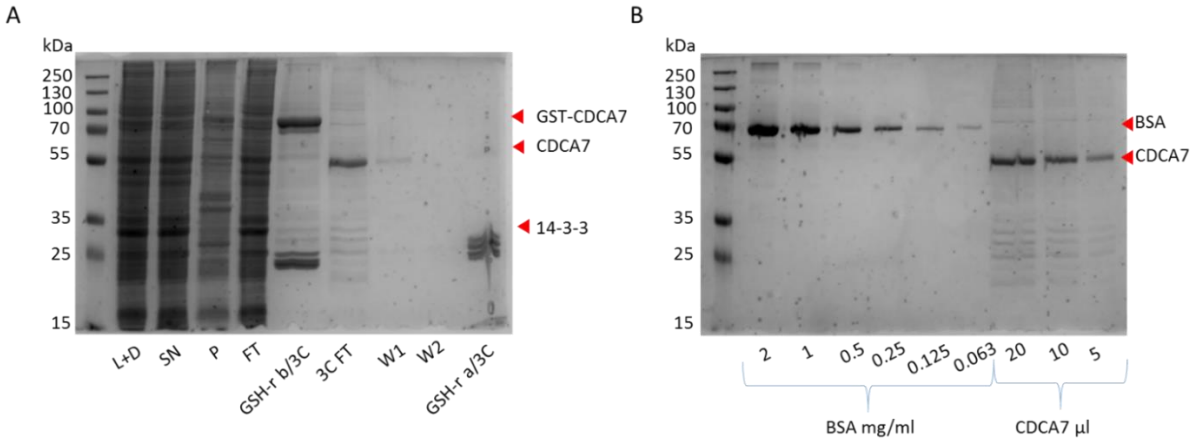


Figure 6.19 Coomassie stained 12% SDS-PAGE of partial CDCA7 purification. A) L+D – lysate incubated with DNAseI for 30', SN – supernatant, P – pellet, FT – flow-through, GSH-r b/3C – GST resin before 3C incubation, W 1-2 – washes, GSH-r a/3C – GST resin after 3C incubation. B) CDCA7 quantification using BSA standard.

Since Nanodrop reading was not reliable due to nucleic acid contamination, I used a BSA standard for CDCA7 quantification (***Fig. 6.19 B***). A BSA titration from 2 to 0.063 mg/ml was ran alongside different loading volumes of CDCA7 on an SDS-PAGE gel. Following a thorough destaining, the gel was scanned, and band intensities quantified using ImageJ software. CDCA7 amounts were derived from the BSA standard curve (not shown). According to this quantification approach, 128 µg of CDCA7 were obtained from 3.9 g of wet pellet, with final concentration 0.043 mg/ml and molarity 1 µM (higher concentration could not be achieved due to the problem with PES concentrator). This relatively low yield demonstrated the need for CDCA7 expression optimisation, however, the lack of time did not allow to do this within the present project. This low yield also prevented a mass spectrometry identification of post-translational modifications, therefore, I could not confirm whether CDCA7 was, indeed, phosphorylated.

6.4 CDCA7 cloning, expression and purification in *E. coli*

6.4.1 CDCA7 cloning for *E. coli* expression

Along with the expression and purification attempts in Sf9 cells, I decided to try expressing CDCA7 in *E. coli*, since yields from bacterial expression are often better than from insect cells. For that expression, I used ligation-independent cloning (LIC) technique to produce three constructs: full length CDCA7 (GST-CDCA7_FL) and two truncated versions – GST-CDCA7_T1

and GST-CDCA7_T2, all with an N-terminal GST tag (***Fig. 6.21***). The truncations were designed based on predicted domain boundaries from UniProt and COILS (see Section 5.1 of Results). The first truncation (GST-CDCA7_T1) was missing the first 63 residues, which comprise a long disordered region (***see Section 6.2***). The truncation removes two N-terminal α-helices and is located in a disordered region, as predicted by PsiPred (***Fig. 6.6***). These regions often contribute to protein instability, since they are one of the main sites for protease activity. Hence, I hypothesised that removing the N-terminus may result in a more stable protein with all putative functional domains intact. However, this truncation still leaves a disordered region leading up to the first coiled coil predicted by COILS and MARCOIL (***Fig. 6.2 and 6.3***), which may also be targeted by proteases. Therefore, various truncation boundaries, in which the optimal region upstream the first coiled coil is remove, can be tested in further studies.

The second truncated version (GST-CDCA7_T2) was lacking a further 87 residues and in which the first coiled coil has been removed (***Fig. 6.21***). The rationale behind this truncation was that it may serve as a control for potential interaction studies using LSH, since the second coiled coil was still in place and it may also participate in the interaction with LSH, hence, removing one of the CCs would allow to dissect their roles in protein-protein interaction. According to PsiPred, the N-terminal part of this truncation falls into a disordered region (***Fig. 6.6***).

I also planned to clone only the zinc finger domain, which would be a crucial negative control for any protein-protein interaction studies but cloning of that construct was unsuccessful (not shown). The disordered region upstream the zinc finger domain could be further tested for the optimal truncation boundaries, with potential incorporation of the upstream α-helix E258-I264 (***Fig. 6.6***), but the exact boundaries would have to be determined empirically. Nevertheless, in the future studies it would be important to check whether this part interacts with either protein or DNA and if it is capable of facilitate chromatin remodelling by LSH. The three constructs in ***Fig. 6.21*** were verified by Sanger sequencing and used for bacterial transformation.

The disorder profile of these truncations from PONDR and IUPred2A would also be consistent with potential functional regions: the first truncation (Δ1-63) would abrogate the disordered N-terminus, which was not predicted to contact any protein bindings regions, the second truncation (Δ1-150) would remove first predicted region for protein-protein interaction and

third truncation (Δ1-271) would have removed the second region (***Fig. 6.5***) leaving only the zinc finger domain.

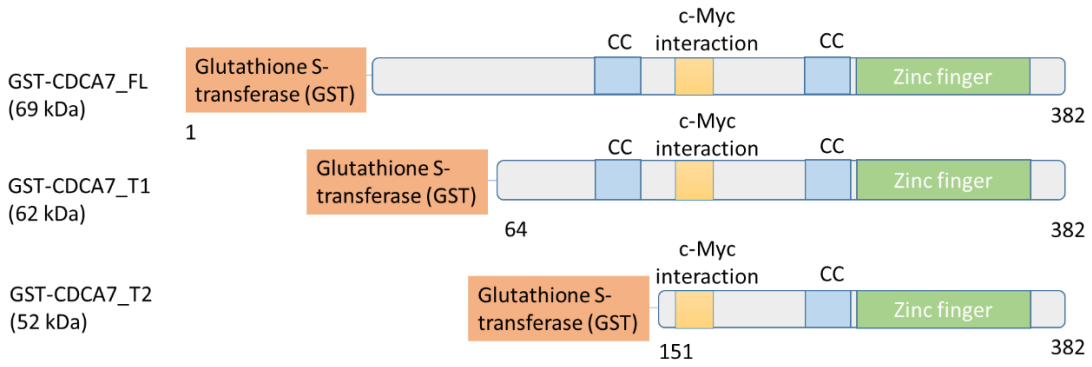


Figure 6.21 Schematic representation of the GST-CDCA7 constructs used in these expression tests and their domain predicted by Pfam. The numbers under the constructs indicate the amino acid residues.

6.4.2 CDCA7 test expression in *E. coli*

One of the main challenges in bacterial expression process is a potentially long expression optimisation process. Selection of an appropriate competent cell strain may have an effect of protein expression. BL21(DE3) is a strain widely used for recombinant protein expression, therefore, it was selected for a test expression, as well as its derivatives C41(DE3) and C43(DE3) that were selected as mutant host strains of BL21(DE3) are suitable for expression of recombinant proteins that can be potentially cytotoxic (Dumon-Seignovert et al., 2004).

Initially, I selected CDCA7_T1 for a test expression (***Fig. 6.22***), and followed a standard protocol, with an overnight induction at 16^{o}C using 0.5 mM IPTG. The expected MW of the protein was 37 kDa without a GST tag and 62 kDa with the tag. A band of that size appeared in a GST pull-down, however, the sample was dominated by a more prominent band of around 35 kDa (with the tag, since it was present in the GST pull-down). This result appeared across all the selected expression strains (***Fig. 6.22***).

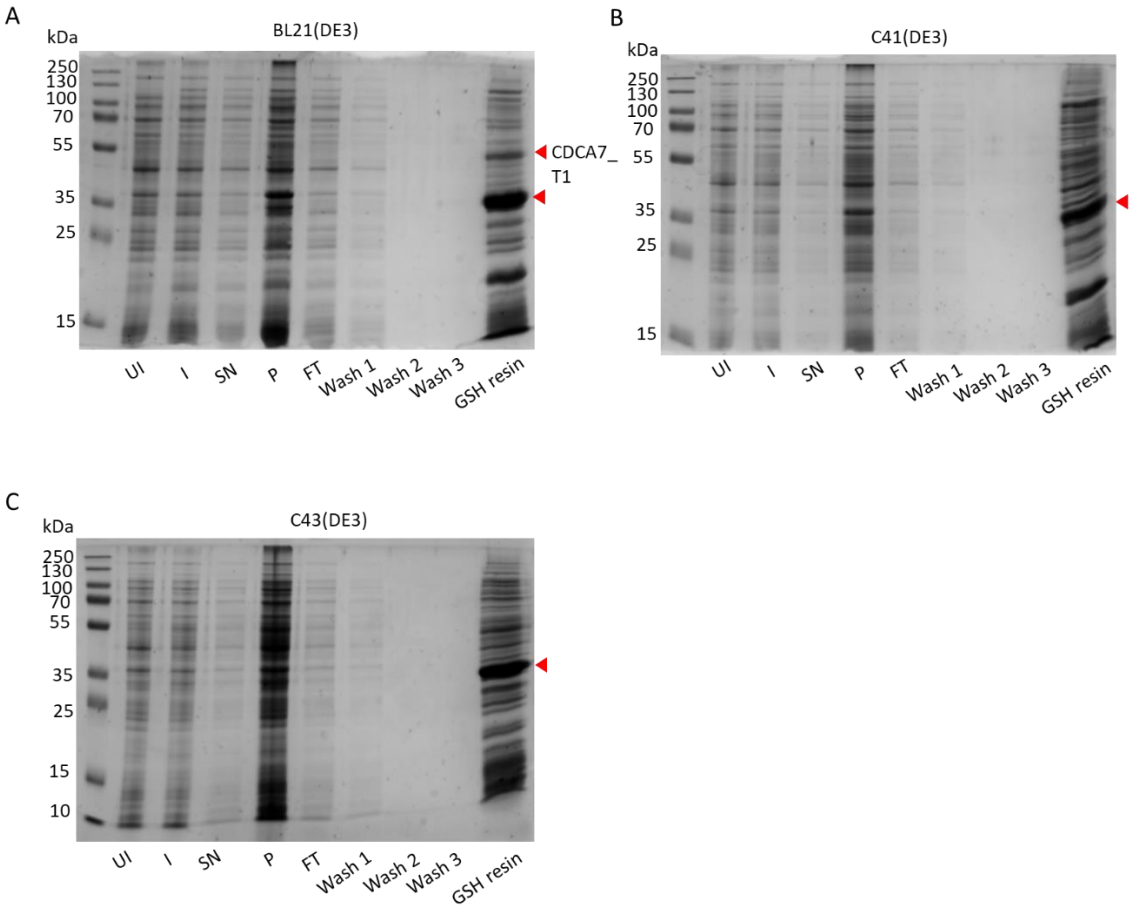


Figure 6.22 CDCA7_T1 test expression GST pull-down after test expression in the indicated strains. UI – uninduced, I – induced, SN – supernatant, P – pellet, FT – flow-through. Unlabelled arrows indicate the most prominently expressed fragment. CDCA7_T1 was identified by a Western blot on Fig. 5.20.

To reveal the identity of the bands, a Western blot was performed on the samples from the test expressions using an anti-CDCA7 polyclonal antibody (***Fig. 6.23***). Unexpectedly, CDCA7_T1 was detected in the lanes corresponding to BL21(DE3) expression, albeit with a molecular weight of around 52 kDa instead of the expected 62 kDa (***Fig. 6.21, 6.23***). A large portion of the protein was present in the pellet. Some soluble CDCA7_T1 was detected in the BL21 (DE3) sample. The large bands visualised by Coomassie staining was not detected by the antibody, indicating that it was either not CDCA7_T1 or that smaller degradation product was lacking the necessary epitopes.

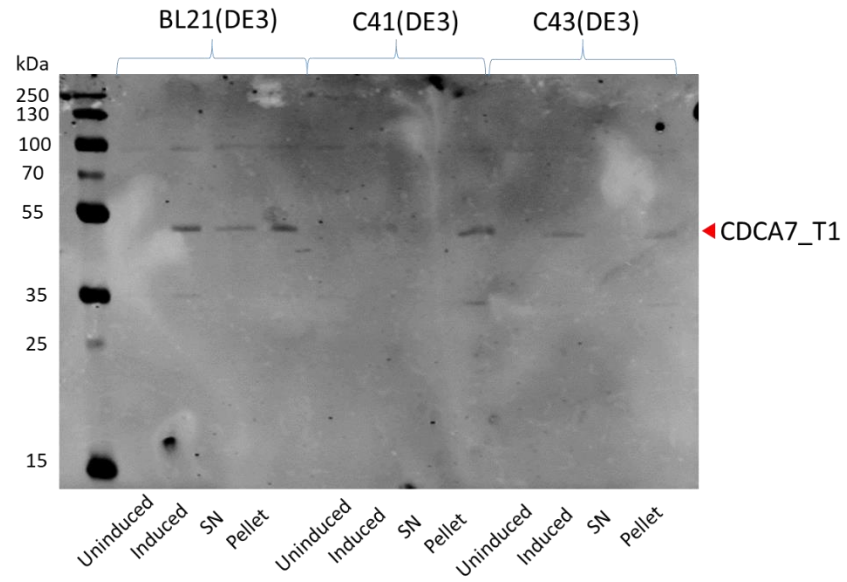


Figure 6.23 Western blot of CDCA7_T1 following test expression in three E. coli strain. Primary antibody – rabbit polyclonal α-CDCA7, secondary antibody – anti-rabbit Licor IR800-conjugated.

Since CDCA7 was detected, although not as an expected fragment, I continued optimisation using a shorter induction time, which could alleviate protein degradation process. This time I tested all three constructs, also adding 0.1% of Triton X-100 (a non-denaturing non-ionic detergent), since previously a lot of CDCA7 was present in the insoluble fraction, which might have been due to incomplete cell lysis. This may be due to only sonication was being used for this preparation, whereas cell disruptor generally provides better yields. Whereas 0.1% Triton X-100 improved samples' purity (***Fig. 6.24***) and it was decided to incorporate it into the lysis and purification process, the bands of incorrect sizes persisted. However, CDCA7_T2 sample contained a smaller band of around 52 kDa, which corresponded to the expected size of the truncated version (25 kDa GST + 27 kDa CDCA7_T2).

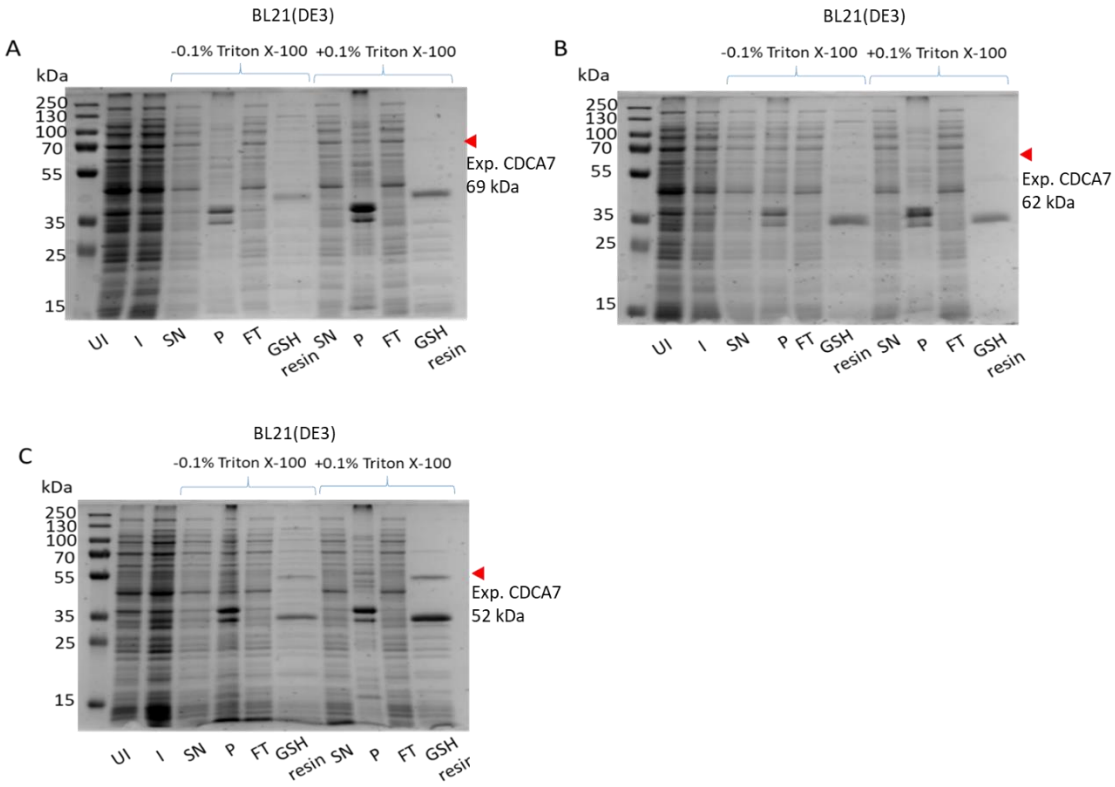


Figure 6.24 GST-CDCA7 full length and the truncations 1 and 2 test expression with and without 0.1% Triton X-100 added to the lysis buffer. UI – uninduced, I – induced, SN – supernatant, P – pellet, FT – flow-through. A) GST-CDCA7_FL was induced in BL21(DE3) at 30°C for 3h; B) GST-CDCA7_T1 was induced in BL21(DE3) at 30°C for 3h; C) GST-CDCA7_T2 was induced in BL21(DE3) at 30°C for 3h. Expected positions of CDCA7 truncations are indicated by red arrows with expected molecular weights.

In order to test more conditions, I also tried expressing the proteins in an *E. coli* strain B843 (DE3) (***Fig. 6.25 B, C***). This strain is auxotrophic, i.e. it cannot synthesise certain organic compounds, such as amino acids, in this respect methionine, which should be added to the growth medium. A supplement of choice in X-ray crystallographic studies is selenomethionine, which, when incorporated into the protein of interest, provides improved diffraction data. The aim of this expression was not to produce a selenomethionine-containing protein, but simply to test which strains can produce desirable proteins, therefore, a standard 2xTY enriched medium was used.

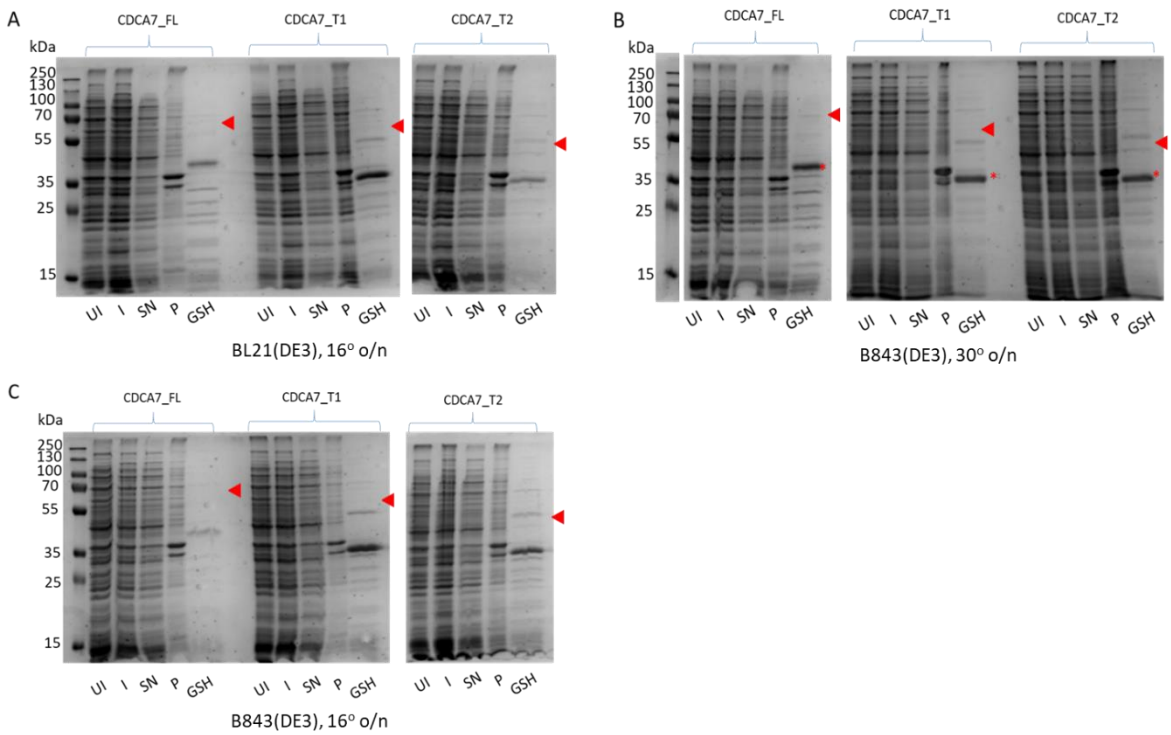


Figure 6.25 GST-CDCA7 expression test at different conditions. UI – uninduced, I – induced, SN – supernatant, P – pellet, GSH – GSH resin. A) Indicated constructs in BL21(DE3) induced with 0.5 mM IPTG at 16°C overnight; B) Indicated constructs in B843(DE3) induced with 0.5 mM IPTG at 30°C for 3h; C) Indicated constructs in B843(DE3) induced with 0.5 mM IPTG at 16°C overnight. The constructs of expected sizes are indicated with red arrows: CDCA7_FL 69 kDa, CDCA7_T1 62 kDa, CDCA7_T2 52 kDa.

This expression test, however, resulted in the bands of around 35-40 kDa on the GST resin lanes (***Fig. 6.25***). No bands of expected sizes were observed with different strains, induction times or incubation temperatures. To elucidate their identity, the bands indicated with a red asterisk on ***Fig. 6.25*** B were excised from the gel, trypsin-digested and processed for MALDI-TOF mass spectrometry. The resulting spectra were compared to the GST-CDCA7 sequence using Protein Prospector and the presence of CDCA7 was confirmed (***Fig. 6.26***), despite the low coverage:

1) **GST-CDCA7_FL**

GST

SPILGYWKIKGLVQPTR**LLLEYLEEKYEEHLYER**DEGDKWRNK**KFELGLEFPNLPYYIDGDVK**LTQSMAIIRYIADKH
NMLGGCPK**ERAEISMLEGAVLDIR**YGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFML
YDALDVVLYMDPMCLDAFPKLVCFKK**RIEAIPQIDK**YLKSSK**YIAWPLQGWQATFGGGDHPPK**SDLEVLFQGPDS

CDCA7

MEARRARQKALKVKNLKDVRYMKLISMETSSSSDDSCDSFASDNFANTRLQLNREGCRTRSQCRHSGPLRVAMKF
PARNTRRAASKKAAPPKPSESSANDSHSDSEEEEEEEEEDGMNFLEKRALNIKQNKAMLAKLMSELESFPGLFSG
RHSLPGHRAKDSKSPRRRTFPGVATRRNPERRTRPLTRSRSRILGSLGALPTEEEEEEEEEEEDKYMLVRQRKSMDSY
MNDDDVPRSRRPGSMTLPHIIRPVEEVTEEEIRNICSNSREKIYNRSLGSTCHQCRQKTTDTKTNCRNPDCWGIRG
QFCGPCLRNRYGEEVKDALLDPNWHCPPCRGICNCSFCRQRDGRCATGVLVYLAKYHGFGNVHAYLKSLKQEFE
MQA

The matched peptides coverage - **13.7%** (84/611AA's).

2) **GST-CDCA7_T1**

GST

SPILGYWKIK**GLVQPTRLLLEYLEEKYEEHLYER**DEGDKWRNK**KFELGLEFPNLPYYIDGDVKLTQSMAIIR**YIADK**H
NMLGGCPKERAEISMLEGAVLDIR**YGVSR**IAYSK**DFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFM
LYDALDVVLYMDPMCLDAFPK**LVCFKK****RIEAIPQIDK**YLKSSK**YIAWPLQGWQATFGGGDHPPK**SDLEVLFQGPD
S

CDCA7_T1

R**HSGPLRVAMKFPAR**NTRRAASKKAAPPKPSESSANDSHSDSEEEEEEEEEEDGMNFLEKRALNIKQNKAMLAKL
MSELESFPGLFSGRHSLPGHRAKDSKSPRRRTFPGVATRRNPERRTRPLTRSRSRILGSLGALPTEEEEEEEEEEEDKY
MLVRQRKSMDSYMNDDDVPRSRRPGSMTLPHIIRPVEEVTEEEIRNICSNSREKIYNRSLGSTCHQCRQKTTDTKT
NCRNPDCWGIRGQFCGPCLRNRYGEEVKDALLDPNWHCPPCRGICNCSFCRQRDGRCATGVLVYLAKYHGFGN
VHAYLKSLKQEFEMQA

The matched peptides coverage - **21.8%** (133/611AA's).

3) **GST-CDCA7_T2**

GST

SPILGYWKIKGLVQPTR**LLLEYLEEKYEEHLYER**DEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHN
MLGGCPK**ERAEISMLEGAVLDIR**YGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLY
DALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLEVLFQGPDS

CDCA7_T2

GRHSLPGHRAKDSKSPRRR**TFPGVATR**RNPERRTRPLTRSRSRILGSLGALPTEEEEEEEEEEEDKYMLVRQRKSMD
SYMNDDDVPRSRRPGSMTLPHIIRPVEEVTEEEIRNICSNSREKIYNRSLGSTCHQCRQKTTDTKTNCRNPDCWGIR
GQFCGPCLRNRYGEEVKDALLDPNWHCPPCRGICNCSFCRQRDGRCATGVLVYLAKYHGFGNVHAYLKSLKQEFE
MQA

The matched peptides coverage - **6.7%** (41/611AA's).

Figure 6.26 Confirmation of full-length (FL), truncation 1 (T1) and truncation 2 (T2) of GST-CDCA7 by MALDI-TOF mass spectrometry with peptide coverage. Identified peptides are in bold.

There are a few possible reasons for the low coverage of the matched peptides. First, trypsin digestion may have been insufficient, therefore, no peptides were present. Second, trypsin

digestion was too efficient. CDCA7 has many lysine and arginine residues, making them highly likely digestion sites (***Fig. 6.27***).

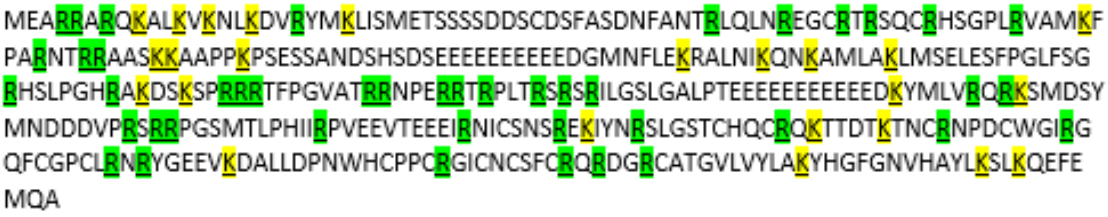


Figure 6.27 Mapping of lysine and arginine residues in Mus musculus CDCA7 full length sequence. Lysines are shown in yellow, arginines are shown in green.

Since the boundaries of the trypsin-digested sample could not be identified by MALDI-TOF, the intact sample was subjected to electrospray ionisation mass spectrometry (ESI-MS) to determine the exact molecular mass of the fragments, which would allow to determine which part of the protein was expressed. However, no meaningful data was obtained (not shown) due to insufficient amount of the protein and its contaminated state.

In the attempt to scale up protein production for a more rigorous purification and for subsequent ESI-MS, I repeated CDCA7 expression scaling up to 1 L culture volumes. This time, instead of cleaving the GST tag with 3C, I tried eluting the protein using 20 mM GSH. This elution was more successful compared with the insect cell sample (***Fig. 6.28 A, B***), however, a large portion of the protein was still retained on the resin, which I then incubated with 3C overnight.

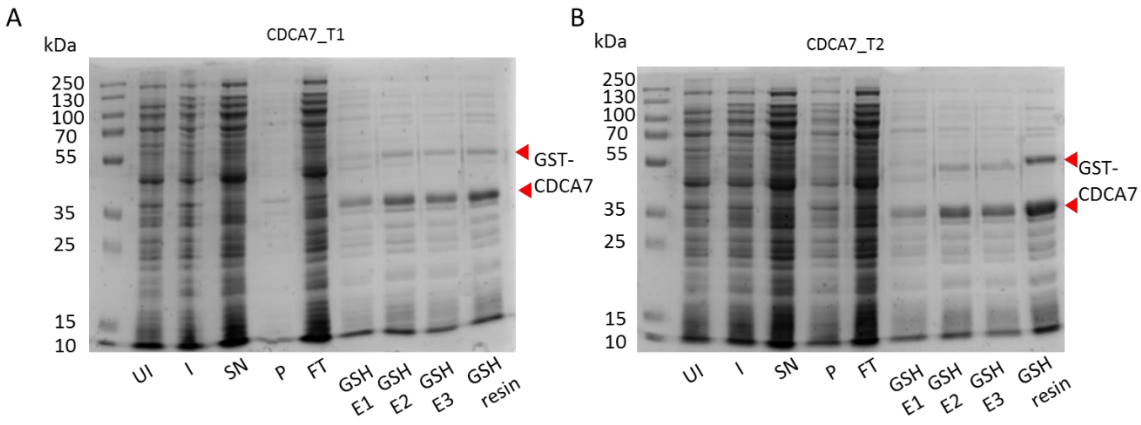


Figure 6.28 Indicated GST-CDCA7 constructs 1L scale up expression. UI – uninduced, I – induced, SN – supernatant, P – pellet, FT – flow-through, GSH E1-3 – elution with 20 mM GSH.

Since the samples from ***Fig. 6.28***, both GSH eluted and cleaved with 3C protease, still had a lot of contaminants present, I attempted to remove the impurities by gel filtration (***Fig. 6.29 A, C and Fig. 6.30 A, B***) using Superdex200 10/30 24 ml column. All samples were concentrated using PES concentrators before applying the samples the column. The shorter version of CDCA7_T1 (35 kDa) eluted at 11 ml in the first peak, indicating that it eluted in a complex with a larger than 100 kDa molecule, presumably DNA, as indicated by Nanodrop absorbance ratio $A_{260/280}$ of 2, confirming nucleic acid contamination (***Fig. 6.29 A, B***). The larger version of CDCA7_T1 (~52 kDa) eluted at around 17 ml in the second peak ($A_{260/280}$ 1.9), in a very dilute state, however, there was another band of 10 kDa of unknown origin, which might have been the main species in the peak 2 (***Fig. 6.29 B***). The discrepancy between the sizes of the two proteins and their elution volumes indicates that the smaller protein may interact with DNA, oligomerise or aggregate. It seems counterintuitive, since I would expect the larger fragment, which presumably retained some ZF domain to interact with DNA and to be present in the peak 1.

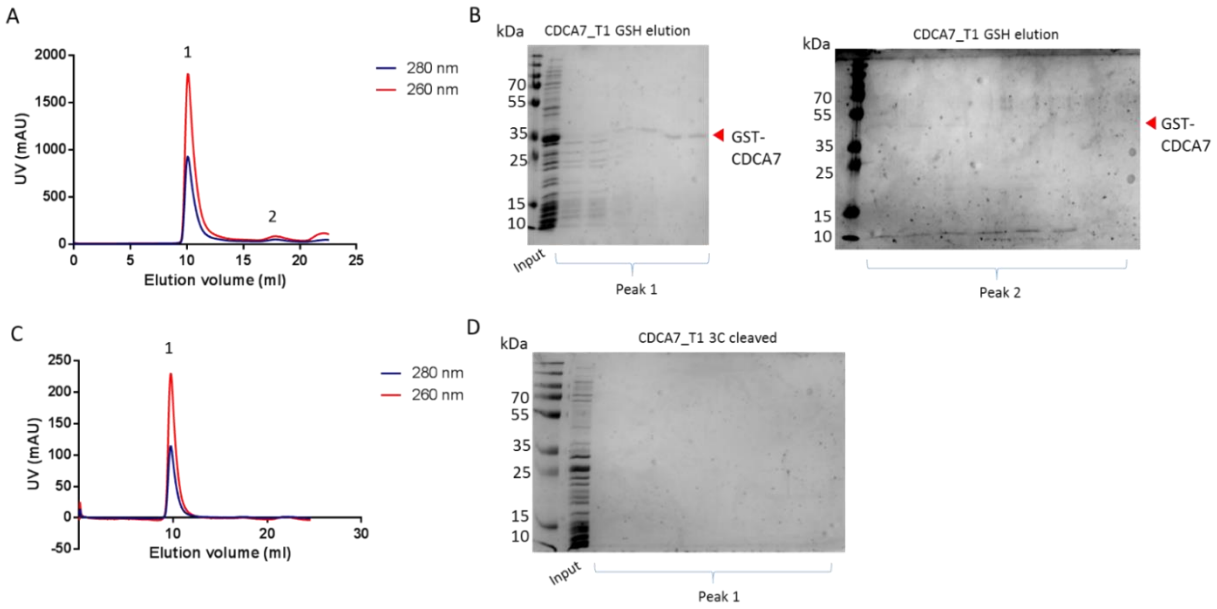


Figure 6.29 GST-CDCA7_T1 and CDCA7_T1 purificaiton by gel filtration. A) Chromatogram of GST-CDCA7 following elution with 20 mM GSH; B) 12% SDS-PAGE of GST-CDCA7 following gel filtration, 0.5% of total sample was loaded; C) Chromatogram of GST-CDCA7 following cleavage with 3C protease; D) 12% SDS-PAGE of 3C-cleaved CDCA7 following gel filtration, 0.5% of total sample was loaded.

A similar situation was observed in CDCA7_T2 gel filtration (***Fig. 6.30 A***), however, it was not possible to discern any CDCA7_T2 constructs on the gel. Therefore, as with the insect cell sample, CDCA7 expressed in bacteria seemed to interact with the resin, and Superdex200 was decided to be omitted from the purification process.

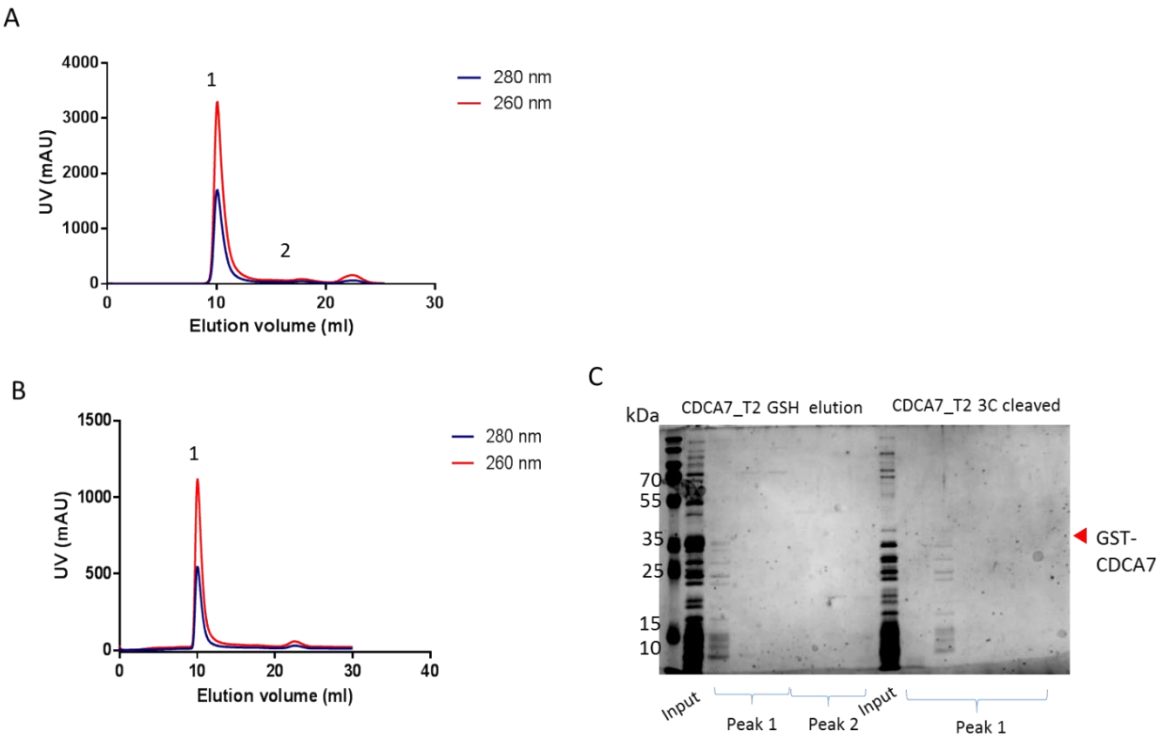


Figure 6.30 GST-CDCA7_T1 and CDCA7_T1 purificaiton by gel filtration. A) Chromatogram of GST-CDCA7 following elution with 20 mM GSH; B) Chromatogram of GST-CDCA7 following cleavage with 3C protease; C) 12% SDS-PAGE of 3C-cleaved CDCA7 following gel filtration, 0.5% of total sample was loaded.

Since it was not possible to express any CDCA7 variants of the correct size, and this time- and labour-consuming did not result in improved yields, I decided to return to the full-length CDCA7 expressed in insect cells and use it for protein characterisation studies.

6.5 Discussion

Little understanding of structural and functional aspects is available for CDCA7, but in the light of the recent findings of its interaction with LSH, this information is crucial for understanding its role in chromatin remodelling (Jenness et al., 2018). Therefore, the first step was to examine its sequence. The N-terminal part of CDCA7 has two coiled coil domains predicted by the servers COILS and MARCOIL (***Fig. 6.2 and 6.3***). These regions are conserved across the CDCA7 orthologues, however, the close homologue CDCA7L does not appear to harbour the second predicted coiled coil (***Fig. 6.7***), indicating a distinct role for that region in

CDCA7. As mentioned earlier, LSH also has a predicted coiled coil in its N-terminus and it is the main candidate region for protein-protein interaction in LSH.

Human CDCA7 was predicted to contain a leucine zipper region upstream its phosphorylation sites (serine-142, threonine-163, serine-190 in human CDCA7) that promote interaction with 14-3-3 proteins (Gill et al., 2013). This leucine zipper location corresponds with the first coiled coil predicted by the both coiled coil prediction servers for murine CDCA7 analysis (***Fig. 6.2 and 6.3***), and since a leucine zipper is a subtype of a coiled coil (Reger et al., 2014; Truebestein and Leonard, 2016), this strengthens the probability that this prediction is valid, highlighting the potential of CDCA7 to participate in homo- and heterodimerisation. The homologue CDCA7L has been found to interact with LEDGF/p75 (Chan et al., 2016), a host factor participating in HIV integration through interaction with DNA via its AT-hook motif (Cherepanov et al., 2003). Therefore, it is not clear whether CDCA7L also interacts with DNA or solely assists LEDGF through protein-protein interaction in stabilising it or in forming contacts with other proteins. CDCA7L-LEDGF complex has been implicated in PI3K/AKT signalling in promoting aggressive medulloblastoma (Chan et al., 2016), however, the roles of CDCA7L domains in this interaction have not yet been dissected.

The C-terminal region of both CDCA7 and CDCA7L was predicted to harbour a zinc finger domain (ZFD) (***Fig. 6.1***) (Bartholomeeusen et al., 2007; Gill et al., 2013). The ZFD of CDCA7L was predicted to belong to the RING ZFD class, as identified through SMART database, and addition of 70-90 μM of $ZnCl_2$ increased pull-down efficiency for LEDGF/p75 interaction (Bartholomeeusen et al., 2007). This domain was termed 4CXXC-R1, where C is a cysteine, X is any amino acid and R1 is an alternative name for CDCA7L (Chen et al., 2005).

About a half of eukaryotic zinc finger domains share a consensus sequence (F/Y)-X-C-X_{2-5}-C-X_3-(F/Y)-X_5-ψ-X_2-H-X_{3-5}-H, where X represents any amino acid and ψ is a hydrophobic residue, which comprises a classical zinc finger domain (Wolfe et al., 2000). The classic C2H2 ZF domain has a common conserved motif TGEKP connecting adjacent ZFs (John H Laity et al., 2000) and transitioning from disordered to ordered state upon binding to DNA (J. H. Laity et al., 2000). Zinc finger proteins rely on CG-rich sequence specificity and sequence-specific DNA deformation with nanomolar binding affinities between ZFD and DNA (Elrod-Erickson et al., 1996; Lisowsky et al., 1999; Paillard, 2004). Some interactions are non-specific and have micromolar affinities (von Hippel and McGhee, 1972). An example of zinc finger-DNA binding was elucidated by X-ray crystallography. The transcription factor EGF1/Zif268 contains the

classical zinc finger domain C2H2 which coordinates three zinc ions and fits α-helices into the major groove, as indicated by its crystal structure (Elrod-Erickson et al., 1996; Pavletich and Pabo, 1991).

It has been observed that RING ZFD target binding molecules are proteins, rather than nucleic acids (Borden and Freemont, 1996; Ravasi, 2003), which is contrary to the findings by Chen et al. that CDCA7 could bind DNA (Chen et al., 2005). The findings from my CDCA7 purification experiments indicate high CDCA7 affinity for DNA, however, this can be due to their non sequence- or base-specific affinity through the extensive positive surface charge of CDCA7 (Gamsjaeger et al., 2007), therefore, the nature of the CDCA7-DNA interaction and its role in LSH-mediated chromatin remodelling is not clear. *In vivo* yeast one-hybrid assay demonstrated that CDCA7L had a putative DNA-binding activity by directly binding the CG-rich sequence 5'-ccg gac gcg cag cCC CGC CCG CCC GCC tac gcg cag-3' found in MAOA core promoter, which contains Sp1 transcription factor binding sites (Chen et al., 2005). This finding is the only one showing the capacity of the ZFD common to CDCA7 and CDCA7L to bind DNA, and it cannot be ruled out that the result was a false positive, which is a common caveat of this method (Reece-Hoyes and Marian Walhout, 2012). Alternatively, the DNA binding activity of CDCA7L may be due to a region distinct from the RING ZFD. Previously, a Polycomb-group related protein Mel-18/bmi-1 containing a RING domain was shown to bind DNA in a sequence-specific manner (Kanno et al., 1995), however, the study used a full-length protein, and the roles of its domains have not been dissected, therefore, it is feasible to suggest that RING domain is not necessarily the one participating in DNA interaction.

Due to the lack of close homologues with structural data, *ab initio* molecular modelling of murine CDCA7 ZFD$^{272-371}$ was performed using the programmes I-TASSER and Phyre2 (***Fig. 6.8***). Among the top threading templates selected by I-TASSER were the crystal structures of the TRIM37 E3 ligase RING domain, the ADD domain of DNA methyltransferase DNMT3A, the E2-E3 ubiquitylation complex and the DNMT3A-DNMT3L complex, which strengthens the link of CDCA7 with RING domain containing proteins. The top model from I-TASSER had average quality parameters (***see Section 6.2***) but was used to examine possible structural characteristics of the ZFD and to compare it with the *ab initio* model from Phyre2 (***Fig. 6.8***). The most obvious discrepancy between the known RING domains and CDCA7 models was the location of cysteine pairs within the structure: if described structures of RING ZFDs showed a cross-brace arrangement of cysteines (Miyamoto et al., 2017), both CDCA7 models

demonstrated sequential localisation of cysteines in two zinc fingers (***Fig. 6.8***) more similar to the arrangement of the classical C2H2 DNA-binding domain or protein-binding LIM domain (Cassandri et al., 2017). However, there is a sequence discrepancy between the classic C2H2 domain and CDCA7 ZFD, since CDCA7 does not obey the rules for containing histidines in zinc-coordinating positions. Histidine residues are located proximally to the cysteines, however, their role in the ZFD arrangement was not clear, as they did not obey previously described sequence patterns (Cassandri et al., 2017; Miyamoto et al., 2017). It is also possible that the cross-brace orientation of the zinc fingers exists in CDCA7 but was not captured in either model. The structures identified by the TM-align server (Zhang and Skolnick, 2005) to be the most similar to the CDCA7 model included the ZFDs of DNMT3A and DNMT3L proteins (***Fig. 6.9***). Their ZFD are predicted to have a GATA-1-like and PHD domain, with the former following a continuous cysteine arrangement coordinating one Zn^{2+} and the latter having a cross-brace arrangement with two zinc atoms (Ooi et al., 2007; Otani et al., 2009). The GATA-1 and PHD ZFDs comprise the ADD (ATRX, DNMT3A, DNMT3L) domain, which binds the histone H3 tail unmethylated at lysine 4 (Ooi et al., 2007; Otani et al., 2009). The DNA-binding domain of DNMT3A is located downstream of the ADD domain and binds DNA with assistance from the methyltransferase-like domain of DNMT3L (Ren et al., 2018; Zhang et al., 2018). Therefore, neither the ZFD of DNMT3A nor DNMT3L are expected to participate in interaction with DNA, which further strengthens the argument against CDCA7 ZFD binding to DNA.

The location of CDCA7 ZFD (***Fig. 6.4***) corresponds to the location of the ICF mutations (***see Section 1.4***). When both arginines were converted to histidines or a cysteine (R285C/H, R315H), these CDCA7 mutants could still maintain interaction with LSH in reticulocyte lysates (Jenness et al., 2018). This indicates either that the ZFD does not participate in LSH interaction or the interaction with LSH can still be maintained by the CDCA7 mutant but it changes LSH conformation. Therefore, LSH can no longer remodel nucleosome. However, an additional argument for CDCA7 ZFD being crucial for DNA interaction comes from a pull-down experiment, where the R285C/H mutations decreased MBP-CDCA7 binding to DNA beads and LSH recruitment (Jenness et al., 2018), indicating that CDCA7 ZFD provides a platform for nucleosome binding. The third ZFD mutation, R315H, abolished DNA binding by CDCA7 *in vitro* but not in chromatin egg extracts (Jenness et al., 2018), which may point to additional binding factors participating *in vivo*.

Zinc finger domains rely on zinc ions coordinated by cysteines and/or histidines. The sulphur atom of cysteine residue or nitrogen on histidine imidazole ring chelate zinc atom providing structural stability (Pace and Weerapana, 2014), therefore, introduction of a neighbouring cysteine or histidine may introduce additional chelating point, thus disrupting the initial tetrahedral geometry of zinc coordination, shifting the domain conformation and impairing the protein function as a result. If introduction of histidines does not directly affect zinc coordination, it could also introduce additional electrostatic points of contact between histidines and nucleobases, making CDCA7-DNA interaction more rigid, however, this mechanism does not explain the experimentally-observed decrease of MPB-CDCA7 in DNA binding (Jenness et al., 2018). Therefore, mutagenesis and structural experiments would be required to dissect the mechanism of CDCA7-DNA binding and the molecular impact of the ICF mutations.

Here, the recombinant murine CDCA7 with an N-terminal GST tag was purified from *Sf9* insect cells, albeit with a range of caveats. First, a significant DNA contamination was observed after GSH resin and 3C cleavage step. This contamination was not affected by incubation with either benzonase or DNAseI. Therefore, it has a potential to complicate any downstream experiments involving DNA binding, since the binding sites on CDCA7 may be already occupied by insect DNA. Second, CDCA7 was always co-eluted with a pair of proteins, which was revealed to be 14-3-3-ε (epsilon) and 14-3-3-ζ (zeta) by MALDI-TOF mass spectrometry. Third, any residual contaminations could not be removed by gel filtration due to apparent interaction of CDCA7 with the column matrix.

Therefore, an up-to-date standard protocol for murine CDCA7 purification includes standard pellet preparation with lysis in high NaCl (500 mM), followed by a 2-3 h incubation with GSH resin and an overnight incubation with 3C protease. This method allowed to obtain CDCA7 in complex with 14-3-3 proteins with an overall sample purity of around 90-95% (by visual estimation). This level of purity allows to perform biophysical assays to characterise the protein, however, they are complicated by the low protein yield. Biochemical assays aiming to reveal interaction patterns of CDCA7 would be complicated by the contaminating DNA.

CDCA7 appears to be a strong DNA binder, as revealed by persistent DNA contamination. This interaction is most likely to be mediated through the zinc finger domain, however, that could only be confirmed by experimental work involving, for instance, C-terminal truncations of CDCA7. The present result indirectly supports the hypothesis that CDCA7 may allow LSH to

perform chromatin remodelling by acting as an additional point of contact with DNA, mimicking the function of DNA-binding domains of the other chromatin remodellers.

As revealed previously, phosphorylated human CDCA7 was interacting with 14-3-3 protein, which resulted in the sequestration of the former into the cytoplasm. Here, a mass spectrometry analysis revealed that heterologous murine CDCA7 and endogenous 14-3-3-ε and – ζ were bound to CDCA7 with high affinity which was not disrupted by 500 mM NaCl. This indicates that CDCA7 is likely to be phosphorylated on serines 197 and 225 and threonines 170 and 203 as identified by similarity in the bioinformatics analysis (see Section 1.4 of Introduction). The PTMs were not confirmed in the scope of the present study due to a low protein yield.

Attempts to remove nucleic acid contamination from CDCA7 sample have not yet been successful. They included increasing salt concentration in lysis and wash buffers, varying nucleases treatment conditions and a chromatography run with heparin column. Additional methods could be tried. For example, polyethyleneimine (PEI) precipitation with subsequent ammonium sulfate precipitation. Even though a heparin column was not successful in removing nucleic acid, ion exchange may still be tried. An anion exchange column may be used. CDCA7 has an overall positive charge, however, when it is bound to DNA, the positively charged residues may be masked by DNA and negatively charged residues more exposed, allowing CDCA7 to bind the anion resin. Otherwise, cation exchange chromatography can be tried. Finally, addition of EDTA to the lysis buffer may release (completely or partially) Mg^{2+} bound to the protein, which may cause conformational change in the protein and release of DNA, however, that can also destabilise the protein.

In the attempt to improve CDCA7 yields, I explored the option of expressing it in bacterial cells. I designed three constructs with various N-terminal truncations that could potentially stabilise the protein. Even though Western blot and mass spectrometry confirmed the presence of CDCA7, none of the expressed fragments had the correct molecular weight, indicating degradation. All three constructs retained the GST tag, as indicated by their ability to bind GSH resin. CDCA7 expressed in insect cells, presumably retained post-translational modifications (PTMs). These modifications may play a crucial role in stabilising the protein, therefore, the lack of PTMs in the bacterially expressed CDCA7 may explain the observed degradation. Further studies of CDCA7 PTMs are likely to shed more light on the CDCA7 structure and its stability.

7. INVESTIGATION OF THE LSH-CDCA7 INTERACTION

7.1 Introduction

Jennes *et al.* (2018) showed that the LSH-CDCA7 interaction is required for nucleosome remodelling and their co-immunoprecipitation experiments demonstrated that LSH and CDCA7 directly interact (Jenness et al. 2018). However, it was not clear whether the proteins form a stable complex and whether the nucleosome core-particle was necessary to mediate their interaction. In their experiments with protein co-depletion from *Xenopus* eggs, LSH depletion did not cause depletion of CDCA7 and *vice versa,* indicating that only a fraction of the proteins may interact (Jenness et al. 2018). Nevertheless, it is feasible that the LSH-CDCA7 complex still exists if only in association with nucleosome. Elucidating the mode of interaction of LSH and CDCA7 *in vitro* would require a stable complex, which would allow to study the structural and molecular aspects of the interaction and to obtain functional insights. If they do not form a stable complex, then additional techniques such as chemical crosslinking and/or gradient fixation could potentially be used. Therefore, I tested interaction using mouse homologues of LSH and CDCA7 in the absence of the nucleosome by a GST pull-down, as well as by co-infection or pellet co-lysis.

7.2 LSH-CDCA7 interaction was detected by a GST pull-down

A GST pull down was chosen over an Ni-NTA pull-down due to the higher specificity of the former and higher protein background of the latter. An Ni-NTA pull-down was only used to detect expression of LSH-6xHis. Lysates were used as a quick way to check if any interaction can be detected without going into a length protein purification process. Additional advantage of a pull-down from lysates is that the proteins of interest may already be bound to any necessary accessory proteins or nucleosome, which facilitate the interaction between LSH and CDCA7.

Therefore, *Sf9* lysates containing either LSH-6xHis or GST-CDCA7 were prepared and incubated with GST resin. The control samples included LSH and GST-CDCA7 (both from lysates), purified GST and these were compared with a mix of LSH-6xHis and GST-CDCA7 from Sf9 lysates. Both the input samples and the protein left on the beads after washes were resolved on 12% SDS-PAGE gel and blotted using antibodies against LSH, CDCA7 and GST. To avoid antibody background and any potential cross-reactivity, I cut the membrane at around 35 kDa mark and blotted LSH and CDCA7 separately from GST. Strong LSH bands were present

on all input lanes (***Fig. 7.1 lanes 1, 3, 4***), compared with the GSH resin samples (***Fig. 7.1 lanes 5-8***), indicating that LSH expression was more efficient than GST-CDCA7. A small amount of LSH was detected on the GSH resin (***Fig. 7.1 lane 5***), compared with input (***Fig. 7.1 lane 1***), indicating that there was some non-specific interaction, perhaps due to insufficiently stringent wash. A strong GST-CDCA7 band was present in the GSH resin lane (***Fig. 7.1 lane 6***), consistent with the affinity of the construct for the resin. The LSH+GST lane (***Fig. 7.1 lane 3***) did not show much GST in the input but it was enriched in the GST resin sample (***Fig. 7.1 lane 3***). A strong LSH band present in the input was weak and similar to the LSH only sample (***Fig. 7.1 lane 3***), indicating that LSH has some non-specific affinity for the resin, but not for the GST tag. The lane containing both LSH and GST-CDCA7 had a strong LSH band in the input (***Fig. 7.1 lane 4***), but barely visible GST-CDCA7 band (***Fig. 7.1 lanes 4, 8***). The respective GST lane had a stronger LSH signal compared to the LSH only and LSH+GST controls (***Fig. 7.1 lane 8***), implicating that LSH was retained on the resin due to interactions other than non-specific interaction with GST resin or GST tag. Therefore, it is possible to hypothesise that LSH interacted with GST-CDCA7, enriched on the resin, in the present pull-down.

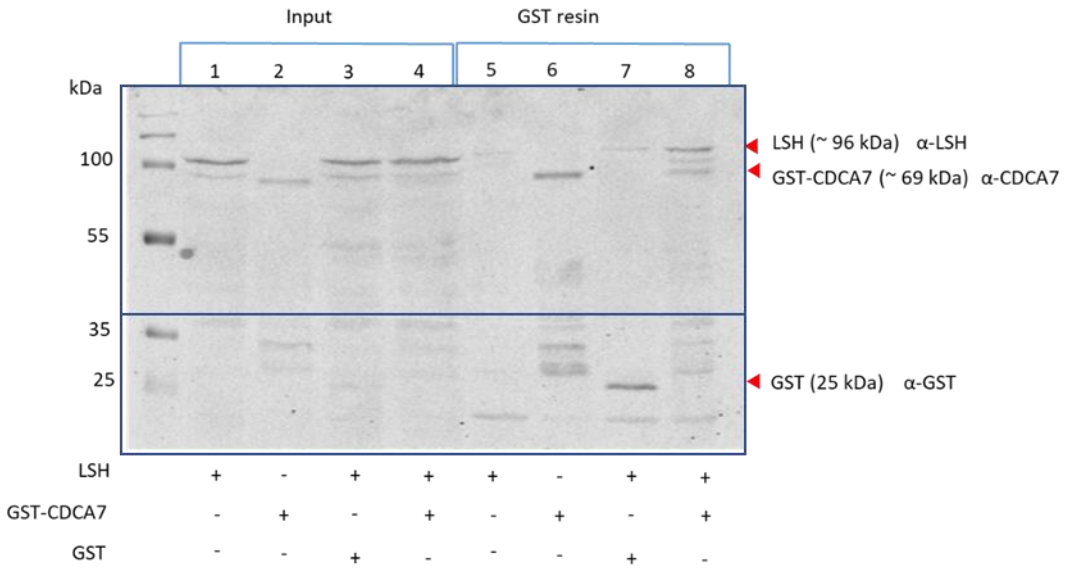


Figure 7.1 GST pull-down of LSH and CDCA7 visualised by Western blot using α-LSH, α-CDCA7 and α-GST antibodies. The blue boxes surrounding the membrane indicate the borders where the membrane was cut in two and probed separately. The top part was probed with primary antibodies α-LSH and α-CDCA7 and secondary IR800-fused α-mouse and α-rabbit, respectively. The bottom part was probed with primary antibody rabbit α-GST and secodary IR800-fused α-rabbit. Extra bands indicate high background signal due to antibody crossreactivity or insufficient membrane blocking or washing.

7.3 LSH and CDCA7 co-expression and co-purification

Since both LSH and CDCA7 had low expression in *Sf9* cells and a significant fraction of both proteins was present as insoluble, I attempted to co-express them to check if the simultaneous production improves their yields and solubility and if they form a nuclear complex that could be purified. The assumption here was that proteins that were shown to interact might form a complex if co-expressed and stabilise each other. I did not clone LSH and CDCA7 as a polycistronic assembly in a single plasmid, instead, I expressed them from individual plasmids, generated baculoviral V0 and V1 and used V1 viral generation for separate expressions of each protein as controls and for co-infection, adding equal volumes of a V1 sample of each to *Sf9* cells (***see Section 2.1.3.6***).

Separate expressions and co-infection were followed by pellet processing and analysis of the soluble fractions by Ni-NTA and GSH pull-downs (***see Section 2.3.4***). All samples were treated with 750 U of benzonase, which could potentially digest the hypothetical bridging nucleosome, however, eukaryotic sample contain large amounts of DNA and without proper nucleic acid digestion downstream purification steps can be problematic. The assumption was that if there was a bridging nucleosome between LSH and CDCA7, its phosphodiester bonds could be protected by the bound proteins from complete digestion. Pull-downs for separately expressed proteins demonstrated expression as indicated by the presence of the bands on the respective resin lanes: LSH-6xHis was present of Ni-NTA resin and GST-CDCA7 was present on the GSH resin (***Fig. 7.2 A, lanes Ni-NTA resin and GSH resin***). When the proteins were GSH probed following co-infection, both proteins were observed in the pellet fraction (**Fig. *7.2 lane P***) indicating their insolubility, without giving any insight whether they formed a complex. No obvious protein bands were detected on the GSH resin lane (***Fig. 7.2***

B), indicating that the proteins were not present in the soluble fraction in sufficient levels to be visualised by Coomassie staining (***Fig. 7.2 B***).

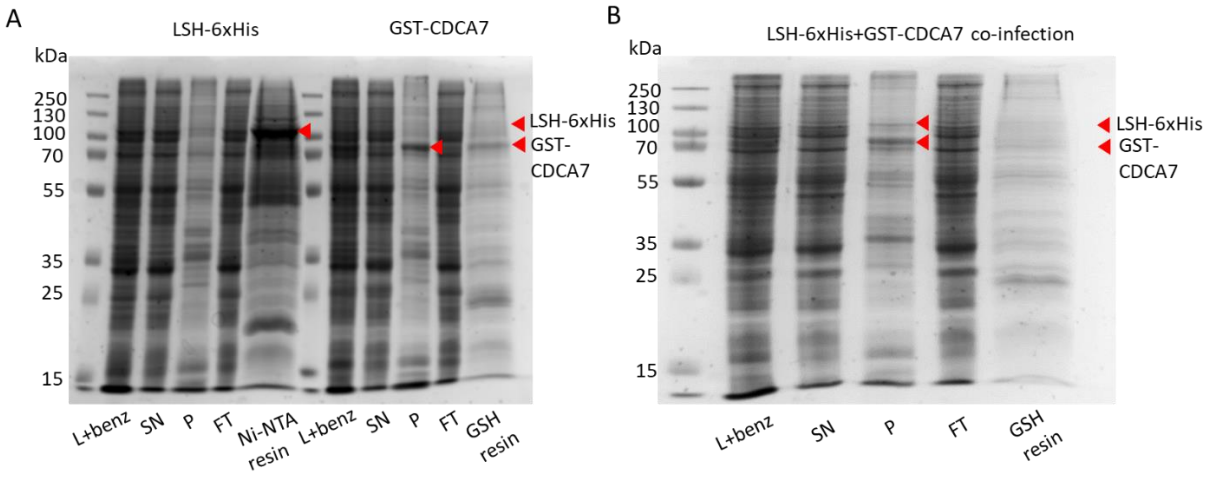


Figure 7.2 Coomassie stained 12% SDS-PAGE of separate expression (A) and co-infection (B) of LSH and GST-CDCA7 in Sf9 cells. L+benz – lysate incubated with 3 µl of 250 U/µl of benzonase;. SN – Sf9 supernatant; P – pellet; FT – flow-through. Wash fractions were not run on these gels. Red arrows point to LSH-6xHis at ~100 kDa and GST-CDCA7 at around 70 kDa.

Following co-infection, I attempted to improve the visibility of the bands by scaling up the cultures. I also tried mixing the pellets containing each protein instead of co-infecting. I used 250 ml of *Sf9* to co-infect it with V1 of both LSH and CDCA7 and 500 ml of Sf9 infected with each protein's virus separately for a subsequent co-lysis.

The pellets of individually expressed proteins that were then mixed together and processed following the standard protocol (***see Section 2.1.3.10***) and the proteins were isolated by a GST pull-down with the subsequent 3C protease cleavage overnight to remove the GST tag. Interestingly, a band corresponding in size to CDCA7 (at 70 kDa) was observed in the 3C flow-through sample, however, the bands corresponding to uncleaved GST-CDCA7 and LSH persisted, indicating that the proteins may show some interaction, which prevented efficient 3C cleavage (***Fig. 7.3***). This assumption was reinforced by the presence of a significant amount of both proteins on GSH resin (***Fig. 7.3 lane 3C o/n GSH***).

When the proteins were expressed by co-infection, GST-CDCA7 was still present in its uncleaved form following incubation with 3C in the 3C flow-through, which should have contained cleaved CDCA7 and LSH. Even though CDCA7 has been efficiently cleaved by 3C

protease, when expressed on its own (***see Section 6.2 and Fig. 7.3***), it appeared to be less prone to cleavage in the presence of LSH (***Fig. 7.3***).

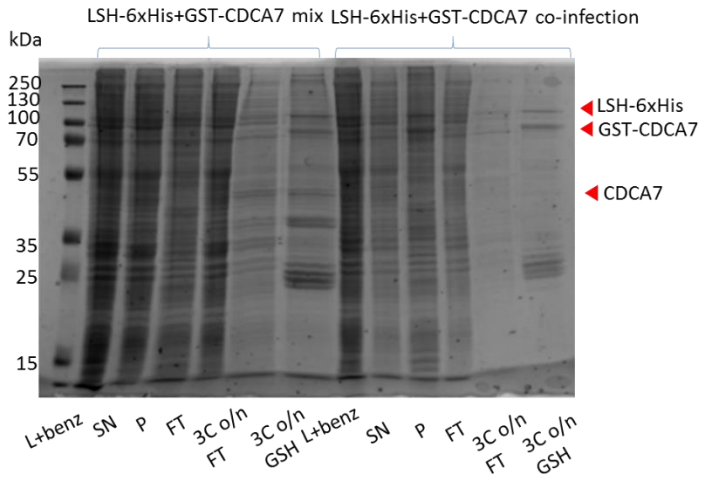


Figure 7.3 Coomassie stained 12% SDS-PAGE of co-lysis and co-infection of LSH and GST-CDCA7 in Sf9 cells. Lanes on the gel are as follows: L+benz – lysate incubated with 3 µl of 250 U/µl of benzonase;. SN – supernatant; P – pellet; FT – flow-through; 3C o/n FT – flow-through following overnight incubation with 3C protease; 3C o/n – GSH resin following overnight incubation with 3C protease.

Since co-infection following the GSH resin purification step did not result in a high yield of the proteins, I decided to try applying the 3C flow-through containing cleaved CDCA7 and potentially a CDCA7-LSH-6xHis complex on a Superdex200 24 ml column. The first labelled peak 1 on the chromatogram had a significant nucleic acid presence (***Fig. 7.4 A, peak 1***) as indicated by $A_{260/280}$ of 1.9. When that peak fraction was run on a 12% SDS-PAGE, it revealed to contain predominantly LSH as indicated by the migration pattern being consistent with the expected mass of LSH (***Fig. 7.4 B***). The peak's elution volume of 10 ml did not correspond to LSH molecular mass (expected elution volume should be around 14.2 ml), which may indicate that LSH was in a complex with a large nucleic acid fragment. A small amount of CDCA7 was detected in peak 2 (***Fig. 7.4 lanes 3, 4, 5***) elution volume of 17 ml), which corresponds to its molecular mass of 44 kDa, after the GST tag cleavage (***Fig. 7.4 A, B***). However, the $A_{260/280}$ corresponding the peak 2 was 1.9, indicating the presence of nucleic acid. If CDCA7 was in a complex with nucleic acid, it would elute earlier due to higher molecular mass of the CDCA7-nucleic acid complex. A possible explanation to this discrepancy may be that there was a DNA fragment similar to CDCA7 in size and steric parameters that eluted together with CDCA7. Peak 3 represented a protein of around 30 kDa (***Fig. 7.4 A, B***), which could correspond to one of the 14-3-3 proteins that usually elute with CDCA7.

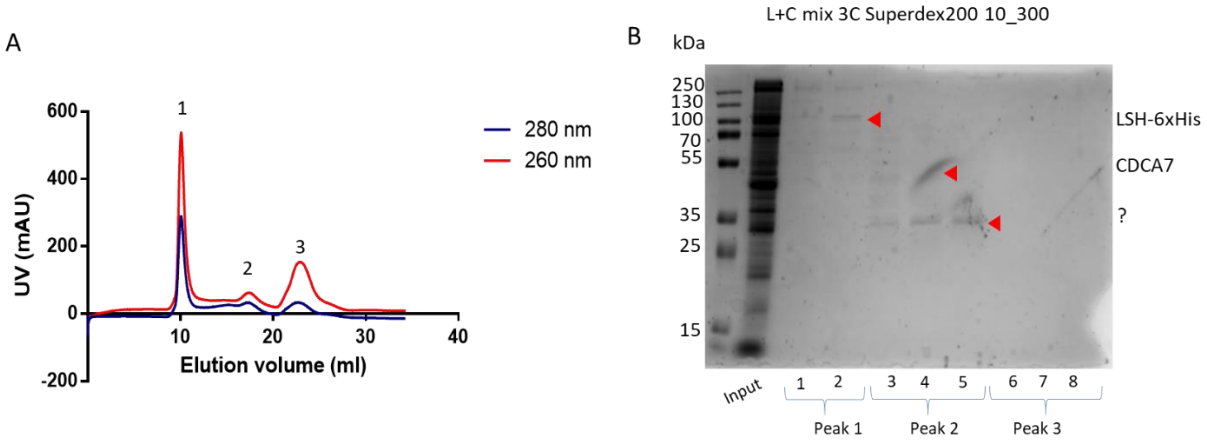


Figure 7.4 Purification of co-lysed samples containing LSH-6xHis and CDCA7. A) Chromatogram of LSH+CDCA7 following gel filtration on a Superdex200 24 ml column; B) Coomassie stained 12% SDS-PAGE of the fractions collected in gel filtration on a Superdex200 24 ml column. Co-lysis of pellets containing each protein expressed individually.

In summary, the LSH-CDCA7 complex was not detected following either co-infection or pellet mixing and co-lysis. Even though LSH was present in the GST pull-down, it may be not due to its association with CDCA7 carrying the GST tag, but due to its non-specific interaction with the GSH resin. As was seen in the GST pull-down from Sf9 lysates, LSH has some affinity for GSH resin (***Fig. 7.1***). These results reinforce the notion that LSH and CDCA7 are likely to interact when a mononucleosome is present and which can bridge the two proteins. The presence of a mononucleosome may trigger conformational changes in either or in both proteins and cause them to form a transient complex, which readily dissociates once remodelling has taken place.

7.4 Discussion

In order to stabilise CDCA7 and/or increase its solubility, I tried co-expressing it or co-purifying it together with LSH, since they were predicted to interact by mass spectrometry and biochemical experiments (Jenness et al, 2018). Neither approach improved CDCA7 solubility or yield, however, 3C protease cleavage of GST tag from CDCA7 N-terminus appeared to be inhibited in the presence of LSH, indicating that LSH may bind CDCA7 either at the N-terminus or its middle part and to induce conformational changes resulting in

protection of the 3C cleavage site. It is unclear, however, whether these results only followed non-specific binding or were a genuine protein-protein interaction event.

During CDCA7 expression, a large part of the expressed protein was pelleted in the insoluble fraction. GST pull-down of the two proteins resulted in larger amount of LSH being bound to GST resin in the presence of CDCA7 compared with LSH only control (***Fig. 6.31***), indicating that LSH is retained on the resin in the presence of CDCA7, therefore, complex formation might have been occurring.

In order to stabilise CDCA7 and/or increase its solubility, I tried co-expressing it or co-purifying it together with LSH, since they were predicted to interact by mass spectrometry and biochemical experiments (Jenness et al., 2018). Neither approach improved CDCA7 solubility or yield, however, 3C protease cleavage of the N-terminal GST tag appeared to be inhibited in the presence of LSH (***Fig. 6.33***), indicating that LSH may bind CDCA7 either at the N-terminus or its middle part and to induce conformational changes resulting in protection of the 3C cleavage site. Since both proteins participate in the nucleosome binding, it is feasible to suggest that any potential complex is stabilised in the presence of the nucleosome, however, further work is required to test this assumption.

In the present study, LSH has been successfully expressed and purified in insect cells, however, CDCA7 could not be completely cleared of contaminating nucleic acid, which complicated any possible downstream application. Therefore, a number of approaches can be utilised to further elucidate the modes of binding and catalysis involving LSH, CDCA7 and a mononucleosome.

Since published results showing active LSH and CDCA7 in chromatin remodelling relied on both proteins from *Xenopus*, and robust expression systems have been established (Jenness et al. 2018), the next step would be to explore those expression systems for producing active proteins in sufficient (mg) amounts for structural and biochemical studies. Production of high yield contamination-free CDCA7 is cardinal for the crucial downstream experiments. Expression system optimisation is also crucial for producing various truncated versions of each protein, which would be indispensable for dissecting their functions.

Dissecting the CDCA7 binding to the nucleosome is a crucial step in elucidation the mechanism of LSH mechanism of chromatin remodelling, therefore, pure CDCA7 free from nucleic acid contaminants should be produced, including the full-length protein and its

truncates, with domains gradually removed from the N-terminus. Such constructs should then be used for a binding shift assay, such as EMSA, as well as an ATP hydrolysis assay, in which LSH-nucleosome, CDCA7-nucleosome, LSH-CDCA7 and LSH-CDCA7-nucleosome complexes should be compared, and the rate of ATP hydrolysis measured. This can be done with a malachite green or thin layer chromatography (TLC) assays.

Following that, protein truncations combination should be tested in a restriction enzyme remodelling assay, wherein a recombinant nucleosome with an engineered PstI restriction site is incubated with the proteins and PstI restriction enzyme. Confirmation of LSH remodelling activity would be required before moving to structural studies.

8. POSSIBLE MODELS FOR THE LSH-CDCA7 COMPLEX ROLE IN CHROMATIN REMODELLING

CDCA7-facilitated nucleosome remodelling by LSH is a crucial mechanism in normal organism development. The main pathological phenotype associated with LSH and CDCA7 is the Immunodeficiency, Centromeric instability, Facial anomaly (ICF) syndrome characterised by congenital defects of DNA methylation (Vukic and Daxinger, 2019) with less than 100 cases described worldwide (van den Boogaard et al., 2017). ICF syndrome manifests in reduction or absence of B cells and antibody titres, altered chromosomal configuration, the presence of characteristic facial features, intellectual disability and early mortality (van den Boogaard et al., 2017; Vukic and Daxinger, 2019; Weemaes et al., 2013). Whereas the most common mutations found in ICF are associated with DNMT3B catalytic activity, 12 cases each have been described for CDCA7 and LSH (Thijssen et al., 2015). LSH and CDCA7 mutations are also associated with a subset of ICF patients demonstrating α-satellite repeat hypomethylation but not subtelomeric repeats (Hu et al., 2019; Toubiana et al., 2018).

Interestingly, other chromatin remodellers have also been linked to B cell and neuronal development. Chd4 of the Nucleosome Remodeling and Deacetylase (NuRD) complex was shown to be crucial for B cell progenitor maturation and is likely to direct their lineage transcription programme through chromatin accessibility, preventing spurious expression of other cell types (Arends et al., 2019). Chd4 was found to regulate B cell development at the gene expression level and Chd4 loss was abolished pre-B cell population but did not appear to affect survival of mature B cells (Yen et al., 2019).

Mutations in the ATPase domain of Chd3 resulted in decreased chromatin remodelling activity *in vitro* and manifested in neurodevelopmental impairment and patients with intellectual disabilities and facial anomalies (Blok et al., 2018). An additional mutation manifesting in this phenotype, but not disrupting the ATPase domain of Chd3, was also present in yeast Snf2 (W1158 in both Chd1 and Snf2) (Liu et al., 2017a). In murine LSH this residue (W679) is also conserved and is located two residues away from Q682R implicated in ICF syndrome (Q682) (***see Section 3.2***) indicating a possible common mechanism in chromatin remodelling. Similarly, mutations across Chd1 protein (including its ATPase and DNA-binding domains) were found in a neurological syndrome with facial dysmorphic features (Pilarowski et al., 2018).

Additionally, Brahma SWI/SNF through its chromatin remodelling activity was shown to provide enhanced accessibility to CSL transcription factor (Pillidge and Bray, 2019), which is a part of the Notch signalling pathway implicated in a variety of developmental processes, including neuronal development (Gaiano and Fishell, 2002) and T cell lineage progression (Laky et al., 2015). Taken together these findings reinforce the role of LSH-mediated chromatin remodelling in immunological and neuronal syndromes and can potentially shed more light on the mechanism of LSH impairment.

Dissection of the causes of ICF syndrome are complicated by the participation of each implicated protein in a range of processes not associated with *de novo* DNA methylation. For instance, LSH has also been found to interact with the maintenance DNA methyltransferase DNMT1 *in vitro* (Dunican et al., 2013). Another ICF implicated zinc finger protein ZBTB24 (~30% of all cases) is a positive regulator of CDCA7 expression (Wu et al., 2016b) through it C2H2 zinc finger motifs (Aktar et al., 2019; Ren et al., 2019). ZBTB24-CDCA7 axis was found to be important for survival of T cells (Qin et al., 2019), providing another link between these proteins and immunodeficiency.

Potentially, LSH may link the ICF phenotype not with DNA methylation but with its role in double-strand break (DSB) repair. LSH may be required for efficient DSB repair in the non-homologues end joining (NHEJ) in class switch recombination mechanism for antibody generation in B cells. Inability of LSH to carry out its chromatin remodelling functions as well as impaired recruitment to chromatin by mutated CDCA7 can undermine remodelling or even remodelling-independent recruitment of NHEJ proteins Ku70 and Ku80 (Unoki et al., 2018).

From the work performed previously and in the scope of this study, it appears that LSH binds the nucleosome, and the interaction is predominantly through binding to DNA rather than the histone octamer, as LSH demonstrated a similar binding pattern (as indicated by the EMSA experiments) between free DNA and a mononucleosome. LSH is likely to bind to the mononucleosome via its ATPase domain, and the exact role of its N- and C-termini have yet to be elucidated. It is known that LSH required CDCA7 for recruitment to chromatin (Jenness et al. 2018). From the EMSA experiments performed in this study, it appears that LSH can bind DNA on its own, however, in the nuclear context this binding and retention of LSH on chromatin is likely to require additional partners, such as CDCA7. It is feasible to suggest that LSH adopts the conformation poised for catalysis only in the presence of CDCA7.

Sequence examination and *ab initio* modelling of CDCA7 confirmed that is possesses a zinc finger domain likely of a non-classical type. Previous studies indicated that it is a RING zinc finger, however, even though a characteristic cross-brace conformation is not present in CDCA7 models according to the *ab initio* models, it is not clear whether CDCA7 indeed binds DNA or this interaction is artefactual, since RING zinc fingers predominantly interact with proteins not nucleic acids. Experiments with protein truncations would be required to elucidate that. For example, LSH with truncated N- and/or C-terminus should be expressed to elucidate the role in the ATPase core in the DNA-binding EMSA experiments. Similarly, CDCA7 should be expressed in truncated versions, where the N-terminal part prior to the first predicted coiled coil and then the second coiled coil should be removed, which would allow to start with the DNA and LSH interaction experiments.

Taken together, the homology modelling and biochemical findings from the present study indicate that there are several options for LSH interaction with the nucleosome and its subsequent role in chromatin remodelling and ICF syndrome. There are two parts in LSH remodelling activity: first, 1) it binds the nucleosome, then 2) it performs chromatin remodelling and allow other proteins perform their roles in chromatin metabolism. In the first step, ***1.A)*** CDCA7 bind dsDNA in core nucleosome or linker DNA through its RING zinc finger domain and then through its coiled coils recruits LSH (Jenness et al., 2018), which then binds the nucleosome at SHL2 through its ATPase domain. This is consistent with the homology modelling data for LSH, showing that LSH is likely to interact with the nucleosome on SHL2 site (***Fig. 8.1 A***). However, it contradicts the *in vitro* binding experiments that indicated that LSH can bind the nucleosome on its own. It is more likely that LSH has a more independent binding more with the nucleosome, however, CDCA7 may induce conformational changes in LSH, which facilitate its remodelling activity.

Alternatively, ***1.B)*** LSH is constitutively present on the chromatin without requiring additional factors for binding (as indicated by my EMSA experiments showing that LSH can independently bind nucleosomes *in vitro* (***Fig. 5.6 and 5.8***) and as indicated by experiments *in vivo* (Kollarovic et al., 2018), bound at SHL2 or linker DNA through its ATPase domain, existing in an equilibrium of nucleosome-bound and apo state. It then recruits CDCA7, which binds LSH via its RING zinc finger, induces conformational changes in LSH and acts as a stabilising and positioning factor (***Fig. 8.1 B***), however, that is inconsistent with the findings from Jenness et al., which demonstrated that CDCA7 recruits LSH to chromatin (Jenness et

al., 2018). CDCA7 may be required to relocate LSH from an indiscriminate position anywhere on dsDNA to SHL2 specifically, which would be consistent with initial indiscriminate binding of LSH and its subsequent "recruitment" on SHL2 by CDCA7. Homology modelling (***see Section 3.2***) demonstrates the most likely positioning of LSH to be SHL2, however, it is unclear whether is locates there itself or is aided by another protein.

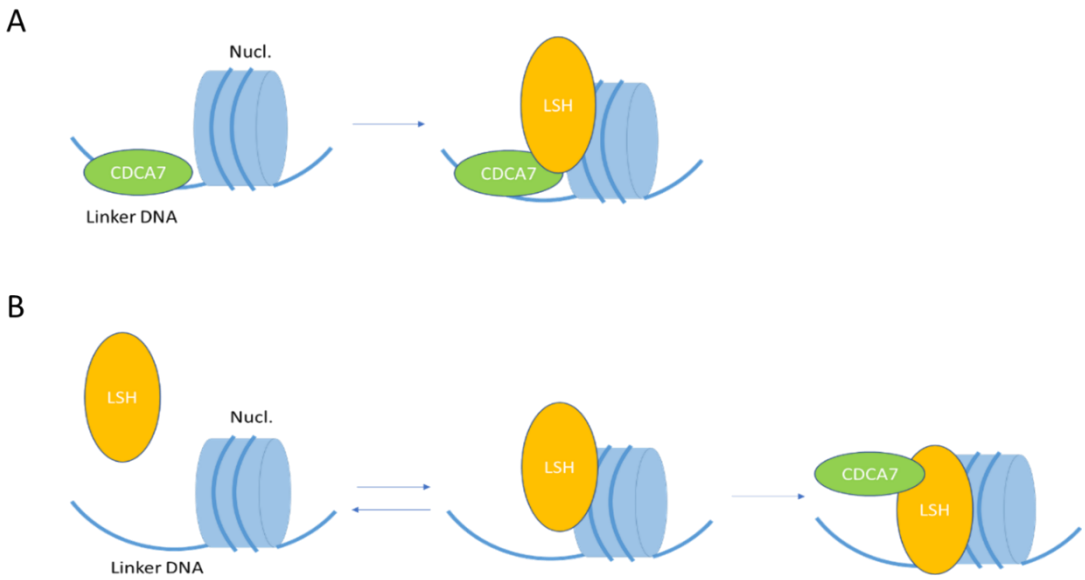


Figure 8.1 Schematic diagram showing possible ways of LSH interaction with the nucleosome and CDCA7. A) CDCA7 initially binds to the nucleosome and recruits LSH to SHL2 site; B) LSH binds the nucleosome independently of other proteins, and then recruits CDCA7, that stabilises LSH on the nucleosome.

Once bound to the nucleosome and adopted the correct position and/or conformation for remodelling, ***2.A)*** LSH performs chromatin remodelling and also acts as a platform for recruiting other proteins, such as Ku70 and Ku80 (Unoki et al., 2018), end-resection machinery (Kollarovic et al., 2018), cohesin-assembly associated protein Scc1 (Rad21 in humans) (Litwin et al., 2017) or transcription factor E2F3 (von Eyss et al., 2012) (***Fig. 8.2 A***). All these functions can still be related to the chromatin remodelling activity of LSH, and their binding may be dependent on nucleosome-free regions exposed by LSH remodelling of chromatin.

Alternatively or complementary to its assistance to the proteins not directly involved with DNA-methylatinon, ***2.B)*** LSH performs chromatin remodelling to provide binding space to DNMT3B for *de novo* DNA methylation (Myant and Stancheva, 2008; Zhu et al., 2006) or DNMT1 for constitutive DNA methylation (Dunican et al., 2013) (***Fig. 8.2 B***). LSH, however, is

likely to play a redundant role in maintenance methylation. There can be further, more intricate mechanism, wherein LSH is displaced from chromatin by DNMTs, which induce its conformational change, or an LSH-CDCA7 complex acts as a platform for DNMTs binding, however, the experimental evidence for that is insufficient. *In vitro* binding assays can be used to test LSH interaction with DNA repair proteins, however, this is a more speculative direction and should not be taken before LSH-CDCA7 interaction in chromatin remodelling is elucidated

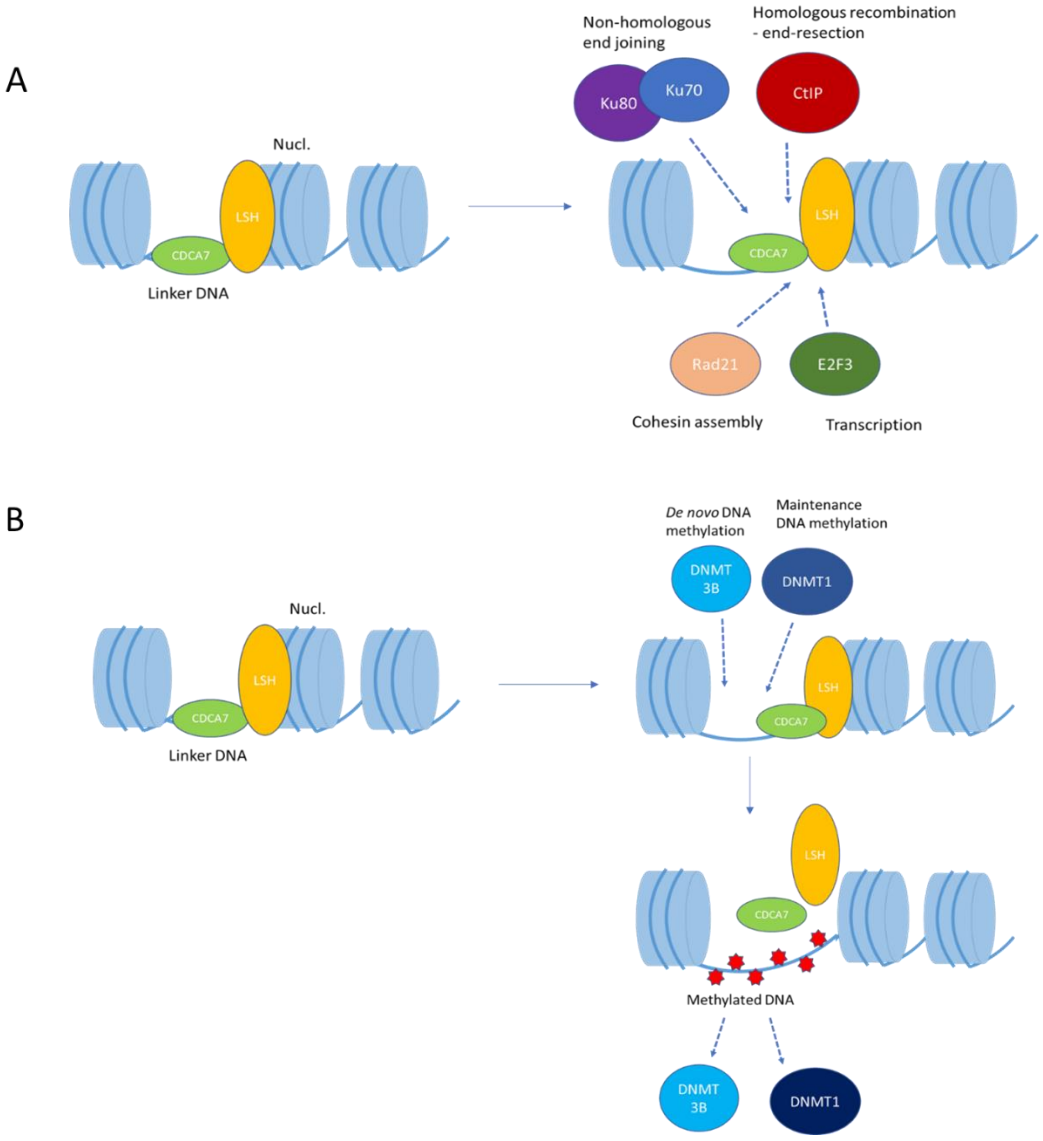


Figure 8.2 Role of the nucleosome-bound LSH in chromatin metabolism. A) Following chromatin remodelling accelerated by CDCA7, LSH acts as a binding platform for the proteins involved in various chromatin-related activities or provides nucleosome-free areas for them; B) Nucleosome-free areas provided by LSH are targeted by DNA-methyltransferases.

If CDCA7 ZFD participates in enhancing LSH ATPase activity not through DNA binding, but through its interaction with LSH, it is still unclear whether LSH binds DNA linker through its N- and/or C-terminal regions or does not interact with the linker at all. Unlike the remodellers that form regularly spaced nucleosomes, LSH does not have a DNA-binding domain, which was reflected in its ability to follow centre-to-end direction, rather than end-to-centre (Jenness et al., 2018). Experiments from Jenness et al. indicate that CDCA7 is likely to bind DNA and this way insure LSH remodelling activity, however, the exact binding mode and conformational changes of both proteins should be elucidated in further experiments.

Experiments with a histone deacetylase (HDAC) inhibitor trichostatin A showed that its application abolished LSH association with chromatin (Yan et al., 2003). This indicates that LSH recognises acetylated chromatin and requires it for binding. To dissect this possibility, octamers with chemically modified histone tails can be produced carrying acetylation marks on various residues. These may involve active (H3K27ac, H3K9ac and H3K14ac) marks (Karmodiya et al., 2012). *In vitro* binding experiments, such as EMSA or surface plasmon resonance (SPR), with modified histone tails could demonstrate difference in affinity for differentially modified nucleosomes. However, a potential caveat with this system may be that LSH could be sensitive to these modifications in the nuclear context but not *in vitro*. Moreover, it is still unclear whether LSH requires CDCA7 for proper positioning, therefore, the LSH-histones interaction may be artefactual.

CDCA7 is a highly disordered (> 60%) protein that has predicted coiled coil regions and a C-terminal zinc finger domain. A number of DNA-binding proteins undergo conformation changes upon binding their nucleotide substrates and protein binding partners (Poddar et al., 2018). CDCA7 has been shown to tightly bind DNA, however, it is not clear whether this binding is specific. Utilising chemical crosslinking to CDCA7 in different states (semi-pure with nucleic acid contamination, free from nucleic acid contamination, incubated with LSH and/or NCP) would allow to gain insights into the possible range of its conformations. The caveat in this approach will be the disordered nature of CDCA7, which can potentially result in various captured contacts reflecting dynamic conformation of the protein.

In the present study, I attempted to chemically crosslink LSH to the mononucleosome using BS^3 and EDC+sulfo-NHS crosslinkers. Neither chemical resulted in a shift different from the LSH only control. One option could be to optimise LSH and NCP concentrations to detect the optimal binding conditions, since some binding was detected by EMSA.

Another option could be to set up a DNA-to-protein UV crosslinking (Flett et al., 2018). In this system, specific sites on free or nucleosomal DNA would be labelled with the photoactivatable nucleotide 5-Iodouracil (5IdU) and a 3' biotin-TEG tag. These nucleotides are then covalently cross-linked to proximal aromatic amino acids using UV irradiation at 312 nm. This system would allow to detect the direct binding to DNA, which is the most likely site of LSH contact.

Gradient fixation method (Kastner et al., 2008; Stark 2010) is a common approach for stabilising molecular complexes in preparation for cryo-EM. Previously, I tried to capture LSH-NCP complex by GraFix, however, LSH appeared to aggregate during centrifugation and could not be resolved on the gel. Even though in the present study, LSH-CDCA7 complex was not observed, it is worth trying to capture the LSH-CDCA7-NCP complex as presumably more stable, being a trimolecular remodelling entity. This approach would require optimisation of the fixative and centrifugation time and speed. If GraFix could be carried out, it would produce a stable complex for structural studies, which would reveal the orientations of LSH and CDCA7 in the nucleosome binding.

Taken together, this work provided further insights into LSH interaction with the nucleosome, analysis of its binding partner CDCA7 and the approaches to their purification for biochemical and structural studies.

9. REFERENCES

Adamkewicz, J.I., Mueller, C.G.F., Hansen, K.E., Prud'homme, W.A., Thorner, J., 2000. Purification and Enzymic Properties of Mot1 ATPase, a Regulator of Basal Transcription in the Yeast *Saccharomyces cerevisiae*. J. Biol. Chem. 275, 21158–21168. https://doi.org/10.1074/jbc.M002639200

Aktar, S., Sasaki, H., Unoki, M., 2019. Identification of ZBTB24 protein domains and motifs for heterochromatin localization and transcriptional activation. Genes Cells 24, 746–755. https://doi.org/10.1111/gtc.12723

Al-Ani, G., Malik, S.S., Eastlund, A., Briggs, K., Fischer, C.J., 2014. ISWI Remodels Nucleosomes through a Random Walk. Biochemistry 53, 4346–4357. https://doi.org/10.1021/bi500226b

Alvaro, D., Lisby, M., Rothstein, R., 2007. Genome-Wide Analysis of Rad52 Foci Reveals Diverse Mechanisms Impacting Recombination. PLoS Genet 3, e228. https://doi.org/10.1371/journal.pgen.0030228

Andrau, J.-C., Van Oevelen, C.J.C., Van Teeffelen, H.A.A.M., Weil, P.A., Holstege, F.C.P., Timmers, H.Th.M., 2002. Mot1p is essential for TBP recruitment to selected promoters during in vivo gene activation. EMBO J 21, 5173–5183. https://doi.org/10.1093/emboj/cdf485

Arends, T., Dege, C., Bortnick, A., Danhorn, T., Knapp, J.R., Jia, H., Harmacek, L., Fleenor, C.J., Straign, D., Walton, K., Leach, S.M., Feeney, A.J., Murre, C., O'Connor, B.P., Hagman, J.R., 2019. CHD4 is essential for transcriptional repression and lineage progression in B lymphopoiesis. Proc Natl Acad Sci USA 116, 10927–10936. https://doi.org/10.1073/pnas.1821301116

Armache, J.P., Gamarra, N., Johnson, S.L., Leonard, J.D., Wu, S., Narlikar, G.J., Cheng, Y., 2019. Cryo-EM structures of remodeler-nucleosome intermediates suggest allosteric control through the nucleosome. eLife 8, e46057. https://doi.org/10.7554/eLife.46057

Armache, K.-J., Garlick, J.D., Canzio, D., Narlikar, G.J., Kingston, R.E., 2011. Structural Basis of Silencing: Sir3 BAH Domain in Complex with a Nucleosome at 3.0 A Resolution. Science 334, 977–982. https://doi.org/10.1126/science.1210915

Auble, D.T., Wang, D., Post, K.W., Hahn, S., 1997. Molecular analysis of the SNF2/SWI2 protein family member MOT1, an ATP-driven enzyme that dissociates TATA-binding protein from DNA. Mol. Cell. Biol. 17, 4842–4851. https://doi.org/10.1128/MCB.17.8.4842

Ausio, J., Seger, D., Eisenberg, H., 1984. Nucleosome core particle stability and conformational change. Journal of Molecular Biology 176, 77–104. https://doi.org/10.1016/0022-2836(84)90383-8

Ayala, R., Willhoft, O., Aramayo, R.J., Wilkinson, M., McCormack, E.A., Ocloo, L., Wigley, D.B., Zhang, X., 2018. Structure and regulation of the human INO80–nucleosome complex. Nature 556, 391–395. https://doi.org/10.1038/s41586-018-0021-6

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., Kouzarides, T., 2001. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature 410, 120–124. https://doi.org/10.1038/35065138

Barbera, A.J., 2006. The Nucleosomal Surface as a Docking Station for Kaposi's Sarcoma Herpesvirus LANA. Science 311, 856–861. https://doi.org/10.1126/science.1120541

Bartholomeeusen, K., De Rijck, J., Busschots, K., Desender, L., Gijsbers, R., Emiliani, S., Benarous, R., Debyser, Z., Christ, F., 2007. Differential Interaction of HIV-1 Integrase

and JPO2 with the C Terminus of LEDGF/p75. Journal of Molecular Biology 372, 407–421. https://doi.org/10.1016/j.jmb.2007.06.090

Basenko, E.Y., Kamei, M., Ji, L., Schmitz, R.J., Lewis, Z.A., 2016. The LSH/DDM1 Homolog MUS-30 Is Required for Genome Stability, but Not for DNA Methylation in Neurospora crassa. PLoS Genet 12, e1005790. https://doi.org/10.1371/journal.pgen.1005790

Bednar, J., Horowitz, R.A., Grigoryev, S.A., Carruthers, L.M., Hansen, J.C., Koster, A.J., Woodcock, C.L., 1998. Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. Proceedings of the National Academy of Sciences 95, 14173–14178. https://doi.org/10.1073/pnas.95.24.14173

Berg, J.M., Tymoczko, J.L., Stryer, L., Stryer, L., 2002. Biochemistry, 5th ed. ed. W.H. Freeman, New York.

Bertin, A., Mangenot, S., Renouard, M., Durand, D., Livolant, F., 2007. Structure and Phase Diagram of Nucleosome Core Particles Aggregated by Multivalent Cations. Biophysical Journal 93, 3652–3663. https://doi.org/10.1529/biophysj.107.108365

Bilokapic, S., Strauss, M., Halic, M., 2018. Cryo-EM of nucleosome core particle interactions in trans. Sci Rep 8, 7046. https://doi.org/10.1038/s41598-018-25429-1

Bird, A.P., 1980. DNA methylation and the frequency of CpG in animal DNA. Nucl Acids Res 8, 1499–1504. https://doi.org/10.1093/nar/8.7.1499

Biswas, M., Voltz, K., Smith, J.C., Langowski, J., 2011. Role of Histone Tails in Structural Stability of the Nucleosome. PLoS Comput Biol 7, e1002279. https://doi.org/10.1371/journal.pcbi.1002279

Blok, Snijders Blok, L., Rousseau, J., Twist, J., Ehresmann, S., Takaku, M., Venselaar, H., Rodan, L.H., Nowak, C.B., Douglas, J., Swoboda, K.J., Steeves, M.A., Sahai, I., Stumpel, C.T.R.M., Stegmann, A.P.A., Wheeler, P., Willing, M., Fiala, E., Kochhar, A., Gibson, W.T., Cohen, A.S.A., Agbahovbe, R., Innes, A.M., Au, P.Y.B., Rankin, J., Anderson, I.J., Skinner, S.A., Louie, R.J., Warren, H.E., Afenjar, A., Keren, B., Nava, C., Buratti, J., Isapof, A., Rodriguez, D., Lewandowski, R., Propst, J., van Essen, T., Choi, M., Lee, S., Chae, J.H., Price, S., Schnur, R.E., Douglas, G., Wentzensen, I.M., Zweier, C., Reis, A., Bialer, M.G., Moore, C., Koopmans, M., Brilstra, E.H., Monroe, G.R., van Gassen, K.L.I., van Binsbergen, E., Newbury-Ecob, R., Bownass, L., Bader, I., Mayr, J.A., Wortmann, S.B., Jakielski, K.J., Strand, E.A., Kloth, K., Bierhals, T., Roberts, J.D., Petrovich, R.M., Machida, S., Kurumizaka, H., Lelieveld, S., Pfundt, R., Jansen, S., Deriziotis, P., Faivre, L., Thevenon, J., Assoum, M., Shriberg, L., Kleefstra, T., Brunner, H.G., Wade, P.A., Fisher, S.E., Campeau, P.M., 2018. CHD3 helicase domain mutations cause a neurodevelopmental syndrome with macrocephaly and impaired speech and language. Nat Commun 9, 4619. https://doi.org/10.1038/s41467-018-06014-6

Borden, K.L., Freemont, P.S., 1996. The RING finger domain: a recent example of a sequence—structure family. Current Opinion in Structural Biology 6, 395–401. https://doi.org/10.1016/S0959-440X(96)80060-1

Borgel, J., Tyl, M., Schiller, K., Pusztai, Z., Dooley, C.M., Deng, W., Wooding, C., White, R.J., Warnecke, T., Leonhardt, H., Busch-Nentwich, E.M., Bartke, T., 2016. KDM2A integrates DNA and histone modification signals through a CXXC/PHD module and direct interaction with HP1. Nucleic Acids Res gkw979. https://doi.org/10.1093/nar/gkw979

Boubrik, F., Bonnefoy, E., Rouvière-Yaniv, J., 1991. HU and IHF: similarities and differences. In Escherichia coli, the lack of HU is not compensated for by IHF. Res. Microbiol. 142, 239–247.

Boyer, L.A., Latek, R.R., Peterson, C.L., 2004. The SANT domain: a unique histone-tail-binding module? Nat Rev Mol Cell Biol 5, 158–163. https://doi.org/10.1038/nrm1314

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. Nat Commun 8, 15616. https://doi.org/10.1038/ncomms15616

Brandani, G.B., Niina, T., Tan, C., Takada, S., 2018. DNA sliding in nucleosomes via twist defect propagation revealed by molecular simulations. Nucleic Acids Research 46, 2788–2801. https://doi.org/10.1093/nar/gky158

Brehm, A., 2000. dMi-2 and ISWI chromatin remodelling factors have distinct nucleosome binding and mobilization properties. The EMBO Journal 19, 4332–4341. https://doi.org/10.1093/emboj/19.16.4332

Brosh, R.M., 2013. DNA helicases involved in DNA repair and their roles in cancer. Nat Rev Cancer 13, 542–558. https://doi.org/10.1038/nrc3560

Brzeski, J., Jerzmanowski, A., 2003. Deficient in DNA Methylation 1 (DDM1) Defines a Novel Family of Chromatin-remodeling Factors. J. Biol. Chem. 278, 823–828. https://doi.org/10.1074/jbc.M209260200

Burkhard, P., Stetefeld, J., Strelkov, S.V., 2001. Coiled coils: a highly versatile protein folding motif. Trends Cell Biol. 11, 82–88.

Burrage, J., Termanis, A., Geissner, A., Myant, K., Gordon, K., Stancheva, I., 2012. The SNF2 family ATPase LSH promotes phosphorylation of H2AX and efficient repair of DNA double-strand breaks in mammalian cells. J Cell Sci 125, 5524–5534. https://doi.org/10.1242/jcs.111252

Butryn, A., Schuller, J.M., Stoehr, G., Runge-Wollmann, P., Förster, F., Auble, D.T., Hopfner, K.-P., 2015. Structural basis for recognition and remodeling of the TBP:DNA:NC2 complex by Mot1. eLife 4, e07432. https://doi.org/10.7554/eLife.07432

Butryn, A., Woike, S., Shetty, S.J., Auble, D.T., Hopfner, K.-P., 2018. Crystal structure of the full Swi2/Snf2 remodeler Mot1 in the resting state. eLife 7, e37774. https://doi.org/10.7554/eLife.37774

Cannuyer, J., Van Tongelen, A., Loriot, A., De Smet, C., 2015. A gene expression signature identifying transient DNMT1 depletion as a causal factor of cancer-germline gene activation in melanoma. Clin Epigenet 7, 114. https://doi.org/10.1186/s13148-015-0147-4

Cassandri, M., Smirnov, A., Novelli, F., Pitolli, C., Agostini, M., Malewicz, M., Melino, G., Raschellà, G., 2017. Zinc-finger proteins in health and disease. Cell Death Discov. 3, 17071. https://doi.org/10.1038/cddiscovery.2017.71

Chan, T.S.Y., Hawkins, C., Krieger, J.R., McGlade, C.J., Huang, A., 2016. JPO2/CDCA7L and LEDGF/p75 Are Novel Mediators of PI3K/AKT Signaling and Aggressive Phenotypes in Medulloblastoma. Cancer Research 76, 2802–2812. https://doi.org/10.1158/0008-5472.CAN-15-2194

Chatterjee, N., Sinha, D., Lemma-Dechassa, M., Tan, S., Shogren-Knaak, M.A., Bartholomew, B., 2011. Histone H3 tail acetylation modulates ATP-dependent remodeling through multiple mechanisms. Nucleic Acids Research 39, 8378–8391. https://doi.org/10.1093/nar/gkr535

Chen, D., Maruschke, M., Hakenberg, O., Zimmermann, W., Stief, C.G., Buchner, A., 2017. TOP2A , HELLS , ATAD2 , and TET3 Are Novel Prognostic Markers in Renal Cell Carcinoma. Urology 102, 265.e1-265.e7. https://doi.org/10.1016/j.urology.2016.12.050

Chen, K., Ou, X.-M., Chen, G., Choi, S.H., Shih, J.C., 2005. R1, a Novel Repressor of the Human Monoamine Oxidase A. Journal of Biological Chemistry 280, 11552–11559. https://doi.org/10.1074/jbc.M410033200

Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E., Debyser, Z., 2003. HIV-1 Integrase Forms Stable Tetramers and Associates with LEDGF/p75 Protein in Human Cells. J. Biol. Chem. 278, 372–381. https://doi.org/10.1074/jbc.M209278200

Chicca, J.J., Auble, D.T., Pugh, B.F., 1998. Cloning and Biochemical Characterization of TAF-172, a Human Homolog of Yeast Mot1. Mol. Cell. Biol. 18, 1701–1710. https://doi.org/10.1128/MCB.18.3.1701

Clapier, C.R., 2002. A critical epitope for substrate recognition by the nucleosome remodeling ATPase ISWI. Nucleic Acids Research 30, 649–655. https://doi.org/10.1093/nar/30.3.649

Clapier, C.R., Cairns, B.R., 2012. Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. Nature 492, 280–284. https://doi.org/10.1038/nature11625

Clapier, C.R., Iwasa, J., Cairns, B.R., Peterson, C.L., 2017. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat Rev Mol Cell Biol 18, 407–422. https://doi.org/10.1038/nrm.2017.26

Clapier, C.R., Kasten, M.M., Parnell, T.J., Viswanathan, R., Szerlong, H., Sirinakis, G., Zhang, Y., Cairns, B.R., 2016. Regulation of DNA Translocation Efficiency within the Chromatin Remodeler RSC/Sth1 Potentiates Nucleosome Sliding and Ejection. Molecular Cell 62, 453–461. https://doi.org/10.1016/j.molcel.2016.03.032

Clapier, C.R., Langst, G., Corona, D.F.V., Becker, P.B., Nightingale, K.P., 2001. Critical Role for the Histone H4 N Terminus in Nucleosome Remodeling by ISWI. Molecular and Cellular Biology 21, 875–883. https://doi.org/10.1128/MCB.21.3.875-883.2001

Clark, A.J., Margulies, A.D., 1965. ISOLATION AND CHARACTERIZATION OF RECOMBINATION-DEFICIENT MUTANTS OF ESCHERICHIA COLI K12. Proceedings of the National Academy of Sciences 53, 451–459. https://doi.org/10.1073/pnas.53.2.451

Coletta, A., Pinney, J.W., Solís, D.Y.W., Marsh, J., Pettifer, S.R., Attwood, T.K., 2010. Low-complexity regions within protein sequences have position-dependent roles. BMC Syst Biol 4, 43. https://doi.org/10.1186/1752-0509-4-43

Corem, S., Doron-Faigenboim, A., Jouffroy, O., Maumus, F., Arazi, T., Bouché, N., 2018. Redistribution of CHH Methylation and Small Interfering RNAs across the Genome of Tomato *ddm1* Mutants. Plant Cell 30, 1628–1644. https://doi.org/10.1105/tpc.18.00167

Dang, W., Bartholomew, B., 2007. Domain Architecture of the Catalytic Subunit in the ISW2-Nucleosome Complex. Molecular and Cellular Biology 27, 8306–8317. https://doi.org/10.1128/MCB.01351-07

Dang, W., Kagalwala, M.N., Bartholomew, B., 2006. Regulation of ISW2 by Concerted Action of Histone H4 Tail and Extranucleosomal DNA. Molecular and Cellular Biology 26, 7388–7396. https://doi.org/10.1128/MCB.01159-06

Dann, G.P., Liszczak, G.P., Bagert, J.D., Müller, M.M., Nguyen, U.T.T., Wojcik, F., Brown, Z.Z., Bos, J., Panchenko, T., Pihl, R., Pollock, S.B., Diehl, K.L., Allis, C.D., Muir, T.W., 2017. ISWI chromatin remodellers sense nucleosome modifications to determine substrate preference. Nature 548, 607–611. https://doi.org/10.1038/nature23671

Dao, H.T., Dul, B.E., Dann, G.P., Liszczak, G.P., Muir, T.W., 2019. A basic motif anchoring ISWI to nucleosome acidic patch regulates nucleosome spacing. Nat Chem Biol. https://doi.org/10.1038/s41589-019-0413-4

Dasgupta, A., Darst, R.P., Martin, K.J., Afshari, C.A., Auble, D.T., 2002. Mot1 activates and represses transcription by direct, ATPase-dependent mechanisms. Proceedings of the National Academy of Sciences 99, 2666–2671. https://doi.org/10.1073/pnas.052397899

De La Fuente, R., Baumann, C., Fan, T., Schmidtmann, A., Dobrinski, I., Muegge, K., 2006. Lsh is required for meiotic chromosome synapsis and retrotransposon silencing in female germ cells. Nat Cell Biol 8, 1448–1454. https://doi.org/10.1038/ncb1513

Deindl, S., Hwang, W.L., Hota, S.K., Blosser, T.R., Prasad, P., Bartholomew, B., Zhuang, X., 2013. ISWI Remodelers Slide Nucleosomes with Coordinated Multi-Base-Pair Entry Steps and Single-Base-Pair Exit Steps. Cell 152, 442–452. https://doi.org/10.1016/j.cell.2012.12.040

Delorenzi, M., Speed, T., 2002. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. Bioinformatics 18, 617–625. https://doi.org/10.1093/bioinformatics/18.4.617

Dennis, K., 2001. Lsh, a member of the SNF2 family, is required for genome-wide methylation. Genes & Development 15, 2940–2944. https://doi.org/10.1101/gad.929101

Dosztányi, Z., 2018. Prediction of protein disorder based on IUPred: Prediction of Protein Disorder Based on IUPred. Protein Science 27, 331–340. https://doi.org/10.1002/pro.3334

Dosztanyi, Z., Meszaros, B., Simon, I., 2009. ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics 25, 2745–2746. https://doi.org/10.1093/bioinformatics/btp518

Dumon-Seignovert, L., Cariot, G., Vuillard, L., 2004. The toxicity of recombinant proteins in Escherichia coli: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). Protein Expression and Purification 37, 203–206. https://doi.org/10.1016/j.pep.2004.04.025

Dunican, D.S., Cruickshanks, H.A., Suzuki, M., Semple, C.A., Davey, T., Arceci, R.J., Greally, J., Adams, I.R., Meehan, R.R., 2013. Lsh regulates LTR retrotransposon repression independently of Dnmt3b function. Genome Biol 14, R146. https://doi.org/10.1186/gb-2013-14-12-r146

Dunican, D.S., Pennings, S., Meehan, R.R., 2015. Lsh Is Essential for Maintaining Global DNA Methylation Levels in Amphibia and Fish and Interacts Directly with Dnmt1. BioMed Research International 2015, 1–12. https://doi.org/10.1155/2015/740637

Dunker, A.K., Obradovic, Z., Romero, P., Garner, E.C., Brown, C.J., 2000. Intrinsic protein disorder in complete genomes. Genome Inform Ser Workshop Genome Inform 11, 161–171.

Dyson, H.J., Wright, P.E., 2005. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6, 197–208. https://doi.org/10.1038/nrm1589

Edayathumangalam, R.S., Weyermann, P., Dervan, P.B., Gottesfeld, J.M., Luger, K., 2005. Nucleosomes in Solution Exist as a Mixture of Twist-defect States. Journal of Molecular Biology 345, 103–114. https://doi.org/10.1016/j.jmb.2004.10.012

Ehrlich, M., Sanchez, C., Shao, C., Nishiyama, R., Kehrl, J., Kuick, R., Kubota, T., Hanash, S.M., 2008. ICF, an immunodeficiency syndrome: DNA methyltransferase 3B involvement, chromosome anomalies, and gene dysregulation. Autoimmunity 41, 253–271. https://doi.org/10.1080/08916930802024202

Elrod-Erickson, M., Rould, M.A., Nekludova, L., Pabo, C.O., 1996. Zif268 protein–DNA complex refined at 1.6å: a model system for understanding zinc finger–DNA interactions. Structure 4, 1171–1180. https://doi.org/10.1016/S0969-2126(96)00125-6

Eustermann, S., Schall, K., Kostrewa, D., Lakomek, K., Strauss, M., Moldt, M., Hopfner, K.-P., 2018. Structural basis for ATP-dependent chromatin remodelling by the INO80 complex. Nature 556, 386–390. https://doi.org/10.1038/s41586-018-0029-y

Fan, H.-Y., He, X., Kingston, R.E., Narlikar, G.J., 2003. Distinct Strategies to Make Nucleosomal DNA Accessible. Molecular Cell 11, 1311–1322. https://doi.org/10.1016/S1097-2765(03)00192-8

Fan, J.Y., Rangasamy, D., Luger, K., Tremethick, D.J., 2004. H2A.Z Alters the Nucleosome Surface to Promote HP1α-Mediated Chromatin Fiber Folding. Molecular Cell 16, 655–661. https://doi.org/10.1016/j.molcel.2004.10.023

Fan, T., 2005. Lsh controls silencing of the imprinted Cdkn1c gene. Development 132, 635–644. https://doi.org/10.1242/dev.01612

Fan, T., Schmidtmann, A., Xi, S., Briones, V., Zhu, H., Suh, H.C., Gooya, J., Keller, J.R., Xu, H., Roayaei, J., Anver, M., Ruscetti, S., Muegge, K., 2008. DNA hypomethylation caused by Lsh deletion promotes erythroleukemia development. Epigenetics 3, 134–142. https://doi.org/10.4161/epi.3.3.6252

Fan, T., Yan, Q., Huang, J., Austin, S., Cho, E., Ferris, D., Muegge, K., n.d. Lsh-deficient Murine Embryonal Fibroblasts Show Reduced Proliferation with Signs of Abnormal Mitosis 7.

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

Felle, M., Hoffmeister, H., Rothammer, J., Fuchs, A., Exler, J.H., Längst, G., 2011. Nucleosomes protect DNA from DNA methylation in vivo and in vitro. Nucleic Acids Res. 39, 6956–6969. https://doi.org/10.1093/nar/gkr263

Ferreira, H., Somers, J., Webster, R., Flaus, A., Owen-Hughes, T., 2007. Histone Tails and the H3 N Helix Regulate Nucleosome Mobility and Stability. Molecular and Cellular Biology 27, 4037–4048. https://doi.org/10.1128/MCB.02229-06

Fitzgerald, D.J., Berger, P., Schaffitzel, C., Yamada, K., Richmond, T.J., Berger, I., 2006. Protein complex expression by using multigene baculoviral vectors. Nat Methods 3, 1021–1032. https://doi.org/10.1038/nmeth983

Flanagan, J.F., Mi, L.-Z., Chruszcz, M., Cymborowski, M., Clines, K.L., Kim, Y., Minor, W., Rastinejad, F., Khorasanizadeh, S., 2005. Double chromodomains cooperate to recognize the methylated histone H3 tail. Nature 438, 1181–1185. https://doi.org/10.1038/nature04290

Flaus, A., 2006. Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. Nucleic Acids Research 34, 2887–2905. https://doi.org/10.1093/nar/gkl295

Flaus, A., Owen-Hughes, T., 2011. Mechanisms for ATPdependent chromatin remodelling: the means to the end. FEBS Journal 17.

Flett, F.J., Ruksenaite, E., Armstrong, L.A., Bharati, S., Carloni, R., Morris, E.R., Mackay, C.L., Interthal, H., Richardson, J.M., 2018. Structural basis for DNA 3'-end processing by human tyrosyl-DNA phosphodiesterase 1. Nat Commun 9, 24. https://doi.org/10.1038/s41467-017-02530-z

Fratini, A.V., Kopka, M.L., Drew, H.R., Dickerson, R.E., 1982. Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG. J. Biol. Chem. 257, 14686–14707.

Fried, M.G., Bromberg, J.L., 1997. Factors that affect the stability of protein-DNA complexes during gel electrophoresis. Electrophoresis 18, 6–11. https://doi.org/10.1002/elps.1150180103

Gaiano, N., Fishell, G., 2002. The Role of Notch in Promoting Glial and Neural Stem Cell Fates. Annu. Rev. Neurosci. 25, 471–490. https://doi.org/10.1146/annurev.neuro.25.030702.130823
Gamarra, N., Johnson, S.L., Trnka, M.J., Burlingame, A.L., Narlikar, G.J., 2018. The nucleosomal acidic patch relieves auto-inhibition by the ISWI remodeler SNF2h. eLife 7, e35322. https://doi.org/10.7554/eLife.35322
Gamsjaeger, R., Liew, C., Loughlin, F., Crossley, M., Mackay, J., 2007. Sticky fingers: zinc-fingers as protein-recognition motifs. Trends in Biochemical Sciences 32, 63–70. https://doi.org/10.1016/j.tibs.2006.12.007
Garner, E., Romero, P., Dunker, A.K., Brown, C., Obradovic, Z., 1999. Predicting Binding Regions within Disordered Proteins.
Geiman, T.M., Durum, S.K., Muegge, K., 1998. Characterization of Gene Expression, Genomic Structure, and Chromosomal Localization ofHells(Lsh). Genomics 54, 477–483. https://doi.org/10.1006/geno.1998.5557
Geiman, T.M., Muegge, K., 2000. Lsh, an SNF2/helicase family member, is required for proliferation of mature T lymphocytes. Proceedings of the National Academy of Sciences 97, 4772–4777. https://doi.org/10.1073/pnas.97.9.4772
Geiman, T.M., Tessarollo, L., Anver, M.R., Kopp, J.B., Ward, J.M., Muegge, K., 2001. Lsh, a SNF2 family member, is required for normal murine development. Biochimica et Biophysica Acta (BBA) - General Subjects 1526, 211–220. https://doi.org/10.1016/S0304-4165(01)00129-5
Gill, R.M., Gabor, T.V., Couzens, A.L., Scheid, M.P., 2013. The MYC-Associated Protein CDCA7 Is Phosphorylated by AKT To Regulate MYC-Dependent Apoptosis and Transformation. Mol. Cell. Biol. 33, 498–513. https://doi.org/10.1128/MCB.00276-12
Gopalakrishnan, S., Van Emburgh, B.O., Robertson, K.D., 2008. DNA methylation in development and human disease. Mutat. Res. 647, 30–38. https://doi.org/10.1016/j.mrfmmm.2008.08.006
Goto, Y., Hayashi, R., Muramatsu, T., Ogawa, H., Eguchi, I., Oshida, Y., Ohtani, K., Yoshida, K., 2006. JPO1/CDCA7, a novel transcription factor E2F1-induced protein, possesses intrinsic transcriptional regulator activity. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression 1759, 60–68. https://doi.org/10.1016/j.bbaexp.2006.02.004
Gottesfeld, J.M., Luger, K., 2001. Energetics and Affinity of the Histone Octamer for Defined DNA Sequences †. Biochemistry 40, 10927–10933. https://doi.org/10.1021/bi0109966
Gutiérrez, J.L., Chandy, M., Carrozza, M.J., Workman, J.L., 2007. Activation domains drive nucleosome eviction by SWI/SNF. EMBO J 26, 730–740. https://doi.org/10.1038/sj.emboj.7601524
Haggerty, T.J., Zeller, K.I., Osthus, R.C., Wonsey, D.R., Dang, C.V., 2003. A strategy for identifying transcription factor binding sites reveals two classes of genomic c-Myc target sites. Proceedings of the National Academy of Sciences 100, 5313–5318. https://doi.org/10.1073/pnas.0931346100
Hakimi, M.-A., Bochar, D.A., Schmiesing, J.A., Dong, Y., Barak, O.G., Speicher, D.W., Yokomori, K., Shiekhattar, R., 2002. A chromatin remodelling complex that loads cohesin onto human chromosomes. Nature 418, 994–998. https://doi.org/10.1038/nature01024
Hamiche, A., Kang, J.-G., Dennis, C., Xiao, H., Wu, C., 2001. Histone Tails Modulate Nucleosome Mobility and Regulate ATP-Dependent Nucleosome Sliding by NURF.

Proceedings of the National Academy of Sciences of the United States of America 98, 14316–14321.
Hamiche, A., Sandaltzopoulos, R., Gdula, D.A., Wu, C., 1999. ATP-Dependent Histone Octamer Sliding Mediated by the Chromatin Remodeling Complex NURF. Cell 97, 833–842. https://doi.org/10.1016/S0092-8674(00)80796-5
Hanson, P.I., Whiteheart, S.W., 2005. AAA+ proteins: have engine, will work. Nat Rev Mol Cell Biol 6, 519–529. https://doi.org/10.1038/nrm1684
Harada, B.T., Hwang, W.L., Deindl, S., Chatterjee, N., Bartholomew, B., Zhuang, X., 2016. Stepwise nucleosome translocation by RSC remodeling complexes 20.
Harrer, N., Schindler, C.E.M., Bruetzel, L.K., Forné, I., Ludwigsen, J., Imhof, A., Zacharias, M., Lipfert, J., Mueller-Planitz, F., 2018. Structural Architecture of the Nucleosome Remodeler ISWI Determined from Cross-Linking, Mass Spectrometry, SAXS, and Modeling. Structure 26, 282-294.e6. https://doi.org/10.1016/j.str.2017.12.015
Hauk, G., McKnight, J.N., Nodelman, I.M., Bowman, G.D., 2010. The Chromodomains of the Chd1 Chromatin Remodeler Regulate DNA Access to the ATPase Motor. Molecular Cell 39, 711–723. https://doi.org/10.1016/j.molcel.2010.08.012
He, X., Yan, B., Liu, Shuang, Jia, J., Lai, W., Xin, X., Tang, C., Luo, D., Tan, T., Jiang, Y., Shi, Y., Liu, Y., Xiao, D., Chen, L., Liu, Shao, Mao, C., Yin, G., Cheng, Y., Fan, J., Cao, Y., Muegge, K., Tao, Y., 2016. Chromatin Remodeling Factor LSH Drives Cancer Progression by Suppressing the Activity of Fumarate Hydratase. Cancer Res 76, 5743–5755. https://doi.org/10.1158/0008-5472.CAN-16-0268
Heinemann, U., Alings, C., Bansal, M., 1992. Double helix conformation, groove dimensions and ligand binding potential of a G/C stretch in B-DNA. EMBO J. 11, 1931–1939.
Hellman, L.M., Fried, M.G., 2007. Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. Nat Protoc 2, 1849–1861. https://doi.org/10.1038/nprot.2007.249
Hong, J., Feng, H., Wang, F., Ranjan, A., Chen, J., Jiang, J., Ghirlando, R., Xiao, T.S., Wu, C., Bai, Y., 2014. The Catalytic Subunit of the SWR1 Remodeler Is a Histone Chaperone for the H2A.Z-H2B Dimer. Molecular Cell 53, 498–505. https://doi.org/10.1016/j.molcel.2014.01.010
Hu, H., Chen, C., Shi, S., Li, B., Duan, S., 2019. The gene mutations and subtelomeric DNA methylation in immunodeficiency, centromeric instability and facial anomalies syndrome. Autoimmunity 52, 192–198. https://doi.org/10.1080/08916934.2019.1657846
Huang, A., Ho, C.S.W., Ponzielli, R., Barsyte-Lovejoy, D., Bouffet, E., Picard, D., Hawkins, C.E., Penn, L.Z., 2005. Identification of a novel c-Myc protein interactor, JPO2, with transforming activity in medulloblastoma cells. Cancer Res. 65, 5607–5619. https://doi.org/10.1158/0008-5472.CAN-05-0500
Huang, J., 2004. Lsh, an epigenetic guardian of repetitive elements. Nucleic Acids Research 32, 5019–5028. https://doi.org/10.1093/nar/gkh821
Huang, Y.-T., Mason, J.O., Price, D.J., 2017. Lateral cortical Cdca7 expression levels are regulated by Pax6 and influence the production of intermediate progenitors. BMC Neurosci 18, 47. https://doi.org/10.1186/s12868-017-0365-0
Huttlin, E.L., Jedrychowski, M.P., Elias, J.E., Goswami, T., Rad, R., Beausoleil, S.A., Villén, J., Haas, W., Sowa, M.E., Gygi, S.P., 2010. A Tissue-Specific Atlas of Mouse Protein Phosphorylation and Expression. Cell 143, 1174–1189. https://doi.org/10.1016/j.cell.2010.12.001
Hwang, W.L., Deindl, S., Harada, B.T., Zhuang, X., 2014. Histone H4 tail mediates allosteric regulation of nucleosome remodelling by linker DNA. Nature 512, 213–217. https://doi.org/10.1038/nature13380

Jackson, S.P., Bartek, J., 2009. The DNA-damage response in human biology and disease. Nature 461, 1071–1078. https://doi.org/10.1038/nature08467
Jarvis, C.D., n.d. A novel putative helicase produced in early rnurine lymphocytes 5.
Jeddeloh, J.A., Stokes, T.L., Richards, E.J., 1999. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nat Genet 22, 94–97. https://doi.org/10.1038/8803
Jenness, C., Giunta, S., Müller, M.M., Kimura, H., Muir, T.W., Funabiki, H., 2018. HELLS and CDCA7 comprise a bipartite nucleosome remodeling complex defective in ICF syndrome. Proc Natl Acad Sci USA 115, E876–E885. https://doi.org/10.1073/pnas.1717509115
Jeong, W.-J., Park, J.-C., Kim, W.-S., Ro, E.J., Jeon, S.H., Lee, S.-K., Park, Y.N., Min, D.S., Choi, K.-Y., 2019. WDR76 is a RAS binding protein that functions as a tumor suppressor via RAS degradation. Nat Commun 10, 295. https://doi.org/10.1038/s41467-018-08230-6
Ji, Q., Ma, J.-W., Liu, R., Li, X., Shen, F., Huang, L., Hui, L., Ma, Y., Jin, B., 2019. CDCA7L promotes glioma proliferation by targeting CCND1 and predicts an unfavorable prognosis. Mol Med Report. https://doi.org/10.3892/mmr.2019.10349
Jiménez-P, R., Martín-Cortázar, C., Kourani, O., Chiodo, Y., Cordoba, R., Domínguez-Franjo, M.P., Redondo, J.M., Iglesias, T., Campanero, M.R., 2018. CDCA7 is a critical mediator of lymphomagenesis that selectively regulates anchorage-independent growth. Haematologica 103, 1669–1678. https://doi.org/10.3324/haematol.2018.188961
Jimenez-Useche, I., Ke, J., Tian, Y., Shim, D., Howell, S.C., Qiu, X., Yuan, C., 2013. DNA Methylation Regulated Nucleosome Dynamics. Sci Rep 3, 2121. https://doi.org/10.1038/srep02121
Jin, B., Li, Y., Robertson, K.D., 2011. DNA Methylation: Superior or Subordinate in the Epigenetic Hierarchy? Genes & Cancer 2, 607–617. https://doi.org/10.1177/1947601910393957
Jin, C., Felsenfeld, G., 2007. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes & Development 21, 1519–1529. https://doi.org/10.1101/gad.1547707
Joazeiro, C.A.P., Weissman, A.M., 2000. RING Finger Proteins. Cell 102, 549–552. https://doi.org/10.1016/S0092-8674(00)00077-5
Johnson, Pollard, Earnshaw, 2017. Cell Biology. Elsevier. https://doi.org/10.1016/C2014-0-00272-9
Jones, D.T., 1999. Protein secondary structure prediction based on position-specific scoring matrices 1 1Edited by G. Von Heijne. Journal of Molecular Biology 292, 195–202. https://doi.org/10.1006/jmbi.1999.3091
Kagalwala, M.N., Glaus, B.J., Dang, W., Zofall, M., Bartholomew, B., 2004. Topography of the ISW2–nucleosome complex: insights into nucleosome spacing and chromatin remodeling. EMBO J 23, 2092–2104. https://doi.org/10.1038/sj.emboj.7600220
Kalashnikova, A.A., Porter-Goff, M.E., Muthurajan, U.M., Luger, K., Hansen, J.C., 2013. The role of the nucleosome acidic patch in modulating higher order chromatin structure. J. R. Soc. Interface 10, 20121022. https://doi.org/10.1098/rsif.2012.1022
Kanno, M., Hasegawa, M., Ishida, A., Isono, K., Taniguchi, M., 1995. mel-18, a Polycomb group-related mammalian gene, encodes a transcriptional negative regulator with tumor suppressive activity. EMBO J. 14, 5672–5678.
Karlin, S., Brocchieri, L., Bergman, A., Mrazek, J., Gentles, A.J., 2002. Amino acid runs in eukaryotic proteomes and disease associations. Proceedings of the National Academy of Sciences 99, 333–338. https://doi.org/10.1073/pnas.012608599

Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., Uchtenhagen, H., Urlaub, H., Herzog, F., Peters, J.M., Poerschke, D., Lu, R., 2008a. GraFix : sample preparation for single- particle electron cryomicroscopy 5, 53–55. https://doi.org/10.1038/NMETH1139

Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., Uchtenhagen, H., Urlaub, H., Herzog, F., Peters, J.M., Poerschke, D., Lührmann, R., Stark, H., 2008b. GraFix: sample preparation for single-particle electron cryomicroscopy. Nat Methods 5, 53–55. https://doi.org/10.1038/nmeth1139

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10, 845–858. https://doi.org/10.1038/nprot.2015.053

Kevin Bryan Myant, K.B.M., 2008. Functional Analysis of the SWI/SNF family protein LSH. University of Edinburgh.

Keyes, W.M., Pecoraro, M., Aranda, V., Vernersson-Lindahl, E., Li, W., Vogel, H., Guo, X., Garcia, E.L., Michurina, T.V., Enikolopov, G., Muthuswamy, S.K., Mills, A.A., 2011. ΔNp63α Is an Oncogene that Targets Chromatin Remodeler Lsh to Drive Skin Stem Cell Proliferation and Tumorigenesis. Cell Stem Cell 8, 164–176. https://doi.org/10.1016/j.stem.2010.12.009

Kollarovic, G., Topping, C.E., Shaw, E.P., Chambers, A.L., 2018. The human HELLS chromatin remodelling protein promotes end resection to facilitate homologous recombination within heterochromatin (preprint). Biochemistry. https://doi.org/10.1101/504043

Kopel, V., 1996. Unwinding of the third strand of a DNA triple helix, a novel activity of the SV40 large T-antigen helicase. Nucleic Acids Research 24, 330–335. https://doi.org/10.1093/nar/24.2.330

Kouzarides, T., 2007. Chromatin Modifications and Their Function. Cell 128, 693–705. https://doi.org/10.1016/j.cell.2007.02.005

Krajewski, W.A., 2018. Effects of DNA Superhelical Stress on the Stability of H2B-Ubiquitylated Nucleosomes. Journal of Molecular Biology 430, 5002–5014. https://doi.org/10.1016/j.jmb.2018.09.014

Laity, John H, Dyson, H.J., Wright, P.E., 2000. DNA-induced α-helix capping in conserved linker sequences is a determinant of binding affinity in Cys2-His2 zinc fingers. Journal of Molecular Biology 295, 719–727. https://doi.org/10.1006/jmbi.1999.3406

Laity, J. H., Dyson, H.J., Wright, P.E., 2000. Molecular basis for modulation of biological function by alternate splicing of the Wilms' tumor suppressor protein. Proceedings of the National Academy of Sciences 97, 11932–11935. https://doi.org/10.1073/pnas.97.22.11932

Laky, K., Evans, S., Perez-Diez, A., Fowlkes, B.J., 2015. Notch Signaling Regulates Antigen Sensitivity of Naive CD4+ T Cells by Tuning Co-stimulation. Immunity 42, 80–94. https://doi.org/10.1016/j.immuni.2014.12.027

Längst, G., Bonte, E.J., Corona, D.F.V., Becker, P.B., 1999. Nucleosome Movement by CHRAC and ISWI without Disruption or trans-Displacement of the Histone Octamer. Cell 97, 843–852. https://doi.org/10.1016/S0092-8674(00)80797-7

Laurent, B.C., Yang, X., Carlson, M., 1992. An essential Saccharomyces cerevisiae gene homologous to SNF2 encodes a helicase-related protein in a new family. Mol. Cell. Biol. 12, 1893–1902. https://doi.org/10.1128/MCB.12.4.1893

Lee, D.W., Zhang, K., Ning, Z.-Q., Raabe, E.H., Tintner, S., Wieland, R., Wilkins, B.J., Kim, J.M., Blough, R.I., Arceci, R.J., n.d. Proliferation-associated SNF2-like Gene (PASG): A SNF2 Family Member Altered in Leukemia 11.

Lee, S.Y., Skolnick, J., 2007. Development and benchmarking of TASSERiter for the iterative improvement of protein structure predictions. Proteins 68, 39–47. https://doi.org/10.1002/prot.21440

Leonard, J.D., Narlikar, G.J., 2015. A Nucleotide-Driven Switch Regulates Flanking DNA Length Sensing by a Dimeric Chromatin Remodeler. Molecular Cell 57, 850–859. https://doi.org/10.1016/j.molcel.2015.01.008

Levendosky, R.F., Bowman, G.D., 2019. Asymmetry between the two acidic patches dictates the direction of nucleosome sliding by the ISWI chromatin remodeler. eLife 8, e45472. https://doi.org/10.7554/eLife.45472

Levendosky, R.F., Sabantsev, A., Deindl, S., Bowman, G.D., 2016. The Chd1 chromatin remodeler shifts hexasomes unidirectionally. eLife 5, e21356. https://doi.org/10.7554/eLife.21356

Levy, Y., Onuchic, J.N., Wolynes, P.G., 2007. Fly-Casting in Protein−DNA Binding: Frustration between Protein Folding and Electrostatics Facilitates Target Recognition. J. Am. Chem. Soc. 129, 738–739. https://doi.org/10.1021/ja065531n

Li, Meijing, Xia, X., Tian, Y., Jia, Q., Liu, X., Lu, Y., Li, Ming, Li, X., Chen, Z., 2019. Mechanism of DNA translocation underlying chromatin remodelling by Snf2. Nature 567, 409–413. https://doi.org/10.1038/s41586-019-1029-2

Lindstrom, M.S., Jin, A., Deisenroth, C., White Wolf, G., Zhang, Y., 2007. Cancer-Associated Mutations in the MDM2 Zinc Finger Domain Disrupt Ribosomal Protein Interaction and Attenuate MDM2-Induced p53 Degradation. Molecular and Cellular Biology 27, 1056–1068. https://doi.org/10.1128/MCB.01307-06

Linke, K., Mace, P.D., Smith, C.A., Vaux, D.L., Silke, J., Day, C.L., 2008. Structure of the MDM2/MDMX RING domain heterodimer reveals dimerization is required for their ubiquitylation in trans. Cell Death Differ 15, 841–848. https://doi.org/10.1038/sj.cdd.4402309

Lisowsky, T., Loguercio Polosa, P., Sagliano, A., Roberti, M., Gadaleta, M.N., Cantatore, P., 1999. Identification of human GC-box-binding zinc finger protein, a new Krüppel-like zinc finger protein, by the yeast one-hybrid screening with a GC-rich target sequence. FEBS Letters 453, 369–374. https://doi.org/10.1016/S0014-5793(99)00754-1

Litwin, I., Bakowski, T., Maciaszczyk-Dziubinska, E., Wysocki, R., 2017. The LSH/HELLS homolog Irc5 contributes to cohesin association with chromatin in yeast. Nucleic Acids Research 45, 6404–6416. https://doi.org/10.1093/nar/gkx240

Liu, J.C., Ferreira, C.G., Yusufzai, T., 2015. Human CHD2 Is a Chromatin Assembly ATPase Regulated by Its Chromo- and DNA-binding Domains. J. Biol. Chem. 290, 25–34. https://doi.org/10.1074/jbc.M114.609156

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

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

Lorch, Y., Zhang, M., Kornberg, R.D., 1999. Histone Octamer Transfer by a Chromatin-Remodeling Complex. Cell 96, 389–392. https://doi.org/10.1016/S0092-8674(00)80551-6

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

Lowary, P T, Widom, J., 1998. New DNA Sequence Rules for High Affinity Binding to Histone Octamer and Sequence-directed Nucleosome Positioning.

Ludwiczak, J., Winski, A., Szczepaniak, K., Alva, V., Dunin-Horkawicz, S., 2019. DeepCoil—a fast and accurate prediction of coiled-coil domains in protein sequences. Bioinformatics 35, 2790–2795. https://doi.org/10.1093/bioinformatics/bty1062

Ludwigsen, J., Pfennig, S., Singh, A.K., Schindler, C., Harrer, N., Forné, I., Zacharias, M., Mueller-Planitz, F., 2017. 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., M Auml Der, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J., 1997a. Crystal structure of the nucleosome core particle at 2.8|[thinsp]||[Aring]| resolution. Nature 389, 251.

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

Luger, K., Rechsteiner, T.J., Richmond, T.J., 1999. Expression and Purification of Recombinant Histones and Nucleosome Reconstitution 119.

Luk, E., Ranjan, A., FitzGerald, P.C., Mizuguchi, G., Huang, Y., Wei, D., Wu, C., 2010. Stepwise Histone Replacement by SWR1 Requires Dual Activation with Histone H2A.Z and Canonical Nucleosome. Cell 143, 725–736. https://doi.org/10.1016/j.cell.2010.10.019

Lungu, C., Muegge, K., Jeltsch, A., Jurkowska, R.Z., 2015. An ATPase-Deficient Variant of the SNF2 Family Member HELLS Shows Altered Dynamics at Pericentromeric Heterochromatin. Journal of Molecular Biology 427, 1903–1915. https://doi.org/10.1016/j.jmb.2015.03.014

Lupas, A.N., Gruber, M., 2005. The Structure of α-Helical Coiled Coils, in: Advances in Protein Chemistry. Elsevier, pp. 37–38. https://doi.org/10.1016/S0065-3233(05)70003-6

Lyons, D.B., Zilberman, D., 2017. DDM1 and Lsh remodelers allow methylation of DNA wrapped in nucleosomes. eLife 6, e30674. https://doi.org/10.7554/eLife.30674

Machida, S., Takizawa, Y., Ishimaru, M., Sugita, Y., Sekine, S., Nakayama, J., Wolf, M., Kurumizaka, H., 2018. Structural Basis of Heterochromatin Formation by Human HP1. Molecular Cell 69, 385-397.e8. https://doi.org/10.1016/j.molcel.2017.12.011

Maertens, G.N., 2006. Transcriptional co-activator p75 binds and tethers the Myc-interacting protein JPO2 to chromatin. Journal of Cell Science 119, 2563–2571. https://doi.org/10.1242/jcs.02995

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, 562–566. https://doi.org/10.1038/nature09321

Makharashvili, N., Paull, T.T., 2015. CtIP: A DNA damage response protein at the intersection of DNA metabolism. DNA Repair 32, 75–81. https://doi.org/10.1016/j.dnarep.2015.04.016

Mansfield, R.E., Musselman, C.A., Kwan, A.H., Oliver, S.S., Garske, A.L., Davrazou, F., Denu, J.M., Kutateladze, T.G., Mackay, J.P., 2011. Plant Homeodomain (PHD) Fingers of CHD4 Are Histone H3-binding Modules with Preference for Unmodified H3K4 and Methylated H3K9. J. Biol. Chem. 286, 11779–11791. https://doi.org/10.1074/jbc.M110.208207

McKnight, J.N., Jenkins, K.R., Nodelman, I.M., Escobar, T., Bowman, G.D., 2011. Extranucleosomal DNA Binding Directs Nucleosome Sliding by Chd1. Molecular and Cellular Biology 31, 4746–4759. https://doi.org/10.1128/MCB.05735-11

Meehan, R.R., Dunican, D.S., Pennings, S., Meehan, R.R., 2015. Lsh Is Essential for Maintaining Global DNA Methylation Levels in Amphibia and Fish and Interacts Directly with ... Lsh Is Essential for Maintaining Global DNA Methylation Levels in Amphibia and Fish and Interacts Directly with Dnmt1. https://doi.org/10.1155/2015/740637

Meier, F., Heinzmann, G., 2017. Field-Flow Fractionation: A powerful technology for the separation and advanced characterization of proteins, antibodies, viruses, polymers and nano-/microparticles. www.chemie.de.

Meneghini, M.D., Wu, M., Madhani, H.D., 2003. Conserved Histone Variant H2A.Z Protects Euchromatin from the Ectopic Spread of Silent Heterochromatin. Cell 112, 725–736. https://doi.org/10.1016/S0092-8674(03)00123-5

Mészáros, B., Erdős, G., Dosztányi, Z., 2018. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Research 46, W329–W337. https://doi.org/10.1093/nar/gky384

Mészáros, B., Simon, I., Dosztányi, Z., 2009. Prediction of Protein Binding Regions in Disordered Proteins. PLoS Comput Biol 5, e1000376. https://doi.org/10.1371/journal.pcbi.1000376

Miller, J., McLachlan, A.D., Klug, A., 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. 4, 1609–1614.

Minguez, P., Letunic, I., Parca, L., Garcia-Alonso, L., Dopazo, J., Huerta-Cepas, J., Bork, P., 2015. PTMcode v2: a resource for functional associations of post-translational modifications within and between proteins. Nucleic Acids Research 43, D494–D502. https://doi.org/10.1093/nar/gku1081

Miyamoto, K., Uechi, A., Saito, K., 2017. The zinc finger domain of RING finger protein 141 reveals a unique RING fold: Zinc Finger Structure of RING Finger Protein 141. Protein Science 26, 1681–1686. https://doi.org/10.1002/pro.3201

Mizuguchi, G., 2004. ATP-Driven Exchange of Histone H2AZ Variant Catalyzed by SWR1 Chromatin Remodeling Complex. Science 303, 343–348. https://doi.org/10.1126/science.1090701

Moore, L.D., Le, T., Fan, G., 2013. DNA Methylation and Its Basic Function. Neuropsychopharmacol 38, 23–38. https://doi.org/10.1038/npp.2012.112

Mueller-Planitz, F., Klinker, H., Ludwigsen, J., Becker, P.B., 2013. The ATPase domain of ISWI is an autonomous nucleosome remodeling machine. Nat Struct Mol Biol 20, 82–89. https://doi.org/10.1038/nsmb.2457

Muñoz, S., Minamino, M., Casas-Delucchi, C.S., Patel, H., Uhlmann, F., 2019. A Role for Chromatin Remodeling in Cohesin Loading onto Chromosomes. Molecular Cell 74, 664-673.e5. https://doi.org/10.1016/j.molcel.2019.02.027

Myant, K., Stancheva, I., 2008. LSH Cooperates with DNA Methyltransferases To Repress Transcription. Molecular and Cellular Biology 28, 215–226. https://doi.org/10.1128/MCB.01073-07

Myant, K., Termanis, A., Sundaram, A.Y.M., Boe, T., Li, C., Merusi, C., Burrage, J., Heras, J.I. d. L., Stancheva, I., 2011. LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. Genome Research 21, 83–94. https://doi.org/10.1101/gr.108498.110

Na, J.-H., Lee, W.-K., Yu, Y., 2018. How Do We Study the Dynamic Structure of Unstructured Proteins: A Case Study on Nopp140 as an Example of a Large, Intrinsically Disordered Protein. IJMS 19, 381. https://doi.org/10.3390/ijms19020381

Nasmyth, K., 2011. Cohesin: a catenase with separate entry and exit gates? Nat Cell Biol 13, 1170–1177. https://doi.org/10.1038/ncb2349

Nguyen, Vu Q., Ranjan, A., Stengel, F., Wei, D., Aebersold, R., Wu, C., Leschziner, A.E., 2013. Molecular Architecture of the ATP-Dependent Chromatin-Remodeling Complex SWR1. Cell 154, 1220–1231. https://doi.org/10.1016/j.cell.2013.08.018

Nguyen, Vu Q, Ranjan, A., Stengel, F., Wei, D., Aebersold, R., Wu, C., Leschziner, A.E., 2013. Molecular Architecture of the Complex SWR1. Cell 154, 1220–1231. https://doi.org/10.1016/j.cell.2013.08.018

Niina, T., Brandani, G.B., Tan, C., Takada, S., 2017. Sequence-dependent nucleosome sliding in rotation-coupled and uncoupled modes revealed by molecular simulations. PLoS Comput Biol 13, e1005880. https://doi.org/10.1371/journal.pcbi.1005880

Niu, J., Chen, T., Han, L., Wang, P., Li, N., Tong, T., 2011. Transcriptional activation of the senescence regulator Lsh by E2F1. Mechanisms of Ageing and Development 132, 180–186. https://doi.org/10.1016/j.mad.2011.03.004

Nodelman, Ilana M., Bleichert, F., Patel, A., Ren, R., Horvath, K.C., Berger, J.M., Bowman, G.D., 2017. Interdomain Communication of the Chd1 Chromatin Remodeler across the DNA Gyres of the Nucleosome. Molecular Cell 65, 447-459.e6. https://doi.org/10.1016/j.molcel.2016.12.011

Nodelman, Ilana M, Bleichert, F., Patel, A., Ren, R., Horvath, K.C., Berger, J.M., Bowman, G.D., Bleichert, F., Patel, A., Ren, R., Horvath, K.C., Berger, J.M., 2017. Interdomain Communication of the Chd1 Chromatin Remodeler across the DNA Gyres of the Nucleosome Article Interdomain Communication of the Chd1 Chromatin Remodeler across the DNA Gyres of the Nucleosome. Molecular Cell 65, 447-459.e6. https://doi.org/10.1016/j.molcel.2016.12.011

Oldfield, C.J., Cheng, Y., Cortese, M.S., Brown, C.J., Uversky, V.N., Dunker, A.K., 2005. Comparing and Combining Predictors of Mostly Disordered Proteins †. Biochemistry 44, 1989–2000. https://doi.org/10.1021/bi047993o

Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., Cheng, X., Bestor, T.H., 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448, 714–717. https://doi.org/10.1038/nature05987

Osthus, R.C., Karim, B., Prescott, J.E., Smith, B.D., McDevitt, M., Huso, D.L., Dang, C.V., 2005. The Myc Target Gene *JPO1/CDCA7* Is Frequently Overexpressed in Human Tumors and Has Limited Transforming Activity *In vivo*. Cancer Res 65, 5620–5627. https://doi.org/10.1158/0008-5472.CAN-05-0536

Otani, J., Nankumo, T., Arita, K., Inamoto, S., Ariyoshi, M., Shirakawa, M., 2009. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L domain. EMBO Rep 10, 1235–1241. https://doi.org/10.1038/embor.2009.218

Ou, X.-M., Chen, K., Shih, J.C., 2006. Monoamine oxidase A and repressor R1 are involved in apoptotic signaling pathway. Proceedings of the National Academy of Sciences 103, 10923–10928. https://doi.org/10.1073/pnas.0601515103

Owen-Hughes, T., Workman, J.L., 1996. Remodeling the chromatin structure of a nucleosome array by transcription factor-targeted trans-displacement of histones. The EMBO Journal 15, 4702–4712. https://doi.org/10.1002/j.1460-2075.1996.tb00847.x

Ozturk, N., Singh, I., Mehta, A., Braun, T., Barreto, G., 2014. HMGA proteins as modulators of chromatin structure during transcriptional activation. Front. Cell Dev. Bio. 2. https://doi.org/10.3389/fcell.2014.00005

Pabo, C.O., Sauer, R.T., 1984. Protein-DNA Recognition. Annu. Rev. Biochem. 53, 293–321. https://doi.org/10.1146/annurev.bi.53.070184.001453
Pace, N., Weerapana, E., 2014. Zinc-Binding Cysteines: Diverse Functions and Structural Motifs. Biomolecules 4, 419–434. https://doi.org/10.3390/biom4020419
Paillard, G., 2004. Looking into DNA recognition: zinc finger binding specificity. Nucleic Acids Research 32, 6673–6682. https://doi.org/10.1093/nar/gkh1003
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, 200–213. https://doi.org/10.1016/j.cell.2010.12.021
Pappin, D.J.C., Hojrup, P., Bleasby, A.J., 1993. Rapid identification of proteins by peptide-mass fingerprinting. Current Biology 3, 327–332. https://doi.org/10.1016/0960-9822(93)90195-T
Parry, D.A.D., 1982. Coiled-coils in ?-helix-containing proteins: analysis of the residue types within the heptad repeat and the use of these data in the prediction of coiled-coils in other proteins. Biosci Rep 2, 1017–1024. https://doi.org/10.1007/BF01122170
Patel, A., Chakravarthy, S., Morrone, S., Nodelman, I.M., McKnight, J.N., Bowman, G.D., 2013. Decoupling nucleosome recognition from DNA binding dramatically alters the properties of the Chd1 chromatin remodeler. Nucleic Acids Research 41, 1637–1648. https://doi.org/10.1093/nar/gks1440
Pavletich, N., Pabo, C., 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science 252, 809–817. https://doi.org/10.1126/science.2028256
Peng, K., Radivojac, P., Vucetic, S., Dunker, A.K., Obradovic, Z., 2006. Length-dependent prediction of protein intrinsic disorder. BMC Bioinformatics 7, 208. https://doi.org/10.1186/1471-2105-7-208
Pilarowski, G.O., Vernon, H.J., Applegate, C.D., Boukas, L., Cho, M.T., Gurnett, C.A., Benke, P.J., Beaver, E., Heeley, J.M., Medne, L., Krantz, I.D., Azage, M., Niyazov, D., Henderson, L.B., Wentzensen, I.M., Baskin, B., Sacoto, M.J.G., Bowman, G.D., Bjornsson, H.T., 2018. Missense variants in the chromatin remodeler *CHD1* are associated with neurodevelopmental disability. J Med Genet 55, 561–566. https://doi.org/10.1136/jmedgenet-2017-104759
Pillidge, Z., Bray, S.J., 2019. SWI/SNF chromatin remodeling controls Notch-responsive enhancer accessibility. EMBO Rep 20. https://doi.org/10.15252/embr.201846944
Pjura, P.E., Grzeskowiak, K., Dickerson, R.E., 1987. Binding of Hoechst 33258 to the minor groove of B-DNA. Journal of Molecular Biology 197, 257–271. https://doi.org/10.1016/0022-2836(87)90123-9
Prescott, J.E., Osthus, R.C., Lee, L.A., Lewis, B.C., Shim, H., Barrett, J.F., Guo, Q., Hawkins, A.L., Griffin, C.A., Dang, C.V., 2001. A Novel c-Myc-responsive Gene, JPO1, Participates in Neoplastic Transformation. J. Biol. Chem. 276, 48276–48284. https://doi.org/10.1074/jbc.M107357200
Qin, X.-Y., Feng, J., Chen, G., Dou, X.-W., Dai, X.-Q., Dong, H.-L., Gong, F.-Y., Xiao, F., Zhao, Y., Gao, X.-M., Wang, J., 2019. ZBTB24 regulates the apoptosis of human T cells via CDCA7/TRAIL-receptor axis. Biochemical and Biophysical Research Communications 514, 259–265. https://doi.org/10.1016/j.bbrc.2019.04.147
Qiu, Y., Levendosky, R.F., Chakravarthy, S., Patel, A., Bowman, G.D., Myong, S., 2017. The Chd1 Chromatin Remodeler Shifts Nucleosomal DNA Bidirectionally as a Monomer. Molecular Cell 68, 76-88.e6. https://doi.org/10.1016/j.molcel.2017.08.018
Raabe, E.H., Abdurrahman, L., Behbehani, G., Arceci, R.J., 2001. An SNF2 factor involved in mammalian development and cellular proliferation. Dev. Dyn. 221, 92–105. https://doi.org/10.1002/dvdy.1128

Racki, L.R., Yang, J.G., Naber, N., Partensky, P.D., Acevedo, A., Purcell, T.J., Cooke, R., Cheng, Y., Narlikar, G.J., 2009. The chromatin remodeller ACF acts as a dimeric motor to space nucleosomes. Nature 462, 1016–1021. https://doi.org/10.1038/nature08621

Ranjan, A., Wang, F., Mizuguchi, G., Wei, D., Huang, Y., Wu, C., 2015. H2A histone-fold and DNA elements in nucleosome activate SWR1-mediated H2A.Z replacement in budding yeast. eLife 4, e06845. https://doi.org/10.7554/eLife.06845

Raschle, M., Van Komen, S., Chi, P., Ellenberger, T., Sung, P., 2004. Multiple Interactions with the Rad51 Recombinase Govern the Homologous Recombination Function of Rad54. J. Biol. Chem. 279, 51973–51980. https://doi.org/10.1074/jbc.M410101200

Ravasi, T., 2003. Systematic Characterization of the Zinc-Finger-Containing Proteins in the Mouse Transcriptome. Genome Research 13, 1430–1442. https://doi.org/10.1101/gr.949803

Reece-Hoyes, J.S., Marian Walhout, A.J., 2012. Yeast one-hybrid assays: A historical and technical perspective. Methods 57, 441–447. https://doi.org/10.1016/j.ymeth.2012.07.027

Reger, A.S., Yang, M.P., Koide-Yoshida, S., Guo, E., Mehta, S., Yuasa, K., Liu, A., Casteel, D.E., Kim, C., 2014. Crystal Structure of the cGMP-dependent Protein Kinase II Leucine Zipper and Rab11b Protein Complex Reveals Molecular Details of G-kinase-specific Interactions. J. Biol. Chem. 289, 25393–25403. https://doi.org/10.1074/jbc.M114.575894

Ren, J., Briones, V., Barbour, S., Yu, W., Han, Y., Terashima, M., Muegge, K., 2015. The ATP binding site of the chromatin remodeling homolog Lsh is required for nucleosome density and de novo DNA methylation at repeat sequences. Nucleic Acids Research 43, 1444–1455. https://doi.org/10.1093/nar/gku1371

Ren, R., Hardikar, S., Horton, J.R., Lu, Y., Zeng, Y., Singh, A.K., Lin, K., Coletta, L.D., Shen, J., Lin Kong, C.S., Hashimoto, H., Zhang, X., Chen, T., Cheng, X., 2019. Structural basis of specific DNA binding by the transcription factor ZBTB24. Nucleic Acids Research 47, 8388–8398. https://doi.org/10.1093/nar/gkz557

Ren, W., Gao, L., Song, J., 2018. Structural Basis of DNMT1 and DNMT3A-Mediated DNA Methylation. Genes 9, 620. https://doi.org/10.3390/genes9120620

Rippe, K., Schrader, A., Riede, P., Strohner, R., Lehmann, E., Langst, G., 2007. DNA sequence- and conformation-directed positioning of nucleosomes by chromatin-remodeling complexes. Proceedings of the National Academy of Sciences 104, 15635–15640. https://doi.org/10.1073/pnas.0702430104

Romero, P., Obradovic, Z., Kissinger, C., Villafranca, J.E., Dunker, A.K., 1997. Identifying disordered regions in proteins from amino acid sequence, in: Proceedings of International Conference on Neural Networks (ICNN'97). Presented at the International Conference on Neural Networks (ICNN'97), IEEE, Houston, TX, USA, pp. 90–95. https://doi.org/10.1109/ICNN.1997.611643

Rondelet, G., Dal, T., Willems, L., Wouters, J., 2016. Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. Journal of Structural Biology 194, 357–367. https://doi.org/10.1016/j.jsb.2016.03.013

Roussel, L., Erard, M., Cayrol, C., Girard, J.-P., 2008. Molecular mimicry between IL-33 and KSHV for attachment to chromatin through the H2A–H2B acidic pocket. EMBO Rep 9, 1006–1012. https://doi.org/10.1038/embor.2008.145

Rudnizky, S., Bavly, A., Malik, O., Pnueli, L., Melamed, P., Kaplan, A., 2016. H2A.Z controls the stability and mobility of nucleosomes to regulate expression of the LH genes. Nature Communications 7, 12958.

Saha, A., 2002. Chromatin remodeling by RSC involves ATP-dependent DNA translocation. Genes & Development 16, 2120–2134. https://doi.org/10.1101/gad.995002
Saha, A., Wittmeyer, J., Cairns, B.R., 2005. Chromatin remodeling through directional DNA translocation from an internal nucleosomal site. Nat Struct Mol Biol 12, 747–755. https://doi.org/10.1038/nsmb973
San Filippo, J., Sung, P., Klein, H., 2008. Mechanism of Eukaryotic Homologous Recombination. Annu. Rev. Biochem. 77, 229–257. https://doi.org/10.1146/annurev.biochem.77.061306.125255
Santos-Rosa, H., Schneider, R., Bernstein, B.E., Karabetsou, N., Morillon, A., Weise, C., Schreiber, S.L., Mellor, J., Kouzarides, T., 2003. Methylation of Histone H3 K4 Mediates Association of the Isw1p ATPase with Chromatin. Molecular Cell 12, 1325–1332. https://doi.org/10.1016/S1097-2765(03)00438-6
Sartori, A.A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., Jackson, S.P., 2007. Human CtIP promotes DNA end resection. Nature 450, 509–514. https://doi.org/10.1038/nature06337
Schuster-Böckler, B., Lehner, B., 2012. Chromatin organization is a major influence on regional mutation rates in human cancer cells. Nature 488, 504–507. https://doi.org/10.1038/nature11273
Schwanbeck, R., Xiao, H., Wu, C., 2004. Spatial Contacts and Nucleosome Step Movements Induced by the NURF Chromatin Remodeling Complex. J. Biol. Chem. 279, 39933–39941. https://doi.org/10.1074/jbc.M406060200
Segal, E., Widom, J., 2009. From DNA sequence to transcriptional behaviour: a quantitative approach. Nat Rev Genet 10, 443–456. https://doi.org/10.1038/nrg2591
Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S., Yokoyama, S., 2006. Structural Basis for RNA Unwinding by the DEAD-Box Protein Drosophila Vasa. Cell 125, 287–300. https://doi.org/10.1016/j.cell.2006.01.054
Shen, F.-Z., Li, X.-S., Ma, J.-W., Wang, X.-Y., Zhao, S.-P., Meng, L., Liang, S.-F., Zhao, X.-L., 2019. Cell division cycle associated 7 like predicts unfavorable prognosis and promotes invasion in glioma. Pathology - Research and Practice 215, 50–56. https://doi.org/10.1016/j.prp.2018.10.023
Shoemaker, B.A., Portman, J.J., Wolynes, P.G., 2000. Speeding molecular recognition by using the folding funnel: The fly-casting mechanism. Proceedings of the National Academy of Sciences 97, 8868–8873. https://doi.org/10.1073/pnas.160259697
Sichuan Xi, Theresa M. Geiman, V.B., 2009. Lsh Participates in DNA Methylation and Silencing of Stem Cell Genes 27, 2691–2702. https://doi.org/10.1002/stem.183.Lsh
Siedlecki, J., Zimmermann, W., Weissbach, A., 1983. Characterization of a prokaryotic topoisomerase I activity in chloroplast extracts from spinach. Nucleic Acids Research 11, 1523–1536. https://doi.org/10.1093/nar/11.5.1523
Simon Varzandeh, 2016. Structural and Biochemical insights into the ATP-dependent Chromatin Remodeler LSH. University of Edinburgh.
Sirinakis, G., Clapier, C.R., Gao, Y., Viswanathan, R., Cairns, B.R., Zhang, Y., 2011. The RSC chromatin remodelling ATPase translocates DNA with high force and small step size: Remodeller motor translocates DNA with high force. The EMBO Journal 30, 2364–2372. https://doi.org/10.1038/emboj.2011.141
Smith, C.L., Peterson, C.L., 2005. A Conserved Swi2/Snf2 ATPase Motif Couples ATP Hydrolysis to Chromatin Remodeling. Molecular and Cellular Biology 25, 5880–5892. https://doi.org/10.1128/MCB.25.14.5880-5892.2005
Song, X., Zhao, X., Huang, Y., Xiang, H., Zhang, W., Tong, D., 2015. Transmissible Gastroenteritis Virus (TGEV) Infection Alters the Expression of Cellular MicroRNA

Species That Affect Transcription of TGEV Gene 7. Int. J. Biol. Sci. 11, 913–922. https://doi.org/10.7150/ijbs.11585
Sørensen, H., 2010. Towards universal systems for recombinant gene expression. Microb Cell Fact 9, 27. https://doi.org/10.1186/1475-2859-9-27
Stanley, N., Esteban-Martín, S., De Fabritiis, G., 2015. Progress in studying intrinsically disordered proteins with atomistic simulations. Progress in Biophysics and Molecular Biology 119, 47–52. https://doi.org/10.1016/j.pbiomolbio.2015.03.003
Stark, H., 2010. GraFix : Stabilization of Fragile Macromolecular Complexes for Single Particle Cryo-EM, Methods in Enzymology. Elsevier Masson SAS. https://doi.org/10.1016/S0076-6879(10)81005-5
Strohner, R., Wachsmuth, M., Dachauer, K., Mazurkiewicz, J., Hochstatter, J., Rippe, K., Längst, G., 2005. A “loop recapture” mechanism for ACF-dependent nucleosome remodeling. Nat Struct Mol Biol 12, 683–690. https://doi.org/10.1038/nsmb966
Sun, L.-Q., 2004. Growth retardation and premature aging phenotypes in mice with disruption of the SNF2-like gene, PASG. Genes & Development 18, 1035–1046. https://doi.org/10.1101/gad.1176104
Sundaramoorthy, R., Hughes, A.L., Singh, V., Wiechens, N., Ryan, D.P., El-Mkami, H., Petoukhov, M., Svergun, D.I., Treutlein, B., Quack, S., Fischer, M., Michaelis, J., n.d. Structural reorganization of the chromatin remodeling enzyme Chd1 upon engagement with nucleosomes 28.
Tan, F., Lu, Y., Jiang, W., Wu, T., Zhang, R., Zhao, Y., Zhou, D.-X., 2018. DDM1 Represses Noncoding RNA Expression and RNA-Directed DNA Methylation in Heterochromatin. Plant Physiol. 177, 1187–1197. https://doi.org/10.1104/pp.18.00352
Tao, Y., Xi, S., Shan, J., Maunakea, A., Che, A., Briones, V., Lee, E.Y., Geiman, T., Huang, J., Stephens, R., Leighty, R.M., Zhao, K., Muegge, K., 2011. Lsh, chromatin remodeling family member, modulates genome-wide cytosine methylation patterns at nonrepeat sequences. Proceedings of the National Academy of Sciences 108, 5626–5631. https://doi.org/10.1073/pnas.1017000108
Termanis, A., Torrea, N., Culley, J., Kerr, A., Ramsahoye, B., Stancheva, I., 2016. The SNF2 family ATPase LSH promotes cell-autonomous *de novo* DNA methylation in somatic cells. Nucleic Acids Res 44, 7592–7604. https://doi.org/10.1093/nar/gkw424
Thijssen, P.E., Ito, Y., Grillo, G., Wang, J., Velasco, G., Nitta, H., Unoki, M., Yoshihara, M., Suyama, M., Sun, Y., Lemmers, R.J.L.F., de Greef, J.C., Gennery, A., Picco, P., Kloeckener-Gruissem, B., Güngör, T., Reisli, I., Picard, C., Kebaili, K., Roquelaure, B., Iwai, T., Kondo, I., Kubota, T., van Ostaijen-Ten Dam, M.M., van Tol, M.J.D., Weemaes, C., Francastel, C., van der Maarel, S.M., Sasaki, H., 2015. Mutations in CDCA7 and HELLS cause immunodeficiency–centromeric instability–facial anomalies syndrome. Nat Commun 6, 7870. https://doi.org/10.1038/ncomms8870
Tian, Y., Huang, C., Zhang, H., Ni, Q., Han, S., Wang, D., Han, Z., Li, X., 2013. CDCA7L promotes hepatocellular carcinoma progression by regulating the cell cycle. International Journal of Oncology 43, 2082–2090. https://doi.org/10.3892/ijo.2013.2142
Tosi, A., Haas, C., Herzog, F., Gilmozzi, A., Berninghausen, O., Ungewickell, C., Gerhold, C.B., Lakomek, K., Aebersold, R., Beckmann, R., Hopfner, K.-P., 2013. Structure and Subunit Topology of the INO80 Chromatin Remodeler and Its Nucleosome Complex. Cell 154, 1207–1219. https://doi.org/10.1016/j.cell.2013.08.016
Toubiana, S., Velasco, G., Chityat, A., Kaindl, A.M., Hershtig, N., Tzur-Gilat, A., Francastel, C., Selig, S., 2018. Subtelomeric methylation distinguishes between subtypes of

Immunodeficiency, Centromeric instability and Facial anomalies syndrome. Human Molecular Genetics 27, 3568–3581. https://doi.org/10.1093/hmg/ddy265
True, J.D., Muldoon, J.J., Carver, M.N., Poorey, K., Shetty, S.J., Bekiranov, S., Auble, D.T., 2016. The Modifier of Transcription 1 (Mot1) ATPase and Spt16 Histone Chaperone Co-regulate Transcription through Preinitiation Complex Assembly and Nucleosome Organization. J. Biol. Chem. 291, 15307–15319. https://doi.org/10.1074/jbc.M116.735134
Truebestein, L., Leonard, T.A., 2016. Coiled-coils: The long and short of it. BioEssays 38, 903–916. https://doi.org/10.1002/bies.201600062
Turner, C.E., Miller, J.T., 1994. Primary sequence of paxillin contains putative SH2 and SH3 domain binding motifs and multiple LIM domains: identification of a vinculin and pp125Fak-binding region. J. Cell. Sci. 107 (Pt 6), 1583–1591.
Unoki, M., Funabiki, H., Velasco, G., Francastel, C., Sasaki, H., 2018. CDCA7 and HELLS mutations undermine nonhomologous end joining in centromeric instability syndrome. Journal of Clinical Investigation 129, 78–92. https://doi.org/10.1172/JCI99751
van den Boogaard, M.L., Thijssen, P.E., Aytekin, C., Licciardi, F., Kıykım, A.A., Spossito, L., Dalm, V.A.S.H., Driessen, G.J., Kersseboom, R., de Vries, F., van Ostaijen-ten Dam, M.M., Ikinciogullari, A., Dogu, F., Oleastro, M., Bailardo, E., Daxinger, L., Nain, E., Baris, S., van Tol, M.J.D., Weemaes, C., van der Maarel, S.M., 2017. Expanding the mutation spectrum in ICF syndrome: Evidence for a gender bias in ICF2: VAN DEN BOOGAARD ET AL. Clin Genet 92, 380–387. https://doi.org/10.1111/cge.12979
Velasco, G., Grillo, G., Touleimat, N., Ferry, L., Ivkovic, I., Ribierre, F., Deleuze, J.-F., Chantalat, S., Picard, C., Francastel, C., 2018. Comparative methylome analysis of ICF patients identifies heterochromatin loci that require ZBTB24, CDCA7 and HELLS for their methylated state. Human Molecular Genetics 27, 2409–2424. https://doi.org/10.1093/hmg/ddy130
von Eyss, B., Maaskola, J., Memczak, S., Möllmann, K., Schuetz, A., Loddenkemper, C., Tanh, M.-D., Otto, A., Muegge, K., Heinemann, U., Rajewsky, N., Ziebold, U., 2012. The SNF2-like helicase HELLS mediates E2F3-dependent transcription and cellular transformation: E2F3 needs HELLS for tumour cell proliferation. The EMBO Journal 31, 972–985. https://doi.org/10.1038/emboj.2011.451
von Hippel, P.H., McGhee, J.D., 1972. DNA-Protein Interactions. Annu. Rev. Biochem. 41, 231–300. https://doi.org/10.1146/annurev.bi.41.070172.001311
Vukic, M., Daxinger, L., 2019. DNA methylation in disease: Immunodeficiency, Centromeric instability, Facial anomalies syndrome. Essays in Biochemistry 63, 773–783. https://doi.org/10.1042/EBC20190035
Walker, M., 2001. Drug Target Discovery by Gene Expression Analysis Cell Cycle Genes. CCDT 1, 73–83. https://doi.org/10.2174/1568009013334241
Wang, H., Ye, L., Xing, Z., Li, H., Lv, T., Liu, H., Zhang, F., Song, Y., 2019. CDCA7 promotes lung adenocarcinoma proliferation via regulating the cell cycle. Pathology - Research and Practice 215, 152559. https://doi.org/10.1016/j.prp.2019.152559
Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Peng, W., Zhang, M.Q., Zhao, K., 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet 40, 897–903. https://doi.org/10.1038/ng.154
Warren DeLano, 2002. PyMOL: An Open-Source Molecular Graphics Tool. CCp4 Newsletter.
Watanabe, S., Tan, D., Lakshminarasimhan, M., Washburn, M.P., Erica Hong, E.-J., Walz, T., Peterson, C.L., 2015. Structural analyses of the chromatin remodelling enzymes INO80-C and SWR-C. Nat Commun 6, 7108. https://doi.org/10.1038/ncomms8108

Weemaes, C.M., van Tol, M.J., Wang, J., van Ostaijen-ten Dam, M.M., van Eggermond, M.C., Thijssen, P.E., Aytekin, C., Brunetti-Pierri, N., van der Burg, M., Graham Davies, E., Ferster, A., Furthner, D., Gimelli, G., Gennery, A., Kloeckener-Gruissem, B., Meyn, S., Powell, C., Reisli, I., Schuetz, C., Schulz, A., Shugar, A., van den Elsen, P.J., van der Maarel, S.M., 2013. Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. Eur J Hum Genet 21, 1219–1225. https://doi.org/10.1038/ejhg.2013.40

Wei, Y., Mizzen, C.A., Cook, R.G., Gorovsky, M.A., Allis, C.D., 1998. Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in Tetrahymena. Proceedings of the National Academy of Sciences 95, 7480–7484. https://doi.org/10.1073/pnas.95.13.7480

Whitehouse, I., Stockdale, C., Flaus, A., Szczelkun, M.D., Owen-Hughes, T., 2003. Evidence for DNA Translocation by the ISWI Chromatin-Remodeling Enzyme. Molecular and Cellular Biology 23, 1935–1945. https://doi.org/10.1128/MCB.23.6.1935-1945.2003

Whitfield, M.L., Sherlock, G., Saldanha, A.J., Murray, J.I., Ball, C.A., Alexander, K.E., Matese, J.C., Perou, C.M., Hurt, M.M., Brown, P.O., Botstein, D., 2002. Identification of Genes Periodically Expressed in the Human Cell Cycle and Their Expression in Tumors. MBoC 13, 1977–2000. https://doi.org/10.1091/mbc.02-02-0030

Winger, J., Nodelman, I.M., Levendosky, R.F., Bowman, G.D., 2018. A twist defect mechanism for ATP-dependent translocation of nucleosomal DNA. eLife 7, e34100. https://doi.org/10.7554/eLife.34100

Wolfe, S.A., Nekludova, L., Pabo, C.O., 2000. DNA Recognition by Cys2His2 Zinc Finger Proteins. Annu. Rev. Biophys. Biomol. Struct. 29, 183–212. https://doi.org/10.1146/annurev.biophys.29.1.183

Wollmann, P., Cui, S., Viswanathan, R., Berninghausen, O., Wells, M.N., Moldt, M., Witte, G., Butryn, A., Wendler, P., Beckmann, R., Auble, D.T., Hopfner, K.-P., 2011. Structure and mechanism of the Swi2/Snf2 remodeller Mot1 in complex with its substrate TBP. Nature 475, 403–407. https://doi.org/10.1038/nature10215

Woodbine, L., Brunton, H., Goodarzi, A.A., Shibata, A., Jeggo, P.A., 2011. Endogenously induced DNA double strand breaks arise in heterochromatic DNA regions and require ataxia telangiectasia mutated and Artemis for their repair. Nucleic Acids Research 39, 6986–6997. https://doi.org/10.1093/nar/gkr331

Wu, H., Thijssen, P.E., de Klerk, E., Vonk, K.K.D., Wang, J., den Hamer, B., Aytekin, C., van der Maarel, S.M., Daxinger, L., 2016a. Converging disease genes in ICF syndrome: *ZBTB24* controls expression of *CDCA7* in mammals. Hum. Mol. Genet. 25, 4041–4051. https://doi.org/10.1093/hmg/ddw243

Wu, H., Thijssen, P.E., de Klerk, E., Vonk, K.K.D., Wang, J., den Hamer, B., Aytekin, C., van der Maarel, S.M., Daxinger, L., 2016b. Converging disease genes in ICF syndrome: *ZBTB24* controls expression of *CDCA7* in mammals. Hum. Mol. Genet. 25, 4041–4051. https://doi.org/10.1093/hmg/ddw243

Wu, S., Skolnick, J., Zhang, Y., 2007. Ab initio modeling of small proteins by iterative TASSER simulations. BMC Biol 5, 17. https://doi.org/10.1186/1741-7007-5-17

Wu, Y., Brosh, R.M., 2012. DNA helicase and helicase-nuclease enzymes with a conserved iron-sulfur cluster. Nucleic Acids Research 40, 4247–4260. https://doi.org/10.1093/nar/gks039

Wyatt, P.J., 1993. Light scattering and the absolute characterization of macromolecules. Analytica Chimica Acta 272, 1–40. https://doi.org/10.1016/0003-2670(93)80373-S

Xi, S., Zhu, H., Xu, H., Schmidtmann, A., Geiman, T.M., Muegge, K., 2007. Lsh controls Hox gene silencing during development 104, 12754–12758.

Xia, X., Liu, X., Li, T., Fang, X., Chen, Z., 2016. Structure of chromatin remodeler Swi2/Snf2 in the resting state. Nat Struct Mol Biol 23, 722–729. https://doi.org/10.1038/nsmb.3259

Xiao, D., Huang, J., Pan, Y., Li, H., Fu, C., Mao, C., Cheng, Y., Shi, Y., Chen, L., Jiang, Y., Yang, R., Liu, Y., Zhou, J., Cao, Y., Liu, S., Tao, Y., 2017. Chromatin Remodeling Factor LSH is Upregulated by the LRP6-GSK3β-E2F1 Axis Linking Reversely with Survival in Gliomas. Theranostics 7, 132–143. https://doi.org/10.7150/thno.17032

Yager, T.D., van Holde, K.E., 1984. Dynamics and equilibria of nucleosomes at elevated ionic strength. J. Biol. Chem. 259, 4212–4222.

Yan, L., Wang, L., Tian, Y., Xia, X., Chen, Z., 2016. Structure and regulation of the chromatin remodeller ISWI. Nature 540, 466–469. https://doi.org/10.1038/nature20590

Yan, L., Wu, H., Li, X., Gao, N., Chen, Z., 2019a. Structures of the ISWI–nucleosome complex reveal a conserved mechanism of chromatin remodeling. Nat Struct Mol Biol 26, 258–266. https://doi.org/10.1038/s41594-019-0199-9

Yan, L., Wu, H., Li, X., Gao, N., Chen, Z., 2019b. Structures of the ISWI–nucleosome complex reveal a conserved mechanism of chromatin remodeling. Nat Struct Mol Biol 26, 258–266. https://doi.org/10.1038/s41594-019-0199-9

Yan, Q., 2003. Lsh, a modulator of CpG methylation, is crucial for normal histone methylation. The EMBO Journal 22, 5154–5162. https://doi.org/10.1093/emboj/cdg493

Yan, Q., Cho, E., Lockett, S., Muegge, K., 2003. Association of Lsh, a Regulator of DNA Methylation, with Pericentromeric Heterochromatin Is Dependent on Intact Heterochromatin. Molecular and Cellular Biology 23, 8416–8428. https://doi.org/10.1128/MCB.23.23.8416-8428.2003

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER Suite: protein structure and function prediction. Nat Methods 12, 7–8. https://doi.org/10.1038/nmeth.3213

Ye, J., Osborne, A.R., Groll, M., Rapoport, T.A., 2004. RecA-like motor ATPases—lessons from structures. Biochimica et Biophysica Acta (BBA) - Bioenergetics 1659, 1–18. https://doi.org/10.1016/j.bbabio.2004.06.003

Ye, L., Li, F., Song, Y., Yu, D., Xiong, Z., Li, Y., Shi, T., Yuan, Z., Lin, C., Wu, X., Ren, L., Li, X., Song, L., 2018. Overexpression of CDCA7 predicts poor prognosis and induces EZH2-mediated progression of triple-negative breast cancer. Int J Cancer 143, 2602–2613. https://doi.org/10.1002/ijc.31766

Yen, W.-F., Sharma, R., Cols, M., Lau, C.M., Chaudhry, A., Chowdhury, P., Yewdell, W.T., Vaidyanathan, B., Sun, A., Coffre, M., Pucella, J.N., Chen, C.-C., Jasin, M., Sun, J.C., Rudensky, A.Y., Koralov, S.B., Chaudhuri, J., 2019. Distinct Requirements of CHD4 during B Cell Development and Antibody Response. Cell Reports 27, 1472-1486.e5. https://doi.org/10.1016/j.celrep.2019.04.011

Yu, W., McIntosh, C., Lister, R., Zhu, I., Han, Y., Ren, J., Landsman, D., Lee, E., Briones, V., Terashima, M., Leighty, R., Ecker, J.R., Muegge, K., 2014. Genome-wide DNA methylation patterns in LSH mutant reveals de-repression of repeat elements and redundant epigenetic silencing pathways. Genome Res. 24, 1613–1623. https://doi.org/10.1101/gr.172015.114

Zemach, A., Kim, M.Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., Zilberman, D., 2013. The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. Cell 153, 193–205. https://doi.org/10.1016/j.cell.2013.02.033

Zeng, W., Baumann, C., Schmidtmann, A., Honaramooz, A., Tang, L., Bondareva, A., Dores, C., Fan, T., Xi, S., Geiman, T., Rathi, R., de Rooij, D., De La Fuente, R., Muegge, K.,

Dobrinski, I., 2011. Lymphoid-Specific Helicase (HELLS) Is Essential for Meiotic Progression in Mouse Spermatocytes1. Biology of Reproduction 84, 1235–1241. https://doi.org/10.1095/biolreprod.110.085720

Zentner, G.E., Henikoff, S., 2013. Mot1 Redistributes TBP from TATA-Containing to TATA-Less Promoters. Molecular and Cellular Biology 33, 4996–5004. https://doi.org/10.1128/MCB.01218-13

Zhang, W., Xu, C., Bian, C., Tempel, W., Crombet, L., MacKenzie, F., Min, J., Liu, Z., Qi, C., 2011. Crystal structure of the Cys2His2-type zinc finger domain of human DPF2. Biochemical and Biophysical Research Communications 413, 58–61. https://doi.org/10.1016/j.bbrc.2011.08.043

Zhang, Y., Skolnick, J., 2005. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Research 33, 2302–2309. https://doi.org/10.1093/nar/gki524

Zhang, Y., Skolnick, J., 2004. Scoring function for automated assessment of protein structure template quality. Proteins 57, 702–710. https://doi.org/10.1002/prot.20264

Zhang, Y., Smith, C.L., Saha, A., Grill, S.W., Mihardja, S., Smith, S.B., Cairns, B.R., Peterson, C.L., Bustamante, C., 2006. DNA Translocation and Loop Formation Mechanism of Chromatin Remodeling by SWI/SNF and RSC. Molecular Cell 24, 559–568. https://doi.org/10.1016/j.molcel.2006.10.025

Zhang, Z.-M., Lu, R., Wang, P., Yu, Y., Chen, D., Gao, L., Liu, S., Ji, D., Rothbart, S.B., Wang, Y., Wang, G.G., Song, J., 2018. Structural basis for DNMT3A-mediated de novo DNA methylation. Nature 554, 387–391. https://doi.org/10.1038/nature25477

Zhou, H., Di Palma, S., Preisinger, C., Peng, M., Polat, A.N., Heck, A.J.R., Mohammed, S., 2013. Toward a Comprehensive Characterization of a Human Cancer Cell Phosphoproteome. J. Proteome Res. 12, 260–271. https://doi.org/10.1021/pr300630k

Zhou, R., Han, L., Li, G., Tong, T., 2009. Senescence delay and repression of p16 INK4a by Lsh via recruitment of histone deacetylases in human diploid fibroblasts. Nucleic Acids Research 37, 5183–5196. https://doi.org/10.1093/nar/gkp533

Zhu, H., Geiman, T.M., Xi, S., Jiang, Q., Schmidtmann, A., Chen, T., Li, E., Muegge, K., 2006. Lsh is involved in de novo methylation of DNA. EMBO J 25, 335–345. https://doi.org/10.1038/sj.emboj.7600925

Zofall, M., Persinger, J., Bartholomew, B., 2004. Functional Role of Extranucleosomal DNA and the Entry Site of the Nucleosome in Chromatin Remodeling by ISW2. Molecular and Cellular Biology 24, 10047–10057. https://doi.org/10.1128/MCB.24.22.10047-10057.2004

Zofall, M., Persinger, J., Kassabov, S.R., Bartholomew, B., 2006. Chromatin remodeling by ISW2 and SWI/SNF requires DNA translocation inside the nucleosome. Nat Struct Mol Biol 13, 339–346. https://doi.org/10.1038/nsmb1071